# DEVELOPMENT OF AN ARCGIS INTERFACE AND DESIGN OF A GEODATABASE FOR THE SOIL AND WATER ASSESSMENT TOOL

A Thesis

by

MILVER ALFREDO VALENZUELA ZAPATA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2003

Major Subject: Civil Engineering

**DEVELOPMENT OF AN ARCGIS INTERFACE AND DESIGN OF A GEODATABASE FOR THE SOIL AND WATER ASSESSMENT TOOL**

A Thesis

by

MILVER ALFREDO VALENZUELA ZAPATA

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Approved as to style and content by:

| | |
|---|---|
| Francisco Olivera<br>(Chair of committee) | Ralph Wurbs<br>(Member) |
| Raghavan Srinivasan<br>(Member) | Paul N. Roschke<br>(Head of Department) |

December 2003

Major Subject: Civil Engineering

## ABSTRACT

Development of an ArcGIS Interface and Design of a Geodatabase

for the Soil and Water Assessment Tool. (December 2003)

Milver Alfredo Valenzuela Zapata, B.S., Pontificia Universidad

Católica del Ecuador

Chair of Advisory Committee: Dr. Francisco Olivera

This project presents the development and design of a comprehensive interface coupled with a geodatabase (ArcGISwat 2003), for the Soil and Water Assessment Tool (SWAT). SWAT is a hydrologically distributed, lumped parameter model that runs on a continuous time step. The quantity and extensive detail of the spatial and hydrologic data, involved in the input and output, both make SWAT highly complex. A new interface, that will manage the input/output (I/O) process, is being developed using the Geodatabase object model and concepts from hydrological data models such as ArcHydro. It also incorporates uncertainty analysis on the process of modeling. This interface aims to further direct communication and integration with other hydrologic models, consequently increasing efficiency and diminishing modeling time. A case study is presented in order to demonstrate a common watershed-modeling task, which utilizes SWAT and ArcGIS-SWAT2003.

# DEDICATION

To God, my wife and my family.

## ACKNOWLEDGMENTS

This research was conducted at Texas A&M University and was supported by the U.S. Army Corp of Engineers (Forth Worth District) through the Texas Agricultural Experiment Station (TAES) as part of the project "Strengthening water resources planning for Texas: Development and application of a decision support system for water resources modeling for the Brazos River Basin."

I would like to thank my committee members: Dr. Francisco Olivera, Dr. Raghavan Srinivasan and Dr. Ralph Wurbs for their trust, guidance and contributions throughout this research.

I also wish to thank my fellow classmates and friends for their active help and contribution: Srikanth Koka, Ashish Awral, Rajeev Raina, Zubin Zukheswalla, Janghwoan Choi, and Joshua Peschel.

I am thankful for my parents' support along this journey called life.

Special thanks to my wife, whose love and patience illuminated my road.

# TABLE OF CONTENTS

Page

## LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

# LIST OF TABLES

# 1 INTRODUCTION

Non-point source of pollution (NPS) has been given increasing importance in the assessment of the quality of water bodies in United States (Environmental Protection Agency 2003a). Quantifying the effect of non point source pollution is a difficult task. To address this difficulty, several hydrologic models have been developed. In order to give a comprehensive representation of how the system is behaving, the majority of these models rely on each other's results. Therefore, direct communication between hydrologic models is essential.

Links between hydrologic models can be established by using the same data model. However, each hydrologic model has its own data model. The solution to this problem has been to create an interface, which serves as an intermediary to translate one data model into the other. This process results in a series of model links and thus an inefficient system of communication due to the required number of links.

The ideal solution is to have every data model translated into a common standard data model. This new standard could be based on the Geodatabase data model and some concepts of the ArcHydro data model (Maidment, 2002). The Geodatabase data model represents the new paradigm for organizing data and facilitating its analysis in GIS. ArcHydro is a GIS-based hydrologic data model, whose structure is meant to support

---

This thesis follows the style and format of *Journal of Hydrologic Engineering*. ASCE.

geospatial hydrologic features and temporal data. The power of ArcHydro relies on the predefined relationships between spatial features and temporal data (Maidment, 2002). ArcHydro is a step in the process of obtaining this common standard for dealing with hydrologic data.

This project presents the methodology followed for the development of an interface for SWAT that works under a Geographic Information Systems (GIS) environment. This interface extracts, analyzes and manages the required input and the produced output of geospatial information, under the structure of the geodatabase object model. It also follows the current paradigm on Information and Technology: *Component Object Model* (COM), ensuring that its application can be expanded easily with the addition of components from different sources.

SWAT is a comprehensive hydrologic model that has been used and validated during the last decade to simulate the hydrology in both large and small watersheds (Arnold and Allen, 1996, Saleh *et al,* 2000, Eckhardt and Arnold, 2001, Muttiah and Wurbs, 2001).

Due to its physical base, SWAT requires detailed information from extensive input datasets derived from a hydrological system. The effectiveness of the SWAT model relies on the availability and completeness of these input datasets. Therefore, estimation of parameters for missing input data is a usual practice that can lead to the incorporation of uncertainty in the model results. Consequently a long and tedious process of calibrating the model's output with observed data is indispensable. Moreover, when there is scarcity of observed data, the user runs the risk of over or under estimating the effects of the input parameters that can produce impaired output data.

There is the necessity of the inclusion of risk analysis on all the uncertainties in the parameterization of a watershed modeled using SWAT. The new interface incorporates a module for explicitly considering the uncertainty in the parameterization of SWAT, leading to a risk assessment analysis of the impact of point and non-point sources of pollution.

# 2 LITERATURE REVIEW

The experience with other hydrologic-hydraulic models can be used to illustrate the work that has been done in building GIS-based interfaces.

The *Hydrologic Modeling System* (HMS), developed by the Hydrologic Engineering Center (HEC) of the U.S. Army Corp of Engineers (USACE), is a set of computer models that work together to simulate the precipitation-runoff-routing processes on a watershed (HEC-HMS, 2000). HMS input data is based on three components: the basin model, the meteorological model and the control specifications (HEC-HMS, 2000). HMS has a main geospatial input component in the basin model. The parameters of the basin model, can be estimated using GIS. Olivera (2001) presented a methodology for the extraction of hydrologic information for supporting modeling with HMS. This methodology established main concepts, as well as summarized and combined previous work in the field (Olivera *et al,* 1998; Hellweger and Maidment, 1999). The product of this methodology is *CRWR-PrePro*: a pre-processor that is used to build the basin model for HMS. This is an example of an interface, which performs the following tasks: (1) extracts part of the required input information using GIS methods, (2) analyzes this data and performs the computation of required parameters, and (3) organizes the data and builds the required input model in a specific format (*i.e.* the format of the HMS input files). Similarly, in other hydrologic computer models, the input consists of text files that follow a specific code. Although this schema is conceptually basic, it has been used by most of the front-end pre-processors for hydrologic models, as a way to link them to GIS (Byars *et al,* 2000).

The *River Analysis System* (HEC-RAS) is a hydraulic model developed by HEC-USACE. It is a one-dimensional flow model that is used for flow profile calculations of steady and unsteady water flow. Its friendly Graphic User Interface (GUI) guides the user through three modules: (1) hydraulic analysis computations module, (2) database storage and management module and (3) reports and graphics module (HEC, 2001). An interchange of data between HEC-RAS and GIS for the development of *Digital Terrain Models* (DTM) was proposed by Tate *et al,* (2002). In this level of linkage with GIS, the information is handled *after* the model has produced its output. It was observed that HEC-RAS stores its one dimensional spatial data in a unique format that doesn't allow a quick two or three dimensional representation. Although newer versions of HEC-RAS (*i.e. 3.0, 3.1*) permit the import of georeferenced data, the hydraulic database keeps the same design. Therefore, this singular database format complicates further communication with other non-HEC models.

*HEC-GeoRas* is an interface for HEC-RAS that was designed as an ArcView 3.x extension. It encircles completely HEC-RAS in a GIS environment performing not only the common front-end preprocessor tasks but also back-end post-processor actions like interpreting and analyzing the HEC-RAS output data and determining whether or not the output has "hydraulic correctness" (HEC, 2000).

The *Better Science Integrating Point and Nonpoint Sources* (BASINS) is a GIS-based environmental analysis system that works under the "watershed approach" paradigm. It was developed by the U.S. Environmental Protection Agency (EPA) to provide support for its TMDL program (EPA, 2003b). BASINS bundles, in a single package, a complete set of components that includes national databases (*i.e.* land use, soils, stream networks), assessment tools for databases, pre-processors, models and postprocessors. The models are classified into water quality models (Toxiroute, QUAL2E) and hydrologic models (SWAT, HSPF) (EPA, 2003c).

The BASINS interface works as a customized ArcView 3.x extension that establishes a direct link between the embedded models and GIS. It was a milestone that highlighted the importance of GIS for integrating hydrologic models.

The models included in BASINS were already tested and validated previous to their incorporation. Thereby, modeling efficiency has improved. BASINS drawbacks are based on the complexity of its database structure. Since every model keeps its own data model, an intricate net of sub-databases scatter relevant information and make it difficult to use the data outside the BASINS framework. It can also be noticed that an improvement of components or the addition of new ones might be a complicated task since it has to be in compliance to the BASINS data model.

SWAT and the interfaces that link it with GIS have evolved since their creation in the early 90's.

Srinivasan and Arnold (1993) created an interface for linking the Geographic Resources Analysis Support System (GRASS) with the early SWAT 1990 version. GRASS is public domain GIS software developed by the United States Corp Of Engineers (USACOE). It is a front-end preprocessor built in a modular structure that aims at building input files for SWAT. It is important to recognize that this preprocessor was built in C programming language, the same as the GIS software, GRASS. Thus, an update of any of the components of the linkage or the addition of other modules can be effectively applied without complications. In addition, this interface also presents detailed tools for database access.

Bian *et al,*(1996) adapted the 1994 version of SWAT for ARC/INFO . This interface runs as a preprocessor under UNIX environment. It was built on Arc Macro Language (AML). Its conceptual model shows a centralized database that manages the information required by SWAT. It does not exploit all the capacities that ARC/INFO has in order to extract all the required hydrological parameters from spatially distributed datasets. For example, it does not include a module for watershed and stream network delineation, that can be obtained using ARC/INFO commands over a Digital Elevation Model (DEM).

Di Luzio *et al* (1998) developed an ArcView Interface for the 1996 SWAT model. It follows a methodology similar to that presented by Olivera (2001), for stream and watershed delineation, but it adds capabilities for managing Land use and Soil datasets. Developed in Avenue programming language, it is the first pre and post-processor for the SWAT model.

Di Luzio *et al* (2002) extended the ArcView Interface, developing a comprehensive ArcView extension (AVSWAT), which improved the extraction of hydrological information from raster data. It is important to note the set of tools for watershed delineation and hydrological response units (HRU) definition. Thus, AVSWAT was adapted and integrated into the BASINS package.

In addition to discussing efficient ways to build input files for a model, this thesis stresses the importance of organizing and storing Input and Output data in a standard data model.

The inclusion of risk assessment in hydrologic studies is a new tendency that has been adopted by various federal agencies. The US Army Corp of Engineers (USACOE) has included it as a requirement for all the funded flood damage reduction studies. USACOE has also developed procedures to model the uncertainties in hydrologic, hydraulic and economic studies (U.S. Army Corp of Engineers, 1996).

Wurbs *et al* (2001) defined a methodology to model the uncertainties in the parameters used to calculate the Average Annual Damage value of floods. This methodology was adopted and applied in the development of new hydrologic software like the Hydrologic Engineering Center's Flood Damage Analysis (HEC-FDA). The software uses a probabilistic approach based on Monte Carlo simulations, a statistical sampling analysis method. HEC-FDA aims at giving the user the possibility of making better informed decisions (HEC, 2003).

SWAT has an extensive literature regarding the calibration process. Santi *et al* (2001) derived a method for calibrating a watershed, modeled with SWAT. It focused on certain sensitive parameters that may affect the calculated values for flow, sediment and nutrient transport. Observed data was indispensable for these matters.

Eckhardt and Arnold (2001) developed a procedure that considers the uncertainties in the parameterization of a watershed modeled by SWAT-G (a modified version of SWAT which has been adapted for the European topography). Based on a genetic algorithm, the procedure automatically varies the sensitive parameters to produce an output that matches observed data. It was computationally demanding (6 days of computer run for a small watershed) and the results were acceptable but not optimum (Eckhardt and Arnold , 2001).

Eckhardt *et al* (2003) developed a method for SWAT-G that aims at deterministically assessing the impact of the land use change on the model of a hydrologic system by means of incorporating the uncertainty in the parameters related to land use. Normal probability density functions where assigned to them. Thousands of Monte Carlo simulations where performed and the output was analyzed to converge to a single value that determines whether the land use change is going to affect the hydrologic model or not.

This thesis aims also at producing a comprehensive methodology that will take into account all the parameterization uncertainties that might affect a watershed model. Rather than trying to isolate each parameter to analyze their individual effect on the output, this thesis stresses the importance of considering the combined effect that all the parameters have on the process of modeling. The expected results from this methodology are probability density functions that can help in the development of risk-based strategies for the assessment of point and non point sources of pollution.

# 3 METHODOLOGY

Fig. 1.1 shows the methodology followed for the development of the SWAT interface. This methodology builds a procedure that can be applied for any hydrologic model. This thesis constitutes a particular case.

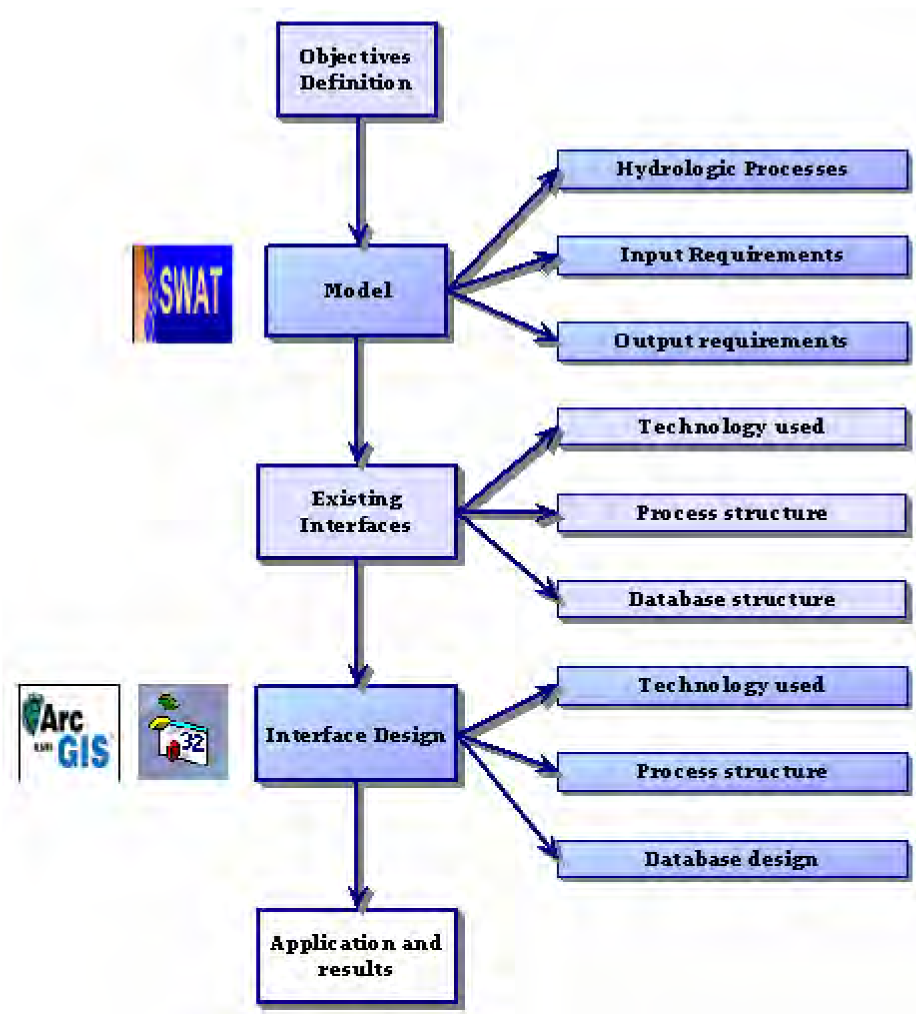



**FIG. 1.1.** Methodology followed on the development of the SWAT interface.

## 3.1 OBJECTIVES

The objectives may be determined by deciding the degree of complexity that we would like to achieve. Shamshi (1998) defined three levels of complexity regarding the information exchange between GIS and a hydrologic model (SWMM for his case): *interchange*, *interface* and *integration*. Integration is the most complex level. It involves the modification of the model structure to include GIS. On the other hand, the *interface* level can perform the actions of preprocessing, model calling, and post processing. It is an independent device, used to interpret and manage the information exchange between the system and the model. In the lower level, we have the *interchange* that might be reflected in the use of selected information that goes either from GIS to the model or from the model to GIS.

In the case of SWAT, the model itself has been used and validated for the last decade, which demonstrates its efficiency (Saleh *et al, 2000*). In addition, the application of previous interfaces has proven to reduce the modeling time, which increases the modeling efficiency (Di Luzio, *et al,* 2002). For the development of this application, our objective is to achieve an interface level.

Interface efficiency is going to rely on an organized source of data to manage the tasks of preprocessing and postprocessing. Therefore a parallel objective is to create a data model that is going to work with the interface to produce a comprehensive analysis of the chosen system.

## 3.2 SWAT

As mentioned previously, SWAT is a physically based hydrologic model. It is important to understand the complexity of the physical processes that SWAT involves. Several papers have analyzed and validated its procedures, and rather than citing these examples, I would like to summarize and highlight the main topics that were covered in order to design the interface. Each one of these processes is explained in detail in Neitsch *et al* 2000.

### 3.2.1 SWAT hydrologic processes and parameterization

The system that SWAT models is the watershed and the movement of water, nutrients, sediment and pesticides in it. Watershed can be defined as the contributing area in which water is accumulated and moves towards a point that has been defined as outlet This movement of water is based on the hydrologic cycle. This is the notation or basic rules that the SWAT model is based on. Whatever water is coming in, it is going out on a single point.

These are some characteristics of SWAT and some of its components:

- SWAT could be categorized as a hydrologic distributed model, because it considers the geospatial variations of parameters and processes.
- It could be also categorized as a lumped parameter model, since it *aggregates* unique combinations of soils and land use, with specific land management practices (known as Hydrologic Response Units) to simulate the hydrology on a watershed. The hypothesis is that these unique combinations are going to *respond* in a similar way.
- It runs in a continuous time step and it can simulate extended periods of time.
- It is a physically based model.
- The hydrologic processes are divided in a land phase and a stream phase. The land phase controls the transport of water, sediment and agricultural chemicals to the channel system. The stream phase routes the flow and constituents through the main channel to the watershed outlet.
- Land phase calculations are based on the mass balance of the components of the hydrologic cycle: precipitation, soil moisture, surface runoff, evapotranspiration, infiltration and groundwater flow. Highlighted land phase processes: Infiltration – surface runoff, calculated by the SCS curve number method, a non physically based method (for daily time step) and the option of the calculation of the infiltration using Green – Ampt's equation (for sub daily time step).

- Highlighted stream phase processes: flood routing computed using either variable storage coefficient method or the Muskingum routing method; nutrient routing, adapted from the QUAL2E model.
- The hydrology is simulated in each Hydrologic Response Unit.

**TABLE 3.1.** Summary of highlighted processes on SWAT

| Phase | Process/Component | Model/Equation |
|---|---|---|
| LAND PHASE | Precipitation | Weather Generator: generated precipitation using the Nicks model. |
| | Air temperature and solar radiation. | Generated from normal distribution and adjusted from a continuity equation. |
| | Wind speed. | Modified exponential equation. |
| | Infiltration | Green & Ampt or difference between rainfall and surface runoff using SCS CN method. |
| | Potential Evapotranspiration | Hargreaves or Priestley Taylor or Penman-Monteith equations. |
| | Lateral subsurface flow | A Kinematic storage model |
| | Surface runoff volume | SCS Curve Number method or Green & Ampt. Equations. |
| | Peak runoff rate | Modified Rational formula. |
| | Transmission losses | Lane’s method. |
| | Land cover/plant growth | Modified EPIC model. |
| | Erosion | Modified Universal Soil Loss Equation. |
| | Pesticide movement | Adapted from GLEAMS model. |
| STREAM PHASE | Flood routing | Variable storage coefficient method or Muskingum routing method. |
| | Sediment routing | Function of peak channel velocity. |
| | Nutrient routing | Adapted from QUAL2E |

Table 3.1 is a summary table of SWAT highlighted processes and the equations or models used for the parameterization of the processes.

### 3.2.2 SWAT input requirements and organization

SWAT organizes its input in a series of text files, which contain the required information for the three parameterization levels: basin, sub-basin and HRU. Usually there is one text file for each feature in a level. For example, the text files with an extension *.sub (subbasin level) detail the subbasin parameters. If there are 25 subbasins on the basin, there are going to be 25 textfiles with a *.sub extension. The name of each textfile is going to be the number of subbasin.

Table 3.2 is a summary table of the required SWAT input text files and a brief description (summarized from Neitsch *et al* 2000).

**TABLE 3.2.** Summary of SWAT input files

| File | Description | Optional? |
|---|---|---|
| Watershed configuration file (.fig) | Routing network in watershed. | Required |
| Control input/output file (file.cio) | Names of input files | Required |
| Input control code file (.cod) | Length of simulation, printing frequency, selected options | Required |
| Basin Input file (.bsn) | Watershed level parameters. | Required |
| Precipitation input file (.pcp) | Daily measured precipitation for a gage(s) | Optional |
| Temperture input file (.tmp) | Daily measured maximum and minimum for a gage(s) | Optional |
| Solar radiation input file (.slr) | Solar radiation for measuring gage(s) | Optional |
| Wind speed input file (.wnd) | Daily average wind speed for a measuring gage(s) | Optional |
| Relative humidity input file (.hmd) | Daily relative humidity values for gage(s) | Optional |
| PET input file (.pet) | Daily PET values for the watershed | Optional |

**TABLE 3.2. (continued)**

| File | Description | Optional? |
| --- | --- | --- |
| Land cover/plant growth database file (crop.dat) | Plant growth parameters for land covers in watershed. | Required |
| Tillage database file (till.dat) | Amount and depth of mixing caused by tillage operations in the watershed. | Required |
| Pesticide database file (pest.dat) | Information on mobility and degradation for the watershed. | Required |
| Fertilizer database file (fert.dat) | Information on nutrient content of all fertilizers and manures simulated on watershed. | Required |
| Urban database file (urban.dat) | Build up and wash off of solids in urban areas in the watershed. | Required |
| Subbasin input file (.sub) | Subbasin level parameters | Required |
| Weather generator input file (.wgn) | Statistical data needed for the climatic generation. | Required |
| Pond/wetland input file (.pnd) | Information of the impoundments in the subbasin. | Optional |
| Water use input file (.wus) | Consumptive water use in the subbasin | Optional |
| Main channel input file (.rte) | Parameters governing water and sediment movement in the main channel of the subbasin. | Required |
| Watershed water quality input file (.wwq) | Parameters used to model QUAL2E transformations in main channel. | Optional |
| Stream water quality input file (.swq) | Parameters used to model pesticide and QUAL2E nutrient transformations in the main channel of the subbasin. | Optional |
| HRU input file (.hru) | HRU level parameters | Required |
| Management input file (.mgt) | Management scenarios nad specifies land cover simulated in the HRU. | Required |
| Soil input file (.sol) | Information abput the physical characteristics of the soil in the HRU. | Required |
| Soil chemical input file (.chm) | Information about initial nutrient and pesticide levels of the soil in the HRU. | Required |
| Groundwater input file (.gw) | Information about the shallow and deep aquifer. | Required |
| Reservoir input file (.res) | Parameters for modeling the movement of water and sediment through a reservoir. | Optional |
| Lake water quality input file (.lwq) | Parameters used to model the movement of nutrients and pesticides through a reservoir. | Optional |
| Point source input file | Loadings to the channel network from a point source | Optional |

Since SWAT is categorized as a distributed model, it needs spatially distributed properties and processes within the boundaries of the watershed. This model's characteristic determines the use of geospatial data as the main source of its input. Most of the parameters that populate these text files have a spatial component. Therefore, the use of Geographic Information Systems (GIS) proves to be a valid tool for analysis and calculation of the required input parameters. Thus, the elements that constitute the three levels of parameterization can be defined spatially using GIS technology and spatially distributed data. Watershed and subbasins can be delineated spatially performing an analysis over a Digital Elevation Model (DEM), similar to that detailed by Olivera (2001). A DEM is based on an array of square cells that models the topography of the system. Each cell has the value of the elevation of the terrain at its center. Data that follows a distributed pattern of square cells containing numeric or coded information is called *raster data*.

After defining the watershed and subbasins, HRU's can be determined on each subbasin using raster data for soils and land use/land cover data. If the soils are modeled on three dimensions, an almost complete water budget can be performed on each HRU.

Consequently, the rest of the parameters can be either derived from spatial data or introduced manually by the user and linked to the spatial data using GIS techniques.

It is important to stress that SWAT has a temporal component on some of its input variables (*i.e.* water use, reservoir operations). A data model is necessary to store, retrieve, connect and allow further analysis of the input data. The interface and the geodatabase design must contemplate all of its requirements and furthermore foresee necessities from other models.

### 3.2.3 Output organization

SWAT simulates the modeled conditions of the system and produces a series of temporal data that are organized with the same level of parameterization as the input data. The results are stored in textfiles. The following table (Table 3.3) provides a brief description of each one (summarized from Neitsch *et al* 2000).

**TABLE 3.3.** Summary of SWAT output files

| File | Description |
|---|---|
| Input summary file (input.std) | Summary tables of important input values. |
| Output summary file (output.std) | Watershed average loadings from the HRU's to the streams. Also average annual HRU and subbasin values for a few parameters. |
| HRU output file (.sbs) | Summary information for each HRU in the watershed. |
| Subbasin Output file (.bsb) | Summary information for each subbasin in the watershed. The values are the total amount or weighted average of all HRUs within the subbasin. |
| Main channel output file (.rch) | Summary information for each routing reach in the watershed. |
| Impoundment output file (.wtr) | Summary information for ponds, wetlands and depressional/impounded areas in the HRUs. |
| Reservoir output file (.rsv) | Summary information for the reservoirs in the watershed. |

The main output files: HRU, sub-basin and main channel, reflect the land phase/stream phase hydrologic subdivision in which the SWAT model is based.

An interface must process and help to assess the validity of the output information. In this way, it should provide tools and means for analyzing and comparing this data with observed data. SWAT has a strong physical component and it requires a big quantity of data that may or may not be available.

### 3.3 EXISTENT INTERFACES: AVSWAT

Instead of starting our work from scratch, it is recommendable to take the best of the previous work, evaluate and improve it. Proof of this statement is one of the conclusions of the Watershed Allocation Model project (Wurbs, 2001) in which it was decided to analyze the current water allocation/hydrologic models that were developed in the previous decades instead of building new ones. Most of the selected models had a framework similar to that of the SWAT model (they use the same programming language and follow the same technology). Extending this statement to the interfaces of the model, it is important to extract the concepts and structure that can be applied using the current paradigms and technology and extend them to improve their characteristics.

As the literature review indicates, there are several interfaces that have been developed for SWAT. Among them, AVSWAT (Di Luzio *et al,* 2002) is one of the most comprehensive interfaces. It allows the user to build a project from scratch, build the required input data and analyze the output data from SWAT. It constitutes a good example of Shamshi's definition of interface, performing actions for preprocessing and postprocessing.

### 3.3.1 Technology overview

AVSWAT was chosen as a starting point for building the new SWAT interface. AVSWAT links the ArcView 3.2 GIS software with SWAT 2000 version. Avenue programming language was used to develop the interface, which works as an ArcView extension.

#### *3.3.1.1 ArcView 3.x*

ArcView 3.x gives the user a number of capabilities for managing data in several data structures. The basic types of data structures that ArcView 3.x manages are: vector, raster, Triangulated Irregular Networks (TIN's) and image data. AVSWAT takes advantage of two types of formats: vector data represented by the shapefile and raster data, represented by grids.

A shapefile is a way to store data that is represented by the basic elements: points, lines and polygons, storing their geometry in a shape with geographic coordinates. It represents non topological data. This means that the features have no spatial relationships between them. The attributes of a feature are stored in a dBase table that is linked to the shape, having a one to one relationship between a feature and a record in the table (ESRI, 1998).

A grid is an array of discrete, uniform pixel blocks defined as cells. Each cell has a geographic representation; this means that it has a unique *x, y* location defined by its row and column location. A grid dataset may represent continuous data (i.e. elevation, usually defined with floating point grids) or discrete data (i.e. land use/land cover, usually represented with integer grids). Integer grids may have a linked table that summarizes the frequency of the values of the grid. This table is named Value Attribute Table (VAT). Cells with the same value define a zone, whether they are adjacent or not.

*3.3.1.2 Avenue Programming Language*

Avenue is an object oriented programming language developed for the ArcView 2.x and 3.x series. It is a scripting language that is used to create applications under the ArcView 2.x/3.x environment. It manages *scripts* that are built to achieve the following objectives, according to the ArcView 3.x help: *a)* automate tasks, *b)* add new capabilities to ArcView and *c)* build complete applications.

ArcView 2.x and 3.x series are built under the framework of a big set of Avenue scripts called *system scripts*. The user can make use of these scripts to automate tasks or modify them to suit their necessities. It can also build complete applications based on their object model diagram.

An ArcView extension called Dialog Builder can be used to design Graphic User Interfaces (GUI) and link the scripts to specific events on the *dialogs*.

The advantages that can be highlighted about AVENUE are:

- As a scripting language, it comes rooted in ArcView without charge.
- The scripts can be coded on any text editor.
- It is an object oriented programming language.
- Samples and support easily available.

The main disadvantages of Avenue programming language:

- It works only under the ArcView 2.x and 3.x series.
- It can't build independent applications outside the ArcView's framework. This means that it is difficult to couple or extend its capabilities using external object components or other languages.
- Difficult debugging process.
- It is no longer supported by the ESRI.

### 3.3.2 Process structure, code and GUI

The process structure on AVSWAT follows the order of the levels of parameterization in SWAT: watershed, subbasin and HRU. Consequently, the objective of the first module is to delineate the focused watershed (the model of the system), it's respective subbasins and calculate their hydrologic parameters. In order to facilitate the analysis, AVSWAT uses a DEM as the model of the terrain of the system. Olivera (2001) detailed a methodology for the delineation of watershed, definition of stream networks and extraction of hydrologic information from DEMs. The delineation and definition of subbasins and streams are accomplished using the eight-pour-point flow direction algorithm (Jenson and Domingue, 1988). AVSWAT follows a similar procedure in its first module.

In AVSWATS's first module, subbasins are delineated either from outlets (defined by the end points of the generated stream links) or from user-defined points (sustained as points of interest, i.e. gage stations) (Di Luzio *et al,* 2000). In any case just one stream is defined per watershed. This is a limitation set by the SWAT model and not by the AVSWAT interface. This relationship facilitates the process of routing in the stream network since the drainage area that contributes to overland flow to a specific stream can unambiguously be defined.

The user can also define the level of detail of the streams using a threshold drainage area. This means that the user can modify what would be the necessary drainage area of a point in order to be considered part of the stream network (the "blue lines" on the map).

The second module defines the HRU parameterization level. Based on the concept of HRU, it calculates areas and percentages for each unique combination of land use and soil per subbasin. It accomplishes this objective by working with land use and soil type grids. Land use grids can be obtained from any available source, and can be used with the condition of being integer raster datasets. This means that the value that each cell is going to store is going to be an integer land use code. The SWAT interface only supports one specific modified version of the State Soil Geographic (STATSGO) database (NRCS 1995). This modified version arranges the STATSGO information into soil layers. However, the user can always define its own soil database, provided it is compliant with the modified version of the STATSGO database.

A GIS operation called "overlay" was applied on soils and land use. It generates cross -tabulated areas of the two datasets within the boundaries of each subbasin. After this process, it allows the user to establish filters for the HRU's calculated areas based on percentages of land use over a subbasin and soils over land use class. These filters aim to define the most relevant HRUs on each subbasin. In Neitsch *et al* (2000), it is recommended not to exceed 10 HRUs per subbasin; otherwise, it is preferable to define a greater number of subbasins by modifying the threshold number.

The third module deals with the calculation of the weather parameters. It allows the user to input custom weather data organized in tables that follow a specific format. Alternatively, the SWAT weather generator model can calculate the required time series data for the weather module.

The fourth module generates the required parameter values for the SWAT input files based on custom databases and default values. It interactively guides the user to replacing each parameter's value and consequently generating input text files based mainly on user's defined data and default values.

The fifth module details the actions for postprocessing the information simulated by SWAT. AVSWAT organizes it on dBase tables and gives the user tools for creating graphs for the time series and also for calibrating the model varying sensitive parameter values.

The AVSWAT code is organized as a unique list of Avenue scripts that detail the requests for the ArcView objects. It has not been built in a modular way. Inputs and outputs for each script cannot be easily identified. Reuse of scripts on other applications can not be done since all the scripts are fixed to a rigid framework application design.

AVSWAT presents its GUI as a customized ArcView project. Developed under Avenue, it uses so-called *Dialogs* for the interaction with the user. The dialogs were developed using the Dialog Builder extension. These items can only be used under the ArcView 3.X framework.

### 3.3.3 Database structure: I/O management

The management of the input and output from SWAT is based on a structure of Windows folders that differentiates the SWAT general required data from a specific SWAT project data. The SWAT specific project data is divided into folders that contain all the vector data, the raster data, tables and report documents. Vector data in AVSWAT uses the shapefile data structure, while raster data uses the GRID format. The tables summarizing data for the creation of text files are generated in DBase format and serve as an interface between the text files parameter values and the user.

There is no explicit definition of relationships between geospatial data and input/output data.

## 3.4 INTERFACE AND GEODATABASE DESIGN

The new interface and geodatabase package created for the SWAT hydrologic model is called ArcGIS SWAT. It was developed under the new paradigms in GIS and spatial data management. It is based on the software developed by the Environmental Systems Research Institute (ESRI): ArcGIS.

### 3.4.1 Technology overview

#### *3.4.1.1 ArcGIS 8.x*

Previously, ESRI used to manage two types of products separately: ArcView (with its series 1.x to 3.x) and ArcInfo Workstation. ArcInfo Workstation was developed in the early 80s and aimed to introduce a new paradigm in GIS. It links computer displayed graphics with attribute tables, supported with sophisticated geographic analysis commands that allows the user to perform advanced analysis on geographic data. ArcInfo used to run on minicomputers and later on it was adapted to work under a UNIX environment and afterwards under Windows operating systems.

ArcView, on the other hand, was developed as a *viewer*. It offered tools for displaying the data created or managed by ArcInfo. It lacked the capabilities of ArcInfo for analysis, but it showed a user-friendly environment that made it a popular product. On the subsequent versions, analysis capabilities were added to the product maintaining its friendly interface. But the two products were developed using different technologies with different supporting teams. For the new versions of ArcView and ArcInfo, ESRI reengineered all its products using the current paradigms on the Information/Technology field, that are explained later. ESRI decided to combine the two products under the same architecture and platform, using the same technology. ArcGIS 8.x series is the result of these innovations. ArcGIS 8.x is compile of four components: ArcReader, ArcView, ArcEditor and ArcInfo . These components are presented as a scalable system, meaning that they offer the same technology and architecture, but they differ in the complexity of the problems that they can analyze (Fig. 3.1). There are several advantages on sharing the same technology and architecture. For example, they share the same interface and tools. Problems can grow in terms of complexity (from ArcView to ArcInfo, where the former manages simpler data structures and the latter manages complex ones) or in terms of number of users and complexity of the database systems (from single users to multiple users with internet publishing).

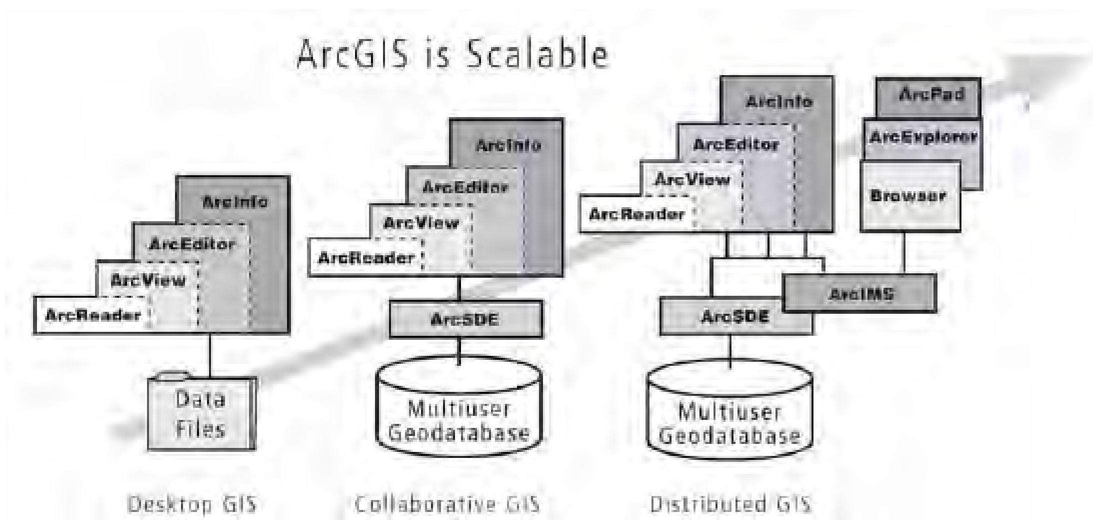



**FIG. 3.1.** ArcGIS 8.x scalable system (ESRI, 2003b).

ArcGIS new technology is characterized by its compliance with new standards on the IT industry (ESRI, 2003b):

- Component Object Model (COM)
- Extensive Markup Language (XML)
- Structured Query Language (SQL)

ArcGIS-SWAT 2003 takes advantage of these standards on its design and structure. It works under ArcView 8.x. Since ArcView 8.x is the basic package in the scalable system, more users will have access to it. Working under a scalable system, ArcEditor and ArcInfo users can also have access to the same package.

*3.4.1.2 Component Object Model (COM)*

COM is a new technology that defines a standard for writing programs. Developed by Microsoft Corporation, it allows the communication between applications even if they were built in different programming languages. For example, a module of an application written in a programming language X can be used on an application written on a programming language Y if both implement COM technology. In this way, programs can be written in a modular way (component objects) so they can be used in conjunction with different applications from different sources.

Besides the several new features that ArcGIS 8.x presents, the new interface is based on the *customization* feature of ArcGIS. ArcGIS uses Visual Basic for Applications (VBA) as the base programming language for its applications. Thus, ArcGIS SWAT 2003 takes advantage of the COM technology and is written in Visual Basic 6.0 using the Object Model developed by ESRI, which is COM compliant. The Visual Basic code keeps a modular structure so that other processes can be added easily.

ArcGIS SWAT 2003 is built as an *Extension* for ArcGIS 8.x series. An extension is a complete set of tools that *extends* the capabilities of ArcGIS to suit the needs of a specific problem. The implementation of an extension gives ArcGIS SWAT 2003 a complete control over the map document, making it a SWAT project.

