International Journal of Geospatial and Environmental Research

Volume 4 | Number 1 Article 2

2-2017

Advancement of a Soil Parameters Geodatabase for the Modeling Assessment of Conservation Practice Outcomes in the United States

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

Mike J. White
Agriculture Research 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; White, Mike J.; Arnold, Jeffrey G.; Williams, Jimmy R.; and Kiniry, James R. (2017) "Advancement of a Soil Parameters Geodatabase for the Modeling Assessment of Conservation Practice Outcomes in the United States," *International Journal of Geospatial and Environmental Research*: Vol. 4 : No. 1 , Article 2.
Available at: https://dc.uwm.edu/ijger/vol4/iss1/2

Advancement of a Soil Parameters Geodatabase for the Modeling Assessment of Conservation Practice Outcomes in the United States

Abstract

US-ModSoilParms-TEMPLE is a database composed of a set of geographic databases functionally storing soil-spatial units and soil hydraulic, physical, and chemical parameters for three agriculture management simulation models, SWAT, APEX, and ALMANAC. This paper introduces the updated US-ModSoilParms-TEMPLE, which covers the entire United States and is organized as a framework of 22 nested and hydrologically-ordered regional geographic databases with internal spatial segmentation drainage-defined at a conveniently manageable tile (Watershed Boundary Dataset's, WBD, 8-digit Subbasin) level. Spatial features are stored in multiple formats (raster and vector) and resolutions (10-meter and 30-meter), while being in direct relationship with the table of attributes storing the models' parameters.

A significant number of former parameter voids, determined by the local incompleteness of the source datasets, were filled using a methodology leveraging upon the hierarchy of the Soil Taxonomy information and the geographic location of the gaps. The functionality of each geographic database was extended by adding customized tools, which streamline the incorporation into geoprocessing workflows, the aggregation and extraction of data sets, and finally the export to other model support software user environments. These tools are attached and conveniently distributed along with detailed metadata documentation within each of the developed regional geographic databases. The system hosting this framework is developed using a proprietary software format (ESRI® File Geodatabase), however, a companion version of the framework of 8-digit tiles is also developed and provided using openly accessible formats. The experience shared in this paper might help other efforts in developing hydrology-oriented geographical databases.

Keywords

Geodatabase, Soil, Crop Model, Hydrology Modeling, SSURGO, SWAT, APEX, ALMANAC, ArcSWAT, WinAPEX

Acknowledgements

Soil Survey Staff, Natural Resources Conservation Service (NRCS), United States Department of Agriculture (USDA). National Agricultural Statistics Service (NASS), United States Department of Agriculture (USDA). United States Geological Survey (USGS). Environmental Protection Agency (EPA).

1. INTRODUCTION

US-ModSoilParms-TEMPLE (United States Temple Models' Soil Parameters) is a database composed by a set of hydrology-structured geographic databases functionally storing soil-spatial units and soil-related parameters for three agriculture management simulation models, SWAT (Soil and Water Assessment Tool) (Arnold and Fohrer 2005); APEX (Agricultural Policy Environmental Extender) (Williams et al. 2006); and ALMANAC (Agricultural Land Management Alternatives with Numerical Assessment Criteria) (Kiniry et al. 1992). US-ModSoilParms-TEMPLE covers the entire United States in the areas of availability of the main source information, such as the Soil Taxonomy-based Soil Survey Geographic Database (SSURGO) (USDA-NRCS 2016) database. The core of this development was largely automated in a GIS-based methodology (Di Luzio et al. 2014), which operates from extracting the involved tables and combining source spatial layers, to jointly storing derived spatial units with associated soil physical, hydraulic and chemical model parameters in the final database products.

Besides the advancement of this development, a user-community review of an earlier version of the database has highlighted the following concerns and/or desired improvements:

a) SSURGO source datasets are partially complete and updated yearly. In fact, the main source spatial units, such as the Map Unit (MU) polygons, are continuously extended, yet in some instances entirely or partially missing within a specific Soil Survey Areas (SSAs) part of the entire set covering the country. In a number of instances, available MU polygons are associated with partially complete soil record tables. Both situations contribute to the spottiness of the derived model attributes.
b) Spatial features and models' parameter tables part of US-ModSoilParms-TEMPLE can be conveniently incorporated into geo-processing workflows. However, the functionalities of the database are expected to be extended and serve the needs of both the beginner user as well as the expert developer.
c) Conventional metadata associated with US-ModSoilParms-TEMPLE may not fulfill the practical needs of a variety of users. Similarly to other interdisciplinary datasets, such needs range based on user's skill level as well as background and disciplines of interest. Metadata distributed within the database need to address different purposes, from detailing the development to efficiently using the information and associated tools.
d) Proprietary software formats traditionally lead the development and serve the needs of the GIS-oriented hydrologic modeling community. However, the growing Open Source (OS) community calls for geographic data sets easily implementable in the respective environments.

