A Conterminous United States Multilayer Soil Characteristics Dataset for Regional Climate and Hydrology Modeling

Douglas A. Miller* and Richard A. White

Earth System Science Center, The Pennsylvania State University, University Park, Pennsylvania

Received 7 March 1997; accepted 19 January 1998.

ABSTRACT: Soil information is now widely required by many climate and hydrology models and soil–vegetation–atmosphere transfer schemes. This paper describes the development of a multilayer soil characteristics dataset for the conterminous United States (CONUS-SOIL) that specifically addresses the need for soil physical and hydraulic property information over large areas. The State Soil Geographic Database (STATSGO) developed by the U.S. Department of Agriculture–Natural Resources Conservation Service served as the starting point for CONUS-SOIL. Geographic information system and Perl computer programming language tools were used to create map coverages of soil properties including soil texture and rock fragment classes, depth-to-bedrock, bulk density, porosity, rock fragment volume, particle-size (sand, silt, and clay) fractions, available water capacity, and hydrologic soil group. Interpolation procedures for the continuous and categorical variables describing these soil properties were developed and applied to the original STATSGO data. In addition to any interpolation errors, the CONUS-SOIL dataset reflects the limitations of the procedures used to generate detailed county-level soil

* Corresponding author address: Douglas A. Miller, Earth System Science Center, The Pennsylvania State University, University Park, PA 16802.
E-mail address: miller@essc.psu.edu

survey data to the STATSGO map units. CONUS-SOIL is available in several formats and may be accessed via the World Wide Web.

KEYWORDS: Unsaturated zone; Soil moisture; Land/atmosphere interactions; Hydroclimatology; Hydrologic budget

1. Climate and hydrology model requirements for soil information

Over the past several decades the climate and hydrology modeling communities have been developing increasingly sophisticated parameterizations of the interaction between the land surface and the atmosphere in so-called soil–vegetation–atmosphere transfer schemes (SVATS). A major requirement of these process descriptions is an understanding of the surface and subsurface nature of the soil environment. The soil controls the downward movement of water in the subsurface and the amount of water available for evapotranspiration. Knowledge of soil physical and hydraulic properties is, therefore, a key element in correctly modeling land surface atmosphere exchange processes. Unfortunately, a lack of information about soils at the regional scale has impeded the improvement of SVATS components in small-scale (large area) atmospheric and hydrologic models. Various approaches to modeling soil surface and subsurface control on water and energy distribution in these predominantly one-dimensional models have been developed.

Early soil hydrology treatments in general circulation models considered the soil mainly as a storage reservoir, or ‘‘bucket,’’ with a holding capacity equal to 15 cm of water in the upper 1 m of soil. The bucket fills with water when precipitation is greater than evaporation. The soil is allowed to evaporate water at some potential rate until a critical threshold (usually 75% of the field capacity of the soil), whereupon evaporation then proceeds as a fraction of the potential rate. Runoff is not accounted for in these models (Manabe, 1969).

The inadequacy of the bucket model in describing soil water processes led numerous researchers to adopt deterministic solutions that more closely simulate water movement processes in the soil profile (Dickinson et al., 1993; Abramopolous et al., 1988; Wilson et al., 1987). These models are based on Darcy's law of water movement, which relates the flux of downward infiltrating water (q) moving proportionally to the forces of gravitation and the matric potential (ψ):

$$q = K\frac{\partial \psi}{\partial z} + K, \tag{1}$$

where z is depth and K is the hydraulic conductivity. Darcy's equation can be combined with the equation for continuity,

$$\frac{\partial \theta}{\partial t} = -\frac{\partial q}{\partial z}, \tag{2}$$

where θ is the volumetric soil water content and t is time, to fully account for the diffusion and gravitation components of water movement. The result is formally known as the Richards equation:

$$\frac{\partial\theta(z)}{\partial t} = \frac{\partial}{\partial z}\left[K(q)\frac{\partial\psi(q)}{\partial q}\frac{\partial q}{\partial z}\right] - \frac{\partial K(q)}{\partial q}\frac{\partial q}{\partial z}, \tag{3}$$

which describes the flow of water in the unsaturated zone as a function of soil water content and its vertical gradient. Solution of Eq. (3) requires an understanding of the water retention characteristic; the relationship between the water content, θ; the matric potential, ψ; and the hydraulic conductivity, K.

Laboratory and field methods for determining these soil hydraulic properties are time consuming and expensive. Pedo-transfer functions (PTFs) have been developed to derive hydraulic characteristics from data (e.g., soil texture, particle-size distribution, bulk density, porosity) gathered in the course of traditional soil surveys. Tietje and Tapkenhinrichs (Tietje and Tapkenhinrichs, 1993) differentiate among the methods for deriving PTFs. Point regression analysis predicts the soil water content at specified matric potentials using multiple linear regression. Input (regressor) variables include, for example, particle-size distribution, bulk density, organic matter content, and porosity. Physical methods rely on the relationship between particle-size distribution and porosity and their relationship to water content and matric potential through mass conservation and capillarity. Functional parameter regression methods use multiple nonlinear regression analysis in conjunction with physical variables and closed form functions for the relationship between matric potential and water content. The most frequently applied models that require functional parameters are based on the work of Brooks and Corey (Brooks and Corey, 1964), Campbell (Campbell, 1974), Clapp and Hornberger (Clapp and Hornberger, 1978), and van Genuchten (van Genuchten, 1980). Table 1 summarizes the parameters for these power law–based models. A number of studies have tested and evaluated these PTFs using a wide variety of soil physical property information in a range of settings (Arya and Paris, 1981; Kern, 1995b; Saxton et al., 1986; Rawls et al., 1991; Vereecken et al., 1990). These studies have relied on the availability of standard soil survey and characterization information for the input variables.

Despite the implementation of soil hydrologic processes in SVATS there has been a dearth of spatial information on soil physical and hydraulic properties for regional climate and hydrology applications. Webb et al. (Webb et al., 1993) produced a global dataset, at 1° by 1° spatial resolution, of soil profile physical properties by combining the Food and Agriculture Organization of the United Nations/United Nations Educational, Scientific and Cultural Organization Soil Map of the World with the World Soil Data File of Zobler (Zobler, 1986). This data specifies the top and bottom depths and percentages of sand, silt, and clay of individual soil horizons for 106 soil types. Kern (Kern, 1995a) used the U.S. Department of Agriculture–Natural Resources Conservation Service (USDA–NRCS) National Soil Geographic Database Major Land Resource Areas as a geographic base and the 1992 National Resources Inventory and the associated Soil Interpretations Record (SIR) to estimate, for the continuous 48 United States, geographic patterns of soil-water-holding capacity. He used empirically developed regression models from the literature in conjunction with soil physical properties, including soil classification; rock fragment; sand, clay, and organic matter contents; depth to indurated layers; and depth-to-bedrock to obtain spatially distributed estimates of soil water retention. Lathrop et al. (Lathrop et al., 1995) reported the use of STATSGO available water holding capacity data in a forest ecosystem model for the northeast United States. They reported significant challenges regarding the within-unit heterogeneity of STATSGO data and the need to quantify this variability for use in ecosystem modeling. Zheng et al. (Zheng et al., 1996) compared available water capacity estimated from topography to STATSGO data in Montana.

This paper describes the development of a multilayer soil characteristics dataset (CONUS-SOIL) for use in regional and continental-scale climate and hydrology models over the conterminous 48 United States. This effort is unique in that, for the first time, the science community will have access to a dataset of soil physical and hydraulic properties specifically designed for modeling applications.

2. Data and methods

2.1. State Soil Geographic Database (STATSGO)

The USDA–NRCS, through the National Cooperative Soil Survey (NCSS), is in the process of developing soil geographic databases at three scales: local, regional, and national. At the regional level, the State Soil Geographic Data Base was released in 1992 for use in river basin, multicounty, multistate, and state resource planning. This database was created by generalizing available soil survey maps, including published and unpublished detailed soil surveys, county general soil maps, state general soil maps, state major land resource area maps, and, where no soil survey information was available, Landsat imagery (Reybold and Tesselle, 1989).

STATSGO consists of georeferenced digital map data and associated digital tables of attribute data. The compiled soil maps were created using the U.S. Geological Survey (USGS) 1° by 2° topographic quadrangles (1:250,000 scale, Albers equal area projection) as base maps, which were then merged on a state basis. The District of Columbia is included with the data for Maryland. The full STATSGO database is available from the NRCS on CD-ROM and is also available online over the Internet at the following location: http://www.ncg.nrcs.usda.gov/statsgo ftp.html (Soil Survey Staff, 1994a).

