The Pennsylvania State University

The Graduate School

College of Engineering

IMPLEMENTING THE SHALE HILLS WATERSHED MODEL IN APPLICATION OF PIHM

A Thesis in

Civil Engineering

by

Wenfang Li

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2010

The thesis of Wenfang Li was reviewed and approved* by the following:

Christopher J. Duffy
Professor of Civil Engineering
Thesis Advisor

Patrick M. Reed
Associate Professor of Civil Engineering

Michael N. Gooseff
Assistant Professor of Civil Engineering

Peggy Johnson
Department Head, Civil and Environmental Engineering
Professor of Civil Engineering

* Signatures are on file in the Graduate School.

ABSTRACT

This thesis addresses the problem of model implementation at the Shale Hills experimental watershed using PIHM (Penn State Integrated Hydrologic Model). Ever since the 1970s, interdisciplinary teams have been working in Shale Hills watershed to study a wide range of earth science problems. However, a watershed model was not constructed until Qu and Duffy (2007) proposed the PIHM model. In recent years, the PIHM had a major update by Kumar and Bhatt where they added new flux components to the channel flow, implemented macropore effects, throughfall drainage, evaporation from ground and transpiration from the canopy. At the same time, PIHM was extended to include national databases which are referred to as A-priori data. This research was performed with NSF funding through the Susquehanna River Basin Project, the Critical zone Observatory project and the RTH_NET project. The focus of this thesis is to implement a new version of PIHM at Shale Hills using data sets recently acquired through the Critical Zone Observatory Project. These new data sets include: 3 meter digital elevation data, a new bedrock elevation coverage, the latest soil classification data from SSURGO with site specific extensions to SSURGO made by H. Lin's group, and the National Land Cover Data distribution dataset. The updated model is calibrated through a trial and error process, using the 1974 artificial irrigation experiment by Lynch et al. The model successfully reproduces the runoff at the watershed outlet during a sequence of 6 rainfall events. It shows that the Horton overland flow and subsurface storm flow are the main drive for the runoff peak in the channel. The model also simulates groundwater

levels, recharge, transpiration, etc. The model represents a preliminary calibration which will be implemented in real-time model with current data. Operating the model in real-time will allow the continuous calibration using CZO experimental data and provide feedback to scientists. Finally, Bhatt and Kumar have implemented a GIS interface for PIHM and this was used at Shale Hills for setting up the model new data coverages. A tutorial for the PIHMgis and the PIHM model is included as part of this thesis.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Notation

a	Macropore Fraction
C_p	Air heat capacity
C_{o2r}	Overland Flow to River Coefficient
d	Infiltration depth
d_{ij}	The distance between the center of elements i and j
et	Evaporation rate
et_0	Evapotranspiration rate
et_{canopy}	Evaporation from canopy
et_{ol}	Evaporation from overland flow
F_{ch}	Exchange flow rate between the channel and the aquifer
F_{inf}	Infiltration rate
F_g	Groundwater flow rate
F_{o2r}	Overland Flow to River Rate
F_{riv}	Channel Flow Rate
F_{surf}	Surface Flow rate
F_{us}	Flow rate of saturated-unsaturated zone / recharge
g	Gravity
G	Radiation absorbed by the ground
G_{tf}	Throughfall drainage rate
h_{riv}	Channel water depth

h_{ovl}	Overland flow depth
h_s	Ground water depth
h_u	Unsaturated zone soil moisture depth
IS	Interception Storage
k	Throughfall coefficient
K_{eff}	The effective conductivity of soil matrix and macropore
K_{mat}	Conductivity of soil matrix
K_{mac}	Conductivity of macropore
K_s	Saturated conductivity
K_u	Unsaturated conductivity
LAI	Leaf Area Index
n	Manning's coefficient
P	Precipitation
R_n	Total radiation
SP	Snow package water equivalent depth
v	Vegetation fraction
z	Surface Elevation
z_b	Bedrock elevation
z_r	River bed elevation
z_{rb}	River bank elevation
z_u	Elevation of the top soil layer
α	van Genuchten coefficient
β	van Genuchten coefficient
ε_s	Saturated vapor pressure at air temperature
ε_a	Vapor pressure
Δw	Snow Melt Rate

ψ_{int}	Interception storage
$\psi_{\text{int max}}$	Canopy water storage capacity
ρ_a	Air Density

ACKNOWLEDGMENTS

I would like to thank my thesis advisor Dr. Christopher Duffy for his guidance and support throughout my master's study. I would also like to thank Dr. Patrick Reed for the valuable suggestions at the beginning of the thesis research, and Dr. Michael Gooseff for taking the time to serve on my committee. For all my friends in Penn State, thank you for sharing all the memories in this nice and quiet college town. Finally, for both of my parents, thank you for always being so supportive, so caring even if I am on the other end of the world.

Chapter 1

Introduction

1 .1 Literature Review

The thesis addressed the problem of constructing a model of Shale Hills experimental watershed using PIHM (Penn State Integrated Hydrologic Model). Mathematical models have been widely accepted as an important tool for theoretical and experimental research in hydrology science in recent years. Duffy (1996) proposed a two-state model by integrating local conservation equation with respect to the soil moisture storage of the saturated and unsaturated states. The model was tested in the Shale Hills watershed, and showed that soil moisture and saturated storage are controlling factors for the hydrologic response of first order watershed (Tchaou, 1999). Qu and Duffy (2007) developed a multi-process, multi-scale, spatially distributed, physically based hydrologic model, in which, major processes (surface overland flow, subsurface flow, channel routing, interception, snowmelt, evaporation and evapotranspiration) are fully coupled using the semi-discrete finite volume method. In the model, the TIN (triangular irregular

network) is generated by Delaunay triangulation. The model is solved by an implicit sequential solver from SUNDIALS (suite of nonlinear and differential algebraic equations solver), which is developed in LLNL (Lawrence Livermore National Laboratory). The model is also known as the first version of PIHM (Penn State Integrated Hydrologic Model). Qu and Duffy (2007) implemented the approach to simulate the Shale Hills field experiment, and "successfully simulated observed groundwater levels, as well as runoff at the outlet and internal points within the watershed". Kumar and Bhatt revised the code, and enhanced it by adding the macropore phenomena, throughfall drainage, evaporation of canopy and ground process to the model, which was the second generation of PIHM code. On the other hand, however, the large number of physical parameters that need to input to the model becomes a big barrier for the wide application for the model (Bhatt et al. 2009). Bhatt (2009) integrated the PIHM model to an open-source Geographical Information System (GIS), which was known as PIHMgis. PIHMgis applies the Qgis interface for the preprocessing of topography, geology, soil, vegetation and climate data, which greatly reduced the effort of using the model.

Shale Hills experimental watershed is one of the Critical Zone Observatories. It is a typical V-shape, forested, small catchment lying in the Valley and Ridge physiographic province of East United States. The Shale Hills experimental watershed is a first-order, 19.8 acre watershed approximately 14 miles southwest of State College in Barree Township, Huntingdon County, and is part of the Stone Valley Experimental Forest owned by Pennsylvania State University. The climate is transiting through the seasons. Lynch (1976) performed artificial rainfall experiments in the 1974, by the irrigation system installed in the watershed. Water budgets are collected to investigate the effects of antecedent soil moisture on storm flow volumes and timing. Lin (2006) revisited the site in 2003 and 2004 for a better understanding of temporal and spatial

distribution of soil moisture at the Shale Hills watershed. Based on local measurements, a 3-m DEM is refined from 10-m DEM for a better representation of the swales and ridges. Lin (2006) also identified the 5 soil classes and detailed the soil map for the Shale Hills experimental watershed. From Lin's research, Blairton and Ernest are located on the valley floor along the stream, Rushtown is found in the swales, while Berks and Weikert are found on the slopes and ridges. Species of maple, oak, hickory, which are typical deciduous trees in Shale Hills watershed, distribute on the slopes and ridges. Hemlocks are found on the valley floor near the west, Pines are located up on the ridges of the southern slope.

1.2 Motivation and Objective

Since Qu and Duffy (2007) implemented the Shale Hills watershed in the first version of PIHM, the PIHM model had a major update by Kumar and Bhatt. The 1D channel flow is updated by multiple channel flow components; and macropore effects, evaporation from ground and transpiration are newly integrated. It is hoped that more information can be obtained from the model, to help us understand the hydrologic mechanism in Shale Hills area and also helped us in experimental planning.

Although Bhatt had integrated the GIS interface to the PIHM model as its pre-processing part, the use of the PIHMgis is still quite subtle. As part of the thesis, a tutorial for the PIHMgis and PIHM model is included to improve the wide application of the PIHMgis and PIHM model to other watersheds.

Newly obtained data from Henry Lin's group is ready to update the soil map, the surface elevation map and the bedrock elevation map. As the experiments go on in the Shale Hills watershed, the ultimate goal of the work is to establish a real-time hydrologic model to help us understand the hydrologic behavior and facilitate the studies in the watershed.

Chapter 2

Methodology and Approach

2.1 PIHM model

2.1.1 Semi-discrete volume method

The hydrological processes within a watershed are represented by a mixture of PDEs (Partial Differential Equations) and ODEs (Ordinary Differential Equations). In PIHM, semi-discrete finite volume method is applied to approximate the physical processes governed by PDEs on unstructured grid cells, and reduced them to ODEs, since SUNDIALS package has the state-of-art ODE solver. This section detailed the steps of the semi-discrete finite volume method.

Eq. 1 is a general equation of mass conservation; where A is an arbitrary scale (e.g. Mass fraction of storage), V is the velocity vector, and S is the local source/sink term for the process. Consider the delta prism used in PIHM model, the velocity is decomposed into horizontal components U= {u, v}, and vertical component {w}.

$$\frac{\partial A}{\partial t}+\nabla\cdot(AV)=S \tag{1}$$

$$\frac{\partial A}{\partial t}+\nabla\cdot(AU)+\frac{\partial Aw}{\partial z}=S \tag{2}$$

First, integrate Eq.2 over the depth of the prism, from z_a to z_b. Notice that the 2 boundaries are free surfaces. Then consider a thin layer $\varepsilon \to 0$ at upper and lower boundary respectively, the integral of Eq. 4 will go away, and flux from the 2 boundaries of the thin layer will have same terms, thus, we can define Q_a and Q_b as in Eq.5 and Eq.6, the flux through the upper and lower boundaries of the prism. Substitute them into Eq. 4, we'll get Eq. 7.

$$\int_{z_a}^{z_b}\frac{\partial A}{\partial t}dz+\int_{z_a}^{z_b}\nabla\cdot(AU)dz+\int_{z_a}^{z_b}\frac{\partial Aw}{\partial z}dz=S \tag{3}$$

$$\frac{\partial}{\partial t}\int_{z_a}^{z_b}Adz-A_{z_b}\frac{\partial z_b}{\partial t}+A_{z_a}\frac{\partial z_a}{\partial t}+\nabla\int_{z_a}^{z_b}AVdz-(VA)_{z_b}\nabla z_b+(VA)_{z_a}\nabla z_a+(wA)_{z_b}-(wA)_{z_a}=\int_{z_a}^{z_b}Sdz \tag{4}$$

$$Q_a=A_{z_a}\frac{\partial z_a}{\partial t}+(VA)_{z_a}\nabla z_a-(wA)_{z_a} \tag{5}$$

$$Q_b=A_{z_b}\frac{\partial z_b}{\partial t}+(VA)_{z_b}\nabla z_b-(wA)_{z_b} \tag{6}$$

$$\frac{\partial}{\partial t}\int_{z_a}^{z_b}Adz+\nabla\int_{z_a}^{z_b}AVdz-Q_b+Q_a=\int_{z_a}^{z_b}Sdz \tag{7}$$

Define that $\omega=\int_{z_a}^{z_b}Sdz$ and $\overline{A}=\int_{z_a}^{z_b}Adz$ Eq.2 will be changed as

$$\frac{\partial \bar{A}}{\partial t} + \nabla(V\bar{A}) = \omega + Q_b - Q_a \tag{8}$$

Integrate Eq. 8 over the area Ω to complete the volume integration, Γ is the perimeter of the area, N is the unit vector of the normal direction

$$\frac{\partial}{\partial t}\int_A \bar{A} d\Omega + \int_\Gamma N(V\bar{A}) d\Gamma = \int_A (Q_b - Q_a + \omega) d\Omega \tag{9}$$

$$\frac{d\bar{\bar{A}}}{dt} = \sum_{k=1}^{2} Q_k - \sum_{i=1}^{m} Q_i \tag{10}$$

Writing equation (9) in semi-discrete finite volume form yields Eq. 10, where Q_i is net volumetric flux through the sides i = 1, 2, 3 of the control volume, and Q_k is the net volumetric flux across the upper and lower boundaries k = 1, 2. The $\bar{\bar{A}}$ can be a representation of all the processes in the control volume. Thus the finite volume method guarantees mass conservation for each control volume, and that the semi-discrete representation reduces all equations to a standard ODE form.

2.1.2 Domain Decomposition and Fully Coupled Model Architecture

The watershed domain is decomposed to unstructured mesh by Delaunay triangulation with constrains of the stream, hydraulic structures, etc. Then each triangular is vertically projected to the depth to the bedrock to form a prism, which is the finite volume element in the model. The prism is then divided into surface and subsurface

layers, and the subsurface layer is subdivided into saturated zone and unsaturated zone. The hydrological processes are independent of the domain decomposition. Each hydrological process is related to a single layer inside the prism. For example, overland flow and channel flow are in the surface layer. Thus the whole set of processes reside in the prism. The ODE system is fully coupled with no time lagging or iterative linking of processes. The independence of model kernel makes it convenient to be modified, in order to customize the application for a particular watershed.

2.1.3 Physical Processes

The main physical processes in the model include surface overland flow, channel flow, subsurface flow, interception, snow accumulation, evaporation and evapotranspiration, and these processes are connected by the throughfall drainage, infiltration, surface overland flow to river, saturated-unsaturated flow, the macropore infiltration and macropore stormflow is also included in the model.

In this section , we will briefly explain all the physical processes in the PIHM model, especially those that enhanced in the second version. The followed Kumar (2009) and Qu's(2007) description.

I Throughfall Drainage.

The throughfall drainage is defined as

$$G_{tf} = \begin{cases} ke^{(b\frac{\psi_{\text{int}}}{\psi_{\text{int max}}})} & 0 \le \psi_{\text{int}} < \psi_{\text{int max}} \\ ke^{b} & \psi_{\text{int}} \ge \psi_{\text{int max}} \end{cases} \qquad (11)$$

Where $k = 3.91\times10^{-5} mm/\min$, and b ranges from 3.0 to 4.6 suggested by Rutter and Morton (1977). $\psi_{\text{int max}}$ is the canopy water storage capacity, which relies on the LAI as $\psi_{\text{int max}} = K \cdot LAI$, and K is assumed 0.2 mm.

II Evapotranspiration

Evapotranspiration is comprised of evaporation from bare land, from the river, from overland flow, from interception, and transpiration from the canopy. Pennman equation is used for estimating the evaporation from overland flow, river, vegetation interception, and Pennman-Monteith equation is used for the potential evapotranspiraton from soil and plant. And the evapotranspiration is a sum of them all.

$$et = \frac{\Delta(R_n - G) + \rho_a C_p(\varepsilon_s - \varepsilon_a)}{\Delta + \gamma} \tag{12}$$

$$et_0 = \frac{\Delta(R_n - G) + \rho_a C_p \dfrac{(\varepsilon_s - \varepsilon_a)}{r_a}}{\Delta + \gamma(1 + \dfrac{r_s}{r_a})} \tag{13}$$

III Snow Melt

Snowmelt Process is represented by Eq. 14, (Dingman, 1994) and where T_m is a critical temperature, above which, the snow begins to melt. The melting factor M is assumed to be 1.8~3.7 $mm/^oC$.

$$\Delta w = \begin{cases} M(T_a - T_m), & T_a > T_m \\ 0 & T_a \le T_m \end{cases} \tag{14}$$

IV Infiltration

Infiltration is represented by Eq. 15, K is the conductivity of the top soil surface layer, h_{ovl} and h_u are the overland flow depth and equivalent water depth of the unsaturated zone, z and z_u are the elevation of the ground surface and the bottom of the unsaturated zone.

$$\mathrm{F}_{\inf} = K\frac{(h_{ovl}+z)-(h_u+z_u)}{d} \tag{15}$$

V Surface Overland Flow

The surface overland flow between 2 neighboring cells is calculated by Eq. 16.

$$F_{surf} = \frac{h_{ovl}^{\frac{2}{3}}}{n}\cdot\nabla(h_{ovl,i}+z_i)^{-\frac{1}{2}}\cdot\frac{(h_{ovl,i}+z_i)-(h_{ovl,j}+z_j)}{d_{ij}} \tag{16}$$

VI Surface Overland Flow to River

The $(h+z)_u$ and $(h+z)_d$ are chosen from the overland water head and the river water head, the choice depends on which is higher, thus also decides the direction of the flow, the z_{rb} is the depth of the river bank, which is also a restriction for the flow.

$$F_{o2r} = C_{o2r}\cdot\frac{2}{3}\cdot\sqrt{2g}\left((h+z)_u-\max\left[(h+z)_d,z_{rb}\right]\right)^{\frac{1}{2}} \tag{17}$$

VII Channel –Aquifer Flow

The exchange of water between the river and the aquifer is calculated as Eq. 18. It is used in calculating the flow through the river bank and river bed.

$$F_{ch} = K_{eff}\cdot\frac{(h_{riv,i}+z_{r,i})-(h_s+z_b)}{d_{ij}} \tag{18}$$

VIII Channel Flow

The Channel Flow is calculated by Eq. 19, to represents river flux from upstream to downstream.

$$F_{riv} = \frac{h_{r,i}^{\frac{2}{3}}}{n} \cdot (\nabla(h_{r,i} + z_{r,i}))^{-\frac{1}{2}} \cdot \frac{(h_{r,i} + z_{r,i}) - (h_{r,j} + z_{r,j})}{d_{ij}} \quad (19)$$

VIIII Saturated-Unsaturated Zone Flow

The flow between the saturated and unsaturated zone is calculated by Richard's equation in the vertical direction, and the unsaturated zone conductivity is estimated by van Genuchten equation.

$$F_{us} = \frac{K_u K_s(\alpha(z - z_b - h_s) - 2(-1 + S^{\frac{\beta}{1-\beta}})^{\frac{1}{\beta}})}{\alpha(K_u h_s + K_s(z - z_b - h_s))} \quad (20)$$

$$K_u = S^{1/2}(1 - (1 - S^{\frac{\beta}{\beta-1}})^{\frac{\beta-1}{\beta}})^2 \quad (21)$$

$$S = \frac{h_u}{z - z_b - h_s} \quad (22)$$

X Macropore Storm Flow

Root holes, crack forms macropores in the soil layers. Although the existence of macropores only occupies a small fraction compared to the soil matrix, but its large conductivity creates a relatively fast fluid flow. Thus, within the mode β is the volumetric fraction of the macropore, K_{mat} and K_{mac} are the conductivity for the soil matrix and macropore respectively.

$$K_{eff} = K_{mat}(1 - a) + K_{mac} a \quad (23)$$

XI Ground Water Flow

The ground water flow is governed by Darcy's law as in Eq. 24, where the K_{eff} is the effective conductivity. z_b is the bedrock depth, and h_s is the depth of saturated zone.

$$F_g = K_{eff} \cdot \frac{\left((h_{s,i} + z_{b,i}) - (h_{s,j} + z_{b,j})\right)}{d_{ij}} \tag{24}$$

XI The water balance for different processes

As stated in Chapter one, the computation element – the prism has been divided into several layers in order to accommodate all the processes inside the model. And in each layer, the water budgets inside each processes are our main concerns.

i) The interception is the water that is intercepted by the leaves and branches of the canopy during a precipitation. The main influence comes from the precipitation, the throughfall process, and the evaporation from canopy.

$$\frac{d(IS)}{dt} = P \cdot v - F_{tf} - et_{canopy} \tag{25}$$

ii) Snow package is accumulated during precipitation, when the temperature is below a critical value, and is melted . f_s is a coefficient that determines the percentage of snow fall during a precipitation. T_a is the air temperature, $T_r = 1^oC$, and $T_s = -3^oC$.

$$\frac{d(SP)}{dt} = (1 - f_s) \cdot P - \Delta w \tag{26}$$

$$f_s = \begin{cases} 1.0 & T_a < T_s \\ \dfrac{T_r - T_a}{T_r - T_s} & T_s \le T_a \le T_r \\ 0 & T_a > T_s \end{cases} \quad (27)$$

iii) The surface overland flow depth is influenced by many factors, the precipitation, the evaporation from overland flow, the throughfall drainage, the snow melt, the infiltration, the surface flow in and out of the cell, and flow to river, when it the cell is neighboring a river element.

$$\frac{dh_{ovl}}{dt} = (1-v)\cdot P + F_{tf} - F_{\inf} - et_{ol} + \Delta w + /- F_{surf} + \| F_{o2r} \| \quad (28)$$

iv) The unsaturated zone is controlled by the infiltration and the recharge to the saturated zone

$$\frac{dh_u}{dt} = F_{\inf} - F_{us} \quad (29)$$

v) Saturated Zone is mainly controlled by the infiltration rate and the groundwater flow from cell to cell.

$$\frac{dh_s}{dt} = F_{us} + \sum_{j=1}^{3} \frac{1}{A} F_{g,j} \quad (30)$$

vi) The water stage in the channel is influenced by the upstream and downstream stream flow, the overland flow to river, the water recharge to the bed, and the flow exchange through the river bank.

$$\frac{dh_{riv}}{dt} = F_{riv,i} - F_{riv,i+1} + F_{o2r} - F_{ch} \quad (31)$$

2.2 PIHMgis

PIHMgis merged the open source Qgis interface to PIHM model as the pre-processing for the data model. The main developer of the code was G. Bhatt while M. Kumar developed the underlying data-model for PIHM_GIS. The procedure for developing physical models at a watershed such as Shale Hills mainly involves delineating the watershed model and creating the input files for the PIHM model. Raster Processing, Vector Processing, Domain Decomposition, Data Model Loader are each carried out sequentially to incorporate the topographic data of soil properties, hydrogeology, land cover, climate data into the model. Figure 2.1 shows the framework of the PIHMgis.

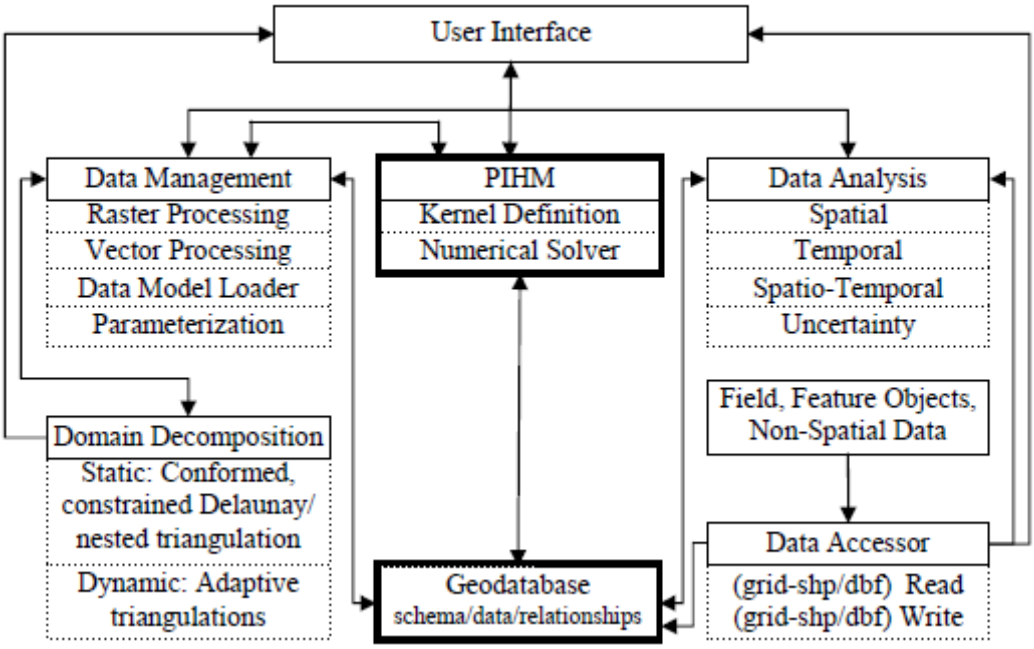


Figure 2.1 PIHMgis Architectural Framework (Bhatt et al. 2008)

Next the steps and issues of implementing PIHM are described. Raster processing delineates the watershed and defines the stream from the DEM file of a

watershed. Fill Pits identifies the pitfalls inside a grid, and raise their elevation so that it can drain off the edge of the grid. Flow Grid outputs a flow direction grid, and a flow accumulation grid. D8 algorithm identifies the steepest slope to neighboring cells for a particular cell, and encoded it as 1 - east, 2 – northeast, 3 –north, 4 – northwest, 5 – west, 6 –southwest, 7 – south, 8 – southeast for each direction in the flow direction grid. Flow accumulation grid contains the accumulated number of cells upstream of a cell. Stream Grid is extracted from the flow accumulation grid, cells that have the value larger than the given threshold is defined as stream, and marked as 1, while the other cells are marked with NoData. It is a raster equivalent of the streams. Link divides the stream grid into segments at junctions. A unique integer is assigned to each segment starting with 1, while the rest of the grid is marked with NoData. The Link grid is converted to polyline features, which represents the drainage network for the watershed. Flow direction grid is used to ensure that the segments are topographically correct. Catchment grid characterizes the cells that drain to a particular stream segment with integer starting with 1. Catchment Polygon creates the vector feature for the catchment cells which draining to a single outlet.

Vector Processing helps to define constraints for domain decomposition. Constraints vary from stream, watershed boundary, land cover classification, soil classification. Before merging them into one file, they have to be converted to polylines. Polygon to polyline enables us to extract the boundary of polygons, and make polylines. Simplify line is used to simplify a polyline by removing small fluctuations or extraneous bends while preserve its essential shape. This step is crucial for quality and efficient in domain decomposition. Split line splits polylines at each vertex, turning a single polyline feature into a multiple line feature depending upon the number of vertices present in the original polyline. Vector Merge is the final step of Vector processing. It merges all the

layers into one shape file. The merged shape file acts as constraints in domain decomposition process.

Domain decomposition applies Delaunay Triangulation to decompose the modeling domain into triangular irregular network (TIN). A terrain can be better represented by an irregular mesh if all the critical terrain and hydrographic points are taken into account while performing domain decomposition. This may include watershed boundary, different types of contours (e.g. hypsometry, soil), stream network, hydraulic structures (e.g. dams, gages) for generating those points. Read Shape Topology prepares a .poly file which acts as input for running the TRIANGLE in the next section. All the node and line information from the input shape file (vector merged shape file obtained in the earlier section) is transformed into the poly file. A .poly file represents a PSLG, as well as some additional information. PSLG stands for Planar Straight Line Graph. By definition, a PSLG is just a list of vertices and segments. A .poly file can also contain information about holes and concavities, as well as regional attributes. 'TRIANGLE' is one of the efficient implementation of this algorithm which provides several flexibility and options to a user. It is also computationally efficient. It takes planner straight line graph (PSLG) as input. The algorithm works in such a way that it refines the Delaunay triangulation by inserting carefully placed vertices until the generated mesh meets a provided quality and size criterion. TIN Generation read in the .ele and .node file and generates a polygon shapefile with all the triangle elements. Data model loader reshapes the structure of GIS maps into a geo-database which is specifically designed according to the PIHM data model. The files prepared constitute spatial and relational attributes of the modeling domain. PIHM requires input in the form of 11 separate files: (1) mesh file; (2) att file; (3) soil file; (4) lc file; (5) riv file; (6) forc file; (7) ibc file; (8) para file; (9) calib file; and (10) geol; (11) init file.

Chapter 3

A-priori data for Shale Hills watershed

A-priori data refers to surface elevation, bedrock elevation, soil mapping, land cover mapping, and climate data in PIHM model. In this Chapter, the author looked into the Shale Hills CZO experimental data from 1970s to present. To develop the datasets needed for the model.

3.1 Surface elevation

Figure 3.1 shows a surface elevation map of the Shale Hills experimental watershed. The dataset is 3m accuracy. It is refined by the local measurement data based the 10m DEM file that was downloaded from SSURGO (Soil Survey Geographic Dataset).

