An Enhanced and Automated Approach for Deriving a priori SAC-SMA Parameters from the Soil Survey Geographic Database

Yu Zhang[1,*], Ziya Zhang[1,2], Seann Reed[1], Victor Koren[1]

Abstract

This paper presents an automated approach for processing Soil Survey Geographic (SSURGO) Database and National Land Cover Database (NLCD), and deriving gridded a priori parameters for the National Weather Service (NWS) Sacramento Soil Moisture Accounting (SAC-SMA) model from these data sets. Our approach considerably extends methods previously used in the NWS and offers automated and geographically invariant ways of extracting soil information, interpreting soil texture, and spatially aggregating SAC-SMA parameters. The methodology is comprised of four components. The first and second components are SSURGO and NLCD preprocessors. The third component is a parameter generator producing SAC-SMA parameters for each soil survey area on an approximately 30-m grid mesh. The last component is a postprocessor creating parameters for user-specified areas of interest on the Hydrologic Rainfall Analysis Project (HRAP) grid. Imple-

*Phone: 301-713-0640 Fax: 301-789-2875

Email address: yu.zhang@noaa.gov (**Yu Zhang**)

[1]National Oceanic and Atmospheric Administration, National Weather Service , Office of Hydrologic Development, Silver Spring, MD, 20910

[2]University Corporation for Atmospheric Research, Boulder, CO 80307

mented in open-source software, this approach was employed in creating a set of SAC-SMA parameter and related soil property grids spanning 25 states, wherein it was shown to greatly reduce the derivation time and meanwhile yield results comparable to those based on State Soil Geographic Database (STATSGO). The broad applicability of the methodologies and associated intermediate products to hydrologic modeling is discussed.
Keywords: hydrology, model, parameter, SSURGO, SAC-SMA, NLCD

1. Introduction

Distributed modeling of hydrologic processes, by accounting for spatial variations in meteorological forcing and land surface properties, holds the promise of augmenting the hydrologic prediction capabilities currently provided by lumped modeling frameworks (Beven, 1985; Smith et al., 2004). Recent years have witnessed a growing number of efforts in developing and testing distributed modeling frameworks (Woolhiser et al., 1990; Arnold et al., 1993; Ogden et al., 2004). Some of the frameworks were comparatively evaluated in the Distributed Model Inter-comparison Project (DMIP; Smith et al. (2004)). A common issue confronting the implementation of various distributed models concerns determining an initial set of spatially distributed parameters commonly referred to as a priori parameters (Koren et al., 2003), from physiographic properties such as soil and land cover (see Di Luzio and Arnold (2004), Bandaragoda et al. (2004), and Anderson et al. (2006) for related model needs). These a priori parameters can serve to facilitate model calibration and evaluation(Kuzmin et al., 2008).

Heretofore, advances in remote sensing and measurement technologies have brought increasingly refined soil and land cover information. Such information, when properly used, can potentially improve the physical representation of a priori model parameters and thereby expedite calibration and deployment of distributed models. The challenge, however, is that the growing data volume and complexity requires coherent, reproducible, and efficient a priori parameter estimation methods and tools. This paper presents a suite of such products conceived for a specific hydrologic model but nevertheless having potential applicability to a broad range of models. The model of

interest is the distributed model developed by Office of Hydrologic Development (OHD) of the National Weather Service (NWS). This model, hereafter referred to as the *NWS-DHM* or simply the *DHM*, is aimed at providing enhanced river forecasts and other products and services to the nation (Carter, 2002). Interested readers are referred to Koren et al. (2003, 2004) for a complete description of the NWS-DHM and to Reed et al. (2007) for flash flood applications of this model. In the DMIP (Smith et al., 2004), NWS-DHM performed favorably against the NWS lumped model and other distributed models in several catchments (Reed et al., 2004).

At present, the water balance component of DHM is the Sacramento Soil Moisture Accounting model (SAC-SMA; Burnash et al. (1973) and Burnash (1995)). Koren et al. (2000) laid out an a priori estimation framework in which the values of 11 SAC-SMA parameters were related to observable soil properties. This framework was employed in deriving a priori parameters from a gridded version of State Soil Geographic Database (STATSGO; Miller and White (1998)), and the results were used in the DMIP. More recently, Anderson et al. (2006) developed a set of procedures utilizing this framework to derive SAC-SMA parameters from the Soil Survey Geographic database (SSURGO[3]) and the National Land Cover Dataset (NLCD[4]), two public domain data sources that offer by far the most detailed soil and land cover characterizations on a national scale.

The approach described by Anderson et al. (2006) offers consistent ways of aggregating soil properties over multiple soil texture components and hori-

[3] `http://grass.osgeo.org/`
[4] `http://www.mrlc.gov`

zons in the SSURGO data. In past hydrologic simulation experiments, the parameters derived via this approach have been shown to outperform those based on STATSGO/NLCD in terms of simulation accuracy for a number of catchments (Zhang et al., 2006). Yet, at least three major limitations exist for this approach which to a great extent hinder its wider application. Firstly, the said approach relies exclusively on the user to manually parse and interpret texture. This can be a daunting and error-prone task given the wide variation in SSURGO soil texture types and encoding, and the fact that a generic texture code may correspond to multiple texture descriptions depending on the locations. Secondly, the said approach lacks a coherent way of aggregating parameters over larger areas (i.e., at a watershed or a region scale) from multiple survey areas. Thirdly, most of the procedures were done via graphical user interface (GUI). The sheer volume of the SSURGO and NLCD data makes it difficult, if not impossible, to apply these GUI-based procedures to a regional scale (i.e., for one or more states). In addition to these, a minor, but practically important issue is that the application of this approach is predicated on the access to proprietary software, such as ESRI-Arcview and Microsoft-Access. In light of these limitations, an enhanced approach was developed. This approach retains the basic mechanisms of Anderson et al. (2006) but is based on systematic, and consistent underlying methodologies that are applicable to most, if not all, geographic settings. Among the notable features of the new approach are automated preprocessors for the SSURGO and NLCD data, a standardized method of interpreting soil texture, and a coherent method for spatially aggregating soil properties and parameters. These enhancements allow a user to easily reproduce the

results obtained by others. Implemented exclusively via open source software packages, this new approach can be implemented and tested by users both within and outside NWS without the restriction of commercial software licensing requirements [5].

In this paper we describe this enhanced approach and demonstrate its efficacy by applying it without any modification to 25 states that encompass the geographic domains of six NWS River Forecast Centers (RFCs). The creation of the end product, a set of mosaicked parameters spanning 25 states (referred to as 25-state parameter mosaics hereafter), offers the opportunity to comprehensively assess the potential hydrologic benefits of incorporating the high resolution SSURGO and NLCD data sets. While the approach was intended for the NWS DHM, its methodologies, as well as some of the intermediate products, are broadly applicable to other hydrologic models whose parameters can be derived from the SSURGO, NLCD, or a combination of both. Examples of such models include the Soil and Water Assessment Tool (SWAT; Wang and Melesse (2006)), Gridded Surface/Subsurface Hydrologic Analysis (GSSHA; Ogden et al. (2004)), TOPNET (network version of the TOP model; (Bandaragoda et al., 2004)), and the TIN-based Real-time Integrated Basin Simulator (tRIBS; Ivanov et al. (2004)). The constituent methodologies for digesting and infusing the SSURGO and NLCD data complement the more generic parameter estimation interface provided by some of the afore-mentioned models (see, e.g., Di Luzio et al. (2002) and Ogden et al. (2004)).

[5]The US government does not specifically endorse any of the software packages mentioned herein.

The remainder of the paper is organized as follows: Section 2 reviews the a priori parameter estimation framework. Section 3 describes the SSURGO and NLCD data sets and the parameter derivation procedures. Section 4 examines the resulting parameter grids and Section 5 summarizes the work.

2. The a priori Parameter Estimation Framework

The formulation of the SAC-SMA model was originally presented in Burnash et al. (1973). In the model, soil horizons are generalized into a relatively thin *upper zone* (UZ) and a *lower zone* (LZ), with water stored in each zone further partitioned into *free water* that drains by gravity and *tension water* held by the suction of the soil matrix. The free water storage of the lower zone is subdivided into *supplemental* and *primary storages*, which account for faster and slower draining groundwater flows, respectively. *Percolation* is allowed from the upper to the lower zone. During a rainfall event, the runoff rate is determined jointly by rainfall, UZ storages and the percolation rates (the last is a function of UZ free water saturation and LZ saturation; see Koren et al. (2003, 2004), Anderson et al. (2006) and references therein).

The framework of Koren et al. (2000) offers a means for estimating 11 soil-related SAC-SMA parameters (Table 1) from soil and land cover data. The soil data set provides three sources of information 1) hydrologic soil group (HSG), 2) texture class and 3) vertical soil profile. In Anderson et al. (2006), HSG is used jointly with land cover to determine the Soil Conservation Service (SCS, now Natural Resource Conservation Service, or NRCS) curve number (CN), which is needed for estimating initial rainfall abstraction (I_a). Soil texture is used to determine hydrologic soil group (HSG) that

would form the basis for estimating SAC-SMA parameters.

The curve number is assigned on the basis of HSG and land cover type in accordance to the empirical relations published in Chapter 9 of NRCS-ARS (2010) assuming dry antecedent condition (which has been shown to yield more accurate streamflow simulations). There are four primary HSGs and each corresponds to an estimated runoff potential, i.e., A (low), B (moderate), C (moderately high) and D (high) (see Chapter 7 of NRCS-ARS (2010)). Besides the four classes, there are also dual classifications (e.g., "A/D", "B/D" and "C/D"). It is assumed here that a) it is the variation in soil moisture level that results in the pairwise difference in drainage characteristics (say, between "**A**" and "**A**/D"), and b) such a difference is automatically accounted for by soil moisture states generated through NWS-DHM simulations. Under these assumptions, no additional distinction needs to be made in the curve number between two HSGs sharing the same primary designator, e.g., "**A**" and "**A**/D". This is reflected in the curve number assignment in Table 2.

The curve number thus estimated is then used to derive the thickness of the upper zone Z_{up}. The method of estimation developed by Koren et al. (2000) is based on the NRCS approach documented in McCuen (1982). Essentially, the curve number determines the abstraction I_a for a soil column via the formula $I_a = 5.08(CN/1000 - 10)$ (Chap. 10 of NRCS-ARS (2010)). I_a is assumed to be 20% of additional water needed to saturate a soil column initially at its field capacity. Therefore, Z_{up} is uniquely determined by I_a, soil porosity θ_s and field capacity θ_{fld}. For a multi-horizon soil column, Z_{up} needs to satisfy the following condition:

$$I_a = \int_0^{Z_{up}} (\theta_s(z) - \theta_{fld}(z))dz \tag{1}$$

where z is the depth from surface. The schematic of the estimation is shown in Figure 2. In practice, one adds up the free water storage of each horizon iteratively until the sum equals or exceeds the initial abstraction I_a (Fig. 2). In the latter case, the horizon where I_a is exceeded is split in such a way that allows the free water storage of the upper zone to equal I_a. The lower zone extends from this depth to the depth of bedrock, or to the upper edge of an impermeable layer when such a layer exists above the bedrock (Fig. 2). The 11 SAC-SMA parameters can be computed given the soil properties for the upper and lower zones which include porosity (θ_s), field capacity (θ_{fld}), wilting point (θ_{wp}) and saturated hydraulic conductivity (K_s). (see Appendix A, and Koren et al. (2000) for additional details). Among these, three parameters, namely, UZTWM, UZFWM, LZTWM can be computed directly by adding corresponding quantities of each soil horizon, whereas estimating the rest requires vertically averaged soil properties (Appendix B).

3. Methodology and Implementation

The enhanced SAC-SMA parameter estimation approach, similar to its predecessor by Anderson et al. (2006), and those developed for models such as GSSHA (Downer and Ogden, 2006) and SWAT (Di Luzio et al., 2002), employs Geographic Information System (GIS) for mapping and infusing soil and land cover data. In this approach, parameter derivation is performed in three phases (see schematic in Fig 1). The first phase entails preprocessing the SSURGO and NLCD, extracting information relevant to parameter

derivation, and saving the results in formats usable in later phases. This phase is implemented in two software components, i.e, SSURGO and NLCD *preprocessors* (Fig 1). The second phase entails generating parameters for each soil survey area, and the corresponding software component is referred to as *parameter generator* (Fig 1). The third and final phase entails post-processing the 30-m parameter grids to yield mosaicked, gridded data sets on the Hydrologic Rainfall Analysis Project (HRAP; Greene and Hudlow (1982), Reed and Maidment (1999)) grid mesh, and the associated software component is termed *postprocessor* (Fig. 1).

