This is to certify that the
dissertation entitled

A PROCESS-BASED DISTRIBUTED HYDROLOGIC MODEL AND
ITS APPLICATION TO A MICHIGAN WATERSHED

presented by

Chaopeng Shen

has been accepted towards fulfillment
of the requirements for the

Ph. D. degree in Environmental Engineering

Major Professor's Signature

12/18/2009

Date

A PROCESS-BASED DISTRIBUTED HYDROLOGIC MODEL AND ITS APPLICATION TO A MICHIGAN WATERSHED

By

Chaopeng Shen

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Environmental Engineering

2009

ABSTRACT

A PROCESS-BASED DISTRIBUTED HYDROLOGIC MODEL AND ITS APPLICATION TO A MICHIGAN WATERSHED

By

Chaopeng Shen

The PAWS (Process-based Adaptive Watershed Simulator) model is a novel distributed hydrologic model that is based on solving partial differential equations (PDE) for physical conservation laws of the hydrologic cycle. The objective is to create an efficient physically-based modeling framework to describe the linkages between processes at different scales and to improve the applicability of physically-based models. The model simulates evapotranspiration, overland flow, channel flow, unsaturated soil moisture, groundwater flow, depression storage, vegetation growth and snowpack. PAWS focuses on the dynamic surface- subsurface interactions and integrated responses by efficiently coupling runoff and groundwater flow to the vadose zone processes governed by the Richards equation. This novel approach solves a long-standing bottleneck in PDE-based subsurface flow modeling by removing the computational limitations while maintaining physically consistent solutions. Surface flow is solved by an efficient Runge-Kutta Finite Volume (RKFV) scheme.

We follow the Freeze and Harlan (1969) blueprint in that we believe each component of the model should be verifiable by itself. All flow components have been independently verified using analytical solutions and experimental data where applicable. PAWS utilizes readily available data from national databases. The model is applied to a medium-sized watershed in Michigan achieving high performance metrics in terms of streamflow prediction at two gages during the calibration period and the verification period. The baseflow flow periods are described particularly well. Starting from a rough initial estimate of the groundwater heads, the model describes the observed groundwater heads well (R^2=0.98). The annual hydrologic fluxes are close to those estimated by a calibrated SWAT model. The model is considerably less expensive than previous physically-based models of similar complexity.

The model is able to elucidate the complex interactions of processes in space and time. Such detailed, quantitative and mechanistic descriptions cannot be produced by conceptual models. The watershed is found to be a subsurface-dominated system with saturation excess being the main runoff generation mechanism. Infiltration, recharge and ET are also found to be strongly related to topography and groundwater flow. The large seasonal variation of energy input drives the strong annual cycle and markedly different responses in streamflow.

To my father, Hongxun Shen, and my mother, Jieying Li,
who have given me the best they have, and
all the love in the world

Acknowledgements

I would like to express my sincere appreciation to my research and academic advisor, Dr. Mantha S. Phanikumar, for his diligent, insightful and incredibly careful guidance throughout my graduate studies, without which this dissertation would have been impossible. I still have much to learn from his relentless enthusiasm, rigorous professionalism and uncompromising pursuit of quality. I am grateful that I had the opportunity to work with him on such a grand project.

I would like to thank my committee members: Dr. Joan B. Rose, Dr. Shu-Guang Li and Dr. Ashton M. Shortridge for their invaluable comments, suggestions and advice to improve this thesis to its final form.

I would like to express my sincere appreciation for NOAA Center for excellence for Great Lakes and Human Health which has provided the financial support for my Ph.D. study.

I would like to thank lab mate Niu Jie for his assistance with data processing and GUI development. I also thank Mehmet Oztan, Hassan Abbas and Dr. Huasheng Liao in Dr. Shu-Guang Li's group for their assistance with groundwater data processing.

I would like to acknowledge the folks at MSU High Performance Computing Center (HPCC) for the numerous help they offered. Special thanks go to Andrew Keen and Dr. Dirk Colbry.

I want to take the chance to thank the many who have helped with experimental data

collection reported in this dissertation, including Dr. Aslam Irfan, Dr. Shawn McElmurry and Heather Stevens, to name a few.

I would pay special thanks to Dr. Susan Masten who has initially given me the chance to study at MSU. Otherwise this dissertation would have been non-existent.

Finally I want to thank Miss Wenqian Shan for her endless care and dedicated support, which has given me much of the motivation at all stages of my Ph.D. study.

Table of Contents

List of Tables x

List of Figures xii

Chapter 1. Background and literature review 1
- 1.1. Motivation for a new hydrologic model 1
- 1.2. Review of hydrologic models 4

Chapter 2. Development of the Hydrologic Model: Mathematical Bases and Test Cases 12
- 2.1. Model Processes 13
- 2.2. Discretization and Representations 15
- 2.3. Hydrologic Processes and Solution to each flow domain 20
  - 2.3.1. Evapotranspiration 20
  - 2.3.2. Vegetation 29
  - 2.3.3. Overland flow 31
  - 2.3.4. Channel flow 43
  - 2.3.5. Unsaturated vadose zone model 48
  - 2.3.6. Saturated groundwater flow model 54
- 2.4. Interactions between domains 61
  - 2.4.1. Coupling of the Unsaturated Richards equation and the groundwater flow equation: 61
  - 2.4.2. Coupling of the vadose zone and surface flow: 75
  - 2.4.3. Interaction between overland flow and channel flow 79
  - 2.4.4. Interaction between groundwater and channel flow 84
- 2.5. Data preparation steps general to all watersheds 86
  - 2.5.1. Digital elevation data 86
  - 2.5.3. Soils 91
  - 2.5.4. Climatic data 91
- 2.6. Model Flow Diagram 93
- 2.7. Test cases 95
- 2.8. Summary and Conclusions 112

Chapter 3. Development of the Hydrologic Model: Model Applications and Comparisons 114
- 3.1. The Red Cedar River watershed model 114
  - 3.1.1. Study site and input data 114
  - Groundwater data 128

3.1.2. Model Calibration 130
3.2. SWAT Model for comparison 131
3.2.1. Brief summary of SWAT mathematical bases 132
3.2.2. SWAT Model Setup 136
3.3. Results and Discussions 139
3.3.1. Model Evaluation 139
3.3.2. Model performance evaluations 139
3.3.3. Additional model results and hydrologic system of the RCR watershed 151
3.3.4. Understanding the hydrology 153
3.4. Summary and Conclusions 167
3.5. Limitations and future research 168
3.6. Software package 168

Chapter 4. Estimating Longitudinal Dispersion and Surface Storage in Streams Using Acoustic Doppler Current Profilers 170
4.1. Introduction 170
4.2. Description of Sites 179
4.3. Materials and Methods 182
4.3.1. Transient Storage Modeling 188
4.3.2. Multi-resolution Wavelet Decomposition of ADCP Data 191
4.3.3. Estimating the Longitudinal Dispersion Coefficient 196
4.4. Results 200
4.4.4. Evaluation of Channel Features and Potential for Hyporheic Exchange in the RCR 200
4.4.5. Evaluation of TS Model Parameters for RCR 202
4.4.6. Results from ADCP Surveys and Wavelet Analysis 207
4.4.7. Estimating Longitudinal Dispersion in Rivers 215
4.5. Conclusions 229

Appendix A . Supplemental Tables 232

Appendix B. User's Manual for the model Graphical User Interface 238
B.1. Creating and Running the model in interactive mode 238
B.1.1. Installing and starting the model 239
B.1.2. Loading the data 240
B.1.3. Setting up the grid 243
B.1.5. Setting up solution schemes and time steps 247
B.1.6. Final model set up 249
B.1.7. Saving and Loading model 250
B.1.8. Running the model 250
B.2. Running the model in non-interactive mode 252

References....254

List of Tables

Table 2.1. Modeled Processes 13

Table 2.2. Major state variables and their symbols 15

Table 2.3. Supported land use types and their representations 18

Table 2.4. Simplied decision table for the atmospheric transmittance τ from [*Spokas and Forcella*, 2006] 25

Table 2.5. Soil Parameters for the test case: Infiltration into very dry soils 103

Table 2.6. Soil Parameters for the test case: Vauclin experiment 108

Table 3.1. Climatic Data sources and data availability. The dates in YYYYMMDD format are the dates when a station starts to have records; Frac is the fraction of the RCR watershed that is controlled by this weather station; PRCP PCT is the percentage of valid precipitation records from 1998 to 2007 119

Table 3.2. Land use percentages from the NLCD database and the MDNR databse 122

Table 3.3. SWAT Land use class percentages after apply thresholds. URLD: Low Intensity Urban; URMD: Medium Intensity Urban; WETF: Forested Wetland; URHD: High Intensity Urban; HAY: Hay/Forage Crops; AGRR: General Agriculture; FRSD: Deciduous Forest 138

Table 3.4. Performance metrics evaluating the model applied to the RCR watershed 142

Table 3.5. Metric of Groundwater heads comparison to wellogic data 147

Table 3.6. Annual Average Fluxes compared to SWAT 151

Table 4.1. Parameters in the Transient Storage model estimated for four different flow rates 205

Table 4.2. Comparison of the Relative Sizes of the Transient storage zones estimated using tracer data and independently using the ADCP data for Reach A 213

Table A.1. List of input data to the watershed model and format 232

Table A.2. An example transformation matrix from MDNR dataset to model classes 233

Table A.3. Summary of several watershed-scale hydrologic models 236

List of Figures

(Images in this dissertation are presented in color)

Figure 2.1. Definition sketch of the model. T: transpiration, p: precipitation, EO Evaporation from overland flow/stream, EB: evaporation from bare soil; I: infiltration; R: Recharge 14

Figure 2.2. Illustration of Arakawa-C grid 37

Figure 2.3. Relationship between K, θ and h at different parameter values with the van Genuchten formulation for a hypothetic soil type (a) at different α; (b): at different λ; (c) at different N. THE in the figure heading means θ. The base parameter for this comparison is α= 2.49(1/m), N = 1.507, $\theta_S = 0.43$, $\theta_r = 0.01$; $K_S = 0.175$ (m/day); λ = -0.14. Unit in the figure is the same as the base parameter set 49

Figure 2.4. sketch of the Vauclin 1979 test problem. HW(x) is the water table location at x, z is measured 0 at the bottom. 65

Figure 2.5. Illustration of Assumptions 1 and 2 66

Figure 2.6. h(x,z) as a function of z and DR values 70

Figure 2.7. sketch of a river cell in the model. (a). cross-sectional view, (b) plane view 82

Figure 2.8. Calculation of River/Land exchange 83

Figure 2.9. Discrepancies between NED and DEM when aggregated into the same grid 87

Figure 2.10. the river bed elevations estimated based on the DEM and NED as compared to the groundwater head 90

Figure 2.11. Rainfall pattern of the Midwest from 92

Figure 2.12. Program Flow Chart for the proposed model 94

Figure 2.13. Outflow hydrograph of the inclined plane test problem compared to the analytical solution 97

Figure 2.14. Sketch of the V-Catchment test problem 99

Figure 2.15. Solution to the V-Catchment test problem: Comparison of River outflow hydrograph with analytical solution and the finite element (FE) solution reported in [DiGiammarco]. (a) River outflow (b) Plane side outflow 100

Figure 2.16. Simulated channel flow vs measured data by [*Irfan*, 2002]. Red circle: measured data; blue solid line: Simulated 102

Figure 2.17. Infiltration into very dry soil test problem at t = 6hr. circle: Analytical solution; solid line: Δx= 0.075cm; dashed line: Δx= 0.25cm; dashed line with cross: Δx= 0.6cm. 104

Figure 2.18. Pumping near impervious wall test problem 106

Figure 2.19. Solution to the pumping near impervious wall test problem: transient comparison; 107

Figure 2.20. Comparison of the proposed model with experimental results from [*Vauclin et al.*, 1979] 109

Figure 2.21. Comparison of soil moisture profile obtained using the approach in the proposed and traditional method, θ_S=0.3, EBC=Equation Boundary Condition (Present approach); DBC=Dirichlet Boundary Condition (Conventional approach) 111

Figure 3.1. Location of the Red Cedar River Watershed 116

Figure 3.2. Elevation map of the Red Cedar River watershed 117

Figure 3.3. Locations of weather stations used for climatic input data 118

Figure 3.4. Basin-average Annual precipitation for the years from 1988 to 2008 (mm) 119

Figure 3.5. NLCD Land use/Land cover map for the RCR watershed 121

Figure 3.6. River system of the RCR watershed with USGS flow gages 123

Figure 3.7. Geology maps of the RCR watershed. a. Land Systems; b. Bedrock Geology 125

Figure 3.8. Soil property maps of the RCR watershed as reported in the STATSGO database. a. Sand percent; b. Soil Available Water Capacity; c. Saturated Hydraulic Conductivity (K_{sat}) 127

Figure 3.9. Watersheds delineated for the SWAT2005 model using the ArcSWAT GIS interface 138

Figure 3.10. Comparison of the observed and simulated daily flow at USGS gage 04112500 at East Lansing. Upper: Entire period from 1998/09/01 to 2005/12/31, Middle and lower: Close-up look of the hydrograph 141

Figure 3.11. Comparison of the observed and simulated daily flow at USGS gage 04111379 at Williamston 142

Figure 3.12. Comparison of the simulated streamflow from PAWS and SWAT and observed USGS streamflow data. Upper: from 1999 to 2005, Lower: Close-up look of water year 2001 144

Figure 3.13. Comparison of the simulated streamflow from PAWS and SWAT and observed USGS streamflow data in log-scale 146

Figure 3.14. (a) 5 year averaged simulated groundwater, (b) observed heads interpolated from wellogic database, (c) initial groundwater heads used to start the model 148

Figure 3.15. Observed vs simulated groundwater heads. (a) Observed data vs 5 year averaged value, (b) Observed data vs the rough initial guess. X-axis is the observed data obtained from Kriging interpolation of the wellogic data 150

Figure 3.16. Pie chart showing the comparison of the average annual fluxes as a percentage of total precipitation between PAWS and SWAT 152

Figure 3.17. Hydrograph partitioning into streamflow generation mechanisms: Infiltration excess, saturation excess and groundwater contribution. X-axis are dates in YY/MM/DD format. SatE: Saturation Excess, InfE: Infiltration Excess, Qgc: Groundwater contribution, Qout: Red Cedar River outflow 154

Figure 3.18. Temporal dynamics of fluxes and state variables. Upper: Monthly mean fluxes (mm). Lower: Daily fluxes (close up of 2004). Inf: Infiltration; Qoc: Overland contribution to channel; Qgc: Groundwater contribution to channel; Rchrg: Recharge; InfE: Infiltration Excess; SatE: Saturation Excess; Dperc: Deep Percolation 155

Figure 3.19. Basin average state variables (a) Soil Water Content (change from initial state), (b) Snow Water Equivalent. X-axis marks the beginning of each year (YY). 157

Figure 3.20. Daily fluxes of recharge and groundwater baseflow. (a) Rising limb of baseflow, (b) falling limb of baseflow. Again, dates are in YY/MM/DD format. Unit of the figure is m 159

Figure 3.21. Average annual runoff generation (mm) (a) Saturation Excess (b) Topographic Index 163

Figure 3.22. Average Annual Infiltration Excess 164

Figure 3.23. Annual Infiltration (a) and recharge map (b) 165

Figure 3.24. Annual ET (a) and EvapG (b) map. EvapG is evaporation from the ground. 166

Figure 4.1. Red Cedar River and the sampling locations 180

Figure 4.2. Red Cedar River and the study region showing the sampling locations of the Grand River tracer study 182

Figure 4.3. Comparison of the numerical solution with the analytical solution of De Smedt et al [2005] 191

Figure 4.4. Schematic diagram illustrating the concept of discrete wavelet decomposition. S denotes the original signal (the 2D image of the velocity field measured by ADCP). L and H denote the low-frequency approximations and high-frequency details, respectively. Suffixes denote wavelet scale levels 195

Figure 4.5. Comparison of observed and simulated tracer concentrations for four slug releases conducted during summer 2002. (a) Q=2.49 m^3/s. Sampling locations at x=0.87km (Bogue Street Bridge) (b) Q = 14.41 m^3/s, sampling locations are, namely, Farm Lane Bridge (x = 1.40km). Kellogg Bridge (x = 3.10km) and the Kalamazoo Bridge (x=5.08km), respectively. (c) Q = 16.82km^3/s, (d) Q= 19.06m^3/s 203

Figure 4.6. Observed mean velocity fields at two different stations in reach A in the Red Cedar River obtained using a 1200 kHz ADCP (a and b) Q = 5.49m^3/s (8 November 2003). (c and d) Q = 19.89 m^3/s (19 March 2006). Note that during the high event the adjacent low lying areas near the banks were filled with relatively stagnant water which were not available during low discharge 209

Figure 4.7. Multiresolution wavelet approximations for the images shown in Figure 4.6. After completing the wavelet analysis, the low-frequency content at wavelet levels 1 and 2 (denoted by L1 and L2) was plotted in the physical space. (a and b) Hagadorn Bridge. (c and d) Farm Lane Bridge. 211

Figure 4.8. Sample ADCP transect data used for computing the longitudinal dispersion coefficients. 218

Figure 4.9. Vertical velocity profiles in the Ohio and St. Clair Rivers showing the effect of fitting a power law (black lines) and logarithmic profiles (red lines). The raw data from the ADCP is shown using symbols..220

Figure 4.10. Effects of smoothing (logarithmic, power-law or no-smoothing) and velocity projection methods (rotating velocities in the streamwise direction or projecting them in a direction normal to the transect track) on the dispersion estimates from ADCP 225

Figure 4.11. Comparisons between ADCP and tracer estimates of the dispersion coefficient: (a) Box plots denote the variability in D estimated using the ADCP method within a given river reach. Tracer estimates based on the ADE and the transient storage modeling are shown using different symbols. (b) Comparisons between ADCP and tracer estimates plotted on top of similar results reported in the literature. 227

Figure 4.12. Comparison of ADCP estimates of the dispersion coefficient with tracer estimates and results from empirical relations. (b) Box plots showing the deviation between estimates of D based on the ADCP method and empirical relations 228

Figure 4.13. Depth-averaged longitudinal (u) velocity profiles plotted as a function of the transverse coordinate y 229

Figure B.1. GUI after the model is started..240

Figure B.2. Data GUI. (a) before loading data (b) after loading data files 242

Figure B.3. GUI after the model is started..244

Figure B.4. Discretization GUI 246

Figure B.5. Solvers GUI. (a) after it is opened (b) after dt file is loaded 248

Figure B.6. Loading solvers data into solvers functions GUI....................................249

Chapter 1. Background and literature review

1.1. Motivation for a new hydrologic model

In the twenty-first century the management of water resources has gained unprecedented scientific as well as strategic importance. On one hand, the growing human population raises larger and larger demand for usable and available water resources. Satisfying such a need can be a serious challenge in many parts of the world. On the other hand, humans are equipped with the ability greater than ever to harness water in nature for their own end, and they often opt to do so. These interventions have led to some fundamental modifications to the hydrosphere. In the past century we have essentially transformed the Earth's ecosystem, even on the grandest scale, into a human-dominated one [*Vitousek et al.*, 1997]. While many of the large-scale hydraulic projects proved essential for the economic development and well-being of people and their positive impacts should not be downplayed or understated, the hydrological, ecological and environmental consequences are usually not fully understood. Today the Aral Sea in Central Asia, once the fourth largest inland lake in the world, has been desiccated to 1/10 of its original size with its entire ecological system destroyed. Studies reveal that this overwhelmingly owns to the diversions of water from the sea's tributaries for irrigation expansion [*Cretaux et al.*, 2005; *Micklin*, 1986; 1988]. The impacts of such interventions are so far reaching, profound and irreversible that all aspects of the hydrologic system need to be carefully assessed.

The surface-subsurface water interactions are often overlooked. Due to the slow response of the subsurface water system, impacts of large-scale projects - benign, malign or neutral - can take a long time to manifest themselves. The Aswan Dam in Egypt and its reservoir, Lake Nasser, still witness newly discovered impacts that could not be predicted 30 years ago when the dam was constructed. Some have observed reduced carbon emission and more sustainable development [*Prasad et al.*, 2001; *Strzepek et al.*, 2008], while others reported erosion, salinization, and pollution that induced decline in agricultural productivity and loss of land and coastal lagoons [*Stanley and Warne*, 1993].

These challenges are further complicated by the more and more pronounced trend of climate change. Climate change is expected to exacerbate current water stresses [*IPCC*, 2007]. Semi-arid regions and drought-affected areas are projected with high confidence to suffer decreased water resources due to climate change. For example, the Lakes Mead and Powell created by the Hoover Dam, the lifeblood of U.S. southwest, are estimated to dry up by 2021 if no changes to the current water usage are made [*Barnett and Pierce*, 2008; 2009]. The study has attributed this result to global warming and current operating conditions.

Understanding water fluxes is also important from a human health perspective. A variety of pollutants including chemical and biological agents pose threats to human and ecosystem health [*U.S.EPA*, 2000]. A well-known USGS study that involved 139 streams in 30 US states between 1999 and 2000 found pharmaceuticals, hormones and a number of emerging contaminants in 80 % of the streams sampled [*Kolpin et al.*, 2002]. Similarly a majority of rivers sampled in Michigan tested positive for the presence of viable enteric viruses [*Jenkins et al.*, 2005]. Therefore understanding factors that influence the fate and transport of contaminants in rivers and streams is extremely important from the point of predicting human health risks and protecting the public. The flow generation process governs the source and form of contamination. Pollutants can reach streams via point source discharge, non-point source (overland flow), or subsurface seepage [*Jamieson et al.*, 2004].

All these challenges call for resolute but well-informed decision making. Sound decision making is best aided by good understanding of the complex and interrelated hydrologic systems. Better understanding comes most elucidated with the ability to explicitly describe the hydrologic processes in space and time. This is where a reliable and verifiable process-based hydrologic model with good predictive power could play an important role.

The present study attempts to create a process-based hydrologic model that finds a good balance between process modeling and applicability. This model should link processes that occur at different scales and illustrate the interactions among the hydrologic domains, including, surface water, soil water, groundwater, river, canopy and atmosphere. It is hoped that this model will help address some of the challenges posed by human intervention as well as climate change.

1.2. Review of hydrologic models

Many hydrologic models have been historically developed to study different hydrologic systems. Several models that are recently cited are reviewed in Table A.3. Models are often developed with specific scientific objectives in mind. Thus each model has its own strengths in some areas and may be inadequate in some other areas. For example, SWAT is designed as a long term water balance and non-point source pollution simulator [*Arnold and Allen*, 1996; *Arnold and Fohrer*, 2005]. Thus simulating short term point source pollution with SWAT may not be advantageous. Before developing our own model it is helpful to understand the current state of research.

Broadly, there are two categories of models: conceptual and physically-based (or mechanistic). Generally, in conceptual hydrologic models (CHM), the modeler forms hypotheses, either from experience or his own perceptions, about the hydrologic processes and proposes mathematical formulations to represent these processes with sometimes strong simplifying assumptions. CHM are often based on empirical relations and conceptual state variables that cannot be always measured. Physically-based hydrologic models (PBHM), on the other hand, are derived deductively from established physical principles with appropriate assumptions and physically meaningful/measurable parameters [*Beven*, 2002]. Historically there have been heated discussions about the relative advantages and disadvantages of the two approaches. In 1966, the paper [*Freeze and Harlan*, 1969] laid out a blueprint for physically based hydrologic modeling, writing out the equations for different flow processes and the linkages via common boundary conditions. The development of such a physically based model in the years followed, however, has been limited by the computational power, data availability, understanding of the complex hydrologic system and, to a lesser extent, the accumulation of mathematical technique. [*Beven*, 2002] has challenged that blueprint and provided an alternative blueprint to hydrologic modeling that is based on lumped conceptual models. These models do not use the process theory to build a model structure *a priori*, but rely on observed data to define an appropriate model structure (described by Beven as 'hypothesis testing'). He also stressed on the importance of quantifying the uncertainty of the models. With the advancement of computer power, Geographic Information System (GIS) and readily available databases, recently published models lean more and more toward the physically-based approach [*Karvonen et al.*, 1999]. However, to date there does not seem to be a conclusion to that debate and the two schools of models continue to be created and advanced.

In fact, both types of models have their own advantages and disadvantages. Some of the notable conceptual models include TOPMODEL [*Beven and Kirkby*, 1979], DLBRM [*Croley and He*, 2005], Sacramento Soil Moisture Accounting Model (SAC-SMA) [*Burnash*, 1995], HEC-HMS [*HEC*, 2000], VIC-3L [*Liang and Xie*, 2001; *Reed et al.*, 2004], etc (also see [*Borah and Bera*, 2003; *Reed et al.*, 2004]).

Usually, conceptual models require less physical input as its components are idealized. The models tend to be structurally simple, computationally inexpensive and more easily operational. However, they need long term monitoring data to calibrate. The parameters generally cannot be applied for ungauged watersheds. Moreover, the conceptualization process blurs the underlying dynamics and extension of conceptual models beyond the range of calibration is questionable [*Beven*, 1985]. It is not rare that a CHM that describes completely different physics from the study region and still get fair results after optimization, but the physics can be far from reality. In fact, a large number of papers have been published to quantify the uncertainties associated with the conceptual models (e.g. see [*Beven*, 2006; *Beven and Binley*, 1992; *van Griensven and Bauwens*, 2003; *Vrugt et al.*, 2005; *Yang et al.*, 2008], etc). The large uncertainties undermine the reliability of the model, especially since estimated parameters cannot be interpreted.

A review of current physically-based models, on the other hand, reveals that there is still much room for improvement, especially in achieving the right balance between process descriptions and computational demands. PBHM generally tend to be data intensive and computationally expensive and thus their applicability tends to be limited. The reported results from PBHMs are often simulations for small areas during short periods of time. The published comparisons with observations indicate that these models, as commented by [*Ivanov et al.*, 2004a], are yet to emerge as the preferred tool for prediction and analysis. Some well-documented physically based hydrologic models include the MIKESHE model [*DHI*, 2001], Soil & Water Assessment Tool (SWAT) [*Arnold and Fohrer*, 2005], Water and Energy Transfer Process (WEP) model [*Jia et al.*, 2001; *Jia et al.*, 2006], CASC2D/GSSHA [*Downer and Ogden*, 2004b] and tRIBS [*Ivanov et al.*, 2004b], etc. We must mention here that some regard SWAT as a semi-physically based model since considerable amount of empiricism is included in the model, as highlighted in section 3.2. Although almost all PBHMs still carry some level of empiricism, the models can be constrained much better by real data since most parameters are physically based.

Here we review six watershed-scale and one macro-scale (VIC) PBHM. Table A.3 in the appendix lists, in alphabetical order, seven models that are currently being published and cited in the literature. This is by no means a complete list of hydrologic models that are of interest, but it does give a representative coverage. In its completeness, watershed hydrologic models should incorporate several flow domains that cover various flow paths possible after rain drops to the ground: overland flow, unsaturated subsurface flow (the vadose zone model), saturated subsurface flow (groundwater flow) and channel flow. Water also exists in canopy interception, snowpack, biomass and depression storage. Besides the flow domain, one of the processes of utmost importance is the evapotranspiration, which on average is estimated to take out 70% of the rainfall in North America [*Jensen et al.*, 1990]. In order to account for the seasonal dynamics of a watershed, a reasonable vegetation growth module should also be included. A sufficient PBHM should contain all of the relevant processes. Table A.3 also lists the spatial discretization method and solution schemes to the flow problems.

The InHM [*VanderKwaak and Loague*, 2001] is a processe-based model that integrates overland flow with the mixed form three-dimensional Richards equation for the subsurface. InHM simulations in showed that both the Horton and Dunne overland flow mechanisms can be important streamflow generation processes. The InHM simulations also suggested that accurate accounting of soil water storage can be as important as exhaustive characterization of spatial variations in near-surface permeability. However, the InHM is very computationally demanding such that it can only applied at the plot-scale.

The GSSHA model, developed for the Army Corps of Engineers, evolved from CASC2D [*Downer and Ogden*, 2004b]. CASC2D uses Green and Ampt infiltration method [*Green and Ampt*, 1911] with redistribution (GAR) for moisture accounting. However, it has been found that the processes modeled in the CASC2D model cannot adequately describe watersheds where saturation excess is important [*Downer et al.*, 2002]. This research highlights the importance of incorporating the correct processes in physically-based models and applying hydrologic models in the settings where they are applicable. In GSSHA, the subsurface flow component is improved by coupling the Richards equation with the groundwater flow equation. The height of the water table is provided as the lower boundary condition to the Richards equation. A constantly changing discretization is used to cope with the rise and fall of the groundwater table. As will be shown later in section 2.4.1, this coupling approach will cause soil moisture profile to be inconsistent with the location of the groundwater table. Moreover, it is difficult to evaluate the potential of GSSHA or CASC2D as a large-scale long term analysis tool as most of the published results are for small catchment areas (20 km^2 and 3.64 km^2 in [*Downer and Ogden*, 2004a] and 3 km^2 in [*Downer and Ogden*, 2004b]) in a short time frame (<200 days). The comparison is also limited to streamflow measurements.

The WEP was originally developed by [*Jia et al.*, 2001] and applied in Japan to a small size urban watershed. Later, it was expanded into a large scale version WEP-L and applied to a very large basin — the Yellow River Basin in China [*Jia et al.*, 2006]. The WEP model uses the mosaic approach which allows sub-grid heterogeneity to be parameterized. Three layers of soils are defined in the model and water is allowed to move only vertically in the soil. The soil layers are connected to the unconfined aquifer which is modeled using the Boussinesq equation. The proposed model shares some similarities in the structure with the WEP model. The results reported for both the Japanese watershed and the Yellow River basin are promising. Unfortunately, this model is not available so direct comparison is not possible. It is felt that applications to some medium-sized watersheds with different hydrologic settings and geologic configurations may help further illustrate the general applicability of the WEP model.

[*Panday and Huyakorn*, 2004] details the processes included in the MODHMS model, which is a commercial package. Overland flow exchange with channel is modeled as flow over a rectangular weir. A similar approach is used by the proposed model. The subsurface is modeled with a variably saturated 3D flow model. 3D modeling is generally considered to be computationally too expensive for watershed-scale modeling. Since in this paper no application to real watershed is presented, we cannot assess the applicability of this approach.

tRIBs uses a unique Triangular Irregular Network (TIN)-based discretization and thus is thought to be capable of more properly capturing the spatial heterogeneity in topography, land use or soils. However, the mapping of data into the triangular element ('Voronoi Regions' in the original paper) requires special mesh generation and processing procedures, which can be non-trivial for other modelers. The tRIBs model has been applied to several medium-sized watersheds [*Ivanov et al.*, 2004a] and obtained reasonable results. More recently, a sophisticated weather generator [*Ivanov et al.*, 2004b] and a vegetation growth module have been added to the model. The impact of topographic controls has been studied in detail, prompting new directions of catchment hydrology [*Ivanov et al.*, 2008a; b].

The MIKE-SHE model is a commercial hydrologic model that contains comprehensive modules. The model was developed in Europe. It has a 1D Richards equation for the unsaturated zone linked to the groundwater aquifers. 3D groundwater flow is simulated. Procedurally, 3D saturated flow modeling is not very much different from 2D modeling. The difficulty is mainly with the unsaturated vadose zone. To be able to simulate transport process in a 8.7 km^2 watershed, [*Thompson et al.*, 2004] created a MIKESHE model with a 30×30m grid.

The PAWS model proposed in this dissertation attempts to improve the applicability of PBHM by finding the right balance between processes, data and computation. The model follows the Freeze and Harlan blueprint in that we believe each component of the model should be verifiable by itself. PAWS uses accurate and computationally efficient schemes to solve the physically-based governing equations. The descriptions of the unsaturated soil water domain by the Richards equation allow dynamic interactions among different components. In the next chapter, we will develop the mathematical details.

Chapter 2. Development of the Hydrologic Model: Mathematical Bases and Test Cases

Mathematical models are, by definition, simplifications and abstractions of the real world. The model designer chooses, using his own judgment, the processes that are important to certain objectives and endpoints, and builds mathematical representations of the processes. For the hydrologic modeling in our context, the processes involved start at the moment precipitation reaches the ground (or canopy) and ends when water exits the system via channel, overland flow boundary, groundwater flow boundary or evapotranspiration.

In this chapter we discuss the mathematical basis of PAWS. We lay out our general conceptual representation of the watershed and the modeled processes in the first section. We then describe the governing equations for each hydrologic unit and how the 3D physical space of the watershed is discretized into computational grids in the next section. In the third section the numerical solution schemes to solve the equations or general calculation steps are described. In the last section, we compare our numerical solution for each component with the analytical solution for one or more test cases to ensure the accuracy of the numerical code. Where applicable, experimental data are compared with numerical solutions to test the model performance.

2.1. Model Processes

The processes modeled in PAWS are graphically presented in Figure 2.1 and summarized in Table 2.1. The eight compartments where most calculations take place are, respectively, surface ponding layer, canopy storage layer, impervious cover storage layer, overland flow layer, snowpack, soil moisture, groundwater aquifers and channels. The major state variables are summarized in Table 2.2.

Table 2.1. Modeled Processes

Processes	**Governing Equations**
Snowfall Accumulation and melting	Mass and energy balance (UEB) [*Luce and Tarboton*, 2004]
Canopy interception	Bucket model, storage capacity related to Leaf Area Index (LAI)
Depression Storage	Specific depth
Runoff	Manning's formula + Kinematic wave formulation+Coupled to Richards equation
Infiltration/Exfiltration	Coupled to Richards equation
Overland flow	2D Diffusive wave equation
Overland/Channel Exchange	Weir formulation
Channel network	Dynamic wave or diffusive wave
Evapotranspiration	Penman Monteith + Root extraction
Soil Moisture	Richards equation
Lateral Groundwater Flow	quasi-3D
Recharge/Discharge (Vadose zone/Groundwater interaction)	Non-iterative coupling inside Richards equation
Stream/Groundwater interaction	Conductance/Leakance
Vegetation Growth	Simplified Growth Cycle

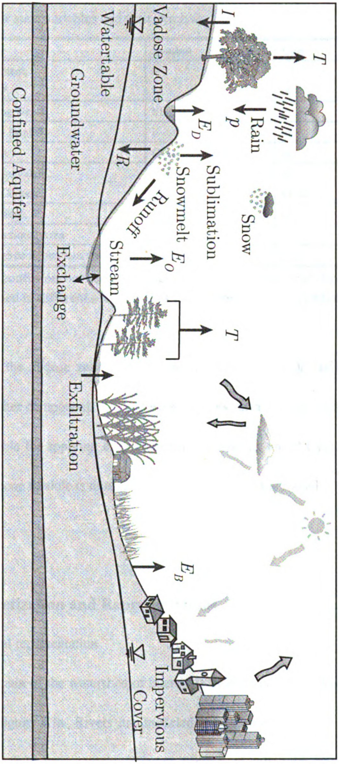


Figure 2.1. Definition sketch of the model. T: transpiration, p: precipitation, EO Evaporation from overland flow/stream, EB: evaporation from bare soil; I: infiltration; R: Recharge

Table 2.2. Major state variables and their symbols

State Variable	Symbol	Unit .
Soil Water Content, Pressure Head*	θ h	(-), m
Surface Ponding Depth	hl	m
Overland Flow Depth*	h	m
Canopy Storage	CS	m
Snow Water Equivalent, Snow Energy Content	SWE, U	m, kJ/m2
Groundwater Head	H	m
River Cross Sectional Area	A	m2
Ponding storage on impervious cover	hi	m

*** There is no confusion between soil pressure head and overland flow depth since they are described in different sections, thus they share the use of symbol h.**

It is clear that the vadose zone plays a central role in the model as it connects to almost every other component. The vadose zone links surface water with groundwater and is responsible for applying evapotranspiration. Having an efficient, flexible, and robust vadose zone module is thus critical to the success of the model.

2.2. Discretization and Representations

2.2.1. Horizontal representation

The spatial domain of the watershed of interest is discretized into a structured grid as illustrated in Figure 2.1a. Rivers are modeled as separate objects in a network of channels. Each river may contain different number of cells with possibly variable spatial step sizes. Within each river cell, the river characteristics such as river width, bottom elevation, water depth, are assumed to be uniform. The intersecting length of river cells with land cells are pre-computed and stored in sparse matrices. These matrices are used to convert the exchange fluxes from land grid to river cell grid as will be explained in 2.4.3. This flexible strategy allows river and land discretizations to be independent and the coupling can be done regardless.

Within each land cell, sub-grid heterogeneity is modeled using a method similar to the mosaic approach reported in [*Jia et al.*, 2001]. Depending on the grid size, a cell may possess a mixture of land use/land cover types. Since different land use/land cover classes respond differently to hydrologic events, the treatment of sub-cell heterogeneity is important. However, nationally-prepared, readily available datasets contain too many plant types for a model to simulate. A cell may contain many classes each taking up only a small fraction of the cell area. Moreover, some classes in the database are a complex group of several plant species. For example, low intensity urban commonly has non-negligible amount of soil-vegetation cover. Considering these factors, land use data are re-classified into model classes which are represented by certain plant types. Table 2.3 summarizes the representative plant types (RPT) currently modeled. Table A.2 provides an example transformation matrix from the land use classification provided by the Michigan Department of Natural Resources [*MDNR*, 2010b] to the model classes. For instance, the first row of the table which reads 'low intensity urban' is divided into 40% of impervious, 20% of Deciduous forest (modeled by Oak), and 40% of grass cover. The urban information is obtained from [*NRCS*, 1986] Then, the fractional areas of each RPT inside a cell are summed up. Then the number of land use classes that are going to be modeled (nRPT) should be determined. Considerations include cell sizes, the complexity of the landscape, and the computational cost. Smaller cell size would need correspondingly smaller nRPT. Further, the first nRPT largest RPTs are selected as the model classes in the cell. Other types would be removed and have their areas assigned to the selected RPTs. Rather than simply allocating the areas that belong to the removed types proportional to the areas of the selected RPTs, they are assigned to the RPT in the same group (as in Table 2.3). This processing logic attempts to represent land use classes as close to the original land use as possible, while reducing the demand for computational resources.

Currently, five characteristics completely characterize a certain RPT: crop coefficient for evapotranspiration (K_c), canopy height (h_c), rooting depth (Root), Leaf Area Index (LAI) and growth periods (T_g). These characteristics are found from literature [*Breuer et al.*, 2003; *FAO*, 1998; *Neitsch et al.*, 2005].

Table 2.3. Supported land use types and their representations

Land use Class	Representation	Group
Water	Water	Water
Urban or Industrial	Impervious	Impervious
Evergreen Forest	Pine	Tall Vegetation
Deciduous Forest	Oak	Tall Vegetation
Bush and shrub	Shrub	Short Vegetation
Herbaceous	Grass	Short Vegetation
Row Crops	Corn	Agricutural
Bare Soil	Bare	Bare
Forage Crops	Alfalfa	Agricultural

The soils data may come from STATSGO or SSURGO in the United States. In each cell, as we only model one soil column, the most dominant soil type is used. For the elevation, either the mean or the median value within a cell is taken as the elevation for this cell, if the grid size is larger than the resolution of the digital elevation dataset provided.

2.2.2. Vertical representation

The model structure in the vertical direction is given in Figure 2.1b. Water on the ground is separated into two domains, the ponding water layer and the flow domain. The ponding domain will be subject to infiltration and evaporation whereas only water in the flow domain can move from one cell to another (including river cells). Runoff from ponding to flow domain can be calculated using several different methods described in section 2.4.2. Depending on the climatic conditions, a layer of snow, quantified by the Snow Water Equivalent (SWE) and snow cover fraction

(Afrac), may also exist on the ground. After taking out the canopy interceptions, rainfall is joined by snowmelt to add to the ponding layer. Except for the nRPT land use classes on the surface as described earlier, we model one soil column for each cell. The soil column is responsible for computing infiltration from surface ground (the ponding domain), soil evaporation, root extraction and percolation into unconfined aquifer. Water is assumed to move only vertically in the soil column but once in the aquifers it can move laterally. The unconfined aquifer has a thickness equal to its water depth. It can exchange water and energy with the soil column. Water can percolate further down to deeper aquifers via a layer of aquitard and contribute to or gain from river water through river bed materials. Pumping activities can also directly extract water from the aquifer layers.

2.3. Hydrologic Processes and Solution to each flow domain

The four flow domains that the model considers are overland flow, channel flow, soil moisture and saturated groundwater flow. They will be described one after another in this section. The other important component that will be discussed in detail is evapotranspiration.

The mass balance equation for the ponding layer on the ground is written as:

$$\frac{\partial h_1}{\partial t} = P - CS_{new} + SNOM - E_g - Inf - F_g \tag{2.1}$$

Where h_l is the water depth in the surface ponding layer (m), $P\text{-}CS_{new}$ is the precipitation (m/day) reaching the ground after subtracting canopy storage, *SNOM* is the rate of snowmelt (m/day), E_g is the rate of evaporation on the ground (m/day), *Inf* denotes infiltration (m/day), and F_g is the runoff to the overland flow domain.

2.3.1. Evapotranspiration

Evapotranspiration is the collective term of evaporation and transpiration and is a major component of the hydrologic cycle. In the proposed model, we apply the reference evapotranspiration (RET)/adjustment approach to calculate ET. First the reference ET is computed using climate input data for a type of reference plant, and then it is adjusted for different plants and soil moisture conditions. Multiple definitions of potential evapotranspiration and reference evapotranspiration exist in the literature. In particular, the term 'potential evapotranspiration' has caused a great deal of confusion. In this dissertation, we follow the definitions of 'reference ET' given in [*FAO*, 1998]: The evapotranspiration rate from a reference surface, not short of water.