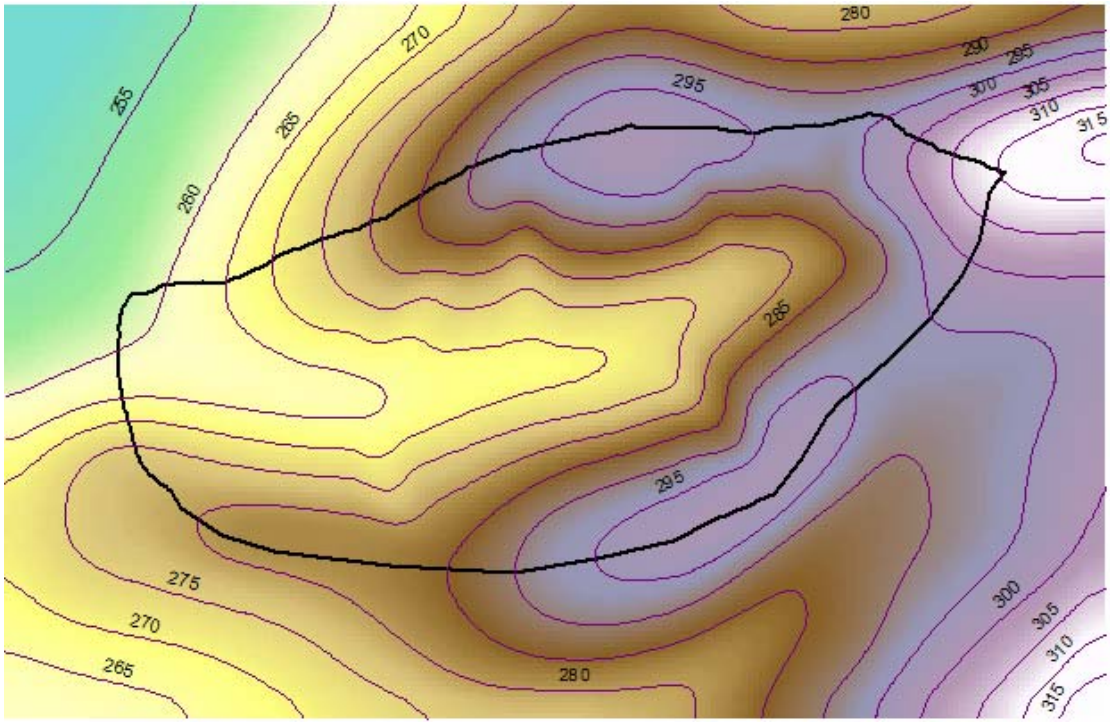


Figure 3.1 Surface Elevation (Contours in meters)

3.2 Bedrock Elevation

The bedrock elevation map is developed based on the data from the 2004 experiments. 77 sites distributed throughout the watershed as is shown in Figure 3.2 collected data for the depth of the ground surface to the bedrock. The data was input into ArcGIS. Second order local polynomial interpolation in Geostatistical Analysis was applied to develop the map for the bedrock depth. The bedrock depth grid is subtracted from the surface elevation grid developed in 3.2 to get the bedrock elevation of the Shale

Hills watershed. The anisotropy is considered in the data interpolation with a major semi-axis along the direction of the stream. (200 for major semi-axis, 100 for minor semi-axis)

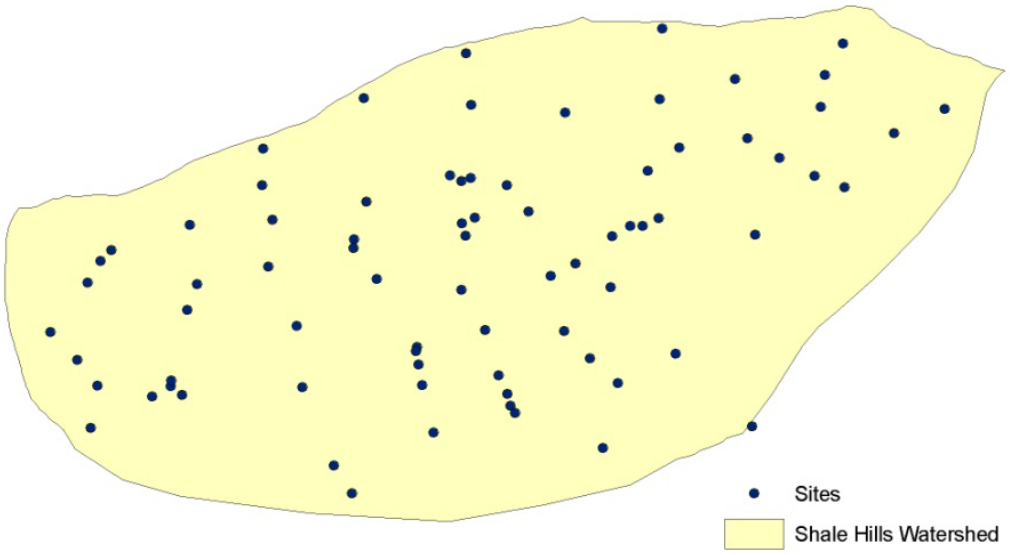


Figure 3.2 Sites Distribution

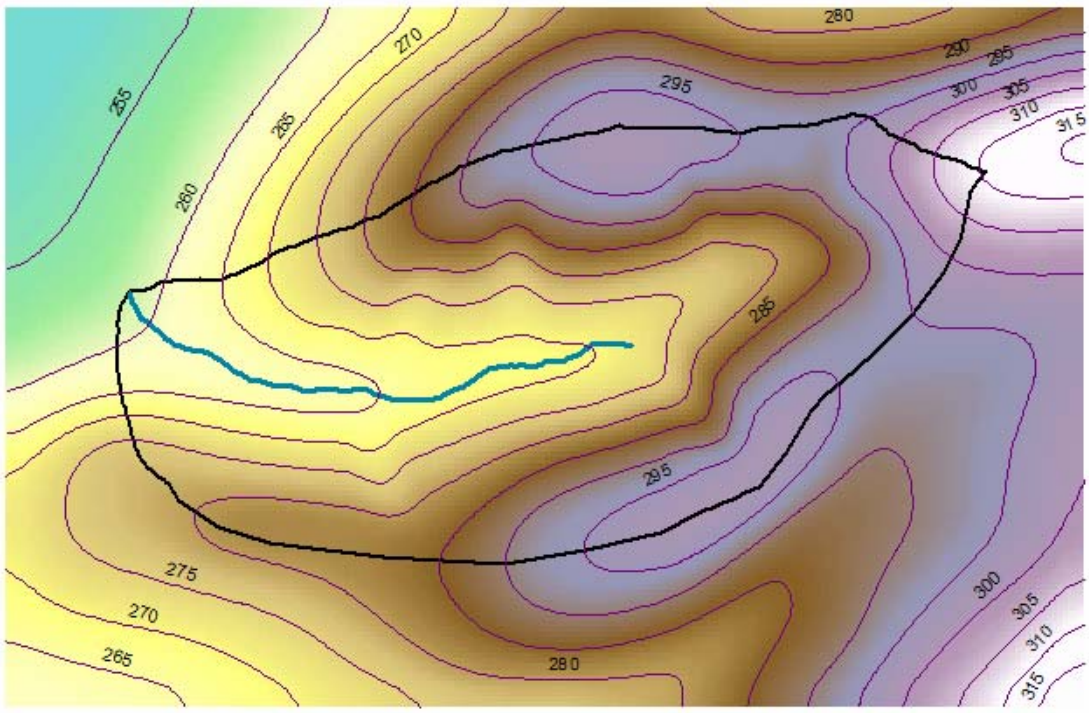


Figure 3.3 Bedrock Elevation(Contour in meters)

3.3 Soil Mapping

The spatial distribution and the properties of soil are considered as crucial factors for the PIHM model. Lin identified 5 soil categories for Shale Hills watershed in his research, which are Berks, Blairton, Ernest, Rushtown and Weilkert according to the hillslope and the soil moisture distribution.

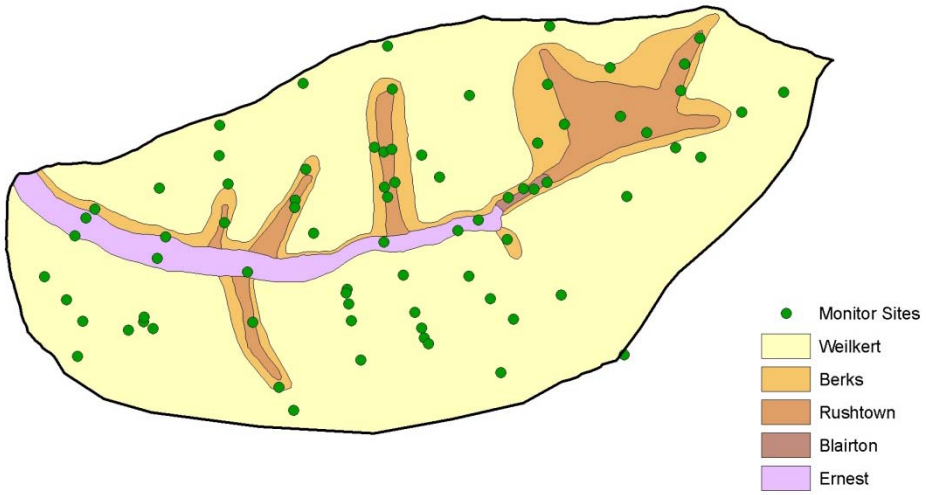


Figure 3.4 Shale Hills Catchment, Monitoring Sites and soil mapping.

Another map is also found on the SSURGO database as shown in Figure 3.5. As for the Shale Hills experimental watershed, there are 4 major soil types which are BMF, ErB, BkC, and BID. The soil types can be found in the soil data mart in the SSURGO server. Since BID and BkC occupies only a small portion of the watershed, thus, they are omitted from our model. Because of the coarse grid, however, it is not a good

representation of the soil mapping in our Shale Hills model. But it still provide us with information as reference to the soil properties. Table 3.1 listed the data from the SSURGO soil data mart. The percentages of silt, sand and clay was calculated in Rosetta (Schaap, 1999), to predict the soil hydraulic parameters. However, a new soils dataset from Lin recently became available with more detailed soil coverage so that was used in this thesis.

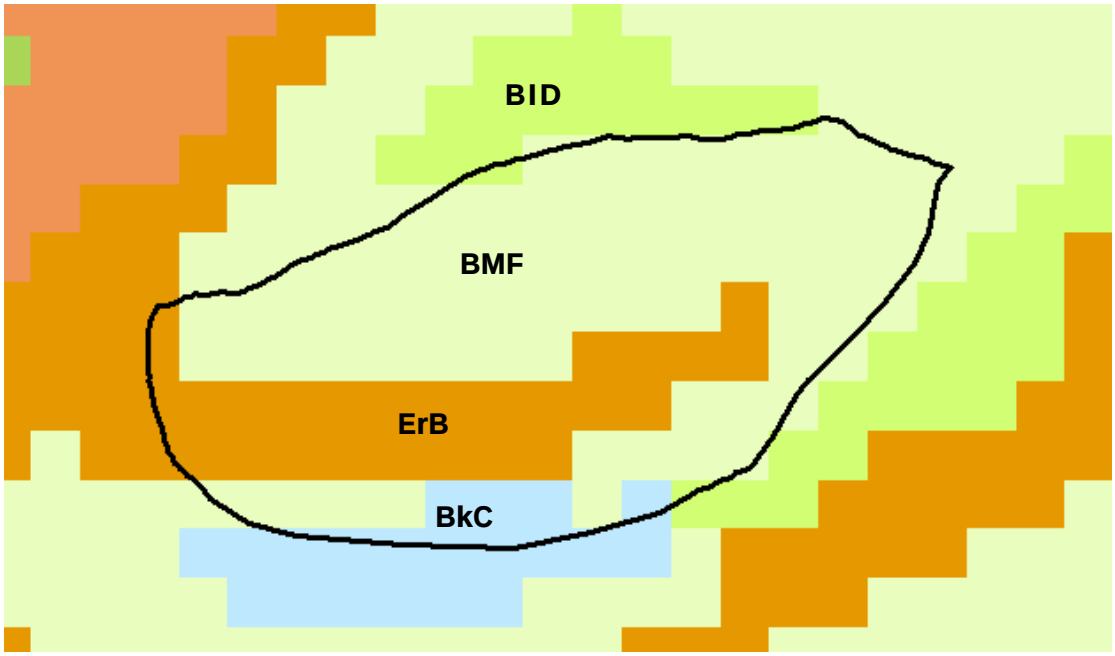


Figure 3.5 Soil Classifications from SSURGO

	Depth(cm)	*Sand Pct*	*Silt Pct*	*ClayPct*
BMF				
Berks	0-4	20-	35-	5-23
	4-19	20-	35-	5-32
	19-	20-	35-	5-32
	24-	---	---	---

Weikert	0-7	15-	30-	15-
	7-17	15-	30-	15-
	17-	---	---	---
ERB				
Ernest	0-8	15-	50-	20-
	8-26	10-	50-	20-
	26-	10-	50-	20-
	51-	10-	50-	20-

Table 3.1 Percentages of sand, silt and clay (SSURGO)

	Sand%	Silt %	Clay%
Weilkert	15-50	30-70	15-27
Estimated	30	50	20
Ernest	15-35	50-80	15-27
Estimated	23	65	12

Table 3.2 Estimated sand silt and clay percentage for soils.

	Resi. Prosity	Alpha	Beta
Weilkert	0.0653	0.595525	1.589278
Ernest	0.0557	0.426973	1.705297

Table 3.3 Estimated soil properties from Rosetta

	KsatH	KsatV	Porosity
Weilkert	40.32	10.8	0.638
Berks	2.678	0.8928	0.37
Rushtown	4.248	1.25	0.382
Blairton	9.3456	0.2592	0.373
Ernest	2.6784	0.864	0.369

Table 3.4 Soil property data from Lin (2006)

Table 3.2 shows the estimated sand, silt, and clay percentages from the SSURGO dataset. Since the SSURGO dataset is too coarse for our problem, only 2 kinds of soil can be identified. These data are put into Rosetta, and get the values in Table 3.3 for residual porosity and coefficient for van Genuchten parameters. On the other hand, the 2004 experiment also provided us with the porosity, horizontal and vertical saturated conductivity as shown in Table 3.4. Besides, the macropore hydraulic conductivity is estimated 100 times of the matrix hydraulic conductivity value as the A-priori value. And the percentages for macropore to soil matrix are assumed 1% to 30%, as a range for calibration.

Chapter 4

Getting Data into Model

A Short PIHMgis Tutorial (Shale Hills Watershed)

4.1 Introduction

After getting the data ready, the next step is to get the data into the model, the grid files needs to be converted to data files that the model can read. Bahtt et al (2008) coupled the PIHM model with Qgis (Quantum Geographic Information System) to facilitate the preprocessing of the data model with minimum data redundancy and optimal retrievability. Bahtt (2009) developed a V-shape ideal catchment tutorial for the PIHMgis system. In this chapter, we will navigate through the process of developing PIHM data model for Shale Hills experimental watershed, and explain briefly the algorithm in each step. In this process, realistic problems like identifying the watershed, correction of the mismatch of the stream and the boundary, will be demonstrated.

4.2 Software Initialization

Choose the right version of PIHMgis for the operating system you are using, install it onto your computer according to the instructions. PIHMgis applies QuantumGIS as the base GIS framework. Then launch the QuatumGIS. If the PIHMgis plugin tab does not show (most time for first time users), load it by selecting Plugins >> Plugin Manager, then checking the box in front of PIHMgis. Click OK to exit, and the PIHMgis will show up as in Figure 4.1.

Adding a raster or vector layer can be accomplished by selecting the items in the Layer dropdown menu.

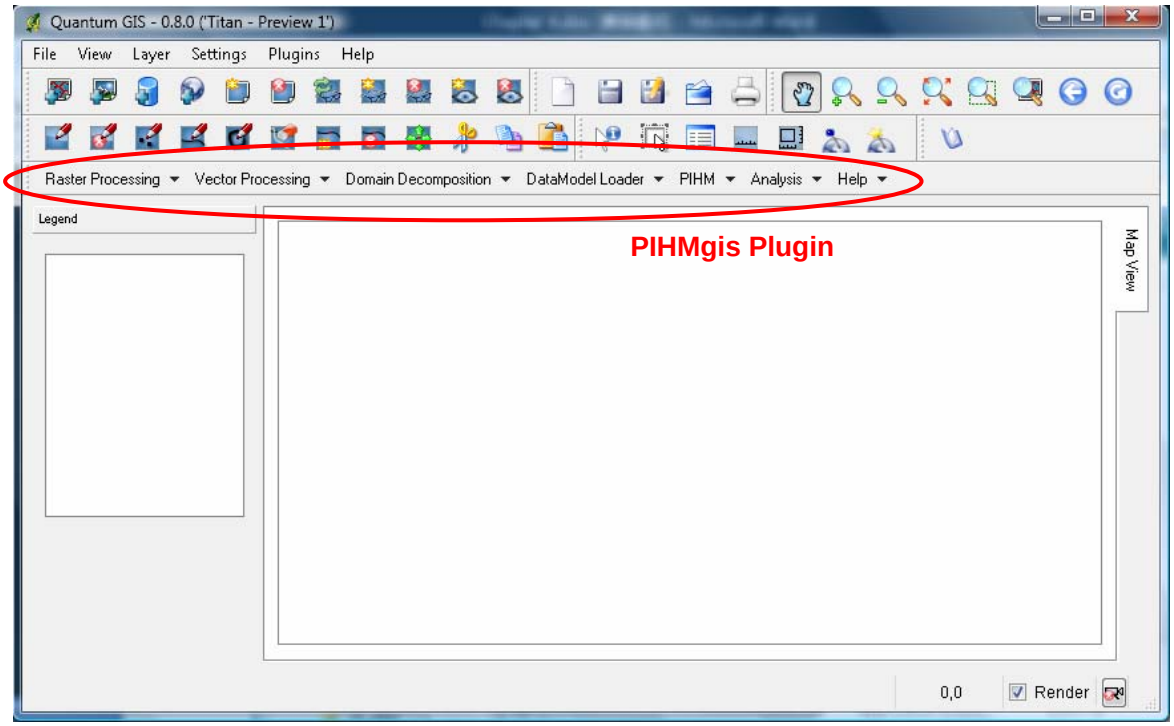


Figure 4.1 PIHMgis Interface (PIHMgis_v2.2beta)

4.3 Data Preparation

Getting ready the data files for the data model. Topographic files: surface elevation, bedrock elevation. Series of soil, geology, land cover, and time series precipitation, temperature, solar radiation, relative humidity, vapor pressure, wind velocity, melt factor, source or sink, boundary conditions, macropore distribution, initial conditions for interception storage, snow, unsaturated and saturated zone, etc. should all be mapped to grids, thus PIHMgis can get access to. Make sure all your grid files are in the same coordinate system, all the spatial units are in meters other than degree or feet. Data files of the properties of the soil, geology and land cover should also be available to fulfill the process.

The first step is to create folders like Raster Processing, Vector Processing, Domain Decomposition, Data Model to store processed files separately to avoid confusion.

4.4 Raster processing

In the Raster processing step delineation of the watershed and the stream from the DEM of Shale Hills watershed is carried out. There are 7 steps to complete the raster processing: Fill pits, Flow Grid, Stream Grid, Link Grid, Stream Polyline, Catchment Grid, Catchment Polygon. All these can be found in the pull-down menu of Raster Processing.

4.4.1 Fill Pits

Fill Pits fills the pits in the grid file. If a cell surrounds by cells with higher elevations, the water will be trapped in the cell, and cannot flow out. This can cause problems for the flow routing process. The Fill Pits function identifies these locations and removes it by raising their elevations to the point where they can drain off the edge of the DEM.

Select Fill Pits in the Raster Processing dropdown menu to start the Fill Pits dialog, browse to asc file in the surface elevation folder, and save the output file in the Raster Processing folder as Pitfilled.

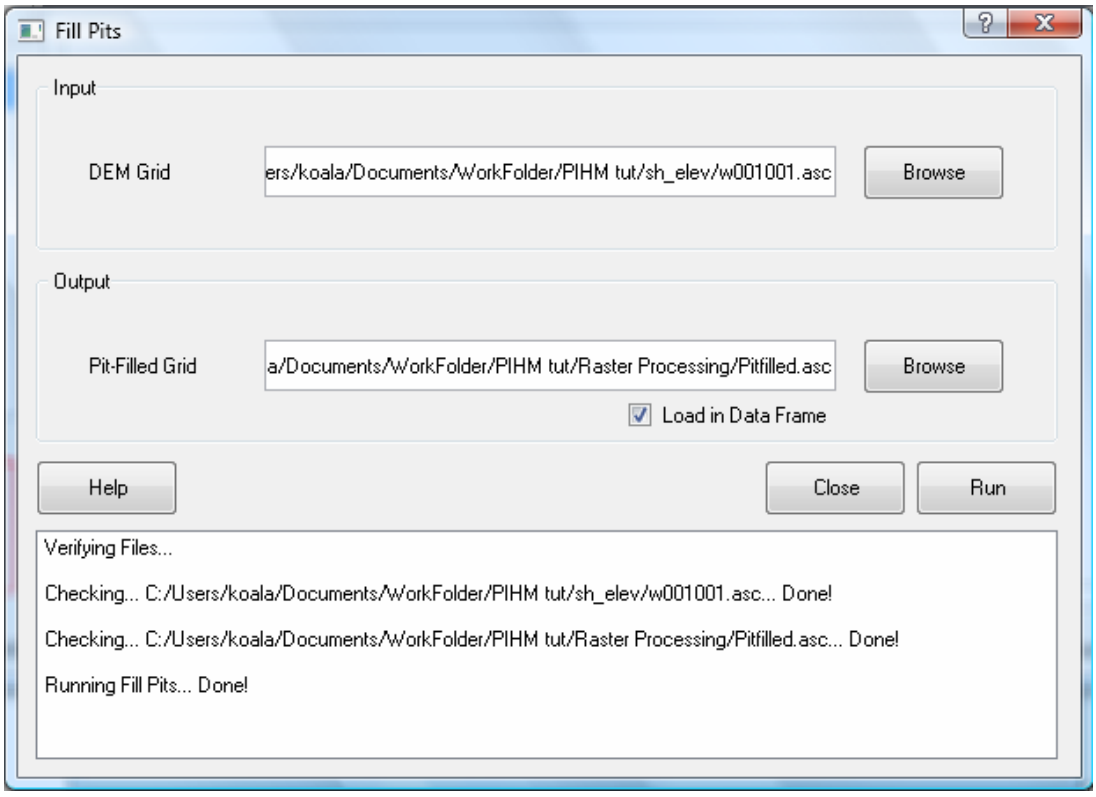


Figure 4.2 Fill Pits dialog

Checking the box in the dialog will load the raster automatically to the Qgis window after it finishes. Or load the raster layer from the Layers menu.

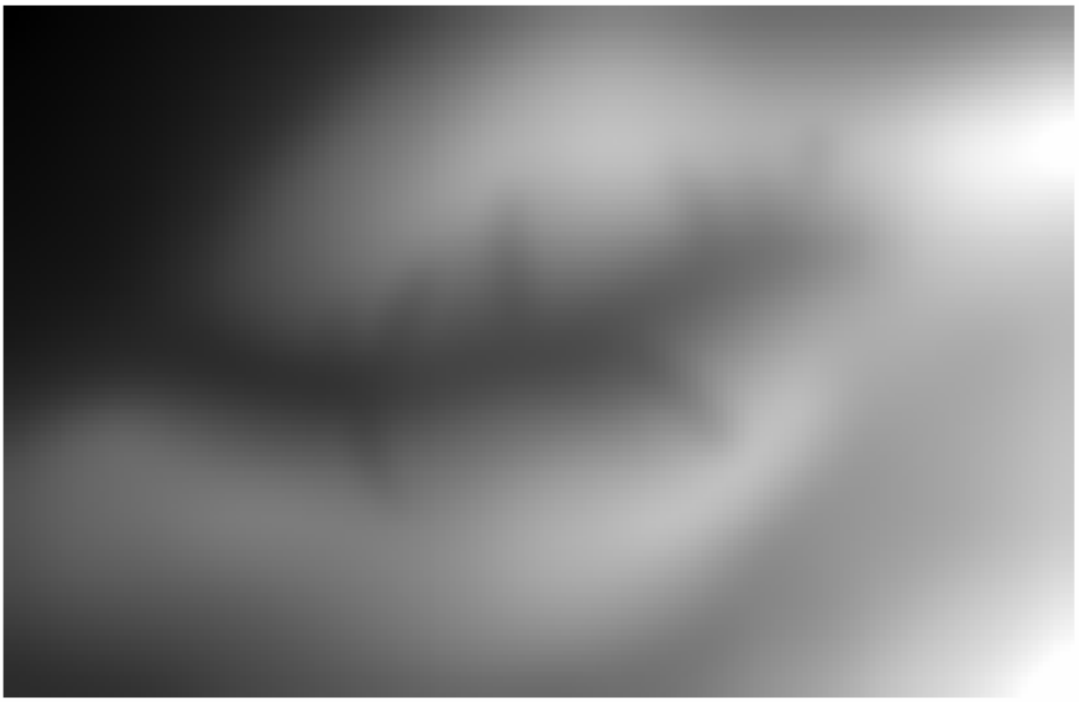


Figure 4.3 Pits-filed grid

4.4.2 Flow Grid

Flow direction and flow accumulation grid will be created in this step. Flow Direction outputs an encoded grid with the neighboring cell direction to which the steepest slope is found using D8 algorithm (O'Callaghan and Mark,1984). The encoding is 1 - east, 2 – northeast, 3 –north, 4 – northwest, 5 – west, 6 –southwest, 7 – south, 8 – southeast for each direction.

Flow accumulation outputs an accumulation grid that contains the accumulated number of cells upstream of a cell, for each cell in the input grid using a recursive procedure explained in (Mark, 1988).

Select Flow Grid in the Raster Processing dropdown menu to start the Flow Grid dialog, browse to pits-filled file just created, and save the output file in the Raster Processing folder as FlowDir and FlowAcc as shown in Figure 4.4. After it is done, close the dialog, and the flow direction and flow accumulation grid will show in the display window as in Figure 4.5 and Figure 4.6.

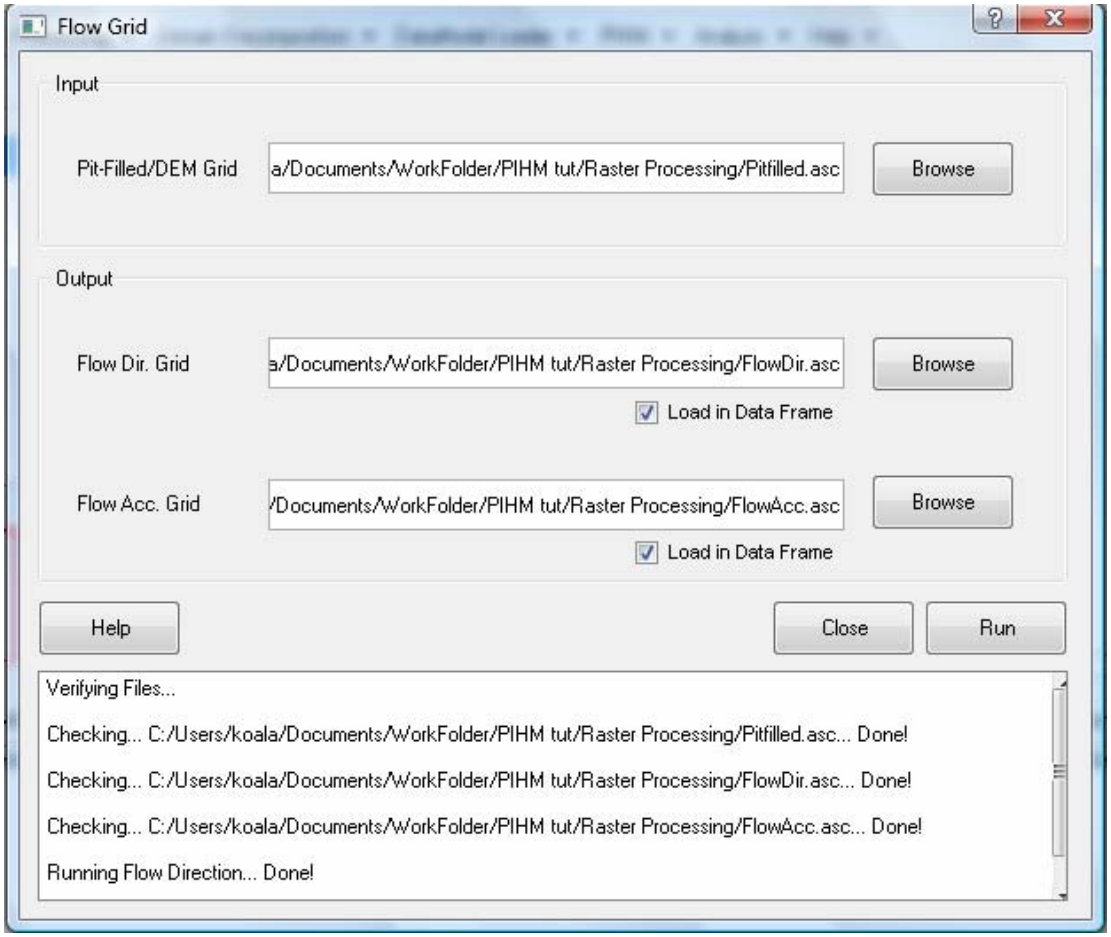


Figure 4.4 Flow Grid dialog

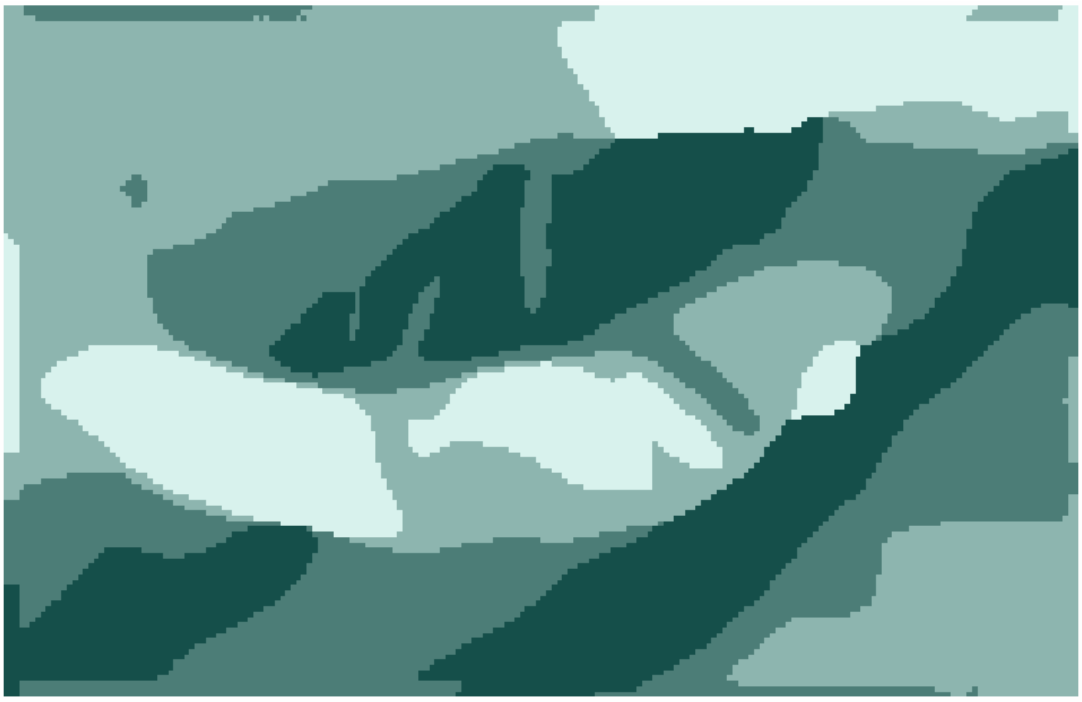


Figure 4.5 Flow direction (Different color stands for different direction of flow)

Figure 4.6 Flow Accumulation (Blue lines stand for the identified stream)

4.4.3 Stream Grid

Stream Grid is a raster equivalent of the stream network. Extracting from the flow accumulation grid, the cells that have values equal or greater than the threshold value are marked with 1 and defined as stream. Threshold implies the minimum number of cells draining to a particular cell that should be classified as strea. The rest of the cells in the grid are assumed NoData.

Select Stream Grid from Raster Processing dropdown menu, and browse to the flow accumulation grid file, FlowAcc. The threshold is set to be 1800 for the Shale Hills area. Note that the cell size is ~3 by 3 in our grid file, if the cell size changes, the threshold should change accordingly. For different watersheds, the threshold vary according to the needed resolution. The grid resolution can also reflect how the user wants to simplify the computational domain. Next name the output file StrGrid. The stream grid is shown in Figure 4.8 with the pit-filled grid file as the background, to show the location of the streams.