While core updates of the US-ModSoilParms-TEMPLE database are planned to follow the traditional yearly update of the source data by USDA-NRCS (United States Department of Agriculture-Natural Resources Conservation Service), our goal is to advance conceptual and practical value to this derived database. In the next sections of this paper, we highlight and discuss the latest updates and related methods introduced to address the concerns of the user-community listed above.

2. MATERIALS AND METHODS

The data processing environment is the ESRI (Environmental Systems Research Institute) ArcGIS version 10.1 along with ArcPy and the Python (Python.org) version 2.7. SSURGO data source update comprises more than three-thousand (3,261) complete (including both digital map and tabular data) SSA packages acquired on October 2015 from the USDA-NRCS web site at http://websoilsurvey.sc.egov.usda.gov/. The SSURGO MU polygon boundaries in Geographic coordinates, formerly provided using the North American Datum of 1983 (NAD83), are newly provided using the World Geodetic System 1984 (WGS84) datum, which we retained for the polygons of the final regional geographical databases.

2.1 FILE GEODATABASES

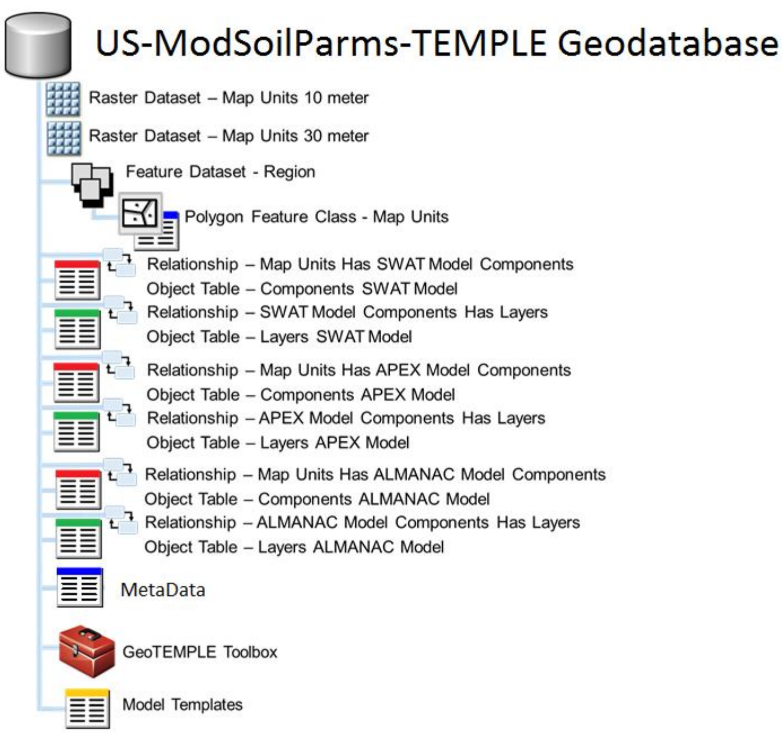

Figure 1. Primarily objects of the US-ModSoilParms-TEMPLE geodatabases.

Each geographic database is primarily developed as an ESRI ArcGIS File Geodatabase (FGDB). The updated internal framework of the geodatabase hosts the 8-digit-based tiles (8-digit is the original coding used to identify *Subbasins* within WBD) (USDA-NRCS et al. 2016), each one including: a) the soil MUs features as Raster Datasets (10-meter and 30-meter) and as polygon Feature Classes grouped within a regional Feature Dataset; the models' parameters at the component and layer level; b) the relationships linking the MU polygons to the respective models' tables and records; c) a metadata table; d) software tools and external software model templates (Figure 1).