3.1. SSURGO Preprocessor

The project for creating the Soil Survey Geographic Database was undertaken by the NRCS to provide digitized soil maps for the entire United States at resolutions considerably higher than those for STATSGO (mapping scale from 1:12,000 to 1:63,360 for SSURGO compared to approximately 1:250,000 for STATSGO). The project was originally set to complete in 2008, and at present SSURGO maps are available for the majority of counties for most of the states[6].

Figure 3 depicts the hierarchical structure of the SSURGO database. The data sets are organized by *survey area*s, where a survey area is usually equivalent to a county. Within each survey area, soil patches sharing similar characteristics are lumped into one *map unit* (Figs. 3 and 4). Each map unit is uniquely identified using *map unit symbol* (MUSYM). Within each map unit are multiple *soil components* with varying percentage areal coverage

[6] `http://soildatamart.nrcs.usda.gov/StatusMaps/SoilDataAvailabilityMap.pdf`

(Fig. 3). Each soil component encompasses one or multiple *horizons*, and at each horizon one or multiple soil texture classes may be present (Fig. 3). At a given horizon, the soil element corresponding to a unique texture class is hereafter referred to as a *subcomponent*. For each survey area, the locations of map units (represented as polygons) are provided in spatial files, and the soil texture and associated properties at various depths are provided in tabular files.

The SSURGO preprocessor performs the following tasks. In a preliminary step, it standardizes tabular file names and removes irregularities. In the second step, it delineates the effective depth of the soil column, filters out the dominant component for subsequent curve number computation, and extracts subcomponent level physical properties for each polygon. In the third step, it computes the horizon-averaged physical properties, and augments the attribute of each polygon by adding the hydrologic property. The details of each step follow below.

As opposed to the previous approach of Anderson et al. (2006), which involves using a Microsoft-Access interface provided by NRCS, the first step of the present approach employs an automated procedure for directly accessing the SSURGO spatial and tabular data. Due to the observation that the tabular headings are often inconsistent with the standard ones documented in the metadata[7], the former are automatically modified wherever such an inconsistency arises. As an example, in some survey areas texture group information is stored in a file incorrectly named "chtexgrp.txt", this file is

[7] `http://soildatamart.nrcs.usda.gov/documents/SSURGOMetadataTableColumns.pdf`

renamed to "chtexturegrp.txt" according to SSURGO metadata.

In the second step, some simplifications are made to digest the complex, multi-component, multi-horizon structure of soil texture within a map unit. The enhanced approach follows the strategy devised by Anderson et al. (2006), which was shown to produce parameter values that led to improved simulation results (Anderson et al., 2006). This strategy consists of the following two elements (see illustration in Fig. 6). First, for a map unit, whenever multiple texture components are present in a map unit, only the dominant ones, i.e., those associated with the highest percentage areal coverage, are used. Second, in determining variables such as depth, only the representative values are used (SSURGO provides the lower, representative and upper values). These simplifications vastly reduce the complexity of the soil textures. The effective depth of a soil column (i.e., the depth to SAC-SMA lower zone, or Z_{max}) can then be delineated as the depth to the bedrock (or to a restrictive layer, e.g., dense glacial till and fragipan, if such a layer is present) in the dominant texture component. In cases where the depth to restrictive layer is zero, the parameters are left as "missing" (the gaps can be filled by interpolation). The hydrologic property would be the one for the dominant component.

Physical properties are determined on the basis of soil texture. Though SSURGO contains fields of measured values for quantities such as θ_s and K_s, such fields are often left unpopulated (also noted in Anderson et al. (2006)). As a result, the mapping from texture to physical property is performed via a table (Table 5) constructed from empirical relationships documented in Clapp and Hornberger (1978) and Cosby et al. (1984); associated details

can be found in Anderson et al. (2006)). Considerable difficulty, however, remains in simplifying complex SSURGO soil textures into the categories given in Table 5. In the prior approach, classification was done entirely by the user, who would first interpret the soil texture for a given survey area, and then create a lookup table manually to relate the local texture classes to standard ones (see Table 5 for the latter). The need of subjective decisions from the user, along with the labor-intensiveness of table creation, hinder the application of such procedures to regional and national scale implementation. The new approach employs an automatic soil identification algorithm that addresses these shortcomings. As illustrated in Figure 7, this algorithm relies on two SSURGO fields as the basis for interpreting texture, namely "TEXTURE" and "MUNAME", with the former as the primary and the latter the secondary reference. To elaborate, for each subcomponent, the algorithm first examines the value of the field "TEXTURE". If the value points unambiguously to a known texture class, then the simplified texture class is assigned accordingly (Scenario A in Fig. 7). On the other hand, in some cases the "TEXTURE" field contains only a generic descriptor (such as "SPM", or slightly decomposed plant material) and meanwhile the description field "MUNAME" provides an identifiable texture name (Scenario B of Fig. 7). In these situations the latter would be used in lieu of the former to determine the simplified texture (Scenario B in Fig. 7). If neither field provides the needed information, a symbol of "O", which represents "Other", is assigned (Scenario C in Fig. 7). This objective procedure removes the ambiguity surrounding texture determination and renders the procedure reproducible.

In the third step, as in Anderson et al. (2006), soil properties are arithmetically averaged for each horizon by combining the embedded subcomponents (Fig. 6). To simplify the process, for a given horizon only those subcomponents with identifiable texture are used. Also in this step, the hydrologic property for each map unit is added to the original SSURGO attribute table, and the augmented table will be used in the curve number computation.

The SSURGO preprocessor is implemented primarily in a set of scripts written in *R*, an open source statistical package (Becker et al., 1988). The regularization of tabular heading is done via perl scripts and automation of the process is realized using a collection of shell scripts. The data flow diagram is shown in Figure 5 and the detailed functionalities are provided in Table 4. For each map unit, this implementation takes SSURGO tabular data and attribute table from the shapefile as inputs, automatically performs the tasks as mentioned above, and produces the horizon-averaged soil properties and an augmented attribute table containing HSG information (Fig5) that will be used in the curve number computation.

3.2. NLCD Preprocessor

The National Land Cover Database (NLCD) was created from Land Remote-Sensing Satellite (Landsat) images by the Multi-Resolution Land Characteristics (MRLC) Consortium [8]. Two versions are currently available: the NLCD 1992 and NLCD 2001. Methodologies for creating the data sets can be found in Vogelmann et al. (2001) and Homer et al. (2004), respec-

[8] `http://www.mrlc.gov`

tively. In both versions the land cover is represented on approximately 30-m grids in the coordinates of Albers Equal Area (AEA). The NLCD 1992 files are partitioned along state boundaries (California and Texas are sub-divided into multiple sub-regions, and data for Alaska is unavailable). The NLCD 2001, by contrast, is organized by zones (data available for Alaska). There are 14 super-zones for the conterminous United States (CONUS).

The NLCD employs a modified version of the Anderson land-use and land-cover classification system (Anderson et al., 1972) that contains nine broad categories[9]. The sub-categories differ slightly between the NLCD 1992 and 2001 under Shrubland, Herbaceous Upland Natural/Semi-Natural Vegetation, and Wetlands (Table 2). To account for these differences, NWS developed separate methods for each version of the NLCD. Only the one for the NLCD 2001 is described here, as the latter one provides more recent land cover information and is deemed to be of closer relevance to the forecasting missions of NWS.

Preprocessing of the NLCD is done in three steps (Fig. 8). In the first step the raw NLCD 2001 data sets are ingested, wherein any value beyond the valid range (1-99) is set to null (Fig. 8). Then zonal NLCD 2001 data sets are mosaicked and divided along state boundaries (Fig. 8). The NLCD data for each state is subsequently reprojected into geographic coordinates to match the SSURGO data (Fig. 8). The NLCD preprocessor is implemented in a series of SHELL scripts that uses *Geographic Resources Analysis Support System* (GRASS; [10]) as the GIS platform. The descriptions of the scripts can

[9]see details at `http://landcover.usgs.gov/classes.php/#similar`
[10]`http://grass.osgeo.org/`

be found in Table 6.

3.3. Parameter Generator

The parameter generator takes the outputs from the SSURGO and NLCD preprocessors and produces 30-m SAC-SMA parameters and associated zonal averaged soil properties. While the algorithms are mostly derived from the previous approach (Anderson et al., 2006), the processes were considerably improved and automated. The generator consists of the following steps (illustrated in Fig. 9). First, SSURGO spatial files are ingested into a GIS framework, wherein the HSG data from the SSURGO Preprocessor are rasterized to match the NLCD grid (Fig. 9). Then the curve number is computed by coupling HSG and the NLCD land cover (Fig. 9). Subsequently, SAC-SMA parameters are computed for each polygon based on HSG and polygon-averaged curve number (Fig. 9). In the end, polygon-based parameters and properties are rasterized to 30-m grids (Fig. 9). The details of the curve number and parameter generation are provided below.

The curve number is computed for each 30-m cell following the algorithm outlined earlier in the paper on the basis of the NLCD land cover and collocated HSG. The curve number is then used jointly with the horizon-averaged soil properties to compute the upper zone depth (Z_{up}). Note that Z_{up} computed following Eqn. (1) could be zero or beyond Z_{max}. In these situations, the parameters are left missing. Otherwise, upper and lower zone SAC-SMA parameters can then be computed via the equations provided in Koren et al. (2000) (reproduced in Appendix A). To facilitate future corroboration and modification of the parameters, a suite of 30-m byproducts are also generated in this step. These data sets include zonal averaged values of porosity (θ_s^u and

θ_s^l, where u and l denote upper and lower zones, respectively), field capacity (θ_{fld}^u and θ_{fld}^l), wilting point (θ_{wp}^u and θ_{wp}^l) and saturated hydraulic conductivity (K_s^u and K_s^l) for upper and lower zones. Each quantity is computed by vertically averaging the respective values within the designated zone. For example, the upper zone averaged saturated hydraulic conductivity, K_s^u, is simply:

$$K_s^u = \frac{1}{Z_{up}} \int_0^{Z_{up}} K_s(z)dz \tag{2}$$

An additional auxiliary data set created in this process is the HRAP coordinates for each 30-m cell, which will be used subsequently in creating coarser resolution parameters.

The parameter generator is wrapped in a set of GRASS/SHELL scripts, which rely on embedded R scripts for parameter computations. The scripts and their descriptions can be found in Table 7.

Postprocessor

The postprocessor combines 30-m parameter grids for all survey areas of interest and produces SAC-SMA parameter values on 1/4, 1/2 and full HRAP grids over one or multiple states. A difficulty in creating these data sets, which the previous approach (Anderson et al., 2006) was not able to address, arises from the situations where a pixel along survey area boundaries overlaps with multiple survey areas. For such a pixel, survey-area level parameter estimation results in multiple pixel-mean values, each based solely on the values within one survey area. These pixel-mean values need to be combined according to the degree of overlap between the pixel and survey

areas as well as the number of missing values in the sub-pixel cells. The present approach addresses this issue via a three-step process. In the first step, all embedded 30-m cells are identified for each 1/4 HRAP pixel within a state (or a part of a state), then the valid parameter values (i.e., not missing) for these cells, which can be located within one or multiple survey areas, are arithmetically averaged. The number of of embedded 30-m cells with valid values is recorded in an auxiliary grid (also on 1/4 HRAP scale). In the second step, this auxiliary grid is used jointly with corresponding parameter grids for designated states in creating the multi-state mosaics, with the former serving as the weights in averaging the latter, i.e.,

$$\bar{\chi}_{i,j} = \frac{\sum_k n_{i,j}^k \chi_{i,j}^k}{\sum_k n_{i,j}^k} \tag{3}$$

where $\chi_{i,j}^k$ and $n_{i,j}^k$ denote the parameter value and the number of valid 30-m cells within each 1/4 HRAP pixel (i, j) over a survey area k, and $\bar{\chi}_{i,j}$ is the weighted average across multiple survey areas.

In the third step, the 1/4 HRAP based parameters are aggregated onto 1/2 and full HRAP grids, wherein the number of embedded cells within each embedded pixels are once again used as the weight (the number of sub-pixel cells varies depending on the number of missing values).

The postprocessor is implemented in a set of programs written in C++ and SHELL scripts. Their descriptions and the associated data flow diagram are provided in Table 8 and Figure 10, respectively.

4. Derivation of a Multi-State a priori Parameter Set

The automated approach was employed in deriving a multi-state parameter product for an area that covers the geographic domains of six RFCs, i.e., CNRFC (California-Nevada), CBRFC (Colorado Basin), WGRFC (West Gulf), ABRFC (Arkansas-Red Basin), LMRFC (Lower Mississippi) and SERFC (Southeast). This region encompasses 25 states and a total of 1713 survey areas.

