INTEGRATION OF STREAM AND WATERSHED DATA FOR HYDROLOGIC MODELING

A Thesis

by

SRIKANTH KOKA

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

May 2004

Major Subject: Civil Engineering

INTEGRATION OF STREAM AND WATERSHED DATA FOR HYDROLOGIC MODELING

A Thesis

by

SRIKANTH KOKA

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 (Chair of Committee)	Ralph Wurbs (Member)
Clyde Munster (Member)	Paul N. Roschke (Head of Department)

May 2004

Major Subject: Civil Engineering

ABSTRACT

Integration of Stream and Watershed Data for Hydrologic Modeling. (May 2004)

Srikanth Koka, B.E., Osmania University,

Hyderabad, India.

Chair of Advisory Committee: Dr. Francisco Olivera

This thesis presents the development of a hydrologic model in the vector environment. Establishing spatial relationship between flow elements is the key for flow routing techniques. Such a relationship is called hydrologic topology, making each flow element know which other elements are upstream and which are downstream. Based on the hydrologic topology established for the flow elements, tools were developed for flow network navigation, drainage area estimation, flow length calculation and drainage divide determination. To apply the tools, data required might be obtained from different sources, which may lead to certain problems that have to do with wrong flow direction of stream lines and, mismatches in location of stream lines with respect to the corresponding drainage area polygons. Procedures to detect such inconsistencies and to correct them have been developed and are presented here. Data inconsistencies correction and parameter computation methods form the basis for the development of a routing model, which would be referred as hydrologic model. The hydrologic model consists of an overland flow routing module, two options for channel routing and a reservoir routing module. Two case studies have been presented to show the application of the tools developed.

DEDICATION

To God, my parents, brother, relatives and friends.

ACKNOWLEDGMENTS

I am thankful to my parents for supporting me at all times and for giving me a word of confidence every once in a while.

I would like to especially thank Dr. Francisco Olivera for his guidance and trust. Working under him, I have learnt a lot of things, which will surely help in my career.

I would also like to thank my other committee members: Dr. Clyde Munster 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: Ashish Agrawral, Rajeev Raina, Milver Valenzuela and Paramjit Chibber.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

LIST OF TABLES

TABLE **Page**

1 INTRODUCTION

1.1 MOTIVATION

Water resources related problems such as mapping of flood plains require realistic flow predictions. Accurate prediction of flows requires accurate representation of the hydrologic processes occurring in the system. An effective way to improve this accuracy is by employing spatially-distributed models. The advancement of computer technology and the relative ease in the availability of spatial data has made it possible to efficiently process spatial data for deriving physical parameters needed by hydrologic models. The data visualization and analyzing capabilities of Geographic Information Systems (GIS) present a convenient platform for hydrologic modeling. Currently, a number of hydrologic models are interfaced with GIS. The principal tasks in such models consist of discretization of the watershed system into units of uniform properties, extraction of hydrologic parameter information, and interfacing with hydrologic models.

There are different formats in which the spatial data are available. The raster based representation has the structure of a grid in which each cell stores the value of the property it represents. For example, in a Digital Elevation Model (DEM), cells represent elevation values. Another type of surface data representations is Triangulated Irregular Network (TIN) in which the surface is represented as a set of connected points forming

---

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

triangles. Apart from these surface based representations, the third type is the vector data. Vector data is used to represent geographic objects that have shape and size.

A hydrologic system consists of streams and their corresponding drainage areas. As these objects have shape and size, they are better represented in vector data format. This kind of data is derived from traditional paper maps, photographs or by surveys (Garbrecht et al. 2001). So, development of distributed models by employing vector based hydrologic information seems to be a good option because they represent the real world objects. Additionally, inherent topology (i.e. connectivity and adjacency information) that vectors possess and the existence of spatial relationships among them, make the vector - based environment very conducive for hydrologic modeling.

Significant amount of resources have been devoted to developing models in the raster-based environment. Raster based techniques, though attractive and simple to develop, have some limitations that have to do with the accuracy in determination of drainage patterns. More often than not, such surface models require vector data to be imposed on them to accurately delineate streams and their corresponding drainage areas (Neitsch et al. 2000). In the absence of such required vector data, high resolution raster files are needed. The resolution enhancement in raster for better capturing the landscape, however, leads to computational challenges, resulting in diminished efficiency. On the other hand, vector data file size increase is not as dramatic upon betterment of resolution, resulting in almost the same or reasonable computational speeds.

The above stated reasons have been taken into account to propose the development of a distributed hydrologic model in the vector domain.

1.2 OBJECTIVES

The objectives of this thesis can be summarized as follows:

1. Integrate stream and watershed vector data for hydrologic modeling
2. Analyze watershed and stream networks
3. Develop and implement a spatially-distributed flow routing model based on watershed and stream network data

Integration of stream and watershed data is a process of making data compatible with the model requirements such as orienting each stream line such that it flows towards the outlet on a unique path, and solving problems resulting due to improper location of stream lines with respect to their drainage area polygons. In the hydrologic model, a network of lines is used for representing streams and a set of polygons for their corresponding drainage areas. The model proposed imposes certain conditions, which the data have to cater to. Some of these conditions can be satisfied by using the techniques explained, though it not always required applying them if the data is compatible. Briefly summarizing the model data requirements;

1. The stream network should be dendritic in shape, which means that lines in the stream network should all be oriented such that the flows are directed towards the outlet of the watershed, the stream network should be free of loops i.e. there should exist a unique flow path from every stream to the outlet.
2. Each polygon representing a drainage area should be associated with one and only one outlet stream.

Inconsistencies recognized consisted of streams crossing over the drainage divides, two streams posing to be potential outlets and some other problems. The corresponding remedial procedures have been developed to treat these data inconsistencies and make the data suitable for the hydrologic model.

Once the data is made suitable for modeling, the next objective is that of determining hydrologic topology and computing of relevant network parameters for streams and sub-basins. Techniques for drainage area determination, flow length calculation and drainage divide determination are presented. Also presented are the procedures for network tracing.

The final objective of thesis was to develop a distributed hydrologic model. To capture the spatial variability in the hydrologic properties of a watershed system, soils, and land use information were intersected with the sub-basins. The result was the delineation of land units of uniform physical properties within each sub-basin called sources. Though this model does not offer a way to compute losses due to infiltration, models like Green and Ampt could easily be coupled with this model to compute excess precipitation. In a case study, to compute excess precipitation, Soil and Water Assessment Tool (SWAT) was employed. An overland flow transport model was developed using a diffusion wave model called 'Source-to-Sink', to route the surface runoff depths generated at various sources to the respective sub-basin outlets as flows. To route these flows through the channels to the watershed outlet, the Muskingum and Pure-lag method were used. Flow routing through reservoirs was achieved by employing the Storage-Outflow technique.

1.3 STUDY AREA

To evaluate the procedures developed, two case studies are presented. United States Geological Survey (USGS) has divided the United States into hydrologic units. There are 4 levels of classification. These classifications are based on the amount of geographic area covered in each unit. The first level of classification divides the nation into 21 regions. Further these regions are divided into 222 sub regions, which are further divided into 352 accounting units. Finally, these accounting units are further divided into 2150 cataloging units. Region 12 as classified by USGS has been chosen as the first study area. For this region the watershed data was obtained from fourth level classification of United States by USGS. The corresponding streams data were extracted from National Hydrographic Dataset (NHD) (USGS, 2003b). NHD is a comprehensive digital spatial dataset that contains information about surface water features in the United States. The first case study was aimed at applying and validating the procedures dealing with data inconsistencies, hydrologic topology and parameter calculation.

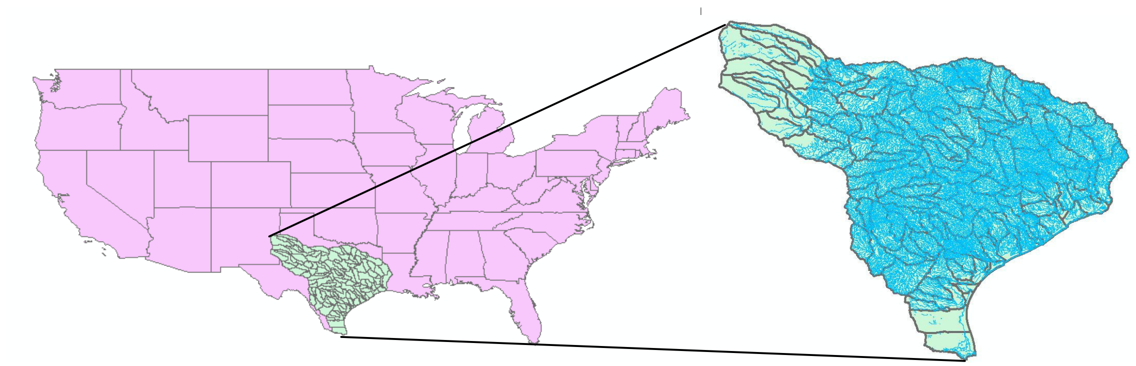


FIG. 1.1. Study Area: Region 12

A second case study is presented to show the application of hydrologic topology, network parameter computation techniques. The parameters thus obtained were then used as input for the hydrologic model. Bull Creek watershed upstream of USGS gauging station number 08154700 located near loop 360 in North Austin, Texas, was chosen. The drainage area of the watershed at the gage station is 22.3 sq.miles. The corresponding stream and watershed data were delineated from 10m DEM obtained from USGS (USGS, 2003a). Soils data STATSGO (State Soil Geographic Data Base) was obtained from United States Department of Agriculture (USDA, 2003). Land Use information was obtained from USGS, available as National Land Cover Data (NLCD). 15 minute gage-adjusted radar rainfall estimates were obtained from NEXRAIN

Corporation. The datasets mentioned in this section will be elaborately described later in section 4.

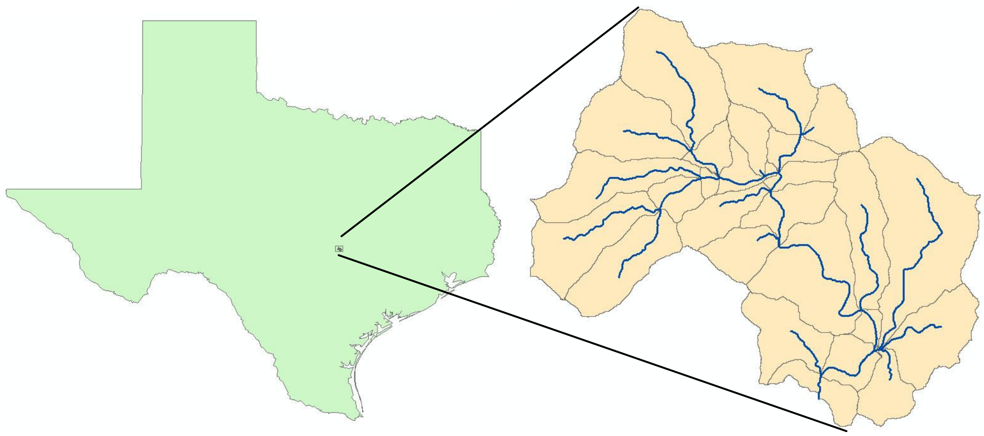


FIG. 1.2. Study Area: Bull Creek Watershed

1.4 OUTLINE

The thesis consists of 6 sections. Section 1 presents the motivation, objectives, description of the case studies and a brief outline. Literature review which explains some of the methods previously developed and which are related to this thesis constitutes section 2. Section 3 gives an overview of various concepts and terminology related to GIS, GIS Data models, Component Object Models (COM). Various algorithms developed to treat the input data, procedures related to hydrologic network parameter calculation, overland and channel flow routing technique descriptions constitute section 4. Section 5 describes the application and results of the models by employing two case studies. Section 6 consists of conclusions.

2 LITERATURE REVIEW

This section is divided into two parts. First part sheds light on hydrologic models of relevance to this thesis and the second part discusses about some GIS based methods for deriving hydrologic parameters needed for the hydrologic models.

2.1 HYDROLOGIC MODELS

Hydrologic models in general can be classified as lumped models or distributed models. In lumped models, spatial variability in hydrologic parameters or meteorological related data are not accounted for, meaning are averaged or assumed uniform over the system, whereas, in distributed models spatial variability is explicitly accounted by assuming uniformity over smaller modeling units by sub dividing the bigger system based on physical properties. In most of the distributed hydrologic models, these units are delineated by combining topography, soil properties, land use properties and other pertinent properties. Distributed models are especially useful, for example, when impacts of land use change are to be studied or for analyzing spatially varying flood responses. As the topic of distributed modeling is of importance to this thesis, a discussion of related back ground is provided.

2.1.1 HEC-HMS (Hydrologic Modeling System)

HMS is a comprehensive hydrologic model developed by Hydrologic Engineering Center (HEC) of United States Army Corps of Engineers (USACE). It is an event – based overall lumped model (HEC, 2000). HMS offers several options to model various physical processes occurring in a watershed system. One such process is the direct runoff computations. Most of runoff models available with HMS are lumped in nature except for two which are distributed. Most of the lumped runoff models derive their roots from the Unit Hydrograph (UH) concept.

This model provides a lumped model option called Clark's UH. To overcome its lumped character, a modified version called ModClark method was developed for HMS (Daniel and Arlen 1998). ModClark's method requires that watershed be further divided into sub-areas by intersecting it with a grid. Each of these sub-areas is assigned individual lag time, instead of one value for the whole watershed, as in the case of Clark's UH. The precipitation excess at each sub-area is transported to the watershed outlet using the corresponding lag time. Thus the inflow contributions due to all the sub-areas to linear reservoir are computed. These flows are then routed through a linear reservoir (only a single value for storage coefficient being defined for all the sub areas) to obtain the hydrograph at the outlet, which will later be routed through the channels. ModClark's technique though tries to overcome the lumped character of Clark's UH, has certain limitations. The limitations that have been recognized are:

1) It is not a physically based model.

2) Dispersion effects that occur along the flow path are not captured, because the model assumes one single linear reservoir for the whole watershed.

2.1.2 SWAT (Soil and Water Assessment Tool)

SWAT was developed by USDA-ARS (Agriculture Research Service). Prediction of impacts of land management practices on water, sediment and agricultural chemical yields was the aim for the development of this physically based model. SWAT simulates complete hydrologic cycle of a watershed system. "The hydrologic cycle is simulated in two phases: land phase and routing phase. The land phase hydrology controls the amount of water, sediment, nutrient and pesticide loadings. The routing phase consists of defining the movement of water, sediments, etc through the channel network of the watershed" (Neitsch et al. 2000). As overland flow computation of the land phase is of importance to this thesis, a discussion of the techniques employed in SWAT is provided.

To capture heterogeneity in physical properties the basin is divided into sub-basins, each one of them corresponding to a stream and then the sub-basins are discretized into sub-areas called HRUs, which are unique in terms of soils and land use. Physical processes modeled in SWAT can be classified based on the scale (i.e. sub-basin or HRU) assumed for lumping. For example, the precipitation input in SWAT can be provided only at the sub-basin level, whereas, runoff computations can be done at HRU level. The Green-Ampt infiltration method is one of the options that this model offers to compute excess precipitation at the HRU level, the other one being NCRS curve number method. Though the excess precipitation values are computed at the HRU level, an average of excesses over the sub-basin is assumed for overland flow computations. The overland flow routing model assumes a linear reservoir scheme. The equation corresponding to this scheme is given by:

$$Q_{surf} = \left(Q'_{surf} + Q_{surf,i-1}\right).\left(1 - \exp\left[\frac{-surlag}{t_{conc}}\right]\right) \qquad (2.1)$$

Where, Q_{surf} is the amount of surface runoff discharged to the main channel on a given time step (mm H_2O), Q'_{surf} is the amount of surface runoff generated in the subbasin on a given time step (mm H_2O), $Q_{surf,i-1}$ is the surface runoff stored or lagged from the previous time step (mm H_2O), $surlag$ is the surface runoff lag coefficient, and t_{conc} is the time of concentration for the subbasin (hrs).

The runoff model incorporated in SWAT is lumped at the sub-basin level, because it computes an average value for spatial varied surface runoff. This is one of the limitations of the SWAT model. To take the advantage of the distributed nature of surface runoff, this thesis explains a methodology that could be extended to this model, which will enhance the capabilities of handling overland flows. Such an extension could make the SWAT model work at any time step as assigned by the user, noting that presently SWAT model works only at daily time step. Another advantage that can be derived by employing a distributed model is quantifying the effectiveness of Best Management Practices (BMP). BMPs are a set of structural and non-structural measures suggested by EPA (Environmental Protection Agency) for controlling non-point source pollution caused due to urban storm runoff discharging into rivers. Modeling such control measures requires that the surface runoff model with embedded water quality model be distributed in nature because the BMPs can be spatially distributed themselves with locations being not on the streams. Literature on how to extend a distributed model to SWAT for capturing BMPs has been explained in the methodology section but has not been implemented as a part of this thesis.

2.1.3 Source-to-Sink Method

Unit hydrograph method has been developed to characterize a watershed's response to unit instantaneous impulse (unit depth of excess precipitation, uniformly distributed). A technique known as source-to-sink algorithm (Olivera et al. 2000) has been developed to extend the concept of unit hydrograph to sub areas formed by discretization of watershed. Response functions to unit impulse of sub areas are defined individually based on spatially distributed flow velocity and dispersion coefficient parameters. Then the responses for excess precipitation depths per sub area are convolved at the sub-basin outlet. The responses of all the sub areas are then summed up to obtain a hydrograph. This is the overall procedure employed by this technique. The main advantage of this model is that, excess precipitation generated at each sub-area can be routed to any point of interest in the watershed. This character makes it very attractive way for modeling BMPs.

Published literature (Olivera et al. 2000; Olivera and Maidment 1999) suggests that the Source-to-sink method had been applied for areas using their surface representations such as DEMs. In this thesis, instead of using raster cells, sub areas (polygon vectors) defined as land units of uniform properties (combination of land use and soils per sub basin) have been employed for overland flow routing.

2.1.4 Others Methods

Wang et al. (2002) developed a mathematical overland flow routing model for simulating transport of water, sediment and other pollutants on hill slopes using Saint-Venant equations. In this method the hill slope is divided into several sub planes. Each sub plane is associated with a set of first order non-linear differential equations, which represent continuity and storage release equations required to relate inflows and outflows. Chezy's equation parameters are then used to derive analytical solutions to these equations.

Liu and Todini (2002) present a comprehensive physically-based rainfall-runoff model which is the advanced version of TOPKAPI to study the impacts of land use and climatic changes on the hydrological regimes in the river basins. This model is based on lumping of kinematic wave assumption in the soil, on the surface and in the drainage network. The parameterization for this model is obtained from DEM, soil map, and land use map. The model is structured around 5 modules to represent evaporation, snow melt, soil water, surface water and channel water components. A non-linear reservoir model is assumed for overland flow for each grid cell. The kinematic wave approximation for overland flow combined with Manning's formula is used to solve the non-linear equations.

The benefits of using GIS spatial analysis capabilities to derive parameter coupled with a hydrologic model to solve water resources management problem is presented by Xu et al. (2001). The hydrologic processes are simulated by discretizing the watershed based on topography, soil and land use properties. These blocks of uniform land units are conceptualized as tanks: surface tank, sub-surface tank, ground water tank and river tank. Each of these blocks is assumed to have a prescribed slope and direction while simulating overland flow. A simplified kinematic wave model combined with Manning's equation is used to compute flows at the entry of the channel. River tanks then route these flows to produce a hydrograph at the pertinent points of the watershed. Finite-difference methods are employed to solve the governing equations over the gridded representation of the land surface and input.

There have been numerous distributed models based on the premises of modeling hydrologic processes by adopting tank concept. In such models, spatially varying information such as topography, soils, land use and meteorological data are required. Such heterogeneity in models can be accomplished by using remote sensed data. Biftu et al. (2001) present a hydrologic model that is semi distributed and physically based. Using remote sensed data the watershed is discretized into cells of uniform properties. The overland flow routing technique model is based on the assumption of kinematic wave approximation. A unit impulse is applied on each cell and the individual response function is computed using the kinematic wave approximation. The flow is routed from cell to cell based on eight possible directions. After computing response functions for all the cells, an average response function of the sub-basin is derived which is used to compute the lateral inflows into the channel. Muskingum-Cunge flow routing method is used to route the flows through the drainage network. Gainnoni et al. (2003) also employed remote sensed data in their hydrological model. The main purpose of this model was to compute discharges at any location on the drainage network for analyzing flood responses. The surface runoff was assumed to move over the hill slopes with overland velocity and then through the channel with channel velocity. These land phase velocities can then be used to compute flows at any location on the drainage network, at any time.

CASC2D was developed for superior predictions of runoff and stream flows compared with the standard lumped parameter model like HEC1 (Charles et al. 2000).

This model assumes Hortonian runoff predictions techniques. Surface runoff that is spatially distributed is routed to the channels using a 2-dimensional diffusion wave overland flow model, the velocities for which are computed using Manning's equation. The inflows to the channels are then routed to produce outflows by solving 1-dimensional diffusion wave model. Though this model offers a strong physical base for hydrologic modeling, can be applied only to areas where runoff production mainly assumes Hortonian patterns. A model which is based on 1 dimensional kinematic wave approximations has been developed to simulate both Hortonian and Non-Hortonian types of runoff predictions.

2.2 GIS – HYDROLOGIC PARAMETERS

Automating the process of watershed characterization has been made possible by the availability of geospatial data and software systems that process such data. A number of interfaces have adopted an automated approach to interface with hydrologic models by deriving input information. One way of classifying them can be based on the data structure employed to represent the relevant terrain and other hydrologic information like soils and land use. The first type of data structure which is most explored, is the raster. In a raster model, land surface is represented as a matrix of cells, in which each cell represents an elevation value (e.g. DEM). These Gridded structures have also been used to represent spatial information like soils, land use etc also. The second type of data structure and the one explored in this thesis project is the vector-based representation of the hydrologic elements. The geospatial datasets that assume the vector data structure represent the hydrologic elements as features (streams as lines and watersheds as polygons) with associated tabular information (attribute tables) for feature properties. With this back ground about the data models, discussions on some of the techniques previously developed which relate to hydrologic parameter extraction are discussed.

2.2.1 DEM-Based Methods

Since the time Jenson and Domingue (1988) have developed the D-8 algorithm, it has been incorporated in a number of watershed parameterization and hydrologic models. The concept of this method is that each cell in a DEM is assumed to flow to one of the eight neighboring cells according to the direction of steepest slope. Though a number of other flow direction determination methods that are not based on D-8 algorithm have been developed, because of its simplicity, the D-8 algorithm has been employed in a number of DEM-based models. Among the noted ones is the Watershed Delineator developed by ESRI which can be used for delineating streams and watersheds (Djokic et al. 1997). Later, based on the same D-8 algorithm, CRWR-PrePro was developed at Center for Research in Water Resources (CRWR), to create input files for HEC-HMS (Hellweger and Maidment 1999, Olivera 2001). The capabilities of this model like terrain analysis, topologic analysis, watershed delineation helps create basin model for HEC-HMS models (Olivera 2001). HEC-GeoHMS is a preprocessor similar to

CRWR-PrePro (HEC 2000). AVSWAT, an ArcView interface developed with the aim of creation of input files for SWAT, whose watershed delineation function is based on D-8 algorithm. This model also incorporates in itself a parameter calculation function (Neitsch et al. 2000). To overcome the limitation of D-8 algorithm like flow direction being restricted to one of the eight possible directions, Tarboton (1997) explained a method to determine the flow direction based on a single angle among the infinite possible directions, which was called $D\infty$ method. Several other researchers came up with alternatives for determining flow directions (Costa-Cabral.M and S. Burges, 1994, Jenson, 1991).

2.2.2 Vector-Based Methods

This is one environment which has not been explored as much as raster. Among the models developed for computing hydrologic topology based on vector domain, the significant ones include those built on Pfafsetter system for numbering watersheds. Verdin and Verdin (1999), later Furnans and Olivera (2001), explain watershed classification system based on Pfafstetter numbering method to identify the watershed downstream of each watershed to aid in understanding the topological relationships between a basin of interest and other local basins. Such relationships are helpful to know which basin affects which basin. Neighborhood information such as the one discussed could be helpful in tracking constituent particles caused by point and non-point source pollution. So, to find a source of pollution or to model the travel path of pollution, network tracing techniques are needed. Network tracing capabilities come with the ESRI's GIS software. The ArcGIS (system of software products) come in three different versions: ArcView, ArcEditor and ArcInfo. ArcEditor includes all the capabilities of ArcView and additional functionalities. Similarly, ArcInfo is fully developed of all the versions i.e. it has capabilities of the other two. The reference to the versions of ArcGIS software was provided because the network tracing tools that come with ArcGIS work only with ArcInfo. These methods come as extensions, namely Utility Network Analyst (MacDonald 2001). The ArcInfo version is the expensive of all the ArcGIS versions. So to avoid the extra cost incurrence and complexities involved in constructing the geometric network, it is very useful to develop such methods for analyzing the vector data in the ArcView version. In an attempt to answer this problem, a set of tracing tools has been developed.

Lot of exploration has been conducted in the developing raster based methods. Compared to raster, the vector domain has not been fully exploited given that the vector represents the real world flow element. Some of the advantages of modeling with vector compared to raster are discussed here. Stream meandering is a common occurrence in natural streams. In order to capture such properties of streams, high resolution DEM would be required. In the event that high-resolution datasets are required, vector datasets have an advantage because they would present the same amount of information that raster would, but the vector file size increase is not as significant as that of raster. The evolution of DEMs suggests that there has been a significant increase in the resolution of the datasets and at the expense of unmanageable file sizes. They have evolved from a resolution of 3 arc-seconds (i.e. approximately 90m) in 1990s to a resolution of 10m. Although the raster based terrain models can process the DEMs for delineating the streams, some of them have the option called Burning-In-Streams, where vector streams are burnt into the DEM so that streams delineated from DEM match the real streams (Olivera 2001). This suggests that the streams delineated from a DEM may not be representing the real streams. In such a scenario it is rather better to model with vector data. So, use of vector data as oppose to raster data, has the advantage that each element represents a real-world flow element and, consequently sets as better ground for physically-based modeling, not to mention it is overall accurate.

3 CONCEPTS AND TERMINOLOGY

This section discusses the concepts and terminology related to GIS (Geographic Information System) that are used in this thesis. The descriptions are grouped into three sections, namely, GIS, types of data models (sometimes referred as data structures) and COM (Component Object Model).

3.1 GEOGRAPHIC INFORMATION SYSTEMS

GIS is a system of software used for storage, retrieval, mapping, and analysis of geographic data. Though GIS has evolved through last four decades, much of the impetus on development of GIS was given only in last 2 decades. In 1969 the Environmental System Research Institute (ESRI) was founded by Jack Dangermond, for developing software for GIS. ESRI's GIS, which started as a command line interface has evolved into user-friendly Windows applications. One of the older versions of desktop GIS, called ArcView1.x has displaying and printing capabilities only, whereas their future counterparts like ArcView 2.x and ArcView 3.x had additional capabilities such as limited data analysis, data development and customization capabilities. The latest versions of desktop GIS software are categorized under ArcGIS. These versions are built based on latest software technologies like Component Object Models and relational databases using RDBMS (Relational Database Management System). ArcGIS is classified into 3 versions: ArcView 8.x, ArcEditor 8.x, and ArcInfo 8.x. The classification of ArcGIS has been based on the number available functionalities that a version possesses. ArcInfo and ArcEditor have more capabilities compared to ArcView. Each of these versions contains three components: ArcMap, ArcCatalog and ArcToolBox. ArcMap is used for viewing and analyzing spatial data, ArcCatalog can also be used for viewing the spatial data but is mainly used for management of spatial data and ArcToolBox is used for data conversion and also management of spatial data. The following flow chart diagram gives an idea of the classification and components of ArcGIS.

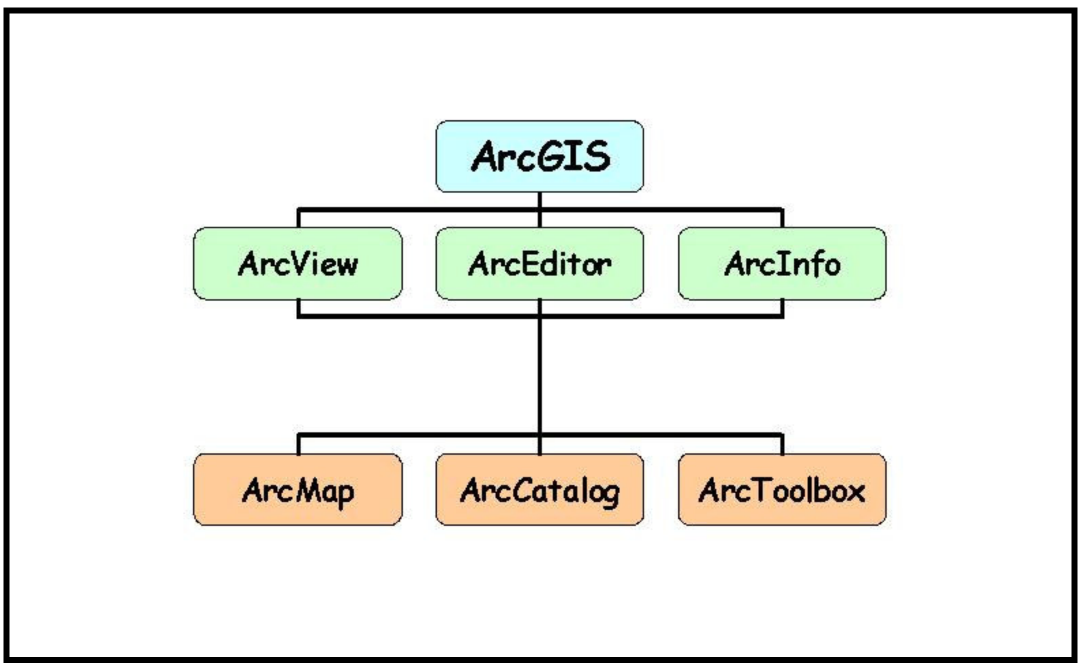


FIG. 3.1. Classification of ArcGIS

3.2 DATA MODELS

Although there has been a little discussion about the GIS data models in the literature review section, a more detailed discussion is provided here. Data models are a way to store and manage geographic data. The choice of data model depends on the type of geographic elements being represented. Discrete geographic objects like buildings, rivers, roads, etc., that have size and shape, are usually represented using Vector data model, whereas continuous geographic features like elevation, population density, etc., are represented using surface models like Rasters or TINs (Triangulated Irregular Networks). Because vector data is of importance to this thesis a detailed description of vector is given.