ArcGIS SWAT 2003 extension is build as a series of Dynamic Link Libraries (DLLs) that are COM compliant. DLLs are applications that are called only when they are needed as opposed to an executable file (*.exe) that stays in memory the time they are executed. Consequently, memory is saved and we can allow having more complicated processes. In addition, as they are COM compliant DLLs, they can be reused on other COM applications.

*3.4.1.3 The Geodatabase data model*

A geodatabase is a database that stores geographic features inside of a *Relational Database Management System* (RDBMS), which manages the geographic data (MacDonald, 2001). The Geodatabase model is an object-oriented model designed for vector data, which supports an integrated topology (MacDonald, 2001).

In order to explain the elements of the Geodatabase data model it is necessary to make a quick review of basic terms related to Unified Modeling Language.

#### 3.4.1.3.1 The Unified Modeling Language

The Unified Modeling Language (UML) "is a language for specifying, visualizing, constructing, and documenting the artifacts of software systems, as well as for modeling business and other non-software systems. The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems" (Object Management Group, 1997)

UML was developed to create a universal standard in modeling systems, using object oriented concepts. UML was developed by Rational Software with the fusion and modification of the leading methods in the industry: Booch, OOSE/Jacobson and OMT (Object Management Group, 1997).

The basic terms that are commonly used in UML are (adapted from Booch *et al,* 1999):

- *Object:* Abstract representation of a real thing. An object is the model of a real thing.


**FIG. 3.2.** Left: Amazon river. Right: UML representation of the Amazon River.

- *Class:* The blueprint for creating objects that share characteristics and behavior. An object is "simply an instance of a class" (Booch, *et al,* 1999). Objects that come from the same class share characteristics and behavior.

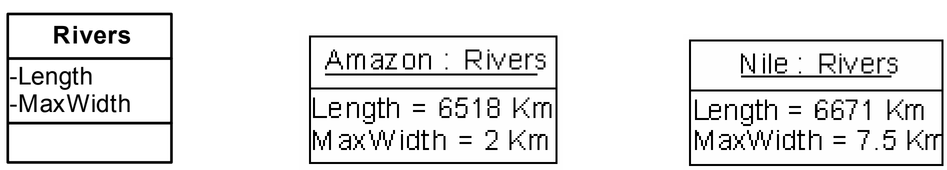



**FIG. 3.3.** The Rivers class and objects as "instance" of the class.

- *Meta Classes:* can be defined as "a class' class" (Hathaway, 1996) or as a parent class. The *children classes* of a meta class are going to *inherit* the characteristics and behavior of the *parent's class*.

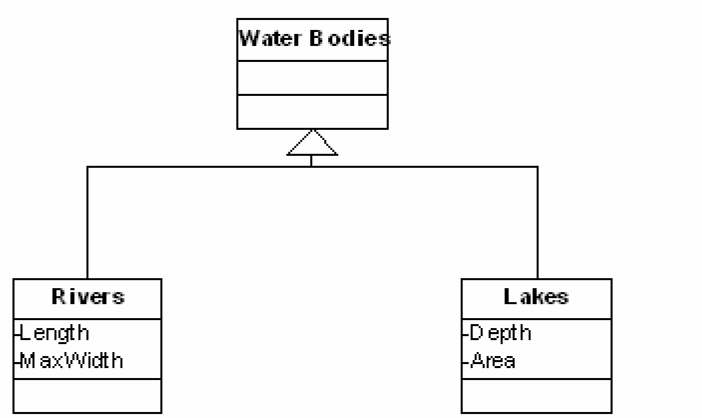



**FIG. 3.4.** Meta class example: Water Bodies.

In Fig. 3.4, it is possible to say that the children classes Rivers and Lakes are going to inherit the characteristics of their parent class Water Bodies. In UML, this can be depicted as a "Generalization" arrow that symbolizes the inheritance.

- *Association*: is a type of relationship between classes that defines the connections between their object instances. Booch *et al* (1999) defined it as a "structural relationship that describes a set of links, in which a link is a connection among objects; the semantic relationship between two or more classifiers that involves the connections among their instances"

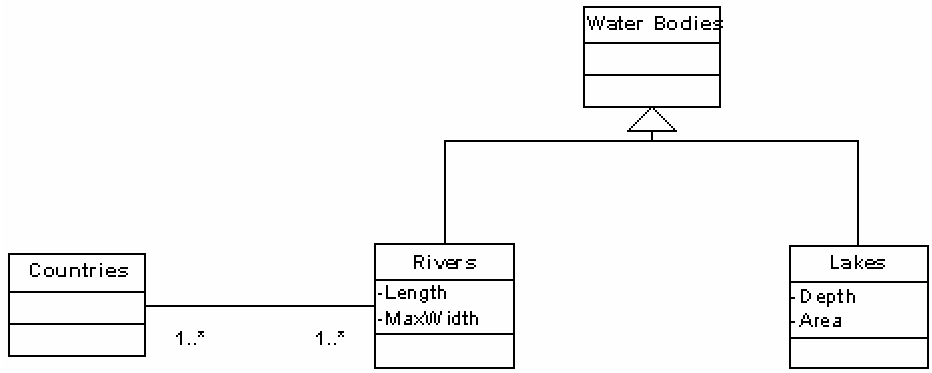



**FIG. 3.5.** Association.

In Fig. 3.5, there is an association between the Countries class and the Rivers class. This association or relationship is called *binary association* because there are only two classes involved. The end of the association lines establish the *multiplicity* or elements that can be involved. In this case, there can be many countries that have many rivers.

- *Aggregation:* Establishes a whole-part relationship. For example, an aggregation of provinces (*parts*) make a country (*whole*). If the life time of the part-elements are dependant on the life time of the whole-element then it is called *composition*.

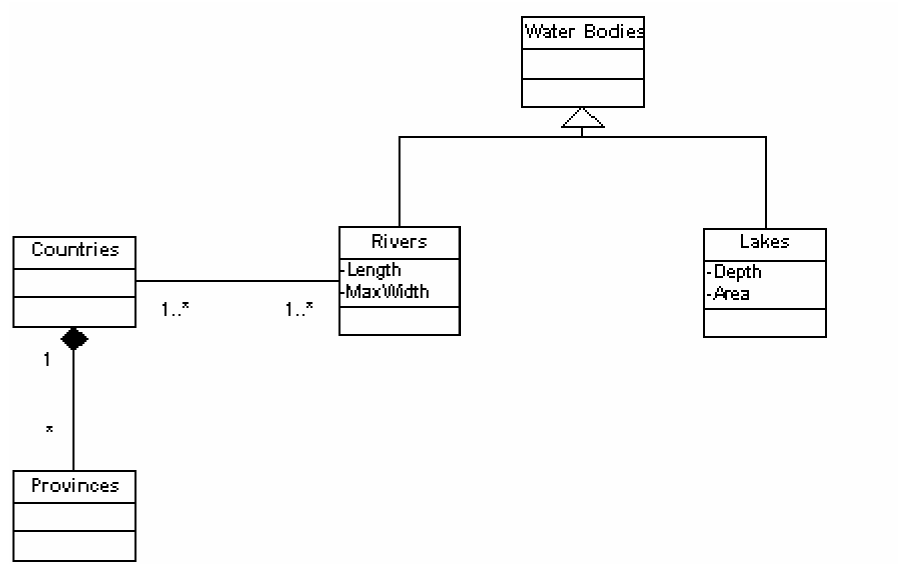



**FIG.. 3.6.** Composition.

The Fig. 3.6 shows an example of composition: without the country there cannot be provinces (black diamond at the end of the line). A white diamond at the end of the line would imply an aggregation.

### 3.4.1.3.2 Geodatabase data model principal elements

On the top level of the hierarchy of the Geodatabase data model we find the *geodatabase* (represented by the *Workspace* box on Fig. 3.7). It is considered as a "collection of datasets, feature classes, object classes and relationship classes" (Zeiler, 1999).

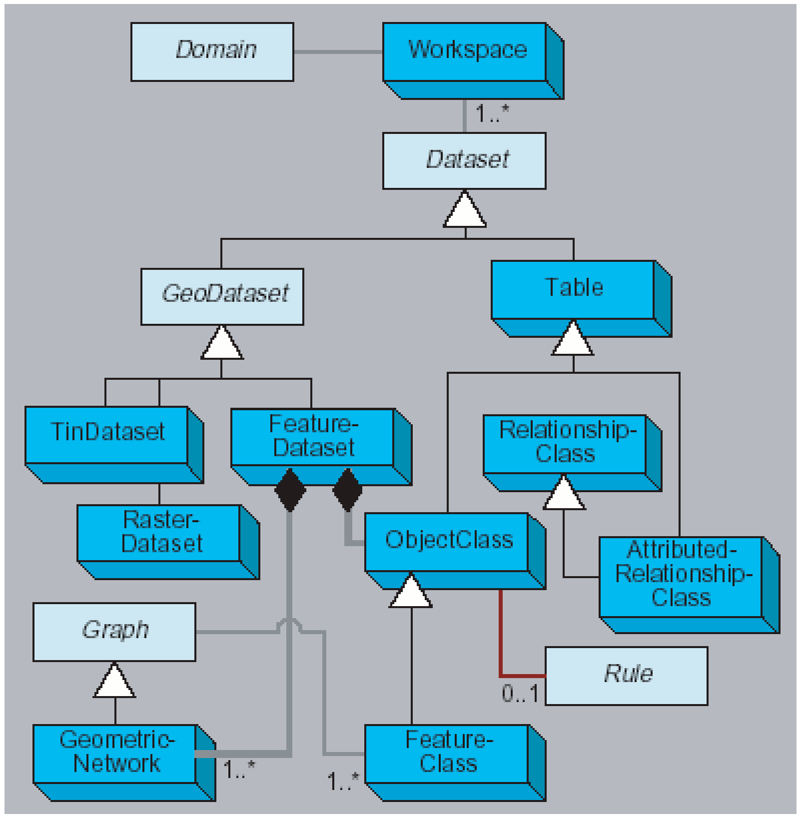



**FIG. 3.7.** Part of the geodatabase data model, build with UML (Zeiler, 1999).

There are several types of geodatabases, but the main classification is between Personal Geodatabases (PGDB) and Enterprise Geodatabases. The difference between them is the number of users that can edit them simultaneously. Enterprise Geodatabases can be accessed at the same time by several users. Enterprise Geodatabases can be only accesed through ArcSDE, a software created by ESRI that acts as an interface between the client side (the user) and the server side of a database (*i.e.* Oracle, Informix). Enterprise Geodatabases support all the datasets presented in Fig. 3.7 (Tin dataset, Feature dataset, Raster dataset) while PGDB's only support Feature datasets.

Feature datasets are considered a "collection of feature classes, graphs and relationship classes that share the same spatial reference" (Zeiler, 1999). The Spatial Reference is composed of projection and extent. On the same fashion, a feature class is a collection of features that share the same type of geometry. The geometry of the features can only be of the following types: point, line and polygon. The type of geometry is specified as an attribute of the feature class called *shape*. Features are the simplest objects that have a geographic location and a shape. Feature classes can be part of a collection (feature dataset) or outside of it (called standalone feature class).

An object class is a standalone table without a shape attribute in which each row is an object and each column is an attribute of the object (Zeiler, 1999).

A relationship class in UML is defined as an *association* between classes. It is a collection of relationships between objects in two classes (Zeiler, 1999). The classes can be feature classes or Object Classes. Only ArcInfo and ArcEditor can create and edit relationship classes in a geodatabase. ArcView has read-only access to relationship classes. A relationship class can also be understood as a table that stores the key values that are used to relate the objects in two classes. If this table contains more columns, then it is an *attributed relationship class*.

ArcView has a way to emulate the characteristics of the relationship class, using what is called *memory relationship class*. This type of association is created on the Random Access Memory (RAM) memory and stored in the map project.

*Cardinality* is the definition of the multiplicity ends of the association line. In this way, there can be cardinalities of *one to one* (one object in a class corresponds to one object in another class) *one to many* (one object in a class corresponds to many objects in another class) and *many to many* (many objects in a class are related to many objects in another class). There are two types of memory relationship classes: join and relate. Join is a relationship with a cardinality of *one to one*. Cardinalities *one to many* and *many to many* can only be produced with the *relate* memory relationship class.

A *Geometric Network* is a combination of two feature classes: Junction and Edge. These feature classes define the connectivity in a network. Geometric networks can only be created and edited with ArcInfo, therefore ArcView can't create or edit geometric networks.

A summary of the differences in capabilities between ArcView, Arc Editor and ArcInfo with respect to geodatabases is showed on the Table 3.1.

**TABLE 3.4.** Personal geodatabase and multiuser geodatabase comparison (ESRI, 2003b)

| | Personal Geodatabase With ArcView | Personal GeodatabaseWith ArcEditor or ArcInfo | Multiuser Geodatabase |
|---|---|---|---|
| **Number of concurrent editors** | One | One | Many |
| **Create and edit simple features (points, lines, areas, static annotation)** | √ | √ | √ |
| **Define and use attribute domains** | √ | √ | √ |
| **Set database schema 1** | √ (1) | √ | √ |
| **Versioning (long transactions)** | | | √ |
| **Store raster data** | | | √ |
| **Create and edit features with subtypes or dimension features** | | √ | √ |
| **Establish behavior (topology,relationships, geometric networks, feature-linked annotation, etc.)** | | √ | √ |
| **Create and edit custom features** | | √ | √ |
| **Database size** | ≤ 250K features (2) | ≤ 250K features (2) | Unlimited |
| **Requires ArcSDE** | | | √ |
| **Supported databases** | Microsoft Jet | Microsoft Jet | Oracle, Microsoft SQL Server,IBM DB2,IBM Informix |

(1) Limited to simple features in a personal geodatabase.
(2) This is an approximate limit affected by two factors—file size and computer memory. Microsoft Jet 4.0 used by the personal geodatabase has a 2 GB file size limit. In addition, a personal geodatabase is a single file that is loaded into computer memory. Therefore, performance can become unacceptable even for file sizes less than 2 GB. The recommended 250,000 feature limit

### *3.4.1.4 Hydrologic data models: ArcHydro*

The geodatabase data structure can be used to create custom data models that suit the needs of specific disciplines. ESRI, in partnership with key sectors of different industries interested in the application of GIS data models, have created a series of custom data models derived from the geodatabase data structure. The main objective of these data models is to facilitate modeling and analysis based on a standard data storage format (ESRI, 2003a). In this process, ESRI is not interested in the creation of a new common standard for each industry. ESRI stresses the importance of a common data model as a "key to making better decisions based on available geographic information" (ESRI, 2003a), but the intention of ESRI and partners (called a *consortium*) is just to give the user the necessary tools for start an ArcGIS project. In this fashion, *consortiums* developed data models detailed on Table 3.5.

**TABLE 3.5.** Custom data models

| Consortiums Data Models |
|---|
| Basemap |
| Biodiversity |
| Census-Administrative Boundaries |
| Defense-Intel |
| Energy Utilities |
| Energy Utilities - MultiSpeak TM |
| Environmental Regulated Facilities |
| Forestry |
| Geology |
| Historic Preservation and Archaeology |
| Hydro |
| International Hydrographic Organization (IHO) S-57 |
| Land Parcels |
| Local Government |
| Marine |
| Petroleum |
| Pipeline |
| Telecommunications |
| Transportation |
| Water Utilities |

The ArcGIS Hydro data model (or ArcHydro) was developed by a consortium between the University of Texas at Austin, ESRI and others. It was developed for supporting the tasks of modeling and analyzing water resources problems.

ArcHydro was designed to be “a standardized way of describing that data [Hydrographic and Hydrologic]” (Maidment *et al,* 2002). Thus, a schema that reflects temporal and geospatial hydrologic data was created to support surface water hydrology and hydrography modeling at any scale. Its benefits can be better seen on complex projects where an overwhelming amount of data may lower the efficiency and possibly jeopardize the results of a model simulation. ArcHydro was build to work with independent hydrologic simulation models.

Several versions of the ArcHydro data model have been developed in order to consider all possible cases. There can be identified two data model structures: ArcHydro and ArcHydro Framework. ArcHydro is composed of five modules: Network, Drainage, Hydrography, Channel and Time Series. ArcHydro Framework is a simpler version of ArcHydro and it is used for simple applications and models that don’t require all the custom objects that ArcHydro considers. For most of the applications, an ArcHydro Framework with a Time Series module proves to be enough to support modeling of hydrologic features. The ArcHydro Framework with Time Series module is discussed below.

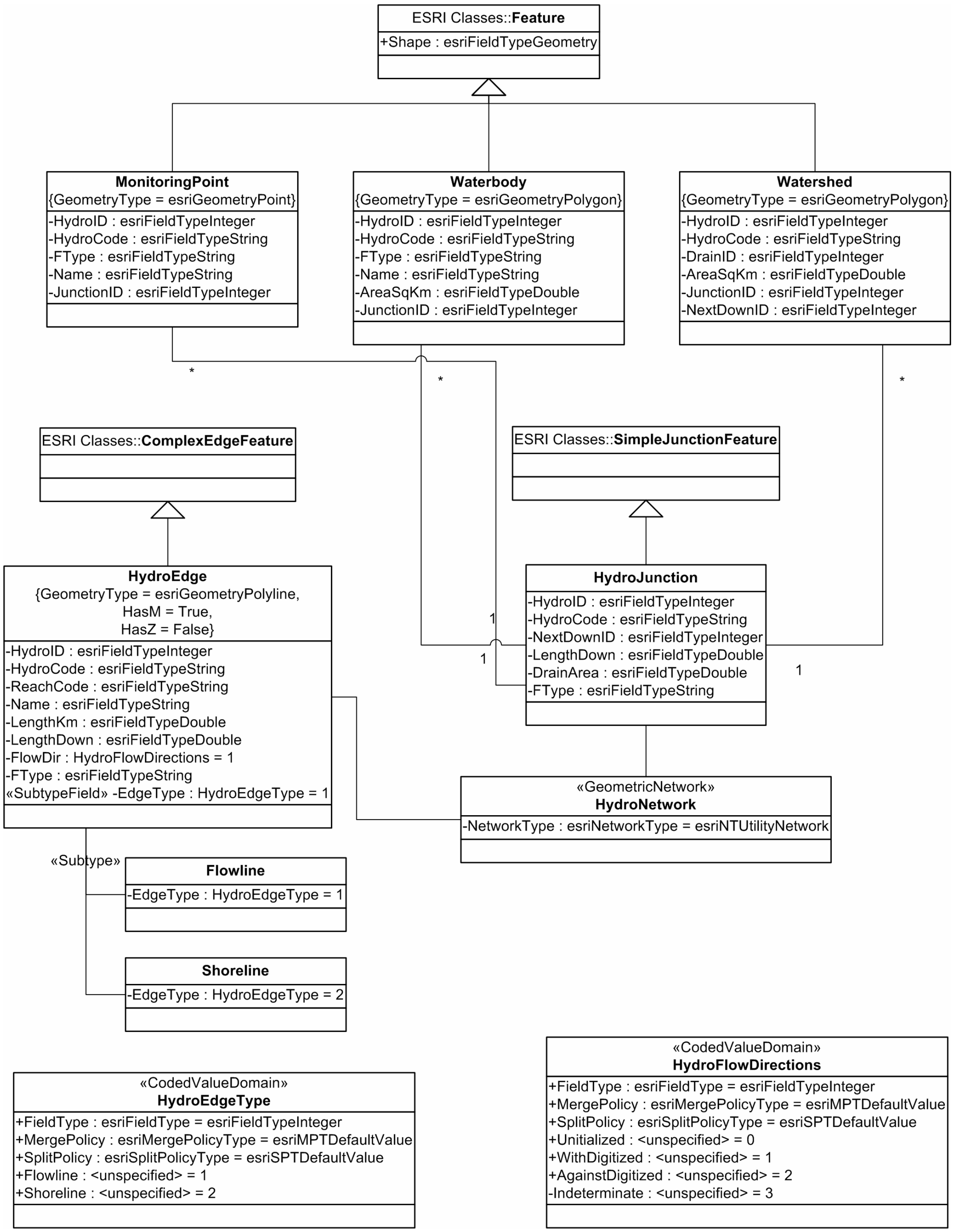



**FIG. 3.8.** ArcHydro Framework UML (Maidment *et al* 2002).

Fig 3.8 shows the UML of the ArcHydro Framework. This UML shows the classes and their relationships. All the classes are *generalized* (inherited) from the ESRI object class. In this way, every class contains an Object Identification number (OID) that identifies them as unique objects within the geodatabase. The ESRI feature classes: Feature, ComplexEdgeFeature and SimpleJunctionFeature are child classes of the ESRI object class. Therefore, even if in this UML it is not specified, each class has an OID attribute inherited from the Object Class. ESRI feature classes have, additionally, a *shape* attribute that stores the geometry and spatial location of a feature. This field is what describes an ESRI feature class. Consequently child classes inherit this attribute.

ComplexEdgeFeature and SimpleJunctionfeature classes are the components of a *geometric network*. A geometric network is a logical network that has topological relationships (MacDonald, 2001). As indicated on the previous section, they can be created and maintained by ArcEditor and ArcInfo, but not by ArcView.

The ArcHydro custom objects are children classes of the ESRI classes. In each one, the ESRI geometry type has to be defined. Each ArcHydro child class is explained as follows (Maidment *et al*, 2002):

- MonitoringPoints: their geometry makes them a point feature class. It is a simple feature collection that stores features that represent any measuring point on a watershed (*i.e.* stream flow gages, rain gages, water quality stations). It can also represent hydrologic points like reservoirs, water withdrawal points or other structures that can alter the natural regime of the water.

- Waterbody: defined as a polygon feature class. It stores simple polygon objects that represent water bodies such as lakes or inundated areas.
- Watershed: represents the drainage areas within the boundaries of the selected system. These features are part of polygon feature class.
- HydroEdge: it is a complex feature class that is going to store line features that model the river streams, the blue lines in paper maps.
- HydroJunction: a point feature class that stores points of interest in the geometric network. Usually contains the *junctions* of the geometric network but it can also contain important features as outlets of a watershed.

A *subtype* is a sub classification inside the feature class. Subtypes are used to classify homogeneous data inside a feature class. In this fashion, HydroEdge can be sub classified as Flowline and Shoreline. Subtypes can only be defined using ArcInfo.

*Coded Value Domains* are classes that describe the only possible attribute values for features. Each value (text or numeric) is represented by an integer number. For example in the coded value domain HydroEdgeType (Fig. 3.8), there are only two types of *edges: flowlines* and *shorelines*. The integer 1 represents a *flowline* and the integer 2 represents a *shoreline*. This coded value domain is being used by the HydroEdge subtypes to enforce the user to choose between these two options.

On a similar way, coded value domain HydroFlowDirections contains values that describe how the line features were built or digitized. 0, 1, 2 or 3 integer numbers are used by the model to distinguish between each type. This coded value domain is used on the HydroEdge feature class on its attribute named *FlowDir* (Fig. 3.8) with a value equal to 1. This obligates the user to introduce line features whose flow direction is *with digitized* (goes in the same direction as the line was digitized).

As discussed before, the relationships between the feature classes are represented as association lines on the UML. This is a representation of a relationship class. All the relationship classes on ArcHydro are based on the *HydroID* attribute. The HydroID attribute is present on all the ArcHydro feature classes. It is an integer number that uniquely identifies the features within the geodatabase. This is the key for all the relationship classes between the classes in the ArcHydro data model. Relationship classes are created between HydroJunction (that has an origin key of HydroID); with MonitoringPoint, Watershed and Waterbody (that have a foreign key of JunctionID). This means that the features contained in the Watershed, Waterbody and MonitoringPoint feature classes store the HydroID of the HydroJunction feature class in their JunctionID attribute. For example, a watershed-outlet relationship (a watershed delineated from an outlet), has a logical relationship that can be modeled by this relationship class. The watershed feature will contain the HydroID number of the outlet feature in its JunctionID attribute (Fig. 3.9).

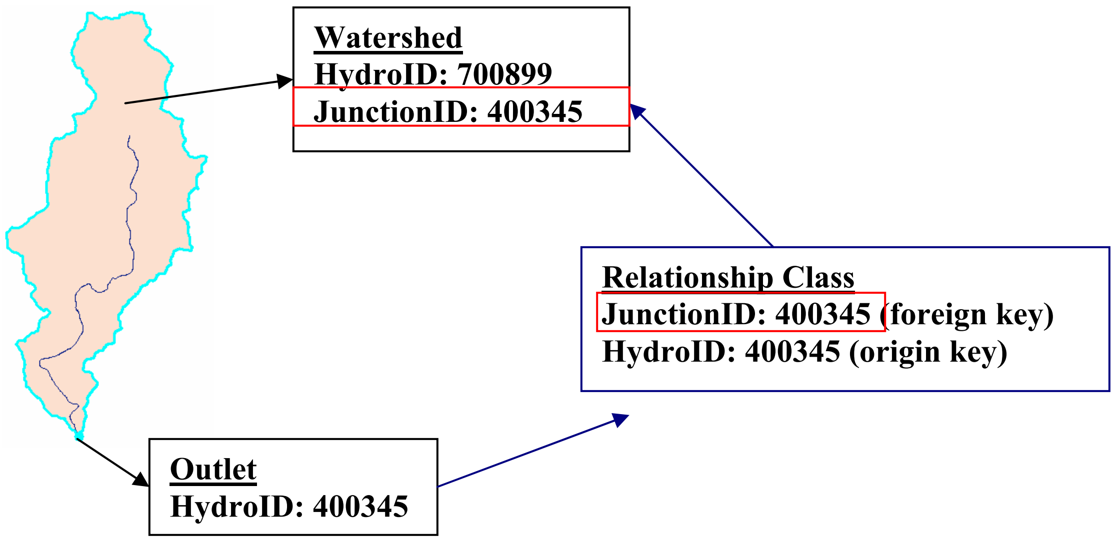



**FIG. 3.9.** Watershed-outlet topological relationship.

In addition, it is important to stress the importance of the *direction* of the relationship class. In the ArcHydro Framework, HydroJunction is the origin class and Watershed, MonitoringPoint and Waterbody are the foreign classes. All of the relationship classes have a cardinality of one to many. Hence, there can be several Watershed features related to one HydroJunction feature; several MonitoringPoint features related to one HydroJunction feature; or several Waterbody features related to one HydroJunction feature.

*HydroCode* is a class attribute that contains a string that helps identify the feature. For example it can contain an agency code that will help locate this feature on a data model different than ArcHydro.

*NextDownID* can be used in HydroJunction for tracing purposes on the Geometric Network.

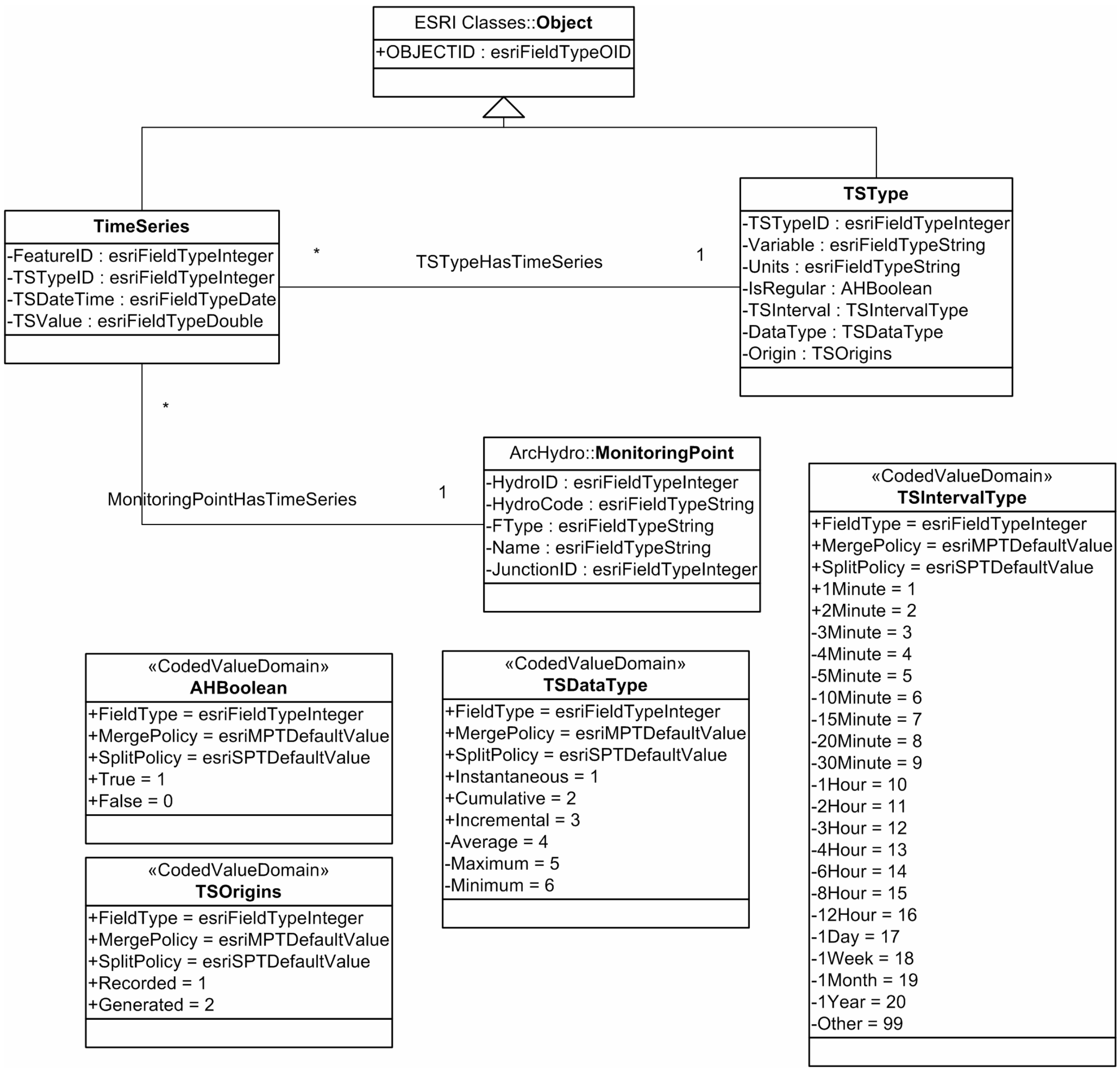



**FIG. 3.10.** Time Series UML in ArcHydro data model (Maidment *et al*, 2002).

In Fig. 3.10, the Time Series module of ArcHydro is described as a UML model. The concept is that a feature's time series can be defined by 3 dimensions. The axes of these 3 dimensions are: *FeatureID, TSDateTime and TSTypeID*, representing space, time and variable measured (Maidment *et al*, 2002). The FeatureID attribute will contain the HydroID number of the particular feature in the MonitoringPoint feature class to which the time series object relates; therefore, it is a direct link to the spatial location. The TSDateTime attribute contains the time when the variable was measured (with accuracy up to the millisecond if necessary). TSTypeID is an attribute that indicates which variable is measured and what are the characteristics of the variable.

*TimeSeries* and *TSType* are children object classes of the ESRI's object class; therefore, an OID field is inherited on each class.

TSType is related to TimeSeries through a one-to-many relationship class that has TSTypeID as origin and foreign keys. Thus, a variable can exist on many objects (rows) of a TimeSeries class (table). TSType class attribute values are enforced by coded value domains:

- AHBoolean: defines if the time series variable's interval is regular or not.
- TSOrigins: defines if the time series variable comes from an observation (recorded) or a models output (generated).
- TSDataType: defines which type of time series is the variable describing (i.e. incremental, maximum, minimum).
- TSIntervalTime: defines the time series variable's interval.

### 3.4.2 ArcGIS SWAT 2003 interface process structure

ArcGIS SWAT interface follows the logic of AVSWAT for calculating parameters and creating input files for the SWAT model: following the parameterization levels (Fig. 3.11) previously defined on the SWAT model description (Di Luzio et al, 2002).

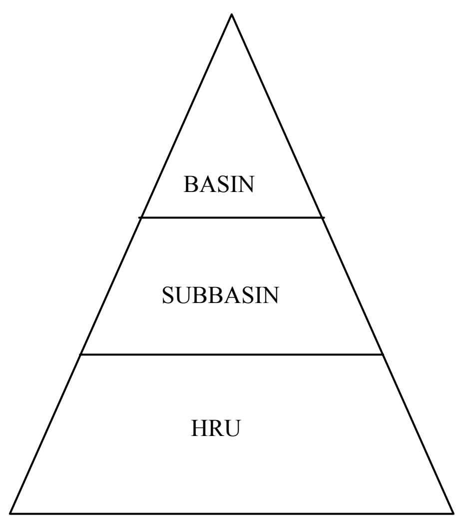



**FIG. 3.11.** Parameterization levels.

The creation of the input files for the SWAT model follows the next processes: watershed delineation, HRU distribution, weather definition, and writing and editing SWAT input. The case study, which will be presented later in section 4, is a practical example of these processes.

### *3.4.2.1 Watershed delineation*

The objective of the watershed delineation module is to delineate a basin with its corresponding subbasins and streams. In addition, it calculates the parameters that can be extracted from a DEM following a methodology similar to that of CRWR-PrePro (Olivera, 2001), and with a similar *Graphic User Interface* (GUI) as AVSWAT .

Table 3.6. summarizes the required input for the watershed delineation module.

**TABLE 3.6**. Input data for the watershed delineation module

| Input data | Data type | Remarks |
|---|---|---|
| DEM | GRID | Integer or real numbers for elevation values. The units that are used to define the map resolution and the elevation are not required to be identical. For example, the map resolution may be in meters while the elevation may be in feet. The map resolution must be defined in one of the following units: meters, kilometers, feet, yards, miles, decimal degrees.The elevation must be defined in one of the following units: meters, centimeters, yards, feet, inches. |
| Dem Mask | GRID or Feature or Draw Manually | The interface allows a mask to be superimposed on the DEM. The interface differentiates the mask grid into areas classified as category 0 (no data) and areas classified as any category > 0. Areas of the DEM grid for which the Mask grid has a value of 0 will not be processed for stream delineation. |
| Stream Network (Used for "burning" operation") | Digitized stream(Feature)<br>Reach File V1(Feature)<br>RF2(Feature)<br>NHD(Feature) | The digitized stream network<br>The oldest set of U.S. stream network delineations<br>The second generation set of U.S stream network delineations<br>The latest set of U.S stream network delineation |
| Threshold value | Integer | Used to define the minimum drainage area that a cell must have in order to be considered part of the stream network. |
| Subbasins outlet location table | dBase Table | The subbasin outlet location table is used to specify the location of: additional subbasin outlet locations (for example, stream gaging locations).The use of a location table to import locations for subbasin outlets is recommended when the user plans to compare observed or measured data with SWAT results. |

The output of the Watershed Delineation module is composed of six datasets: W*atershed, Reach, Outlet, MonitoringPoint, LongestPath and Basin*. These datasets will be explained in depth in the context of the geodatabase design.

*3.4.2.2 HRU definition*

The HRU definition is the result of the process described below.

**3.4.2.2.1 Soils definition**

In this step, a GRID dataset of soils of the watershed is defined. The type of dataset that the interface expects is the AVSWAT/BASINS modified version of STATSGO, but the user can provide his/her own datasets in any vector or raster data structure. The supported data types are GRID, shapefile, personal geodatabase feature class and coverage. The interface performs all necessary data conversions. This means that any of the mentioned supported data types are converted to the GRID format. Additionally, the interface allows the user to merge and clip several datasets in order to fill the study area characterized by the basin boundary.

The user can perform a reclassification of the GRID dataset with the aid of lookup tables, or table modifications aided by the interface (explained in detail on section 4).

The result of the soils definition is a GRID dataset with a joined text file table that contains the STATSGO soil codes for each GRID value.

**3.4.2.2.2 Land use/Land Cover definition**

A GRID of Land use/ Land Cover (LULC) of the watershed is defined in this step. Any vector data type or GRID is supported. The interface performs automatically any necessary data conversion from vector format to GRID format. It also allows processes for merging and clipping the datasets using the watershed dataset boundaries. If the user provides his/her landuse GRID dataset, it has to be integer. A reclassification, with the aid of the interface must be done by the user in order to ensure that the resulting grid (with the resolution of the DEM) is coded with the SWAT databases codes.

The result of the LULC definition is a GRID dataset with a joined text file table that contains the SWAT codes for each GRID value. For a complete list of the SWAT codes, please refer to Neitsch, *et al* (2000) appendix A.1.

**3.4.2.2.3 Overlay of Soils with LULC**

In this step, a set of one table per subbasin, containing cross tabulated areas of LULC classes versus soils classes (operation known as *overlay* of two grids) is calculated and summarized on two tables (UNCOMB, LUSO) and one text file report.

UNCOMB table contains the area of each HRU per subbasin while LUSO contains the total areas of each LULC class and soil class on each subbasin.

#### 3.4.2.2.4 HRU spatial definition

The HRUs are spatially defined based on the following GIS and non-GIS processes:

1. A *combinatorial and* operation (CAnd) is executed between 3 GRID datasets: subbasins, Land use and Soils, using ArcGIS Spatial Analyst's raster calculator. CAnd calculates unique combinations on a cell-by-cell basis and assigns them a value on a new grid, maintaining also the original values as *attributes* in the Value Attribute Table (VAT). The syntax used was:

   *[subbasin grid] CAnd [Land use grid] CAnd [Soils grid]*

   The output is a HRU grid containing unique combinations of subbasin, land use and soils per subbasin, with a VAT that stores the three codes.
2. The HRU GRID is converted to vector format. The output is a feature class that contains a record per HRU isolated polygon.
3. The HRU feature class is dissolved based on the GRIDCODE of the HRU grid. The output is a feature class with a complex polygon feature per row. Each complex polygon corresponds to each GRID value (not the SWAT code). This feature class is called FullHRU.

4. Two lookup dictionaries containing grid value vs. SWAT code are defined for land use and soil datsets. These dictionaries are build from the GRID's joined text file tables. Based on these dictionaries, SWAT codes are copied to the FullHRUs feature class and dissolved based on the SWAT codes. It is now called PolyHRU.

SWAT doesn't require georeferenced HRU datasets as PolyHRU and FullHRU to carry out simulations. It just requires a table containing a list of the HRU's that exist in a certain subbasin, because it lumps them on each subbasin to perform the simulation. There are several advantages of using their location in the input and output viewpoint of the SWAT simulation.

From the input viewpoint, important input parameters can be better estimated considering an HRU spatial distribution. Most of these parameters are summarized on the HRU text file that SWAT uses as input (see Neitsch, 2000). One of these parameters, the HRU mean slope is now calculated automatically using GIS methods. Previously, it was generalized by AVSWAT as the subbasin mean slope. The mean slope is calculated with the following procedure:

1. Clip the DEM using the Basin feature class.
2. Calculate the slope percentage grid using the SLOPE function over the clipped DEM. The SLOPE function of the Spatial Analyst extension works over a 3-cell by 3-cell neighborhood applying the following algorithm:

$$rise/run = SQR(SQR(dz/dx)+SQR(dz/dy)) \tag{3.1}$$

3. The mean slope of each HRU is calculated using Zonal Statistics. A zonal statistics operation requires the following input: a zone dataset (the HRUs grid), a value dataset (the slope grid) and a statistic (mean value). This means that within each HRU zone, the mean slope is calculated using the slope grid.

On the output viewpoint of the SWAT simulation, the user knows how the values of the output parameters are spatially distributed over the subbasins. For example, the user can know how the runoff is distributed over the subbasin.

