EPA/600/R-02/046
ARS/137460
April 2002

Quality Assurance and Quality Control in the Development and Application of the Automated Geospatial Watershed Assessment (AGWA) Tool

Mariano Hernandez, Scott Miller, Darius Semmens, David C. Goodrich, and W.P. Miller
USDA-Agricultural Research Service
Southwest Watershed Research Center
2000 East Allen Road
Tucson, Arizona 85719

and

William G. Kepner
U.S. Environmental Protection Agency
National Exposure Research Laboratory
Environmental Sciences Division
944 East Harmon Avenue
Las Vegas, Nevada 89119

Table of Contents

1 Introduction

In June 1997, the United States Environmental Agency (EPA), National Exposure Research Laboratory (NERL), Landscape Science Program and the United States Department of Agriculture, Agricultural Research Service (ARS) entered into an Interagency Agreement for the purpose of improving ecosystem risk assessment via characterization research, process modeling, and long-term monitoring studies.

One of the project tasks within the interagency agreement was the development of a computer application tool for assessing the hydrologic impacts of land cover change in semi-arid watersheds at different scales. At the outset of the project, a detailed evaluation of existing hydrological models was conducted to select suitable models for multi-scale watershed assessments. It was concluded that for multi-scale modeling it was necessary to select two models that perform successfully at prescribed scales. Therefore, for studies to be conducted at the basin scale, the SWAT model was selected and for studies at the watershed or subwatershed scale the KINEROS model. The use of continuous time and event based, distributed parameter hydrologic models has provided several opportunities to improve watershed modeling accuracy. However, it has also placed a heavy burden on users with respect to the amount of work involved in parameterizing the watershed models and adequately representing the spatial variability of the watershed in particular. Recent developments in Geographical Information Systems (GIS) have alleviated some of the difficulties associated with managing spatial data. However, the user must still choose among various parameterization approaches that are available within each model. At small watershed scales, preparation of model input and parameter files as well as examination of spatially distributed model output is a manageable task. As watershed size increases, spatial data preparation, handling, and interpretation becomes very labor intensive. The Automated Geospatial Watershed Assessment (AGWA) tool, a GIS-based interface, was developed to automatically derive, from DEMs, spatially distributed parameters such as contributing area, slope, average flow length and from data layers of soils and land cover parameters such as hydraulic conductivity and curve numbers. In addition, one strong feature of the computer tool is to display the results of the analysis in a spatially distributed format; this feature will assist in interpretation of model results.

Successful multi-scale watershed assessment requires the use of technically and scientifically sound data collection, information processing, interpretation methods, and proper integration of these methods. As computer codes are essential building blocks of modeling-based management, it is crucial that before such codes are used as planning and decision-making tools, their credibility is established through systematic testing and evaluation of the codes' characteristics.

Developing efficient and reliable software and applying such tools in watershed modeling requires a number of steps, each of which should be taken conscientiously and reviewed carefully. Taking a systematic, well-defined and controlled approach to all steps of the model (software) development and application process is essential for successful implementation of the model. Quality Assurance (QA) provides the mechanisms and framework to ensure that decisions are based on the best available data and analyses.

The following sections provide background information on QA and define its role in watershed modeling. They present a functional and practical methodology, written from the perspective of the model user in need of technical information on which to base decisions. An important part of quality assurance is code testing and performance evaluation.

1.1 Purpose

The purpose of this report is to document the procedures followed to ensure that AGWA conforms to the design objectives and specifications, and that it correctly performs the incorporated functions. These procedures include parameterization of the hydrologic models, the application of coding standards and practices for the development of the GIS-based interface, testing of its functional design, and evaluation of its performance characteristics.

AGWA is a graphical user interface for the KINEROS and SWAT models developed as an ArcView extension. The main purpose of the AGWA tools is to assist in the assessment of the effects of land cover effects of land cover change and land use on watershed response across multiple scales.

1.2 Report Organization

The structure of this document reflects EPA's quality assurance guidelines for modeling development and application projects (EPA, 1991). This report begins with background information on quality assurance in hydrologic modeling. Chapter 3 describes briefly the main components of each hydrologic model, applications and limitations of each model. Chapter 4 deals with data source and input/output quality assurance. This chapter includes quality assurance and quality control for digital elevation models, soil and land use databases, and precipitation data for hydrologic modeling. Chapter 5 describes software development and code testing. Finally, Chapter 6 presents a summary of activities and conclusions.

2 Quality Assurance Plan

2.1 Quality Assurance Definitions

Quality assurance in hydrologic modeling is the procedural and operational framework put in place by the organization managing the modeling study to ensure adequate execution of all project tasks included in the study, and to ensure that all modeling-based analysis is verifiable and defensible (Taylor, 1985).

The two major elements of quality assurance are quality control (QC) and quality assessment (QA). Quality control refers to the procedures that ensure the quality of the final product. These procedures include the use of appropriate methodology in developing and applying computer simulation codes, adequate verification and validation procedures, and proper usage of the selected methods and code. Quality assessment is applied to monitor the quality control procedures and to evaluate the quality of the studies (van der Heijde, 1987).

2.2 Model Development Process

Model development is closely related to the scientific process of acquiring new, quantitative knowledge about nature through observation, hypothesizing, and verifying deduced relationships resulting in the establishment of a credible theoretical framework for the observed phenomena. The fundamental understanding of a hydrologic system thus is the product of research synthesized by theory (van der Heijde et al., 1988).

The object of such research is a prototype of a natural system containing selected elements of a real-world-element hydrologic system. The selection of a particular prototype system for study is driven primarily by management needs and the researcher's personal interest (Figure 1). The conceptual model of the selected hydrologic system forms the basis for determining the causal relationships among various components of the system and its environment. These relationships are defined mathematically, resulting in a mathematical model. If the solution of the mathematical equations is complex or when many repetitious calculations are required, the use of computers is essential. This requires the coding of the solution to the mathematical problem in a programming language, resulting in a computer code. The conceptual formulation, mathematical descriptions, and the computer coding constitute the prototype model (Figure 1).

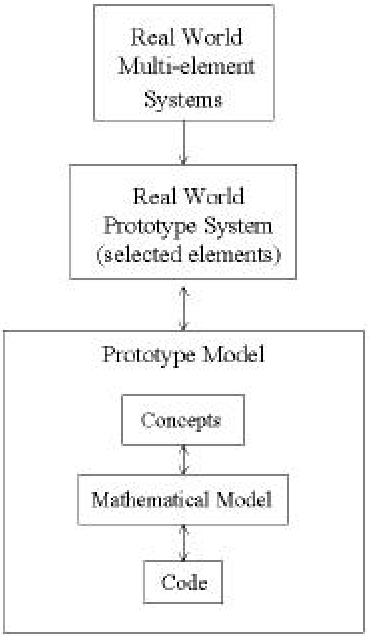

Figure 1. Model development concepts (van der Heijde et al., 1988)

Before a model or software product is used as an evaluation tool, its credibility must be established through systematic testing of the model's accuracy and evaluation of the model's performance characteristics. Of the major approaches, the evaluation or review process is rather qualitative in nature, while code testing results can be expressed using quantitative performance measures. Performance characteristics may be expressed in terms of reliability, efficiency of coded algorithms, and resources for model setup. Performance characteristics need to be determined for the full range of parameters and stresses that the code is designed to simulate. It is also important to test the code to determine the consequences if the code is used beyond its original design criteria, or beyond the range of applications for which it has already been tested. Through extensive and systematic code testing and model evaluation, confidence in the applicability of the code will increase.

Code testing is aimed at detecting programming errors, testing embedded algorithms, and evaluating the operational characteristics of the code through its execution of carefully selected examples, test problems, and test data sets. It is important to distinguish between code testing and model testing. Code testing is limited to establishing the correctness of the computer code with respect to the criteria and requirements for which it is designed. Model testing is more inclusive than code testing, as it represents the final step in determining the validity of the quantitative relationships derived for the real-world prototype system the model is designed to simulate (Figure 1).

In this report, code validation is defined as the process of determining how well the AGWA code's theoretical foundation and computer implementation describe actual system behavior in terms of the degree of correlation between calculated and independently observed responses of the reference hydrologic system for which the code has been developed.

In this report, code verification is defined as the process of demonstrating the consistency, completeness and accuracy of the AGWA code with respect to its design criteria by evaluating the functionality and operational characteristics of the code.

3 Model Description

Key components of AGWA are the hydrological models used to evaluate the effects of land cover and land use on watershed response. In this section, a description of the basic structure of each model is provided as well as their simplifying assumptions, strengths, and weaknesses. Additionally, guidelines are provided for correctly applying the hydrological models to capture the spatial heterogeneities of the watershed to represent the dominant processes at different scales. The KINEROS and SWAT models are able to process complex watershed representations in order to explicitly account for spatial variability of soils, rainfall distribution patterns, and vegetation.

3.1 KINEROS

KINEROS is an event-oriented, physically based model describing the processes of interception, infiltration, surface runoff, and erosion from small agricultural and urban watersheds (Smith et al., 1995). In this model, watersheds are represented by discretising contributing areas into a cascade of one-dimensional overland flow and channel elements using topographic information. The infiltration component is based on the simplification of the Richard's equation posed by (Smith and Parlange, 1978)

$$f_c = K_s \frac{e^{F/B}}{\left(e^{F/B} - 1\right)} \tag{1}$$

$$B = G \cdot e \cdot \left(S_{\max} - SI\right) \tag{2}$$

Where f_c is the infiltration capacity (L/T), K_s is the saturated hydraulic conductivity (L/T), F is the infiltrated water (L), B is the saturation deficit (L), G is the effective net capillary drive (L), ε is the porosity, $S_{\max}$ is the maximum relative fillable porosity, and

SI is the initial relative soil saturation. Runoff generated by infiltration excess is routed interactively using the kinematic wave equations for the overland flow and channel flow, respectively stated as:

$$\frac{\partial h}{\partial t}+\frac{\partial a\cdot h^{m}}{\partial x}=r_i(t)-f_i(x,t) \tag{3}$$

$$\frac{\partial A}{\partial t}+\frac{\partial Q(A)}{\partial x}=q_l(t)-f_{c_i}(x,t) \tag{4}$$

Where h is the mean overland flow depth (L), t is the time (T), x is the distance along the slope (L), α is the 1.49 $S^{1/2}/n$, S is the slope, n is the Manning's roughness coefficient, m is 5/3, $r_i(t)$ is the rainfall rate (L/T), f_{c_i} (x, t) is the infiltration rate (L/T), A is the channel cross-sectional area of flow (L^2), Q(A) is the channel discharge as a function of area (L^3/T), $q_l(t)$ is the net lateral inflow per unit length of channel (L^2/T) and f_{c_i} (x, t) is the net channel infiltration per unit length of channel (L^2/T). These equations, and those for erosion and sediment transport, are solved using a four-point implicit finite difference method (Smith et al., 1995). Unlike excess routing, interactive routing implies that infiltration and runoff are computed at each finite difference node using rainfall, upstream inflow, and the current degree of soil saturation. This feature is particularly important for accurate treatment of transmission losses with flow down dry channels. To explicitly account for space-time variations in rainfall patterns the model computes, for each overland flow element, the rainfall intensities at the element centroid as a linear combination of intensities at the three nearest gages forming a piece-wise planar approximation of the rainfall field over the watershed (Goodrich, 1991). The interpolated centroid intensity is applied uniformly over the individual model element.

3.1.1 Application of KINEROS

In numerous modeling studies, the KINEROS model has been applied to the Walnut Gulch Experimental Watershed administrated by the USDA, Agricultural Research Service (Renard et al., 1993). This is a semi-arid watershed, with 11 nested subwatersheds that range in area from 2.3 to 148 km^2, and an additional 13 small watershed areas ranging from 0.004 to 0.89 km^2. Spatial variability in rainfall is assessed using a network of 85 gages. At a small scale, Goodrich et al. (1995) and Faures et al. (1995) applied KINEROS to the 4.4 Lucky Hills LH-104 subwatershed to examine the importance of different antecedent soil moisture estimates and the effects of wind and rainfall pattern on the predicted discharges. At this scale, both studies conclude that an adequate representation of the rainfall pattern is crucial to achieve accurate runoff prediction in this environment. Goodrich et al. (1994) also looked at the sensitivity of runoff production to pattern of initial water content at the larger scale of the WG-11 subwatershed (6.31 km^2). They suggested that a simple basin average of initial moisture content will normally prove adequate and that, again, knowledge of the rainfall patterns is far more important. Michaud and Sorooshian (1994) compared three different models at the scale of the whole watershed, a lumped curve number model, a simple distributed curve number model, and the more complex distributed KINEROS model. The modeled events were 24 severe thunderstorms with a rain gage density of one per 20 km^2. Their results suggested that none of the models could adequately predict peak discharge and runoff volumes, but that the distributed models did somewhat better in predicting time to runoff initiation and time to peak. The lumped model was, in this case, the least successful.