In a vector data model, geographic objects are represented as features that have shape and size. The three basic types of vector shapes are points, lines and polygons, although there are variations of these types. For example, rivers and roads are represented as line features, buildings are represented as polygon features and cities are represented as point features or, in some cases as polygons. Apart from geometry, each feature can also be attributed with non-spatial information. Only features of one shape type can be collected together for storage. These storage types can be classified as file-based storage (e.g. shapefiles and coverages) or DBMS (Database Management System) (e.g. geodatabases). In coverages, not only is the geometry of the feature stored, but also the connectivity and adjacency (also called Topology) is explicitly stored, the associated data structure is called topologic data structure. Shapefiles and geodatabases do not explicitly store topologic information, owing to cartographic data structure. The type of data structure exploited in this thesis is the cartographic data structure. Before going into further discussion about data structures, a brief discussion on geodatabases is provided. In geodatabases, features can be grouped into feature classes. Each feature class contains features of the same shape type. Further, feature classes with the same extent and map projection can be grouped in a feature dataset. Feature datasets can only have feature classes, relationship classes but not tables. Tables (Object classes), do not contain shape field and should always be stored outside feature datasets. Feature classes are special type of object classes containing shape type. Relationship class contains relationship information among the classes of the geodatabases. Relationship classes are established using key fields. With this background on various types of data structures, discussion of various feature types would be presented.

From now on, shapefiles, feature classes, etc. would be referred to as datasets. Point datasets are a collection of points representing geographic objects that can be treated as points in space. Each point feature in point dataset contains a set of X, Y coordinates. Stream gauging stations, for example, can be represented as points. This is an example of point dataset, whereas geographic elements like rivers can be represented as linear features. A line feature in a dataset is a straight line segment connecting two points. Usually not all geographic elements are straight, so they are represented using a polyline. A polyline can be described as sequence of points connected by line segments, the end points are called nodes and the intermediate points are called vertices (see Figure 3.2). Polylines also have fixed orientation, i.e. they flow in one direction. Usually, orientation depends on the direction of digitizing a feature. The node at which polylines start is called FromNode and the point at which they end is called ToNode. Streams of a river system are represented as polylines in geospatial datasets. These polylines that would represent streams would be referred to as stream lines from now on for convenience sake. In a dataset all the stream lines are collected to represent a river network. Stream network representing natural river system usually have a tree structure. Such networks are called Dendritic Networks. In dendritic networks, the stream lines flow towards a single outlet point (called sink). The following figure shows an example of a general dendritic network and the corresponding table shows the attribute information about identification number for each feature and corresponding from node and to node information.

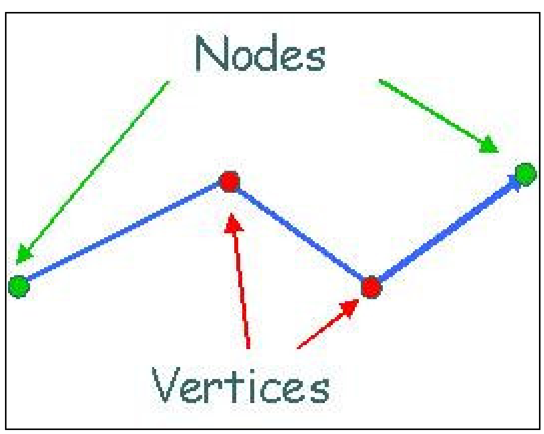
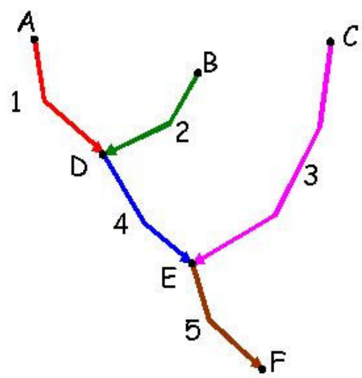


FIG. 3.2. Polyline Schematic



Link	From-Node	To-Node
1	A	D
2	B	D
3	C	E
4	D	E
5	E	F

F is the network sink.

FIG. 3.3. Example Dendritic Network

In river systems, meandering is a common occurrence. The increase in length of a stream caused by meandering is not accurately captured in geospatial datasets due to resolution limitations. To account for the inaccuracies in the measurement of other feature properties such as length, each feature could be attributed with a correction factor called Linear Weight.

Apart from the streams in a river system, their corresponding contributing areas could be represented as polygons in GIS. Polygons can be imagined as a set of points connected by a sequence of lines, similar to polylines, but their from-node and to-node coincide forming a regular or irregular shape that has a property called area. The polygons representing drainage areas of streams are referred to as sub-basins in this thesis and the complete set of sub-basins would be referred as basin. Each sub-basin polygon can correspond to one or more than one streams but not vice versa. To account for spatial patterns of precipitation, each sub-basin polygon could be assigned a property through attribute tables called Area Weight.

3.3 ARCOBJECTS, COM AND CUSTOMIZATION

Deriving hydrologic related information from digital spatial datasets needs spatial analysis. These analyses can be performed using the functionalities available with GIS software, but some times the analysis process may be too complex or may need to be repeated and it is better to develop custom functionalities. This process is called Customization. A technology framework known as ArcObjects is a set of components that can be used for customizations (achieved by programming). ArcObjects is based on COM (Component Object Model) technology developed by Microsoft. According to Microsoft 2000, COM is a standard that enhances software interoperability by allowing different software components to communicate directly. COM representations are usually done through the use UML (Unified Modeling Language) (Ungerer and Goodchild 2002). There are several programming options for customizing GIS using COM-compliant languages like Visual Basic, Visual C++ or Delphi. The tools for this thesis project were developed in Visual Basic. In this thesis, COM based programming was used to interact with the GIS software and for some spatial analysis.

4 METHODOLOGY

This section deals with the descriptions of methodologies developed to accomplish the objectives of the thesis. The discussions move in a logical sequence starting from the procedures that process raw data, to the procedures for hydrologic modeling. The methods for processing raw vector data consists of those that deal with making data consistent with model requirements which were explained previously in the introduction section. These data requirements would be explained in detail when the procedures that handle them are presented. Once the model data compatibility is achieved, the data would be suitable for deriving the physical parameters and for hydrologic modeling. The procedures that were developed for determining hydrologic topology and for computing physical parameters are explained in the subsequent sections. Also included in this section are the methods for tracing upstream and downstream in a network. The data that includes the geometries of flow elements, accompanying parameters derived along with the meteorological data are used for hydrologic modeling. Descriptions of methods incorporated in the development of hydrologic model are explained next. This part includes explanation of techniques for discretization of watershed into smaller land units based, for overland flow routing, for channel routing and for reservoir routing. The techniques are explained in a general manner and are supported with examples if necessary, to make understanding easier.

4.1 PREPROCESSING GIS DATA

4.1.1 Dendrification of Stream Network

The basis for the concept of routing water in hydrologic models has been the idea that water at every point in the watershed flows towards the outlet. In vector stream networks, flow direction of a stream is defined by its orientation. When stream vector data are extracted by methods such as digitization of paper maps, the direction of digitization is very important. This is so because, the direction of digitization defines the flow direction of stream vectors, which is called Line Orientation. In the NHD (National Hydrographic Datasets) stream data, it has been observed that some of the digitized lines have a direction opposite to that of their natural flow. The methods for detecting such disorientations and the corresponding correction methods have been explained. The correct orientation of stream lines is one part of the dendritic shape requirement, the other being the stream networks be free of loops. The occurrence of loops in stream datasets has been observed in NHD. The following figure shows an example of a non-dendritic network, in which red colored line represents the outlet stream; blue colored lines represent the streams that are not part of loops, green colored lines represent the streams that are part of loops. In figure 4.1, arrows show the line orientations and the arrows that are highlighted by circles are incorrectly oriented.

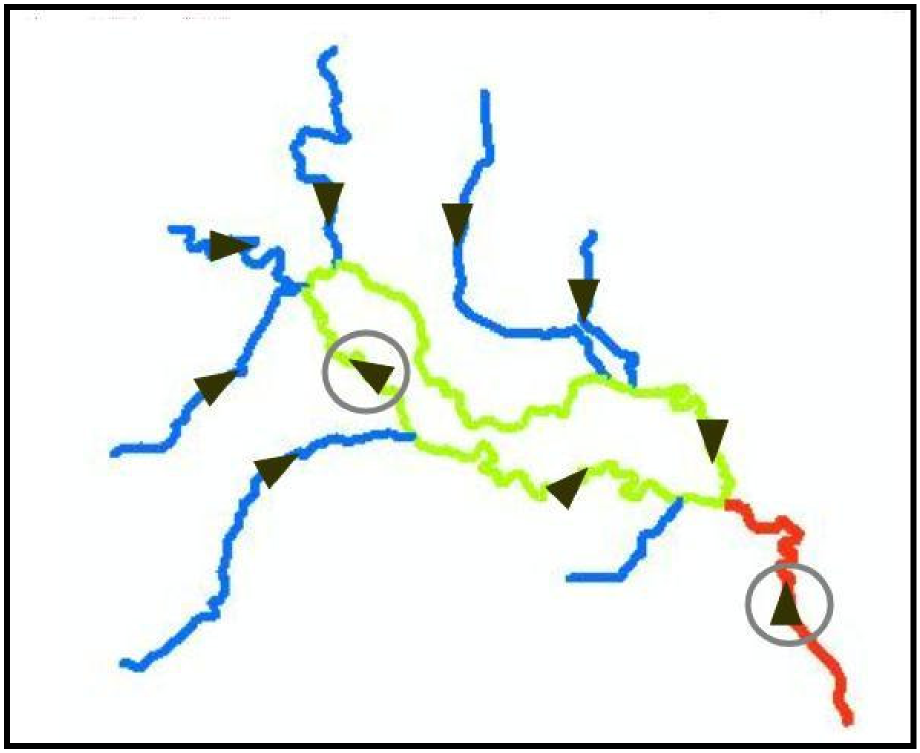


FIG. 4.1. Non-Dendritic Network

'Dendrification' is a term coined in this thesis to refer to the process of converting a non-dendritic network into a dendritic network. By converting a non-dendritic stream network into a dendritic stream network, the lines in the network are oriented in such a way that they flow towards the outlet and the loops are removed so that flow between any two nodes in the network has a unique path. Note that non-dendritic networks can also be referred as Graphs and dendritic networks as Trees.

The procedures for dendrification basically rely on navigation of the network in a bottom-up fashion (i.e. starting from the outlet stream of the network, visiting all the elements (stream lines) in the network above it). While navigating in the network, certain decision schemes are applied for correcting the line orientation and for removal of loops. As each line in the network will be associated with an identification number, a list of identification numbers is maintained to keep track of all lines visited at every instance of execution of the dendrification algorithm. This algorithm works as two processes: the first one deal with correction of line orientation and loops, and the second deals with navigation. The algorithm starts at a user specified outlet stream line on the map. The outlet line's orientation is then checked and corrected if needed. The check for orientation is accomplished by examining its to-node. If the to-node is shared by any other line in the network means that its orientation is wrong, otherwise it is correctly oriented. Once the outlet has been treated for orientation, its identification number is added to the list. The elements in the list are processed one-by-one in the order they are added. To start with, the first element in the list, which is the outlet, is checked and marked as 'Parent' for the first iteration. Knowing the identification number of the present parent, a spatial query is applied to extract the corresponding feature from the map. Once the feature is found on the map, a spatial query is applied again on its from-node to find out which other features in the map share that node. The newly discovered features are marked as 'Children' and checked for their line orientation. The check for orientation of the children is done by comparing their to-nodes with the from-node of the parent. If the to-node of a child is same as the from-node of the parent, the child is correctly oriented (does not need orientation correction), otherwise is incorrectly oriented, and it is flipped on the map. This check is conducted on all the children just discovered and their identification numbers are added to the list. Similarly, subsequent iterations are done by moving element-by-element in the list. From the second iteration onwards, the check for line orientation is done in the same way as explained for the first iteration, but the check for loops is also applied. The existence of loops can be inferred by the following logic. If an element existing in the list is rediscovered by a new parent, the presence of a loop in the network is confirmed. Before discussing on how loops are tackled, it is important to describe the two types of loops that occur in non-dendritic networks. The first type of loops will be referred to as 'Balanced loops' and the other one will be called 'Unbalanced loops'. Balanced loops occur when a feature is rediscovered in the network and its to-node is shared by both the parents. For an example of balanced loop, refer to figure 4.2 (red colored line refers to child and blue lines refer to parents). To remove such kinds of loops, the parent that exits on the longest flow path, computed from the child to the outlet is deleted, while maintaining the connectivity of elements in the network. Unbalanced loops are those in which, a feature rediscovered shares its from-node with to-node of one parent and shares its to-node with from-node of the other parent. In such cases, the child is deleted from the network to remove the loop, while still maintaining the connectivity of the elements in the network. For an example of unbalanced loop refer to figure 4.3 (red colored line refers to child and blue lines refer to parents). Thus at each iteration checks for orientation and loops are done. Iterations continue until an element is reached such that it has no children and is the last element in the list.

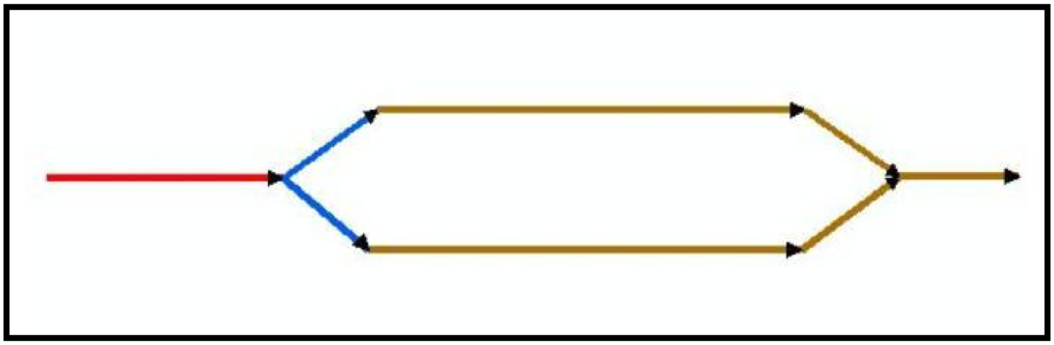


FIG. 4.2. Balanced Loop

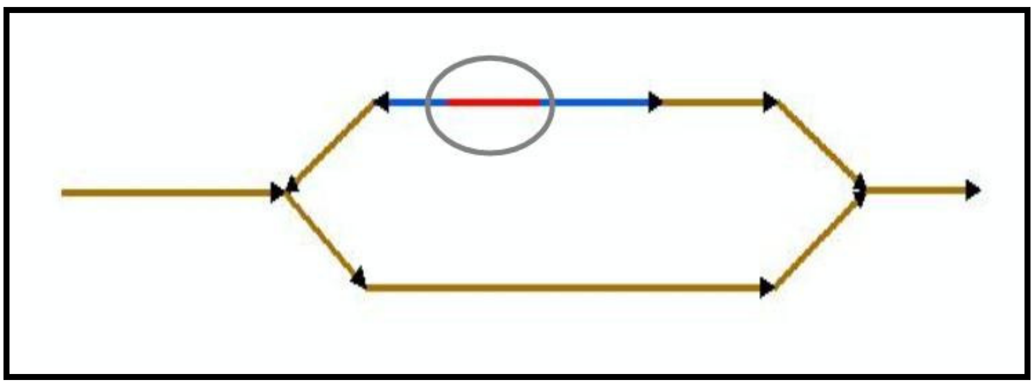


FIG. 4.3. Unbalanced Loop

The end product of application of the dendrification method is a network that is dendritic, in which all the lines flow towards the outlet, also making the network free of loops.

4.1.2 Intersection and Building Topology

To capture the heterogeneity among various parameters involved, when modeling a watershed in a distributed manner, it is required that the watershed system be represented with a set of streams and a set of corresponding drainage areas (called sub-basins) in which each stream can be associated with a sub-basin. To make each line belong to a single polygon, the stream network should be intersected with the polygons. Intersection of lines with polygons results in a new dendritic network, in which the lines are split at the polygon boundaries in such a way that each resulting segment lies within one polygon only. During this process, the lines in the new network are attributed with an identification number of the sub-basins they are contained by. The inbuilt functions for available with ArcGIS software packages, have been used to automate intersection. (ESRI 2003).

Apart from location based association among flow elements (i.e. between streams and sub-basins), the neighborhood based association among flow elements (between streams) is important, as flow routing requires transport of water from one point in the watershed to the other through the streams. The neighborhood association consists of attributing each stream in the network, with the corresponding downstream and upstream line information, so that every stream knows which other streams are flowing into them and to which streams they are flowing into. It should be remembered that in a dendritic network, each line can have only one downstream line, but can have more than one upstream line and, all the lines are connected through nodes. Taking advantage of streams sharing nodes makes it easy to attribute them with their neighborhood information. In order to attribute streams with their neighborhood information, each line in the network is attributed with their upstream and downstream node numbers. The process of attributing the lines with upstream and downstream node numbers is called building topology. The functionalities for building topology are not available with commercial GIS software packages, but the procedures for doing the same have been obtained from ESRI (ESRI 2003).

4.1.3 Treatment of Data Inconsistencies

When modeling in the vector environment, data required may have to be obtained from different sources. For example, stream data may be obtained from NHD and watershed data may be obtained from HUCs (USGS 2003d). Since for hydrologic analysis, it is necessary to compute physical parameters, streamlines are overlaid on the watershed polygons to establish location based relationship between them, i.e. making streams know as to which sub-basin they belong to. In order to achieve the location based association between streams and sub-basins, lines are intersected with polygons. The discussion on intersection was provided in the last paragraph. When the lines are intersected with the polygons, certain problems might arise, what will be called 'Data inconsistencies' here. These inconsistencies consist of streams crossing over the divides, two lines posing to be potential outlets of a sub-basin, etc. The reason that could be attributed to these inconsistencies is the data available at different sources are developed at different scales. In case of NHD and HUCs data, the NHD data was developed at 1:100,000 scale, whereas HUCs are available at a scale of 1:250,000 (USGS, 2003d). To understand the type of inconsistencies that might occur, an overlay of NHD streams data and HUCs data has been studied and the following types have been recognized:

Case 1: Small upstream parts of some streams cross over the drainage boundaries (Refer to Figure 4.4).

Case 2: A streamline outletting into a sub-basin, after some distance downstream, flows back into the sub-basin where the flow started. (Refer to figure 4.5).

Case 3: Two or more streamlines offer to be potential outlets of a single sub-basin. (Refer to Figure 4.6).

Case 4: Streamlines outletting into wrong sub-basins. (Refer to figure 4.7).

During the intersection and building topology, streams are attributed with their from-node (i.e., upstream node) and to-node (i.e., downstream node) identification numbers and are also attributed with sub-basin identification numbers. Due to inconsistencies between lines and polygons, as mentioned in the previous paragraph, some of the lines in the network may be attributed with incorrect polygon identification numbers. In order to detect and correct these inconsistencies, concerning methods are presented in the next few paragraphs. These methods rely basically on segment's knowledge of their downstream segment number and, on polygon's knowledge of their potential outlet segment numbers. The term potential outlet segment numbers is used because, as mentioned previously, due to inconsistencies between stream and basin data, more than one stream poses to be potential outlets of a sub-basin, which needs to be corrected. The procedures developed to derive such information form the first part of the algorithm dealing with the data inconsistencies. In this part, the algorithm iterates through all the segments in the network. During an iteration a segment is chosen randomly, its downstream node number is read and a query is performed on the upstream node column of the network table, such that a segment is found whose upstream node number is same as the downstream-node number of the segment under question. The number of the segment found is then attributed to the segment under question as downstream segment number. This process is repeated until all the segments are processed. During these iterations information needed for polygons is also derived. Due to inconsistencies, as each polygon can have more than one potential segment as its outlet, a collection (called ‘OutColl’) will maintain the information. Each element in the collection will in turn contain a list. This list will correspond to a polygon, and will contain the outlet segment numbers of that polygon. In order to populate the collection, during an iteration, a condition will be applied to find out whether the polygon number of the segment under question is same as that of its downstream segment’s. If they are different then, the segment represents potential outlet of the polygon, so its number is added to the list in the collection corresponding to the polygon. Thus at the end of last iteration all the streams would be populated with their neighborhood information and the collection would be populated with as many lists containing outlet segment numbers as polygons, concluding the first part of the algorithm.

The second part of the algorithm deals with detection and correction of data inconsistencies. These inconsistency rectification procedures require two user input threshold values; one for trace length, and the other for the number of streams. The length threshold would be referred to as 'ThreshLength' and the stream count threshold as 'ThreshCount' in the descriptions. The use of these threshold values would be described in the inconsistency treatment methods. These values are determined by the user, by examining the data.

Taking into account the threshold values, the algorithm runs though the elements in the OutColl two times. In the first run, it iterates through all the elements in the 'OutColl'. The number of iterations would be equal to the number of polygons, as each element in the collection corresponds to a sub-basin. During an iteration, an element is picked up from the collection, which is a list of outlet segment numbers of the sub-basin under question. The list count is noted, if it is greater than one, then case 1 (see figure 4.4) or case 2 (see figure 4.5) or case 3 (see figure 4.6) is assumed to exist, otherwise the current iteration is skipped. To decide which one of these three cases exist the following logic is applied. To make the understanding of procedures easier, the explanations are provided with the help of examples.

Case 1: While looping through the outlet segment numbers in a list, if it is found that a segment corresponding to a number exists in the list such that it has no upstream segments, the occurrence of this case is confirmed. To determine whether a segment has any upstream segments, a query is performed so as to find the segments whose downstream node is same as the upstream node of the segment under question. If there exists at least one, then this case is assumed not to have occurred, otherwise the occurrence of case 1 is confirmed. In the figure 4.4a, it can be seen that segment attributed 'A' located in polygon 'A' should actually be attributed B, only if its geometric length is less than the user defined threshold for length. So, to rectify this inconsistency, in the network table, segment A will be attributed as a segment of polygon B (figure 4.4b) and will be removed from the outlet attribute of polygon A.

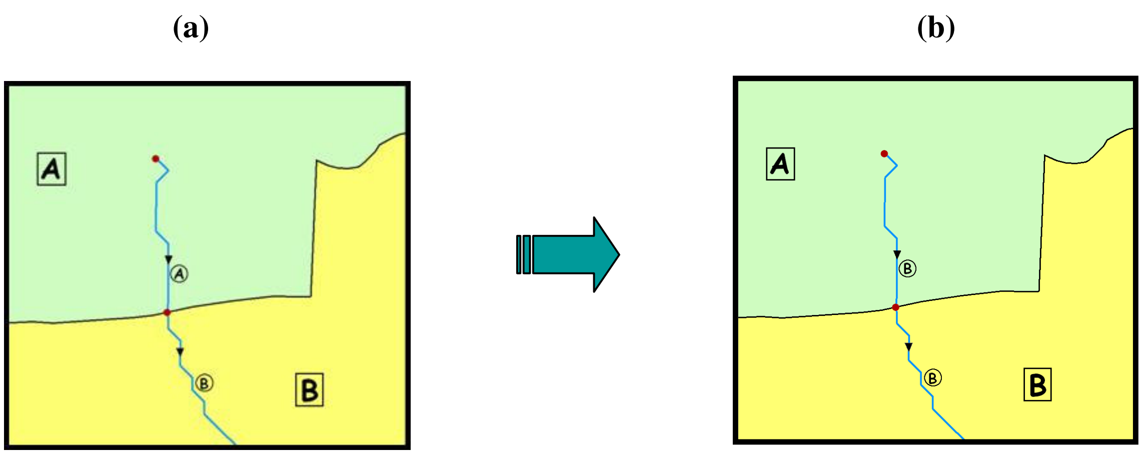


FIG 4.4. Stream Line Flowing in Sub-watershed B Has a Small Upstream Part in Sub-Watershed A. (a) Problem; (b) Result

Case 2: If Case1 fails to exist, a check for case 2 is performed. In this case, the threshold values, which are user defined, come into picture. An outlet segment number is read from the list and a downstream trace is performed on the corresponding segment, until a segment is found whose sub-basin number is same as sub-basin number of segment under question, where the number segments along the trace is less than or equal to ThreshCount. If a match for sub-basin number occurs, it implies the existence of case2. For example, in figure 4.5a, the segment lying upstream in sub-basin ‘A’ drains into sub-basin ‘B’ and then drains back into sub-basin ‘A’. To solve this kind of problem all the segments lying along trace are updated with new sub-basin number. The new sub-basin number would be that of the segment which is under question. The segment number of the segment under question is deleted from the collection. In the example, the segments lying in sub-basin ‘B’ are assigned ‘A’ as their sub-basin numbers. From the list corresponding to polygon ‘A’, segment ‘A’ is removed (Fig 4.5b).

(a) **(b)**

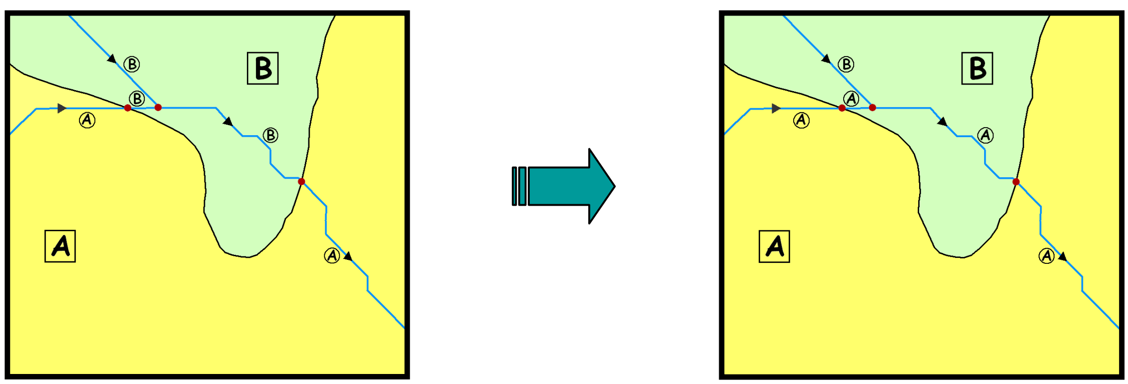


FIG. 4.5 Stream Line Flowing in Sub-watershed A Draining into Sub-watershed B and Then Draining Back into Sub-watershed A. (a) Problem; (b) Result

Case 3: Even after checking for existence of case 1 and case 2, if there stills exits two outlet segment numbers, the existence of case 3 is assumed. In order to confirm that this case persists, downstream traces are performed on both the outlet segments. The numbers of the segments that fall on these two traces are recorded in two separate lists; let them be called ‘SegList1’ and ‘SegList2’. ThreshLength and ThreshCount values provide the boundary conditions for stopping the execution of the traces. During a trace, the number of segments visited and the sum of their lengths at every instance a new segment is visited are recorded. Before a segment visited, if one these values has crossed the corresponding threshold values, the tracing is stopped and the problem of the sub-basin is assumed to be not solvable. The numbers of all such sub-basins are provided as a list, so that the user can edit them manually. If the threshold values are not crossed, case 3 exists. To solve this type of inconsistency, each element in 'SegList1' is compared with all the elements of 'SegList2' by iterating through the elements of 'SegList1' in the sequence they were added to the list. The first element (segment number) that is common to both the lists is chosen, the feature corresponding to this number is queried and it is broken into two new segments. Among the new segments formed, the one which comes first in the trace will be the new outlet segment of the sub-basin under question. For example, in figure 4.6a, it can be observed that sub-basin 'A' has two potential outlets (say 'A1', 'A2'). In this case, these two segments flow into sub-basin 'B'. To solve this inconsistency, the first outlet segment ('A1') is traced downstream and all the numbers of the segments visited along the trace are added to 'SegList1'. So, 'SegList1' now consists of {'A1', 'B1', 'B3'}. Similarly for the second outlet segment ('A2') also, 'SegList2' is built, which consists of {'A', 'B2', 'B3'}. Now on comparing elements of both these lists it can be observed that 'B3' is the segment that falls on both the traces. So, the segment corresponding to 'B3' is broken into two new segments, say 'B31' and 'B32'. Among these two new segments, 'B31' which will be visited first in the trace, will now form the unique outlet of polygon 'A', so in the

OutColl the two previous outlet segment numbers are replaced by the newly found outlet segment number and all the segments along are updated with new sub-basin numbers (Fig 4.6b).

