The Soil Database of China for Land Surface modeling

Table of content

1. Introduction

A comprehensive and high-resolution gridded soil characteristics dataset was developed for use in the land surface modeling. The dataset includes soil physical and chemical attributes: pH value, organic matter fraction, cation exchange capacity, root abundance, total nitrogen (N), total phosphorus (P), total potassium (K), alkali-hydrolysable N, available P, available K, exchangeable H^+, Al^{3+}, Ca^{2+}, Mg^{2+}, K^+ , Na^+, horizon thickness, soil profile depth, sand, silt and clay fractions, gravel, bulk density, porosity, structure, consistency and soil color. The dataset was developed based on 8,979 soil profiles and the soil map of China (1:1,000,000). We employed the polygon linkage method considering the distance between soil map polygon and soil profile to capture the spatial variation of the soil type [*Shangguan et al.*, 2012]. The linkage between the profile attribute database and soil map was established under the framework of Genetic Soil Classification of China (GSCC), which avoids uncertainty in taxon referencing, and makes direct use of the abundant soil profile information in GSCC. The resolution is 30 arc-seconds (about 1 km at the equator).The vertical variation of soil property was captured by eight layers to the depth of 2.3 m (i.e. 0- 0.045, 0.045- 0.091, 0.091- 0.166, 0.166- 0.289, 0.289- 0.493, 0.493- 0.829, 0.829- 1.383 and 1.383- 2.296 m).

2. Data description

Here we take sand content file ("SA.nc") as an example to show the data. The dataset takes the NetCDF Climate and Forecast Metadata Convention (CF-1.0). The extent is 73-136°E and 18-54°N. The following is the metadata:

```
dimensions:
        lon = 7560 ;
        lat = 4320 ;
        depth = 8 ;
variables:
        float lon(lon) ;
                lon:long_name = "longitude" ;
                lon:units = "degrees_east" ;
        float lat(lat) ;
```

```
                    lat:long_name = "latitude" ;
                    lat:units = "degrees_north" ;
        float depth(depth) ;
                    depth:long_name = "depth to the bottom of a soil layer" ;
                    depth:units = "centimeter" ;
        int NO(lat, lon) ;
                    NO:long_name = "map polygon code of the data" ;
                    NO:units = "" ;
                    NO:missing_value = 0 ;
        int SC(lat, lon) ;
                    SC:long_name = "soil class code of the data" ;
                    SC:units = "" ;
                    SC:missing_value = 0 ;
        float SA(lat, lon, depth) ;
                    SA:missing_value = -999 ;
                    SA:units = "% of weight" ;
                    SA:long_name = "sand content" ;
        int RSA(lat, lon, depth) ;
                    SC:units = "" ;
                    SC:missing_value = 0 ;
                    SC:long_name = " quality control information of sand content " ;
// global attributes:
                    :Conventions = "CF-1.0" ;
```

Quality control information (QC) was provided in numerical symbols. The symbol '11' indicates that the map unit is non-soil; otherwise, numerical symbols have 6 digits. Table 1 shows the codes of the digits. The linkage level (*d1*) represents the soil classification level at which the linkage is performed. The texture consideration (*d2*) represents whether the soil texture is considered in the linkage. The sample size level (*d3* and *d6*) represents how many soil profiles are used to represent a soil map unit or soil polygon. We provide two kinds of sample size levels: *d6* is taken from [*Batjes*, 2002], and *d3* is set according to the linkage level (*d1*) because there is more variation of soil properties at higher soil type levels, which needs more samples to be representative. The search radius flag (*d4*) represents whether the search radius is in the initial radius (15km). The map unit level (*d5*) represents the soil classification level of soil map unit. The digit *d5* is related to the detail level of the soil categorical map and the other digits are related to the linkage method. The importance of the above factors is assumed to decrease in the following order: the linkage level (*d1*), soil texture consideration (*d2*), sample size (*d3* or *d6*) and search radius (*d4*). For each factor, the code is better when the corresponding numerical number is smaller.