Goodrich et al. (1997) have used data from the entire watershed to investigate the effects of storm area and watershed scales on runoff coefficients. They concluded that, unlike humid areas, there is a tendency for runoff response to become more nonlinear with increasing watershed scale in this type of semi-arid watershed as a result of the loss of water into the bed of ephemeral channels and the decreasing relative size of rainstorm coverage with watershed area for any individual event.

According to Syed (1999), modeling a medium size watershed (~150 km^2) using the kinematic wave approximation along with a coarse resolution DEM of the order of 80 m with vertical accuracy of tens of meter is acceptable. For watersheds of this size, this implies that USGS level I, 30 m DEM data available throughout the continental United States is adequate. For smaller watersheds of the order of several hectares better vertical accuracy is desired especially when using high horizontal resolution (small grid spacing) DEMs.

3.1.2 Limitations of the Kinematic Wave Approximation

There is one important limitation of using the kinematic approximation to the fully dynamic flow equation; the kinematic wave equation cannot reproduce the effects of a downstream boundary on the flow. Essentially the effects of any disturbance to the flow will generate a kinematic wave, but the equation can only predict the downstream movement of these waves. Thus a kinematic wave description cannot predict the backwater effects of an obstruction to the flow for a surface flow (Beven, 2000).

3.1.3 Basin representation with kinematic wave elements

The contribution to the flood hydrograph from pervious and impervious areas within a single watershed is modeled in the kinematic wave method by using different types of elements as shown in Figure 2. The kinematic wave elements shown are overland flow planes and a main channel. In general, watershed runoff is modeled with kinematic wave elements by taking an idealized view of the basin. Rather than trying to represent every overland flow plane and every possible channel, watersheds are depicted with overland flow planes and channels that represent the average conditions of the basin. Various levels of complexity can be obtained by combining different elements to represent a watershed. The simplest combination of elements that could be used to represent a watershed is two overland flow planes and a main channel. The overland flow planes are used to separately model the overland flow from pervious and impervious surfaces to the main channel. Flow from the overland flow planes is input to the main channel as a uniform lateral inflow. The complexity of a watershed can be modeled by combining various levels of channel elements.

The procedure for representing a watershed using overland flow and channel elements is shown in Figure 3. Using topographic maps and other geographic information, a watershed is configured into an interconnected system of stream network components (Figure 3a). The watershed is subdivided into a number of subwatersheds in order to configure the stream network (Figure 3b). In performing the subdivision, the following are taken into account: (1) the study purpose and (2) the spatial variability of precipitation and runoff response characteristics. The purpose of the study serves to pinpoint the areas of interest and, therefore, the location of watershed boundaries. The spatial variability aids in the selection of the number of subwatersheds. Each subwatershed is intended to represent an area of the basin that, on the average, has the same hydraulic and hydrologic properties. Usually, the assumption of uniform precipitation and infiltration over a subwatershed becomes less accurate as the subwatershed size increases. The flow routing structure is delineated by intersecting the channels with the overlying planes to define the individual plane and channel elements in an 'open-book' model structure (Figure 3c). The abstract routing scheme used in KINEROS is presented in Figure 3d.

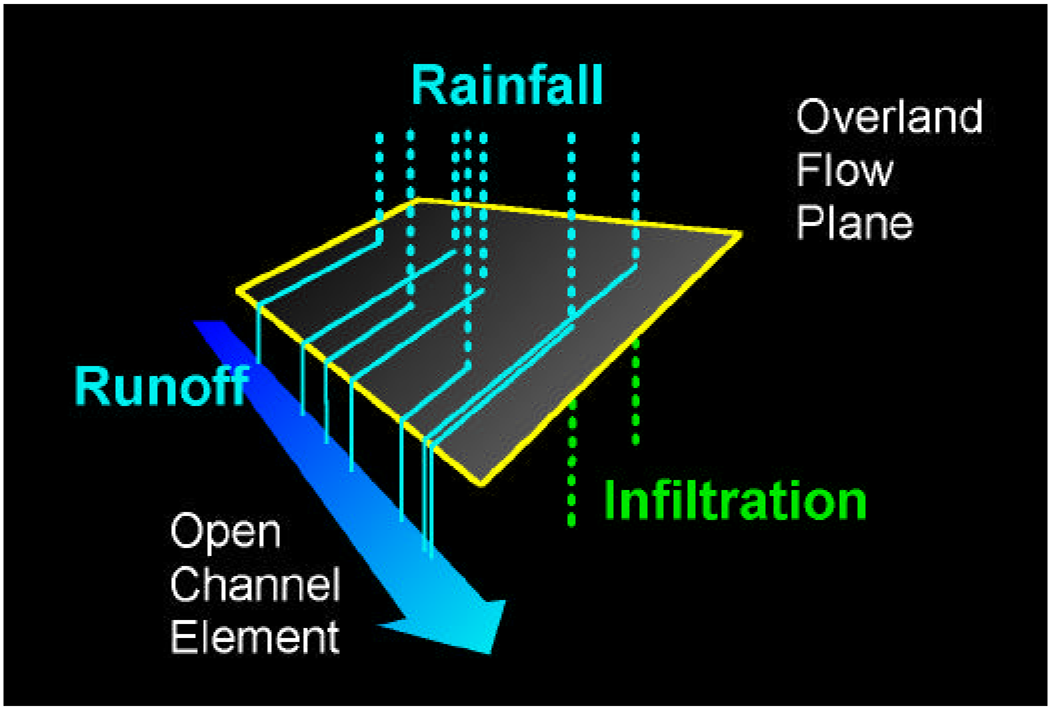

Figure 2. Kinematic wave elements.

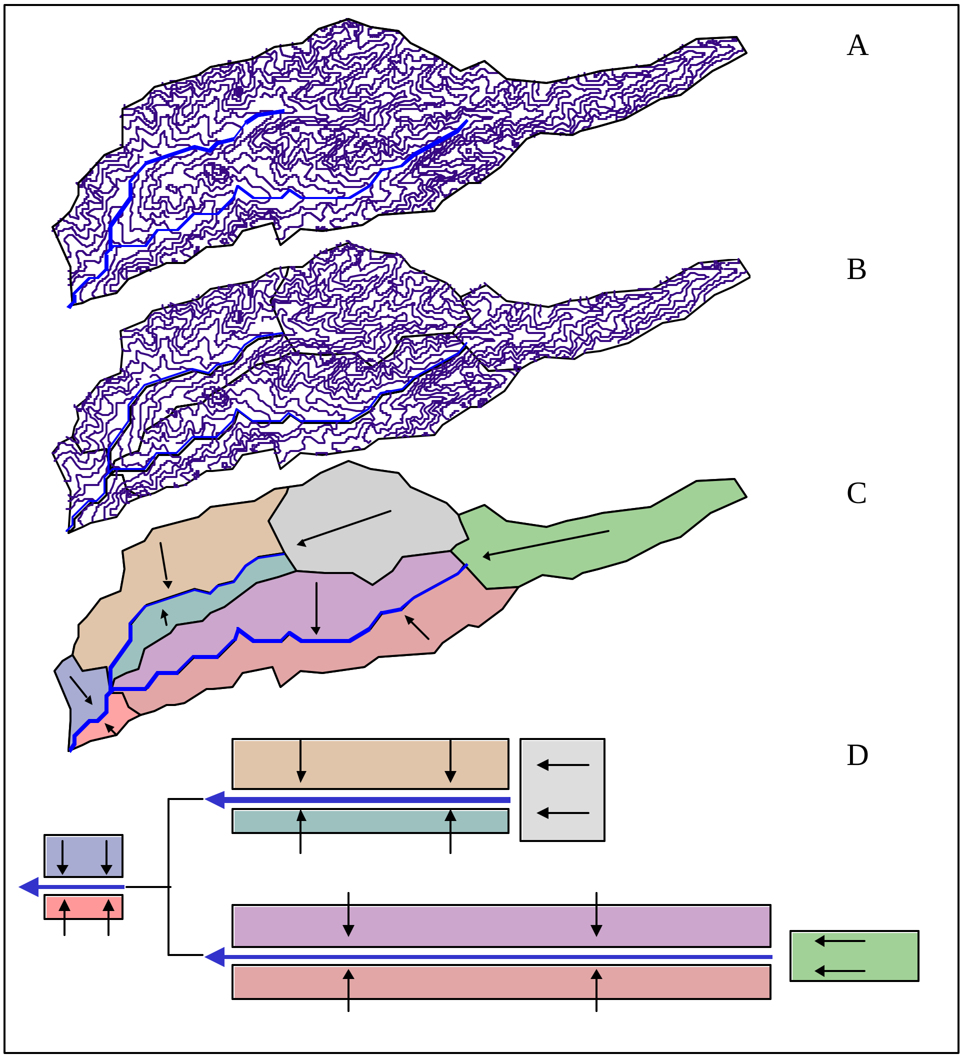

Figure 3 Delineation of planes and channels in AGWA for KINEROS hydrologic modeling. The raw topographic map (A) is used to define the channels, shown in blue, and subwatershed divides, as shown in (B). The flow routing is shown in (C). The abstract routing scheme used in KINEROS is presented in (D).

3.1.4 Estimation of kinematic wave parameters

The parameters that have the strongest influence on runoff from a land cover perspective for KINEROS are saturated hydraulic conductivity, canopy cover, and Manning's roughness coefficient (n). The procedures for determining the hydrologic parameter values for the model are described as follows.

Saturated Hydraulic Conductivity

Saturated hydraulic conductivity (Ks) is of particular relevance to rainfall-runoff modeling in semi-arid regions and is the most critical parameter for accurately simulating runoff using KINEROS (Goodrich, 1991). Rawls et al. (1982) developed a technique for estimating Ks from soil texture; a look–up table based on this work is contained in the original KINEROS documentation (Woolhiser et al., 1990). Soil texture is determined from the STATSGO database, and an area-weighted estimate of Ks is derived from the KINEROS look-up tables for each watershed discretized or subwatershed. This initial estimate is reduced by half to account for entrapped air (Bouwer, 1966), and further reduced by Ks*(1-volumetric rock content) to account for the decrease in pore space caused by the presence of rocks (Woolhiser et al., 1990). Finally, this reduced Ks value is adjusted for the effects of vegetation by a power function suggested by (Stone et al., 1992): $Ksf = Ks * e^{(0.015 * \% \text{ canopy cover})}$. This power function relates vegetation cover and runoff by increasing infiltration with increasing vegetal cover. KINEROS accounts for the small-scale spatial variability of infiltration through an estimate of the coefficient of variation for Ks with the assumption that Ks is log-normally distributed. Estimates of these coefficients are obtained from (Jury, 1985).

Canopy Cover

During a rainfall event on vegetated surfaces, some portion of the rainfall will be retained on the vegetation by tension forces. This portion of the rainfall does not contribute to infiltration or runoff; therefore, an interception depth should be subtracted from the rainfall before infiltration or runoff is performed. In KINEROS, a total depth of interception may be specified for each runoff element, based on the vegetation or other surface condition. This amount is taken from the earliest rainfall pulse until the potential interception depth is filled. The modified rainfall pulse data then becomes input to the soil surface. Woolhiser et al. (1990) provide general estimates for interception by vegetation type.

Manning's Roughness Coefficient

Manning's roughness coefficient (n) is a principle factor in the determination of runoff velocity and, consequently, runoff depth. KINEROS uses Manning's equation in the determination of coefficients for solving the kinematic wave equations for routing water across overland flow elements and channels. A survey of published literature is used to determine estimated values for Manning's (n) based on the land cover classification. Where multiple land covers characterized a given subwatershed element, an area weighted (n) value is used.

3.2 SWAT

SWAT is a river basin, or watershed, scale model developed to predict the impact of land management practices on water, sediment, and agricultural chemical yields on large, complex watersheds with varying soils, land use, and management conditions over long periods of time (Arnold et al. 1994). The model combines empirical and physically based equations, uses readily available inputs, and enables users to study long-term impacts.

The hydrology model is based on the water balance equation

$$SW_t = SW + \sum_{i=1}^{t} \left(R_i - Q_i - ET_i - P_i - QR_i \right) \qquad (5)$$

Where SW is the soil water content minus the 15-bar water content, t is the time in days, and R, Q, ET, P, and QR are the daily amounts of precipitation, runoff, evapotranspiration, percolation, and return flow, respectively; all the units are in mm. Since the model maintains a continuous water balance, complex basins are subdivided to reflect differences in ET for various crops, soils, etc. Thus, runoff is predicted separately for each sub area and routed to obtain the total runoff for the basin. This increases accuracy and gives a better physical description of the water balance.