#### 3.4.2.2.5 HRU filters

On this step, the user has the capability of filtering the HRUs. The *filtering* operation is used to eliminate small/non-representative HRU polygons and therefore, reduce the number of HRUs. It can be done by two methods:

- Dominant HRU and
- Percentage of land class use area over subbasin area (the area percentage that the land use class covers the subbasin) and a percentage of soil area over land use class area (the area percentage that a soil class covers a land use class).

With these filters, the unnecessary HRUs are eliminated and the remnant HRUs are normalized. This means that new areas are calculated for each HRU, in order to make the sum equal to the subbasin area.

If any of these filters is to be applied, a new mean slope is calculated for each HRU based on the weighted average of the new HRU area, according to the following formula:

$$New\ HRU\ Mean\ Slope = \frac{\sum_{i=1}^{n} MeanSlope_i \times Normalized\ Area_i}{\sum_{i=1}^{n} Normalized\ Area_i} \qquad (3.2)$$

*3.4.2.3 Weather definition*

The input parameters for SWAT are defined in this module. Using a weather station dataset, the interface calculates automatically the nearest weather station to each subbasin and summarizes the parameters on several tables. In addition, the user can provide the following custom data: rainfall, solar radiation, temperature, wind speed, weather or relative humidity. The custom data can be provided on the following formats: text file, personal geodatabase standalone table or dBase table.

If the user doesn't provide the custom data, SWAT generates the data with its weather generator model.

*3.4.2.4 Writing and editing SWAT input data*

In this module, a personal geodatabase stand alone table per SWAT input file (Arnold et al, 2002) is created. The idea of AVSWAT of creating a dBase interface between the input text files and the user is expanded with the incorporation of these tables inside the geodatabase as standalone tables that can be related with spatial data following the geodatabase design. The design of the geodatabase and the tables are explained furthermore in the following chapters.

### 3.4.3 Geodatabase design

*3.4.3.1 General considerations.*

The database structure and it's interaction with the interface is designed in a different way compared to AVSWAT. This constitutes the core of the improvements of this new interface.

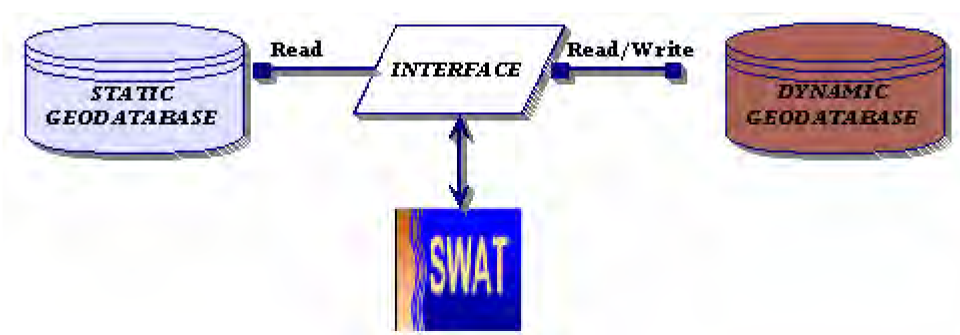



**FIG. 3.12.** Interface-SWAT-geodatabase relationship.

Fig. 3.12 shows the basic components of the relationships between the interface/geodatabase structure and with the SWAT model.

Two geodatabases have been designed to establish the relationship between the interface and the SWAT model: The *Static* and the *Dynamic geodatabase*s. The *Static geodatabase* manages spatial and tabular data that are not project specific. The *Dynamic geodatabase* is created for each project. It stores all the spatial and tabular data produced by the interface processes and also the data that the user inputs for the model calculations. Fig. 3.13 details the interaction between the geodatabases and the interface at every process stage. The design and content of the two geodatabases are explained in the subsequent sections.

The *Dynamic geodatabase* supports read and write options. The *Static geodatabase* is read-only. The user can modify its information using provided specialized tools for its edition, since the stored information is used by all projects.

There are three ways in which the structure of a geodatabase can be designed: (1) Creating the *schema* (structure of the geodatabase) in ArcCatalog (a component of the ArcGIS 8.x series), (2) Importing existing data to an already existing empty geodatabase and (3) Designing it with VISIO and CASE (Computer-Aided Software Engineering). The CASE tools help the user in the definition of *custom object classes,* their *behavior, relationships* and the schema in which they exist (MacDonald, 2001).

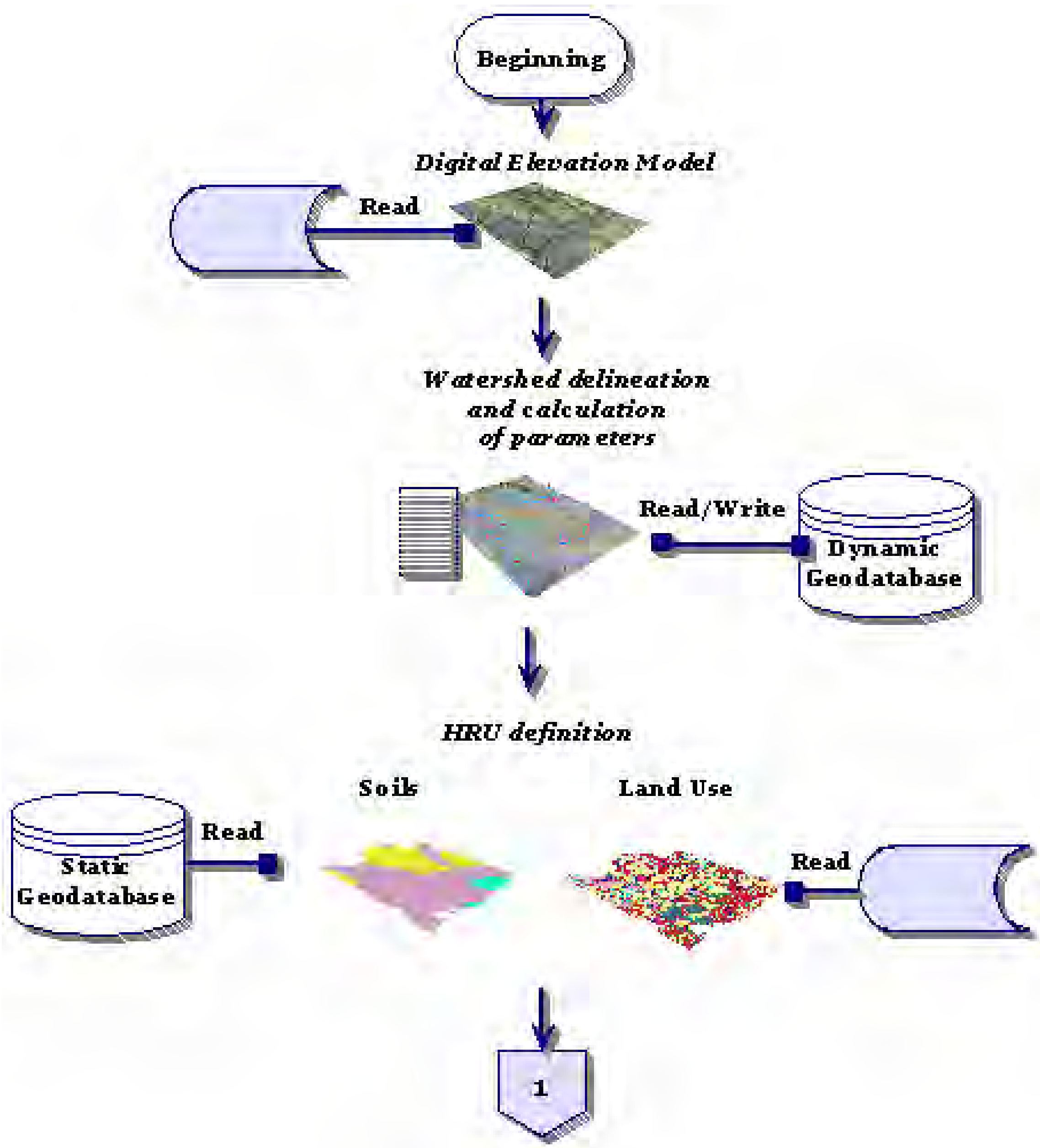



**FIG. 3.13.** Interaction between geodatabases and interface processes.

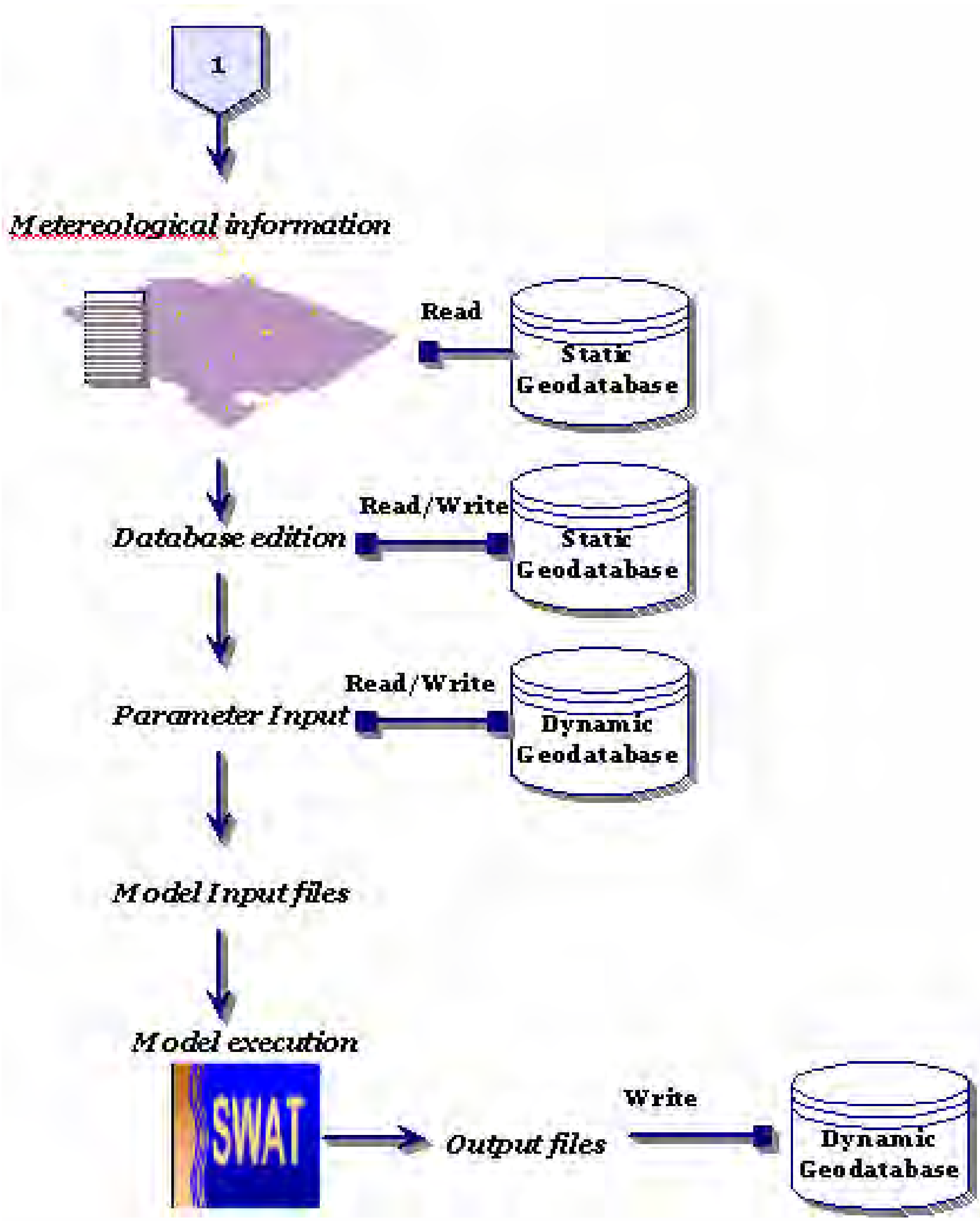



**FIG. 3.13.** Continued.

Several considerations were made for the design of the Dynamic and Static geodatabases:

1. It was decided that the geodatabases should be Personal Geodatabases (PGDBs). This means that only one user can edit a geodatabase at a time and that the format is a customized Access database. As consequences, we have:
    - The user is limited to the use of Arc GIS desktop in personal computers (working under Windows operating system). Enterprise databases are not supported by the interface.
    - Raster data is stored on Operating System folders, similar to AVSWAT.
2. It was decided to make the PGDBs fully compatible with ArcView GIS 8.x. Procedures that can be performed only with ArcEditor or ArcInfo 8.x were excluded from the design of the interface and geodatabases. In this way:
    - Geometric Networks are not supported by the interface.
    - Relationship classes are not created and/or maintained by the interface. Instead Memory Relationship Classes are defined automatically by the interface on every project.
    - CASE tools option for the creation of the geodatabase schema was not used. The interface creates the geodatabase from scratch using ArcObjects.

- More people can use the interface due to the reduced costs of ArcView compared with ArcEditor or ArcInfo. This point is important since the SWAT model is of public domain.

3. It is *ArcHydro compliant*.

   ArcHydro compliant is a term that implies ArcHydro compatibility. ArcHydro is a data model that can be implemented with ArcEditor and ArcInfo 8.x series. In this way, the ArcView user cannot apply the schema of the ArcHydro data model on his/her data.

   Without geometric networks and relationship classes, and with the addition of several components, the final design was far from the original ArcHydro framework design. Certain valuable elements and concepts of ArcHydro have been kept:

   - The concept of a MonitoringPoint dataset keeps the same meaning, concept and relationships on the Dynamic geodatatabase.
   - The concept of a unique hydrologic identifier for all the features within the geodatabase has also been used. This is the key for most of the relationships in the Dynamic geodatabase. Although we don't use the same algorithm used by ArcHydro (for its generation), the final use remains the same.
   - The TimeSeries package is kept with the same concept and relationships.

The way in which ArcHydro schema was designed was using VISIO and CASE Tools. VISIO is Microsoft software that allows the user the creation of UML models. ArcInfo/ArcEditor's CASE Tools come with ESRI VISIO templates that contain all the ESRI classes and interfaces already designed in UML. The finished UML model is then exported as an XML file format. From ArcCatalog, the CASE Tools wizard allows to import the XML file and convert it to the geodatabase applying the schema designed in UML. This process can't be carried out under the ArcView framework.

In the case of ArcGIS SWAT geodatabases, they were created from scratch using ArcObjects (for the Dynamic geodatabase) and ArcCatalog (for the Static geodatabase). The UML design, tough, was done using VISIO 2002. The Visual Basic code follows the UML model logic, for ArcGIS SWAT geodatabases creation. The designed UML models allow easier creation of geodatabases using ArcInfo. The resulting schema would have a more efficient design since the relationship classes are explicitly defined.

*3.4.3.2 Dynamic geodatabase design*

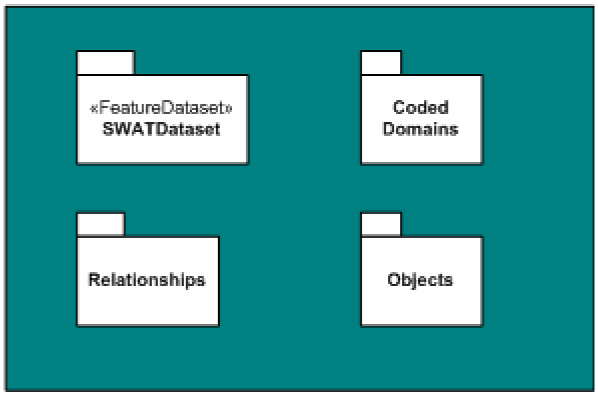



**FIG. 3.14.** Core model packages for the Dynamic geodatabase.

A package is a basic element for grouping *objects* on a UML model. Like the *folders* in Windows group files, they group objects. A specific type of package is ESRI's feature datasets, defined previously as a collection of feature classes. The UML model of the Dynamic geodatabase has one feature dataset called *SwatDataset* The other three packages in Fig. 3.14, were created for organizational purposes. The coded domain values for the Dynamic geodatabase are stored under the *Coded Domains* package. In the same fashion, the relationships (relationship classes) between all the classes of the UML model are defined under the *Relationships* package. Similarly, all the object classes (that are not children of the ESRI's feature class, otherwise they will have the *shape* field that is a characteristic of a feature class) are stored in the *Objects* package.

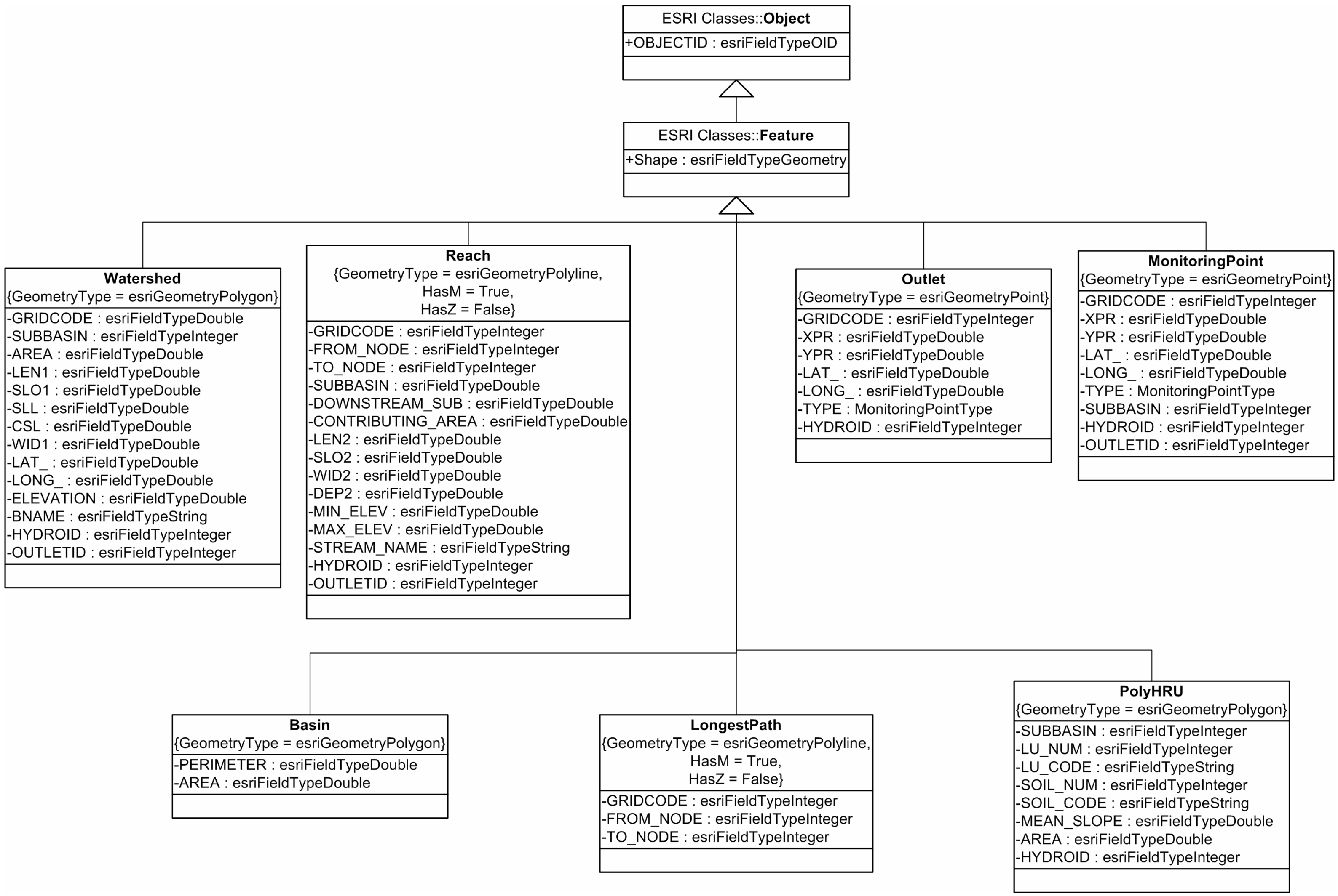



**FIG. 3.15.** The SwatDataset UML model.

#### 3.4.3.2.1 The SwatDataset feature dataset

Fig. 3.15 details the UML model of the SwatDataset All the classes in this model are child classes of the ESRI's object class and feature class. Thus, all the custom feature classes: Watershed, Reach, Outlet, MonitoringPoint, Basin, LongestPath and PolyHRU have the OID attribute and the Shape attribute.

All the feature classes, but PolyHRU, are created at runtime in the watershed delineation module. Therefore, they contain all the possible hydrologic information that can be derived from a DEM. The FullHRU dataset is not included inside the SwatDataset feature dataset, since it is not used by the SWAT model. It is included in the Dynamic geodatabase as a stand-alone feature class.

The Watershed feature class contains subbasin features with the attributes that the class shows in Fig. 3.15. It is defined as a polygon feature class. Table 3.7. summarizes and lists its attributes. All of the attributes are populated in the watershed delineation module. For a more detailed explanation on the meaning of the attributes, refer to Neitsch, *et al*, 2000.

The Reach feature class is defined as an ESRI's polyline that can contain measures (HasM=true) and can only be represented in two dimensions (HasZ=false). The reach feature class contains the stream features defined after the *threshold* and *custom outlets definition* on the watershed delineation module. Its attributes are listed in Table 3.8.

**TABLE 3.7.** Watershed feature class attributes

**watershed**

| Field Name | Description |
|---|---|
| GRIDCODE | key field |
| Subbasin | Subbasin number |
| Area | Subbasin area[hectares] |
| Slo1 | Subbasin slope[%] |
| Len1 | Stream reach(longest path within the subbasin) length[meters] |
| Sll | Field slope length[meters] |
| Csl | Stream reach(longest path within the subbasin) slope [meters] |
| Wid1 | Stream reach(longest path within the subbasin) width [meters] |
| Dep1 | Stream reach(longest path within the subbasin) depth [meters] |
| Lat | Latitude of the subbasin labelpoint |
| Long_ | Longitude of the subbasin labelpoint |
| Elev | Elevation of the subbasin labelpoint[meters] |
| Bname | String avaiable for labeling the TOC |
| HydroID | Unique ID for Watershed |
| OutletID | Outlet ID(key field) |

**TABLE 3.8.** Reach feature class attributes

**reach**

| Field Name | Description |
|---|---|
| GRID_CODE | Key Field |
| FROM_NODE | Arcmap Internal Field |
| TO_NODE | Arcmap Internal Field |
| Subbasin | Subbasin number |
| Downstream_sub | Subbasin number receiving surface water from the subbain |
| Contributing_area | Cumulated drainage area[hectares] |
| Len2 | Stream reach length[%] |
| Slo2 | Stream reach slope[%] |
| Wid2 | Stream reach width[meters] |
| Dep2 | Stream reach depth[meters] |
| MinEl | Mininum elevation of the stream reach[meters] |
| MaxEl | Maxinum elevation of the stream reach[meters] |
| HydroID | Unique ID for reach |
| OutletID | Outlet ID |

The Reach feature class differs from HydroEdge feature class (ArcHydro class) in that it is a simple feature (as opposed to the ArcHydro HydroEdge complex features) that can be managed and edited within ArcView 8.x series.

The Outlet feature class is defined as an ESRI point feature class. It contains features that model outlet points either generated automatically by the interface's delineation or added by a table provided by the user. Differs from HydroJunction feature class (ArcHydro class) in that it doesn't contain junctions (junction type), that are required for the creation of the geometric network. It can be edited and managed within ArcView. A list of the attributes of the outlets is shown on Table 3.9.

**TABLE 3.9.** Outlet feature class attributes

**outlet**

| Field Name | Description |
|---|---|
| GRIDCODE | Key field |
| Xpr | X coordinate in the current projection |
| Ypr | Y coordinate in the current projection |
| Lat | Latitide-decimal degrees |
| Long_ | Longitude-decimal degrees |
| Type | Monitoring Point Type |
| HydroID | Unique ID for Outlet |

The *Monitoring Point Type* detailed on Table 3.9 will be described on detail on the Coded Value Domains chapter.

The tracing capabilities of the ArcHydro data model were based on the geometric network and the *NextDownID* attributes of the HydroJunction and Watershed feature classes. In the Dynamic geodatabase, tracing is based on the Downstream_sub attribute of the reach feature class. This can be done because within the SWAT model there is a one-to-one relationship between streams, watersheds and outlets.

MonitoringPoint is defined as an ESRI's point feature class. It reflects the same concept as the ArcHydro MonitoringPoint class. For the specific purpose of SWAT, it stores: reservoirs, inlets, outlets (a copy of the outlet feature, with the difference that this feature can be related to TimeSeries), Point Source Discharges (PSD), and any custom gauging station. Its attributes are explained briefly in the Table 3.10.

**TABLE 3.10.** Monitoring Point feature class attributes

| **Monitoring point** | |
|---|---|
| Field Name | Description |
| GRID_CODE | Key field |
| Xpr | X coordinate in the current projection |
| Ypr | Y coordinate in the current projection |
| Lat | Latitide-decimal degrees |
| Long_ | Longitude-decimal degrees |
| Type | Monitoring Point type |
| Subbasin | Number of subbasin |
| HydroID | Unique ID for Monitoring Point |
| OutletID | Oulet ID |

*Basin* is a polygon feature class that will contain the whole watershed feature. This means that it will contain only one feature. It has two attributes that are populated on the watershed delineation module: area and perimeter. AVSWAT allowed the user to define more than one *basin* in the project, but the SWAT model works only with one basin at a time, therefore it is better to define a project per basin. In this new version, the user is expected to define one per project. Consequently the feature class attribute table will contain just one row.

*LongestPath* will contain the longest flow path feature of each subbasin. The *longest flow path* is the distance that goes from the outlet of the subbasin to the most remote hydraulic point in the subbasin. This means that it is the longest path that a drop of water will follow from a point in the subbasin boundary to the outlet. It is defined as a feature class that will contain polylines representing the longest path followed by the water. It may or may not coincide with the streams features. The longest flow path is used to calculate hydrologic parameters such as the time of concentration in the subbasin.

*PolyHRU* will contain complex polygon features defined with the already explained procedure (section 3.4.2.2.4. HRU spatial definition). It is a custom feature class that wasn't included in the ArcHydro framework but it can be found a similar in the full version of the ArcHydro data model under the name of HydroResponseUnits. It has no similar representation in AVSWAT. AVSWAT calculated areas for the HRUs using a cross tabulation algorithm. This algorythm used an ArcView command for the automatic calculation of the areas of unique combinations of two grids. It didn't create a georeferenced dataset that can exploit the advantages of GIS analysis. ArcGIS SWAT allows the user to take advantage of the relationships that exist between the georeferenced dataset with input and output data of the SWAT model.

The relationships that exist between the classes are going to be explained in detail in the *Relationships* chapter.

### 3.4.3.2.2 Coded Value Domains

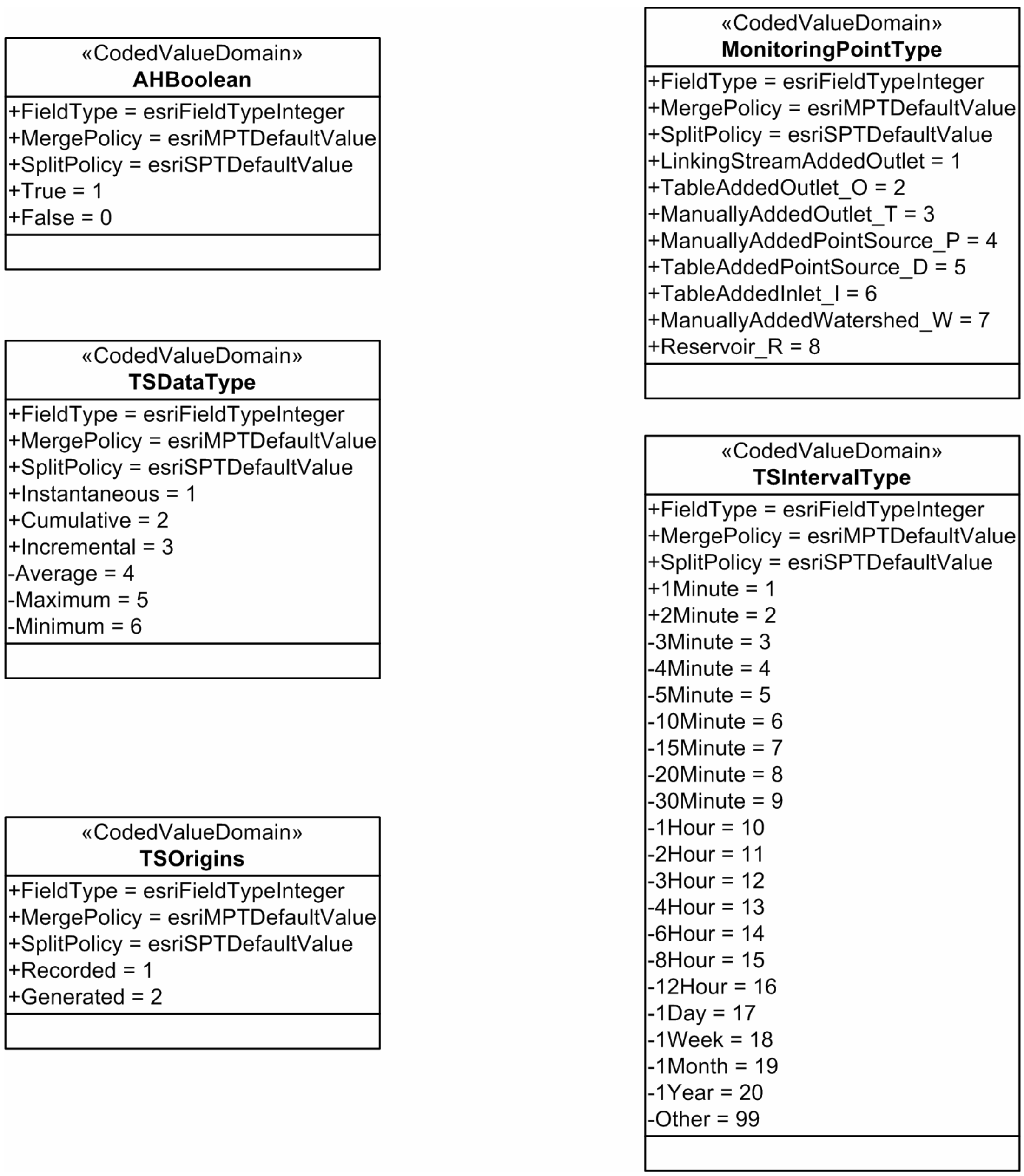



**FIG. 3.16.** Coded value domains for the Dynamic geodatabase.

The coded value domains (Fig. 3.16) contain the same domains that ArcHydro uses for its Time Series module with the addition of a type for the MonitoringPoint features. This type will help to identify the type of MonitoringPoint feature, without the creation of subtypes that are limited to ArcInfo and ArcEditor. As it was explained in the UML theory, a coded value domain allows the user to chose from a predefined list. The model recognizes the choices from a discrete list. In the *MonitoringPointType* list, there are 8 possible choices with integer numbers from 1 to 8. The number 1 (*LinkingStreamAddedOutlet)* will represent the points that were added automatically by the computer to identify the end of the stream lines. Number 2 (*TableAddedOutlet*) represents the outlets that were added to the project as a table (dBase, PGDB table, Text File supported) that contained their coordinates. Number 3 (*ManuallyAddedOutlet*) represents the outlet points that were manually added using ArcGIS SWAT editing tools. Number 4 (*ManuallyAddedPointSource*) represents the Point Source Discharge (PSD) points that were added manually using ArcGIS SWAT editing tools. Number 5 (*TableAddedPointSource*) represents the PSD points that were added using a table (dBase, PGDB table, Text File supported) that contains their coordinates. Number 6 (*TableAddedInlet*) represents the *inlet* points that were added using a table (same formats supported) containing their coordinates. Number 7 (*ManuallyAddedWatershed*) represents the points that were manually added using ArcGIS SWAT editing tools. An inlet is a point that summarizes how much water is draining into the subbasin, from an upstream subbasin.

#### 3.4.3.2.3 Objects

Figures 3.18., 3.19. and 3.22. show the UML model of the object classes in the Dynamic geodatabase. The object classes are derived from the ESRI object class. They are subdivided into object classes created for the SWAT input, object classes created for the SWAT output and the object classes for the Time Series.

The object classes on Appendix B, Fig. B.1 follow the concept of AVSWAT of being an interface between the user and each input text file. In the Dynamic geodatabase, this concept is extended using personal geodatabase stand alone tables (the object classes on the UML) that are explicitly related to spatial features in order to facilitate a pre analysis of the data that serves as input for SWAT.

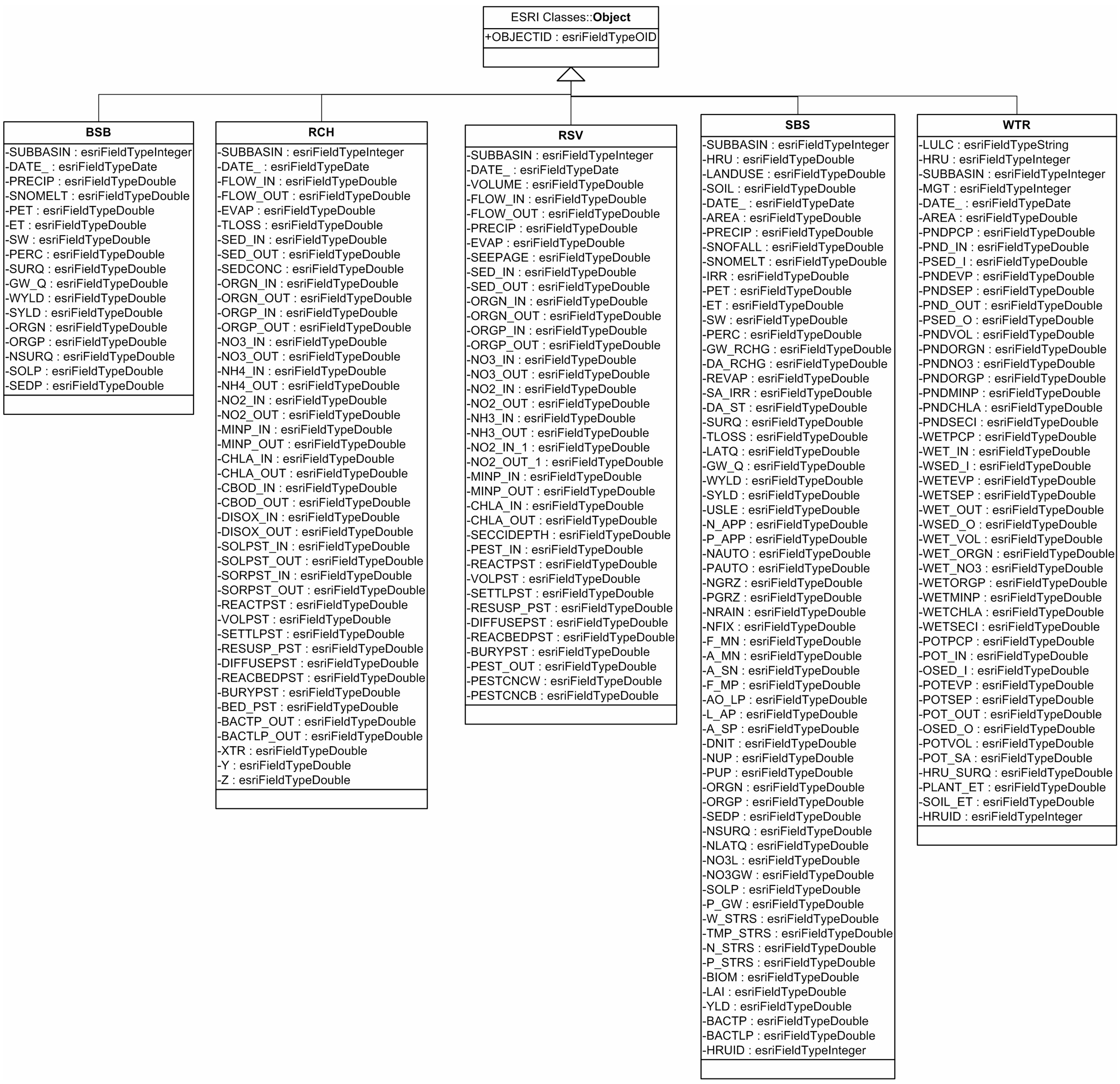



**FIG. 3.17.** UML model for the SWAT output object classes.

Similarly, the Dynamic geodatabase, with the model presented on the Fig. 3.18 and the relationships already explained, simplifies the analysis of the output data.

Each attribute on the object classes of the figures 3.17. and 3.18. represents a parameter to be included in an input text file. An explanation of each SWAT input parameter can be seen on Netsch, *et al*, 2000.

**3.4.3.2.4 Relationships**

Relationships in the Dynamic geodatabase are automatically created by the interface, using memory relationship classes. The UML model for the relationships is shown in Appendix B, Fig. B.2.

The base of the relationships is three global codes and one local code. The global code uniquely identify an object or feature within a geodatabase and a local code uniquely identifies an object or feature within a feature class. The *HydroID* attribute is a global code (uniquely defined within the geodatabase) that exists on every object of every feature class in the Dynamic geodatabase (the limitations of the *HydroID* global code are detailed on Appendix A). The second global code is the *OutletID*. The *OutletID* global code is used on a feature class to be related to the Outlet feature class. The *OutletID* contains the *HydroID* of the related outlet.

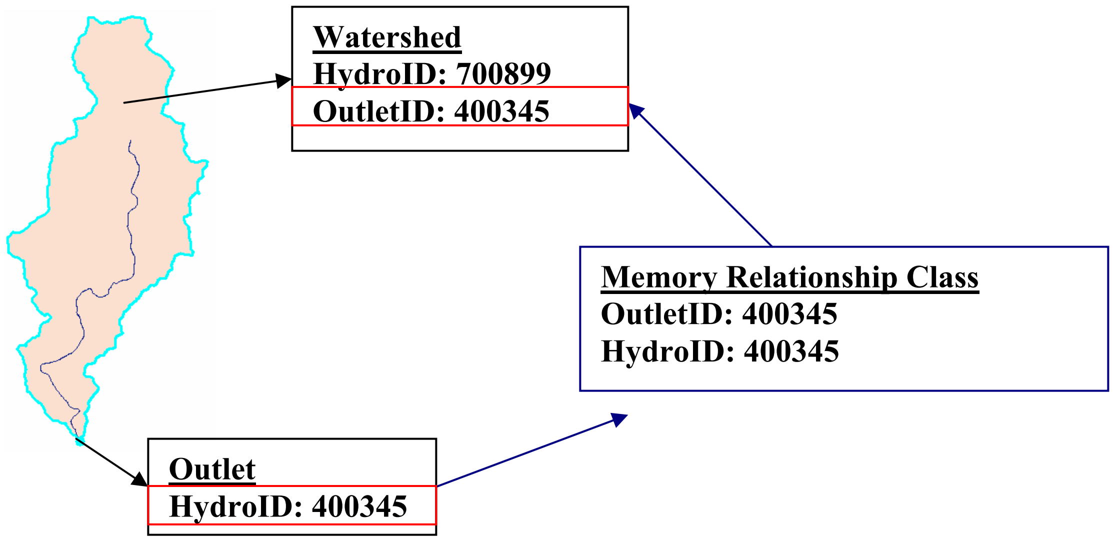



**FIG. 3.18.** Global codes example.

For example in Fig. 3.20 we have an outlet feature with a HydroID 400345 (this code is not repeated in the whole geodatabase). If the watershed with HydroID 700899 would be related to its outlet, it needs and OutletID field that is populated with the code 400345 and the memory relationship class (join or relate) could be established between the OutletID field on the watershed feature class with the HydroID field of the Outlet feature class. These two fields are often called *key* fields and depending on the direction of the relationship they are called *origin* and *foreign key*.