For this effort, the NLCD 2001 data for 14 super-zones were obtained and processed. The preprocessors and the parameter generator were initially developed and tested for a few survey areas in Missouri, and were then applied without modifications to other states. In the testing phase, data from the SSURGO and NLCD 1992 data were used to produce parameter value via both the manual and automated approach and the results were compared (note that the manual procedures are only directly applicable to NLCD 1992 data). The polygon-specific results were most identical, and the only differences were owing to a change in the delineation of the depth to the bottom of the lower zone (the earlier Avenue script uses hzdepb_r and resdept_r from the secondary component when either of these variables (not both) is missing in the primary component, whereas the preprocessor in the automated approach uses the non-missing value from the primary component in order to be consistent with the documentation).

Preliminary comparisons were performed between the parameter values derived using the SSURGO and NLCD 2001 data against those based on

STATSGO soil data and Global Land Cover Characterization (GLCC[11] data. Such comparisons indicate that the parameters derived via the two means (SSURGO + NCLD 2001 vs. STATSGO + GLCC) exhibit similar geographic features. Such features include low UZTWM values across central and southern Florida where sandy soils have a relatively low capacity to hold water in tension (Fig. 12). For UZFWM and UZTWM, the values from the two sources are broadly comparable, with the ratio of the SSURGO to the STATSGO means close to 1 (0.99 for UZFWM and 0.96 for UZTWM). The UZFWM and UZTWM values from the two sources are also considerably correlated (0.59 and 0.56, respectively; Figure 13). Yet, considerable differences are also evident (Figs. 12 and 13). These differences can be attributed to a combination of the following factors. First, the SSURGO and STATSGO-based soil texture may differ substantially. Our analysis indicates that, for some areas (e.g., northern Oklahoma), these differences were a major contributing factor to large discrepancies in the UZTWM and UZFWM values. Second, the NLCD and GLCC land cover may differ due to differences in sensor, resolution and classification methods. The land cover difference translate into those in curve number and parameter values. Third, the parameter derivation method is nonlinear. To elaborate, the average of parameter values computed individually over sub-areas using distinct curve number and soil properties does not necessarily equal the parameter value computed using the averaged curve number and soil properties over the entire area. Additional investigations on the impacts of these differences on hydrologic simulations are underway, which would shed light on whether and

[11] `http://edc2.usgs.gov/glcc`

how the use of SSURGO would lead to improvements to the accuracy in streamflow predictions.

The automation built in each of the components (preprocessors, parameter generator, and postprocessor) led to a sharp reduction of processing time. The absence of licensing restrictions on ESRI-Arcview and Microsoft Access further improved efficiency by enabling simultaneous parameter derivation for geographic areas on multiple workstations. In deriving the 25-state mosaic, the entire process of parameter derivation took about a week to complete when running on three 32-bit linux workstations (not dedicated to this task). Since the previous approach developed by Anderson et al. (2006) required manual intervention, lacked precise mechanism for inter-state mosaicing, and relied on multiple platforms, a precise comparison of the performance of the two approaches was not feasible. Nevertheless, past experience suggests that excluding the time for downloading SSURGO and preprocessing the NLCD data, on the average about six hours were required for deriving the HRAP-based product for a single survey area. By contrast, applying the present approach for deriving the HRAP-based product for 92 survey areas took only about 0.24 hour for each survey area. This translates into a 96% reduction in processing time. A break-down of the approximate time needed in each component is shown in Table 9.

5. Discussions

One of the requisites of implementing spatially distributed hydrologic model is a spatially varying a priori parameters that are representative of land surface hydrologic properties. Despite the diversity of current modeling

frameworks, soil and land cover remain two of the most important sources of observations that modelers rely on in deriving these a priori parameters. Yet, the volume and complexity of these data hindered the derivation of parameters over regional and national domains, which in turn delayed comprehensive evaluations of the parameters and refinement of the parameter derivation methods.

This paper documents an enhanced and automated a priori approach that allows users to fast and reliably digest the SSURGO and NLCD data sets, and derive parameters for SAC-SMA that are used in operational hydrologic forecast on a national scale. Though based on previously established methods, this approach replaces the subjective and ad hoc elements of the previous approach by Anderson et al. (2006) with standardized and automated procedures, and offers new algorithmic features such as automatic texture identification and coherent, sequential spatial aggregation. An important outcome of the standardization and automation is reproducibility. The first aspect of the reproducibility is that one, by applying the software tools according the documented procedures, is able to reproduce parameters derived by others. The second aspect is that the software tools can be applied unchanged to other geographic settings. In the derivation of the 25-state parameter mosaics, the algorithms were developed at two counties in Missouri but were applied without alteration to 1713 survey areas over 25 states in the Coterminous US. Preliminary assessments indicated that the newly derived parameter values are quite reasonable, as these values are closely correlated with parameters based on the STATSGO/NLCD. Another important outcome is the vastly improved efficiency. The creation of the

25-state parameter mosaics took only about 4% of the time that it used to take using the manual approach.

Though developed for a specific model, i.e., SAC-SMA, a large portion of the automated approach is directly applicable to similar parameter estimation tasks for other models. While guidance and tools for soil and land cover based parameter estimation have been available for some of the models (see, e.g., Di Luzio et al. (2002) and Downer and Ogden (2006)), few provided specifics of digesting and infusing the SSURGO and NLCD. In cases where such specifics were provided Anderson et al. (2006); NRCS (2008), the methodologies consisted of mostly manual procedures. The present approach, by offering efficient, geographic invariant ways of processing the SSURGO and NLCD data, complements the strength of existing methodologies. In fact, some of the products of the SSURGO and NLCD preprocessors, such as porosity and hydraulic conductivity (computed for each horizon survey polygon and for each horizon), can be ingested in models such as GSSHA (provided the soil layers is similarly delineated). Another useful product is the curve number, a common input to a variety of curve-number based models, including TR-20 (NRCS, 1983), Storm Water Management Model (SWMM, Heaney et al. (1975)), and SWAT (Arnold et al., 1993). The ingestion of some of the intermediate products (e.g., hydraulic conductivity and curve number) in other models would permit a closer examination of how structural differences of models affect their relative performance. Furthermore, these data sets also offer modelers an opportunity to address the scale-dependent impacts of infusing fine-scale soil and land cover on the accuracy of hydrologic model simulations (relative to, say, those based on STATSGO).

Such an effort is, in fact, already underway at the NWS.

On the software side, our implementation is based on open-source software packages, and the scripts and source codes can be easily modified by a user. Though tested only on a linux system, these programs can be ported to any platform where the underlying packages can be installed (R, GRASS and Perl are available on windows, Mac OS, so are the compilers). The modularity of the design allows it to be tailored to suit the particular needs of a modeler. As an example, one can retain the Microsoft-Access interface provided by NRCS for extracting soil information and feed it to the rest of the SSURGO preprocessor. In areas where either the SSURGO or the NLCD is not available, one can compute the curve number externally using local data, and ingest the results in the Parameter Generator.

Despite the afore-mentioned promises, the current approach is not without its own limitations. Soil survey results, after all, consist of subjective elements which often require human expertise to properly interpret and digest. The simplification of SSURGO data documented in the current approach, while necessary, may leave out information useful for determining soil properties and the associated parameters. As an example, the effective soil column is assumed to end at the restricted layer, though such a layer itself might be highly permeable. Similarly, minor texture components, which are eliminated in the soil preprocessor, may play a determining role in shaping the aggregate soil properties of a map unit. In addition, the current approach utilizes solely laboratory based soil properties as a first cut solution, though the values provided in SSURGO, when they do exist, may be more accurate representations of polygon-specific soil properties. Moreover, for simplic-

ity arithmetic averaging was used in computing multi-horizon effective K_s, though harmonic average in theory might be more appropriate. In view of these shortcomings, additional investigations have been proposed, which include, but are not limited to: a) experimenting retaining the minor texture components in computing soil properties, b) exploring additional sources of information for refining the texture mapping, c) integrating the available in situ soil properties given in SSURGO and examining the effects on parameter values and hydrologic simulations; and d) estimating the uncertainty of parameters arising from the simplifications.

6. Summary

This paper presents an enhanced and automated approach for deriving a priori parameters for the NWS distributed hydrologic model from Soil Survey Geographic database and National Land Cover Dataset. The approach was implemented entirely in open source software packages, notably R and GRASS. It consists of four elements: i) SSURGO preprocessor; ii) NLCD preprocessor; iii) parameter generator and iv) parameter postprocessor. These elements offer systematic and reproducible ways of acquiring, processing and computing parameters. In deriving a set of 25-state a priori parameter grids, this approach was shown to substantially reduce the time for parameter estimation (to 4% of the time previously needed). The methodologies and the associated software, in particular those for deriving curve number, identifying soil texture, and spatially aggregating parameter values, can be adopted to facilitate the implementation of other hydrologic models that can utilize the SSURGO and NLCD data.

Acknowledgment

We are indebted to Thomas Adams at Ohio River Basin Forecast Center (OHRFC) who contributed sample GRASS scripts for this project and provided many insightful comments to the manuscript. We would also like to acknowledge James Wickham at MRLC, Charles Larson at USGS, and Sam Brown at USDA for their generous support.

Appendix A

Zonal-averages of soil property χ for upper and lower zones (denoted by χ^u and χ^l respectively) are defined as follows:

$$\chi^u = \int_0^{Z_{up}} \chi(z)dz \tag{A.1}$$

and

$$\chi^l = \int_{Z_{up}}^{Z_{max}} \chi(z)dz \tag{A.2}$$

where Z_{up} and Z_{max} are the depths to the bottom of the upper and lower zones, respectively. In these equations,χ can be porosity θ_s, field capacity θ_{fld}, wilting point θ_{wp}, and saturated hydraulic conductivity K_{sat}. The SAC parameters for the upper and lower zones can be defined correspondingly (the definitions of these parameters can be found in Koren et al. (2000)).

$$UZTWM = \int_0^{Z_{up}} \theta_{fld}(z) - \theta_{wp}(z)dz \tag{A.3}$$

$$UZFWM = \int_0^{Z_{up}} \theta_s(z) - \theta_{fld}(z)dz \tag{A.4}$$

$$UZK = 1 - \left(\frac{\theta^u_{wp}}{\theta^u_s}\right)^n \tag{A.5}$$

$$LZTWM = \int_{Z_{up}}^{Z_{max}} \theta_{fld}(z) - \theta_{wp}(z)dz \tag{A.6}$$

$$LZFSM = \left(\frac{\theta^l_{wp}}{\theta^l_s}\right)^n \int_{Z_{up}}^{Z_{max}} \theta_s(z) - \theta_{fld}(z)dz \tag{A.7}$$

$$LZFPM = \left[1 - \left(\frac{\theta^l_{wp}}{\theta^l_s}\right)^n\right] \int_{Z_{up}}^{Z_{max}} \theta_s(z) - \theta_{fld}(z)dz \tag{A.8}$$

$$LZSK = \frac{UZK}{1 + 2(1 - \theta^l_{wp})} \tag{A.9}$$

$$LZPK = 1 - e^{-\frac{1}{\mu}(1+\beta)\pi^2 K_s D_s^2 (Z_{max} - Z_{up})\delta t} \tag{A.10}$$

$$PFREE = \left(\frac{\theta^l_{wp}}{\theta^l_s}\right)^n \tag{A.11}$$

$$REXP = \left(\frac{\theta^l_{wp}}{\theta_{wp,sand} - 0.001)}\right)^{1/2} \tag{A.12}$$

where μ is defined as follows:

$$\mu = 3.5(\theta^l_s - \theta^l_{fld})^{1.66} \tag{A.13}$$

$$ZPERC = \frac{LZTWM + LZFSM(1 - LZSK) + LZFPM(1 - LZPK)}{LZFSM\ LZSK + LZFPM\ LZPK} \tag{A.14}$$

Appendix B

The electronic supplement is made up of an instruction file and five archives: nlcd_proc.tgz, param_ssurgo.tgz, grass_prog.tgz, and util_prog.tgz, and grass_dir.tgz. The first archive contains the scripts needed for processing NLCD data. The second archive contains the column title table and the scripts for SSURGO preprocessor, parameter generator and postprocessor. The third and fourth archives provide the GRASS extensions and utility programs for postprocessing, respectively. The fifth archive provides an example of GRASS directory structure and the GRASS projection files for NLCD, SSURGO, and parameter products.