Surface runoff is estimated with a modification of the SCS curve number method (U. S. Department of Agriculture, 1986).

$$Q = \frac{\left(R - 0.2S \right)^2}{R + 0.8S} \qquad R>0.2S \qquad (6)$$

$$Q = 0 \qquad R<=0.2S$$

Where Q is the daily surface runoff (mm), R is the daily rainfall (mm), and S is the retention parameter. The retention parameter, S, varies (1) among watersheds because soils, land use, management, and slope and (2) with time because of changes in soil water content. The parameter S is related to curve number (CN) by the SCS equation (U. S. Department of Agriculture, 1986).

$$S = 254 \left(\frac{100}{CN} - 1 \right) \qquad (7)$$

The constant 254 in equation (7) gives S in mm. The curve number varies non-linearly from 1, dry condition at wilting point, to the wet condition at field capacity and approaches 100 at saturation.

3.2.1 Application of the SWAT model

SWAT is currently being utilized in several large basin projects. SWAT provides the modeling capabilities of the HUMUS (Hydrologic Unit Model of the United States) project (Srinivasan et al., 1993). The HUMUS project simulates the hydrologic budget and sediment movement for the approximately 2,100 hydrologic unit areas that have been delineated by the USGS. Findings of the project are being utilized in the Resource Conservation Act (RCA) appraisal conducted by the Natural Resources Conservation Service. Scenarios include projected agricultural and municipal water use, tillage and cropping system trends, and fertilizer and animal waste use management options. The model is also being used by NOAA to estimate nonpoint source loadings into all U. S. coastal areas as part of the National Coastal Pollutant Discharge Inventory. The U. S. EPA is incorporating SWAT into the BASINS interface for assessment of impaired water bodies.

3.2.2 Limitations of the Curve Number Method

The curve number approach to predicting runoff generation has been the subject of a number of critical reviews (e.g. Hjelmfelt et al., 1982; Bales and Betson, 1982). Further work is required to clarify under what conditions the method gives satisfactory predictions. Mishra and Singh (1999) show that their generalized version of the method gives better results than the original formulation, as it should, since it has two additional fitting parameters. Hjelmfelt et al. (1982) suggest that the curve number, rather than being considered as a characteristic for a given soil-land cover association, might better be considered as a stochastic variable. Their analysis of the annual maximum storms for two small catchments in Iowa suggested that the storage capacity parameter, Smax, derived for individual storms was approximately log normally distributed with a coefficient of variation on the order of 20 percent. The 10 and 90 percent quartiles of the distributions corresponded well to the modified curve numbers for dry and wet antecedent conditions, following the standard SCS procedure based on the preceding five-day rainfall. However, they found no strong correlation between curve number and antecedent condition for the individual storms, suggesting that interactions with individual storm characteristics, tillage, plant growth and temperature were sufficient to mask the effect of antecedent rainfall alone.

Despite its limitations, the Curve Number method has recently been used quite widely since the tabulated curve number values provide a relatively easy way of moving from a GIS data set on soils and vegetation to a rainfall-runoff model.

3.2.3 Basin representation with SWAT

For modeling purposes, a watershed may be partitioned into a number of subwatersheds or subbasins. The use of subbasins in a simulation is particularly beneficial when different areas of the watershed are dominated by land uses or soils characteristically different enough to impact hydrology. By partitioning the watershed into subwatersheds, the user is able to relate different areas of the watershed to one another spatially. The number of subwatersheds chosen depends on the size of the watershed, the spatial detail of available input data, and the amount of detail required to meet the goals of the project. Figure 4 illustrates a watershed delineation for subwatershed 11 of the Walnut Gulch Experimental Watershed for SWAT. In Figure 4A the raw topography are used to define the channel network, shown in blue, and watershed divides in B. The flow routing structure is delineated by linking the channels with the surrounding uplands to define the individual subwatershed and channel elements (C). The abstract routing scheme used in SWAT is presented in (D). As opposed to the 'open-book' structure with left and right lateral contributing areas for KINEROS shown in Figure 3D, AGWA does not split the subwatershed elements into more than one unit for SWAT.

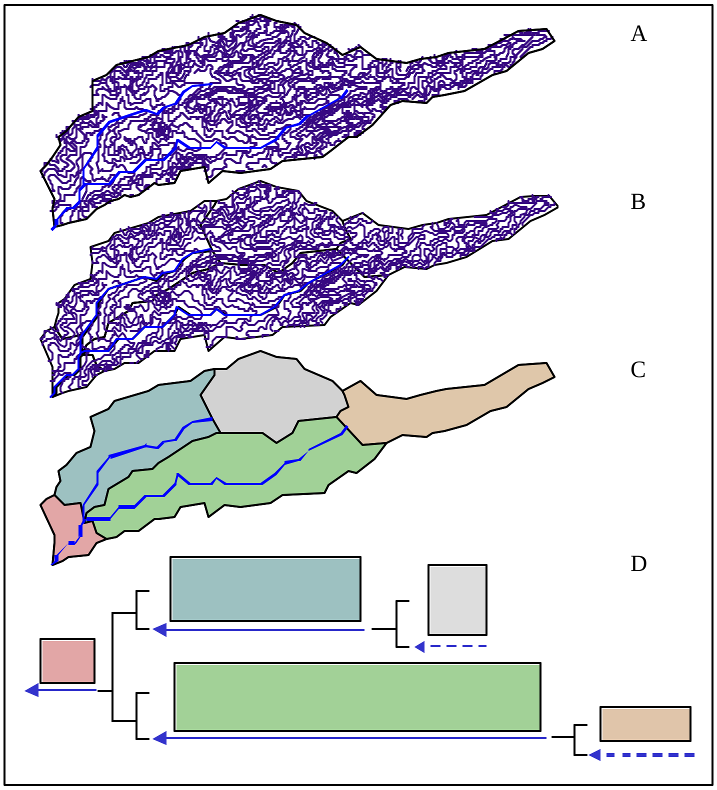

Figure 4. Delineation of planes and channels in AGWA for SWAT hydrologic modeling. Topography map (A), Subwatershed boundaries (B), flow routing structure (C), and the abstract routing scheme used in SWAT (D).

3.2.4 Factors used to estimate Curve Number values

The major factors that determine the CN are hydrologic soil group, hydrologic condition, cover type, treatment, and antecedent runoff condition.

Hydrologic soil groups

Infiltration rates of soils vary widely and are affected by subsurface permeability rates. Soils are classified into four hydrologic soils groups according to their minimum infiltration rate, which is obtained for bare soil after prolonged wetting. The soils in the area of interest may be identified from a soil survey report, which can be obtained from local NRCS offices or soil and water conservation district offices. In the AGWA tool the hydrologic group classification is determined from the STATSGO soil database description.

Cover type

There are a number of methods for determining cover type; the most common are field reconnaissance, aerial photographs, and land use maps. The SWAT manual addresses most cover types, such as vegetation, bare soil, and impervious surfaces (www.brc.tamus.edu/swat/).

Treatment

Treatment is a cover type modifier to describe the management of cultivated agricultural lands. It includes mechanical practices, such as contouring and terracing, and management practices, such as crop rotations and reduced or no tillage.

Hydrologic condition

Hydrologic condition indicates the effects of cover type and treatment on infiltration and runoff and is generally derived from estimates of plant density and residue cover on sample areas. Good hydrologic conditions indicate that the soil usually has a low runoff potential for that specific hydrologic soil group, cover type, and treatment. Some factors to consider in estimating the effect of cover on infiltration and runoff are (a) canopy or density of lawns, crops, or other vegetative areas; (b) amount of year-round cover; (c) amount of grass or legumes in rotations; (d) percent of residue cover; and (e) degree of surface roughness. In the AGWA tool, the hydrologic condition is determined from the STATSGO soil database description.

4 Data Source/Quality/Input-Output

4.1 Digital Elevation Model Data

Digital elevation models are generally produced by photogrammetric techniques from stereo-photo pairs, stereo-satellite images, or interpolation of digitized contour elevation data. The U.S. Geological Survey, Earth Science Information Center, offers a variety of digital elevation data products (U. S. Geological Survey, 1990). These include the 7.5-minute grid DEM data, 1 degree grid DEM data, regular angular 30-minute grid DEM data, and contour DLGs corresponding to maps of various scales. The USGS 7.5-minute DEM data have a grid spacing of 30 by 30 meters, are cast on Universal Transverse Mercator (UTM) projection, and are produced from contour overlays or from automated or manual scanning of National Aerial Photography Program stereo-photographies. DEMs provide coverage in 7.5 by 7.5 minute blocks, each providing the same coverage as a standard USGS 7.5-minute map series quadrangle (U. S. Geological Survey, 1990). Elevation values are provided in either feet or meters.

DEM data are classified into one of three levels of quality. Level 1 classification is generally reserved for data derived from photogrammetric compilation of stereo imagery from the National High-Altitude Photography Program or National Aerial Photography Program. A vertical Root Mean Square Error (RMSE) of 7 meters is the targeted accuracy standard, while a RMSE of 15 meters is the maximum permitted. Level 2 classification is for elevation data sets that have been processed or smoothed for consistency and edited to remove identifiable systematic errors. DEM data derived from hypsographic and hydrographic data digitizing are entered into Level 2 classification; an RMSE of one-half of the original map contour interval is the maximum permitted. There are no errors greater than one contour interval in magnitude. Level 3 classification is derived from DLG data by using selected elements from both hypsography (contours, spot elevations) and hydrography (lakes, shorelines, drainage). If necessary, ridge lines and hypsographic effects or major transportation features are also included in the derivation. A RMSE of one-third of the contour interval is the maximum permitted. There are no errors greater than two-thirds of the contour interval in the magnitude.

4.2 Selection of DEMs for Hydrologic Modeling

The two important aspects in the selection of a DEM for hydrologic modeling are the quality and resolution of the DEM data. Quality refers to the accuracy of the elevation data, and resolution refers to the horizontal grid spacing and vertical elevation increment. Quality and resolution must be consistent with the scale and model of the physical process under consideration and within the study objectives. For many applications of physically processed based environmental models the USGS 30 by 30 meter DEM data (Level 1 and 2) has a relatively low accuracy standard and a rather coarse resolution with documented shortcomings (Syed, 1999; Garbrecht and Starks, 1995; Ostman, 1987). In particular, surface drainage identification is difficult in low relief landscapes, as is derivation of related information such as slope and landform curvature. No firm guidelines are available for selection of DEM characteristics. DEM selection for a particular application is generally driven by data availability, experience and test applications.

4.3 SOILS: Data Sources

Soils maps for the State Soil Geographic (STATSGO) (www.statlab.iastate.edu/soils-info/nssc/) database are made by generalizing the detailed soil survey data. The mapping scale for a STATSGO map is 1:250,000. The level of mapping is designed for broad planning and management uses covering state, regional, and multi-state areas. STATSGO data are available for the conterminous U. S., Hawaii, and Puerto Rico. Digitizing is done by line segment format in accordance with Natural Resources Conservation Service (NRCS) digitizing standards. The base map used is the USGS 1:250,000 topographic quadrangles. The number of soil polygons per quadrangle map is typically between 100 and 400. The minimum area mapped is about 1,544 acres. Each STATSGO map is linked to the Soil Interpretations Record attribute database. The attribute database gives the proportionate extent of the component soils and their properties for each map unit. The STATSGO map units consist of 1 to 21 components each. The SIR database includes over 25 physical and chemical soil properties, interpretations, and productivity. Information that can be queried from the data base include available water capacity, soil reaction, salinity, flooding, water table, bedrock, and interpretations for engineering use of cropland, woodland, rangeland, wildlife, and recreation development.

4.4 LAND USE: Data Sources

Land use/cover information is used in hydrologic modeling to estimate the value of surface roughness or friction as it affects the velocity of the overland flow of water. Land use information is also useful as an indicator of the amount of rainfall infiltration on a surface. The land use information, coupled with the hydrologic characteristics of the soils of a land surface, can also provide measures of expected percolation and water holding capacity. The amount of expected runoff from vegetated land use types, such as forest, is affected not only by the surface and soil physical properties, but also by the uptake capacity of the flora present. The North American Landscape Characterization (NALC) classification was used to derive hydrologic parameter values for different land use/cover. NALC data consist of remote sensing imagery that comprises three or more registered Landsat MSS images corresponding to the 1990s, 1980s, and 1970s time periods. On average, a NALC triplicate consists of one scene from 1990s and 1980s, and two from the 1970s for each path/row.