The third global code is *HRUID* and, as *OutletID*, it allows the relationship between any feature or object class with the PolyHRU feature class.

*Subbasin* is a local code that stores the subbasin number generated automatically by ArcGIS. This means that this number is unique just within the feature class and it can be repeated in the geodatabase. It works only for relationship purposes. If an object class is related to the watershed feature class, both of them would have the same attribute name to connect them: *Subbasin*.

There is a logical relationship between Watershed, Reach and Outlet. The relationships between these three feature classes are based on the logic of SWAT. A subbasin can have one reach only and a reach can have one outlet only. This logic creates a one to one relationship between them, as opposed to the one-to-many relationships defined in AVSWAT and the ArcHydro data model.

The relationship between Watershed and PolyHRU is one to many. This relationship can prove beneficial in the input and output analysis. They are based on the *Subbasin* local code.

The relationships starting at PolyHRU are one-to-one since the HRU is the smallest hydrologic unit of the model.

The name of the relationships reflects the direction in which the relationship is being build. For example, the OutletHasWatershed relationship shows a relationship that goes from Outlet (origin class) to watershed (foreign class). Table 3.11 summarizes the relationships between the object and feature classes of the Dynamic geodatabase.

**TABLE 3.11.** Table of relationships between classes in the Dynamic geodatabase

**Relationships**

| Name | Relationship type |
|---|---|
| OutletHasReach | one-to-one |
| OutletHasWatershed | one-to-one |
| OutletHasMonitoringPoint | one-to-many |
| WatershedHasWGN | one-to-many |
| WatershedHasSWQ | one-to-many |
| WatershedHasRSV | one-to-many |
| WatershedHasRCH | one-to-many |
| WatershedHasBSB | one-to-many |
| WatershedHasWUS | one-to-one |
| WatershedHasInlet | one-to-one |
| WatershedHasNPSD | one-to-one |
| WatershedHasPND | one-to-one |
| WatershedHasRTE | one-to-one |
| WatershedHasSUB | one-to-one |
| WatershedHasSUBWGN | one-to-one |
| WatershedHasReservoir | one-to-one |
| WatershedHasCHM | one-to-many |
| WatershedHasGW | one-to-many |
| WatershedHasHRU | one-to-many |
| WatershedHasHRUS | one-to-many |
| WatershedHasSBS | one-to-many |
| WatershedHasPolyHRU | one-to-many |
| WatershedHasWTR | one-to-many |
| WatershedHasSOL | one-to-many |
| WatershedHasMGT1 | one-to-many |
| WatershedHasMGT2 | one-to-many |
| WatershedHasPSD | one-to-many |
| PolyHRUHasHrus | one-to-one |
| PolyHRUHasHru | one-to-one |
| PolyHRUHasGW | one-to-one |
| PolyHRUHasCHM | one-to-one |
| PolyHRUHasSBS | one-to-one |
| PolyHRUHasWTR | one-to-one |
| PolyHRUHasMGT1 | one-to-one |
| PolyHRUHasMGT2 | one-to-one |
| PolyHRUHasSOL | one-to-one |
| MonitoringPoinHasPSD | one-to-one |
| MonitoringPointHasReservoir | one-to-one |

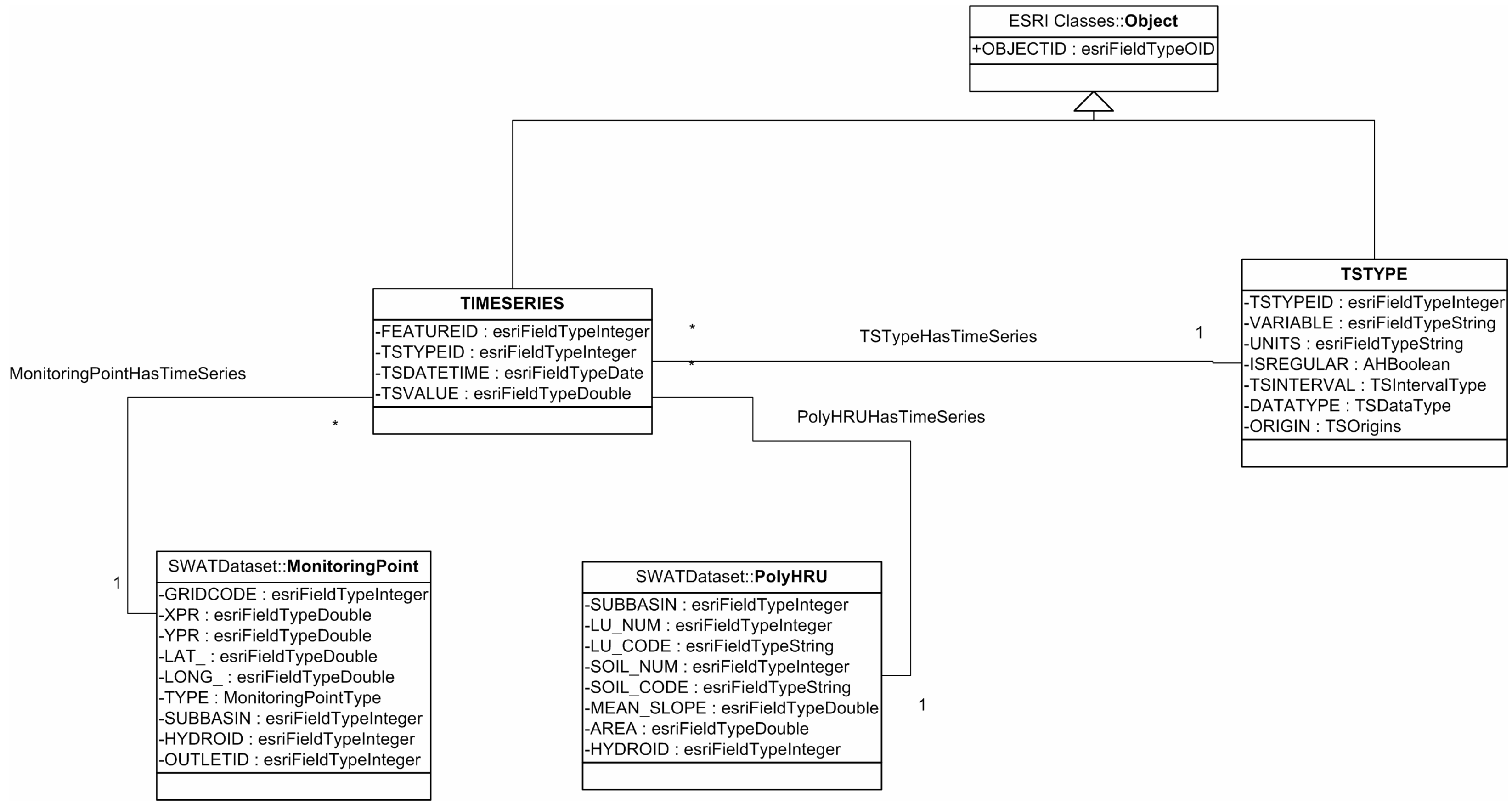



**FIG. 3.19.** UML model for the Time Series in the Dynamic geodatabase.

**3.4.3.2.5 Time Series on the Dynamic geodatabase**

Time Series on the Dynamic geodatabase (Fig. 3.21) have a similar design to the ArcHydro data model. However, there is the addition of the PolyHRU feature class that is linked to the TimeSeries object class using the HydroID. SWAT gives as output a considerable amount of temporal information linked to each HRU. Therefore a relationship between these two classes was necessary for the Dynamic geodatabase.

*3.4.3.3 Static geodatabase design*

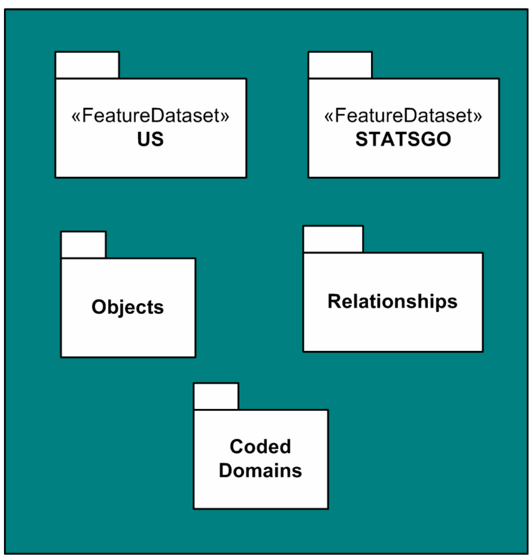



**FIG. 3.20.** Static geodatabase core model packages.

The purpose of the Static geodatabase is to serve as a supporting structure for every SWAT project. This supporting structure provides basic data from custom and SWAT databases. It is organized in 5 packages as Fig. 3.22 shows.

#### 3.4.3.3.1 US Feature dataset

The *US* Feature dataset includes a USGS weather stations feature class for the whole United States. Each weather station is populated with the attributes detailed on the UML model in Appendix B, Fig. B.3. An in-depth explanation of these parameters can be found at Neitsch et al, 2000.

The feature dataset can also contain simple feature classes that show information for the whole United States.

#### 3.4.3.3.2 STATSGO Feature dataset

The feature dataset called STATSGO contains the STATSGO soils database. Fig. 3.24 shows the classes in this package.

The modified version of STATSGO database for ArcGIS SWAT is composed of 3 groups of elements:

1. A Feature dataset element as shown on Fig. 3.24. The name of the feature dataset is the State Name (i.e. Texas). There is one feature dataset per state.

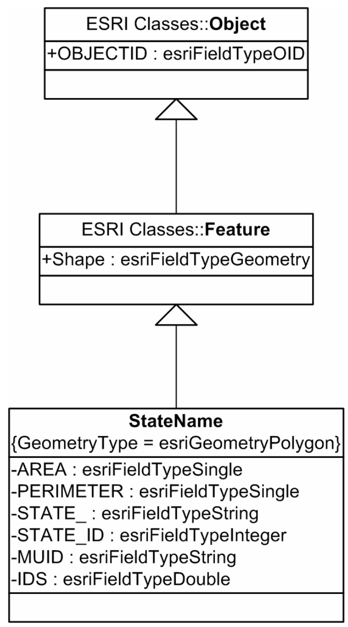



**FIG. 3.21.** UML model for the STATSGO dataset

2. A summary of all soils in the United States with their respectives codes: MUID, S5ID and Name (State Soil Geographic Database, 1995). This summary table is located outside the feature dataset, in the main root of the Static geodatabase (named SOILUS).
3. A group of tables that represent each soil within a state. Each table is named with the MUID code (i.e. TX002 for the soil 002 in Texas). These tables are located outside the feature dataset in the main root of the Static geodatabase. The number of tables will vary depending on the state. For example, there are 633 tables that represent 633 MUID codes for Texas. Pennsylvania has 104 tables.

The STATSGO dataset is not included in the US dataset because most of the time the information for all the United States is not necessary. This ensures the performance and efficiency of the Static geodatabase since its capacity is limited to 250000 features (ESRI, 2003b). A tool for importing and deleting STATSGO state information has been included in the ArcGIS SWAT interface. Each STATSGO soil table contains: (STATSGO, 1995)

- The number of components of the map unit (MUID) called Sequence Number (Seqn).
- Up to ten different layers containing soil characteristics and properties.
- A soil interpretation record (s5id)
- A percentage of the component in the map unit (comppct)
- A soil series name associated with the component or sequence number

| MUID | SEQN | SNAM | S5ID | CMPPCT | NLAYERS |
|---|---|---|---|---|---|
| PA003 | 1 | SHEFFIELD | OH0073 | 48 | 4 |
| PA003 | 2 | PLATEA | OH0075 | 20 | 4 |
| PA003 | 3 | PLATEA | OH0075 | 13 | 4 |
| PA003 | 4 | PLATEA | OH0075 | 10 | 4 |
| PA003 | 5 | HOLLY | OH0032 | 5 | 4 |
| PA003 | 6 | PIERPONT | OH0074 | 4 | 4 |

**FIG. 3.22.** Detail from a STATSGO table (MUID table).

Fig. 3.25 shows 6 fields from a modified STATSGO table. We can notice that the MUID value is the same for all rows. This is expected since the table is for the MUID PA003 (the table's name is also PA003). The sequence number details up to 6 components for this MUID. SNAM represents the name of the component and it can be repeated, but depending on the sequence number, it can have different hydrologic properties (since it can be located on a different region) and number of layers. CMPPCT values represent the percentages of the component in the MUID and they sum 100%. NLAYERS tell the number of layers that the certain component has. For this specific case all of the components have 4 layers. The table has several other fields also to the right, detailing the hydrologic properties for each layer. In this example, there is information to the right, up to the 4$^{th}$ layer. The rest of the fields, from the 5$^{th}$ layer to the 10$^{th}$ layer are present but empty. All the MUID tables have the same number of fields.

**3.4.3.3.3 Object classes**

The UML model of Appendix B, Fig. B.4 details the objects in the Static geodatabase. The UML model includes an example of one STATSGO MUID table.

The object classes are classified in the following groups:

1. STATSGO MUID tables
2. Parameter range tables. These tables contain the ranges for each input parameter for the SWAT input. There is one parameter range table per SWAT input text files group. For example WUSRNG table contains the ranges for the parameters involved on the Water Use (WUS) text files group. It is called group since all the text files required as input for the Water Use have a common extension (*.wus) and there is one text file per subbasin (for the specific case of Water Use).

3. Lookup tables. These tables are used to translate two common land use datasets: NLCD and USGS (Anderson) to SWAT codes. It also contains the TSType table that is copied to each Dynamic geodatabase (each SWAT project).
4. Management Tables. These tables have the information of the SWAT databases.
5. User defined Weather stations are also in the Management Tables group. This table contains the same attributes as the ones in the US weather stations feature class.

#### 3.4.3.3.4 Relationships

The relationships established inside the Static geodatabase are defined by the UML model on Fig. 3.27. These relationships are produced between the STATSGO feature classes and the tables. A feature class is related with the SOILUS summary table, with a relationship one-to-many. This means that each feature can be related to several records on the SOILUS table. There is another relationship that has been established but it can't be modeled using UML and is explained as follows.

Inside each State feature dataset there is one feature class with the same state name. This feature class contains a number of polygon features that represent the state soils (ie. Soils with MUID's: TX120, TX345, TX098). Each of these soil features has a *related* table, but the relationship is based on the table's name (MUID of the state soil). This table contains the information of all the components, each one of them with up to ten layers stored in as fields. Thus, a soil polygon with a MUID TX345 has a related table with name TX345. There is no relationship class, and no memory relationship class, but the interface is taking advantage of this linkage when retrieving data for each soil.

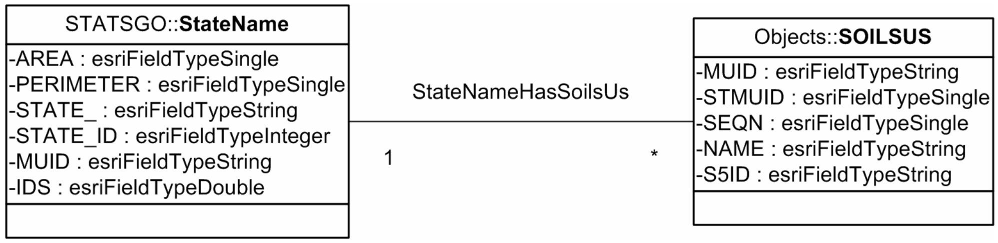



**FIG. 3.23.** UML model for the Static geodatabase relationships.

### 3.4.3.3.5 Coded value domains

Since the Static geodatabase contains the TSType table, it also contains the coded value domains that pertain to the TSType Table, as shown in the Fig. 3.28.

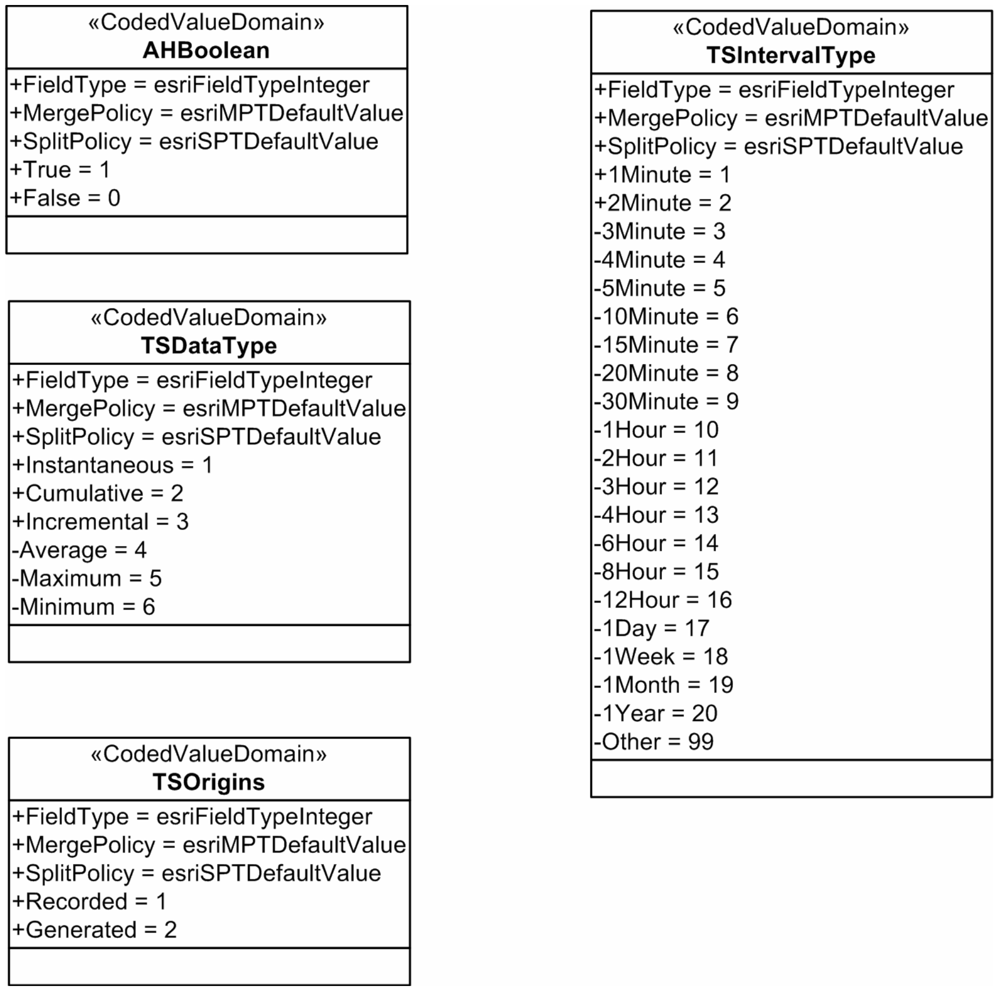



**FIG. 3.24.** UML model for the coded value domains of the Static geodatabase.

### 3.4.4 Managing SWAT output: Uncertainty Analysis

*3.4.4.1 Overview*

As described in this thesis, SWAT is a complex hydrologic model that requires a great amount of input data. These data consists of a number of physical, chemical, biological and empirical parameter values that need to be measured, calculated and/or estimated (parameterization). However, the methods for measuring parameter values are not exact (they carry an error); formulas and algorithms for calculating parameter values have their own errors; and finally, there is scarcity of data. When there is paucity of data (lack of knowledge of the system), it is necessary an estimation of parameter values based on "engineering judgement". Regardless of how the parameter values are obtained, they carry a degree of *uncertainty*. If the input parameter values are *uncertain*, the output values are *uncertain*. If it were possible to quantify the *uncertainty* of the parameter values, then it would be possible to quantify the uncertainty of the model output.

The SWAT Parameter Analysis Tool (SPAT), included in the interface, was developed to aid the user in quantifying the uncertainty of the output values of the SWAT model.

*3.4.4.2 SPAT and Monte Carlo simulation*

SPAT is based on the Monte Carlo Simulation technique. The Monte Carlo simulation technique "describes how uncertainty is passed from input variables to the output answer" (Wright, 2003). In this way, Monte Carlo simulation takes probability density functions (pdfs) of input variables and obtains pdfs of output results.

The methodology of SPAT can be outlined in the subsequent steps: (1) definition of the sensitive parameters and pdfs of their values; (2) application of the Monte Carlo simulation technique; and (3) Output analysis.

*3.4.4.3 Definition of the sensitive parameters and pdfs*

According to Neitsch *et al* (2000), there are 27 parameters that significantly affect the SWAT's output (Table 3.12 summarizes the 27 parameters and their location in the SWAT input files). These parameters are adjusted when calibrating the model. Some parameters, however, might be irrelevant for specific applications and purposes. Therefore, only those parameters that are relevant and that carry uncertainty are selected for the Monte Carlo simulations.

Once the selection of parameters has been done, it is necessary to establish what Wurbs *et al* (2001) defined as probability descriptors. Three values for each parameter must be defined: a low and high reasonable values and a best estimate.

**TABLE 3.12** SWAT sensitive parameters (Neitsch *et al,* 2000)

| # | Parameter | Definition | Location |
|---|---|---|---|
| 1 | USLE_C | Minimum value of USLE C factor for water erosion applicable to the land cover/plant. | Crop.dat |
| 2 | SMFMX | Maximum melt rate for snow during (mm/°C/day) where ?C pertains to the air temperature. | .bsn |
| 3 | SMFMN | Minimum melt rate for snow during the year (occurs on winter solstice) (mm/°C/day) (ref. the air temperature) | .bsn |
| 4 | SPCON | Linear parameter for calculating the maximum amount of sediment that can be reentrained during channel sediment routing | .bsn |
| 5 | SPEXP | Exponent parameter for calculating sediment reentrained in channel sediment routing | .bsn |
| 6 | NPERCO | Nitrogen percolation coefficient. | .bsn |
| 7 | PPERCO | Phosphorus percolation coefficient. | .bsn |
| 8 | PHOSKD | Phosphorus soil partitioning coefficient. | .bsn |
| 9 | SOL_LABP | Initial labile (soluble) P concentration in surface soil layer (kg/ha). | .chm |
| 10 | SOL_ORGN | Initial organic N concentration in surface soil layer (kg/ha). | .chm |
| 11 | SOL_ORGP | Initial organic P concentration in surface soil layer (kg/ha). | .chm |
| 12 | SOL_NO3 | Initial NO3 concentration (mg/kg) in the soil layer . | .chm |
| 13 | ALPHA_BF | Baseflow alpha factor (days). | .gw |
| 14 | GWQMN | Threshold depth of water in the shallow aquifer required for return flow to occur (mm). | .gw |
| 15 | GW_REVAP | Groundwater "revap" coefficient. | .gw |
| 16 | REVAPMN | Threshold depth of water in the shallow aquifer for "revap" to occur (mm). | .gw |
| 17 | ESCO | Soil evaporation compensation factor. | .hru |
| 18 | SLOPE | Average slope steepness (m/m). | .hru |
| 19 | SLSUBBSN | Average slope length (m). | .hru |
| 20 | TLAPS | Temperature laps rate (°C/km). | .sub |
| 21 | CH_COV | Channel cover factor. | .rte |
| 22 | CH_EROD | Channel erodibility factor. | .rte |
| 23 | CH_K2 | Effective hydraulic conductivity in main channel alluvium (mm/hr). | .rte |
| 24 | BIOMIX | Biological mixing efficiency. | .mgt |
| 25 | USLE_P | USLE equation support practice (P) factor. | .mgt |
| 26 | CN2 | SCS runoff curve number for moisture condition II. | .mgt |
| 27 | SOL_AWC | Available water capacity of the soil layer (mm/mm soil). | .sol |

The uncertainty of the results due to uncertainty in the input parameters is quantified by assigning pdfs according to the probability descriptors. Eckhardt *et al* (2003) related normal pdfs to land use change in a hydrologic system. For our interface, it was assumed that the default pdf for modeling the uncertainty of the parameters was the normal distribution, especially if the best estimate of the parameter is the mean of the range. When the best estimate of a parameter doesn't fall close to the mean, then the chosen pdf was a skewed triangular distribution. The pdfs are used for the generation of a set of random numbers with specific statistical properties (such as mean and variance). Its generation is explained as follows.

#### 3.4.4.3.1 Random Number generation

There are several methods for generating random numbers. The most common used are *mechanical methods* like dices, cards, balls, astragalus (six side dices used by ancient greeks and egyptians), *random tables* (tables published by people claiming to have gathered the numbers at random, for example from telephone guides) and *arithmetic algorithms* (used nowadays by computers).

The set of random numbers that a Monte Carlo simulation requires is often produced using arithmetical methods. However, as pointed out by John Von Neumann (1951), "*Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin*". This statement refers to the fact that it is nearly impossible to obtain "truly" random numbers with arithmetical methods, because no algorithm can provide numbers totally unrelated. What can be obtained with arithmetical methods are *pseudo-random numbers*. That is, a set of numbers that has the "appearance" of random numbers, but "showing repeatable patterns". (Computational Science Education Project, 2003)

The random numbers that SPAT generates were produced using the *Marsenne-Twister* algorithm (Matsumoto and Nishimura, 1998). The result of this algorithm, a series of pseudo-random numbers, has better statistical characteristics than other methods previously used (Matsumoto, 1998; Cantu-Paz, 2002)

**3.4.4.3.2 PDF Association with parameters**

The Marsenne-Twister algorithm, written in C++ was inserted in a DLL Visual basic project to be used as a COM object, therefore working together with the Visual Basic code.

The C++ code provided two subroutines:

- A subroutine to create uniform random numbers.
- A subroutine created by Shawn Cockus (2002) for the generation of Gaussian random numbers based on the Box-Mueller method (Press *et al,* 1992) for obtaining Gaussian numbers from a uniform distribution of random numbers. It uses the variance, the mean and the number of random numbers to be generated as parameters.

The triangular distribution is obtained from a series of uniformly distributed random numbers (obtained using the *Marsenne-Twister algorithm*) with the following algorithm (Fig. 3.29):

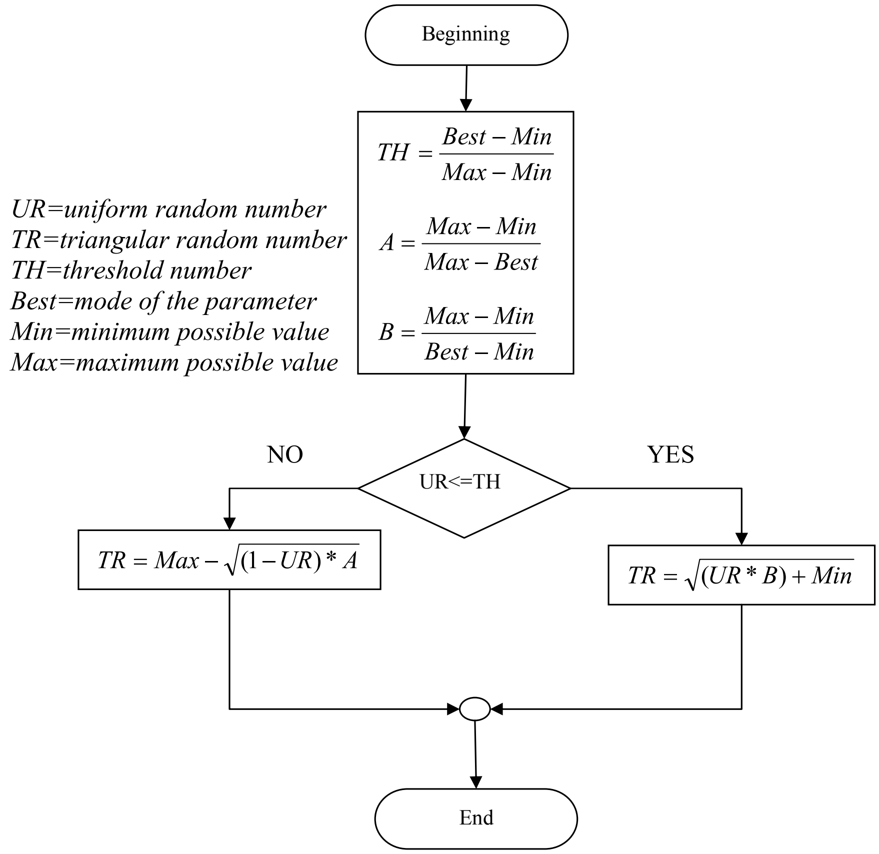



**FIG. 3.25.** Triangular distribution algorithm.

This set of random numbers is used by the Monte Carlo simulation routine. Fig. 3.30 shows the two probability density functions.

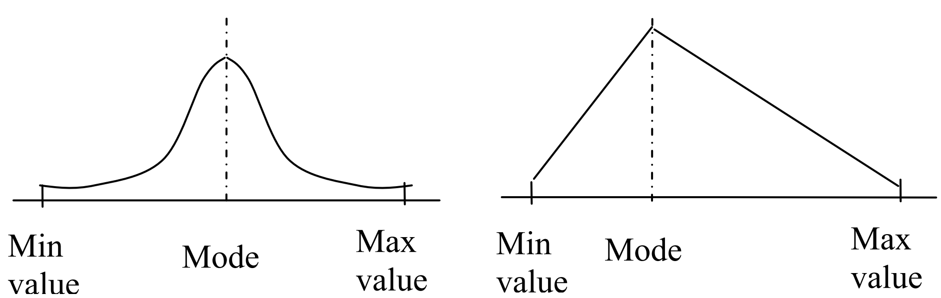



**FIG. 3.26.** Normal and triangular probability density functions of the descriptors.

In this way, a random numbers table is created by SPAT. Each column of this table represents a set of random parameter values that follow their specific pdf (chosen from normal or triangular). This table acts as a database of random numbers. The supported format is DBase.

The case study, discussed in the next chapter, shows an example of this table.

*3.4.4.4 Monte Carlo simulation application*

In this section, SPAT takes generated random parameter values (according to their previously defined pdfs) from the random numbers database (defined on the previous section) and apply them iteratively thousands of times, to the SWAT model as input parameters. A random numbers database is used because each parameter is going to require a different quantity of random numbers, depending on the level of parameterization. For example, if the parameter level is *basin*, SPAT is going to use one random number for each simulation (since SWAT only works with one basin at a time). In the same way, if a parameter level is *subbasin*, and there are 10 subbasins in the model, SPAT uses 10 random numbers per simulation. Similarly, if the parameter level is *HRU*, and there are 50 HRUs, SPAT uses 50 random numbers per simulation. Following the example, if we would be reading these values from a text file table (a textfile that has the columns evenly spaced or tab delimited or comma delimited), we would be enforced to read it line by line; consequently, for a *basin* parameter level, we would be using 1 out of 50 random numbers each simulation.

Using a Dbase table, SPAT has a pointer and a counter on every column, taking the required quantity of random numbers for each parameter level, using less random numbers at a time. This would ensure the statistical properties of the set of random numbers since we are not establishing a trend for picking one random number each fixed number of times.

Once the random numbers are pulled out of the database, SPAT modifies the input text files for the specific SWAT model and collects the representative output in text files.

*3.4.4.5 Output Analysis*

The output the Monte Carlo simulation is a series of text files that compile the relevant information from the SWAT output. Using the collected data, frequency histograms can be generated on a spreadsheet like Excel. Analyzing these frequency histograms, we can deduce several properties of any output variable. Among them: range, mode, distribution and cumulative percentage. If these histograms are normalized, they can show the probability of occurrence of a certain parameter value and, thus, quantify the uncertainty.

# 4 APPLICATION, RESULTS AND DISCUSSION

A case study has been analyzed to illustrate the explained methodology. This will show applications of the ArcGIS SWAT interface and the Geodatabase data model design.

**FIG. 4.1.** Seco Creek (source: www.govart.com).

The Upper Seco Creek (Fig. 4.1) watershed is the case study. Input data is gathered from different sources and processed using the ArcGIS SWAT interface for creating the SWAT input text files. The output hydrologic (stream flow) is compared to historic data and furthermore the hydrologic model is calibrated using the interface. Finally, uncertainty analysis is carried out using the SPAT module of the ArcGIS SWAT interface for the quantification of the uncertainty in the SWAT water quality output, due to the uncertainty of the input data.

## 4.1 DESCRIPTION OF STUDY AREA

The Upper Seco Creek watershed is located in South Central Texas (Fig. 4.2.). It is shared by Bandera, Uvalde and Medina counties. This watershed is part of the upstream region of the Nueces Basin. The downstream part of the watershed is located over the recharge zone of the Edwards Aquifer (figures 4.3. and 4.4.). The city of San Antonio relies, mainly, on the Edwards Aquifer as a source of drinking water. Therefore, hydrologic models prove to be beneficial in estimating the water budget of the aquifer.

The selected study area is located upstream of Seco Creek at Miller Ranch near Utopia city, where a USGS gauging station is in place (USGS 08201500). The drainage area of this point is approximately 116 square kilometers ($km^2$ ). This watershed was selected on the base of availability of data, and previous studies, projects and scientific papers that gave enough information to validate this methodology.

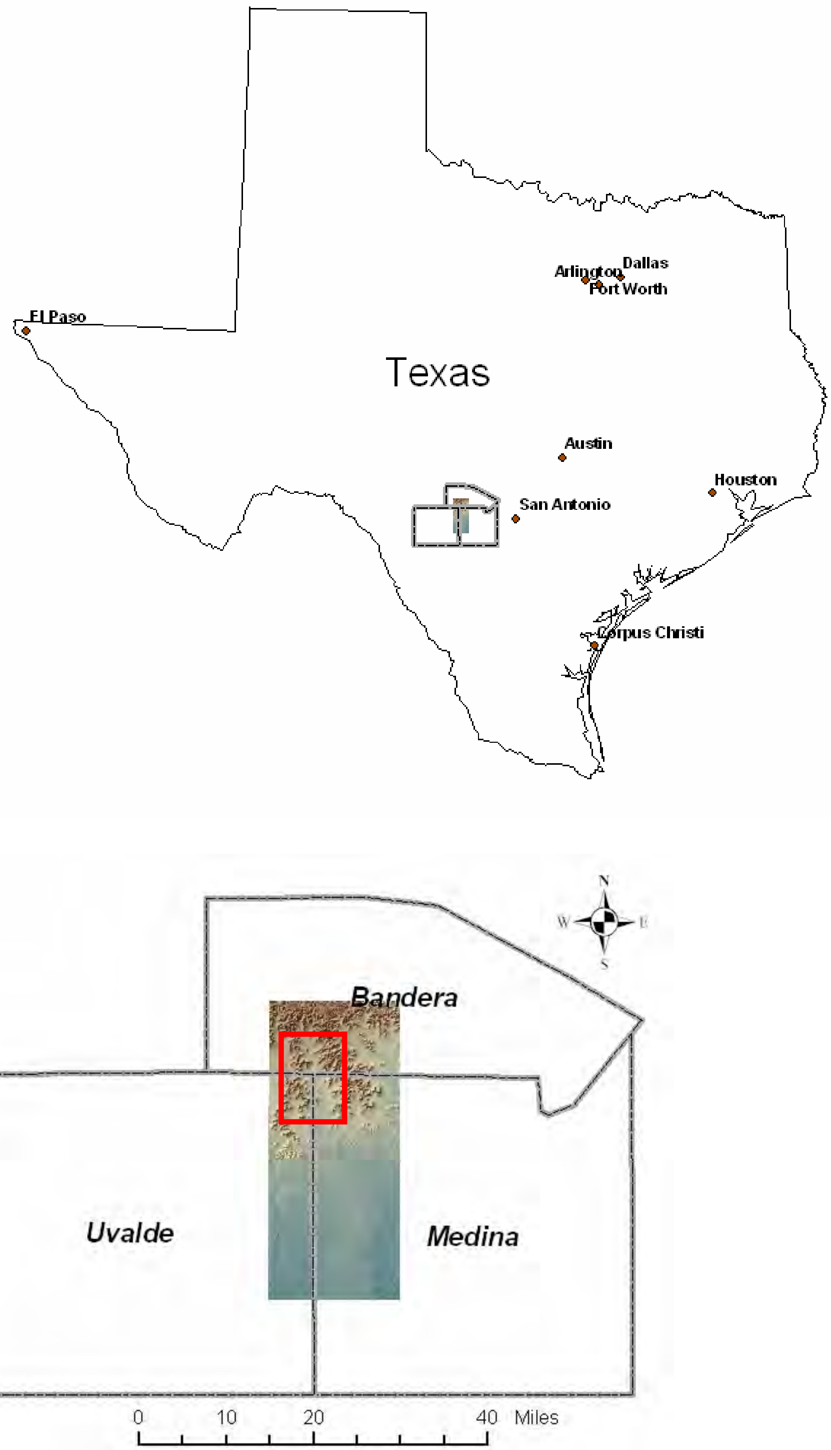



**FIG. 4.2.** Location of Upper Seco Creek study area.

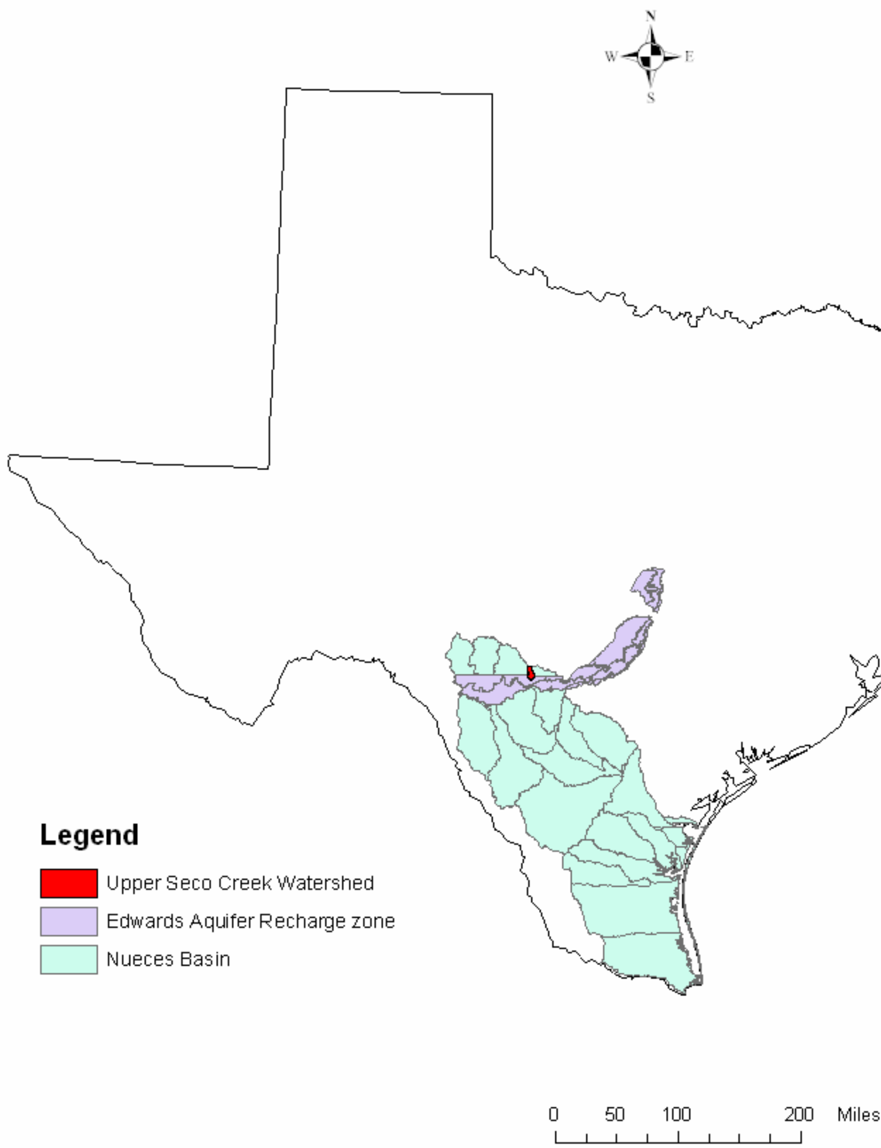



**FIG. 4.3.** Upper Seco Creek, Edwards Aquifer recharge zone and Nueces Basin.

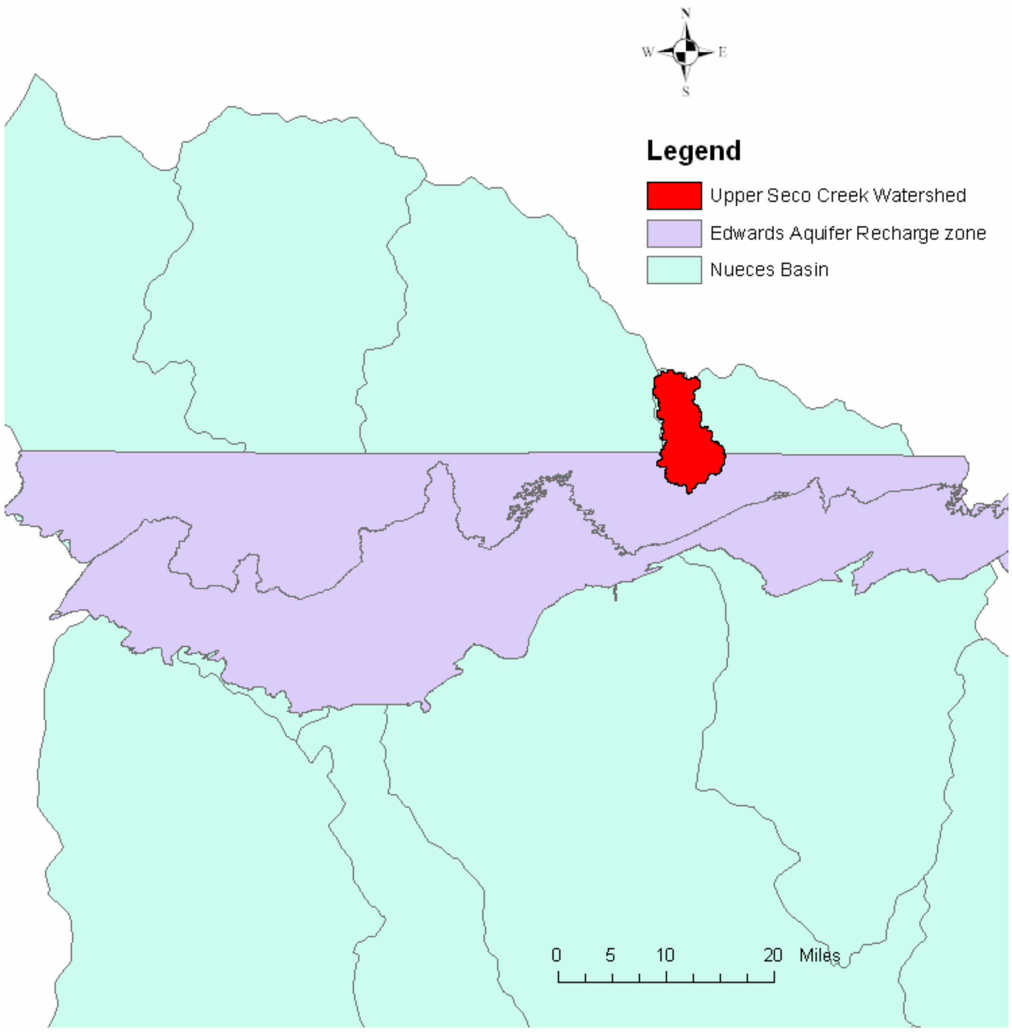



**FIG. 4.4.** Detail of Upper Seco Creek watershed, Edwards aquifer recharge zone and Nueces Basin.

We can highlight four useful sources of information:

1. The *Seco Creek Water Quality Demonstration Project*. This project was developed by the State of Texas and the U.S. Department of Agriculture and involves several state and federal agencies. The project aims to demonstrate the benefits of Best Management Practices (BMPs) to the inhabitants of the region in order to support its use (Brown and Raines, 2002).
2. The *Corpus Christi Bay National Estuary Program*. This program has a project that intends to characterize non point sources of pollution from the Nueces River Basin (Baird *et al* 1996).
3. Srinivasan and Arnold (1994) run the same case study for applying their integration of SWAT with GRASS (Geographic Resources Analysis Support System). This paper acts as a comparison document that helps to evaluate our thesis methodology and its application.
4. USGS hydrologic and environmental records of the selected gage.