References

Anderson, J. R., Hardy, E. E., Roach, J. T., 1972. A land-use classification system for use with remote-sensor data. U. S. Geological Survey Circular 671, 16 pp.

Anderson, R. M., Koren, V., Reed, S., 2006. Using SSURGO data to improve Sacramento model a priori parameter estimates. Journal of Hydrology 320, 103–116.

Arnold, J. G., Allen, P. M., Bernhardt, G., 1993. A comprehensive surface-groundwater flow model. Journal of Hydrology 142, 47–69.

Bandaragoda, C., Tarboton, D. G., Woods, R., 2004. Application of TOPNET in the distributed model intercomparison project. Journal of Hydrology 298, 178–201.

Becker, R. A., Chambers, J. M., Wilks, A. R., 1988. The New S Language. Chapman & Hall, London, 702 pp.

Beven, K., 1985. Distributed models. In: Anderson, M. G., Burt, T. P. (Eds.), Hydrological Forecasting. Wiley, New York, NY, pp. 405–435.

Burnash, R. J. C., 1995. The NWS river forecast system – catchment modeling. In: Singh, V. P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Littleton, Colorado, pp. 311–366.

Burnash, R. J. C., Ferral, R. L., McGuire, R. A., 1973. A generalized streamflow simulation system - conceptual modeling for digital computers. Technical report, United States Department of Commerce, National Weather

Service and State of California, Department of Water Resources, Sacramento, 204 pp.

Carter, G. C., 2002. Infusing new science into the national weather service river forecast system. In: Proceedings of 2nd Federal Interagency Hydrologic Modeling Conference. Las Vegas, NV, 10 pp.

Clapp, R. B., Hornberger, G. M., 1978. Empirical equations for some soil hydraulic properties. Water Resources Research 14 (4), 601–604.

Cosby, B. J., Hornberger, G. M., Clapp, R. B., Ginn, T. R., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soil. Water Resources Research 20 (6), 682–690.

Di Luzio, M., Arnold, J. G., 2004. Formulation of a hybrid calibration approach for a physically based distributed model with NEXRAD data input. Journal of Hydrology 298, 136–154.

Di Luzio, M., Srinivasan, R., Arnold, J. G., Neitsch, S., 2002. ArcView Interface For SWAT2000: User's Guide, URL: `ftp://ftp.brc.tamus.edu/pub/swat/doc/swatav2000.pdf`.

Downer, C. W., Ogden, F. L., 2006. Gridded Surface Subsurface Hydrologic Analysis (GSSHA) User's Manual, Version 1.43 for Watershed Modeling System 6.1. Vicksburg, MS, 207 pp.

Greene, D., Hudlow, M., 1982. Hydrometeorologic grid mapping procedures. In: Proceedings of International Symposium on Hydrometeorology. American Water Resource Association, Denver, CO, 22 pp.

Heaney, J. P., Huber, W. C., Sheikh, H., Medina, M. A., 1975. Urban stormwater management modeling and decision-making. Technical Report EPA-670-2-75-022, United States Environmental Protection Agency, Washington, D.C., 185 pp.

Homer, C., Huang, C., Yang, L., Wylie, B., Coan, M., 2004. Development of a 2001 national land-cover database for the United States. Photogrammetric Engineering & Remote Sensing 70, 829–840.

Ivanov, V. Y., Vivoni, E. R., Bras, R. L., Entekhabi, D., 2004. Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: a fully-distributed physically-based approach. Journal of Hydrology 298, 80–111.

Koren, V., Reed, S. M., Smith, M., Zhang, Z., Seo, D.-J., 2004. Hydrology Laboratory Research Modeling System (HL-RMS) of the US National Weather Service. Journal of Hydrology 291, 297–318.

Koren, V., Smith, M., Duan, Q., 2003. Use of a priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall-runoff models. In: Duan, Q., Sorooshian, S., Gupta, H., Rosseau, H., Turcotte, R. (Eds.), Calibration of Watershed Models, Water Science and Applications. Vol. 6. American Geophysical Union, Washington, DC, pp. 239–254.

Koren, V. I., Smith, M., Wang, D., Zhang, Z., 2000. Use of soil property data in the derivation of conceptual rainfall-runoff model parameters. Proceedings of the 15th Conference on Hydrology, American Meteorological Society,, Long Beach, CA, pp. 103–106.

Kuzmin, V., Seo, D.-J., Koren, V., 2008. Fast and efficient optimization of hydrologic model parameters using a priori estimates and stepwise line search. Journal of Hydrology 353, 109–128.

McCuen, R. H., 1982. A Guide to Hydrologic Analysis Using SCS Methods. Prentice-Hall Inc., Englewood Cliffs, New Jersey, 145 pp.

Miller, D., White, R., 1998. A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. Earth Interactions 2, 1–26.

NRCS, 1983. Computer program for project formulation. Hydrology Unit, Technical Release 20. US Department of Agriculture, Soil Conservation Service, 294 pp.

NRCS, 2008. Creation of Johnson County SWAT SSURGO Soil Input Dataset: Using the AvSwat Model with SSURGO Data in Johnson County, Iowa. US Department of Agriculture, Soil Conservation Service, URL: `http://www.nriai.nrcs.usda.gov/technical/analysis/swatsurgo/creation.html`.

NRCS-ARS, 2010. Hydraulics and Hydrology - Technical References NRCS National Engineering Handbook. US Department of Agriculture, Natural Resources Conservation Service, URL: `http://directives.sc.egov.usda.gov/viewerFS.aspx?id=2572`.

Ogden, F. L., Downer, C. W., Meselhe, E., 2004. GSSHA: A model for simulating diverse streamflow generating processes. Journal of Hydrologic Engineering 9 (3), 161–174.

Reed, S. M., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., DMIP Participants, 2004. Overall distributed model intercomparison project results. Journal of Hydrology 298, 27–60.

Reed, S. M., Maidment, D. R., 1999. Coordinate transformations for using NEXRAD data in GIS-based hydrologic modeling. Journal of Hydrologic Engineering 4 (2), 174–182.

Reed, S. M., Schaake, J., Zhang, Z., 2007. A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. Journal of Hydrology 337, 402–420.

Smith, M. B., Seo, D.-J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F., Cong, S., 2004. The distributed model intercomparison project (DMIP): motivation and experiment design. Journal of Hydrology 298, 4–26.

Vogelmann, J., Howard, S., Yang, L., Larson, C., Wylie, B., Van Driel, J., 2001. Completion of the 1990s national land cover data set for the conterminous United States. Photogrammetric Engineering & Remote Sensing 67, 650–662.

Wang, X., 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.

Woolhiser, D. A., Smith, R. E., Goodrich, D. C., 1990. A kinematic runoff and erosion model: documentation and user manul. Technical Report ARS-

77, Agriculture Research Service, U.S. Department of Agriculture, Tucson, AZ, 130 pp.

Zhang, Z., Koren, V., Reed, S. M., Smith, M., Moreda, F., 2006. Comparison of simulation results using SSURGO-based and STATSGO-based parameters in distributed hydrologic modeling. 3rd Federal Interagency Hydrologic Modeling Conference, Reno, NV, 8 pp. URL: http://acwi.gov/hydrology/mtsconfwkshops/conf_proceedings/3rdFIHMC/8D_ZhangZi.pdf

List of Figures

1 Schematic of processing procedures. xiii

2 Schematic for method of estimating Z_{up}. xiv

3 Hierarchy of SSURGO data (uniquely defined by Map Unit Symbol, or MUSYM). xv

4 Spatial pattern of a map unit labeled "23B" (given by map unit symbol, or MUSYM). This map unit contains seven non-contiguous polygons. xvi

5 Data flow diagram for SSURGO Preprocessor. xvii

6 Simplification of texture components in SSURGO Preprocessor. For each map unit, only dominant component (associated with the highest percentage coverage) is retained. In this component, soil properties for various subcomponents within each horizon are arithmetically averaged. xviii

7 Illustration of texture mapper. xix

8 Data flow diagram for NLCD 2001 Preprocessor. xx

9 Data flow diagram for Parameter Generator. Among inputs, NLCD data is output from NLCD 2001 preprocessor (Fig. 8), whereas augmented attribute and mu_prop.dbf are outputs from SSURGO preprocessor (Fig. 5). xxi

10 Data flow diagram for Parameter Postprocessor. xxii

11 States for which SAC-SMA parameters were derived. Superimposed are geographic domains of six River Forecast Centers (RFCs), namely, California-Nevada (CNRFC), Colorado Basin (CBRFC), Arkansas Red River Basin(ABRFC), West Gulf (WGRFC), Lower Missisippi (LMRFC) and Southeast (SERFC). xxiii

12 Comparisons of 25-state composite a priori grids (HRAP resolution) of UZFWM (top) and UZTWM (bottom) based on SSURGO (left) and STATSGO (right). Left blank are areas where SSURGO data is currently unavailable, or no meaningful soil texture can be derived from SSURGO data, or curve number reaches 100 (impervious surface). xxiv

13 Comparisons of SSURGO and STATSGO-based UZFWM and UZTWM for areas shown in Figure 12. Superimposed are correlations (denoted by ρ). xxv

List of Tables

1 SAC-SMA Parameters xxvi

2 NLCD 2001 Classes and Curve Number xxvii

3 Highlights of Differences between Previous and Current Approaches xxviii

4 SSURGO Preprocessor xxviii

5 Soil properties xxix

6 NLCD Preprocessor xxix

7 Parameter Generator xxx

8 Parameter Postprocessor xxx
9 Processing Time for Previous and Present Approach (per survey area) xxx

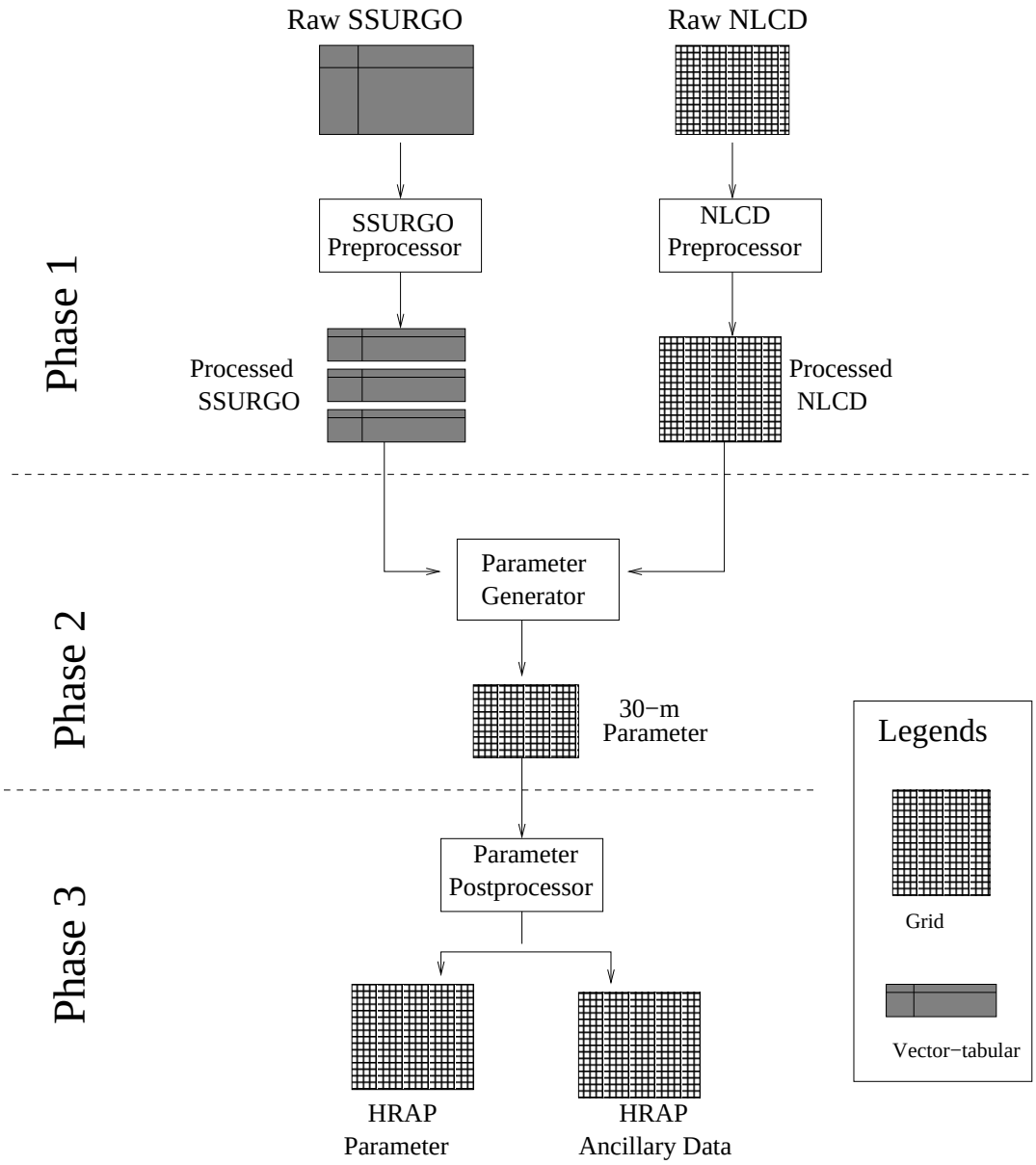

Figure 1: Schematic of processing procedures.