(a) (b)

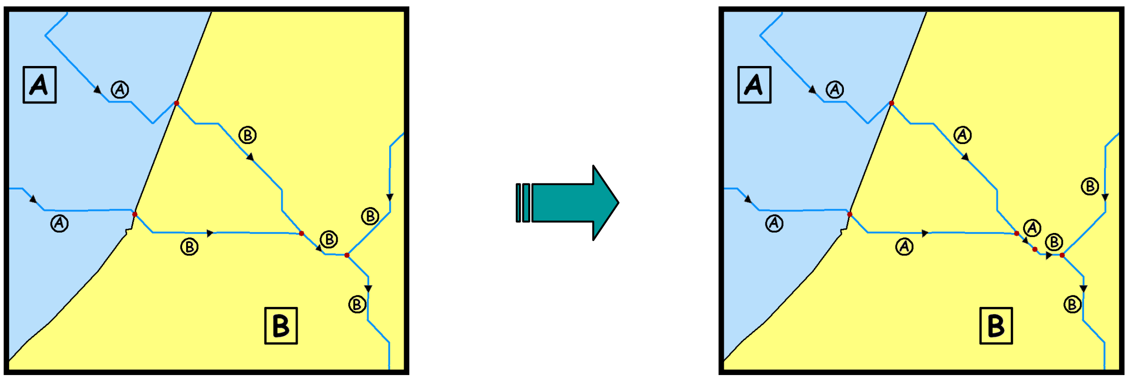


FIG. 4.6 Two Stream Lines in Sub-watershed A Outletting into Sub-watershed B. (a) Problem; (b) Result

This concludes the procedure for solving the problem of two or more segments in a polygon posing to be potential outlets. Even after checking for all the three cases presented above, if still there are more than one outlet segments for a polygon, the problem is reported to the user at the end of execution of the algorithm, so that the unsolved problems can be solved manually.

Case 4: Once all the above cases have been dealt with for a sub-basin, the occurrence of case 4 is checked. In this case, some sub-basins may outlet into wrong sub-basins. To detect this case, ThreshLength and ThreshCount values are used. During an iteration, the outlet segment is traced downstream such that either the length of the trace or number of traced segments have crossed the corresponding threshold values, or a segment is found whose sub-basin number is different from the number of the sub-basin into which the sub-basin under question is outletting. If one of the threshold values is crossed, the outlet segment number of the sub-basin is not altered; else the most downstream segment on the trace would be the new outlet. For example, see figure 4.7a, in which the segment in polygon ‘A’ outlets into polygon ‘B’ and flows for distance shorter than the threshold value and then outlets into polygon ‘C’. Based on the logic explained, segment ‘A’ will be traced downstream, until the distance traversed along the path is greater than ThreshLength, or the number of segments traced is greater than ThreshCount, or the segment ‘C’ is reached. If one of the threshold values is crossed, the algorithm does not do anything other wise, first segment in polygon ‘B’ is assigned ‘A’ as the polygon number and is also made the outlet of polygon ‘A’. Remaining segments are assigned ‘C’ as the polygon number. In the figure shown below segment ‘B’ will be the new outlet of sub-basin ‘B’.

(a) **(b)**

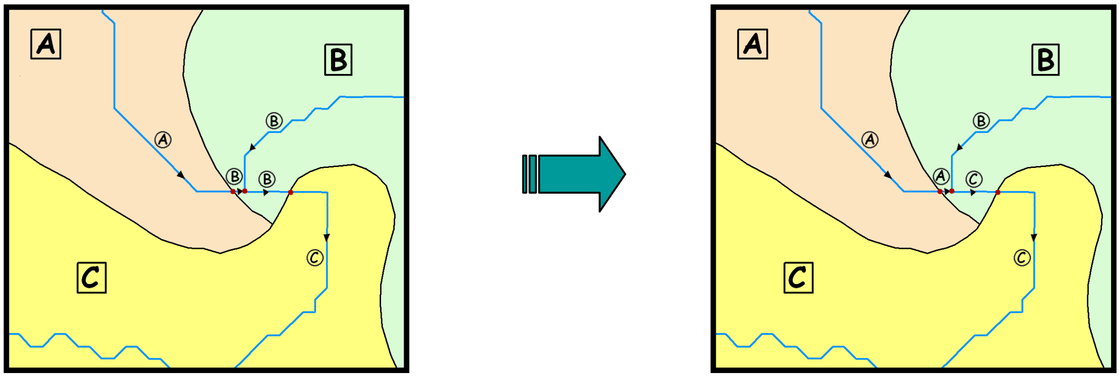


FIG. 4.7 Sub-watershed A Draining into Sub-watershed B Instead of Draining into Sub-watershed C. (a) Problem; (b) Result

4.2 HYDROLOGIC NETWORK PARAMETER DERIVATION

This section of the section focuses on the techniques developed for calculating hydrologic network parameters and on algorithms developed for network tracing. The tools developed for calculating network parameters, estimate drainage areas, calculate flow lengths and determine drainage divides. The tools developed for network tracing; provide capabilities for tracing streams, upstream and downstream in the network.

These methods assume that the flow system is represented by two datasets, one containing set of lines forming a dendritic network to represent streams and the other containing a set of polygons that represent the drainage areas (Fig 4.8). The lines and polygons possess relationship based on location (i.e., line inside polygon), whereas lines with other lines possess neighborhood (i.e. upstream and downstream) relation.

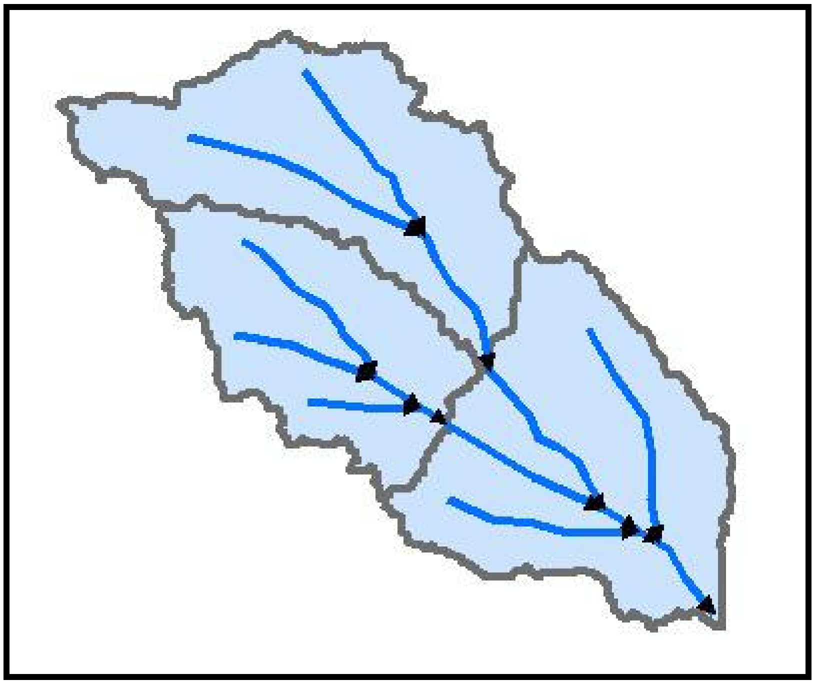


FIG. 4.8 Example Stream Network

In the previous three sections, methods which are related to making data consistent with the model requirements were described. Two of them perform intersection and build topology for streams; as a result, streams will be:

- Attributed with unique identification numbers (see field LineID in the table 4.1).
- Attributed with their geometric lengths (see field Length in the table 4.1).
- Attributed with their upstream-node and downstream-node numbers (see fields TNODE and FNODE respectively in the table 4.1).
- Attributed with the sub-basin polygon identification number they belong to (see field WShedID in the table 4.1),

TABLE 4.1. Stream network table

Attributes of TestStreams1

FID	Shape	LineID	Length	FNODE	TNODE	LenWt	WShedID	WShedArea
0	Polyline	1	28906.61776	1	2	1.25	10	7895000000
1	Polyline	2	7777.924049	3	1	1.25	10	7895000000
2	Polyline	3	275.325245	4	3	1.25	10	7895000000
3	Polyline	4	275.325245	4	3	1.25	10	7895000000
4	Polyline	5	12596.10823	5	4	1.25	10	7895000000
5	Polyline	6	63008.50842	6	3	1.25	10	7895000000
6	Polyline	7	56010.71975	7	5	1.25	10	7895000000
7	Polyline	8	59296.81178	8	5	1.25	10	7895000000
8	Polyline	9	74817.39972	9	1	1.25	10	7895000000
9	Polyline	10	50006.26758	10	8	1.25	12	7258000000
10	Polyline	11	82416.20111	11	10	1.25	12	7258000000
11	Polyline	12	66307.02138	12	10	1.25	12	7258000000
12	Polyline	13	11428.60846	13	7	1.25	11	5324000000
13	Polyline	14	36268.66181	14	13	1.25	11	5324000000
14	Polyline	15	17909.31446	15	13	1.25	11	5324000000
15	Polyline	16	47242.49761	16	15	1.25	11	5324000000
16	Polyline	17	62973.41682	17	15	1.25	11	5324000000

Record: 4 Show: All Selected Records (0 out of 17 Selected.) Options

- Optionally, attributed with the length weight (see field LenWt in the table 4.1),

and, the polygons will be:

- Attributed with unique identification numbers (see field WShedID in the table 4.2).
- Attributed with their geometric areas (see field WShedArea in the table 4.2),
- Optionally, attributed with the area weight (see field AreaWt in the table 4.2),

TABLE 4.2. Watershed table

Attributes of TestWshed

FID	Shape	WShedID	WShedArea	AreaWt
0	Polygon	10	7895000000	1.350000
1	Polygon	12	7258000000	1.350000
2	Polygon	11	5324000000	1.350000

Record: 1 Show: All Selected Records (0 out of 3 Selected.)

Though the model provides methods to extract all these information from datasets, user has an option to use these methods directly without applying the data treatment methods such as dendrification, intersection, building topology and treatment of inconsistencies. This is valid only if stream and watershed data cater to the model requirements and the attribute information requirements.

Apart from the attribute information mentioned above, lines and polygons can also be attributed with optional information that will be used to capture features of data that are not captured in the spatial datasets. Each segment in the network can be attributed with a weight factor that can be used, for example, to account for additional stream length caused by meandering, which is not captured in the line dataset (see field LenWt in the table 4.1). Other applications of the length weight factor are related to, for example, the calculation of moments of the response functions of runoff sources for use in source-to-sink model. In this case, the first moment (i.e. flow time) is calculated using weight equal to $1/v$ – where, v is the flow velocity and second moment is calculated using a weight equal to D/v^3 – where, D is the flow longitudinal dispersion coefficient.

Also, each polygon can be assigned a weight factor that can be used to weight the upstream areas differently based on spatially distributed pattern or presence of a certain pollutant source. (See field AreaWt in the table 4.2).

The procedure starts by making the streams know who their neighbors are, then goes onto calculating the parameters pertaining only to streams, next and finally calculates the parameters that are common to both streams and watersheds. To make the computations efficient, an intermediate parameter called 'Rank' has been introduced. In turn, to make the calculations of rank efficient, segments in the network are attributed with their downstream segment numbers. Once the downstream segment numbers have been assigned to all the streams, their upstream segment information is computed. As each segment can have only one downstream segment and as they can have more than one upstream segment, segments have one value stored in the network table for their downstream segment number and have a collection of values for upstream segment numbers. Although this information is implicitly available through nodes of the segments, the network table is attributed with this information explicitly making it easier and faster to access the information when ever needed. For a given segment, a downstream segment is the one whose upstream node number is equal to its downstream node number. A reverse logic is used for finding the upstream segments for any given segment, which is, for any given segment, an upstream segment is the one whose downstream node number is equal to its upstream node number. A single loop is used to run through all the elements in the network. During a single step, a segment is chosen randomly, its downstream node number is read and a query is applied on the upstream node number list to find the segment that has the same number as the number read. The segment thus found is the downstream segment; its identification number is attributed to the segment under question (see DSLine field in Table 4.2.c). Once all the segments are attributed with their downstream segment information, the upstream segment information is computed and then stored in collections, the logic for computing which is as follows: For each segment in the network, reading its downstream segment number, a query is performed on the segment identification number column, until the segment which bears this identification number is found. Once such a segment is found, its upstream segment list is updated with the identification number of the segment under question. This process is performed for all the segments. These lists are stored in a collection, in which a list corresponds to an array of numbers. The elements in the collection are accessed using, what is called a 'Key'. Unlike arrays, in which elements are accessed by indexes, in collections elements are accessed by keys (identification numbers for segments are used as keys). An advantage of using collections is that information access becomes efficient.

With segments being attributed with their downstream segment numbers, they are then used in computation of ranks for the segments. Rank of a segment is defined by:

$$R = 1 + \max \{R_{u1}, R_{u2}, R_{u3} \ldots\ldots\ldots\ldots, R_{ui}, \ldots\ldots R_{un}\} \qquad (4.1)$$

Where R is the rank of the segment, and $R_{u1}, R_{u2}, R_{u3} \ldots\ldots\ldots\ldots, R_{ui}, \ldots\ldots R_{un}$ are the ranks of 'n' segments upstream of it. Although exceptions could be found, in most cases the value of n is 2, referring to two streams, upstream of given stream. From the physical point of view, rank of a segment is equal to one plus the number of segments of the upstream flow path with the greatest number of segments. Note that the rank indicates only the number of upstream segments, but does not identify the specific upstream flow path that contains these segments.

The algorithm for computing rank starts by finding all the segments that represent headwaters, which constitutes the first iteration. These are found by querying the downstream segment column of the network table. The segments whose identification numbers are missing in the downstream segment number column represent headwaters. All such segments are assigned a value of one for their rank attribute. In the next iteration, the algorithm loops through all the headwater segments. During a single execution, a head water segment is traced downstream and all the segments that fall on the trace are attributed a value equal to one plus rank of its immediate upstream segment found on the trace. The trace is continued until the outlet segment is reached, or a segment is visited whose rank is greater than or equal to one plus the rank of its immediate upstream line located along the trace. Here outlet segment refers to the most downstream segment in the network (Fig 4.9).

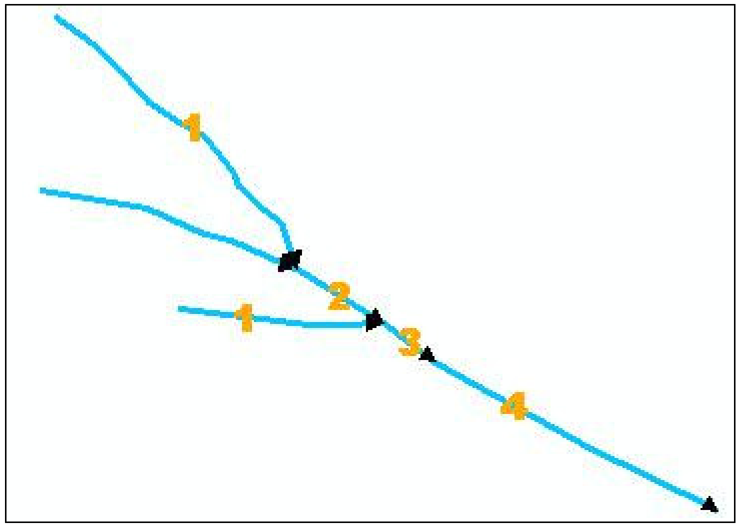


FIG. 4.9 Rank Attribute

The network table is then sorted with respect to the rank attribute. The advantage of sorting the network table by the rank is, for a given segment all its upstream segments are found above it in the table and all the downstream segments are found in the table below it. By doing so, parameters that vary in the upstream to downstream direction can be calculated in the table from top to bottom without leaving any calculations pending and parameters varying in the opposite direction are calculated in the bottom to top direction(see Rank field in Table 4.3).

Taking the advantage of the sorting the table with respect to rank, parameter calculation methods would now be presented. One of them is the computation of upstream flow lengths of the segments in the stream network. Upstream flow length of a segment is the length of the longest upstream flow path out of all upstream flow paths, flow paths starting at its downstream node. The upstream flow length of a segment is defined as

$$U = \max\{U_{u1}, U_{u2}, U_{u3}\ldots\ldots\ldots\ldots\ldots, U_{ui},\ldots\ldots\ldots U_{un}\} + L \qquad (4.2)$$

Where, U is the upstream flow length of a segment, $U_{u1}, U_{u2}, U_{u3}\ldots\ldots\ldots\ldots\ldots, U_{ui},\ldots\ldots\ldots U_{un}$ are the upstream flow lengths of n segments located immediately upstream of the segment and L is the length of the segment. Notice that the upstream flow length of a rank one segment is the length of the segment itself. Additionally, notice that segments located upstream of a given segment are always above in the table, and are assigned their upstream flow lengths prior to the calculation of the upstream flow length of the segment in question (see USLength field in Table 4.3).

Similarly, weighted upstream flow lengths of the segments can be computed, the only difference is that instead of adding the lengths of the segments of the longest upstream flow path, their weighted lengths are added. The weighted upstream flow length is calculated as

$$WU = \max\{WU_{u1}, WU_{u2}, WU_{u3}\ldots\ldots\ldots\ldots\ldots, WU_{ui},\ldots\ldots\ldots WU_{un}\} + L \text{ x } W_L \qquad (4.3)$$

Where, WU is the weighted upstream flow length of a segment, $WU_{u1}, WU_{u2}, WU_{u3}\ldots\ldots\ldots\ldots\ldots, WU_{ui},\ldots\ldots\ldots WU_{un}$ are the weighted upstream flow lengths of n segments located immediately upstream of the segment, L is the length of the segment and W_L is the length weight of the segment itself (see WTUSLength field in Table 4.3).

Downstream flow length of a segment is the length of the downstream flow path from segment to the network outlet. Downstream length is measured from the downstream node of the segment, so the length of the segment itself is not included in the calculation. Since there is only one downstream segment for a segment in the network, there exists only downstream flow path. The downstream flow length of a segment is defined as:

$$D = D_d + L_d \tag{4.4}$$

Where, D is the downstream flow length of a segment, D_d is the downstream flow length of the segment located immediately downstream of the segment and L_d is the length of the segment located immediately downstream. Note that the downstream flow length of the segment with highest rank, which is the network outlet, is zero, which is the last record in the network table. Starting at the last record in the network table, the calculations move all the way up until the downstream length of all the segments is completed (see DSLength field in Table 4.3).

Similarly, weighted downstream flow lengths of the segments can be computed, the only difference is that instead of adding the lengths of the segments of the downstream flow path, their weighted lengths are added. The weighted downstream flow length is measured from the downstream node of the segment. The weighted downstream flow length is calculated as:

$$WD = WD_d + L_d \times W_{Ld} \tag{4.5}$$

Where, WD is the weighted upstream flow length of a segment, WD_d is the weighted downstream flow length of the segment located immediately downstream of the segment, L_d is the length of the segment located immediately downstream and W_{Ld} is the length weight of the segment located immediately downstream itself (see WTDSLength field in Table 4.3).

This concludes the part of the methodology that deals with techniques for computing parameters, that are only stream based. To this point, all parameters calculated are based on the network topology and geometry. However, by relating the segments of the network with the polygons of the watershed dataset, additional information concerning drainage areas can be obtained. To establish a one-to-one relation between stream segments and watershed polygons, the outlet segment of each polygon is identified by flagging the segment of highest order out of all segments with the same polygon identification number. The logic behind this is as follows: While looping through the segments of the network, a segment is chosen, its polygon number is compared with its downstream segment's polygon number. If they are equal, it is assigned a value of 0 for its outlet segment attribute; otherwise it is assigned a value of 1 for its outlet segment attribute (see Outlet field in Table 4.3). Figure 4.10 shows the stream network in which the outlet segments are selected red in color. Also, the outlet segment is attributed with the downstream polygon number, which is the polygon number of its downstream segment. This attribute is transferred to all other segments in the subbasin by querying the polygon number column in the network table (see DSWShdID field in Table 4.3).

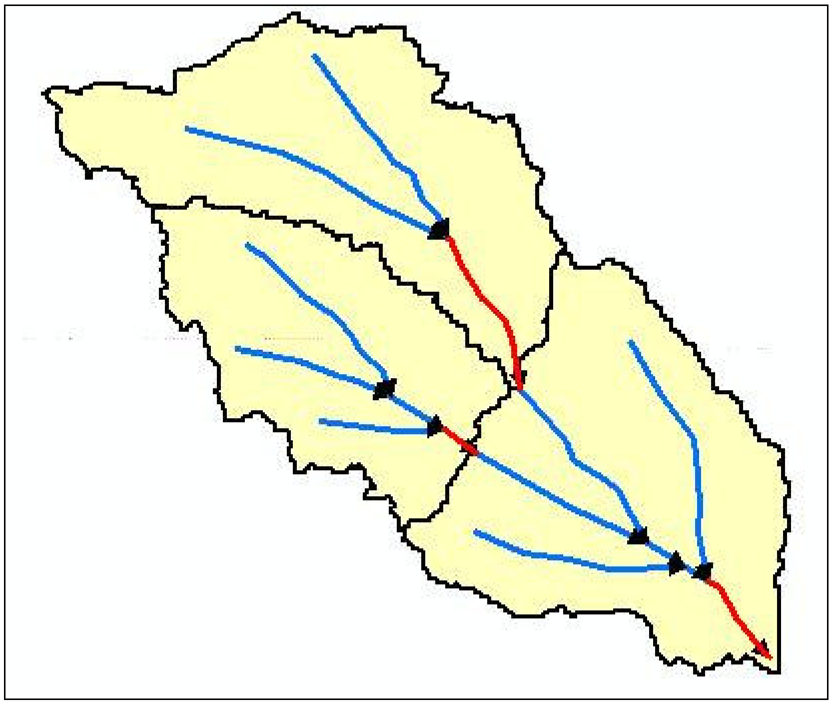


FIG. 4.10 Outlet Segments

Calculation of total drainage areas can now be done based on the area of the watershed polygons and the network topology. The total drainage area is the sum of the areas of all polygons located upstream of a segment; however, the area of the watershed polygon in which the segment itself is located is added only at the outlet segments (see AccumArea field in Table 4.3). The total drainage area is calculated as:

$$TA = \sum_{i=1}^{n} TA_{u-i} + A \tag{4.6}$$

Where TA is the total drainage area of the segment, TA_{u-1}, TA_{u-2}, … TA_{u-i} … and TA_{u-n} are the total drainage areas of the n segments located immediately upstream of it, and A is the area of the watershed polygon in which the segment is located.

The weighted total drainage area is similar to the total drainage area, but adding weighted areas (i.e., the product of the area by the area weight) instead of areas. The weighted total drainage area is the sum of the weighted areas of all polygons located upstream of a segment; however, as in the previous case, the weighted area of the watershed polygon in which the segment itself is located is added only at the outlet segments (see WTAccumArea field in Table 4.3). The weighted total drainage area is calculated as:

$$WTA = \sum_{i=1}^{n} WTA_{u-i} + A \times W_A \quad (4.7)$$

Where WTA is the weighted total drainage area of the segment, WTA_{u-1}, WTA_{u-2}, … WTA_{u-i} … and WTA_{u-n} are the weighted total drainage areas of the n segments located immediately upstream of it, and A and W_A are the area and the area weight of the watershed polygon in which the segment is located.

TABLE 4.3. Stream network table attributed with hydrologic parameters

Attributes of HydroEdge

LineID	DSLine	Rank	USLength	WTUSLeng	DSLength	WTDSLeng	Outlet	DSWShedID	AccumArea	WTAccuArea
3	2	1	275.325245	344.156556	36684.54180	45855.67726	0	0	0	0
6	2	1	63008.50842	78760.63553	36684.54180	45855.67726	0	0	0	0
9	1	1	74817.39972	93521.74965	28906.61776	36133.27220	0	0	0	0
11	10	1	82416.20111	103020.2513	158859.0546	198573.8183	0	10	0	0
12	10	1	66307.02138	82883.77673	158859.0546	198573.8183	0	10	0	0
14	13	1	36268.66181	45335.82726	116995.3035	146244.1293	0	10	0	0
16	15	1	47242.49761	59053.12201	134904.6179	168630.7724	0	10	0	0
17	15	1	62973.41682	78716.77102	134904.6179	168630.7724	0	10	0	0
10	8	2	132422.4686	165528.0858	108852.7870	136065.9838	1	10	7257999975.	9798300140.27
15	13	2	80882.73129	101103.4141	116995.3035	146244.1293	0	10	0	0
8	5	3	191719.2804	239649.1006	49555.97529	61944.96911	0	0	7257999975.	9798300140.27
13	7	3	92311.33975	115389.1746	105566.6950	131958.3688	1	10	5324000060.	7187400208.47
7	5	4	148322.0595	185402.5743	49555.97529	61944.96911	0	0	5324000060.	7187400208.47
5	4	5	204315.3887	255394.2358	36959.86705	46199.83381	0	0	1258200003	16985700348.7
4	2	6	204590.7139	255738.3924	36684.54180	45855.67726	0	0	1258200003	16985700348.7
2	1	7	212368.6380	265460.7975	28906.61776	36133.27220	0	0	1258200003	16985700348.7
1	0	8	241275.2557	301594.0697	0	0	1	0	2047699998	27643950469.3

Record: 0 Show: All Selected Records (0 out of 17 Selected.) Options

Based on the relation established between segments and polygons through the polygon number, the parameters like the total drainage area, weighted drainage area, outlet segment number and downstream polygon number can be transferred to the polygons (see AccumArea, WTAccumArea, Outlet, DSWShedID fields respectively in Table 4.4).

TABLE 4.4. Watershed table attributed with hydrologic parameters

Attributes of Watershed

WShedID	Outlet	DSWShedID	AccumArea	WTAccumArea
10	1	0	20476999986	27643950469.3097
12	10	10	7257999975.7	9798300140.2797
11	13	10	5324000060.4	7187400208.47405

Record: 0 Show: All Selected Records (0 out of 3 Selected.)

Moreover, as oppose to the calculation of flow lengths, network tracing consists of identifying the actual segments located upstream or downstream of a given segment. Because of the large number of segments involved in the tracing of each segment, tracing has to be performed “on the fly” for a limited number of user-defined segments. Downstream tracing consists of populating a collection with the number of the downstream segments of the selected segment. The process starts by adding the number of the immediate downstream segment, and then adding the number of the downstream segment of the just added segment, and so on (Fig. 4.11).

Likewise, upstream tracing consists of populating a collection with the numbers of all the upstream segments of the selected segment. If a number of segments are selected, and at least one of them is upstream of one of the others, the upstream trace collection of the downstream segment will not include the segments already included in the collection of the upstream one. The algorithm is based on a technique called ‘Breadth First Search’ Technique (Weiss 1994). The algorithm starts by adding the identification numbers of the selected segments to a collection. It starts visiting elements in the order they are added to the list. During a visit to an element (i.e. segment identification number) in the list, its immediate upstream segment number list is appended to the collection. This process is performed until the last segment in the collection, which does not have any upstream segment numbers, is visited (Fig. 4.12).