For each triplicate set, the 1980s image was rectified, and then used as template to co-register 1970s and 1990s images. Image control points were selected from the 1980s images, and corresponding map control points were obtained from maps or a library of ground control points for use in developing the geometric transformations model. The result was an image registration procedure that only involved one step of resampling (Lunetta et al., 1993). The final database development task involved the mosaicing and projection transformation of the DEM data. The DEM data were derived from the Defense Mapping Agency digital terrain and elevation data, which were digitized from standard 1:250,000 scale topographic maps. Complete coverage existed for the United States and Mexico for 60 by 60 meter pixels in a UTM projection.

4.5 Precipitation Data for Hydrologic Modeling

Confidence in the hydrologic modeling effort depends, to a large extent, on the availability of high quality rainfall and runoff data for model calibration and verification. Traditionally, rainfall estimated from sparse rain gage networks has been considered a weak link in watershed modeling. The purpose of this section is to document available data sources and limitations of available data for event-based hydrologic modeling.

Many sources of rain gage data are available. However, the likelihood of obtaining rain gage data for a particular watershed is small because of the sparse nature of the national rain gage network. Rainfall data are archived by the NOAA National Climatic Data Center (NCDC) located in Asheville, North Carolina (www.ncdc.noaa.gov). The NCDC is charged with archiving rainfall and meteorological data from a number of sources.

Relevant available precipitation data from NCDC include: daily parameters such as maximum/minimum temperature, precipitation, and snowfall/snow depth. Some stations have additional data such as evaporation and soil temperature. Hourly rainfall rates are recorded at the National Weather Service meteorological stations. These stations are sparsely located around the U. S. The period of record for these data is quite variable, with few stations installed before 1970. The NCDC is very efficient at archiving available precipitation data sources, and performing quality control on the data (American Society of Civil Engineers, 1999). Unfortunately, the precipitation data from the NCDC is not available free on-line.

In Walnut Gulch, rainfall observations from more than eighty gages are available. These are standard weighing type gages that record the cumulative depth of precipitation continuously as a line trace on a revolving chart driven by an analog clock. The chart completes one revolution in 24 hours and remains in place for seven days before it is replaced with a fresh chart. These charts are manually checked and inferred for starting and ending times of rainfall events. Weekly rain gages (one chart revolution per 7 days) are also used to infer storm start times.

5 Software Development and Code Testing

In this section the process of developing and testing the main components of the AGWA tool are described. To accomplish the goals set out for the AGWA tool, several algorithms were developed such as derivation of flow paths and subsequent subdivision of the study watershed into channels and plane elements for input into the KINEROS and SWAT models. These processes rely upon core ArcView utilities (Environmental Systems Research Institute, 1996) to perform the initial watershed subdivision, but AGWA introduces several unique processes into the watershed discretization routines. To account for the spatial distribution of rainfall, an algorithm based on the Thiessen polygons concept was written. A detailed description of this algorithm is provided below and those specifically developed as part of AGWA development beyond well documented ArcView utilities.

5.1 Derivation of Land Surface Drainage and Channel Network from DEM's

Surface drainage and channel network configuration are important landscape attributes for hydrologic modeling of runoff processes. Both attributes can be determined from field surveys, stereo photos, and detailed topographic contour maps. However, these approaches are resource and time consuming, particularly for large watersheds.

An accurate definition of drainage networks in hydrologic analyses is important because the network indirectly determines the hillslope travel distance and network link lengths, both of which affect the simulated hydrologic response of a watershed. A drainage network can be extracted from a DEM with an arbitrary drainage density or resolution (Tarboton et al., 1991). The characteristics of the extracted channel network depends extensively on the definition of channel sources on the digital landscape. Once the channel sources are defined, the essential topology and morphometric characteristics of the corresponding downstream drainage network are implicitly pre-defined because of their close dependence on channel source definition. Thus, the proper identification of channel sources is critical for extraction of a representative drainage network from DEMs.

One of the primary tasks of AGWA is the derivation of flow paths and subsequent subdivision of the study watershed into channel and plane elements for input into the KINEROS and SWAT hydrologic models. This process relies upon core ArcView utilities to perform the initial watershed subdivision, but AGWA introduces several unique processes into the watershed discretization routines. In the process of generating subwatershed elements, the core ArcView algorithms often generate orphaned elements that can confuse hydrologic routing. Consequently AGWA contains a subroutine that snaps together larger elements, thereby erasing the small, spurious elements. The other significant addition to core functionalities provided by AGWA is the automated determination of upland (0-order) watershed elements and the splitting of the primary watershed into lateral elements for KINEROS.

These subroutines are robust, but they may fail when the underlying digital elevation model (DEM) is overly coarse and does not adequately represent topographic detail. Furthermore, since AGWA uses a threshold approach to generate stream channels, which then serve to determine the size and locations of watershed elements, it is possible to select a threshold level that creates an error in the discretization subroutines. In short, there are cases where these subroutines are not capable of fulfilling their tasks, and these problems are illustrated in the AGWA user manual so the user may determine the cause and find a remedy for their particular problem. Error trapping is performed to determine the cause of the failure, with the results presented to the user. The subwatershed routines have been tested on a wide range of spatial scales (ranging from < 5 ha to > 1800 km2) and within a variety of topographic and bio-geophysical provinces (southern Arizona, Nevada, montane Colorado, and upstate New York), with consistently good results.

Watershed discretization is dependent upon the presence of a high quality DEM. This DEM is used to generate a flow direction map, within which each cell is assigned a numerical value indicating the direction of flow. This raster map is used to create a flow accumulation map, also a raster, within which each cell is assigned a value corresponding to the number of cells that contribute flow to it. Thus, a cell residing on a watershed divide will have a flow accumulation value of 0, while the watershed outlet will be assigned a value equal to the number of cells in the entire watershed. Converging cells rapidly accumulate flow and high values are used to determine the locations of stream channels. In AGWA, a user specifies a threshold of flow accumulation; each cell containing a value higher than the threshold is designated as a channel, while the remainder is considered to belong to plane elements. If the user selects a low threshold the number and length of stream channels will be relatively high, while a high threshold results in the creation of fewer stream channels (Figure 5). This is an important step in AGWA since the watershed subdivision is based on the presence of stream channels; more channels means that more watershed elements will be created.

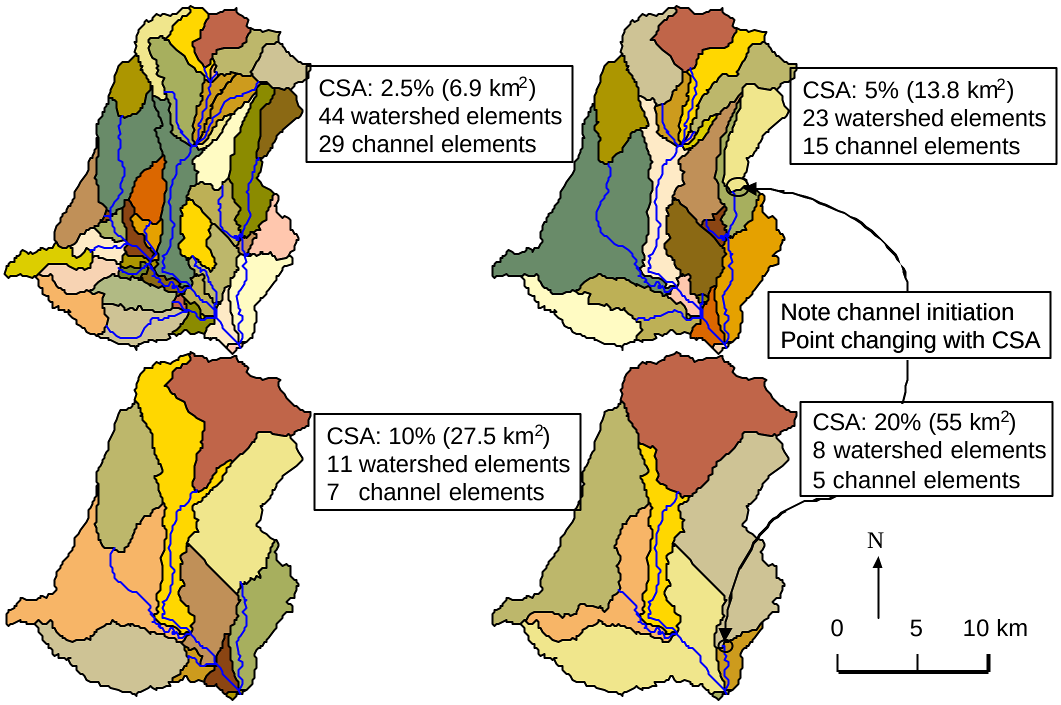

Figure 5. Illustration of the impact of threshold value (shown as CSA, or contributing source area) on channel formation and watershed discretization. Increasing the threshold results in fewer and/or shorter channels, and often in fewer and larger watershed elements.

Figure 5 is an illustration showing several finalized watersheds; all four examples were successfully created using the snapping and subdivision routines described earlier. Some discussion is warranted regarding the subroutines and the approach used by AGWA to produce these results. There are 23 important steps and subroutines used to generate the final watershed products, but the principles can be reduced to a few key steps. First, the watershed is delineated and channels created using core ArcView utilities based on the user-defined threshold. These raster maps are transformed into vector maps (shape files) so that they may be intersected with one another for the purposes of determining the channel routing sequence.

When watersheds are created, they are often unattached to their parent stream channel, so AGWA snaps the channels and watersheds together to ensure hydrologic connectivity. This is a source of problems for some watersheds; if the channels are very small (less than one grid cell), more than one watershed element may be snapped, and connected to, a channel element. This interferes with the routing routines. If this occurs, the user is alerted to the problem and asked to alter the threshold so that the channel is either enlarged or removed entirely. Once the connectivity is completed the channels are numbered and their routing linkages recorded in a database file. Watersheds that contribute runoff to a given channel are assigned routing numbers following a rule-based numeric system: all channels are assigned numbers ending in the number “4”, while the upland watersheds are assigned the same prefix, but their suffix will end in “1” and lateral elements will be assigned numbers with the same channel prefix, but ending in “2” or “3”. This strict scheme ensures that routing will be unique and the connectivity among watershed and channel planes is ensured. If more than one watershed is assigned a value associated with a channel, this is flagged as an error and the user is prompted to try again using a slightly different threshold.

It is at this stage that the significant differences between the AGWA approach and the generalized techniques provided by ArcView are manifest. Upland watersheds are created by AGWA as shown in Figure 6 for both SWAT and KINEROS. These watersheds are also transformed into vector shape files and are cleaner in appearance and contain fewer, if any, orphaned watershed elements. The next step is to subdivide the main watershed planes into two lateral elements for running KINEROS; SWAT requires that no such step be taken. Figure 7 illustrates the effect of subdividing the main watershed into two lateral elements for simulating overland flow in KINEROS.

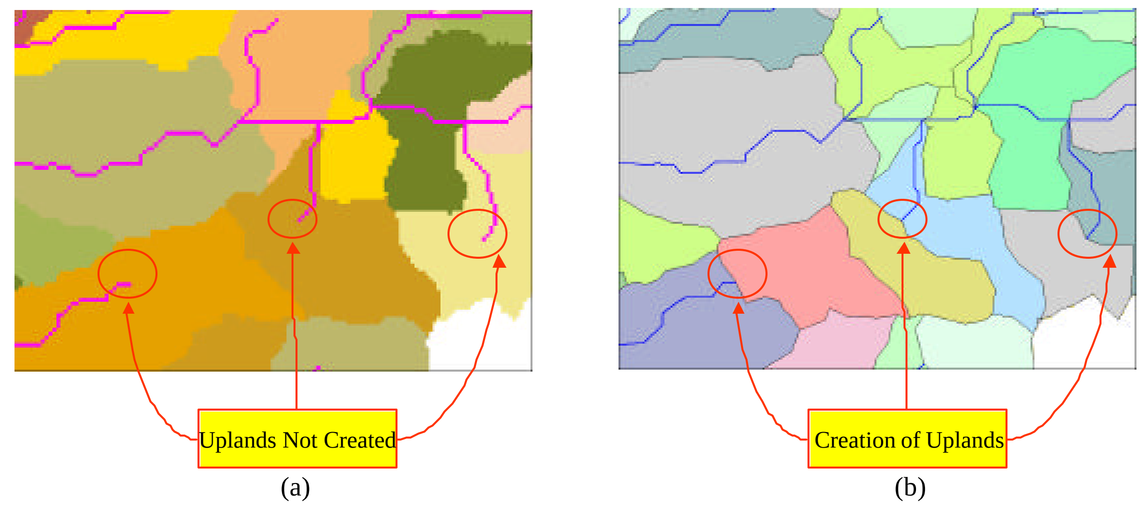

Figure 6. Upland definition as determined by AGWA. The various colored polygons and grid cells represent overland flow planes; channels are included with uniform colors. Part (a) illustrates the watershed definition provided by the core ArcView utilities; part (b) shows the derivation of upland (0-order) watersheds. Note that the AGWA watersheds have been transformed into vector shape files as evidenced by the smooth boundaries.