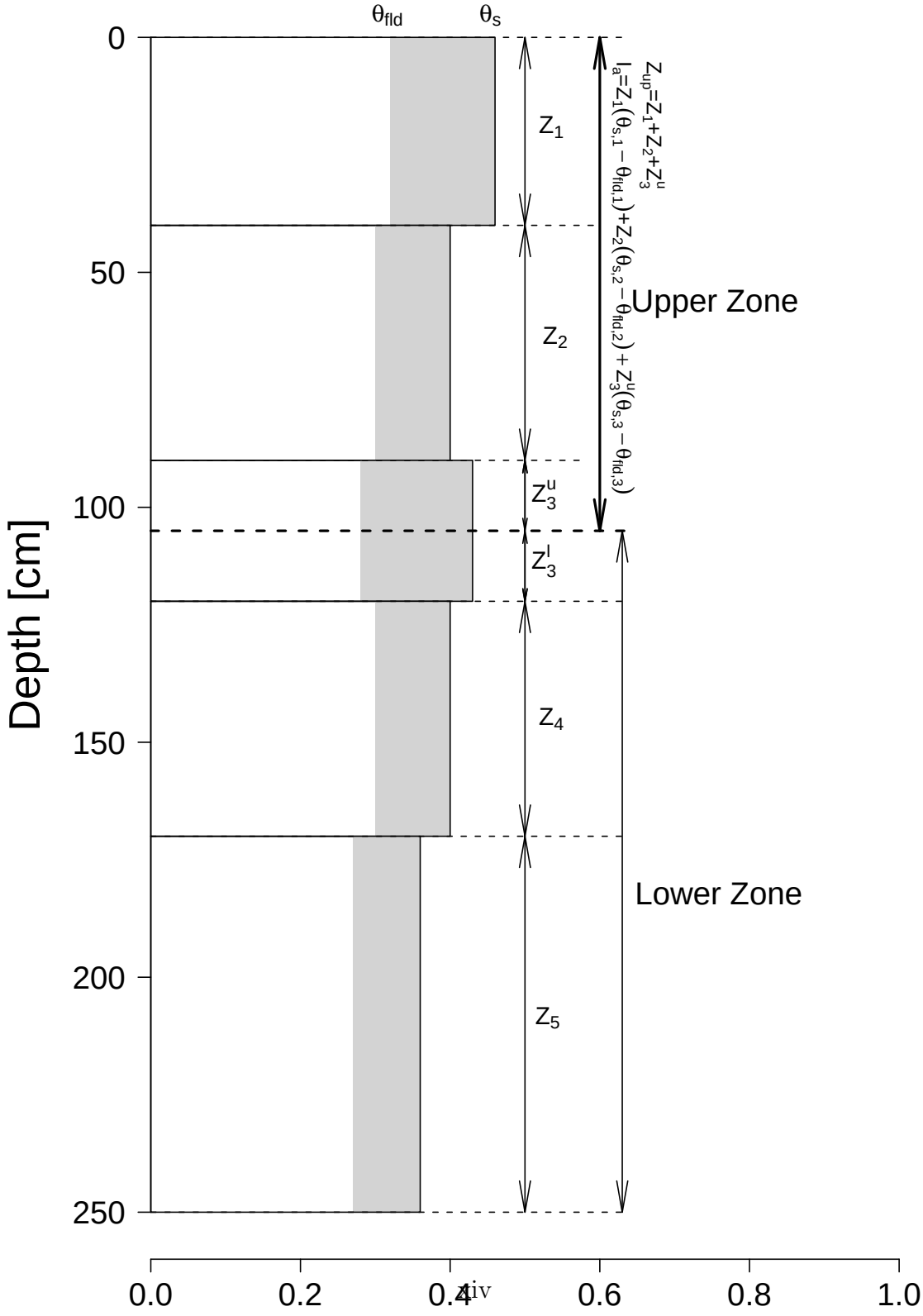

Figure 2: Schematic for method of estimating Z_{up}.

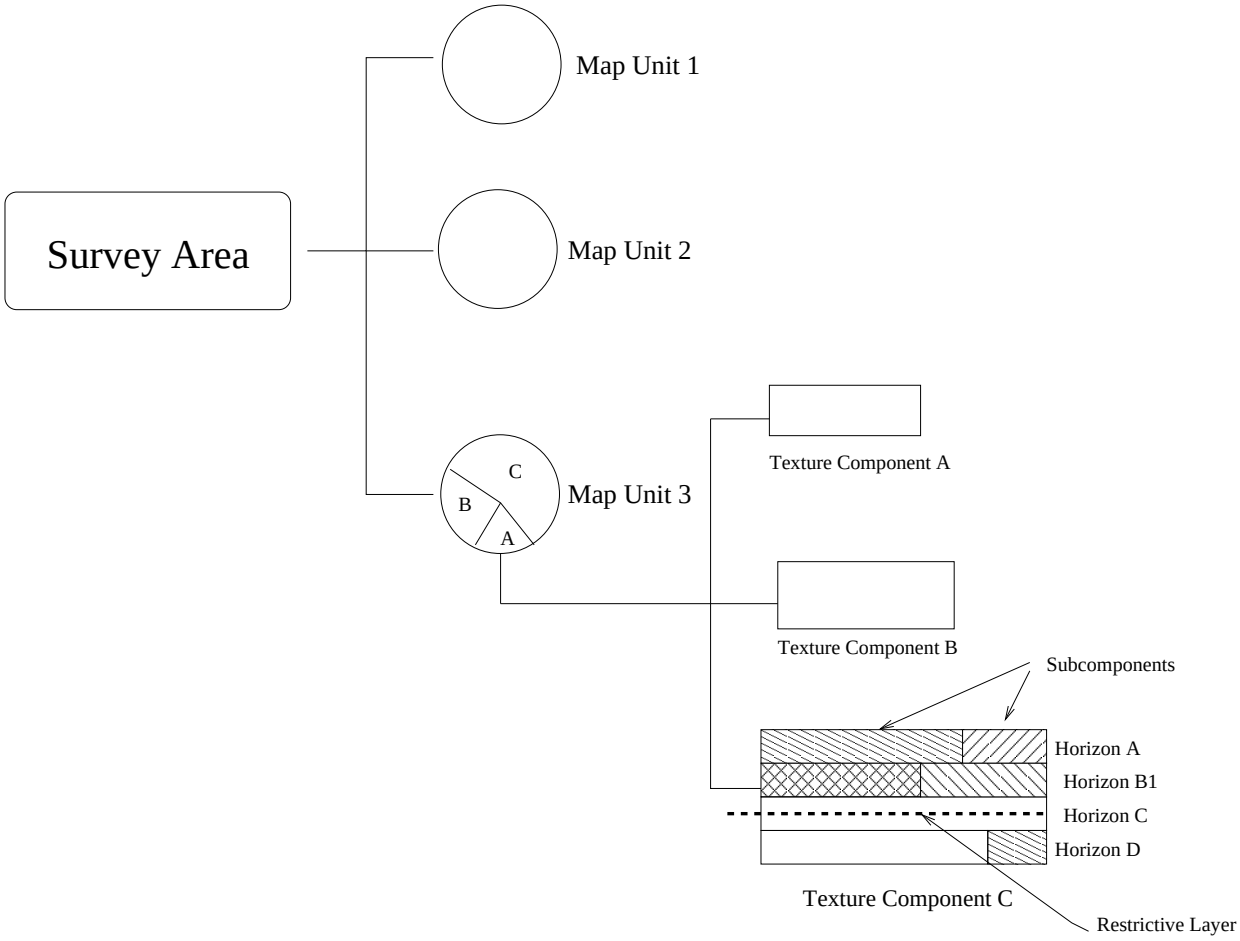

Figure 3: Hierarchy of SSURGO data (uniquely defined by Map Unit Symbol, or MUSYM).

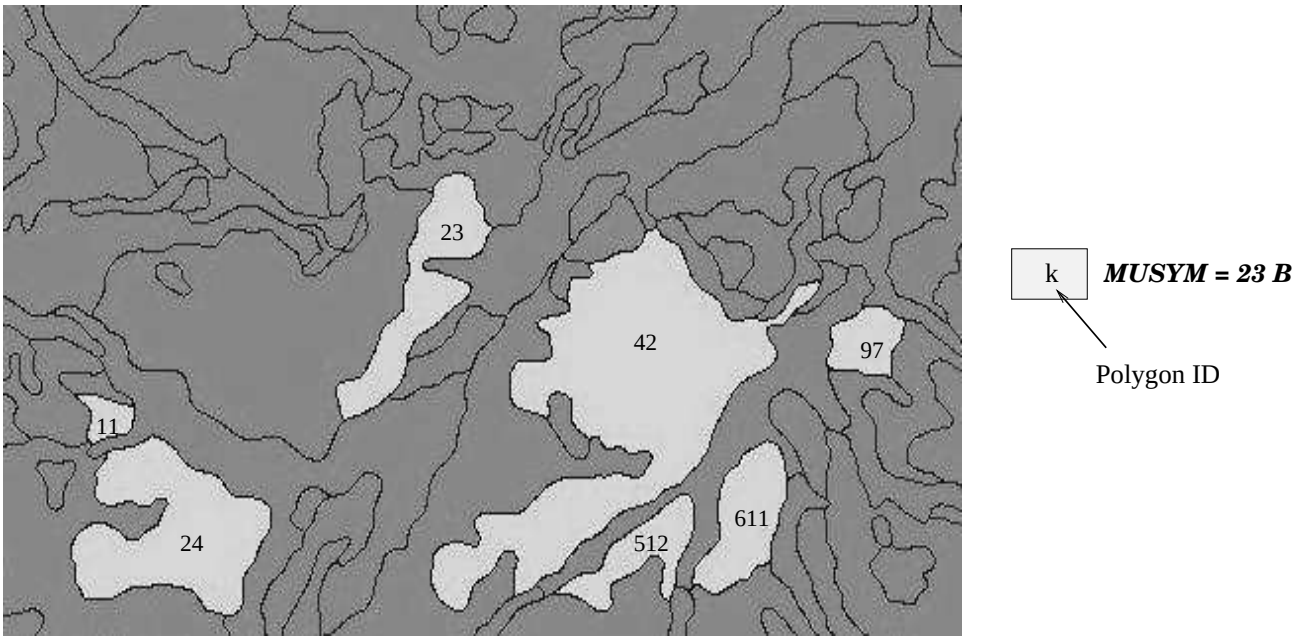

Figure 4: Spatial pattern of a map unit labeled "23B" (given by map unit symbol, or MUSYM). This map unit contains seven non-contiguous polygons.