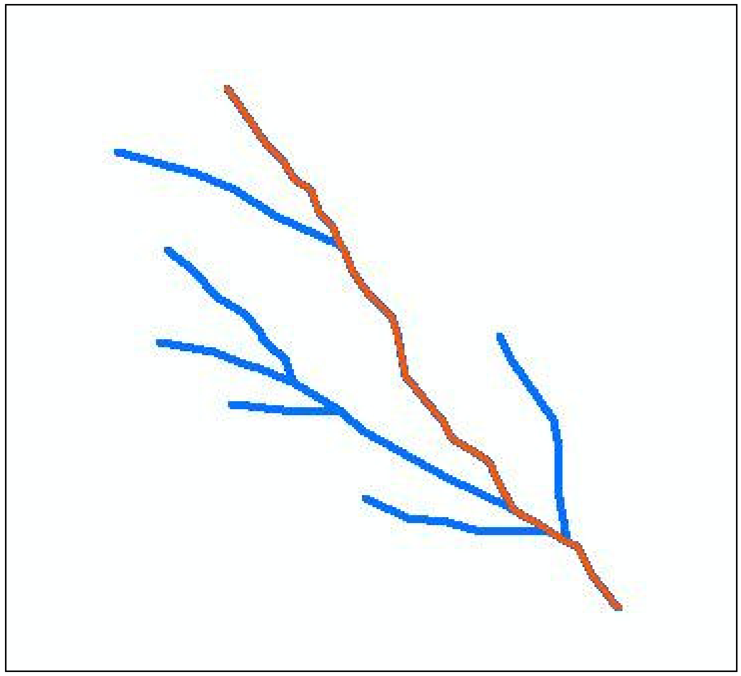


FIG. 4.11 Downstream Trace

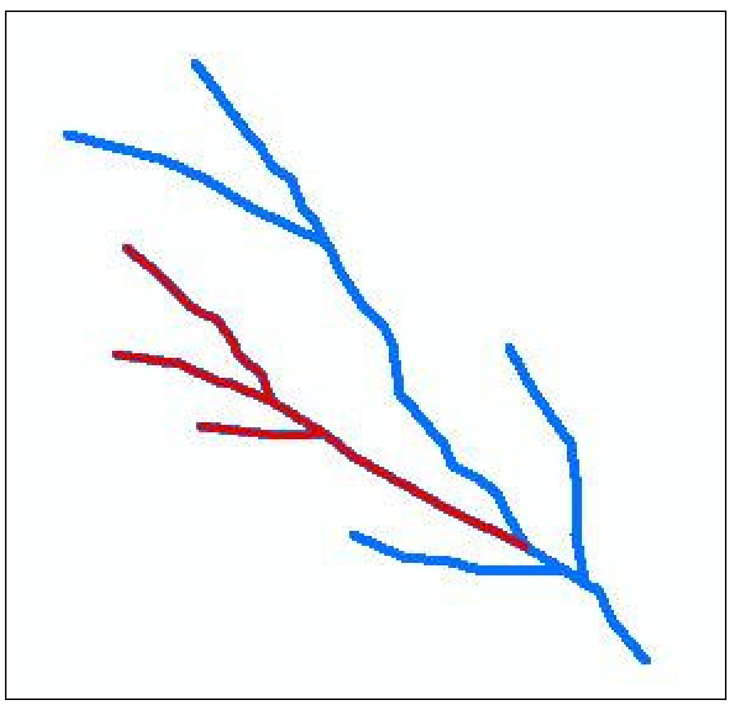


FIG. 4.12 Upstream Trace

4.3 HYDROLOGIC MODEL DESCRIPTION

Physical processes occurring in hydrologic systems are represented using hydrologic models, the input to these models being the meteorological and hydrologic quantities and output of the system is the response of the system to these inputs. Hydrologic models can be classified based on the level of simplification of hydrologic variable information. These variables vary both in the three space dimensions and the time dimension. So, it is complex to model a system in all these dimensions. To overcome this complexity, some variables are spatially or temporally averaged. Based on the spatial averaging of the variables, the models are classified into lumped or distributed models. Although some models are referred to be distributed, they assume uniformity in the variables at some spatial scale. Lumped models assume uniformity at a higher spatial scale compared to their counterpart. This part of the section discusses about the development of a distributed hydrologic model. Methods for routing runoff over land, through the channels and reservoirs are presented in this section. Though this model does not provide capabilities to capture infiltration processes, the abstractions can be computed by coupling the model with already existing ones. A procedure on how this can be accomplished is shown with the help of an example in the next section. In order to make the model distributed, the watershed, which is already discretized into sub-basins, is further divided into smaller land units, what are called ‘Sources’ here. Sources refer to the land units of uniform properties, in soils and land use per each sub-basin.

Sources are delineated by intersecting soil, land use and sub-basin polygons. The surface runoffs generated at these sources are then routed to the respective sub-basin outlets (called 'Sinks') as flows. This process is called overland flow routing. A technique known as 'Source-to-Sink' has been incorporated to model overland flow (Olivera et al. 2000). Once the flows at the sinks are computed, they are routed through the channels and reservoirs to obtain flows at the watershed outlet, the point in the watershed after which no more routing is required. Muskingum routing method and pure lag method are employed for channel routing. Storage-outflow method is employed for reservoir routing.

As mentioned previously, the flow system is represented by a set of lines for streams, a set of polygons for sub-basins, and a set of polygons for Sources. For each source, it is assumed that runoff data is provided. Apart from the hydrologic topology parameters, parameters such as average flow velocities and dispersion coefficients are assumed to have been attributed to both streams and sub-basins in their attribute tables. Also, the stream lines are assumed to be attributed with their Muskingum X parameter (see section 4.3.2.1). To model flow through reservoirs, a set point features to represent them should be provided. Apart from this, it is also assumed that for each reservoir, its characteristics such as storage versus outflow are provided. These are the set of data and attribute information required for hydrologic modeling.

4.3.1 Overland Flow Routing

The transport of excess precipitation from all points inside a sub-basin to the corresponding outlet is called overland flow routing. In the hydrologic model developed, overland flow processes are captured using the ‘Source-to-Sink’ technique. Instead of computing the response of a sub-basin using a single function, as in unit hydrograph method, response function for each source is defined separately. The response functions are based on overland flow velocities, overland dispersion coefficients, stream flow velocities and stream dispersion coefficients. Response function computed for a source is then used to compute flows at the sink (sub-basin outlet) by convolving the runoff depths over the response values. The flows thus produced at the sink due to all the sources are then added to obtain the complete response of individual sub-basins.

To delineate sources, soils, land use and sub-basin polygons are intersected. Intersection is a process by which new polygons are generated, that are geo-referenced. The methods for intersection are available with GIS software packages and these procedures are used for the automation of the same. The sources thus delineated along with the corresponding runoff data form the input for the overland flow model.

Mathematically describing the source-to-sink technique, the response function of a source, is defined by

$$u_j(t) = \frac{1}{2t\sqrt{\pi(t/t_j)/\prod_j}} \exp\left\{\frac{-[1-(t/t_j)^2]}{4(t/t_j)/\prod_j}\right\} \tag{4.8}$$

Where t_j is the average flow time and $\prod_j$ is the representative Peclet number for the flow path from source j to sink i.

$$t_j = \frac{L_{ov}}{V_{ov}} + \sum_k \frac{L_k}{V_k} \tag{4.9}$$

Where, L_{ov} is the length of the overland flow path, V_{ov} is the overland flow velocity, L_k and V_k are length and velocity of stream flow path 'k' (see figure 4.13).

$$\prod_j = \left[\frac{L_{ov}}{V_{ov}} + \sum_k \frac{L_k}{V_k}\right]^2 \Bigg/ \left[\left(\frac{D_{ov}}{V_{ov}^3}\right)L_{ov} + \sum_k \left(\frac{D_k}{V_k^3}\right)L_k\right] \tag{4.10}$$

Where, D_{ov} is the dispersion coefficient of the overland flow path, D_k is the dispersion coefficient of the stream flow path k (Fig. 4.13).

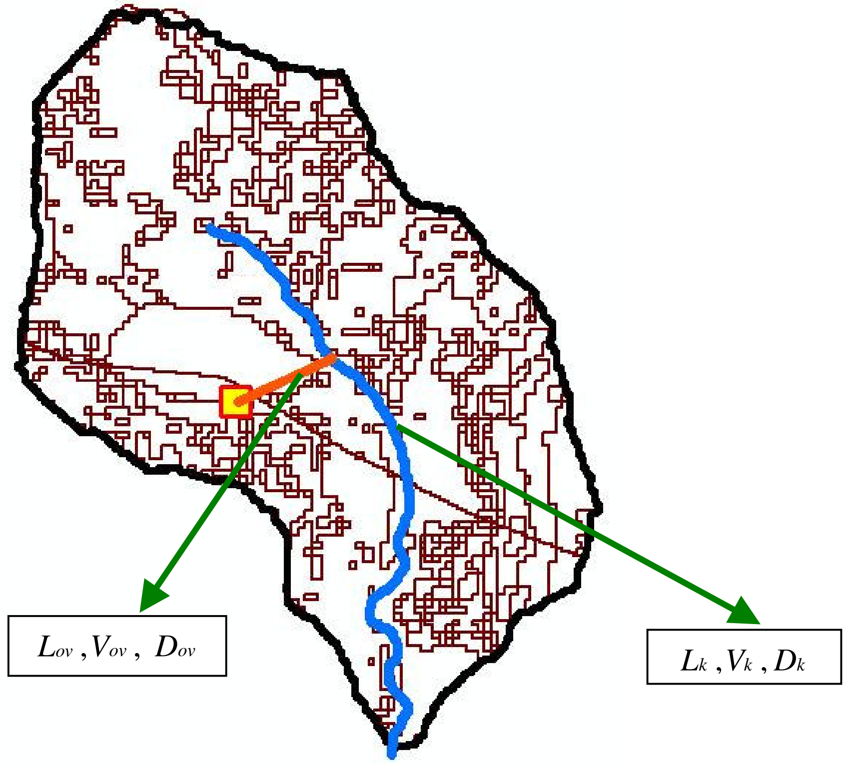


FIG. 4.13 Source-to-Sink Method

Computation of total response as flow, of a source 'j' at sink 'i', is given by the following convolution equation.

$$Q_j(t) = A_j R_j(t) * u_j(t) \qquad (4.11)$$

Where, A_j is the area of the source 'j', $R_j(t)$ is the excess precipitation depth at source 'j' at time t and $u_j(t)$ is the unit-impulse response function of source 'j'.

The total response of a sub-basin is then computed by:

$$Q_i(t) = \sum_j Q_j(t) \qquad (4.12)$$

Where, $Q_i(t)$ is the flow at the sink of sub-basin i at time‘t’, $Q_j(t)$ is the flow due to source ‘j’ at time‘t’.

These are the equations that constitute the building blocks for the overland flow section of the hydrologic model. Section 2 included the discussion about Source-to-sink model and was mentioned that this method can be extended to modeling BMPs. Though this was not included in the present model, a brief discussion would be provided next. Olivera (1996) described the inclusion of loss factor, which when multiplied to the response function could be used to account for evaporation and infiltration.

$$u_j(t) = \frac{\Phi_j}{2t\sqrt{\pi(t/t_j)/\prod_j}} \exp\left\{\frac{-[1-(t/t_j)^2]}{4(t/t_j)/\prod_j}\right\} \tag{4.13}$$

where, if λ_k [1/T]is the loss parameter, then Φ_j is the dimensionless loss factor defined by,

$$\Phi_j = \exp\left\{-\sum_k \left(\frac{\lambda_k}{v_k}\right) L_k\right\}$$

Similarly, to model loss of contaminant concentration along a stream flow path due to BMPs and flow time, λ_k will act as the loss parameter.

The algorithm for modeling overland flow can be explained in two phases. The first phase deals with the computation of Source-to-sink parameters for various elements in the flow system. The second part consists of methods for the routing the runoff itself.

Going into details of the algorithm, the first phase of the algorithm consists of two loops- outer loop and the inner loop. The outer loop is meant for iterating through all the sub-basins, while the inner loop is meant for iterating through all the sources that belong to the sub-basin chosen in the outer loop. Simultaneously the sources are attributed with the parameters needed for routing excess precipitation in the next phase of the algorithm.

The algorithm starts in the outer loop, picks up a sub-basin, and performs a query on all the source polygons to choose the ones that lie within the sub-basin under question. Once the list of source polygons is compiled, it runs through these sources, which will constitute the inner loop. During an iteration in the inner loop, a source is chosen randomly, a spatial query is performed on the map so as to find the nearest stream to the source. The identification number of this nearest stream and the shortest distance to it are attributed to the source as the closest stream line and as overland flow length (L_{ov}) respectively. Once this is achieved, the source under question is also attributed with stream flow length (L_k) which is the distance from the point where the overland flow path ends to the downstream node of the outlet stream. The overland flow path and stream path lengths are then used for computing the Source-to-sink parameters (average flow time (t_j) and representative Peclet number ($\prod_j$)) for the source under question using equations 4.9 and 4.10. Thus at the end of an inner loop, all the sources belonging to a sub-basin are dealt with and at the end of the outer loop all the sub-basins would have been dealt with, ending the first phase of the overland flow routing algorithm.

The second phase of the algorithm as already mentioned deals with the routing procedure. Even this phase runs in two similar loops, outer loop dealing with the sub-basins and the inner loop dealing with the sources. Similar to the first phase, at an execution of the outer loop, a sub-basin is picked randomly. The algorithm then performs a query on the sources to find those that belong to the sub-basin under question. During the execution of an inner loop, a source is chosen randomly, knowing its source-to-sink parameters, its response function is computed using equation 4.8. The response function thus computed is then used for convolving the excess precipitation values to obtain the complete response of a source using equation 4.11. This process is repeated till the inner loop is broken, by then the responses of all the sources in a sub-basin would have computed. All such responses are then summed up to obtain the flows at the outlet of the sub-basin under question using equation 4.12. Finally, when the outer loop is broken, i.e. all when all the sub-basins are processed, overland flow hydrographs would have been obtained for all sub-basins, which then need to be routed through the channels and reservoirs to the outlet of the watershed.

4.3.2 Channel Routing

Flow routing in channels is a process of computing outflow hydrograph at the downstream end of a channel, given the inflow hydrograph to the channel at the upstream end. Generally speaking, channel routing techniques can be broadly classified into hydrologic routing and hydraulic routing, though in the model hydrologic routing techniques are used. Hydrologic routing is also referred to as storage routing, because a continuity equation in conjunction with the relationship between storage and discharge are used. "Hydraulic routing involves simultaneously computing both stage and discharge as a function of location and time" (Wurbs and James, 2002).

The flows at the sub-basin outlets, due to overland flow, are routed to their downstream sub-basin outlets using either of the two different types of channel routing techniques provided: Muskingum routing technique or Pure Lag method. The choice of technique to be employed is based on the channel flow time and analysis time step. If the channel flow time is less than the analysis time step, channel routing is conducted using Pure Lag method; otherwise routing is done using Muskingum method (Fig.4.14). Before getting into the details of how these methods are implemented in the model, theoretical discussions of the routing techniques are presented.

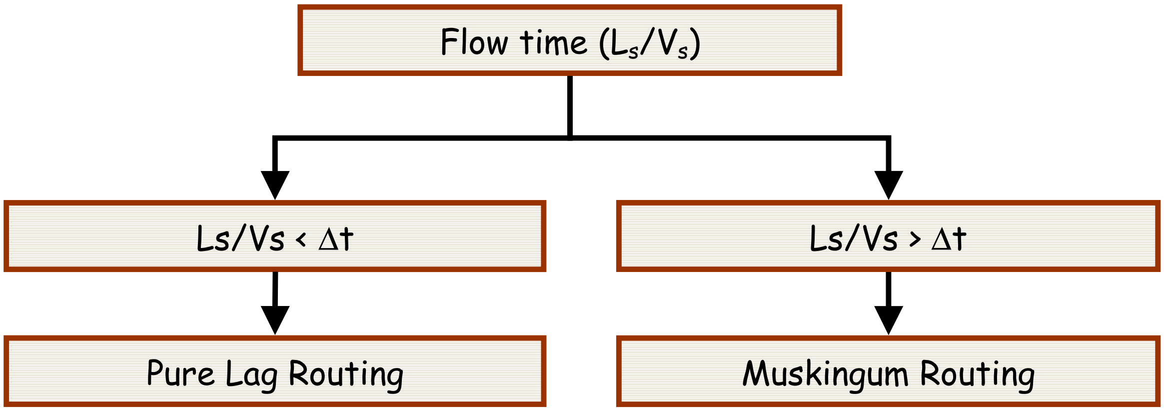


FIG. 4.14 Routing Method Determination

4.3.2.1 Muskingum Routing Method (Wurbs and James, 2001)

The Muskingum method is based on the continuity equation and the relationship between storage, inflows and outflows. The method assumes that storage volume in a channel at an instant in time is a linear function of weighted inflow and outflow, which is given by the following equation:

$$S = K\left[xI + (1-x)\,O\right] \tag{4.14}$$

Where, S is the storage, I and O are the inflow and outflow at the beginning and at the end of a time step, x and K are the Muskingum parameters. x is a weighting factor that lies between 0.0 to 1.0 representing the relative importance of inflow over outflow and, K represents channel travel time. K is nothing but the average flow time in channel, which is the ratio of length of the channel to the average velocity in the channel. The outflow (O_2) at the end of a time step is related to inflow at the end of the time step

(I_2), to inflow at the start of the time step (I_1) and outflow at the start of the time step (O_1) by:

$$O_2 = C_1 I_2 + C_2 I_1 + C_3 O_1 \quad (4.15)$$

Where, if, Δt is the analysis time step then,

$$C_1 = \frac{0.5\,\Delta t - Kx}{K - Kx + 0.5\,\Delta t} \quad (4.16)$$

$$C_2 = \frac{0.5\,\Delta t + Kx}{K - Kx + 0.5\,\Delta t} \quad (4.17)$$

$$C_3 = \frac{K - Kx - 0.5\,\Delta t}{K - Kx + 0.5\,\Delta t} \quad (4.18)$$

$$C_1 + C_2 + C_3 = 1 \quad (4.19)$$

Though these equations can be directly applied to all channels, for channels that longer, numerical instability may arise. Here long channels refer to those that fail the condition: $K/3 < \Delta t < K$. Once a channel is determined to be long, it is sub-divided into sub-channels based on the following equation (Olivera 1998):

$$n = \text{int}\left[\frac{L}{3\Delta t V}\right] + 1 \quad (4.20)$$

Where n is the number of sub-channels, L is the length of the channels, Δt is the analysis time step, V is the average flow velocity in the channel.

4.3.2.2 Pure Lag Method

This is a simple channel routing method used for shorter channels, using which the inflow hydrograph is lagged in time so as to obtain the outflow hydrograph (Olivera, 1998). In other words the inflow hydrograph is shifted along the time axis, by an amount equal to the lag time. Lag time is the same as average flow time (K) in a channel. This method assumes only pure translation to take place, but does not consider storage effects in channels. Due to this reason, the shape of the hydrograph is conserved. As mentioned previously, this method is used for channels, whose average flow time to be less than the analysis time step. Mathematically the model can be represented as:

$$O(t + t_{lag}) = I(t) \qquad (4.21)$$

Where, $O(t)$ is the outflow hydrograph, $I(t)$ is the inflow hydrograph, t_{lag} is the lag time.

4.3.3 Reservoir Routing

The method for reservoir routing follows the procedure explained by Wurbs and James, 2001. Reservoir routing consists of routing inflow hydrograph at the upstream side of the reservoir to its downstream side. The method employed in the model to route flows through reservoirs is a variation of hydrologic routing technique, called Storage-Outflow routing. This technique assumes that storage is function of outflow, instead of storage being a function of both inflow and outflow, as in the case of Muskingum method. As input, for each reservoir, the relationship between storage (S) and outflow (O) is needed. Apart from this relationship, initial storage of the reservoir is required. In this technique, continuity equation is combined with the inputs to produce the outflow hydrograph. The equations below represent the mathematical form of the continuity equation.

$$\frac{S_2 - S_1}{\Delta t} = \frac{I_2 + I_1}{2} - \frac{O_2 + O_1}{2} \tag{4.22}$$

Where, Δt is the analysis time step, S_1, S_2 are the storages at the start and end of the analysis time step, I_1, I_2 are the inflows at the start and end of the analysis time step and O_1, O_2 are the outflows at the start and end of the analysis time step.

The algorithm for routing starts by building a relationship between $2S/\Delta t + O$ Vs O, using the given relationship between storage and outflow. This relationship will later be used in routing flows through the reservoirs. At every time step inflows at the beginning and the end of the time step are known. Also known are the storage and outflow at the start of the time step. The remaining unknowns that are storage and outflow at the end of the time step are computed by rearranging the terms of equation 4.23 as shown below.

$$\left[\frac{2S_2}{\Delta t}+O_2\right]=I_2+I_1+\left[\frac{2S_1}{\Delta t}-O_1\right] \tag{4.23}$$

At the end of a time step, the left hand side value is unknown but the one on the right hand side is known. Now using the left side value the outflow at the end of a time step can now be found by using the earlier developed relationship, $\frac{2S}{\Delta t}+O$ Vs O. This process is continued until the outflows from reservoir become zero.

4.3.4 Routing Algorithm

In the section that dealt with the overland flow routing method, it was explained how the excess precipitation at the sources is routed to the sub-basin outlets as flows. Thus, each sub-basin's outlet is associated with a set of flow values that form the inflows to the channels and reservoirs. During the computation of hydrologic network parameters for streams and sub-basins, streams would be attributed with their Rank, Downstream Line Identification number and Outlet {1, 0} values. Rank attribute is useful in determining the order in which the streams should sorted so that in the list for a stream all its upstream streams located above it and the downstream ones are located below it. This is necessary to route flows correctly because it makes sure that before routing the flow from a subbasin outlet, all the flows upstream to it have reached it. Downstream Line Identification number is useful in knowing which stream is downstream of a given stream. Outlet attribute is useful because it provides the starting point of routing flows from a sub-basin to the downstream sub-basin.

During the first phase of the execution of the algorithm, a query is performed on the stream lines such that those that have a value of 1 for Outlet attribute are sorted in the ascending order of Rank values. In other words outlet streams are sorted in the ascending order of their Rank values. Once this is achieved, in the next phase of the algorithm, an outlet with least rank value is picked, a check is performed to see if reservoir exists, if a reservoir exists the flows are routed through the reservoir and then are routed to the next downstream sub-basin and they are added to the flows of this sub-basin. If reservoir does not exist, the flows are directly routed to the downstream sub-basin. This procedure is continued until the outlet with the greatest rank is reached, meaning the flows in all the sub-basins are routed to the watershed outlet.

5 APPLICATION, RESULTS AND DISCUSSIONS

Two case studies have been presented in this section to evaluate the tools developed. For the first case study, which deals with application of tools developed for dendrification, treatment of data inconsistencies, hydrologic topology and network parameters, region 12 has been chosen as the study area. For the second case study, which deals with the application of methods dealing with hydrologic topology, network parameters and hydrologic modeling, Bull Creek watershed upstream of USGS gauging station number 08154700 located near loop 360 in North Austin, Texas, was chosen.

5.1 CASE STUDY-1

5.1.1 HUC Data

To evaluate the procedures developed for dendrification, treatment of data inconsistencies and hydrologic topology computations, Region 12 was chosen (Fig 5.1). Major part of the region lies in the state of Texas, though other states like New Mexico and Louisiana share a very little part of it. The watershed and streams data were obtained from USGS (USGS, 2003d).

The cataloging units for region 12 were downloaded from a USGS website (http://water.usgs.gov/lookup/getspatial?huc250k). The data downloaded was in interchange format (.e00), which contained HUC data in coverage format. ArcToolBox was used to import the e00 files. Once the coverages were imported, ArcMap was then used to import the shapefile from the coverage.

The downloaded HUCs data came in a Geographic Decimal Degrees projection with datum NAD 83. To maintain consistent projection for both HUCs and NHD, Universal Transverse Mercator was used as the spatial reference system. ArcToolBox was used to reproject the data. The new projection characteristics are as follows:

- Projection: Universal Transverse Mercator (UTM) zone 14.
- Datum: North American Datum 1983 (NAD 1983)
- Geographic System: North American Geographic Coordinate System 1983

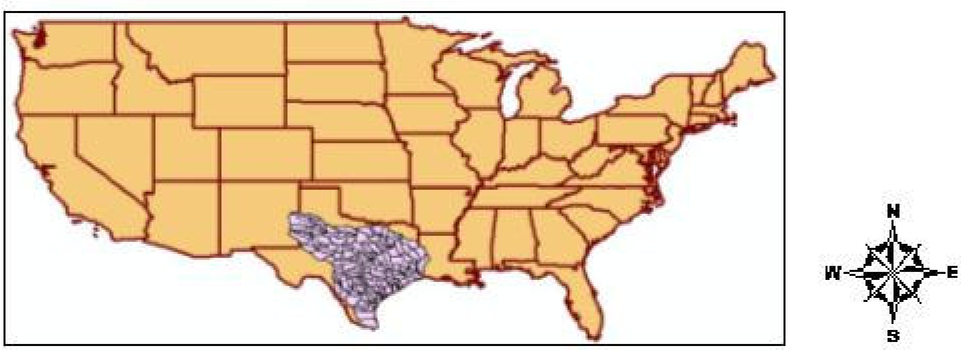


FIG. 5.1. Location of Region 12

The HUCs dataset is accompanied by pertinent attribute information such as region number, sub-region number, accounting unit number and cataloging unit number for each polygon (see fields REGION, SUBREGION, ACCTUNIT, HYDROUNIT in Table 5.1.). To attribute each polygon with a unique number, these numbers were combined as HUC number (see field HUC in Table 5.1.). Though there is some other information given in the attribute table, but was not used for analysis.

TABLE 5.1. Attribute table for Region 12 HUCs

Attributes of region12 polygon

FID	Shape	AREA	PERIMETER	REGION1	REGION1	HUC	REGIO	SUBREGI	ACCTUNI	HYDROU
2	Polygon	4150414336	471908.09375	2	1713	12050005	12	05	00	05
3	Polygon	9774125056	536077.5	3	1719	12050001	12	05	00	01
4	Polygon	4064963840	394982.5625	4	1736	12050002	12	05	00	02
5	Polygon	4345378816	468712.625	5	1763	12050006	12	05	00	06
6	Polygon	6147979264	440868.75	6	1806	12080001	12	08	00	01
7	Polygon	2764903680	330820.625	7	1810	12050003	12	05	00	03
8	Polygon	5605476864	531577.125	8	2253	12050007	12	05	00	07
9	Polygon	7066213888	699754.75	9	2255	12050004	12	05	00	04
10	Polygon	4800330240	367190.75	10	1823	12030103	12	03	01	03
11	Polygon	6437430272	458888.625	11	2258	12060101	12	06	01	01
12	Polygon	1858903680	280036.09375	12	1832	12030104	12	03	01	04
13	Polygon	5090848256	377569.03125	13	1837	12030101	12	03	01	01
14	Polygon	4490822144	523297.09375	14	1842	12080006	12	08	00	06
15	Polygon	6850724864	421053.5	15	1843	12080003	12	08	00	03
16	Polygon	3361653504	313541.375	16	1844	12030106	12	03	01	06
17	Polygon	6781150208	666044	17	1846	12080004	12	08	00	04
18	Polygon	8167997952	493249.09375	18	1850	12060201	12	06	02	01
19	Polygon	3545782016	315513.1875	19	1857	12010001	12	01	00	01
20	Polygon	2812622592	255383.9375	20	2259	12060103	12	06	01	03
21	Polygon	1655281024	229519.1875	21	1866	12060104	12	06	01	04
22	Polygon	6980989440	439960.96875	22	1868	12080002	12	08	00	02
23	Polygon	3564656384	457761.96875	23	1871	12030105	12	03	01	05

Record: 0 Show: All Selected Records (0 out of 122 Selected.) Options

5.1.2 NHD Data

"The National Hydrographic Dataset is a comprehensive set of digital spatial data that contains information about surface water features such as lakes, ponds, streams, rivers, springs and wells" (USGS, 2003b). The data was compiled by USGS in collaboration with EPA, which supersedes previously developed river reach files. Other than NHD, different versions of reach files are RF1 and RF3.

USGS and EPA maintain a separate website for downloading NHD data, from which the NHD data for Region 12 was obtained (http://nhd.usgs.gov/). These data is available in coverage format. There are three different options for downloading data: High, Local and Medium, which are based on resolution. As local and high versions were not completely developed for region 12, medium resolution data was downloaded. As NHD data was available one file per cataloging unit, a number of files had to be merged to create a single file for region 12. In region 12, as there are 122 cataloging units, 122 NHD files have been processed.

All of the 122 files were downloaded in interchange format which were then imported into coverages. Each coverage file contained a line coverage feature class representing streams. These line coverage classes were then exported as shapefiles using ArcMap. Once the shapefiles were extracted, they were then merged using Geoprocessing wizard’s merge tool. This is how one single shapefile containing NHD streams was generated for region 12.

Apart from the inconsistencies in NHD data discussed in the literature, there were some others which had to be solved manually. Among the type of problems faced consisted of some streams which were not broken at the confluences, two or more stream lines overlapping each other, and streams disconnected from the network. These problems were frequently found in the bay regions, so the data pertaining to the bay areas was not considered for analysis, i.e., related NHD lines and HUCs polygons were deleted from the dataset.

The downloaded data had a geographic projection: GCS_North_American_1983. The data was then projected using ArcToolBox, into the coordinate system which is the same as that of the HUC dataset. The new projection has the following spatial reference properties.

- Projection: Universal Transverse Mercator (UTM) zone 14.
- Datum: North American Datum 1983 (NAD 1983)
- Geographic System: North American Geographic Coordinate System 1983

Though the data contained some attribute information, the information was not used for application.

At this point, the data needed for testing dendrification, data inconsistencies and hydrologic topology tools was compiled. The next step was to apply these tools on the data. As it is not easy to comprehend the results obtained for complete region 12, San Antonio basin was selected to explain the application and results. Though the results are explained only for San Antonio river basin, the complete set of results for region 12 has been provided in electronic format.