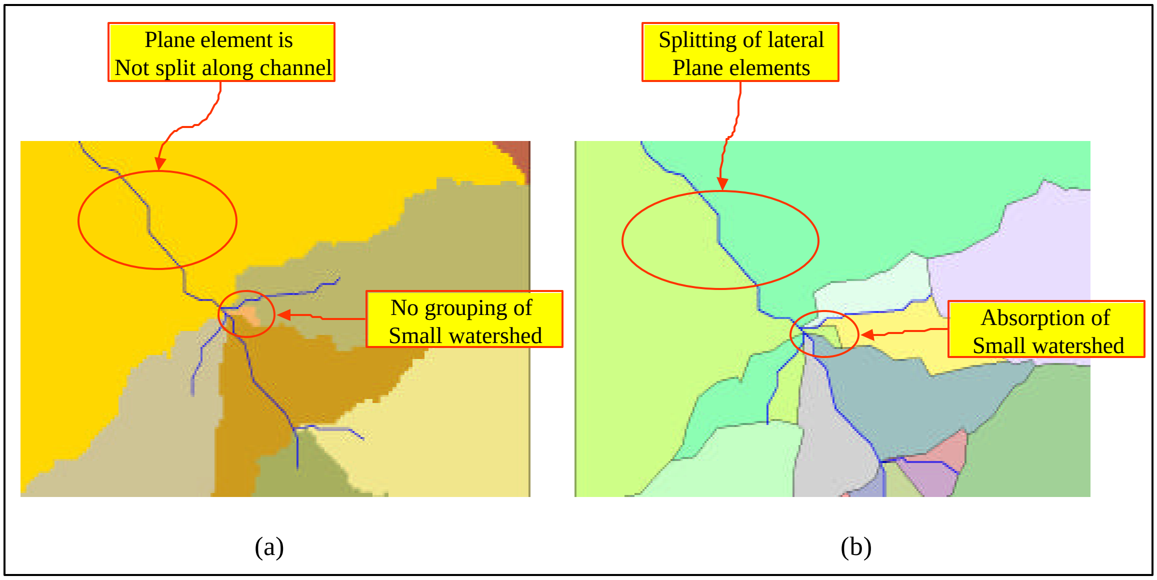

Figure 7. Illustration of watershed discretization for KINEROS using AGWA. Part (a) shows the basic delineation determined using ArcView core utilities; part (b) shows the discretization after the upland elements have been formed and the main watershed have been split into lateral planes using the stream channel as a bisector.

As shown in Figures 6 and 7 is the process of snapping and transforming data from raster to vector form, which alters the appearance of the watershed elements; there is a significant smoothing effect on watershed and channel boundaries. Small watershed elements are absorbed, as shown in Figure 7 where the orphaned element is assigned a numeric value that belongs to another watershed element. These processes serve several functions: they (1) reduce the errors associated with determining the routing scheme for hydrologic modeling; (2) remove orphaned watersheds that have little or no connectivity to channels; (3) ensure connectivity among watershed elements and channels; and (4) reduce errors in the DEM assignation of stream channel location that may occur at stream junctions.

These techniques are not entirely foolproof, and certain circumstances will cause them to fail. When the DEM grid resolution or vertical accuracy is inappropriate to adequately describe topographic detail, flow routing may fail. DEM errors resulting from common pre-processing mistakes, such as incorrect edge matching and datum differences, are described in the user manual. However, the basic premise under which AGWA operates is that the user has the basic GIS expertise to generate accurate and appropriate GIS data sets required for use by AGWA. Error trapping routines have been incorporated into the watershed discretization routines so that if AGWA encounters a situation where it cannot correctly determine the routing sequence it will alert the user to the problem and stop. Several suggestions for working around common difficulties are presented in the user manual. It is the goal of AGWA developers to create a robust software capable of operating under a wide range of topographic characteristics, and significant attention has been paid to making these routines robust, but it is acknowledged that there will be cases where AGWA cannot perform the expected tasks.

As stated previously, these subroutines have tested on a wide range of watersheds. These test watersheds have varied in scale from < 5 km^2 to over 1800 km^2 and the DEM quality has ranged from coarse (USGS 30m resolution data) to highly detailed (synthetic aperture radar derived DEM's with 2.5m resolution). Some general conclusions may be stated based from these investigations. The quality and scale of the DEM are the primary factors controlling the success of the discretization routines. Increasing DEM resolution increases the ability of AGWA to discriminate stream channel locations and accurately depict channel intersections and better distinguish flow paths. More errors occur in flat landscapes and those where flow paths converge in highly acute angles. Flow paths may be indiscriminate in flat landscapes, and acute channel junctions lead to confusion when the space between two channels is less than 2 times the DEM resolution. Steep, elongated watersheds cause trouble when the user selects a small threshold. The problem associated with small stream channels was detailed above, and elongated watersheds are especially prone to this problem because the side channels will, by definition, be relatively small. Choosing very small thresholds (< 30 acres) will yield more problems than selecting a less complex watershed. An investigation into the minimum threshold required by AGWA for a 10m USGS DEM revealed that the chances of failure dramatically increase below 20 acres, and a software limit was introduced that prevents the user from selecting a threshold less than 20 acres.

The user is advised to follow several general principles to maximize the chances for success in watershed discretization. First, acquire the highest resolution and most accurate DEM possible and properly assemble these various data into a seamless product. Second, determine a suitable threshold for the DEM resolution; poor quality DEMs require a more generalized watershed discretization. Third, remain flexible, and if AGWA runs into a problem, attempt a subsequent iteration with a slightly changed threshold value. We have found that many of the problems that AGWA cannot account for can be overcome with slight adjustment by the user. However, it should be stated the problems discussed in this section are relatively rare, and most users are unaware of the adjustments made by AGWA to account for the vagaries of GIS data.

5.2 Delineation of subwatersheds with two DEM resolutions

It is important to note that watershed delineation is dependent on the resolution of the Digital Elevation Map (DEM) and the resultant stream channels created by AGWA. For instance, in the case of Walnut Gulch, one can achieve significantly variant results through the use of higher or lower resolution DEMs. While changes can be widespread, the most visible changes occur in the area of the watershed and the construction of the stream channels.

The commonly accepted area of the Walnut Gulch Experimental Watershed is approximately 148 km^2 based on digitization of outlines from contour maps (Renard et al., 1993). However, this value is not reproduced with the use of a 30m DEM. AGWA delineates the watershed from a specified point in an active point theme located on the stream network. The resultant outline of the watershed contains 160,865 cells, each cell having an area of 900 m^2. Therefore, we obtain an area approximately equal to 145 km^2. This value can be somewhat corrected through the use of a higher quality and more detailed DEM. In a similar test, a 10m DEM gives an area of just under 148 km^2.

The stream channels created by AGWA are also affected by DEM resolution. While the general form and flow of the channels might be similar, the actual location of a stream could move on the order of meters or kilometers. Figure 8 shows the difference in stream positioning at the outlet location of Walnut Gulch. Notice the location of our point outlets. These outlet points must be post-processed by the user to ensure that the points fall on the stream channels in question.

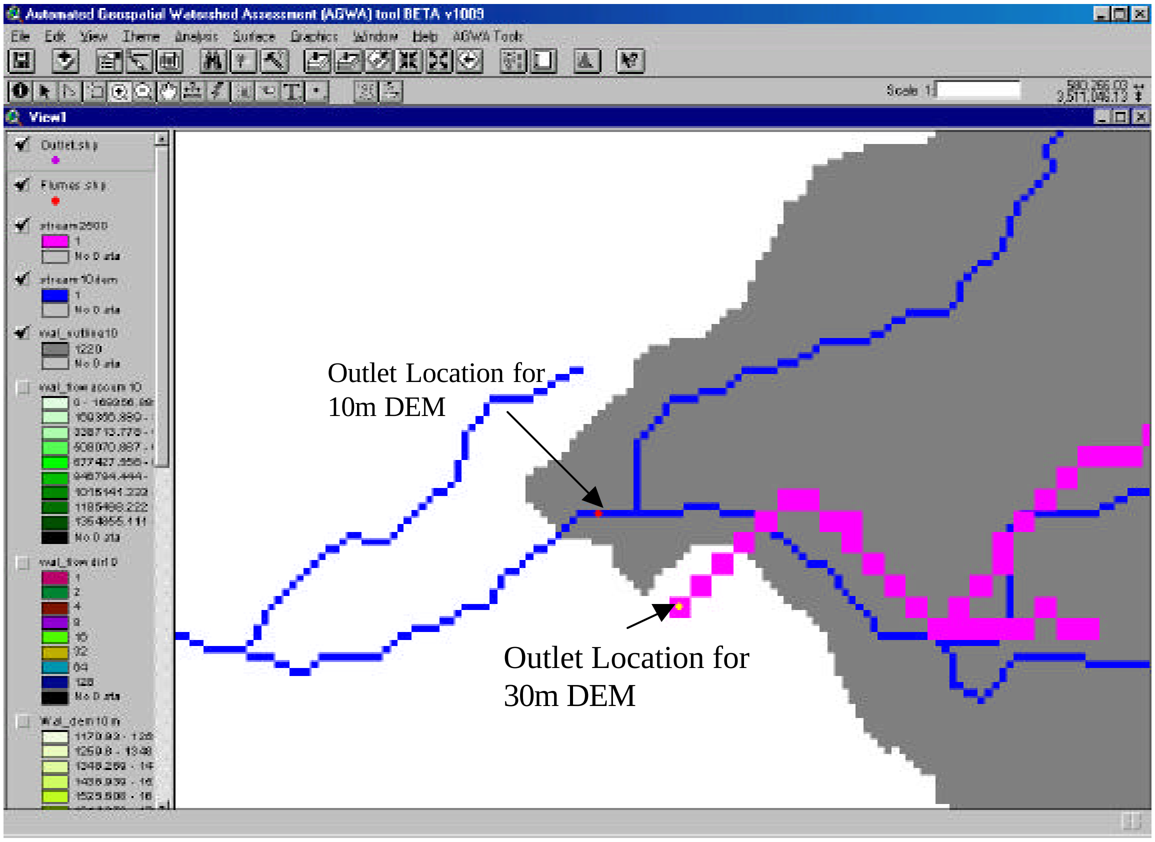

Figure 8. Blue stream was constructed using 10m DEM while the pink was created with a 30m DEM. Notice the theoretically same outlet located at different points.

While some might expect a drastically different stream network to develop, we see that rather than a completely different pattern, a 10m DEM simply provides a more extensive and detailed network, as evidenced by Figure 9

.

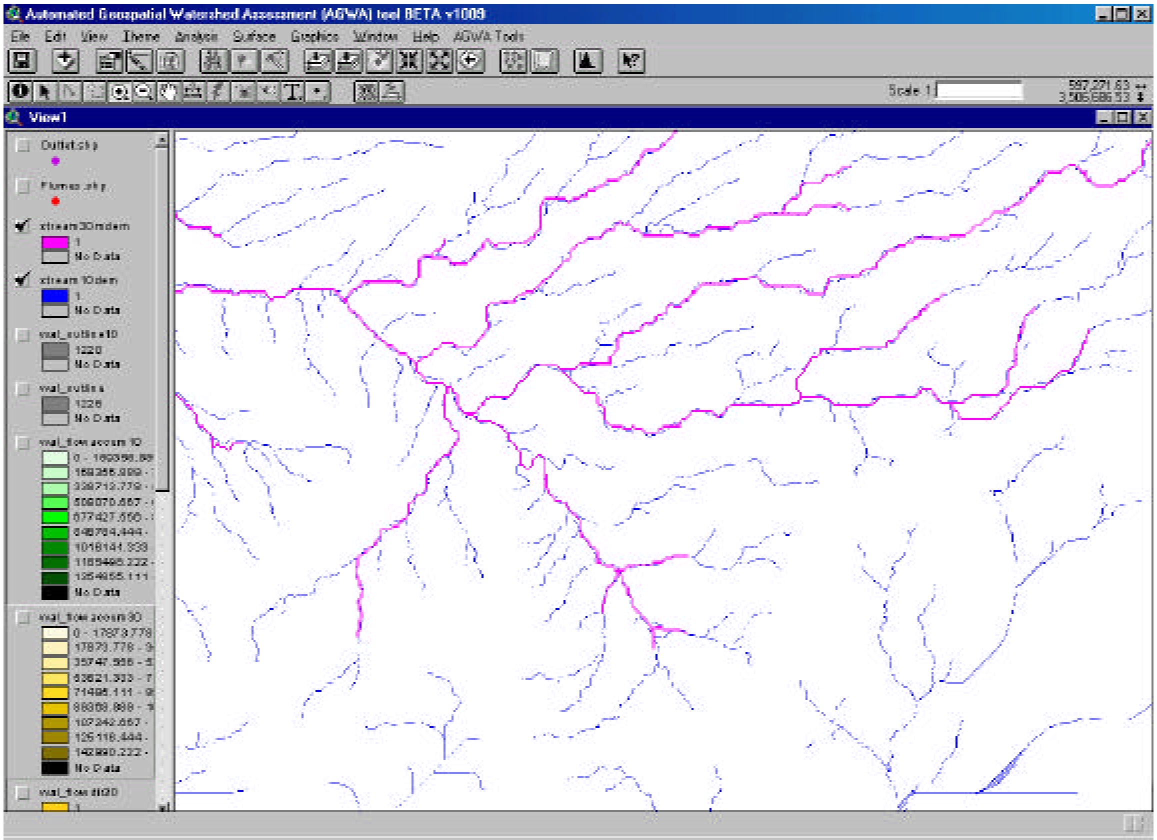

Figure 9. A more extensive stream network is constructed using a 10m DEM (in blue) in this portion of the Walnut Gulch Watershed as compared to the network derived from the 30 m DEM (in pink).

