International Journal of Geospatial and Environmental Research

Volume 1 | Number 2 Article 6

March 2015

A Soil Parameters Geodatabase for the Modeling Assessment of Agricultural Conservation Practices Effects in the United States

Mauro Di Luzio
Blackland Research Center, Texas A&M AgriLife Research, madilu@sbcglobal.net

Martin L. Norfleet
Natural Resource Conservation Service, United States Department of Agriculture

Jeffrey G. Arnold
Agriculture Research Service, United States Department of Agriculture

Jimmy R. Williams
Blackland Research Center, Texas A&M AgriLife Research

James R. Kiniry
Agriculture Research Service, United States Department of Agriculture

Recommended Citation
Di Luzio, Mauro; Norfleet, Martin L.; Arnold, Jeffrey G.; Williams, Jimmy R.; and Kiniry, James R. (2015) "A Soil Parameters Geodatabase for the Modeling Assessment of Agricultural Conservation Practices Effects in the United States," *International Journal of Geospatial and Environmental Research*: Vol. 1 : No. 2 , Article 6.
Available at: https://dc.uwm.edu/ijger/vol1/iss2/6

A Soil Parameters Geodatabase for the Modeling Assessment of Agricultural Conservation Practices Effects in the United States

Abstract
Soil parameters for hydrology modeling in cropland dominated areas, from the regional to local scale, are part of critical biophysical information whose deficiency may increase the uncertainty of simulated conservation effects and predicting potential. Despite this importance, soil physical and hydraulic parameters lack common, wide-coverage repositories combined to digital maps as required by various hydrology-based agricultural water quality models.

This paper describes the construction of a geoprocessing workflow and the resultant hydrology-structured soil hydraulic, physical, and chemical parameters geographic database for the entire United States, named US-SOILM-CEAP. This database is designed to store a-priori values for a suit of models, such as SWAT (Soil and Water Assessment Tool), APEX (Agricultural Policy Environmental EXtender) and ALMANAC (Agricultural Land Management Alternatives with Numerical Assessment Criteria), which are commonly used for the across scale assessment of agricultural hydrology and conservation practice scenarios. The Soil Survey Geographic (SSURGO) database developed by the U.S. Department of Agriculture provided the main source data for this development. Additional spatial information, a geographic information system platform and Python computer programming language code were used to create hydrology-based tile coverage of the areal soil units linked to the specific and detailed attributes required by each model.

The created repository adds value to the source soil survey data, while maintaining and extending the detailed information necessary for the across scale and combined application of the models. Ultimately, our multi-model database provides a comprehensive product achieving joined informational-mapping-geoprocessing functionality with the explicit maintenance of the original conceptual links between soil series and composing soil layers, allowing for efficient data retrieval, analysis and service as input for modeling conservation effects.

Keywords
Geodatabase, Soil, Crop Model, Hydrologic Model, SSURGO, SWAT, APEX, ALMANAC

Acknowledgements
This research was supported by the United States Department of Agriculture - Natural Resource Conservation Service (USDA - NRCS) Conservation Effect Assessment Project (CEAP).

1. INTRODUCTION

Soil parameters are fundamental input for hydrology-based model simulations in cropland areas and ultimately for the assessment of the effects of conservation practices and agricultural nonpoint pollution. These models, hereafter referred to as CAGNPMs, rely comprehensively on hydraulic, physical and chemical attributes of the soil to establish essential soil-water-plant-nutrients-management relationships across the agricultural landscape. It is well recognized that proper application of CAGNPMs and other similar models across the world suffer from the lack of reliable and consistent soil data information. Historically, unavailable or roughly estimated model parameters have considerably biased land degradation assessments, environmental impact studies and associated planning for sustainable land management interventions (Batjes et al. 2007). Several studies have reported the sensitivity of CAGNPMs' simulations to soil input information, scale, and resolution (Di Luzio et al. 2005; Geza and McCray 2008; Mednick 2010; Perez–Quezada et al. 2003; Peschel et al. 2006; Singh et al. 2011; Wang and Melesse 2006; Xie et al. 2003). Such importance is implicit for site-specific and field applications, while it is often inadequately explained and/or undermined in the case of watershed and basin-scale assessments (Geza and McCray 2008; Kumar and Merwade 2009; Mednick 2010; Moriasi and Starks 2010; Peschel et al. 2006). This is most likely a consequence of the partial understanding of the soil-hydrological functions at these scales combined with the equifinality effect (Beven 2006), by which comparable simulation statistics are obtained at gaged outlets despite the usage of different scales of soil parameter sets. Nevertheless, quality soil parameter data sets are expected to provide spatial consistency and realistic representation of hydrologic and agronomic processes in order to address proper land management and the assessment of conservation practices (Delgado et al. 2011).

Traditional hydrologic modeling at experimental plot/field scale relies on quality detailed soil profile data, pedotransfer rules and other analytical tools to derive soil properties, which are then translated into model parameters. Moving from local to larger scale, soil sampling, remote sensing and other evolving spatial and non-spatial techniques are additionally applied to identify and quantify soil patterns, thereby extending the coverage of local knowledge. Over large domains and across the world, a viable method for deriving soil parameters for hydrology modeling remains to be the usage of soil series information identified as soil taxonomic units mapped by expert surveyors with the support of aerial photography (Grunwald 2009). In the United States, a number of taxonomy-based nationwide digital spatial databases are developed by the Department of Agriculture-Natural Resources Conservation Service (USDA-NRCS), formerly the Soil Conservation Service (SCS), which collects, stores, maintains, and distributes soil survey information for privately owned lands in the United States. These publicly available databases provide a source of spatially-referenced soil information, which offers a cost-effective method for accounting soil parameters for hydrology modeling. The most detailed database is the Soil Survey Geographic Database (USDA-NRCS 2012).

A number of similar approaches were developed to derive a-priori parameters starting from SSURGO and ending with the application of a variety of hydrology models and CAGNPMs over watershed areas. Commonly, these approaches are specifically designed to create the required parameter dataset for target application watersheds (e.g. Di Luzio et al. 2004; Frankenberger et al. 2011; Sheshukov et al. 2011). This is partially due to the source data complexity and the challenging, time-consuming effort required to process considerable volumes of data for larger landscape scales than the single soil survey domain. The lack of adequate database frameworks, capable of efficiently storing and combining digital soil and other maps and various levels of soil numerical attributes has been a considerable limiting factor that has impeded the development of repositories of the a-priori parameters required by CAGNPMs models. Altogether, these aspects have prevented CAGNPMs' users from having access to readily available, up-to-date and common soil-derived parameters across large geographic domains.

In the recent past, similar efforts were embraced for large scale non-CAGNPMs models. USSOILS (Schwarz and Alexander 1995) was originally compiled to support a national model of water quality. USSOILS aggregates the STATSGO (USDA-SCS 1994) layer and component information up to the level of a soil unit by depth-averaging, over the entire soil column, median properties within a component and then area-averaging component values across a soil unit. CONUS-SOIL (Miller and White 1998) was developed using STATSGO to establish a data set of combined map coverage and associated soil characteristics for application in regional climate and hydrology models (e.g. soil-vegetation-atmosphere transfer schemes, SVATS) in the lower 48 states. Soil properties estimated at six depth increments at the nominal scale of the data is 1:250,000. The National Weather Service (NWS) derived a-priori parameters for the spatially lumped Sacramento Soil Moisture Accounting (SAC-SMA) model using STATSGO, and recently (Zhang et al. 2011) updated it using SSURGO databases with the support of the National Land Cover Dataset (Homer et al. 2004). A set of 25-state a-priori parameter grids resulting from the aggregation of parameter values compose the resulting database. To address different aspects of cropping systems, Yang et al. (2011) developed a system to automatically store SSURGO attributes and generate generic soil attribute maps for different soil properties and at different soil depths with display scales at national, state, and county levels.

This paper describes the development of a geospatial database containing soil parameters pertinent to a target set of CAGNPMs (SWAT, APEX, and ALMANAC) and the entire United States by adding value to the source SSURGO database. As described in the sections below, our approach provides specific benefits over approaches embraced in the past that address current and future requirements of the CAGNPMs' user community. In the first section we introduce source data, models, and GIS elements used in the work. In the following section we present the results, and in the final section we discuss some highlights.