Table 1.　Quality control information of the derived soil properties.

Digit[a]	Name	Code
d1	Linkage level	1: family; 2: subgroup; 3: great group; 4: order; 5: (non-)acid; 6: Andosols; 7: Histosols. Andosols and Histosols are separated for their rather specific behavior.
d2	Texture consideration	0: texture was considered in the linkage; 1: texture was not considered in the linkage.
d3	Sample size level 1	1: 3N or more; 2: N-(3N-1); 3: 0-(N-1); 4: no data. N is the target sample size and has different value due to the linkage level. At (non-)acid level, N = 400; at order level, N= 200; at great group level, N = 40; at subgroup level, N = 10; and at family level, N = 5.
d4	Search radius flag	0: search radius is in the initial radius (15 km); 1: search radius is larger than the initial radius but in the soil map extent (i.e. at least one soil profile in the same soil type of a map unit are not included in the linkage); 2: the linkage takes place in the whole map.
d5	Map unit level	1-7: The meanings of the numbers are the same as those in *d1*.
d6	Sample size level 2	1: 30 or more; 2: 15-29; 3: 5-14; 4: 1-4; 5: no data.

[a] The quality control information is composed of 6 digits. From left to right are *d1*-*d6*.

The coordinate system is Krasovsky_1940, and the parameters are:
Semimajor Axis: 6378245.000000000000000000
Semiminor Axis: 6356863.018773047300000000
Inverse Flattening: 298.300000000000010000

3. Data usage

The data in NetCDF file format can be used by multiply software. Here we give three example softwares, i.e. Panoply, NCL and R.

3.1 Panoply

This software is recommended to a fast visual look at the data. It can be downloaded here (www.giss.nasa.gov/tools/panoply). This software can also be used to export data (in the file menu) as csv or txt files.

3.2 NCAR Command Language (NCL)

Here is an example of NCL script to use the data:

```
load "$NCARG_ROOT/lib/ncarg/nclscripts/csm/gsn_code.ncl"
load "$NCARG_ROOT/lib/ncarg/nclscripts/csm/gsn_csm.ncl"
```

```
begin

SAdata = addfile("SA.nc","r")
lat = SAdata->lat
lon = SAdata->lon
SA = SAdata->SA
RSA = SAdata->RSA
;printVarSummary(SA)

SA@_FillValue = -999

wks = gsn_open_wks("pdf","SA")
gsn_define_colormap(wks,"rainbow+white+gray")

res      = True
res@gsnAddCyclic = False

res@mpLimitMode = "LatLon"
res@mpMaxLatF = 54.0
res@mpMinLatF = 18.0
res@mpMaxLonF = 136.0
res@mpMinLonF = 73.0

res@cnFillOn=True
res@cnLinesOn=False

res@lbLabelAutoStride=True
res@lbBoxLinesOn=False

res@gsnSpreadColors=True
res@gsnSpreadColorStart=50
res@gsnSpreadColorEnd=-3

res@cnFillMode = "RasterFill"
res@cnLevelSelectionMode="ManualLevels"
res@cnMinLevelValF=0.0
res@cnMaxLevelValF=90.0
res@cnLevelSpacingF = 5.0

plot = gsn_csm_contour_map(wks,SA(0,:,:),res)

end
```

Note that workspace reallocation would exceed maximum size 32556688, the easiest way to increase the size is to put a line like the following into your ~/.hluresfile:

```
*wsMaximumSize : 500000000
```

3.3 R language

The NetCDF files can be used by loading "RNetCDF" package, and the corresponding maps can be drawn by loading “raster” package. The following is an example:

```
rm(list=ls(all=TRUE))
setwd("D:\\NC\\data")    # The directory of NetCDF file
library("RNetCDF")
library(raster)

cnfile<-"SA.nc"
q3<-open.nc(cnfile, write=FALSE)
print.nc(q3)

r <- raster(ncol=7560, nrow=4320,xmn=73, xmx=136, ymn=18, ymx=54)
#read value
tmp<-var.get.nc(q3,"SA",c(1,1,1),c(7560,4320,1))
#plot maps
tmp<-tmp[,4320:1]
values(r)<-as.vector(tmp)
plot(r, asp=1)
close.nc(q3)
```

3.4 ArcGIS

The NetCDF files should be converted into single layer (depth) NetCDF file first. This can be done by NCL, R language or other tools. Then use the “Make NetCDF Raster Layer” in Arctool to make an in-memory raster layer. Then right click on the layer name in ArcMap. The data can be export to raster format through “data” -> “export data”.