Map units in STATSGO are a combination of associated phases of soil series. A soil series is the lowest level in the U.S. system of taxonomy (Soil Survey Staff, 1993) and the most homogeneous with regard to properties. A phase of a soil series is based on attributes and factors that affect soil management. Map unit composition was derived from a statistical analysis of transects across detailed soil survey maps. Percentages of the map unit components were based on the length of the map units crossed. The total number of transects was based on the size, number, and complexity of the detailed soil map delineations. Details of the exact procedures for determining map unit composition may be found in the

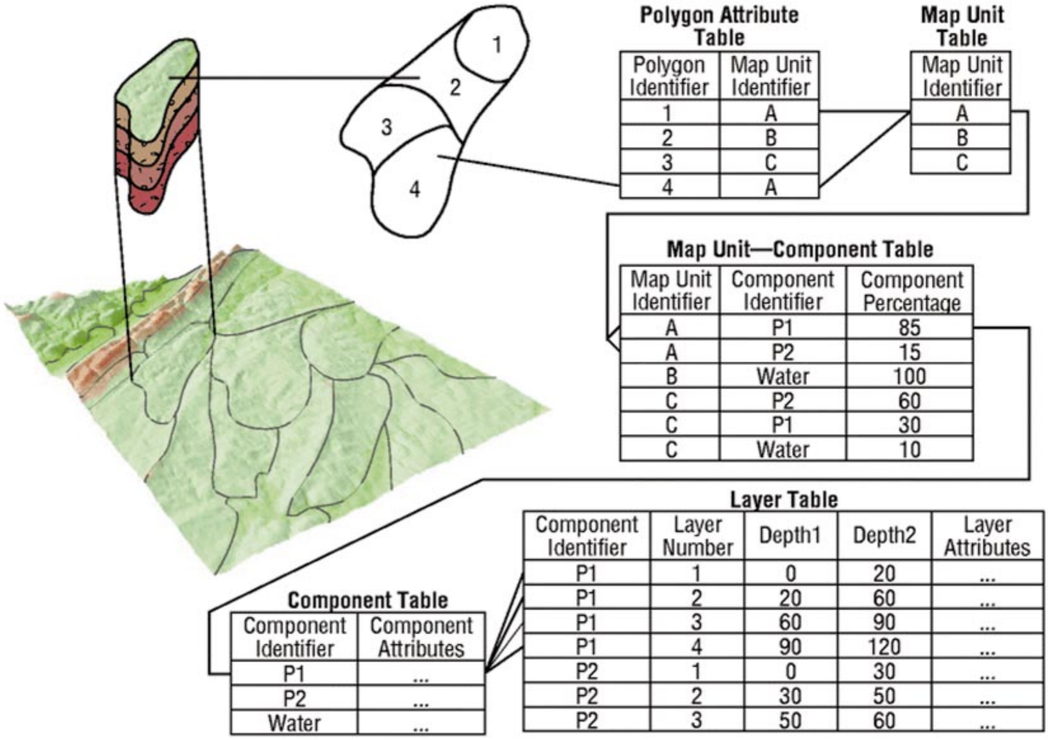

Figure 1. Schematic diagram of the STATSGO database structure.

STATSGO Data Users Guide (Soil Survey Staff, 1994b). The approximate minimum map unit area delineated is documented as 625 hectares (1544 acres) with minimum linear dimensions of 1.25 km. The number of map units delineated on each 1° by 2° quadrangle is typically 100–200 but may range up to 400 (Soil Survey Staff, 1994b).

Figure 1 is a schematic diagram depicting several of the elements in the STATSGO database. Each map unit is specified digitally by one or more georeferenced polygons and may consist of as many as 21 components (a component being a phase of a soil series), which are linked to the NCSS SIR. Attributes include characteristics such as map unit area percentage, surface texture and texture modifier, slope range, and flooding category. STATSGO attribute information is organized, on an individual state basis, into a set of 15 tables that are linked to the georeferenced polygons. The tables are logically organized to contain related items. The Map Unit, Component, and Layer tables provide the relevant data items for climate and hydrology applications. The data items used in developing the soils dataset, along with table locations and definitions, are provided in Table 2.

2.2. Data processing

The STATSGO data, as distributed by NRCS, present significant challenges for climate and hydrology modelers who would like to use soils information, but who are unfamiliar with soil survey and spatial data formats. These challenges include the following:

1. Many climate and hydrology models require a uniform grid cell or raster format, while the STATSGO map units are defined as polygons in a vector geographic information system (GIS) environment.
2. The number, thickness, and depth to top and bottom of soil layers vary widely from one soil component to another, even within the same map unit.
3. Although the relative amount of each component within a map unit is specified (variable COMPPCT), no information is provided about the location of the component within the map unit.
4. For approximately half the components, the minimum and maximum depth-to-bedrock (ROCKDEPL and ROCKDEPH) both have the value 152 cm (60 in.); in the great majority of these cases, this indicates that this was the maximum depth to which soil was normally examined and bedrock was not actually encountered.

Our goal in extracting information from the STATSGO database was to produce derived map coverages that would meet the needs of modelers regardless of their familiarity with, or access to, GIS resources.

The STATSGO data for each state are available as a vector (polygon) map coverage and a set of tables in the format generated by the Arc/Info GIS software from Environmental Systems Research Institute, Inc. Therefore, Arc/Info was initially used to import the STATSGO data from the NRCS CD-ROM. All derived soil physical and hydraulic properties were compiled on a map unit basis and entered into Arc/Info tables linked to the map units. The vector-format (polygon) data were subsequently rasterized to uniformly spaced grids.

Grid-based models typically require that vertical soil profiles be divided into the same layers at each grid point; the great range and diversity of STATSGO layer thicknesses for different components cannot be handled by such models (Table 3). Accordingly, all data from the STATSGO Layer tables were interpolated to a set of standard layers. These layers provide a consistent framework for the CONUS-SOIL dataset (Table 4).

The selection of standard layers represents a trade-off between the desire to retain as much structural information as possible while avoiding an unmanageably large number of layers. Since the STATSGO layers tend to be thinner near the top of the soil profile, and this is also the region where many models require the most detail, the top two standard layers were assigned thicknesses of 5 cm (approximately 2 in.). Progressively larger thicknesses were chosen as the depth increased, yielding the 11 standard layers. For the majority of STATSGO components, soil properties were apparently not sampled to depths greater than 152 cm (60 in.). The maximum depth reported for any component layer was 250 cm (98 in.); only about 2.5% of components have layers extending below 203 cm (80 in.). Accordingly, the bottom two standard layers contain meaningful data only for a minority of map units.

The interpolation of data from the STATSGO layers for each component to the standard layers was performed in the following way. The top and bottom depths of each standard layer were compared to the top and bottom depths of each layer for the STATSGO component. If the standard layer was entirely contained within a single STATSGO layer, the standard layer was assigned the data value for that layer. Otherwise, all STATSGO layers that were fully or partially included within the standard layer were identified, and the amount of overlap between each of these layers and the standard layer was determined. For continuous-valued variables (such as bulk density and available water capacity), an average value for the standard layer was computed by weighting the values for each of the contributing STATSGO layers by the amount of overlap. For categorical variables (e.g., soil texture and rock fragment classes), the overlap amounts were used to compute the relative amount of each category within the standard layer.

If the bottom of the deepest STATSGO layer was above the mean depth-to-bedrock specified for the component, the STATSGO layer was assumed to extend down to bedrock. For a number of components, the deepest STATSGO layer extended below the maximum rock depth specified for the component. In these cases, bedrock was assumed to begin at the bottom of the deepest STATSGO layer. For the large proportion of components for which the depth-to-bedrock was above the bottom of the deepest standard layer, 250 cm (98 in.), the portion of the standard layer(s) below the bedrock depth was assigned a value of the physical or hydraulic property appropriate to solid rock. Because a rock depth of 152 cm (60 in.) is frequently used to indicate the maximum depth of the soil profile, this assignment is often misleading for the bottom two standard layers.

Since no information is provided about the location of each component within the map unit, all physical and hydraulic variables were aggregated over all components of a given map unit. For continuous-valued variables, average values for the map unit were computed by weighting the values for each component by the value of COMPPCT. For categorical variables, the relative amounts of each category in each component were multiplied by COMPPCT and summed; the category having the largest total was chosen as the category for the map unit, even if it represented less than half the area of the map unit. These two procedures are henceforth identified as the ‘‘continuous-’’ and ‘‘categorical-’’ aggregation procedures.