5.3 Areal Rainfall Representation

The arithmetic-mean method is the simplest method of determining areal average rainfall. It involves averaging the rainfall depth recorded at a number of gages. This method is satisfactory if the gages are uniformly distributed over the area and the individual gage measurements do not vary greatly about the mean.

If some gages are considered more representative of the area in question than others, then relative weights may be assigned to the gages in computing the areal average. The Thiessen method assumes that at any point in the watershed the rainfall is the same as that at the nearest gage so the depth recorded at a given gage is applied out to a distance halfway to the next station in any direction. The relative weights for each gage are determined from the corresponding areas of application in a Thiessen polygon network. Boundaries of the polygons are formed by the perpendicular bisectors of the lines joining adjacent gages. If there are j gages, and the area within the watershed assigned to each is A_j, and P_j is the rainfall recorded at the j^{th} gage, the areal average precipitation for the watershed is

$$\bar{P} = \frac{1}{A}\sum_{j=1}^{J} A_j P_j \qquad (8)$$

Where the watershed area $A = \sum_{j=1}^{J} A_j$. The Thiessen method is generally more accurate than the arithmetic mean method; however, it is inflexible because a new Thiessen network must be constructed each time there is a change in the gage network, such as when data are missing from one of the gages. Also, the Thiessen method does not directly account for orographic influences in rainfall.

The option to create distributed rainfall files for SWAT in AGWA uses Thiessen precipitation weighting to generate SWAT input files. The user must have three items to complete this process: the watershed discretization; a point theme of rain gage locations; and an unweighted daily precipitation database file. Specific requirements for the point theme and unweighted precipitation file are described in the AGWA User's Manual. Generating the weighted precipitation file proceeds in four steps: 1) constructing the Thiessen polygons, 2) intersecting the Thiessen polygons with the watershed discretization (subwatershed elements), 3) computing gage weights for each subwatershed, and 4) using the weights to compute weighted depths for each subwatershed for each day, which are written to a *.pcp file. Each of these steps in the weighting process are described and illustrated below.

Generating Thiessen Polygons:

A hypothetical watershed being represented by a single plane element and 30 rain gages were regularly placed in a square grid to test the algorithms (Figure 10). According to the Thiessen polygon method, the rain gages are joined with straight lines in order to form a pattern of triangles. Perpendicular bisectors to the sides of these triangles are drawn to enclose each station within a polygon circumscribing an area of influence. Since the rain gages lie on a regular square grid, then the Thiessen polygons are equal, regular cells with each side equal to the grid spacing (Figure 11).

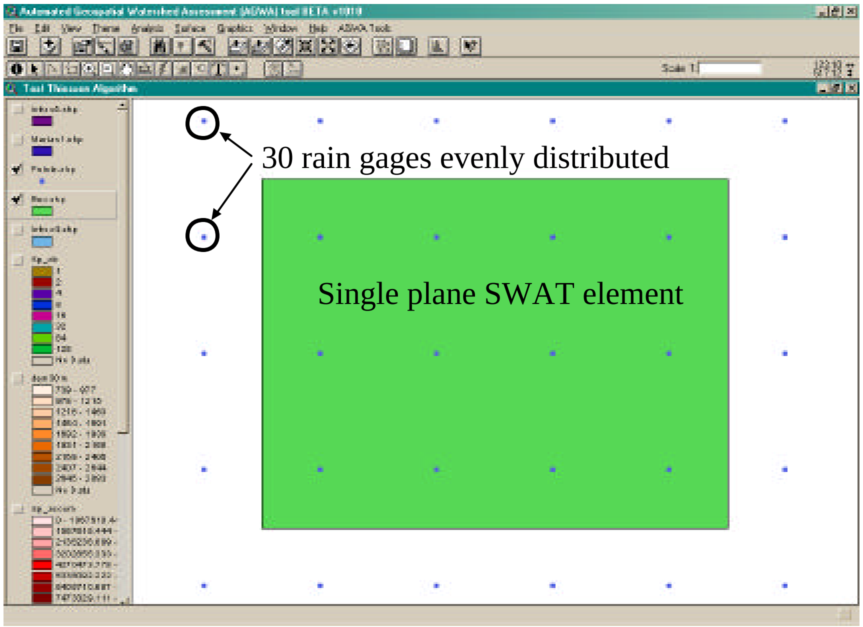

Figure 10. Hypothetical watershed with 30 rain gages regularly placed.

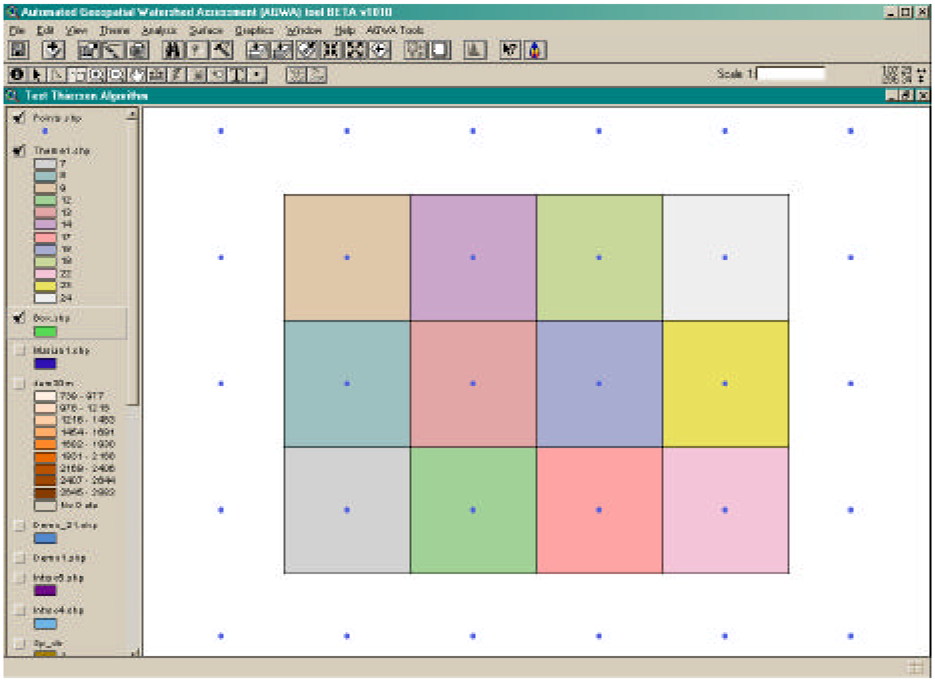

Figure 11. Weighted area using the Thiessen polygon algorithm developed in AGWA.

It is important to note that polygons are generated only for gages with gage IDs that are found in both the point theme attribute table and the unweighted precipitation database file. In this manner AGWA avoids generating polygons for rain gages that do not have data. The user is notified which gages are and are not used via a text box which pops up prior to creating the Thiessen polygons.

Intersecting Subwatershed Elements and Thiessen Polygons

The process of intersecting the two polygon themes (Thiessen and subwatershed) involves using a predefined function of ArcView Spatial Analyst (Environmental Systems Research Institute, 1996). As such, it is deemed unnecessary to demonstrate its accuracy in this report. An example of the resulting intersection theme, however, is presented in Figure 12. The following demonstration of the remaining two steps in the precipitation weighting process are based on this configuration.

Computing Gage Weights:

Once the Thiessen polygons and watershed discretization have been intersected AGWA uses the attribute table of the intersection theme to compute the weight influence of each rain gage in each subwatershed and writes the results to a database file called weights.dbf. More specifically, AGWA divides the area of each polygon by the summed area of all polygons with a common subwatershed ID to get the area weighted influence of each gage in a subwatershed as a percentage. To confirm that these computations proceed as expected, they have been reproduced in a spreadsheet and are illustrated in Table I.

Computing Weighted Precipitation Depths

The final step in creating a distributed rainfall input file for SWAT is computing the weighted precipitation depths for each subwatershed on each day during the simulation period. This is accomplished in AGWA by first selecting all records in weights.dbf (Table II) for a subwatershed. For each gage in this list the gage depth (GD) for that day from the unweighted precipitation file (Table III) is multiplied by the appropriate gage weight (GW) from weights.dbf. The weighted depths from each gage are then summed to get the total weighted depth (WD) for the subwatershed that day.

$$WD = \sum \left(GD * GW \right) \tag{9}$$

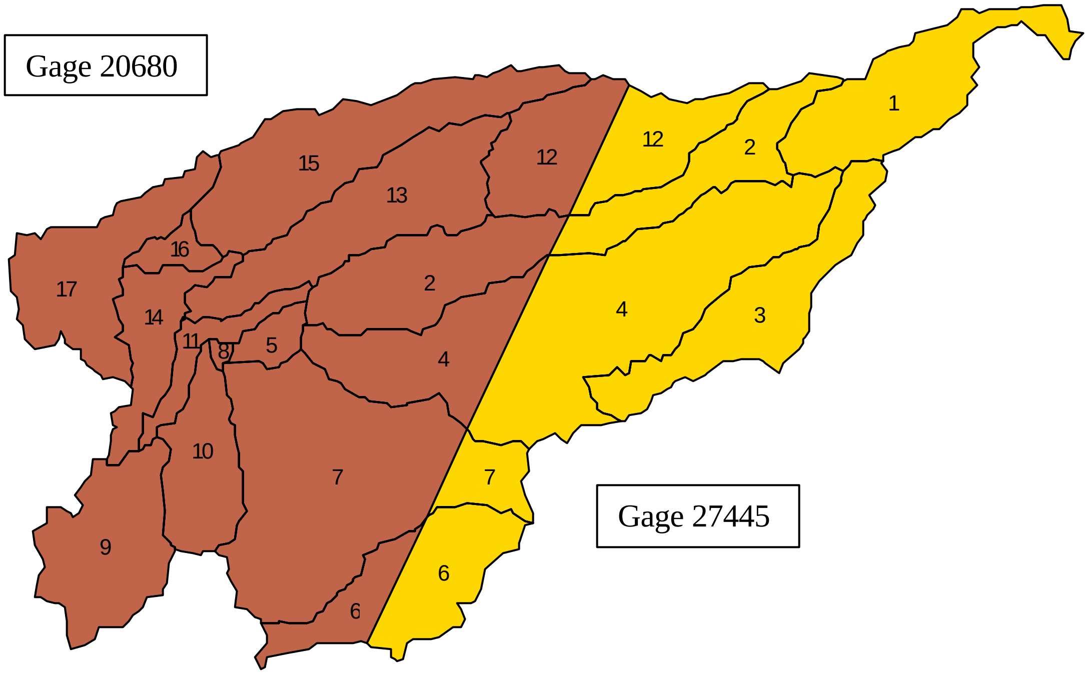

Figure 12. An example watershed configuration with numbered subwatershed elements that has been intersected with Thiessen polygons. Gage and subwatershed numbers correspond with the example.

Table I. Demonstration of the area weighting computations in AGWA.