The reference Evapotranspiration is computed using the Penman-Monteith (PM) equation:

$$\lambda ET = \frac{\Delta\left(R_n - G\right) + \rho c_p\left(e_z^0 - e_z\right) / r_a}{\Delta + \gamma\left(1 + r_c / r_a\right)} \tag{2.2}$$

where ET is the reference evapotranspiration on a day (mm/day) for a given reference plant, λ is the latent heat of vaporization (MJ/kg), R_n is the net radiation (MJ/day), G is the ground heat flux (MJ/day), Δ is slope of the saturation vapor pressure-temperature curve (kPa/°C), ρ is the air density (kg/m^3), c_p is the specific heat of moisture at constant pressure (1.013×10^3 MJ/kg/°C), e_z^0 is the saturation vapor pressure at height z (kPa), e_z is the actual vapor pressure at height z (kPa), r_a is the aerodynamic resistance (s/m), r_c is the canopy resistance (s/m), and γ is the psychrometric constant (kPa/°C). λ, γ, Δ and e_z^0 are calculated as

$$\begin{aligned}
\lambda &= 2.501 - 2.361\times10^{-3}T_a \\
\gamma &= \frac{c_p P_a}{0.622\lambda} \\
e_a^0 &= \exp\left[\frac{16.78T_a - 116.9}{T_a + 237.3}\right] \\
\Delta &= \frac{4098 e_a^0}{\left(T_a + 237.3\right)^2}
\end{aligned} \tag{2.3}$$

where T_a is air temperature in Celsius (°C), P_a is air pressure (kPa) calculated for a given elevation Elev (m):

$$P_a = 101.3 - 0.01152 Elev + 0.544 \times 10^{-6} Elev \tag{2.4}$$

The actual vapor pressure is calculated from relative humidity, which is often provided in the form of dew point temperature from climatic data sources:

$$\begin{aligned} e &= RH \cdot e_a^0 \\ RH &= \frac{\exp\left[\dfrac{16.78 T_{dew} - 116.9}{T_{dew} + 237.3}\right]}{e_a^0} \end{aligned} \tag{2.5}$$

where RH is relative humidity and T_{dew} is dew point temperature (°C).Given the maximum and minimum temperatures T_{mx} (°C) and T_{mn} (°C) in a day, the sub-daily temperature is computed [*Campbell*, 1985], assuming highest temperature at 3PM local time:

$$T = T_{av} + \frac{T_{mx} - T_{mn}}{2} \cos\left(0.2618\left(t - 15\right)\right) \tag{2.6}$$

in which hr is the hour of the day (local time), T_{av} is average temperature of the day (°C).

Solar radiation data are normally scarce but can be adequately estimated from sun-earth relationships and other available climatic data. Thus a solar radiation calculation scheme was developed to fill the data gaps. The net radiation, R_n, is computed at the weather station sites and applied to the model domain using nearest neighbor interpolation. The net radiation is the sum of absorbed incoming shortwave radiation, incoming and outgoing long-wave radiation:

$$R_n = (1-\alpha)R_s + H_{li} + H_{le} \tag{2.7}$$

Where R_n is the net radiation (MJ/day), α is the albedo, Rs is the incoming shortwave radiation, H_{li} is the incoming long-wave radiation (MJ/day), H_{le} is the outgoing long-wave radiation (MJ/day). All terms are non-negative except for H_{le}.

The shortwave solar radiation at a given time is calculated by summing up the direct radiation (or beam radiation) and the diffuse radiation:

$$R_s = S_b + S_d \tag{2.8}$$

where S_b is the beam radiation and S_d is the diffuse radiation. S_b is calculated from the following formula:

$$S_b = I_{toa}\tau_s = I_{SC}\cos\theta_z\tau_s \tag{2.9}$$

in which, I_{SC}=1367 w/m^2 is the solar constant, τ_s is the atmospheric correction factor, $I_{toa}=I_{SC}\cos\theta_z$ is the top of atmosphere solar irradiance corrected for incident angle, θ_z is the zenith angle between the sun and a given surface on earth:

$$\cos\theta_z = \sin\phi\sin\delta + \cos\phi\cos\delta\cos\left[\frac{\pi\left(t-t_{sn}\right)}{12}\right] \tag{2.10}$$

where φ is the latitude of the site, δ is the solar declination angle, t is the time of the day and t_{sn} is the local solar noon time of the day. The solar declination angle can be calculated from[*Campbell and Norman*, 1998]:

$$\sin\delta = 0.3979\sin\left[4.869 + 0.0172J + 0.03345\sin(6.2238 + 0.0172J)\right] \tag{2.11}$$

where J is the Julian day of the year. The term 0.03345sin(6.2238+0.0172J) accounts for the eccentricity correction factor for the Earth's orbit.

Much research has been done to estimate the atmospheric correction factor τ_S. The methods range from complex models that need comprehensive climate input to simple empirical relations that use only temperature. To create an efficient method for estimating hourly incoming solar insolation based on limited climatic input (namely, daily precipitation, maximum and minimum temperature), [*Spokas and Forcella*, 2006] proposed the following simplified formula for τ_S:

$$\tau_s = \tau^m \tag{2.12}$$

where τ is the atmospheric transmittance given in Table 2.4. m is the optical mass number:

$$m = \frac{P_a}{101.3 \cos\theta_z} \tag{2.13}$$

in which, P_a is the atmospheric pressure (kPa) at the site.

The incoming long-wave radiation is emitted from particles and gases in the air. It is estimated as:

$$H_{li} = \varepsilon_a \sigma T_k^4 \tag{2.14}$$

where ε_a is the air emittance, σ the Stephen-Boltzman constant (4.903×10-9 $\mathrm{MJm^{-2}K^{-4}d^{-1}}$), and T_K is air temperature in Kelvin. The air emissivity is evaluated considering cloud cover and using the Satterlund parameterization [*Luce and Tarboton*, 2004; *Satterlund*, 1979]:

$$\varepsilon_a = C_f + \left(1 - C_f\right)1.08\left[1 - \exp\left(-\left(10e_a\right)^{\frac{T_a}{2016}}\right)\right] \tag{2.15}$$

In which, e_a is the vapor pressure in the air (kPa), Cf is the Bristow and Campbell transmission factor, which measures how close the air transmittance is to clear sky conditions:

$$C_f = 1 - \frac{\tau}{\tau_{mx}} \tag{2.16}$$

In which τ and τ_{mx} are actual and maximum air transmittance. They are provided by the solar radiation algorithm described above, using Table 2.4 proposed by [*Spokas and Forcella*, 2006].

Table 2.4. Simplied decision table for the atmospheric transmittance τ from [*Spokas and Forcella*, 2006]

Conditions	**Value of τ**
No precipitation and Δ T>10C	τ =0.70
No precipitation on present day, but precipitation fell the previous day	τ =0.60
Precipitation occurring on present day	τ =0.40
Precipitation today and also the previous day	τ =0.30

If it is needed, the topographic effect may also be considered as in [*Piedallu and Gegout*, 2007].

The diffuse radiation in Eq. (2.8) can then be evaluated as [*Campbell and Norman*, 1998]:

$$S_d = 0.3\left(1 - \tau^{m}\right)I_{toa} \tag{2.17}$$

which assumes 30% of scattered/absorbed solar radiation by the atmosphere is then reflected to the ground. The above model is easy to use due to its simplicity and limited data demands .

The outgoing long-wave radiation is calculated as:

$$H_{le} = f_c \varepsilon_s \sigma T_K^4 - H_{li} \tag{2.18}$$

in which, fc is a factor describing the effect of cloud, and ε_s is the emittance of ground surface, their calculations are from [*Neitsch et al.*, 2005],[*Jensen et al.*, 1990]:

$$\begin{aligned} f_c &= 0.9\frac{\tau}{\tau_{mx}} + 0.1 \\ \varepsilon_s &= -\left(0.34 - 0.139\sqrt{e_a}\right) \end{aligned} \tag{2.19}$$

This completes the estimation of net radiation. So far, the only missing terms from Eq. (2.2) are r_a, the aerodynamic resistance and r_c, the canopy resistance.

The reference ET calculated using alfalfa as the reference crop is used as the total ET demand. It is first used to evaporate any water in the canopy storage and on the ground. The ET demand on the impervious fraction of the cell is used to evaporate depression storage on the impervious fraction. This ET demand, if not depleted, is not used in any other way. The ET demand on the snow-covered fraction is also discarded as the energy is assumed to be used for the melt of snow, which is computed by other methods (reference for these equations):

$$\begin{aligned} dET &= rET_{alfalfa}\left(1 - f_{snow}\right)\left(1 - f_{imp}\right) \\ dET &= dET - E_c - E_s \end{aligned} \tag{2.20}$$

where dET is the remaining ET demand, $rET_{alfalfa}$ is the reference ET computed for alfalfa. f_{imp} is the fraction of impervious cover, f_{snow} is the fraction of snow-covered area, E_c is the evaporation from canopy storage and E_s is the evaporation of water on the ground surface:

$$\begin{aligned} E_c &= \max(dET, CS) \\ E_s &= \max(dET - E_c, h) \end{aligned} \tag{2.21}$$

where h is the water on the ground surface (m). The transpiration demand for each plant type is calculated by multiplying dET by the crop coefficient:

$$dTP_i = Kc_i \cdot dET \tag{2.22}$$

In which, dTP_i is the transpiration demand for the i-th plant type, Kc_i is the crop coefficient for the i-th plant type (alfalfa-based) and is a function of plant species and growth stages. Common Kc values can be found in literature studies, e.g. [*FAO*, 1998].

Then we distribute transpiration demand to the soil layers considering root distribution and soil moisture constraints. A uniform root zone distribution function is assumed for all plant types:

$$g_{i,j} = \frac{rt_j}{\sum_{j=1}^{rmx} rt_j} \tag{2.23}$$

where $g_{i,j}$ is the root zone distribution function for the i-th plant type in the j-th layer, rt_j is the root density in the j-th layer. Vegetation roots experience difficulty in extracting water when the soil moisture is low. The amount of water that roots can extract depends on the soil moisture and this relationship is called the root efficiency.

The root efficiency function is taken from [*Lai and Katul*, 2000]:

$$\eta(\theta) = \left(\frac{\theta - \theta_w}{\theta_s}\right)^{\Upsilon/(\theta-\theta_w)} \tag{2.24}$$

in which η is the root efficiency function, θ is the soil moisture content, θ_S is the saturated water content, θ_W is the wilting point and Υ is an empirical parameter. [*Braud et al.*, 2005] has examined this root efficiency function and found it performed well compared to a field-measured soybean dataset.

At this point, we can calculate the transpiration for the j-th soil layer:

$$TP_j = \sum_{i=1}^{nRPT} g_{i,j}\eta_j\delta_i \cdot dET \tag{2.25}$$

where j is the vertical soil layer number, TP_j is the transpiration in soil layer j, nRPT is the number of RPT in the current land cell and δ_i is the leaf-cover fraction of the i-th RPT. The leaf-cover fraction of the RPT is calculated using Beer Lambert law:

$$\delta_i = f_i \cdot \left(1 - \exp(-0.5LAI_i)\right) \tag{2.26}$$

in which, f_i is the land use fraction of the i-th RPT, LAI_i is the leaf area index of this RPT.

The evaporation, on the other hand, is applied to the area that is not covered by leaves:

$$E_j = \eta^e g_j^e \left(1 - \sum_{i=1}^{nRPT} \delta_i\right) dET \tag{2.27}$$

where j is the vertical soil layer number

$$g_j^e = \frac{1}{2} \tag{2.28}$$

and the water constraint function is:

$$\eta_j^e = \begin{cases} 1, \quad \theta \geq \theta_f \\ \exp\left(2.5\dfrac{\theta-\theta_f}{\theta_f-\theta_w}\right), \quad \theta < \theta_f \end{cases} \tag{2.29}$$

2.3.2. Vegetation

At the current stage of model development, a simple vegetation growth module is included. The LAI, rooting depth (RMX), crop ET coefficient (Kc) and canopy height (hc) are updated daily according to a piecewise linear function described in [*Dingman*, 2002]:

$$V = \begin{cases} V_{\min} & JD < JP(1) \\ V_{\min} + \dfrac{JD - JP(1)}{JP(2) - JP(1)}\left(V_{\max} - V_{\min}\right) & JP(1) \leq JD < JP(2) \\ V_{\max} & JP(2) \leq JD < JP(3) \\ V_{\max} - \dfrac{JD - JP(2)}{JP(3) - JP(2)}\left(V_{\max} - V_{\min}\right) & JP(3) \leq JD < JP(4) \\ V_{\min} & JP(1) \leq JD \end{cases} \tag{2.30}$$

where V is either LAI, RMX, Kc or hc. V_{max} is the maximum value of the variable and V_{min} its minimum value. JD is the Julian day, and JP is the Julian days of the control points, JP(1), JP(2) , JP(3) and JP(4) corresponding, respectively, to the starting day of growth, the day on which the plant reaches its maximum canopy, the day on which the leaves begin to wilt, and the day on which canopy comes back to its minimum. Maximum and minimum LAI, RMX and h_c have been obtained from literature values [*Breuer et al.*, 2003; *Neitsch et al.*, 2005]. Kc for the alfalfa based reference ET is found from [*FAO*, 1998].

Vegetation Canopy Storage

The maximum Canopy storage of a cell is updated each day using its LAI [*Noilhan and Planton*, 1989]:

$$Wr_{mx} = 2 \times 10^{-4} \sum_{i=1}^{nRPT} \delta_i LAI_i \tag{2.31}$$

Where Wr_{mx} is the maximum canopy storage (m), and δ and nRPT is defined above. The canopy interception is then calculated using a bucket model:

$$CS_{new} = \min(Wr_{mx} - CS, P) \tag{2.32}$$

in which, CS is the current canopy storage (m), CS_{new} is the new interception during the time step (m), and P is the precipitation during the time step (m).

2.3.3. Overland flow

Overland flow is an important contributor to channel flow. The magnitude of overland flow is an extremely important factor that determines erosion and the amount of solids carried into the river system. It is of great significance to the transport of land applied chemicals to surface waters and thus to water quality and human health issues.

Rainfall, snowmelt, or occasionally subsurface exfiltration are the main causes of overland flow. Based on the source of the excess water, overland flow is conventionally categorized into infiltration excess (Hortonian) [*Horton*, 1933; *Leach et al.*, 1933] and saturation excess [*Dunne and Black*, 1970]. In infiltration excess, the rate of rainfall exceeds the infiltrating capacity of the receiving surface. Hortonian runoff is often more significant in arid and semi-arid regions due to the absence of well developed soil and vegetation covers [*Lange et al.*, 2003]. The infiltrating capacity can be further reduced by surface sealing that is a result of sudden wetting, heavy rainfall on bare soils or other reasons [*Assouline*, 2004]. The saturation excess, on the other hand, results from the underlying soil being saturated and unable to accept any more infiltration. However, it is becoming more and more widely accepted that subsurface flow plays an important role in runoff generation even in cases where Hortonian runoff is traditionally thought to be dominant.

Overland flow is characterized by fast, ephemeral flows over rough, uneven surfaces and over temporally discontinuous flow domains. In most hydrologic models, overland flow is considered as a thin film of water which evenly covers the entire surface of the flow area (e.g. [*Downer and Ogden*, 2004b; *Jia et al.*, 2001]). However, water tends to rapidly concentrate into multiple rivulets and the ideal thin film flow generally does not exist except for the first couple of hundred meters. The National Resources Conservation Services (NRCS) describes the three stages of overland flow [*Division*, 1986] as: (1) sheet flow, the flow over plane which occurs in the headwater of streams, normally less than 300 feet of travel length; (2) shallow concentrated flow; and (3) open channel, well developed flow paths such as gullies, pipes, rills and ditches. Although many models do attempt to explicitly model major rivers, the existence of these well-defined flow paths is so ubiquitous that it is practically impossible to model all of them. Figure 3.6 shows well-developed channels that have been registered in the National Hydrography Dataset (NHD) in a 1000 km^2 region in Michigan. We observe that the river network forms a dense web. Explicitly modeling these flow webs has far surpassed current available computational resources. Thus the overland flow models are inevitably describing the bulk effects.

After infiltration is computed in the vadose zone model, the water left on the ground may become surface runoff. PAWS contains two compartments: the runoff compartment (or the overland flow layer) and the surface storage compartment (or the surface ponding layer). The runoff compartment describes sheet flow, shallow concentrated flow, flows in pipes, furrows, ditches and other pathways. The surface storage compartment models the layer of water that interacts with soil. Once water moves into a flow pathway, it no longer infiltrates over the entire surface like the surface storage does. As a result, we treat water in the runoff compartment as non-infiltrating unless the water flows back into the storage.

Although the thickness of the overland flow is generally within the order of a couple of centimeters, which qualifies for the laminar flow regime description, the highly uneven, rough and variable surface favors a turbulent treatment [*Gunduz and Aral*, 2005; *Singh*, 1996]. Due to this reason, the depth-integrated Saint Venant Equations (SVE) for shallow, gradually varied unsteady flow, sometimes called the shallow water equations or the Dynamic Wave Equations, can be used:

$$\begin{aligned} &\frac{\partial h}{\partial t} + \frac{\partial (hu)}{\partial x} + \frac{\partial (hv)}{\partial y} = s \\ &\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -g\frac{\partial h}{\partial x} + g(S_{0x} - S_f) \\ &\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -g\frac{\partial h}{\partial y} + g(S_{0y} - S_f) \end{aligned} \tag{2.33}$$

Where, h is the overland flow water depth [L], u and v are the x- and y-direction water velocities (m/s), g is the gravitational acceleration [L/T^2], s is the source term as precipitation, exfiltration from subsurface, or sink term as infiltration or evaporation,

S_0 is the slope ([L/L]), S_f is the frictional slope ([L/L]).

The full non-linear Saint Venant equations govern the conservation of volume (1st equation, this is also called the continuity equation) and momentum (2nd and 3rd equations) of shallow water flow. This set of equations includes gravity (i.e. the slope of the underlying surface), pressure gradient and local and convective accelerations as the sources of momentum, and frictional loss as the resistance to flow. In overland flow routing, we regard the effect of wind as negligible.

Elegant mathematical theories of these equations have been developed and details are available in [*Singh*, 1996] . By recognizing the dominant role of gravity in overland flow and the complex nature of surface roughness characteristics, Equations (2.33) can be simplified by dropping terms. The Diffusive Wave Equation is obtained by ignoring local and convective acceleration (that is inertia) terms:

$$\begin{aligned} &\frac{\partial h}{\partial t}+\frac{\partial (hu)}{\partial x}+\frac{\partial (hv)}{\partial y}=s \\ &0=-g\frac{\partial h}{\partial x}+g(S_{0x}-S_f) \\ &0=-g\frac{\partial h}{\partial y}+g(S_{0y}-S_f) \end{aligned} \tag{2.34}$$

The Diffusive Wave (DW) Equation retains gravity and pressure gradient and thus can describe backwater and flooding.

The simplest form of the Saint Venant equation is the Kinematic Wave (KW)

Equation, which further ignores pressure gradient from the Equation (2.34):

$$\frac{\partial h}{\partial t}+\frac{\partial (hu)}{\partial x}+\frac{\partial (hv)}{\partial y}=s$$
$$S_{0x}=S_f \qquad (2.35)$$
$$S_{0y}=S_f$$

This equation assumes steady state flow conditions (i.e, surface slope is equal to friction slope). For a given surface slope, the flow direction is always pre-defined. As a result, the KW Equation cannot describe backwater [*Borah and Bera*, 2003] and flooding effects.

For the closure of the DW and KW equations, an additional equation is needed. This equation must relate the state of flow to the flow resistance forces. The formulation that has been most widely used is the Manning's formula,

$$S_f=\left(\frac{nu}{h^{2/3}}\right)^2 \qquad (2.36)$$

where n is the manning's roughness coefficient [$L^{-1/3}T$], h is the flow depth [L] and u is the mean flow velocity [LT^{-1}]

In the proposed model, three schemes for solving the overland flow have been included to meet the different needs, namely, a Runge-Kutta Finite Volume scheme (hereby abbreviated as RKFV), a Semi-Implicit Semi-Lagrangian (SISL) scheme, and a shock-capturing scheme based on the Weighted Essentially Non-Oscillatory

(WENO) method [*Shu and Osher*, 1989; *Shu*, 1997]. The RKFV can solve the KWE and DWE. It is sufficient for modeling overland flow in most cases. The SISL scheme is a powerful scheme for the SWE and is most useful for coastal areas, estuaries, shallow lakes and wetlands where surfaces are mostly submerged, water systems are complex and the effect of topography is less predominant. The WENO scheme which solves the DWE and SWE is reserved for future research. Its primary applications are modeling extreme hydrological events, including flashflood and dam-break scenarios. As one of the most important requirements for water routing, all three schemes are mass-conservative.

The RKFV scheme

The Runge-Kutta Finite Volume (RKFV) scheme is designed specifically to solve the Diffusive Wave equation for long term surface flow routing. This scheme has been found to be accurate, very efficient and stable. Combining Manning's formula Eq. (2.36), the Diffusive Wave equation (2.34) is re-arranged as:

$$\frac{\partial h}{\partial t} = -\frac{\partial (hu)}{\partial x} - \frac{\partial (hv)}{\partial y} + s$$

$$u = \frac{1}{n} h^{2/3} \left(S_{0x} - \frac{\partial h}{\partial x} \right)^{1/2} = -\mathrm{sgn}\left(\frac{\partial \eta}{\partial x} \right) \frac{1}{n} h^{2/3} \left| \frac{\partial \eta}{\partial x} \right|^{1/2} \tag{2.37}$$

$$v = \frac{1}{n} h^{2/3} \left(S_{0y} - \frac{\partial h}{\partial y} \right)^{1/2} = -\mathrm{sgn}\left(\frac{\partial \eta}{\partial y} \right) \frac{1}{n} h^{2/3} \left| \frac{\partial \eta}{\partial y} \right|^{1/2}$$

where η is the free surface elevation [L], η = E+h, E being the ground surface elevation. Due to the transient nature of overland flow, wetting and drying are frequent phenomena that must be described naturally. The Arakawa-C grid is used for the discretization of Equation (2.34), Illustrated in Figure 2.2, the Arakawa-C grid defines velocities on the cell boundaries, whereas water depth is defined at the cell center. The water depth at the interface is explicitly calculated, depending on the water depth on two sides of the boundary.

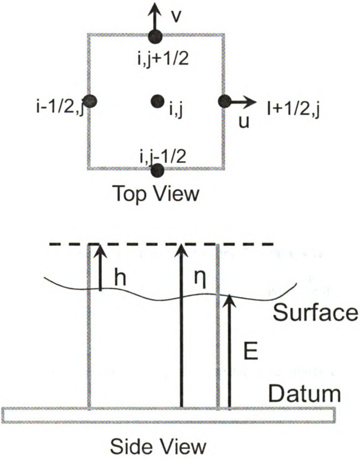


Figure 2.2. Illustration of Arakawa-C grid.

With the above sketch, Eq. (2.37) can discretized in space as:

$$\frac{\partial h}{\partial t} = -\left[\begin{array}{l}\dfrac{h_{i-1/2,j}u_{i-1/2,j} - h_{i+1/2,j}u_{i+1/2,j}}{\Delta x} \\ +\dfrac{h_{i,j-1/2}u_{i,j-1/2} - h_{i,j+1/2}u_{i,j+1/2}}{\Delta y}\end{array}\right] + s_{i,j}$$

(2.38)

Given the water depth on two sides of the boundary as h_L and h_R, and the free surface elevation as η_L and η_R, we use the following logic to determine the flow depth at the boundary, $h_{1/2}$:

$$\eta_L > \eta_R ? \begin{cases} Y, h_L > 0 ? \begin{cases} Y, h_L > h_R ? \begin{cases} Y, h_{1/2} = 0.5 * (h_L + h_R) \\ N, h_{1/2} = h_L \end{cases} \\ N, h_{1/2} = 0 \end{cases} \\ N, h_R > 0 ? \begin{cases} Y, h_R > h_L ? \begin{cases} Y, h_{1/2} = 0.5 * (h_L + h_R) \\ N, h_{1/2} = h_R \end{cases} \\ N, h_{1/2} = 0 \end{cases} \end{cases} \tag{2.39}$$

The interface flow depth in the y direction can be calculated using the same approach. The first condition, $\eta_L > \eta_R ?$, is to decide the upwind direction. For the DWE, the free surface elevation always determines the direction of flow. The second conditional statement is to decide if the upper stream cell is dry. The third selection, $h_L > h_R$?, is to ensure the interface flux does not exceed what is available for outflow during the explicit updating.

Once $h_{1/2}$ is obtained, we use $h_{1/2}$ in Eq. (2.37) to calculate interface velocities u and v, and fluxes.

$$\begin{aligned} u_{i,j+1/2} &= -\operatorname{sgn}\left(\eta_{i,j+1} - \eta_{i,j}\right) \frac{1}{n} h_{i,j+1/2} \left| \frac{\eta_{i,j+1} - \eta_{i,j}}{\Delta x_{i,j+1/2}} \right|^{1/2} \\ v_{i+1/2,j} &= -\operatorname{sgn}\left(\eta_{i+1,j} - \eta_{i,j}\right) \frac{1}{n} h_{i+1/2,j} \left| \frac{\eta_{i+1,j} - \eta_{i,j}}{\Delta y_{i+1/2,j}} \right|^{1/2} \end{aligned} \tag{2.40}$$

Then equation (2.38) is used to update the solution. The procedure is computationally very efficient.

To improve accuracy and stability, Equation (2.38) is marched in time with the explicit Runge-Kutta method. For any given partial differential equation:

$$\frac{\partial U}{\partial t} = f\left(U\right) \tag{2.41}$$

The standard Runge Kutta method can be written as:

$$\begin{aligned} U_i^* &= U^n + \Delta t\left(\sum_{s=0}^{r} c_s F_s\right) \\ F_r &= f\left(U^n + \sum_{s=0}^{r-1} b_{rs} F_s\right) \end{aligned} \tag{2.42}$$

Where U is the unknown variable and F is the time derivative computed at a certain time level. Coefficients c_S and b_{rs} can be looked up from the Butcher's table. Here we employ a third order version of the TVD RK method. More specifically:

$$\begin{aligned} F_0 &= f\left(U^n\right) \\ F_1 &= f\left(U^n + \Delta t F_0\right) \\ F_1 &= f\left(U^n + \frac{1}{4}\Delta t\left(F_0 + F_1\right)\right) \\ U^{n+1} &= U^n + \frac{\Delta t}{6}\left(F_0 + 4F_2 + F_1\right) \end{aligned} \tag{2.43}$$

A second order accurate in time TVD RK is given as:

$$\begin{aligned} F_0 &= f\left(U^n\right) \\ F_1 &= f\left(U^n + \Delta t F_0\right) \\ U^{n+1} &= U^n + \frac{\Delta t}{2}\left(F_0 + F_1\right) \end{aligned} \tag{2.44}$$

This is also a form of predictor/corrector scheme as described in the literature. Second order accuracy in space is achieved with the above scheme when the $h_{1/2} = 0.5 * (h_L + h_R)$ branch in Procedure (2.39) is invoked.

The modified Semi-Implicit Semi-Lagrangian (SISL) scheme

The SISL scheme was originially developed by [*Casulli*, 1990; 1999] and then advanced by [*Martin and Gorelick*, 2005]. Our development of the scheme largely follows that of [Martin], but some changes were made to the scheme. In order to use this scheme, the Chezy's formulation is employed for the friction loss term. This scheme solves the fully nonlinear Saint Venant equations with the inclusion of wind, Coriolis and eddy viscosity effects [*Martin and Gorelick*, 2005]:

$$\frac{\partial h}{\partial t} + \frac{\partial (hu)}{\partial x} + \frac{\partial (hv)}{\partial y} = s$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = -g\frac{\partial \eta}{\partial x} + \varepsilon\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + \frac{\gamma_T\left(U_a - u\right)}{H} + g\frac{\sqrt{u^2 + v^2}}{Cz^2}u + fv$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} = -g\frac{\partial \eta}{\partial y} + \varepsilon\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) + \frac{\gamma_T\left(V_a - v\right)}{H} + g\frac{\sqrt{u^2 + v^2}}{Cz^2}v + fu$$

(2.45)

For the above hyperbolic PDE, solving the advective part of the equation, namely, the highlighted portion:

$$\boxed{\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = \varepsilon\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right)} + fv - g\frac{\partial \eta}{\partial x} + \frac{\gamma_T\left(U_a - u\right)}{H} + g\frac{\sqrt{u^2 + v^2}}{Cz^2}u$$

$$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}=\varepsilon\left(\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}\right)+fu-g\frac{\partial \eta}{\partial y}+\frac{\gamma_T\left(V_a-v\right)}{H}+g\frac{\sqrt{u^2+v^2}}{Cz^2}v$$

normally poses the greatest challenge to the stability and efficiency of the numerical scheme. Whereas the treatment of other terms generally admits much larger time steps. This provides the motivation for the usage of operator splitting and the semi-Lagrangian scheme to solve the advective components. Let us follow [*Martin and Gorelick*, 2005] and denote the advective scheme as F and the solution to other terms as S. Then Equation (2.45) is solved in two fractional steps:

$$\begin{bmatrix} \overset{*}{u} \\ \overset{*}{v} \end{bmatrix} = F\left(\begin{bmatrix} u^n \\ v^n \end{bmatrix}\right)$$

$$\begin{bmatrix} h^{n+1} \\ u^{n+1} \\ v^{n+1} \end{bmatrix} = S\left(\begin{bmatrix} h^n \\ \overset{*}{u} \\ \overset{*}{v} \end{bmatrix}\right) \tag{2.46}$$

The F operator is a semi-Lagrangian scheme involves two steps. In the first step, the current grid points $(X,Y)=x\otimes y$ are traced back through Δt of the travel time along the travel paths that are defined by the flow field of (u^N,v^N), and we thus find the departing location of (X,Y) as (X_{SL},Y_{SL}). [*Martin and Gorelick*, 2005] used a semi-analytical approach. Although this is also implemented in the code, after some initial comparisons, we found the classic 4-stage Runge Kutta method [*Zheng*, 2002] to be more stable.

In the second step, we calculate the velocities (u_{sl}^N,v_{sl}^N) at locations (X_{SL},Y_{SL}).

This can be done using a simple bilinear interpolation from (u^N, v^N). For a point (xi,yi) such that $x_1 \leqq xi < x_2$ and $y_1 \leqq yi < y_2$ where x_1, x_2, y_1, y_2 are coordinates of the closest cell centers in our grid, the bilinear interpolation of velocity u can be written as (2.47).

$$
\begin{aligned}
u_{sl}(x_i, y_i) \approx & \frac{u^N(x_1, y_1)}{(x_2 - x_1)(y_2 - y_1)}(x_2 - x_i)(y_2 - y_i) \\
& + \frac{u^N(x_2, y_1)}{(x_2 - x_1)(y_2 - y_1)}(x_i - x_1)(y_2 - y_i) \\
& + \frac{u^N(x_1, y_2)}{(x_2 - x_1)(y_2 - y_1)}(x_2 - x_i)(y_i - y_1) \\
& + \frac{u^N(x_2, y_2)}{(x_2 - x_1)(y_2 - y_1)}(x_i - x_1)(y_i - y_1).
\end{aligned}
\tag{2.47}
$$

Essentially, we trace back the points along the flow lines and find out their values at the previous time step. We must note here that this is in fact not the most correct implementation of semi-Lagrangian method, as a classic SL method traces back along the characteristics, which are not necessarily the flow lines.

We notice that the SISL scheme is considerably more complex in theory and much more computationally expensive than the RKFV scheme. Although the time step can be larger, it is not obvious whether the gain with larger time step would offset the much more computations involved. Also, the semi-Lagrangian scheme may crash at large discontinuities. The semi-Lagrangian step is useful only at places where flow is continuous (so that flow path tracking is meaningful). In overland flow, land is constantly drying and wetting, and flow is transient and discontinuous. This makes the

RKFV scheme more favorable for long term simulations, whereas the SISL scheme may be of more interest for areas with shallow lakes and estuarine systems. We will compare the performance of the two schemes in section 2.7.

2.3.4. Channel flow

The channel flow model is based on the one dimensional Saint Venant equations. Similar to the overland flow, the channel flow can also be described by, from full complexity to simplified forms, the Dynamic Wave, Diffusive Wave and Kinematic Wave equations. To account for the variation of flow area along channels, the conservation of mass and momentum takes a slightly different form. The full nonlinear Saint Venant equation for channel routing, or the dynamic wave equation, reads:

$$\begin{aligned} &\frac{\partial A}{\partial t} + \frac{\partial (uA)}{\partial x} = rb + q_l + q_{gc} + q_t \\ &\frac{\partial (uA)}{\partial t} + \frac{\partial}{\partial x}\left(A\left(g\frac{\eta}{2} + u^2\right)\right) = gA\left(S_0 - S_f\right) + \frac{Q_r}{A} \end{aligned} \tag{2.48}$$

Or, one may write it in a matrix form:

$$\begin{bmatrix} A \\ uA \end{bmatrix}_t + \begin{bmatrix} uA \\ A\left(g\frac{\eta}{2} + u^2\right) \end{bmatrix}_x + \begin{bmatrix} -\left(rb + q_l + q_{gc} + q_t\right) \\ -\left(gA\left(S_0 - S_f\right) + \frac{Q_r}{A}\right) \end{bmatrix} = 0 \tag{2.49}$$

where A is the cross-sectional area of the channel (m^2), u is the flow velocity (m/s), η is the river stage (also called free surface elevation, m), r is precipitation (m/s), ql is lateral inflow from overland flow ($m^3/m^2/s$), q_{gc} is the groundwater contribution ($m^3/m^2/s$), S_0 is the slope and S_f is the friction slope. The computation of Q_{gc} will be described in section 2.4.2. As with the overland flow, several solvers have been implemented in PAWS, including a Runge-Kutta Finite Volume scheme that is similar to the overland RKFV scheme and a second order MacCormack scheme. The RKFV scheme solves the simplified Diffusive Wave Equation, in which the momentum conservation equation is replaced with:

$$u = \frac{1}{n} h_c \left(S_{0x} - \frac{\partial h}{\partial x} \right)^{1/2} = -\operatorname{sgn}\left(\frac{\partial \eta}{\partial x} \right) \frac{1}{n} h_c \left| \frac{\partial \eta}{\partial x} \right|^{1/2} \tag{2.50}$$

Where η is the free surface elevation (m), n is the manning's roughness coefficient and h_c is the channel flow depth (m). Kinematic wave Equation is found to be unusable for channel routing, as it cannot describe backwater effects and its topography-driven nature is unsuitable for the stream networks where the elevation variation is mild. The RKFV scheme serves as the long-term channel routing scheme and its details have been given in previous section. The difference is that the temporal derivative of area is calculated using:

$$\frac{\partial A}{\partial t} = -\frac{1}{\Delta x} \left[A_{i-1/2} u_{i-1/2} - A_{i+1/2} u_{i+1/2} \right] + s_i \tag{2.51}$$

Where, for a rectangular channel

$$A_{i+1/2} = w_{i+1/2} hc_{i+1/2} \tag{2.52}$$

$W_{i+1/2}$ is the channel width at cell interfaces. The channel width is an input value to the model and may be obtained from field or aerial observations or regression techniques. $hc_{i+1/2}$ is determined from hc_i and hc_{i+1} and the free surface elevations on both sides again using the procedure (2.39). The formulation in (2.51) also admits other channel geometries, in which case different functional relationships between A and h must be provided. However for the current research, rectangular channels are deemed sufficient. Other widely used models such as SWAT [*Neitsch et al.*, 2005] use rectangular channels only. We employ an operator splitting approach such that the groundwater contribution q_{gc} is solved implicitly as described in section 2.4.2.

Boundary conditions

Internal Boundary Conditions

The river's physical domain is discretized into M one-dimensional cells whose mass/stage is defined on each cell's center. A ghost cell is added to both ends of the grid. A ghost cell is a frictional cell whose existence is only for the purpose of implementing the boundary conditions. When the equation is solved, only cells inside the physical domain are updated. The boundary conditions are implemented by setting up the values of the ghost cells to reflect the correct physical inflow/outflow conditions. When a river drains into a downstream river, we call the receiving one the main river and the contributing one a tributary. The main river's cell at which the tributary meets is identified as the confluence cell. The stage of the confluence cell is taken as a Dirichlet boundary condition for the tributary:.

$$\begin{gathered} E_{M+1,t}=E_{c,m} \\ h_{M+1,t}^{N \sim N+1} = h_{c,m}^{N} \end{gathered} \tag{2.53}$$

Where $E_{N+1,t}$ is the bed elevation of the N+1-th cell (a ghost cell) of the tributary river, $E_{c,m}$ is the bed elevation of the main river at the place the tributary confluences the main river. $h_{M+1,t}^{N\sim N+1}$ is the stage of the head of the ghost cell of the tributary river during the time step t^N to t^{N+1}, while $h_{c,m}^N$ is the head of the confluence cell on the main river at the end of t^N.

The river models are run in an upstream-downstream cascade sequence, so the contributions from tributaries are always computed before the model is run for the main river. The tributary inflows are converted into a source term to the main river:

$$s_i^{N\sim N+1} = s_i^{N\sim N+1} + \frac{1}{\Delta x_i \Delta t} \sum_{j=1}^{nt} \int_{t^N}^{t^{N+1}} q_j dt \tag{2.54}$$

In which, s is the source term throughout the new time step of the confluence cell on the main river $[L^2/T]$, Δt and Δx are the temporal and spatial steps of the main river and $\int_{t^N}^{t^{N+1}} q_j dt$ is the accumulative inflow of its j-th tributary during this time step.

The above boundary condition implicitly assumes that the downstream river stage does not change much during the time step, which is generally valid since the main river is a larger river. But its limitation is that during high flow periods the time step of the river network may need to be reduced to maintain stability. Each river is checked against its own Courant number restriction and the smallest time is used for the entire river network. This ensures that the rivers communicate frequently enough and the Dirichlet downstream boundary is properly updated. At the highest flows, this setting has constrained the time step of the river network to be around 1 to 2 minutes. However, this is not unreasonable as the durations of high flows are often not long.

For regular flows, the river network uses a Δt of 10 minutes. It is possible to have negative inflow, i.e., backwater from the main river into tributaries. This situation is handled without the need for any further modification by the above internal boundary conditions.

At the final outlet of the river network, if downstream river stage record is provided, it can be used in PAWS. Otherwise, a simple free outflow condition can be used. Since the physical condition is not available, this method extrapolates the head of the ghost cell from its internal cells:

$$hc_{M+1} = 2hc_M - hc_{M-1} \tag{2.55}$$

If the upstream end of the river is a headwater, a Dirichlet value of $h_0 = 0$ is simply given to the ghost cell. It is possible to read in an inflow record file if measured data are available. Moreover, sometimes it is desirable to split a river into several domains. In this case, the upstream river inflow can be treated as a tributary and use the BC detailed above.

As described in the previous section, some researchers [*Gunduz and Aral*, 2005; *Ivanov et al.*, 2004a], have argued for finite element methods for overland flow modeling. Their main motivation is the flexibility of the finite element mesh to resolve the topography, hydrologic response units and rivers. Although the concern is legitimate, we follow a structured grid approach for its simplicity and generality.

2.3.5. Unsaturated vadose zone model

Vertical water moisture movement in the soil compartment is described by the mixed form of the Richards's Equation [*Celia et al.*, 1990; *van Dam and Feddes*, 2000]

$$C(h)\frac{\partial h}{\partial t}=\frac{\partial}{\partial z}\left[K(h)\left(\frac{\partial h}{\partial z}+1\right)\right]+W(h) \tag{2.56}$$

In which, K(h) is the unsaturated hydraulic conductivity (L/T), h is the soil water pressure head (L), W(h) is the volumetric source and sink term, including evaporation, plant root extraction and bypass flow and z is the vertical coordinate (positive upward) (L). The two state variables, pressure head h and water content θ are linked via C(h)= $\partial\theta / \partial h$, the differential water capacity (L^{-1}). This equation assumes that for a given type of soil, there is a one-to-one relationship between the pressure head and the water content (water retention curve) and thus we can solve only for one of the variables.