2. MATERIALS AND METHODS

2.1 SOIL GEOGRAPHIC DATABASE

The USDA-NRCS develops soil geographic databases at three scales: local, regional and national. The SSURGO database provides the local, most detailed level of soil geographic data collected over several decades and developed through the National Cooperative Soil Survey (NCSS) in accordance with its mapping standards. SSURGO is built at a range of scales between 1:12,000 and 1:24,000 as defined by the underlining orthophotos used as source maps to identify repeatable patterns of soil inventory on the landscape.

SSURGO contains geo-referenced digital map data and associated digital tables of attributes data. Digital maps are sets of areas, namely Map Units (MUs), each one delineated and uniquely identified on a soil map based on properties, interpretations, and productivity. The area of the minimum size map unit delineated ranges between 1 to 10 acres. Each MU may consist of one or more non-contiguous polygons and may include one to three major non-georeferenced constituents, namely Components (COMPs), which identify phases of soil series as well as minor inclusions. Inclusions represent soil areas too small to be geographically identified separately. A soil series is the lowest level in the U.S. system of taxonomy (Soil Survey Staff 1998) and the most homogeneous with regard to factors that distinguish soil management, such as vertical profile similar in arrangement and in differentiating characteristics. In the database, each component-soil series is accounted as percentage of the MU area and includes specific characteristics across vertical levels of the profile, namely Horizons (Hs), at which each one is conventionally segmented. The MUs are linked to information about the composing soils and their properties derived from characteristics stored in the National Soil Information System (NASIS). Information is deployed as a set of tables grouping thematic items, linkable mutually and ultimately to the provided digital maps using designed string IDs. Metadata explain the relationships of tables, table fields and their respective data type, and detailed column descriptions.

At the time of this work, comprehensive composing parts of the SSURGO database were available to be acquired at the level of single Soil Survey Area (SSA) package. Each SSA geographically comprises either a single county, multiple counties or parts of multiple counties. Each SSA-based digital map includes mainly all the MU polygons falling within the represented area and other minor spatial entities. The associated information in tabular format counts sixty-eight (68) tables. The Map Unit, Component, and Horizon table provide the most relevant data items for this application. The sub-set of tabular data, seven (7) tables, used in this work and the respective definitions, are provided in Table 1. Within Figure 1 (left side) a schematic diagram depicts main elements and links of the SSURGO database features contained within each SSA package used in this work. Figure 1 highlights the conceptual and operational relationships among the items part of the database, such as: a) any geographic area of interest (e.g. SSA), includes one or more geographically outlined MUs (1-to-many); b) each MU may include one or more (maximum three) non-georeferenced major components (1-to-many) and other inclusions; c) each component/soil series is characterized by one or more vertically outlined horizons (1-to-many); and d) each horizon may include volume constituents of the horizon above 2 mm, Rock Fragments (RFs), within a set of approximately two hundred (191) distinguished types.

Table 1. List and description of the seven SSURGO tables used in this work.

Table	Short Description (USDA-NRCS, 2012)	Level of Detail
LEGEND	Provides the soil survey area that the legend is related to, and related information.	Soil Survey Area
MAPUNIT	Map units included in the referenced legend and data related to the map unit as a whole.	Map Unit
COMPONENT	Map unit components identified in the referenced map unit, and selected properties of each component.	Component
CHORIZON	Horizon(s) and related data for the referenced map unit component.	Horizon
CHFRAGS	Mineral and organic fragments that generally occur in the referenced horizon.	Horizon Fragment
CHTEXTUREGRP	Range of textures for the referenced horizon as a concatenation of horizon texture and texture modifier(s).	Horizon
MUAGGATT	Variety of soil attributes and interpretations that have been aggregated from the component level to a single value at the map unit level to express a consolidated value or interpretation for the map unit as a whole.	Map Unit

More than three-thousand (3,151) complete (including both digital map and tabular data) SSA packages were acquired via ftp on November-December 2012 from the following USDA-NRCS web sites: a) the Geospatial Data Gateway (http://datagateway.nrcs.usda.gov/); and b) the Soil Data Mart (http://soildatamart.nrcs.usda.gov/). The geographic coverage of the acquired data included the Conterminous United States (CONUS), Alaska and the Territories, Commonwealths, and Island Nations. The missing areas are either under completion or are lands not served by the USDA-NRCS (Figure 3). The CONUS included 3,098 SSAs with complete data. The MU boundaries in geographic coordinates at the North American Datum of 1983 (NAD83) were provided in ESRI shapefile format (Environmental Systems Research Institute, 1998). The tabular data were delivered as pipe (|) delimited ASCII files, which are easily loadable into Microsoft Access using a provided template and/or into other software. In SSURGO, a large volume of numerical parameters are generally provided as minimum (L), representative (R), and maximum (H).

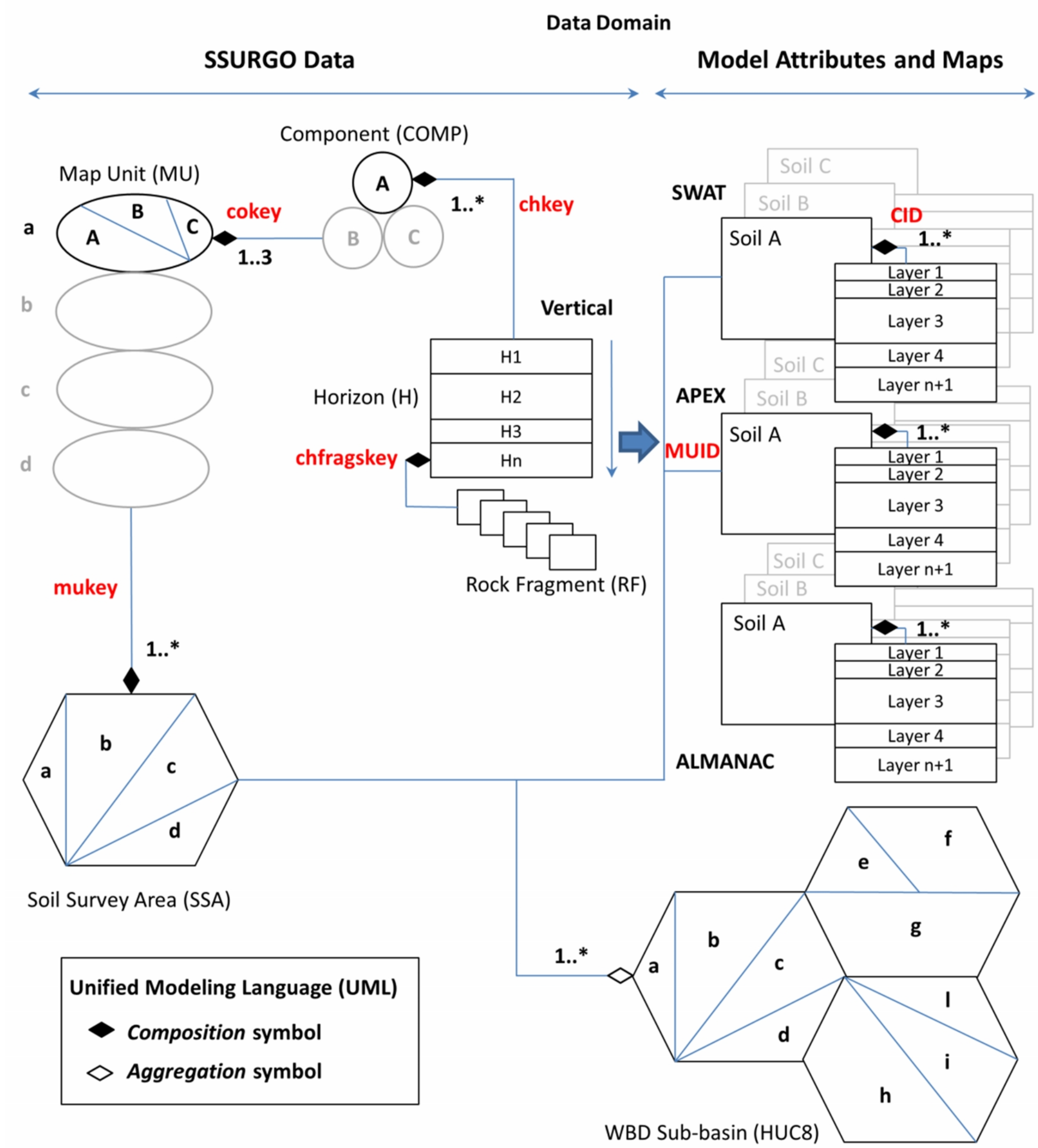

Figure 1. Structure of the essential information in the source SSURGO/NASIS database and of the derived geodatabase. Essential keys of both structures are highlighted in red.