2.2 LOCAL GAP-FILLING

We extended the former methodology (Di Luzio et al. 2014) by processing the records, here intended as parameter sets stored in tables, of the constructed models' attributes. Representative replacement records for the three models (SWAT, APEX, and ALMANAC) were pre-compiled at two (component and layer) levels as follows:

a) At the component level, the records of the entire country were grouped in multiple ways based on a hierarchical set of Soil Taxonomy attribute values (Soil Survey Staff 1999). The grouping attributes are:
   1) *Taxorder*, such as "The highest level in Soil Taxonomy";
   2) *Taxsuborder*, such as "The second level of Soil Taxonomy. The suborder is below the order and above the great group";
   3) *Taxpartsize*, such as "Particle-size refers to grain-size distribution of the whole soil and is not the same as texture";
   4) *Taxgrtgroup*, such as "The third level of Soil Taxonomy. The category is below the suborder and above the subgroup";
   5) *Taxsubgrp*, such as "The fourth level of Soil Taxonomy. The subgroup is below great group and above family";
   6) *Taxclname*, such as "A concatenation of the Soil Taxonomy subgroup and family for a soil (long name)".

   The set of records sharing the same group attribute value (Group Item records), were scored based on the number of missing parameters and areal frequency. For each Group Item a single record was selected as the most representative based on the resulting score (minimum number of missing parameters and largest area). In addition, representative records were built for a wide number of records referring to land mediums, inappropriately to be defined as soils (e.g. water, rock, etc.) or poorly defined soils (e.g. badlands, gullies, pits, etc.).
b) At the layer level, the records for the entire country were grouped based on the *Texture* attribute obtained adding a modifier defined "to denote the presence of a condition of a component other than sand, silt, or clay" (Soil Survey Staff 1999). The representative record for each group item was selected based on completeness (minimum number of missing parameters).

Component-level records were reviewed following these ordered steps:

a) Empirical step: Each record carrying a partially populated set of parameters was associated to matching records sharing the same name (*Component Name* attribute) and part of the same Soil Survey Area (same *SSA* ID). If a match was not found, the search was extended to all the SSAs of the same State. Each of the matching records was scored by counting the number of missing parameters. The record of the matching component with the highest score (minimum number of missing parameters) was used to substitute the entire record or the specific parameter missing in the record under review. In this instance, the update involved the missing values of the associated records at the layer level as well. A particular case occurred when the *Component Name* matched one of the Soil Taxonomy identified Group Item values (see points ‘b’ above). In this case, the record or the specific parameter of the matching representative record was used to substitute the one under review.

b) Soil Taxonomy-base step: each missing parameter of the records carrying Soil Taxonomy information was refilled with the representative record value of the respective Group Item. The ordered search for the matching Group Item started with the most populated group (*Taxclname*) and eventually ended with the least populated one (*Taxorder*) (Figure 2). If a matching group item was not found, a specific default value was assigned to the component level parameter.

Each layer-level parameter was reviewed, and when missing, was refilled with the parameter value associated to the respective Texture-group record earlier defined. If a matching group was not found, a specific parameter default value was assigned to the parameter.