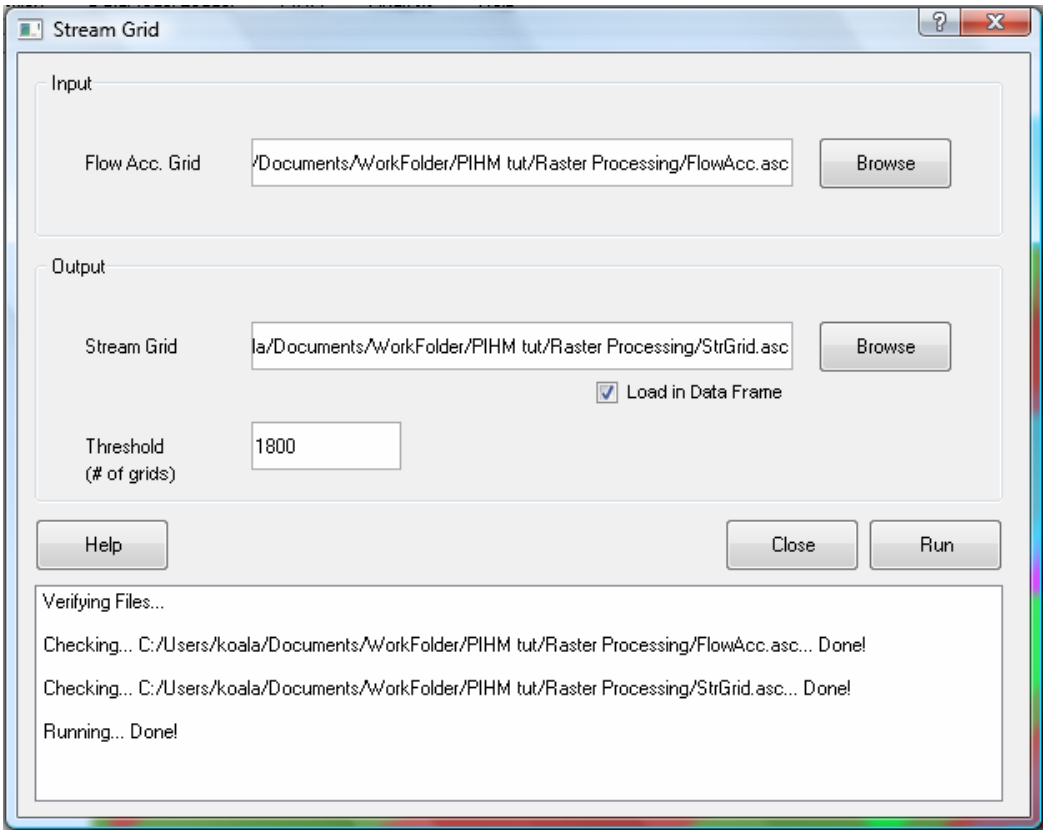


Figure 4.7 Stream Grid dialog

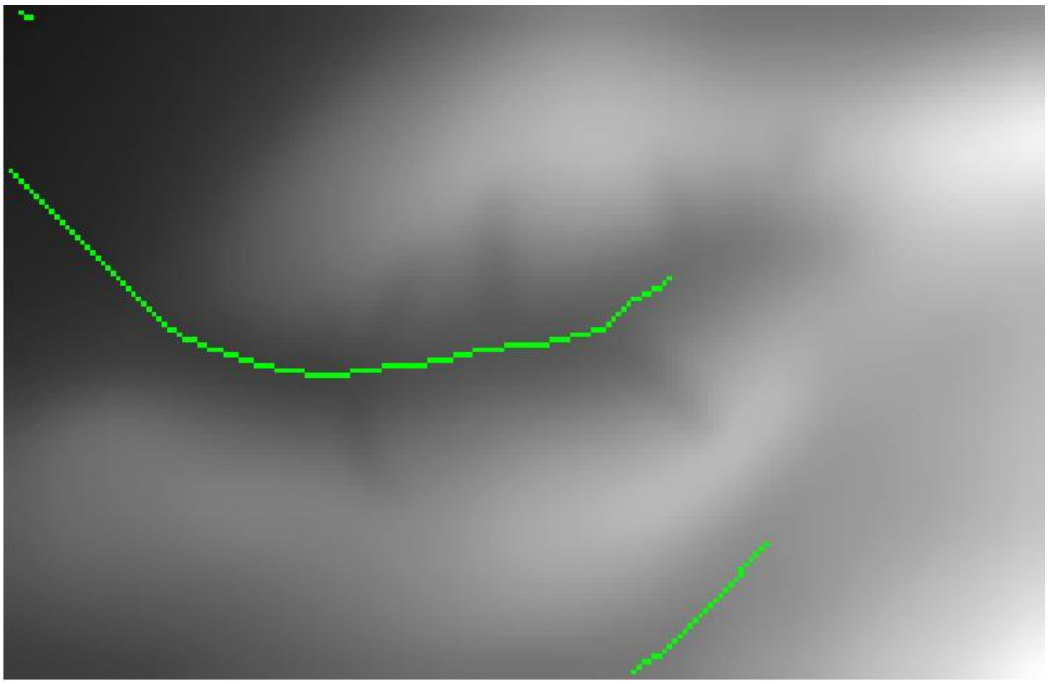


Figure 4.8 Stream Grid

4.4.4 Link Grid

Link grid separates the stream grid segments at the junctions. Each link grid segment is assigned a unique integer value starting with 1. The rest of the grid assumes No Data value similar to that of steam gird.

Select Link Grid from Raster Processing dropdown menu to start the Link Grid Dialog, select the StrGrid and FlowAcc files as the input files, and save the output file in the Raster Processing folder named Link. Close the dialog to see the Link Grid.

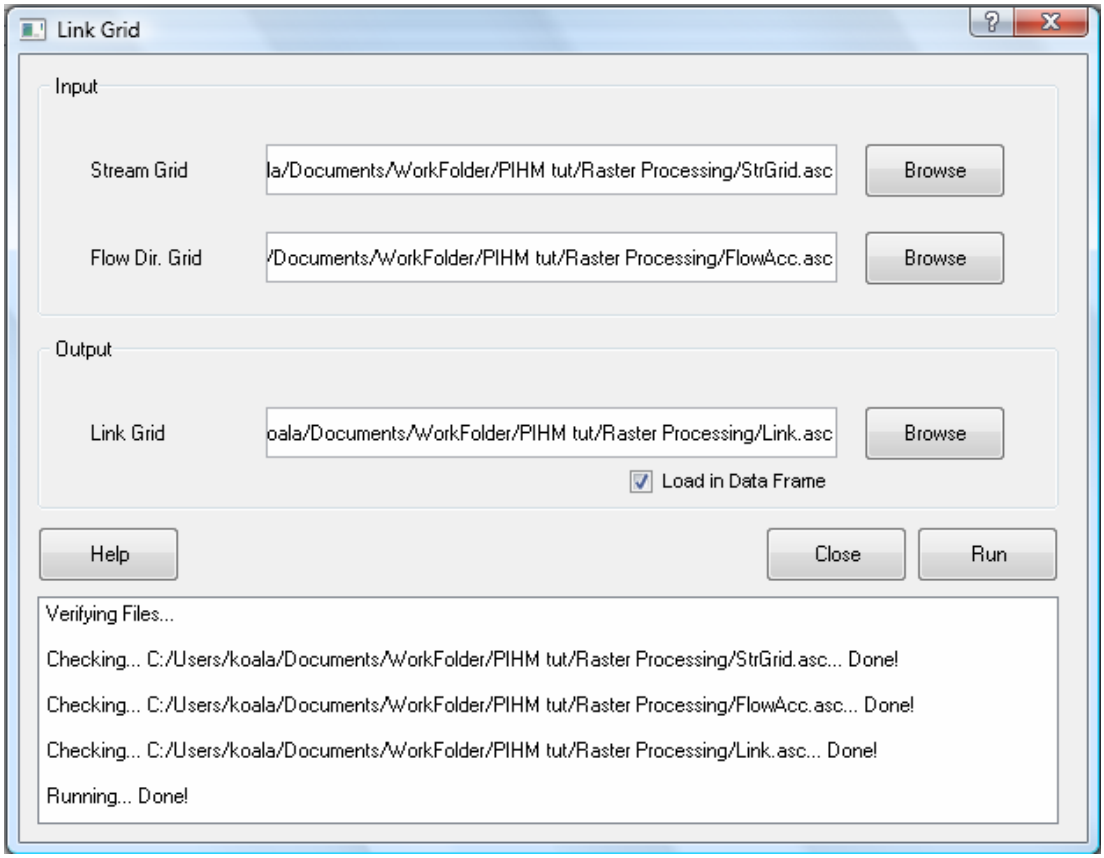


Figure 4.9 Link Grid Dialog

The 3 different streams already have different values as 1, 2, and 3 for each of them in Figure 4.10.

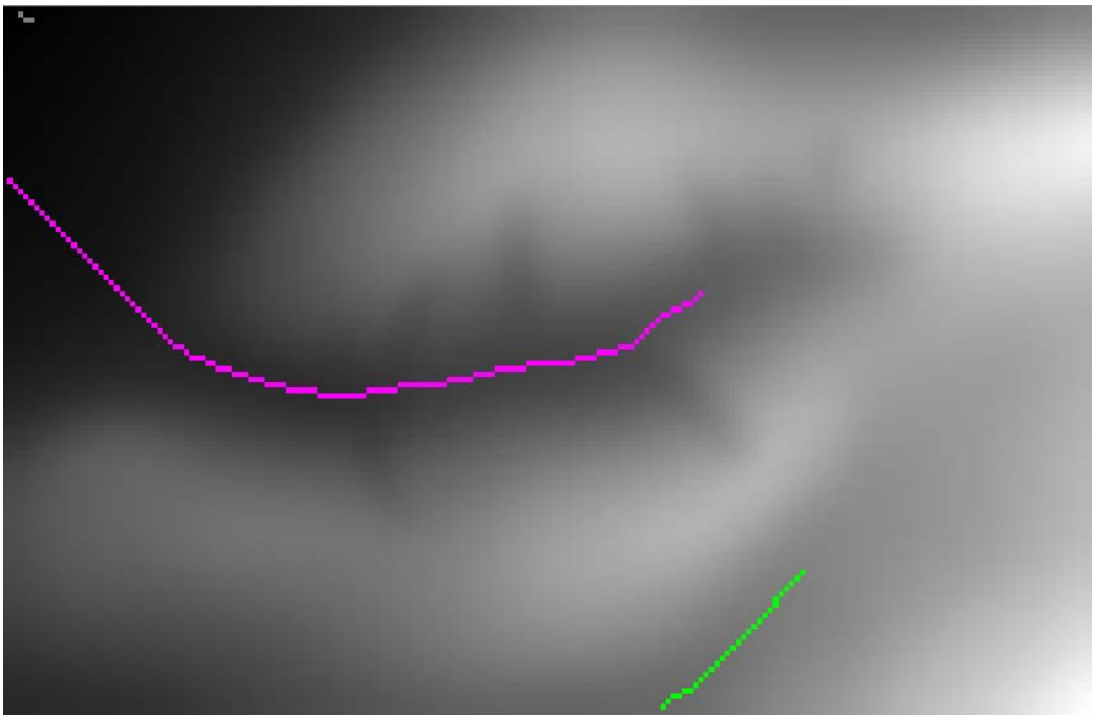


Figure 4.10 Link grid with pit-filed grid as background

4.4.5 Stream Polyline

Stream Polylines are the drainage network for the region of interest obtained by the conversion of the link grid to the vector format from the raster. Each link segment forms an individual stream segment and connected at the junction points. Flow direction is used to ensure that the segments are topographically correct (i.e. From-Node and To-Node are consistent with the flow direction).

Select Stream Polyline from Raster Processing to start the dialog, browse to the files StrGrid and FlowDir as the input, save the output file as ‘StrPoly.shp’ to Raster Processing folder. Close the dialog to see the results.

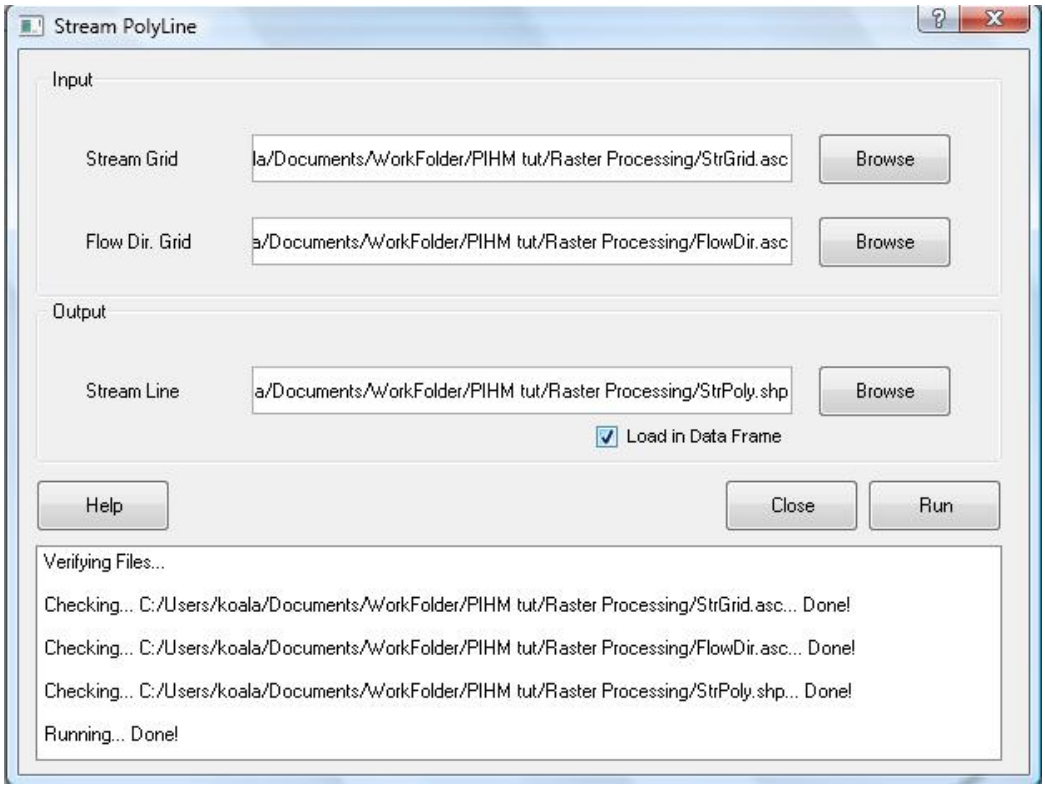


Figure 4.11 Stream Polyline dialog

4.4.6 Catchment Grid

All the grids draining to a particular stream polyline element are grouped into one type of catchment grid. Catchment grids are marked according to the stream polyline nomenclature with integer numbers starting with 1.

Select Catchment Grid from Raster Processing dropdown menu, select Link file as the stream grid, and FlowDir for the flow direction, and name the output file as CatGrid, and save it in the Raster Processing folder. Run the process, and close the dialog to see the results.

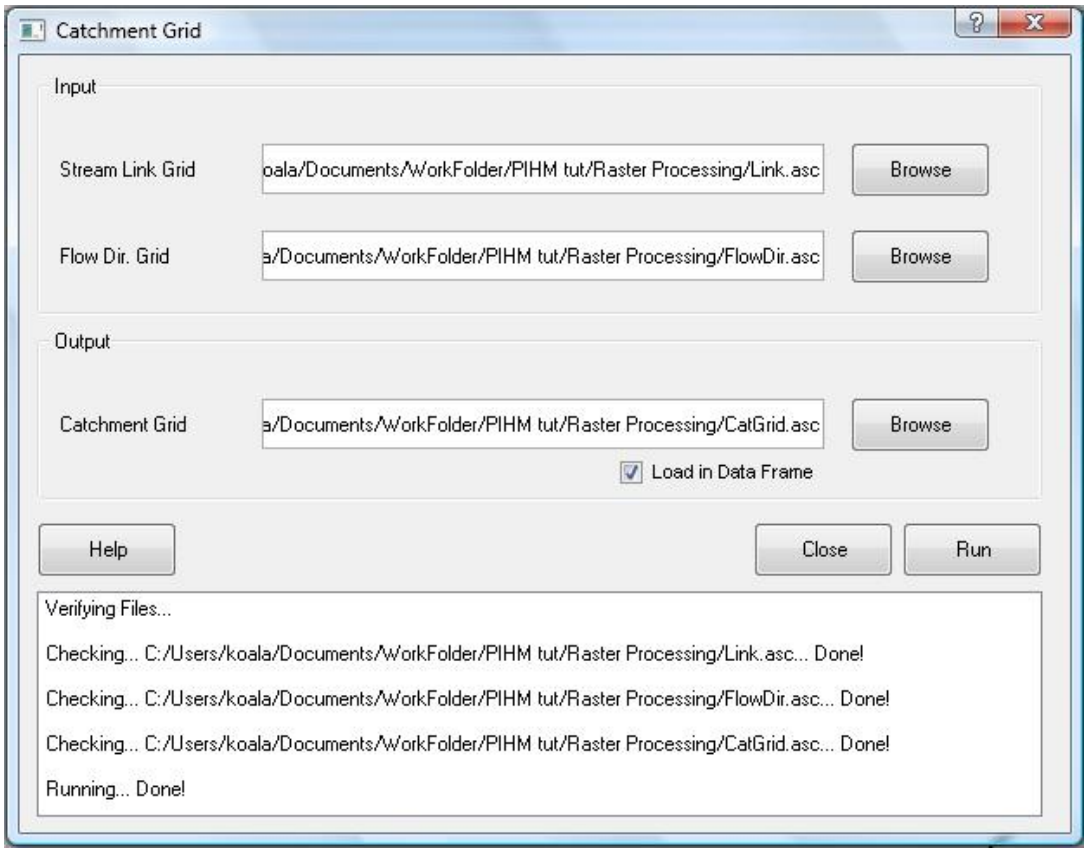


Figure 4.12 Catchment Grid dialog

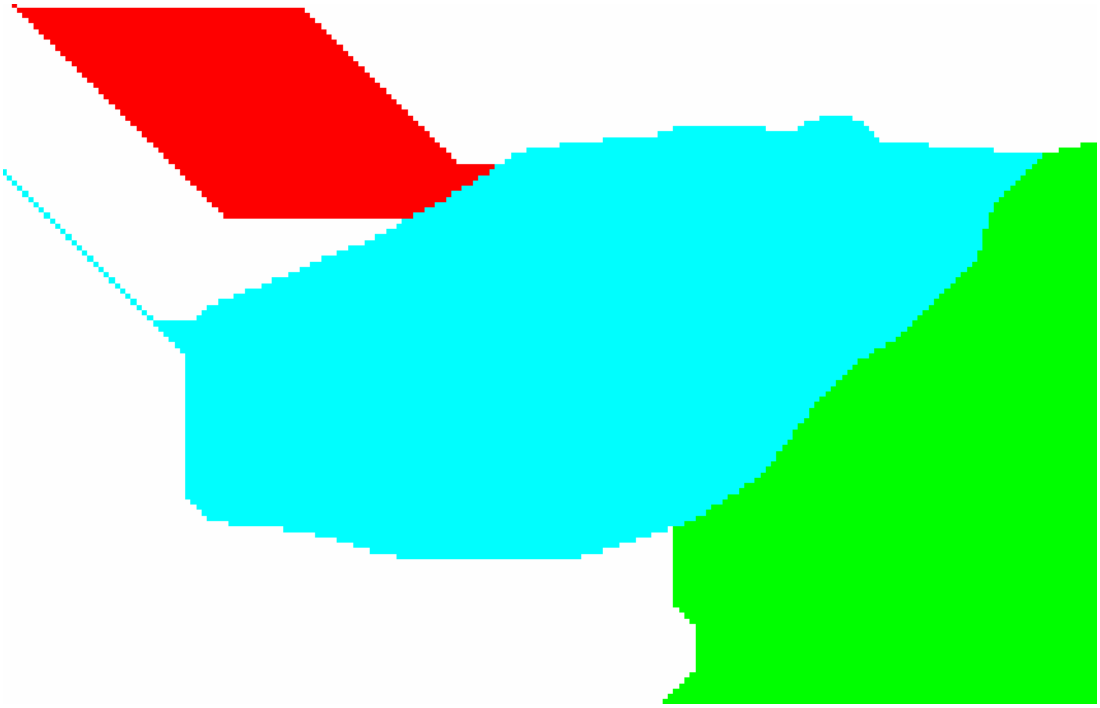


Figure 4.13 Catchment Grid

4.4.7 Catchment Polygons

Catchment Polygons are the vector representation of the catchment grid. Similar to the catchment grid a catchment polygon bounds the region which has a single drainage outlet.

Select Catchment Polygon from Raster Processing Dropdown menu, and select CatGrid as the input file, save the output file as CatPoly in the Raster Processing folder, then click on Run. If you right click on the layer labels in the left column, and go into the properties, in the symbology tab, change the legend type from Single Value to Unique Value, you will get Figure 4.15 showing all the delineated catchments in different colors.

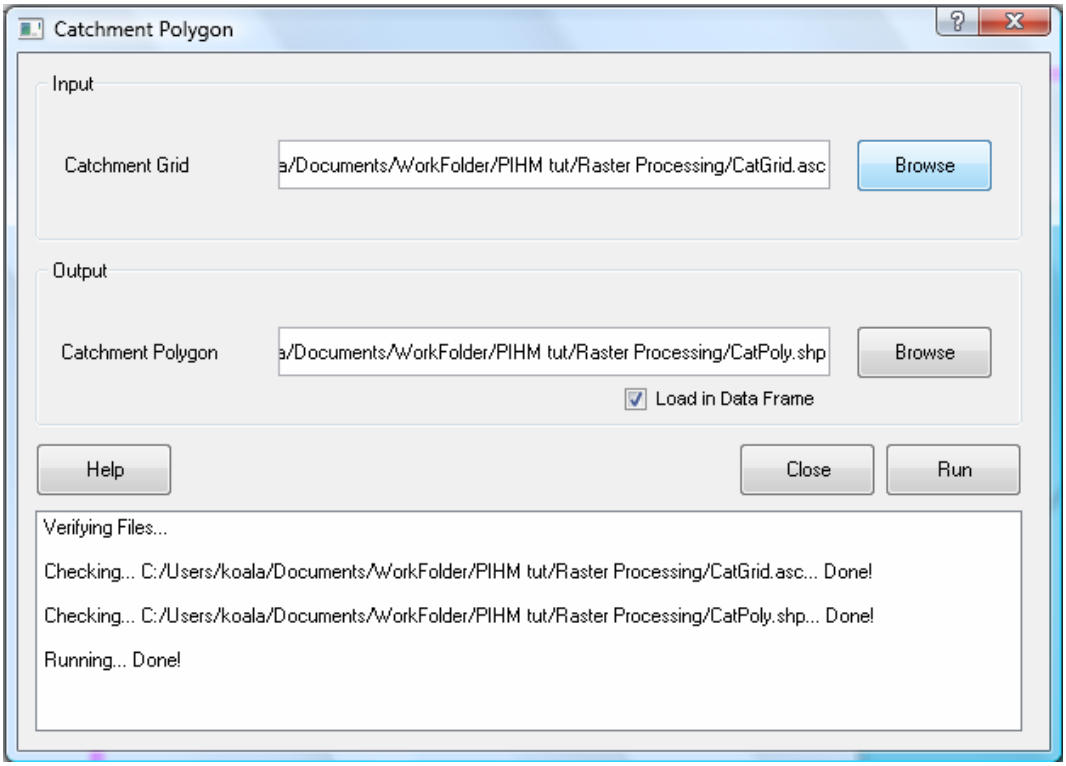


Figure 4.14 Catchment Polygon Dialog

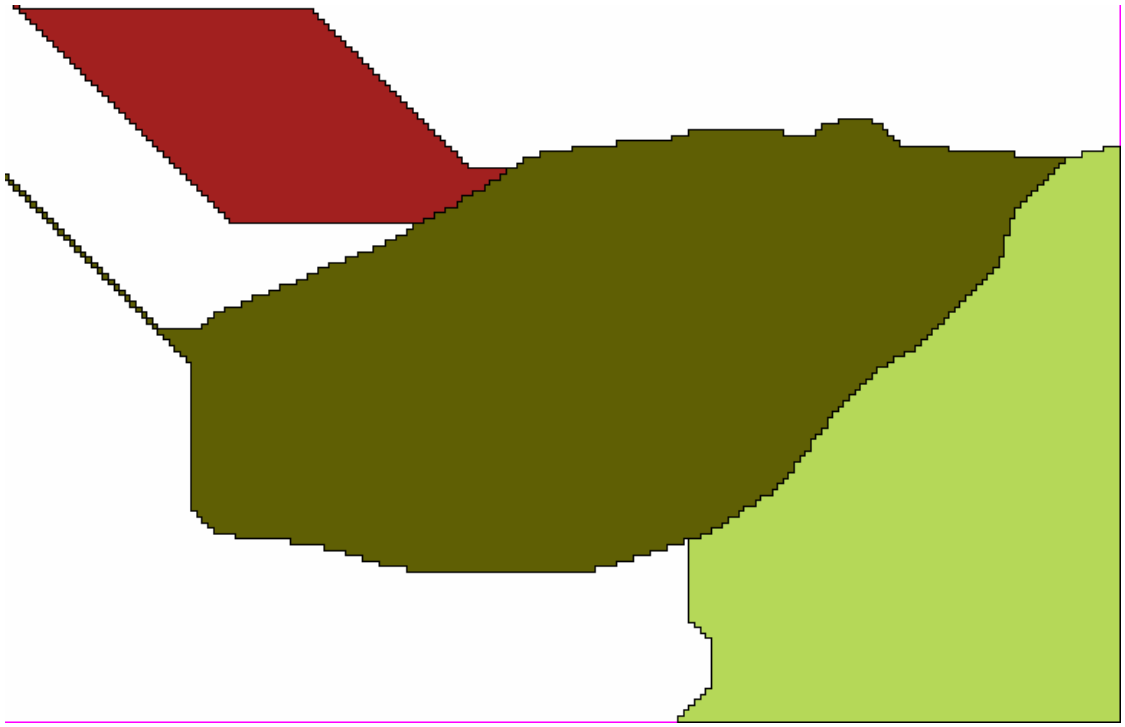


Figure 4.15 Catchment Polygons

4.4.8 Identify the Watershed

There are 3 catchments identified in the Raster Processing following the steps through 4.3.1 to 4.3.7, as shown in Figure 4.15. However, to investigate the shale Hills watershed, we are only interested in the biggest catchment in the middle. Thus in this step, we will delete the extra catchments, and leave the only one as shown in Figure 4.16.

Notice that, this step is done in ArcGIS using the Edit toolbar. PIHMgis is on the stage to integrate this function into it. If you are using a newer version, read the documents to see if it has been done.

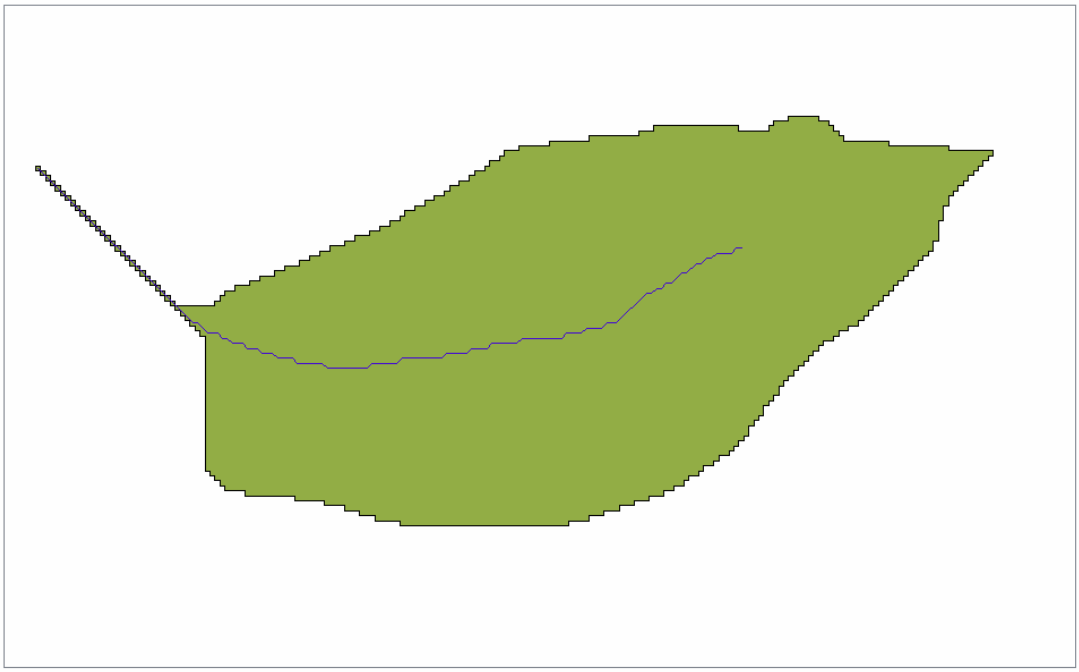


Figure 4.16 Shale Hills catchment

4.5 Vector Processing

Vector Processing consists of a set of operations which enables efficient discretization of the modeling domain. Stream polyline and catchment polygon generated earlier can be used primarily for this purpose. However, other hydrologic constraint such as soil coverage, land cover type coverage can also be incorporated. Eventually, it prepares a GIS layer which is used as input constraint for the purpose of domain decomposition of the domain. Before starting the process, copy the shapefiles of stream and catchement, to Vector Processing folder, so that files created in this step will be saved in the same folder.

4.5.1 Polygon To Line

The geo-data feature can exist as point, line or polygon objects. In order to merge all the data features together before it can be used by a domain decomposition code, the object properties for all the features should be same. Thus polygons are converted to polylines before they can be merged with already existing line features like river.

Select Polygon to Line from Vector Processing dropdown menu to start the dialog, click on the "+" button, and browse to the CatPoly.shp. The system will automatically name the output files as CatPoly_Polyline.shp as is shown in Figure 4.17. The boundary of the catchment is extracted from the catchment polygon as polylines.