2.2 Models

Target CAGNPMs selected for this work provide assessment solution at the river/watershed through field and plant community level. They are the result of a long history of research and model development at Temple, Texas (Williams et al. 2008) and are, among other tasks, an integral component of the developing national USDA NRCS

Conservation Effect Assessment Project (CEAP) (Duriancik et al. 2008; Mausbach and Dedrick 2004). CEAP was established to estimate conservation benefits at the national, regional and watershed levels, and to provide research and assessment on optimized conservation practices for managing agricultural landscapes, with the overarching goal of environmental quality protection and enhancement.

The SWAT model (Arnold and Fohrer 2005) is a river basin and watershed hydrology scale model operating on a daily time-step to simulate water, sediment, nutrient, pesticide and fecal bacteria yields in agriculture-dominated landscapes and draining channels. SWAT targets and requires the segmentation of the watershed into a number of sub-watersheds based on topography and into hydrologic response units (HRUs), which are unique combinations of soil, land use, slope, and land management. The channel and reservoir component routes the flows, sediment, nutrients, pesticides, and bacteria. SWAT includes components for the simulation of typical agricultural structures and practices, such as tillage, fertilizer and manure application, subsurface drainage, irrigation, ponds and wetlands, as well as edge-of-field buffers.

The APEX model (Williams et al. 2006), evolution of the EPIC model (Williams et al. 1989), is a field- and farm-scale, daily time-step model that simulates all of the basic biological, chemical, hydrological and meteorological processes of farming systems and their interactions. Soil erosion is simulated over time, including wind erosion, sheet and rill erosion, and the loss of sediment beyond the edge of the field. The nitrogen, phosphorus and carbon cycles are simulated, including chemical transformations in the soil that affect their availability for plant growth or for transport from the field. Exchange of gaseous forms between the soil and the atmosphere is also simulated, including losses of gaseous nitrogenous compounds.

ALMANAC (Kiniry et al. 1992) is a field-scale model specialized in simulating the crop growth of a wide range of plant species and their competition. The model provides the simulation of processes such as soil water balance, nutrient cycling and plant-growth in conjunction to peculiar functions for the interception by leaves, dry matter production and partitioning of biomass into grain. ALMANAC, with its additional capability of accurately simulating long-term mean crop yields for diverse environments and in extreme climatic conditions, is valuable for risk assessment and management evaluation.

At their highest level of detail these models lump the areal soil information and conceptualize it in a one-dimensional representation of the profile. Therefore, soil parameters are partitioned in two categories:

a) The first includes parameters, which for their nature characterize the soil as a whole or else aggregated unit. A set of single parameter values represent it. Examples of parameters in this category include the SCS (now NRCS) soil hydrologic group, number of layers. etc.
b) The second category includes parameters characterizing layers of the vertical profile. Examples of parameters in this category include: bulk density, available water capacity, saturated hydraulic conductivity, sand, silt, clay, soil pH, organic carbon, rock fragment content, etc.

The right side of Figure 1 summarizes this concept, which prolongs the conceptual and informational relationship of the SSURGO database described above. The three models include a number of common parameters, frequently with distinct naming and/or units of measure. SWAT, APEX and ALMANAC employ sixteen (16), sixty (60), and fort-two (42) soil parameters, respectively.

2.3 Geodatabase and Python

The File Geodatabase (FGDB), one of the most recent ESRI ArcGIS Desktop database options, was adopted for this work (ESRI 2009). A FGDB can host collections of geographic datasets of various types held in a common file system. The three primary dataset types are: Tables (collection of rows sharing the same fields); Feature classes (a table with a field for the geographic feature such as point, line, or polygon); and Raster datasets (pixel-based representing continuous geographic phenomena). FGDB is basically a collection of various types of GIS datasets held in a file system folder instead of a relational database. The creation of a FGDB starts by building a number of these fundamental dataset types, which can be enriched with more advanced capabilities (such as by adding topologies, networks, or subtypes) to model GIS behavior, maintain data integrity, and work with an important set of spatial relationships. Other key factors for this work are: a) scalability to handle very large data sets (1-256 TB); b) data structure optimized for performance and storage. FGDBs outperform and use about one-third of the feature geometry storage required by shape files and personal geodatabases. File geodatabases also allow users to compress vector data to a read-only format to reduce storage requirements even further.

Python is a free, open source, cross platform, interpreted, object oriented programming language. It was created and released in 1991 (van Rossum 2013) after which a community of contributors joined to develop the subsequent versions. Python version 2.7 was used in this work. This scripting language is distributed with ArcGIS 10.1 and allows accessing and extending of its basic functionality, automating workflows and application development. These features are provided by ArcPy, a package companion of ArcGIS, which introduces numerous modules, classes, and functions for the integration with Python. Our work leverages on this capability and the creation of code grouped in batch processes to access and operate the built-in geoprocessing routines and other tools offered by ArcGIS along with its Spatial Analyst extension (ESRI 2013). Spatial Analyst is the software package that provides raster analysis and management functionality to ArcGIS.

2.4 National Watershed Boundary Dataset (WBD)

In the United States, science-based hydrologic principles have led to the establishment of a base-line drainage boundary framework of nested hydrologic units (HU), accounting for all land and surface areas in the country. HU boundaries for the United States have been mapped by the U.S. Geological Survey (USGS) at a scale of 1:250,000 using a 4-level hierarchical Hydrologic Unit Code (HUC). Natural Resources Conservation Service (NRCS) has enhanced the USGS datasets following the National Interagency Guidelines in a GIS format at a minimum scale of 1:24,000. This effort has recently culminated in the National Watershed Boundary Dataset (WBD), new topographically-based hydrologic unit boundaries coincident to and computationally integrated with both the National Hydrography Dataset (NHD) and the National Elevation Dataset (NED). WBD includes 22 regions (identified by 2-digit numbers). Regions 01-18 compose the CONUS area, while Alaska (19), Hawaii (20), Caribbean (21), and South Pacific Islands (22) are the remainder. WBD embeds 222 sub-regions (identified by 4-digit numbers), 386 basins (formerly named Accounting Unit, identified by 6-digit numbers), and 2,297 sub-basins (formerly Cataloging Units, identified by 8-digit numbers), which are all based on surface hydrologic features. The average sub-basin, or 8-digit HUC, area is approximately four-thousand-five-hundred (4,543) square kilometers, the maximum area is eighty-four-thousand (8,455) square kilometers and the smallest is approximately one-hundred (115) square kilometers. Newly introduced levels, namely watershed (5th level, 10-digit) and sub-watershed (6th level, 12-digit), were not used in our work. The WBD GIS coverage was obtained from the NRCS Geospatial Data Gateway on November 2012 and represents the near-complete version of the data set.

2.5 Crop Data Layer, Cultivated Layer and National Land Cover Data Set

A land cover product for the United States is currently generated annually by the USDA's National Agricultural Statistics Service (NASS), namely the Cropland Data Layer (CDL). The CDL is typically a more than one-hundred (133)-class, 30 m resolution raster-based grid spanning the CONUS, with agricultural cover types as the focus and in less detail those non-agricultural (Boryan et al. 2011; Johnson and Mueller 2010). The data were produced using satellite imagery from the Landsat 5 TM sensor, Landsat 7 ETM+ sensor, and the Disaster Monitoring Constellation (DMC) DEIMOS-1 and UK2 sensors collected during the current growing season along with agricultural training data from the Farm Service Agency (FSA). NASS used a multi-year (2007-2011) complete CDL coverage for the CONUS and different sets of rules (models) for merging pixels to develop a new national Cultivated Layer (CL) (Boryan et al. 2012). The resulting geospatial data set represents the most recent, highest resolution and most accurate characterization of U.S. cultivation available. Precisely, the dataset distinguishes at 30 meter resolution the cultivated from non-cultivated land. These data sets were obtained from the NASS data server http://nassgeodata.gmu.edu/CropScape/.