Although the Arc/Info GIS environment is convenient for storing map unit boundaries and related tables, it proved cumbersome for performing the actual computations. Perl (Practical Extraction and Report Language) with its flexible approach to processing text and files (Wall and Schwartz, 1991) was found to provide the capabilities needed for processing the attribute tables for the 48 states. A Perl subroutine was written to access the Arc/Info data structure and extract variables of interest from the STATSGO database. Individual Perl scripts were then written to process, on a state basis, the desired soil physical and hydraulic properties (White and Miller, 1998). The nature of the STATSGO data, as revealed in initial tests of the software, required the use of error checks and flags for inconsistencies. Outputs from the Perl scripts included an ASCII file with the desired physical or hydraulic property information and a report file with messages regarding problems and inconsistencies encountered during processing. Short Perl scripts were also used to pack the ASCII output into binary formats suitable for

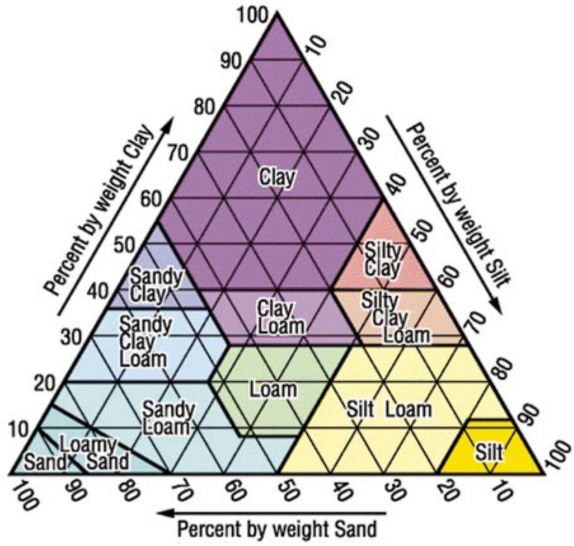

Figure 2. Standard USDA soil texture triangle.

import by the Arc/Info GIS software. This enabled use of the Arc/Info relational database facilities to establish links between the processed attribute information and the original STATSGO map unit polygons.

3. Soil physical and hydraulic properties

3.1. Soil texture and rock fragment classes

Soil texture and the type and amount of rock fragments in the soil impact the infiltration and redistribution of water within the soil profile. In the STATSGO database the dominant soil texture and rock fragment information, for each layer, are contained in a combined variable, TEXTURE1.

The categorical nature of these attributes precludes the creation of a weighted average variable from the map unit component percentages. Instead, the dominant soil texture and rock fragment classes were determined for each of the 11 standard layers using the categorical aggregation procedure.

Table 5 and Table 6 list the texture and rock fragment classes respectively and their assigned numerical codes. Water, organic materials (peat, muck, etc.), bedrock, and other nonsoil surfaces cannot be placed in one of the 12 standard soil texture classes (Figure 2). These classes have been combined into ‘‘water,’’ ‘‘organic materials,’’ ‘‘bedrock,’’ and, ‘‘other’’ and are also listed in Table 5 and Table 6. Including these classes in the final data product provides maximum flexibility for making decisions with regard to the use of the data. Additional lookup tables may be used to interpret these classes. Standard GIS software packages, if available to CONUS-SOIL users, often have functions that can be used to aggregate classes. Otherwise, standard computing languages like FORTRAN and C can be used to create software for the aggregations.

3.2. Depth-to-bedrock

Scientists using climate and hydrology models would benefit from a knowledge of the depth of soil and/or unconsolidated material that lies between the land surface and the geologic substratum. The STATSGO Component table contains maximum (ROCKDEPH) and minimum (ROCKDEPL) values for the depth-to-bedrock (DTB). A mean value for the DTB of a map unit may be computed by weighting the mean of the ROCKDEPL and ROCKDEPH values for each component by the percentage of each component within the map unit:

$$\mathrm{DTB} = \sum_{1}^{n} (\tfrac{1}{2}[\mathrm{ROCKDEPL} + \mathrm{ROCKDEPH}] \times \mathrm{COMPPCT}). \qquad (4)$$

As noted previously, the DTB calculations in CONUS-SOIL are complicated by the nature of soil survey practice and the subsequent compilation of the SIR that was used to attribute the STATSGO map unit polygons. Soil survey methods in the United States, as in many other countries, are heavily influenced by the practice of soil classification. In the United States, Soil Taxonomy (Soil Survey Staff, 1975) is the official system for the classification of soils and is used heavily in the soil mapping process. Soil Taxonomy has a control section in the soil profile from which soil morphological, chemical, and physical properties are used in classification. Generally, this control section ranges from 25 to 100 cm below the surface. In areas where soils are very deep or have subtle lower boundaries, classification and mapping without complete knowledge of the lower boundaries between soil and geologic material produces a soil survey that is unlikely to contain reliable information on DTB (Buol et al., 1989). STATSGO, having been derived from the original detailed soil surveys of the United States, contains this bias. The Component (COMP) table in STATSGO appears to use a value of 60 for both the ROCKDEPL and ROCKDEPH variables to indicate that bedrock was not encountered within 152 cm (60 in.) of the surface.

3.3. Bulk density and porosity

The nature and structure of the soil volume is a major controlling factor in the movement and storage of water in the subsurface. Soils may be thought of as a three-phase system consisting of air, water, and solid materials that vary continually over space and through time (Hillel, 1980a). A knowledge of the relative distribution of these quantities improves predictions about infiltration and subsurface water movement. The soil bulk density (BD) is the ratio of the mass of soil material to the total volume of solids plus pores. The BD is often used as a predictive variable in PTFs that predict soil water retention properties (Tietje and Tapkenhinrichs, 1993).

The occurrence of BD in the STATSGO database follows the internal standard of providing the maximum (BDH) and minimum (BDL) values on a layer basis for each component in a STATSGO map unit. The mean value for BD for each STATSGO component layer was computed as

$$\mathrm{BD} = \tfrac{1}{2}[\mathrm{BDL} + \mathrm{BDH}]. \qquad (5)$$

The values for individual components were then interpolated to the 11 standard soil layers and aggregated over all components of the map unit using the continuous aggregation procedure.

Porosity is a measure of the volume of air- and water-filled pores in the soil. The STATSGO database does not contain a direct measure of soil pore space. However, a value for porosity can be calculated from the BD and the particle density (PD, the ratio of the mass of all solid particles to the volume of the solids) according to

$$\text{porosity} = 1 - [\text{BD/PD}], \tag{6}$$

where PD is generally assumed to be particle density of 2.65 g cm^{-3} (Hillel, 1980a). The BD values determined from Eq. (5) were used to calculate a porosity for each layer of the STATSGO map unit components unless bedrock was present, in which case a porosity value of zero was assigned. These values were then interpolated to the standard layers and aggregated over all components of the map unit in the same way as the BD values. COMPPCT was then used to weight the component contribution to the final map unit standard layer porosity.

Somewhat more than 25% of all map units contained one or more nonwater components that specified BDH and BDL both equal to 0 for all layers. These components were omitted from the computations of mean bulk density and porosity unless the depth-to-bedrock was specified as zero, in which case the component was included and assigned bulk density and porosity values appropriate to bedrock.

3.4. Rock fragment volume

Infiltration is significantly affected by the size and amount of rock fragments (>2 mm fraction) in the soil. Soil physical properties, primarily bulk density and porosity, are modified by fragments >2 mm. The presence of rock fragments changes the soil pore space amount and structure, which modifies the size and distribution of pathways for water movement through the soil (Brakensiek and Rawls, 1994).

A ratio can be devised to indicate the fraction by volume of the <2 mm soil fabric, within the layers of a component (Soil Survey Staff, 1996; Bliss et al., 1995). This ratio is

$$R = \frac{V_{\text{fines}}}{V_{\text{fines}} + V_{\text{rock}}}, \tag{7}$$

where V_{fines} is the volume of fine soil material (<2 mm fraction) and V_{rock} is the volume of rock material. This may be expressed on a mass basis as

$$R = \frac{M_f/\text{BD}}{M_f/\text{BD} + M_{Tr}/\text{PD}} \tag{8}$$

$$R = \frac{R_{ft}/\text{BD}}{R_{ft}/\text{BD} + R_{rt}/\text{PD}}, \tag{9}$$

where M_f is the mass of the fine soil fraction, M_r is the mass of rock fragments

(>2 mm), R_{ft} is the ratio of the mass of soil fines to the total mass of the soil, and R_{rt} is the mass fraction of rock.