Table 4.1. details and summarizes some characteristics of the region.

**TABLE 4.1.** Upper Seco Creek watershed characteristics (Summarized from Brown and Raines, 2002)

| Definition | Value |
|---|---|
| Mean monthly temperture | 68 degrees (°F) |
| Mean monthly temperture range | 51 °F - 84°F |
| Mean annual rainfal at Utopia | 32.6 in. |
| Basin slope range | 8% - 12% |
| Major land use | Rangeland (88%) |
| Uplands and hillslope soil hydrologic classes | B to D |
| Floodplains and terraces soil hydrologic classes | C, D |
| Major geologic element | Glen Rose Limestone |

## 4.2 INPUT DATA

### 4.2.1 Digital Elevation Model

Several 7.5 min DEM tiles were downloaded from a USGS supported website (http://data.geocomm.com/dem/). These DEMs were derived from hypsographic data (contour lines each 20 or 40 ft) and photogrammetric methods. The hypsographic data correspond to topographic maps of a 1:24.000 scale. This scale corresponds to a grid of 30m cell size, but USGS has interpolated the elevation information to produce a finer resolution (although it is not new data) of 10m cell size (USGS, 2003). These DEM tiles have the following geographic characteristics:

- Projection: Universal Transverse Mercator (UTM) zone 14.
- Datum: North American Datum 1983 (NAD 1983)
- Geographic System: North American Geographic Coordinate System 1983
- X_RESOLUTION: 10
- Y_RESOLUTION: 10
- XY_UNITS: Meter
- Z_RESOLUTION: .010
- Z_UNITS: Meter

All the tiles were converted to the ESRI GRID format since they are in *USGS DEM* format. The conversion was made using ArcToolbox 8.2. Then, the grids were merged using the *Spatial Analyst* under ArcGIS. Some no-data cells inside the study area were interpolated using the *Spatial Analyst* using the following expression:

*con(isnull([GRIDNAME]), focalmean([GRIDNAME], rectangle, 5,5), [GRIDNAME])*

This expression interpolates the value of a cell, based on a 5 by 5 cells neighborhood mean.

The resulting DEM is illustrated on Fig. 4.5. The range of elevation is approximately 460m. The upper part of the basin is characterized by high hills while the downstream region shows plains, directly located over the recharge zone of the Edwards aquifer.

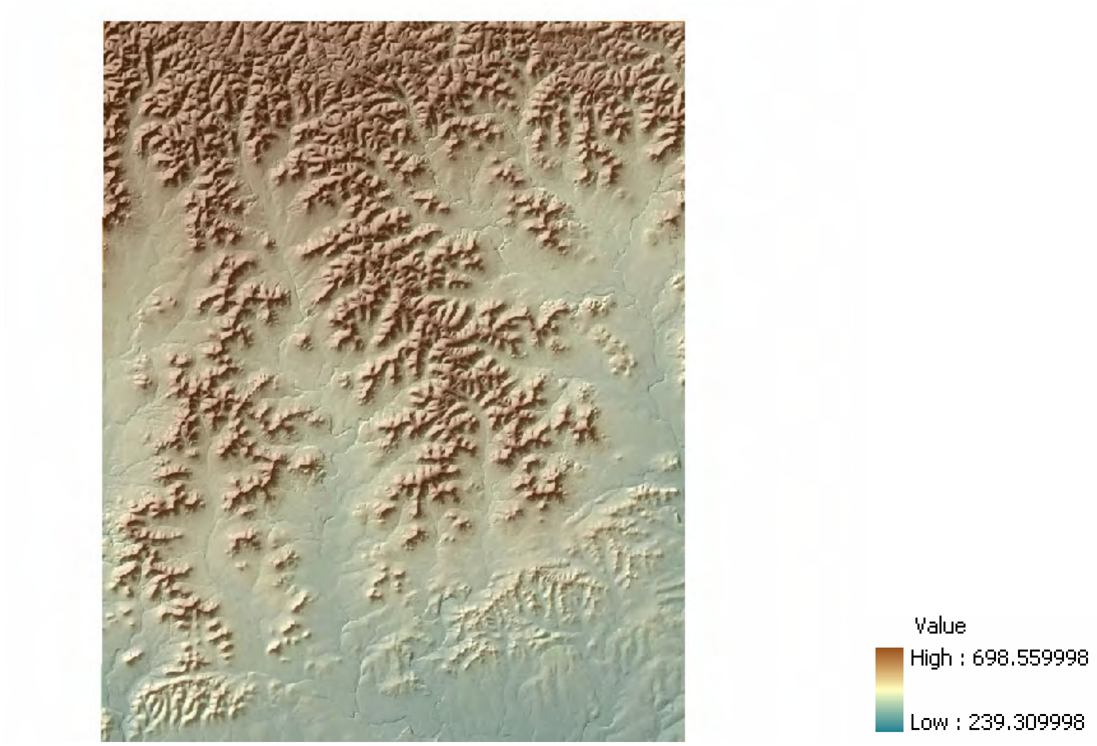



**FIG. 4.5.** DEM of the Seco Creek watershed study area.

### 4.2.2 Land use

The Land use data used for this case study is part of the National Land Cover Data (NLCD) (USGS, 2002). The data can be downloaded using the USGS Seamless Data Distribution System (http://seamless.usgs.gov/viewer.htm) where the user can select the area of interest and download a package containing a dataset in an ESRI GRID format.

NLCD was compiled using Landsat satellite TM imagery. The imagery themes were interpreted using aerial photographs and processed to obtain a final resolution of 30m cell size.

The downloaded data comes in a Geographic Decimal Degrees projection with datum NAD 83. With the purpose of matching the DEM projection, the NLCD data was re-projected using ArcToolbox.

The classification of the data relies on a modified Anderson Level 2 land use and land cover classification system (Anderson *et al* 1976). It contains 21 classes described in Table 4.2.

**TABLE 4.2.** NLCD classes (USGS, 2002)

| Water |
|---|
| 11 Open Water |
| 12 Perennial Ice/Snow |
| **Developed** |
| 21 Low Intensity Residential |
| 22 High Intensity Residential |
| 23 Commercial/Industrial/Transportation |
| **Barren** |
| 31 Bare Rock/Sand/Clay |
| 32 Quarries/Strip Mines/Gravel Pits |
| 33 Transitional |
| **Vegetated; Natural Forested Upland** |
| 41 Deciduous Forest |
| 42 Evergreen Forest |
| 43 Mixed Forest |

**TABLE 4.2. (Continued)**

| Shrubland |
|---|
| 51 Shrubland |

| Non-natural Woody |
|---|
| 61 Orchards/Vineyards/Other |

| Herbaceous Upland |
|---|
| 71 Grasslands/Herbaceous |

| Herbaceous Planted/Cultivated |
|---|
| 81 Pasture/Hay |
| 82 Row Crops |
| 83 Small Grains |
| 84 Fallow |
| 85 Urban/Recreational Grasses |

| Wetlands |
|---|
| 91 Woody Wetlands |
| 92 Emergent Herbaceous Wetlands |

### 4.2.3 Soils

Initially, the Soil Survey Geographic Database (SSURGO) was considered as soil data source, considering the small drainage area of the Upper Seco Creek watershed (less than 50 square miles). SSURGO is a database that has a greater resolution compared to the STATSGO database (1:24000 in SSURGO vs. 1:250000 in STATSGO). SSURGO can be used on the ArcGIS interface as an alternate source for soil data, but it is required to comply with the format that the interface is expecting. A SSURGO extension tool (Peschel, *et al* 2003) was used to convert the original format of SSURGO to the one expected by the ArcGIS SWAT interface. Two simulations of the SWAT model, one using SSURGO and another using STATSGO data were run. Both simulations gave similar results. Therefore, there was no resolution constrain in the soil data for this specific area and it was decided the use of the STATSGO dabatase.

As explained in the methodology, the modified version of the STATSGO database is stored in a Feature dataset in the Static geodatabase. It has an accuracy of a 1:250.000 scale. The original projection has the following characteristics:

- Map_Projection_Name: Albers Conical Equal Area
- Horizontal_Datum_Name: North American Datum 1983
- Ellipsoid_Name: Geographic Reference System 80
- Units: meters

It was used the ArcGIS capability *projecting on the fly*. This capability automatically reprojects feature data to the current data frame projection (the projection in which the user is working). In this case study, the Texas modified STATSGO dataset was automatically reprojected to the DEM projection, without modifying the original source data.

### 4.2.4 Precipitation stations

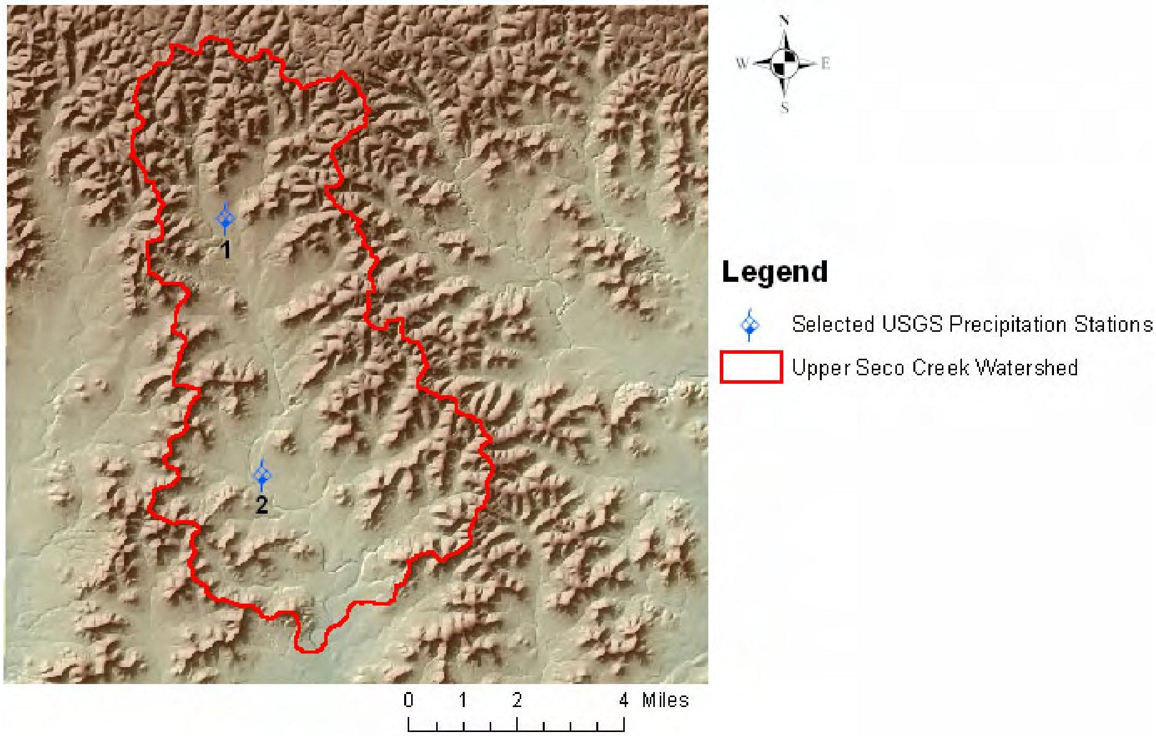



**FIG. 4.6.** Selected USGS precipitation stations.

Fig. 4.6 shows the location of the selected precipitation stations. These precipitation stations are part of the *Seco Creek water quality demonstration project*, where several rainfall gage stations were installed. Table 4.3. details the location and period of record of each station.

**TABLE 4.3.** Daily rainfall stations in the Upper Seco Creek Basin

| Site # | Station # | Station Name | Latitude | Longitude | Period of record (water years) |
|---|---|---|---|---|---|
| 1 | 2941250992554 | Seco Creek rain gage #1 | 29° 41' 25" | 99° 25' 54" | 1991-1998 |
| 2 | 2937170992513 | Seco Creek rain gage #2 | 29° 37' 17" | 99° 25' 13" | 1991-1998 |

### 4.2.5 Flow gauging station

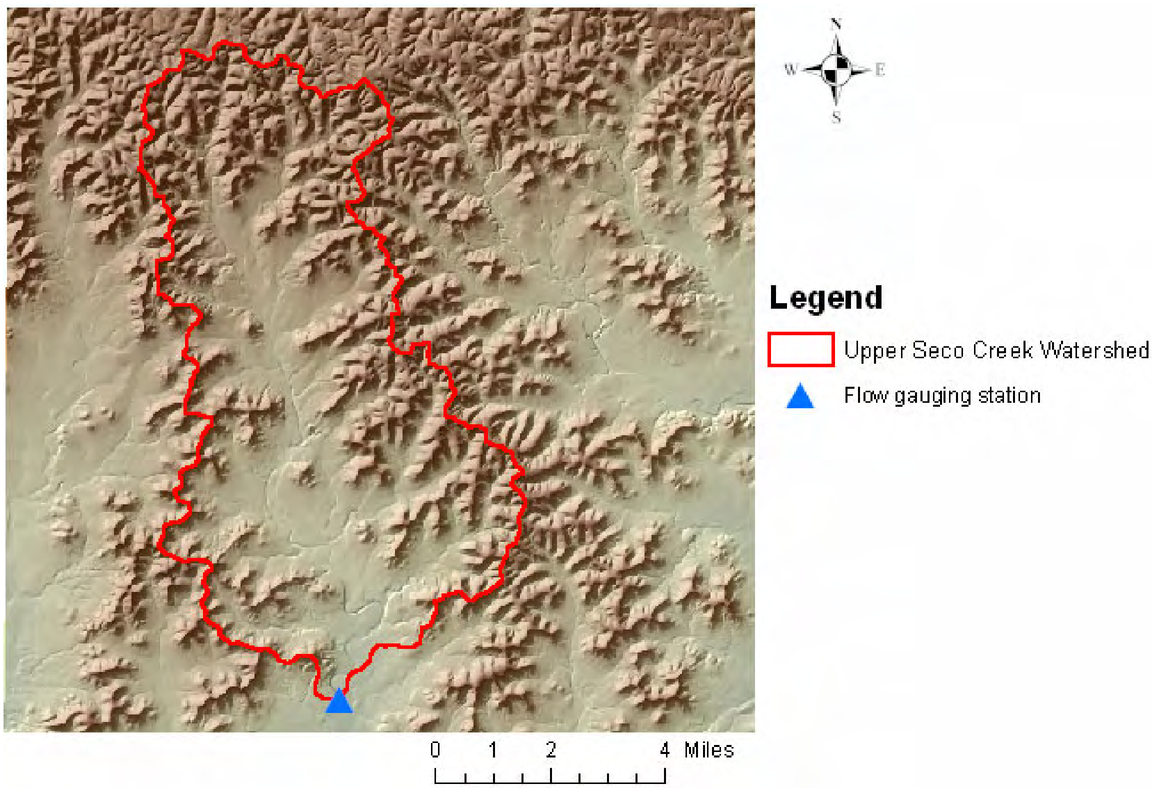



**FIG. 4.7.** Flow gauging station at Miller Ranch.

The selected gauging station, USGS # 08201500 is a real time/daily/monthly streamflow gage, located at Miller Ranch, near the city of Utopia, Medina County. Its latitude is 29° 41’ 25” and its longitude is 99°25’ 54”. The gage datum is 1,265.80 feet above sea level. Its period of record is 1961-05-01 to 2001-09-30. Its location served as a point for setting up the main outlet of the delineated basin (Fig. 4.7).

### 4.2.6 Curve Number

An ESRI GRID, containing curve number values for moisture condition II, has been used on this case study on the HRU management files (*.mgt). The SWAT model requires the curve number in order to apply the SCS curve number method for calculating runoff volume.

The grid was developed by the Blacklands Research Center in Temple, Texas. It is a 250m cell size grid that covers the conterminous US. It has the following characteristics:

- Projection: Albers Equal Area
- Spheroid: Clarke 1866
- Datum: Clarke 1866

The grid was reprojected using ArcToolbox, to match the DEM projection. Fig. 4.8 shows the curve number grid and the case study area.

It can be seen that our case study area is dominated by low curve numbers. Therefore it is expected high infiltration rates on the region.

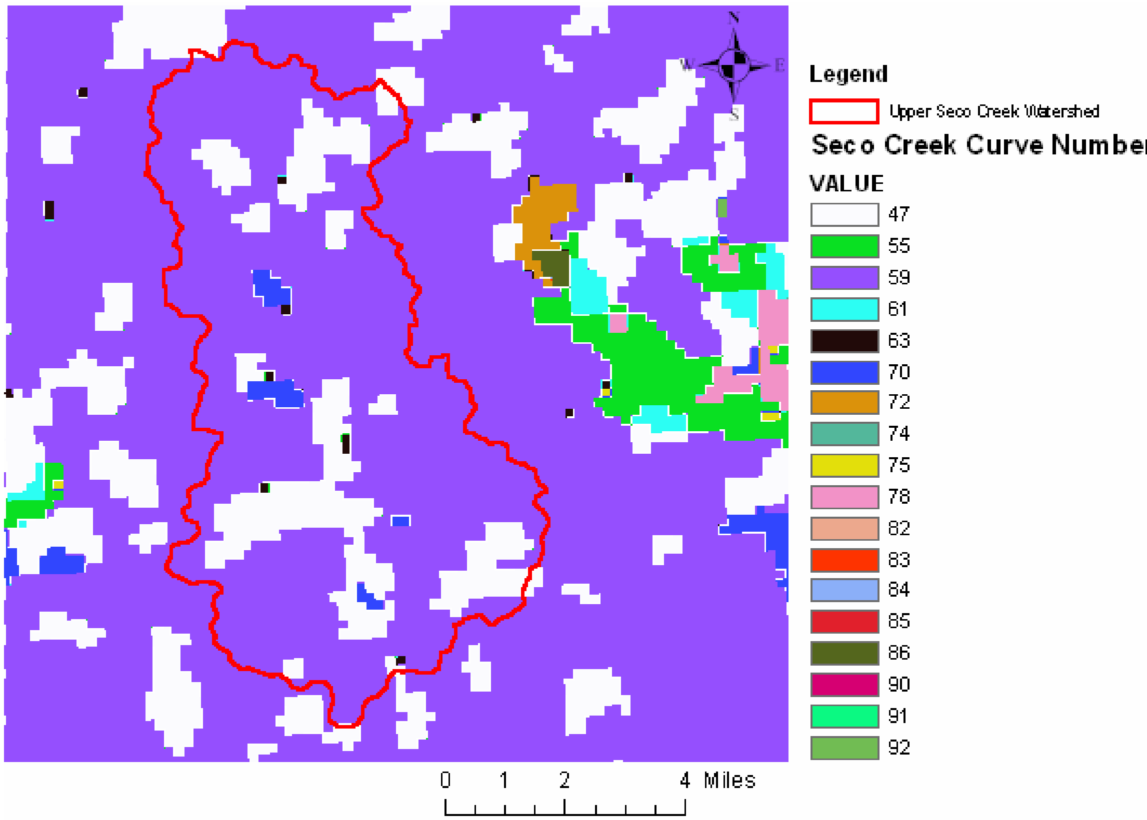



**FIG. 4.8.** Curve number grid for moisture condition II.

## 4.3 ARCGIS SWAT INTERFACE APPLICATION

### 4.3.1 ArcGIS SWAT extension toolbar

Once the extension has been activated, the user has access to a customized toolbar (Fig. 4.9) containing the menus that will aid in pre and post processing input and output data of the SWAT model.


**FIG. 4.9.** ArcGIS SWAT extension toolbar.

The menus are explained as follow:

- *Swat definition*: It helps the user in defining Dynamic and Static geodatabases. It contains GIS tools for extracting hydrologic information from the geographic input data.
- *Write Input Tables*: it organizes extracted and default parameters into personal geodatabase tables. From each table a set of input text files is created.
- *Edit Input*: a series of editors that aid the user in modifying the calculated/default parameters from either static or Dynamic geodatabases.
- *Simulation*: Defines the configuration options of the SWAT model and executes it. After this step uncertainty analysis can be carried out.

### 4.3.2 Project Setup

The project Setup can be accessed from the SWAT definition menu (Fig. 4.10). It helps in defining the name and location of the Dynamic geodatabase (*Personal Geodatabase Name* and *Work directory* on Fig. 4.10). Then, it creates an empty dynamic geodabatase and the folder structure that will hold the grids during the process.

Additionally, the project Setup window queries the user for the location of the Static geodatabase (*Swat Geodatabase* on Fig. 4.10).

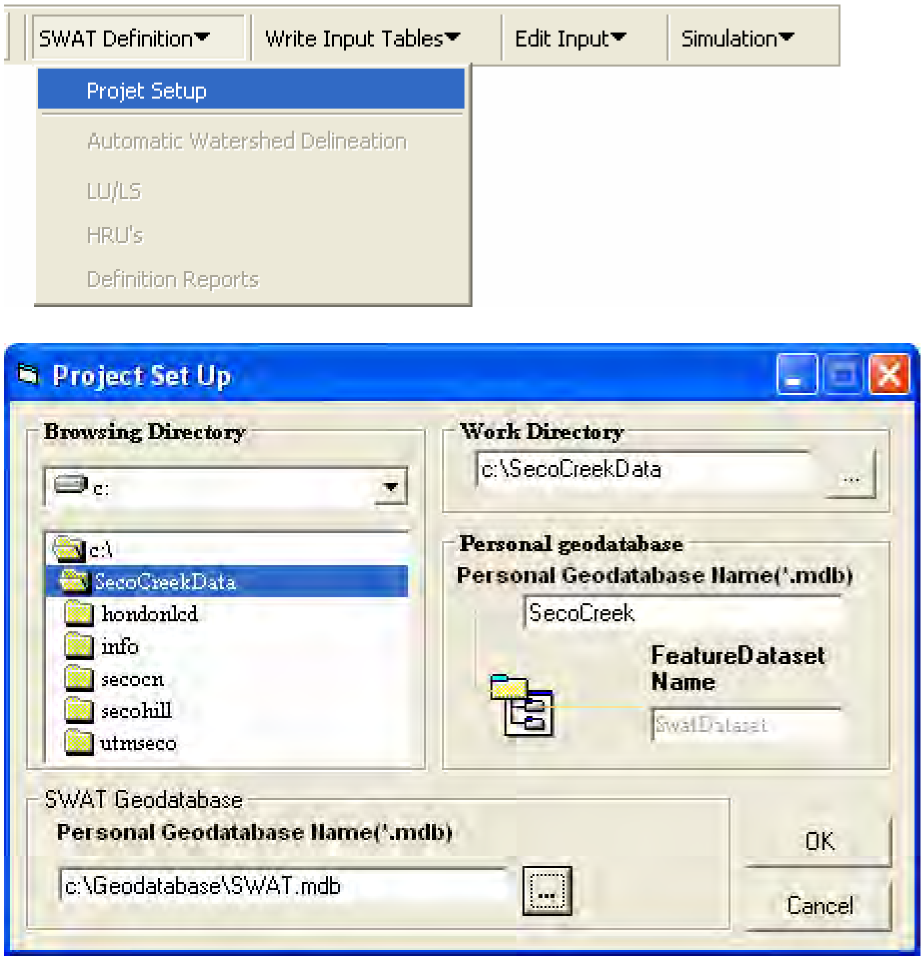



**FIG. 4.10.** Project setup.

### 4.3.3 Watershed Delineator.

Following the methodology of chapter 3, the Watershed Delineator (Fig. 4.11) helps the user in defining basin, subbasins and streams datasets.

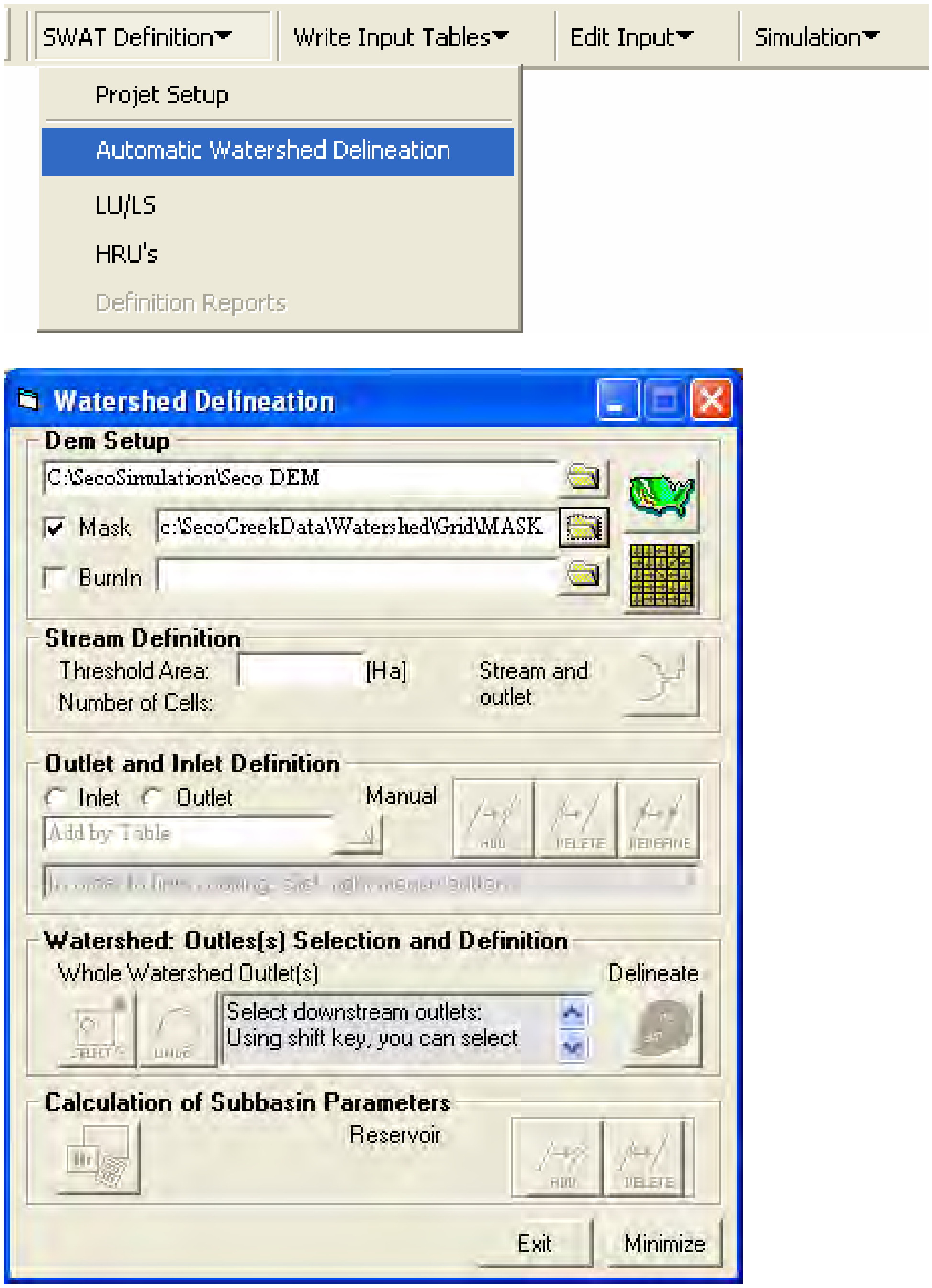



**FIG. 4.11.** Watershed Delineator.

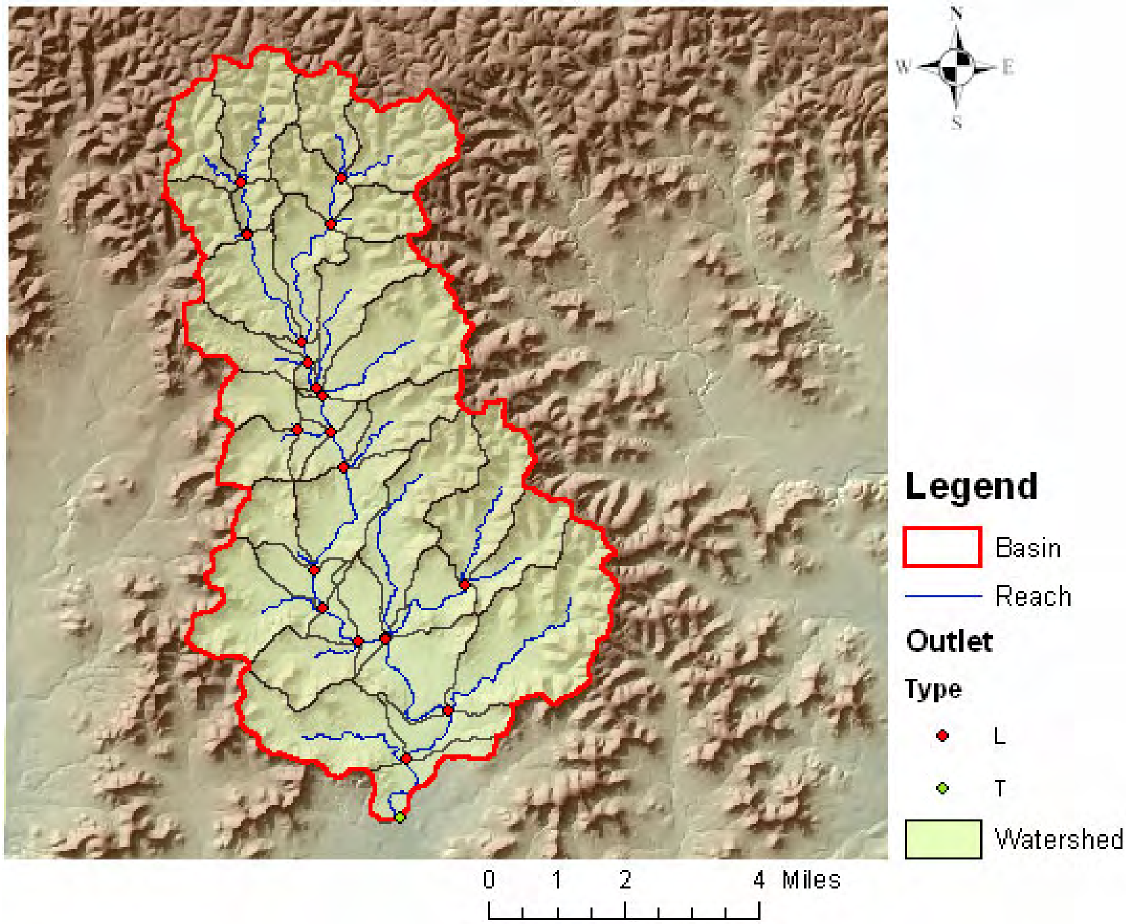



**FIG. 4.12.** Upper Seco Creek stream and watershed delineation.

In the Seco Creek case study, the selected stream threshold is 200 cells that represent 2 Ha. of minimum drainage area. 39 subbasins and streams were delineated based on this threshold.

Fig. 4.12 shows the delineation of the Upper Seco Creek region. It can be recognized that only one outlet has been defined for the entire basin (Fig. 4.12 green point) and there is one reach and one outlet per subbasin. These are the basic relationships in which the SWAT model relies.

Fig. 4.13 shows a screen capture in which 4 spatial relationships can be illustrated. The outlet with HydroID *100039* is related to the *watershed, reach* and *monitoring point* datasets that contain an *OutletID* field with the same number. Thus, spatial relationships have been established on the Seco Creek watershed, and now any parameter value can be traced to the outlets.

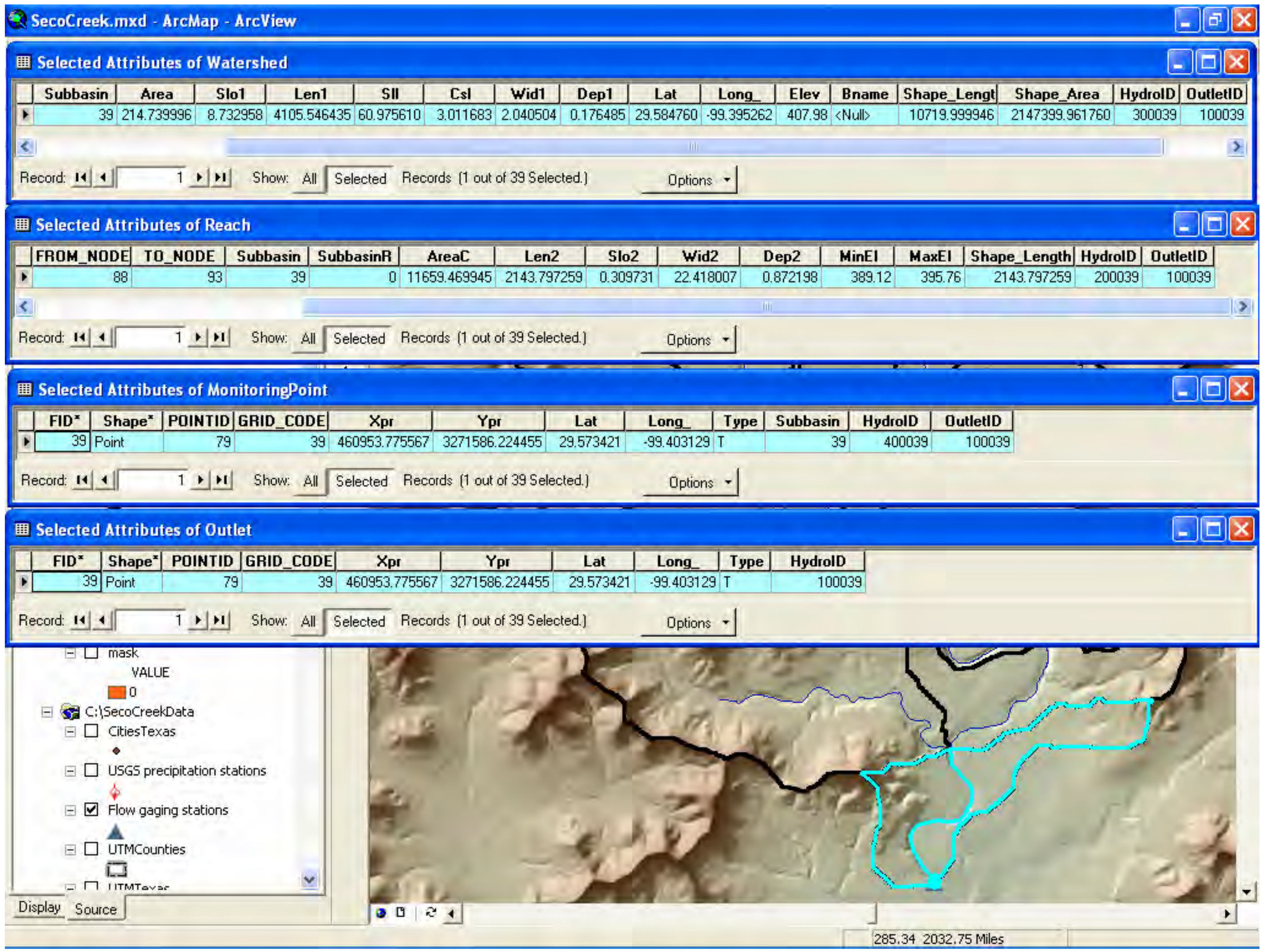



**FIG. 4.13.** Watershed delineation relationships.

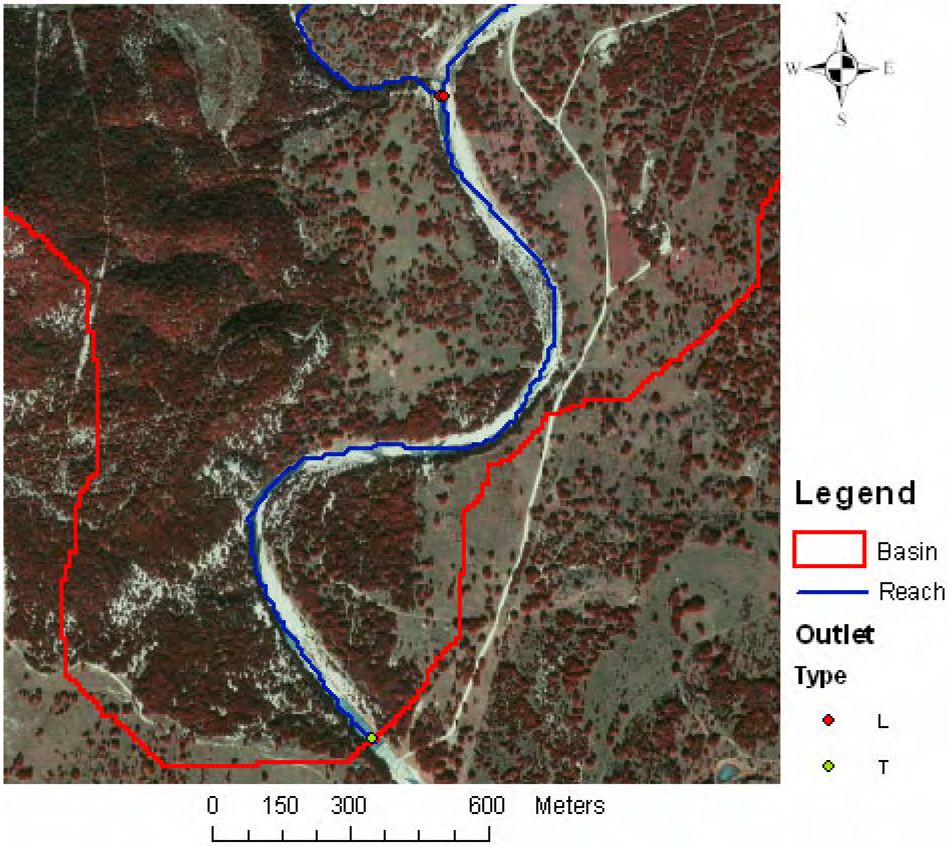



**FIG. 4.14.** Delineation compared to a Digital Orthophoto Quadrant.

Goodness of the delineation can be evidenced when comparing it to a Digital Orthophoto Quadrant (DOQ) of 1m resolution, like in Fig. 4.14. This figure shows the main outlet of the basin.