The National Land Cover Data Set (NLCD) for the year 2001 (Homer et al. 2004) is a 16-class (additional four classes are used only in Alaska) land cover classification at a spatial resolution of 30 meters, and is based primarily on the unsupervised classification of Landsat Enhanced Thematic Mapper Plus (ETM+) circa 2001 satellite data. NLCD was obtained from the Multi-Resolution Land Characteristics Consortium (MRLC) at http://www.mrlc.gov/ to characterize areas outside the CONUS, such as region 19-21.

3. RESULTS

We designed and implemented a largely automated methodology for the construction of the geodatabase objective of this work. Figure 2 shows an outline of the methodology, which includes both GIS and standard database processing techniques to bring together maps (SSURGO MUs) and derived model parameters for each sub-basin (HUC8) of each region defined by the WBD data set in the U.S. For a particular sub-basin of concern, the involved MUs were obtained using GIS functions, while the associated tabular data were converted into an intermediate database containing the "raw" source parameters associated to them. Proper transfer functions were applied to derive model-specific parameters from the intermediate database. These include the usage of CDL/CL/NLCD data to obtain MU-specific land use data and to generate land use-biased parameters. The final step includes the design of the map and model parameter schema and the creation and assemblage of the target FGDBs.

The overall result of the implementation comprises 22-region (2-digit HUC)-wide FGDBs. We assembled each FGDB by merging the SSURGO polygons and the derived model parameters into composing 8-digit-wide tiles.

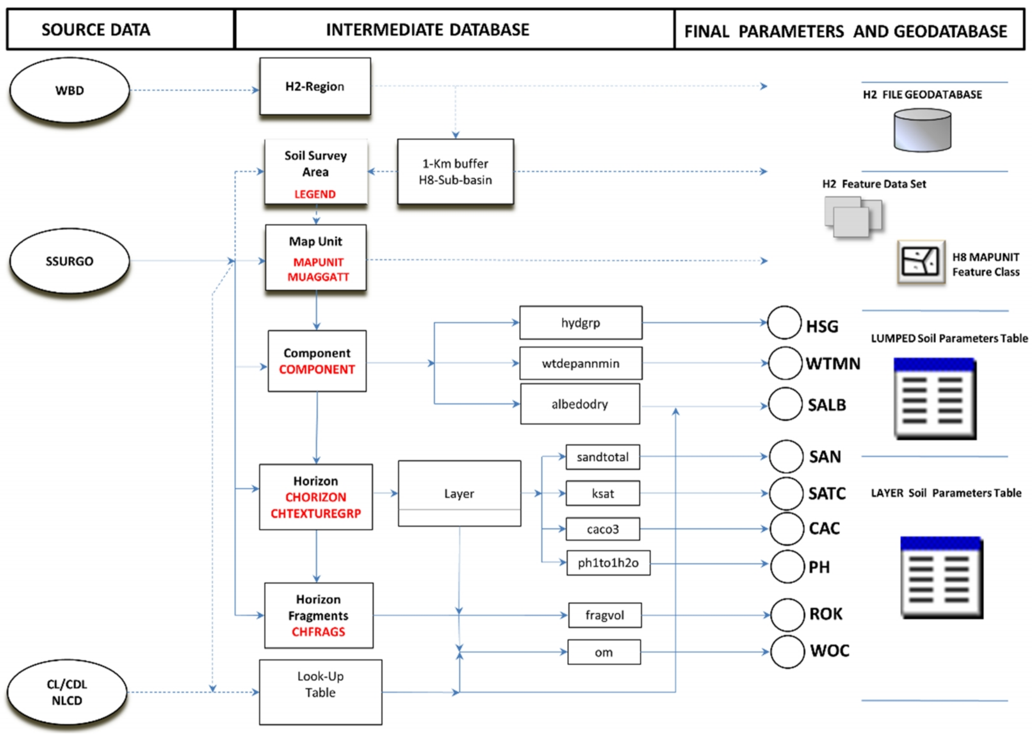

Figure 2. Flowchart of the geodatabase construction (spatial features and APEX model subset) from the input sources. Input SSURGO tables are highlighted in red. Shaded areas indicate inclusion of GIS components. A solid line indicates database processing. A dashed line indicates GIS processing. See the Acronyms section for meaning of each APEX parameter of the shown sub-set.

More precisely, each tile is outlined by a 1-km buffered 8-digit polygon. Each FGDB is constructed using two alternative modalities to represent each 8-digit tile, such as:

1) The geographic part is a Feature Class (ArcGIS format for vector data) resembling the pertinent MU vector polygons stored with the same coordinate system (Geographic NAD83) of the acquired SSA-based packages. The model attributes are then stored using distinct FGDB tables for each model.
2) A GeoTIFF file (an open source format for raster data) was obtained by a raster conversion of the same MU polygons, using the same 30-meter cell size, alignment, and USGS Albers Equal Area (AEA) projection coordinate system of the CDL/CL/NLCD data sets (a distinguished coordinate system was used for Alaska, Hawaii, and South Pacific regions). The model attributes were stored using dBASE tables.

In both cases, polygons and raster cell groups carry the same newly created unique integer Map Unit Identifier (MUID), which distinguishes each MU and provides a link between the respective polygons and/or the raster cell zones and the derived Lumped model parameters tables. These tables are also linked to the Layer model attributes using the newly constructed unique integer Component Identifier (CID). For the first modality, we built actual links (relationships) as described below.

We implemented the methodology by building Python subroutines that execute the following tasks to:

a) Extract and store the geographic and tabular information from each SSA-based package. This package includes the respective digital map polygons in ESRI shapefile format and the seven (7) selected SSURGO tables described in Table 1 and hierarchically located within the workflow depicted in Figure 2.
b) Build map component of the tiles. For each region, this task established the areal portion of each SSA overlapping the respective 8-digit buffered polygon, clipped the intersected MU polygons, merged into the respective tile Feature Class polygons, and created the respective raster tiles by projecting and rasterizing the polygons.
c) Build attribute tables component of the tiles. For each of the three models, the resulting tile includes two tables: one storing parameters at the component level (Lumped soil series/component) and the other at the respective layer level. The name of the field attributes were maintained consistent with the name of the parameters listed in the respective model operative manual and code. Source variables are extracted from the seven imported tables and manipulated in different ways (e.g. by a weighted average across the soil profile) based on their thematic connection to the respective recipient model tables. While only a few source variables match the specifications of each model, the remainder of the parameters required manipulations such as: unit of measure update, conversion (e.g. Van Bemmelen factor 1.724 applied to convert source variable organic matter into organic carbon), aggregation, and cross combination. For each parameter, we derived its value from the single or combined R value of the source data. The first stage of this database processing included a rule for updating missing R value as: mean calculated as $R = ½ (L+H)$ if the latest two values were available. More importantly, as a result of the experience gathered during the development of the CEAP national project for cropland (USDA-NRCS, 2014), fundamental water

holding parameters common across the three models such as field capacity, wilting point and available water capacity (provided by SSURGO), were substituted with values calculated using the soil texture and soil carbon content based relationship by Rawls et al. (1982). In addition, in the occurrence of excessive depth of the first horizon, two composing layers were created by carving and merging the properties of the first two original SSURGO horizons (Figure 1 shows the occurrence of n horizons resulting into n+1 layers). A couple of model parameters, such as albedo and organic carbon, required additional land use information by MU. In that regard, we pre-processed look-up tables storing the spatial distribution of the CDL, CL, and/or NLCD grid categories within each map unit polygon. This information allowed establishing most likely values for the parameters from the available range of the source data based on the land use intensity impact on them. We provided details about the derivation of each single model parameter in metadata documentation included with the database. The remaining missing data were flagged, rather than substituted with default values.

d) Assemble the regional FGDBs. With this task, we created and populated regional FGDBs, each including the proper map and tabular tiles organized based on the geographic and hierarchical structure established by the 2-digit and 8-digit WBDs. Figure 3 depicts the hierarchical structure of the geographic and associated tabular data (see also Figure 4) embedded in the geodatabases. Within the FGDB, Feature Classes were conveniently organized in Feature Data Sets, while raster data, in unmanaged modality, were organized in Mosaic data sets.
e) Create and store data relationships. This task allowed storing in the FGDBs, and for each 8-digit Feature Class tile, two types of relationships: the first represents a spatial relationship between geographic entities (Feature Class containing the MU polygons) and nongeographic entities (the table of models attributes at the component/lumped level); the second one represents a relationship only between nongeographic entities (the table of models attributes at the Lumped/component and the one at the Layer level). Figure 4 depicts explicitly the links established via the newly created MUID and CID. We also retained the NASIS-SSURGO keys highlighted in Figure 1 by storing them in the derived tables.
f) Store Metadata and Tools. We used the FDGBs to store essential information about the database, such as details about the derivation of the various parameters, unit of measures, internal and mutual map and table organization, and specifications that can help the user and/or the developer apply the data in a convenient manner. We added to each FGDB a set of tools based on Python scripts that can be used to manipulate the data as needed. In particular, an application proposed as an ArcGIS tool, allows merging multiple tiles, including maps, tables and relationships, into single units at the desired scale.