The value of R_{ft} is calculated as

$$R_{ft} = (R_{fs})(R_{st}), \tag{10}$$

where R_{fs} is the sampled fraction of soil material that passes a No. 10 sieve and is calculated directly from the STATSGO database by taking the mean of NO10H and NO10L variables from the Layer table,

$$R_{fs} = [(\mathrm{NO10L} + \mathrm{N10H})/200]. \tag{11}$$

Here R_{st} is the mass fraction of sampled soil, which is calculated arithmetically as a remainder after accounting for rock fragments between 3 and 10 in. (R_{yt}) in size and rock fragments greater than 10 in. (R_{zt}):

$$R_{st} = 1 - R_{yt} - R_{zt}. \tag{12}$$

A range of values for both R_{yt} and R_{zt} are given in the Layer table; the means of the upper and lower ends of the specified ranges were used:

$$R_{yt} = \tfrac{1}{2}[\mathrm{INCH3L} + \mathrm{INCH3H}] \tag{13}$$

$$R_{zt} = \tfrac{1}{2}[\mathrm{INCH10L} + \mathrm{INCH10H}]. \tag{14}$$

The mass fraction of rock is determined by subtracting the mass fraction of fine material, R_{ft}, from the total for all materials:

$$R_{rt} = 1 - R_{ft}. \tag{15}$$

The value of R computed using Eq. (5) and Eqs. (7)–(15) is then used to determine the volume of rock fragments, VRF, in a given component layer:

$$\mathrm{VRF} = 1 - R. \tag{16}$$

The mean volume of rock fragments for each standard layer of each map unit is obtained by interpolating the values for each component to the standard layers, and then averaging the values for all components of the map unit, using the continuous aggregation procedure.

As noted above, more than 25% of map units had some nonwater components specifying BDH and BDL both equal to 0. This leads to division by zero in Eq. (9). In many of these cases, no values were specified for NO10H and NO10L. For components with BD equal 0, the rock volume was assumed to be zero unless NO10L and NO10H was explicitly given as zero or the depth-to-bedrock was zero. In these cases the component layer was assumed to be all rock.

There were also quite a few component layers that contained ambiguous information, including almost 1% of all layers for which values of INCH3 and INCH10 were specified, but NO10H or NO10L were not provided. All components containing ambiguous layers were omitted from the computation of mean rock values unless the depth-to-bedrock was zero, in which case the component was assumed to be 100% rock.

3.5. Sand, silt, and clay fractions

Frequently, climate and hydrology models require information about soil physical properties in the form of a continuous distribution rather than discrete classes.

This is frequently the case with the percentages of sand, silt, and clay within the soil profile. The STATSGO database contains information on clay content and sieved samples on a component layer basis that could potentially allow the subsequent determination of sand and silt fractions. Preliminary examination of the database indicated that calculations based on these variables would not provide consistent results. Specifically, percentages of the sand, silt, and clay fractions should total to 100% percent of the soil material less than 2 mm. Initial calculations using these variables for the STATSGO data for Pennsylvania revealed that these fractions rarely summed to 100%.

As an alternative to using the clay content and sieve information, the USDA soil texture diagram (Figure 2) was used to determine the midpoint values of sand, silt, and clay for each of the soil texture classes. These values, shown in Table 7, were then used to determine percentages of sand, silt, and clay in each STATSGO component layer. These percentages were then interpolated to the 11 standard layers for CONUS-SOIL and aggregated over the components for each map unit. Many map units contain nonsoil components (organic materials, water, bedrock, or other), and the lowest standard layers for most components contained bedrock. As a result, the sum of the computed sand, silt, and clay fractions was often less than 100%. To make the results representative of the actual soil components, the sand, silt, and clay fractions were normalized to make them total 100% (before rounding) if at least 50% of the nonwater components for a given standard layer were soil; otherwise the fractions were all set to zero. The final output from these calculations is packaged as three separate map coverages (sand, silt, and clay fractions) for each standard layer in CONUS-SOIL.

3.6. Available water capacity

Information about soil water storage can be used to aid in the determination of the water and energy balance at the land surface. Traditionally, a quantity known as the available water capacity (AWC) has been used to indicate the amount of water in the soil profile that is available for use by plants. AWC, as defined in STATSGO, is that amount of water held in the soil between field capacity and permanent wilting point. Field capacity is the amount of water that remains after an initially saturated soil is allowed to drain for 2–3 days. Permanent wilting point is the root zone soil wetness at which the wilted plant can no longer recover turgidity even when it is placed in a saturated atmosphere for 12 h (Hillel, 1980b).

AWC is provided in STATSGO as minimum (AWCL) and maximum (AWCH) values in units of inches of water per inch of soil on a layer basis for each map unit component. AWCL and AWCH were used to generate a mean AWC for each component layer. In principal, the mean AWC for each layer could then be summed to create the total AWC for the component. This would in many cases be misleading, since the depth of the lowest layer varies from map unit to map unit, with 152 cm (60 in.) being the most common depth. Accordingly, AWC for each component was computed for three different profile depths, 100 cm (39 in.), 150 cm (59 in.), and 250 cm (98 in.). These values were then averaged over all components of the map unit, weighted by their percentage contribution. Map unit components consisting entirely of water were omitted from the average computation.

3.7. Hydrologic soil group

Many hydrologic models use simplified empirical formulations for infiltration to determine storm runoff volumes from land surfaces. These models often rely on an empirical index, the runoff curve number (CN), which rates the hydrologic response of various combinations of soil type and land use/land cover. USDA–NRCS has classified all of the soil series in the United States into one of four hydrologic groups. The classes and their definitions are shown in Table 8. The STATSGO Component table contains the HYDGRP variable, which represents these classes. In addition to the standard classes A, B, C, and D, HYDGRP may also take on the following combinations of classes: A/D, B/D, and C/D.

A table of HSG values was created on a map unit basis for CONUS-SOIL by determining the component contributions of each HSG class within a map unit. The mixed classes were conservatively assigned to HSG D. The percentage of each HSG class in a map unit was placed in a table linked to the map unit. When the HSG coverage is combined with additional information on land use/land cover, a weighted CN value for each map unit can be calculated.

4. Dataset structure

The soil physical and hydraulic properties described above are required in formats amenable to climate and hydrology models. These models generally work on a regular two-dimensional grid cell structure with one value assigned to a cell for any one particular property. The original vector format (Figure 1) of STATSGO is quite robust for efficient storage of multiple attributes associated with the soil map unit polygons. In order to provide the end user with the most functional dataset, however, the original vector format must be converted to a uniform grid. A grid resolution of 1 km was selected. This grid resolution is compatible with the minimum linear dimension (1.25 km) of the original STATSGO map units and matches the resolution of other readily available land surface characteristics datasets including topography, land cover, and satellite vegetation indices.

Some modelers use GIS technology and prefer to work with the vector data structure and perform their own grid and map projection conversions. Others may prefer to work with a previously gridded map product imported into their GIS environment. Finally, many climate and hydrology modelers have their own internal format conventions and would prefer to use a simply formatted dataset that can be read with commonly available software or computer languages. To meet the needs of the latter two groups, the facilities of the Arc/Info GIS software were used to rasterize each of the previously described soil physical and hydraulic properties from the vector (polygon) map coverage to a 1-km resolution grid. To provide maximum user flexibility, the original Albers equal area projection was also converted to the Lambert azimuthal projection, which has been used for a number of other datasets covering the conterminous United States, such as the land cover and topographic data distributed by the USGS Earth Resources Ob-

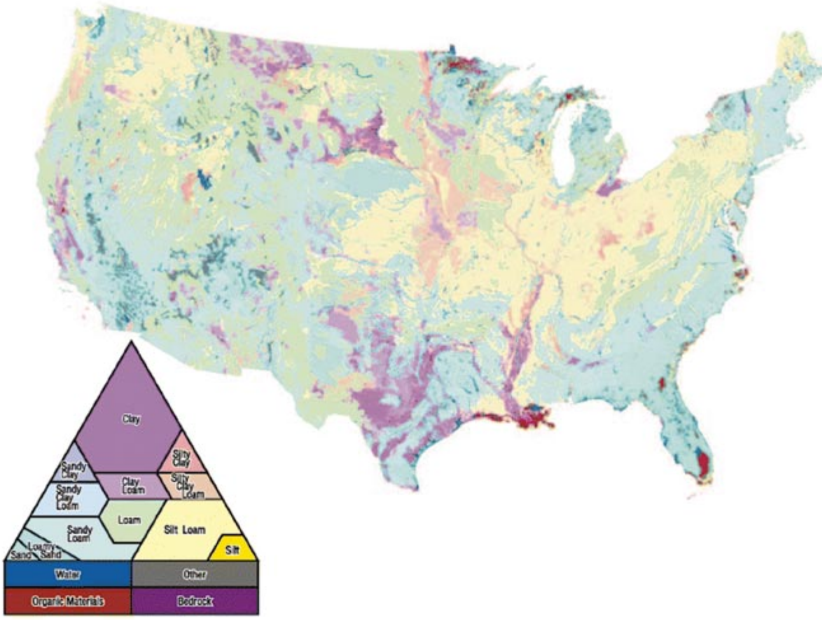

Figure 3. CONUS-SOIL surface texture.

servation Satellite Data Center (Loveland et al., 1991). Some climate and hydrology models work with geographic coordinates (latitude and longitude) rather than equal area projections. Accordingly a third gridded representation of the dataset was created for these applications, using a grid spacing of 30 arcseconds. To meet the needs of users who do not have access to GIS facilities, the gridded datasets were also converted to simple two- and three-dimensional arrays of binary values. FORTRAN and C language subroutines were written to enable users to read these arrays quickly and easily. A summary of the available data formats and map projections is provided in Table 9. In addition to the digital spatial data in CONUS-SOIL, a series of cartographic products depicting the various data layers have been created (Figure 3–Figure 11).