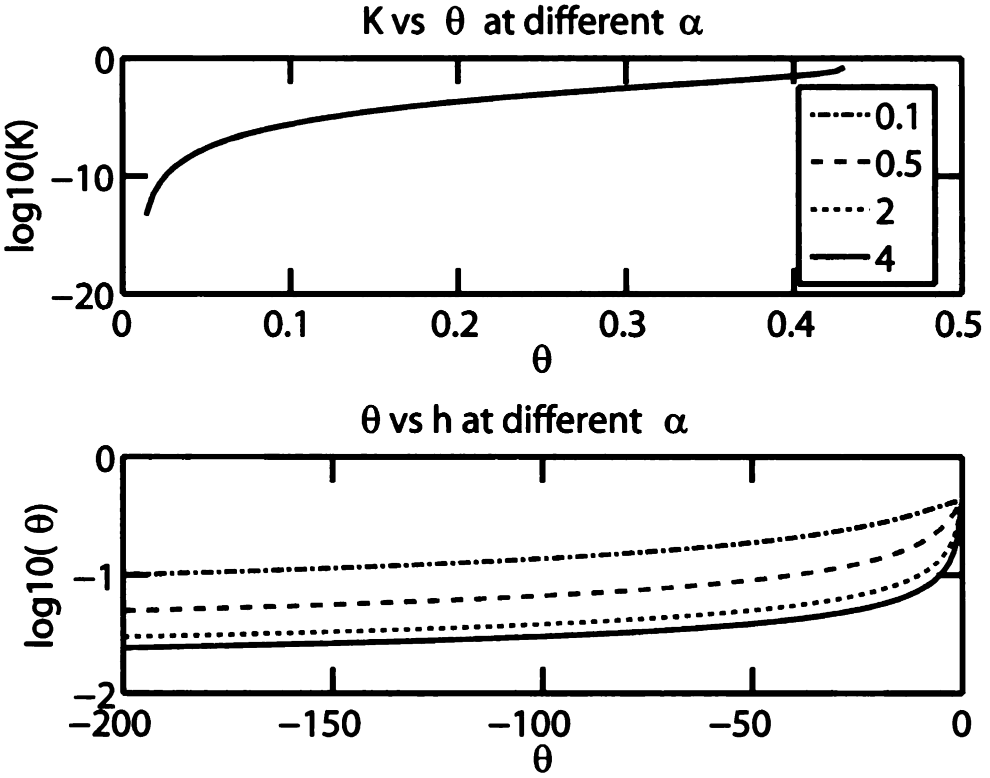
The solution to the Richards' Equation is of central importance to the overall performance of the watershed model, because this compartment links infiltration and percolation, and serves as the vital linkage between surface water and groundwater, and governs the actual evapotranspiration. The unsaturated hydraulic conductivity and pressure head are both functions of the soil moisture content. We choose the Mualem-van Genuchten (vG) formulation:

$$
\begin{aligned}
S &= \frac{\theta(h) - \theta_r}{\theta_s - \theta_r} = \left(1 + \left|\alpha h\right|^n\right)^{-(n-1)/n} \\
K(h) &= K_S S^{\lambda}\left[1 - \left(1 - S^{n/(n-1)}\right)^{(n-1)/n}\right]^2
\end{aligned}
\qquad (2.57)
$$

Where S is the relative saturation, θ is the soil moisture content, θ_S is the saturated water content, θ_r is the residual water content, n is a measure of the pore-size distribution, α is a parameter related to the inverse of the air entry suction and λ is a pre tortuosity/connectivity parameter [*van Genuchten*, 1980].

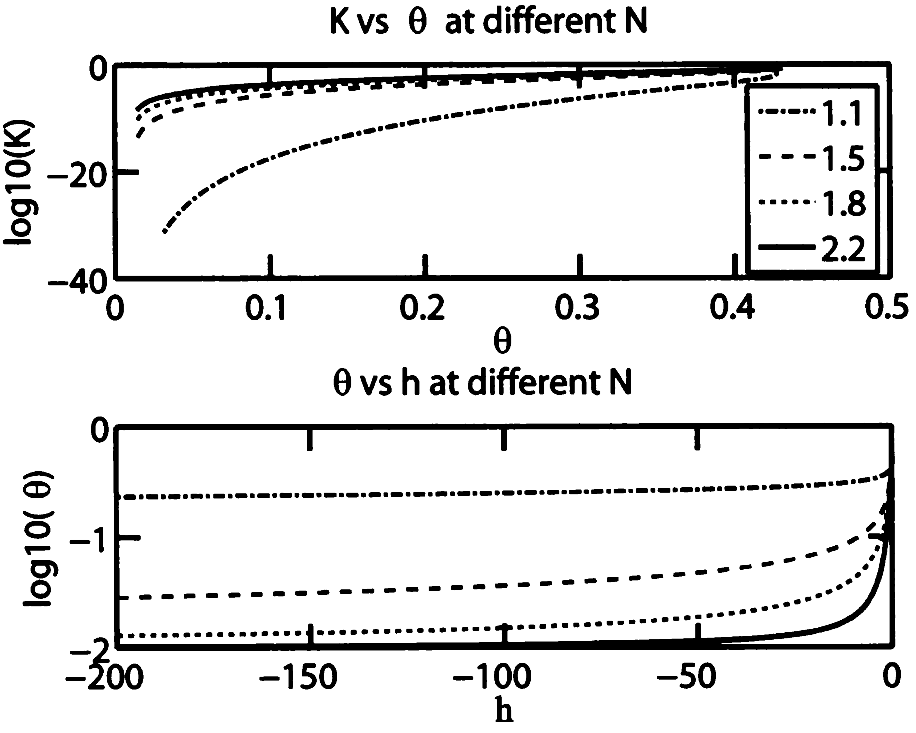
Eq. (2.57) is widely used in literature. However, with this formulation, the equation is highly nonlinear and is known to produce results of a very transient nature. Figure 2.3 shows relationship between K, θ and h at different n, λ and α values.

Figure 2.3. Relationship between K, θ and h at different parameter values with the van Genuchten formulation for a hypothetic soil type (a) at different α; (b): at different λ; (c) at different N. THE in the figure heading means θ. The base parameter for this comparison is α= 2.49(1/m), N = 1.507, $\theta_S = 0.43$, $\theta_r = 0.01$; $K_S = 0.175$ (m/day); λ = -0.14. Unit in the figure is the same as the base parameter set

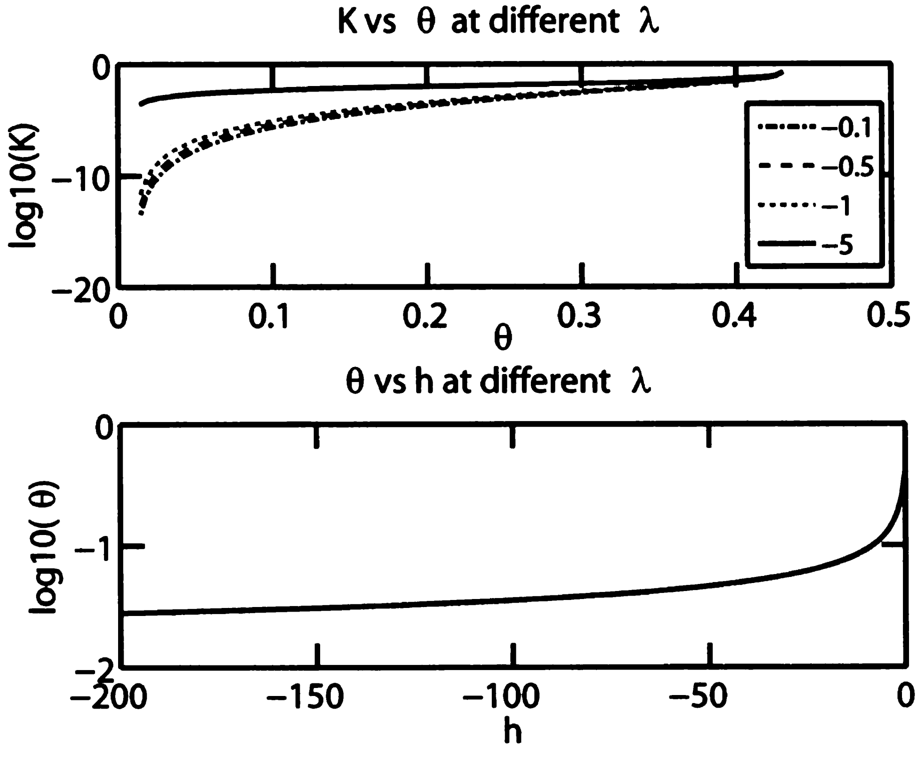


(a)

Figure 2.3 (cont'd.)



(b)



(c)

In this figure, a base parameter set is used (see figure caption) while one of the parameters, either n, λ or α, is replaced by the values indicated by the figure legend. Then the K as a function of θ and 0 as a function of h are plotted for the altered parameter sets. We observe that K can vary by many orders of magnitude at different θ. All of n, λ and α greatly influence the relationships, in which, n is the most sensitive.

A naive finite difference discretization in time for the LHS of equation (2.56) would read:

$$\frac{\theta^{j+1}-\theta^{j}}{\Delta t}=\frac{C_i^{j+1/2}\left(h_i^{j+1}-h_i^{j}\right)}{\Delta t} \tag{2.58}$$

Where θ^{j+1} is the water content at the next time level. (To avoid confusion with the connectivity parameter n, in this section we used j to denote the time levels and use i for vertical cell index, i is positive downward). However, due to the high non-linearity of the differential water capacity term, temporal discretizations similar as above are known to produce serious mass balance errors [*Lehmann and Ackerer*, 1998; *van Dam and Feddes*, 2000]. Here we employ the widely used implicit iteration scheme by [*Celia et al.*, 1990].

$$\theta_i^{j+1}-\theta_i^{j}=C_i^{j+1,p}\left(h_i^{j+1,p}-h_i^{j,p}\right)+\theta_i^{j+1,p-1}-\theta_i^{j} \tag{2.59}$$

In which p is the iteration number and C is evaluated as:

$$C_i^{j+1,p} = \frac{\partial \theta}{\partial h} = -\frac{\left(\theta_s - \theta_r\right)}{\left(1 + \left|\alpha h^{j,p}\right|^n\right)^{\frac{2n-1}{n}}} \frac{\left|\alpha h^{j,p}\right|^n (n-1)}{h^{j,p}} \tag{2.60}$$

In other words, the differential water capacity C(θ) is updated at every iteration using the newly obtained $h^{j,p}$. The iteration repeats until both h and θ converge. This stable implicit scheme can almost guarantee mass balance error to be close to rounding off errors [*van Dam and Feddes*, 2000]. When we apply Equation (2.59) in to (2.56) we obtain the full discretization formula:

$$\frac{1}{\Delta t}\left[C_i^{j+1,p}\left(h_i^{j+1,p} - h_i^{j+1,p-1}\right) + \theta_i^{j+1,p-1} - \theta_i^j\right] = \left[\frac{K_{i-1/2}^j}{\Delta z_i}\left(\frac{h_{i-1}^{j+1,p} - h_i^{j+1,p}}{\Delta z_{i-1/2}} + 1\right) - \frac{K_{i+1/2}^j}{\Delta z_i}\left(\frac{h_i^{j+1,p} - h_{i+1}^{j+1,p}}{\Delta z_{i+1/2}} + 1\right)\right] + W_i^j \tag{2.61}$$

We can write this into a tri-diagonal form:

$$\begin{aligned}
&a_i^j h_{i-1}^{j+1,p} + b_i^{j,p} h_i^{j+1,p} + c_i^j h_{i+1}^{j+1,p} = d_i^{j,p} \\
&a_i^j = -\frac{K_{i-1/2}^j}{\Delta z_i \Delta z_{i-1/2}}, c_i^j = -\frac{K_{i+1/2}^j}{\Delta z_i \Delta z_{i+1/2}} \\
&b_i^{j,p} = -(a_i^j + c_i^j) + \frac{C_i^{j+1,p}}{\Delta t} \\
&d_i^{j,p} = \frac{C_i^{j+1,p}}{\Delta t} h_i^{j+1,p-1} + W_i^j + \frac{K_{i-1/2}^j - K_{i+1/2}^j}{\Delta z_i} - \frac{\theta_i^{j+1,p-1} - \theta_i^j}{\Delta t}
\end{aligned} \tag{2.62}$$

We observe that a_I and c_i do not change from iteration to iteration so the iteration index p is dropped from the notation. The above system can be solved efficiently using the Thomas algorithm [*Press et al.*, 1997].

2.3.6. Saturated groundwater flow model

The unconfined aquifer is conceptualized as a series of vertical layers. In each vertical layer, we solve the 2-dimenional groundwater equation:

$$S\frac{\partial H}{\partial t}=\frac{\partial}{\partial x}\left[T\left(\frac{\partial H}{\partial x}\right)\right]+\frac{\partial}{\partial y}\left[T\left(\frac{\partial H}{\partial y}\right)\right]+R+W-Dp \qquad (2.63)$$

Where S is the storativity (dimensionless), T is the transmissivity of the aquifer $[L^2/T]$, T=Kb where K is the saturated hydraulic conductivity [L/T] and b is the saturated thickness of the aquifer [L], H is hydraulic head [L], R is recharge or discharge (L/T), W is the source and sink term due to pumping or root extraction [L/T] (inflow as positive) and Dp is percolation into deeper aquifers [L/T]. The above equation is discretized using standard backward-in-time, center-in-space finite difference scheme:

$$\frac{S_{i,j}}{\Delta t}\left(H_{i,j}^{n+1}-H_{i,j}^{n+1}\right)=\left[\frac{T_{i-1/2,j}}{\Delta x_{i,j}}\left(\frac{H_{i-1,j}^{n+1}-H_{i,j}^{n+1}}{\Delta x_{i-1/2,j}}\right)-\frac{T_{i+1/2,j}}{\Delta x_{i,j}}\left(\frac{H_{i,j}^{n+1}-H_{i+1,}^{n+1}}{\Delta x_{i+1/2,j}}\right.\right.$$

$$+\left[\frac{T_{i,j-1/2}}{\Delta y_{i,j}}\left(\frac{H_{i,j-1}^{n+1}-H_{i,j}^{n+1}}{\Delta y_{i,j-1/2}}\right)-\frac{T_{i,j+1/2}}{\Delta y_{i,j}}\left(\frac{H_{i,j}^{n+1}-H_{i,j+1}^{n+1}}{\Delta y_{i,j+1/2}}\right)\right]+R_{i,j}^{n}+W_{i,j}^{n}-D \qquad (2.64)$$

Here I and j are y and x coordinate indices, respectively, n is the time level.

$W_{i,j}^{n}=\int_{t_n}^{t_{n+1}}W_{i,j}dt$ is the integral source term throughout the time step.

We solve for the percolation term implicitly:

$$Dp^{n}_{i,j} = -K_l \frac{\left(H^{n+1}_{i,j} - H^{n}_{l}\right)}{\Delta z_l} \tag{2.65}$$

where K$_l$ is the hydraulic conductivity [L/T] of the aquitard beneath the current aquifer, Δz_l is the thickness of the aquitard, H^n_l is the hydraulic head in the aquifer below. Equation (2.64) becomes:

$$\frac{S_{i,j}}{\Delta t}\left(H^{n+1}_{i,j} - H^{n+1}_{i,j}\right) = \left[\frac{T_{i-1/2,j}}{\Delta x_{i,j}}\left(\frac{H^{n+1}_{i-1,j} - H^{n+1}_{i,j}}{\Delta x_{i-1/2,j}}\right) - \frac{T_{i+1/2,j}}{\Delta x_{i,j}}\left(\frac{H^{n+1}_{i,j} - H^{n+1}_{i+1,j}}{\Delta x_{i+1/2,j}}\right)\right]$$
$$+\left[\frac{T_{i,j-1/2}}{\Delta x_{i,j}}\left(\frac{H^{n+1}_{i,j-1} - H^{n+1}_{i,j}}{\Delta x_{i,j-1/2}}\right) - \frac{T_{i,j+1/2}}{\Delta x_{i,j}}\left(\frac{H^{n+1}_{i,j} - H^{n+1}_{i,j+1}}{\Delta x_{i,j+1/2}}\right)\right]$$
$$-K_l \frac{\left(H^{n+1}_{i,j} - H^{n}_{l}\right)}{\Delta z_l} + R + \int_{z_b}^{z_t} W_{i,j} dz$$

(2.66)

The x-direction interface transmissivity is taken as:

$$\frac{T_{i-1/2,j}}{\Delta x_{i,j}} = \frac{2T_{i,j}T_{i-1,j}}{\Delta x_{i,j}T_{i,j} + \Delta x_{i-1,j}T_{i-1,j}} \tag{2.67}$$

y direction interface transmissivity is defined in a similar fashion. Re-arranging Eq (2.64) we obtain:

$$a_{i,j}H^{n+1}_{i-1,j} + b_{i,j}H^{n+1}_{i,j-1} + c_{i,j}H^{n+1}_{i,j} + d_{i,j}H^{n+1}_{i+1,j} + e_{i,j}H^{n+1}_{i,j+1} = f_{i,j} \tag{2.68}$$

Where

$$\begin{aligned} a_{i,j} &= -\frac{T_{i-1/2,j}}{\Delta y_{i,j}\Delta y_{i-1/2,j}}, b_{i,j} = -\frac{T_{i,j-1/2}}{\Delta x_{i,j}\Delta x_{i,j-1/2}} \\ d_{i,j} &= -\frac{T_{i+1/2,j}}{\Delta y_{i,j}\Delta y_{i+1/2,j}}, e_{i,j} = -\frac{T_{i,j+1/2}}{\Delta x_{i,j}\Delta x_{i,j+1/2}} \\ c_{i,j} &= -(a_{i,j} + b_{i,j} + d_{i,j} + e_{i,j}) + \frac{S_{i,j}}{\Delta t} + \frac{K_l}{\Delta z_l} \\ f_{i,j} &= \frac{S_{i,j}}{\Delta t} H_{i,j}^n + W_{i,j}^n + R + \frac{K_l H_l^n}{\Delta z_l} \end{aligned} \tag{2.69}$$

This gives rise to a sparse, symmetric, penta-diagonal and positive definite system which can be efficiently solved by many advanced matrix solvers. We employ the Conjugate Gradient method (described in [*Leveque*, 2007]) here due to its proven efficiency, stability and readily-available codes [*Press et al.*, 1997].

The implementation of boundary condition is straightforward in the model. Cells outside the watershed boundary are marked as inactive, and their conductivity values are set to zero. Thus no flow exists at the watershed boundaries. It is known that groundwater divides don't always coincide exactly with the watershed boundaries. However unless we are provided with more data regarding groundwater flow, this is the best approach we can use.

The nonlinear drainable porosity unconfined aquifer flow model

The above solver deals with the linear groundwater flow equation, whereas in reality the storage coefficient of the unconfined aquifer is nonlinear. [*Hilberts et al.*, 2005] has studied the effect of nonlinear storage on the specific yield of the unconfined aquifer. The storage coefficient of the unconfined aquifer was found to be not a constant but a function of soil water retention characteristics and depth to water table. The finding is also consistent with previous experimental studies (e.g. [*Healy and Cook*, 2002]). The concept of drainable porosity is based on the hydrostatic equilibrium state. This condition certainly deviates from reality because the soil is constantly wetting and drying and hydrostatic equilibrium condition rarely exists, but it represents a significant improvement over the common Bousinessq approximation, which assumes the specific yield as simply the porosity. The stored moisture in a soil column can be obtained by integrating the moisture profile under hydrostatic equilibrium, using a modified van Genuchten formulation:

$$\begin{aligned} S(H) &= (Z-H)(\theta_s-\theta_r)\left(1+\left(\alpha(H-Z)\right)^n\right)^{-1/n} + (\theta_s-\theta_r)H \\ S'(H) &= (\theta_s-\theta_r)\left\{1-\left(1+\left(\alpha(H-Z)\right)^n\right)^{-\left(\frac{n+1}{n}\right)}\right\} \end{aligned} \tag{2.70}$$

The flow equation for the unconfined aquifer is then written as:

$$\frac{\partial ST(H)}{\partial t} = \frac{\partial}{\partial x}\left[T\left(\frac{\partial H}{\partial x}\right)\right] + \frac{\partial}{\partial y}\left[T\left(\frac{\partial H}{\partial y}\right)\right] + R + W - Dp \tag{2.71}$$

We see that equation (2.71) reverts back to the ordinary groundwater flow equation (2.63) if the function ST(H) is a linear function of H. However, when we use the nonlinear storage function in (2.70), this equation becomes nonlinear and need to be solved using an iterative approach such as the Newton iteration. One important consideration on any method we choose is mass conservation. We enforce conservation by demanding that the water storage/water table position curve must be single valued and the solution to the PDE must fall on this curve. To do this we discretize equation (2.71) in a conservative manner:

$$
\begin{aligned}
\frac{S\left(H_{i,j}^{n+1}\right) - S\left(H_{i,j}^{n}\right)}{\Delta t} &= \left[\frac{T_{i-1/2,j}}{\Delta x_{i,j}}\left(\frac{H_{i-1,j}^{n+1} - H_{i,j}^{n+1}}{\Delta x_{i-1/2,j}}\right) - \frac{T_{i+1/2,j}}{\Delta x_{i,j}}\left(\frac{H_{i,j}^{n+1} - H_{i+1,j}^{n+1}}{\Delta x_{i+1/2,j}}\right)\right] \\
&+ \left[\frac{T_{i,j-1/2}}{\Delta y_{i,j}}\left(\frac{H_{i,j-1}^{n+1} - H_{i,j}^{n+1}}{\Delta y_{i,j-1/2}}\right) - \frac{T_{i,j+1/2}}{\Delta y_{i,j}}\left(\frac{H_{i,j}^{n+1} - H_{i,j+1}^{n+1}}{\Delta y_{i,j+1/2}}\right)\right] \\
&+ R_{i,j}^{n} + W_{i,j}^{n} - K_l \frac{\left(H_{i,j}^{n+1} - H_l^{n}\right)}{\Delta z_l}
\end{aligned}
\tag{2.72}
$$

Re-arranging terms and writing in a minimization form:

$$
\begin{aligned}
G\left(H_{i,j}^{n+1}\right) &= \frac{S\left(H_{i,j}^{n+1}\right)}{\Delta t} + \frac{K_l H_{i,j}^{n+1}}{\Delta z_l} - \frac{S\left(H_{i,j}^{n}\right)}{\Delta t} - R_{i,j}^{n} - W_{i,j}^{n} - K_l \frac{H_l^{n}}{\Delta z_l} \\
&+ \left[\frac{T_{i,j-1/2}}{\Delta y_{i,j}}\left(\frac{H_{i,j-1}^{n+1} - H_{i,j}^{n+1}}{\Delta y_{i,j-1/2}}\right) - \frac{T_{i,j+1/2}}{\Delta y_{i,j}}\left(\frac{H_{i,j}^{n+1} - H_{i,j+1}^{n+1}}{\Delta y_{i,j+1/2}}\right)\right] \\
&- \left[\frac{T_{i-1/2,j}}{\Delta x_{i,j}}\left(\frac{H_{i-1,j}^{n+1} - H_{i,j}^{n+1}}{\Delta x_{i-1/2,j}}\right) - \frac{T_{i+1/2,j}}{\Delta x_{i,j}}\left(\frac{H_{i,j}^{n+1} - H_{i+1,j}^{n+1}}{\Delta x_{i+1/2,j}}\right)\right]
\end{aligned}
\tag{2.73}
$$

The Newton iteration approach involves computing the Jocabian of G(H):

$$\mathbf{J}\left(H\right)=\begin{bmatrix} \frac{\partial G_{i,j}}{\partial H_{i,j}} & \frac{\partial G_{i,j}}{\partial H_{i+1,j}} & 0\ldots & \frac{\partial G_{i,j}}{\partial H_{i,j+1}} & & \\ \frac{\partial G_{i+1,j}}{\partial H_{i,j}} & \frac{\partial G_{i+1,j}}{\partial H_{i+1,j}} & \frac{\partial G_{i+1,j}}{\partial H_{i+2,j}} & 0\ldots & \frac{\partial G_{i+1,j}}{\partial H_{i,j+1}} & \\ & & \ddots & \ddots & \ddots & \\ \frac{\partial G_{l,k}}{\partial H_{l,k01}} & 0\ldots & \frac{\partial G_{l,k}}{\partial H_{l-1,k}} & \frac{\partial G_{l,k}}{\partial H_{l,k}} & \frac{\partial G_{l,k}}{\partial H_{l+1,k}} & 0\ldots \\ & & & \ddots & \ddots & \ddots \\ & & \frac{\partial G_{n,n}}{\partial H_{n,n-1}} & 0\ldots & \frac{\partial G_{n,n}}{\partial H_{n-1,n}} & \frac{\partial G_{n,n}}{\partial H_{n,n}} \end{bmatrix} \tag{2.74}$$

which can be evaluated as:

$$\begin{aligned} \frac{\partial G_{i,j}}{\partial H_{i-1,j}} &= -\frac{T_{i-1/2,j}}{\Delta x_{i,j}\Delta x_{i-1/2,j}} \\ \frac{\partial G_{i,j}}{\partial H_{i+1,j}} &= -\frac{T_{i+1/2,j}}{\Delta x_{i,j}\Delta x_{i+1/2,j}} \\ \frac{\partial G_{i,j}}{\partial H_{i,j-1}} &= -\frac{T_{i,j-1/2}}{\Delta y_{i,j}\Delta y_{i,j-1/2}} \\ \frac{\partial G_{i,j}}{\partial H_{i,j+1}} &= -\frac{T_{i,j+1/2}}{\Delta y_{i,j}\Delta y_{i,j+1/2}} \\ \frac{\partial G_{i,j}}{\partial H_{i,j}} &= \frac{S'\left(H_{i,j}^{n}\right)}{\Delta t}+\frac{K_l}{\Delta z_l} \\ &-\left(\frac{\partial G_{i,j}}{\partial H_{i-1,j}}+\frac{\partial G_{i,j}}{\partial H_{i+1,j}}+\frac{\partial G_{i,j}}{\partial H_{i,j-1}}+\frac{\partial G_{i,j}}{\partial H_{i,j+1}}\right) \end{aligned} \tag{2.75}$$

The Newton update is obtained by solving the linear system of equations:

$$
\begin{aligned}
&\mathbf{J}\left(H^{p}\right)\delta^{p} = -\mathbf{G}\left(H^{p}\right)\\
&G_{i,j}\left(H^{p}\right) = \frac{S\left(H^{p}_{i,j}\right) - S\left(H^{n}_{i,j}\right)}{\Delta t} + \frac{K_{l}\left(H^{p}_{i,j} - H^{n}_{l}\right)}{\Delta z_{l}} - R^{n}_{i,j} - W^{n}_{i,j}\\
&-\left[\begin{matrix}\dfrac{\partial G_{i.j}}{\partial H_{i-1,j}}\left(H^{p}_{i-1,j} - H^{p}_{i,j}\right) + \dfrac{\partial G_{i,j}}{\partial h_{i+1,j}}\left(H^{p}_{i+1,j} - H^{p}_{i,j}\right) + \\ \dfrac{\partial G_{i,j}}{\partial H_{i,j+1}}\left(H^{p}_{i,j+1} - H^{p}_{i,j}\right) + \dfrac{\partial G_{i,j}}{\partial H_{i,j-1}}\left(H^{p}_{i,j-1} - H^{p}_{i,j}\right)\end{matrix}\right]\\
&H^{p+1} = H^{p} + \delta^{p}
\end{aligned}
\tag{2.76}
$$

where p is the iteration number. Since J is symmetrical and positive definite, it can be solved, again, using the conjugate gradient method. We keep computing the solution to equation (2.76) until H converges. Newton iteration converges quadratically when the initial guess is close enough to the solution. Normally it converges in 2~4 iterations.

2.3.7. Snowpack model.

The UEB snowpack model [*Luce et al.*, 1999; *Luce and Tarboton*, 2004] is adapted as the snow module of the proposed model. Some coding errors in the original package have been corrected. The UEB model keeps track of snow water equivalent and energy content of the snowpack:

$$
\begin{aligned}
&\frac{dU}{dt} = Q_{sn} + Q_{li} + Q_{p} + Q_{g} - Q_{le} + Q_{h} + Q_{e} - Q_{m}\\
&\frac{dSWE}{dt} = P_{r} + P_{s} - M_{r} - E
\end{aligned}
\tag{2.56}
$$

In the energy balance equation (all per unit of horizontal area i.e. in kJ $m^{-2}hr^{-1}$) terms are: Q_{sn}, net shortwave radiation; Q_{li}, incoming longwave radiation; Q_{p}, advected heat from precipitation; Q_g, ground heat flux; Q_{le}, outgoing longwave radiation; Q_h, sensible heat flux; Q_e, latent heat flux due to sublimation/condensation; and Q_m, advected heat removed by meltwater. In the mass balance equation (all in m/hr of water equivalence) terms are: SWE: snow water equivalent (m), P_r, rainfall rate (m/hr); P_s, snowfall rate (m/hr); M_r, meltwater outflow from the snowpack (m/hr); and E, sublimation from the snowpack (m/hr) [*Luce and Tarboton*, 2004].

2.4. Interactions between domains

The above sections (Sec. 2.3) complete the descriptions of individual solvers. The hydrologic system is a fully coupled system such that problems in each domain cannot be studied without interactions with other domains. In this section we discuss how these different domains are coupled in the model. We will see that the vadose zone plays a central role in the model by solving runoff and recharge (also ET, although this has already been discussed in Sec. 2.3) together with soil moisture. We will detail how we decompose the 3D subsurface flow into a combination of a 2D and an array of 1D problems, while efficiently obtaining physically consistent solutions. This has been a major roadblock in physically-based hydrologic modeling.

2.4.1. Coupling of the Unsaturated Richards equation and the groundwater flow equation:

Water movement in the three-dimensional subsurface domain can be described by a single three-dimensional Richards' equation:

$$
\begin{aligned}
&\left(C\left(h\right)+S_s\right)\frac{\partial h}{\partial t}=W(h)\\
&+\left[\frac{\partial}{\partial z}\left(K_z(h)\left(\frac{\partial h}{\partial z}+1\right)\right)+\frac{\partial}{\partial x}\left(K_x(h)\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_y(h)\frac{\partial h}{\partial y}\right)\right]
\end{aligned}
\tag{2.56}
$$

where h is the pressure head [m], K_X, K_Y, K_Z are hydraulic conductivities [md^{-1}], C(h)=$\partial\theta / \partial h$ is the differential water capacity [m^{-1}], S_S is the specific storage [m^{-1}], W(h) is the volumetric source or sink term [d^{-1}]. Due to the strong non-linearity of unsaturated conductivity and differential water capacity, a large matrix resulting from 3-D discretization must be solved iteratively. Applying this approach on a watershed scale is computationally unfeasible. There is some research going in the direction of directly solving this 3D equation [*Maxwell and Kollet*, 2008]. We note that even with the help of massively-parallel super computers, the solution to Eq (2.56) in forward mode is still difficult, not to mention long-term simulations in the context of a watershed hydrologic model and model calibrations.

To cope with this difficulty, we seek to lower the dimensionality of Eq. (2.56) by separating it into the one dimensional Richard's equation that governs the unsaturated portion above the unconfined aquifer:

$$
C\left(h\right)\frac{\partial h}{\partial t}=\frac{\partial}{\partial z}\left[K(h)\left(\frac{\partial h}{\partial z}+1\right)\right]+W(h)
\tag{2.56}
$$

and the two-dimensional groundwater flow equation that governs flow in the saturated unconfined aquifer:

$$S\frac{\partial H}{\partial t}=\frac{\partial}{\partial x}\left[T\left(\frac{\partial H}{\partial x}\right)\right]+\frac{\partial}{\partial y}\left[T\left(\frac{\partial H}{\partial y}\right)\right]+R+W-D \qquad (2.63)$$

To avoid confusion, here we use H to represent the hydraulic head (and thus also the location of the water table of the unconfined aquifer). Eq. (2.63) has been used ubiquitously to describe flow in unconfined aquifers [e.g., Modflow]. We want to find a method that solves the two above equations and couples their solutions to approximate the 3D system.

Many previous studies have attempted such a simplification. One approach is to solve Equations (2.56), taking the solution to equation (2.63) from the last time step as the lower boundary condition. Several examples of this approach are [*Downer and Ogden*, 2004b; *Jia et al.*, 2001; *Twarakavi et al.*, 2008]. This method implicitly assumes that the head in the unsaturated zone does not change throughout the time step. However, such a scheme is not accurate when recharge from the vadose zone or lateral flow of groundwater changes the location of the water table. When recharge/discharge is large during one time step, the method may cause numerical instability because it neglects the head change of the water table induced by recharge itself. In addition, a theoretical inconsistency could occur as a head difference is always required to maintain recharge to the aquifer. We will illustrate this in the next section in more detail.

Another method is to iteratively solve the two equations until their solutions converge. [*Stoppelenburg et al.*, 2005; *van Dam et al.*, 2008] coupled the two equations via the deep percolation flux at the bottom of the saturated domain and the vadose zone model. When iteration is involved on the equation level, the method can be much

more expensive. Moreover, the coupling is ad-hoc as apparently, if an impervious layer sits at the bottom and the deep percolation is zero, such a coupling is not possible. Here we seek a method that does not use equation-level iterations.

Before we discuss the new coupling method, we introduce below a test case to better illustrate the method. This case closely resembles the widely known experiment first reported in [*Vauclin et al.*, 1979] and later frequently studied in the literature (e.g. [*Dogan and Motz*, 2005; *Twarakavi et al.*, 2008]). As shown in Fig 2.4, the flow domain consists of a rectangular soil slab with width L and height Z. Due to homogeneity in the y direction, the problem reduces to a 2D problem with x and z axes. z is 0 at the bottom of the unconfined aquifer (saturated flow domain) and positive upward. Initially the soil has established hydrostatic equilibrium with the water level on the left and right boundaries. Then a constant inflow is applied at the center of the top surface from X_L to X_R (X_L=-X_R). The rest of the top surface is covered to prevent evaporation. Water level on the boundaries is maintained at a steady height of H_0. After the wetting front reaches the water table, the pore spaces are filled up and the water table starts to rise. We can consider the saturated zone as being 'recharged'. As the head difference is created, water flows to the sides, brings up the water table at locations away from the inflow zone and finally exits from the boundaries. After some time, a steady state will be reached when the water table level becomes static (if the inflow rate does not exceed a certain threshold). Analyzing this steady state situation helps shed light into the problem. At steady state, we must have inflow rate that is equal to the outflow rate. Denoting the water table height as HW(x), we must have:

$$K\left(\frac{\partial h}{\partial z}+1\right)\Bigg|^{z=HW(x)} = Q_{in} \qquad (2.77)$$

Keeping in mind that HW is a function of the location x, we will simply write HW in the rest of the section for shorter notation.

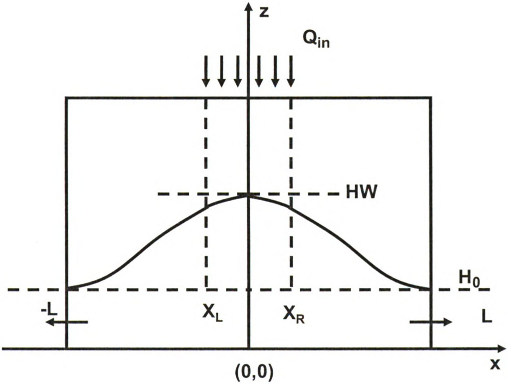


Figure 2.4. sketch of the Vauclin 1979 test problem. HW(x) is the water table location at x, z is measured 0 at the bottom.

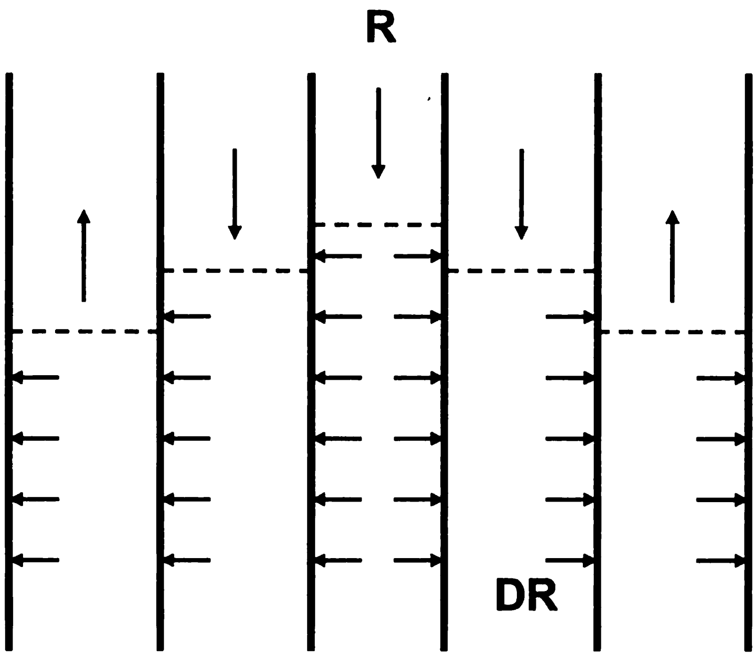


Figure 2.5. Illustration of Assumptions 1 and 2

The method is based on two assumptions describe below and illustrated in Fig 2.5. The first assumption is that water moves only vertically in the unsaturated zone. Denoting the lateral flow into a unit volume of the soil column as q_r, with the unit $[T^{-1}]$, this can be written as:

Assumption 1:

$$q\left(x,z\right)=0\text{, z>HW} \tag{2.78}$$

This assumption is used by many and can be regarded as simplifying Equation (2.56) by replacing $\frac{\partial}{\partial x}\left(K_x(h)\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_y(h)\frac{\partial h}{\partial y}\right)$ with q=0 in the unsaturated portion of the flow domain. The rationale is that in the unsaturated domain gravity dominates over lateral moisture diffusion.

The next assumption is that the lateral flow is uniformly distributed along the saturated thickness of the unconfined aquifer:

Assumption 2:

$$\frac{\partial q_r}{\partial z} = 0 \text{ and } \int_0^{HW(x)} q_r(x,z)dz = DR(x), \text{z<=HW} \tag{2.79}$$

where DR(x) is the integrated lateral drainage term (m/day), positive for inflow. With respect to Eq. (2.63), we see that:

$$\begin{cases} q_r = \frac{\partial}{\partial x}\left[K\left(\frac{\partial H}{\partial x}\right)\right] + \frac{\partial}{\partial y}\left[K\left(\frac{\partial H}{\partial y}\right)\right] \\ DR = \frac{\partial}{\partial x}\left[T\left(\frac{\partial H}{\partial x}\right)\right] + \frac{\partial}{\partial y}\left[T\left(\frac{\partial H}{\partial y}\right)\right] \end{cases} \tag{2.80}$$

And we can denote the drainage flux, simply solved from (2.79), as

$$q_r(z) = \frac{DR}{HW}, z \leq HW \tag{2.81}$$

Assumption 2 is a natural result of the Dupuit-Forchheimer (D.F.) assumption, which assumes the flow lines to be parallel. This condition has been used by many models to apply groundwater flow terms in the Richards Equation, e.g. [*van Dam et al.*, 2008]. However, the original D.F. assumption also states that the hydraulic gradient is zero in the vertical direction of the saturated zone:

$$H(x,z) = HW, \text{z<=HW} \tag{2.82}$$

The paradox of the D.F. assumption is well known (e.g. [*Kirkham*, 1967]). Although often providing acceptable solutions to unconfined aquifer flow problems, Eq. (2.82) creates a theoretical inconsistency for our coupling approach here and therefore needs modifcation. To see this, we need only look at the mass balance of a unit volume at steady state, using Eq. (2.56), (2.80) and assuming W=0:

$$\left(C\left(h\right)+S_s\right)\frac{\partial h}{\partial t}=0=\left[\frac{\partial}{\partial z}\left(K_z\frac{\partial H}{\partial z}\right)+q\right] \tag{2.83}$$

As a result:

$$-q_r=\frac{\partial}{\partial z}\left(K_z\frac{\partial H}{\partial z}\right) \tag{2.84}$$

Thus Eq. (2.82) cannot be directly applied. For the steady state case, consider deep percolation (q_{bot}, m/day, positive for downward percolation), the above equation forms a second order Boundary Value Problem (BVP) that can be quite easily solved. The boundary conditions are:

$$\begin{cases} K_z\left(\dfrac{\partial h}{\partial z}+1\right)\Bigg|^{z=0} = q_{bot} \\ K_z\left(\dfrac{\partial h}{\partial z}+1\right)\Bigg|^{z=HW} = DR+q_{bot} \\ h(x,HW)=0 \end{cases} \tag{2.85}$$

And the solution is:

$$h(x,z)'=\left(1-\frac{q_{bot}}{K_z}\right)\left(HW-z\right)+\frac{1}{2}\frac{DR}{K_z}\left(HW^2-z^2\right) \tag{2.86}$$

where h(x,z)' is the steady state pressure head (m), K_z is the vertical saturated hydraulic conductivity (m/day), HW is the water table location(m), z is the vertical coordinate and DR is the integrated lateral drainage flux (m/day), positive for inflow. As we can see, h is a quadratic function of HW when head loss due to drainage is considered instead of the linear relationship assumed in Eq. (2.82). Figure 2.6 shows h as a function of z at different DR values. Clealy h(x,z) is less than HW-z when DR is negative (water draining away). Although (2.86) is derived from the steady state case, we will still use it for the unsteady case to provide a functional relationship between the pressure head at z and the water table location HW.

Previous studies have directly used Eq. (2.82) to provide a Dirichlet boundary condition to the Richards equation. An important implication of this approach is that, even at steady state, the saturated/unsaturated interface found in the soil profile must be higher than HW because a head difference must exist in the soil profile to supply the required recharge to the saturated zone. Thus it is almost certain that the moisture profile will be erroneous because the saturated/unsaturated interface should be placed right at the groundwater table. We will show that Eq. (2.86) allows the moisture distribution in the vadose zone to agree with the water table location and thus provides an improvement over the previous studies.