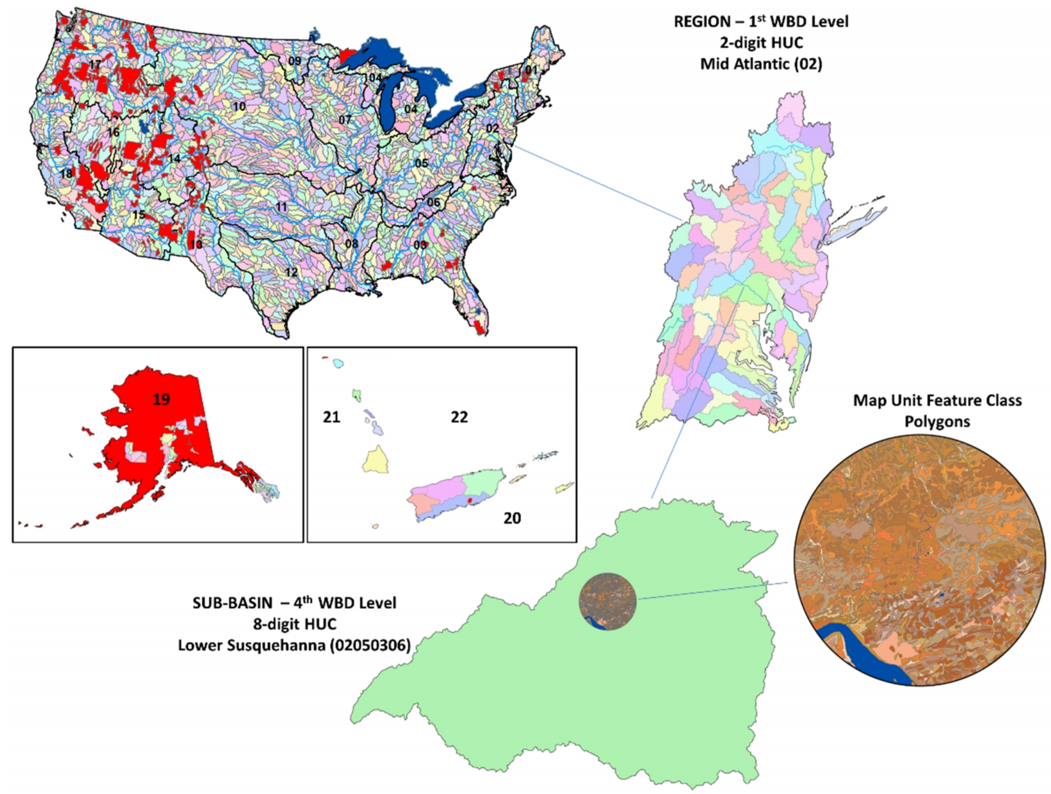

Figure 3. Geographic coverage and structure of the constructed geodatabases in the United States. Zoom-in of Region 02 and Sub-basin HUC8 02050306. Highlighted in red areas with currently unavailable SSURGO data.

Finally, the constructed database, both the vector and raster map version, includes information and their linkage for a total number of approximately 291,000 MUs and 898,000 soil series, respectively. The storage volume for each regional FGDB (vector and raster) is summarized in Table 2. The volume of the originally acquired SSURGO zipped packages was around 54 GB; the total storage of the constructed vector database resulted in 39.7 GB, while the raster version resulted in 13.3 GB. We applied a specific FGDB compression function to the vector Feature Classes and the associated tables of parameters, and we obtained a compression ratio between 2 and 5, which accomplished a reduction of the final total feature class FGDB to a final storage volume of 14.6 GB.

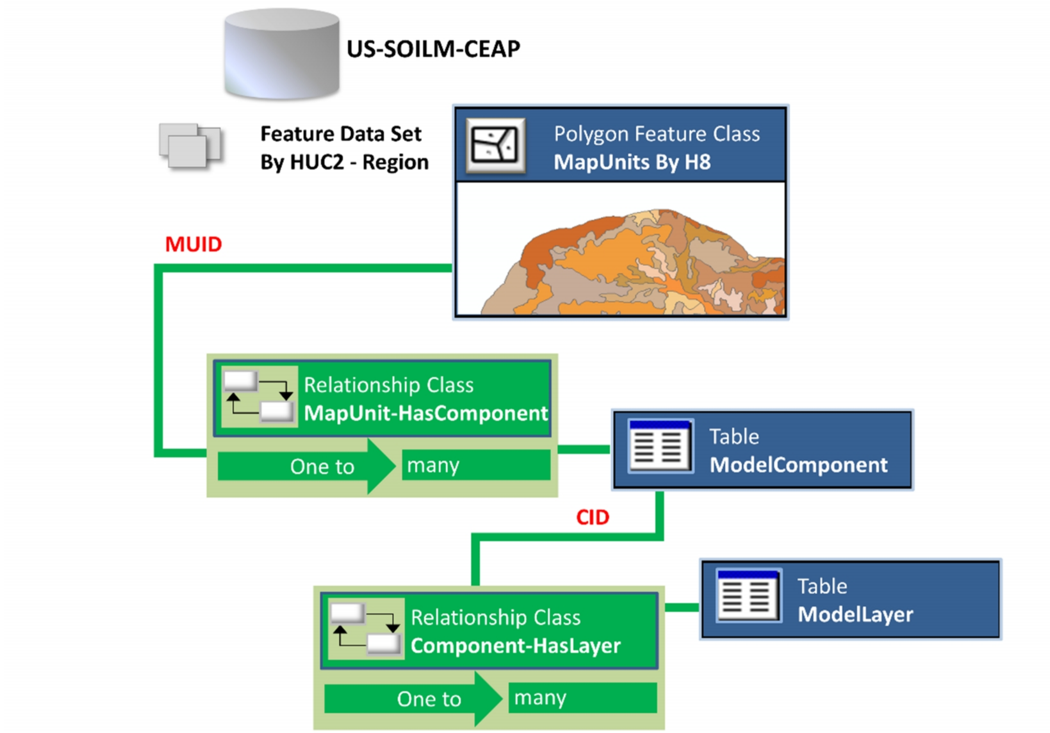

Figure 4. Relationships stored within the constructed US-SOILM-CEAP File Geodatabase.

4. DISCUSSION

We developed a geographical database repository of soil parameters for CAGNPM models with the purpose of retaining and expanding all the useful details, and the original areal and vertical grain of the source SSURGO soil series. In this way, we avoided the usual simplified representation of each map unit area with a single dominant component covering the largest area of the respective map unit, or alternatively with a standard component including parameters established using areal and vertical aggregation approaches. These practices are often applied a priori to simplify the data and to streamline the necessary one-to-one association of each map unit area to a single set of soil parameters. Promising approaches were developed (Gatzke et al. 2011; Yang et al. 2011); however, their extended capability of providing the best soil functional response remains to be proven for CAGNPMs' model applications across different watersheds and scales. Modelers working in small areas and/or designing precision conservation applications value the availability of the information for the entire soil series spectrum as derived from the surveys. Large-scale applications require the same ample information to optimize simulations and to correctly manage the response complexity and uncertainty that may increase with the dimension of the application domain. In this regard, we considered it useful for the modeler to be aware of the data originally missing, rather than filling the gaps with unproven methodologies.

Table 2. Regional distribution of the source data and constructed File Geodatabases.