The World Wide Web (WWW) provides an ideal medium for delivery of spatial data products. Consequently we have developed a WWW server for CONUS-SOIL that allows easy access to all elements of the dataset including all spatial and tabular data, documentation, and cartographic products. CONUS-SOIL may be retrieved via FTP from the following World Wide Web address: http://www.essc.psu.edu/soil_info/. This paper serves as formal reference when including CONUS-SOIL in published work.

5. Nature and limitations of CONUS-SOIL

CONUS-SOIL represents the first focused attempt at creating an easily usable high-resolution soil physical and hydraulic properties dataset for regional climate

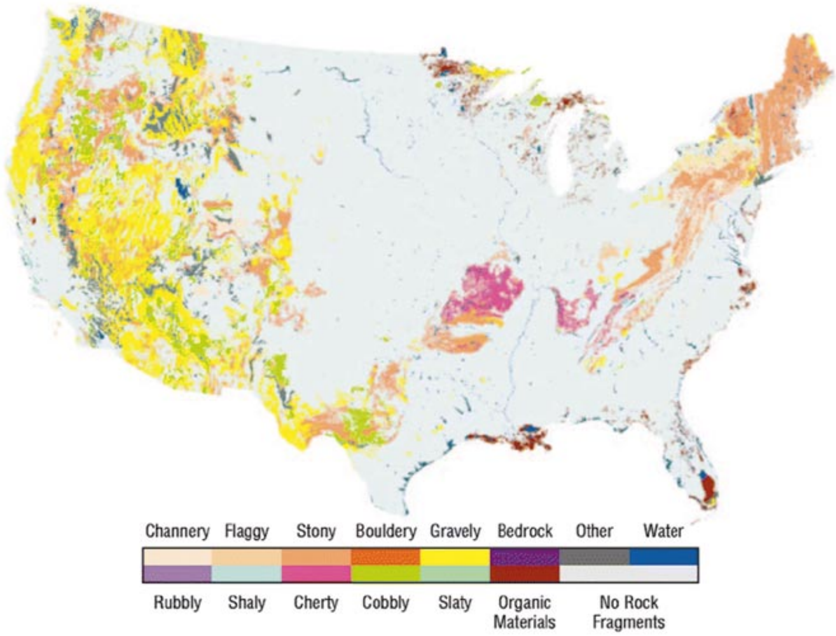

Figure 4. CONUS-SOIL surface rock fragment.

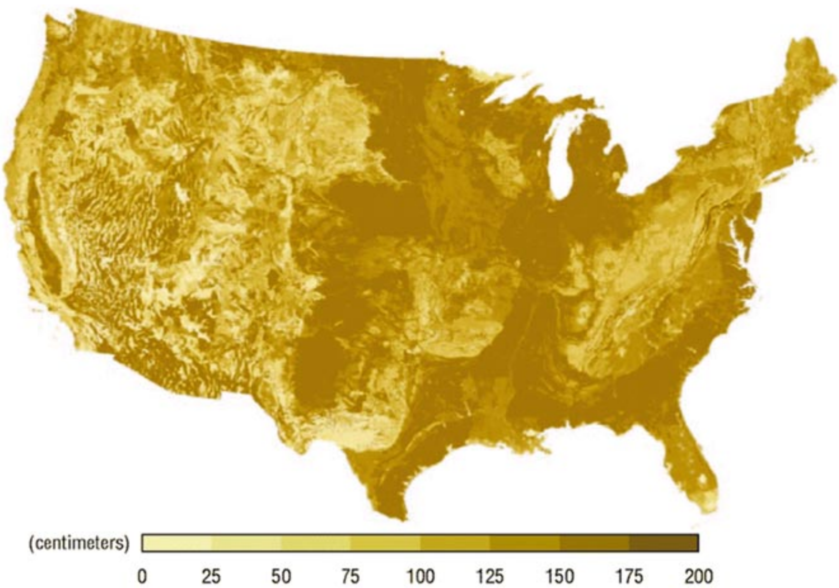

Figure 5. CONUS-SOIL depth-to-bedrock.

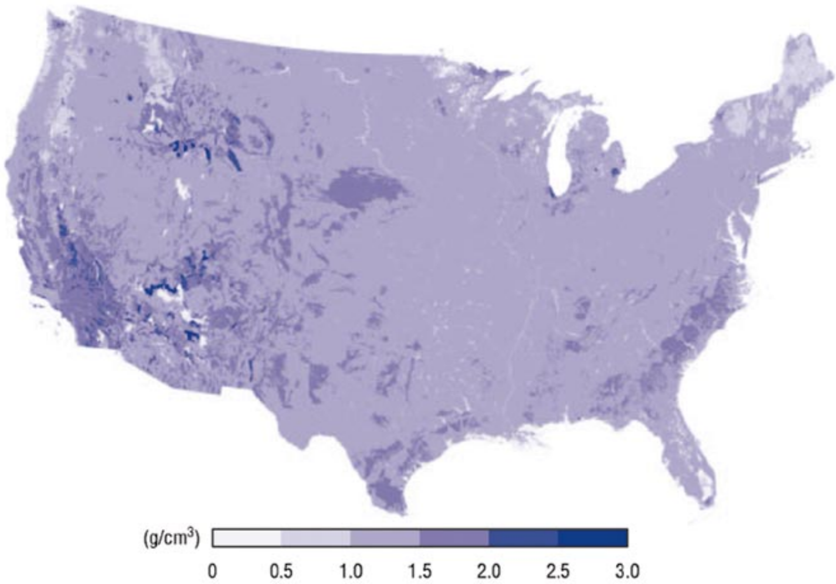

Figure 6. CONUS-SOIL bulk density.

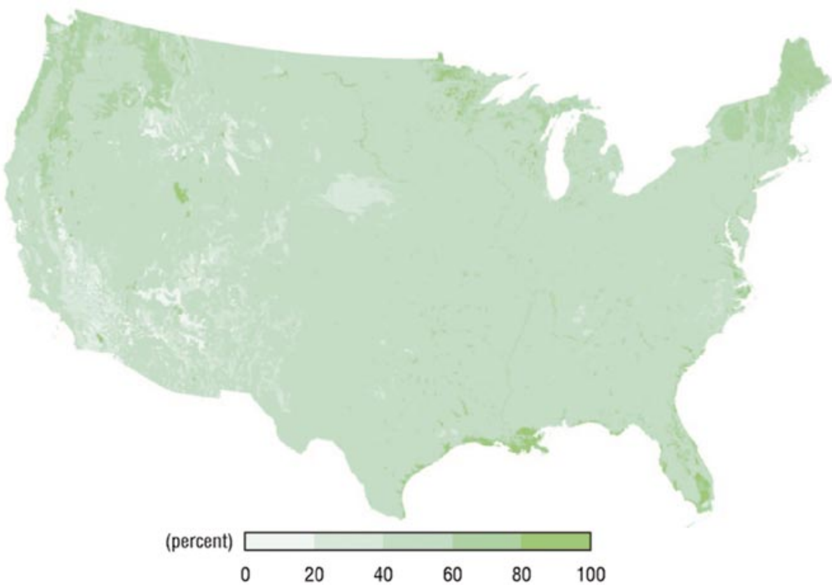

Figure 7. CONUS-SOIL porosity.

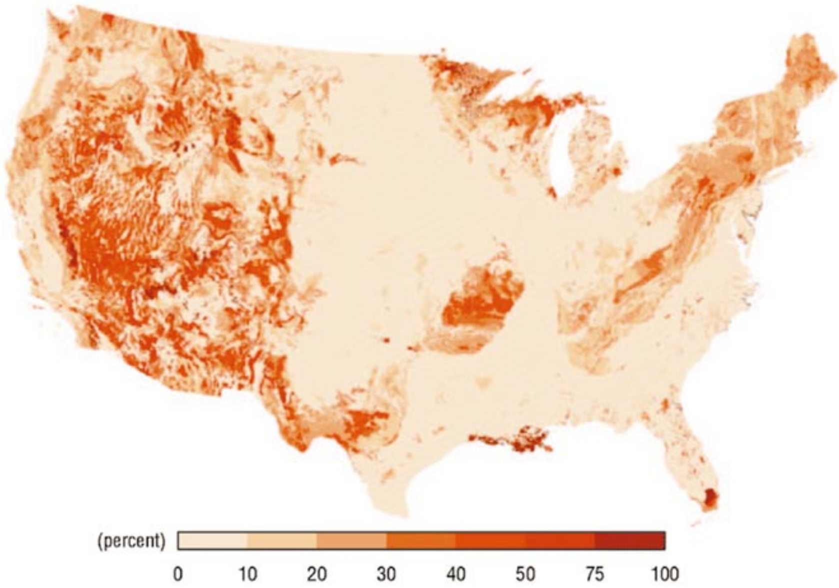

Figure 8. CONUS-SOIL rock fragment volume.