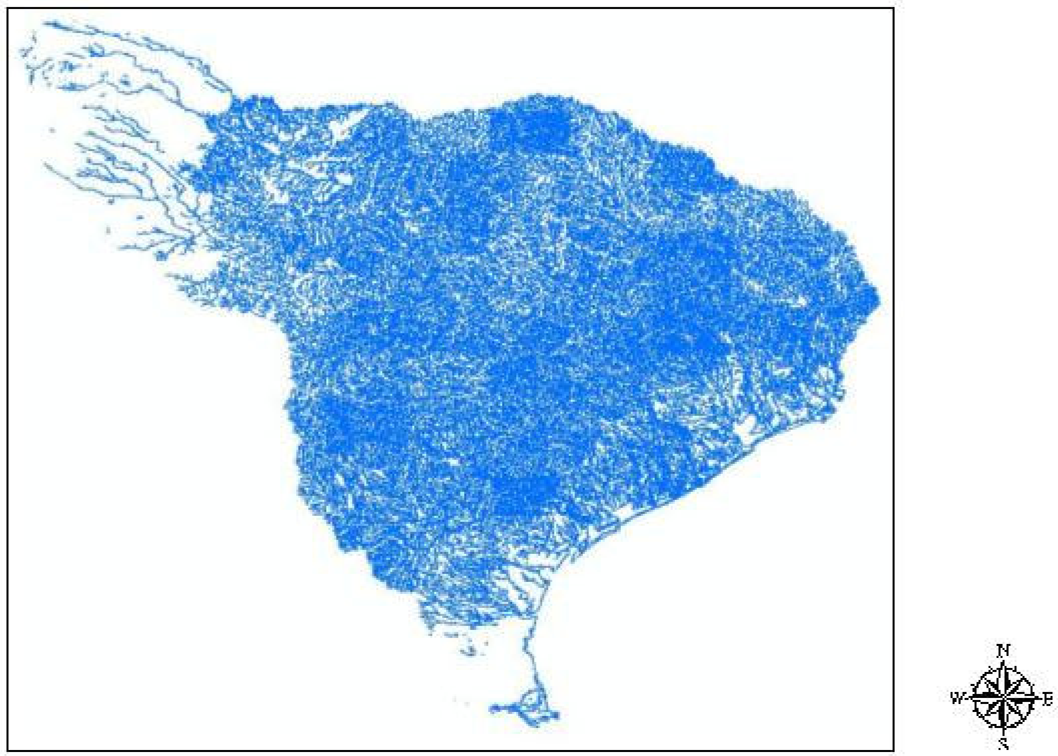


FIG. 5.2. NHD Streams for Region 12

5.1.3 San Antonio River Basin Data

San Antonio river basin is located in South Central Texas (Fig. 5.3.). This basin consists of Medina River, Medio creek, Leon creek, Salado creek, Cibolo creek and San Antonio River. There are four HUCs corresponding to this basin, bearing numbers 12100301, 12100302, 12100303 and 12100304. As can be observed from the pattern of these numbers, 12 corresponds to region number, 10 corresponds to sub-region number

03 corresponds to accounting unit number which are common to all HUCs. The last two digits in each of these numbers correspond to their Cataloging units.

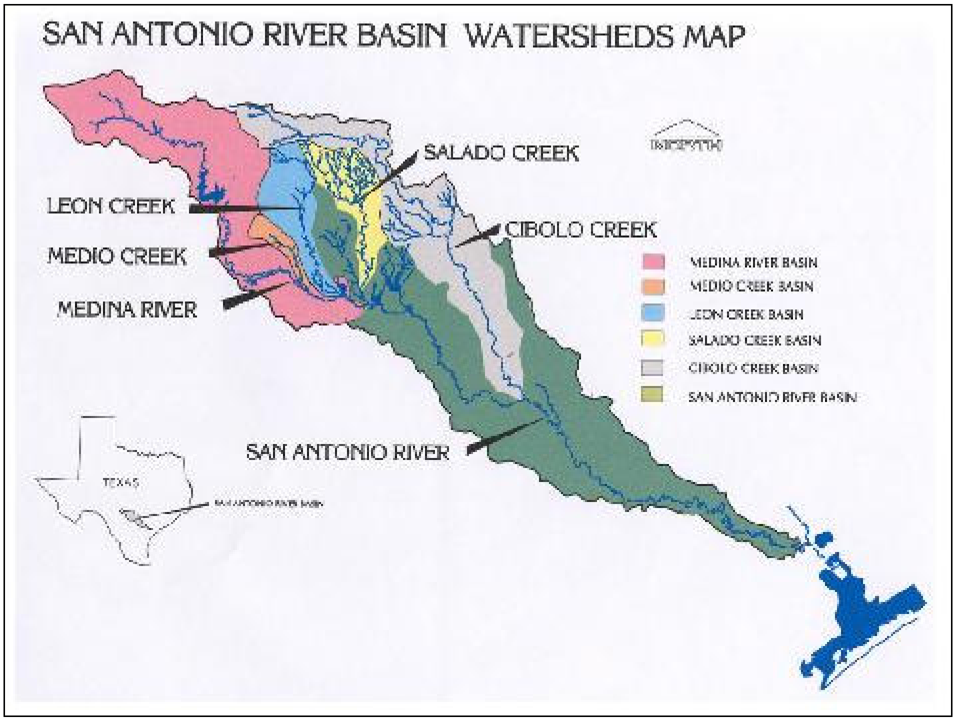


FIG 5.3. San Antonio River Basin (http://www.sara-tx.org/)

The extraction of HUC data for San Antonio River basin from already compiled region 12 data was accomplished using ArcMap, by simple export operation on the selected HUCs, basis for which was their cataloging unit numbers. The stream network, i.e. NHD streams data for the basin was obtained by intersecting the region 12 stream network with the four HUC polygons.

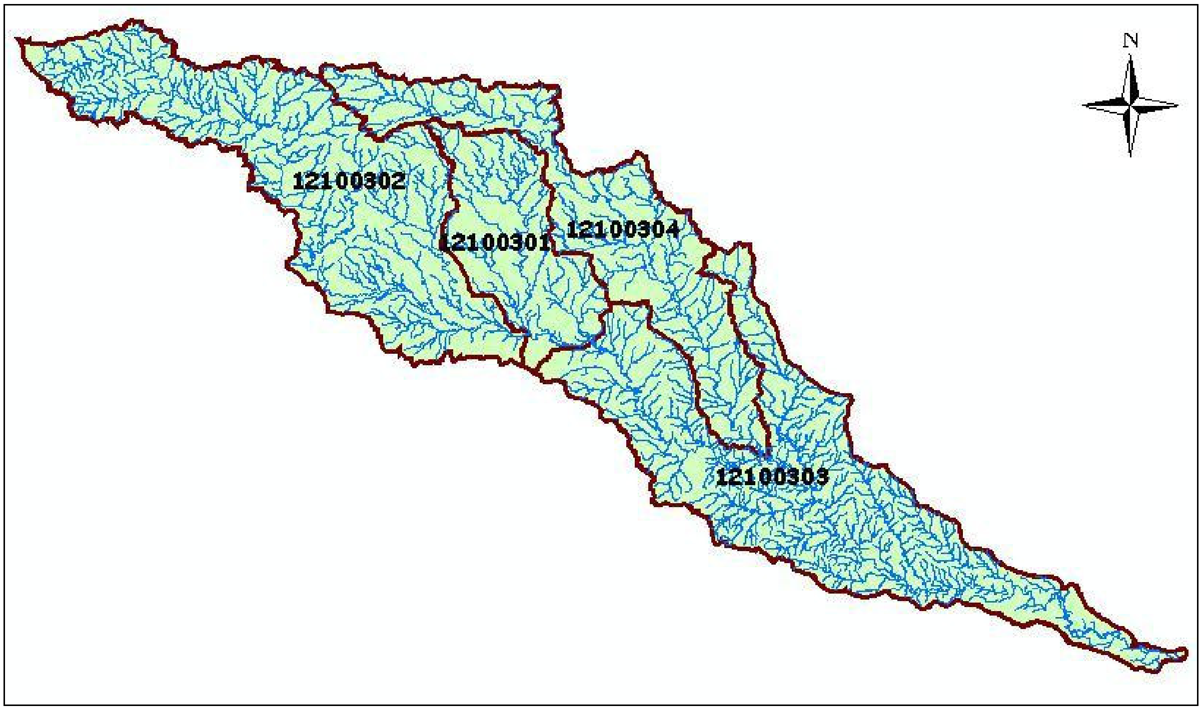


FIG. 5.4. San Antonio River Basin Data

5.1.4 Dendrification

In the NHD stream network obtained from USGS, disorientation of streams and the existence of loops was detected. So the dendrification tool was used to make the network dendritic. Dendrification tool has two options, one for correcting only loops (i.e. braided streams) and the other for correcting both loops and stream line orientations (Fig 5.5). The choice of option depends upon whether line orientation for all the streams is correct or not. When using 'Correct Braided Streams' option, the user should be absolutely sure that all the stream lines are correctly oriented.

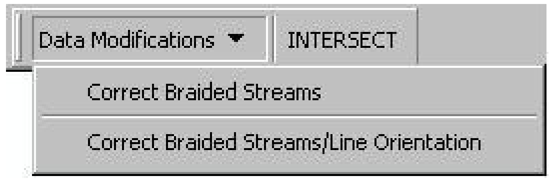


FIG. 5.5. Dendrification Tools

To apply these tools, the user is required to select all the possible outlets in the stream network. In the case study of San Antonio basin, there was only one outlet stream to be selected. Figures 5.6 and 5.7 have two images. In each of these figures, the image located above the arrow shows features of the network that make it non-dendritic and those below show those parts of the stream network after dendrification tools have been applied on the network. In figure 5.6 the red colored circles highlight the loops and in figure 5.7, the red colored circle highlights the outlet and the square in the image above the arrow represents wrongly oriented lines and the square in the image below the arrow represents the correctly oriented lines.

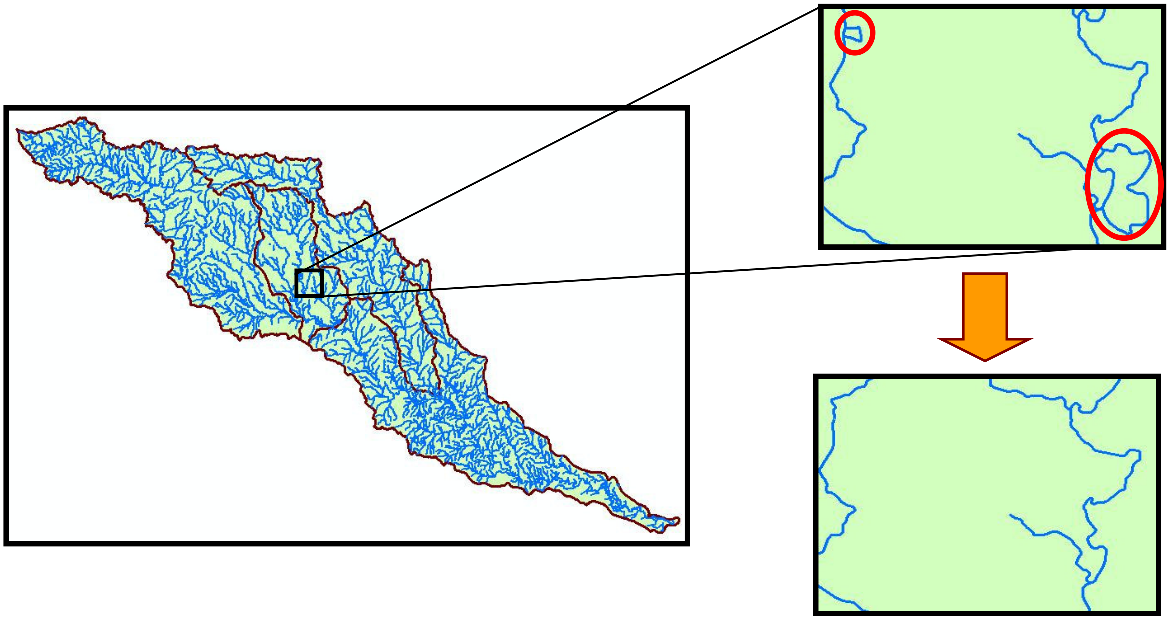


FIG. 5.6. Loop Correction

In figure 5.7, in both the images, the outlet stream is highlighted with red colored circles, which are correctly oriented. The image above the arrow shows streams with wrong orientations being highlighted by red colored square. Upon application of dendrification tool on the network, all the lines highlighted by red colored squared are correctly oriented as shown in the image below the arrow mark.

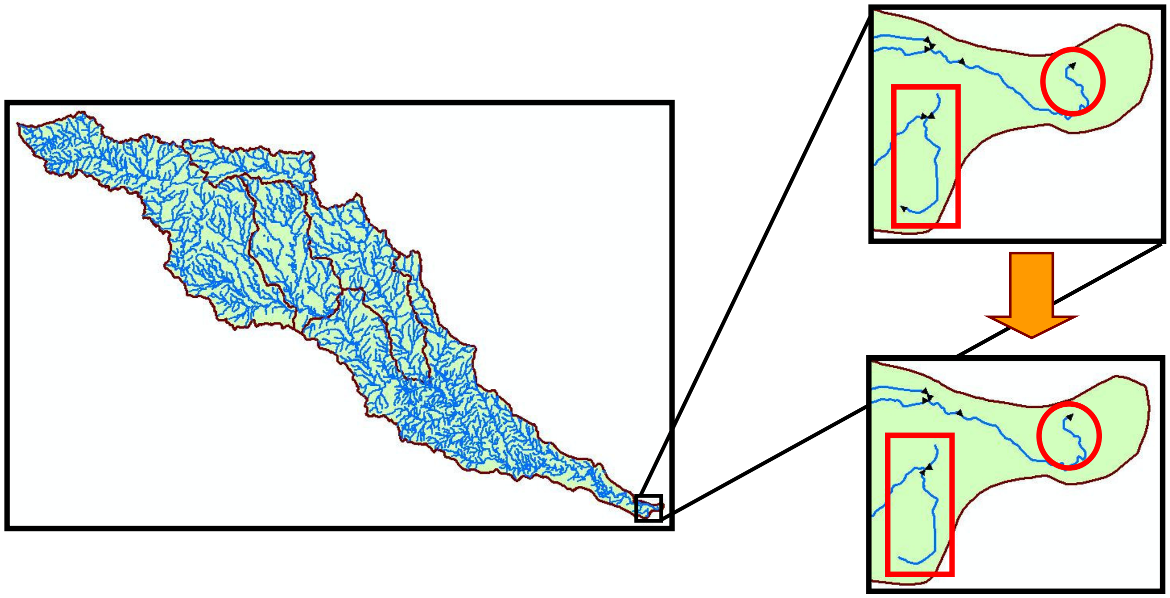


FIG. 5.7. Line Orientations Correction

This completed the first step towards making the data compatible with the model requirements, which is dendrification. At this stage the NHD stream network was made free of loops and the stream lines were correctly oriented, so that they flow towards the outlet stream. The next step in the application was to intersect the NHD network with HUCs and to build topology for the streams in the network so that a stream line lies completely within a polygon and each line knows it from-node and to-node numbers.

5.1.5 Intersection and Building Topology

The stream network was treated for loops and line orientation, so that it is converted into a dendritic network. The NHD stream network was intersected with the HUC polygons to make each line know as to which polygon it belonged to. This was done using the intersection tool that automated the process of intersection Due to intersection, the number of lines increased as some of them were broken at the HUC boundaries to form a new network, which were also attributed with the HUC number. After the intersection, the tool also attributes each line with a unique identification number (see Recno field in Table 5.2), from-node number (see FNODE field in Table 5.2), to-node number (see TNODE field in Table 5.2) and geometric length (see Length field in Table 5.2), which is called building topology. In this application, the units for length are chosen to be meters.

TABLE 5.2. Stream network attribute table after intersection and building topology

Attributes of SanAnnNHD_prj3

FID	Shape*	Recno	Length	FNODE	TNODE	HUC
0	Polyline	1	442.726505	1	2	12100303
1	Polyline	2	2244.725491	3	1	12100303
2	Polyline	3	3844.983021	4	5	12100303
3	Polyline	4	5227.335422	5	2	12100303
4	Polyline	5	721.310470	6	5	12100303
5	Polyline	6	137.750128	7	6	12100303
6	Polyline	7	2720.064449	2	8	12100303
7	Polyline	8	448.107458	9	8	12100303
8	Polyline	9	107.940736	8	10	12100303
9	Polyline	10	6263.864326	11	12	12100303
10	Polyline	11	1238.493277	13	11	12100303
11	Polyline	12	2895.090024	10	13	12100303
12	Polyline	13	206.617938	14	13	12100303
13	Polyline	14	2014.776480	15	9	12100303
14	Polyline	15	516.005005	16	17	12100303
15	Polyline	16	2453.694647	18	15	12100303
16	Polyline	17	2341.811979	19	18	12100303
17	Polyline	18	5066.146473	20	14	12100303
18	Polyline	19	3076.940770	17	21	12100303
19	Polyline	20	5065.803659	22	18	12100303
20	Polyline	21	2046.810113	21	22	12100303
21	Polyline	22	644.502591	23	24	12100303
22	Polyline	23	8958.278124	24	21	12100303

Record: 4 Show: All Selected Records (0 out of 2278 Selected.) Options

5.1.6 Treatment of Data Inconsistencies

Before the procedures developed for applying data inconsistency tools could be applied, it was important to define the threshold value for the maximum number of elements to trace and the maximum threshold length. In this application, the threshold value for number of elements was chosen to be 10 and the threshold value for length was chosen to be 3000m. When the methods for treating inconsistencies were applied, two comma delimited files were created. One of them provides a list of numbers of streams posing to be potential outlets, before the treatment of inconsistencies and the other provides the list of sub-basin numbers for which the outlet stream number could not be determined.

In the San Antonio basin case study it was found that there were two sub-basins bearing identification numbers 12100303 and 12100304 that had more than one stream to resolve for, to find the real outlet stream and the other two had one outlet stream. For these two sub-basins, there were 11 potential outlet streams. A table was generated with sub-basin and the streams information. Table 5.3 shows the sub-basin numbers in the field that is titled WShedID and the potential outlet streams in the field that is titled StreamIDs.

TABLE 5.3. Potential outlet streams table

WShedID	StreamIDs
12100303	10
	679
12100304	1217
	1219
	1220
	1223
	1247
	1259
	1284
	1312
	1345

After the data inconsistency treatment program was applied, it was found that the number of candidate outlet streams was reduced from 11 to five instead of four. Three sub-basins were correctly assigned their outlet stream number but for one of them. The identification number of the lone sub-basin, for which outlet could not be resolved was provided in a comma delimited file as well as in the sub-basin attribute table. Sub-basin, bearing number 12100304 (see table 5.4) was the one for which data inconsistency could not be solved, so -1 was assigned to its outlet stream number attribute (see the field NewOut in table 5.4).

TABLE 5.4. Data inconsistencies table

Attributes of SanAnHUC_Prj1

FID	Shape*	HUC	NewOut
0	Polygon	12100303	10
1	Polygon	12100301	1010
2	Polygon	12100304	-1
3	Polygon	12100302	1538

Record: 1 Show: All Selected Records (1 out of 4 Selected.)

When the problem was analyzed, it was found that there existed a special case, which could not be handled by the algorithm. There were two candidate stream lines crossing the sub-basin boundary; one was the real outlet stream and the other one was similar to case 1 type of data inconsistency, but instead, the stream that crossed over had another stream upstream of it (Fig 5.8). In Fig 5.8, it can be seen that stream bearing number 1247 has a stream bearing number 1248, upstream of it. This is the reason why it could not be treated as case 1 by the algorithm. This data inconsistency was solved manually by editing the table.

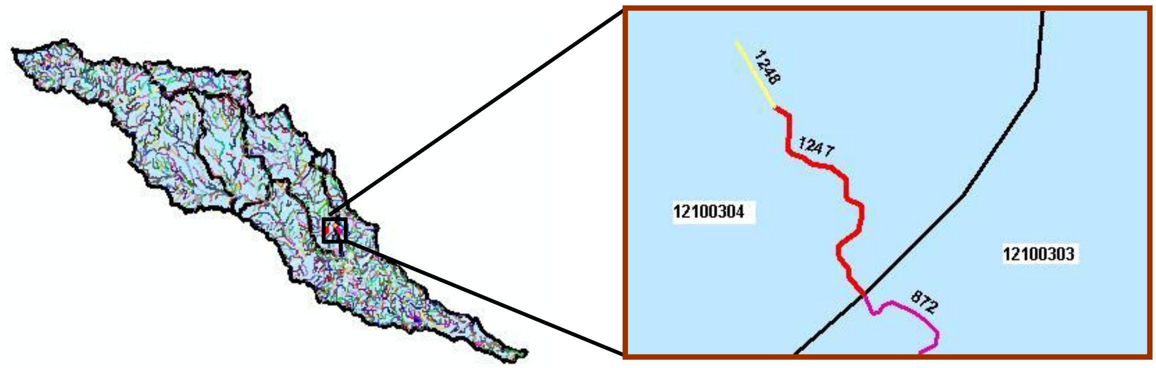


FIG. 5.8. Irresolvable Data Inconsistency Case

5.1.7 Hydrologic Topology and Network Parameters

Hydrologic topology tools were then applied to the San Antonio river basin. As previously discussed, the NHD stream data and HUC sub-basins data were made suitable for hydrologic parameter functions. Upon completion of these data preprocessing steps, each stream in the network was assigned most of the required attributes ((i.e., Stream ID, Watershed ID, Length, Upstream Node and Downstream Node for the stream lines, and Watershed ID and Area for the watershed polygons) except the optional attributes. The optional attributes such as Length weight for all lines and Area weight for all polygons were assumed equal to 1. These threshold values were not required to be attributed to the line and polygon features because the program takes value of 1 by default, unless specified.

Because of the difficulty to display results of the hydrologic topology tools for the whole watershed, maps showing patterns in the calculated parameters of both streams and sub-basins are provided.

The intermediate parameter called rank was first calculated for the streams in the network. It can be observed from Fig 5.9, the outlet stream of the basin has the highest rank, which is equal to 205 (darkest and most downstream line), the head waters have a rank equal to 1 and the rest lie between 1 and 205.

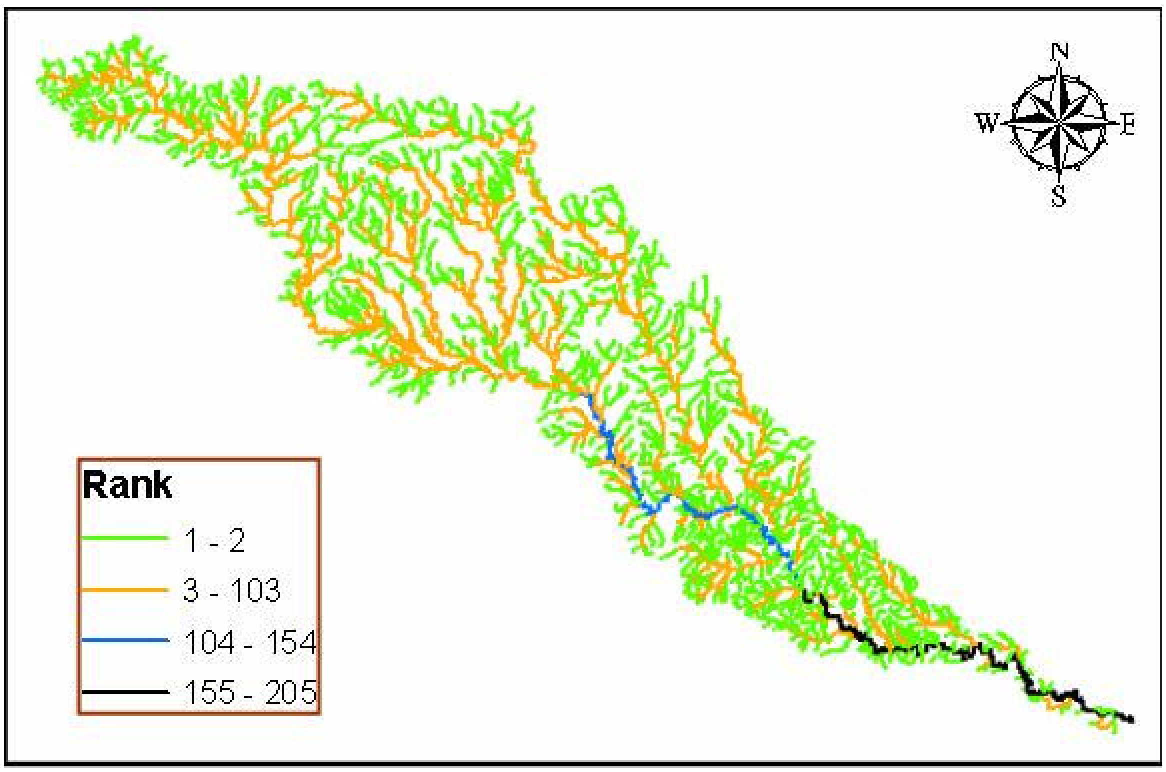


FIG. 5.9. Map Showing the Rank Pattern

In the case study, a length weight equal to one was used, so, upstream length and the weighted one have same value. Upstream lengths of stream lines vary similar to the rank values and it can be observed by comparing Fig 5.9 and Fig 5.10, i.e. the values increase in upstream to downstream fashion. The outlet stream of the whole basin has the highest value for upstream length, which is equal to 595.9 Km (i.e. 595900 m) and each head water stream has an upstream length equal to its own length.

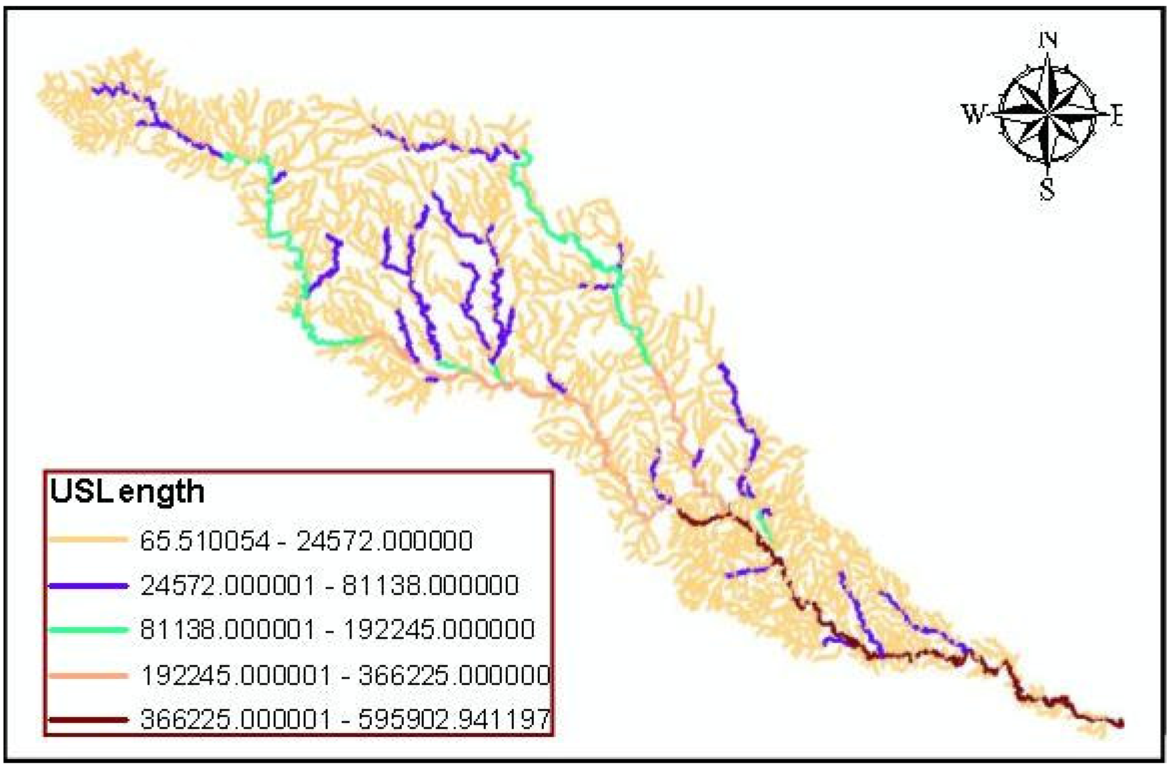


FIG. 5.10. Map Showing Upstream Lengths (in meters)

Similarly, the streams were also attributed with downstream lengths and the weighted counterparts (Fig 5.11). Downstream lengths increase in downstream to upstream fashion. The outlet stream of the whole basin has the lowest value for downstream length, which is equal to 0 as length of the stream is not considered when computing downstream lengths. Head water streams have highest set of values for downstream lengths. Apart from length based parameter computations, streams that represent outlets of sub-basins were flagged, by attributing a value of 1 to them and 0 to the others. Using these values the algorithm also computed accumulated area and its weighted counterpart values for the streams, which were then transferred to the sub-basins (Fig 5.12).