SWS ID	Station ID	Area (m^2)	Computed weight	Sum to 1?	AGWA computed weight	Difference (AGWA - Excel)
1	27445	6401590.1	1.000		1.000	0.000
2	20680	5740600.8	0.562	1.000	0.562	0.000
2	27445	4476205.8	0.438		0.438	0.000
3	27445	6277639.0	1.000		1.000	0.000
4	20680	5542077.4	0.295	1.000	0.295	0.000
4	27445	13275690.6	0.705		0.705	0.000
5	20680	998725.4	1.000		1.000	0.000
6	20680	2264974.3	0.362	1.000	0.362	0.000
6	27445	3984728.9	0.638		0.638	0.000
7	20680	13878808.3	0.885	1.000	0.885	0.000
7	27445	1805464.9	0.115		0.115	0.000
8	20680	164240.3	1.000		1.000	0.000
9	20680	6015917.7	1.000		1.000	0.000
10	20680	3905974.1	1.000		1.000	0.000
11	20680	1677341.2	1.000		1.000	0.000
12	20680	4190969.0	0.539	1.000	0.539	0.000
12	27445	3589124.2	0.461		0.461	0.000
13	20680	7466539.1	1.000		1.000	0.000
14	20680	3249121.9	1.000		1.000	0.000
15	20680	8406013.1	1.000		1.000	0.000
16	20680	922381.4	1.000		1.000	0.000
17	20680	6033997.8	1.000		1.000	0.000

The frequent presence of "nodata" or missing values in data derived from National Weather Service (NWS) archives complicates the weighting calculations. AGWA interprets all negative rainfall depths to represent missing data, and uses the following logic to account for these values as it cycles through all the gages for a subwatershed:

- If all the gages have data, then $WD = WD + GD * GW$ as described above.
- If all the gages for a day have either zero values or no data then, WD = 0 for all subwatersheds. No computations are made that day.
- If one or more gages in the watershed have no data (and at least one non-zero value) then:
  - If all gages for a subwatershed have no data then find the closest gage with data:
    - If only one gage in the watershed with data that day, then WD = depth from the gage with data for all subwatersheds.
    - If the number of gages with data is > 1 and < the number of gages in the watershed, then WD = depth from the gage with data that is closest to the centroid of the subwatershed.
  - If some but not all of the gages intersecting a subwatershed have data:
    - If GD < 0, then WD = WD (i.e. the weighted depth for a subwatershed is not affected by no data values).
    - If GD > 0 then $WD = WD + GD * \frac{GW}{1 - BW}$ where BW = bad weight, the sum of all the gage weights for a subwatershed for which the corresponding gage depths are missing data (GD < 0). Thus, in this situation the gages with missing data are excluded from the weighted depth calculations entirely.

Table II. Subwatershed gage weights

SWS ID	Gage ID	Weight	SWS ID	Gage ID	Weight
1	G27445	1.000	8	G20680	1.000
2	G20680	0.562	9	G20680	1.000
2	G27445	0.438	10	G20680	1.000
3	G27445	1.000	11	G20680	1.000
4	G20680	0.295	12	G20680	0.539
4	G27445	0.705	12	G27445	0.461
5	G20680	1.000	13	G20680	1.000
6	G20680	0.362	14	G20680	1.000
6	G27445	0.638	15	G20680	1.000
7	G20680	0.885	16	G20680	1.000
7	G27445	0.115	17	G20680	1.000

Table III. Unweighted daily precipitation (mm) – NWS raw data.

DAY	G20680	G27445
1	7.0	190.0
2	0.0	10.0
3	-999.0	10.0
4	-9999.0	-999.0
5	0.0	-999.0
6	0.0	0.0

5.4 Design Storms

The AGWA tool has a function to create design storm input data for KINEROS for select events in southern Arizona. Historical rainfall records from the USDA-ARS Walnut Gulch Experimental Watershed have been analyzed to determine a number of return-period events at the point scale (Osborn et al., 1985). Following Osborn et al. (1985) we have included a rainfall generator for the following the 5, 10, and 100 year return periods for the 30 and 60 minute storms, yielding a data set of 9 events.

Applying point estimates for design storms across larger areas tends to lead to the overprediction of runoff due to the lack of spatial heterogeneity in input data. An area-reduction method developed by Osborn et al. (1980) has been implemented in the AGWA tool to reduce rainfall input. Osborn et al. (1980) developed curves relating the increase in watershed area to the factor by which rainfall must be reduced to more closely mimic reality. These relationships have been re-created within AGWA, such that the rainfall data are reduced according to the overall size of the watershed being simulated before being input to the KINEROS model.

5.5 Validation and Sensitivity Analysis

EPA uses hydrologic models to manage watersheds around the country for which poor or no monitoring data exist. Without long-term records it is impossible to produce coherent and defensible management goals. The lack of real observations means that instead of credible, watershed–specific information, models are forced to rely on default values. These defaults may be based purely on expert judgment or the outcome of limited field experiments or simulations.

The calibration procedure aims at estimating parameter values that cannot be assessed directly from field data. However, calibration may produce parameter datasets that can achieve the same degree of simulation matching to monitoring data. Because models contain many variables, there are an unlimited number of scenarios that will yield the same result.

During the calibration procedure, different accuracy criteria can be used to compare the simulated and measured data. This allows us to define an objective measure of the goodness of fit associated with each set of model parameters and estimate the parameter values which provide the best overall agreement between model output and measured data. The performance criteria were related to annual runoff volume measured at the outlet of the watershed with a graphical assessment of observed and simulated supported by the efficiency coefficient, E, developed by Nash and Sutcliffe (1970).

$$E = \frac{\sum_{i=1}^{n}(x_i - m)^2 - \sum_{i=1}^{n}(x_i - e_i)^2}{\sum_{i=1}^{n}(x_i - m)^2} \tag{10}$$

Where x_i is the observed runoff volume of the i^{th} year, m the observed average, and e_i the estimated annual runoff volume. E values over 0 indicate the efficiency of the model is better than the average of observed runoff volume. A value of 1 indicates a perfect model fit. A negative value of E indicates the model is performing more poorly than simply using the average of the observed data.

The calibration procedure was undertaken using a nonlinear parameter estimator program PEST (Doherty, 1994). PEST can adjust model parameters in order that the discrepancies between the pertinent simulated runoff volume numbers and the corresponding observed measurements are reduced to a minimum. It does this by taking control of the hydrologic model and running it as many times as is necessary in order to determine this optimal set of parameters. PEST uses a nonlinear estimation technique known as the Gauss-Marquardt-Levenberg method (Marquardt, 1963).

PEST requires that the upper and lower bounds be supplied for each parameter. This information is vital because it informs PEST of the range of permissible values that each parameter can take; hence preventing the calibration from producing a solution with non-realistic parameter values.

The SWAT model was calibrated by reducing the discrepancies between model outputs (annual runoff volume) and field observations to a minimum in the weighted least squares sense. The differences between field measurements and model outputs were encapsulated in an objective function defined as the weighted sum of squared deviations between field observations and corresponding model outputs.

Once the model is calibrated with plausible parameter values, the model is validated to ensure that it can be used for prediction. In the validation process, the model is tested against data different from those used for the calibration. This implies the application of the calibrated model without changing parameter values that were set during the calibration period. The model is validated if its accuracy and predictive capability in the validation period have been proven to lie within acceptable limits or to provide acceptable errors as specified in the performance criteria.

Average annual simulated runoff volume for a 15 year run (1966 – 1980) was calibrated against average annual measured runoff volume at the outlet of the watershed. Observed and modeled (calibrated) volume is shown in Figure 13. Total water yield for the entire watershed was 2.88 mm and 2.60 mm simulated by SWAT. The efficiency coefficient yielded a value of 0.68.

Twelve years of runoff volume data outside the calibration period were available for validation. The efficiency coefficient yielded a value of 0.30 with mean annual measured and simulated runoff volume of 1.99 mm and 1.50 mm, respectively. Figure 14 shows the measured and predicted annual time series from 1981 to 1992.

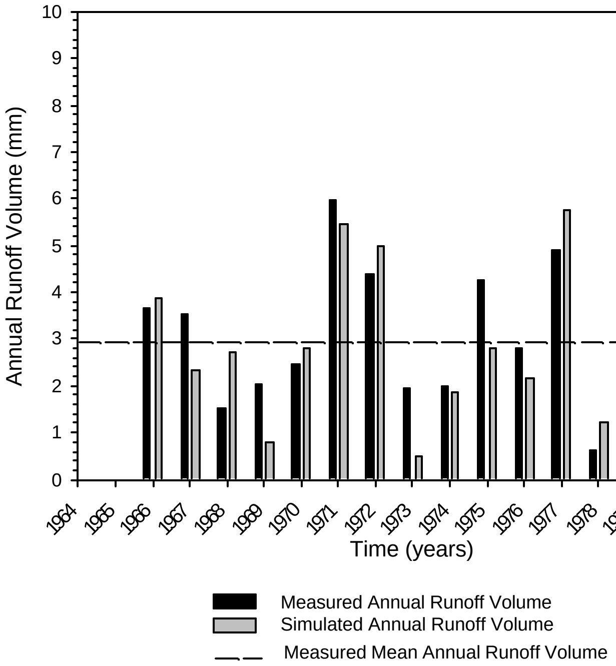

Figure 13. Measured and simulated annual runoff (calibration) for the Walnut Gulch Experimental Watershed.

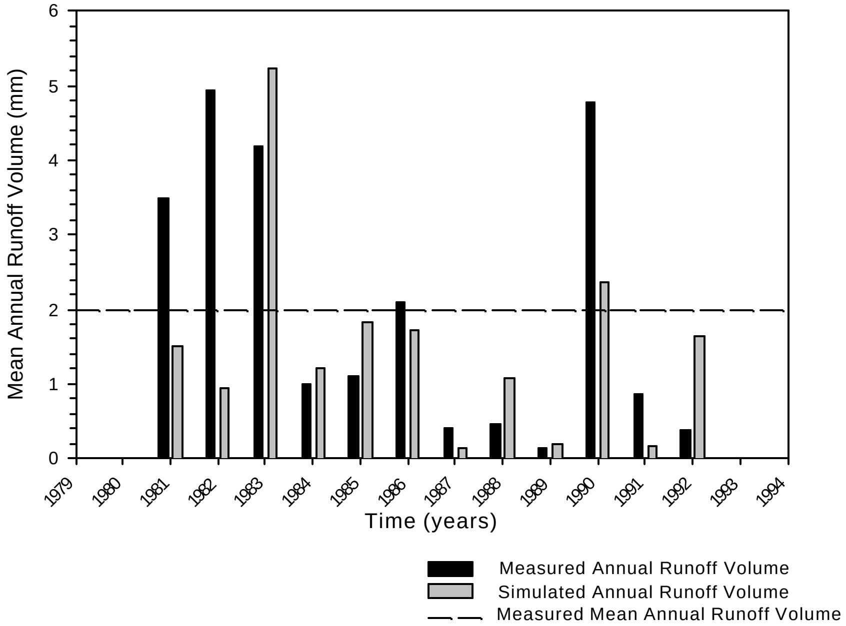

Figure 14. Measured and simulated annual volume (validation) for the Walnut Gulch Experimental Watershed.

The initial CN set obtained from Urban Hydrology for Small Watersheds (U. S. Department of Agriculture, 1986), the best-fit CN parameter set, and the maximum and minimum CN sets are presented in Table IV.

Table IV. Maximum and minimum CN values that minimize the objective function.

Element ID	Initial CN	Best-fit CN	Maximum	Minimum
1	80	65	74	62
2	79	84	85	83
3	80	81	81	81
4	85	62	62	62
5	85	85	85	85
6	80	69	69	67
7	85	62	62	62
8	80	64	64	62
9	85	62	72	62

San Pedro River Basin

Average annual simulated runoff volume for a 14 year run (1960-1973) was calibrated against average annual measured runoff volume at the Charleston USGS stream gages in the Upper San Pedro River Basin. Observed and modeled (calibrated) flow is shown in Figure 15. Total water yield for the watershed at the USGS stream gages was 11 mm and 10.17 mm simulated by SWAT. The efficiency coefficient yielded a value of 0.44.

The results of the calibration indicate that by minimizing the differences between measured and simulated runoff volume using eight rain gages that lie within the basin, the performance of the model is fairly low, that is, an efficiency coefficient of 0.44. Consequently, when no calibration is undertaken the performance of the model will drop.

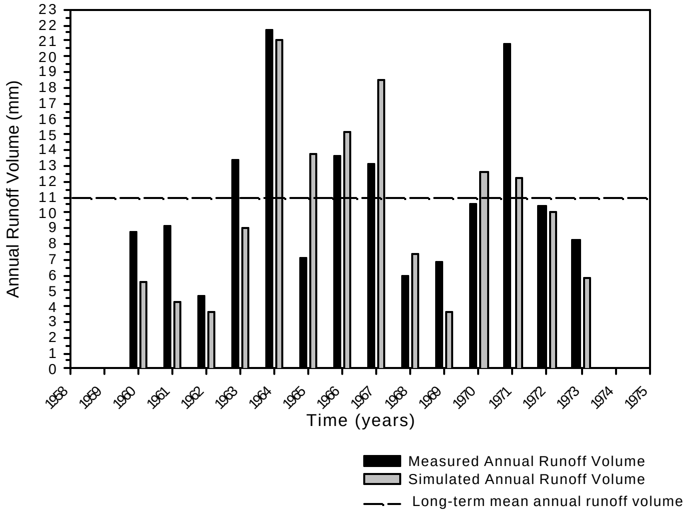

Figure 15. Measured and simulated annual runoff (calibration) volume for the San Pedro River Basin using SWAT.