The following is an example of R language to convert the data into single layer (depth) NetCDF file:

```
rm(list=ls(all=TRUE))
setwd("J:\\nc")    # The directory of NetCDF file
library("RNetCDF")


cnfile<-"SA.nc"    # original file name
q3<-open.nc(cnfile, write=FALSE)
print.nc(q3)
```

```
tname<-"SA"    # soil property name
#read value
i=1 #change the value to extract different layer(depth)
tmp<-var.get.nc(q3,tname,c(1,1,i),c(7560,4320,1))#the i layer

t1<-seq(length=7560,from=73.00417,by=0.00833)
t2<-seq(length=4320,from=18.00417,by=0.00833)


q2<-create.nc(paste("SA",i,".nc",sep=""))

#define variables and attributes of    global, dimension and coordinates
dim.def.nc(q2, "lon",7560)
dim.def.nc(q2, "lat",4320)

att.put.nc(q2, "NC_GLOBAL", "Conventions", "NC_CHAR", "CF-1.0")
var.def.nc(q2, "lon", "NC_FLOAT", "lon")
var.def.nc(q2, "lat", "NC_FLOAT", "lat")
att.put.nc(q2, "lon", "long_name", "NC_CHAR", "longitude")
att.put.nc(q2, "lon", "units", "NC_CHAR", "degrees_east")
att.put.nc(q2, "lat", "long_name", "NC_CHAR", "latitude")
att.put.nc(q2, "lat", "units", "NC_CHAR", "degrees_north")

var.def.nc(q2,tname, "NC_FLOAT", c("lon","lat"))
att.put.nc(q2,tname, "missing_value", "NC_FLOAT",-999.0)
att.put.nc(q2, tname, "units", "NC_CHAR", "% of weight")
att.put.nc(q2, tname, "long_name", "NC_CHAR", "sand content")
var.put.nc(q2,"lat", t2, 1, NA)
var.put.nc(q2,"lon", t1, 1, NA)
var.put.nc(q2,tname,tmp,c(1,1),c(7560,4320))


close.nc(q3)
close.nc(q2)
```

4. Citation

Details about the dataset are in the peer-reviewed paper.

Full acknowledgement and referencing of all sources must be included in any documentation using any of the material contained in the China Dataset of Soil Properties for Land Surface Modeling, as follows:

Shangguan, W., Y. Dai, B. Liu, A. Zhu, Q. Duan, L. Wu, D. Ji, A. Ye, H. Yuan, Q.

Zhang, D. Chen, M. Chen, J. Chu, Y. Dou, J. Guo, H. Li, J. Li, L. Liang, X. Liang, H. Liu, S. Liu, C. Miao, and Y. Zhang (2013), A China Dataset of Soil Properties for Land Surface Modeling, Journal of Advances in Modeling Earth Systems, *5*, 212-224, doi:10.1002/jame.20026.

5. Reference

Batjes, N. H. (2002), Soil parameter estimates for the soil types of the world for use in Dglobal and regional modelling (Version 2.1), ISRIC Report 2002/02c, International Food Policy Research Institute (IFPRI) and International Soil Reference and Information Centre (ISRIC), Wageningen.

Shangguan, W., Y. Dai, B. Liu, A. Ye, and H. Yuan (2012), A soil particle-size distribution dataset for regional land and climate modelling in China, Geoderma, 171-172, 85-91.

6. Contact

If you have any questions or feedbacks when using the data sets, please email: shgwei@mail.sysu.edu.cn (Dr. Wei Shangguan).