There is yet another challenge facing the approximation of (2.56) with (2.56) and (2.63) which is the storage term S in (2.63). The specific yield of the unconfined aquifer is a loosely defined concept. Non-existent in the full 3D Richards equation, its use in the groundwater flow equation is largely for practical purposes. The specific yield of a Boussinesq aquifer in theory is the effective porosity, or drainable porosity. However, aquifers often take exceedingly long time (sometimes several years) to release water for this value to be meaningful [*King*, 1899]. Some interesting discussions of the drainable porosity concept can be found in [*Hilberts et al.*, 2005]. However, the value of the specific yield, if defined as the water yield as a result of unit decline of the head, is also dependent on the transient moisture conditions of the unsaturated soil column above and the depth to water table. Since considerable uncertainty remains as to the evaluation of S [*Healy and Cook*, 2002], the specific yield is often kept as a calibration parameter in groundwater models to match observations. We propose linking the specific yield to the drainable porosity defined in [*Hilberts et al.*, 2005] as:

$$S_y = \sigma f \tag{2.87}$$

where f is the drainable porosity, because field capacity represents the ability of the soil to hold water under normal field conditions.

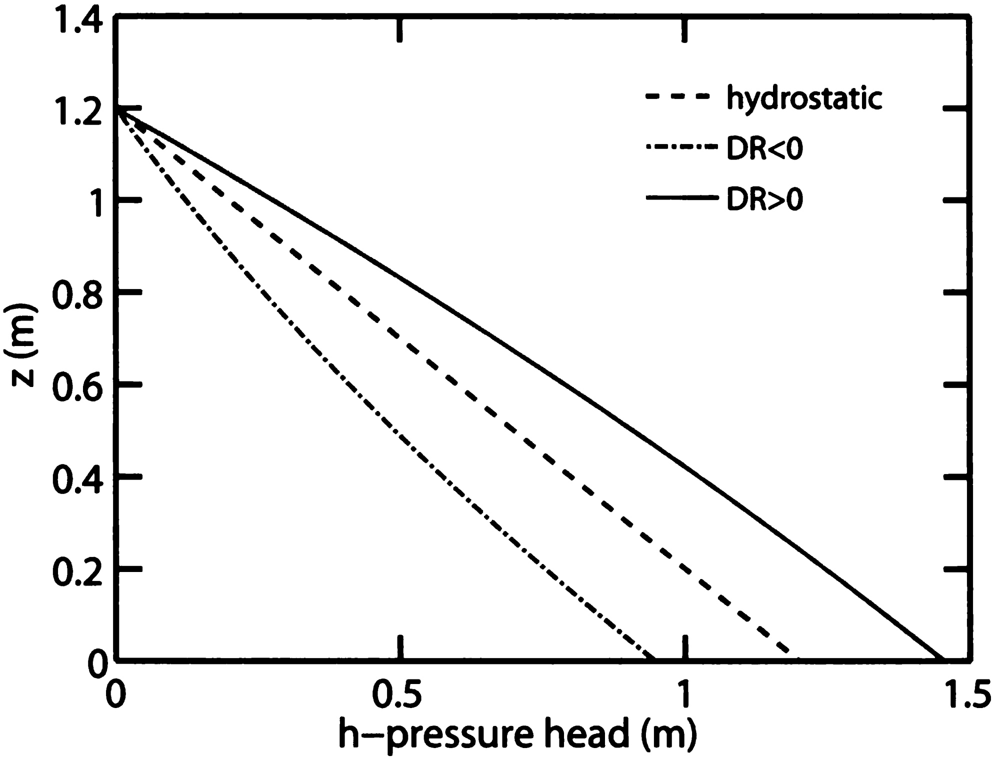


Figure 2.6. h(x,z) as a function of z and DR values.

With the above two assumptions and equation (2.86) we now detail our coupling scheme. First, assumption 1 allows us to discretize the 3D unsaturated flow domain into an array of 1D columns connected to the 2D saturated flow domain at the bottom, as shown in Figure 2.5b. Each one of these columns can now be governed by Eq. (2.56). As the groundwater fluctuates, the water table may progress or retreat from soil column cells, but must share at least one cell with the soil columns. Assuming elastic storage S_S is minor compared to the specific moisture capacity C(h), the 3 dimensional Richards equation (2.56) is re-written for the soil column as:

$$C(h)\frac{\partial h}{\partial t} = \frac{\partial}{\partial z}\left(K_z(h)\left(\frac{\partial h}{\partial z} + 1\right)\right) + q_r + W(h) \tag{2.88}$$

This equation is simply the 1D Richards Equation (2.56) with the addition of a drainage flux term q_r, thus it is treated the same way as described in Section 2.3.5. With Assumptions 1 and 2, we have:

$$q_r(x,z) = \begin{cases} 0 & z > HW \\ \dfrac{DR}{HW(x)} & z \le HW \end{cases} \tag{2.89}$$

This last cell of the soil column serves as the linkage between the two flow domains. We use this cell to represent the unconfined aquifer and bring the dynamics of the groundwater aquifer into the soil column. The upper boundary of this cell is marked z_u. The center of this cell should always be below the water table. A mass balance equation is written for this cell appears as shown below:

$$
\begin{aligned}
S\frac{\partial h}{\partial t} = \int_{0}^{z_u}\left[\frac{\partial}{\partial x}\left(K_x\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial y}\left(K_y\frac{\partial h}{\partial y}\right)\right]dz\, + \\
\left(K_z\left(\frac{\partial h}{\partial z}+1\right)\Bigg|^{z_u} - K_z\left(\frac{\partial h}{\partial z}+1\right)\Bigg|^{0}\right) + \int_0^{z_u} W dz
\end{aligned}
\qquad (2.90)
$$

Applying Eq. (2.80) and re-arranging terms, this equation actually looks similar to Eq. (2.63):

$$
S\frac{\partial h}{\partial t} = K_z\left(\frac{\partial h}{\partial z}+1\right)\Bigg|^{z_u} - K_z\left(\frac{\partial h}{\partial z}+1\right)\Bigg|^{0} + \frac{z_u}{HW}DR + \int_0^{z_u} W dz \qquad (2.91)
$$

By writing down this equation, it seems we have conceptualized the bottom cell as a layer that has some storage capability. However, as we have neglected S_S, the storage is obviously 0 in this cell. The use of S is to mimic the reaction of the unconfined aquifer, whose stage will rise after receiving recharge. Under this context, the RHS of the equation (2.91) represents the net inflow into the unconfined aquifer. Thus we can understand S in another way by linking to the specific yield of the unconfined aquifer:

$$
\begin{aligned}
S\frac{dh}{dt} = S_y\frac{dHW}{dt} = S_y\frac{dHW}{dh}\frac{dh}{dt} \\
\text{Thus, } S = S_y\frac{dHW}{dh}
\end{aligned}
\qquad (2.92)
$$

And dHW/dh can be obtained from differentiating Eq. (2.86).

The next step is to solve Eq. (2.91) for this cell together with the unsaturated zone Eq. (2.88). Applying a semi-implicit temporal discretization to Eq. (2.91), we obtain:

$$\frac{S_{nz}}{\Delta t}\left(h_{nz}^{j+1} - h_{nz}^{j}\right) = K_{nz-1/2}\left(\frac{h_{nz-1}^{j+1} - h_{nz}^{j+1}}{\Delta z_{nz-1/2}} + 1\right)$$
$$-K_l\left(\frac{h_{nz}^{j+1} - h_l^{j}}{\Delta z_l} + 1\right) + \frac{z_u}{HW} DR + W_{nz}^{j} \tag{2.93}$$

in which, nz is the index of the bottom cell, S_{nz} is the storativity (dimensionless) as in Eq. (2.92), h_l^N is the pressure head of the confined aquifer that is sitting beneath the unconfined aquifer, K_l is the hydraulic conductivity of the aquitard between the unconfined and confined aquifers. We treat h_l^N and DR explicitly because their temporal changes tend to be much smaller than the rest of the terms. To be consistent with Eq. (2.62), here we have used j to denote the time level. This equation is linked to the state of the groundwater by the inclusion of h_{nz}^N, which we take from Eq. (2.86):

$$h_{nz}^{j} = h(x, z_{nz})' = \left(1 - \frac{q_{bot}}{K_z}\right)\left(HW - z_{nz}\right) + \frac{1}{2}\frac{DR}{K_z}\left(HW^2 - z_{NZ}^2\right) \tag{2.94}$$

Note that both HW and lateral flow term DR have already been computed in the groundwater flow model in (2.64), lagged by one time step. Thus Eq. (2.93) presents a closure to the vadose zone model (2.61). Writing (2.93) in the same form as (2.62):

$$\begin{aligned} &a_{nz}^{j} h_{nz-1}^{j+1.p} + b_{nz}^{j} h_{nz}^{j+1.p} = d_{nz}^{j} \\ &a_{nz}^{j} = -\frac{K_{nz-1/2}^{j}}{\Delta z_{nz-1/2}}, b_{nz}^{j} = \frac{S_{nz}}{\Delta t} - a_{nz}^{j} + \frac{K_l}{\Delta z_l} \\ &d_{nz}^{j} = \frac{S_{nz}}{\Delta t} h_{nz}^{j} + \left(\frac{h_l^N K_l}{\Delta z_l}\right) + \left(K_{nz-1/2} - K_l\right) + \frac{z_u}{HW} DR + W_{nz}^{j} \end{aligned} \tag{2.95}$$

as in (2.62), p denotes iteration number. Note the dimension of (2.95) is different from

(2.62). After this equation is solved, we calculate the recharge as the water leaving the bottom of the soil column:

$$R = \int_{z_u}^{HW} q_r dz + K_{nz-1/2} \left(\frac{h_{nz-1}^{j+1} - h_{nz}^{j+1}}{\Delta z_{nz-1}} + 1 \right) \tag{2.96}$$

If the vadose zone model and the groundwater model share the same temporal step size, Eq. (2.96) can be directly fed into (2.63). However since we allow adaptive time steps for the vadose zone model, R needs to be integrated over time and divided by the time step of the groundwater flow model to be consistent in dimension. The coupling procedure can be summarized as follows:

1. At the beginning of each time step, extract water table depth HW, and drainage flux DR from the groundwater flow solution from last time step
2. Use Eq. (2.94) and (2.89) to calculate h_{nz} and q_r.
3. Solve the system resulting from (2.88) and (2.91) as described in Section 2.3.5.
4. Calculate R with (2.96), integrate it over time, and divide the result by the time step of the groundwater solver
5. With the predicted R, solve (2.63) as described in 2.3.6
6. Calculate DR
7. Go to step 1 and repeat

In essence, we are using the last cell (in some computational areas called the ghost cell) to emulate the behavior of the groundwater system and make a prediction, albeit a more reliable one, about the flux between the vadose zone and the saturated zone.

Thus iteration on the equation level is avoided. Although the head in this last cell is updated while solving the Richards' Equation in the vadose zone, its final state is not important. Only the flux through the interface is used later to calculate recharge in the groundwater model (2.64). The more accurate the emulation, the less the error would be. Our advantage at hand is that the groundwater system is well-known to be steady, low frequency one such that equation (2.80) calculated from last time step is usually a good approximation to DR. This way we have decoupled two systems to enable large scale simulation, while retaining the salient nature of interactions between the two compartments.

Here we must comment that the coupled 1D/2D system will not completely reproduce the behavior of the 3D equations, especially in places where lateral diffusion of soil moisture is important on the scale of interest. However, the coupled system is a viable alternative in watershed scale modeling where fully 3D solvers are not plausible, even in the foreseeable future. In later sections we validate this approach using an experiment dateset.

2.4.2. Coupling of the vadose zone and surface flow:

Similar to the lower boundary condition, we write a separate mass balance equation for the upper boundary cell, which is the ground surface storage layer:

$$\frac{dh_1}{dt} = \left(P - E_1\right) - K_{1/2}\left(\frac{h_1 - h_2}{\Delta z_{1/2}} + 1\right) - F_g \tag{2.97}$$

where, h_1 is the surface ponding depth, P is precipitation (m/day), E_1 is the surface evaporation (m/day), F_g is the surface runoff (m/day) and $K_{1/2}$ is the surface hydraulic conductivity, which is calculated as the geometric mean of the saturated current-state unsaturated conductivity of the first soil layer (layer 2), considering the fraction of pervious area:

$$K_{1/2} = \left(1 - f_{imp}\right)\sqrt{K_1 KS_1} \tag{2.98}$$

The surface runoff is the contribution from surface ponding to various flow paths. If we assume the total length of overland flow paths in the cell as l, similar to [*Panday and Huyakorn*, 2004], the surface runoff contribution can be computed as

$$F_g\left(h_1\right) = \frac{Q_g}{A} = \frac{\left(h_1 - h_0\right) u_l l}{86400A} \tag{2.99}$$

Where Q_g is the contributing discharge from surface storage to overland flow paths (m/day), u_l is the flow velocity (m/s), A is the area of the cell (m^2), h_0 is a minimum depth of water for surface runoff to occur (m) and the constant 86400 is for unit conversion from m/s to m/day. Here we conceptualize the runoff contributing process as flow on a rectangular plane and let l be calculated as:

$$l = \frac{A}{d_l} \tag{2.100}$$

where d_l is the average distance to the nearest flow paths (m). We use the Manning's formula to compute flow velocity and apply the kinematic wave concept (S_0=S_f):

$$u_l = \frac{1}{n}\left(h_1 - h_0\right)^{2/3} S_0^{1/2} \tag{2.101}$$

where S_0 is the average slope of the cell and n is the manning's roughness coefficient. Then:

$$F_g\left(h_1\right)=\frac{Q_g}{A}=\frac{86400}{nd_l}\left(h_1-h_0\right)^{5/3}S_0^{1/2} \qquad (2.102)$$

h_0 is the depth of surface depression storage, and may also be interpreted as part of the initial abstractions. Eq. (2.97) is a nonlinear equation that can be solved iteratively together with the rest of the soil profile. In order to achieve faster convergence, we partially linearize the F_g term and the above equation is re-written as:

$$F_g\left(h^{j+1,p}\right)=\begin{cases} SN\cdot\left(h_1^{j+1,p}-h_0\right)\left(h_1^{j+1,p-1}-h_0\right)^{2/3}, & h_1^{j+1,p-1}>h_0 \\ 0, & h_1^{j+1,p-1}\le h_0 \end{cases}$$
$$SN=\frac{86400S_0^{1/2}}{nd_l}$$

(2.103)

where j and p are time level and iteration number as discussed in (2.61), k_o is the surface runoff coefficient. With this runoff formulation, Eq. (2.97) is discretized as:

$$\begin{aligned} &b_1^j h_1^{j+1,p}+c_1^j h_2^{j+1,p}=d_1^j \\ &c_1^j=-\frac{K_{1/2}^j}{\Delta z_{1/2}}, b_1^j=\frac{1}{\Delta t}-c_1^j+k_o \\ &d_1^j=W_1+\frac{h_1^j}{\Delta t}-K_{1/2}+k_o h_0 \\ &k_o=SN\cdot\left(h_1^{j+1,p-1}-h_0\right)^{2/3} \end{aligned} \qquad (2.104)$$

When saturation excess occurs, i.e., the soil column is saturated and $Q_{in}>0$, the groundwater and the surface water are connected. In this case, the model will solve the coupled equations (2.97) and (2.91):

$$\frac{dh_1}{dt} = \left(P - E_1\right) - K_T\left(\frac{h_1 - h_{nz}}{\Delta z_T} + 1\right) - F_g\left(h\right) + \int_{z_t}^{z_0} W dz$$
$$S\frac{\partial h_{nz}}{\partial t} = K_T\left(\frac{h_1 - h_{nz}}{\Delta z_T} + 1\right) - K_z\left(\frac{\partial h}{\partial z} + 1\right)\Bigg|^{z_b} + Q_{Dr} + \int_{z_b}^{z_t} W dz$$

(2.105)

where, Δz_T is the total depth of the vadose zone, K_T is the total average conductivity of the vadose zone, and other notations are explained with Eq. (2.97) and (2.91). To improve efficiency, we assume that when saturation excess occurs, root extractions can be instantaneously supplied by the excess water, thus the collective source term in the soil profile, $\int_{z_t}^{z_0} W dz$, is lumped into the top cell equation. Using the same modeling approach discussed above, we have:

$$\frac{\left(h_1^{N+1} - h_1^{N}\right)}{\Delta t} = \left(P - E_1\right) - K_T\left(\frac{h_1^{N+1} - h_{nz}^{N+1}}{\Delta z_T} + 1\right) - F_g\left(h_1^{N+1}\right) + \sum_{i=2}^{nz-1} \Delta z_i W_i$$
$$S\frac{\left(h_{nz}^{N+1} - h_{nz}^{N}\right)}{\Delta t} = K_T\left(\frac{h_1^{N+1} - h_{nz}^{N+1}}{\Delta z_T} + 1\right) - K_l\left(\frac{h_{nz}^{N+1} - h_l^{N}}{\Delta z_l} + 1\right) + DR + W_{nz}$$

(2.106)

By solving for h_{nz} from the second equation:

$$h_{nz}^{N+1} = \frac{\frac{S}{\Delta t}h_{nz}^{N} + \frac{K_T}{\Delta z_T}h_1^{N+1} + \frac{K_l}{\Delta z_l}h_l^{N} + K_T - K_l + Q_{Dr} + W_{nz}}{\frac{S}{\Delta t} + \frac{K_T}{\Delta z_T} + \frac{K_l}{\Delta z_l}}$$

(2.107)

and substituting it back into the first one, we obtain a nonlinear equation:

$$h_1^{N+1} = p_1\left[p_2 - F_g\left(h_1^{N+1}\right)\right] \tag{2.108}$$

With

$$\begin{aligned}
p_1 &= \frac{1}{\frac{1}{\Delta t} + \gamma_2 - \gamma_2\gamma_4} \\
p_2 &= \frac{h_1^N}{\Delta t} + \left(P - E_1\right) + \sum_{i=2}^{nz-1} \Delta z_i W_i - K_T + \gamma_4\left(\gamma_1 h_{nz}^N + \gamma_3 h_l^N + \gamma_5\right) \\
\gamma_1 &= \frac{S}{\Delta t}, \gamma_2 = \frac{K_T}{\Delta z_T}, \gamma_3 = \frac{K_l}{\Delta z_l}, \gamma_4 = \frac{\gamma_2}{\gamma_1 + \gamma_2 + \gamma_3}, \gamma_5 = K_T + Q_{Dr} + W_{nz} - K_l \\
K_T &= harmonicMean(K)
\end{aligned} \tag{2.109}$$

This equation can be quickly solved by either Picard or Newton iterative methods. Convergence is so fast with Newton iteration that it seldom takes more than 3 iterations.

2.4.3. Interaction between overland flow and channel flow

A zoomed-in sketch helps illustrate the exchange between the river and the land (Figure 2.7a). From the point of numerical accuracy, it is perhaps best to use an unstructured grid for the overland flow, enhancing the resolution and fit to the river orientation near the streams. However, many other concerns dictate over the computational accuracy in the flow model, e.g., hydrological response unit, integration of different model components, input data resolution and model generality. Therefore we choose to employ a structure grid for land discretization. Here we denote a reach of a river that is completely inside a land cell as a 'reach segment' (s). A segment enters from one edge of a land cell one and leaves from another (maybe the same edge) as shown in (Figure 2.7a). Because we also allow significant flexibility for the river discretization, the edges of river cells and segments usually do not correspond.

Two transformation matrices are pre-computed and stored to facilitate the calculation of exchange flux. TM records the fraction of each reach segment that belongs to each river cells; TN records the fraction of each river cell that belongs to each segment:

$$
\begin{aligned}
&\mathbf{TM} = \begin{bmatrix} TM_{1,1} & \cdots & TM_{1,ns} \\ & \ddots & \\ TM_{nr,1} & \cdots & TM_{nr,ns} \end{bmatrix} \\
&\mathbf{TN} = \begin{bmatrix} TN_{1,1} & \cdots & TN_{1,nr} \\ & \ddots & \\ TN_{ns,1} & \cdots & TN_{ns,nr} \end{bmatrix} \\
&TM_{j,i} = \frac{r_j \cap s_i}{s_i}, \left(i \in \{1,2,\ldots,ns\}, j \in \{1,2,\ldots,nr\}\right) \\
&TN_{j,i} = \frac{s_j \cap r_i}{r_i}, \left(i \in \{1,2,\ldots,nr\}, j \in \{1,2,\ldots,ns\}\right)
\end{aligned}
\tag{2.110}
$$

Where r_i, i= are the river cells, nr is the number of river cells of the river, s_i, i=(1,...,ns) are the reach segments, ns is the number of reach segments, and $\cap$ is the spatial join operator. TM and TN are both sparse matrices. With these two matrices, it is possible to project variables from river cell grid to the reach segment and vice versa. The length-weighted river stages of the segments can be calculated from the stages defined on the river cells:

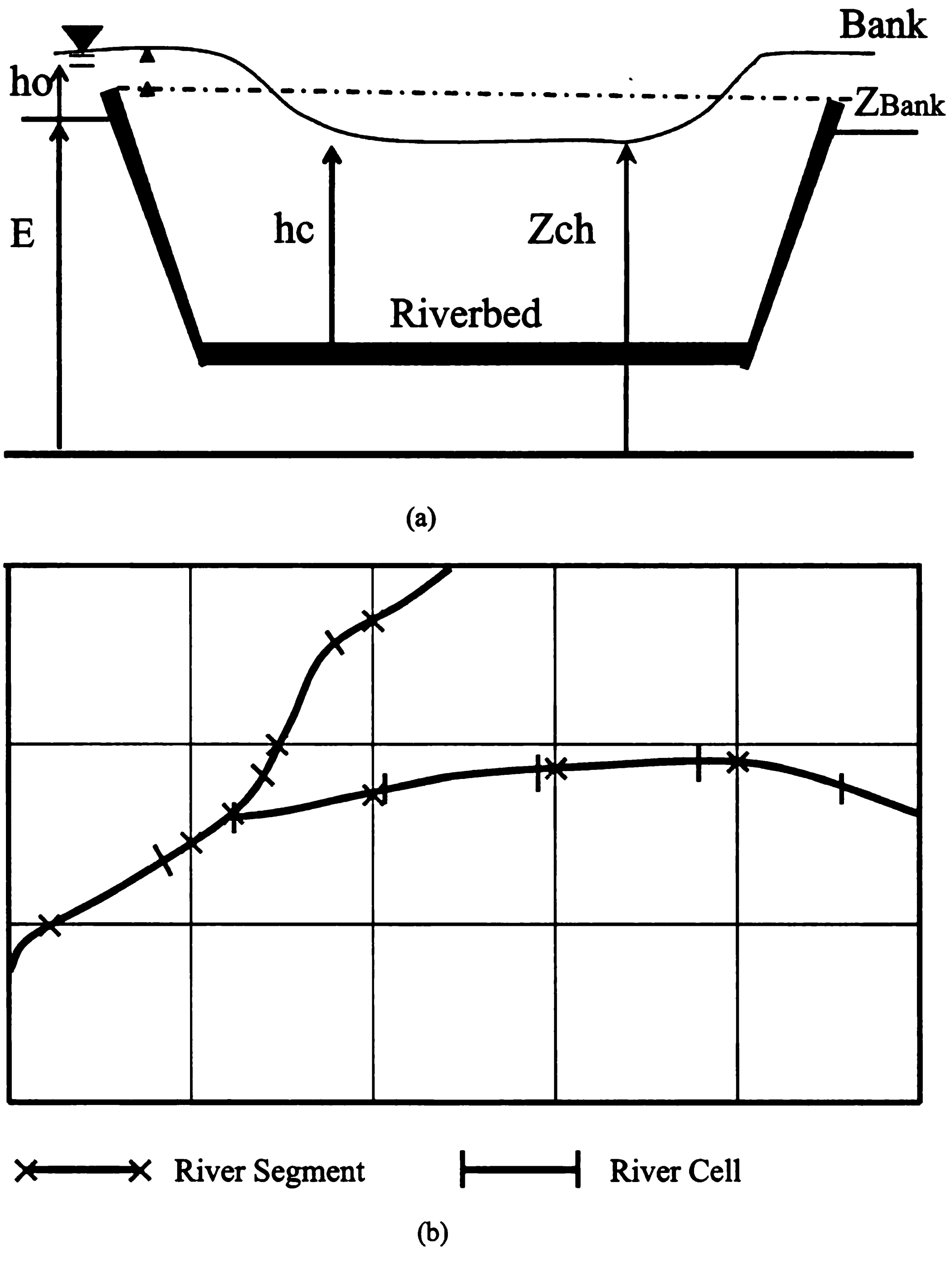
$$\left[Zch_1, Zch_2, \ldots, Zch_{ns}\right] = \left[\eta_1, \eta_2, \ldots, \eta_{nr}\right] \times \mathbf{TM} \tag{2.111}$$

Where Z_{ch} is the river stage of the reach segments, η are the free surface elevation of the river cells. After the exchange flux is computed for the reach segments using the method described below, it can be distributed to the river cells by:

$$\begin{bmatrix} q_{oc,1} \\ \vdots \\ q_{oc,nr} \end{bmatrix} = \mathbf{TM} \times \begin{bmatrix} q'_{oc,1} \\ \vdots \\ q'_{oc,ns} \end{bmatrix} \tag{2.112}$$

in which q'_{oc} is the exchange flux (m^3/day) computed for the reach segments, q_{oc} is the lateral flow into river from overland flow for each river cell. An occasional situation is that one river may have several segments inside one land cell as it meanders in and out of the edges of the same cell. A correction pass is used to remove the redundantly calculated flux and ensure mass balance.

It has been proposed in [*Panday and Huyakorn*, 2004] that interaction between overland flow and channel can be modeled by the equations for flow over a wide rectangular weir. Similar to this approach, we developed an efficient and stable procedure to compute river/land exchanges on a physical basis. The cross-sectional and plane sketches of a river cell are given in Fig 2.7. Zbank is the elevation of the bank (m), ho is the depth of the overland flow (m), E is the average elevation of the cell (m), Z_o is the average free surface elevation of the cell (m) (Z_o = E+h_o), h_c is the channel flow depth and Z_{ch} is the stage of the channel (m).



(a)

(b)

Figure 2.7. sketch of a river cell in the model. (a). cross-sectional view, (b) plane view

The exchange mass M (m^3) between land and channel is computed with the following procedure:

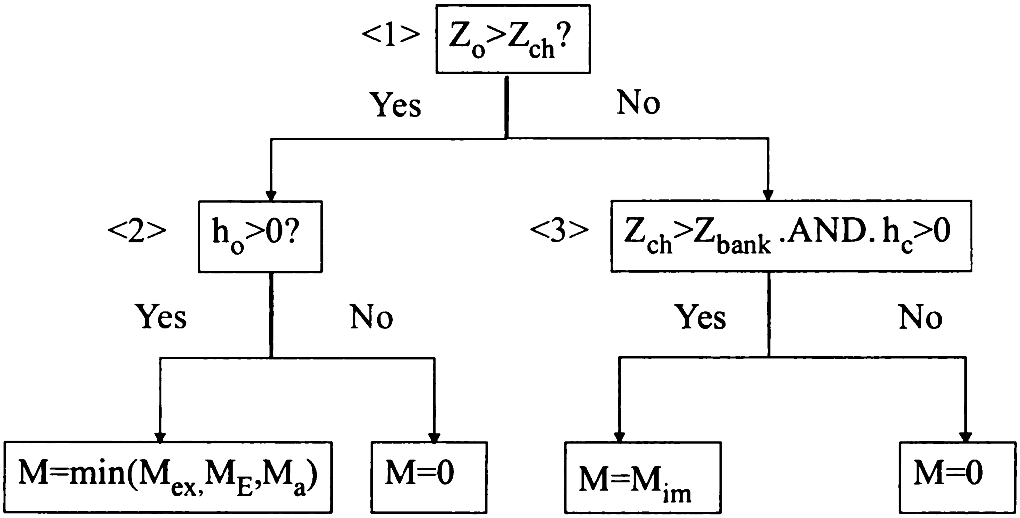


Figure 2.8. Calculation of River/Land exchange

In the above diagram, the first condition <1> is to determine the direction of the flow. If it is from the land to the river ($Z_o > Z_{ch}$) and there is water on the land (condition <2>), we first attempt to compute the contribution explicitly using the diffusive wave equation:

$$M_{ex} = \mathrm{sgn}(\frac{\partial \eta}{\partial x})\frac{2L_c\Delta t}{n}d_{oc}^{5/3}\left|\frac{\partial \eta}{\partial x}\right|^{1/2} = \frac{2L_c\Delta t}{n}h_o^{5/3}\left|\frac{Z_o - \max(Z_{ch}, Z_{Bank})}{(\Delta x / 2)}\right|^{1/2}$$

(2.113)

However, this flux cannot exceed the amount of currently available water on the land cell:

$$M_a = Ar \cdot h_o \tag{2.114}$$

Also, there is an equilibrium state, at which the river stage will be the same as the land free surface elevation. Denoting this stage as Z*, the mass transfer equation is written as:

$$\left(Z^{*}\text{-}Z_{ch}\right)A_{b}=\left(Z_{o}\text{-}Z^{*}\right)Ar \tag{2.115}$$

Where Ar is the area of the land cell, Ar=Δx*Δy and A_b is the area of the river cell that spans this land cell. Then we can find the Z*:

$$\begin{aligned}&\left(Ar+A_{b}\right)Z^{*}=Z_{o}Ar+Z_{ch}A_{b}\\&M_{E}=\left(Z^{*}\text{-}Z_{ch}\right)A_{b}=\left(\frac{Z_{o}Ar+Z_{ch}A_{b}}{\left(Ar+A_{b}\right)}\text{-}Z_{ch}\right)A_{b}=\frac{(Z_{o}\text{-}Z_{ch})A_{b}}{(1+A_{b}/Ar)}\end{aligned} \tag{2.116}$$

And thus the exchange mass will be the minimum of M_{ex}, M_E and M_a.

On the other hand, if the river stage rises higher than the land free surface elevation (and also the bank elevation), flooding would occur, which is solved using an implicit approach to enhance stability. Again using the diffusive wave formulation, the mass exchange function can be written in the form of two ordinary differential equations:

$$\begin{aligned}&Ar\frac{dZ_{o}}{dt}=-M_{im}=-\frac{2L}{n}\frac{Z_{ch}-Z_{o}}{\Delta x/2}\left(Z_{ch}-Z_{Bank}\right)^{5/3}\\&A_{b}\frac{dZ_{ch}}{dt}=-Ar\frac{dZ_{o}}{dt}\end{aligned} \tag{2.117}$$

Again these two equations can be solved by Picard or Newton iteration. The resulting scheme is very stable.

2.4.4. Interaction between groundwater and channel flow

The interactions between groundwater flow and channel flow is solved immediately after the channel flow step in section 2.3.4. We use the concept of operator splitting to couple the river flow and exchange with groundwater. After the river flow model is solved explicitly with Runge-Kutta approach, an implicit step is solved to calculate the exchange between streams and groundwater using the conductance concept [*Gunduz and Aral*, 2005]. The governing equation for the fractional step is written as:

$$\frac{dh}{dt} = \begin{cases} K_r \dfrac{h_g - h_r}{\Delta Z_b} & h_g > \left(z_b - \Delta Z_b\right) \\ K_r \dfrac{\left(z_b - \Delta Z_b\right) - h_r}{\Delta Z_b} & h_g \leq \left(z_b - \Delta Z_b\right) \end{cases} \tag{2.118}$$

in which, z_b is the river bed elevation (m), ΔZ_b is the thickness of the river bed material (m), K_r is the river bed conductivity (m/s) and h_g is the groundwater table elevation (m). The above equation can be discretized implicitly as:

$$\frac{h^{N+1} - h^*}{\Delta t} = K_r \frac{h_g^* - h^{N+1}}{\Delta Z_b} \tag{2.119}$$

h* is the computed from the solution to equation (2.51) and h* is max(h_g, z_b-Δz_b). We can solve h^{N+1} from this equation:

$$h^{N+1} = \frac{K_r\left(h_g^* - h^*\right)}{\dfrac{\Delta Z_b}{\Delta t} + K_r} \tag{2.120}$$

and then the lateral flow from groundwater is computed as:

$$q_{oc} = w\left(h^{N+1} - h^*\right) \tag{2.121}$$

q_{gc} is the lateral inflow from groundwater (m^2/s), w is the river width (m).

2.5. Data preparation steps general to all watersheds

A list of required input to create and drive a model is given in Table A.1. All of inputs except for the groundwater data are readily available for most of the U.S and can be processed by the program without much additional effort. Some data processing steps are general to all watersheds. They are described in this section. Creating the conceptual groundwater model (i.e. how many layers to model) requires some knowledge of the geological configuration. Required quantitative groundwater inputs include conductivities and thickness of the aquifers, which are not available in any readily-usable national database. Therefore the groundwater data may come from different sources for different regions and the processing is slightly more ad-hoc. For the model application in this dissertation, groundwater data source and processing are given in Sec. 3.1.1.

2.5.1. Digital elevation data

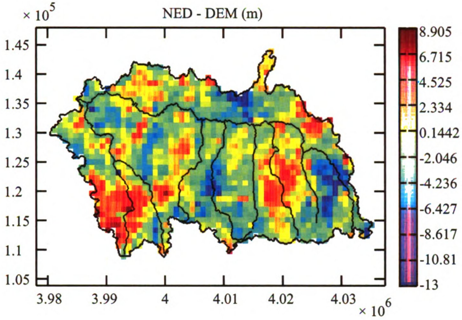


Figure 2.9. Discrepancies between NED and DEM when aggregated into the same grid

The digital elevation data are widely available in many formats and resolutions. In the U.S., the most commonly used elevation data include Digital Elevation Model (DEM) and National Elevation Dataset (NED). The DEM normally has a resolution of approximately 90m. The 30m resolution NED is available for most of U.S. and in some regions even with 10m resolution. Figure 2.9 shows the differences of the model grid elevation as a result of aggregating the DEM and the NED data using the mean method. As we can see, local discrepancies between the NED and DEM can be at times large. A general pattern is that NED is higher on the highlands than the DEM but comparatively lower on the lowlands.

2.5.2. River data processing

The River network is established from the National Hydrography Dataset (NHD). The NHD contains geographic information of streams as a collection of stream segments. Each segment is a short stream reach that records the spatial coordinates of its curve vertices with an attribute table detailing information including the ‘from’ and ‘to’ node of each river segment, the stream order and stream names if applicable. However, the dataset in its original form cannot be utilized by the model as all segments are listed as a relational table, not grouped by actual streams and do not exactly follow upstream-downstream sequence. So some effort is required to sort the data into a usable format. First, some pre-processing is done to identify the rivers that are going to be modeled, by using the stream order and stream name information. Unused river segments are removed from the data. The stream segments with the same stream order and same stream names are tagged with the same river id (RID). Then river segments with the same RID are extracted and grouped for further processing. Within one group of segments that belongs to a river, the segment that has no upstream reach is the head of the river. After we find this segment, we sequentially find its discharging segments using the information in the ‘to’ field until a clean list of segments from upstream to downstream is completed and the X-Y coordinates of the vertices are also sorted. Next, the river network is established by going through all rivers and recording their tributaries, downstream rivers, and locations at which they are joined by their tributaries. Finally, we establish a list that sorts all the rivers from upstream to downstream. This list is used by the watershed model so that it always runs the river models in an upstream-to-downstream sequence.

River bed elevation is found by subtracting the bankful river depth from the local elevation values (taken as the bank elevation). However, due to its relatively large grid size, DEM often has difficulty capturing thin features like streams and creeks. It is found that the elevation data extracted from the DEM tend be too high for the evaluation of bank elevations.

The bankful river depth data are not available in most cases and therefore must be established from empirical relations and field measurements. For the Red Cedar River we have measurements of water depth at bankful conditions from Acoustic Doppler Current Profiler (ADCP) surveys at various locations, as well as operational USGS instruments. These measurements are used to fit to the empirical relation of [*Bjerklie*, 2007]:

$$D = c\frac{w^{0.39}}{S0^{0.24}} \tag{2.122}$$

where w is the river width, S0 is the slope of the river and c is a fitting parameter. c is found to be 0.027 from the Red Cedar River measurements and applied to all reaches of the rivers. While it is possible to use other approaches to estimate the bankful river depth, the simple relation seems to work well with some verification measurements taken on Red Cedar River and its impact on the watershed modeling is limited.

Fig 2.10 shows the river bed elevations estimated based on the DEM and NED using the above approach for the Red Cedar River, as well as the groundwater table provided by well head observations from the wellogic database [*GWIM*, 2006; *Simard*, 2007], later described in the groundwater input data section. As we can see, the NED-based estimates closely correspond to the groundwater heads while the DEM based values generally tend to be much higher, especially at reaches with large elevation drop. Close examination of the discrepancies reveals that the places where the two estimates differ the most are often places where the river feature is missing from the DEM map due to its resolution. Therefore, we conclude that NED-based values are much more accurate.

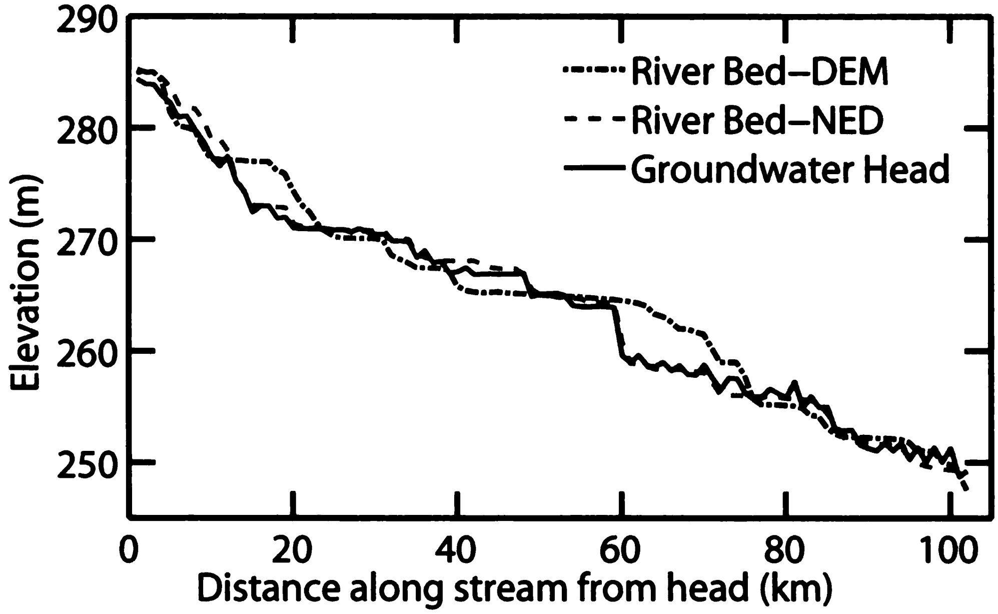


Figure 2.10. the river bed elevations estimated based on the DEM and NED as compared to the groundwater head

2.5.3. Soils

With the initial guess of the groundwater table, hydrostatic equilibrium is assumed as the initial guess for soil moisture. At hydrostatic state, we must have

$$h(x,y,z) = z - H(x,y) \tag{2.123}$$

where h(x,y,z) is the pressure head at any coordinate (x,y,z) in the subsurface domain, H(x,y) is the water table elevation of a grid cell at location (x,y). With this initial state and the van Genuchten formulation (2.57), the initial soil moisture condition anywhere in the watershed can be evaluated.

2.5.4. Climatic data

The Theissen polygon method is used to assign climatic data to cells. At the time of discretization, each climatic station records the indices of cells that they control. When a data record is missing, the first approach is to borrow data from adjacent stations. If no valid data exists on this day in all neighboring stations, the monthly mean value of the region is used.

As the model is based on an hourly time step while climatic data are often available in daily format, there must be ways to disaggregate the data into hourly time steps. Where daily precipitation from the NCDC is used, historical rainfall pattern for the Midwest is used to distribute the precipitation into hours according to [*NRCS*, 1986]. Michigan belongs to the type II (intensive rain) rainfall zone, and the rainfall hyetograph is obtained in the form of cumulative rainfall distribution Pfrac as shown in Figure 2.11a:

$$Pfrac = \int_0^t P\left(t\right)dt \qquad (2.124)$$

where P(t) is the rainfall rate (m/day). The rainfall rate in hour is then evaluated as:

$$P\left(hr\right) = P_{day} \times f = P_{day} \times \frac{Pfrac(hr+1) - Pfrac(hr)}{1/24} \qquad (2.125)$$

where P_{day} is the precipitation in a day (m/day), P(hr) is the rainfall rate in the given hour (m/day), $\sum_1^{24} f\left(k\right) = 1$ and is discretely valued as shown in Figure 2.11b.

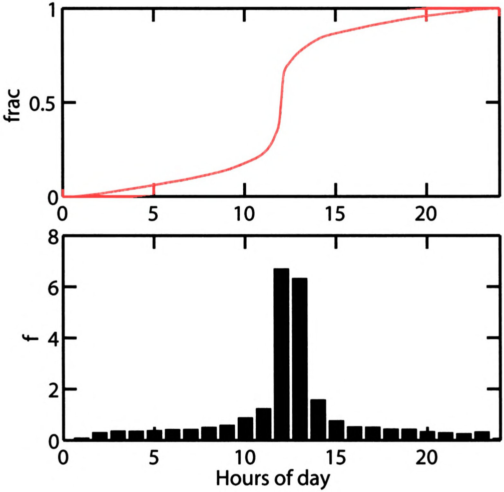


Figure 2.11. Rainfall pattern of the Midwest from

As discussed previously, given maximum temperature T_{mx} (°C) and minimum temperature T_{mn} (°C) in a day, the sub-daily temperature is parsed using [*Campbell*, 1985], assuming highest temperature at 3PM local time:

$$T = T_{av} + \frac{T_{mx} - T_{mn}}{2} \cos\left(0.2618\left(t - 15\right)\right) \tag{2.126}$$

in which hr is the hour of the day, T_{av} is average temperature of the day (°C).