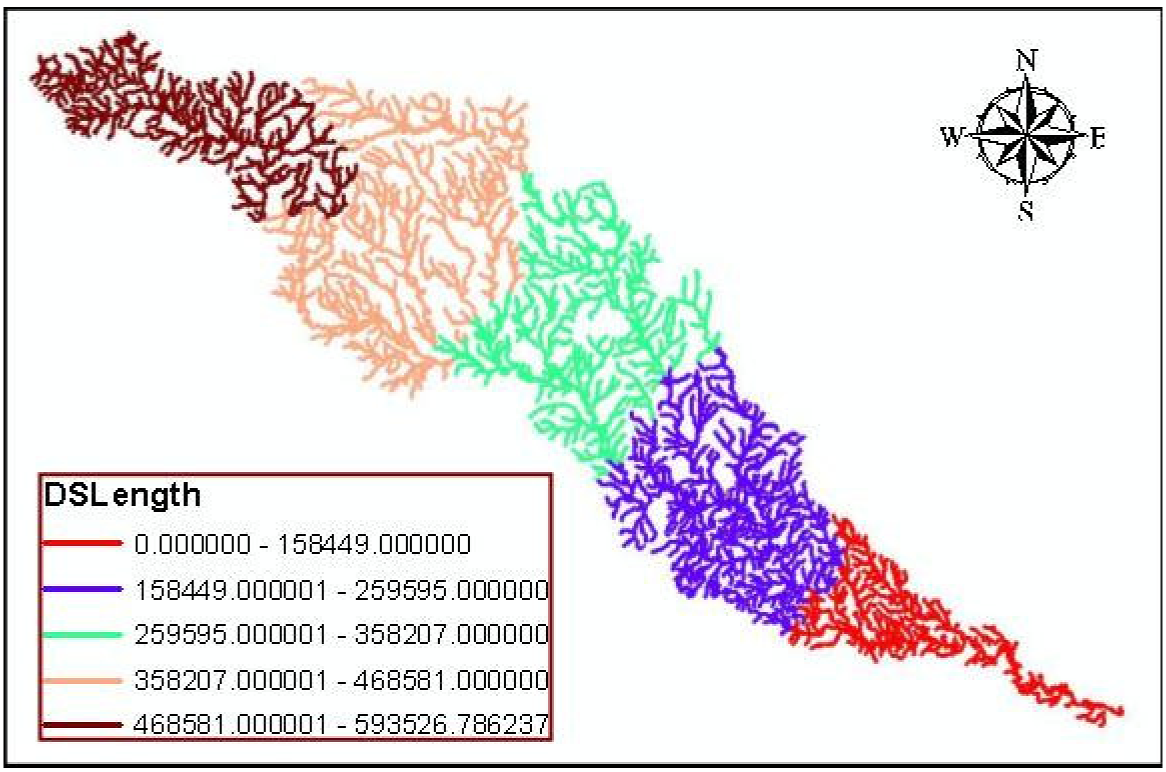


FIG. 5.11. Map Showing Downstream Lengths (in meters)

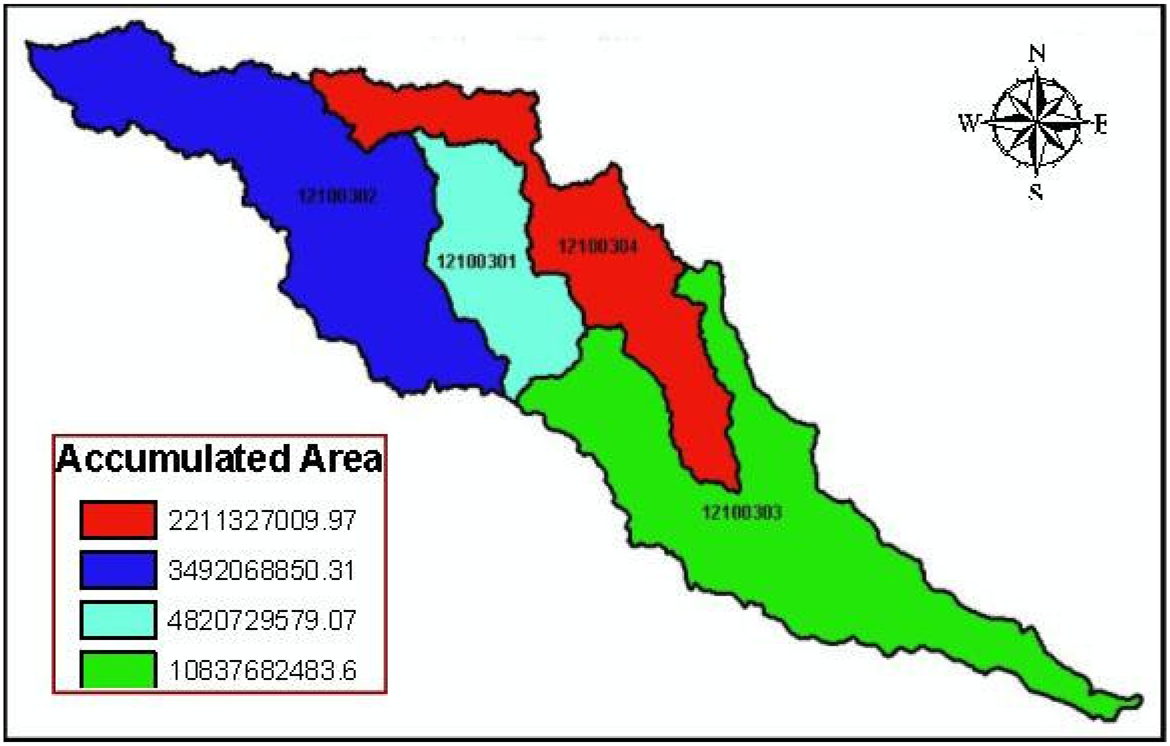


FIG. 5.12. Map Showing Accumulated Areas (in Sq.meters)

Once network parameters are computed for all the elements in basin, streams shapefile and the sub-basin shapefile were exported as feature classes to a feature dataset contained in a geodatabase (MacDonald 2001) (Fig 5.13). The design of this geodatabase has been according to the norms set up by ArcHydro. ArcHydro is a standard GIS-based hydrologic data model for storing geospatial and temporal data designed by University of Texas at Austin, Texas and ESRI (Maidment et al. 2002). The geodatabase can have any name, but there are standard names for the feature dataset and feature classes. Feature dataset is named 'ArcHydro', the streams feature class is named 'HydroEdge' and sub-basins feature class is named as 'Watershed'. The advantage of

ArcHydro is that, because it is a standard way of storing data, the information becomes exchangeable between different programs.

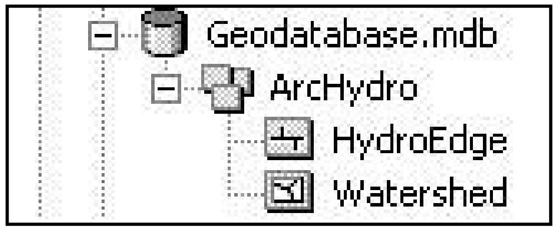


FIG. 5.13. ArcHydro Feature Dataset

Now that the network parameters were computed, the next step was to test tracing tools, namely, downstream trace and upstream trace.

Before going into details of application of tracing tools, it is important to mention the various options available with the program to perform tracing (Fig 5.14). If the user wishes to skip the parameter computation part and wants to just perform tracing, it can be done if, for each stream in the network, its downstream line number is known. There are different ways in which user can provide the input stream line information for which tracing is needed. For tracing downstream, input can be provided from the console as line numbers or by selecting the features from the map, whereas for tracing upstream, including the input options of downstream tracing has an additional option, which is inputting the line numbers in a text file. Once the input is provided, tracing operation can be performed and the resultant trace lines are selected automatically, which can be exported into new datasets if needed. These are various options for input and output.

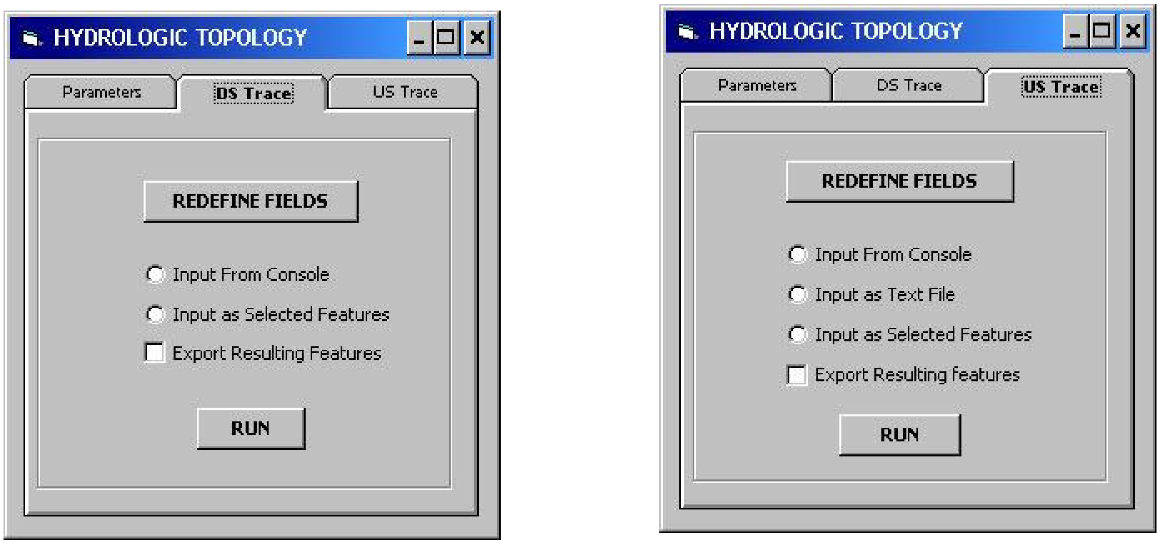


FIG. 5.14. Tracing Tool Wizards

In the application to the San Antonio basin, tracing was performed on few features. For tracing downstream, three lines were selected bearing numbers: 1504, 1762 and 1008. Fig 5.15 shows the result of application of downstream trace tool.

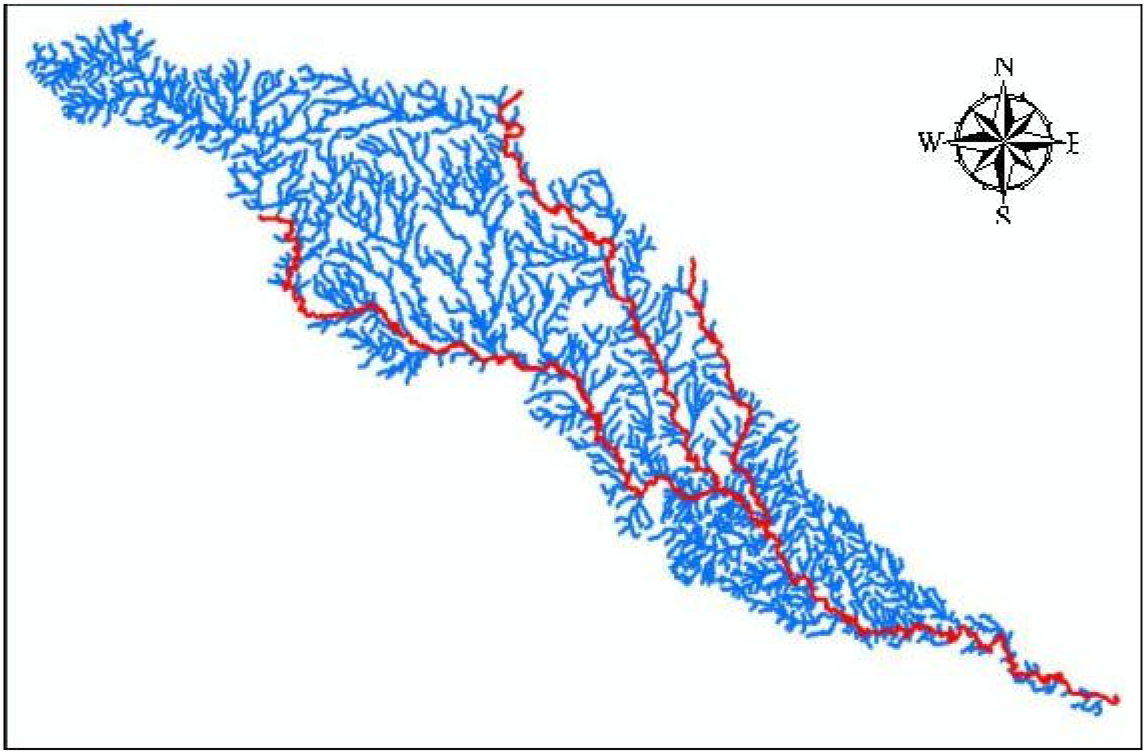


FIG. 5.15. Downstream Trace of Three Lines in the Network

For tracing upstream, three lines were chosen bearing numbers: 1945, 1583 and 900. Fig 5.16 shows the result of application of upstream trace tool.

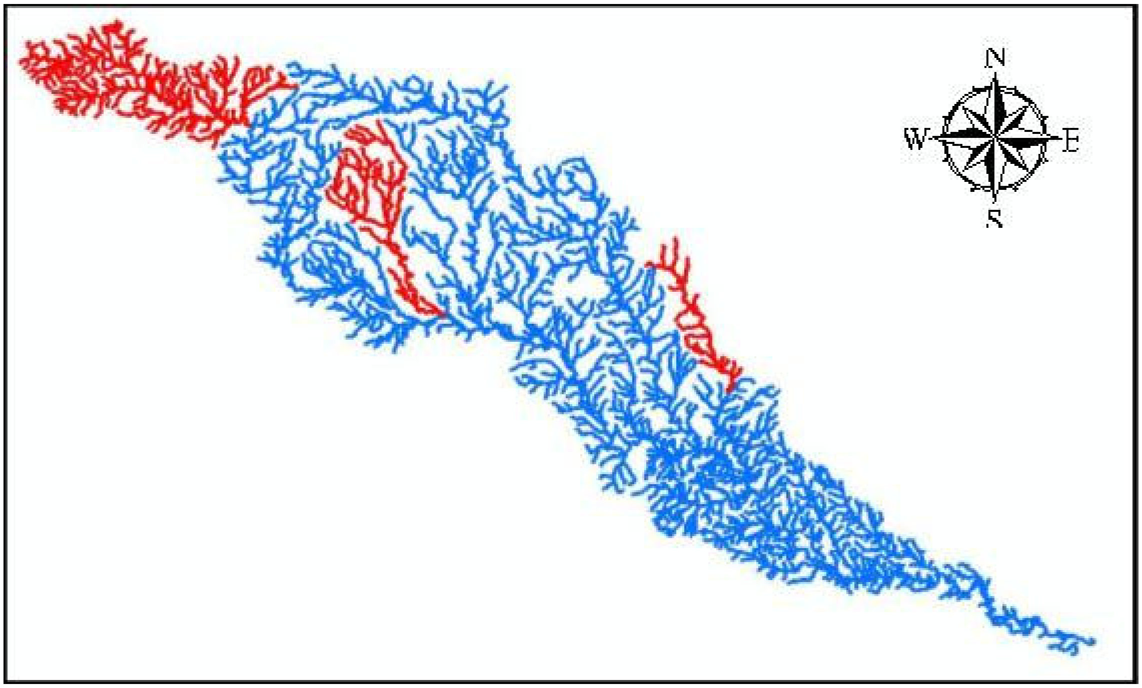


FIG. 5.16. Upstream Trace of Three Lines in the Network

This concludes the first part the section, which deals with the application of tools dealing with data compatibility, hydrologic topology and network tracing. The next case study aims at showing the application of hydrologic model to Bull creek watershed.

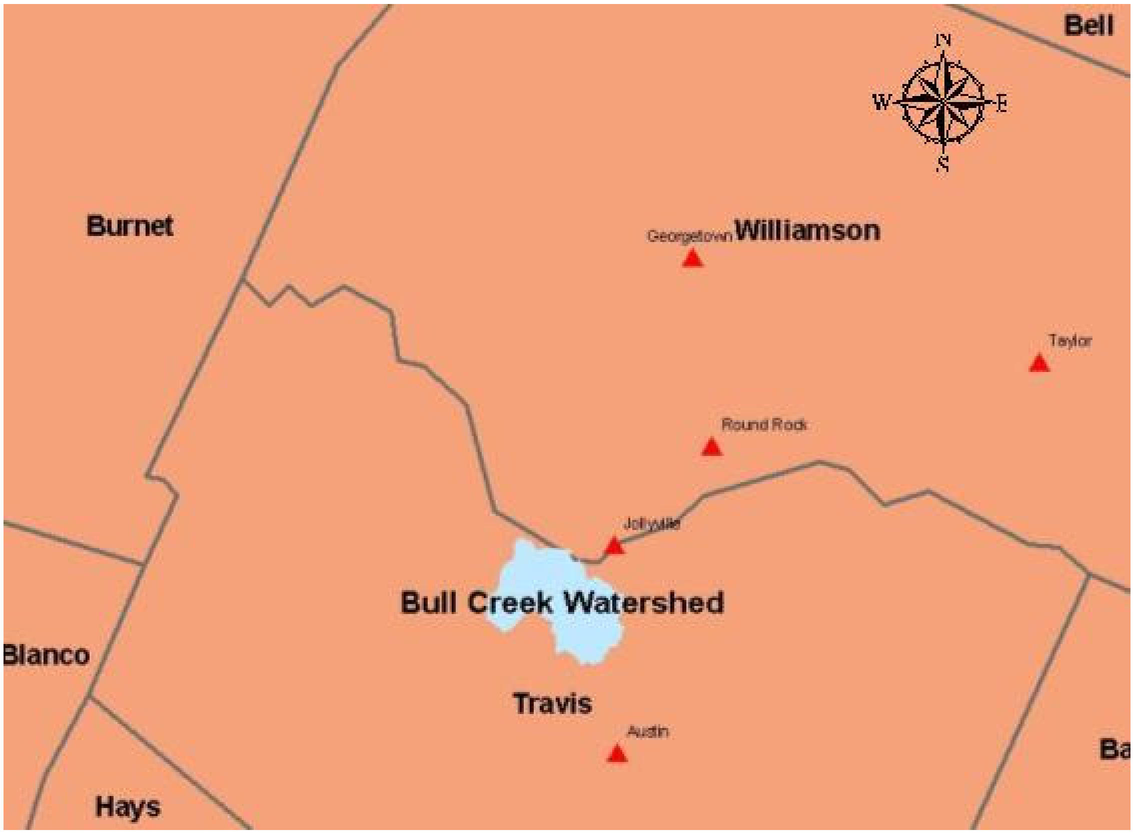


FIG. 5.17. Bull Creek Watershed County

5.2 CASE STUDY -2

This case study shows the application of hydrologic model to Bull Creek watershed. Bull creek watershed is located north west of Austin, which has a total area of 32 square miles. The part of watershed, which is upstream of highway 360, was chosen for the case study because there is a USGS gauging station at this location (USGS 08154700). The chosen part of the watershed has an area of 22.3 square miles. The entire watershed lies in Travis County (Fig 5.17.).

In this case study, the Soil and Water Assessment Tool (SWAT) was used for computing the excess precipitation needed by the hydrologic model. The data required consisted of the digital elevation model, land use data, soils data and precipitation. To create the input files needed by SWAT, the preprocessor called AVSWAT was used (Di Luzio et al. 2000). A flow chart showing the application methodology is shown in Fig 5.18a and Fig 5.18b.

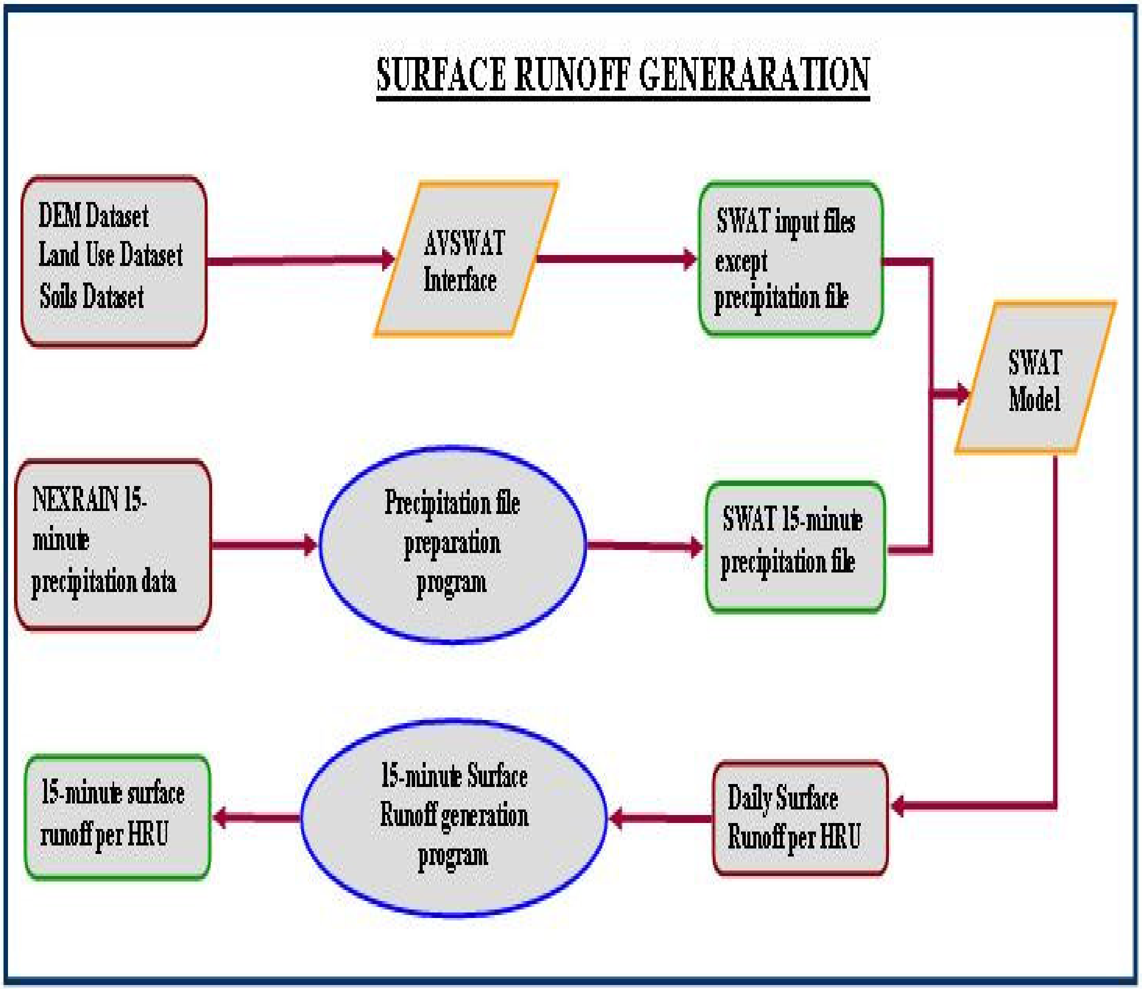


FIG. 5.18a. Surface Runoff generation Using SWAT

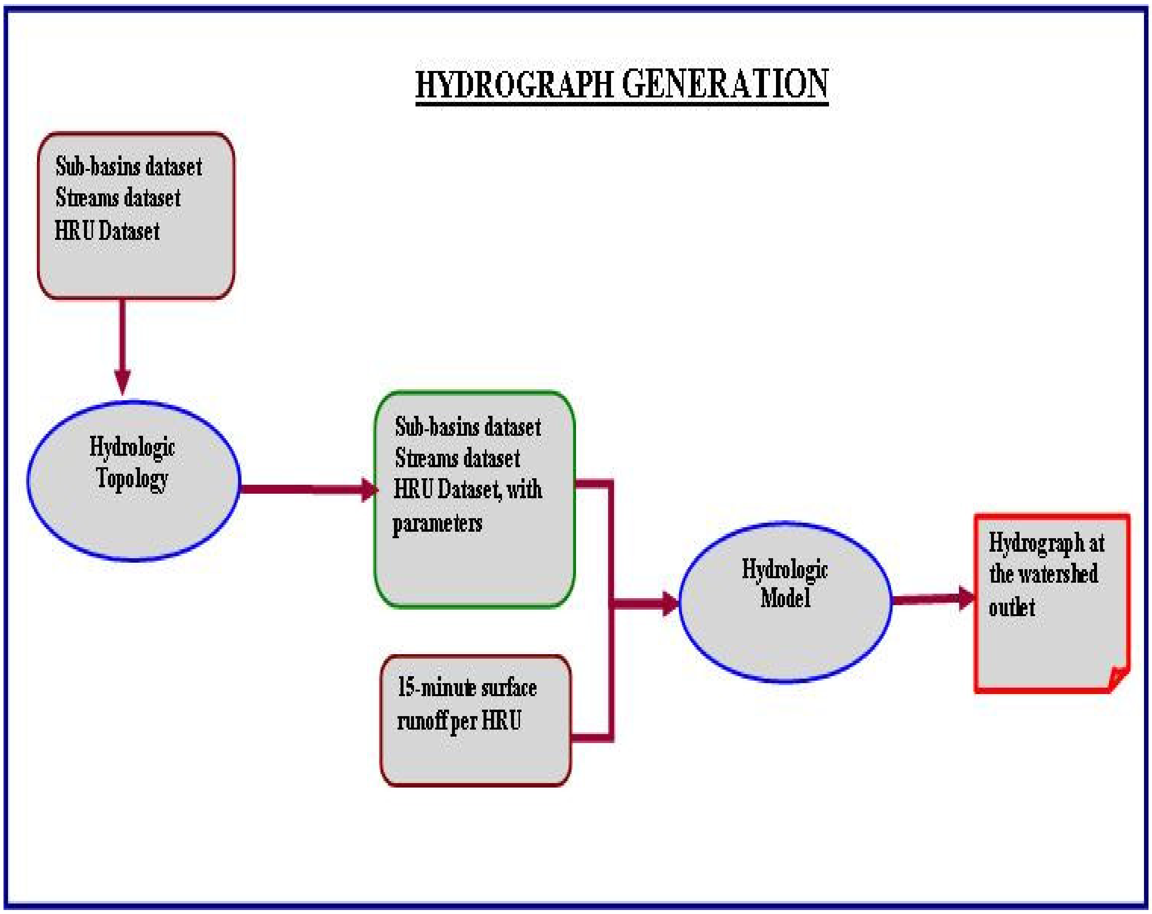


FIG. 5.18b. Hydrograph Generation Using the Hydrologic Model

5.2.1 Data Description

5.2.1.1 Digital Elevation Model

The streams and basin information was obtained using a 10m DEM (Fig 5.19). The elevation data was downloaded from GeoCommunity website (http://data.geocomm.com/). This site contains a number of geospatial datasets, which includes USGS products also. Three 7.5 minute tiles had been downloaded to cover the whole study area. These tiles came with a resolution of 10m (USGS, 2003a). These tiles had the following spatial reference properties:

- 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

As the tiles downloaded were in USGS DEM format, they were converted to ERSI GRID format using ArcToolBox. All the DEMs were then merged to produce a single DEM file using Spatial Analyst extension's Map Calculator function in ArcGIS.

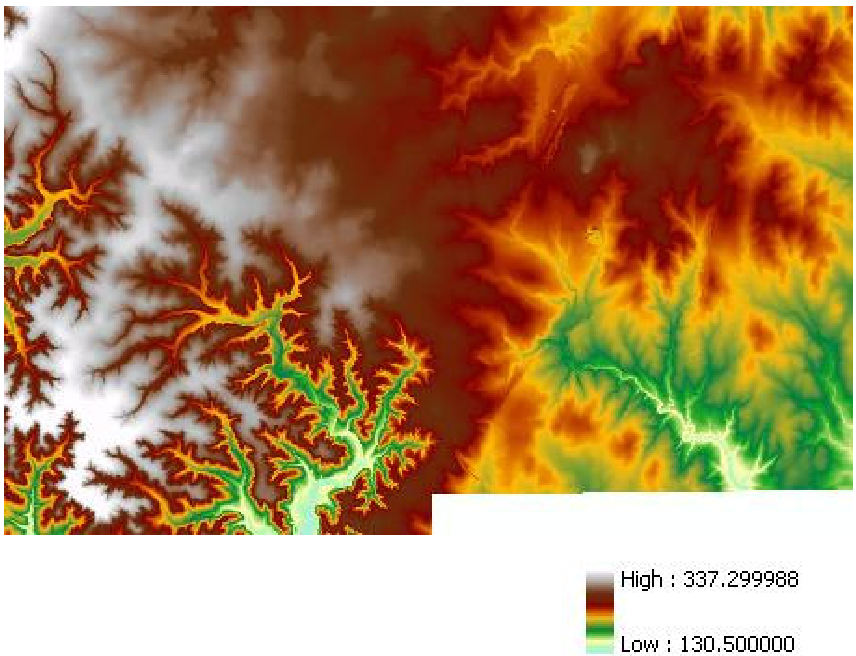


FIG. 5.19. 10m DEM for Bull Creek Baisn.

5.2.1.2 Soils Data

STATSGO Soils data was obtained from USDA (http://www.ncgc.nrcs.usda.gov/). 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

In this case study, the Texas modified STATSGO dataset was reprojected using ArcToolBox to Universal Transverse Mercator (UTM) zone 14.

5.2.1.3 Land Use Data

For the case study, National Land Cover Dataset (NLCD) was chosen to as Land use data (USGS, 2003c). The data was downloaded using the USGS Seamless Data Distribution System (http://seamless.usgs.gov/viewer.htm) in ESRI GRID format.

The downloaded data comes in a Geographic Decimal Degrees projection with datum NAD 83. To maintain consistency in spatial reference properties among the datasets used in the case study, the land use data was re-projected using ArcToolBox. The new projection used was Universal Transverse Mercator (UTM) zone 14.

5.2.1.4 Precipitation Data

Sub-daily precipitation data was used for the case study. 15 minute precipitation data was obtained from NEXRAIN Corporation (http://www.nexrain.com/). NEXRAIN data is gage-adjusted radar rainfall. NEXRAIN data is available in two formats: NEXRAIN-1k and NEXRAIN-2k. NEXRAIN-1k dataset consists of 5 min interval rainfall data, available at 1km radar resolution and NEXRAIN-2k dataset consists of 15 min interval rainfall data, available at 2km radar resolution For the hydrologic model NEXRAIN-2k data was used. The dataset came in polygon shape file format containing radar rainfall pixels forming a mesh, in which each pixel was attributed with a unique number. This dataset was accompanied by a folder containing rainfall data files. Each file contained rainfall data for all the pixels for a single day. Each column in a file contained rainfall data pertaining to a pixel, headed by a pixel number.