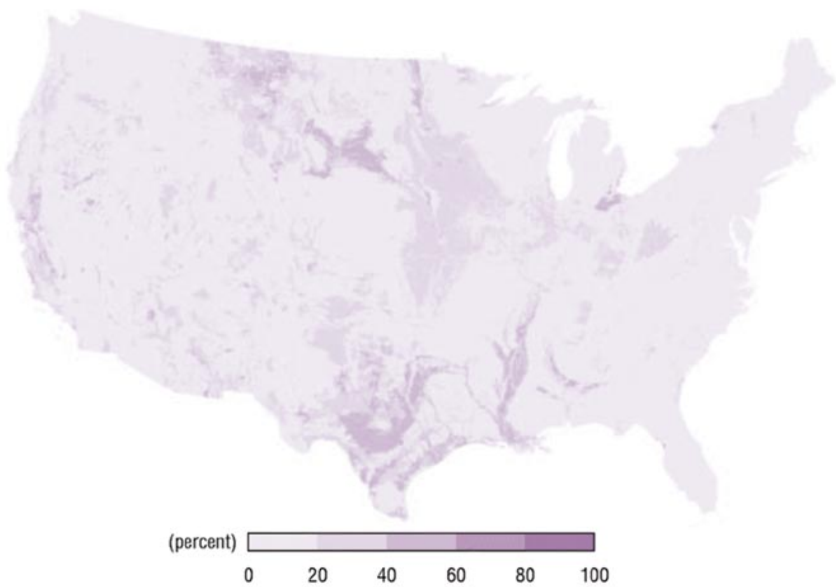

Figure 9a. CONUS-SOIL sand fraction.

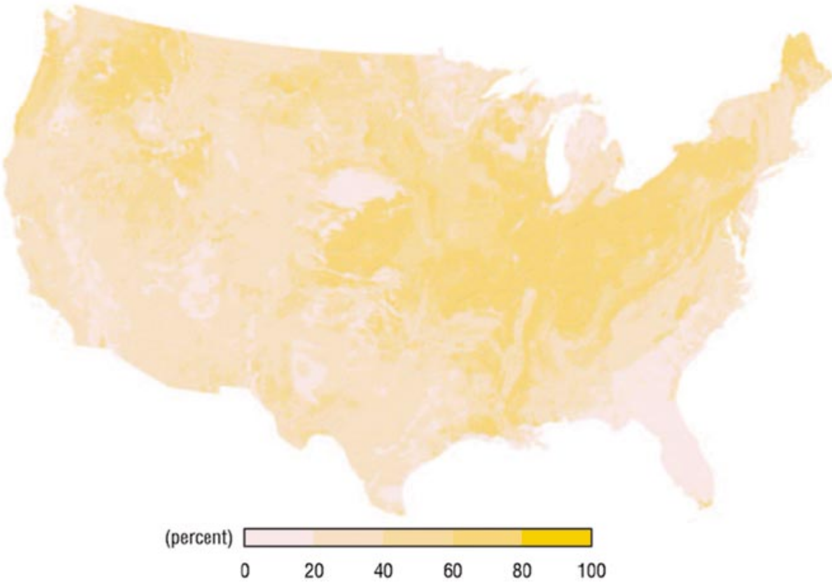

Figure 9b. CONUS-SOIL silt fraction.

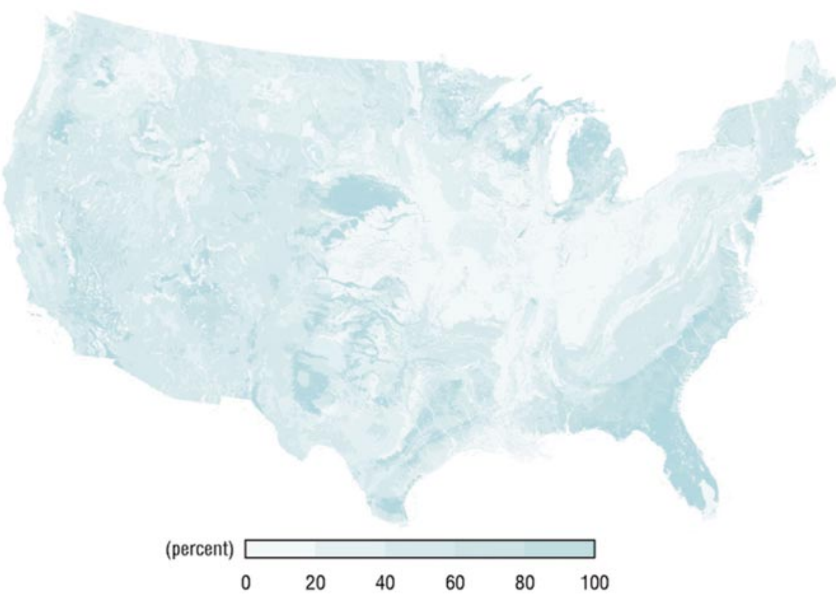

Figure 9c. CONUS-SOIL clay fraction.

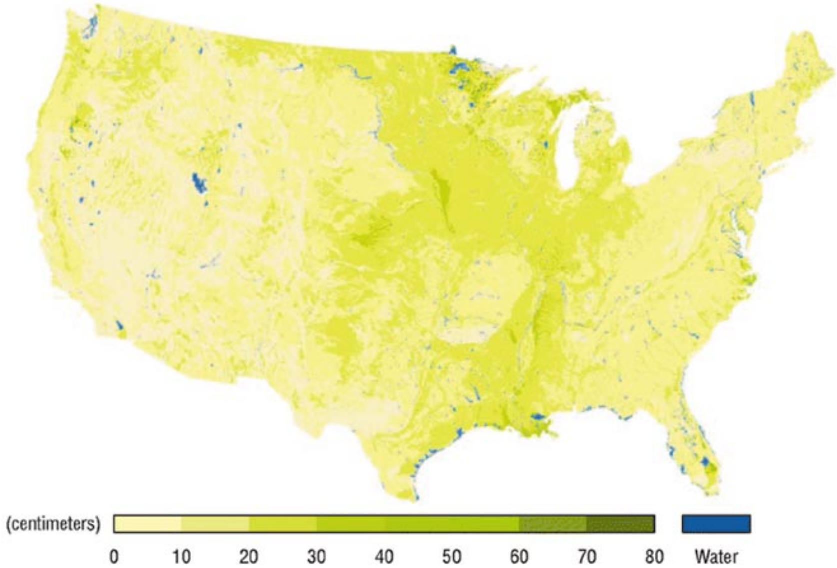

Figure 10. CONUS-SOIL available water capacity.

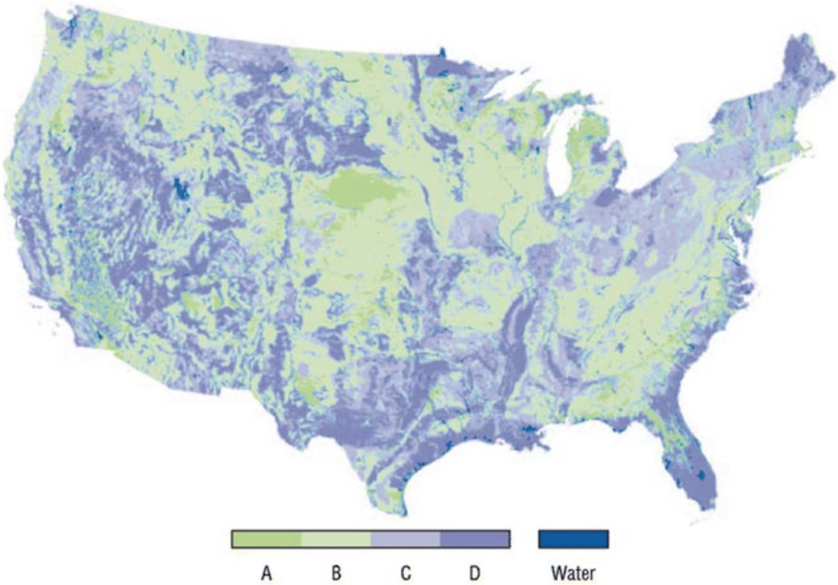

Figure 11. CONUS-SOIL hydrologic soil group.

and hydrology modeling applications in the conterminous United States. As CONUS-SOIL is a derivative of STATSGO, it inherently contains the same limitations and assumptions that are ‘‘built-in’’ to that product. To gain an overall understanding of the limitations of CONUS-SOIL requires that one consider the STATSGO data itself.