2.6. Model Flow Diagram

Figure 2.12 shows the flow diagram for the proposed model. The model marches using the calendar days. The hydrologic processes, including Evaporation, transpiration, snowpack and unconfined aquifer are updated on an hourly basis. The vadose zone model solves infiltration, runoff and depression storage together with soil moisture. It uses an hour as a base time step but adjusts it depending on flow condition and convergence rate, similar to [*van Dam and Feddes*, 2000]. The overland flow and river network also have the ability to adaptively select time steps to ensure stability and computational efficiency.

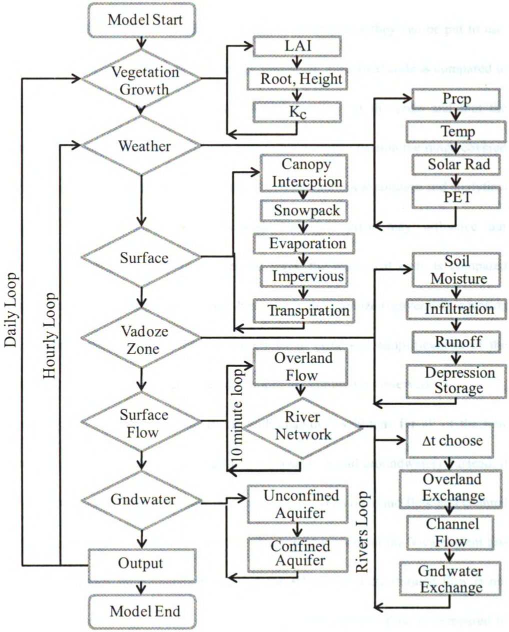


Figure 2.12. Program Flow Chart for the proposed model

2.7. Test cases

Numerical codes must be carefully tested and verified before they can be put to use. There are two levels of code testing. In the first level, numerical code is compared to available analytical solution to ensure there is no bug in the code and that the numerical scheme solves the PDE with acceptable accuracy (within the range covered by the test problem). In this level the match between numerical solution and analytical solution should be very good otherwise numerical error may influence our interpretation of the results. In the second level, the numerical code is compared against experimental data to examine how much the idealized governing equation approximates the real-world. Due to experimental error and simplifications of the processes, this is the step where more deviations are commonly observed.

In this section, we present comparisons with analytical solutions for all of the four flow domains (overland flow, channel flow vadose zone and groundwater). Instead, it is compared to field measurements. The coupling between overland flow and channel flow is tested together with overland and channel flow codes in the V-catchment test problem. Whereas since analytical solution for 3D saturated/unsaturated flow is not available, the coupling between vadose zone and groundwater flow is compared to experimental data reported in [*Vauclin et al.*, 1979].

Flow over an inclined plane

The first test case for the overland flow model is flow over an inclined plane. The plane is of length 200 meters and has a slope of 0.001 in the longitudinal direction but is level in the lateral direction. The Manning's coefficient n for the slope is 0.03. A Δx of 10m and a Δt of 3 seconds are used. A uniform rainfall rate of 1/60 mm/s for the duration of 1 hour is applied starting from the beginning of the solution. The Orlanski free outflow boundary condition (reference) is used. An analytical solution is available for the kinematic wave (KW) equation but not for the Diffusive wave (DW) and Dynamic wave (DyW) equations. We solve the DW with the RKFV scheme, solve the DyW with the SISL scheme and compared them with the KW analytical solution in Figure 2.13. The time of concentration (tc) for the plane is 1896s. This is the time when outflow equals rainfall rate and the hydrograph reaches a plateau. For this problem, there is a small difference between the solution of KW and DW on the rising limb around the time of concentration. This is also observed in the comparison reported in [*Gottardi and Venutelli*, 2008]. The DW has a more smoothed corner as opposed to the sharp edge predicted by the KW. This is not fully attributed to numerical dispersion but rather due to the fact that kinematic wave neglects the backward pressure from downstream. Thus it over-predicts the outflow rate before time of concentration as compared to the full DyW. The solutions from DW and DyW, on the other hand, are almost indistinguishable.

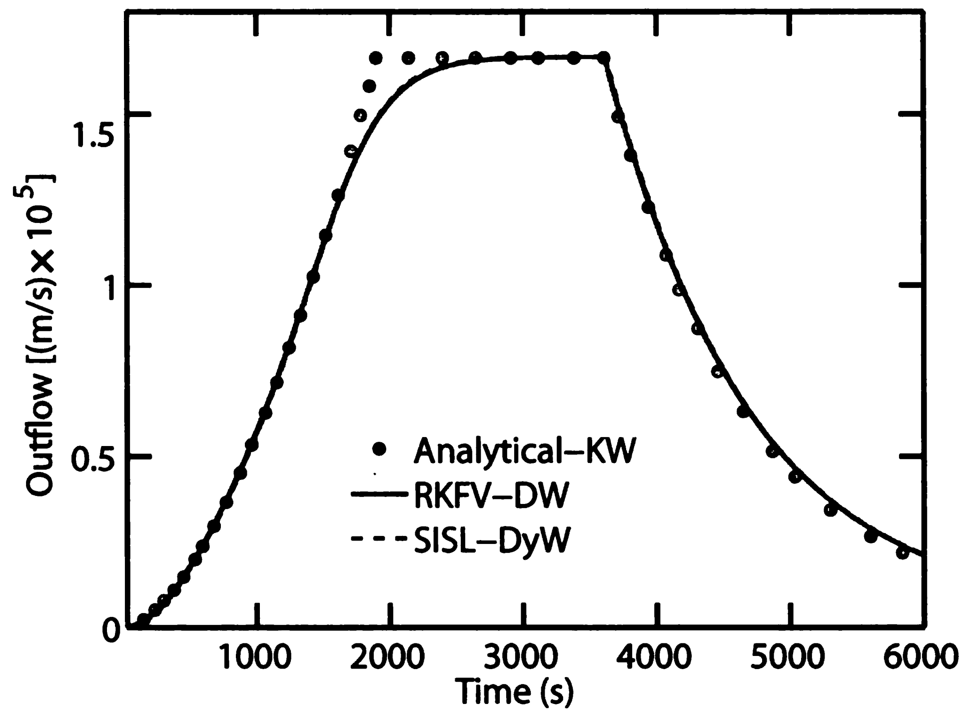


Figure 2.13. Outflow hydrograph of the inclined plane test problem compared to the analytical solution

V-catchment

The V-catchment problem is a standard test case for overland flow models [*DiGiammarco et al.*, 1996; *Stephenson and Meadows*, 1986]. The domain consists of two inclined planes draining into a sloped channel. The sketch of the problem is shown in Figure 2.14. Both of the planes are 800m in the lateral direction and 1000m in the longitudinal direction and the slope is 0.05. The channel has a slope of 0.02. The Manning's coefficient n used for comparison with literature is 0.015 for the planes and 0.15 for the channel. Due to symmetry, only half of the domain needs to be modeled. The plane is solved using the 2D RKFV scheme and the channel is solved on a separate channel grid, using the channel RKFV scheme. The solutions obtained from using the SISL scheme are almost identical and thus will not be reported here. In order to test not only the overland and river flow models but also the land/river exchange scheme of the proposed scheme, we use exactly the same procedure described in section 2.4.3. The land grid uses a Δx of 50m and a Δt of 5 seconds. Because the channel has a milder slope and large n, a $\Delta t=10s$ is used. The channel outflow rate is shown in Figure 2.15a, as compared to the analytical solution and Finite Element (FE) solution reported in [*DiGiammarco et al.*, 1996]. The analytical solution is not available for the declining limb of the hydrograph. We observe that the RKFV scheme matches the analytical solution very well during the initial rising period, slightly better than the FE solution. The plane-side outflow rate, the rate at which land contributes to the river, is shown in Figure 2.15b. There is a very small oscillation around $t = 30$ min due to the mass exchange scheme. This is also present in FE solution, only stronger. In watershed-scale modeling, such a small feature hardly has any impact just like its influence is not seen in Figure 2.15a. Also note the declining limb is captured noticeably better by the RKFV. Considering the simplicity and computational efficiency of the RKFV, these results make the scheme very favorable.

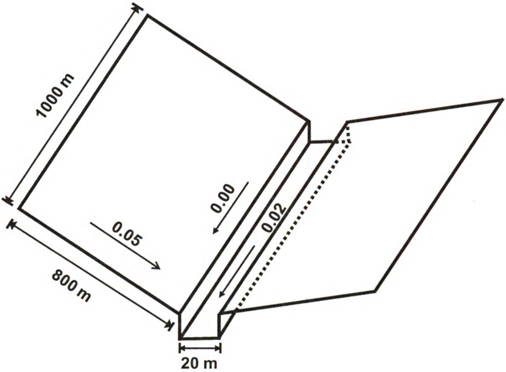


Figure 2.14. Sketch of the V-Catchment test problem

Red Cedar River measured hydrodynamic data

The river flow code has also been tested by comparing with analytical solutions of the inclined plane and V-catchment test problems. The solution is identical to the overland flow solver reported above and thus are not shown here. In order to test the model structure and the applicability of the diffusive equation to local rivers, we test out numerical model against field collected data.

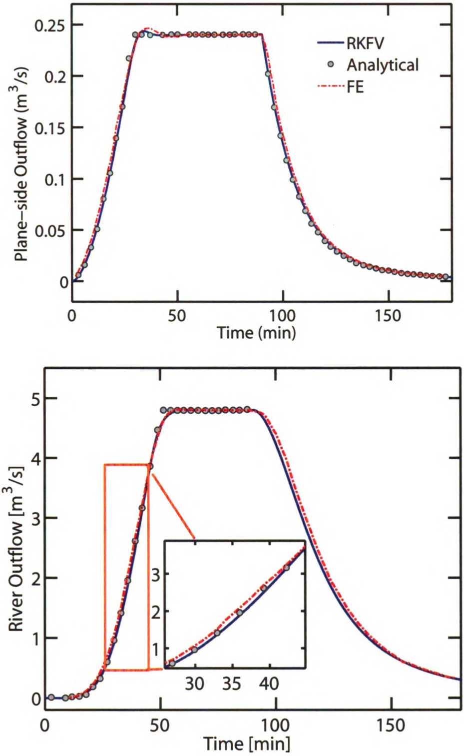


Figure 2.15. Solution to the V-Catchment test problem: Comparison of River outflow hydrograph with analytical solution and the finite element (FE) solution reported in [DiGiammarco]. (a) River outflow (b) Plane side outflow

[*Irfan*, 2002] measured hydrodynamic data (Discharge, Stage and Velocity) over a period of 8 days on the Red Cedar River (RCR). The rising limb and part of the falling limb of a precipitation event is recorded during this period. The bathymetry of the river reach, as well as the river width, is also provided. The data are fed to the river flow module of the proposed model. The comparisons between simulated and observed data at the Library Bridge are shown in Figure 2.16. The model accurately simulates the discharge and stage data. There is some discrepancy between the numerical solution and the measured velocity data. The difference may be due to the presence of a weir at the Library Bridge. In addition, the model assumed rectangular channel geometry while in reality the channel cross-section is slightly parabolic and there are flood plains at various locations of the study reach. The flood plains are expected to attenuate the flood peak and thus reduce the peak flow velocity. However, incorporating the presence of local floodplains considerably complicate the model and increases the data requirement and computational burden. Thus this is not considered here

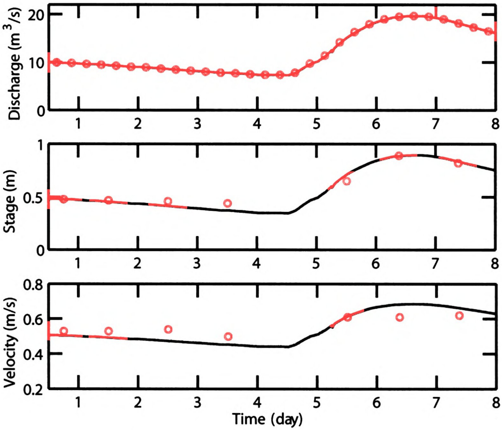


Figure 2.16. Simulated channel flow vs measured data by [*Irfan*, 2002]. Red circle: measured data; blue solid line: Simulated

Infiltration Test Problem

This widely-used test problem demonstrates the ability of the 1D Richards' equation to predict infiltration into very dry soils [*Celia et al.*, 1990; *Haverkamp et al.*, 1977]. It is one of the few analytical solutions available to the Richards equation. The soil column has a total length of L=0.3m, with the initial condition:

$$h(z,0) = -10m \tag{2.127}$$

And Dirichlet boundary conditions on both ends:

$$\begin{aligned} h(0,t) &= -0.75m \\ h(L,t) &= -10m \end{aligned} \tag{2.128}$$

The soil parameters for this test case are given in Table 2.5

Table 2.5. Soil Parameters for the test case: Infiltration into very dry soils

Parameter	θ_r *(-)*	θ_S *(-)*	α *(m^{-1})*	*n (-)*	K_S *(m/day)*	λ*(-)*
Value	0.102	0.368	3.35	2	7.9661	0.5

The solutions (pressure head) obtained with three different spatial sizes (Δx= 0.075, 0.25 and 0.6 cm) and the analytical solution at the final simulation time t = 6hr are presented in Figure 2.17. As noted by many other studies, the vertical resolution does have an impact on the solution quality [*Downer and Ogden*, 2004a]. The solver performs very well at Δx= 0.075cm and 0.25 spatial resolution. And the quality of the solution at Δx= 0.25 is very similar to that reported in [*Lehmann and Ackerer*, 1998]. But we do observe some deviation when Δx gets larger. The upper boundary condition serves as an infinite supply of water. We observe a 'piston' type of wetting front that is found in homogeneous porous media propagating without changing much of its shape. A coarse spatial resolution is hard to maintain the sharp wetting front predicted by the homogeneous Richards equation. One of the reasons is that the solver, as other predominantly used schemes, is first order accurate in space and time. More importantly, at the wetting front the soil hydraulic conductivity changes dramatically due to the strong nonlinearity of the soil water retention and hydraulic models.

Solving the Richards equation accurately using large spatial size remains a big challenge. However, since in the environment there are much larger uncertainties, e.g., the existence of macropores and the parameterization of soil water properties, we believe the numerical error with the solver will be dominated by other sources of errors and uncertainties.

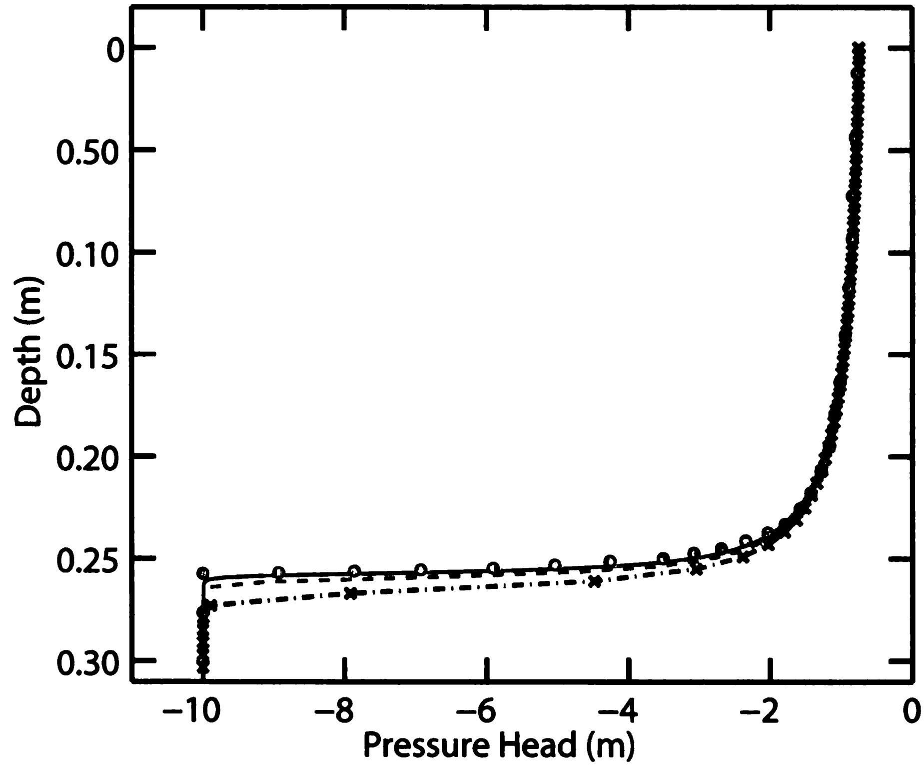


Figure 2.17. Infiltration into very dry soil test problem at t = 6hr. circle: Analytical solution; solid line: Δx= 0.075cm; dashed line: Δx= 0.25cm; dashed line with cross: Δx= 0.6cm.

Puming near impervious wall

The classic Theis solution [*Freeze and Cherry*, 1979] is used to check the accuracy of the groundwater solver. This test problem shows that when an impervious wall is present, the drawdown in the confined aquifer will be greater because water supply is limited (Figure 2.18a). The analytical solution to this problem can be obtained by modifying the classic Theis solution, considering the symmetry of the problem and setting a 'shadow' well in the mirror position of the pumping well. The solution form is given by [*Freeze and Cherry*, 1979]:

$$\begin{aligned} h_0 - h(x,y,t) &= \frac{Q}{4\pi T}\left[W\left(u_r\right) + W\left(u_i\right)\right] \\ u_r &= \frac{\left[\left(x-a\right)^2 + y^2\right]}{4Tt} \\ u_i &= \frac{\left[\left(x+a\right)^2 + y^2\right]}{4Tt} \\ W\left(u\right) &= \int_u^{\infty} \frac{e^{-u}du}{u} \end{aligned} \tag{2.129}$$

In which, h(x,y,t) is the groundwater hydraulic head (m) at location (x,y) and at time t [T], h_0 is the initial head [L], S is the aquifer storage coefficient [-], T is the transmissivity of the confined aquifer $[L^2T^{-1}]$, Q is the pumping rate $[L^3T^{-1}]$ and a is the distance of the pumping well from the impervious wall [L]. W(u) is the well function which can be numerically integrated.

The problem setup is illustrated in Figure 2.18b. A well situated 47.5m from the impervious wall and an observation well is located 50m away from it, further away from the wall. Other related parameters are:

Pumping rate Q=1000 m^3/day

Initial hydraulic head h_0=20m

Confined aquifer storativity: S = 0.0002

Aquifer Thickness b=20m

Hydraulic Conductivity K=50m/day

$\Delta x = \Delta y = 5$m

The time-series comparison with analytical solution at the observation well location is presented in Figure 2.19. As we can see, the solutions match perfectly.

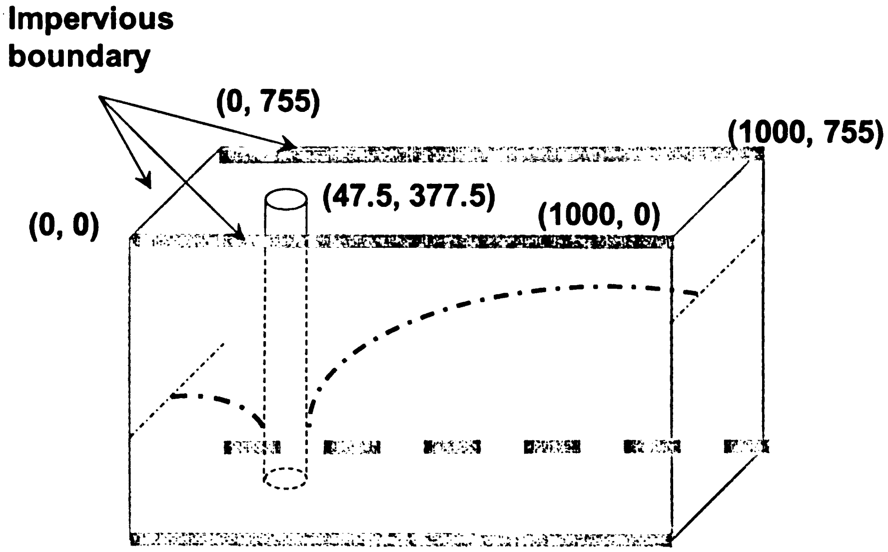


Figure 2.18. Pumping near impervious wall test problem

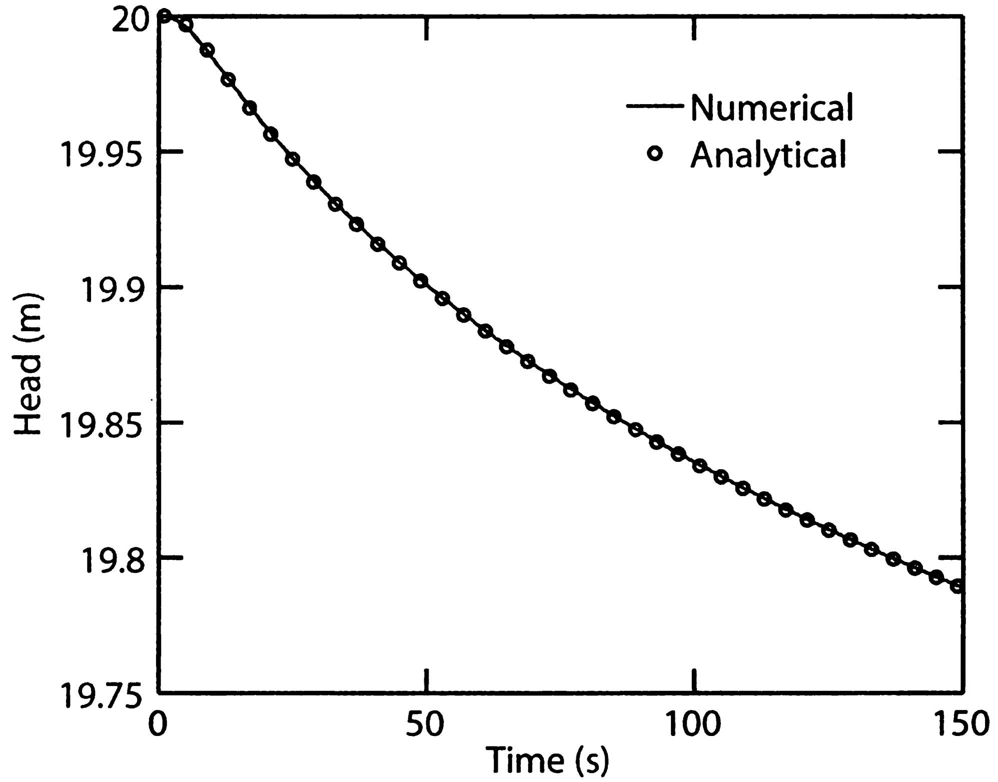


Figure 2.19. Solution to the pumping near impervious wall test problem: transient comparison;

Coupled Saturated/Unsaturated model

Previously, we described a novel approach to couple the unsaturated/saturated model, but its effectiveness to replace the 3-dimensional equation must be validated. As analytical solution to such a coupled system is not available, we can only compare our results to the results obtained from experimental study, or solutions from a 3-dimensional Richards model. Here we first present a comparison with the experimental results from [*Vauclin et al.*, 1979]. This dataset has been employed by many researchers to verify their 3D Richard's equation code, e.g. [*Dogan and Motz*, 2005] [*Clement et al.*, 1994]. The experiment consists of a soil slab 6.00 by 2.00

meters. Initially the soil slab has established hydrostatic equilibrium with the water table at 0.65 meters from the bottom. At the center 1m on the surface of the soil slab, a constant flux of 3.5 m/day flux is applied. On the left and right boundary, water level is maintained at 0.65m. As in other studies, we model only the right half of the domain due to the symmetrical nature of the experiment setup. Thus no flow boundary is applied to the left boundary. The soil retention and conductivity data published in the original data in [*Vauclin et al.*, 1979] is fitted to the Van Genuchten formulation by [*Clement et al.*, 1994; *Dogan and Motz*, 2005] and provided here in Table 2.6. For easier discussion we name the soil columns that are directly receiving inflow as recharge columns and the other columns as passage columns.

Table 2.6. Soil Parameters for the test case: Vauclin experiment

Parameter	θ_r *(-)*	θ_S *(-)*	α *(m^{-1})*	*n (-)*	K_S *(m/day)*	λ*(-)*
Value	0.01	0.30	3.30	4.1	8.4	0.5

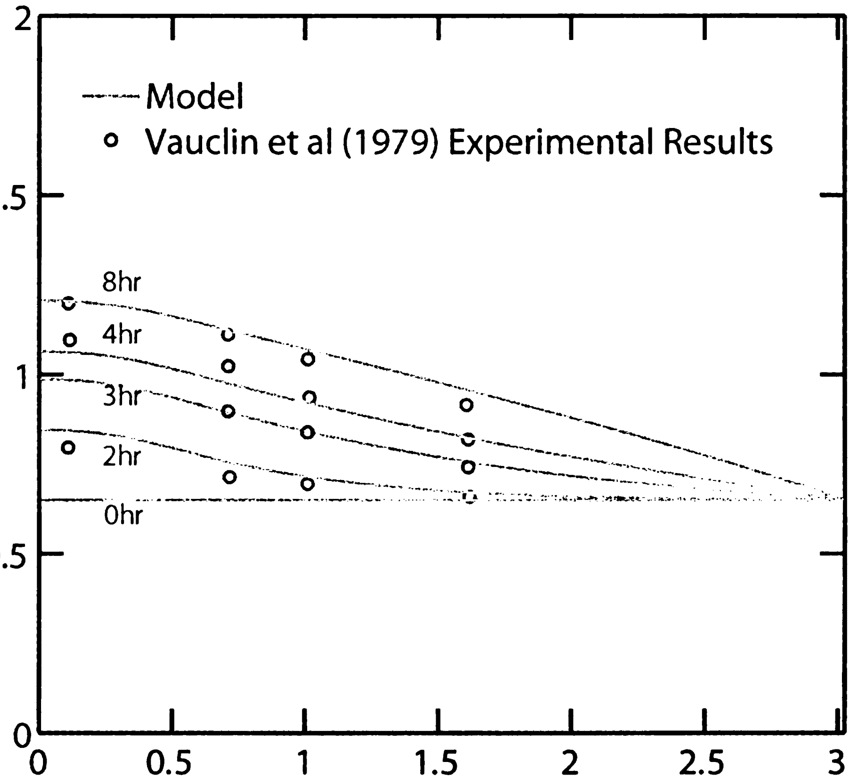


Figure 2.20. Comparison of the proposed model with experimental results from [*Vauclin et al.*, 1979]

The water table simulated using the proposed coupling method is shown in Figure 2.20. It is observed that the solution at hours 3, 4 and 8 fit well to the experimental data, whereas at hour 2 it has over-estimated the head at the left boundary. This over-estimation can be explained by the deviation of the coupled system to the actual 3D system. At the beginning stages of recharge, the passage columns are dry as compared to the recharge columns. Lateral moisture diffusion causes the inflow to be re-distributed to the passage columns before it reaches the unconfined aquifer and the unsaturated part stores a portion of the inflow. However, in the coupled system, water can move laterally only after it enters the unconfined aquifer. Thus at early stages the recharge has been over-estimated. However as the water table rises, the drainage term DR increases, the unconfined aquifer is able to transport more and more water, and the system gets closer and closers to steady state. The solution matches the experimental results better at later stages. This diffusion effect is expected to be less noticeable when the method is applied to a watershed-scale model, where the ratio of lateral flow flux to recharge is much smaller as compared to this experimental scenario.

Figure 2.21 shows a comparison of the soil moisture profile at hour 8 between the proposed coupling method and the traditional method, in which the hydraulic head of the unconfined aquifer is provided as Dirichlet boundary condition to the soil column. We notice that the soil moisture profiles of the recharge columns (x=0.025) are much higher with the conventional method than the present approach. Whereas in one of the passage columns (x=0.925) the soil moisture is lower than the steady state value. Both deviations are explained by the formulation of Eq. (2.86). With the present approach, the water table location (point at which soil moisture becomes saturated) agrees well with the head in the unconfined aquifer (in Fig 2.20).

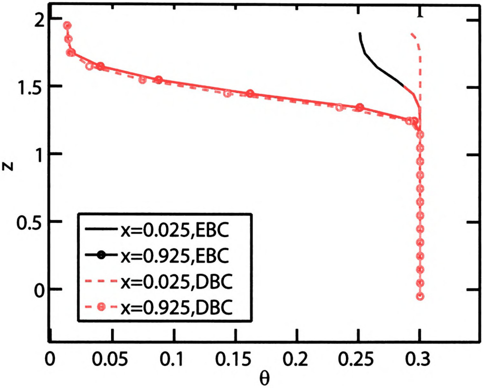


Figure 2.21. Comparison of soil moisture profile obtained using the approach in the proposed and traditional method, θ_S=0.3, EBC=Equation Boundary Condition (Present approach); DBC=Dirichlet Boundary Condition (Conventional approach)

Snowpack model.

The UEB snowpack model [*Luce et al.*, 1999; *Luce and Tarboton*, 2004] is adapted as the snow module of the proposed model. This model has been applied in the mountainous region of North East U.S. Here we apply the snow model using the driving data from Michigan.

Mass Balance Properties of the model

All schemes used in this model are mass-conservative schemes. However since the convergence of some iterative matrix solvers (e.g. the Conjugate Gradient method) is decided when error is less than a certain threshold, it is possible to induce a small amount of mass balance error. The mass balance error of the entire model and each component is closely monitored. After 10 years of simulation, a total mass balance error of 10^{-8} m is generated, which is insignificant compared to around 8m of rainfall during the period of time.

2.8. Summary and Conclusions

We follow the Freeze and Harlan (1969) blueprint for physically-based hydrologic modeling in that we believe each component of the model simulates a physical domain and should be verifiable by itself. All flow components, including overland flow, channel flow, vadose zone and groundwater flow have been independently verified by analytical solutions. Field measurements have been used to verify the channel flow code. The novel coupling mechanism between vadose zone and saturated groundwater flow has been applied to the Vauclin (1979) experimental data and the results were found to be promising. This novel approach solves a long-standing bottleneck in PDE-based subsurface flow modeling by removing the computational limitations while maintaining physically consistent solutions. Surface flow is solved by an efficient Runge Kutta Finite Volume scheme. After testing individual components of the model, we are ready to apply the whole model to a real-world watershed to evaluate the performance of PAWS as an integrated modeling system for the hydrologic cycle. This is accomplished in the next chapter.

Chapter 3. Development of the Hydrologic Model: Model Applications and Comparisons

In this chapter, we apply the PAWS model to a medium-sized watershed in Michigan and evaluate its performance as an integrated watershed model. The model outputs including streamflow and groundwater predictions are compared with observations, and average annual fluxes are compared with that estimated by an accepted annual budget simulator-the SWAT model. Through these comparisons we wish to establish the credibility of the model. Then we employ the model to provide insights to the dynamics of the study watershed and illustrate the interactions between the hydrologic components in space and time.

3.1. The Red Cedar River watershed model

3.1.1. Study site and input data

The Red Cedar River watershed is located in the Grand River watershed in Michigan (Figure 3.1). The National Elevation Dataset (NED) for the watershed is shown in Figure 3.2. The total area of the watershed amounts to 1169km2. The watershed has a relatively low relief with the maximum elevation recorded as 324m and a minimum of 249m.

The watershed is characterized by the humid continental climate with ample precipitation and distinct temperatures in different seasons. Figure 3.4 shows the annual precipitation recorded for the watershed. Daily or sub-daily weather data are obtained from various sources including National Data Climatic Center [*NCDC*, 2010] and Michigan Automated Weather Network [*MAWN*, 2010]. The locations of the stations are shown in Figure 3.3. The availability of data from these stations is given in Table 3.1. Although the MAWN stations contain much more data fields, most MAWN stations are operational only after 2006, with the exception of msuhort, which started from 1996. Solar radiation data are available only from the MAWN site. Thus the values for the NCDC sites are copied from nearest MAWN site when they are available.

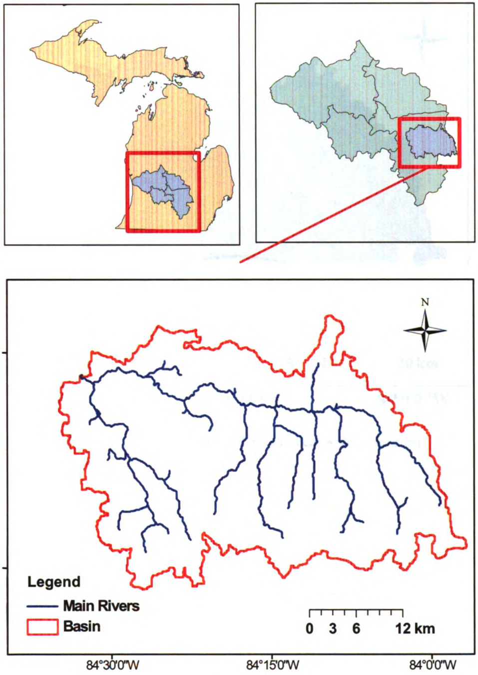


Figure 3.1. Location of the Red Cedar River Watershed

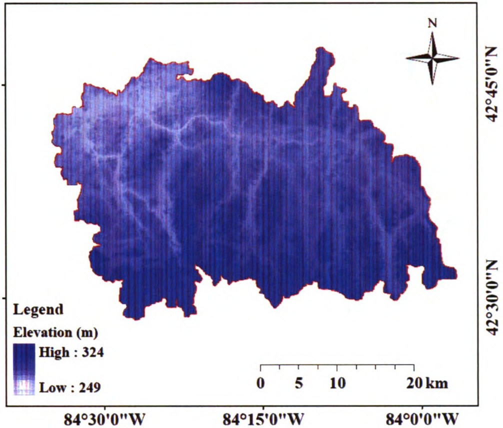


Figure 3.2. Elevation map of the Red Cedar River watershed

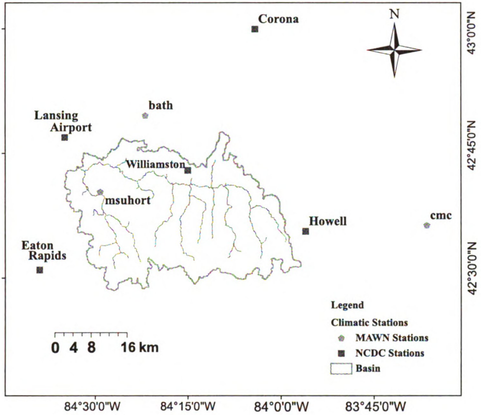


Figure 3.3. Locations of weather stations used for climatic input data

The MAWN stations also record relative humidity and wind velocity. Such data are present only with station 204641 in the NCDC list, which is an airport station. A suite of tools have been developed to process weather data from both types of sources and parse them into the format required by the model. When a particular precipitation/temperature record is missing, it is borrowed from the nearest available station. For solar radiation, missing data are filled using the algorithm detailed in Section 2.3.1.

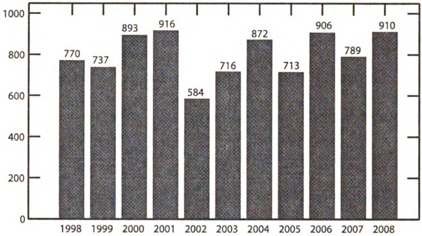


Figure 3.4. Basin-average Annual precipitation for the years from 1988 to 2008 (mm)

Table 3.1. Climatic Data sources and data availability. The dates in YYYYMMDD format are the dates when a station starts to have records; Frac is the fraction of the RCR watershed that is controlled by this weather station; PRCP PCT is the percentage of valid precipitation records from 1998 to 2007

Station ID	99002	99003	202437	201818	209006	204641	203947
NAME	MSUHORT	BATH	Eaton Rapids	Corunna	Williamston	Lansing Airport	Howell
Type	MAWN	MAWN	NCDC	NCDC	NCDC	NCDC	NCDC
Frac	32.9%	2.0%	2.4%	0.0%	41.2%	0.5%	21.1%
PRCP	19960801	20010801	19840101	20010401	19840701	19840101	19911001
TMAX	19960801	20010801	NA	20011201	NA	19840101	20040101
TMIN	19960801	20010801	NA	20011201	NA	19840101	20040101
WIND	19960801	20010801	NA	NA	NA	19840101	NA
RHMD	19960801	20010801	NA	NA	NA	19840210	NA
SRAD	19960801	20010801	NA	NA	NA	NA	NA
PRCP PCT	100.0%	64.2%	88.4%	67.3%	83.6%	99.9%	99.9%

The National Land Cover Data (NLCD) are shown in Figure 3.5 and the percentages of different land use classes are given in Table 3.2. The land use of the Red Cedar

River wat
to 56% o
scattered
percenta
are ofte
deciduou
Natural
Informati
Landsat
3.2b). Th
but differ
knowled
that the
or the fl
the MD
model c

River watershed is predominantly agricultural. Row crops plus forage crops stand up to 56% of the watershed. Urban areas are in the Northwest while forested areas are scattered throughout the watershed. According to NLCD, there is a significant percentage of wetlands present in the watershed (around 13%). However, these areas are often classified as either low land shrub, herbaceous open land or lowland deciduous forest in the land cover dataset provided by the Michigan Department of Natural Resources (MDNR) retrieved from the Michigan Center for Geographic Information [*MDNR*, 2010b]. The MDNR data is derived from classification of Landsat Thematic Mapper™ imagery and has a more detailed classification (Table 3.2b). The MDNR and NLCD data agree well on the percentages of agricultural land but differ quite significantly on the wetland, forest and herbaceous percentages. Little knowledge exists about which one of these two databases is more accurate. We found that the 'wetland' areas in the NLCD are mostly in close neighborhood to the streams or the flood plain. Upon field inspections of some conflicting areas we determine that the MDNR database is to be used for the modeling. The data are reclassified into model classes following the procedure outlined in section 2.2.1.

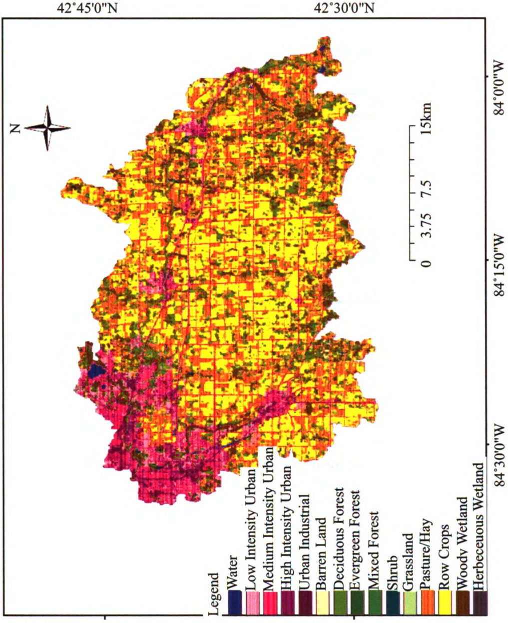


Figure 3.5. NLCD Land use/Land cover map for the RCR watershed

Table 3.2. Land use percentages from the NLCD database and the MDNR databse

Land use Classes	NCLD	MDNR
Urban	18.17%	10.29%
Deciduous Forest	9.01%	15.02%
Evergreen Forest	0.63%	4.64%
Row Crops	35.44%	33.63%
Forage Crops	22.10%	22.14%
Herbaceous and Bushes	0.80%	11.97%
Bare	0.29%	0.19%
Wetland	12.99%	1.56%
Water	0.56%	0.55%

The watershed, as others in Michigan, is very rich in the number of rivers and has a complex surface water system. The main rivers from the National Hydrography Dataset (NHD) in the watershed are shown in Figure 3.6. Majors rivers of the watershed are extracted from the NHD. These are the river objects that will be solved in the proposed model. Other smaller creeks, ditches, furrows, rills, etc, are modeled as the overland flow domain as discussed in section 2.3.3. The Red Cedar River, which is a tributary of the Grand River, is the major stream in the domain, followed by the Sycamore creek. Due to the low relief of the terrain, the rivers generally have a very mild slope.

The two stream gaging stations operating in the RCR are USGS 04112500 at farmlane, East Lansing and 04111379 at Williamston. Since East Lansing is the downstream gage, we focus our calibration at this gage. However, the comparison at Williamston is also important as it shows if the model is working properly inside the watershed.

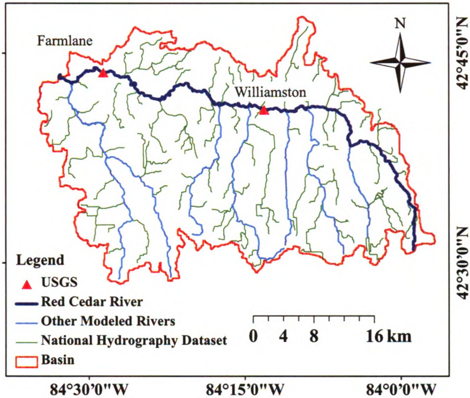


Figure 3.6. River system of the RCR watershed with USGS flow gages

Geology of Michigan including that of the present day Red Cedar River watershed is the result of extensive glaciation during four major glacial periods including the Wisconsian, Illinoian, Nebraskan, and Kansian periods. Glaciation during the Wisconsian period (the last of these periods, approximately 11,000 years ago) is mainly responsible for the development of Michigan's geology, soils, and topography. In the RCR watershed the geology includes glacial till (poorly sorted material including pebbles and boulders), glacial outwash (finer material deposited by glacial melt water), lacustrian material (fine materials deposited in glacial meltwater) and alluvial material (recently deposited material from rivers and streams) as shown in Figure 3.7 [*Corner et al.*, 1999; *MDEQ*, 2010; *MDNR*, 2010a]. These deposited materials, as well as organic material, are the parent materials of the soils in the watershed. Limestone, shale, sandstone, coal, and other sedimentary rocks arranged in almost horizontal layers compose the bedrock surface of the Red Cedar Watershed. These features contributed to the relatively high conductivities of soil materials.