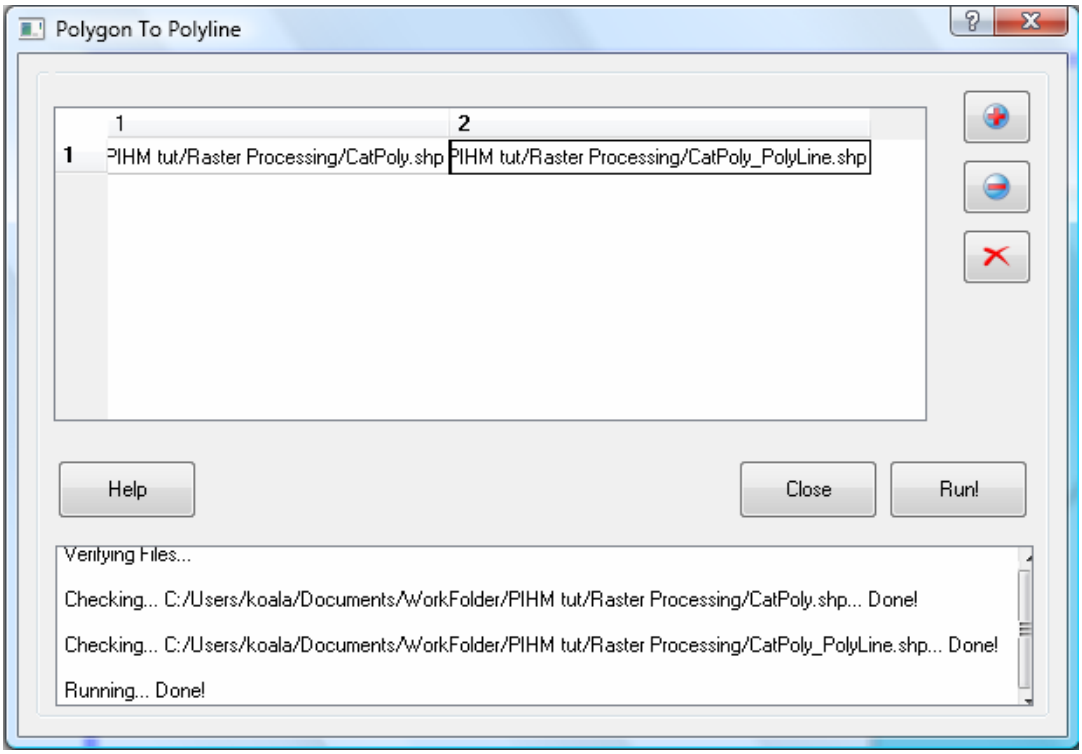


Figure 4.17 Polygon to Polyline Dialog

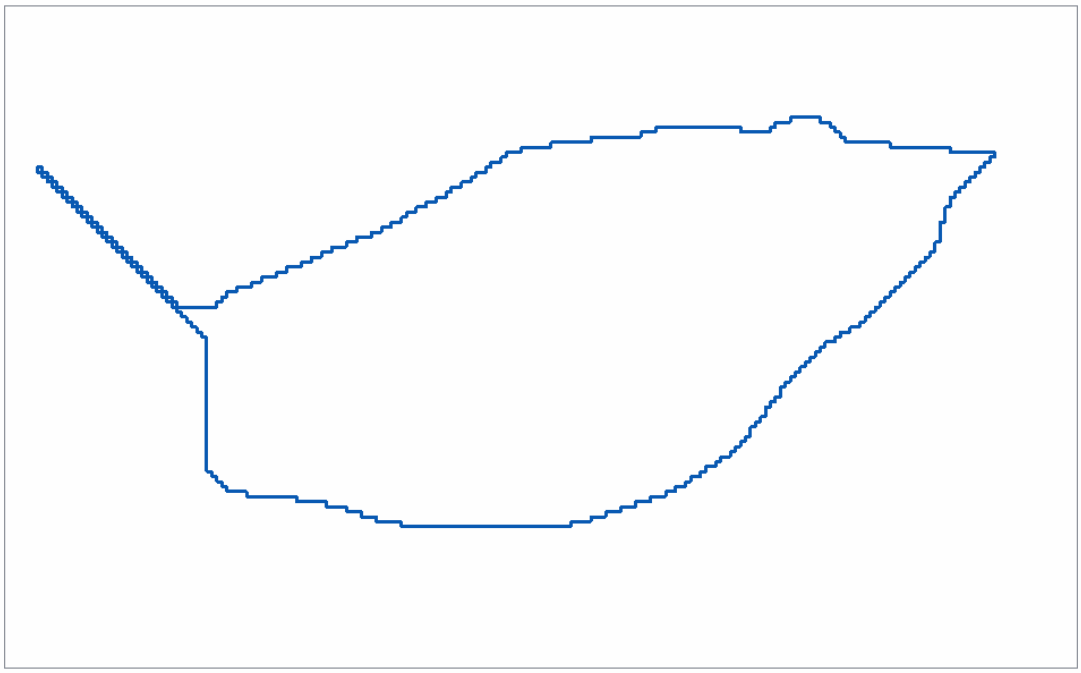
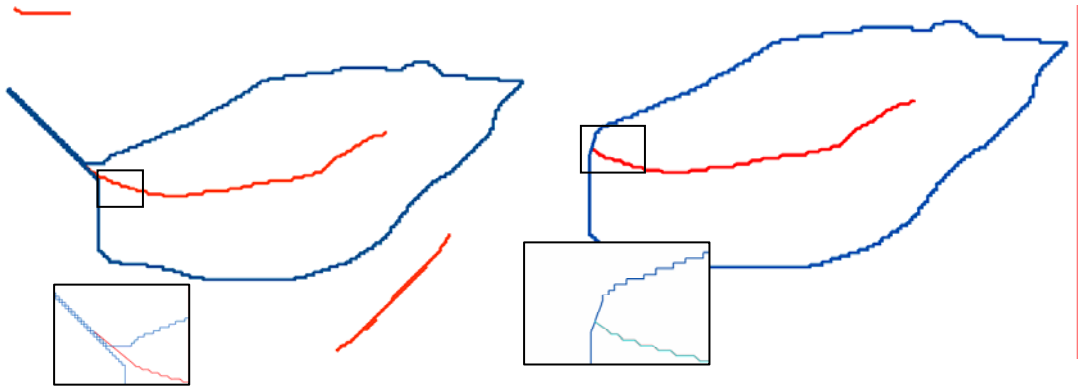


Figure 4.18 Polylines converted from the catchment polygon

4.5.2 Correction of the stream and the boundary.

In order to get a correct watershed boundary and the stream, the polyline shapefiles have to be corrected. In the stream file, streams that are not inside the catchment boundary should be deleted, and the “tail” in the watershed should be modified. More importantly, the end of the stream has to be snapped to the watershed boundary, so that a correct TIN can be generated in the following steps. This step is done in ArcGIS, check the documents to see if it’s available in newer version of PIHMgis.



(a) (b)

Figure 4.19 Stream and catchment boundary before fixed (a) before correction, (b) after correction.

4.5.3 Simplify Line

Simplify line is used to simplify a polyline by removing small fluctuations or extraneous bends from it while preserving its essential shape. This step becomes particularly crucial for quality and efficient domain decomposition as an unsimplified feature can have unnecessarily large number of nodes in it which in turn determine the number of triangulations generated. Needless to say, the larger the number of decomposed triangle elements is, the larger the computational requirement of the numerical model simulation will be.

Select Simplify Line from Vector Processing dropdown menu. Browse to the modified shapefile CatPoly_Polyline.shp and StrPoly.shp, set the tolerance to be 10. The output file is named CatPoly_Polyline_Simp.shp andStrPoly_Simp.shp" automatically, and saved in the Vector Processing folder. Figure 4.20 shows the dialog, and Figure 4.21 shows the simplified shapefile. If the file did not show in the display windows, select the Add a Vector Layer from Layers dropdown map to load them.

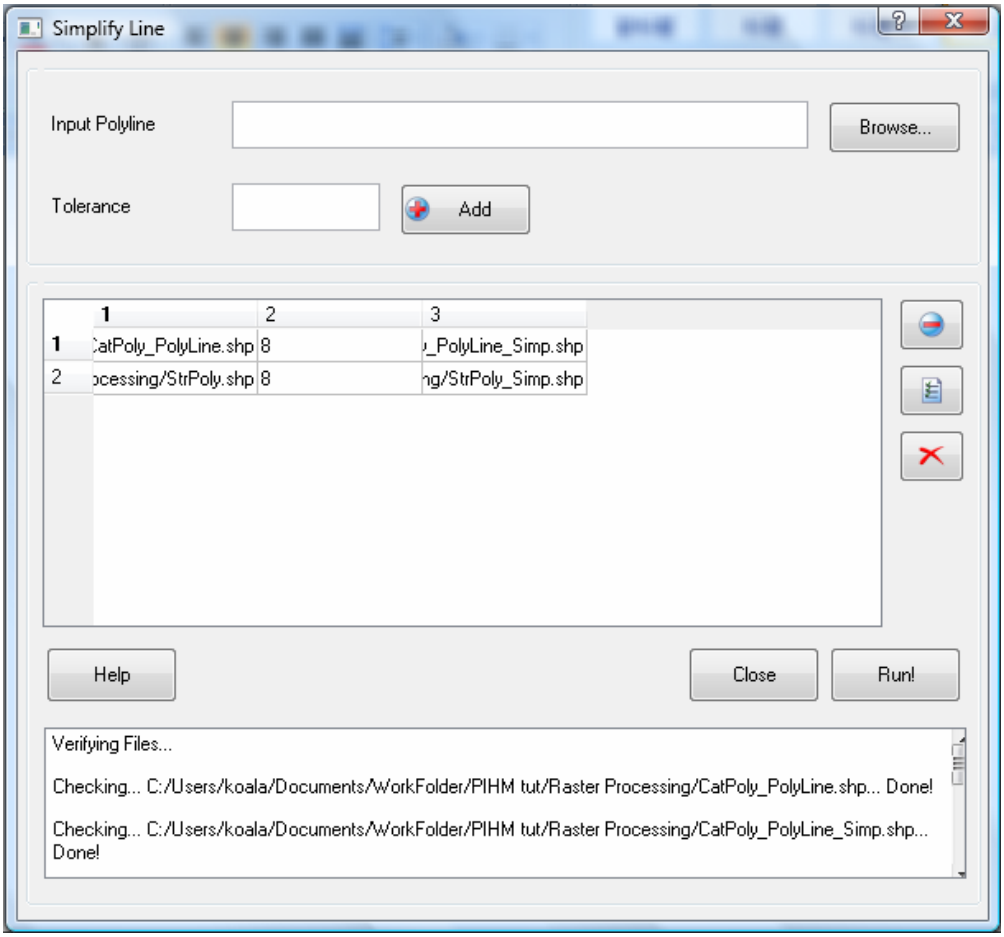


Figure 4.20 Simplify Line Dialog

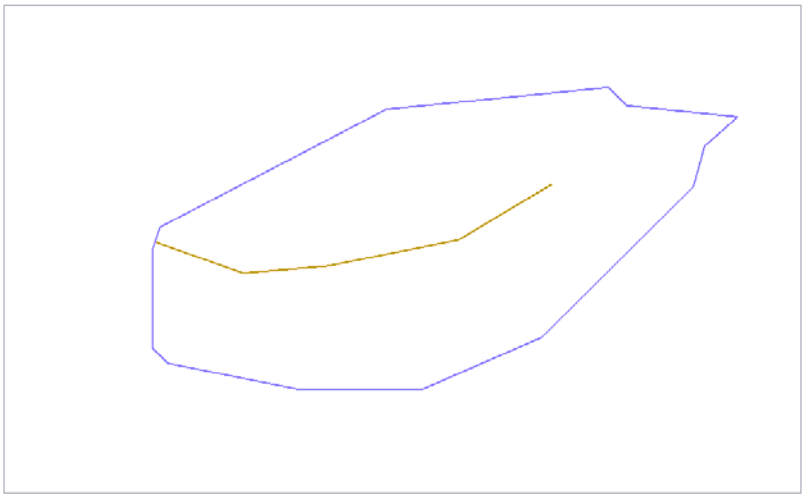


Figure 4.21 Simplified polylines

4.5.4 Split Line

Before merging all the features together it is necessary to have them all in one feature type. Split line splits polylines at each vertex. Therefore, this turns a single polyline feature into a multiple line feature depending upon the number of vertices present in the original polyline.

Select Split Line from Vector Processing dropdown menu, add the 2 simplified files created. StrPoly_Simp_Split.shp and CatPoly_Polyline_Simp_Split. .shp will be created in the folder.

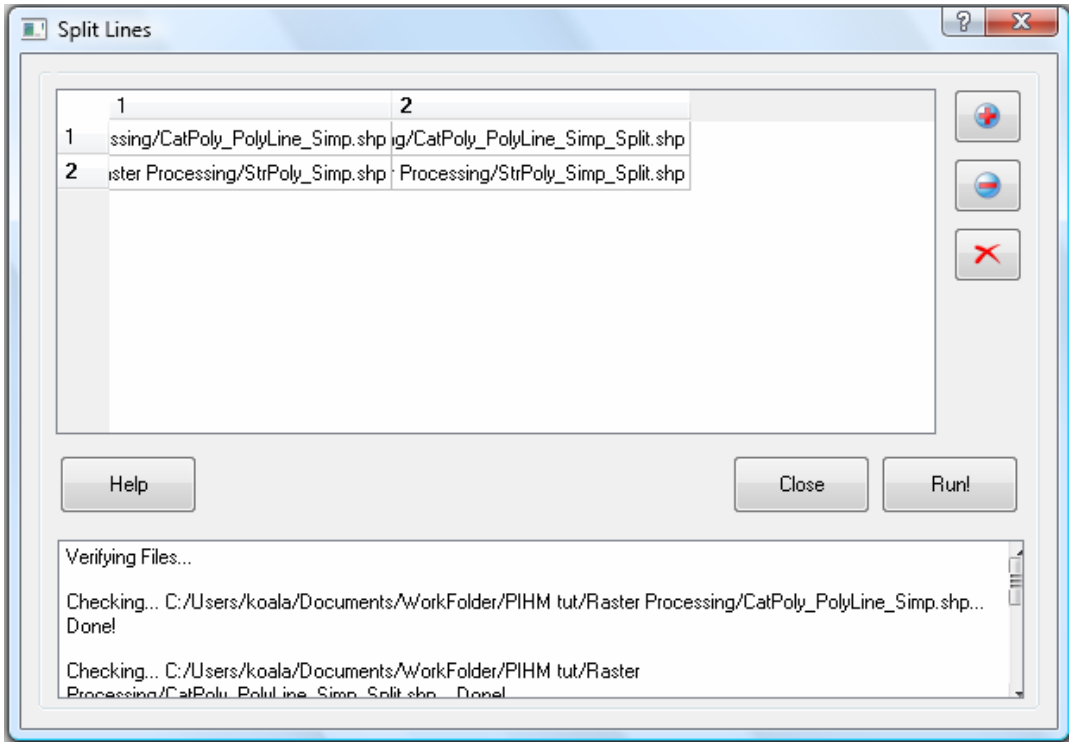


Figure 4.22 Split Lines dialog

4.5.5 Vector Merge

Vector Merge is the final step of Vector processing. It merges all the layers into one shape file. The merged shape file acts as constraints in domain decomposition process.

Select Vector Merge from Vector Processing dropdown menu, add the 2 files created in the Split Lines step. Save the output file as merge.shp in the Vector Processing folder. Cluster tolerance is a double type number. Any two nodes separated by the distance less than the cluster tolerance will be merged into one. Other constrain files like soil classification, lake, landcover type can also be added as constrains for domain decomposition and mesh generation, but make sure that the files are line features.

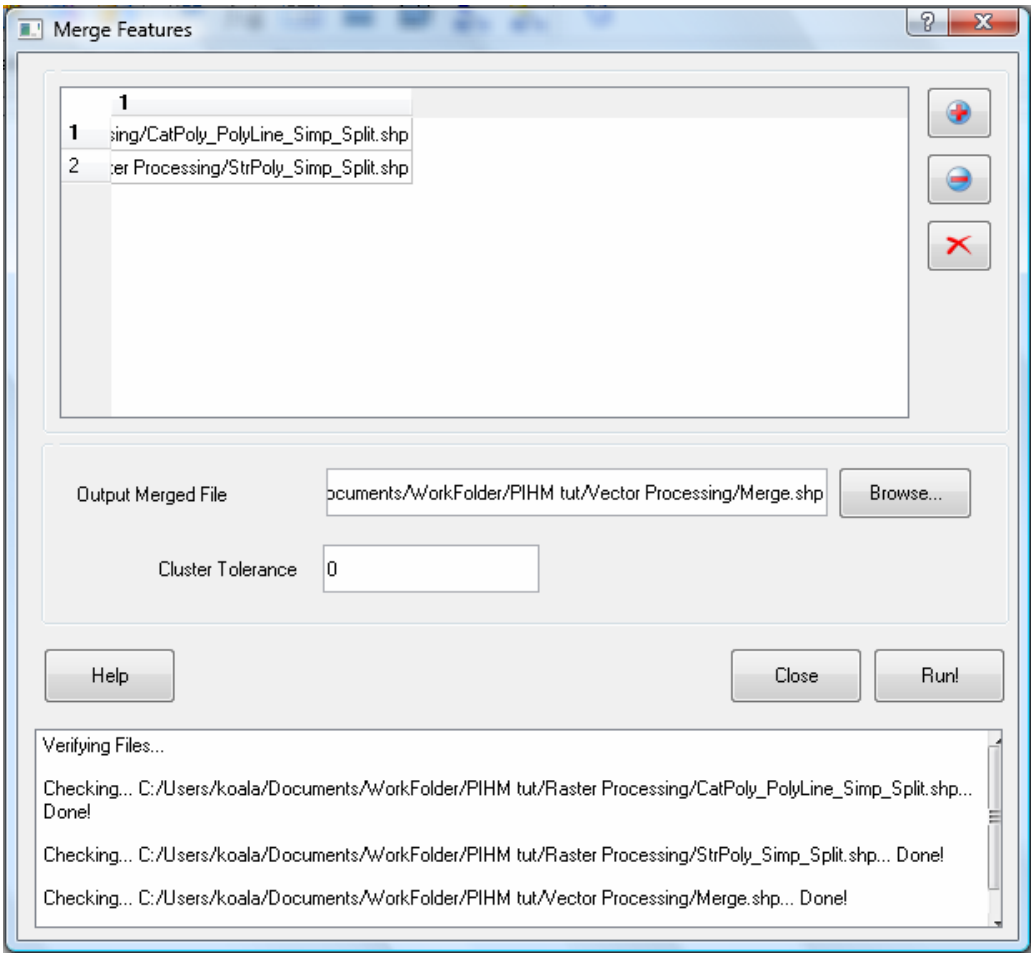


Figure 4.23 Vector Merge dialog

4.6 Domain Decomposition

Domain decomposition applies Delaunay Triangulation (Delaunay, 1934) to decompose the modeling domain into triangular irregular mesh or triangular irregular

network (TIN). A terrain can be better represented by an irregular mesh if all the critical terrain and hydrographic points are taken into account while performing domain decomposition. This may include watershed boundary, different types of contours (e.g. hypsometry, soil), stream network, hydraulic structures (e.g. dams, gages) for generating those points.

4.6.1 Read ShapeTopology

This step prepares a .poly file which acts as input for running the TRIANGLE in the next section. All the node and line information from the input shape file (vector merged shape file obtained in the earlier section) is transformed into the poly file.

A .poly file represents a PSLG, as well as some additional information. PSLG stands for Planar Straight Line Graph. By definition, a PSLG is just a list of vertices and segments. A .poly file can also contain information about holes and concavities, as well as regional attributes.

Select Read Shape Topology in the Domain Decomposition dropdown menu to start the dialog. Choose the merge file created in vector processing, name the output file as Sh, and save it in the Domain Decomposition folder. After running the process, a new file with the extension .poly will appear in the folder.

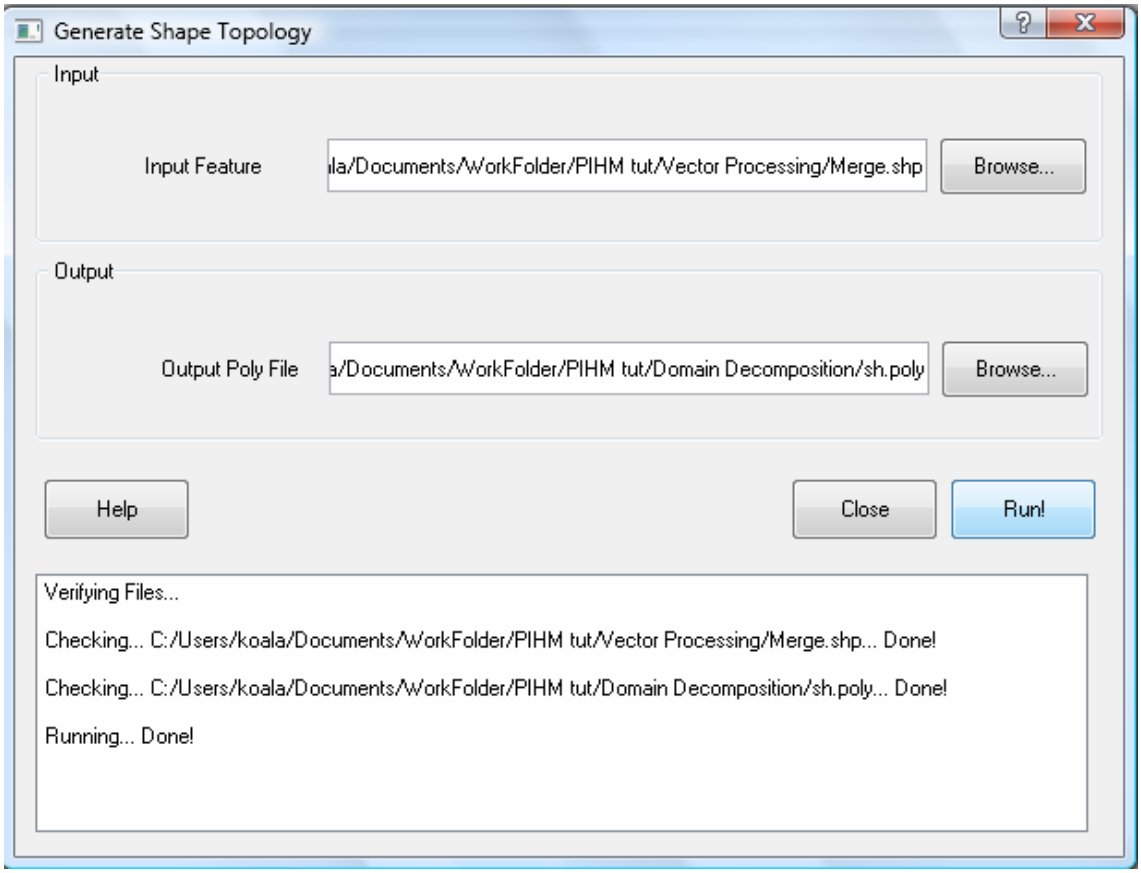


Figure 4.24 Generate Shape Topology dialog

4.6.2 Run TRIANGLE

'TRIANGLE' developed by Shewchuk (2001) is one of the efficient implementation of this algorithm which provides several flexibility and options to a user. It is also computationally efficient. It takes planner straight line graph (PSLG) as input. The algorithm works in such a way that it refines the Delaunay triangulation by inserting carefully placed vertices until the generated mesh meets a provided quality and size criterion.

Select Run Triangle from Domain Decomposition menu, choose the sh.poly, check the first 3 options, and the set the angle within 0~24. A maximum tolerance for the

area is set to be 200 in our case. If more constrains are needed in the process, click the question mark for information. Table is a list of options. A set of files named sh.1.ele, sh.1.node, sh.1.neigh, shavers.1.poly will be created in the Domain Decomposition folder.

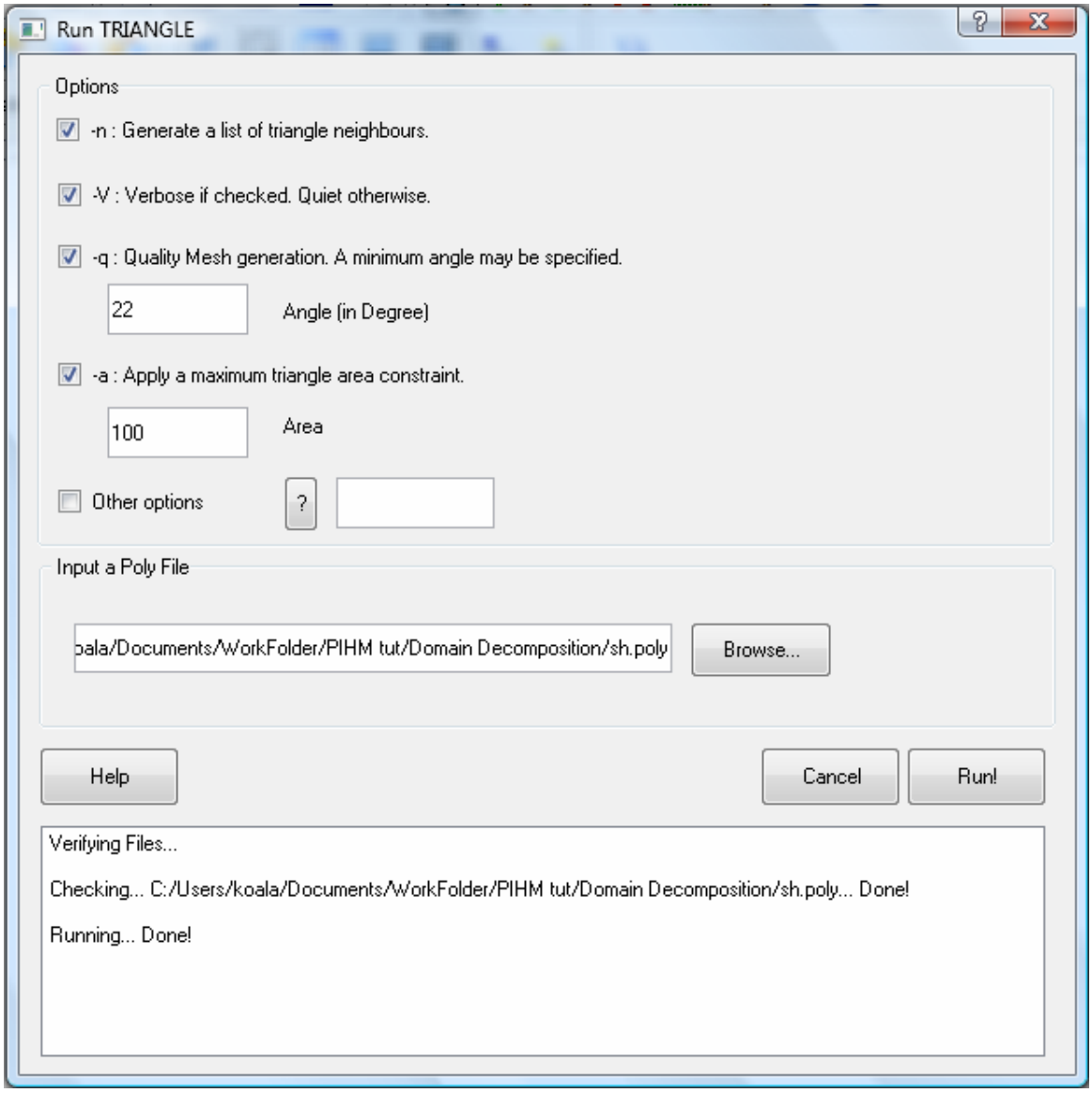


Figure 4.25 Run Triangular dialog

4.6.3 TIN Generation

TIN Generation read in the .ele and .node file and generates a polygon shapefile with all the triangle elements.

Select TIN Generation from Domain Decomposition dropdown menu, browse tothe sh.ele and sh.node files, and name the file sh.shp. A shapefile representing the TIN network of the Shale Hills watershed domain is created.

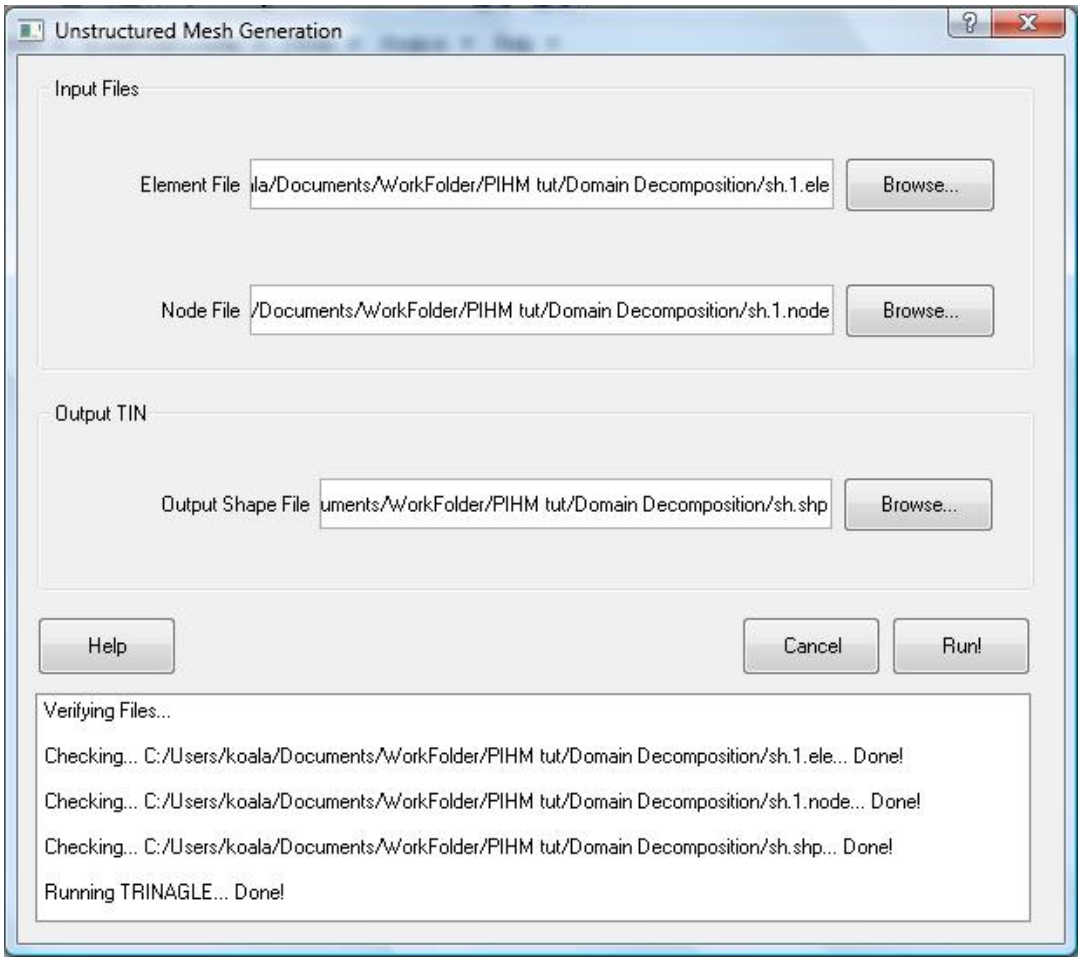


Figure 4.26 Unstructured Mesh Generation

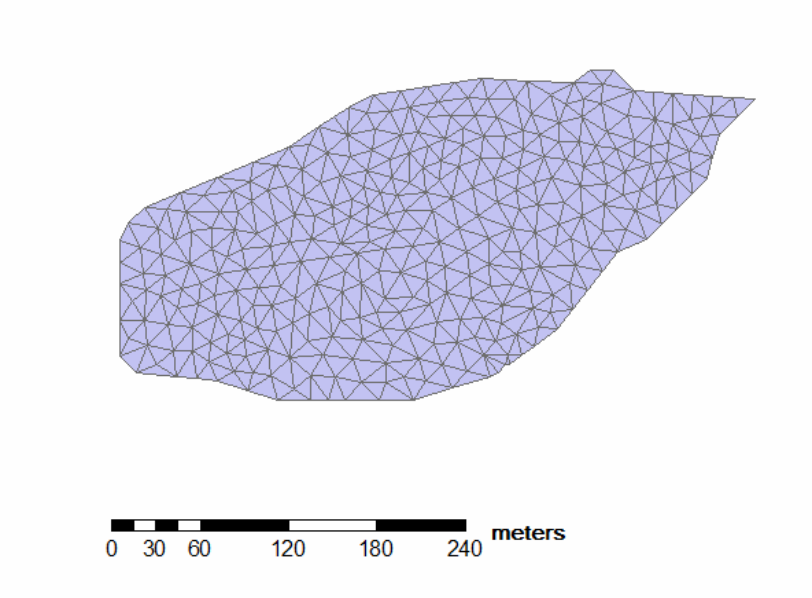


Figure 4.27 Shale Hills Triangular Irregular Network . The model has 598 cells and 332 nodes.