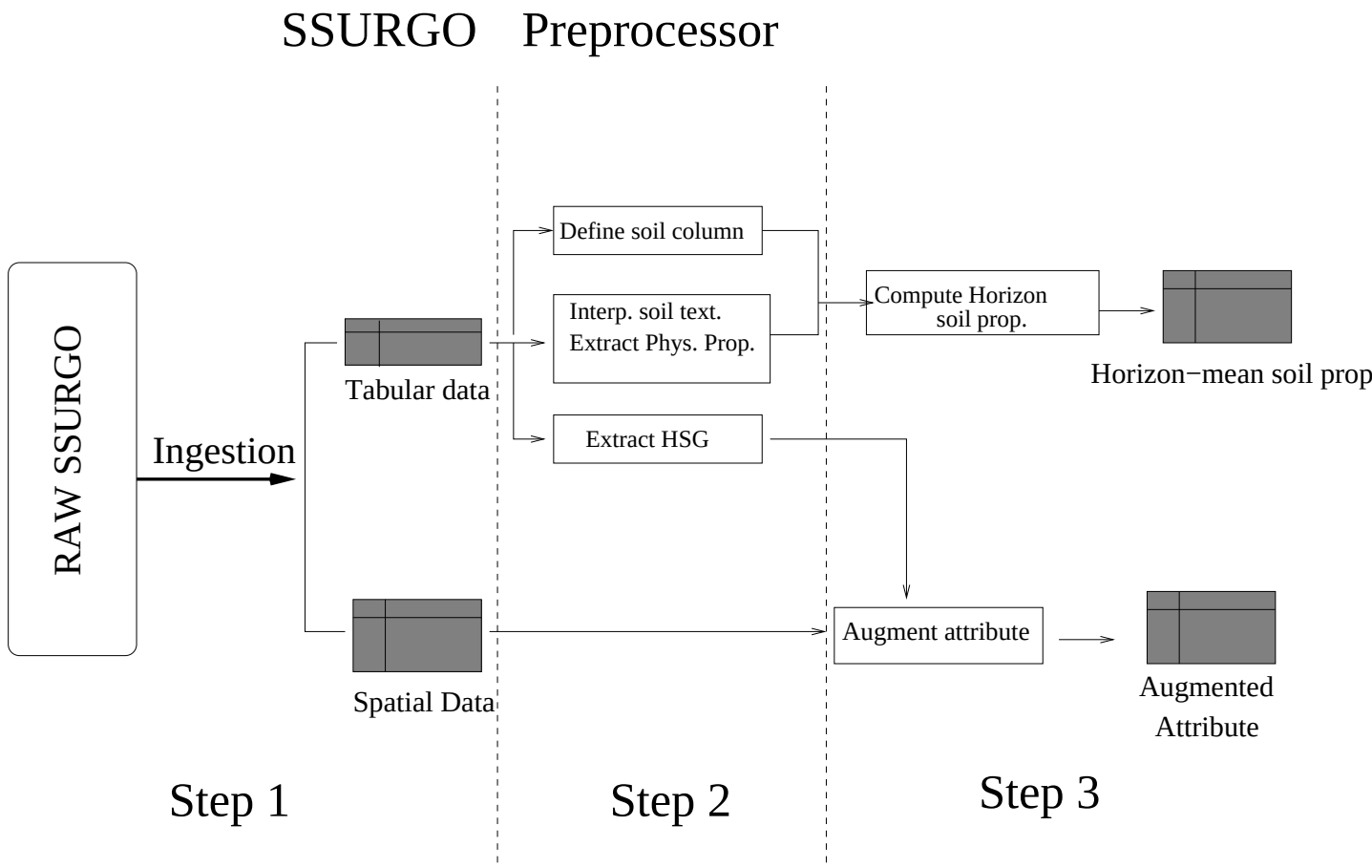

Figure 5: Data flow diagram for SSURGO Preprocessor.

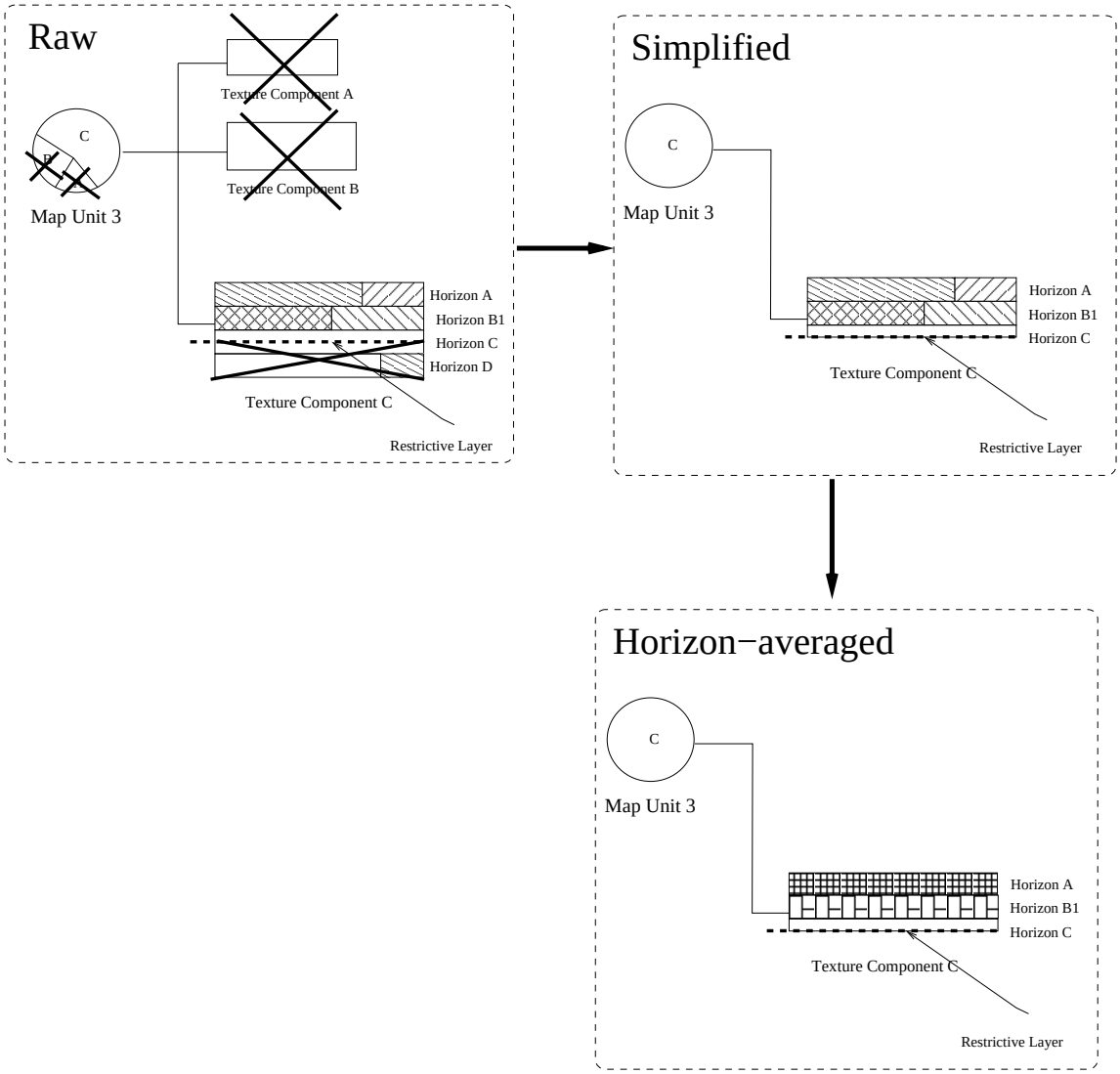

Figure 6: Simplification of texture components in SSURGO Preprocessor. For each map unit, only dominant component (associated with the highest percentage coverage) is retained. In this component, soil properties for various subcomponents within each horizon are arithmetically averaged.

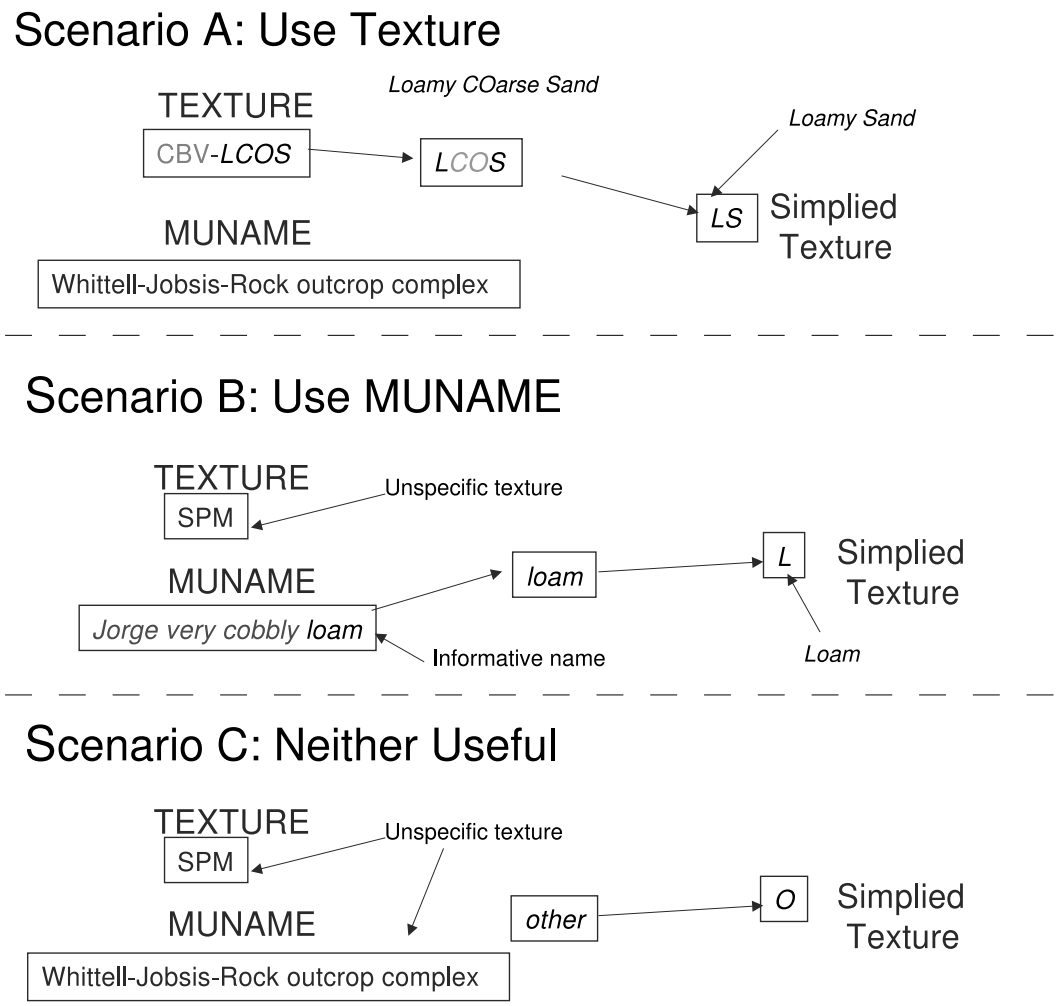

Figure 7: Illustration of texture mapper.

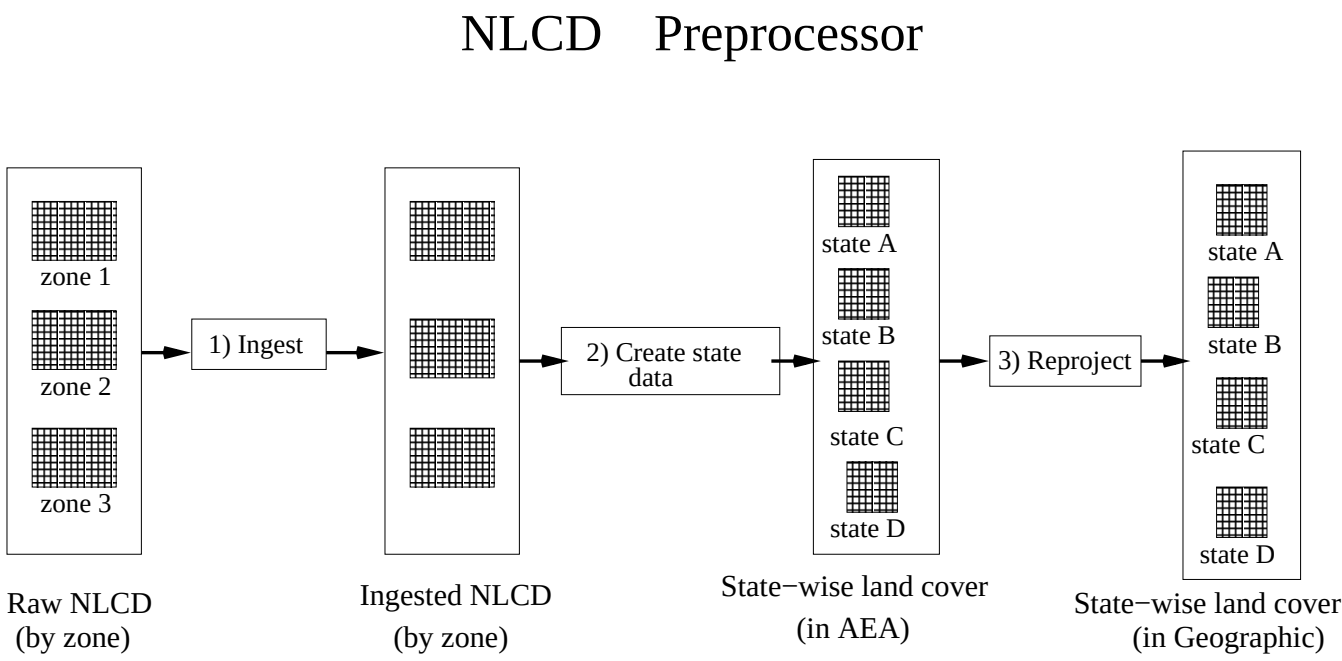

Figure 8: Data flow diagram for NLCD 2001 Preprocessor.