STATSGO was created by generalizing more detailed soil survey maps. Generalization often obscures the limitations inherent in detailed soil survey products. These surveys are created by physically sampling less than 1/1000th of the total area of the soil survey (Hudson, 1992). Soil map units are delineated on the basis of landscape features like topography, geology, and vegetation type. These units may be quite internally heterogeneous, with as much as 50% of the map unit having soil properties that differ significantly from the map unit description. The generalization process used to create small-scale (large area) maps from detailed survey information necessarily lumps much of this map unit variability into even more general map units. The within-unit variability is often larger than the between-unit variability for any particular STATSGO map unit (Lathrop et al., 1995). Lack of specific information about detailed map unit composition leads to even further uncertainty when generalization to regional scales is undertaken.

The guidelines for STATSGO development are given in the STATSGO data user’s guide. Each NRCS state office was responsible for creation of their state’s STATSGO map. Although the prescribed procedures are standard approaches taken in creating general soil maps from more detailed soil surveys, the data indicate an apparent lack of uniformity across states in the exact approaches used to create STATSGO. An informal survey of soil scientists responsible for STATSGO creation partially confirms this suspicion.

As noted under the description of the individual soil properties, the STATSGO data for individual map units are often incomplete or ambiguous. The most common problems encountered while processing the STATSGO data were the following:

1. The specification of 60 in. as the bedrock depth for somewhat more than 50% of all components. In most cases this indicated the maximum depth to which the soil profile was examined, whether or not bedrock was actually encountered.
2. The specification of a bulk density value of 0.0 for at least one layer of more than 25% of the nonwater components.
3. An apparent misinterpretation of the INCH3L/INCH3H variable for some layers. These variables were sometimes interpreted as being the weight fraction of all rock fragments larger than 3 in., rather than the fraction of rocks between 3 and 10 in., as evidenced by the occurrence of about 270 layers for which the sum of INCH3L and INCH10L exceeds 100%.

These uncertainties in STATSGO are propagated even further when soil physical and hydraulic properties are used in conjunction with pedo-transfer functions or directly as model input parameters. Quantification of variability in map unit properties during the soil survey process or systematic approaches using new observations in conjunction with completed soil surveys will be the only way to improve our understanding of the true limitations of soil survey information for modeling applications. Meanwhile, CONUS-SOIL represents a ‘‘best-available’’ regional- to continental-scale digital soil properties dataset for use in a wide range of environmental modeling applications.

CONUS-SOIL has been obtained by approximately 35 U.S. universities, 12 federal agencies, and nearly a dozen foreign investigators. These scientists are actively testing CONUS-SOIL in various modeling applications and early indications are that it is a useful and relevant dataset. The true value of CONUS-SOIL will be revealed as modelers perform systematic and quantitative comparisons of their model's performance with and without CONUS-SOIL.

6. Summary

The STATSGO database was used to develop a multilayer soil characteristics dataset for application in regional climate and hydrology models. Researchers have struggled for years with the lack of adequate soils information at scales that will support regional modeling of climatic and hydrologic processes. The development of this dataset is a first step in providing realistic and useful data about the physical properties of soils that can then be used with a range of empirical approaches for determining the subsequent hydraulic nature of the soil environment.

Future work will focus on extending the CONUS-SOIL dataset to a North American product that will include Canada and Mexico. Efforts will be made to improve the soil physical and hydraulic properties in CONUS-SOIL by using additional sources of soil survey and characterization information.

Acknowledgments. The authors gratefully acknowledge support from NASA, under Grant NAGW-2686 (Earth Observing System), and from the NOAA Office of Global Programs, under Contract NA56GP0239. Norman Bliss, Lou Steyaert, John Leese, John Schaake, Mercedes Lakhtakia, and Gary Petersen all provided helpful input in the development of the CONUS-SOIL database. Philip Kolb's graphic and cartographic expertise improved the CONUS-SOIL dataset and this paper.

References

Abramopoulos, F., C. Rosenzweig, and B. Choudhury. 1988. Improved ground hydrology calculations for global climate models (GCMs): Soil water movement and evapotranspiration. *J. Climate,* **1**, 921–941.

Arya, L. M. and J. F. Paris. 1981. A physicoempirical model to predict the soil moisture characteristic from particle-size distribution and bulk density data. *Soil Sci. Soc. Am. J,* **45**, 1023–1030.

Bliss, N. B., S. W. Waltman, and G. W. Petersen. 1995. Preparing a soil carbon inventory for the United States using geographic information systems. *Soils and Global Change,* R. Lal, J. Kimble, E. Levine, and B. A. Stewart, CRC Press, Boca Raton, FL, 275–295.

Brakensiek, D. L. and W. J. Rawls. 1994. Soil containing rock fragments: Effects on infiltration. *Catena,* **23**, 99–110.

Brooks, R. H. and A. T. Corey. 1964. Hydraulic properties of porous media. *Hydrology Paper 3,* Colorado State University, Fort Collins, CO, 27–110.

Buol, S. W., F. D. Hole, and R. J. McCracken. 1989. *Soil Genesis and Classification,* 3d ed., Iowa State University Press, Ames, IA, 446–110.

Campbell, G. S. 1974. A simple method for determining unsaturated conductivity from moisture retention data. *Soil Sci,* **117**, 311–314.

Clapp, R. B. and G. M. Hornberger. 1978. Empirical equations for some soil hydraulic properties. *Water Resour. Res,* **14**, 601–604.

Dickinson, R. E., A. Henderson-Sellers, and P. J. Kennedy. 1993. Biosphere-Atmosphere Transfer Scheme (BATS): Version 1 as coupled to the NCAR Community Climate Model. *NCAR/TN 387+STR,* Climate and Global Dynamics Division, National Center for Atmospheric ResearchBoulder, CO, 72–604.

Hillel, D. 1980a. *Fundamentals of Soil Physics,* Academic Press,New York, 410–604.

Hillel, D. 1980b. *Applications of Soil Physics,* Academic Press,New York, 385–604.

Hudson, B. 1992. The soil survey as paradigm-based science. *Soil Sci. Soc. Am. J,* **56**, 836–841.

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.

Lathrop, R. G., J. D. Abner, and J. A. Bognar. 1995. Spatial variability of digital soil maps and its impact on regional ecosystem modeling. *Ecol. Modell,* **82**, 1–10.

Loveland, T. R., J. W. Merchant, D. O. Ohlen, and J. F. Brown 1991. Development of a land-cover characteristics database for the conterminous United States. *Photogramm. Eng. Remote Sens,* **57**, 1453–1463.

Manabe, S. 1969. Climate and circulation. Part 1: The atmospheric circulation and hydrology of the Earth's surface. *Mon. Wea. Rev,* **97**, 739–774.

Rawls, W. J., T. J. Gish, and D. L. Brakensiek. 1991. Estimating soil water retention from soil physical properties and characteristics. *Adv. Soil Sci,* **16**, 213–234.

Rawls, W. J., L. R. Ahuja, D. L. Brakensiek, and A. Shirmohammadi. 1993. Infiltration and soil water movement. *Handbook of Hydrology,* D.R. Maidment, McGraw-Hill,New York, 5–51.

Reybold, W. U. and G. W. Tesselle. 1989. Soil geographic data bases. *J. Soil Water Conserv,* **43**, 226–229.

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

Soil Survey Staff. 1975. Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys. *U.S. Department of Agriculture Handbook No. 436,* U.S. Government Printing Office, Washington, DC, 754–1036.

Soil Survey Staff. 1993. Soil survey manual. *U.S. Department of Agriculture Handbook No. 18,* U.S. Government Printing Office, Washington, DC, 437–1036.

Soil Survey Staff. 1994a. State Soil Geographic Database (STATSGO) collection for the conterminous United States. USDA–NRCS, National Soil Survey Center, Lincoln, NE, CD-ROM.

Soil Survey Staff. 1994b. State Soil Geographic Database (STATSGO) data users guide. *USDA Natural Resources Conservation Service Misc. Publ. 1492,* U.S. Government Printing Office, Washington, DC, 88–1036.

Soil Survey Staff 1996. Soil Survey Laboratory methods manual. *Soil Survey Investigations Rep. 42, Version 2.0,* National Soil Survey Center, Soil Conservation Service, U.S. Department of Agriculture, Lincoln, NE, 693–1036.

Tietje, O. and M. Tapkenhinrichs 1993. Evaluation of pedo-transfer functions. *Soil Sci. Soc. Am. J,* **57**, 1088–1095.

van Genuchten, M. Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J,* **44**, 892–898.