4.7 Data Model Loader

Data model loader reshapes the structure of GIS maps into a geo-database which is specifically designed according to the PIHM data model. The files prepared constitute spatial and relational attributes of the modeling domain. PIHM requires input in the form of 11 separate files: (1) mesh file; (2) att file; (3) soil file; (4) lc file; (5) riv file; (6) forc file; (7) ibc file; (8) para file; (9) calib file; and (10) geol; (11) init file. For more details of the files, check the file format documents of the pihm model.

In order to finish this step, grid files of soil, geology, land cover, precipitation, temperature, humidity, vapor pressure, melt factor, solar radiation, macropore distribution, source or sink distribution, etc will be needed. Get these files ready, and make sure that they are in the same spatial coordinate system, and the spatial unites are in meters.

4.7.1 Mesh File Generation

Select Mesh File Generation from Data Model loader Generation, select element file, node file and neighbor file created in last step. For the bedrock elevation, browse to the DEM folder and select the w001001.adf, for the bedrock elevation, browse to the Gridfiles-> bedrock elevation and select the w001001.adf. You can save your output file into Datamodel (a new folder) as shalehills.mesh.

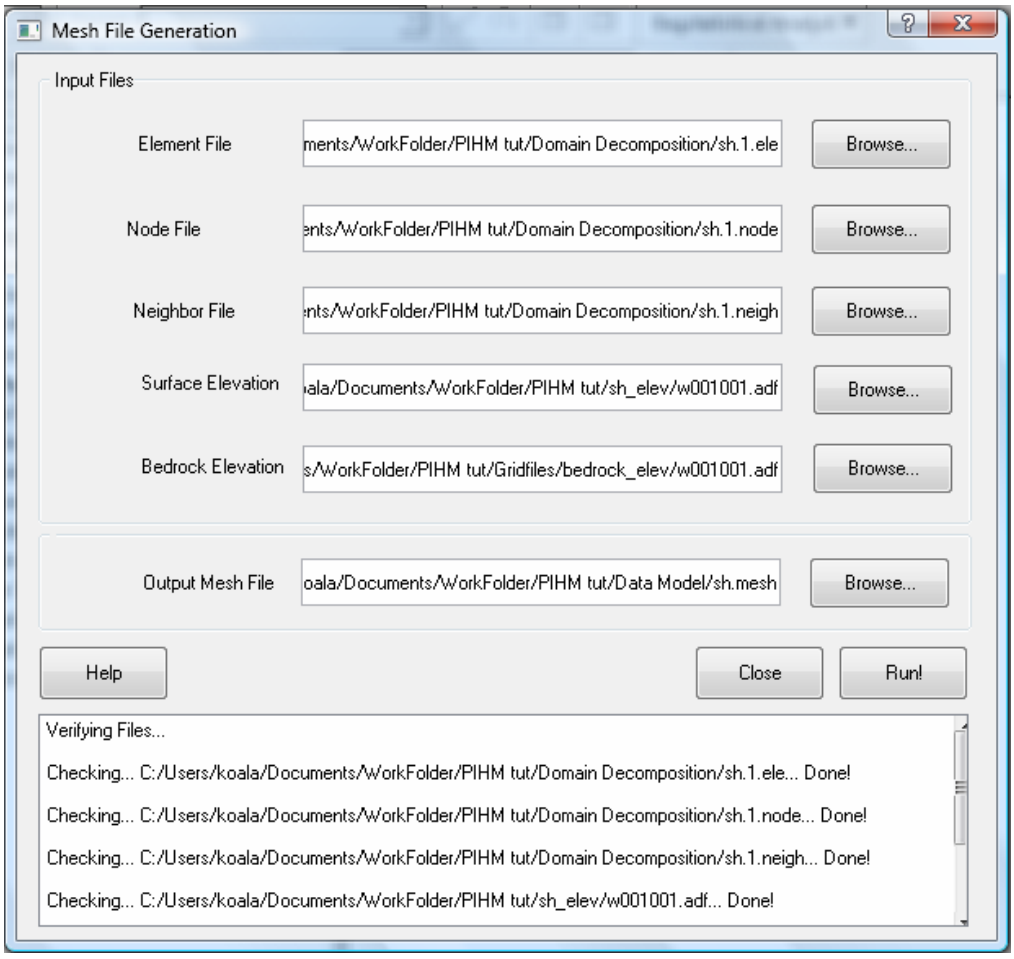
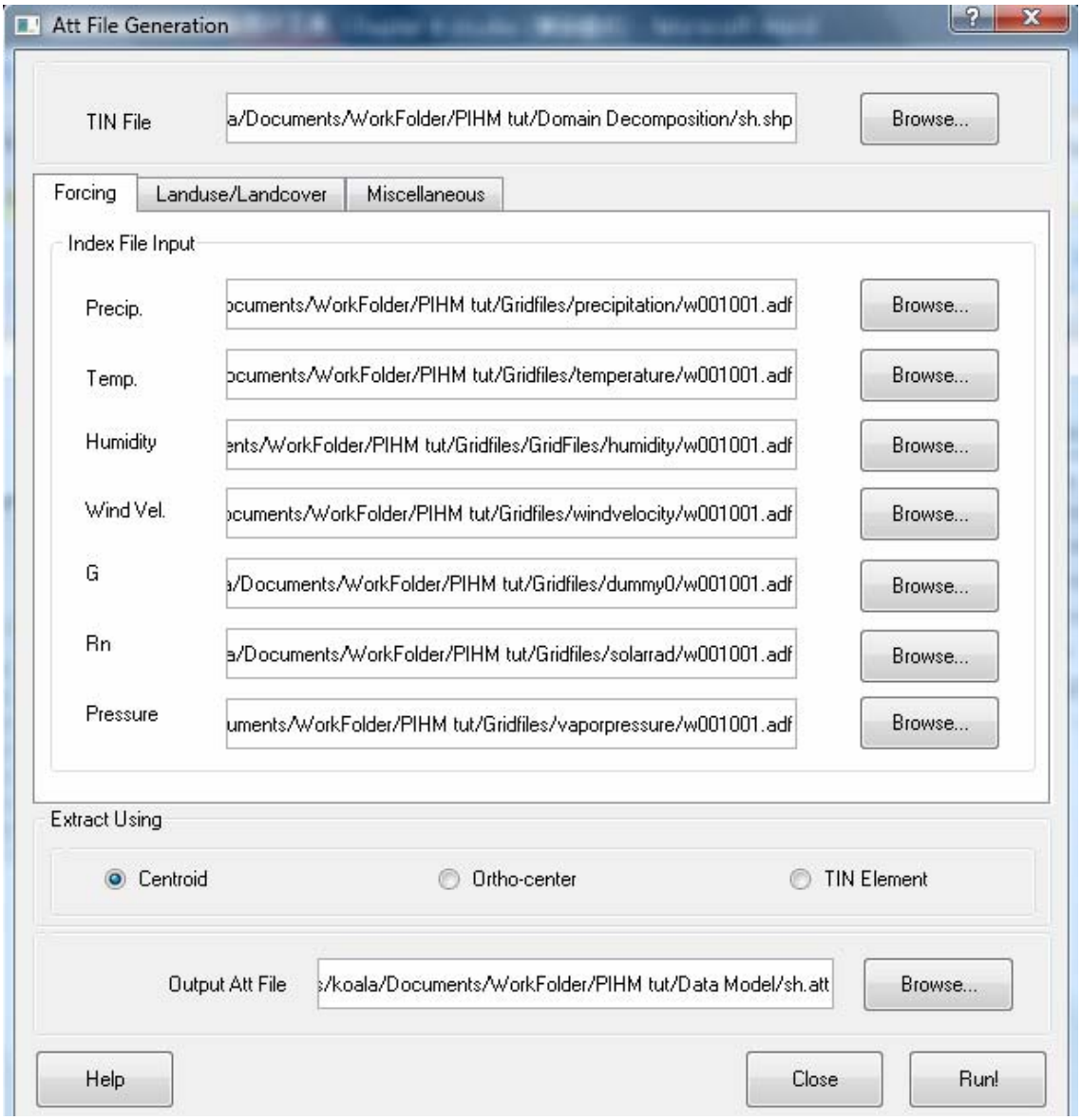


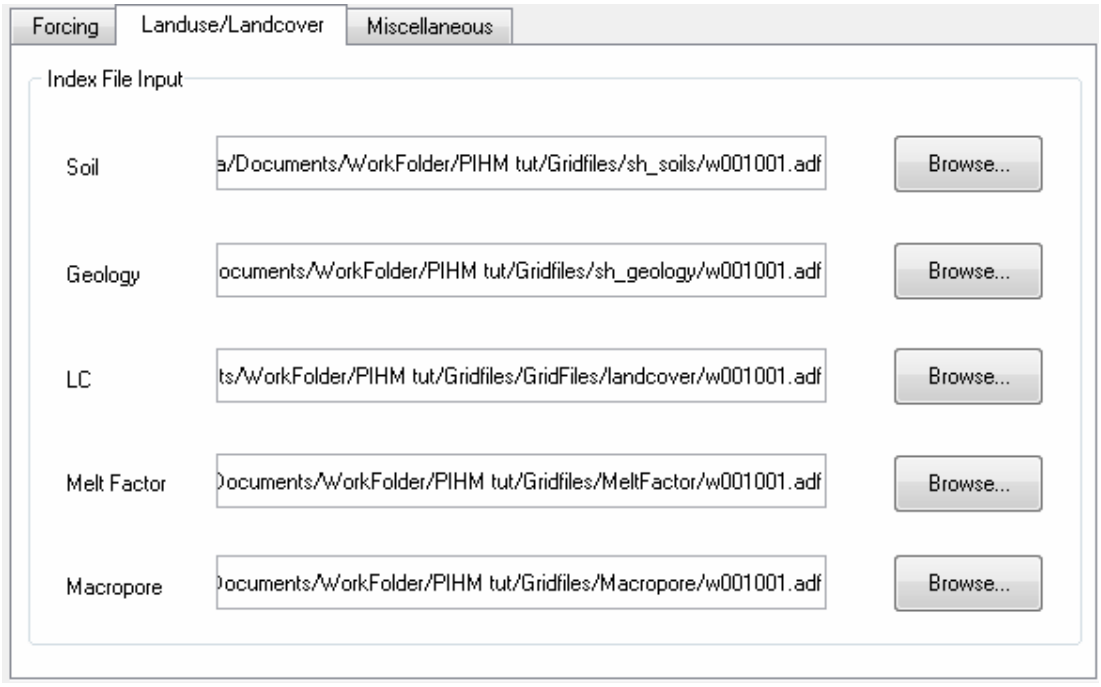
Figure 4.28 Mesh File Generation dialog

4.7.2 Attribute File Generation

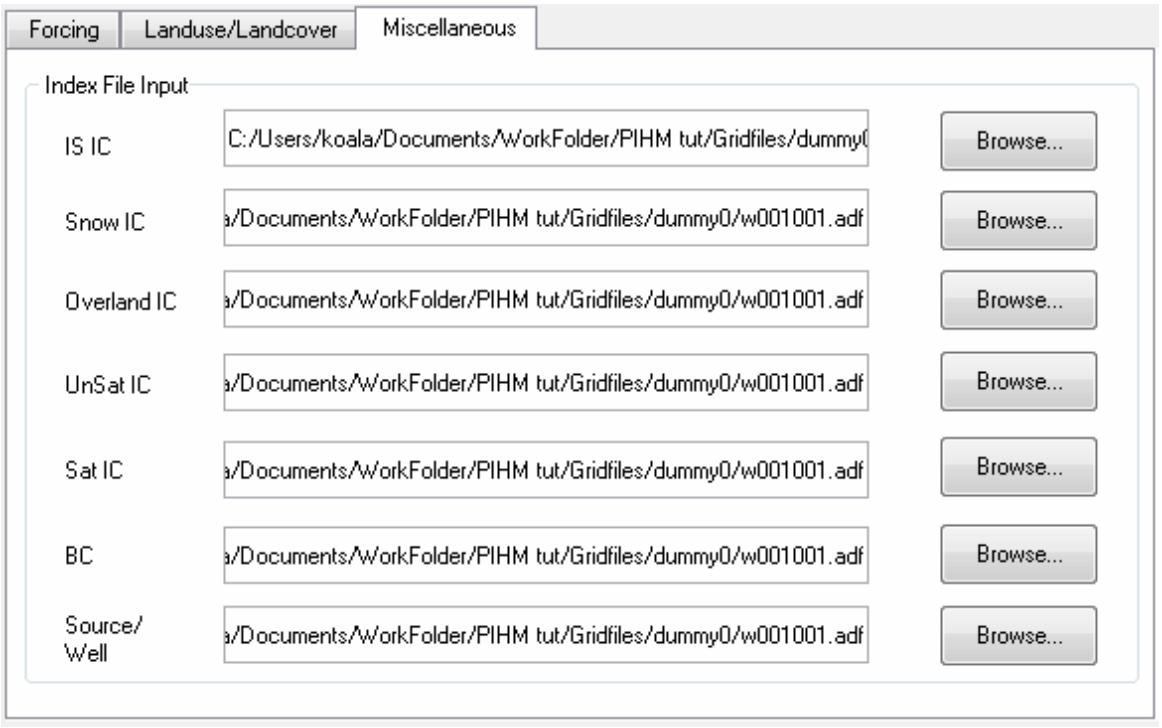
Chose the right file for each item in different tabs according to Figures 4.29 a, b and c



(a)



(b) .



(c)

Figure 4.29 Attribute file generation dialog

4.7.3 River File Generation

Browse to the split stream file as the river file input, and add the ele, node, neigh file into the dialog as shown in Figure 4.30. However, after generating the file, the .riv file need to be edited. Information for shape, material, boundary conditions, etc has to be added into the file. For more information, check the Input File Format for PIHM model.

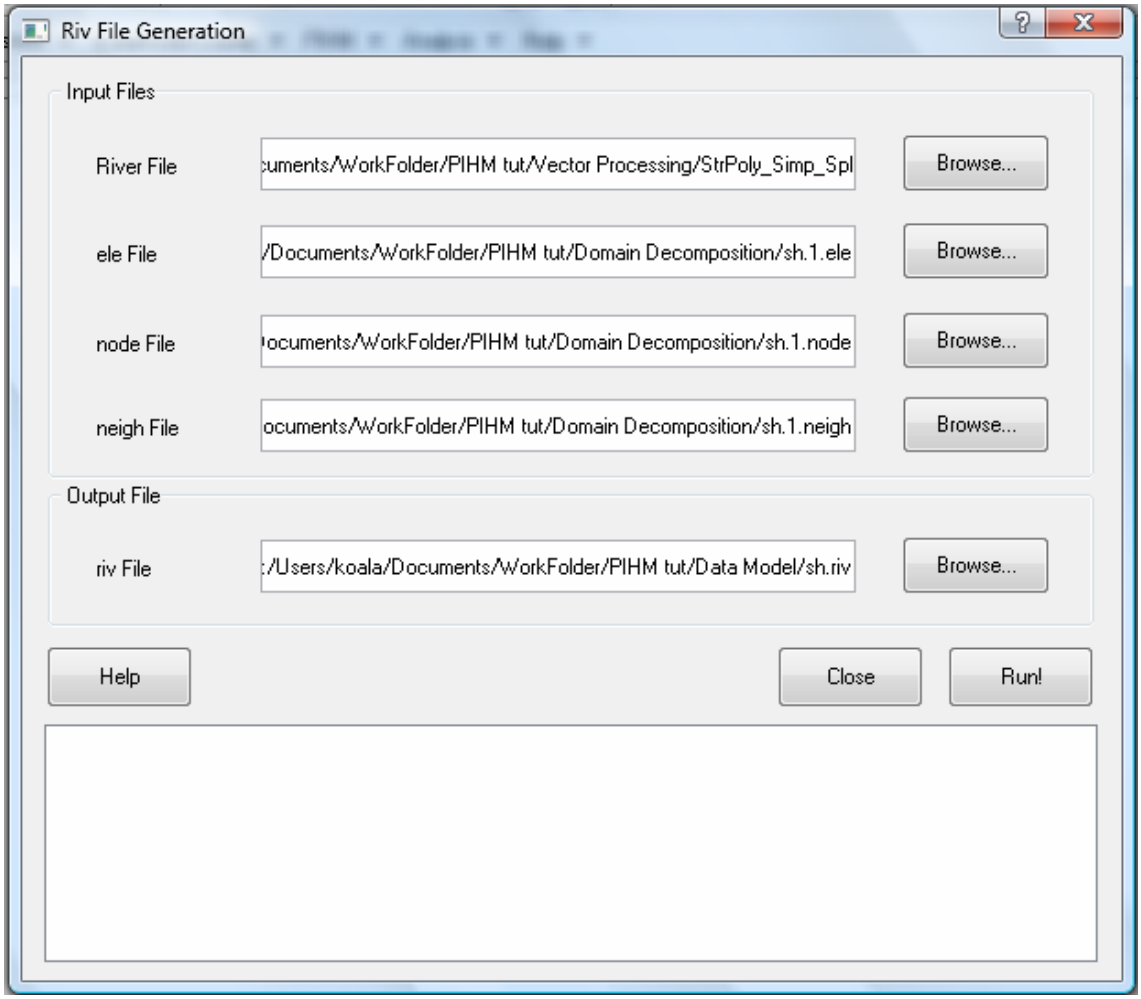
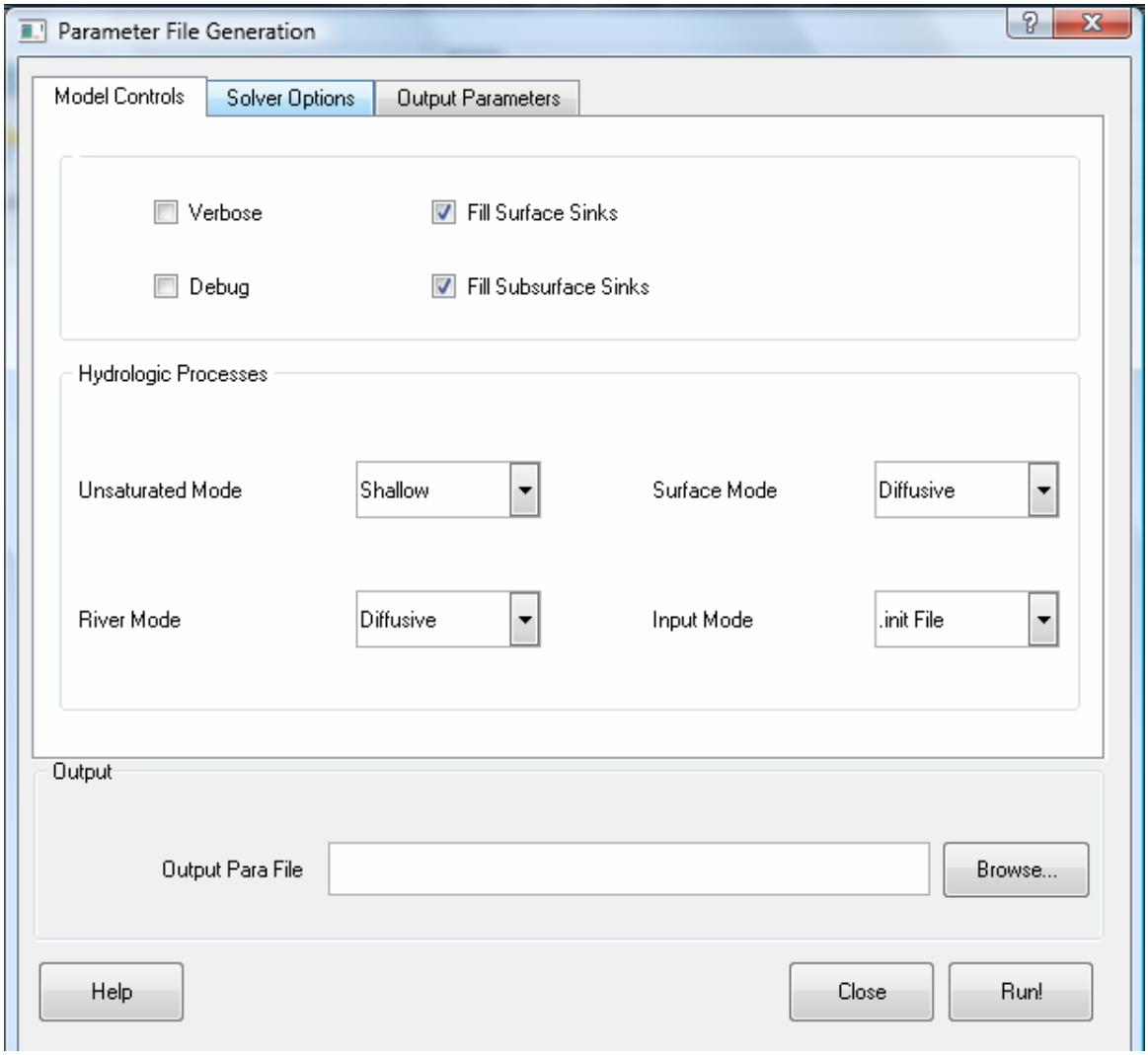


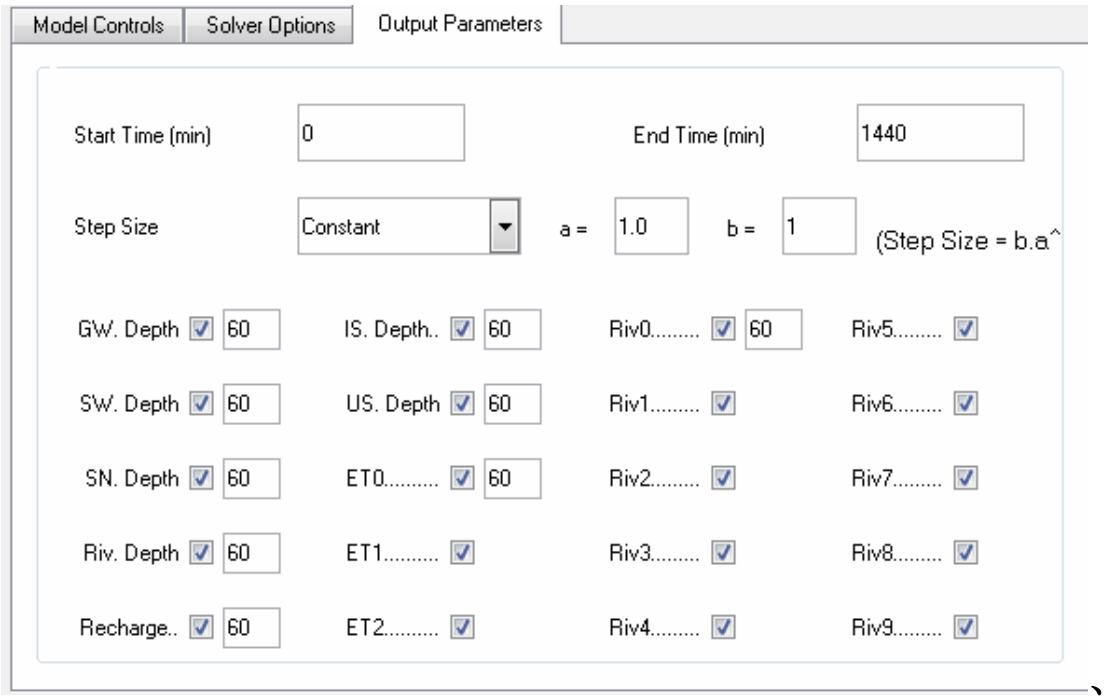
Figure 4.30 River File generation dialog

4.7.4 Parameter File Generation

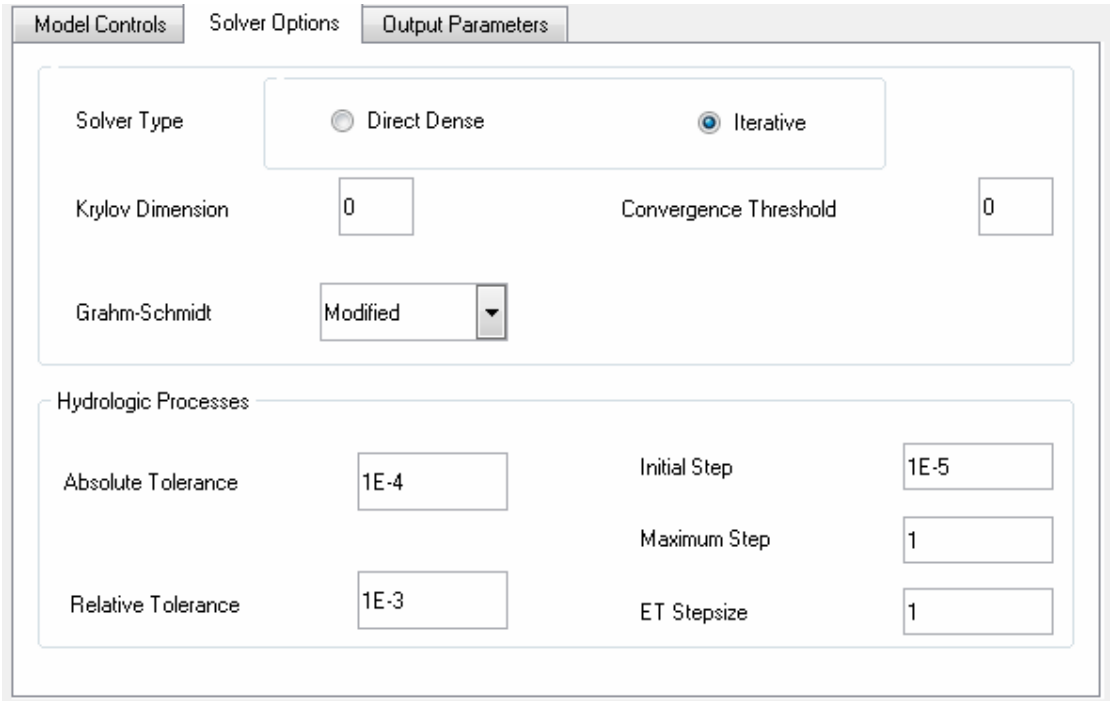
The para file controls the way the calculation is carried out, and it also give the time step of the output and the necessary files that needed to be printed out. At this point, use the default. Then click on Run.



(a)



(b)



(c)

Figure 4.31.Parameter File Generation dialog

4.7.5 Completion of data model for PIHM

In order to build up the data model for PIHM, soil, geology, landcover, forcing, calibration, initial condition, boundary condition files should also be built according to the input file format.

In Chapter 3, we have talked about the way to get the soil, geology, land cover, climate data for the model. Put these data into files will finish the data model for the Shale Hills watershed.

4.8 Discussion

In this chapter, we navigate through the process of getting data into model. After the data process, we have a set of files that for the PIHM model. However, model will not produce results instantly. Because of the inaccuracy of the measurement data, and limitations of the model, the A-priori data need to be calibrated to get a more realistic model

Chapter 5

Land Cover Classification with Topographic Wetness Index

5.1 Introduction

The distribution of the vegetation has the dependence on water, or more precisely, the soil moisture. In the other way around, the topographic distribution of soil moisture is one of the determining factors of the distribution of the vegetation distribution. In this chapter, a trial classification of the land cover in Shale Hills watershed is made based on the distribution of the soil moisture, which is quantified by the topographic wetness index.

5.2 Wetness Index.

The topographic wetness index which is developed by Beven and Kirkby (1979), is supposed to correlate to soil moisture few days after rainfall event, when all gravitational water is drained. It is defined as

$$wi = \ln(\frac{A_s}{\tan B}),$$

where A_s is the upslope contributing area (m^2), and B is the slope of the topography.

5.2 Shale Hills watershed

The climate in Shale Hills watershed represents a transition between the Midwestern continental climate which is relatively dry and the eastern coastal climate which is more humid (Lynch, 1976). Experiments were designed to record trees, in Shale Hills watershed, with diameters of its trunk larger than 10cm by the eco-hydrological research group from Penn State. 22 species are recorded in this area, which are red maple, sugar maple, yellow birch, American hickory, bitternut hickory, pignut hickory, shagbark hickory, mocker nut hickory, beech, white ash, yellow-poplar, cucumber tree, black gum, table mountain, eastern white pine, Virginia pine, black cherry, white oak, chestnut oak, red oak, black oak, American basswood, eastern hemlock. The records have been classified as evergreen and deciduous and is displayed in Figure 5.1.

5.3 Wetness Index Calculation and the Land Cover Classification

The calculation of wetness index is performed in ArcGIS using the raster calculator. D8 algorithm in Qgis is applied to calculate the flow direction grid, from which created the flow accumulation grid. The spatial distribution of the wetness index is displayed in Figure 5.2.