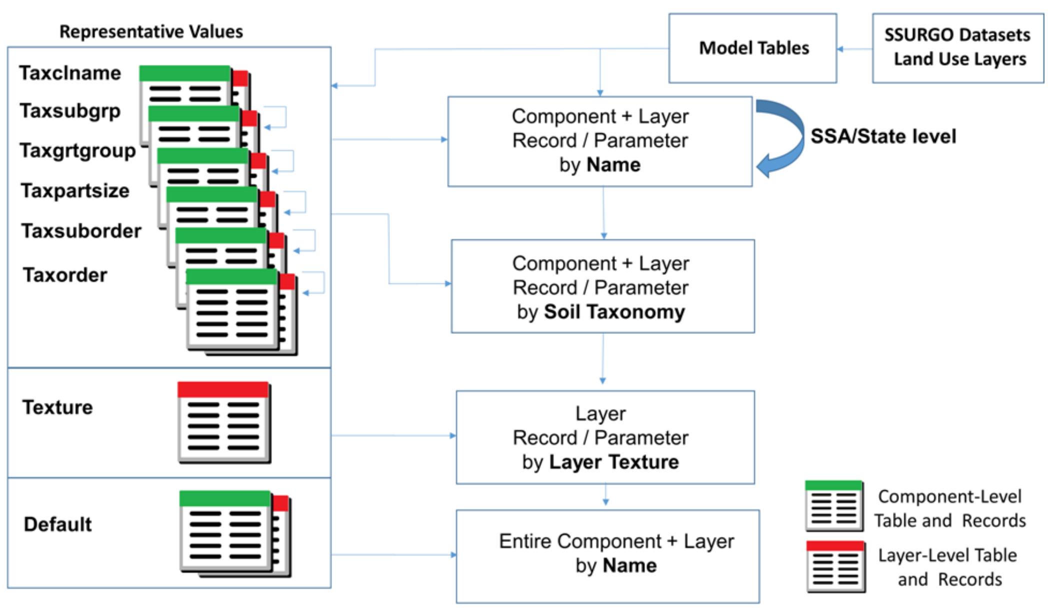

Figure 2. Local gap-filling methodology for US-ModSoilParms-TEMPLE geodatabases.

2.3 TOOLS

We reviewed and expanded a set of post-processing tools adding key functionalities to the geodatabases. The tools are developed as part of a Toolset (namely *Soil Databases*), as part of the customized Toolbox (namely *GeoTEMPLE*) for ArcGIS, and are included within each geodatabase (Figure 3 and 1), as such requiring minimal installation steps.

The functionalities of the added tools include:

a) Merging/Extracting specific or entire tile spatial features (polygon Feature Classes and Raster Datasets) and the respective model tables, while maintaining the original relationships.
b) Exporting features and tables to formats suitable for being consumed by modeling environments such as WinAPEX (Steglich and Williams 2008) and ArcSWAT (Winchell et al. 2009).
c) Streamlining of all the above functionalities for any specific areas and/or location of interest in the country.

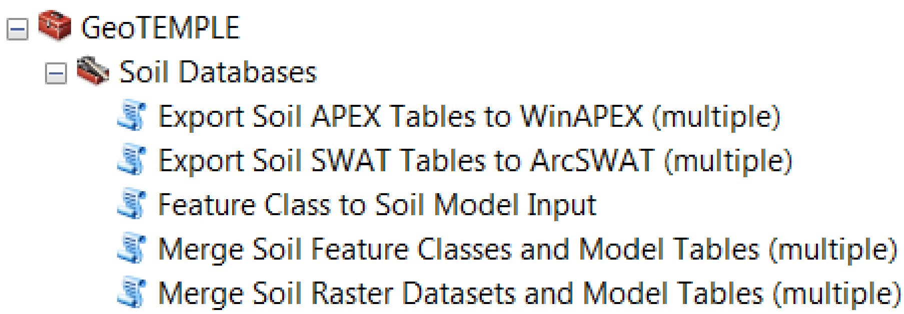

Figure 3. The Soil Databases Toolset distributed with the US-ModSoilParms-TEMPLE File Geodatabases.

Alternatively, to the classical user-friendly screen-driven application, each tool can be directly included in geoprocessing workflows using ArcGIS Model Builder and/or Python along with the ArcPy site-package. Python commands and syntax for the tools were generated and documented in the metadata. The functionalities are available with data stored both at the desk top and the local network level.

2.4 METADATA

We updated and provided metadata as a table attachment to each geodatabase. Both standard FGDC (Federal Geographic Data Committee) metadata and a detailed data User Guide in a PDF file are made part of each regional geodatabase (Figure 1). The first provides essential information for referencing the data, and the second provides technical material and tutorial exercises for a convenient use of the data and tools.

In addition, while basic description of each tool and data are accessible browsing the geodatabase using ArcCatalog, fundamental information about the specific method used to substitute missing parameters are detailed in the database. Strings flagging and describing the data substitution are stored in additional fields and records of the component and layer level tables. Strings are designed to be easily parsed and to trace back the substitution method.

2.5 OPEN SOURCE AND MOSAIC DATA SETS

We reviewed the construction of a file-folder-based framework, namely (US-ModSoilParms-TEMPLE-OS), which contains the 8-digit-tile spatial features and the associated tables using open source file formats. Raster GeoTIFF (Geographic Tagged Image File Format) files are built as the image of geodatabase map unit Raster Dataset tiles at the 10-meter and 30-meter cell size. Polygon ESRI Shapes files correspond to the geodatabase map unit Feature Classes. The model attributes were stored using dBASE tables.