Region	States	Used SSAs	Number of H8s*	Volume Vector** (GB)	Volume Raster*** (GB)
New England	RI, CT, MA, NH, ME, CN	67	47/11	1.34/0.41	0.43
Mid Atlantic	VA, MD, NY, CT, NJ, PA	238	82/6	2.26/0.78	0.71
South Atlantic-Gulf	FL, AL, MS, GA, SC, NC	451	171/32	4.30/1.63	1.13
Great Lakes	OH, IL, WI, IN, CN, MI, NY, PA	230	109/15	3.16/1.14	1.06
Ohio	TN, KY, VA, NC, WV, OH, IN, PA	408	120/0	4.39/1.68	0.89
Tennessee	AL, TN, GA, NC	134	32/0	1.01/0.36	0.23
Upper Mississippi	IL, MO, IA, MN, WI	361	131/0	6.1/2.24	1.12
Lower Mississippi	LA, TN, MS, KY, AR	192	79/3	1.19/0.51	0.33
Souris-Red-Rainy	MN, MT, ND, CN	59	6/40	0.98/0.34	0.43
Missouri	MO, KS, CO, NE, IA, SD, WY, MT	503	303/4	6.95/2.59	2.36
Arkansas-White-Red	TX, LA, OK, AR, NM, KS, CO	314	169/4	2.15/0.86	0.88
Texas-Gulf	TX	198	122/0	1.20/0.51	0.46
Rio Grande	TX, MX, NM	77	43/28	0.38/0.17	0.23
Upper Colorado	NM, CO, AZ, UT, WY	69	18/42	0.43/0.19	0.23
Lower Colorado	AZ, MX, NM	70	38/47	0.36/0.17	0.23
Great Basin	UT, NV, CA	87	28/43	0.38/0.14	0.36
Pacific Northwest	NV, OR, ID, WY, CA, WA	173	60/160	1.80/0.73	0.91
California	CA	125	83/43	1.1/0.40	0.64
Alaska	AK	26	2/158	0.45/0.10	0.21
Hawaii	HI	8	7/1	0.04/0.02	0.16
Caribbean	PR, VI	8	6/1	0.04/0.01	0.17
South Pacific	AS, GU, MP FM, MH, PW	10	1 /2	0.06/0.02	0.20
	Total	3,151	1657/640	39.7/14.6	13.8

* Complete vs incomplete WDB-H8 tiles
** Before compression vs after compression
*** Only pixels defined by the CDL data set

Earlier methods (Di Luzio et al. 2004; Frankenberger et al. 2011; Sheshukov et al. 2011) used ad-hoc procedure tools commonly applied within a GIS or relational database environment to define comprehensive model parameters and digital maps for a specific area of interest and a specific model. Our method constructed a multi-model geospatial database, which provides readily available soil parameters and streamlines each model's applications. This approach is particularly advantageous because it reduces processing time particularly when the applications are extended over large geographic domains. In addition, the integration of models by mosaics of site-specific simulations (e.g. ALMANAC and/or APEX) feeding modeling at the large-scale (e.g. SWAT) can benefit from commonly and consistently elaborated soil attributes. Having a single objective repository of model parameters provides the user community with a homogeneous baseline for model application comparisons. Since the parameters were elaborated in collaboration with scientists knowledgeable about soil characterization in conjunction with the models objective of our work, we established a stewardship of the soil model parameters at the highest detail currently available in the United States. This is particularly beneficial in order to maintain the database in sync with the progressive models evolution and with the update and expansion of the SSURGO and future source records. In addition, the maintenance is facilitated by the procedure that we automated into batch processes and developed with Python. This method provided the capabilities needed for processing the attribute tables as well as the geographic features for the entire United States. For instance, starting from the laborious method here developed using source data at the SSA level, we will be able to update it with minor changes due to the convenient distribution of the data at the state level by NRCS (Soil Survey Staff 2013) that became available to the public when this work was already at an advanced stage of development.

Uniquely, our method created a geographic database hosting vector and alternatively gridded maps. This option was provided earlier by the software described by Di Luzio et al. (2004) and Sheshukov et al. (2011), but which rely on the user to provide each raw SSURGO package input source, or as Frankenberger et al. (2011), using web services to access a confined volume of data. Gridded maps provide the unquestionable advantage when post-processing, combining and analyzing highly spatially variable parameters in the raster domain. We included the gridded tiles linked to the processed table of parameters consistent in cell size (30 meter), alignment and geographical coordinate system (AEA) with the CDL grids, because we expect them to be frequently used in combination. This raster environment is also widespread in environmental and hydrology model applications, because it was inherited from the native LANDSAT datasets and established by the USGS for its areal objectivity across the CONUS. However, since the raster conversion includes an unavoidable approximation, the entire collection of vector maps were retained via the vector version produced in our work. In this manner, the GIS user and/or developer can establish the needed scale of application by creating a new raster environment (cell size, coordinates system, and alignment) and move the data into grid without further approximations. This was successfully accomplished using the FGDB, which allowed us to efficiently store the remarkable size of geographic and tabular data, which is easily transferable across platforms. In the same FGDBs, we were also able to store metadata, documentation, auxiliary maps and tools. In using the FGDB, we had the opportunity to confirm expected advantages when compared to the former ESRI Personal Geodatabase and Shape file implementation (ESRI 2009 and 1998). In particular, the compression of the vector data in a read-only format consistently reduced the storage requirement (see Table 2) that became comparable to the raster version without losing information, query, and/or display performances.

Importantly, we structured the FGDB in hydrologic units-based tiles (8-digits level) defined by the latest WBD national framework. The tile system leverages on the WBD the capability for the aggregation of drainage information at different geographic scales, which facilitates sharing and analysis of land management and natural resource data. The hydrologic unit boundaries, being based on drainage properties, are unlinked to political or other program boundaries, and they are practically useful to water management entities such as state water agencies, water conservation districts, and drinking water suppliers. While we consider the 8-digit HUC tile an effective size for post-processing both for the vector and raster data, we provided functionality to merge the tiles at the desired upper scale or area of interest. This allows efficient extension of the support to crop management and modeling assessments extending from site-specific to large areas. Tiles displaying the raster data of each regional Mosaic data set are shown as a seamless mesh of rasters covering the entire country. We constructed this single FGBD leaving the composing raster unmanaged. In this way, we addressed the concern arising from the usage of a proprietary data format by elaborating and offering the gridded tiles in an open source format (GeoTIFF) and the table of attributes in dBASE formats, which is accessible using open source libraries.

We utilized relationship classes stored in the vector FGDBs to explicitly maintain the conceptual and functional links stored within the original surveys. These are based on the fundamental concept that each basic MU polygon outlines one or more components (soil series), which in turn is characterized by one or more horizons (layers). We actually expanded this concept upward by establishing an implicit spatial relationship between the 8-digit polygons and the MUs polygons, which express the concept that each 8-digit outlines one or more MU polygons. We built these sets of relationships to facilitate the retrieval and analysis of model soil attributes data within the FGDB using proper code, and also directly on the screen. Figure 5 show the results of the application of the Identify Tool within the ArcMap data frame when clicking on a location within the circle (Figure 3). A dialog box lists the 8-digit and MU feature at the clicked location. The bottom panel shows the list and the parameters of the linked components, Layer tables for the three models, and the respective SWAT parameters.