### 4.3.4 Land use and soils definition

The interface (Fig. 4.15) executes the *clip* operation in which the land use dataset is cut using the basin boundary (Fig. 4.16). A reclassification is made over this dataset in order to assign the SWAT land use codes.

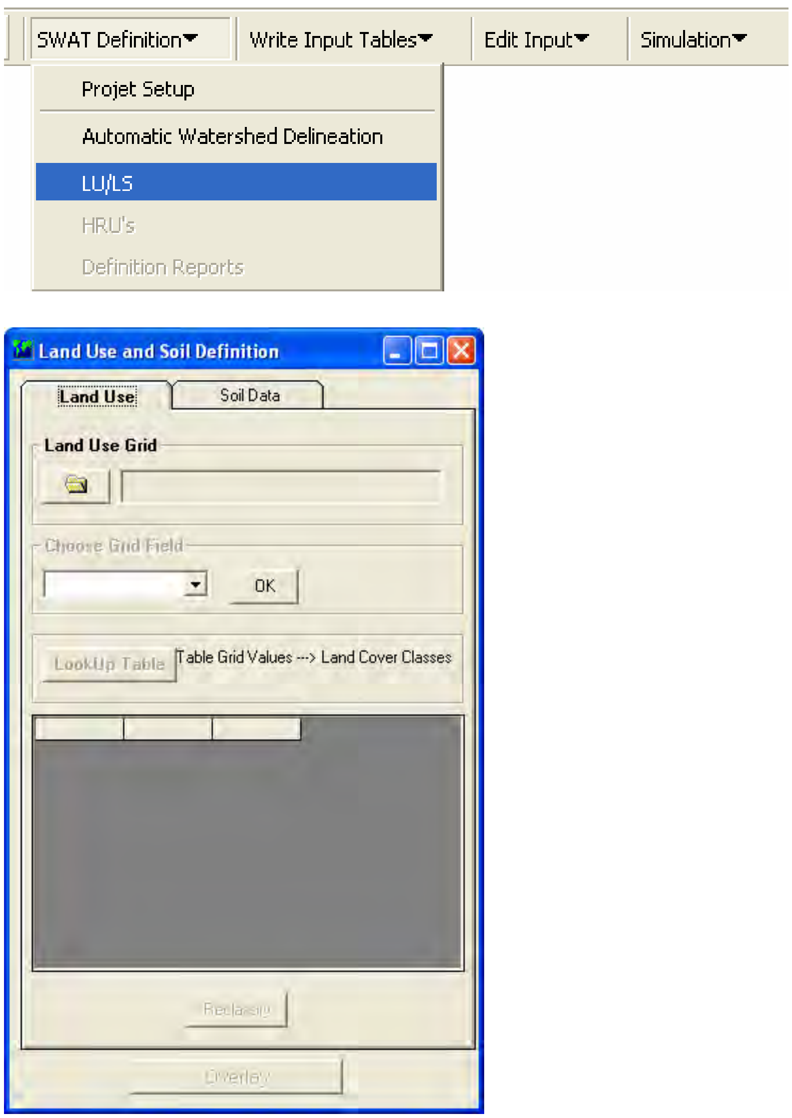



**FIG. 4.15.** Land use and soils definition interface.

Fig. 4.17 shows a detail of the reclassification section in the interface. The column *value* represents the NLCD class codes and the *LandUseSwat* column shows the corresponding SWAT class codes.

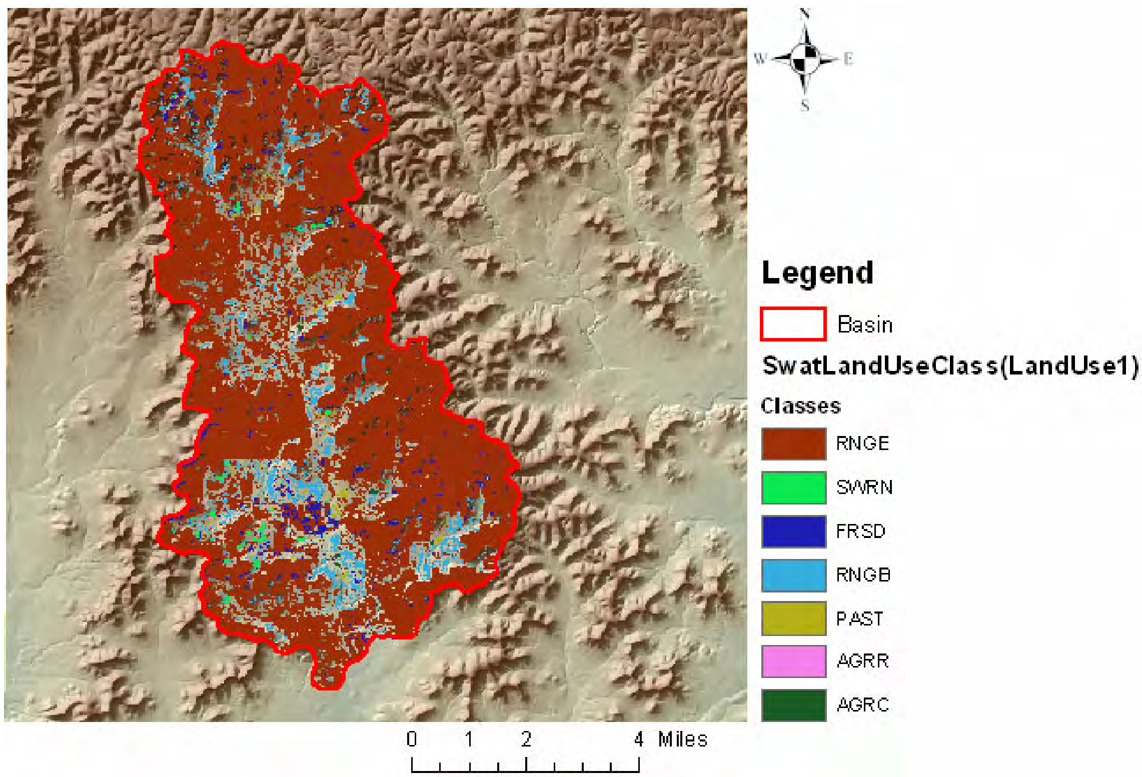
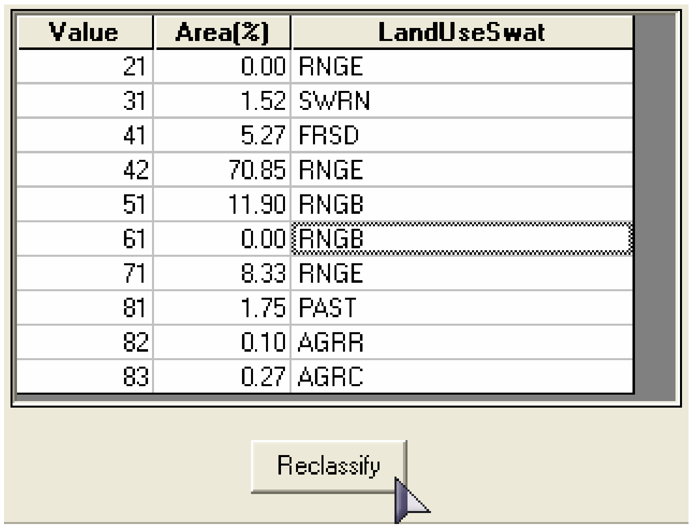



**FIG. 4.16.** Land use map for Upper Seco Creek watershed.

Not every NLCD class code has a corresponding code on the SWAT side. Several NLCD class codes can be represented by one SWAT class code. Moreover, the user has the capability of defining his/her correspondences between codes. This acts as a filter of non representative land use classes. For example, the NLCD code 21 stands for *low intensity residential* and has less than 0.00% (0.0007%) of the watershed area and we chose to add that percentage to the major land use class (range land), by assigning it to the RNGE code.



| Value | Area(%) | LandUseSwat |
|---|---|---|
| 21 | 0.00 | RNGE |
| 31 | 1.52 | SWRN |
| 41 | 5.27 | FRSD |
| 42 | 70.85 | RNGE |
| 51 | 11.90 | RNGB |
| 61 | 0.00 | RNGB |
| 71 | 8.33 | RNGE |
| 81 | 1.75 | PAST |
| 82 | 0.10 | AGRR |
| 83 | 0.27 | AGRC |



**FIG. 4.17.** Land use reclassification detail.

It can be noticed from figures 4.16. and 4.17., that more than 90% of the watershed is covered by rangeland and range brush. Agricultural crops have a minimum percentage of the watershed area.

Fig. 4.18 shows the two STATSGO map units that cover the area of the Upper Seco Creek watershed. 80.65% of the area is covered by MUID TX155 and the rest is covered by MUID TX525 (Fig. 4.19). For this case study, the interface takes the dominant soil in the map unit. An *Eckrant* soil is dominant in MUID TX155 (hydrologic group D) while a *Speck* soil is dominant in MUID TX525 (hydrologic group D). In *Eckrant* soil the clay content is dominant (>40%) while in the *Speck* soil the silt and sand are dominant (>30% each one). MUID TX525 is located mainly under the main streams (where it is expected more infiltration due the sand content) while MUID TX155 is located overland. Thus, it is anticipated a higher runoff on the overland areas than the main channel area.

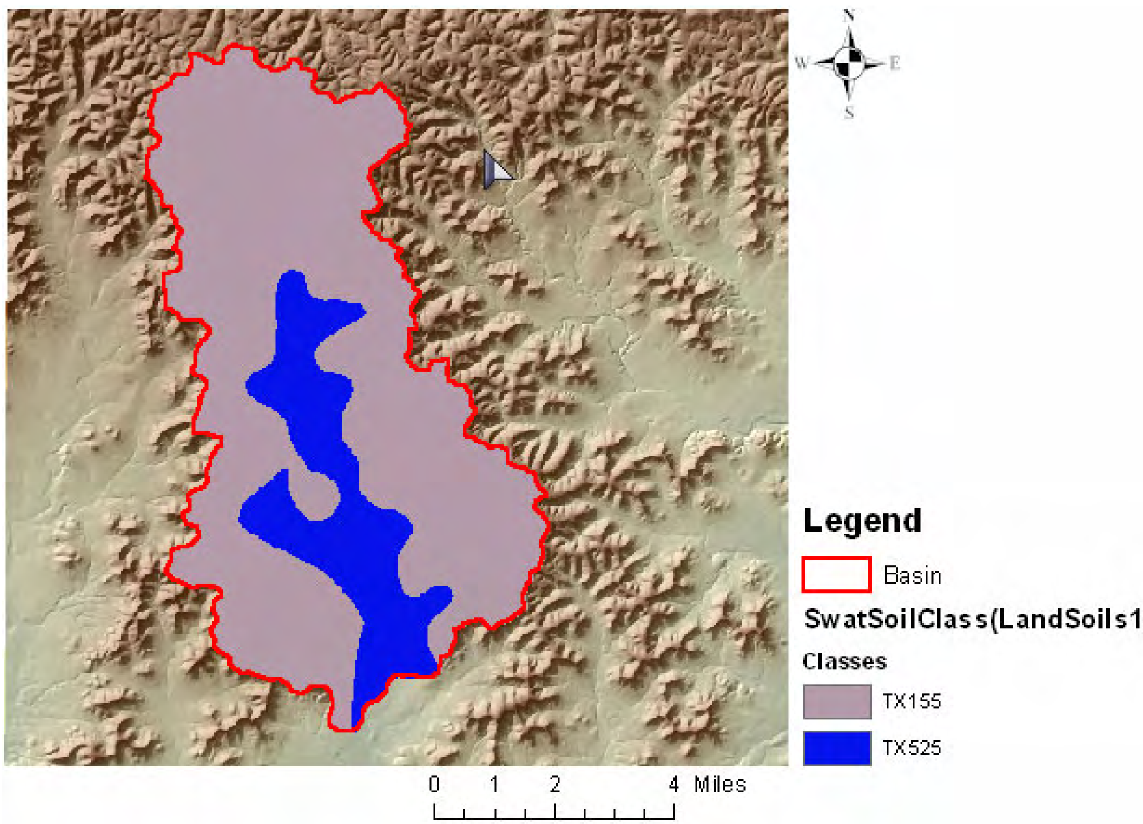
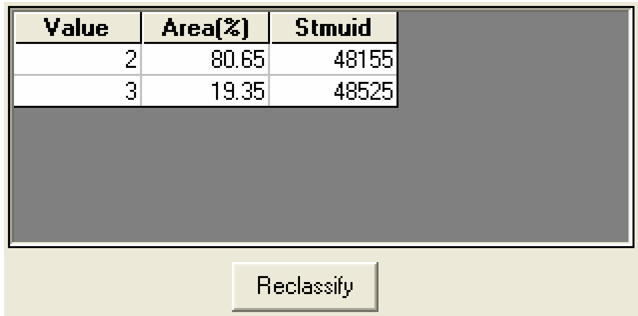



**FIG. 4.18.** Soils map for Upper Seco Creek watershed.



| Value | Area(%) | Stmuid |
|---|---|---|
| 2 | 80.65 | 48155 |
| 3 | 19.35 | 48525 |



**FIG. 4.19.** Soils reclassification detail.

### 4.3.5 Hydrologic Response Units definition and filtering

Using the interface, an *overlay* operation was carried out. This operation has as output, georeferenced HRU's (HRUs that can be located geographically) within the boundaries of the Upper Seco Creek Basin (Fig. 4.20).

Fig. 4.20 shows the downstream portion of the Upper Seco Creek watershed. The highlighted features are a unique combination of the land use RNGB (range brush) with the TX525 soil, within the subbasin #39 (most downstream subbasin).

Through the map it can be observed that this unique combination of land use and soil happens to exist along the main channel and its floodplain. With georeferenced HRUs there is the possibility of a better analysis of floodplain buffers and, thus, flood prevention.

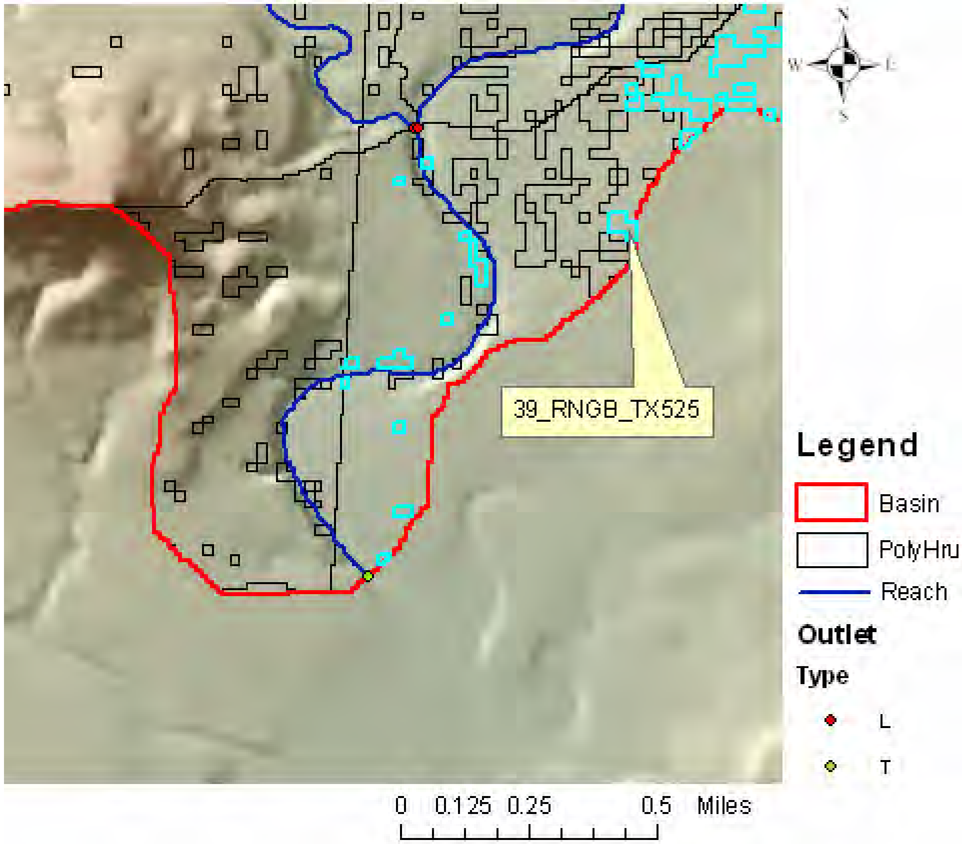
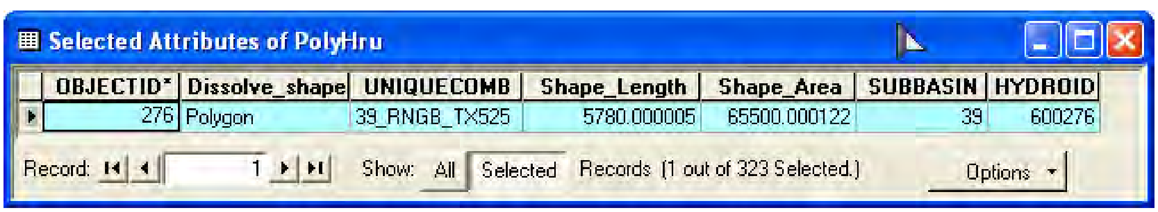



**FIG. 4.20.** Detail of georeferenced HRUs.



Selected Attributes of PolyHru

| OBJECTID* | Dissolve_shape | UNIQUECOMB | Shape_Length | Shape_Area | SUBBASIN | HYDROID |
|---|---|---|---|---|---|---|
| 276 | Polygon | 39_RNGB_TX525 | 5780.000005 | 65500.000122 | 39 | 600276 |

Record: 1 Show: All Selected Records (1 out of 323 Selected.) Options

**FIG. 4.21.** Detail of selected HRU in the PolyHRU feature class.

Fig. 4.21. shows how an specific HRU can be identified in the feature class attribute table. Basic information like the combination codes (*uniquecomb*), shape, area and subbasin are identified. Also, the HydroID field serves as a key to link this

information with several tables of parameter values (input and output of SWAT) and Time Series.

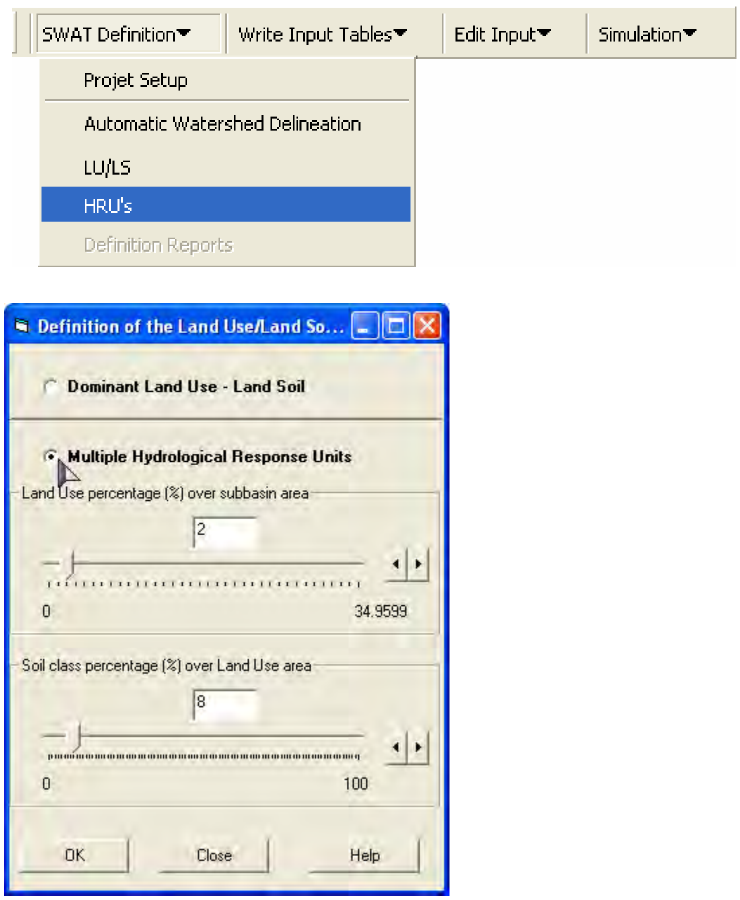



**FIG. 4.22.** HRUs filter interface.

There is also the possibility of filtering the small polygons of non-representative HRUs. For the Upper Seco Creek case study, the HRU's with a land use that has a percentage smaller than the 2% of the subbasin area and a soil class with a percentage smaller than 8% over the land use class, are filtered (Fig. 4.22). 118 non-representative HRUs were filtered out (using this technique) from 323 original HRUs. They represented 0.9 square miles. This area has been shared among the rest of HRU's, normalizing their area. The shapes of the filtered HRU's are not being deleted (they just don't participate in the SWAT simulation) and the shapes of the remaining HRU's are not modified (just the area field).

### 4.3.6 Weather definition.

In this module, several weather variables are defined. The precipitation station data must be given in a specific format. The interface takes these data and assigns them to the Time Series Data table and also to the parameter values table, from which several text files are created.

For this case study, precipitation data from the two USGS stations was provided in the specific format. The rest of the meteorological information was chosen to be simulated by SWAT based on a weather station database that has more than 20 years of record.

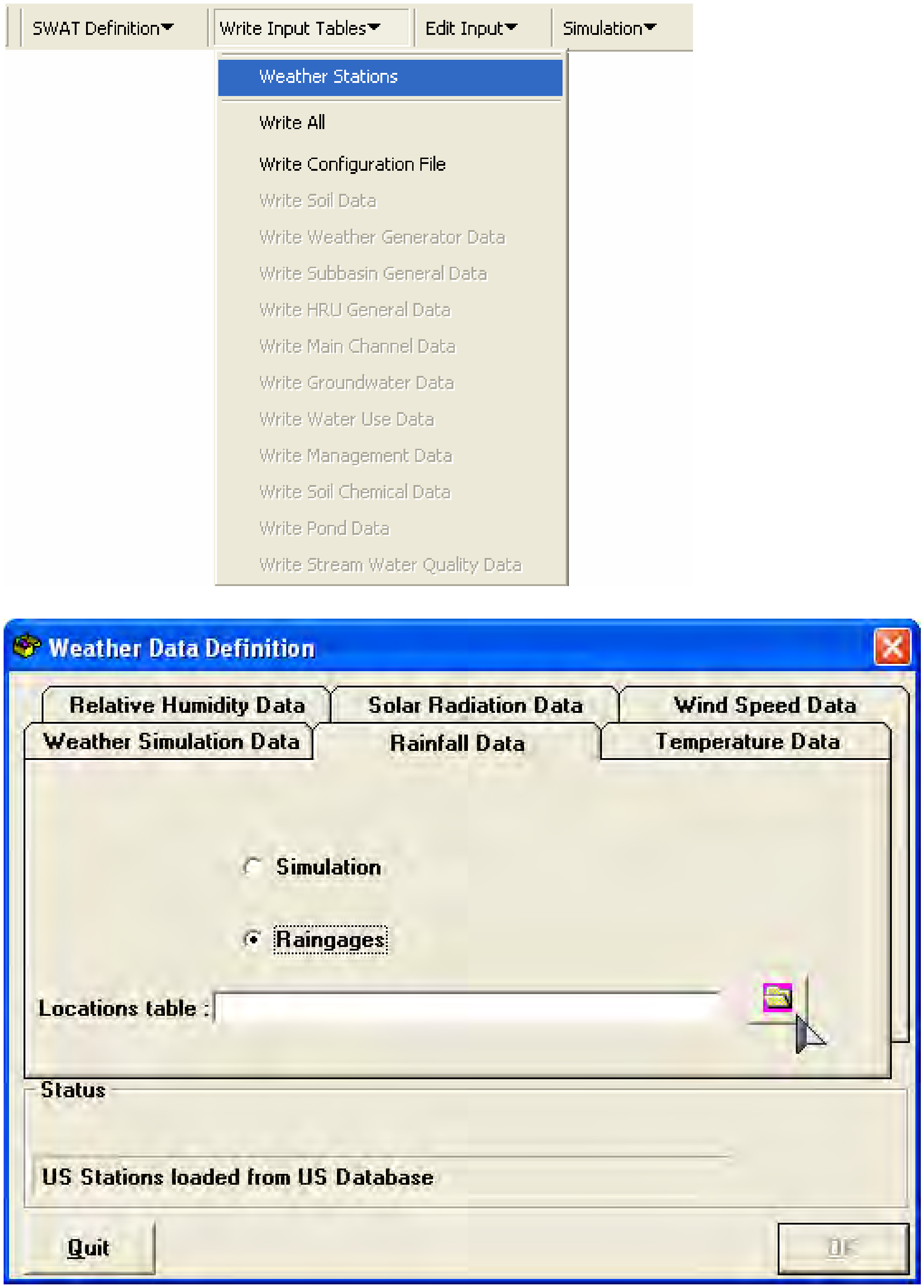
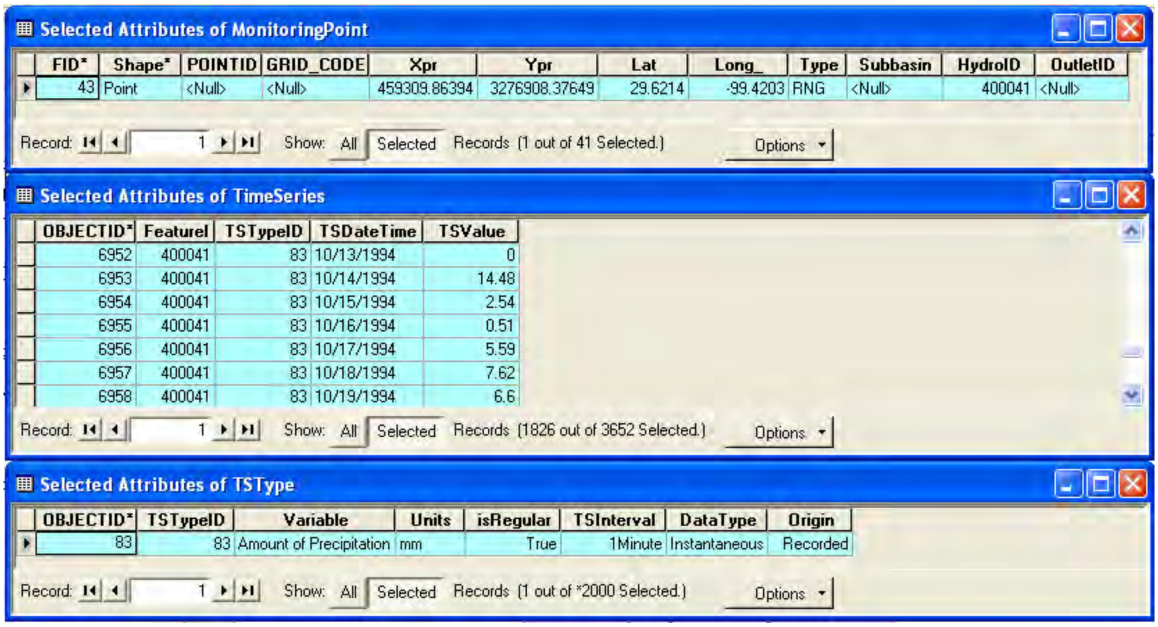



**FIG. 4.23.** Weather definition interface.




| FID* | Shape* | POINTID | GRID_CODE | Xpr | Ypr | Lat | Long_ | Type | Subbasin | HydroID | OutletID |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 43 | Point | <Null> | <Null> | 459309.86394 | 3276908.37649 | 29.6214 | -99.4203 | RNG | <Null> | 400041 | <Null> |

| OBJECTID* | FeatureI | TSTypeID | TSDateTime | TSValue |
|---|---|---|---|---|
| 6952 | 400041 | 83 | 10/13/1994 | 0 |
| 6953 | 400041 | 83 | 10/14/1994 | 14.48 |
| 6954 | 400041 | 83 | 10/15/1994 | 2.54 |
| 6955 | 400041 | 83 | 10/16/1994 | 0.51 |
| 6956 | 400041 | 83 | 10/17/1994 | 5.59 |
| 6957 | 400041 | 83 | 10/18/1994 | 7.62 |
| 6958 | 400041 | 83 | 10/19/1994 | 6.6 |

| OBJECTID* | TSTypeID | Variable | Units | isRegular | TSInterval | DataType | Origin |
|---|---|---|---|---|---|---|---|
| 83 | 83 | Amount of Precipitation | mm | True | 1Minute | Instantaneous | Recorded |

**FIG. 4.24.** Time Series tables of precipitation stations.

Using the Time Series relationships, it is possible to track, query, analyze and create Time Series graphs without leaving the ArcGIS environment.

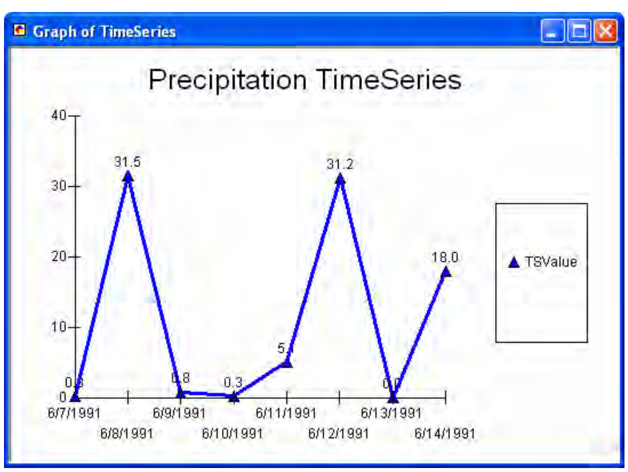



**FIG. 4.25.** Precipitation time series graph built under ArcGIS.

Fig. 4.25 shows a screen capture of a Time Series graph build under ArcGIS from selected Time Series data. It illustrates the precipitation from 6/7/1991 to 6/14/1991 and it was built with the help of Time Series queries.

### 4.3.7 Parameter tables and textfiles

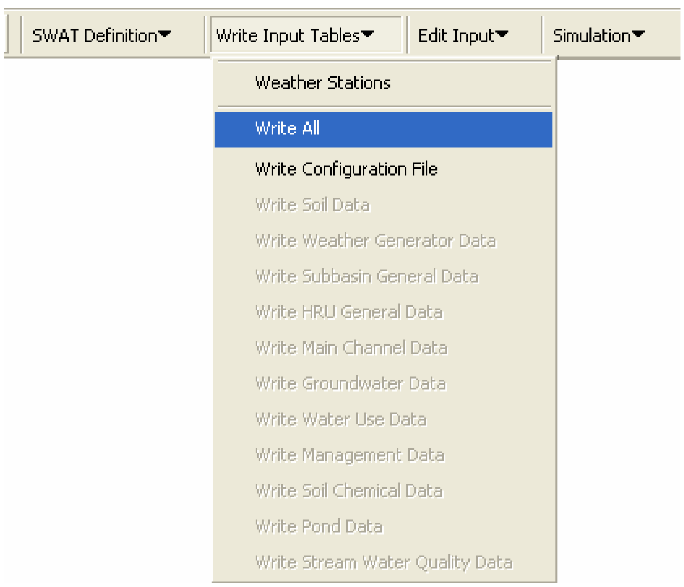



**FIG. 4.26.** Write all function.

The *write all* function (Fig. 4.26) created and populated all the tables detailed on Fig. 3.18 for the case study,  with calculated and default values.

At this point, the SWAT model can be configured and run, obtaining a non calibrated model.

The *Edit Input* menu can be used for calibration or parameter modification (Fig. 4.27). The editing tools contained in this menu allow the user to modify the parameter values of an individual hydrologic element (i.e. subbasin, reach or HRU), and extend the modification to a set of similar hydrologic elements (refer to Fig. 4.28; the similarity can be based on subbasin, land use or soil depending on its parameterization level).

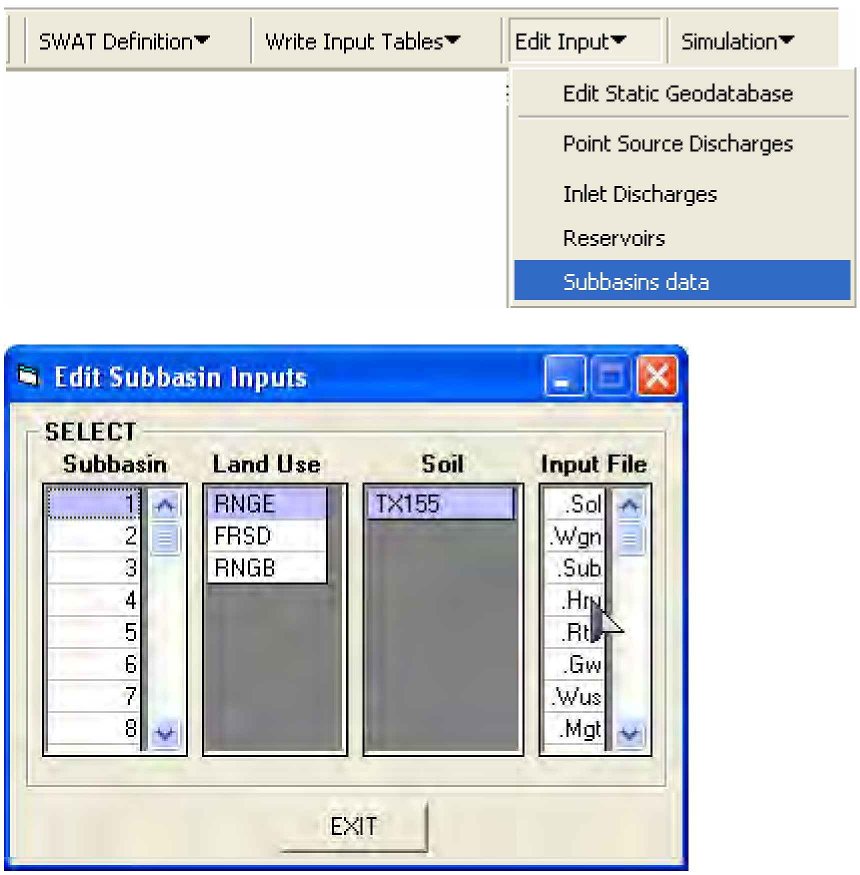



**FIG. 4.27.** Parameter values editor.

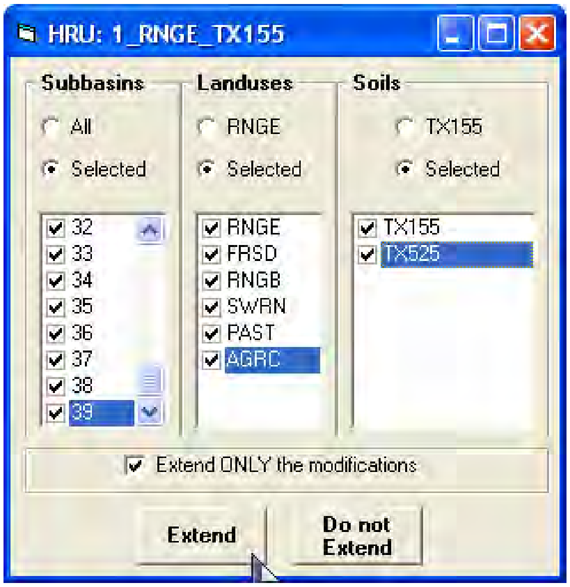



**FIG. 4.28.** Extend modifications tool.

From each personal geodatabase table, the interface creates automatically all the required input text files. For the Upper Seco Creek case study, 1275 required text files were generated (1 basin, 39 subbasins, 205 HRUs and configuration files).

### 4.3.8 SWAT simulation

As Fig. 4.29 indicates, the simulation period goes from July 1$^{st}$ of 1988 to September the 30$^{th}$ 1995. However, we are going to take into consideration a period that coincides with our observed rainfall records. Rainfall records start on January 1991 and finish on September 25$^{th}$ of 1995. We started an earlier simulation at 1988 expecting that the SWAT model stabilizes its initial estimations of water in the system.