The folders resemble the regional geodatabases and contains other six folders, such as:

a) *10m*, which contains the 10-meter resolution map unit raster tiles in GeoTIFF format;
b) *30m*, which contains the 30-meter resolution map unit raster tiles in GeoTIFF format;
c) *Shape*, which contains the map unit polygon tiles in Shape file format;
d) *ALMANAC*, which contains the Component and Layer tables for the ALMANAC model in dBASE format;
e) *APEX*, which contains the Component and Layer tables for the APEX model in dBASE format;
f) *SWAT*, which contains the Component and Layer tables for the SWAT model in dBASE format.

Based on this framework, we also constructed nation-wide ArcMap Mosaics at 10-meter and 30-meter resolution, which represent an alternative way to access the content of national geodatabases.

3. RESULTS AND DISCUSSION

The application of the procedure outlined in section 2.2, refilled the total number of parameter voids (local gaps) shown for each model in Table 1. Notably, for the restricted dominant component, a remarkable number of voids were refilled. The number escalates when considering the complete set of components. However, models are often applied just considering the dominant component as representative of the respective map unit area. For this reason, we are delivering the data in two fashions: one carrying only the dominant component and the associated layers for each model (Lite version); and the second one carrying the complete set of components and the associated layers for each model (Complete version). Both are developed using the 10.1 version of ArcGIS and FGDB, to maintain backward and forward compatibility. The constructed databases include information and their linkage for a total number of approximately 302,000 MUs and 966,000 soil series. The resulting total storage volume for the complete version of the 22 regional FGDB is 26.2 GB, as each regional FGDB now includes both the vectors (Feature Classes) and raster (Raster Datasets) at two resolutions (10-meter and 30-meter). The storage volume for the companion US-ModSoilParms-TEMPLE-OS database is 152 GB. We are developing a companion geodatabase system using larger scale input source to provide the same functionality and model attributes in areas of the country for which source SSURGO information are still spatially incomplete.

Table1. Percentage and total model parameters voids refilled at the Component and Layer level.

Model	Component		Layer	
	All	Dominant	All Components	Dominant Component
SWAT	33.7 (1,302,828)	16.4 (199,217)	15.3 (5,638,393)	9.1 (1,207,865)
APEX	31.4 (1,089,568)	9.3 (233,557)	14.6 (17,539,311)	8.8 (3,787,412)
ALMANAC	24.5 (4,257,519)	21.1 (1,152,174)	19.3 (14,202,484)	13.8 (3,670,560)

The 8-digit tile system was confirmed to facilitate the hydrology-based geoprocessing workflows, while providing an easily manageable size for the detailed polygon Feature Class. In addition, the double-raster option facilitates the current and near future inclusion in geoprocessing workflows preferring ready-to-use Raster Datasets, sharing for the CONUS (Conterminous United States) the same PCS (Projected Coordinate System), alignment, and/or resolution of the often used in conjunction Cropland Data Layer (CDL) (Boryan et al. 2011) and the projected rasters of the Nation Hydrographic Dataset Plus (NHDPlus) (McKay et al. 2012). Raster tiles in Region 19, 20 and 22 are set to a distinct and locally-proper PCS. In the South Pacific Region (Region 22), land use-biased parameters were newly defined with the support of lately collected local land use layers (Liu and Fischer 2007).

The tools and Python scripts included with each FGDB streamline the day-to-day operations using the features and data tables of the entire US-ModSoilParms-TEMPLE. In fact, the built-in Tools manipulate the data as needed by merging multiple tiles and/or any subsets of the Feature Classes along with model attributes and relationships across distinct regional FGDBs. We designed this approach to allow the extraction a single data set just for the desired Area of Interest (AOI), even though it might lay across one or multiple regions and/or tiles. Furthermore, we enhanced the tools by providing the capability to select and/or define target locations around the country and obtain the desired spatial features and soil attributes data with a format designed for being consumed by the common model user interfaces. Currently served interfaces are ArcSWAT and WinAPEX via a template included within each FGDC framework (Figure 1). In this way, the beginner user is shielded from the need to understand the complexity of both the original source data as well as of the simpler US-ModSoilParms-TEMPLE, and with a few steps can obtained the desired data and format. Concurrently, the expert user and programmer can iterate through datasets and features by incorporating the associated scripts and tools for more complex calculations and decision making. The beginner user is now provided with a complete data User Guide, which includes Tools descriptions and Tutorials. The Model Builder and Python programmer can find within the same document the specifications needed to develop more advanced applications of the data.