To make statistics, the species map is overlaid with the wetness index map in Figure 5.3. In this map, the wetness index is grouped in 4 categories, and the statistics for the evergreen and deciduous trees distribution is listed in Table 5.1. The names of the classes are taken from the LDAS (Land Data Assimilation Systems) data base.

	Range of	Evergreen	Deciduous	Ratio	Class
1	3-5	22	246	1:11.2	Deciduous
2	5-7	155	988	1:6.4	Deciduous Needleleaf
3	7-11	78	168	1:2.15	Evergreen
4	11-14.3	13	17	1: 1.3	Water

Table 5.1 Classification of the land cover type based on the wetness index

The soil properties that is needed in the model: maximum LAI, minimum stomatal resistances, reference stomatal resistance, albedo, vegetation fraction, Manning's Roughness. According to the LDAS database online:

http://ldas.gsfc.nasa.gov/LDAS8th/MAPPED.VEG/web.veg.table.html,

the parameters that we picked from the database are as follows.

	Minimum Stomatal Resistance day/m)	Albedo	Maximum LAI	Vegetation Fraction	Reference Stomatal Resistence	manning's coefficient ($daym^{-1/3}$)
Evergreen	0.0018808	0.069	10.76	0.8	6301077	8.10e-7
Deciduous	0.0020255	0.092	10.76	0.8	6221002	8.10e-7
Deciduous Needleaf	0.0019097	0.107295	5.117	0.9	8640000	8.10e-7
Water	0.00202546	0.135	0	0	13477995	4.05e-7

Table 5.2 Land cover properties

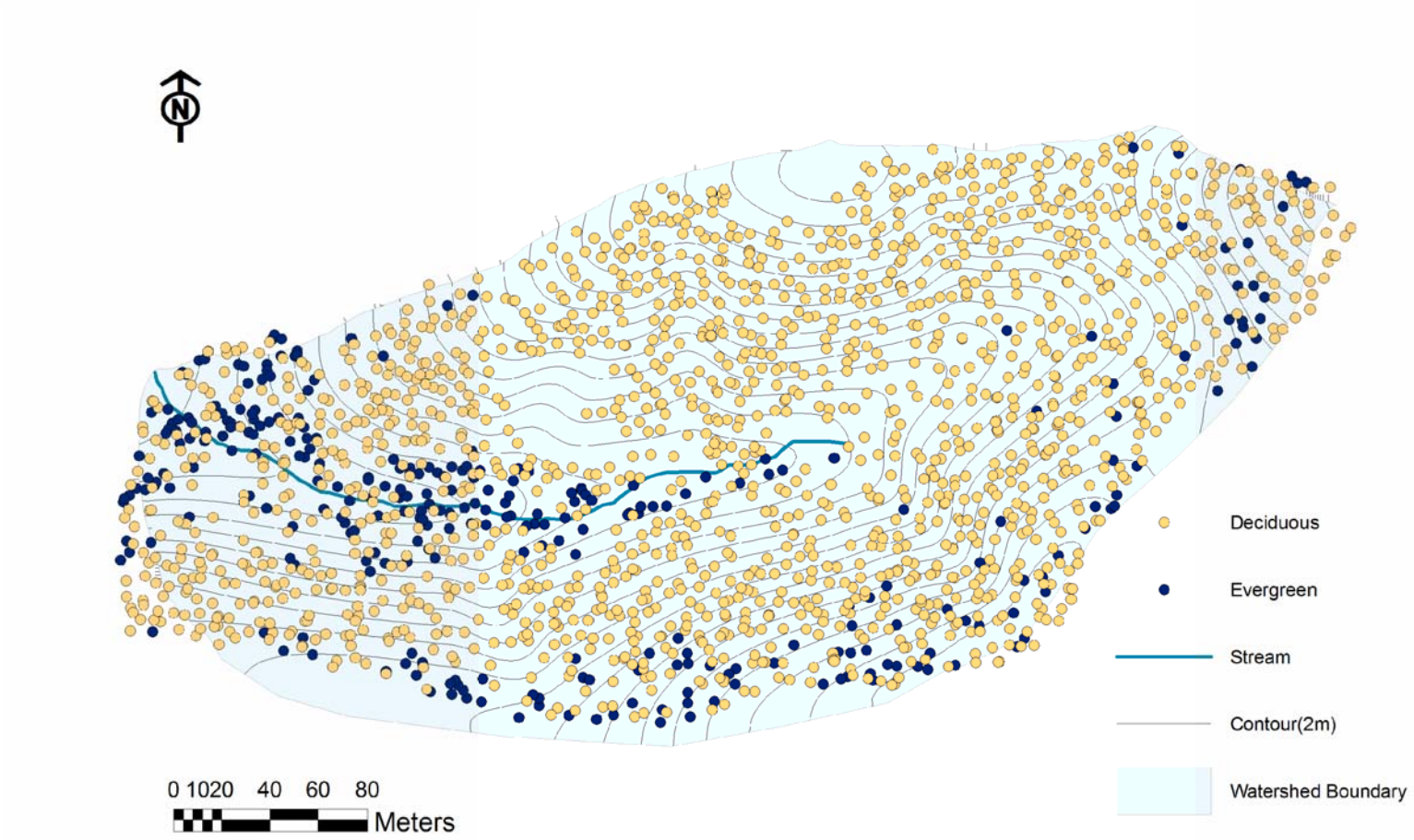


Figure 5.1 Distribution of Evergreen and Deciduous Trees

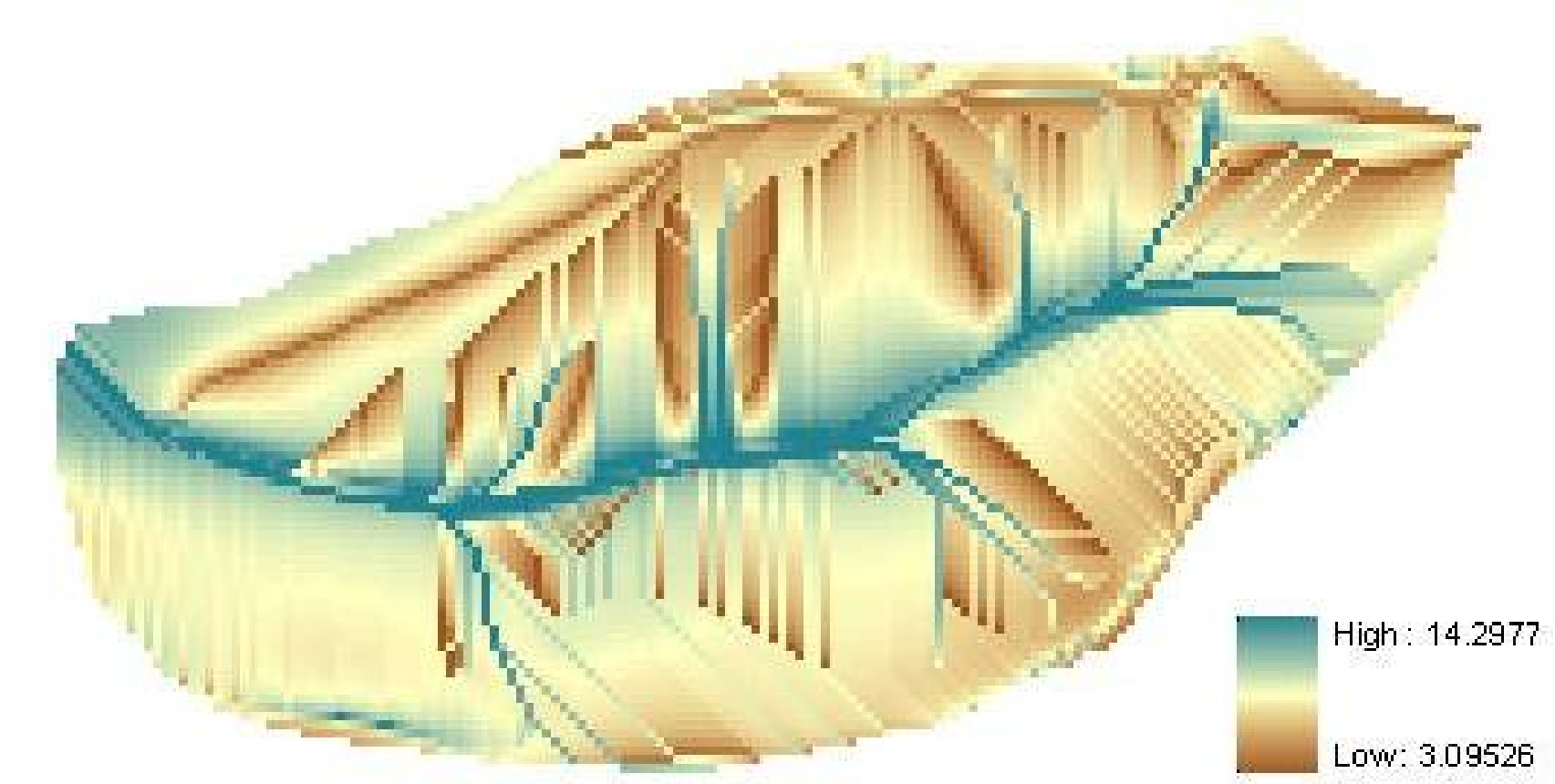


Figure 5.2 Map of Topographic Wetness Index for Shale Hills watershed

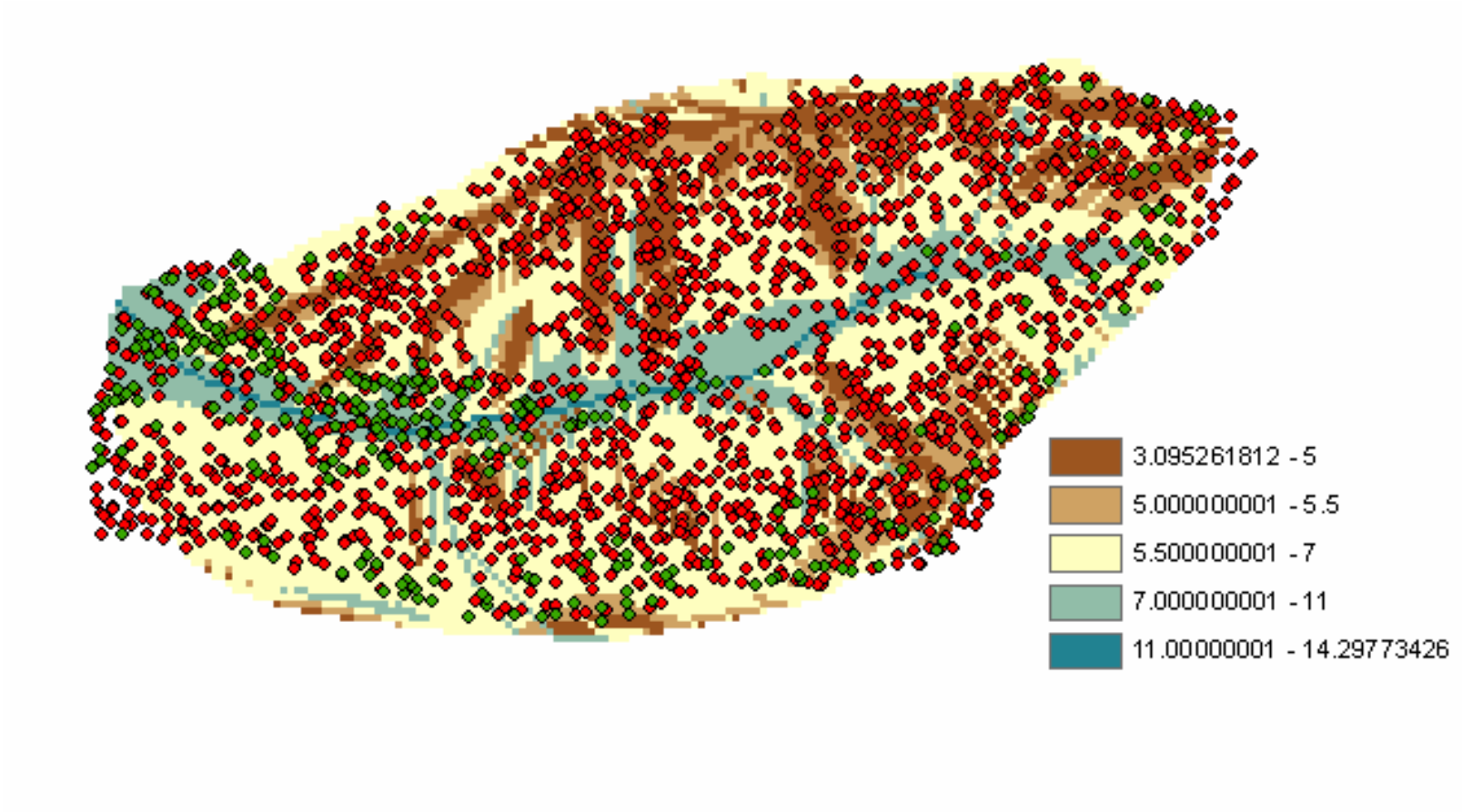


Figure 5.3 Classified Wetness Index with Distribution of Trees

Chapter 6

Results and Discussion

6.1 Model Domain and Experimental data

In 1970s, a spray network which has a precise control of the irrigation rate under the forest canopy was installed on the Shale Hills watershed. 40 piezometers, 40 neutron access tubes monitoring the soil moisture and 4 weirs measuring the stream flow along the river was also installed on the watershed. From July to September, 1974, 9 events of artificial rainfall at the rate of 0.64cm/hr (6hrs) were applied to the watershed (Lynch, 1976). Meanwhile, there were also natural rainfall events during this period. Figure 6.1 displayed the measured precipitation and runoff data at the outlet of the watershed during the experimental period. Since the experiment was carried out in late summer through autumn, which is considered a very dry period of the year. A window of 31 days, from August,1 to August 31 was selected in the simulation to minimize the influence of the natural rainfall. The experiment was designed to test the influence of antecedent soil

moisture on the storm flow, and how it changes the timing and the volume of the storm flow. The experiment is chosen here as a test case for our model, to calibrate the parameters in the model, and to see how the model will response during a series of irrigation events.

The domain was decomposed into 598 triangular elements with 332 nodes, while the stream is divided into 18 segments, which includes the ephemeral reaches upstream. The river is set to be 1m deep and 1.5m wide as rectangular elements with the hydraulic roughness as 0.34 min $m^{-1/3}$. No flow boundary condition is assumed for the groundwater and surface water, and zero depth gradient boundary condition is set at the outlet of the watershed. A prior data for the soil, geology, land cover, and forcing have been discussed in Chapter 3.

6.2 Simulation of Stream Flow

The simulation showed good prediction for the first five peaks and the base flow as shown in Figure 6.3. But the predicted value is lower than the measured data in the 6^{th} event. Each event shows good timing for the starting period of the peak, but not so good for the relaxation period in the last 2 events. As is known, the flow in the stream is coming from the surface flow and subsurface flow, and an interesting question would be what makes the main contribution to the peak of the storm flow. Figure 6.4(a) shows a segment of the stream in our model, and how the water would get into the segment. F_2 and F_3 are flows from overland left and right bank, while F_4, F_5 and F_6 are subsurface

flows that come from left and right bank and the river bed respectively. Four points along the river, as is marked in Figure 6.4(b) are selected, to compare the contributions from subsurface flow and surface flow. It is clear that, from the Figures 6.5, the surface flow and subsurface flow effect sequentially on the storm water peak, the surface flow contributes to the faster peak in the storm water, but it quickly disappears, while the subsurface flows contribute to the slower peak. It also showed that the subsurface contribution has an increasing trend of its peak, while the surface contribution kept almost similar magnitude throughout the six events. From the results, we can assume that the non-linear characteristics of the watershed response to the input of rainfall events are mainly due to the subsurface flow rather than the surface flow. From the results we can also tell that the upstream ephemeral channel segments have lose and period during rainfall and dry period.

6.3 Simulation of Soil Moisture

The soil column was divided into saturated zone and unsaturated zone in our model. The unsaturated zone soil moisture content is determined by the difference of the infiltration and the recharge. The saturated zone is represented by a 2D model, which has the vertical recharge/discharge and the horizontal groundwater flow to neighboring elements and to stream if near a stream segment.

To explicit the problem, the 3rd event was chosen to show the time transition of the groundwater head, the unsaturated zone soil moisture content and the recharge in Figure 6.7. During the event, the averaged unsaturated zone soil moisture content reaches its

peak first, and then the recharge peak follows, the groundwater head reaches its peak as a latter time, and finally the runoff at the watershed outlet. 4 moments are chosen in Figure 6.8 to show the time transition of the groundwater head spatial distribution. During the event, the area that holds high groundwater head does not expand too much corresponding to the rain. It only clustered in the riparian areas. And at the end of the event, the area even shrinks to the downstream, leaving an even drier initial condition for the 4^{th} event. This is interesting, because in Figure 6.2, from both simulated data and experimental data, the 4^{th} peak is lower than the 3^{rd} peak, which explains the change for groundwater head.

Qu and Duffy (2007) has also investigated experimental data for the unsaturated zone and saturated zone soil moisture contents, and presented a relationship as in Figure 6.9(a). 5 points in the riparian area is also picked in Figure 6.9 (b), to compare the relationship for them in the model. In Figure 6.10, the soil moisture contents does show a competing relationship especially when the groundwater head is high as in II and V, and when the groundwater table is low, the model shows the recharge process, in which the unsaturated zone soil moisture grow first, and then follows the saturated zone soil moisture.

Figure 6.11 shows the averaged recharge rate of the most intense recharging period of 10 hours of each event. It is interesting to find that the 6 events showed quite different patterns in both the distribution and the magnitude of the recharging rate. There is a great change between the 4^{th} event and the first 3 events, and the 5^{th} and 6^{th} events seems to be an decayed version of the 4^{th} events. This also enhanced the previous groundwater distribution at the beginning and the end of the 3^{rd} event. It is assumed that, during the first 3 events, the soil moisture has enjoyed a redistribution, so that the pattern begin to retain relatively identical afterwards.

Figure 6.12 is another map showing the time averaged recharge rate map of the 31 days. The most intense recharge rate areas cluster on the valley floor and the east end of the watershed.

6.4 Simulation of Evaporation and Transpiration

Figure 6.13 and Figure 6.14 are the time averaged map of evaporation and transpiration for 31 days respectively. They both show higher values in the riparian area, while at the other areas there seems to be a competing relationship. While the higher evaporation rates reside in the swales, the higher transpiration rates preferred the ridges instead. Figure 6.15 shows the multiple vertical fluxes including the transpiration and evaporation. They are compared at 2 locations as plotted in Figure 6.12, one on the hillslope, the other on the valley floor near the river. The comparison shows that the evaporation from the ground is highly correlated to the surface overland flow. However, on the hillslope, the evaporation is completely depend on the surface water, while on the valley floor, it still kept a higher rate although the surface water goes away. It is believed that, on the valley floor, the groundwater begin to have an effect on the evaporation rate from the ground, which is not the case on the hillslope, because the water table is too deep.

Another point to notice is the transpiration's reliance on the unsaturated zone soil moisture content. The transpiration rate curve follows the unsaturated zone soil moisture curve both on the hillslope and the valley floor. During the rainfall events, the transpiration was almost shut down. Reasons are considered as: 1) The humidity of the air increases; 2) The saturated zone rises.

6.5 Water Budgets

Table 6.1 lists the water budgets for the watershed, the sum of the runoff at the watershed outlet, the transpiration and evaporation is slightly higher than the input of precipitation. This is caused by the interception throughfall process calculated in the model. To reflect the real process, we put the all the water fall on the ground instantly as the amount of irrigation, however, in the model, the interception and through is still calculated, which create more water on the ground by the throughfall process.

Flux	**Event 1**	**Event 2**	**Event 3**	**Event 4**	**Event 5**	**Total**
Precipitation and irrigation(m^3)	3302.359	3516.017	2952.641	2972.073	3108.152	15851.24
Runoff(m^3)	836.27	1459.8	1619	1577.3	2035.6	7527.97
Transpiration(m^3)	2848.4	2572.8	1411.6	921.16	768.09	8522.05
Evaporation(m^3)	76.681	79.399	57.224	50.429	48.391	312.12
Recharge(m^3)	72.143	826.21	1312.8	1364.1	1724.5	5299.75
Baseflow(m^3)	626.58	745.67	928.08	871.52	1131.1	4302.95

Table 6.1 Water Budgets for Shale Hills watershed

6.6 Calibration Strategy

6.6.1 Calibration Process

The calibration of the Shale Hills watershed model needs to take care of the processes of interception, throughfall, infiltration, recharge, groundwater flow, overland flow, the stream flow, etc. Thus, the parameter number sums up to more than 30. In this trial and

error procedure, most of these parameters are set to be constant at present, which does not mean it need not to be calibrated any more, but we have a confidence in these values, and choose the 12 parameters in Table xx as the most unavoidable parameters that needed to be calibrated in the model at first.

The calibration process is accomplished through a trial and error procedure. The original value was set in the property files, and kept as constant through the calibration process. A calibration file is set as multipliers for selected parameters. This facilitates us to reserve the original parameter files, and the calibration file can be a monitor of what we have changed of the original value and how far we are away from the original value.

The vertical conductivity and the infiltration depth of the soil layer has a great effect on the infiltration process, the van Genuchten value is calibrated that, the values compensate for the shutting down of the evaporation from the ground process. And the macropore conductivity, the soil matrix conductivity, the storage coefficient, is calibrated, so that the stream flow peaks and their relaxation time can be matched to the experimental data for the selected 6 events.

Variable Description (Multiplicative Coefficients)	**Original Value**	**Range**	**Optimized Values**
Soil matrix conductivity (horizontal) (m/day)	40.32	1.0~40.32	16.0
Soil matrix conductivity (vertical) (m/day)	10.8	1.0~19	2.5
Macropore conductivity (horizontal) (m/day)	403.2	10~4000	1600
Macropore conductivity (vertical) (m/day)	108	10~4000	250
Storage Coefficient	0.638	0~0.8	0.285
Stream bed roughness(n) ($daym^{-1/3}$)	4.63e-7	1e-7~1e-4	2.31e-5
Surface terrain roughness(n) ($daym^{-1/3}$)	1.16e-6	1e-6~1e-3	4.64e-4
van Genuchten (alpha)	0.595	0.1~10	3
van Genuchten (beta)	1.589	0.1~10	1.8
Infiltration Depth(m)	0.1	0.001~0.5	0.1
Horizontal Conductivity of River walls (m/day)	0.6	1.0~19	48
Vertical Conductivity of the river bed (m/day)	0.1	1.0~19	60

Table 6.2 Calibration Parameters

6.6.2 Sensitivity to Conductivity

In this part, we consider the conductivity as an integral at the optimized value. Then increase all the values by 20% and decrease the values by 20%. The outflow is compared at the outlet of the watershed as shown in Figure 6.16.

The conductivity has a different effect on the first 4 peaks and the last 2 peaks. In the first 4 peaks, increasing the conductivity will decrease the peak value, while in the last 2 peaks, the values keeps almost the same. On the other hand, decreasing the conductivity value has a great effect on the first peak, which is generated by the overland flow.

These sensitivity characteristics help us understand the influence of different parameters during the calibration process to get the optimized parameters.

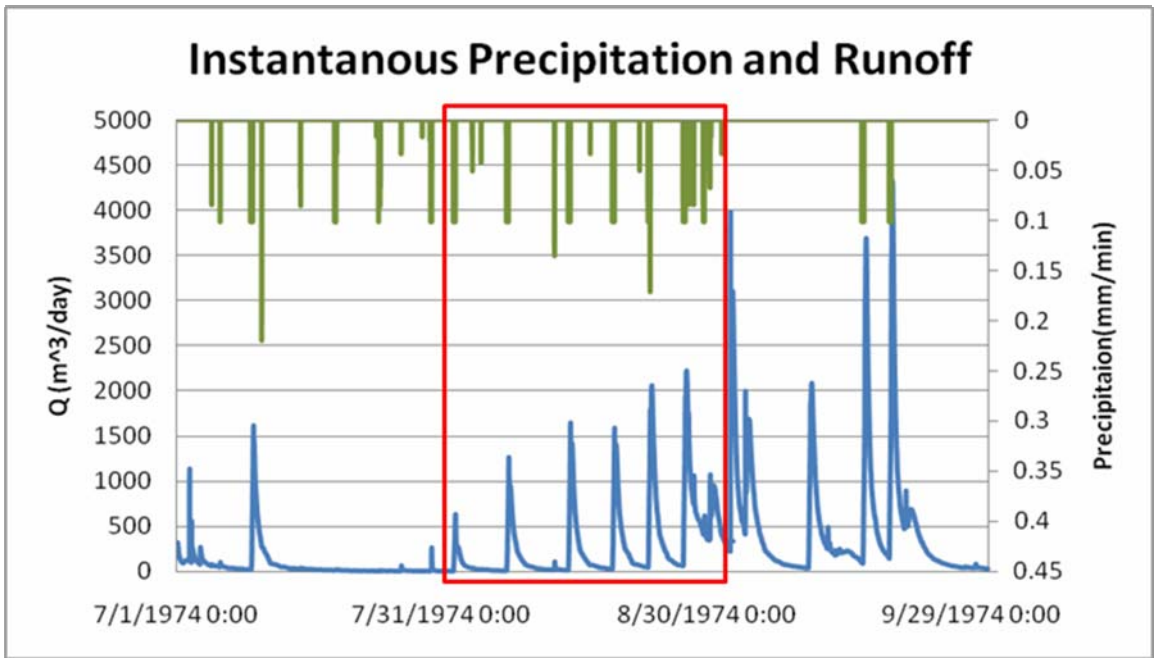


Figure 6.1 Precipitation and Runoff at outlet in 1974.The red window is the model simulation perid for 31 days, which includes 6 events.

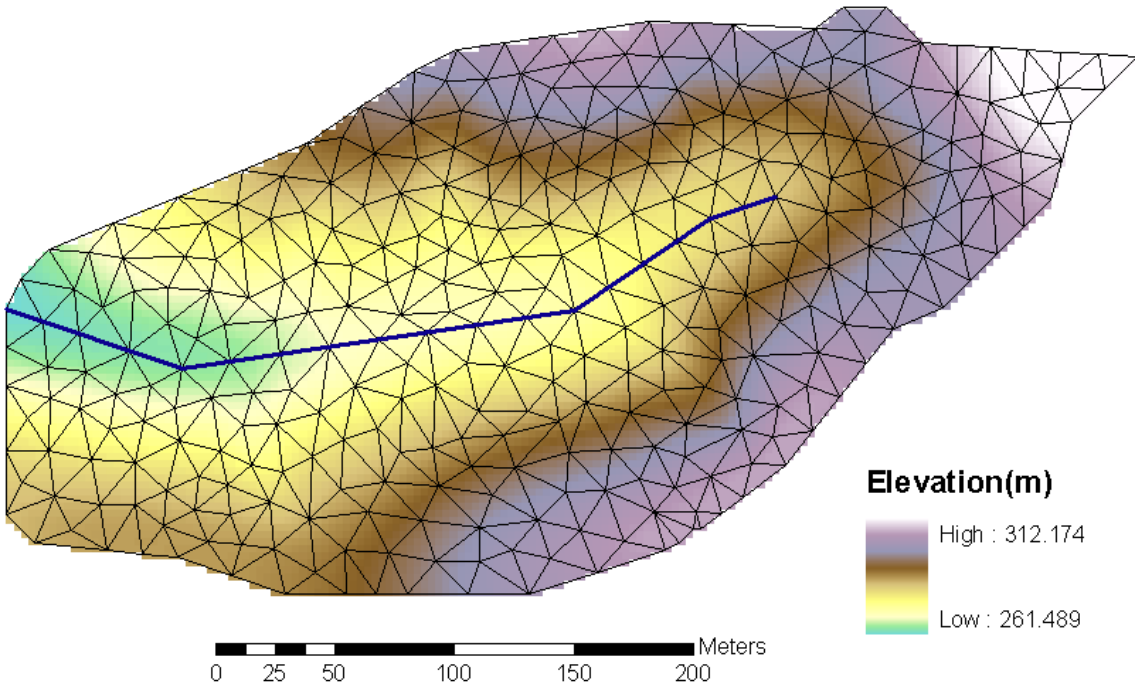


Figure 6.2 TIN network for the calculation, including 598 cells, 332 nodes, and 18 stream segments

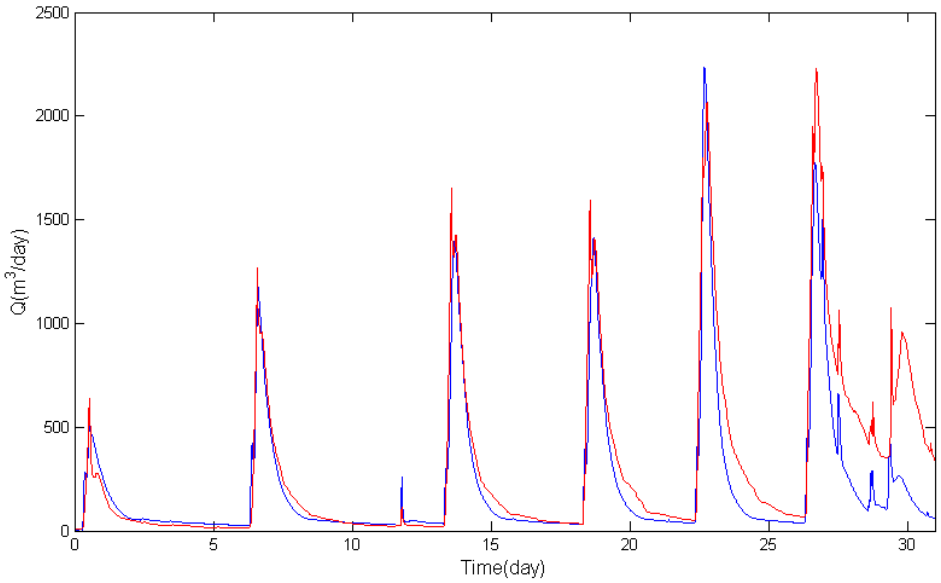
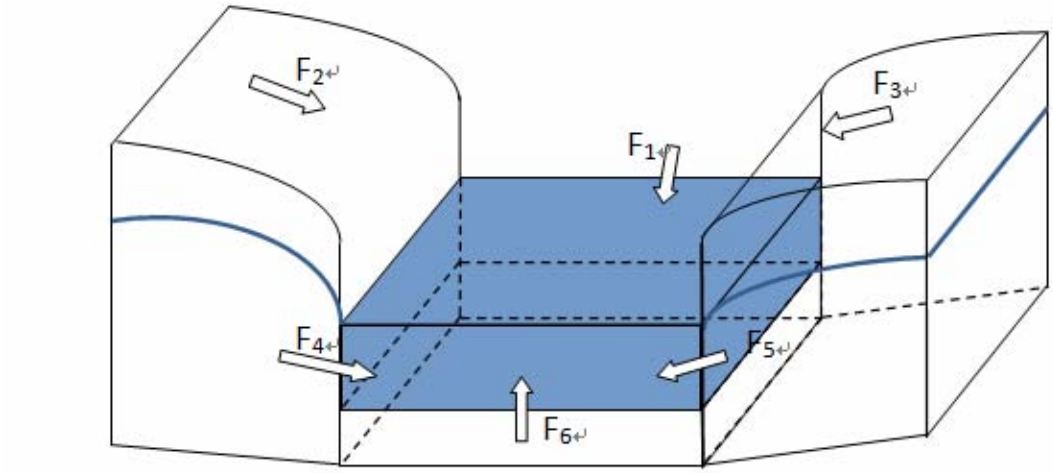
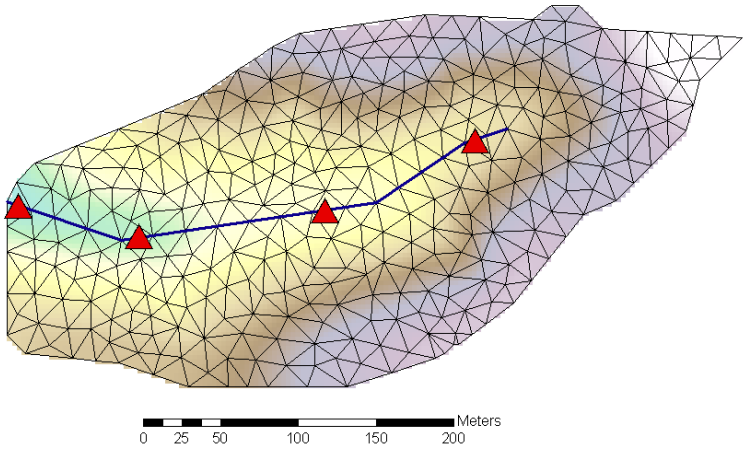


Figure 6.3 Observed (red) verses model (blue) runoff at the outlet Shale Hills watershed for the artificial irrigation experiment in 1974. The simulation showed good prediction of the first five peaks and the base flow. However, the relaxation of the peak that is in the model is much quicker than the observed data. And for the last peak, the simulation failed to reach the peak value from the experimental data.