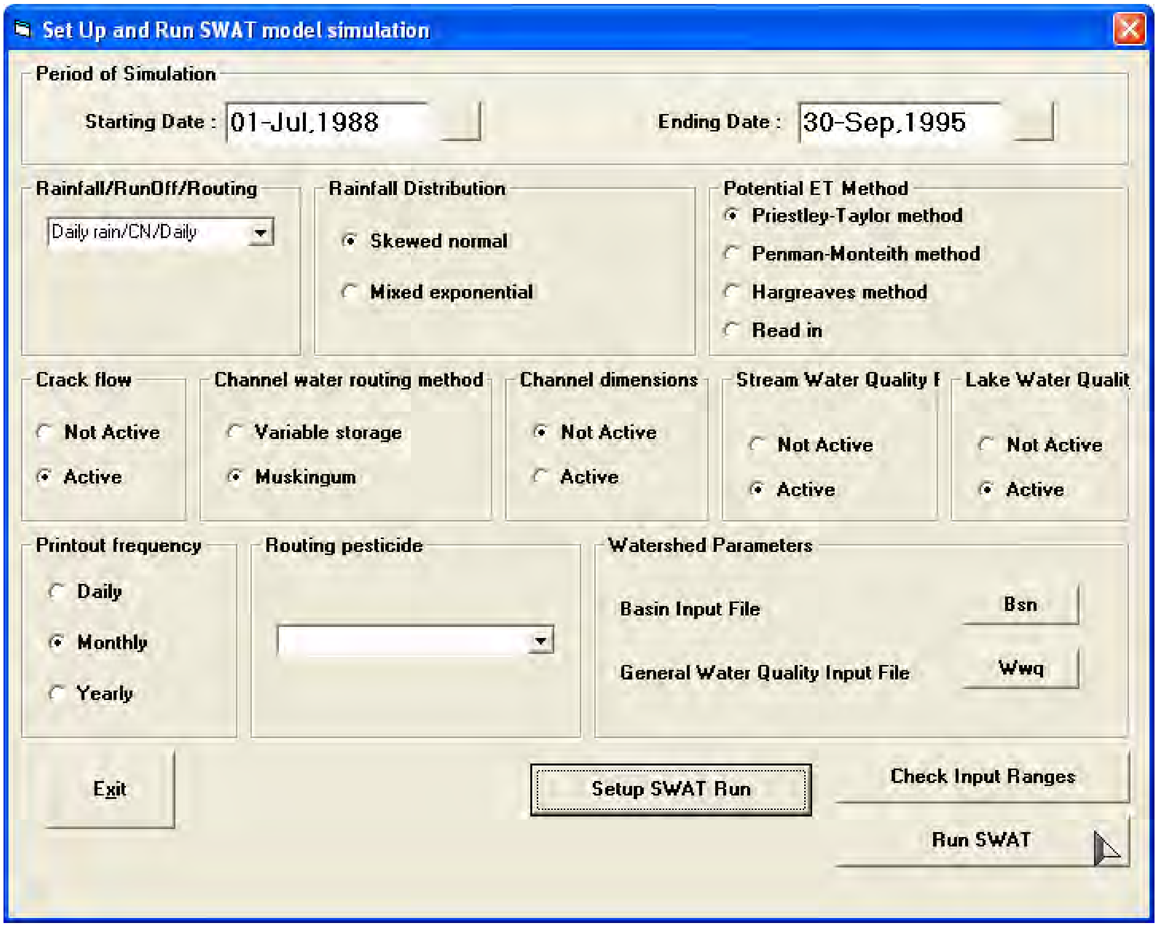



**FIG. 4.29.** SWAT set up configuration window.

Fig. 4.29 shows the configuration of the model. It can be highlighted the following processes and methods:

- Rainfall/Runoff/Routing configuration shows that we chose a rainfall input on a daily basis; SCS Curve number as the method for calculating runoff and a daily routing of the water in the system.
- Potential Evapotranspiration is calculated using the Priestley Taylor Method.

- The *Crack Flow condition* is set to active. This option considers the behavior of the *vertisols* on the runoff calculation. *Vertisols* are a special type of soil, with a clay percentage greater than 30% and produce cracks on the soil surface. This type of soil shrinks during the dry season and swells during the wet season. According to Fig. 4.30, our case study area falls in a region that contains vertisoils.

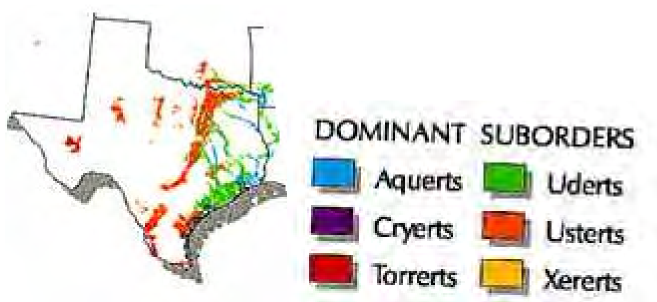



**FIG. 4.30.** Location of Vertisols subtypes on Texas (Source: NRCS).

- The Muskingum method was used for routing the water along the stream channels.
- Stream water quality and lake water quality routines are active.
- The printout frequency of the output is going to be monthly. SWAT prints out also a yearly summary with the results.
- Since we don't have pesticide information, we didn't make use of the pesticide routing.

## 4.4 RESULTS

### 4.4.1 SWAT output and calibration.

The SWAT model produces several text files as output. Some of them detail the time series of the parameters depending on their parameterization level. These text files can be imported into the Dynamic geodatabase and, thus, complete the objective of the interface.

The *basins.rch* output text file contains the time series of the parameter values of the reaches. From this text file we obtain the simulated average monthly stream flow, which can be compared with the average monthly stream flow series recorded on the USGS gage # 08201500. Statistics coefficients are calculated using both series in order to assess the model and the good fit of the simulated series. With the intention of improving these coefficients, various parameter values were calibrated with the aid of the interface's editors and with table operations using GIS techniques. These parameters and their calibrated values are explained on Table 4.4.

Another calibration parameter, not included in Table 4.4., is the SCS curve number. It was noticed that the default curve numbers that the SWAT database (crop table) has for curve numbers, were considerably high for this region. Therefore it was decided to use the curve numbers from the grid detailed on 4.3.6. Moreover, it was also decided to convert those numbers to a soil with moisture condition I, to reflect the dryness at the beginning of the simulation period. After that, the curve number is modified automatically by SWAT, based on soil moisture conditions.

**TABLE 4.4.** Calibration parameters (Description source: Neitsch *et al* 2000)

| Parameter | PGDB table | Description | Parameterization level | Calibrated value |
|---|---|---|---|---|
| ESCO | hru | Soil Evaporation compensation factor. Accounts for the effect of capillarity, crusting and cracks. | HRU | 0.05 |
| SOL_AWC1 | sol | Water capacity of the first soils layer | HRU | 0.9 |
| GWREVAP | gw | Groundwater "revap" coefficient. Accounts for the movement of water from shallow aquifer to the root zone. If it tends to 0 the movement is restricted. | HRU | 0.1 |
| REVAPMN | gw | Threshold depth of water for "revap" or percolation to occur in the shallow aquifer. | HRU | 0 |
| GWQMN | gw | Minimum threhold depth for a shallow aquifer to permit groundwater returning flow. | HRU | 4000 |
| ALPHA_BF | gw | Baseflow alpha factor.Index of groundwater flow response to changes in recharge. | HRU | 0.048 |
| CH_N | rte | Mannings coefficient for the main channels. | HRU | 0.6 |
| CH_N | sub | Mannings coefficient for tributary channels. | SUBBASIN | 0.05 |
| SLSUBBSN | hru | Average slope legth. Distance that the sheet flow is the dominant surface runoff process. | HRU | -20% |
| CH_K | rte | Effective hydraulic conductivity in the main channel. | SUBBASIN | 300 |
| CH_K | sub | Effective hydraulic conductivity for the tributary channels. | SUBBASIN | 0.025 |
| MSK_CO1 | bsn | Muskingum calibration coefficient. Controls the impact of the storage time constant (Km) for low flow. | BASIN | 10 |
| MSK_CO2 | bsn | Muskingum calibration coefficient. Controls the importance of inflow and outflow in determining storage in the reach. | BASIN | 10 |
| MSK_X | bsn | Muskingum weighting factor. It is a function of the wedge storage. | BASIN | 0.1 |

The following statistical coefficients were considered to evaluate the efficiency of the model:

- *Nash-Sutcliffe coefficient (CNS)* (Nash and Sutcliffe, 1970): This coefficient has been used to describe the goodness of fit of hydrologic models.

$$CNS = 1 - \frac{\sum_{i=1}^{n}\left(Qsim_i - Qobs_i\right)^2}{\sum_{i=1}^{n}\left(Qobs_i - \overline{Qobs}\right)^2} \tag{4.1}$$

where:

*Qsim = simulated stream flow for the ith time step.*

*Qobs = observed stream flow for the ith time step.*

$\overline{Qobs}$ *= average of observed stream flow values.*

*n = number of observations.*

- *Square of the Pearson product-moment correlation coefficient (* $r^2$ *):*

Where $r = \dfrac{n\left(\sum_{i=1}^{n} Qobs_i Qsim_i\right) - \left(\sum_{i=1}^{n} Qobs_i\right)\left(\sum_{i=1}^{n} Qsim_i\right)}{\sqrt{\left[n\sum_{i=1}^{n} Qobs_i^{\;2} - \left(\sum_{i=1}^{n} Qobs\right)^2\right]\left[n\sum_{i=1}^{n} Qsim_i^{\;2} - \left(\sum_{i=1}^{n} Qsim\right)^2\right]}}$ (4.2)

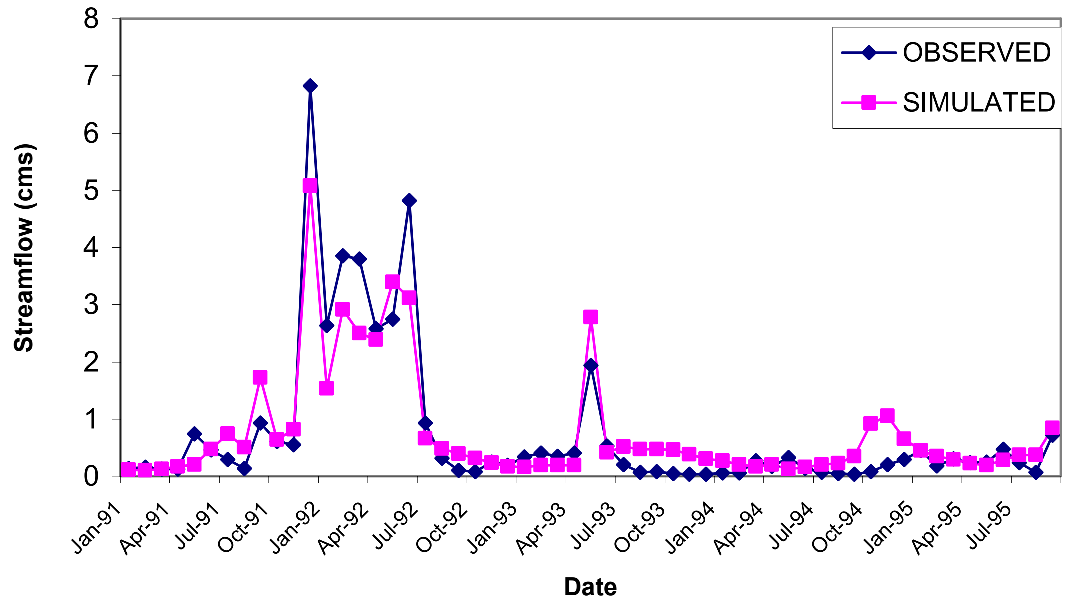



**FIG. 4.31.** Observed and simulated average monthly streamflow in cubic meters per second.

Fig. 4.31 shows the stream flow time series of observed and simulated data, at the outlet of subbasin #39 (USGS gaging station). The series contains seasonal trends and extreme peaks. Overall, the model captures shape and location of extreme peaks, although, underestimating high peaks. Spruill *et al* (2000) in their literature review, pointed cases where SWAT model doesn't describe well peaks in watersheds with scattered precipitation and extreme events. Spruill *et al* also showed a somewhat similar pattern of underestimating and overestimating highs and lows in their hydrographs.

The Nash – Sutcliffe coefficient (CNS) shows a value of 0.84 for the entire simulation period. A value of 0.87 for $r^2$, illustrates the good correlation between observed and simulated stream flows.

### 4.4.2 Uncertainty analysis.

*4.4.2.1 Case study overview*

Several questions are raised once we obtain a SWAT output. For the Seco Creek case study, we obtained time series of different water quality parameters in a time period that goes from January 1991 to September 1995. These parameters are monthly and annually averaged.

But how do we assess the fitness of these time series? USGS gauging station #08201500 has a useful period of record. However, concerning water quality data, it only contains 81 discrete samples (containing the concentration of several water quality indicators) taken on a time period from 1970 to 1995. The samples were taken on different times and in different points of the storm hydrographs. They cannot be used as a source for time series because they are not representative values for any time step. There is scarcity of time series data that can depict the variability, tendencies and/or relationships that these variables might have in a hydrologic system. Without these data, the accuracy of the SWAT simulated time series cannot be determined. Hence, there is *uncertainty* in our results.

In searching for possible solutions for the scarcity of data, Baird *et al* (1996) conducted an analysis of the observed data (from USGS gage #08201500) based on flow rate, time of year and location, but no correlation was reported (between flow rates and water quality concentration values). Nevertheless, this station is still considered valuable for the programs and projects already mentioned. Discharges from non point sources of pollution are being quantified and compared, on the basis of this gage. Baird *et al*, (1996) applied the concept of Event Mean Concentration (EMC) values, to obtain representatives from the recorded data set The concept considers the mean of the events as representative of the data set However, Baird *et al* considered the *median* as a more representative value since it is not affected by occasional lows or highs. The EMC's then are used as input in hydrologic models, to determine annual loadings of several watersheds. They recognize that this approach can only be suitable for relative comparisons between management scenarios.

The methodology presented here proposes modeling the uncertainty in the output caused by the uncertainty in the input data.

For the application and discussion, SPAT is going to be run on the Upper Seco Creek case study. It is expected that the following questions will be answered:

- What is the probability of obtaining a certain output value?
- What is the probability of that value to be exceeded?
- What is the expected (or average) annual value?

The *sediment concentration* parameter has been chosen to illustrate how it is possible to model uncertainty in the Upper Seco Creek Case study.

*4.4.2.2 SPAT configuration*

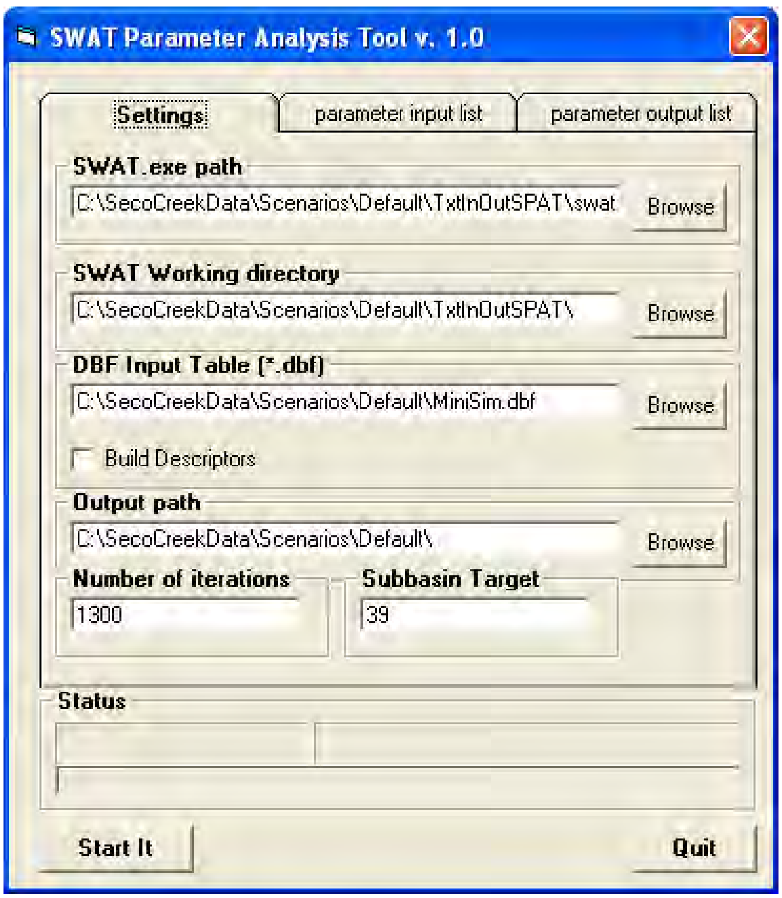



**FIG. 4.32.** SPAT configuration settings.

Fig. 4.32 shows SPAT interface configuration tab. It requires the following data:

- The path to the SWAT model's executable file.

- The SWAT working directory path. This is the location where all the text files for the hydrologic model have been created.
- A Dbase table containing random numbers for each parameter (following the desired distribution). SPAT accepts predefined tables and also has tools for creating them .
- The path where the user wants to store the output from Monte Carlo simulation.
- The desired number of iterations. For our case study there were approximately 1400 iterations of 8 years each one.
- The subbasin target SWAT produces a great amount of data on each run, even on simple hydrologic models. Thus, it is necessary to define a target subbasin. SPAT selects the output information, concerning only to this subbasin. For our case study, it is subbasin 39 (the most downstream).

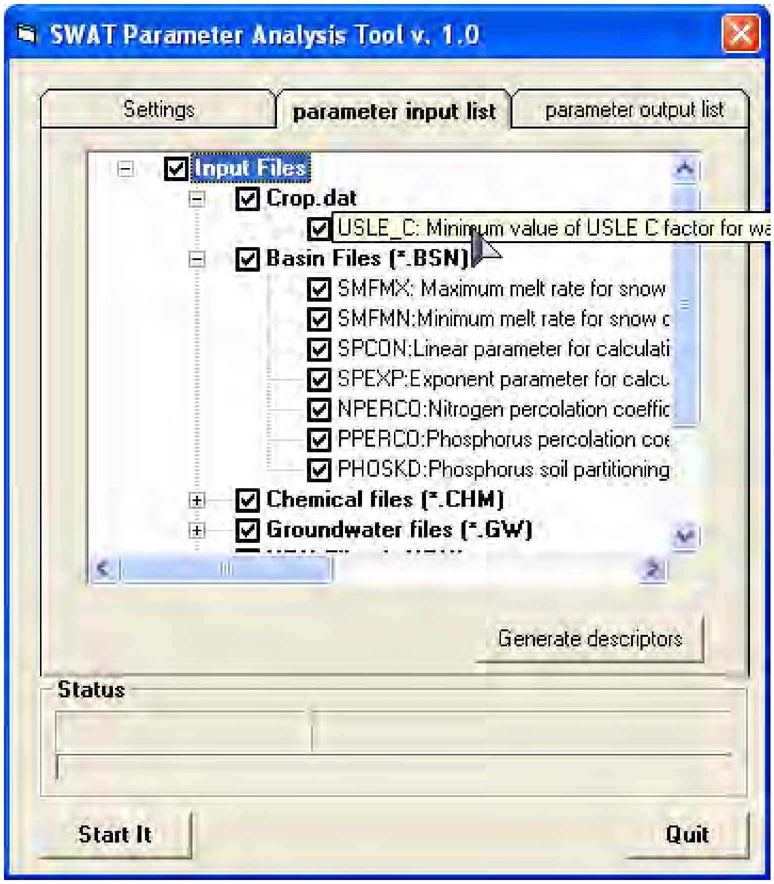
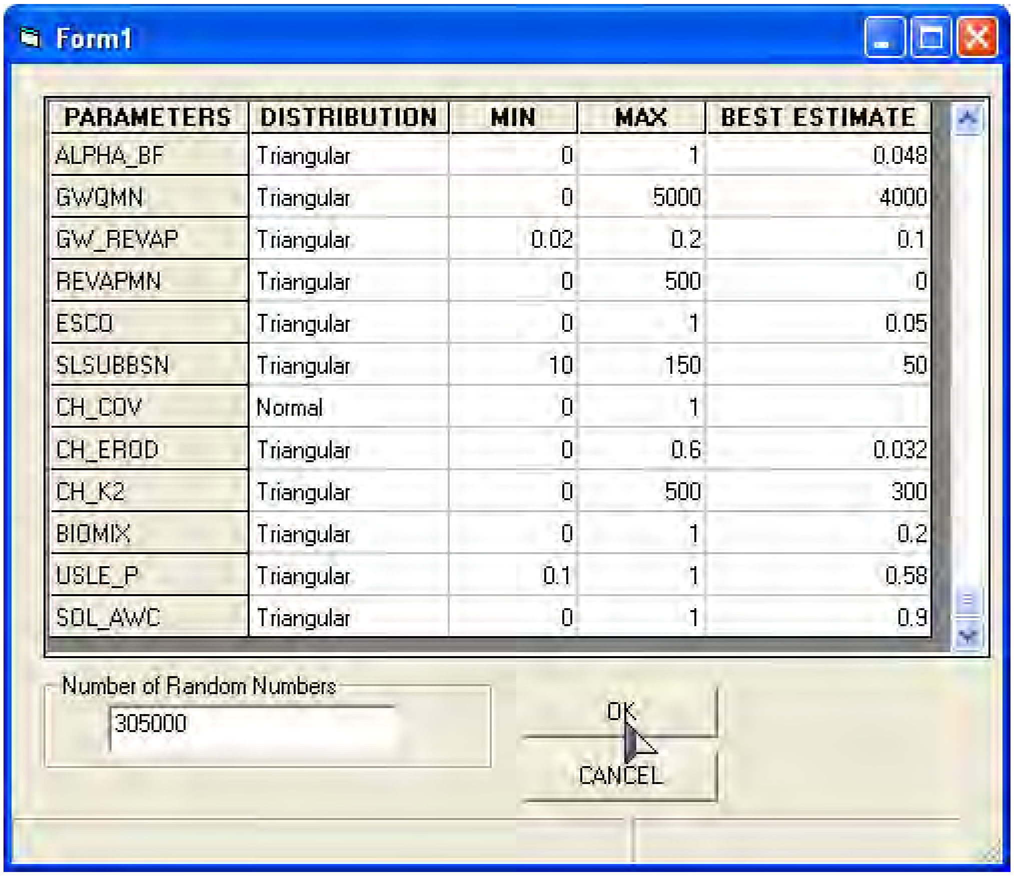



**FIG. 4.33.** SPAT parameter input list.

Fig. 4.33 shows the parameter input list tab. In this section, the user selects the sensitive variables. These variables are the ones to be randomized.

If the user hasn't provided a predefined table of random numbers, the *Generate Descriptors* interface aids the user in building a random numbers table (Fig. 4.34).



| PARAMETERS | DISTRIBUTION | MIN | MAX | BEST ESTIMATE |
|---|---|---|---|---|
| ALPHA_BF | Triangular | 0 | 1 | 0.048 |
| GWQMN | Triangular | 0 | 5000 | 4000 |
| GW_REVAP | Triangular | 0.02 | 0.2 | 0.1 |
| REVAPMN | Triangular | 0 | 500 | 0 |
| ESCO | Triangular | 0 | 1 | 0.05 |
| SLSUBBSN | Triangular | 10 | 150 | 50 |
| CH_COV | Normal | 0 | 1 | |
| CH_EROD | Triangular | 0 | 0.6 | 0.032 |
| CH_K2 | Triangular | 0 | 500 | 300 |
| BIOMIX | Triangular | 0 | 1 | 0.2 |
| USLE_P | Triangular | 0.1 | 1 | 0.58 |
| SOL_AWC | Triangular | 0 | 1 | 0.9 |



**FIG. 4.34.** Generate descriptors interface.

In this interface (Fig. 4.34), the user introduces the descriptors for each parameter that he/she selected from the parameter input list (Fig. 4.33). The user has to determine the number of rows that this table contains (number of random numbers).

18 out of 27 parameters were selected for the Monte Carlo simulation (Table 4.5.). In most of them, the ranges are the ones recommended on Neitsch *et al* (2000). The best estimate has been defined in some cases as the default value that SWAT takes if the user gives no value (case #1 to #6, #16 and #17 on Table 4.5.) ; in other cases, it

was the result of the calibration of the hydrologic model (case #7 to #11, #15 and #18); or simply based on rules of thumb (i.e. case #14 where the channel erodibility is expected to have a  value in an order less than the soil beneath it).

**TABLE 4.5.** Probability descriptors for the Upper Seco Creek case study

| # | Parameter | Distribution | Min | Max | Best Estimate |
|---|---|---|---|---|---|
| 1 | SPCON | Triangular | 0.0000 | 0.0100 | 0.0001 |
| 2 | SPEXP | Triangular | 1.0000 | 1.5000 | 1.0000 |
| 3 | NPERCO | Triangular | 0.0000 | 1.0000 | 0.2000 |
| 4 | PPERCO | Triangular | 10.0000 | 17.5000 | 10.0000 |
| 5 | PHOSKD | Triangular | 100.0000 | 200.0000 | 175.0000 |
| 6 | SOL_LABP | Triangular | 0.0000 | 100.0000 | 5.0000 |
| 7 | ALPHA_BF | Triangular | 0.0000 | 1.0000 | 0.0480 |
| 8 | GWQMN | Triangular | 0.0000 | 5000.0000 | 4000.0000 |
| 9 | GW_REVAP | Triangular | 0.0200 | 0.2000 | 0.1000 |
| 10 | REVAPMN | Triangular | 0.0000 | 500.0000 | 0.0000 |
| 11 | ESCO | Triangular | 0.0000 | 1.0000 | 0.0500 |
| 12 | SLSUBBSN | Triangular | 10.0000 | 150.0000 | 50.0000 |
| 13 | CH_COV | Normal | 0.0000 | 1.0000 | 0.5000 |
| 14 | CH_EROD | Triangular | 0.0000 | 0.6000 | 0.0320 |
| 15 | CH_K2 | Triangular | 0.0000 | 500.0000 | 300.0000 |
| 16 | BIOMIX | Triangular | 0.0000 | 1.0000 | 0.2000 |
| 17 | USLE_P | Triangular | 0.1000 | 1.0000 | 0.5800 |
| 18 | SOL_AWC | Triangular | 0.0000 | 1.0000 | 0.9000 |

Fig. 4.35 shows the parameter output list. The user can select the information that he wants to collect from the output of SPAT. Three files are available for selection: *HRU output file* (.sbs) containing the HRU output information for the selected subbasin; *Subbasin output file* (.bsb) including information of the subbasin level parameters and *Main channel output file* holding information of the main channel of the selected subbasin (subbasin #39). These files keep the same format and variables of the SWAT

output files but contain the data for *x* number of simulations. For more information on the variables contained on each file, please refer to Neitsch *et al* 2000.

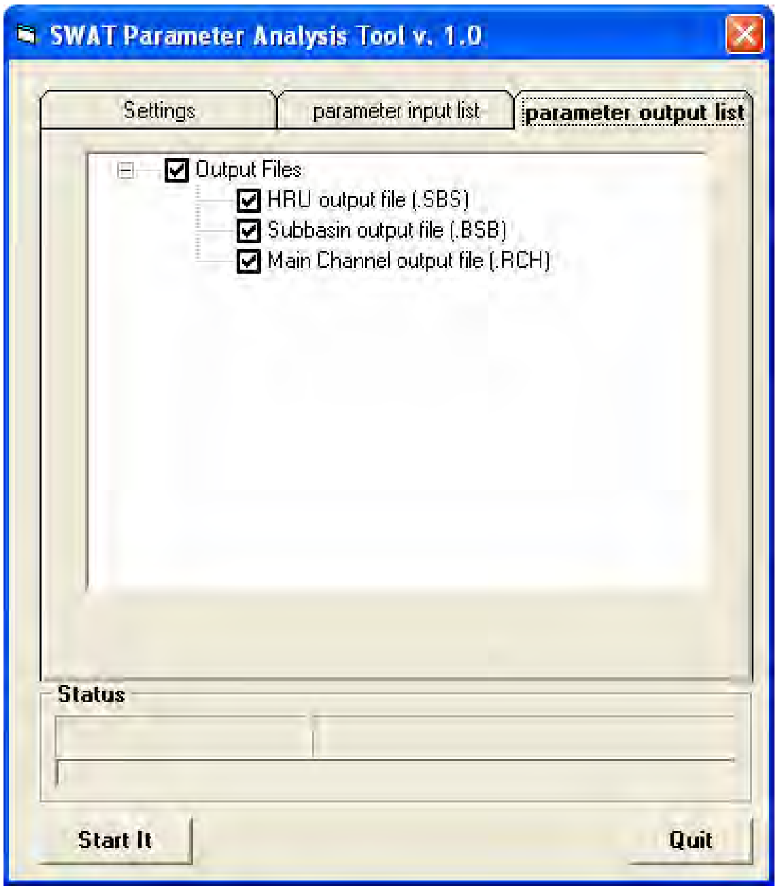



**FIG. 4.35.** SPAT parameter output list.

*4.4.2.3 SPAT output analysis*

For the case study, the *Main channel output file* proves to be useful, since it contains all the water quality parameter values for the main channel of the subbasin #39, the most downstream subbasin. Therefore, it holds the output information for the whole basin.

This output information comprises monthly average values and annual average values for 8 years of simulation (from 1988 to 1995). A period from 1991 to 1995 was chosen for the hydrologic simulation, letting the three previous years as a “warm up” for the model.

As seen on Fig. 4.31, the 1991-1995 period has a considerable variability of average monthly flows. Each year has different characteristics from the previous one. Concentrations of sediments or nutrients would probably vary in a similar way. Therefore, the output of SPAT was analyzed yearly, and compared to the output that SWAT gives without performing uncertainty modeling and EMC’s. The 1992 year has been selected for this analysis, since the stream flow time series (Fig. 4.31) shows several extreme events through the whole year. Determination of annual non-point sources of pollution discharges for this year proves to be of special importance given that high loadings of contaminants from surface runoff are expected.

In order to process SPAT output data, the files are exported to EXCEL and the information is filtered and processed. With the selected data, the following statistical analysis was done:

- Histograms, that will: (1) depict the distribution of the output variable; (2) show the range of the data; (3) identify symmetry or skewness; (4) detect the presence of outliers; (5) identify the mode(s).
- Cummulative histograms that will help us providing information regarding specific statistical parameters like the median.
- Normalized histograms will aid in determining the approximate probability of a value of falling on a certain range (normalized histogram). Ekchardt *et al* 2003 used this method for calculating approximate probabilities of occurrence in the interval.
- Exceedance frequency curves, which will help in determining exceedance probabilities of a certain values.
- Selection of the representative value for the distribution. This would be the *most expected average annual value* for a certain parameter.

In order to build any histogram, it is necessary to define the number of intervals. The number of equally sized intervals (called *bins* in Excel) for the histograms is obtained using the equation detailed on Iman and Conover (1983):

$$2^k \geq n \tag{4.3}$$

Where:

*k = minimum integer that validates the inequality*

*n=number of samples*

Our sample size is 1284, therefore, for our case study the number of intervals (k) is equal to 11.

In order to construct a histogram, it is required an equal interval size. Thus, for each parameter, the interval size will be defined by:

$$Interval\ size = \frac{Max - Min}{k} \tag{4.4}$$

The following are the results for, sediment concentration.

**TABLE 4.6.** SPAT descriptive statistics of sediment concentration (mg/l) population for 1992

| *Statistics* | |
|---|---|
| **Mean** | 1300.508 |
| **Standard Error** | 24.6493 |
| **Median** | 1196 |
| **Standard Deviation** | 883.2571 |
| **Sample Variance** | 780143 |
| **Kurtosis** | -0.4968 |
| **Skewness** | 0.503407 |
| **Range** | 4115.14 |
| **Minimum** | 28.86 |
| **Maximum** | 4144 |
| **Sum** | 1669852 |
| **Count** | 1284 |
| **Confidence Level(95.0%)** | 48.3574 |

Table 4.6. shows the descriptive statistics of the population produced by SPAT for the year 1992. It can be noticed that mean and median show high values of concentrations, if they are compared to the sediment concentration EMC of 245 mg/l.

The negative kurtosis indicates that our histogram if going to be peakedness or flat compared to a normal distribution and the positive skeweness indicates that our histogram is going to be right tailed. The confidence indicates that I'm a 95% *confident* that my mean is 1300.508 ± 48.3574.

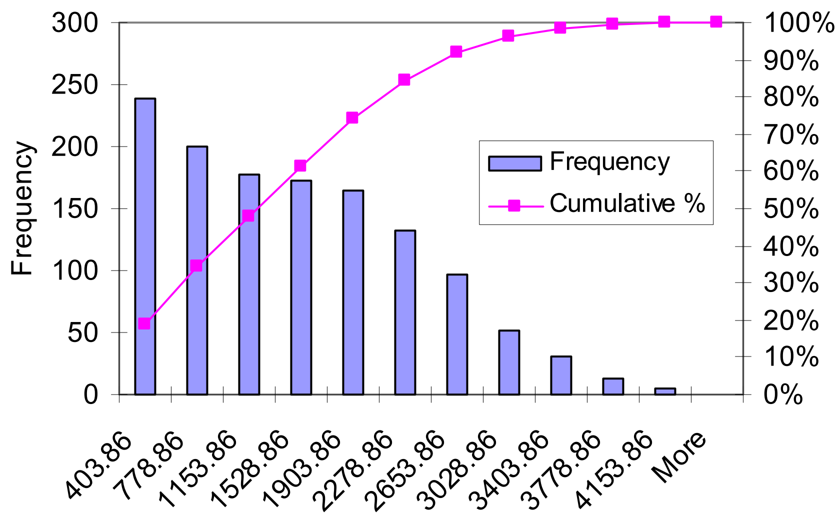



**FIG. 4.36.** SPAT sediment concentration (mg/l) histogram for the 1992 period.

Fig. 4.36 illustrates a right-skewed (tail to the right) histogram. A cumulative histogram is also contained where it can be identified the median at the 50% of the curve.

Several pdfs can be fitted to this histogram shape. The ideal distribution would show linearity on a probability plot. Among the most common skewed distribution families we can highlight the log-normal, half-normal, chi-squared, Cauchy, Weibull etc.

Fig. 4.37 shows the probability plot of a half-normal distribution for the sediment concentration data.

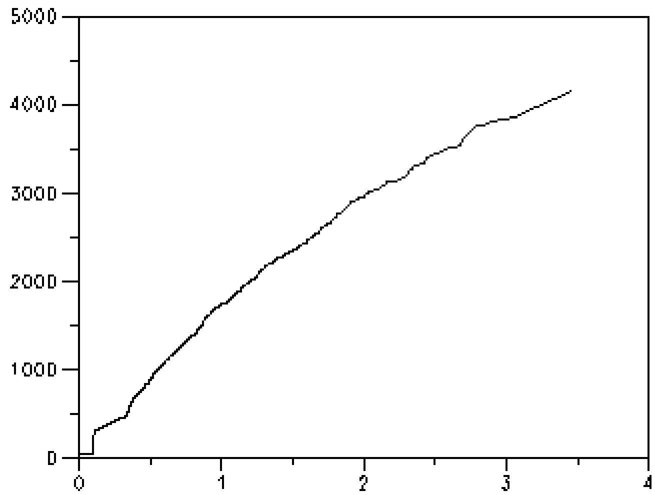



**FIG. 4.37.** Half normal probability plot of SPAT sediment concentration (vertical axis: ordered concentrations in mg/l, horizontal axis: half normal distribution order statistic medians).

The correlation found on the probability plot for the half normal distribution is equal to 0.9906 (it is an estimation of how well the line fits the probability plot). Several other skewed distributions could fit our produced data. For our case study we are going to normalize the histogram and calculate approximate probabilities (calculating areas under the curve) instead of calculating the parameters for a fitted pdf.

Fig. 4.36 also reveals the most frequent value (mode) of 201.93, calculated taking the mid point (*midbin*) of the most frequent bin. It is possible to notice that all of the samples that have been taken from 1970 to 1995 fall in this interval, and that the reported EMC is similar to the reported mode of our histogram. But what would be the probability of finding a value in this interval? This question can be answered, with the aid of a normalized frequency curve with an area equal to 1. The frequency curve is the line that connects the midpoints of the bin bars of a histogram.

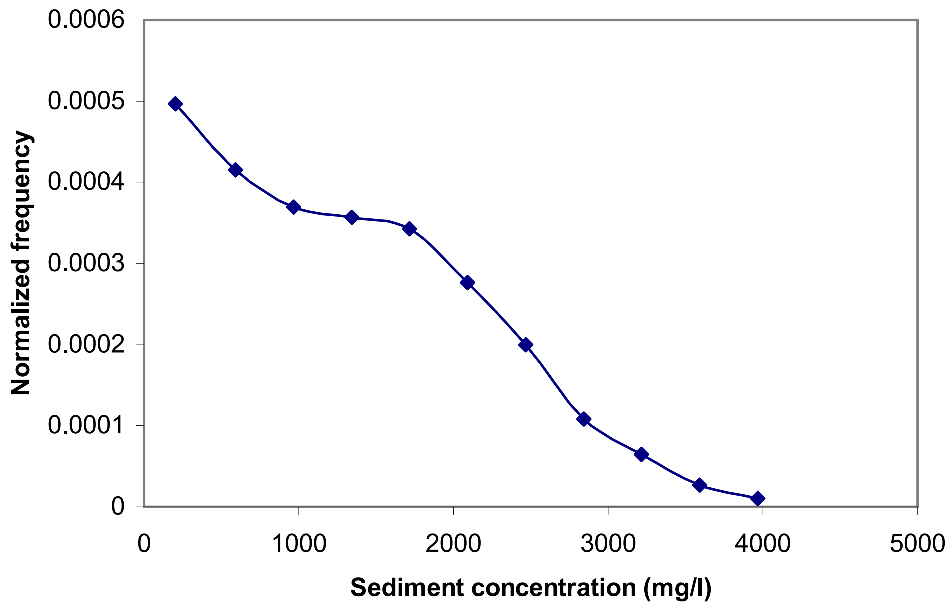



**FIG. 4.38.** Normalized frequency curve of sediment concentration for 1992.

Fig. 4.38 describes the normalized frequency curve. For calculating a probability of finding a value in an interval, we need to integrate numerically the above curve, within the limits of the desired interval. For our case study we would like to test the probability of finding a value in the range of the sampled values (197 mg/l to 295 mg/l)

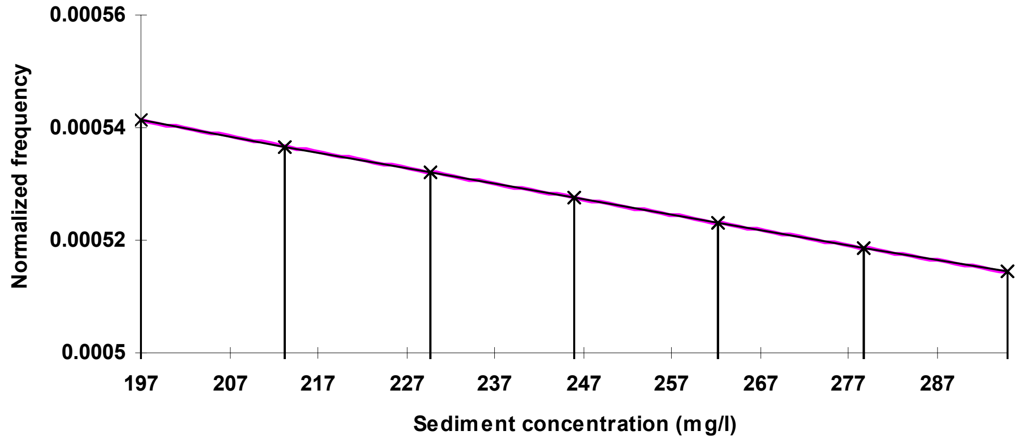



**FIG. 4.39.** Numerically integrated interval.

Fig. 4.39 shows the numerical integration of the normalized frequency curve, in the range from 197 mg/l to 295 mg/l. The used method is dividing the curve into trapezoids. The area gives us a probability of 5.17% chance of finding a value in this range. This probability can be considered low and therefore, this range may not be representative of the year 1992.