We conveniently attached the document to each FGDB as such to be easily accessed at the application environment as a metadata item. In fact, the User Guide stores 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 entire and/or part of the data in a convenient manner. The other essential item of the attachment includes the FGDC metadata, which is necessary information presented as an Extensible Markup Language (XML) document and embedded in each FGDB composing US-ModSoilParms-TEMPLE. The document captures the basic characteristics of the US-ModSoilParms-TEMPLE resource and represents the what, who, when, where, and why of the resource. For practical reasons, primarily listed above, we provide more complete information about the what, why and how of the resource in the User Guide completing the metadata.

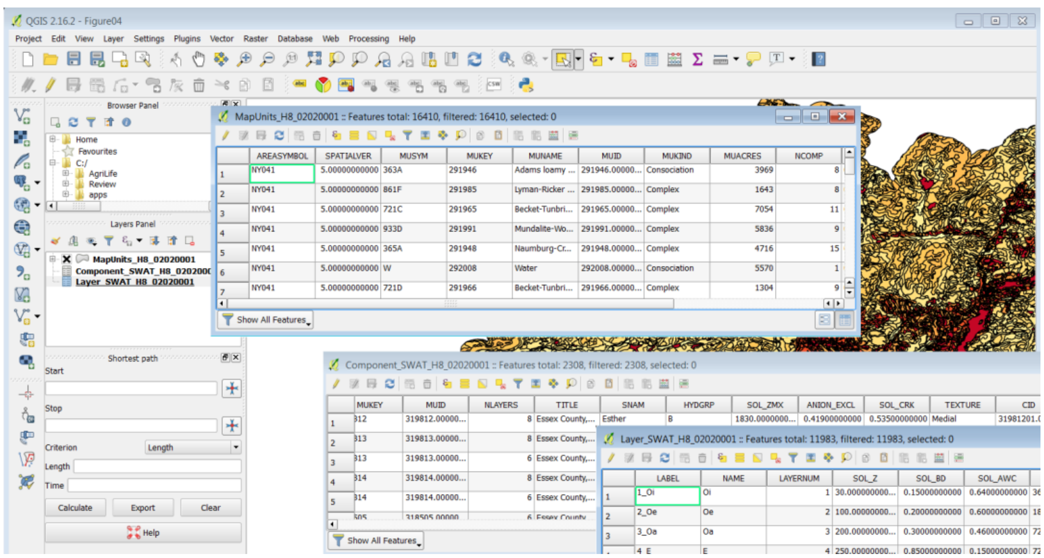

Figure 4. US-ModSoilParms-TEMPLE-OS features within the display of the Nøbedo release of QGIS.

ESRI software provides to Open Source programming languages such as Python and R (www.r-project.org) the capability to access and transfer data back and forth from ArcGIS geodatabases via ArcPy and R-ArcGIS Bridge. In this way, US-ModSoilParms-TEMPLE can be fully analyzed and combined using a large volume of libraries available with Python and R languages. This US-ModSoilParms-TEMPLE-OS version aimed to reach further, being completely independent from the native ArcGIS formats is directly accessible. For example, the datasets are easily useable by QGIS (www.qgis.org), a cross-platform free and open-source desktop GIS, which supports the data formats we used in our development (Figure 4).

As a final option, we used the OS file system to generate ArcGIS Mosaic datasets at the 10-meter and 30-meter resolution to provide a way to have Raster Datasets added to it directly and queries performed on the collections. These objects provide advanced properties for storing and managing raster data that can be directly accessed and easily served over the Internet.

5. CONCLUSIONS

An ensemble of models (SWAT, APEX, and ALMANAC) is historically applied for the hydrology-based assessment of the outcomes from conservation practice implementations, and developed at research laboratories located in Temple, Texas.