(a)



(b)

Figure 6.4 (a) Input components to streams from surface flow and subsurface flow. F_1 is the flux from upstream element, F_2 and F_3 are overland flow from left and right bank, F_4 and F_5 are subsurface flow from left and right bank, and F_6 is the flow exchange between the stream flow and the stream bed. (b) The 4 location chosen to show the input components in Figure 6.5

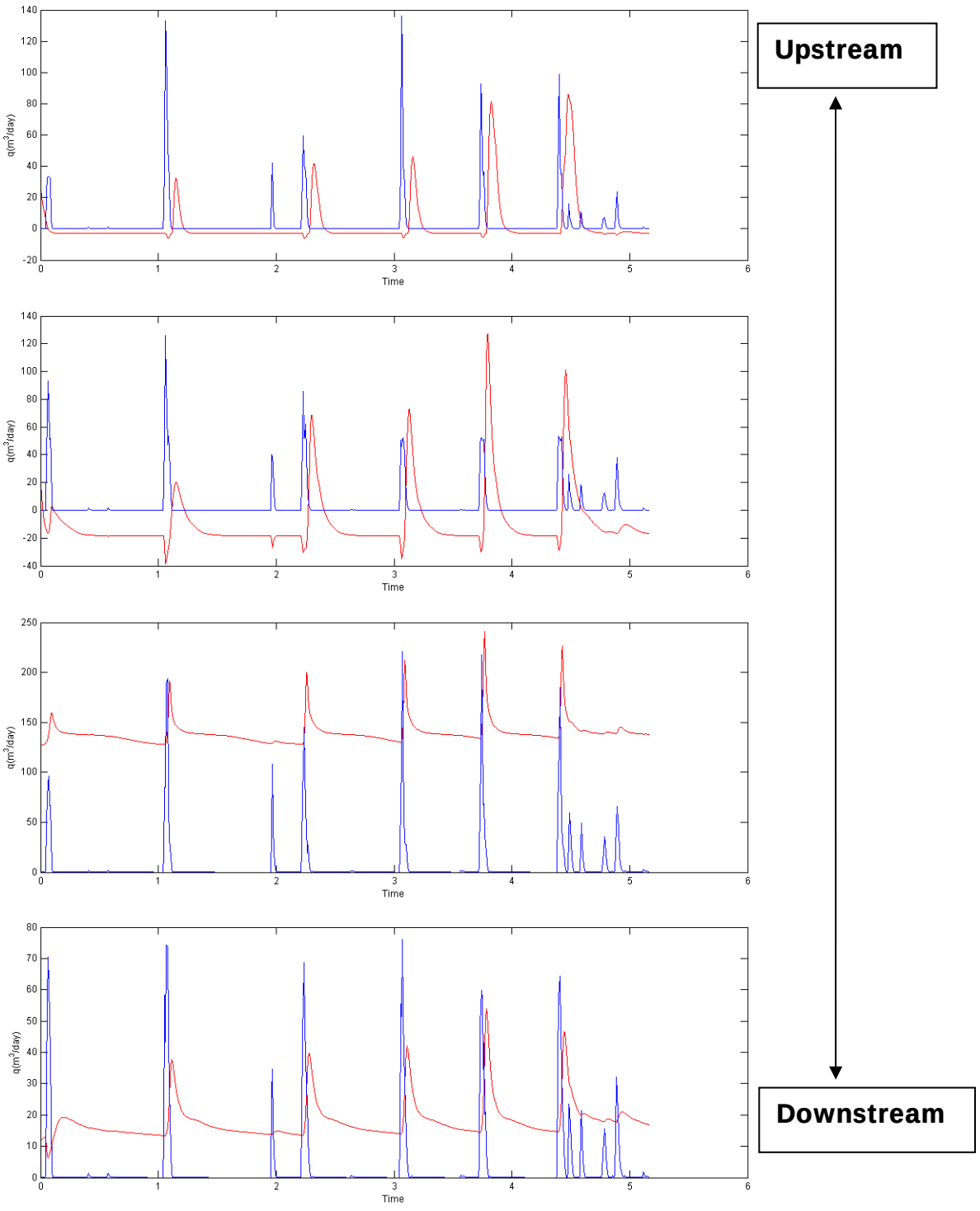


Figure 6.5 Water input components to stream. Surface flow (blue) and subsurface flow(red) are marked in Figure 6.4(a).The 4 locations are displayed in Figure 6.4 (b).

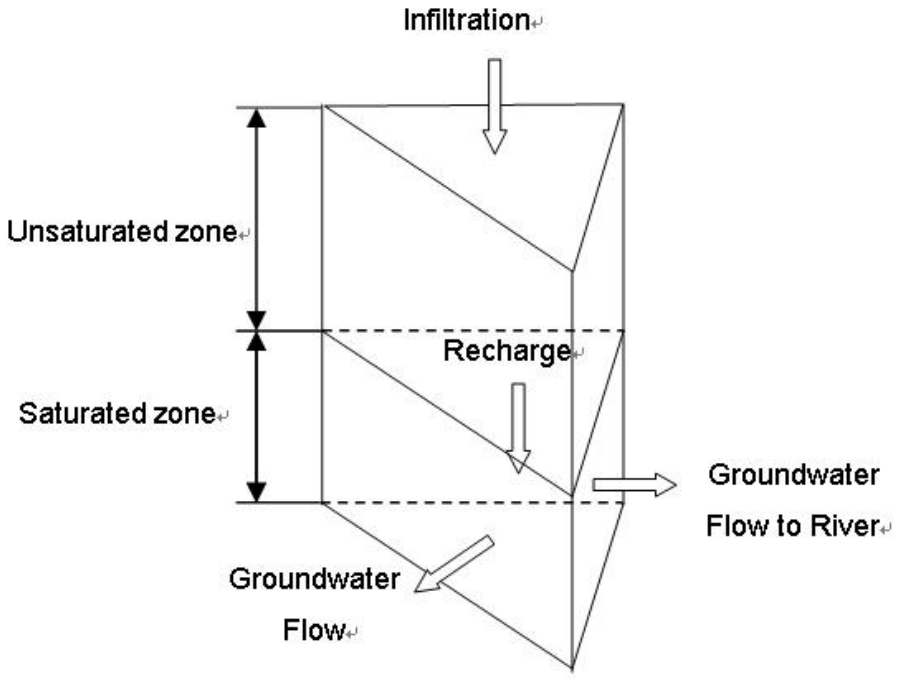
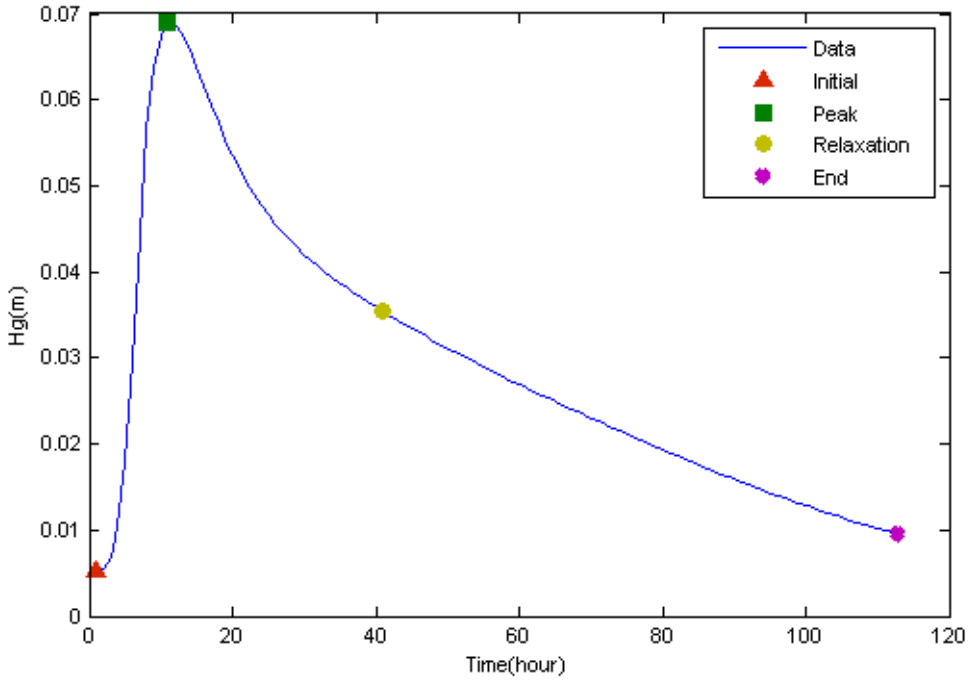
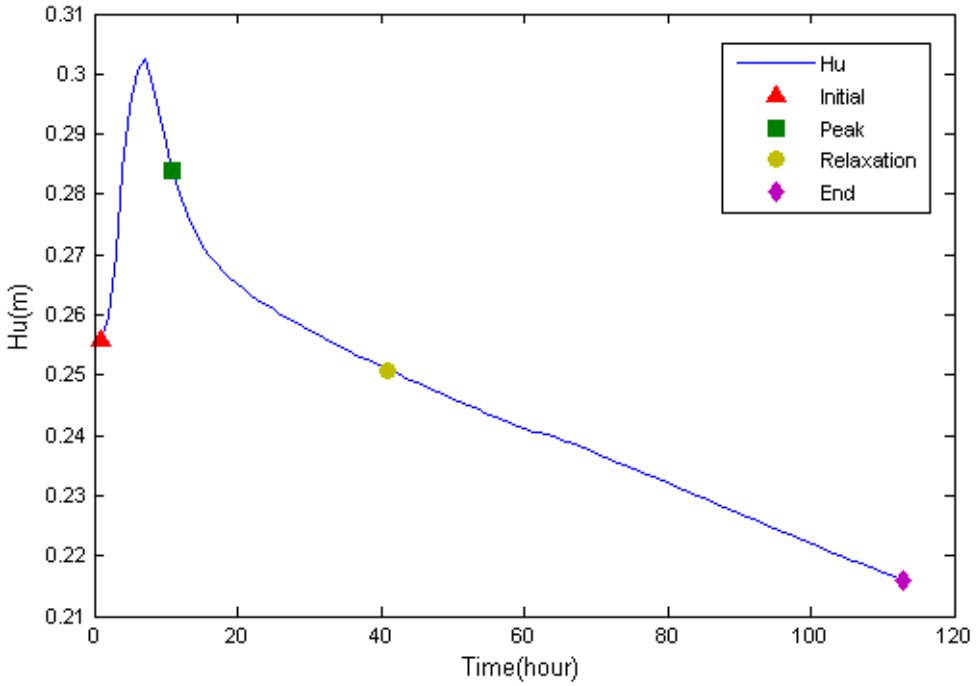


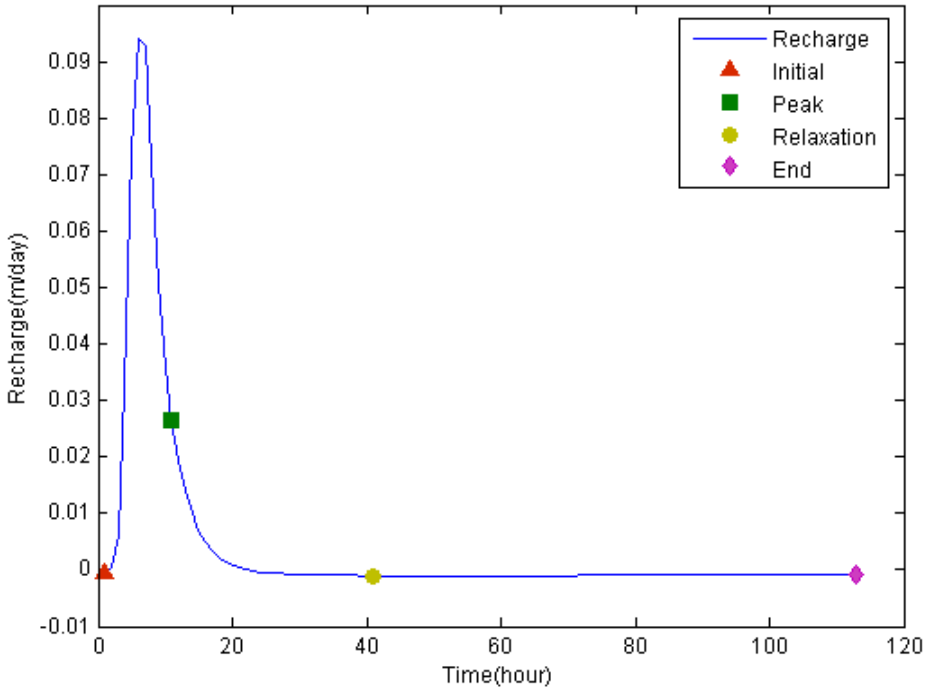
Figure 6.6 Subsurface Flows



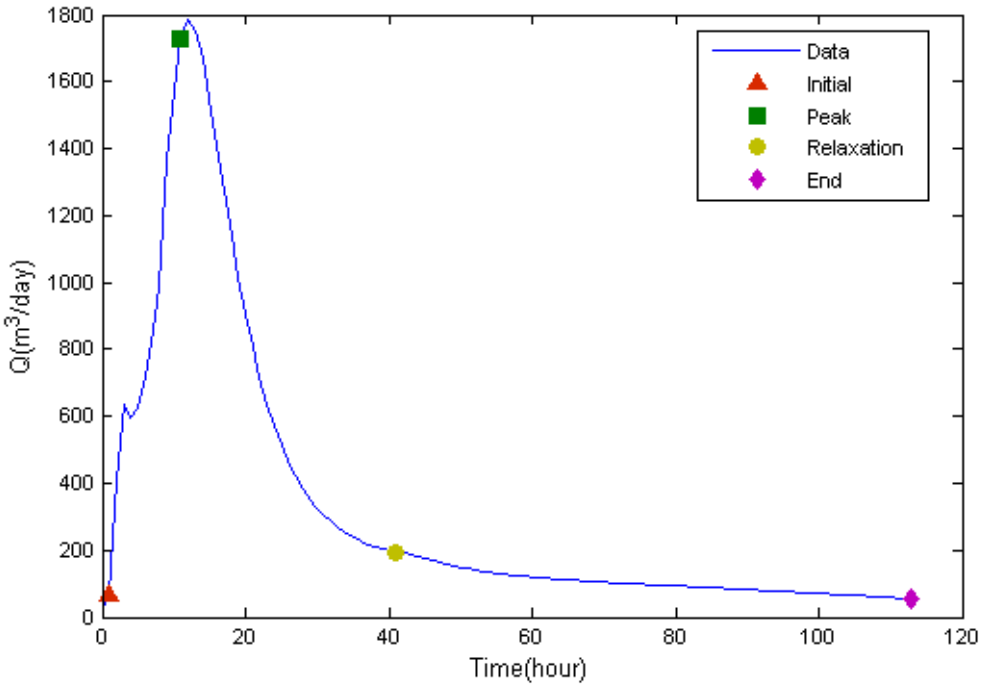
a) Groundwater head



b) Unsaturated Zone soil moisture content



c) Recharge rate



d) Runoff at the outlet

Figure 6.7 Time transition of the spatially averaged groundwater head, unsaturated zone soil moisture content, recharge rate, runoff at the outlet during the 3rd event. The red triangular is the time when the irrigation begins. The 4 markers in each plot stand for 4 identical time that will be plotted in Figure 6.8

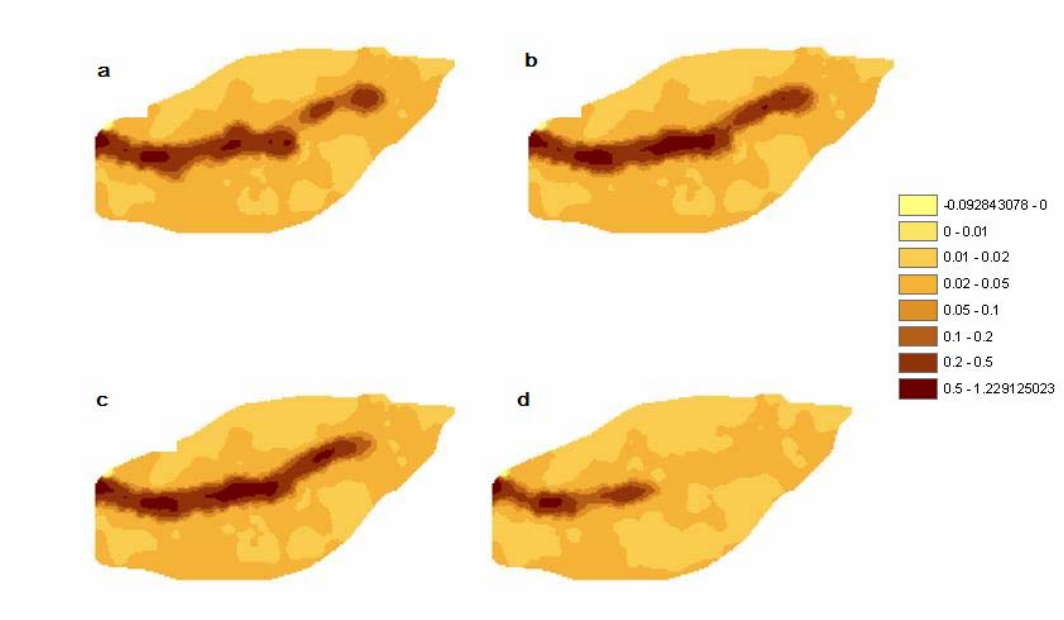


Figure 6.8 Five day period showing the time varying spatial distributions of the groundwater heads during 3rd event. The 4 time is marked in Figure 6.7 as a comparison. a) the initial condition for the event, 6 hours before the peak value of the spatially averaged peak groundwater head. b) The groundwater head reaches peak value. c) 12 hours after the peak, the groundwater has been redistributed, and the recharge from unsaturated zone ends. d) 112hours after the peak, when the relaxation of the flow peak has most of the water redistributed.

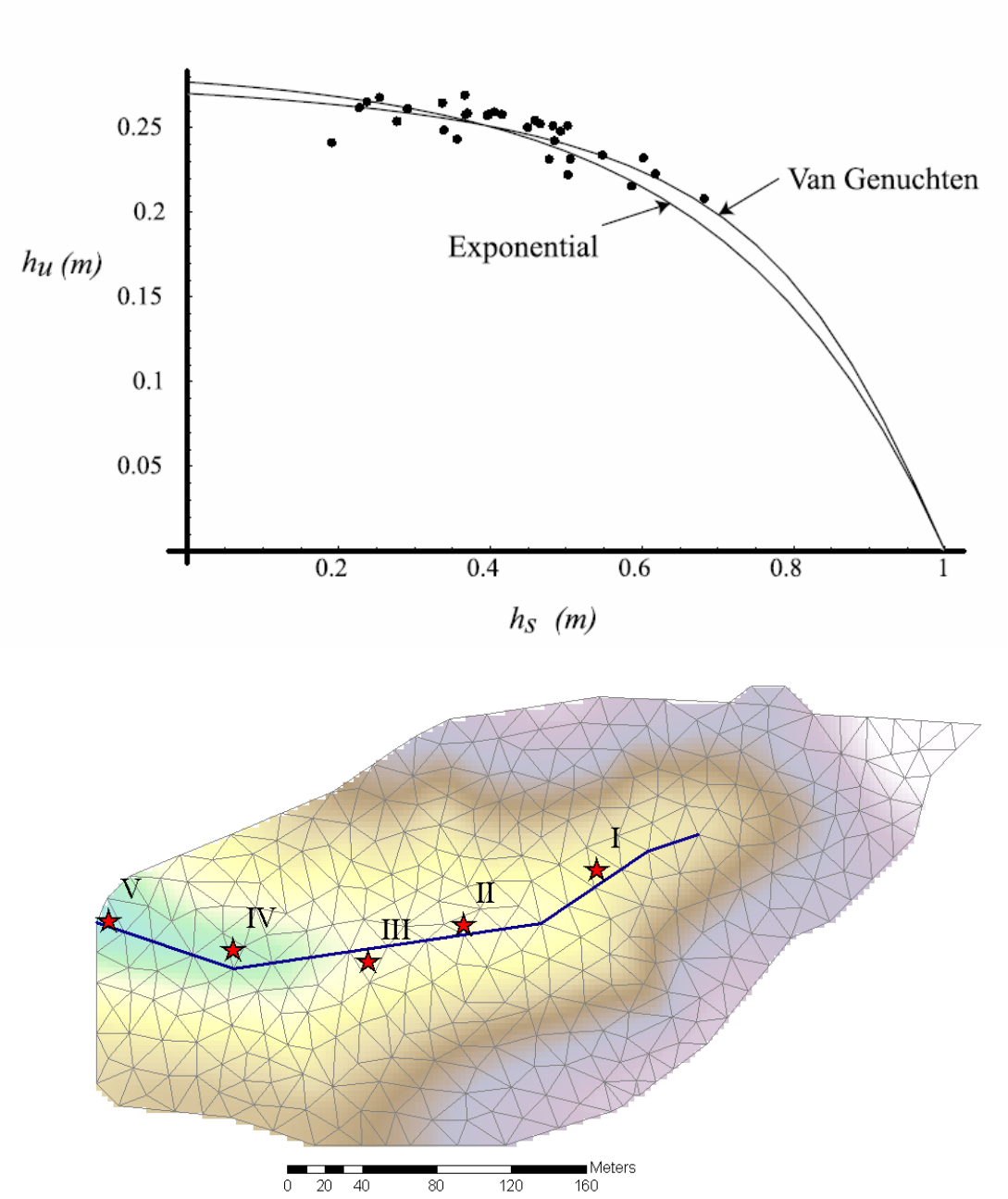
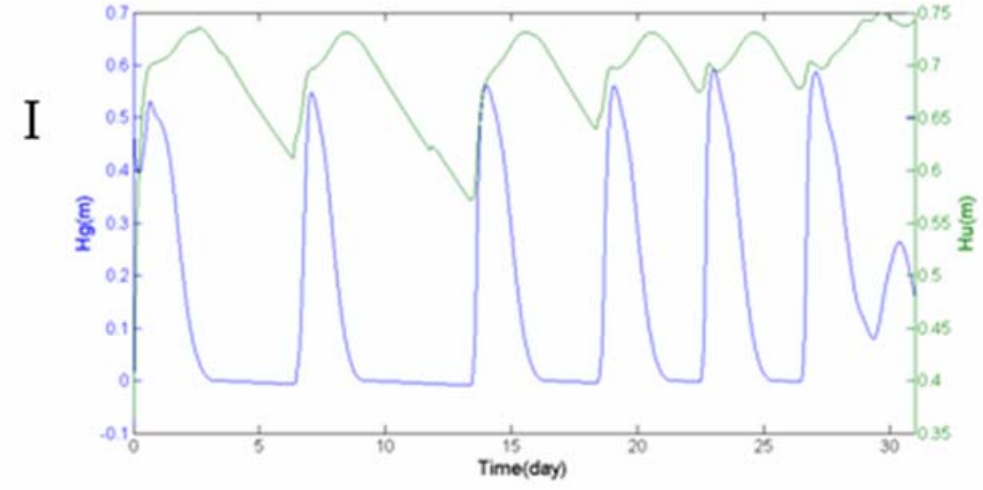
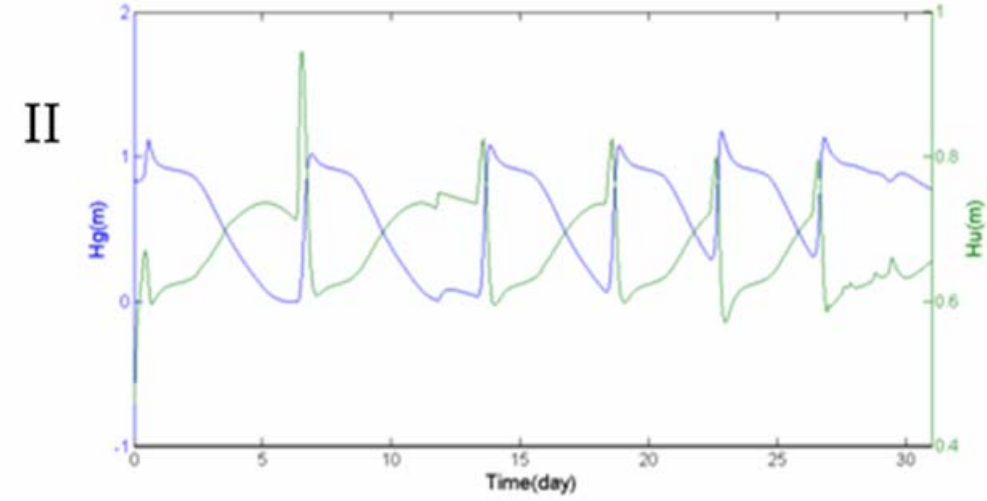


Figure 6.9 a) The experimental data showing the relationship between the saturated zone soil moisture content and the unsaturated zone soil moisture content (Qu and Duffy, 2007). b) 5 location in the riparian area of the watershed chosen for the comparison in Figure 6.10.