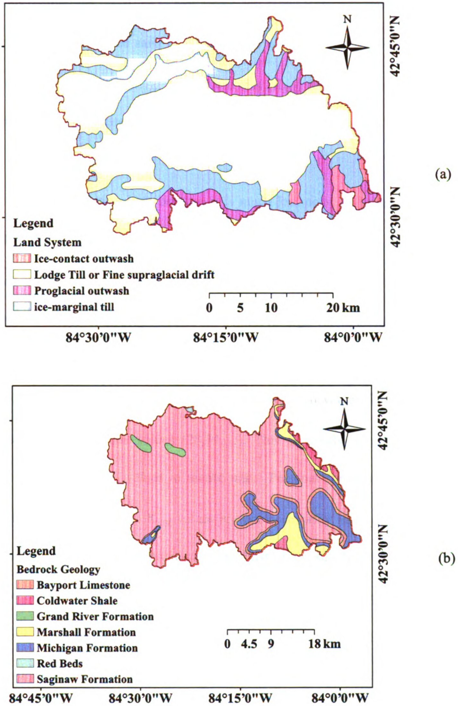


Figure 3.7. Geology maps of the RCR watershed. a. Land Systems; b. Bedrock Geology

As a result of the glacial geology, soils within the watershed are mostly sandy, loamy or muck soils (commonly classified as hydrologic soil group B) mostly share the same common parent materials. The sand percentage map is shown in Fig 3.8a. The soil available water capacity and saturated conductivities (K_{sat}) of the first soil layer as reported in the STATSGO database are shown in Fig 3.8b and Fig 3.8c.

We show the STATSGO data here because it is much less noisy and thus easier to visualize than the SSURGO database. However, for the model simulation, SSURGO data is used as input. K_{sat} is presented here only to give us an impression of the spatial variation of hydraulic properties of the soils. The K_{sat} values from SSURGO or STATSGO cannot be directly used as it is believed that several factors during the generation of these two databases have impacted the reported values. For example, K_{sat} only takes a few unique values (28,34,...). This is clearly a result of unit conversion from integer inch values to metric units. Personal communication with SSURGO staff reveals that these K_{sat} values are not necessarily measured but may also be inferred using a set of procedures established by the Natural Resource Conservation Services (NRCS). Since such critical information as the sources of data is obscure, we choose not to directly use the K_{sat} values in the SSURGO. Ksat is estimated using empirical relations based on the sand, silt and clay [*Rawls et al.*, 1991].

Figure 3.8. Soil property maps of the RCR watershed as reported in the STATSGO database. a. Sand percent; b. Soil Available Water Capacity; c. Saturated Hydraulic Conductivity (K_{sat})

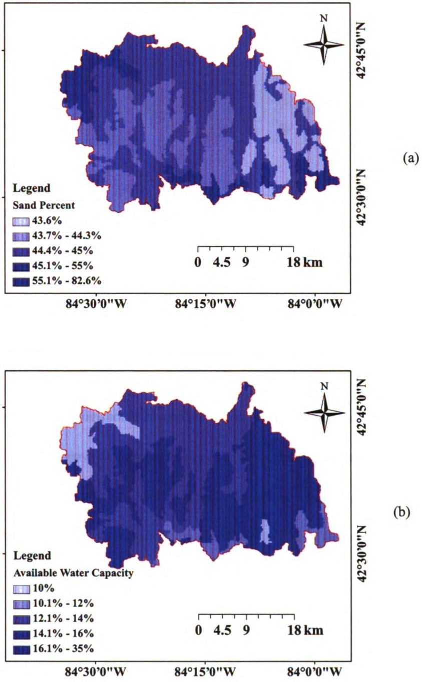
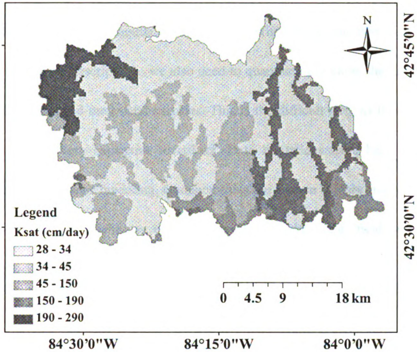


Figure 3.8 (cont'd)



(c)

Groundwater data

Generally, some information should be available about the geological formation of the study domain to employ the proposed model. Some ideas about the aquifer configurations are necessary for constructing the conceptual model. For the case of the RCR watershed, as discussed in section 3.1.1, we know a layer of more permeable glacial deposit with a thickness ranging approximately from 5 meters to 30 meters overlies the bedrock. Although the bedrock has very small permeability, the thickness of the layer is great and the overall transmissivity is not low. This information allows us to divide the groundwater model into two layers. The upper layer is the glacial deposit, taken as the unconfined aquifer; the lower layer is the bedrock, taken as the confined aquifer. A layer of aquitard is assumed to exist between the two layers.

Besides qualitative descriptions, we also need to quantitatively know the thickness of the aquifer layers and their conductivities. This is the difficult step, as these types of data are usually very expensive to collect. Data limitations have been the major hamper of groundwater modeling. Conductivities are known to vary many orders of magnitude over space whereas existing conventional data (e.g. field survey) are normally too scarce.

However, for the RCR watershed and Michigan in general, valuable groundwater -related data have been collected, logged and filed with the Michigan Department of Environmental Quality (MDEQ) when water wells in the state of Michigan were drilled. Reported information include depth to water table, pumping test results, lithological layer descriptions, etc. From these records the local conductivities can be estimated (e.g. [*Mace*, 1997]). The number of wells is so massive that it almost dense covers the entire state of Michigan [*Simard*, 2007]. The large number of wells can partially offset the data quality issue. If the correct statistical procedure is used to remove the noises and outliers, the results are more reliable.

Several studies have attempted to employ this data source, e.g. [*Hill-Rowley et al.*, 2003; *Reeves et al.*, 2003]. However, the difficulty is that, since these data are not recorded by professional surveyors but common household drillers, the quality of the data is mediocre and sometimes poor. Significant noises, if not handled correctly, exist in the data which can result in erroneous interpretations of the aquifers. More information is to be found from [*GWIM*, 2006; *Simard*, 2007].

Fortunately, extensive studies have been underway to denoise and interpolate the data at Michigan State University to produce a smooth, statistically coherent data source. Kriging is identified as the best geostatistical approach to interpolate the data into a spatial field. Part of that research was reported in [*Simard*, 2007]. As a result, groundwater head, conductivity and thickness of the glacial drift layer were processed using the software IGW [*Li and Liu*, 2004; *Li et al.*, 2006].

Nevertheless, at places where such kind of data is not available, the users should not be daunted by the data limitation. Despite all the difficulties, some information can always be obtained from either lithological data or wells pumping tests. Our advantage is that groundwater varies smoothly over space and we do not have a specially high requirement on the quality of the K data. Besides, some parameter adjustment can be applied.

3.1.2. Model Calibration

Flow records at USGS gage 04112500, East Lansing from 2002/09/01 to 2005/12/31 were used to calibrate the model while the years prior to this period were used for verification (or validation). The reason for using water years 2002 to 2005 was that in these three consecutive years the watershed has experienced high, normal and unusually low precipitation periods. Auto-calibration of the model was done using the

Differential Evolution (DE) algorithm [*Chakraborty*, 2008] with the computational resources from the High Performance Computing Center at the College of Engineering, Michigan State University [*HPCC*, 2009]. The original DE code has been modified and implemented into a parallel computing code. Future versions of PAWS will allow calibration on a single CPU.

3.2. SWAT Model for comparison

While developing the proposed model, we also created a Soil and Water Assessment Tool (SWAT) model [*Arnold and Allen*, 1996; *Arnold and Fohrer*, 2005] to compare with the proposed model. The SWAT model is a widely used spatially-distributed parameter model for the assessment and management of water resources, soil conservation, non-point source pollution and soil erosion, etc. It has been applied in many countries for various applications. SWAT was designed as a long-term yield model and is not designed to accurately simulate detailed, single-event flood routing [*Neitsch et al.*, 2005]. In the past 5 years there have been over 400 publications related to SWAT, showing its generally accepted validity, popularity and versatility. Many papers refer to SWAT as a physically-based model while some regard it as a semi-physically-based because there is still significant amount of empiricism. For example, for daily modeling the major infiltration scheme is the SCS curve number method [*Division*, 1986]. The curve number method is built on many years of accumulation of experience in runoff prediction [*NRCS*, 1986]. However, as an empirical approach, it cannot distinguish the effects of different process, e.g. infiltration excess or saturation excess. From the basis on which the SCS curve number method is developed it is also unclear how much water has infiltrated or retained on the ground. For another example, the lack of a physically-based groundwater flow model prevents SWAT from simulating regional groundwater flow and making reliable estimates of surface water-groundwater interactions. The overland flow and channel flow modules, which are based on the reservoir concept, are also regarded by some to be over-simplified and thus presents challenges to the accurate description of transport problems. The empiricism makes the SWAT model less suited to study the impact of human modifications (e.g. dam or other hydraulic structures). Each model has its designed purpose and its own areas of strength, and we seek not to arrive at the conclusions of which one is better. We are interested in comparing the proposed model with the SWAT model primarily because it allows us to see if the fully physically-based modeling approach does have its own advantages.

3.2.1. Brief summary of SWAT mathematical bases

We briefly summarize some of the major SWAT modeling approaches here to highlight the differences between SWAT and PAWS. The theories covered here can all be found in the SWAT theoretical documentation [*Neitsch et al.*, 2005]. The majority of the literature use SWAT as a daily model, thus we only describe the daily modeling options.

SWAT discretizes the study area into subbasins, each one with a main channel and a tributary channel. The subbasin and the channel is the smallest unit with spatial heterogeneity. Inside each subbasin, however, multiple Hydrologic Response Units (HRUs) may be defined. Each HRU is a unique combination of soil type, land use/land cover type and slope (slope added in the SWAT2005 version).

The runoff volume is calculated using the SCS Curve number method. The curve number for each HRU is looked up from tabulated empirical values corresponding to common land use/soil combinations and adjusted for antecedent moisture content. Once the curve number (CN) is determined, the runoff for the HRU is calculated as:

$$\begin{gathered} S = 25.4\left(\frac{1000}{CN} - 10\right) \\ Q_{surf} = \frac{\left(P - 0.2S\right)^2}{\left(P + 0.8S\right)} \end{gathered} \tag{3.1}$$

where S is a retention parameter (mm), P is precipitation (mm) for the day and Q_{surf} is the surface runoff. Due to the empiricism, the curve number approach cannot distinguish the different runoff generation mechanisms (saturation excess, infiltration excess) and does not intend to do so from the design [*Arnold and Fohrer*, 2005; *Borah and Bera*, 2004].

SWAT allows water in the soil layers in excess of field capacity to percolate to lower layers:

$$SW_{ly,excess} = \begin{cases} SW_{ly} - FC_{ly} & SW_{ly} > FC_{ly} \\ 0 & SW_{ly} \le FC_{ly} \end{cases} \tag{3.2}$$

$$w_{perc,ly} = SW_{ly,excess}\left[1 - \exp\left(\frac{-\Delta t}{TT}\right)\right]$$

$$TT_{perc} = \frac{SAT_{ly} - FC_{ly}}{K_{sat}}$$

where TT is the travel time for percolation (day), SW_{ly} is the water content in the soil layer (mm), $SW_{ly,excess}$ is the drainable volume of water on a given day (mm), K_{sat} is the saturated conductivity of the layer, FC_{ly} is the field capacity of the soil layer (mm), $w_{perc,ly}$ is the volume of percolation of the day (mm), SAT_{ly} is the amount of water in the layer when it is fully saturated (mm). As we can see, these equations assume a one-way percolation which does not take into consideration the pressure head (or hydraulic head) of the soil water in the lower layer. Another mechanism termed 'revap' is used to account for groundwater entering the unsaturated soil zone from the aquifer. However, that mechanism is also empirical and controlled by a revap coefficient. The revap coefficient is ranked one of the least sensitive parameters in the model [*van Griensven et al.*, 2006].

The recharge to aquifers on a given day is in turn calculated by lagging the recharge using the method proposed by [*Sangrey et al.*, 1984] and [*Venetis*, 1969]:

$$w_{rchrg,i} = \left[1 - \exp(-\frac{1}{\delta_{gw}})\right] w_{seep} + \exp(-\frac{1}{\delta_{gw}}) w_{rchrg,i-1} \qquad (3.3)$$

$$w_{seep} = w_{perc,ly=n} + w_{crk,btm}$$

in which $w_{rchrg,i}$ is the amount of total recharge to aquifers on a day (mm), w_{seep} is the total amount of groundwater recharge on a given day (mm), δ_{gw} is the delay time or drainage time of the overlying geologic formation (days), $w_{perc,ly=n}$ is the percolation from the lowest soil layer, n, on the day (mm), $w_{crk,btm}$ is the amount of water flow past the lower boundary of the soil profile due to bypass flow on day i.

And the groundwater contribution to streams is calculated as:

$$Q_{gw,i} = Q_{gw,i-1} \exp\left(-\alpha_{gw} \Delta t\right) + w_{rchrg,sh} \left[1 - \exp(-\alpha_{gw} \Delta t)\right],$$

$$\text{if } aq_{sh} > aq_{shthr,q} \tag{3.4}$$

$$Q_{gw,i} = 0, \text{ if } aq_{sh} <= aq_{shthr,q}$$

where $Q_{gw,I}$ is the groundwater flow into the main channl on day I (mm), α_{gw} is an empirical baseflow recession constant, $w_{rchrg,sh}$ is the recharge to the shallow aquifer, aq_{sh} is the amount of water stored in the shallow aquifer, and $aq_{shthr,q}$ is the threshold water level in the shallow aquifer for groundwater contribution to channel to occur (mm). α_{gw} and δ_{gw} are called, respectively, ALPHA_BF and GW_DELAY in the SWAT model codes. In the SWAT literature these two are often used as calibration parameters. Some papers use the hydrograph-baseflow separation technique [Arnald] to identify α_{gw}.

SWAT routes runoff in a subbasin to its main stream using a simple lagging relationship:

$$Q_{surf} = \left(Q_{surf} + Q_{stor,i-1}\right)\left[1 - \exp\left(\frac{-surlag}{t_{conc}}\right)\right] \tag{3.5}$$

where Q_{surf} is the amount of overland flow contribution to the main channel on a given day (mm), Q_{surf}' is the amount of surface runoff generated on the day (mm), $Q_{stor,i-1}$ is the surface runoff stored or lagged from the previous day (mm), surlag is the surface runoff lag coefficient and t_{conc} is the time of concentration for the subbasin (hours). This lagging relationship resembles the solution to a linear ordinary differential equation. Surlag is usually used as a calibration parameter. From the author's experience, the proper value to use with surlag depends heavily on the size of the delineated subbasins. The larger the subbasins are, the smaller surlag should be used. However, the use of this relationship masks the surface flow patterns in the subbasin. For example, the overland flow paths and water velocity are not known using this approach.

SWAT routes water in the river using either the linear reservoir method or the Muskingum Cunge method. For the RCR watershed SWAT model developed, we have used the Muskingum method for river routing. The development of the Muskingum routing method is rather lengthy and is avoided here. It can be found on the water routing chapter in the SWAT theoretical documentation.

3.2.2. SWAT Model Setup

The SWAT model for the RCR watershed was created with the assistance of the ArcSWAT GIS interface for SWAT2005 [*Olivera et al.*, 2006]. Using a 90 meter resolution DEM and the NHD stream network, 53 subbasins were delineated for the RCR watershed (Figure3.9). The NLCD land use/land cover map and STATSGO soils database (1:250,000) [USDA] was used for land use/soils overlay to obtain the distribution of Hydrologic Response Unit (HRU). Each HRU is a unique combination of land use, soil types and slope. As compared to SWAT2000, the 2005 version of

SWAT also calculates the slope for each HRU and overlays different slopes on top of land use/soil combinations, thus creating significantly more HRUs. According to some recent studies, using local slope values for each HRU is of great importance to correctly simulating hydrologic responses and nutrient output. We used three classes for slope: slope<=3%, 3%<slope<=6% and slope>6%. Since the topographic relief of the watershed is very gentle, only a small fraction of areas, mainly the hills, is larger than 6%. To maintain the number of HRUs at a manageable level, SWAT uses customizable thresholds to remove any land use or soil type found in a subbasin that is below their respective thresholds, and assigns their areas proportionally to the dominant classes. We employed a 20% threshold for land use, 10% for soil and 20% for slope. The final land use/soil distributions after these processing steps are given in Table3.3. As a result, a total of 1936 HRUs were obtained. The median area of the HRU is 0.069 km^2.

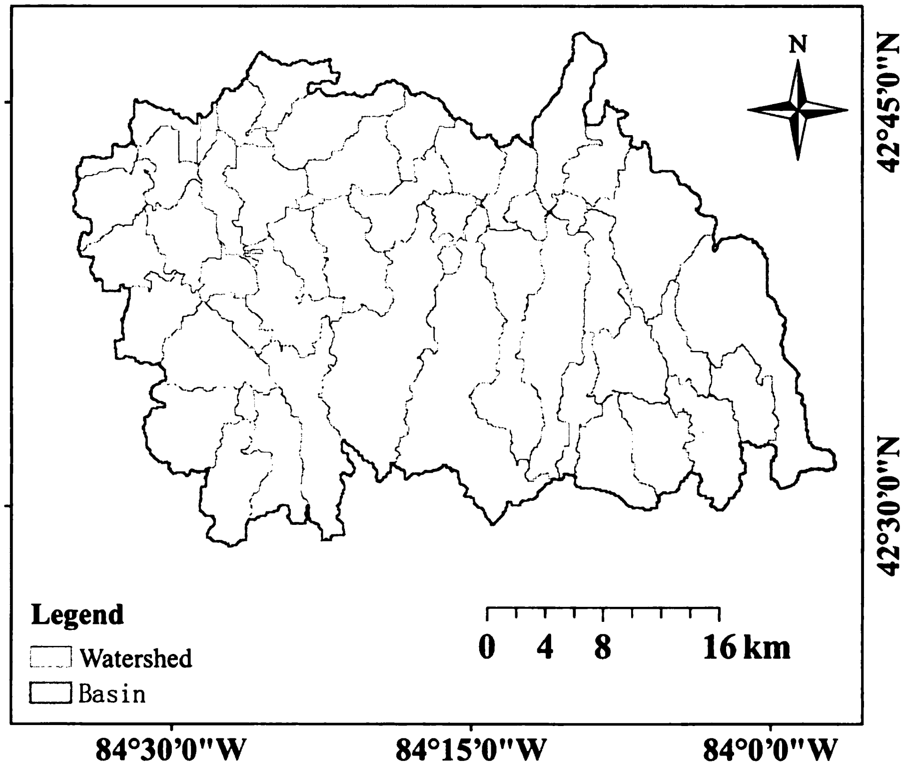


Figure 3.9. Watersheds delineated for the SWAT2005 model using the ArcSWAT GIS interface

Table 3.3. SWAT Land use class percentages after apply thresholds. URLD: Low Intensity Urban; URMD: Medium Intensity Urban; WETF: Forested Wetland; URHD: High Intensity Urban; HAY: Hay/Forage Crops; AGRR: General Agriculture; FRSD: Deciduous Forest

URLD*	URMD	WETF	URHD	HAY	AGRR	FRSD
5.07	5.86%	8.99%	1.55%	29.18%	47.45%	1.91%

*Notes: The SWAT land use codes:

The SWAT model is calibrated against the flow data from USGS gage 04112500 at farmlane, East Lansing using the Shuffled Complex Evolutionary algorithm [*Duan et al.*, 1992; *Duan et al.*, 1994].

3.3. Results and Discussions

3.3.1. Model Evaluation

We first evaluate the performance of the model by comparing it against observations and other models. Then we use the model to shed insights into the hydrologic system of the study watershed. Two forms of observations are included in the comparison. One is the streamflow measurements at the two USGS gaging stations inside the watershed, the other is the groundwater head data interpolated from the well records. Output from the SWAT model which is an established annual budget simulator is also used in the comparison. In addition, the places where saturation excess is produced is compared to the Topographic Index proposed by Beven [*Beven and Kirkby*, 1979].

3.3.2. Model performance evaluations

Model performance is evaluated using the comparison with USGS flow data, groundwater wells records and average annual fluxes reported by SWAT. The metrics we use to measure the performance include coefficient of determination (R^2), root mean squared error (RMSE), the Nash-Sutcliffe model efficiency coefficient (NASH) and the mean error (ME). The RMSE, NASH, RNASH and ME are defined as:

$$RMSE = \sqrt{\sum_{t=1}^{T}\left(Q_o^t - Q_m^t\right)^2}$$

$$NASH = 1 - \frac{\sum_{t=1}^{T}\left(Q_o^t - Q_m^t\right)^2}{\sum_{t=1}^{T}\left(Q_o^t - \overline{Q_o}\right)^2}, RNASH = 1 - \frac{\sum_{t=1}^{T}\left(\sqrt{Q}_o^t - \sqrt{Q}_m^t\right)^2}{\sum_{t=1}^{T}\left(\sqrt{Q}_o^t - \overline{\sqrt{Q}_o}\right)^2}, \quad (3.6)$$

$$ME = \sum_{t=1}^{T}\left(Q_o - Q_m\right)$$

where Q_o is the observed discharge, Q_m is the modeled discharge, T is the number of observations and $\overline{Q_o}$ is the mean of the observed discharges. The RNASH is simply the NASH formula applied to the square-root transformed data. This transform reduces the relative importance of large

The daily observed and simulated flows at USGS gage 04112500 (farmlane, East Lansing) for the 10 years from 1998 to 2007 are compared in Figure 3.10. The proposed model did a very good job at streamflow predictions, especially with the baseflow/ low flow periods. The performance metrics of the calibration, verification and entire periods are provided in Table 3.4. The Nash-Sutcliffe for the calibration period is 0.694 while for the verification period it is 0.585. A NASH greater than 0.60 for daily comparison is generally considered good performance for a watershed of this size. By examining many models and literature, it was found that hydrologic models applied to larger-sized watersheds (>5000km^2) tend to achieve higher NASH at more downstream gages because input errors (e.g. error with precipitation records) and model structural error may cancel out.

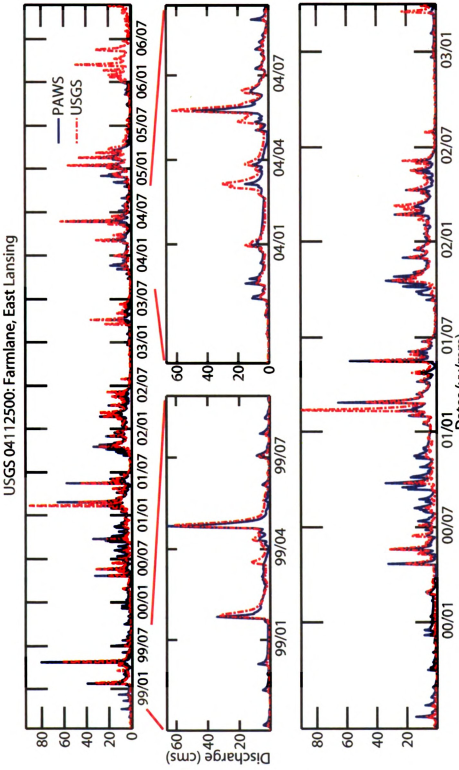


Figure 3.10. Comparison of the observed and simulated daily flow at USGS gage 04112500 at East Lansing. Upper: Entire period from 1998/09/01 to 2005/12/31, Middle and lower: Close-up look of the hydrograph

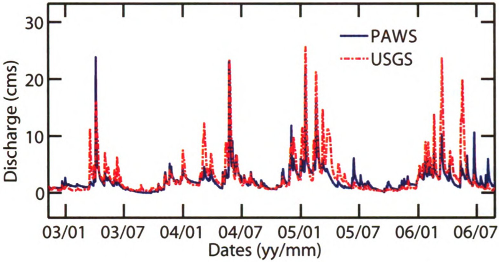


Figure 3.11. Comparison of the observed and simulated daily flow at USGS gage 04111379 at Williamston

Table 3.4. Performance metrics evaluating the model applied to the RCR watershed

		NASH	RMSE	RNASH	ME	R^2
East Lansing	SWAT Calibration	0.665	4.033	0.621	-0.698	0.675
	SWAT Verification	0.675	4.261	0.561	-0.065	0.680
	PAWS Calibration	0.694	3.855	0.693	-1.122	0.742
	PAWS Verification	0.585	4.816	0.661	-0.232	0.591

Close inspection of the hydrograph indicates that the mismatches are generally due to the underestimation of streamflow peaks during snowmelt periods. The reason for the underestimation is attributed to the UEB snowmelt module employed in the proposed model not melting snow rapidly enough to produce large runoff that is apparent from the hydrograph. Since the UEB model is intended as a model with no parameters for adjustment, adjusting this aspect of the simulation has been difficult. Further in-depth research is needed to understand the exact reason for this under-estimation and to improve the results.

We note the NASH coefficient of PAWS ranks higher than SWAT during calibration period but lower during verification period. This reduction of NASH in verification period is almost entirely due to the under prediction of a single snowmelt peak in 2001. This is shown in Figure 3.12, in which the SWAT simulated flow is also plotted. Although the baseflow leading into the peak as well as the recession periods are nicely captured by PAWS, the peak around 2001/02/12 is largely missed, whereas SWAT does a better job at this peak, thus ranking higher in the verification period. The low flow periods are better visualized in a log-transformed as in Figure 3.13. Due to its simplified groundwater flow structure, SWAT produces baseflow that very different in shape from the observed data, while PAWS generally does a good job.

Study has shown that outliers can significantly influence sample values of NASH [*McCuen et al.*, 2006]. Due to its mathematical formulation, NASH puts much more weight on a few large peak values than the low flow period. Low flow periods are when pollutant concentrations tend to be high and thus the risk to human health is greater. The baseflow period is also very important to capture as it best describes the subsurface dynamics of the watershed. From these viewpoints we argue that NASH is not the entire story and it may be more appropriate to look at the indicator RNASH, which put weight more evenly on high and low flow periods. As we can see from Table 3.4, PAWS consistently ranks higher than SWAT in terms of RNASH, both in calibration and verification period, indicating a more realistic subsurface representation.

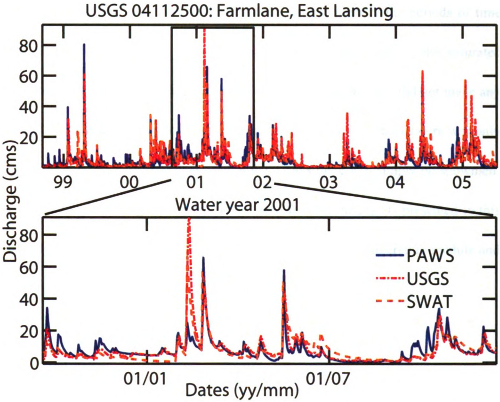


Figure 3.12. Comparison of the simulated streamflow from PAWS and SWAT and observed USGS streamflow data. Upper: from 1999 to 2005, Lower: Close-up look of water year 2001

Figure 3.11 shows the hydrograph comparison at the inner gage Williamston. The model performs equally well at this gage as East Lansing. These results give us more confidence that the model not only predicts flow satisfactorily at a downstream gage, but also captures the underlying hydrologic dynamics of the watershed.

The under-prediction of snowmelt peak may also be due to the freezing soil effect, which reduces the conductivity of the frozen soil layers. Since at current stage there is still not a soil temperature module, soil freezing cannot be physically-described. Since there are some soil temperature measurement data from MAWN, it has been attempted to use the measured soil temperature data to identify the periods of time when soil is frozen. When the soil temperature was less than 0, the saturated conductivity of the soil was reduced to 1/10. However, this attempt did not make any material difference to the hydrograph because the measured soil temperature was only less than 0 in a few very cold days, and in these days there was simply no snowmelt. As such the soil freezing mechanism is not activated for the results presented in this dissertation. Future research should consider including a soil temperature module and a better parameterization of the freezing process.

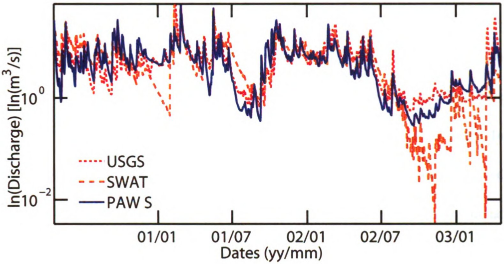


Figure 3.13. Comparison of the simulated streamflow from PAWS and SWAT and observed USGS streamflow data in log-scale

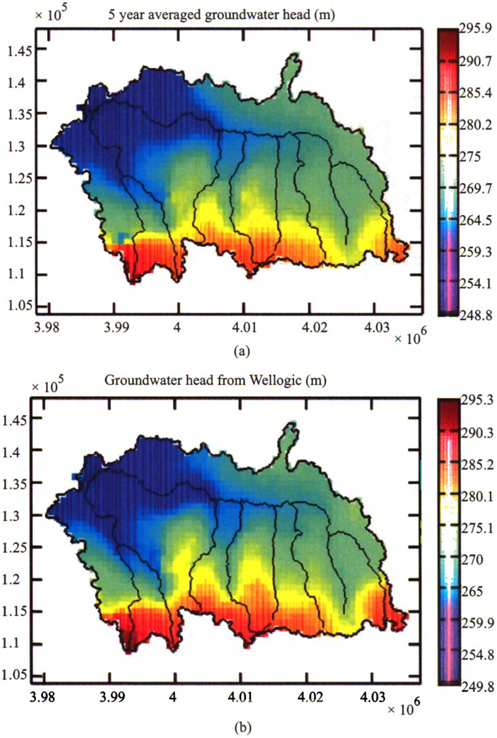
The model was started in 1998 with a rough initial estimate of groundwater heads. The initial guess was simply taken as the mean elevation of the ground surface and bedrock. If the model works properly, inaccuracies in the initial head would be mostly removed during the years prior to 2001. The resulting head is determined mostly by internal and boundary forcing. The groundwater model should converge to the steady state solution and fluctuate around it. We examine if the final state and the time-averaged state is reasonably close to the observed values. The 'observed' groundwater heads are a groundwater head map interpolated from well records. Maps of simulated groundwater heads (averaged from 2001 to 2005) are compared with the heads interpolated from wellogic database in Fig 3.14. Since the groundwater table fluctuates continuously, the 5-year average head is used for the comparison. The statistics of the comparison is given in Table 3.5. Excellent results have been achieved by the model to reproduce the observed data as evidenced by the high R^2 value. However, since the initial guess is bounded by the DEM, a large part of the variation in the data may have already been explained by the initial guess. It is thus legitimate to remain skeptical about the true ability of the model to find the correct states. To address this, we show the initial guess in Fig 3.14c and its comparison with observed data in Fig 3.15b. The statistics are also given in Table 3.5. We note that the initial data is negatively-biased as the mean error (ME) is -4.8 meters. Both the 5-year average state and the final state have much smaller bias. The R^2 and NASH have improved substantially after the simulation. We see that with rough initial estimates of head as initial conditions, the model was able to evolve to a much more realistic state, demonstrating that the internal and boundary forcings have been realistic.

From Figure 3.15 there seem to be some systematic under-estimation of groundwater heads toward the higher range. This can be attributed to the implied no-flow condition at watershed boundaries in the model. Groundwater divide does not necessarily coincide with watershed boundary, and surface water features on the other side of the boundary such as lakes and rivers can have influence on the groundwater heads.

Table 3.5. Metric of Groundwater heads comparison to wellogic data

	NASH	*RMSE*	*ME*	R^2
Initial Data	0.48	6.22	-4.81	0.83
Final Stage at 2005/12/31	0.94	2.07	-1.00	0.96
5 year Average	0.97	1.59	-0.81	0.98

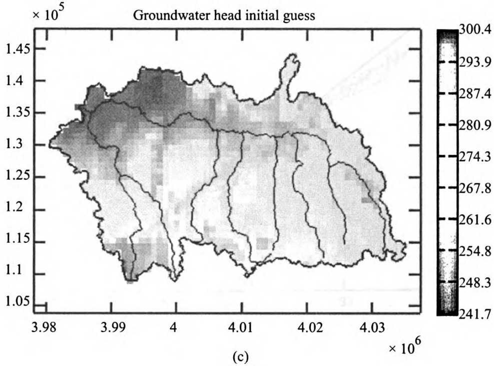
Figure 3.14. (a) 5 year averaged simulated groundwater, (b) observed heads interpolated from wellogic database, (c) initial groundwater heads used to start the model



(a)

(b)

Figure 3.14 (cont'd)



(c)

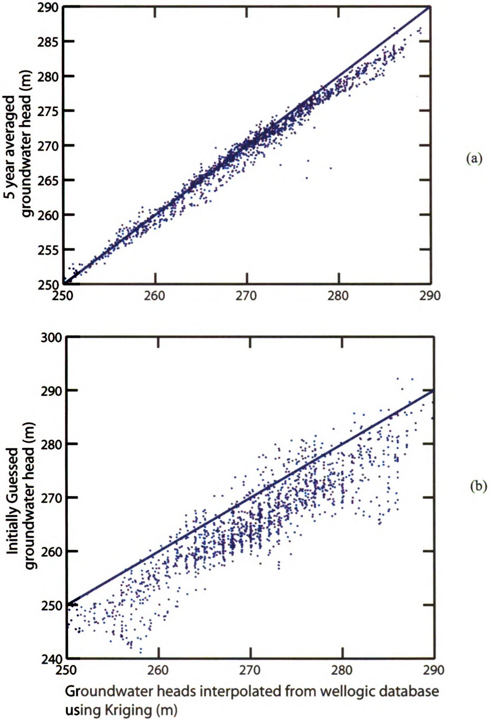


Figure 3.15. Observed vs simulated groundwater heads. (a) Observed data vs 5 year averaged value, (b) Observed data vs the rough initial guess. X-axis is the observed data obtained from Kriging interpolation of the wellogic data

3.3.3. Additional model results and hydrologic system of the RCR watershed

Table 3.6. Annual Average Fluxes compared to SWAT

Fluxes	SWAT (mm)	SWAT (% of prcp)	PAWS (mm)	PAWS (% of prcp)
PRCP	761.5	100.00%	760.3	100.00%
Sublimation	3.9	0.51%	9.8933	1.30%
Surface Runoff	87.94	11.55%	106.63	14.00%
Lateral Soil Q	0.63	0.08%	0	0.00%
Groundwater Q	78.61	10.32%	80.447	10.56%
Recharge	100.24	13.16%	82.634	10.85%
ET	552	72.49%	560.7	73.63%
Infiltration	NA	NA	760.3	100.00%

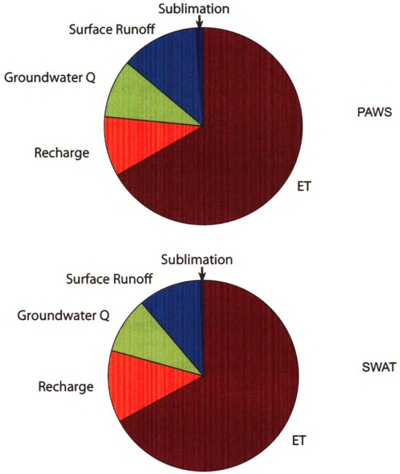


Figure 3.16. Pie chart showing the comparison of the average annual fluxes as a percentage of total precipitation between PAWS and SWAT

Recharge for the proposed model is calculated from the vadose zone model using equation (2.96). As discussed earlier, we take recharge as the flux that leaves the bottom of the soil layers and enters the saturated unconfined groundwater aquifer. Recharge can be negative which means that groundwater is discharging.

The model-estimated annual average of hydrologic fluxes for the years 1998 to 2005 are tabulated in Table 3.6 and graphically presented in Figure 3.16. There is a small difference of the total amount of precipitation applied to the two models because the discretization is different. We see that these fluxes, especially percentages of the precipitation, are in general similar to that estimated by the SWAT model. Infiltration is not directly extractable from SWAT results when SCS curve number method is used (when daily time step is employed) to calculate runoff. The proposed model reports 2% ET higher than the SWAT model while 2.3% less recharge and 2.5% more surface runoff. The SWAT model has been well-established as a long term hydrologic budget simulator. These results give us the confidence that the model is separating the precipitation in a reasonable manner.

3.3.4. Understanding the hydrology

After these comparisons we are finally ready to use the model for our own purposes. Besides streamflow predictions, a physically-based model has the advantage of explicitly elucidating the interactions of components in space and time. With this valuable tool we are able to gain insights of the inner workings of the hydrologic system. With the following discussions we want to answer two questions: (1) Where does the water in the rivers come from? (2) What is the driving force behind the annual cycle in the hydrograph? Why the hydrologic response to rainfall is so different in different seasons?

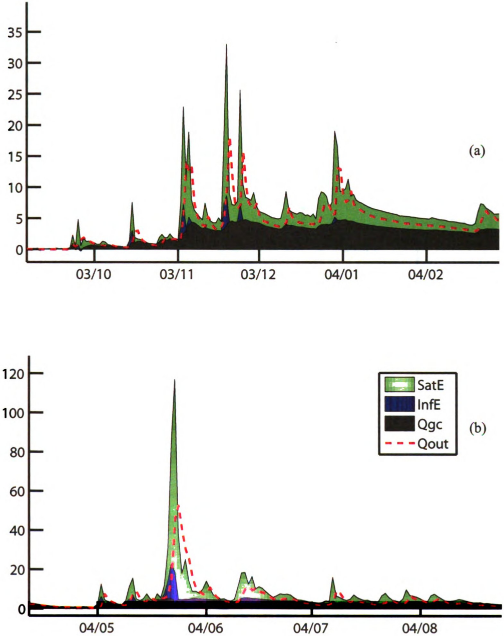


Figure 3.17. Hydrograph partitioning into streamflow generation mechanisms: Infiltration excess, saturation excess and groundwater contribution. X-axis are dates in YY/MM/DD format. SatE: Saturation Excess, InfE: Infiltration Excess, Qgc: Groundwater contribution, Qout: Red Cedar River outflow

Temporal patterns of hydrologic cycle

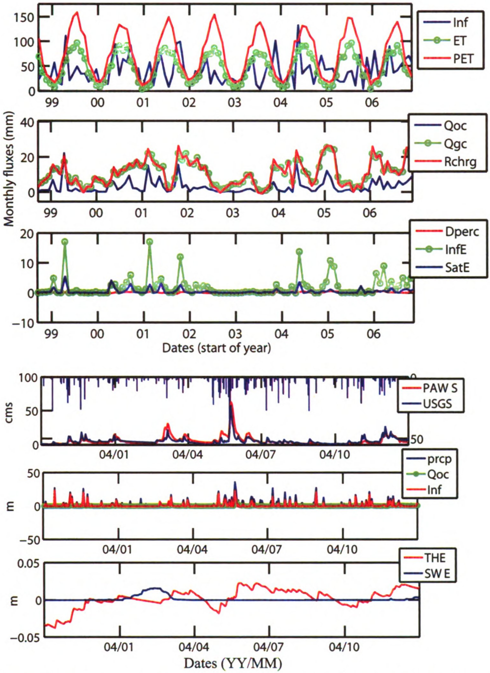


Figure 3.18. Temporal dynamics of fluxes and state variables. Upper: Monthly mean fluxes (mm). Lower: Daily fluxes (close up of 2004). Inf: Infiltration; Qoc: Overland contribution to channel; Qgc: Groundwater contribution to channel; Rchrg: Recharge; InfE: Infiltration Excess; SatE: Saturation Excess; Dperc: Deep Percolation

Figure 3.17 shows the streamflow partitioning into infiltration excess, saturation excess and groundwater contribution. From the model, saturation excess is defined as runoff generated when the whole soil column is saturated. Infiltration excess is calculated as the runoff that is not saturation excess. Due to the presence of other processes (evaporation, overland flow and river flow routing, etc), the partitions do not necessarily add up to be the flow at the gage at any instant of time. However, it stills gives us a good idea of the relative importance of the different processes. We observe clearly that the main contributions to streamflow are saturation excess and groundwater baseflow. The infiltration excess, on the other hand, occurs only at the few extreme rain events. After entering the stream, the flood wave is attenuated and become more smooth. There is some loss of water from runoff generation to final outflow due to evaporation.

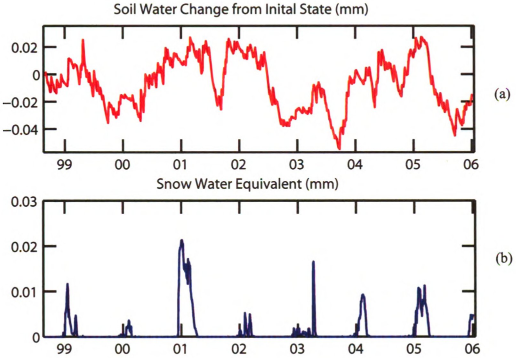


Figure 3.19. Basin average state variables (a) Soil Water Content (change from initial state), (b) Snow Water Equivalent. X-axis marks the beginning of each year (YY).