5.2.1.5 Flow Data

To calibrate and validate the SWAT model and the hydrologic model four different flow data files were required. These flow data files were obtained from USGS (http://waterdata.usgs.gov/nwis). The SWAT model which was used for computing the surface runoff depth, was calibrated and validated on a daily time step, whereas, the flows obtained by hydrologic model were calibrated and validated on a 15 minute time step.

The selected gauging station, USGS # 08154700 is a real time/daily/monthly stream flow gage, located at Bull Ck @ Loop 360, Austin, TX. Its latitude is 30°22'19", and its longitude is 97°47'04". The gage datum is 534.08 feet above sea level (http://waterdata.usgs.gov/tx/nwis/nwismap/?site_no=08154700&agency_cd=USGS). The study basin was delineated, setting up the gauging station as the outlet (Fig. 5.20).

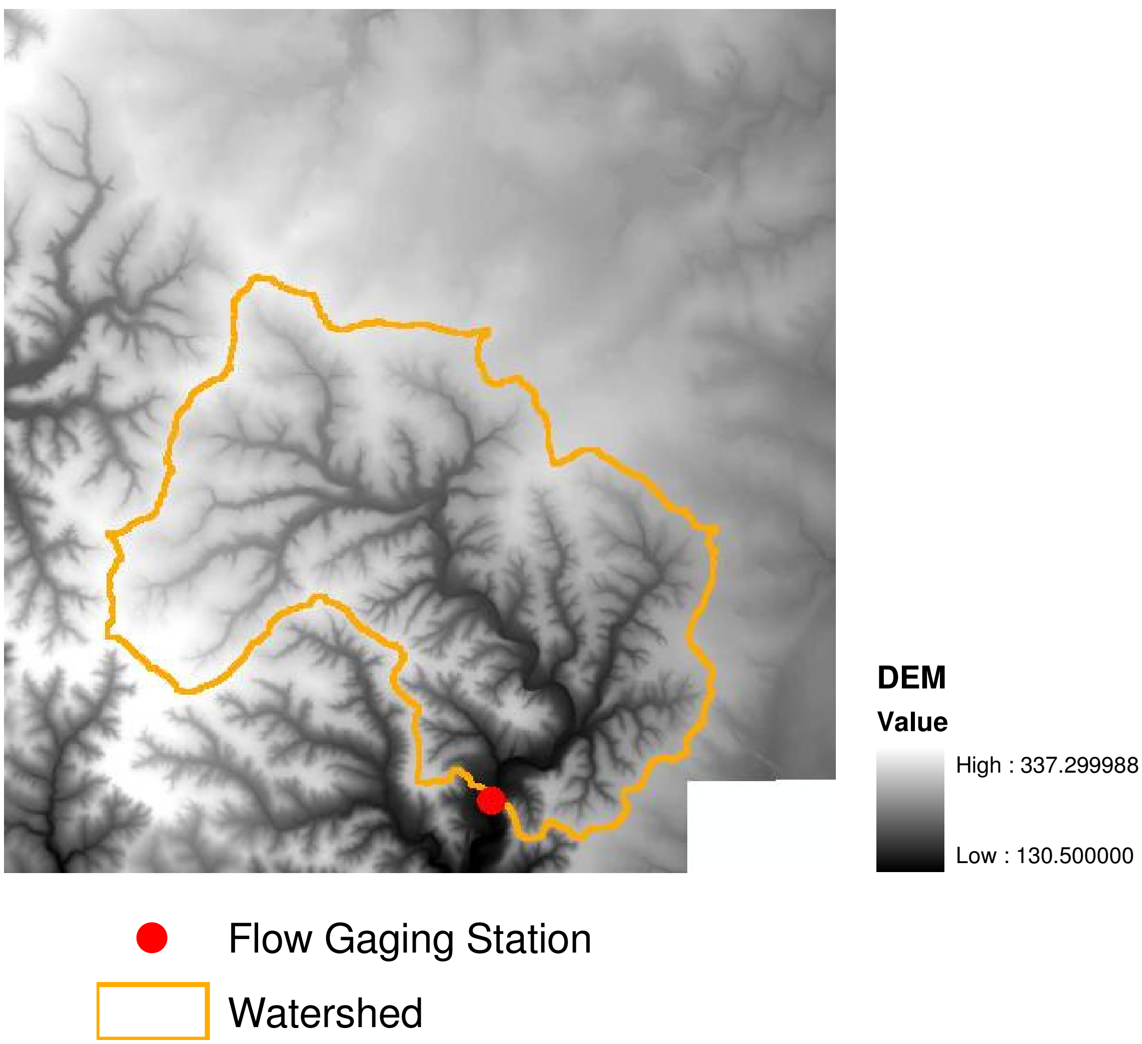


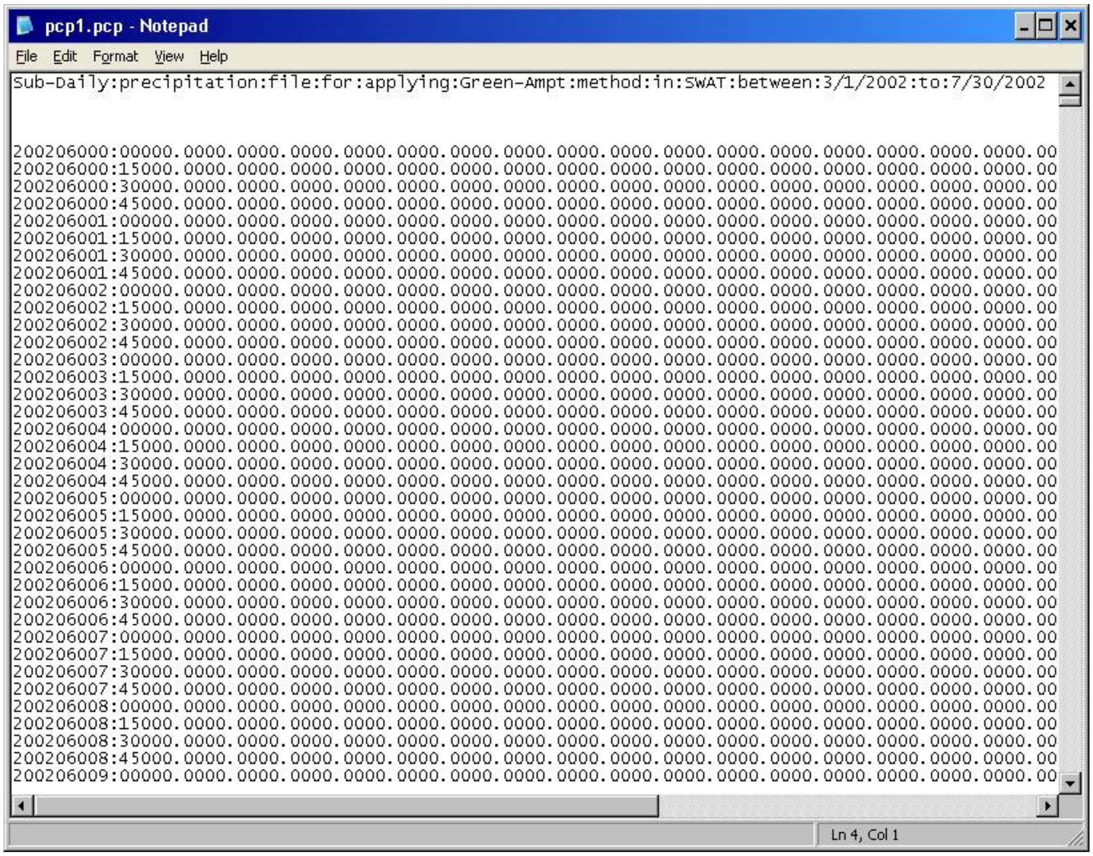
FIG 5.20. Flow Gauging Station at Bull creek@ loop 360, Austin, TX

5.2.2 AVSWAT and SWAT

SWAT was used to obtain the daily surface runoff depths which were later converted to 15 minute time step depths because the precipitation data had a time interval of 15 minute as well. SWAT is a comprehensive water quantity and water quality model and it needs a number of input files (Neitsch et al. 2000). In order to create these input files and to calibrate the SWAT model, a preprocessor called AVSWAT was used (Di Luzio et al. 2000). AVSWAT is an extension of ArcView 3.x. Using this preprocessor, creation of hydrology based input files required DEM, land use data, soils data, gauging station location data and precipitation data. The first step was the delineation of watershed and the hydrologic parameters. For delineating sub-basins a threshold value of 200 hectares was chosen, as suggested by the interface. Total number of sub-basins obtained was equal to 29, each of which contained to a stream.

Once the streams and sub-basins were delineated, the next step was the creation of hydrologic response units (HRUs). A HRU is an intersection of sub-basin polygons with land use and soils polygons. AVSWAT provides an opportunity to the user to filter some of the HRUs, which might be non-representative. For this application, filtering option was ignored and every single HRU was considered. Though AVSWAT creates HRUs, they are not geo-referenced, which means their location is not known to the user. As the hydrologic model developed needs the location of HRUs, a latest version, similar to AVSWAT called ArcGIS, developed at TAMU was used to obtain the georeferenced HRU dataset SWAT (Milver et al. 2003, Olivera et al. 2003). This version is an extension of ArcGIS. Thus, HRU information needed for both SWAT and the hydrologic model were obtained.

The next step was to create the precipitation files required for SWAT. Even though AVSWAT provides the option to create precipitation files by specifying rain gage stations, it works on daily time step and longer. As the option to create them at sub-daily time step was not available, a separate application was created to obtain the precipitation file in a format which is expected by the SWAT model. NEXRAIN rainfall data was available on a 2km resolution mesh, containing radar pixels (NEXRAIN, 2003). The scale at which SWAT model expects rainfall values is sub-basin, i.e. each column in the precipitation file would correspond to a set of rainfall values of a sub-basin (Fig 5.21). So, the precipitation file contained 29 columns, each of which had as many rainfall values as the number of time steps of simulation. Two different precipitation files were created, one for calibration and the other for validation of the model. Calibration of SWAT model was conducted using the rainfall data pertaining to tine period from 3/1/2002 to 7/30/2002 and validation was conducted using the rainfall data of time period between 10/01/01 to 02/27/02. When choosing the dates for rainfall data, factors like availability of rainfall and observed flow data were taken into consideration.




```
Sub-Daily:precipitation:file:for:applying:Green-Ampt:method:in:SWAT:between:3/1/2002:to:7/30/2002


200206000:00000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206000:15000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206000:30000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206000:45000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206001:00000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206001:15000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206001:30000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206001:45000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206002:00000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206002:15000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206002:30000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206002:45000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206003:00000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206003:15000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206003:30000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206003:45000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206004:00000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206004:15000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206004:30000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206004:45000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206005:00000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206005:15000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206005:30000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206005:45000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206006:00000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206006:15000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206006:30000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206006:45000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206007:00000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206007:15000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206007:30000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206007:45000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206008:00000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206008:15000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206008:30000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206008:45000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
200206009:00000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.0000.00
```

FIG. 5.21. 15 minute Precipitation File

As the input files needed for running SWAT were prepared, the next step was to create control specification information. The model provided several options for choosing the time step and method to be used for computing surface runoff depth and a time step for routing.

- Green-Ampt and Sub-hourly were chosen as the options for infiltration method and time step respectively.
- Hourly routing was chosen as the technique for routing the surface runoff.

- Priestley-Taylor method was chosen for computing potential evapotranspiration.
- Variable travel time model was chosen for channel routing.
- Daily output print frequency was chosen as this is the minimum time step at which SWAT can report the results.

5.2.3 SWAT Model Calibration

The SWAT model was first run using the default parameters set by AVSWAT. Calibration was performed by comparing the simulated stream flows at the main basin outlet with those that were observed. As mentioned previously the observed flow data was obtained from USGS (USGS, 2003e). Among various output files produced by SWAT, a file called basin.rch contained the required simulated flow data, which was needed for calibration. The set of flow values chosen for calibration were those that corresponded to the outlet of the main basin. A number of simulations were run in an attempt to match the simulated flows with the observed by adjusting certain parameters taking into account their acceptable range of values. The parameters adjusted consisted of,

- Available water capacity of a soils (SOL_AWC) and which basically was used for balancing the volume of water under the hydrograph. This parameter was adjusted only for the first layer of soil. Before calibration the values ranged from 0.12 to 0.16. After few simulations the value was changed to a single value which is equal to 0.12.

- Manning's n coefficient for tributary channels (CH_N1) was changed from 0.14 to 0.25 at the end of calibration.
- Manning's n coefficient for main channels (CH_N2) was changed from 0.14 to 0.20 at the end of calibration.
- Surface runoff lag time coefficient was adjusted to 0.3 days.

Fig 5.21a shows the graph of observed flows versus simulated flows after final calibration. It can be seen that the model predicts the peak flows, rising and recession curves of the hydrograph well, but fails to simulate the small events. The flow values compared are for time period between 3/01/02 to 7/30/02. The observed flows were compared to the simulated flows visually as the statistical parameters like correlation coefficients cannot be applied as there were long dry periods. This graph was blown out to show the comparison between observed and simulated flows of the event that occurred between 6/25/02 – 7/29/02 (Fig 5.22a and Fig 5.22b). The model was validated for time period between 10/01/01 to 02/28/02 (Fig 5.22c).

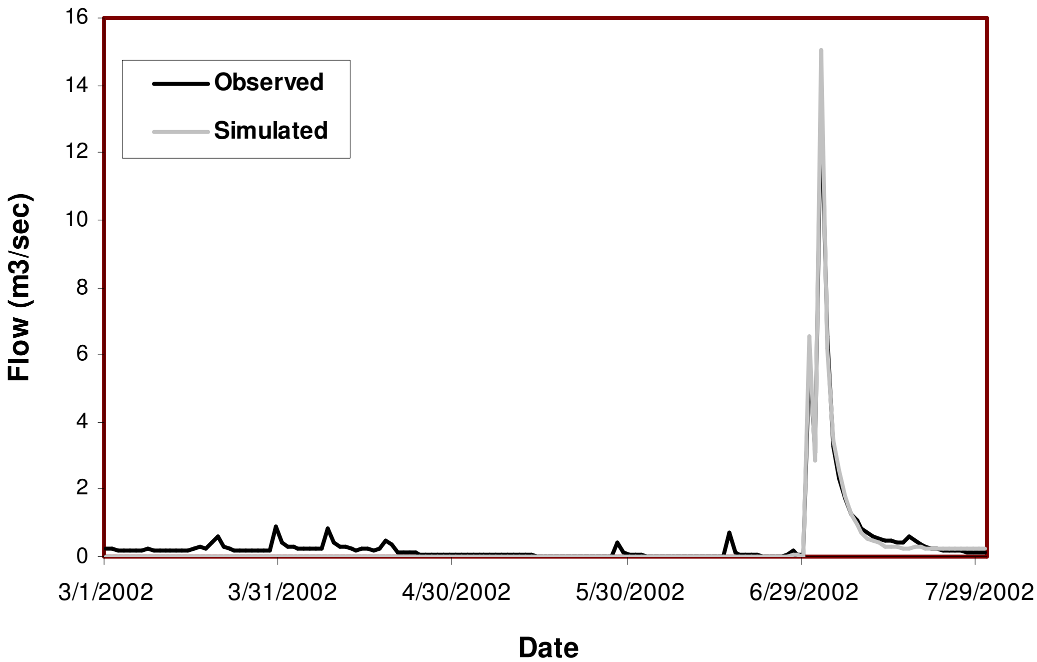


FIG. 5.22a. SWAT Model Calibration (3/1/02 – 7/30/02)

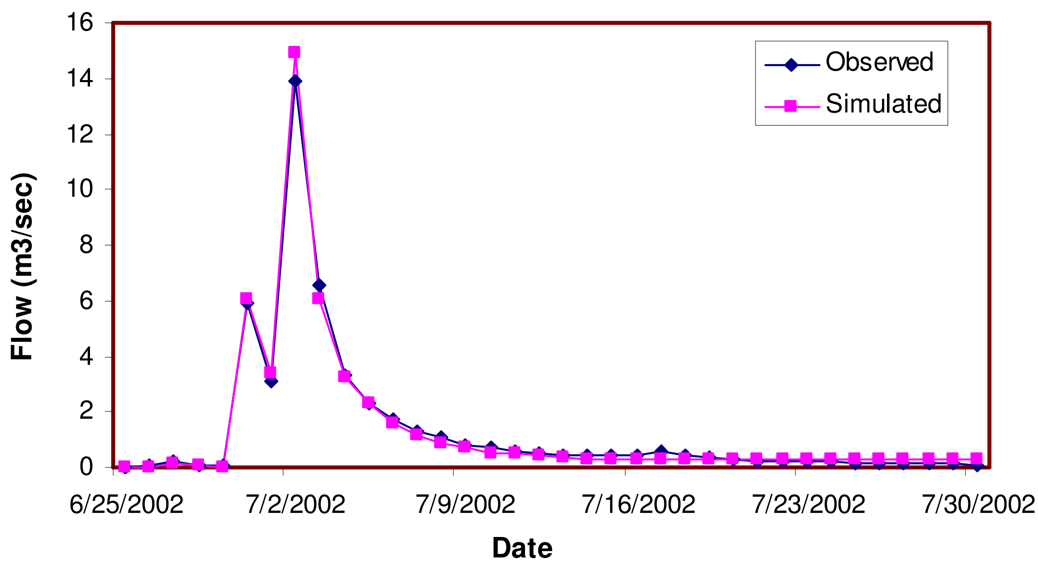


FIG 5.22b. SWAT Model Calibration (6/25/02 – 7/30/02)

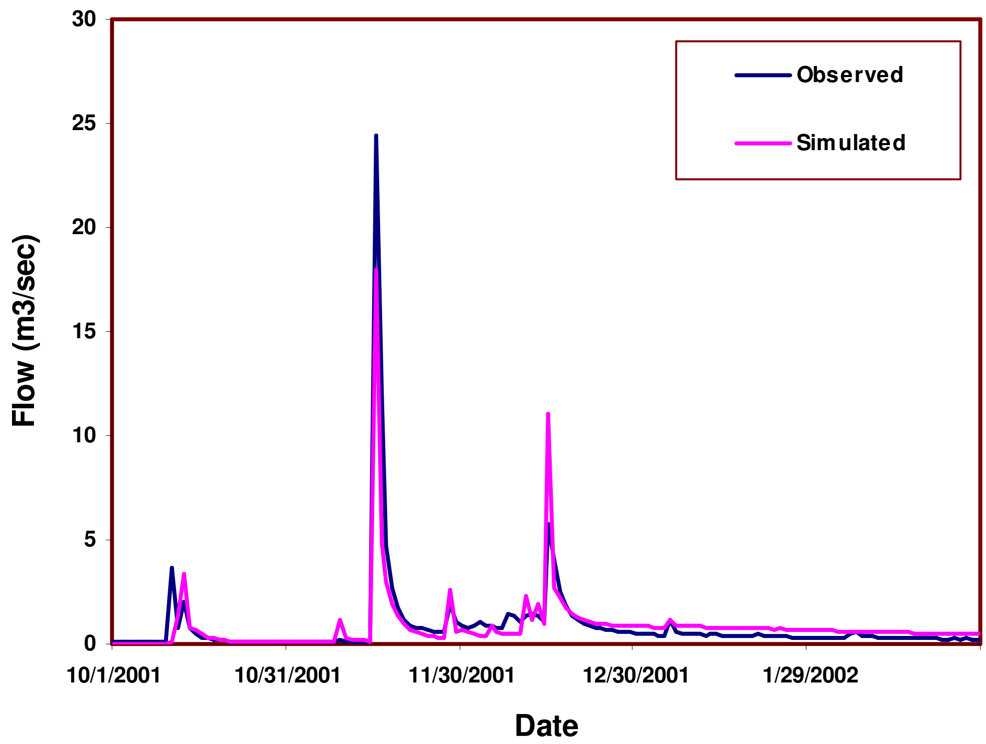


FIG. 5.22c. SWAT Model Validation

Thus, the SWAT model was calibrated and validated. The model not only computes the flows, but also computes surface runoff depths (time series) for each HRU. Though the precipitation input was at a 15 minute time step, it computes surface runoff at a daily time step. The next step was to convert daily surface runoff depths into sub-hourly (15 minute) depths. 15 minute surface runoff depths were needed for calibrating and validating the hydrologic model, developed as part the thesis. The forthcoming sections of this section discuss how this task was accomplished.

5.2.4 15-minute Surface Runoff Depths Generation

The conversion of daily surface runoff depths into 15-minute values was accomplished by creating an Excel macro, which is based on 15 minute precipitation values, total daily surface runoff depth and the incorporation of a constant infiltration depth per day. This is a simple optimization program that adjusts the infiltration value, which when subtracted from the precipitation values produces 15 minute surface runoff depths are obtained. A condition is imposed on the model which makes sure that the sum of all runoff depths per day would be equal to the total daily runoff value reported from the SWAT model. Logically, the model can be described as follows,

For any day, suppose Pi is the precipitation depth during the time step i, Ic is the constant infiltration depth for the day, Ri is the runoff depth during time step i, and $Rswat$ is the daily surface runoff depth computed using SWAT model, then,

Solve for Ic such that,

If Pi > Ic then Ri = Pi - Ic

Else if Pi < Ic then Ri = 0

Subjected to,

$$\sum_i Ri = Rswat\,, Ri <= Pi\,,\ Ri >= 0\,, Ic >= 0$$

The algorithm starts by picking a sub-basin randomly, and then iterates through all the HRUs belonging to it. During an iteration, a HRU belonging to the sub-basin under question is randomly picked, and for each day specified by the user, the optimization program is applied and the runoff values are written to a file in required format. Thus, at the end of the execution of the algorithm on a sub-basin, one runoff file is created, in which each column contained the set of runoff depths for all the time steps for a HRU. In other words, a single file corresponds to a sub-basin and contains as many columns as HRUs. At the end of the whole process, as many runoff files as the number of sub-basins are created. As SWAT model reported daily surface runoff depths in millimeters, the 15 minute surface values obtained were also computed using the above described model in millimeters. At this point, all the input required by the hydrologic model was made ready. Before going into the details of application of the hydrologic model, a brief discussion on the input data and units adopted for this model will be provided. The model computes the geometric properties and the related ones, such as length, area, volume based on the linear units defined by the projection of the spatial data. In the application all the data were defined with the same projection, which is Universal Transverse Mercator (UTM) zone 14. This projection uses meters as linear units, so the lengths were in meters and areas in meter^2. SWAT produced surface runoff depths in millimeters.

5.2.5 Application of Hydrologic Model

The data needed for applying the model includes,

- streams line shapefile,
- sub-basins polygon shapefile,

- HRU polygon shapefile,
- reservoir point shapefile
- surface runoff depth files (one per sub-basin)

The streams dataset, sub-basins dataset, and runoff files were obtained from the previous steps as explained. The HRU dataset has been obtained using the procedure developed as part of the hydrologic model and a reservoir shapefile was provided. Before the hydrologic modeling tools could be applied, as it was required to have attributed the streams and sub-basins with their hydrologic topology parameters, corresponding tools were applied, similar to the previous case study. At the end of application of hydrologic topology tools, stream lines were attributed with parameters such as, Rank (see the field Rank in table 5.5), Outlet (see the field Outlet in table 5.5) and Downstream line number (see the field DSLine in table 5.5). Similarly, sub-basins were attributed with Outlet stream number (see the field Outlet in table 5.6). Apart from these parameters, the model expects that each stream line be attributed with average stream flow velocity (see the field Vel in table 5.5), stream dispersion coefficient (see the field Disp in table 5.5), Muskingum X parameter (see the field MuskX in table 5.5) and the reservoir presence attribute {1 or 0} (see the field Reser in table 5.5). Each sub-basin polygon was attributed with overland flow velocity parameter (see the field Vel in table 5.6) and overland dispersion coefficient parameter (see the field Disp in table 5.6). These parameters were manually attributed. To start with, both overland and stream flow velocities were assumed to be 0.5 m/sec, and both overland and stream flow dispersion coefficients were assumed to be 50 m^2/sec. Though same set of parameters values have been used for all streams in the application, to exploit the model capabilities, each stream could have been attributed with unique set of values, similarly for sub-basins. In other words, for each sub-basin there can be defined only one set of values for overland flow parameters (overland flow velocity and overland dispersion coefficient) and as many sets of stream flow parameter values (stream flow velocity and stream dispersion coefficient) as the number of streams located inside the sub-basin.

TABLE 5.5. Stream routing parameters attributed to streams

Attributes of Streams

FID	Shape	ARCID	GRID_	FROM_	TO_NO	Vel	Disp	Rank	DSLine	Outlet	Reser	MuskX
0	Polyline	1	1	2	5	1.75	200	1	5	1	0	0.2
1	Polyline	2	2	3	5	1.75	200	1	5	1	0	0.2
2	Polyline	3	3	4	6	1.75	200	1	8	1	0	0.2
3	Polyline	4	4	1	6	1.75	200	1	8	1	0	0.2
4	Polyline	5	5	5	8	1.75	200	2	11	1	0	0.2
5	Polyline	6	7	7	9	1.75	200	1	7	1	0	0.2
6	Polyline	7	6	9	8	1.75	200	5	11	1	0	0.2
7	Polyline	8	9	6	10	1.75	200	2	10	1	0	0.2
8	Polyline	9	12	11	10	1.75	200	3	10	1	0	0.2
9	Polyline	10	8	10	9	1.75	200	4	7	1	0	0.2
10	Polyline	11	14	8	13	1.75	200	6	17	1	0	0.2
11	Polyline	12	10	14	11	1.75	200	1	9	1	0	0.2
12	Polyline	13	13	16	13	1.75	200	1	17	1	0	0.2
13	Polyline	14	11	17	11	1.75	200	2	9	1	0	0.2
14	Polyline	15	15	18	17	1.75	200	1	14	1	0	0.2
15	Polyline	16	17	19	20	1.75	200	1	20	1	0	0.2
16	Polyline	17	18	13	20	1.75	200	7	20	1	0	0.2
17	Polyline	18	16	21	17	1.75	200	1	14	1	0	0.2
18	Polyline	19	19	15	22	1.75	200	1	24	1	0	0.2
19	Polyline	20	20	20	22	1.75	200	8	24	1	0	0.2

Record: 1 Show: All Selected Records (0 out of 29 Selected.) Options

TABLE 5.6. Overland flow routing parameters attributed to sub basins

Attributes of SubBasin

FID	Shape*	SUBBASIN	Vel	Disp	Outlet
0	Polygon	1	0.2	50	1
1	Polygon	4	0.2	50	4
2	Polygon	2	0.2	50	2
3	Polygon	3	0.2	50	3
4	Polygon	5	0.2	50	5
5	Polygon	7	0.2	50	6
6	Polygon	9	0.2	50	8
7	Polygon	6	0.2	50	7
8	Polygon	14	0.2	50	11
9	Polygon	8	0.2	50	10
10	Polygon	12	0.2	50	9

Record: 1 Show: All Selected Records (0 out of 29 Selected.)

Up to this point, streams and sub-basins data were obtained. The next step was to compile HRU data. The model was first run using the HRUs obtained from ArcGIS SWAT. There were some problems faced using these HRUs, due to which an alternative was chosen. Following are the reasons due to which an alternative had to be found.

- Some of the HRU polygons were very small, whose significance on the analysis was negligible.
- Some of the HRU polygons were complex, meaning that a single polygon had a number of sub-parts, which were distributed on both sides of the streams, making centroids of the HRUs lie closer to stream. Due to this reason, the overland flow processes were not been captured effectively.

In Figure 5.23, observe the sub-basin numbered 15. It can be seen that both of them are dominated by a single land use type, which is FRSD (see the legend). In sub-basin 15, FRSD land use type covers 70 percent of the sub-basin, accounting to a major part of the surface runoff (Fig. 23). As it can be observed that the stream almost breaks this HRU polygon into two equal parts, the centroid lies closer to it. Existence of such type of HRUs was observed in most of the sub-basins. In order break such polygons into more number of parts and to accumulate the smaller polygons the HRU were broken with a mesh, which is the alternative method.

The alternative method for delineating HRU polygons was to intersect the sub-basin with a mesh. The mesh used had a resolution of 1km. On application of this method, the total number of HRUs (also called sources) dropped from 410 (derived by ArcGIS SWAT) to 213 (Fig. 5.24). As they are not HRUs any more, let them be called 'Sources'.