Another useful graph is the exceedence frequency curve (Fig. 4.40). This graph depicts the probability of a value of being exceeded. It is useful to describe probabilities of several thresholds. It requires sorting the population in descendent order and assigning a rank (*k*) to each element . The probability of exceedance of each element would be calculated using the following formula:

$$P(C > c^*) = \frac{m}{n+1} \tag{4.5}$$

*Where:*

*C = variable*

*c* = variable value*

*m = rank*

*n = population size*

Using this curve, we can know what would be the probability of having a concentration greater than the maximum of the historical sampled data (295 mg/l). This would be approximately 91% chance, showing us that the maximum sampled may not be representative as a maximum for the year 1992.

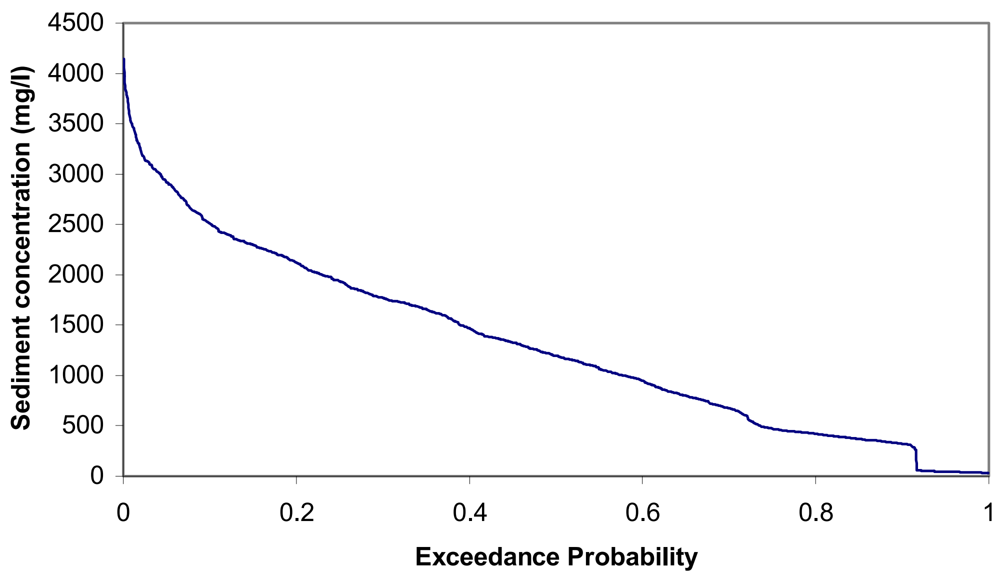
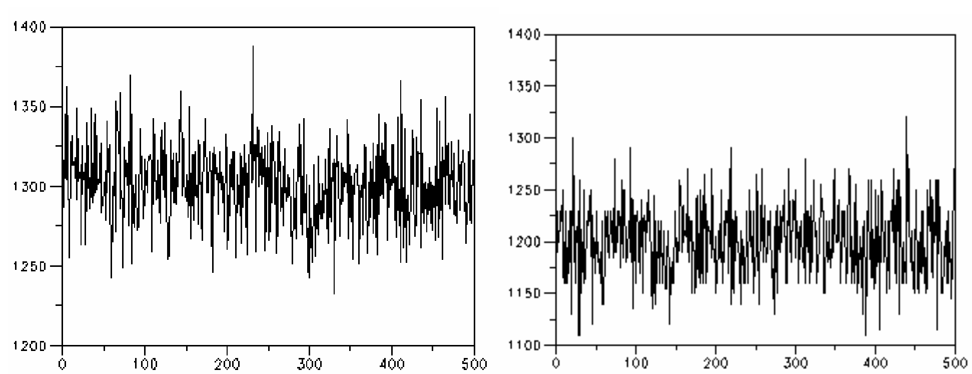



**FIG. 4.40.** Exceedance probability curve of the sediment concentration for 1992.

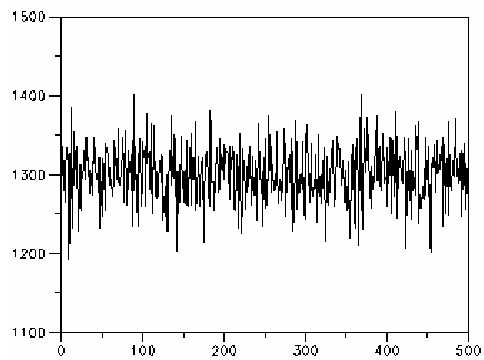
Now, it is necessary to assess what would be the representative value of the obtained distribution. This would be the *most expected average annual concentration for 1992*. Several location indicators can be used: mean, median, mode, trimmed mean, mid-range, etc. It is recommended for a skewed distribution, to report at least three measures of location (National Institute of Standards and Technology –NIST-, 2003).

The *bootstrap plot* is a method that can help in analyzing which measure of location is the most representative for our data. This plot is used to estimate the variability of a location indicator. It builds a sub sample of the data and calculates the measure of location and repeats this process several times (500 to 1000), plotting in a graph the results (NIST, 2003). A bootstrap of the mean would show for example, the variability of the mean when several sub samples are chosen. Several bootstrap plots are shown on Fig. 4.41 and a summary of their statistics, in Table 4.7.

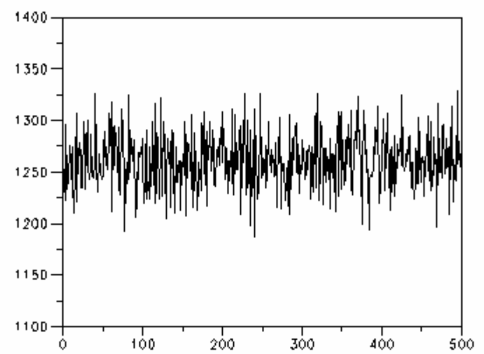



*MEAN* *MEDIAN*

**FIG. 4.41.** Bootstrap plots of mean, median, mid mean and trimmed mean.




*MID MEAN*




*TRIMMED MEAN*

**FIG. 4.41.** Continued.

**TABLE 4.7.** Bootstrap plots statistics summary

| Location Estimator | Mean of bootstrap values | 95% confidence range from | to | Absolute Range |
|---|---|---|---|---|
| MEAN | 1301.18 | 1252.8 | 1347.8 | 95 |
| MEDIAN | 1195.6 | 1122.5 | 1270 | 147.5 |
| MID-MEAN | 1302.49 | 1237.52 | 1370.71 | 133.19 |
| TRIMMEAN | 1263.35 | 1213.57 | 1318.14 | 104.57 |

According to Table 4.7. the mean presented the minimum absolute range for a 95% of confidence. Therefore it can be selected as a representative (or most expected value) for the distribution corresponding to 1992. From the population statistics (Table 4.6.) this value is equal to 1300.51 mg/l.

Another way of calculating the expected average annual concentration is integrating numerically the exceedance frequency curve (Wurbs *et al*, 2001). This would represent a probability weighted average. Dividing the curve into 100 trapezoids, it is possible to obtain a good estimate of the expected average (Fig. 4.42)

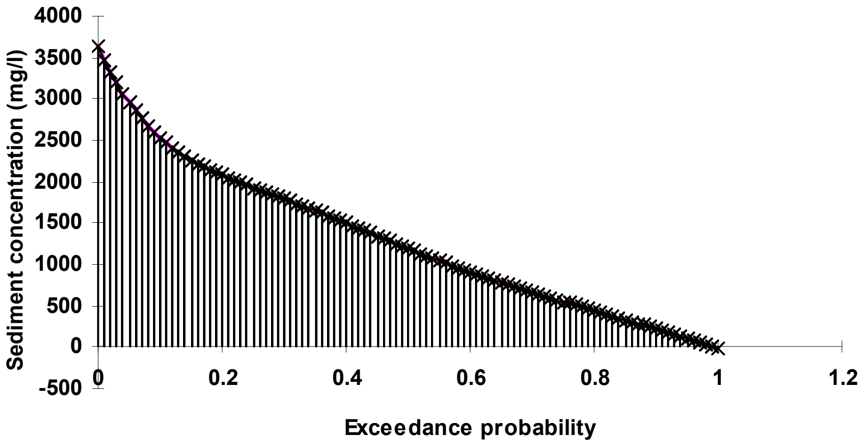



**FIG. 4.42.** Numerical integration of the exceedance probability curve.

The estimated resulting value from the integration is equal to 1300.1 mg/l. This value might demonstrate that the real average concentration for the year 1992 is 5 times higher then a median estimate.

If we would chose the SWAT output value, without explicitly modeling the uncertainties, and considering the default values for sediment related parameters, it produces a value of 0.68 mg/l as an average annual sediment concentration.

# 5 CONCLUSIONS

The main objectives of this thesis have been the development of an ArcGIS interface and the design of a geodatabase for the Soil and Water Assessment Tool (SWAT). A new methodology has been devised for the application of the latest advances of Geographic Information Systems in the Water Resources Engineering field.

SWAT, a hydrological semi-distributed model that has been extensively applied and validated throughout the world, requires a complex series of procedures to extract, manage and categorize the required input information from hydrological datasets. The efficiency of SWAT relies on the availability and completeness of these input datasets. Moreover there is also the necessity of organizing and analyzing SWAT output, since its best application is in long period simulations that yields a great amount of data. Therefore, the development of an interface that performs the pre and post processing operations is justified. Several previous interfaces have been successfully designed and applied. Among them, AVSWAT (Di Luzio 2000) effectively coupled SWAT with the ESRI GIS software ArcView 3.x. and facilitated the use of SWAT when reducing the modeling time. However, new paradigms in GIS emerged, and a new series of software was designed based on the Component Object Model technology. ArcGIS SWAT is the result of the application of new technologies and concepts. ArcGIS SWAT performs pre and post processing operations for the SWAT model.

The base technologies and concepts for the design of ArcGIS SWAT are:

- Component Object Model: a new technology that sets standards for building software.
- ArcGIS 8.x series: the latest software platform created by ESRI, for performing GIS analysis and visualization.
- The Geodatabase data model: the model that details the structure of a Geographic Relational Database Management System (geodatabase).
- Hydrologic data models: custom models derived from the Geodatabase data model that depicts the structure and organization of hydrologic data. Their objective is to improve the performance of hydrologic models and furthermore facilitate the connectivity with other models. ArcHydro set a milestone for being the first hydrologic data model using these technologies and concepts.

The process structure of the ArcGIS SWAT interface was designed based on the advantages of the previous versions. It comprehends a set of user-friendly tools that guide the user throughout the whole process of modeling and simulation without leaving the GIS environment. It performs extensive analysis over a variety of source data and formats. This analysis is based on raster and vector analysis. ArcGIS SWAT is subdivided in the following modules:

- Watershed Delineator: a raster and vector analysis whose objective is to define the location and properties of watersheds and streams.

- HRU definition: prepares land use and soils data for analysis and afterwards it creates hydrologic response units. These are unique combinations of land use and soil within a sub basin. These elements are supposed to behave hydrologically similar and constitute the base for the calculation of the water budget in the analyzed system. ArcGIS SWAT georeferences HRUs (locates them spatially) and calculates their spatial properties. This constitutes a highlighted improvement with respect to previous SWAT interfaces.
- Weather definition: prepares the metereological data extracted from weather stations or provided by the user.
- File Editors: a set of editors that help the user in creating and editing the required input text files for the SWAT model.
- SWAT output analysis: as scarcity of data plays a key role when modeling with SWAT, uncertainty is a factor that should be considered on an in-depth study. ArcGIS SWAT contains a complete module for uncertainty analysis which is based on Monte Carlo simulation technique.

The interface relies on two geodatabases that work collectively. Modeled with UML and applied with ArcObjects, these geodatabases prove to be another highlighted improvement. Their design started with concepts from ArcHydro. However the complexity and the requirements of the SWAT model, demanded a lot more elements, associations, relationships, different organization, creating a final product that reminds the antecessor but is deviated for our purpose.

The two geodatabases are:

- Dynamic geodatabase: sets the structure and organization of the data of each SWAT project, and builds a foundation for further analysis with SWAT or other hydrologic model.
- Static geodatabase: it is the structure and organization of the data that is common for every SWAT project.

A case study has also been presented. The Upper Seco Creek watershed illustrated the benefits of creating a SWAT model with the aid of ArcGIS SWAT. Accurately delineated watersheds and streams, georeferenced HRUs, tools for editing and modifying data, a well defined structure that have all the benefits of a relational database, are some of the proved advantages of ArcGIS SWAT. Average monthly stream flow for a period from 1991 to 1995, was calibrated with ArcGIS SWAT, and produced a square Pearson correlation of 0.87 and a Nash-Sutcliffe coefficient of 0.84, when compared to historic data. Moreover, the uncertainty analysis module, yielded positive results and reinforced the importance of modeling explicitly the uncertainties. For the analysis of sediment concentration for the year 1992, the analysis quantified the uncertainty and produced the most expected result. The results were compared with sampled data over a period from 1970 to 1995.

Every model faces the necessity of being improved and modified. ArcGIS SWAT has some limitations that can also be motive for future work. As known limitations of the interface we can highlight:

- The addition of the new components that execute procedures such as georeferencing HRUs, produced a longer processing time, when compared to other interfaces.
- The delineation of watersheds and streams can only be done over raster data.
- Raster data continue to be saved on a DOS folder structure.
- The interface is limited to Windows users.
- The interface does not accept pre-processed data. The framework is linear. We cannot jump a step in the process. For example, we cannot introduce our delineated watershed and streams.
- Development of a module for the explicit implementation of the modeled relationships (since ArcView cannot define relationship classes, it is necessary for the user to define memory relationship classes like joins or relates) between the datasets would help the user in reduced modeling time.
- A module for exporting SWAT output data to the Dynamic geodatabase is necessary for assisting in further hydrologic analysis.

# REFERENCES


Anderson, R., Hardy, Roach, J., and Witmer, R. (1976) *A Land use and Land Cover Classification System for Use With Remote Sensor Data*, Geological Survey Professional Paper #964.

Arnold, J. G., and Allen, P. M. (1996). "Estimating hydrologic budgets for three Illinois watersheds." *Journal of Hydrology*, 176(1), 57 – 77.

Baird, C. Jennings M., Ockerman, D., Dybala, T. (1996), *Characterization of Nonpoint Sources and Loadings to the Corpus Christi Bay National Estuary Program Study Area*, TNRCC, Austin, TX.

Bian, L., Sun, H., Blodgett, C., Egbert, S., WeiPing, L., LiMei, R., and Koussis, A. (1996). "An integrated interface system to couple the SWAT model and ARC/INFO" *Proc., 3rd Intern. Conf. on Integrating GIS and Environmental Modeling*, U.S. National Center for Geographic Information and Analysis, Santa Fe, NM, January 21 – 26, 1996.
Santa Barbara, CA: National Center for Geographic Information and Analysis. CD.

Booch, G., Jacobson, I., Rumbaugh, James., and Rumbaugh, Jim (1999). *The Unified Modeling Language User Guide*, Addison Wesley Longman, Inc., Reading, MA.

Brown, D., and Raines, T. (2002). *Simulation of Flow and Effects of Best Management Practices in the Upper Seco Creek Basin, South-Central Texas, 1991-98* U.S. Geological Survey, Water-Resources Investigations Report 02-4249, Austin, TX.

Byars, B. W., Srinivasan, R., and Arnold, J. G. (2000) *SWAT/GRASS User Manual,* Baylor University, Waco, TX.
< http://www3.baylor.edu/~Bruce_Byars/swatgrassman.html>

Cantu-Paz, E. (2002), "On random numbers and the performance of genetic algorithms," *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO 2002, pp. 311-318, Morgan Kaufmann Publishers, San Francisco, 2002.

Cockus, S. (2002) *Marsenne-Twister Optimized Code for C*,
< http://www.math.keio.ac.jp/~matumoto/cokus.c>

Computational Science Education Project (CSEP) (2003), *Random Number Generators*,
<http://csep1.phy.ornl.gov/rn/rn.html>

Di Luzio, M., Srinivasan, R., and Arnold, J. G. (2002) "Integration of Watershed Tools and SWAT Model into BASINS" *Journal of the American Water Resour. Assoc. (JAWRA)*, 38(4), 1127 - 1141.

Di Luzio, M., Srinivasan, R., and Arnold, J. G., and S.L. Neitsch (2000) *Soil and Water Assessment Tool, ArcView GIS Interface Manual*, Grassland, Soil and Water Research Laboratory, Agricultural Research Service and Blackland Research Center, Texas Agricultural Experiment Station, Temple, TX.

Di Luzio, M., Srinivasan, R., and Arnold, J. G. (1998) "Watershed Oriented Non-point pollution Assessment Tool. *Proc. of the 7$^{th}$ Int. Conf. on Computers in Agriculture,* ASAE, 233 – 241.

Eckhardt, J., and Arnold, J.G. (2001). "Automatic calibration of a distributed catchment model." *Journal of Hydrology*, 251(1), 103 – 109.

Eckhardt, K., Breauer, L., and Hans-Georg, F. (2003) "Parameter uncertainty and the significance of simulated land use changes effects." *Journal of Hydrology,* 273(1), 164 – 176.

Environmental Protection Agency (EPA). (2003a) *Nonpoint Source Pollution: The Nation's Largest Water Quality Problem*.
<http://www.epa.gov/owow/nps/facts/point1.htm>

Environmental Protection Agency (EPA). (2003b) *Introduction to TMDLs*
< http://www.epa.gov/owow/tmdl/intro.html>

Environmental Protection Agency (EPA). (2003c) *About BASINS 3.0*
< http://www.epa.gov/waterscience/basins/basinsv3.htm>

Environmental Systems Research Institute (ESRI). (2003a)
<http://www.esri.com>

Environmental Systems Research Institute (ESRI). (2003b) "ArcGIS 8: The complete Geographic Information System", *ESRI White paper,* ESRI, Redlands, CA.

Environmental Systems Research Institute (ESRI). (1998) "ESRI Shapefile technical description", *ESRI White paper,* ESRI, Redlands, CA.

Hathaway, B. (1996) *Object-Orientation FAQ*
< http://www.avalon.net/~wbachman/OOFAQ/oo-faq-toc.html>

Hellweger, F. L., and Maidment, D. R. (1999). "Definition and connection of hydrologic elements using geographic data" *Journal of Hydrologic Engineering,* American Society of Civil Engineers (ASCE), 4(1), 10 – 18.

Hydrologic Engineering Center (HEC). (2003) *HEC-FDA* < http://www.hec.usace.army.mil/software/hec-fda/hecfda-hecfda.html>

Hydrologic Engineering Center (HEC). (2000) *HEC-GeoRAS* < http://www.hec.usace.army.mil/software/hec-ras/hecras-hec_georas.html>

*Hydrologic Modeling System (HEC-HMS) – Technical Reference Manual.* (2000). Hydrologic Engineering Center., U.S. Army Corp of Engineers, Davis, CA.

Iman, R. L., and Conover, W. J. (1983) *A Modern Approach to Statistics*, Wiley, New York.

Jenson, S.K. and Domingue, J.O. (1988)**.** **"**Extracting Topographic Structure from Digital Elevation Data for Geographic Information Systems Analysis**".** *Photogrammetric Engineering and Remote Sensing*, 54(11), 1593-1600.

MacDonald, A. (2001) *Building a Geodatabase,* Envirnonmental Systems Research Institute (ESRI), Redlands, CA.

Maidment, D. R., Olivera, F., Nawajish, N., Davis, K., Merwade, V., Grise, S. (2002). *ArcHydro: GIS for Water Resources,* Envirnonmental Systems Research Institute (ESRI), Redlands, CA.

Matsumoto, M (1998) *Marsenne Twister: A Random Number Generator*, < http://www.math.keio.ac.jp/~matumoto/emt.html>

Matsumoto, M. and Nishimura, T., (1998) "Marsenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator", *ACM Trans. on Modeling and Computer Simulation,* 8(1), 3-30

Muttiah, R., and Wurbs, R. (2002). "Scale-dependent soil and climate variability effects on watershed water balance of the SWAT model" *Journal of Hydrology*, 256(1), 264 – 285.

Nash, J. E., Sutcliffe J.E. (1970) "River flow forecasting through conceptual models, Part 1- A discussion of principles", *Journal of Hydrology*, 10(82), 282 - 290.

National Institute of Standards and Technology (NIST) (2003) *Engineering Statistics Handbook*
<http://www.itl.nist.gov/div898/handbook/>

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., Srinivasan, R., Williams, J. R. (2000) *Soil and Water Assessment Tool User Manual*, Grassland, Soil and Water Research Laboratory, Agricultural Research Service and Blackland Research Center, Texas Agricultural Experiment Station, Temple, TX.

Object Management Group (OMG), (1997). *What Is OMG-UML and Why Is It Important?* < http://www.omg.org/news/pr97/umlprimer.html>

Olivera, F. (2001). "Extracting hydrologic information from spatial data for HMS modeling" *Journal of Hydrologic Engineering,* ASCE, 6(6), 524 – 530.

Olivera, R., Reed, S., and Maidment D. R. (1998). "HEC-PrePro v. 2.0: An ArcView pre-processor for HEC's Hydrologic Modeling System." *Proc., 18th Int. ESRI User Conf.*, Environmental Systems Research Institute, San Diego, CA, July 25 – 31, 1998. <http://gis.esri.com/library/userconf/proc98/proceed/TO400/PAP400/P400.HTM>

Peschel, J.M., Lacey R. E. and Haan, P. K. (2003). "A SSURGO Pre-Processor Extension for the ArcView Soil and Water Assessment Tool", *2003 ASAE Annual International Meeting, 27-30 July*, Paper No. 032123.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (1992) *Numerical Recipes in Fortran (Second Edition)*, Cambridge University Press, Cambridge, MA.

*River Analysis System (HEC-RAS) – User's manual.* (2001). Hydrologic Engineering Center., U.S. Army Corp of Engineers, Davis, CA.

Saleh, A., Arnold, P. W., Gassman, P.W., Hauck, L. M., Rosenthal W. D., Williams, J. R., McFarland, A. M. S. (2000). "Application of SWAT for the Upper North Bosque River Watershed" *Transactions of the ASAE*, 42(5), 1077 – 1087.

Santi, C., Arnold, J. G., Williams, J.R., Dugas, W.A., Srinivasan, R., and Hauck, L.M. (2001) "Validation of the SWAT model on a large river basin with point and nonpoint sources" *Journal of the American Water Resources Association,* 37(5), 1169-1188.

Shamshi, U. M. (1998). "Chapter 11: ArcView applications in SWMM modeling." *Advances in Modeling the Management of Stormwater Impacts, volume 6,* W. James, ed., Computational Hydraulics International, Guelph, Ontario. 219 – 233.

Spruill, C. A., Workman, S. R., Taraba J. L. (2000) "Simulation of daily and monthly stream discharge from small watersheds using the SWAT model", *Transactions of the ASAE*, ASAE, 43(6), 1431 – 1439.

Srinivasan, R., and Arnold J.G. (1994). "Integration of a basin-scale water quality model with GIS" *Journal of the American Water Resour. Asoc. (JAWRA)*, 30(3), 453 – 462.

Srinivasan, R., and Arnold J.G. (1993). "Basin scale water quality modeling using GIS". *Proceedings of Application of Advanced InformationTechnologies for the*

*Management of Natural Resources*. Sponsored by ASAE. June 17-19, 1993, Spokane, WA.

Tate, E. C., Maidment, D. R., Olivera, F. and Anderson, D. J. (2002). "Creating a terrain model for floodplain mapping" *Journal of Hydrologic Engineering,* ASCE, 7(2), 100 – 108.

*State Soil Geographic (STATSGO) Database—data User Guide* (PDF format). (1995). Miscellaneous Publ. No. 1492, U.S. Department of Agriculture, Natural Resources Conservation Service (NRCS), Washington, DC.

United States Geological Service (USGS), (2003) *More on DEM's* < http://craterlake.wr.usgs.gov/dem_brief.html>

United States Geological Service (USGS), (2002) *National Land Cover Characterization* < http://landcover.usgs.gov/natllandcover.html>

U.S. Army Corps of Engineers (1996) *Risk-Based Analysis for Evaluation of Hydrology/Hydraulics, Geotechnical Stability, and Economics in Flood Damage Reduction Studies,* Engineering Regulation 1105-2-101, Washington, DC, US Army Corps of Engineers USACE.

Wurbs, R. (2001) "Assessing water availability under a water rights priority system" *Journal of Water Resources Planning and Management,* 127(4), 235-243.

Wurbs, R., Toneatti, S., and Sherwin, J. (2001) "Modelling uncertainty in flood studies" *Water Resources Development,* 17(3), 353-363.

Wright, J. (2003) *Monte Carlo Theory Review* <http://www.drjfwright.com/d/tramctheory.html#mctheory>

Zeiler, M. (1999) *Modeling Our World: The ESRI Guide to the Geodatabase Design*, Environmental Systems Research Institute (ESRI), Redlands, CA.

# APPENDIX A

**TABLE A.1.** Global codes limitations

| Feature Class | Code | Starting number | Maximum # of features |
|---|---|---|---|
| Outlet | HydroID | 100001 | 99999 |
| Reach | HydroID | 200001 | 99999 |
| Watershed | HydroID | 300001 | 99999 |
| MonitoringPoint | HydroID | 400001 | 99999 |
| PolyHRU | HydroID | 600001 | unlimited |

The listed feature classes contain the HydroID code that is being reset to the detailed starting number. Not exceeding the maximum number of features will ensure the HydroID code is unique within the Dynamic geodatabase.

# APPENDIX B

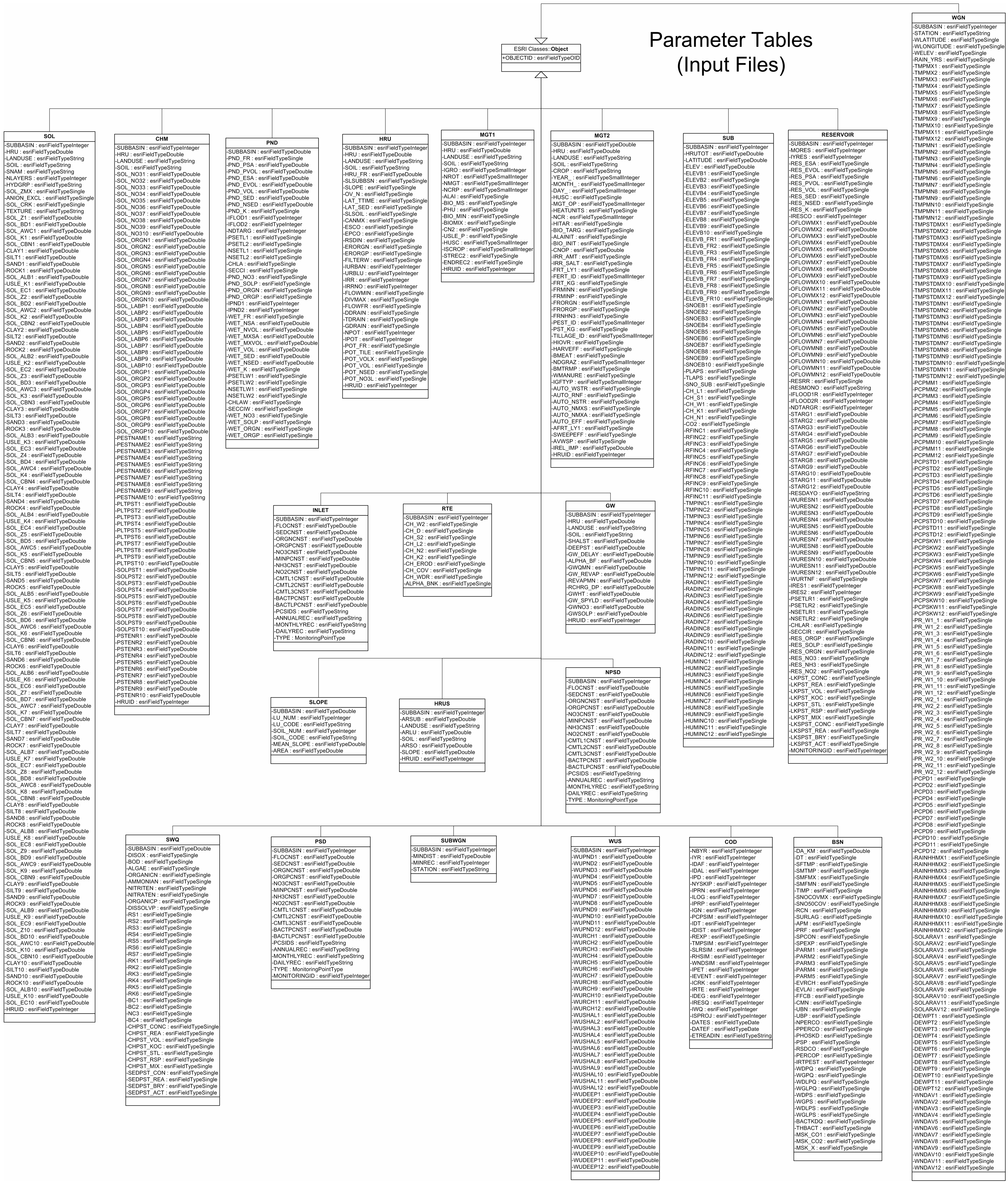



**FIG. B.1.** UML model for the SWAT input object classes of the Dynamic geodatabase.

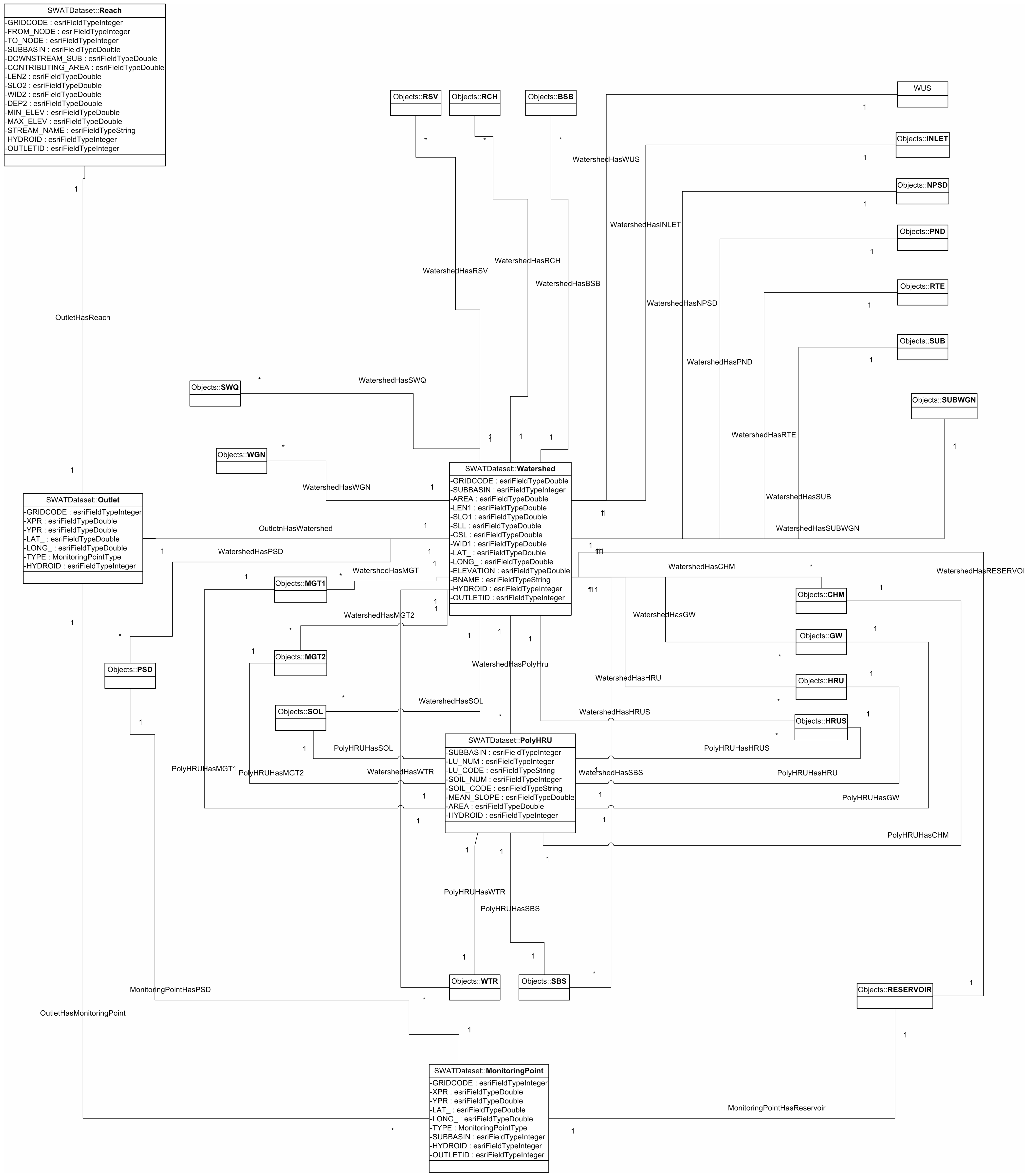



**FIG. B.2.** UML model for the relationships in the Dynamic geodatabase.

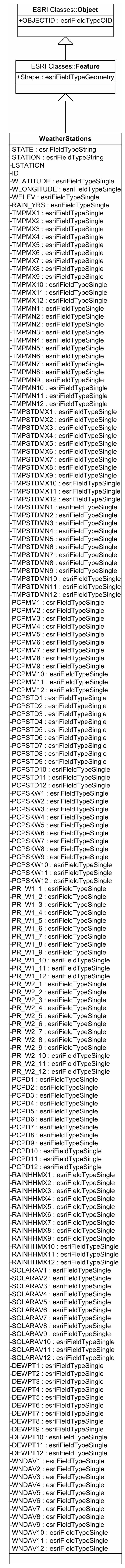



**FIG. B.3.** UML model for the US feature dataset.

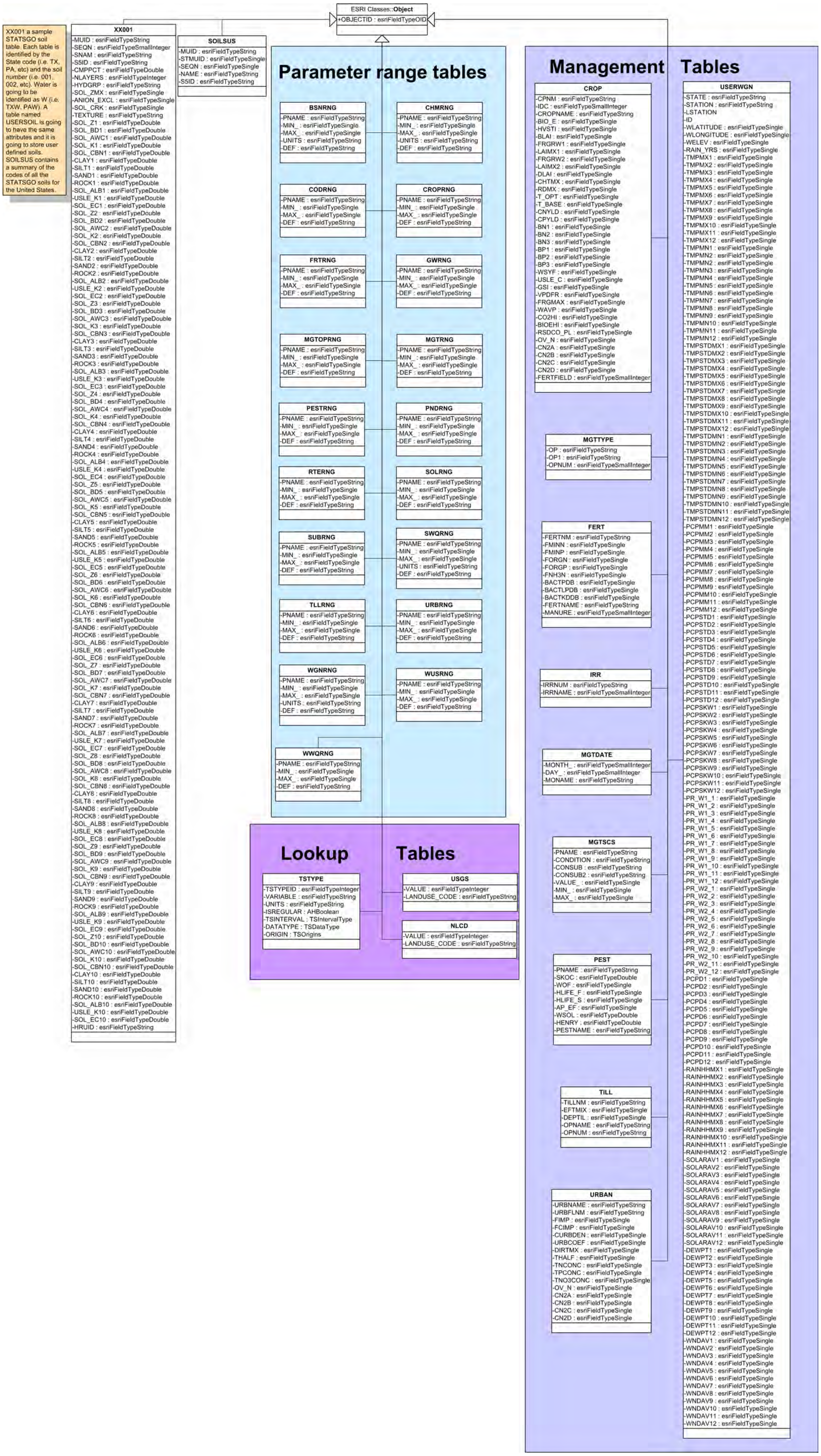



FIG. B.4. UML model for the object classes of the Static geodatabase.

# VITA

Milver Alfredo Valenzuela Zapata

**ADDRESS**

2010 Winrock Blvd. #834, Houston, Texas, 77057
(713) 266-7095
(979) 575-2428

**EDUCATION**

- Master of Science in Civil Engineering. Graduation date: August 2003
  Major: Environmental and Water Resources Engineering.
  (January 2002 – Present)
  Texas A&M University
  GPA: 4.0

- Civil Engineer, (Equivalent to an American B.S. Degree)
  (1995-2001)
  Pontificia Universidad Catolica del Ecuador (P.U.C.E.), Quito - Ecuador.
  GPA: 3.6

**EXPERIENCE**

- **Texas A&M University, College Station, Texas (Jan. 2002 – Aug. 2003)**
  Graduate Research Assistant.
  Development of GIS tools for Water Resources Engineering:
    - Watershed and Stream Delineation Tool
      http://ceprofs.tamu.edu/folivera/GISTools/wsdt/home.htm
    - SWAT interface for ArcGIS 8.x (Master's Thesis)

- **"Cantabria" building project, Quito, Ecuador (May 2001 – Dec. 2001)**
  Resident Civil Engineer. "Cantabria" Building project, Quito – Ecuador.

- **Acuatecnia Co., Quito, Ecuador (Jan. 2001 – May 2001)**
  Sales Manager. Acuatecnia Co. Quito – Ecuador

This thesis was typed by the autor.