Monthly basin-averaged fluxes are presented in Figure 3.18. This figure well explains the seasonal hydrologic cycle of the RCR watershed and the impact of energy input on the system. Generally, starting in Spring, a number of warm days are responsible for melting the snow that accumulated in Winter. The first few precipitation events can be very pronounced in terms of streamflow generation. Because ET has not picked up and snow/ice is melting, the basin on average has high soil moisture content, as shown in Figure 3.19 by the state variable SW. After the effect of snowmelt has waned, the increase in ET gradually takes water from the system and we observe that baseflow starts to decline. From June to Aug, the large ET demand, created by heavy input of solar radiation and high temperature, as well as the vegetation/crop at full canopy, cannot be met by the precipitation and soil moisture. The suction of the soil and vegetation roots start to tap water from the groundwater. We observe rapid decline of soil moisture content in the Basin. Because soil is constantly dry in this period, it can infiltrate much more than in spring. The effects of precipitation events in summer are much more muted. In the fall, PET demand drops, the infiltrated water starts to accumulate and groundwater level rises. On the hydrograph we observe the baseflow begins to increase. The precipitation events induce much higher peaks than in summer. In winter, precipitation comes in the form of snow and the streamflow is provided by the groundwater. Since ET is very small in winter, infiltrated water is stored in the system. The model results show that the RCR watershed is a primarily energy-controlled watershed in terms of hydrologic cycle. Thus potential climate change in the future may profoundly influence the hydrology in the region.

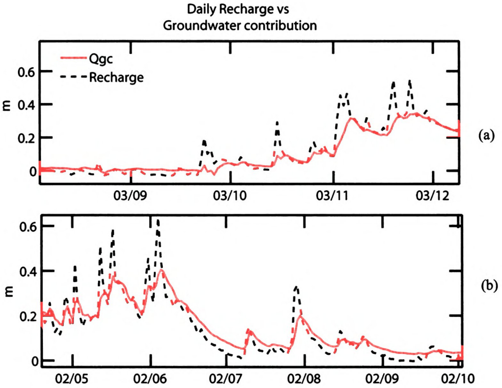


Figure 3.20. Daily fluxes of recharge and groundwater baseflow. (a) Rising limb of baseflow, (b) falling limb of baseflow. Again, dates are in YY/MM/DD format. Unit of the figure is m

We get a peek at the recharge-baseflow relationship by looking at Figure 3.20. As discussed previously, the geological configurations of the RCR watershed and its medium size allows for a rather rapid response of the groundwater system. The time lag between recharge and a response in groundwater baseflow is on the order of days. Thus, on a monthly chart as in 3.18, such a time lag is almost not noticeable. In Figure b, as it goes into summer, the recharge becomes increasingly small and falls below Qgc. The groundwater compartment is thus losing mass.

Spatial patterns of fluxes

Being a spatially distributed process-based model, the proposed model has the ability to examine the spatial variation of hydrologic fluxes due to different land uses, climate input, soil properties and geological configurations. Here we examine a series of spatial distributions of average annual fluxes presented in the form of gridded model output maps. The maps all possess the same Geographic Coordinate System as (GCS) (State Plane NAD1983, Michigan South)

Figure 3.21a shows the annual mean generation of saturation excess in the watershed. This map is compared to the topographic index map in Fig 3.22. The topographic index is a metric developed by [*Beven and Kirkby*, 1979] to evaluate the propensity of areas in the catchment to reach saturation and produce saturation excess. Areas that are prone to saturation tend to be near the stream channels or where groundwater discharge occurs. They can grow in size during rainfall while shrink during dry periods [*Beven*, 1978; *Dunne and Black*, 1970; *Dunne et al.*, 1975]. These areas are referred to as the variable contributing areas. [*Hornberger et al.*, 1998]. The TOPMODEL [*Beven and Kirkby*, 1979] attributes the majority runoff to saturation excess and return flow. The determination of contributing areas in a watershed is of importance from water quality perspective because these are the areas that are most prone to non-point source pollution generation.

The topographic index is calculated as:

$$TI = \ln(a/\tan\beta) \qquad (3.7)$$

Where a is the upslope contributing area per unit contour length(A/c) and tanβ is the local slope (-). Higher TI values correspond to higher likelihood of saturation excess. Although the two maps have different units, their similarity in spatial pattern can be visually examined. Upon visual inspections of Fig 3.22 and Fig 3.21a we observe the two maps agree well. Higher values of saturation excess are indeed found at areas with higher TI. Comparing the two maps with the DEM, it was apparent that most of the areas are cells surrounding the streams, which generally correspond to perennial stream network and flood plain. Local hillslope bottoms are also sources of saturation excess. This finding is shared by [*Ivanov et al.*, 2004a]. These areas are converging regions of the groundwater flow nets and receive inflow from their catchments. However, high saturation excess is also found on the Southwest of the watershed boundaries whose elevation is high. We find the phenomenon is explained by the thickness of the unconfined aquifer in this area (the glacial drift layer), which is significantly smaller than that in the rest of the watershed. The shallow depth to the bedrock limited the volume of water that soils can hold, thus this region gets saturated more easily. This example shows the various factors that control the runoff generation processes.

Infiltration excess, on the other hand, is more dependent on the land use type and soil properties. Fig 3.21b shows the spatially-distributed annual infiltration excess. We can see that the infiltration excess in agreement with the spatial distribution of saturated conductivity in Figure 3.8. Relatively high relief areas also record higher infiltration excess, a typical example being area (1) marked on the map because the large slope values allow runoff to occur more rapidly. On the other hand, infiltration excess is simulated consistently in the urban areas on North East corner of the watershed because of the impervious areas.

The infiltration map in Figure 3.23a looks somewhat similar to the inverse of the saturation excess map. The lower infiltration of the stream cells and hillslope bottoms is explained by frequent groundwater discharging. The urban areas have low infiltration as, intuitively, impervious cover is higher. Over the large area in the center that is agricultural, infiltration is higher on the East than on the West. This is again due to the spatial distribution of soil types.

Quantifying the amount of recharge that the aquifer receives and the spatial distribution can be of great interest to water management and ecological planning. Groundwater contribution to streams in summer has lower temperature than average stream water temperature. It provides the prime habitat for many fish species. The groundwater also serves as one of the major sources of water supply for human consumption in many parts of the world. The annual average recharge map (Figure 3.23b) again confirms the role of regional groundwater flow plays in the hydrologic cycle. The stream cells and hillslope bottoms serve as the exfiltration point. We come to the conclusion that recharge in the region, although highly dependent on both space and time, is very much a topography-controlled phenomenon.

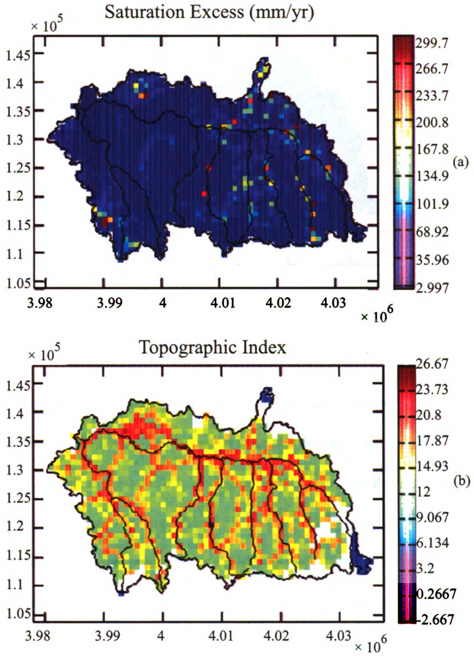


Figure 3.21. Average annual runoff generation (mm) (a) Saturation Excess (b) Topographic Index

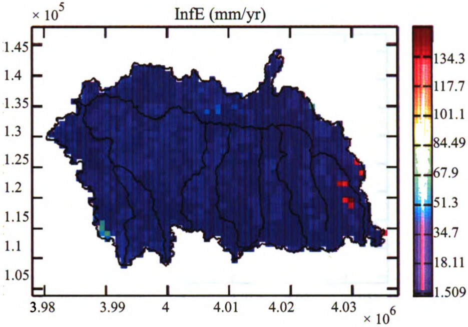


Figure 3.22. Average Annual Infiltration Excess

The total ET and the ground evaporation (EvapG) maps are shown in Figure3.24. The ground evaporation (EvapG) is defined as water that is evaporated from the ground surface. Impervious areas have much higher EvapG than agriculture and forested areas since water cannot infiltrate. The ET is also shown to vary significantly in space, from 140mm to 802 mm inside the watershed. And the variation is apparently related to topography. At hillslope bottoms and lowland areas, regional groundwater flow supplies moisture to the soil and thus actual ET is higher at these areas.

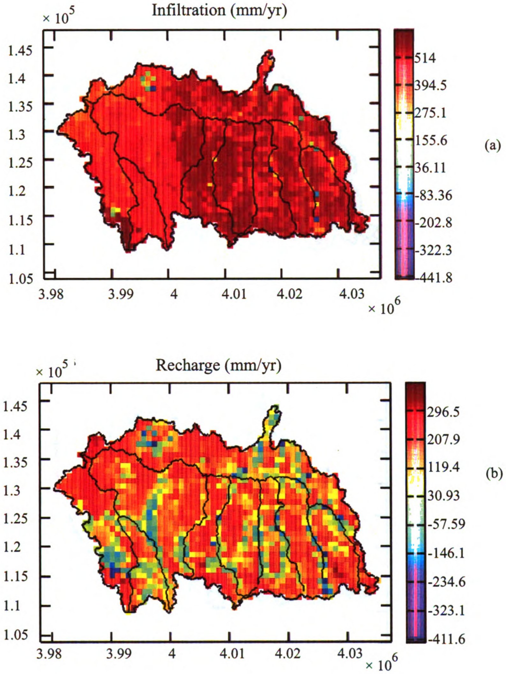


Figure 3.23. Annual Infiltration (a) and recharge map (b)

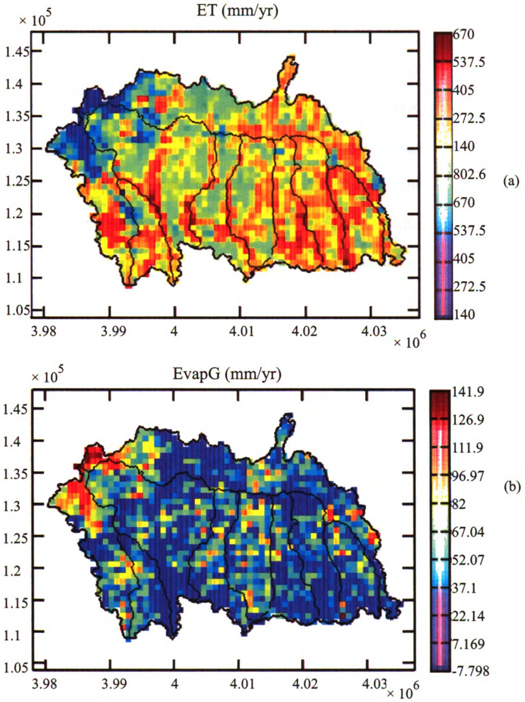


Figure 3.24. Annual ET (a) and EvapG (b) map. EvapG is evaporation from the ground.

3.4. Summary and Conclusions

The newly developed model achieved high performance metrics in terms of streamflow prediction not only during the calibration period (NASH=0.68) but also in the verification period (NASH=0.614). Flow at an internal gaging station is also described well. Starting from a rough initial estimate of the groundwater heads, the model is able to evolve to a more realistic state, and achieve excellent comparisons with the observed groundwater heads. The annual mean hydrologic fluxes, including ET, infiltration, recharge and groundwater contribution are close to those estimated by the SWAT model. All these results indicate that the proposed model is a valid representation of the hydrologic system of the watershed.

The model is able to explicitly detail the complex interactions of processes in the watershed in space and time. The watershed is a subsurface-dominated system with saturation excess found to be the main runoff generation process. The places that produce saturation excess are cells near streams and groundwater discharging areas which agree well with the Topographic Index map computed from method of [Beven]. Infiltration, recharge and ET are also found to be strongly related to topography and groundwater flow.

The geological characteristics of the watershed permit a fast-responding subsurface hydrologic system. The large seasonal variation of energy input introduces a strong annual cycle in the groundwater baseflow and markedly different responses of surface runoff in different seasons. Responses to summer precipitation are often muted while large rain in early spring tends to produce much higher peaks in the hydrograph.

3.5. Limitations and future research

From the above modeling practice and comparison with observed data, we can summarize some limitations of the present model and some future development considerations:

1. Soil moisture lateral flow needs to be added. Although for the study watershed soil lateral flow seems to be a small component (0.08% of precipitation from SWAT), it should be added for the completeness of the model. It may be more important for other watersheds
2. Soil temperature and freezing module need to be added. At least a simple soil temperature should be incorporated to allow the parameterization of soil freezing.
3. The vegetation module needs to be improved for better modeling of the interaction of vegetation and the hydrologic cycle.
4. The model should be applied to larger watersheds and watersheds with different land use/ geology/ soil/ climate.
5. The effect of spatial resolution on the results should be carefully examined

3.6. Software package

A software package with Graphical User Interface (GUI) has been developed for the present model to help interfacing with data and building up the models. Along with this package are tools to visualize model results and calibrate the model.

Due to some iterative approaches used to solve the PDEs (e.g. conjugate gradient for the groundwater flow equation, the celia's iteration for Richards' equation), the computational time can vary depending on the flow conditions and soil properties. It has been found that larger N values of the van Genuchten function (Eq. (2.57)) can lead to more iterations to converge, since the soils properties are more nonlinear with larger N. Also at high flow times the time steps of surface flow modules are reduced to meet stability criterion. Thus the computational time required to complete a model cannot be completely predicted ahead of time. However, for the RCR watershed model, the simulation time varies from 2.5 seconds to 3.5 seconds for each day of simulation.

Chapter 4. Estimating Longitudinal Dispersion and Surface Storage in Streams Using Acoustic Doppler Current Profilers

4.1. Introduction

The ability to accurately describe the transport of conservative as well as reactive solutes in streams and rivers is fundamental to many branches of science and engineering including stream ecology, geomorphology, river engineering, water quality modeling, risk assessment etc. It is well known [*Day*, 1975; *Thackson and Schnelle*, 1970] that the classical Taylor theory [*Taylor*, 1954], which leads to the following advection dispersion equation (ADE), cannot adequately describe tracer transport in natural streams.

$$\frac{\partial C}{\partial t}+u\frac{\partial C}{\partial x}=D\frac{\partial^2 C}{\partial x^2} \tag{4.1}$$

Here C is the solute concentration, u is the mean flow velocity and x and t denote space and time respectively. Extensive tailing and the persistence of skewness in the observed tracer data (which cannot be adequately described using equation (4.1)) led to the development of models based on the concept of transient storage (TS) [*Bencala and Walters*, 1983; *Thackson and Schnelle*, 1970]. The TS models often use conceptualizations based on two distinct zones: the main channel in which advection and dispersion are the dominant processes and the storage zones that contribute to TS.

The governing equations for the TS model appear as follows:

$$\frac{\partial C}{\partial t} = -\frac{Q}{A}\frac{\partial C}{\partial x} + \frac{1}{A}\frac{\partial}{\partial x}\left(AD_S\frac{\partial C}{\partial x}\right) + \frac{q_L}{A}(C_L - C) + \alpha(C_s - C)$$

$$\frac{\partial C_s}{\partial t} = \alpha\frac{A}{A_S}(C - C_s) \tag{4.2}$$

Here C and C_S are the solute concentrations in the main channel and the storage zones, A and A_S are the sizes of the main channel and storage zones, and α is a first-order exchange rate between the main channel and the dead zones. D_S denotes the dispersion coefficient in the TS model. The subscript "S" is used to indicate the fact that the dispersion coefficients estimated using the ADE and TS models generally tend to be different for the same river reach under similar flow conditions. Since the TS model has additional terms for the dead zones, the coefficient D_S represents only the shear-flow contribution while the ADE estimates of dispersion tend to be higher as D includes the effects of dead zones to some extent [*Deng and Jung*, 2009]. By comparing the dispersion coefficients obtained from the TS and ADE equations, we will be able to assess the relative contribution of dead zones (or storage zones) in different stream reaches and this idea will be explored later on in this chapter. Efforts to interpret model parameters and to relate them to physical stream characteristics are often confounded by the inability of current stream tracer techniques to separate TS processes [*Gooseff et al.*, 2005] as well as mathematical difficulties associated with parameter estimation, especially in the presence of competing parameters and false/singular convergence [*Runkel et al.*, 1998]. In this chapter, we propose two approaches for independent, field-based estimation of the longitudinal dispersion coefficient and the contribution of the total storage A_S that is due to surface features such as vegetation, eddies and pools, meander bends etc. This type of storage is called surface storage to separate it from the hyporheic or sub-surface storage (exchange with near-bed sediments or groundwater) found in many stream reaches. By estimating the surface storage in different stream reaches, the importance of hyporheic storage can be assessed. Both techniques reply on the measurement of high-resolution velocity fields within the stream channel. The dispersion coefficient can be estimated by directly integrating the velocity fields using the shear-flow dispersion theory. The size of the surface storage zones can be estimated using wavelet decomposition of the velocity data. Both estimates are expected to lead to improved model parameters and solutions.

Wavelet analysis has been used extensively in the past decade to analyze both time series and spatial data in geophysical and engineering applications. Examples include the analysis of streamflows [*Coulibaly and Burn*, 2004], seismogram data [*Lockwood and Kanamori*, 2006], longterm trends in geomagnetic activity [*De Artigas et al.*, 2006], and ocean waves [*Pairaud and Auclair*, 2005]. Wavelet analysis [*Mallat*, 1989]allows the original signal or image (in our case a two-dimension image of mean velocity in the channel as a function of river depth and width) to be split into different components so that each component can be studied with a resolution that is suitable for its scale. In particular, the signal can be decomposed into slow changing (i.e., coarse or low frequency) features and rapidly varying (i.e., fine or high frequency) features using low-pass and high-pass wavelet filters at different levels (or scales) in a multilevel decomposition. This feature of wavelet analysis is particularly attractive for studying TS processes in streams since we are looking for fast and slow-flowing regions of the river. Previous applications of two-dimensional wavelets include characterization of permeability anisotropy [*Neupauer et al.*, 2006], studies of spatiotemporal dynamics of turbulence [*Guan et al.*, 2003] and analysis of spatial rainfall data [*Kumar and Foufoulageorgiou*, 1993].

The longitudinal dispersion coefficient (D) is an important parameter that describes the transport of solutes in streams and rivers. Accurate estimation of the dispersion coefficient is important from human health and public safety perspectives as the parameter is needed to predict contaminant concentrations near drinking water intakes and receiving water bodies such as lakes or oceans. Once a solute is released into the stream and becomes vertically and laterally well-mixed, longitudinal dispersion is the primary mechanism responsible for spreading the tracer plume and for reducing peak concentrations. The one-dimensional transport of solutes following the initial period of mixing can be described using the advection-dispersion equation (ADE) [*Fischer*, 1979; *Rutherford*, 1994]. The coefficient D in equation (4.1) can be estimated using a number of empirical relations available in the literature [*Deng et al.*, 2001;

Fischer, 1979; *Seo and Cheong*, 1998]; however, these estimates are generally known to exhibit large variability. Dispersion estimates from tracer studies (which usually involve fitting the observed data to a solution of equation (4.1) for appropriate boundary and initial conditions) are generally believed to be more reliable; however, significant time and resources are needed to conduct tracer studies, especially on large rivers. An alternative approach, which is the main focus of this chapter, is to estimate D from the theory of shear flow dispersion [*Fischer*, 1979]:

$$D = -\frac{1}{A}\int_0^B u'(y)h(y)dy\int_0^y \frac{dy'}{D_y h(y')}\int_0^{y'} u'(y'')h(y'')dy'' \tag{4.3}$$

where A is the channel cross-sectional area, y is the transverse coordinate which varies from $y = 0$ at one bank to $y = B$ at the other, $h(y)$ is the depth of flow at a given y location, D_y is the transverse mixing coefficient, $u'(y) = \overline{u}(y) - \overline{U}$ is the deviation from the mean velocity, $\overline{U}$ is the cross-sectional mean velocity and $\overline{u}(y) = \int_0^z u(y,z)dz$ is the depth-averaged velocity at y. Eq. (4.3) is based on several assumptions as described in [*Carr and Rehmann*, 2007; *Fischer*, 1979] including one-dimensional flow (no vertical and transverse gradients in concentration, that is well-mixed). Eq. (4.3) Although the theory for estimating the dispersion coefficient from the velocity field has been around for nearly four decades, the single most limiting factor in the application of Eq. (4.3) in the past was the measurement of detailed velocity distribution and depth across the river. Significant time and resources are needed when instruments designed to make point-measurements (such as the Price-AA and Pygmy current meters) are used to measure the velocity fields in streams. However, this situation has changed with the advent of Acoustic Doppler Current Profilers (ADCP). While the ADCP technology itself is not new, early models of the instruments were geared towards oceanographic applications. Recent ADCP models designed keeping rivers, streams and other shallow inland water bodies in mind can be used to quickly make velocity and bathymetry measurements over long distances with high accuracy and resolution. This new capability introduced the opportunity to address important questions involving inland water bodies [*Dinehart and Burau*, 2005; *García et al.*, 2007; *Phanikumar et al.*, 2007] including the calculation of D with relative ease using Eq. (4.3). However, there are several issues that need to be examined before comparing tracer estimates with ADCP results based on equation (4.3) and these are described below.

The dispersion coefficient estimated from tracer data is often used as the "true" value to evaluate other methods. The tracer estimate represents the bulk, reach-averaged mixing strength in a given reach and is valid only for the particular stream reach and the flow conditions for which the experiment was conducted. The ADCP estimate, on the other hand, represents a "point estimate" along the length of the river (valid for a particular cross-section for which data was obtained). If the channel is fairly uniform and the transect represents the conditions in the entire reach, then the point estimate can be expected to represent reach-averaged conditions. If the river reach exhibits significant heterogeneity in properties, then a reach-averaged estimate can be calculated using a distance-weighted average as suggested by [*Rutherford*, 1994]:

$$\bar{D} = \frac{1}{L}\sum_{i=1}^{n} D_i \Delta x_i \tag{4.4}$$

where $\bar{D}$ is the reach-averaged dispersion coefficient, Δx_i is the length of the sub-reach containing the ADCP transect at i and L is the total length of the river reach. A number of mechanisms including surface storage in dead zones (e.g., vegetation, woody debris) and solute exchange and retention within the hyporheic zone are known to produce an effect that is somewhat similar to the effects of shear-flow dispersion in streams. The transient storage (TS) model [*Runkel*, 1998], which describes the temporal trapping (and release) of solute particles in the dead zones is often used to describe transport when dead zones play an important role and the ADE cannot adequately describe the data. Ignoring lateral flow contributions into the channel due to groundwater, the TS model can be written as shown below [*Bencala and Walters*, 1983; *Runkel*, 1998].

Fisher et al. [*Fischer*, 1979] numerically evaluated Eq. (4.2) for both laboratory flumes as well as natural streams and compared their results with tracer estimates (see Table 5.3 in [*Fischer*, 1979]). Their velocity measurements were obtained using standard current meters. Carr and Rehmann [*Carr and Rehmann*, 2007] recently evaluated Eq. (4.2) using the velocity and bathymetry measurements obtained from ADCPs and compared their estimates with tracer results for ten US rivers. Half of their ADCP estimates are found to be within 50% of the values from tracer studies, and 85% are within a factor of 3. They conclude that the ADCP method is at least as accurate as the best empirical formula considered in their work. While these results are encouraging, many questions remain unanswered. First, if the ADCP estimates are only as good as the best empirical relations (which generally produce estimates within an order of magnitude) then the ADCP method is not very attractive since it is easier to use the empirical relations. Second, in many stream ecosystem studies the primary interest is not in dispersion but in other processes (e.g., biogeochemical processes, mortality/loss rates of bacteria or viruses, storage zone sizes and reaction rates) but accurate estimates of dispersion are still needed to adequately describe these processes. If the ADCP method has the potential to produce dispersion estimates that are as accurate as the tracer method, then the result has important implications for studies involving stream and river ecosystems. The aim of this paper is to explore this question in more detail. The tracer data used in [*Carr and Rehmann*, 2007] spanned a period of nearly three decades (1967 to 1991) while the ADCP measurements coved the period from 2000 to 2004. Changes in both channel cross-section and (local) slope are possible in the period following the tracer studies which could potentially influence the results. In addition, different approaches were used to estimate D from the tracer data - from a routing method [*Rutherford*, 1994] to fitting a line to the variance obtained from the tracer breakthrough curve [*Fischer*, 1979] to modeling the tracer response curve as a scalene triangle [*Jobson*, 1997]. These methods are known to produce vastly different results when applied to the same river reach under similar flow conditions introducing additional sources of uncertainty into the comparison between ADCP and tracer estimates of the dispersion coefficient. The aim of this paper is to contribute additional tracer and ADCP datasets in support of the analysis reported in [*Carr and Rehmann*, 2007] focusing mainly on datasets collected at the same time. Recognizing that the tracer estimates of D have their own sources of error and uncertainty, our aim is to understand if the ADCP method has the ability to produce estimates that are comparable to the tracer method, especially when using multiple/repeated transect data at the same site, a procedure that tends to average errors involved in the data. Due to the randomness of the turbulent flows involved, raw ADCP data usually contain noise which depends on the river characteristics and the operating conditions of the instrument. One of the aims of this paper is to systematically examine the effects of different post-processing methods (e.g., smoothing) on the dispersion results. Finally we will examine the limitations of the ADCP method in order to better understand conditions under which the method can produce reliable dispersion estimates.

Tracer data from several mid-western streams have been used to compare the ADCP estimates of dispersion and surface storage with tracer-based estimates. Data for the Grand River, Michigan is new and was published in [*Shen et al.*, 2008]. Tracer studies on the RCR (a tributary of the Grand) and the wavelet decomposition method for estimating surface storage are described in [*Phanikumar et al.*, 2007]. Tracer and

ADCP data collected by USGS staff on the Ohio River and the St. Clair River are used for comparing the dispersion estimates (tracer versus ADCP).

4.2. Description of Sites

The Red Cedar River (RCR) is a fourth-order stream in south central Michigan, United States that drains a landscape dominated by agriculture and urbanization. It originates as an outflow from Cedar Lake, Michigan, and flows into East Lansing and Michigan State University (MSU). The river then connects with the Grand River in Lansing, Michigan. The total stream length is approximately 70 km. The river and its tributaries drain an area of about 1,230 km2, one fourth of which is drained by Sycamore Creek (Figure 1). More general information is available in previous chapters. Limestone, shale, sandstone, coal, and other sedimentary rocks arranged in almost horizontal layers compose the bedrock surface of the Red Cedar Watershed. The river is characterized by a meandering channel and low stream gradient. A USGS gauging station (04112500) is located at the Farm Lane Bridge (Figure 1) and measures runoff from about 75% of the basin. The study reach was bounded by Hagadorn Bridge on the East and the Kalamazoo Street Bridge on the West (4.1). The RCR meanders through the MSU campus over a stretch of approximately 5 km (our study reach in this paper) and has an average slope of 0.413 m/km. The river width varies considerably, from 16.3 m to 40.4 m, with an average of 28.1 m (4.1). The general trend is that the width decreases for the initial 2400 meters, increases over the next 2100 meters and finally decreases for the remaining 600 meters.

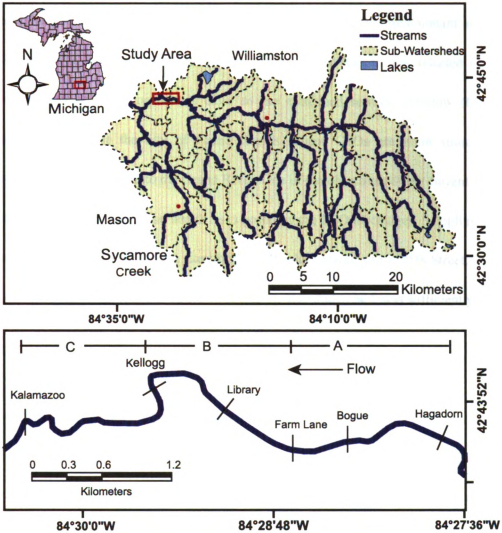


Figure 4.1. Red Cedar River and the sampling locations

The Grand River is a 420 km long tributary to Lake Michigan (4.2) and the tracer study was conducted on a 40 km stretch of the river, starting from the city of Grand Rapids and extending to Coopersville. Surficial geology of the Grand River Basin is dominated by rivers crisscrossing the moraines and outwash plains formed by extensive glaciation during the Pleistocene (Supporting Information). Till plains, moraines, kames, and eskers of the Port Huron system are the predominant surface features. The Ann Street Bridge near downtown Grand Rapids was selected as the injection point. This location is close to the combined sewage overflow (CSOs) outfalls that serve the city of Grand Rapids. These CSOs are point sources of pathogens discharging into the Grand River. One of the objectives of the current study was to examine the potential health risks posed by the CSO discharges. Sampling was carried out at four downstream sites; bridges (Wealthy Street, site 1; 28th Street, site 2; Lake Michigan Drive, site 3; 68th Street, site 4). The study reach was sufficiently long to make watershed influences important. A U.S. Geological Survey (USGS) streamflow gaging station (04119000) is located 1.05 km upstream from the first sampling site. The discharge on the test date was on the end of a recession limb with the values during the experiment gradually declining from 3230 to 3010 cubic feet per second. The study reach is a perennial gaining stream. No precipitation event was reported in the three days prior to the experiment. Contributions from the tributaries and baseflow together, termed lateral inflow, are important to correctly describe the downstream transport of both tracers. Although lateral inflow is often neglected in tracer studies conducted on relatively short river reaches, it is not negligible in our case.

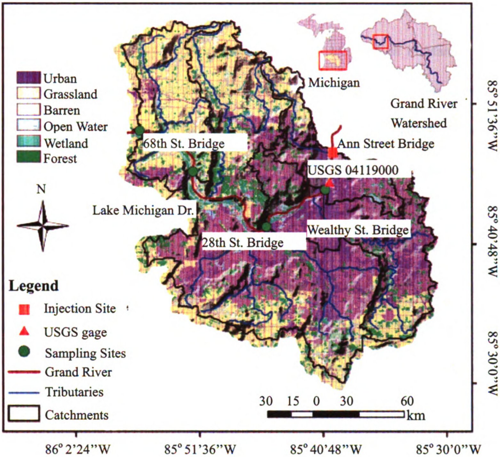


Figure 4.2. Red Cedar River and the study region showing the sampling locations of the Grand River tracer study

4.3. Materials and Methods

Four tracer studies (slug additions) were conducted on the Red Cedar River in summer 2002 on 17 May, 31 May, 6 June, and 21 June. Average discharge (Q) recorded on these four days at the USGS gage near Farm Lane Bridge was 16.82, 14.41, 19.06 and 2.49 m3/s respectively. Discharge and hydrodynamic data were collected from ADCP surveys during the period 2002–2006. ADCP data could not be collected during the tracer tests in 2002 but on two occasions the discharge values were similar to those during the tracer tests. ADCP data collected on 19 March 2006 (Q = 19.89 m3/s) and 29 September 2005 (Q = 2.0 m3/s) were used to compare with results from the tracer studies conducted on 6 and 21 June 2002 respectively. Because of low-flow conditions (long travel times of the dye cloud), the tracer study on 21 June 2002 could not be completed. For this discharge, we present results for only one sampling location. Tracer transport was described using the TS equations [*Bencala and Walters*, 1983]. A one-dimensional hydrodynamic model based on the St. Venant equations was used to examine two of the estimated parameters in the TS equations (mean velocity and dispersion coefficient). For all the four slug injections, the dye was released at the Hagadorn Bridge and samples were collected at the Farm Lane, Kellogg and Kalamazoo Bridges, respectively (4.1). The distances to the three sampling locations from the point of release are: 1400 m (Farm Lane), 3100 m (Kellogg) and 5079 m (Kalamazoo). Flourescein dye was released in the middle 75% of the channel to ensure conditions of instantaneous mixing. Sampling was done at the middle of the cross section for each dye release. Fluorescein solution with a concentration of 179.06 g/L was used for all the tracer tests. The volume of the dye used was based on a desired peak concentration range of 10–20 mg/L at the last sampling point. Samples were analyzed using a Turner10-AU field fluorometer (Turner Designs Inc.,Sunnyvale, CA). Observed tracer data was checked for mass conservation. The mass of the tracer in the dye cloud as it passed a sampling point was calculated using the relation

$$m(t) = Q\int_{0}^{\infty} C(t)\,dt \tag{4.5}$$

where C(t) is the dye concentration and Q is the discharge at the sampling point. The fractional recovery was calculated by m/M where M is the mass injected. The fractionalrecovery values for all the flows (and at different sampling locations) varied from 0.91 to 1.5 with a mean value of 1.29 and a standard deviation of 0.16. To facilitate comparisons with mathematical models and to produce a data set that obeys dye mass balance, fractional recovery corrections were applied to the individual concentration values as reported by [*Atkinson and Davis*, 2000]. Point velocities and river stages were measured on eight consecutive days in 2002 (4–11 April) to calibrate the hydrodynamic model. Price AA current meters and a 16-MHz Sontek acoustic Doppler velocimeter (ADV) (Sontek/YSI Inc., San Diego, CA) were used to measure velocity profiles and discharge in the river. In addition, a vessel-mounted, down-looking 1200 kHz Rio Grande ADCP (Teledyne-RD Instruments, Poway, CA) was used to measure discharge and three-dimensional velocity fields in the river using an Oceanscience2 Riverboat (a low-drag trimaran) equipped with radio modems for real-time deployments. By towing the vessel-mounted ADCP across a river transect (perpendicular to the flow direction), we obtained snapshots of instantaneous velocity fields in the river which were later used for wavelet analysis (described below). Typical boat speeds were in the 0.5–1 m/s range. Several steps were involved before analyzing the ADCP data. The speed at which the ADCP was towed across the river depended on the discharge and the ADCP operating conditions. The ADCP data was exported from WinRiver (the ADCP operating software) for further processing in MATLAB including smoothing and removal of bad ensemble values. Details of ADCP principles and processing are described by [*Dinehart and Burau*, 2005]. The discharge values obtained from ADCP measurements at the Farm Lane Bridge were compared with those from the USGS gauging station on several occasions over a period of four years and an excellent agreement was obtained.

For the Grand River tracer study, Rodamine WT 20% (weight) solution was used. P22 was obtained from Samuel Farrah, University of Florida, Gainsville, Florida and was maintained on the host *Salmonella typhimurium* LT-2 (ATCC 19585). P22 stock was grown by inoculating 100 ml of log-phase *S. typhimurium* host with 1 mL of P22 stock (~ 1011 pfu/mL) and incubated at 37 °C for approximately 3–5 h. After incubation, 0.01 g of lysozyme and 3 mL of 0.2 M sterile EDTA were added to the flask and mixed well. The culture was then centrifuged at 4000 rpm for 10–15 min, and the supernatant was filter sterilized through a 0.45 μmmembrane. P22 stock was stored at 4 °Cuntil used. RWT and P22 solutions were injected into the river (slug release) from the Ann Street Bridge on May 8, 2006 at 7:00 a.m. A total of 8770 g of RWT and 16 L of bacteriophage P22 (4×1011 PFU/ml)were released. At each station, grab samples were collected from just below the surface using manual sampling. Two samples were taken at the same time. One was stored in a dark cooler for RWT analysis. A 5× trypticase soy broth (TSB) was added to the other sample to stabilize the bacteriophage for P22 analysis. All samples were kept on ice and were analyzed within 48 h in the laboratory. Meanwhile, water temperature, pH, suspended solids, and weather data (i.e., ambient temperature, rainfall, wind, etc.) were noted during sampling. A Turner Designs 10 AU field fluorometer (Turner Designs, Inc., Sunnyvale, California) was used to initially detect the dye at the first three sites. The sampling frequency for both tracers was increased after receiving a RWT signal. RWT samples were analyzed in the laboratory using the same 10 AU unit. Water samples were assayed for P22 following the double agar layer procedure [*Adams*, 1959]. Samples from site 1 were diluted to a 10-3 concentration, and between 1 and 2mL of each sample in at least duplicate were assayed for the phage presence on tryptic soy agar. The plates were incubated for 24 h at 37 °C. The detection limit of this method is less than one plaque-forming unit per milliliter. Total suspended solids concentration was determined according to Standard Method 2540-D—total suspended solids dried at 103–105 °C [*Greenberg et al.*, 1992].The river is significantly wider in comparison with the numbers reported inmanyprevious tracer studies. Therefore, to obtain a better idea about the lateral variability at each station, sampling was done at multiple locations (left, right, and center) on each bridge except at site 3, where sampling was done at two locations (left and right) due to the presence of an island in the middle of the channel. Because of the long travel time to site 4, P22 data was not collected. In addition to the manual sampling, a submersible fluorometer (Turner Designs SCUFA) equipped with a data logger and programmed to measure RWT concentrations every 10 s, was deployed at sampling sites 1, 3, and 4. Discharge at all the sampling locations was measured using a Teledyne - RD Instruments (1200 kHz) Rio Grande acoustic doppler current profiler (ADCP).

For estimating the dispersion coefficient using the ADCP technology, Data from a total of 505 ADCP transects collected from seven rivers in the states of Ohio, Indiana and Michigan are used in the present study. In addition, seven tracer studies have been conducted on some of the rivers. Details of the rivers are summarized in Table 1 and maps of the sites showing the locations of the ADCP transects are shown in Figure 1. Out of the seven rivers, simultaneous tracer and ADCP data are collected for three rivers (Ohio River, Grand River and Burns Ditch). For one river (Red Cedar River) tracer data was collected during summer 2002 while ADCP data was collected for similar flows between 2003 and 2006. No tracer data is available for three rivers (Muskegon River, Thornapple River and St. Clair River) but estimates from the ADCP method and empirical relations are shown for comparison. Where conditions warranted we ran multiple transects at the same location and at multiple locations within the same river reach. Details of the tracer study and modeling for the Grand River and Red Cedar River are available in [*Shen et al.*, 2008]. Tracer and ADCP data for the Ohio River (only ADCP data for the St. Clair River) were collected by USGS staff and details are available in [*Holtschlag and Koschik*, 2003; *Koltun et al.*, 2006]. A continuous dye release was conducted on the Portage Burns waterway

(Burns Ditch) on June 21, 2008 using Rhodamine WT. Although the aim of the tracer study is to understand nearshore processes in Lake Michigan, concentration –time data collected within the stream are used to estimate a dispersion coefficient by fitting an analytical solution for continuous release [*Chapra*, 2008] to the data. For the Ohio River, Rhodamine-WT was released at one of the banks and the tracer did not completely mix within the study reach. Breakthrough data is not available in the form of concentration versus time data; however, concentration values are reported within the channel cross-section (at different depths and distances from the bank) for several different locations. We computed the cross-sectional average concentration at different stations and fitted the spatial data to the unsteady ADE to compute a dispersion coefficient. This method will likely introduce some error since the tracer is not fully mixed to justify the use of one-dimensional ADE; however, the estimated dispersion coefficient described the mean concentration values at different stations accurately after the first few sampling locations.

4.3.1. Transient Storage Modeling

Tracer transport in the RCR and GR was described using the TS equations (4.2) which describe transport in the main channel and the storage zones respectively [*Runkel et al.*, 1998]. If Eq. 4.2 applied on a reach basis, then the velocity (Q/A) in (2) is a reach-averaged value and, in general, is not equal to the local (point) velocity u. The TS model equations were solved using a fourth-order accurate compact numerical scheme [*Demuren et al.*, 2001]. Briefly, the spatial derivatives were approximated using a fourthorder scheme with spectral-like resolution and a low-storage fourth-order Runge-Kutta scheme was used for temporal differencing. The resulting tridiagonal matrix system of equations was solved using the Thomas algorithm. A low Courant number of 0.15 and a uniform grid of 1001 points were used for all model runs. The boundary and initial conditions for the model were as follows. The river was assumed to be initially at zero tracer concentration. The upstream boundary was modeled to simulate slug release into the main channel and the storage zone was assumed to be initially solute-free. A no-flux boundary condition was specified at the downstream boundary for the transport equations. Parameters in the TS model (i.e., A, AS, D, a) were estimated using a global optimization procedure, the Shuffled Complex Evolution (SCE) algorithm [*Duan et al.*, 1992] by minimizing the root-mean-square error (RMSE) between the observed data and the model. This algorithm found wide applications in the hydrologic research literature and was shown to be robust and efficient in finding the global minimum. For all the dye studies, optimal parameters were obtained in 3000–6000 iterations on a 512-core Western Scientific Opteron Cluster computing system at MSU. The RCR is generally a gaining stream; however, within the study reach the gain was not significant enough to change the TS parameters. After running two separate optimizations, with and without qL, we decided to use the parameters obtained with qL = 0. The fourth-order accurate compact scheme used to solve the TS equations in this paper was tested extensively and was used to solve similar sets of equations in the past [*Phanikumar and McGuire*, 2004]. To assess the accuracy of the compact scheme for solving the TS equations, we compared our numerical solutions with the analytical solutions reported by [*De Smedt et al.*, 2005]:

$$C(x,t) = \int_0^t \left[\alpha + \left(\frac{x^2 - v^2\tau^2}{4D\tau^2} - \frac{1}{2\tau} - \alpha \right) J\left(\alpha\tau, \frac{\alpha(t-\tau)}{\beta} \right) - \alpha J\left(\frac{\alpha(t-\tau)}{\beta}, \alpha\tau \right) \right] C_1(x,\tau)\, d\tau$$

$$J(a,b) = 1 - e^{-b} \int_0^a e^{-\lambda} I_0\left(2\sqrt{b\lambda}\right) d\lambda = 1 - e^{-a-b} \sum_{n=1}^{\infty} \frac{a^n}{n!} \sum_{m=0}^{n-1} \frac{b^m}{m!} \tag{4.6}$$

where IO is the modified Bessel function of zero-th order, β = (As/A) and C1 (x,t) is the classical solution to the advection-dispersion equation with the same initial and boundary conditions [*Chapra*, 1997]:

$$C_1(x,t) = \frac{M}{A\sqrt{4\pi Dt}} \exp\left(-\frac{\left(x - ut\right)^2}{4Dt} \right) \tag{4.7}$$

where *M* is the mass of tracer released. Comparisons with our numerical solutions obtained using 200 grid points are presented in Figure 5 for different values of a. An excellent agreement is noted between the two solutions.