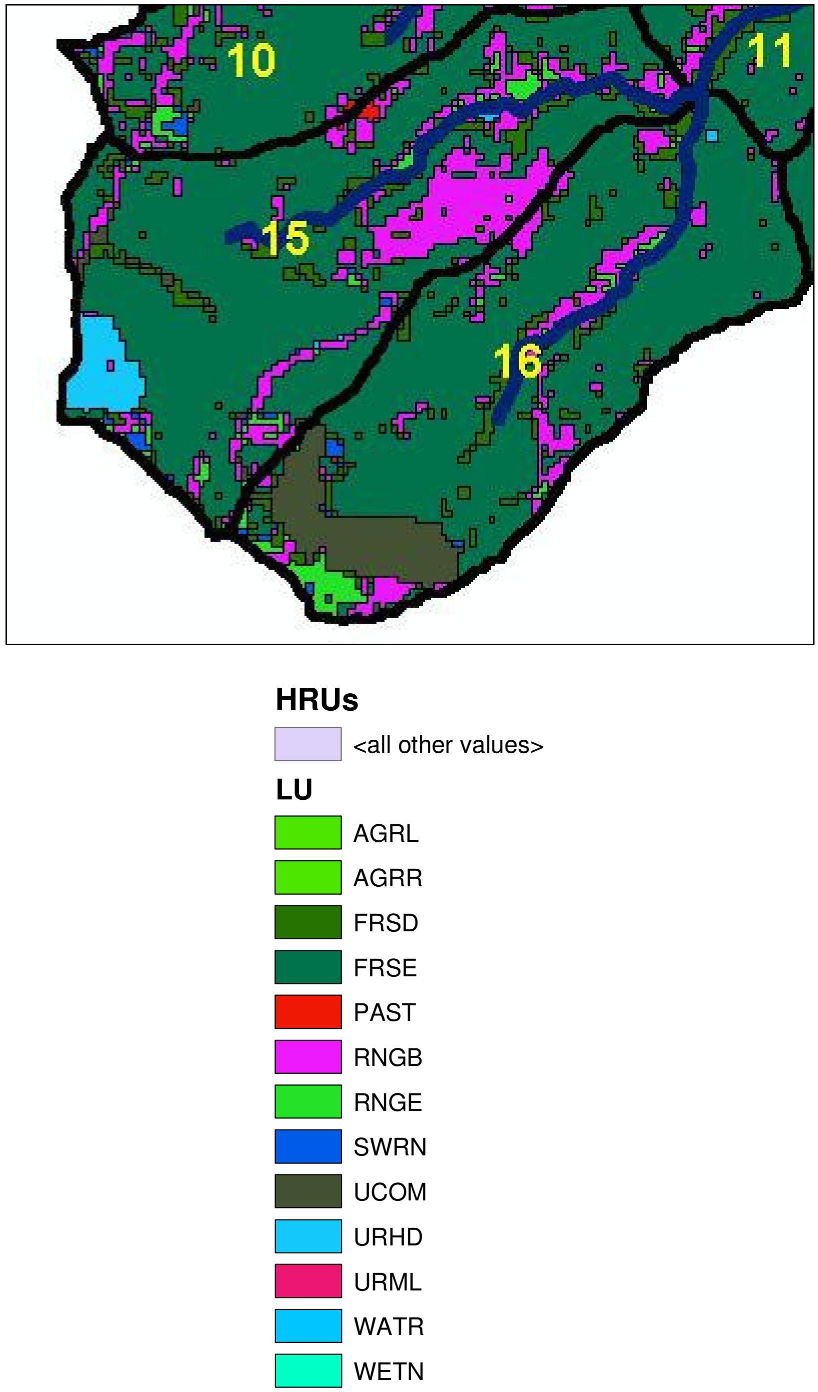


FIG. 5.23 HRUs Delineated by ArcGIS SWAT

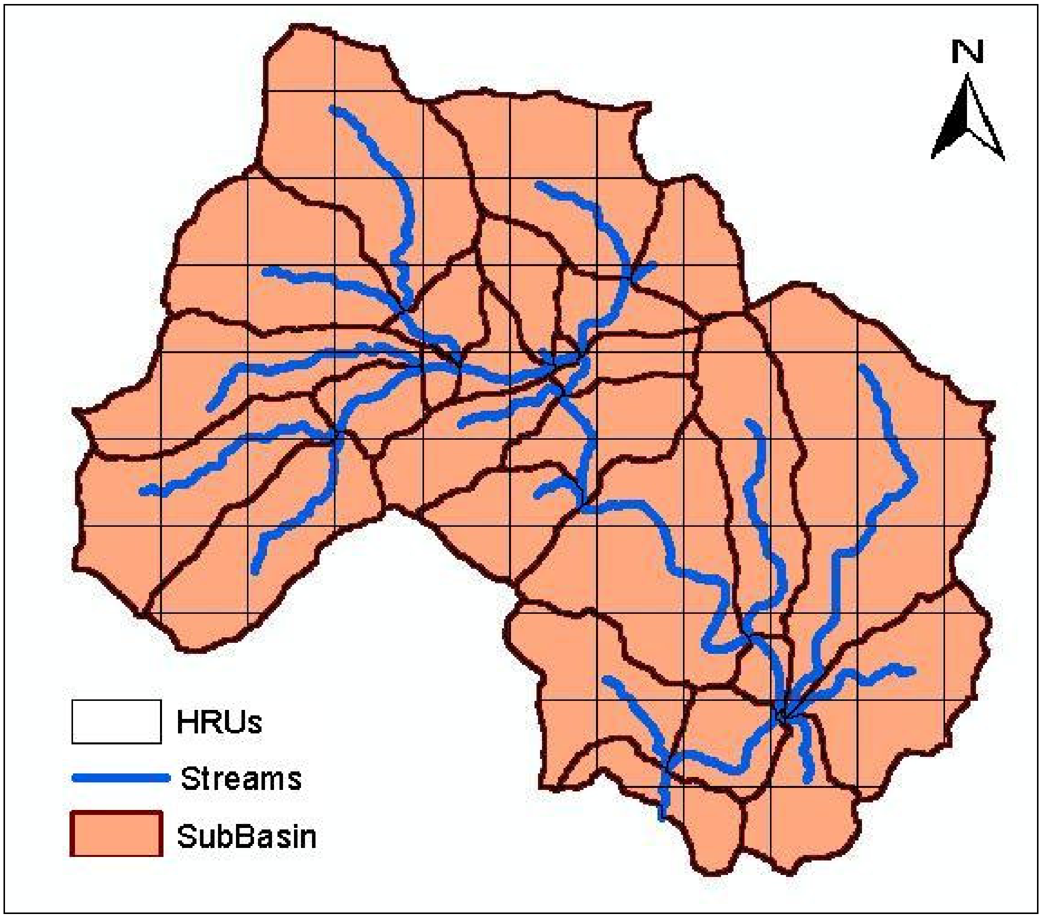


FIG. 5.24. Sources in Bull Creek Watershed

To make computations easier, some intermediate routing parameters, which will be explained later, were attributed to sources and streams. The tools that attribute these parameters are provided with model as explained in the section 4. Upon application of theses procedures, sources were attributed with,

- Closest stream number (see the field Stream in table 5.7)
- Overland flow path length in meters (see the field LO in table 5.7)
- Stream flow path length in meters (see the field LS in table 5.7)

- Average flow time to the sub-basin outlet in seconds (see the field Pa in table 5.7)
- Representative Peclet number (see the field PI in table 5.7).

The streams were attributed with,

- Channel flow time (see the field ChanTim in table 5.8)
- Number of sub-reaches (see the field SReach in table 5.8)
- Channel route type (see the field RType in table 5.8). In this field "Muskingum1" refers to Muskingum flow routing method without dividing the channels into sub-reaches, "Muskingum2" refers to Muskingum flow routing method by dividing channels into a number of sub-reaches and "Lag" refers to Pure Lag routing method.
- Muskingum routing parameters (see the fields Cx, Cy, Cz in table 5.8).

TABLE 5.7. Intermediate overland routing parameters computed for HRUs

Attributes of HRUs

FID	Shape*	Stream	LO	LS	Pa	Pl
0	Polygon	28	789.348147	1021.690495	4530.563874	1.039246
1	Polygon	28	622.414960	968.406214	3665.449779	0.862543
2	Polygon	26	1246.032982	343.137079	6426.243243	1.325453
3	Polygon	29	1167.499999	370.244960	6049.068541	1.253394
4	Polygon	29	979.999998	373.994955	5113.711395	1.067075
5	Polygon	29	13.331154	210.751041	187.084935	0.103916
6	Polygon	29	591.229537	0	2956.147685	0.591230
7	Polygon	27	248.023884	989.116895	1805.329073	0.524225
8	Polygon	28	769.391263	1021.690495	4430.779457	1.019731
9	Polygon	28	507.910366	1021.690495	3123.374970	0.767246
10	Polygon	22	1115.567712	98.284271	5634.001000	1.138078
11	Polygon	22	1222.801295	651.837653	6486.485133	1.375839
12	Polygon	28	491.757046	934.264092	2992.650427	0.727558
13	Polygon	29	323.147606	486.421356	1893.693088	0.443445
14	Polygon	28	456.245123	1021.690495	2865.048757	0.718576
15	Polygon	29	939.014639	373.994955	4908.784600	1.026170
16	Polygon	29	396.519585	590.563494	2320.062780	0.542451
17	Polygon	27	2.399303	203.605814	128.342693	0.259801
18	Polygon	28	53.126103	664.558451	645.378201	0.310990
19	Polygon	23	33.489562	14.142130	175.529028	0.036790
20	Polygon	25	11.533950	73.674516	99.769472	0.034373

Record: 1 Show: All Selected Records (0 out of 213 Selected.)

TABLE 5.8. Intermediate channel routing parameters for streams

Attributes of Streams

ChanTim	RType	SReach	Cx	Cy	Cz
2962.497858	Muskingum2	2	0.094037	0.456422	0.449541
995.895676	Muskingum1	0	0.201185	0.520711	0.278104
1057.772425	Muskingum1	0	0.183955	0.510373	0.305672
1765.705842	Muskingum1	0	0.052003	0.431202	0.516795
1774.732640	Muskingum1	0	0.050837	0.430502	0.518662
1647.463603	Muskingum1	0	0.068161	0.440897	0.490942
1267.071943	Muskingum1	0	0.134311	0.480587	0.385102
1775.510000	Muskingum1	0	0.050737	0.430442	0.518822
2145.096665	Muskingum1	0	0.009686	0.405812	0.584502
1099.396831	Muskingum1	0	0.173086	0.503851	0.323063
1036.623789	Muskingum1	0	0.189694	0.513816	0.296490
213.340140	Lag	0	0	0	1
724.803535	Lag	0	0	0	1

Record: 0 Show: All Selected Records (0 out of 29 Selected.)

At this point, the flow elements were assigned with all the parameters that the routing procedures require. The routing was conducted using the hydrologic model in two parts. First, the overland flow routing procedure was applied, which transports water at HRUs to their respective sub-basin outlets and secondly, the channel routing procedure which transports the flows at all the sub-basin outlets to the main basin outlet. The interface for overland flow routing is as shown in the figure below (Fig 5.25).

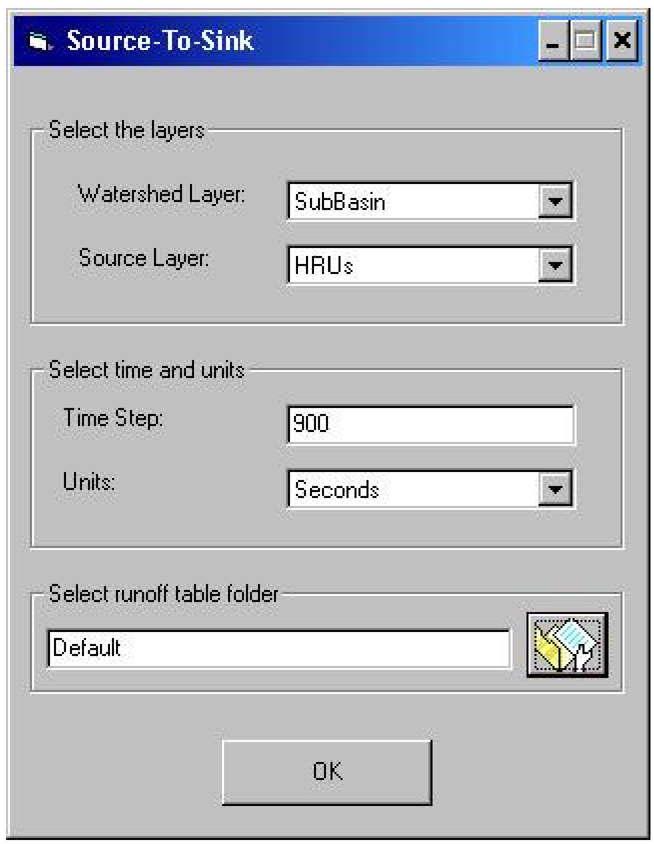


FIG. 5.25. Source-To-Sink Flow Routing Interface

The routing procedure creates a .dbf file, in which each column represents the hydrograph generated at each individual sub-basin outlet. In this application, as runoff units were millimeters, area units were square meters and the time units were seconds, the units of discharge was millimeter meter^2 per second (mm m^2/ sec).

Once hydrographs were generated at the sub-basin outlet, the next step was to route them to the main basin outlet. As there was a USGS gauging station at the basin outlet, 5 minute flows were obtained to calibrate at this point. As the model could predict flow hydrograph of 15 minute duration, observed 15 minute flow data was obtained by averaging the 5 minute flows. These flows included both surface flow and base flow but the model could predict only surface flows. Straight line method was used for separating the base flow from the total flow (Olivera 1996). Two different sets of flow data were obtained from USGS; one was used for calibration and the other for validation. Calibration was done for the time period from June 30th 2002 to July 2nd 2002 (Fig 5.26) and validation from September 6th 2002 to September 8th 2002 (Fig 5.27). The simulated flows were compared to the observed flows visually as the statistical parameters like correlation coefficients cannot be applied as there were long dry periods. After a number of runs of the model, the final parameter values obtained after calibration were:

- Stream flow velocity = 1.75 m/sec
- Stream dispersion coefficient = 200 m^2/sec
- Overland flow velocity = 0.2 m/sec
- Overland dispersion coefficient = 50 m^2/sec

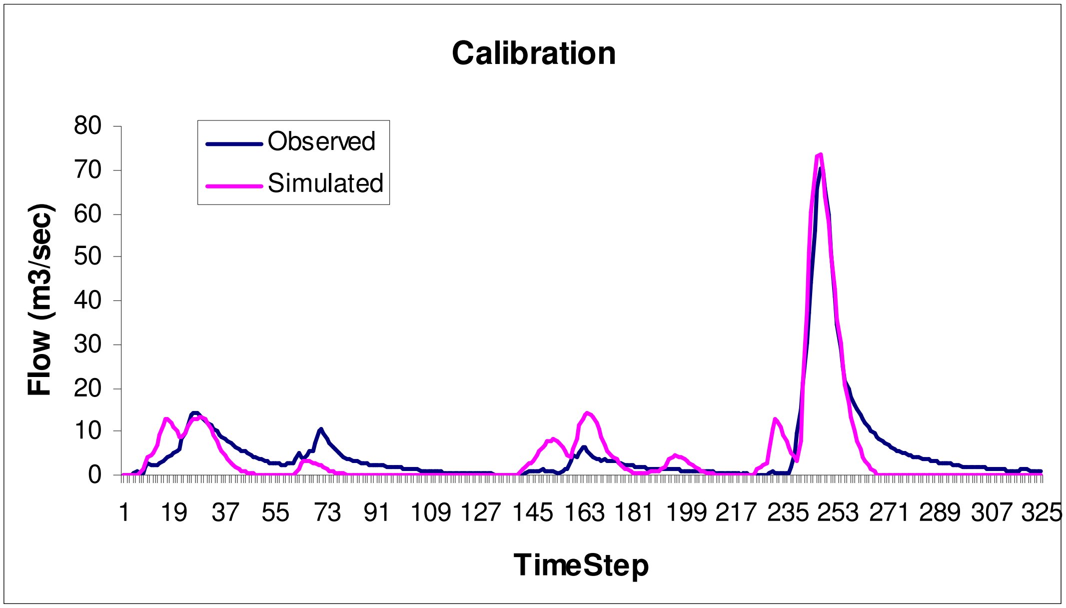


FIG. 5.26 Hydrologic Model Calibration

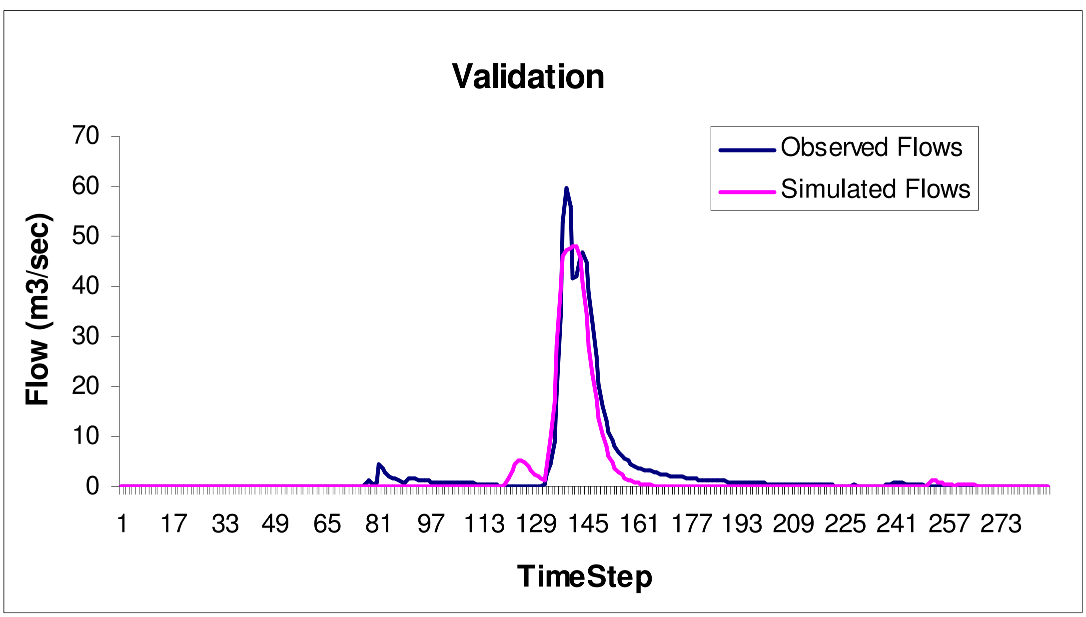


FIG. 5.27 Hydrologic Model Validation

6 CONCLUSIONS

The main objectives of this thesis are the development of a hydrologic model, hydrologic parameter computation methods and methods for treatment of raw vector data, in the vector environment.

As the methods are vector based, streams are represented as lines and the corresponding drainage areas as polygons. The procedures have some restrictions on the structure of the stream network, and on the location of stream lines with respect to their drainage area polygons. These restrictions require that all the stream lines flow towards the outlet on a unique path making the network dendritic. As the datasets might be obtained from different sources, certain problems arise, making them unsuitable for modeling. The problems recognized have been grouped into two categories. The first category of problems consists of some streams possessing illegal flow directions and stream network containing loops. The procedures to detect and solve such problems have been based on network navigation techniques starting at the outlet, moving in a bottom to top fashion and simultaneously applying the correction methods. The second category of problems consists of those which have to do with mismatches in location of some streams with respect to their drainage areas. Four different varieties of problems have been recognized. The first one consists of streams slightly crossing over the drainage divides, the second consists of streams outletting into a sub-basin and flowing back into its sub-basin, the third one consists of two streams posing to be potential outlets and the fourth consists of streams outletting into wrong sub-basins. The procedures to solve these problems have been developed and were presented.

Thus, applications of the correction methods make sure that the data used are made compatible with the model requirements. Once the compatibility issue has been addressed the next module consists of methods that compute hydrologic topology and then derive the network parameters. Among the parameters computed for streams, each of one of them is attributed with Rank, Downstream length, Weighted downstream length, Upstream length, Weighted upstream length, Outlet, Downstream line number, Accumulated area and weighted accumulated area. Here Outlet attribute has a value of 1 or 0 depending whether a stream is outlet or not. Based on the outlet attribute of the streams, the accumulated area and its weighted version values are transferred to the corresponding sub-basins. These constitute the methods dealing with computation of network parameters. Also developed here are the procedures for performing traces on the stream lines, i.e. downstream trace and upstream trace

The next module consists of procedures for hydrologic modeling. A physically based spatially distributed model has been developed. The overland flow routing was conducted using 'Source-to-sink' method; channel routing can be conducted either by Muskingum routing or pure lag method and reservoir routing using storage-outflow method. To represent spatially distributed properties and runoff, sources were derived. The sources are a combination of soils, land use and sub-basins. The algorithm works in two steps, the first step routes the surface runoff depths generated at the sources to their respective sub-basin outlets as flows, using overland flow routing technique and the second step routes the flows generated at the sub-basin outlets to the basin outlet using channel routing and reservoir routing techniques.

Two case studies have been presented to show the application of the methods developed. The first case study consisted of application of data correction methods and network parameter computation methods to San-Antonio basin. The streams and sub-basins data required were extracted from NHD and HUCs respectively for region 12.

The second case study consisted of application of hydrologic topology tools, network parameter tools and hydrologic model tools on Bull creek watershed located in Austin, TX. To derive the surface runoff depths, SWAT was used. 15 min precipitation was used for this application. Though the rainfall time step was 15 minutes, SWAT produced surface runoff depths on a daily basis per each HRU. To convert the daily surface runoff depths into 15 minute depths, a macro was built. In this application HRUs produced by SWAT could have been used as sources, but instead the sub-basins were intersected with a 1km mesh to produce the sources. Finally, the hydrologic model was applied for two different events; one for calibration and the other for validation. Both the calibration and validation produced decent match between the observed and the simulated.

The model proposed could be extended to SWAT for overland flow routing. The SWAT model assumes a linear reservoir technique for routing surface runoff to their respective outlets and reports the values on a daily time step. Incorporating the hydrologic model proposed in the thesis, SWAT overland flow model could be made more physically based and also report the values on a user defined sub-daily time step. Also, the model could be used for modeling Best management practices as the overland flow model proposed can incorporate contaminant losses along the flow travel paths.

As every other model, this model has some limitations, which could be exploited for further research. The limitations could be briefed as,

- The model does not provide an infiltration model
- The model does not provide an Evapo-transpiration model
- The interface is limited to Windows operating system only
- The model needs input in a specific format, any other format will not be accepted
- As mentioned, 1 km mesh was used for obtaining the sources needed by the model. Other mesh resolutions were not tested for.

REFERENCES

Biftu, G.F., and Gan, T.Y. (2001). “Semi-distributed, physically based, hydrologic modeling of the Paddle River Basin, Alberta, using remotely sensed data.” *Journal of Hydrology,* 244, 137-156.

Charles, W.D., Fred, L.O., William, D.M., and Russell, S.H. (2000). “Theory, development, and applicability of the surface water hydrologic model CASC2D”, *Hydrological Processes*, 16, 255-275.

Costa-Cabral, M., and Burges, S. (1994). “Digital elevation model networks (DEMON)- A model of flow over hillslopes for computation and dispersal areas”, *Water Resources Research*, 30(6), 1681-1692.

Daniel, W.K., and Arlen, D.F (1998). “Evolution of Clark’s unit graph method to spatially distributed runoff.” *Journal of Hydrologic Engineering, 3*(1), 9-19.

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

Djokic, D., Ye, Z., and Miller, A. (1997). “Efficient watershed delineation using ArcView and spatial analyst.” *Proc., 17th International ESRI User Conference*, Environmental Systems Research Institute, San Diego, CA.

Environmental Systems Research Institute (ESRI), (2003) *ArcObjects Online*, <http://www.esri.com>

Furnans, F., and Olivera, F. (2001), “Watershed topology: The Pfafstetter system.” *Proc., of the 21st ESRI User Conference*, ESRI, San Diego, CA.

Gainnoni, F., Smith, J.A., Zhang, Y., and Roth, G. (2003). “Hydrologic modeling of extreme floods using radar rainfall estimates.” *Advances in Water Resources*, 26, 195-203.

Garbrecht, J., Ogeden, F. L., DeBarry, P.A., and Maidment, D. R (2001). “GIS and distributed watershed models. I: Data coverages and sources.” *Journal of Hydrologic Engineering,* 6(6), 506-514.

HEC. (2000). *Geospatial Hydrologic Modeling Extension HEC-GeoHMS v. 1.0-User’s Manual,* U.S. Army Corps of Engineers-Hydrologic Engineering Center (HEC), CPD-77, Davis, California.

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

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

Jenson, S.K., and Domingue, J.O. (1988). "Extracting topographic structure from digital elevation data for geographic information system analysis.” *Photogrammetric Engineering and Remote Sensing*, 54(11), 1593-1600.

Jenson, S.K. (1991). “Applications of hydrologic information automatically extracted from digital elevation models.” *Hydrological Processes*, 5(1), 31-44.

Liu, Z.Y., and Todini. E. (2002). “Towards a comprehensive physically-based rainfall-runoff model.” *Hydrology and Earth System Sciences*, 6 (5), 859-881.

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

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

Microsoft (2000), *Microsoft Developer Online Documentation,* <www.microsoft.com>

Neitsch, S. L., Arnold, J. G., Kiniry, J. R., Srinivasan, and Williams, J. R. (2000). *Soil and water assessment tool users manual ver: 2000*, Grassland, Soil and Water Research Laboratory, Agricultural Research Service and Blackland Research Center, Texas Agricultural Experiment Station, Temple, TX.

NEXRAIN Corporation (NEXRAIN), (2003) *NEXRAIN products and services* <http://www.nexrain.com/>

Olivera, F. (1996), “Spatially distributed modeling of storm runoff and non-point source pollution using geographic information systems”, Dissertation, Dept. Of Civil Engineering, University of Texas at Austin.

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, Calif.

Olivera, F., and D. R. Maidment (1999), “Geographic information systems (GIS)-based spatially distributed model for runoff routing”, *Water Resources Research*, 35(4), 1155-1164.

Olivera, F., J. Famiglietti and K, Asante (2000), "Continental scale flow routing using a source-to-sink algorithm", *Water Resources Research*, 36(8), 2197-2207.

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

Olivera, F., Valenzuela, M., and R. Srinivasan (2003), "Linkage of the Arc Hydro Data Model with SWAT", *2nd International Soil and Assessment Tool (SWAT) Conference*, Istitudo di Ricerca sulle Acque (Water Research Institute) and Nazionale delle Ricerche (National ResearchCouncil [of Italy]), Bari, Italy.

Quinn, P., K. Bevan, P. Chevalier and O. Planchon, (1991), "The prediction of hillslope flow paths for distributed modeling using digital terrain models." *Hydrologic Processes*, 5, 56-79

''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, D.C.

Tarboton, D. (1997)., " A new method for the determination of flow directions and upslope areas in grid digital elevation models"., *Water Resources Research,* 33(2), 309-319.

Ungerer, M.J., and Goodchild, M.F. (2002) "Integrating spatial data analysis and GIS: A new implementation using the Component Object Model (COM)." *International Journal of Geographical Information Science,* 16(1), 41-53.

United State Department of Agriculture (USDA), (2003) *National Cartography and geospatial center,* < http://www.ncgc.nrcs.usda.gov/>

United States Geological Service (USGS), (2003a) *National Elevation Dataset*

< http://gisdata.usgs.net/ned/default.asp>

United States Geological Service (USGS), (2003b) *National Hydrographic Datasets*

< http://nhd.usgs.gov/>

United States Geological Service (USGS), (2003c) "*National Land Cover Characterization,* < http://landcover.usgs.gov/natllandcover.html>

United States Geological Service (USGS), (2003d) *Retrieval for Spatial Data Set huc250k,* < http://water.usgs.gov/lookup/getspatial?huc250k>

United States Geological Service (USGS), (2003e) *Surface-water Data for the Nation*

< http://waterdata.usgs.gov/nwis>

Valenzuela, M. (2003), "Development of an ArcGIS interface and design of a geodatabase for the soil and water assessment tool", Masters thesis, Dept. of Civil Engineering, Texas A&M University, College Station, Tex.

Verdin, L. and Verdin, J, (1999). "A topological system for delineation and codification of the Earth's river basins.", *Journal of Hydrology*, 218(1-2), 1-12.

Wang, G., Chen, S., Boll, J., Stockle, C.O., and McCool, D.K. (2002). "Modelling overland flow based on Saint-Venant equations for a discretized hillslope system." *Hydrologic Processes,* 16, 2409-2421.

Weiss, M.L. (1994). *Data Structures and Algorithm Analysis in C++*, Addison-Wesley Publishing Company, Reading, MA.

Wurbs, R.A., and James, W.P. (2002). *Water Resources Engineering,* Prentice Hall, Englewood Cliffs, NJ.

Xu ZX., Ito K., Schultz G.A., and Li J.Y.(2001). "Integrated hydrologic modeling and GIS in water resources management." *Journal of Computing in Civil Engineering*, 15(3), 217-223.

VITA

Srikanth Koka

ADDRESS

I.C.106, Irrummanzil Colony, India-500082
(979) 691-0331
(979) 574-9028

EDUCATION

- Master of Science in civil engineering, May 2004
  Texas A&M University

- Bachelor of Engineering, May 2000
  Osmania University, Hyderabad, India.

EXPERIENCE

- Texas A&M University, College Station, Texas (Jan 2002 – May 2004)
  Graduate Research Assistant.
  Development of GIS tools for Water Resources Engineering (Master's thesis):
    - Hydrologic topology and network parameters computation tools
    - Spatially distributed hydrologic model.

This thesis was typed by the author.