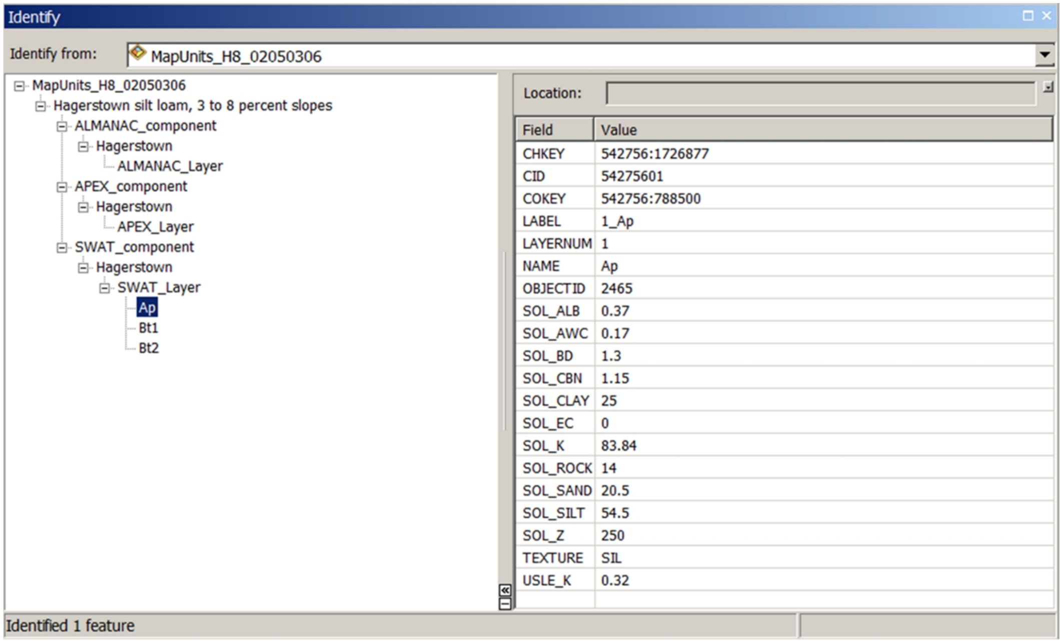

Figure 5. Identify dialog box report returned from the ArcMap Identify Tool applied within the File Geodatabase map shown in the circle of Figure 3.

Ultimately, while all the methods in literature address single models and/or generic key attributes distributed as multilayer surface of soil characteristics, our approach provides the data in the format, unit of measure, field format and names required by our target models. Our created geodatabases provide both an informational repository service and capability to map and baseline for further processing. We used this opportunity to include land use information (e.g. the CDL, CL, and NLCD layers) to pinpoint the most proper value of parameters (organic carbon and soil albedo) from a range of values provided by the data sources. By storing the original NASIS-SSURGO database keys, we not only facilitate post-checking of the data, but also facilitate the joining and/or further addition and/or processing of any other original soil parameter contained within. Our geodatabase is designed to be a crucial component of a developing system of data, namely CEAP HUMUS, and geo-spatial tools, namely GeoCEAP, supporting the integration of farm surveys, agricultural management and hydrological models to quantify the environmental benefits of conservation practices at the national scale as accomplished by the USDA CEAP program.

5. CONCLUSIONS

This paper presents the design and creation of US-SOILM-CEAP, a hydrology-structured geospatial database (geodatabase), storing soil-related parameters for three agriculture management simulation models (SWAT, APEX, and ALMANAC). It covers the entire

United States in the areas of availability of the main source information, such as the taxonomy-based SSURGO database. The core of this development is a largely automated GIS-database processing methodology, which operates from extracting and combining source variables of interest to store soil physical, hydraulic and chemical model parameters in the created geodatabases. The methodology represents a response to the growing challenge resulting from the escalating volume and complexity of the involved data, which requires the development of comprehensible, systematic, reproducible and efficient parameter estimation methods and tools.

In the current version of the methodology and geodatabase(s), a number of aspects provide practical relevance for the modeling community, such as:

- The database establishes an objective data repository and framework for future improvement of the data, the target and other similar models.
- The database simplifies and adds value to an extremely rich and valuable database of detailed soil information such as SSURGO.
- The methodology promises to streamline the expansion and update of the resulting geodatabase once either one or both SSURGO data sets and model inputs evolve. It offers stewardship of the proper parameter calculation at the place where models are developed and at the finest available resolution.
- The database is organized using hydrology-based (WBD polygons) informational tiles rather than units with no administrative (e.g. states or counties) boundaries.
- Tile data organization provides information in a conveniently manageable size, both for post-processing and in combination with other data in a GIS environment (e.g. definition of watershed delineation, Hydrologic Response Units and geo-morphological metrics). These may include the geographic up-scaling and/or the aggregation of parameters.
- The database offers both the advantages of the polygon and raster data sets without increasing storage volume.
- The database maintains and exposes the fine grain information and relationships, which are augmented within the hierarchical data structure.
- The methodology avoids irreversible gridding/aggregations to meet complete and specific requirements of CAGNPM models.
- A gridded realization of the geodatabase is provided using an open source raster format (GeoTIFF) and the table of attributes in dBASE formats, which are accessible using open source software.
- The methodology introduces a new paradigm in SSURGO-derived parameters: the use of auxiliary data to establish the value of specific CAGNPM model parameters.
- The database inherits a number of downstream usage solutions provided by the FGDB implement, including multiplatform, desktop and web service fruition.
- The methodology transfers existing soil physic relationships into the database, which then become functional for the retrieval of parameters and display on the screen.
- The methodology reviews essential SSURGO parameters using auxiliary land use/land cover data and water-holding properties of soil derived from texture classes information.

- The geodatabase could serve as a template for the future development of the next generation of digital soil databases derived from enhanced surveys and field-specific model assessments accomplished with the support of remote sensing information.
- The database includes tools for further geoprocessing, technical documentation and metadata.
- The database will be refreshed once a year and made available on the Web (www.brc.tamus.edu).

Ultimately, our geodatabase designed for modeling agricultural effects of management and conservation practices has the potential to outreach modelers, planners, policy makers and experts to overcome the paucity in essential, wide-area and readily available soil-related model parameters. These aspects address the continuously evolving quest for the optimization of agricultural production and the reduction of consequential environmental foot-prints.

ACRONYMS

AEA	Albers Equal Area projection coordinate system
ALMANAC	Agricultural Land Management Alternatives with Numerical Assessment Criteria
APEX	Agricultural Policy Environmental EXtender
CAC	Calcium Carbonate Content of Soil (%) – APEX model
CAGNPM	Crop and Agricultural Nonpoint Pollution Model
CDL	Cropland Data Layer
CEAP	Conservation Effect Assessment Project
CONUS	Conterminous United States
CL	Cultivated Layer
DMC	Disaster Monitoring Constellation
ESRI	Environmental Software Research Institute
FGDB	File Geodatabase
FSA	Farm Service Agency
GeoTIFF	Geographic Tagged Image File Format
HSG	Soil Hydrologic Group – APEX model
HU	Hydrologic Unit
HUC	Hydrologic Unit Code
HUMUS	Hydrologic Unit Model for the United States
ETM+	Landsat Enhanced Thematic Mapper Plus
MU	Map Unit
NAD83	North American Datum of 1983
NASIS	National Soil Information System
NASS	National Agricultural Statistics Service
NCSS	National Cooperative Soil Survey

NED	National Elevation Dataset
NHD	National Hydrography Dataset
NLCD	National Land Cover Dataset
NRCS	Natural Resources Conservation Service
NWS	National Weather Service
PH	Soil pH – APEX model
ROK	Coarse fragment content – APEX model
SAC-SMA	Sacramento Soil Moisture Accounting
SALB	Soil Albedo – APEX model
SAN	Sand Content – APEX model
SATC	Saturated Conductivity – APEX model
SCS	Soil Conservation Service
SSA	Soil Survey Area
SSURGO	Soil Survey Geographic
SWAT	Soil and Water Assessment Tool
USDA	United States Department of Agriculture
USGS	United States Geological Survey
US-SOILM-CEAP	United States – SOIL Model parameters – CEAP
WBD	National Watershed Boundary Dataset
WOC	Organic Carbon Concentration (%) – APEX model
WTMN	Minimum Depth to Water Table – APEX model

REFERENCES

Arnold, J. G. and Fohrer, N. (2005) SWAT2000: Current capabilities and research opportunities in applied watershed modeling. *Hydrological Processes*, 19(3), 563-572.

Batjes, N.H., Al-Adamat, R., Bhattacharyya, T., Bernoux, M., Cerri. C.E.P., Gicheru, P., Kamoni P., Milne, E., Pal, D.K. and Rawajfih, Z. (2007) Preparation of consistent soil data sets for SOC modelling purposes: secondary SOTER data sets for four case study areas. *Agriculture, Ecosystems and Environment*, 122, 26-34.

Beven, K. (2006) A manifesto for the equifinality thesis. *Journal of Hydrology*, 320, 18-36.

Boryan, C., Yang, Z., Mueller, R. and Craig, M. (2011) Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International*, 26(5), 341-358.

Boryan, C., Yang, Z. and Di, L. (2012) Deriving 2011 Cultivated Land Cover Data Sets Using USDA National Agricultural Statistics Service Historic Cropland Data Layers. IGARSS 2012 Conference, Munich, Germany, July 23-26, 2012.