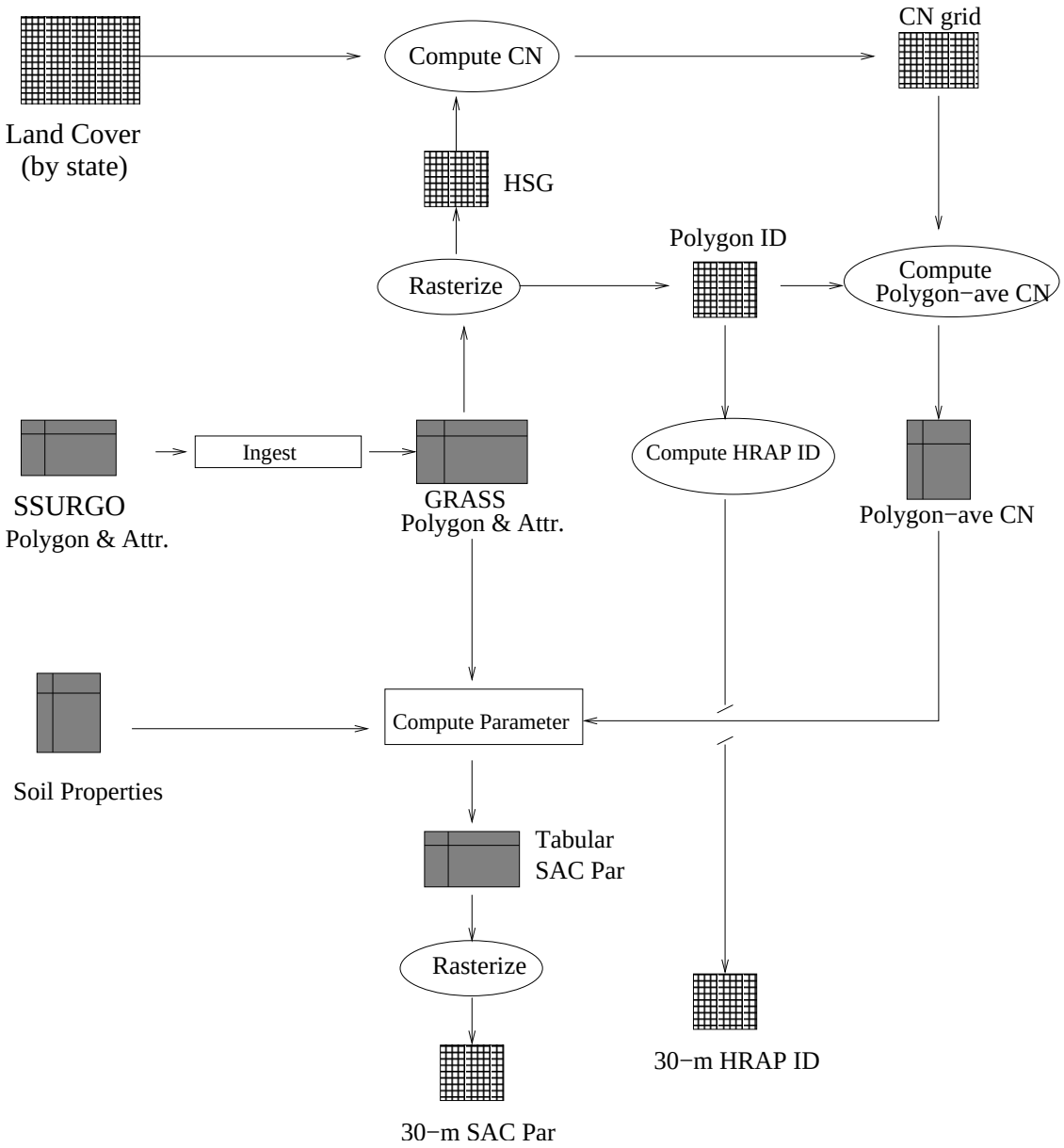

Figure 9: Data flow diagram for Parameter Generator. Among inputs, NLCD data is output from NLCD 2001 preprocessor (Fig. 8), whereas augmented attribute and mu_prop.dbf are outputs from SSURGO preprocessor (Fig. 5).

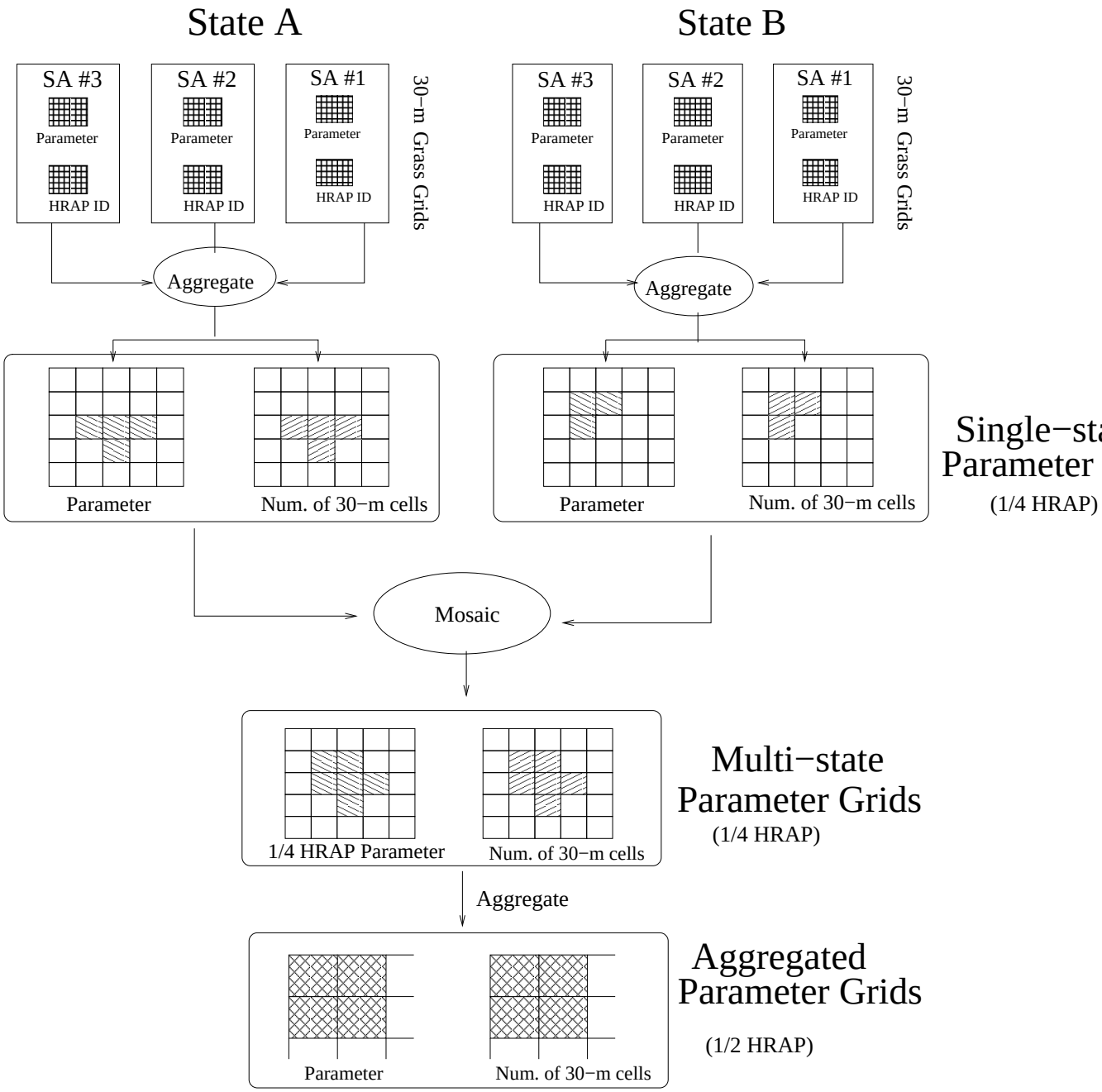

Figure 10: Data flow diagram for Parameter Postprocessor.

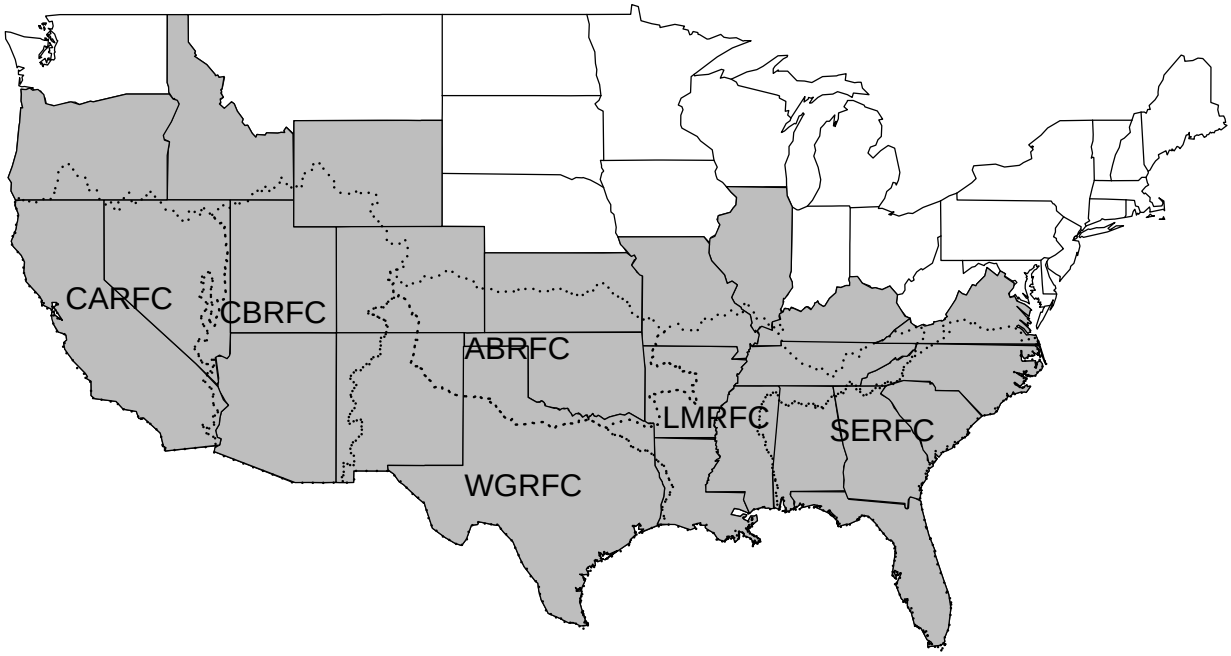

Figure 11: States for which SAC-SMA parameters were derived. Superimposed are geographic domains of six River Forecast Centers (RFCs), namely, California-Nevada (CNRFC), Colorado Basin (CBRFC), Arkansas Red River Basin(ABRFC), West Gulf (WGRFC), Lower Missisippi (LMRFC) and Southeast (SERFC).