We reviewed and extended the construction of a set of geodatabases functionally hosting the spatial features and soil attributes necessary for the simulation using these three agricultural hydrology models across the United States. The primary introduced amendments include:

- The definition of values for the remarkable number of missing model parameters, previously flagged in the earlier development; information about the substitution methodology stored for each substituted parameter.
- The review of the architecture of the FGDBs, which now include Raster Datasets at two resolutions (10-meter and 30-meter), Feature Classes, model tables, metadata, and tools.
- Geoprocessing tools for a wide range of user's skill: from the basic ArcGIS user to the advanced Python programmer; option to store the data either at the desk top as well as the local network.
- Two levels of metadata attached to each geodatabase: XML FGDC metadata and detailed technical documentation within a User Guide and Tutorial document.
- Two main levels of fruition of the data using OS software (e.g. Python and R), through ArcGIS site-package or adopting the full OS Shape file GeoTIFF based framework.

The geographic coverage of the acquired data included the Conterminous United States (CONUS), Alaska and the Territories, Commonwealths, and Island Nations. The geodatabases are planned to be available at Soilandwaterhub.org\GeoTEMPLE and to be updated yearly accordingly to the source data updates by NRCS. A near-future planned update is the inclusion of additional insights from the National Cooperative Soil Survey (NCSS) Soil Characterization Database (ncsslabdatamart.sc.egov.usda.gov) and the extension for the upcoming models' (e.g. SWAT+) updates.

ACRONYMS

ALMANAC	Agricultural Land Management Alternatives with Numerical Assessment Criteria
AOI	Area of Interest
APEX	Agricultural Policy Environmental EXtender
CDL	Cropland Data Layer
CONUS	Conterminous United States
ESRI	Environmental Software Research Institute
FGDB	File Geodatabase
FGDC	Federal Geographic Data Committee
GeoTIFF	Geographic Tagged Image File Format
HUC	Hydrologic Unit Code
MU	Map Unit
NAD83	North American Datum of 1983
WGS84	World Geodetic System 1984
NHDPlus	National Hydrography Dataset Plus
NRCS	Natural Resources Conservation Service
OS	Open Source
SSA	Soil Survey Area
SSURGO	Soil Survey Geographic
SWAT	Soil and Water Assessment Tool
USDA	United States Department of Agriculture
US-ModSoilParms-TEMPLE	Temple Models' Soil Parameters for the United States
XML	Extensible Markup Language
WBD	Watershed Boundary Dataset

REFERENCES

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

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

Di Luzio, M., Norfleet, M. L., Arnold, J. G., Williams, J. R., and Kiniry, J. R. (2014) A Soil Parameters Geodatabase for the Modeling Assessment of Agricultural Conservation Practices Effects in the United States, *International Journal of Geospatial and Environmental Research*, 1(2), Article 6.

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.

Liu, Z., and Fischer, L. (2007) American Samoa Vegetation Mapping Using Very High Spatial Resolution Imagery – Methodology. USDA Forest Service, Pacific Southwest Region, Forest Health Protection. http://www.fs.usda.gov/Internet/FSE_DOCUMENTS/fsbdev3_046054.pdf Last accessed 09/15/2016.

McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A. (2012) "NHDPlus Version 2 : User Guide". Available at ftp://ftp.horizon-systems.com/nhdplus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf Last accessed 09/15/2016.

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.

Steglich, E.M. and Williams, J.R. (2008) Agricultural Policy/Environmental eXtender Model: User's manual. Version 0604 DOS and WINAPEX interface. BREC Report # 2008-16. Temple, TX: Texas AgriLife Research, Texas A&M University, Blackland Research and Extension Center.

USDA-NRCS (2016) Soil Survey Geographic (SSURGO) Database. United States Department of Agriculture, Natural Resources Conservation Service: Washington, DC, USA. Available at http://websoilsurvey.sc.egov.usda.gov/ Last accessed 09/15/2016.

USDA-NRCS, USGS, and USEPA (2016) United States Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS), United States Geological Survey (USGS), and the United States Environmental Protection Agency (USEPA) Watershed Boundary Dataset for the United States: Washington, DC, USA. Available at http://datagateway.nrcs.usda.gov Last accessed 09/15/2016.

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

Winchell, M., Srinivasan, R., Di Luzio, M., and Arnold, J.G. (2009) ArcSWAT 2.3.4 interface for SWAT2005 (ver. 2005): Temple, Texas, GSWR Agricultural Research Service & Texas Agricultural Experiment Station.