Delgado, J. A., Khosla, R. and Mueller, T. (2011) Recent advances in precision (target) conservation. *Journal of Soil and Water Conservation*, 66(6), 167A-170A.

Di Luzio, M., Arnold J.G. and Srinivasan, R. (2004) Integration of SSURGO maps and soil parameters within a geographic information system and nonpoint source pollution model system. *Journal of Soil and Water Conservation*, 59(4), 123-133.

Di Luzio, M., Arnold J.G. and Srinivasan, R. (2005) Effect of GIS data quality on small watershed stream flow and sediment simulations. *Hydrological Processes*, 19, 629-650.

Duriancik, L.F., Bucks, James, D., Dobrowolski, P., Drewes, Y., Diane Eckles, S., Jolley, L., Kellogg, R.L., Lund, D., Makuch, J.R., O'Neill, M.P., Rewa, C.A., Walbridge, M.R., Parry, R. and Weltz, M.A. (2008) The first five years of the Conservation Effects Assessment Project. *Journal of Soil and Water Conservation*, 63(6), 185A-197A.

Environmental Systems Research Institute (ESRI) (1998) ESRI Shapefile Technical Description. http://www.esri.com/library/whitepapers/pdfs/shapefile.pdfs Last accessed May 6, 2013.

Environmental Systems Research Institute (ESRI) (2009) The Top Nine Reasons to Use a File Geodatabase. http://www.esri.com/news/arcuser/0309/files/9reasons.pdf Last Accessed May 9, 2013.

Environmental Systems Research Institute (ESRI) (2013) ArcGIS Spatial Analyst. http://www.esri.com/software/arcgis/extensions/spatialanalyst Last accessed May 9, 2013.

Frankenberger, J.R., Dun, S., Flanagan, D.C., Wu, J.Q. and Elliot, W.J. (2011) Development of a GIS interface for WEPP model application to Great Lakes forested watersheds. ASABE Conference, September 18-21, 2011, Anchorage Alaska. ASABE Publication Number 711P0311cd.

Gatzke, S.E., Beaudette, D.E., Ficklin, D.L., Luo, Y., O'Geen, A.T. and Zhang, M. (2011) Aggregation strategies for SSURGO data: effects on SWAT soil inputs and hydrologic outputs. *Soil Science Society of America Journal*, 75(5), 1908-1921.

Geza, M. and McCray, J.E. (2008) Effects of soil data resolution on SWAT model stream flow and water quality predictions. *Journal of Environmental Management*, 88, 393-406.

Grunwald, S. (2009) Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152, 195-207.

Homer, C., Huang, C., Yang, L., Wylie, B. and Coan, M. (2004) Development of a 2001 National Land-Cover database for the United States. *Photogrammetric Engineering and Remote Sensing*, 70, 829-840.

Johnson, D.M. and Mueller, R. (2010) The 2009 Cropland Data Layer. *Photogrammetric Engineering and Remote Sensing,* 76 (11), 1201-1205.

Kiniry, J.R., Williams, J.R., Gassman, P.W. and Debaeke, P. (1992) A general, process-oriented model for two competing plant species. *Transactions of the ASAE*, 35, 801-810.

Kumar, S. and Merwade, V. (2009) Impact of Watershed Subdivision and Soil Data Resolution on SWAT Model Calibration and Parameter Uncertainty. *Journal of American Water Resources Association*, 45(5), 1179-1196.

Mausbach, M. J. and Dedrick, A. R. (2004) The length we go: Measuring environmental benefits of conservation practices. *Journal of Soil and Water Conservation*, 59, 96-103.

Mednick A.C. (2010) Does soil data resolution matter? State Soil Geographic database versus Soil Survey Geographic database in rainfall-runoff modeling across Wisconsin. *Journal of Soil and Water Conservation*, 65(3), 190-199.

Miller, D.A. and White, R.A. (1998) A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling. *Earth Interactions*, 2, 1-26.

Moriasi, D.N. and Starks, P.J. (2010) Effects of the resolution of soil dataset and precipitation dataset on SWAT2005 streamflow calibration parameters and simulation accuracy. *Journal of Soil and Water Conservation*, 65(2), 63-78.

Peschel, J.M., Haan, P.K. and Lacey, R.E. (2006) Influences of soil dataset resolution on hydrologic modeling. *Journal of American Water Resources Association*, 42(5), 1371-1389.

Perez–Quezada, J. F., Cavero, J., Williams, J.R., Roel, A. and Plant, R. E. (2003) Simulation of within-field yield variability in a four–crop rotation field using SSURGO soil–unit definitions and the EPIC model. *Transactions of the ASAE*, 46(5), 1365-1374.

Rawls, W.J., Brakensiek, D.L. and Saxton, K.E. (1982) Estimation of Soil Water Properties. *Transaction of the ASAE*, 25(5), 1316-1320 and 1328.

Schwarz, G. E. and Alexander, R.B. (1995) State Soil Geographic (STATSGO) database for the conterminous United States, U.S. Geological Survey. Open File Report 95-449: Reston, VA, USA.

Sheshukov, A.Y., Daggupati, P., Douglas-Mankin, K.R. and Lee, M. (2011) High spatial resolution soil data for watershed modeling: 1. Development of a SSURGO ArcSWAT utility. *Journal of Natural and Environmental Sciences*, 2(2), 15-24.

Singh, H.V., Kalin, L. and Srivastava, P. (2011) Effect of soil data resolution on identification of critical source areas of sediment. *Journal of Hydrologic Engineering*, 16(3), 253-263.

Soil Survey Staff (1999) Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys, 2nd edition. Agricultural Handbook 436, United States Department of Agriculture, Natural Resources Conservation Service: Washington, DC, USA.

Soil Survey Staff (2013) Gridded Soil Survey Geographic (SSURGO) Database for Texas. United States Department of Agriculture, Natural Resources Conservation Service. http://datagateway.nrcs.usda.gov/ Last accessed 3/27/2013.

USDA-NRCS (2012) Soil Survey Geographic (SSURGO) Database. U.S. Department of Agriculture, Natural Resources Conservation Service: Washington, DC, USA. http://soildatamart.nrcs.usda.gov Last accessed 11/15/2012.

USDA-NRCS (2014) Conservation Effects Assessment Project (CEAP), National Assessment, Cropland National Assessment. U.S. Department of Agriculture, Natural Resources Conservation Service: Washington, DC, USA. http://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/technical/nra/ceap/na/?cid=nrcs143_014144 Last accessed 12/20/2013.

USDA-SCS (1994) State Soil Geographic (STATSGO) Database. Publication No. 1492. U.S. Department of Agriculture, Soil Conservation Service, National Soil Survey Center: Washington, DC, USA.

van Rossum, G. (2013) Guido van Rossum - Personal Home Page. http://www.python.org/~guido/ Last accessed May 9, 2013.

Wang, X. and Melesse, A.M. (2006) Effects of STATSGO and SSURGO as inputs on SWAT model's snowmelt simulation. *Journal of American Water Resources Association*, 42(5), 1217-1236.

Williams, J. R., Jones, C. A., Kiniry, J. R. and Spanel, D. A. (1989) The EPIC crop growth model. *Transaction of the ASAE,* 32(2), 497-511.

Williams, J. R., Harman, W. L., Magre, M., Kizil, U., Lindley, J. A., Padmanabhan, G. and Wang, E. (2006) APEX feedlot water quality simulation. *Transaction of the ASAE,* 49(1), 61-73.

Williams, J. R., Arnold, J. G., Kiniry, J. R., Gassman, P. W. and Green, C. H. (2008) History of model development at Temple, Texas. *Hydrological Sciences Journal*, 53(5), 948-960.

Xie, Y., Kiniry, J.R. and Williams, J.R. (2003) The ALMANAC model's sensitivity to input variables. *Agricultural Systems*, 78, 1-16.

Yang, Y., Wilson, L.T., Wang J. and Li, X. (2011) Development of an integrated Cropland and Soil Data Management system for cropping system applications. *Computers and Electronics in Agriculture*, 76, 105-118.

Zhang, Y., Zhang, Z., Reed, S. and Koren, V. (2011) An enhanced and automated approach for deriving a priori SAC-SMA parameters from the soil survey geographic database. *Computers and Geosciences*, 37, 219-231.