A sensitivity analysis of KINEROS was performed on 10 variables that affect runoff and sediment yield. Each of the variables was allowed to float within ± 30% of the estimated parameter value from the lookup tables. An original KINEROS parameter file was created for 2 watersheds: watershed 2 on Walnut Gulch and the San Pedro watershed. A program was written to iteratively decrease and increase the individual parameters in ± 5% increments. The same 10-year, 60-minute return period event was used as input. First the uplands were adjusted and the channels left alone. Next, the channels were adjusted and the uplands left alone. Last, both the uplands and channels were adjusted in ± 5% increments. The table shows the maximum percent change resulting from altering the parameters within the ± 30% window. In Table V, the results are presented. Notice that the most sensitive parameter is hydraulic conductivity, which is related to soil and vegetation.

Table V. Sensitivity analyses of KINEROS on subwatershed near Sierra Vista.

	Runoff			Sediment Yield		
Parameter	Max Change Upland	Max Change Channels	Max Change Both	Max Change Upland	Max Change Channels	Max Change Both
Coh	0.00	0.00	0.00	3.06	0.00	3.06
Cov	2.42	0.00	2.42	35.22	0.00	35.22
Dist	0.00	0.00	0.00	0.00	0.06	0.06
G	145.29	2.67	151.15	190.68	1.73	195.70
Ks	228.05	42.05	558.33	299.75	34.11	468.14
Mann (n)	0.00	51.65	51.65	0.00	46.58	46.58
Pave	0.00	0.00	0.00	0.00	0.00	0.00
Por	153.18	9.61	190.52	207.09	5.55	232.39
Rock	86.70	0.00	86.70	118.64	0.00	118.64
Splash	0.00	0.00	0.00	3.80	0.00	3.80

6 Summary and Conclusions

An interface was developed for KINEROS and SWAT models using the ArcView GIS system. It consists of 3 key components: (1) A preprocessor generating subbasin topographic parameters and model input parameters, (2) editing input data and simulation execution, (3) A postprocessor viewing of graphical and tabular results. The preprocessor interface automatically subdivides a basin and then extracts model input data from map layers and associated relational databases for each subbasin. Soils, land use, weather, management, and topographic data are collected and written to appropriate model input files. The output interface allows the user to display output maps by selecting a subbasin from a GIS map.

The evaluation of AGWA was carried out primarily in the automatic extraction of parameter values from readily available databases to generate the input file for the hydrologic models. A key feature in the interface is the delineation of a watershed into subwatersheds and parameterization of each individual subwatershed. Based on the test performed using a 30-m and 10-m DEMs, the results show that the location of the outlet of the watershed may be off of its real location by approximately 20 cells.

When AGWA was applied to the Walnut Gulch Experimental Watershed and the San Pedro River Basin in Arizona, the results showed that the efficiency coefficients for the calibration and validation periods for the SWAT model on the Walnut Gulch were 0.68 and 0.30, respectively. Based on Figures 13 and 14, the performance of the model for annual prediction of runoff volume is relatively poor; however, if the long-term mean annual runoff volume is calculated, the model performance is within 10% error. That is, the observed long-term mean annual runoff volume, based on a 15-year record, is 2.88 mm compared to 2.60 mm simulated by SWAT. Furthermore, for the validation period, the observed long-term mean annual runoff volume is 1.99 mm, based on a 12-yr record, and the simulated by SWAT is 1.50mm. Resulting in an error of estimation of 25%.

The efficiency coefficient for the calibration for the SWAT model on the Upper San Pedro River Basin was 0.44. The poor performance of the model can be attributed to the fact that only eight rain gages were available to characterize the spatial variability of rainfall on a larger watershed. However, the long-term mean annual runoff volume measured at the USGS Charleston stream gage was 11 mm and the simulated by SWAT was 10.17 mm. Therefore, the error of estimation is approximately 8%.

In conclusion, the AGWA tool has tested under different scenarios to ensure that it conforms to the design objectives and specifications and that it correctly performs the incorporated functions. Potential scenarios were identified where AGWA may fail to delineate the watershed. The problem arises when AGWA attempts to derive the flow paths and, subsequently, subdivide the study watershed into channel and plane elements. Error trapping is performed to determine the cause of failure. In addition, these problems are illustrated in the AGWA User Manual so the user may determine the cause and find a solution for their particular problem.

Based on the calibration analysis, it is concluded that the application of the AGWA tool is best suited for scenarios where the user is interested in evaluating the effects of *relative* impacts resulting from land cover change on surface runoff.

References

American Society of Civil Engineers, T. C. 1999. GIS Modules and Distributed Models of the Watershed. American Society of Civil Engineers, Reston, Virginia. pp. 120p

Arnold, J. G., J. R. Williams, R. Srinivasan, K. W. King, and R. H. Griggs. 1994. SWAT: Soil Water Assessment Tool. U. S. Department of Agriculture, Agricultural Research Service, Grassland, Soil and Water Research Laboratory, Temple, TX.

Bales, J., and R. Betson. 1982. The curve number as a hydrologic index. Pages 371-386 *in* V. P. Singh, editor. Rainfall-Runoff Relationships. Water Resources Publications, Highlands Ranch, CO.

Beven, K. J. 2000. Rainfall-Runoff Modelling. John Wiley &Sons Ltd., West Sussex, England.

Bouwer, H. 1966. Rapid field measurement of air entry and hydraulic conductivity as significant parameters in flow systems analysis. Water Resources Research, **2**:279-238.

Doherty, J. 1994. Model Independent Parameter Estimation. *in*. Watermark Computing, Australia.

Environmental Systems Research Institute, 1996. ArcView, Version 3.2. ESRI. Inc., 380 New York Street, Redlands , CA 92373.

Faures, J. M., D. C. Goodrich, D. A. Woolhiser, and S. Sorooshian, 1995. Impact of small scale spatial rainfall variability on runoff modeling. Journal of Hydrology, **173**: 309-326.

Garbrecht, J., and P. J. Starks. 1995. Notes on the use of USGS Level 1 & 0.5 minute DEM coverages for landscape drainage analysis. Photogrammetric Engineering and Remote Sensing, **61** (5):519-522.

Goodrich, D. C. 1991. Basin scale and runoff model complexity. Department of Hydrology and Water Resources Technical Report No. HWR 91-010, University of Arizona, Tucson, AZ. pp. 361.

Goodrich, D. C., J. M. Faures, D. A. Woolhiser, L. J. Lane, and S. Sorooshian. 1995. Measurement and analysis of small-scale convective storm rainfall variability. Journal of Hydrology, **173**:283-308.

Goodrich, D. C., L. J. Lane, R. A. Shillito, S. N. Miller, K. H. Syed, and D. A. Woolhiser. 1997. Linearity of basin response as a function of scale in a semi-arid watershed. Water Resources Research, **33** (12):2951-2965.

Goodrich, D. C., T. J. Schmugge, T. J. Jackson, C. L. Unkrich, T. O. Keefer, R. Parry, L. B. Bach, and S. A. Amer. 1994. Runoff simulation sensitivity to remotely sensed initial soil water content. Water Resources Research, **30** (5):1393-1405.

Hjelmfelt, A. T., L. A. Kramer, and R. Burwell. 1982. Curve numbers as random variables. Pages 365-370 *in* V. P. Singh, editor. Rainfall-Runoff Relationships. Water Resources Publications, Highlands Ranch, CO.

Jury, W. A. 1985. Spatial variability of soil physical parameters in solute migration: A critical literature review. EPRI Topical Report EA 4228, Electric Power Research Institute, Palo Alto, California.

Lunetta, R. S., J. G. Lyon, B. Guindon, and C. D. Elvidge. 1993. North American Landscape Characterization: "Triplicate" Data Sets and Data Fusion Products. *in* R. S. Lunneta and C. D. Elvidge, editors. Remote Sensing Change Detection:

Environmental Monitoring Methods and Applications. Ann Arbor Press, Chelsea, Michigan.

Marquardt, D. W. 1963. An algorithm for least-squares estimation of nonlinear parameters. Journal of the Society of Industrial and Applied Mathematics, **11** (2):431-441.

Michaud, J. D., and S. Sorooshian. 1994. Comparison of simple versus complex distributed runoff models on a midsized semi-arid watershed. Water Resources Research, **30** (3):593-605.

Mishra, S. K., and V. P. Singh. 1999. Another look at SCS-CN method. Journal of Hydrological Engineering, **4**:257-264.

Nash, J. E., and J. V. Sutcliffe. 1970. River flow forcasting through conceptual models. Journal of Hydrology, **10**:282-290.

Osborn, H. B., L. J. Lane, and V. A. Myers. 1980. Rainfall/watershed relationships for Southwestern thunderstorms. Transactions of the American Society of Civil Engineers, **23** (1):82-91.

Osborn, H. B., C. L. Unkrich, and L. Frykman. 1985. Problems of simplification in hydrologic modeling. Hydrology and Water Resources in Arizona and Southwest, **15**:7-20.

Ostman, A. 1987. Accuracy estimation of digital elevation data banks. Photogrammetric Engineering and Remote Sensing, **53** (4):425-430.

Rawls, W. J., D. L. Brakensiek, and K. E. Saxton. 1982. Estimation of soil water properties. Transactions of the American Society of Agricultural Engineers, **25** (5):1316-1320,1328.

Renard, K. G., L. J. Lane, J. R. Simanton, W. E. Emmerich, J. J. Stone, M. A. Weltz, D. C. Goodrich, and D. S. Yakowitz. 1993. Agricultural impacts in an arid environment: Walnut Gulch studies. Hydrological Science and Technology, **9** (1):145-190.

Smith, R. E., D. C. Goodrich, D. A. Woolhiser, and C. L. Unkrich. 1995. KINEROS:A kinematic runoff and erosion model. Pages 697-732 *in* V. P. Singh, editor. Computer Models of Watershed Hydrology. Water REsources Publications, Highlands Ranch, Colorado.

Smith, R. E., and J.-Y. Parlange. 1978. A parameter-efficient hydrologic infiltration model. Water Resources Research, **14** (33):533-538.

Srinivasan, R., J. G. Arnold, R. S. Muttiah, C. Walker, and P. T. Dyke. 1993. Hydrologic Unit Model for United States (HUMUS). Pages 451-456 *in* Advances in Hydro-Science and Engineering. University of Mississippi.

Stone, J. J., L. J. Lane, and E. D. Shirley. 1992. Infiltration and runoff simulation on a plane. Transactions of the American Society of Agricultural Engineers, **35** (1):161-170.

Syed, K. H. 1999. The Impact of Digital Elevation Model Data Type and Resolution on Hydrologic Modeling. PhD. University of Arizona, Tucson, Arizona.

Tarboton, D. G., R. L. Bras, and I. Rodrigues-Iturbe. 1991. On the extraction of channel networks from digital elevation data. Water Resources Research, **5** (1):81-100.

Taylor, J. K. 1985. What is Quality Assurance? Pages 5-11 *in* J. K. Taylor and T. W. Stanley, editors. Quality Assurance for Environmental Measurements. American Society for Testing and Materials, Special Technical Publication No. 867, Philadelphia, Pennsylavania.

U. S. Department of Agriculture, S. C. S. 1986. Urban hydrology for small watersheds. Techical Release 55, Washington, D. C.

U. S. Environmental Protection Agency, 1991. Quality Assurance Guidelines for Modeling, Development and Application Projects, Environmental Research Laboratory, Duluth, Minnesota.

U. S. Geological Survey, Dept. of Interior. 1990. Digital Elevetion Models: Data users guide., Reston, Virginia. pp. 38p.

van der Heijde, P. K. M. 1987. Quality Assurance in Computer Simulations of Groundwater Contamination. Environmental Software, **2** (1):19-28.

van der Heijde, P. K. M., A. I. El-Kadi, and S. A. Williams. 1988. Groundwater Modeling: An Overview and Status Report, Ada, Oklahoma.

Woolhiser, D. A., R. E. Smith, and D. C. Goodrich. 1990. KINEROS: A kinematic runoff and erosion model. U. S. Department of Agriculture, Agricultural Research Service.