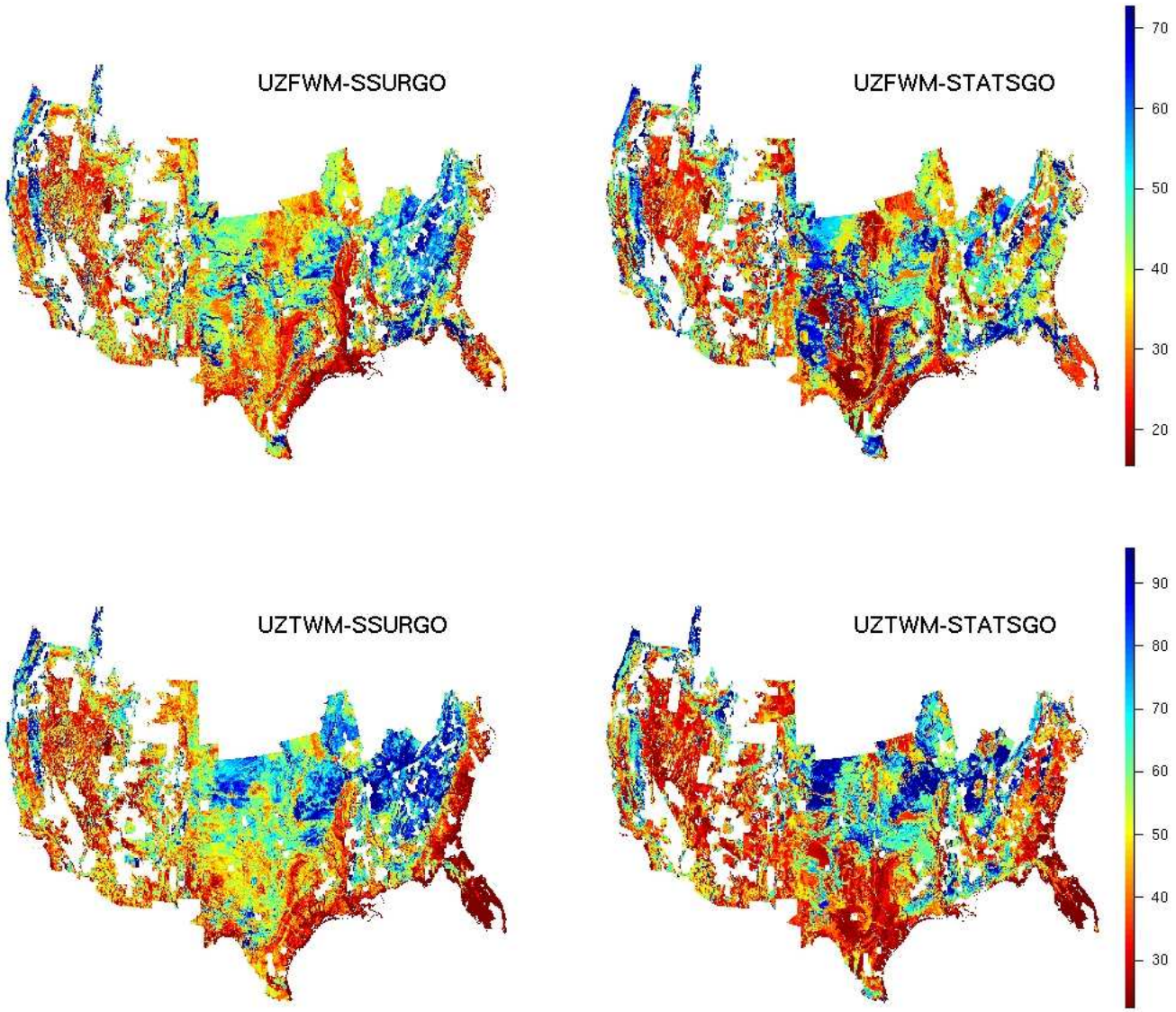

Figure 12: Comparisons of 25-state composite a priori grids (HRAP resolution) of UZFWM (top) and UZTWM (bottom) based on SSURGO (left) and STATSGO (right). Left blank are areas where SSURGO data is currently unavailable, or no meaningful soil texture can be derived from SSURGO data, or curve number reaches 100 (impervious surface).

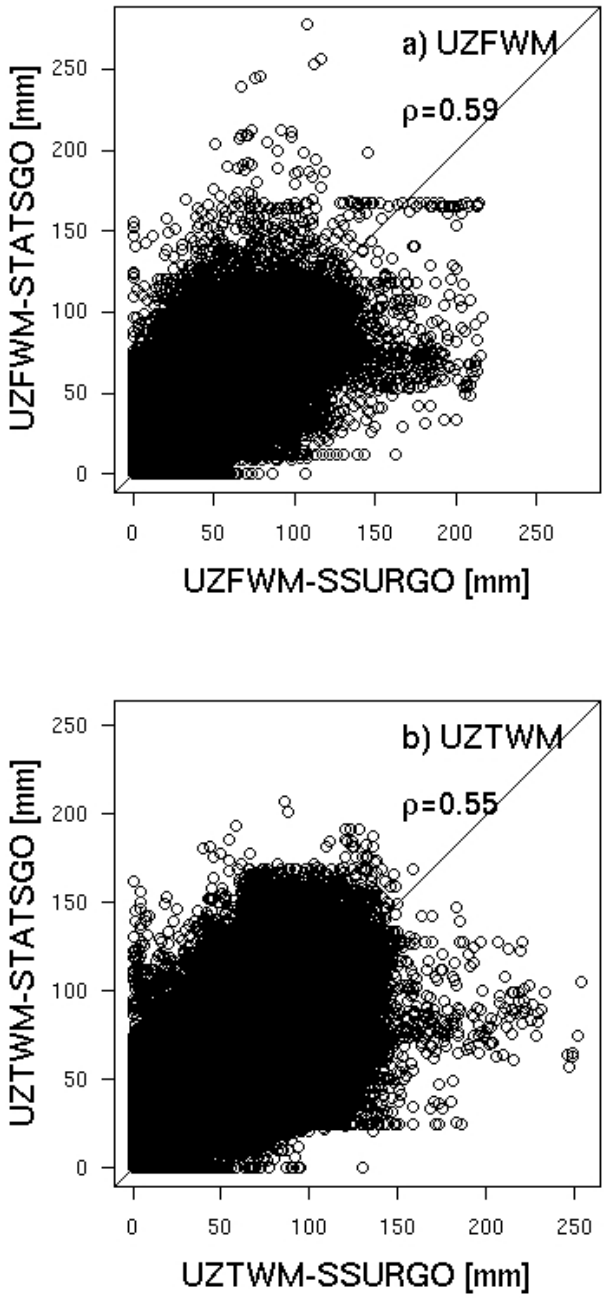

Figure 13: Comparisons of SSURGO and STATSGO-based UZFWM and UZTWM for areas shown in Figure 12. Superimposed are correlations (denoted by ρ).

Table 1: SAC-SMA Parameters

Symbol	Name	Typical Range[a]
UZTWM	Upper zone tension water capacity, mm	10-300
UZFWM	Upper zone free water capacity, mm	5-150
UZK	Interflow depletion rate, day^{-1}	0.1-0.75
ZPERC	Ratio of maximum and minimum percolation rates	5-350
REXP	Shape parameter of the percolation curve	1-5
LZTWM	The lower zone tension water capacity, mm	10-500
LZFSM	The lower zone supplemental free water capacity,mm	5-400
LZFPM	The lower zone primary free water capacity, mm	10-1000
LZSK	Depletion rate of lower zone supplemental free water storage, day^{-1}	0.01-0.35
LZPK	Depletion rate of lower zone primary free water storage, day^{-1}	0.001-0.05
PFREE	Percolation fraction that goes directly to the lower zone free water	0.0-0.8

[a]Ranges are based on lumped model calibration and do not necessarily constrain gridded values.

Table 2: NLCD 2001 Classes and Curve Number

Classes	ID	CN by Hydraulic Group				
		N/A	A,A/D	B,B/D	C,C/D	D
Water						
Water	11	-9999	100	100	100	100
Ice/Snow	12	-9999	95	95	95	95
Developed Areas						
Open Space	21	-9999	29	48	61	69
Low Intensity	22	-9999	40	56	67	74
Medium Intensity	23	-9999	58	70	79	83
High Intensity	24	-9999	70	79	84	87
Barren						
Bare Rock/Sand/Clay	31	-9999	95	95	95	95
Unconsolidated Shore	32	-9999	58	72	81	87
Forested Upland						
Deciduous Forest	41	-9999	19	39	53	61
Evergreen Forest	42	-9999	19	39	53	61
Mixed Forest	43	-9999	19	39	53	61
Shrubland						
Dwarf Scrub - Alaska	51	-9999	34	51	64	77
Shrub/Scrub Areas dominated by shrubs	52	-9999	34	52	64	72
Non-Natural Woody						
Orchards/Vineyards/Other	61	-9999	24	44	57	66
Herbaceous Upland						
Grasslands/Herbaceous	71	-9999	29	48	61	69
Sedge/Herbaceous - Alaska only	72	-9999	28	46	58	67
Lichens - Alaska only	73	-9999	47	61	72	77
Moss- Alaska only	74	-9999	47	61	72	77
Planted/Cultivated						
Pasture/Hay	81	-9999	29	48	61	69
Cultivated Crops	82	-9999	45	57	66	70
Wetland						
Woody Wetlands	90	-9999	100	100	100	100
Palustrine Forested Wetland	91	-9999	100	100	100	100
Palustrine Scrub/Shrub Wetland	92	-9999	100	100	100	100
Estuarine Forested Wetland	93	-9999	100	100	100	100
Estuarine Scrub/Shrub Wetland	94	-9999	100	100	100	100
Emergent Herbaceous Wetlands	95	-9999	100	100	100	100
Palustrine Emergent Wetland (Persistent)	96	-9999	100	100	100	100
Estuarine Emergent Wetland	97	-9999	100	100	100	100
Palustrine Aquatic Bed	98	-9999	100	100	100	100
Estuarine Aquatic Bed	99	-9999	100	100	100	100

Table 3: Highlights of Differences between Previous and Current Approaches

Component	Task	Previous		Current	
		Software	Features	Software	Features
SSURGO Preprocessor	Table Extraction	MS-Access	GUI manual	R	offline, automated
SSURGO Preprocessor	Texture Mapping	MS-Excel Arcview	manual region-specific	R	automated region-independent
NLCD Preprocessor	NLCD Processing	Arcview	GUI manual	GRASS	offline automated
Parameter Generator	Parameter Generation	Arcview	GUI manual	GRASS/R GRASS/R	offline automated
Postprocessor	Postprocess	C/ArcInfo	requires ArcInfo Lib.	C++/GRASS	requires GRASS Lib.

Table 4: SSURGO Preprocessor

Script	Written In	Function
preprocessor.sh	BASH	Create directories and run the following R scripts
std.tname.R	R	Standardizes names of tabular files
hydrologic.R	R	Extracts hydraulic soil groups
physical.R	R	Extracts soil horizons and texture
zmax.R	R	Computes maximum depth of the soil layers
phy_lay_ave.R	R	Computes horizon-averaged soil properties
aug.soil.attr.R	R	Adds drainage group to SSURGO attribute table

Table 5: Soil properties

ID	Symbol	Texture	θ_s	θ_{fld}	θ_{wp}	K_s [mm h^{-1}]	μ
1	S	Sand	0.37	0.15	0.04	634.6	0.29
2	LS	Loamy Sand	0.39	0.19	0.05	562.6	0.23
3	SL	Sandy loam	0.42	0.27	0.09	124.8	0.15
4	SIL	Silt loam	0.47	0.35	0.15	25.9	0.10
5	SI	Silt	0.48	0.34	0.11	20.0	0.12
6	L	Loam	0.44	0.30	0.14	25.0	0.13
7	SCL	Sandy Clay Loam	0.42	0.29	0.16	22.7	0.12
8	SICL	Silty Clay Loam	0.48	0.41	0.24	6.1	0.04
9	CL	Clay Loam	0.45	0.36	0.21	8.8	0.07
10	SC	Sandy Clay	0.42	0.33	0.21	7.8	0.07
11	SIC	Silty Clay	0.48	0.43	0.28	3.7	0.02
12	C	Clay	0.46	0.40	0.28	4.6	0.03
13	O	Other	0.60	0.60	0.53	0.1	0.01

Table 6: NLCD Preprocessor

Script	Written In	Function
import_2001.sh	GRASS/BASH	Imports NLCD data into GRASS
zone_to_state.sh	GRASS/BASH	Derives state-wise NLCD data
reproj.sh	GRASS/BASH	Reproject NLCD to Geographic

Table 7: Parameter Generator

Script	Written In	Function
param.gen.2001.sh	BASH	Shell wrapper that runs the scripts below
import_ssurgo.sh	GRASS/BASH	Imports SSURGO shapefile into GRASS
r.cn.2001	GRASS/C	Computes Curve Number for NLCD 2001
r.cn.ave.poly	GRASS/C	Computes polygon-mean Curve Number
r.ll.hrap	GRASS/C	Computes HRAP ID from lat/lon location
sac_sma.each.R	R	Computes SAC-SMA parameters for each polygon

Table 8: Parameter Postprocessor

Script	Written In	Function
grass2xmrg	GRASS/C++	Merges 30-m parameter grids to 1/4 HRAP resolution
mergeXMRG	C++	Merges multiple 1/4 HRAP-based parameters grids
aggrgXMRG	C++	Aggregates 1/4 HRAP parameters to coarser resolution grids

Table 9: Processing Time for Previous and Present Approach (per survey area)

Task	Previous [h]	Present [h]
SSURGO Preprocessing	1	0.01
Parameter Generation	3	0.22
Parameter Aggregation	2	0.01
Total	6	0.24