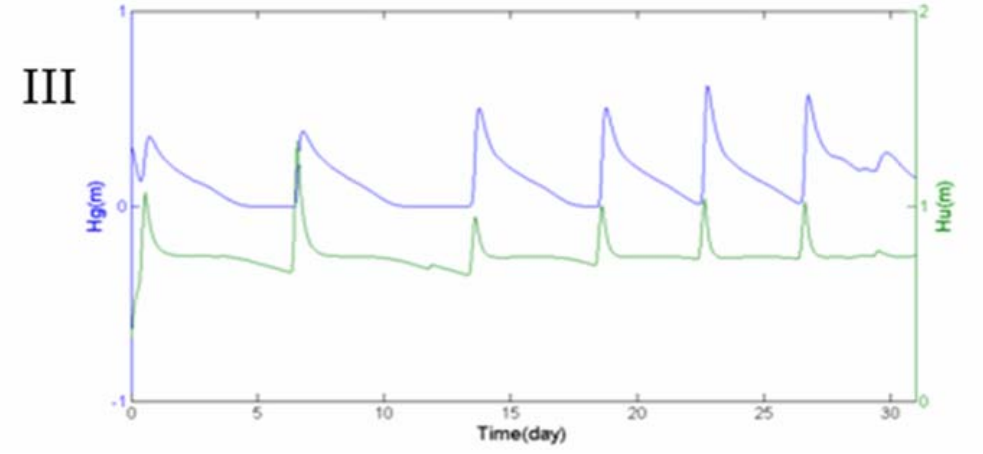
I



II



III



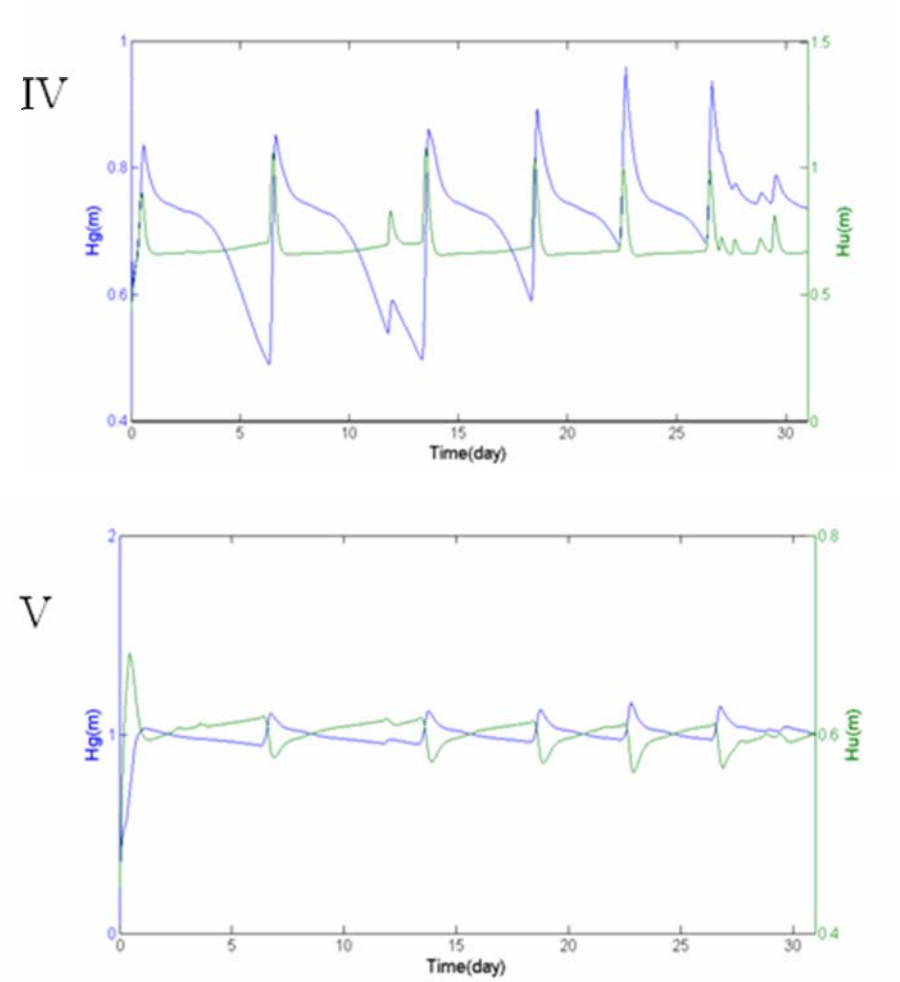


Figure 6.10 Comparison of the soil moisture contents in saturated and unsaturated zone at 5 locations shown in Figure 6.9(b)

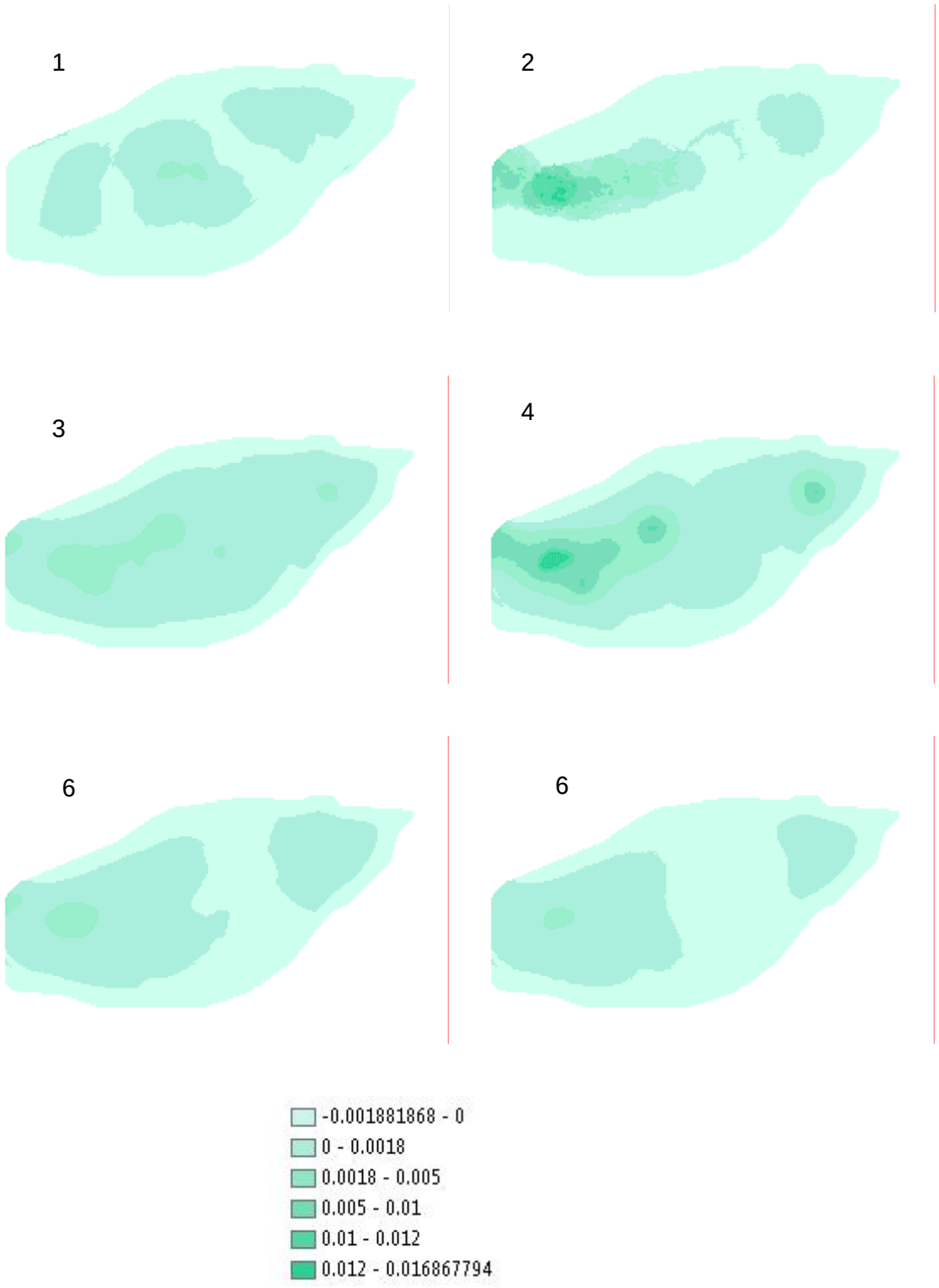


Figure 6.11 Averaged recharge rate. The above 6 figures shows the averaged recharge rate for 6 events, each is an average of the most intense recharge period of 10 hours.(Units are in m/day)

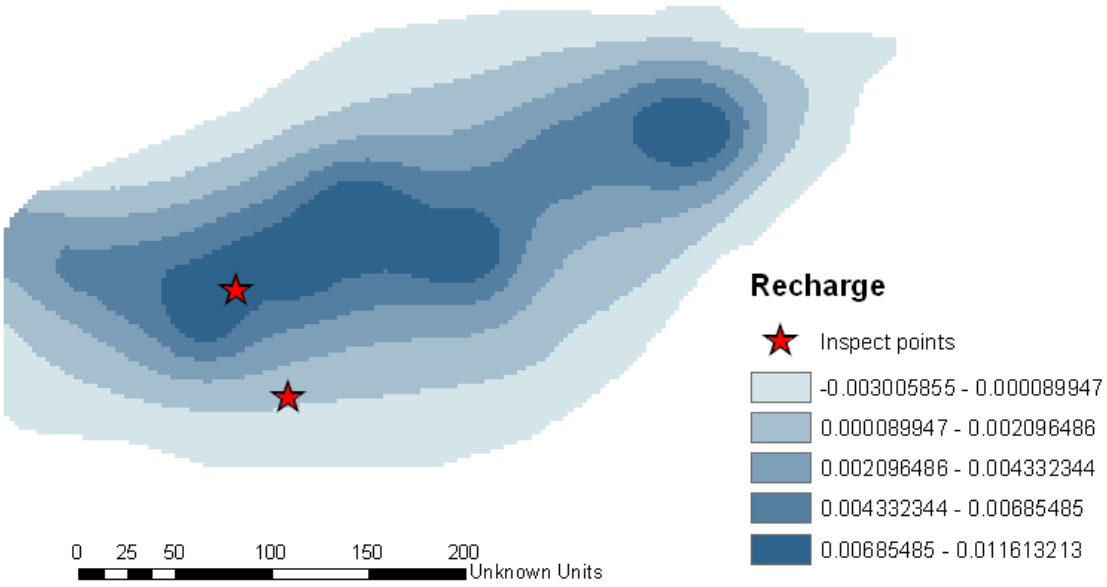


Figure 6.12 Time averaged (31 days) recharge rate. The marker represents the location of comparison of all the vertical flows in Figure 6.15.

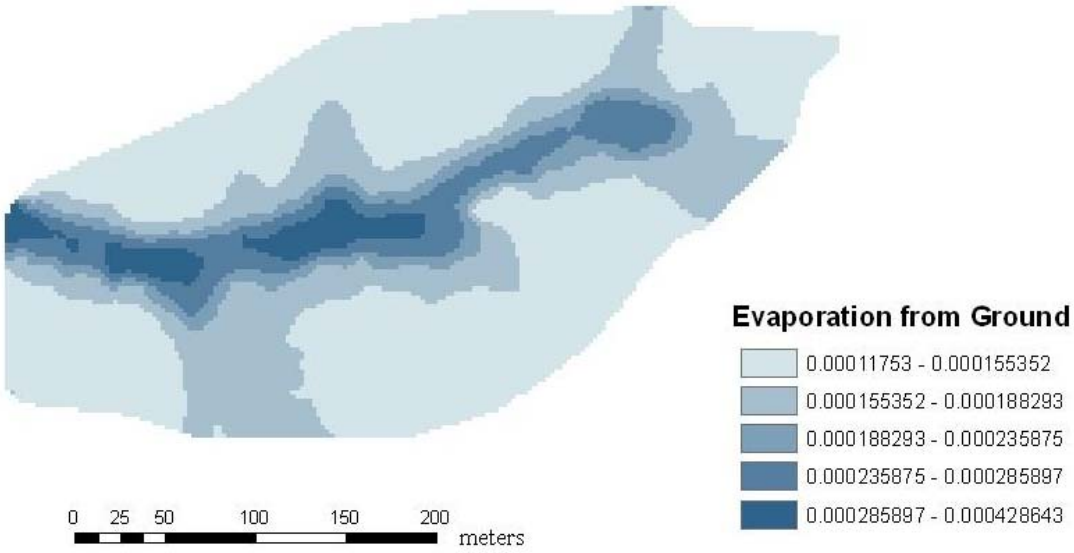


Figure 6.13 Time averaged (31 days) evaporation rate from the ground

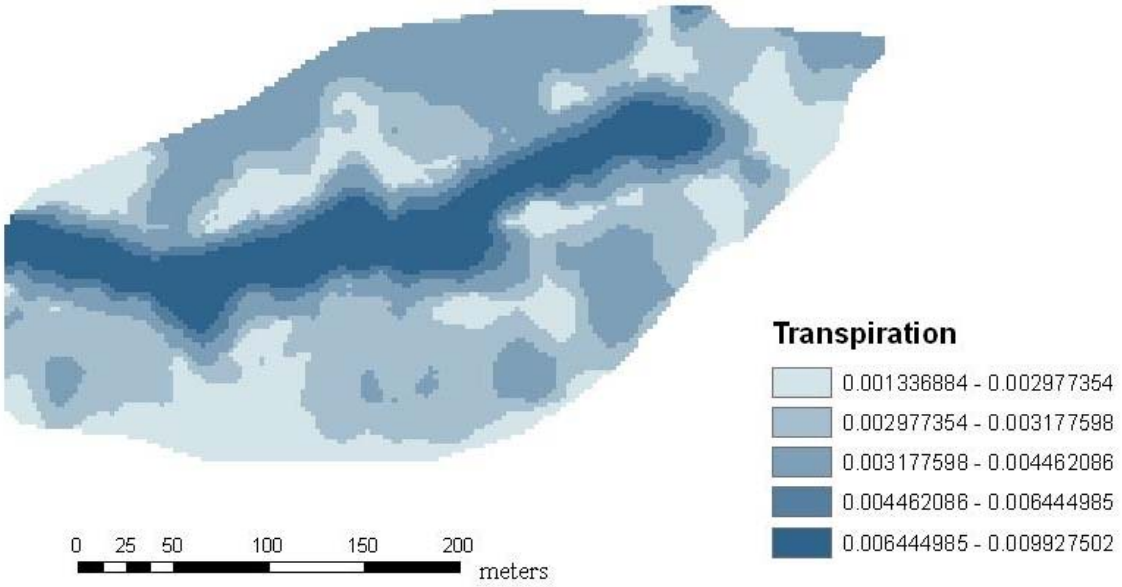
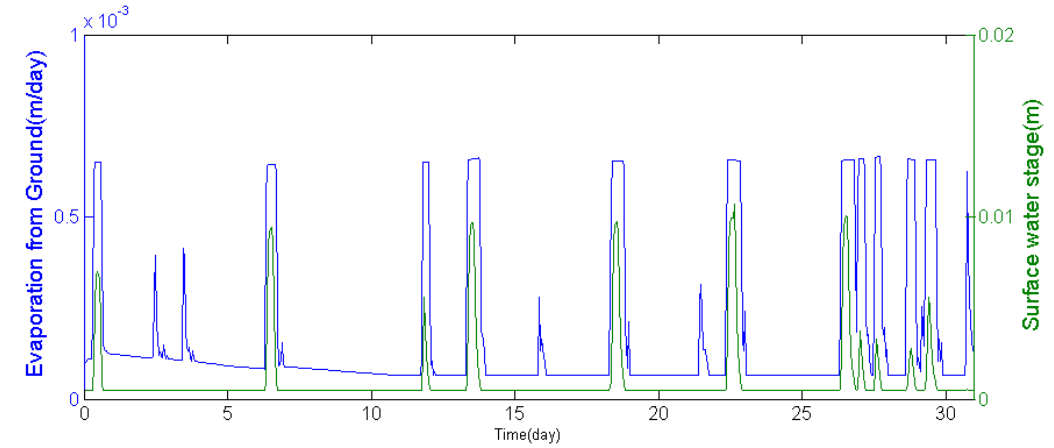
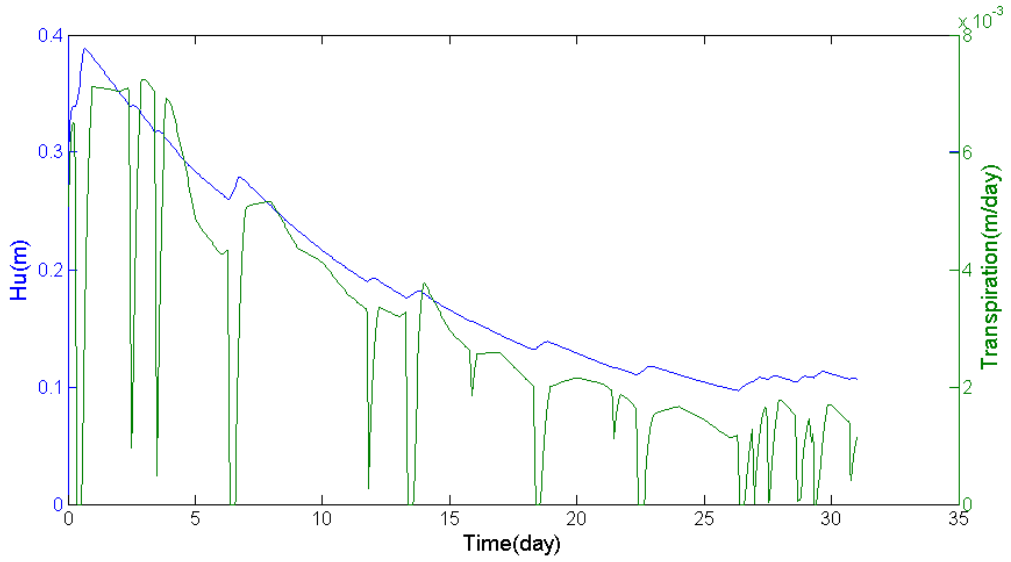
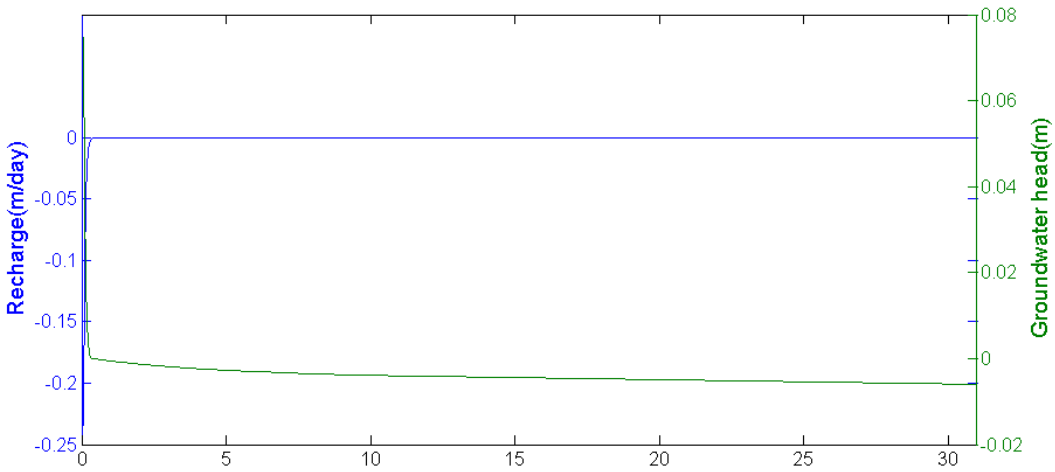


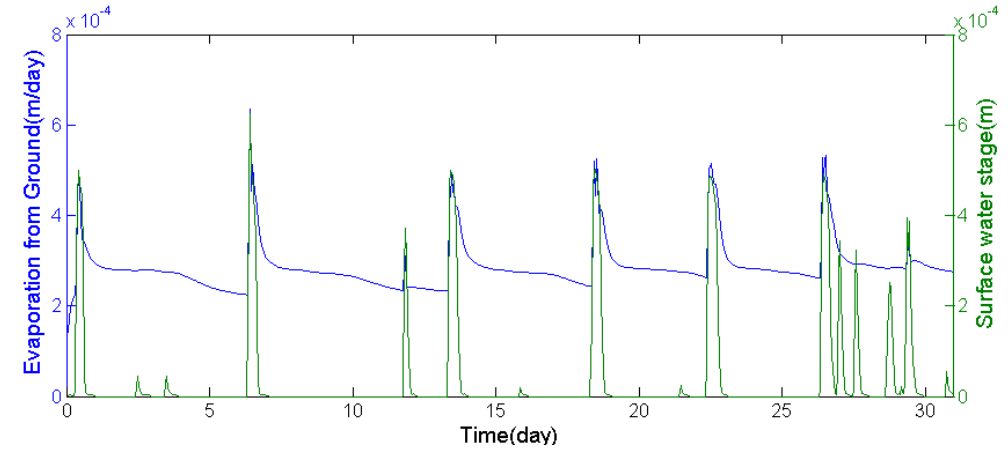
Figure 6.14 Time averaged (31 days) transpiration rate

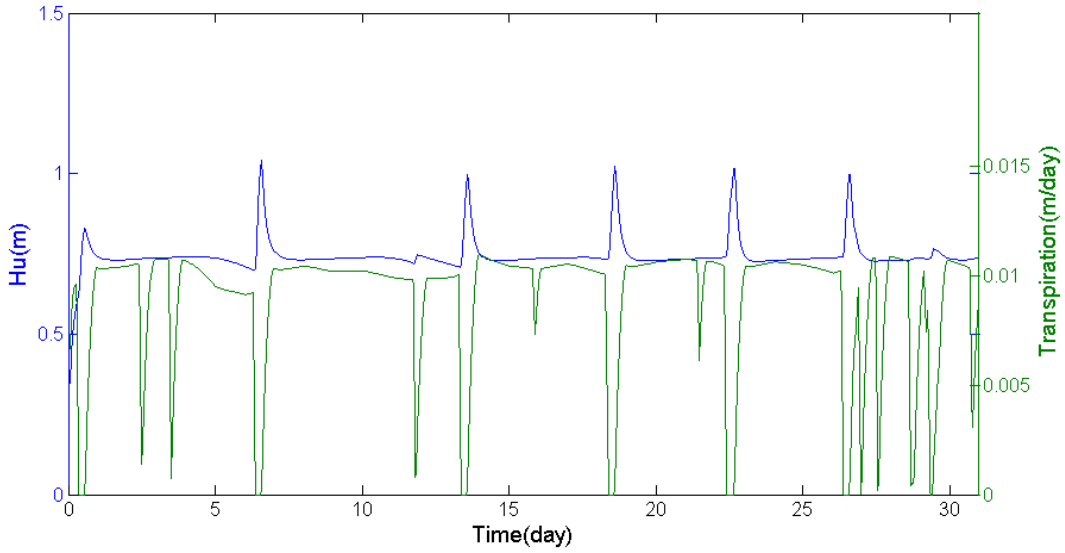


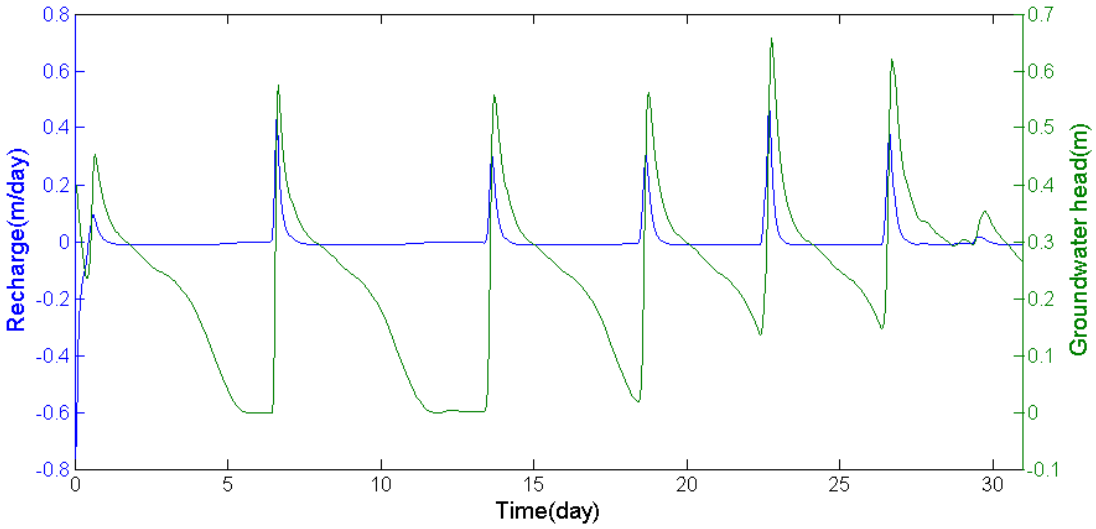




a) Up on the slope







b) Down near the river

Figure 6.15 Multiple vertical fluxes inside the model at 2 locations: a) up on the slope, b) on the valley floor. They are marked as red star in Figure 6.12.

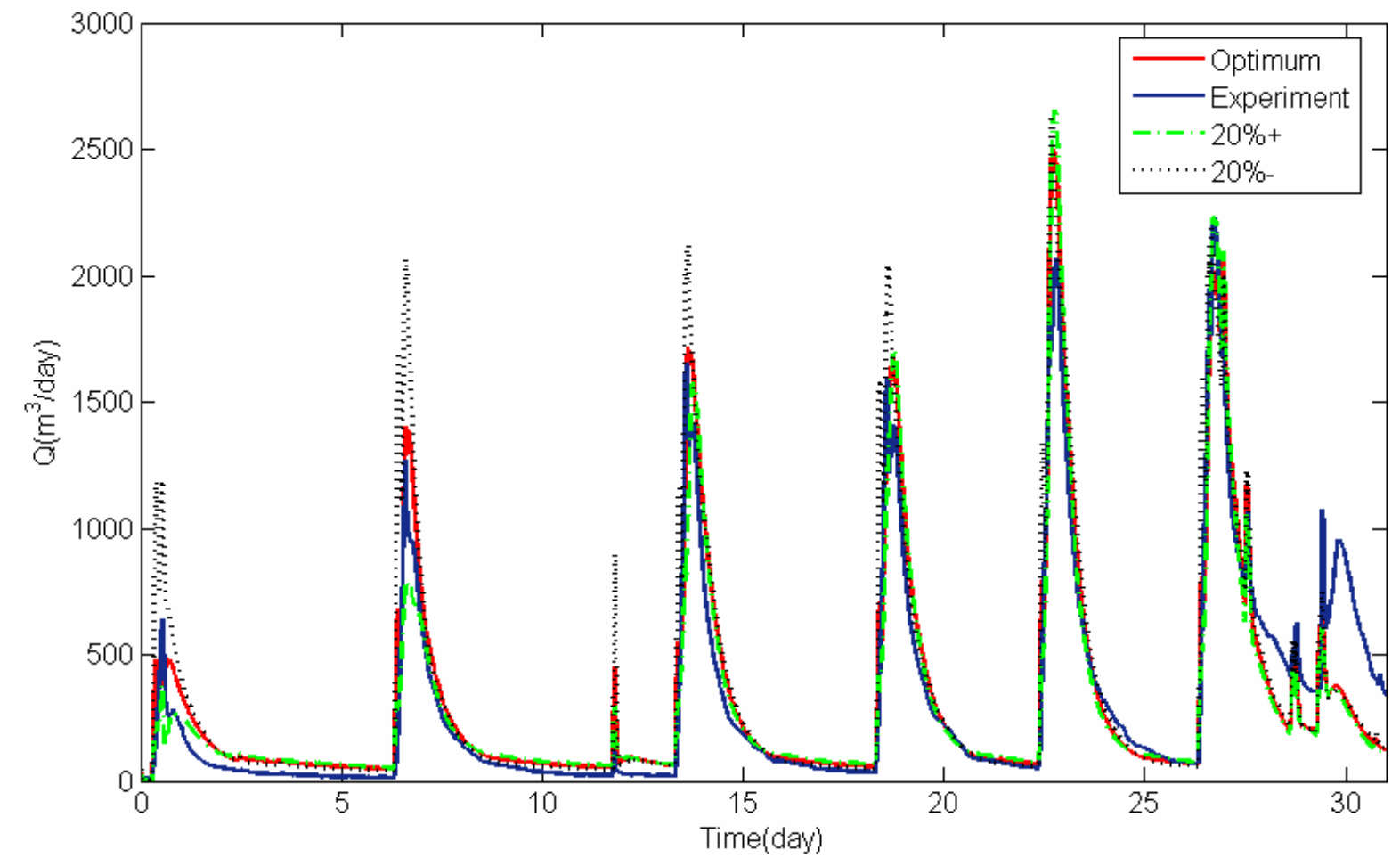


Figure6.16 Sensitivity of the Conductivity

Chapter 7

Conclusions and Future Work

The thesis presents the implementation of the PIHM model to the Shale Hills watershed. Compared to Qu and Duffy(2007) simulation of the Shale Hills watershed, the new model updated the A-priori data by the latest data collected by Dr. Henry Lin's group, and made an initial guesses for the land cover classification within the watershed based on data from Dr. David Eissenstat's group. Taking the advantage of the GIS interface that is newly integrated into the model, the thesis also detailed the steps of delineating the watershed, and prepared the input files for the PIHM model. The model successfully simulated the runoff at the outlet, and showed that, out of the several rainfall runoff generation mechanisms: (1) Horton overland flow; (2) Groundwater sustained base flow; (3) Subsurface storm flow; (4) Saturated Overland Flow; the Horton overland flow and the subsurface storm flow played an important part in creating the peak in the runoff water. On the other hand, the model also made predictions for the groundwater, the evaporation from ground, transpiration, recharge, etc. The results showed that these results are highly correlated to the complex terrain conditions in the watershed, and however, due to the complexity of the patterns, the results are hard to interpret.

In order to enhance the model, there are several aspects that can be considered in future work. 1) The present model applied a weighted average conductivity to integrate the macropore effect into the model. However, this approach ignore the different mechanism of conducting water in the macropores and soil matrix. Wierenga and Van Genuchten (1989) proposed the mobile-immobile transport theory, which assumed that the water flow in the macropores is the dominant flow while the matrix exchanges moisture with the macropores by diffusion. This will help the model get a more reasonable result of the groundwater and soil moisture distribution. And it will also emphasize the importance of the subsurface stormflow. 2) The current model handles the soil column as only 2 layers: the saturated layer and the unsaturated layer. And the soil properties are considered uniform from top to bottom. This however, overlooks the vertical variation of the vertical properties, such as conductivity, the porosity, the macropore distribution, etc. Multiple layers are expected in newer versions of PIHM to better resolve the soil moisture. 3) In the application of the PIHM model, the calibration process is tedious and time-consuming. A semi-automatic calibration system is suggested, which has solid criteria to help the user have a better experience of using the model. 4) Based on the work done in this thesis, and the development of RTH-NET, a real-time modeling system for Shale Hills watershed is ready to be established.

References

Brandes, D., Macropore effects on soil moisture and runoff response at the hillslope scale, numerical experiments and low-dimensional dynamics

Bhatt, G, Kumar, M., Duffy, C.J. (2008) Bridging gap between geolhydrologic data and Integrated Hydrologic Model: PIHMgis, *iEMSs 2008: International Congress on Environmental Modeling and Software.*

Bhatt, G, Tutorial: V-catchment watershed (Tutorial) for PIHMgis

Dingman, S. L. (1994), Physical Hydrology, Prentice-Hall, Upper Saddle River, N. J.

Duffy, C. J.(1996). A two-state integral-balance model for soil moisture and groundwater dynamics in complex terrain. *Water Resourc. Res*; *Vol.*32, *NO.*8, *PAGES* 2421-2432

Kumar, M., Bhatt, G., Duffy, C. J.(2009) The role of physical, numerical and data coupling in a mesoscale watershed model.

Kumar, M., Bhatt, G., Duffy, C. J. An object oriented shared data model for GIS and distributed hydrologic models.

LDAS Database: http://ldas.gsfc.nasa.gov/LDAS8th/MAPPED.VEG/web.veg.table.html,

Lin, H., Kogelmann, K., Walker, C., Bruns, M.A.(2005) Soil moisture patterns in a forested catchment: A hydropedological perspective. *Geoderma* 131 (2006):345–368 (2006*).*

Lin, H.(2006) Temporal stability of soil moisture spatial pattern and subsurface preferential flow pathways in the Shale Hills catchment. *Vadose Zone Journal* 5:317–340 (2006*).*

Lynch, J. A. (1976), Effects of antecedent moisture on storage hydrographs, *Ph.D. thesis*, 192 pp., Dep. of Forestry, Pa. State Univ., University Park.

Lynch, J. A., andW. Corbett (1985), Source-area variability during peakflow in watershed management in the 80's, *J. Irrig. Drain. Eng.*, 300– 307.

Tchaou, M. K. .(1999). A two state dynamical model for rainfall-storage-runoff in a small upland forested watershed, *Ph.D. thesis*, 120 pp., *Dep. of Civ. And Env*., Pa. State Univ., University Park.

Tindall, J. A., Kunkel, J.R. (1999) Unsaturated Zone Hydrology, Prentic-Hall, Upper Saddle, N. J.

Qu, Y.(2004). An integrated hydrologic model for multi-process simulation using semi-discrete finite volume approach. *Ph.D. thesis*, 167 pp., *Dep. of Civ. And Env*., Pa. State Univ., University Park.

Qu,Y.,Duffy, C.J. (2007). A semidiscrete finite volume formulation for multiprocess watershed simulation. *Water Resourc.Res*; *Vol.*43

van Genuchten, M. T. (1980), A closed form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. Am. J.*, 44, 892– 898.