Vereecken, H., J. Maes, and J. Feyen. 1990. Estimating unsaturated hydraulic conductivity from easily measured soil properties. *Soil Sci,* **149**, 1–12.

Wall, L. and R. Schwartz. 1991. Programming Perl. O'Reilly and Associates, Sebastopol, CA, 482–12.

Webb, R. S., C. E. Rosenzweig, and E. R. Levine. 1993. Specifying land surface characteristics in general circulation models: Soil profile data set and derived water-holding capacities. *Global Biogeo. Cyc*, **7**, 97–108.
White, R. A. and D. A. Miller. 1998. Processing soil survey information with Perl. *Comput. Geosci*, in press.
Wilson, M. F., A. Henderson-Sellers, R. E. Dickinson, and P. J. Kennedy. 1987. Sensitivity of the Biosphere–Atmosphere Transfer Scheme (BATS) to the inclusion of variable soil characteristics. *J. Clim. Appl. Meteorol*, **26**, 341–362.
Zheng, D., E. R. Hunt Jr., and S. W. Running. 1996. Comparison of available soil water capacity estimated from topography and soil series information. *Land. Ecol*, **11**, 3–14.
Zobler, L. 1986. A world soil file for global climate modeling. *NASA Tech. Memo. 87802*, 32–14.

Table 1. Soil water retention and hydraulic conductivity relationships (after Rawls et al. (1993)).

Hydraulic soil characteristic	Parameters	Parameter correspondence
Brooks and Corey		
Soil water retention	λ = pore-size index	$\lambda = \lambda$
$\frac{\theta-\theta_r}{\phi-\theta_r} = (h_b/h)^\lambda$	h_b = bubbling capillary pressure	$h_b = h_b$
Hydraulic conductivity	θ_r = residual water content	$\theta_r = \theta_r$
$K(\theta)/K_s = [(\theta-\theta_r)/(\phi-\theta_r)]^n = (S_e)^n$	ϕ = porosity	$\phi = \phi$
	K_s = fully saturated conductivity ($\theta=\phi$)	$K_s = K_s$
	$n = 3 + 2/\lambda$	
Campbell		
Soil water retention	ϕ = porosity	$\phi = \phi$
$\theta/\phi = (H_b/h)^{1/b}$	H_b = scaling parameter with dimension of length	$H_b = H_b$
Hydraulic conductivity	b = constant	$b = 1/\lambda$
$K(\theta)/K_s = (\theta/\phi)^n$	$n = 3 + 2b$	
van Genuchten		
Soil water retention	ϕ = porosity	$\phi = \phi$
$\frac{\theta-\theta_r}{\phi-\theta_r} = [1+(\alpha h)^n]^{-m}$	θ_r = residual water	$\theta_r = \theta_r$
Hydraulic conductivity	α = constant	$\alpha = (h_b)^{-1}$
$\frac{K(\theta)}{K_s} = \left(\frac{\theta-\theta_r}{\phi-\theta_r}\right)^{1/2}\left\{1-\left[1-\left(\frac{\theta-\theta_r}{\phi-\theta_r}\right)^{1/m}\right]^m\right\}^2$	n = constant	$n = \lambda + 1$
	m = constant	$n = \lambda/(\lambda + 1)$

θ = water content
h = capillary suction, cm
$K(\theta)$ = hydraulic conductivity for given water content, cm h^{-1}

Table 2. STATSGO database variables used to develop the multilayer soil characteristics dataset.

STATSGO variable name	Table location	Definition
AWCH	Layer	Maximum value for the range of available water capacity for the soil layer
AWCL	Layer	Minimum value for the range of available water capacity for the soil layer
BDH	Layer	Maximum value for the range in moist bulk density of the soil layer
BDL	Layer	Minimum value for the range in moist bulk density of the soil layer
COMPPCT	Comp	The percentage of a map unit component
HYDGRP	Comp	The hydrologic group for the soil
INCH10H	Layer	The maximum value for the range in percent by weight of the rock fragments greater than 10 in. in size in the soil layer
INCH10L	Layer	The minimum value for the range in percent by weight of the rock fragments greater than 10 in. in size in the soil layer
INCH3H	Layer	The maximum value for the range in percent by weight of the rock fragments 3–10 in. in size in the soil layer
INCH3L	Layer	The minimum value for the range in percent by weight of the rock fragments 3–10 in. in size in the soil layer
LAYDEPH	Layer	The depth to the lower boundary of the soil layer or horizon
LAYDEPL	Layer	The depth to the upper boundary of the soil layer or horizon
MUID	Map unit, Comp, Layer	Map unit identifyer
NO10H	Layer	The maximum value for the range in percent by weight of the soil material in a layer or horizon that is less than 3 in. in size and passes a No. 10 sieve
NO10L	Layer	The minimum value for the range in percent by weight of the soil material in a layer or horizon that is less than 3 in. in size and passes a No. 10 sieve
NO200H	Layer	The maximum value for the range in percent by weight of the soil material in a layer or horizon that is less than 3 in. in size and passes a No. 200 sieve
NO200L	Layer	The minimum value for the range in percent by weight of the soil material in a layer or horizon that is less than 3 in. in size and passes a No. 200 sieve
ROCKDEPH	Comp	The maximum value for the range in depth-to-bedrock
ROCKDEPL	Comp	The minimum value for the range in depth-to-bedrock
SEQNUM	Comp, Layer	Component sequence number
SURFTEX	Comp	Surface soil texture
TEXTURE1	Layer	Dominant soil texture class

Table 3. Frequency of occurrence of layers in the STATSGO database.

Thickness range (in.)	Top layer	All layers
1–3	9%	4%
4	7%	10%
5–10	53%	29%
11–20	22%	25%
>20	8%	32%

Table 4. Standard data layers created for the STATSGO datasets.

Standard layer	Thickness (cm)	Depth to top of layer (cm)	Depth to bottom of layer (cm)
1	5	0	5
2	5	5	10
3	10	10	20
4	10	20	30
5	10	30	40
6	20	40	60
7	20	60	80
8	20	80	100
9	50	100	150
10	50	150	200
11	50	200	250

Table 5. Soil texture classes in CONUS-SOIL.

Class no.	Soil texture class	Class abbreviation
1	Sand	S
2	Loamy sand	LS
3	Sandy loam	SL
4	Silt loam	SiL
5	Silt	Si
6	Loam	L
7	Sandy clay loam	SCL
8	Silty clay loam	SiCL
9	Clay loam	CL
10	Sandy clay	SC
11	Silty clay	SiC
12	Clay	C
13	Organic materials	OM
14	Water	W
15	Bedrock	BR
16	Other	O

Table 6. Rock fragment classes in CONUS-SOIL.

Class no.	Rock texture class	Class abbreviation
1	Bouldery	BY
2	Cobbly	CB
3	Channery	CN
4	Cherty	CR
5	Flaggy	FL
6	Gravelly	GR
7	Rubbly	RB
8	Shaly	SH
9	Stony	ST
10	Slaty	SY
11	Organic materials	OM
12	Water	W
13	Bedrock	BR
14	Other	O

Table 7. CONUS-SOIL texture classes and associated sand, silt, and clay fractions.

Class no.	Soil texture class	Class abbreviation	% sand	% silt	% clay
1	Sand	S	92	5	3
2	Loamy sand	LS	82	12	6
3	Sandy loam	SL	58	32	10
4	Silty loam	SiL	17	70	13
5	Silt	Si	10	85	5
6	Loam	L	43	39	18
7	Sandy clay loam	SCL	58	15	27
8	Silty clay loam	SiCL	10	56	34
9	Clay loam	CL	32	34	34
10	Sandy clay	SC	52	6	42
11	Silty clay	SiC	6	47	47
12	Clay	C	22	20	58
13	Organic materials	OM	0	0	0
14	Water	W	0	0	0
15	Bedrock	BR	0	0	0
16	Other	O	0	0	0

Table 8. USDA–NRCS hydrologic soil groups (from Soil Survey Staff (1993)).

Classification	Type of soil
A (low runoff potential)	Soils with high infiltration capacities, even when thoroughly wetted; chiefly sands and gravels, deep and well drained
B	Soils with moderate infiltration rates when thoroughly wetted; moderately deep to deep, moderately well to well drained, with moderately fine to moderately coarse textures
C	Soils with slow infiltration rates when thoroughly wetted; usually have a layer that impedes vertical drainage, or have a moderately fine to fine texture
D (high runoff potential)	Soils with very slow infiltration rates when thoroughly wetted-iefly clays with a high swelling potential; soils with a high permanent water table; soils with a clay layer at or near the surface; shallow soils over nearly impervious materials

Table 9. CONUS-SOIL data formats.

Formats	Lambert azimuthal	Albers equal area	Latitude–longitude
Arc/Info polygon	●	●	●
Arc/Info grid	●	●	●
Binary array grid	●	●	●