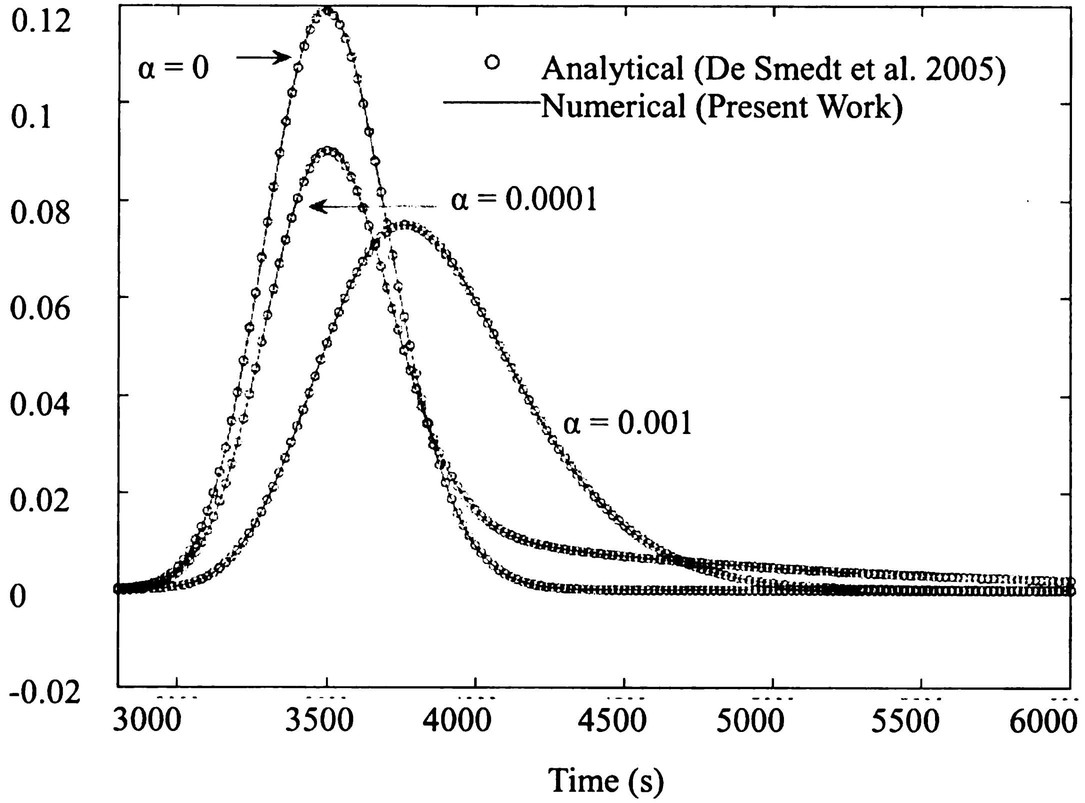


Figure 4.3. Comparison of the numerical solution with the analytical solution of De Smedt et al [2005]

4.3.2. Multi-resolution Wavelet Decomposition of ADCP Data

To examine the estimated parameters and to relate them to the physical characteristics of the river, we used three-dimensional velocity data obtained from ADCP surveys. Given the 3-D velocity field in a river, the dispersion coefficient D can be computed by numerically integrating the velocities [*Fisher et al.*, 1979]. However, we did not follow this approach as earlier studies found that dispersion estimates based on time-dependent velocity fields tend to be highly sensitive to velocity fluctuations [*Palancar et al.*, 2003]. In this paper we focus on the parameters AS and A. Repeated

ADCP surveys in the study reach clearly showed regions of high and low velocities and acoustic backscatter (a well-known measure of suspended solids concentration, SSC) for different cross sections and led us to test the hypothesis that in-stream TS zones can be identified using the three-dimensional velocity fields obtained from ADCP surveys. Earlier studies [*Sukhodolov et al.*, 2004; *Tipping et al.*, 1993] showed that the concentration of suspended particulate matter reduces in the dead zones due to sedimentation of faster sinking fractions of suspended matter in the decelerating flow. Therefore, by identifying regions of decelerating flow or low SSC using ADCP data we may be able to identify the relative importance of dead zones in a river reach. [*Engelhardt et al.*, 2004] noted a correspondence between SSC and mean velocity vectors and this correspondence was also evident in our ADCP data. To identify regions of relatively fast and slow moving water,we use multiresolution wavelet analysis of the twodimensional (y, z) normalized mean velocity fields obtained from ADCP surveys. The continuous wavelet transform (CWT) of a function f(y,z) for a two-dimensional wavelet is defined as the convolution with a scaled and shifted version of the wavelet function Ψ

$$\begin{gathered} W(\mathbf{a},\mathbf{b}) = \frac{1}{\sqrt{a_1 a_2}} \int_{\Re} f(y,z)\Psi\left(\frac{y-b_1}{a_1}, \frac{z-b_2}{a_2}\right) dy dz \\ \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \end{gathered} \qquad (4.8)$$

where a and b are the scale and translation vectors. We used the two-dimensional discrete wavelet transform (DWT) for our analysis. Similar to the 1D wavelet transform, the 2-D transform can decompose a given function into its slow changing (or coarse) features (called approximations) and fine (or rapidly changing) features (called details). For the two-dimensional case, the details can be further decomposed into horizontal, vertical and diagonal details. If φ^y, φ^z denote the scaling functions and ψ^y, ψ^z the wavelet functions for the one-dimensional representation in the y and z directions respectively, then multidimensional wavelet bases can be constructed as the tensor products of the one-dimensional wavelet bases as shown below.

$$\Psi^{\alpha\beta}{}_{m,i,j} = \begin{cases} \psi^y{}_{m,i}\,\psi^z{}_{m,j}, & \alpha = H, \beta = H \\ \varphi^y{}_{m,i}\,\psi^z{}_{m,j}, & \alpha = L, \beta = H \\ \psi^y{}_{m,i}\,\varphi^z{}_{m,j}, & \alpha = H, \beta = L \\ \varphi^y{}_{m,i}\,\varphi^z{}_{m,j}, & \alpha = L, \beta = L \end{cases} \tag{4.9}$$

where *m* denotes the wavelet scale level and *i* and *j* denote the rows and columns of the coefficient matrices. Here *L* and *H* denote the low- and high-frequency content (coarse and fine features) corresponding to the different filter properties of the scaling and wavelet functions in the spectral space. For the function $f(y,z)$, the wavelet $\Psi_{m,i,j}{}^{LL}(y,z)$ captures its coarse features at level *m*, while the remaining three wavelets $\Psi_{m,i,j}{}^{HL}(y,z), \Psi_{m,i,j}{}^{LH}(y,z), \Psi_{m,i,j}{}^{HH}(y,z)$ contain the vertical, horizontal, and diagonal details respectively. The function can be represented as the sum of its approximations and details using the wavelet coefficient matrices as shown below.

$$\begin{aligned} f(y,z) = {} & \sum_{ij} {W^A}_{N,i,j} {\Psi^{LL}}_{N,i,j}(y,z) + \sum_{m=1}^{N} \sum_{ij} {W^V}_{N,i,j} {\Psi^{HL}}_{N,i,j}(y,z) \\ & + \sum_{m=1}^{N} \sum_{ij} {W^H}_{N,i,j} {\Psi^{LH}}_{N,i,j}(y,z) + \sum_{m=1}^{N} \sum_{ij} {W^D}_{N,i,j} {\Psi^{HH}}_{N,i,j}(y,z) \end{aligned} \tag{4.10}$$

where

$$\begin{aligned} {W^A}_{N,i,j} &= \iint {\Psi^{LL}}_{N,i,j}(y,z) f(y,z)\,dydz \\ {W^V}_{N,i,j} &= \iint {\Psi^{HL}}_{N,i,j}(y,z) f(y,z)\,dydz \\ {W^H}_{N,i,j} &= \iint {\Psi^{LH}}_{N,i,j}(y,z) f(y,z)\,dydz \\ {W^D}_{N,i,j} &= \iint {\Psi^{HH}}_{N,i,j}(y,z) f(y,z)\,dydz \end{aligned}$$

Here the W parameters are the wavelet coefficient matrices, N denotes the number of levels in the decomposition, and the superscripts A, V, H and D denote the approximations (the first term in equation (4.10)) and the vertical, horizontal and diagonal details respectively (the last three terms in equation (4.10)). The decomposition (4.10) allows us to examine the different components at multiple levels or scales. Our aim was to extract the coarse features (the first term in equation (4.10)) at different levels as they retain the essential features of the velocity field (the high-frequency content simply adds detail). Since our primary interest was in making a distinction between regions of slow moving or stagnant water (dead zones) and the main channel (two distinct scales), we used two-level decomposition (N = 2). After plotting the single-level reconstructions based on the approximations at levels 1 and 2 in the physical space, (A_S/A) was computed from the area of all the pixels greater than a threshold value T (corresponding to the background pixel value in the two images):

$$\begin{aligned}
\frac{A_S}{A} &\approx \frac{\iint f_1(y,z)dydz}{\iint f_2(y,z)dydz} \\
f_1(y,z) &= \begin{cases} 1 \ \text{ iff } \ {W^A}_{2,i,j}{\Psi^{LL}}_{2,i,j} > T \\ 0 \qquad\qquad \text{otherwise} \end{cases} \\
f_2(y,z) &= \begin{cases} 1 \ \text{ iff } \ {W^A}_{1,i,j}{\Psi^{LL}}_{1,i,j} > T \\ 0 \qquad\qquad \text{otherwise} \end{cases}
\end{aligned} \tag{4.11}$$

The Haar and Daubechies-12 wavelets [*Daubechies*, 1988] were used in our analysis; however, other wavelets produced essentially similar results. After performing the wavelet decomposition and extracting the terms for N = 1 and N = 2 and plotting them in the physical space (i.e., as a function of y and z to identify the channel and the relative locations of the dead zones within the channel), functions in the MATLAB wavelet processing toolbox were used to compute the areas and the ratio in equation (4.11).

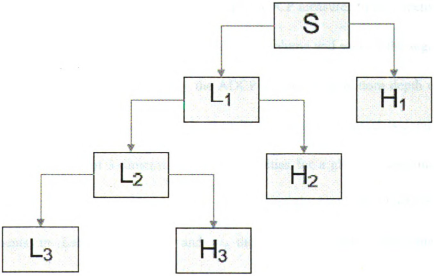


Figure 4.4. Schematic diagram illustrating the concept of discrete wavelet

decomposition. S denotes the original signal (the 2D image of the velocity field measured by ADCP). L and H denote the low-frequency approximations and high-frequency details, respectively. Suffixes denote wavelet scale levels

4.3.3. Estimating the Longitudinal Dispersion Coefficient

Two ADCPs manufactured by Teledyne - RD Instruments, Poway, California (1200 or 600 kHz Workhorse Rio Grande) were used for this study. Data were collected by mounting the ADCP on a trimaran (an OceanScience Riverboat™ with housing for electronics and radio modems to communicate with a land-based laptop) and towing the vessel across the river perpendicular to the direction of flow either from a bridge or behind a small motor boat. During an ADCP ping, the boat will travel a certain distance along the cross-sectional transect and the corresponding water column is called an ensemble. The width of this ensemble depends on the ping rate and the boat velocity (typically 0.2 - 0.5m/s in this work). The ADCP measures three-dimensional water velocities from vertical segments of the water column and each of the segments is referred to as a bin. Simultaneously, the ADCP measures the bottom depth of the river and the boat velocity relative to the river bed. After each transect is completed we obtain a 2D field of 3-dimensional water velocities for a given x (longitudinal) location $\vec{v}(y,z) = (v_N, v_E, v_Z)$ where v_N and v_E are the North and East velocity components in Earth coordinates and v_Z the vertical velocity component at location (y,z) where y and z denote the transverse and vertical coordinates respectively. The measurement of $\vec{v}(y,z)$ by ADCP is equivalent to densely deploying velocimeters throughout the cross-sectional area. The ADCP data is processed carefully after collection including smoothing and removal of bad ensembles. General principles of ADCP operation and post-processing for moving-vessel measurements are described in Simpson [*Simpson*, 2001] and Muste et al. [*Muste et al.*, 2004]. Apart from bad ensembles, some bins in a good ensemble may also occasionally report bad velocity readings (e.g., when data do not meet the echo intensity, correlation or other thresholds or when readings from the four beams differ significantly). In the final datasets, the bad ensembles and bad bins are replaced by using nearest neighbor interpolation.

Noise is almost always present in the ADCP data due to the transient turbulent nature of the flow. In order to reduce the effect of random noise on the dispersion estimates, we have considered smoothing the vertical velocity profile before evaluating Eq. (4.3) to produce a mean velocity field in the river. Muste [*Muste et al.*, 2004] discussed several smoothing techniques including fitting a power law or logarithmic profiles to the data in an ensemble:

$$\frac{u}{\overline{u}} = a\left(\frac{z}{z'}\right)^m \tag{4.12}$$

where z is the vertical coordinate measured from the bottom, z' is the location in the boundary layer at which $u = 0$, k is the von Karman constant, ν is the kinematic viscosity of water and a, m, B are fitting parameters. We will assess the effect of using these two smoothing formulations on the dispersion estimates in a later section. In addition to profile smoothing, we have considered two alternative approaches for orienting the velocities before evaluating Eq. (4.3). In the first approach, which is used in [*Carr and Rehmann*, 2007], the water velocities are projected onto the main direction of river flow (streamwise direction). The main direction of the flow is calculated as:

$$\vec{V} = \sum_{i=1}^{n} \sum_{j=tb(i)}^{db(i)} \vec{v}(i,j) \tag{4.13}$$

where $\vec{V} = \left(\mathrm{V_N}, \mathrm{V_E}\right)$ is the cross-sectional average velocity vector, $tb\left(i\right)$ and $db\left(i\right)$ are, respectively, the top good bin and bottom good bin of the ensemble i. The velocities are then projected as:

$$u = \frac{\vec{v} \cdot \vec{V}}{\left|\vec{V}\right|^2} \vec{V} \tag{4.14}$$

and the width of the ensemble, Δy, is similarly projected to the diagonal direction to $\vec{V}$. In the second approach, u is simply projected to the normal direction of the transect track, $\vec{n}$. This can be evaluated similar to the way the ADCP evaluates discharge:

$$u = \vec{v} \cdot \vec{n} = \frac{q}{a} = \frac{(\vec{v} \otimes \vec{v}_b) \cdot \vec{k} \Delta t}{\Delta y} \tag{4.15}$$

Here $\vec{n}$ is the unit vector normal to the transect, $\vec{k}$ is the unit vector in the vertical direction, q is the fractional discharge, $\vec{v}_b$ is the boat velocity vector, a is the fractional area, Δt and Δy are the elapsed time and distance for the ensemble. One potential advantage of the second formulation is that it can be used when the channel width is changing along the stream. However, its actual performance will be evaluated in the next section. It is clear from equation (4.3) that the success of the ADCP method depends on accurate approximation of the transverse dispersion coefficient D_y. There is no general consensus about the estimation of D_y but an approximate average based on experimental results is given by [*Fischer*, 1979]:

$$D_y \approx C'u^* d \tag{4.16}$$

where C' is a constant, normally taken as 0.145, d is the depth of the channel and u^* is the shear velocity, which is computed as $u^* = \sqrt{gRS}$ where g is the acceleration due to gravity, R is the hydraulic radius and S is the channel slope. The other formula for D_y that was previously used in [*Deng et al.*, 2001] and [*Perucca et al.*, 2009] is:

$$D_y = \theta u^* h(y) \tag{4.17}$$

$$\begin{aligned} \theta &= \left[0.145 + \frac{1}{3520}\left(\frac{u}{U^*}\right)\left(\frac{B}{H}\right)^{1.38}\right] \\ U^* &= \int_0^B u^*(y)\,dy \\ H &= \int_0^B h(y)\,dy \end{aligned} \tag{4.18}$$

where H is the cross-sectionally averaged depth and $h(y)$ is the local depth. After D_y is computed from either (4.16) or (4.17), Eq. (4.3) can be numerically integrated. Repeated ADCP transects have been collected at various locations under different flow conditions in order to assess the variability in the ADCP method.

4.4. Results

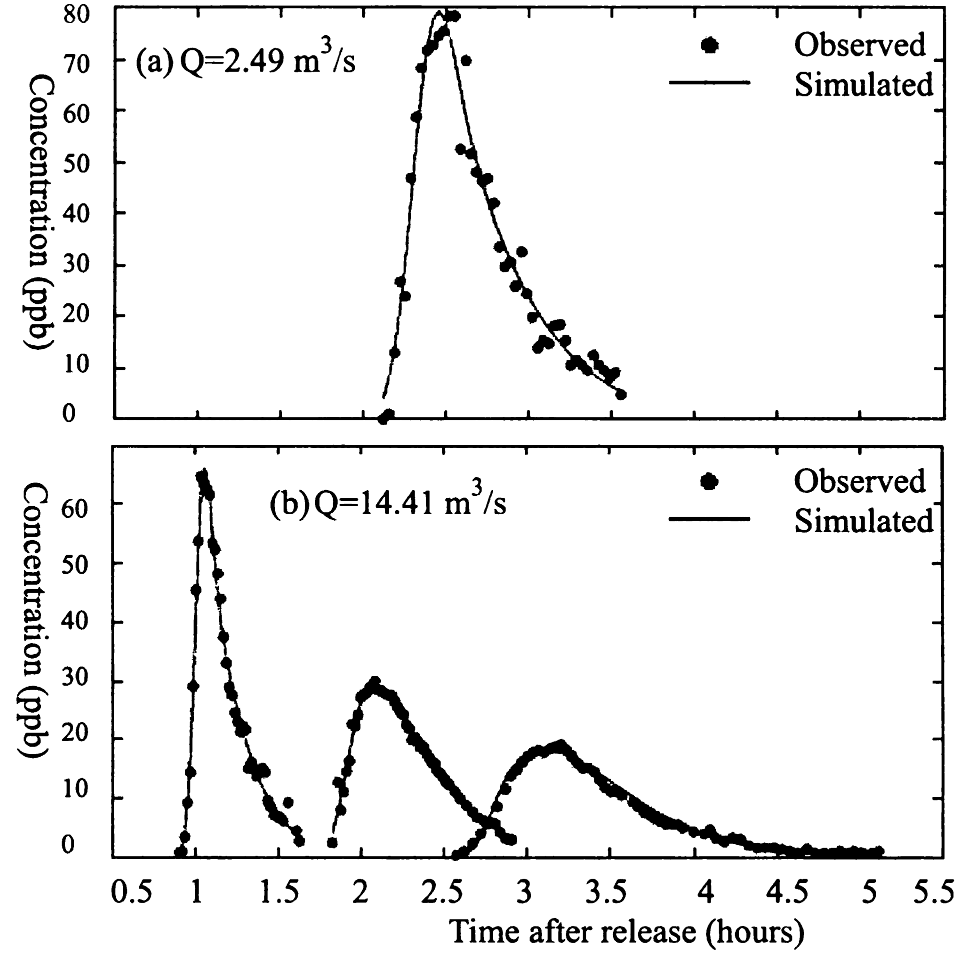
4.4.4. Evaluation of Channel Features and Potential for Hyporheic Exchange in the RCR

To estimate the TS parameters, the 5 km study reach was divided into three test reaches as shown in Figures 4.1 and 4.2. Model parameters were estimated for all three reaches for different flow rates (tracer test dates). Comparison of observed and simulated tracer concentrations is shown in Figure 4.5. Surficial sediments in the study reach consisted of a thin (5 cm or less) layer of sand and gravel underlain by a heavily consolidated clay layer of depth 0.25 m or more [*Uzarski et al.*, 2004]. Thus there was limited potential for hyporheic exchange in the study reach although there were differences between the test reaches. Reach A (between the Hagadorn and Farm Lane Bridges), a relatively straight section of the river, was free of alluvium and surface storage was the primary mechanism contributing to TS in the reach. This reach had extensive vegetation growing near the banks and dead trees within the channel, particularly near the Hagadorn Bridge. Reach B (between the Farm Lane and Kellogg Bridges) was characterized by the presence of a large meander and undifferentiated sand units (mostly fine sand and silt) near the left bank over much of its length. Previous studies indicated that meandering contributes to surface storage by increasing eddies and pools in the reach. In addition, reach B has a weir located near the Library Bridge. The backwaters of this weir extended upstream and created an impoundment resulting in enhanced surface storage. The sand and silt unit hugs the left bank in Reach B and extends all the way into reach C (between Kellogg and Kalamazoo Bridges) and provides the opportunity for enhanced exchange of solutes between the banks and the main channel compared to reach A [*Puzio and Larson*, 1982]. Thus reach B had the potential for both surface storage and hyporheic exchange. In reach C, the river comes down following a north-south course and runs into the Mason esker [*Leverett and Taylor*, 1915], a well-defined ridge of sand and gravel oriented in the south-north direction. The grain size of the sediments decreases radially outward from the centerline of the esker. Since the adjacent regions are mainly loam and other finer material, the gravel acts as a ledge and the river adjusts its gradient producing a meandering channel. Reach C therefore has wide floodplains, exhibits extensive meandering and is marked by the presence of wetlands and swamps near the edge of the river. Meandering of the channel changed the floodplain alluvium and created high-porosity sand and gravel deposits that provide conditions suitable for hyporheic exchange. Variations in velocities and channel widths between the reaches can be used to gain insight into how dispersion changes along the river. Hydrodynamic modeling indicated that although there was a significant variability in the velocities and heads as a function of distance, each subreach can be considered a fair approximation of a channel representing uniform hydraulic properties. Examination of the spatial variations in flows and velocities for the first three dye studies showed that there was a significant difference in the average velocities in reaches A and B but velocities in reaches B and C had similar values. For example, for Q = 16.82 m3/s, the mean velocities for the three reaches were 0.44, 0.75 and 0.71 m/s respectively.

4.4.5. Evaluation of TS Model Parameters for RCR

The estimated parameters of the TS model for the RCR are summarized in Table 6.1. Examination of the estimated parameters showed that, for all the dye studies (with the exception of 1-C in Table 1), the size of the TS zone increased in the downstream direction. This is in agreement with descriptions of channel features and surficial geology presented in the previous section. The size of the TS zones also increased with discharge, from 1.88 m2 for a low flow of 2.49 m3/s to 7.28 m2 for Q = 19.06 m3/s. This is in contrast to the results of [*Morrice et al.*, 1997], who found that the size of the TS zone decreased with increasing discharge in a first-order mountain stream. In the RCR, stream cross-sectional area (A) increased with discharge, an observation also made by [*Morrice et al.*, 1997]; however, at higher discharges the adjacent low-lying areas near the banks were filled with relatively stagnant water which provided additional surface storage that was not available at low discharges. Results from our ADCP surveys (presented in the next section) support this explanation.

Figure 4.5. Comparison of observed and simulated tracer concentrations for four slug releases conducted during summer 2002. (a) Q=2.49 m^3/s. Sampling locations at x=0.87km (Bogue Street Bridge) (b) Q = 14.41 m^3/s, sampling locations are, namely, Farm Lane Bridge (x = 1.40km). Kellogg Bridge (x = 3.10km) and the Kalamazoo Bridge (x=5.08km), respectively. (c) Q = 16.82km^3/s, (d) Q= 19.06m^3/s



(Figure 4.5 contd.)

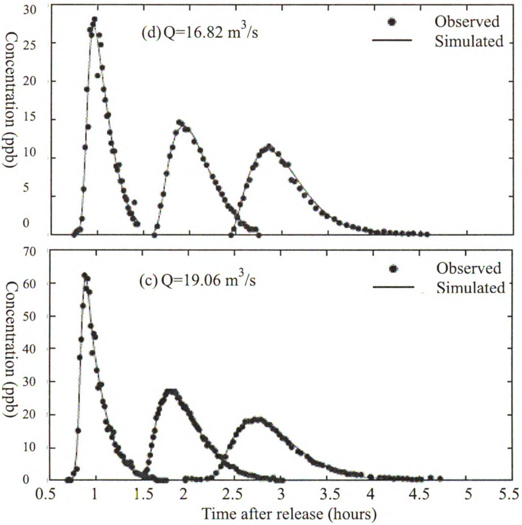


Table 4.1. Parameters in the Transient Storage model estimated for four different flow rates

Dye Study-Reach	Reach Length (km)	Q (m3/s)	A (m2)	As (m2)	D (m2/s)	α (/s)	As/A (-)	Q/A (m/s)	RMSE (-)	T (s)	Fmed (-)	Fmed200 (-)
1-A	1.40	16.82	39.93	6.38	0.75	6.26E-04	0.16	2.64	1.59	1598	3.89	0.64
1-B	1.70	16.82	33.73	6.70	0.79	6.58E-04	0.20	2.51	0.36	1521	5.95	0.85
1-C	1.98	16.82	30.48	5.48	0.79	4.75E-04	0.18	3.07	0.45	2106	4.02	0.46
2-A	1.40	14.41	38.69	5.38	0.58	3.54E-04	0.14	2.68	2.00	2824	2.06	0.32
2-B	1.70	14.41	33.05	5.28	1.45	3.24E-04	0.16	2.73	0.98	3084	2.52	0.32
2-C	1.98	14.41	28.94	5.66	1.45	3.22E-04	0.20	2.54	0.80	3103	3.63	0.41
3-A	1.40	19.06	42.52	6.62	0.69	5.20E-04	0.16	2.88	2.10	1925	3.01	0.48
3-B	1.70	19.06	36.36	7.28	1.19	5.50E-04	0.20	2.62	0.39	1819	5.01	0.69
3-C	1.98	19.06	32.19	7.22	1.19	5.09E-04	0.22	2.64	0.50	1964	5.82	0.69
4-D	0.865	2.49	15.47	1.88	0.31	1.54E-04	0.12	0.16	4.41	6489	6.11	1.89

The rate of exchange a between the main channel and the storage zones increased with discharge Q. A positive relation between a and Q was also noted by [*D'Angelo et al.*, 1993] who attributed it to an increased availability of solute per unit time. We calculated the TS zone residence times as (tS = (AS/A)/a) [*Harvey et al.*, 1996]. Residence times ranged from 255 seconds (a = 6.26 x 10-4s-1) for Q = 16.82 m3/s to 790 seconds (a = 1.5 x $10^{-4}s^{-1}$) for Q = 2.49 m3/s for the RCR. For all dye studies, residence times increased with distance in the downstream direction and generally decreased with discharge Q. We attribute the increase in tS with downstream distance to the presence of sand and gravel deposits in the last two reaches compared to reach A. Our estimated a values are in the same range as the values reported by [*Salehin et al.*, 2003] for the vegetated reach of an agricultural stream in Sweden (a = 6.1 x $10^{-4}s^{-1}$). [*Harvey and Ryan*, 2004] also reported similar range of values (a = 4.7 x 10^{-4} and 5.6 x $10^{-4}s^{-1}$) for a heavily vegetated stream in Arizona. The relatively high tS values determined for reach C are attributed to the alluvium storage and the sediment characteristics (gravel and coarse sand) in this reach. Results from our ADCP surveys (and wavelet decomposition) showed that surface storage in this reach was relatively small indicating that hyporheic exchange was the primary mechanism that contributed to TS in this reach.

The cumulative effect of TS on downstream transport and reach-scale retention of water depends on the parameters AS, a and the flow velocity in the main channel. It is well-known that efforts to interpret TS model parameters often lead to misleading conclusions about the relative importance of TS processes compared to other processes [*Runkel*, 2002] as existing metrics such as the storage zone residence time do not describe the overall effect of the TS parameters described above. [*Runkel*, 2002] proposed the use of a new metric (Fmed) which is the fraction (expressed as percent) of the median reach traveltime that is due to TS.

$$F_{med} = \left[1 - e^{-L\alpha/u}\right] \times \frac{A_s}{A + A_s} \tag{4.19}$$

Stream reaches that substantially influence the downstream transport of solute mass due to TS will have higher values of Fmed and vice versa. Since reach lengths (L) vary significantly in different studies, a Fmed value obtained using a standard reach length of 200 m (Fmed 200) was proposed as a metric to facilitate direct comparison with other streams. Fmed and Fmed 200 values for the Red Cedar River indicate that the importance of TS increases in the downstream direction (4.1). Comparison of the Fmed 200 values with estimates for other streams [*Runkel*, 2002] showed that values for the RCR were at the lower end of the range and were comparable to those for the Snake River, an acidic and metalrich mountain stream in Colorado [*Bencala et al.*, 1990]. Although the RCR is a bigger stream in comparison, the limited potential for hyporheic exchange at this site (due to the consolidated clay layer) was noted by other researchers [*Uzarski et al.*, 2004]. The estimated (AS/A) values and exchange coefficients (a) in several subreaches are comparable for the two streams.

4.4.6. Results from ADCP Surveys and Wavelet Analysis

Figure 4.6 shows the observed bathymetry, channel cross sections and the mean velocity fields at the Hagadorn and Farm Lane Bridges for two different flow rates, Q = 5.49 m^3/s (8 November 2003) and Q = 19.89 m^3/s (19 March 2006). On both days, the flow near the Hagadorn Bridge was highly nonuniform and became relatively uniform with distance in the downstream direction. In addition, channel cross section was W shaped at the Hagadorn Bridge as opposed to the U-shaped cross sections (which favor uniform conditions) at the other bridges. As discharge increased, the river became wider and was marked by the presence of relatively stagnant water near the banks. The relative extent of the low-velocity or stagnant zones decreased in the downstream direction as the flow became more uniform between Hagadorn and Farm Lane Bridges. Since the tracer data was used to estimate reach-averaged values for AS/A, ADCP data collected at multiple stations within a reach can be averaged to compare with the tracer results. If in-channel processes were primarily responsible for storage within a reach, then we expect AS/A estimates from ADCP and tracer data to agree. On the other hand, stream reaches dominated by hyporheic exchange are expected to produce widely different estimates of AS/A from ADCP and tracer data. The approximations to the original velocity fields based on two-level decomposition allowed us to identify TS locations with the channel (4.7). The images marked L1 and L2 show the first term in equation (4.3) for level 1 and level 2 decompositions respectively. The relative locations of the dead zones within the channel given by the wavelet decomposition (L2 approximation in Figure 11) agreed with our observations of relatively stagnant water during our field work. The ratios of the areas (AS/A) calculated based on this decomposition are shown in Table 2 for reach A for two discharge values (2.0 and 19.8 m3/s).

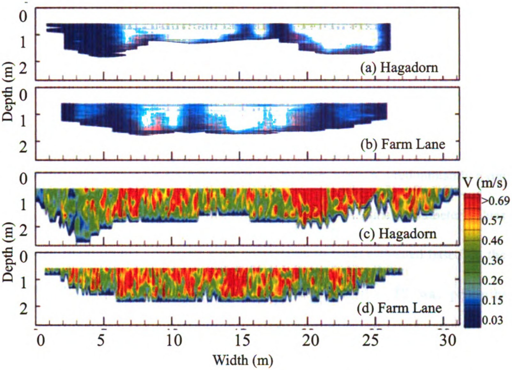


Figure 4.6. Observed mean velocity fields at two different stations in reach A in the Red Cedar River obtained using a 1200 kHz ADCP (a and b) Q = 5.49m^3/s (8 November 2003). (c and d) Q = 19.89 m^3/s (19 March 2006). Note that during the high event the adjacent low lying areas near the banks were filled with relatively stagnant water which were not available during low discharge

The average values obtained from the ADCP data were based on four transects and were found to be in good agreement with results from tracer data. Using more transects will likely improve the ADCP estimates but data from other transects (e.g., at the Bogue Street Bridge) showed that conditions were similar to those at the Farm Lane Bridge. Since our primary focus was on ascertaining whether the sizes of the TS zones independently estimated using the ADCP and tracer data were of a comparable magnitude, we believe that the estimates in Table 4.2 are adequate. In first-order tropical headwater streams differing in channel morphology and hydraulic characteristics, [*Gucker and Boechat*, 2004] compared the sizes of TS zones for different stream morphotypes including straight run, meandering, step-pool, and swamp reaches. They concluded that their (AS/A) estimates were lowest for straight run reaches and highest for swamp reaches. In our case, reach A, predominantly a run reach, had the lowest (AS/A) for all the discharge values. Since parameter values estimated based on the TS model were comparable to those estimated based on the ADCP data for similar discharge values, we conclude that TS was primarily controlled by in-stream processes in reach A. We could not obtain (AS/A) estimates from the ADCP data for reaches B and C as the stream was too shallow to operate our 1200 kHz instrument and obtain good transect data. Since (AS/A) values were highest in reach C (which was consistent with our observations, e.g., the presence of swamps, a meandering channel, wide floodplain), we wanted to test the hypothesis that hyporheic exchange was important in this reach. If this was indeed true, then we expect the (AS/A) estimates from ADCP data to be relatively small compared to the estimates from tracer data. We were successful in obtaining several good transects at the Kalamazoo Bridge (our last sampling point) on 15 October 2006 (Q = 3.35 m3/s). Data from one such transect is shown in Figure 4.5.

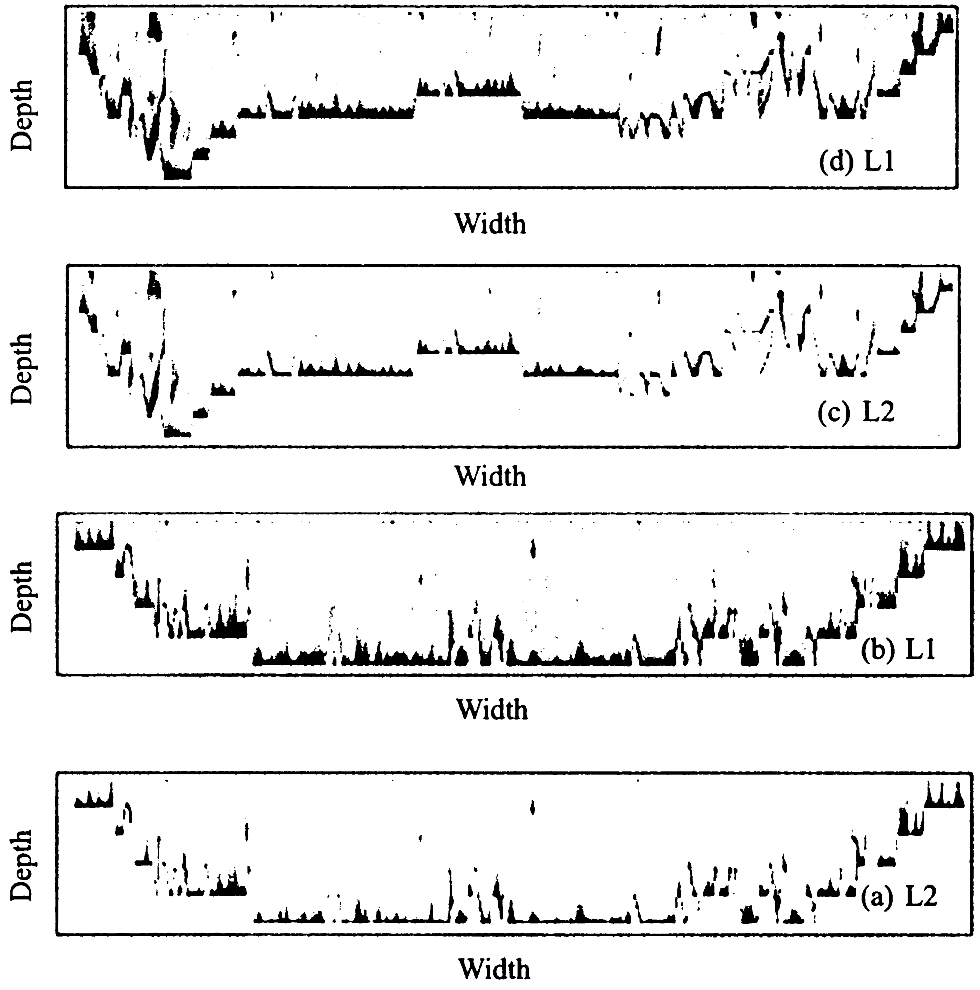


Figure 4.7. Multiresolution wavelet approximations for the images shown in Figure 4.6. After completing the wavelet analysis, the low-frequency content at wavelet levels 1 and 2 (denoted by L1 and L2) was plotted in the physical space. (a and b) Hagadorn Bridge. (c and d) Farm Lane Bridge.

The channel was relatively narrow at this station and showed cross-sectional uniformity in velocity which was an indication that surface storage may be relatively unimportant. Average value obtained from wavelet analysis based on three transects at this site gave (AS/A) = 0.05. Although we do not have estimates from tracer data to compare with this value, our fourth dye study was conducted under similar low-flow conditions (Q = 2.49 m3/s). We obtained (AS/A) = 0.12 for reach D between the Hagadorn and Bogue Street Bridges in Table 1. All other dye studies showed that (AS/A) increased in the downstream direction and with discharge. It is therefore highly probable that the (AS/A) value for reach C on 15 October 2006 was significantly higher compared to the number 0.05 estimated from the ADCP data. An important parameter in assessing the role of storage zones is the Damko¨hler index (Da) calculated from observed data. The parameter Da reflects the relative importance of downstream processes in relation to TS and is computed as the ratio of the time needed for the downstream tracer transport for a certain reach length to the mean tracer residence time in the storage zones [*Harvey and Wagner*, 2000]. [*Schmid*, 2004] analyzed slug release data and concluded that very close or nearly identical results are obtained by the AD model and TS model if Da < 0.6 or Da > 60.0. In our case, the calculated Da based on the estimated parameters of the TS model ranged between 1.5 and 7.6 for all cases, which confirmed the preferred use of TS model over the AD model.

Table 4.2. Comparison of the Relative Sizes of the Transient storage zones estimated using tracer data and independently using the ADCP data for Reach A

Q Tracer (m^3/s)	Q ADCP (m^3/s)	As/A	As/A	Comments
19.06	19.8	0.16	0.16	Farm Lane-1
19.06	19.8	0.16	0.12	Farm Lane-2
19.06	19.8	0.16	0.19	Hagadorn-1
19.06	19.8	0.16	0.21	Hagadorn-2
19.06	19.8	0.16	0.17	Average Value*
2.49	2	0.12	0.11	Farm Lane-1
2.49	2	0.12	0.14	Farm Lane-2
2.49	2	0.12	0.11	Hagadorn-1
2.49	2	0.12	0.17	Hagadorn-2
2.49	2	0.12	0.12	Hagadorn-3
2.49	2	0.12	0.13	Average Value*

By examining the low-frequency contributions (the coarse features) at successive levels in multilevel wavelet decomposition of ADCP data, we were able to identify the relatively stagnant regions in the flow field. The observed velocity fields contained both the mean and the highly oscillatory components of flow. By taking the average of several transects, we were able to quantify the relative magnitudes of surface storage and hyporheic exchange in different test reaches using wavelet analysis. The decomposed states are plotted in the physical space as shown in Figure 11 and image processing was used to estimate the ratio of the two cross-sectional areas in the images. In equation (4.10), the main channel cross-sectional area A can be computed from either the level 0 (i.e., the original image) or the level 1 approximation as shown in Figure 4.7 (they produced identical values for our data sets). The highfrequency subbands in the wavelet decomposition (e.g., the horizontal, vertical and diagonal details) were not shown as they did not contain information useful for our present analysis. In a study of transient storage and hyporheic flow along the Willamette River in Oregon, [*Fernald et al.*, 2001] indicated that they used a boat-mounted ADCP to measure discharges and main channel cross-sectional areas (A) at their sampling locations, although they did not show any comparisons between observed (ADCP) and estimated cross-sectional areas. In the present work, we did not directly compare the cross-sectional areas measured using the ADCP to the A estimated from our TS modeling for the following reason. Depending on the mode of operation and due to the time of delay required to transmit and receive acoustic signals, ADCP data usually have a "blank distance" close to the transducer in which velocity measurements are not available. In addition, there are difficulties in making measurements close to the banks. These limitations become more pronounced in shallow environments. Recent ADCP models specially designed for shallow environments may be more suitable for the type of applications described in this paper. Instead of directly using the cross-sectional areas (A) obtained from the ADCP, we decided to focus on the ratio of the areas (AS/A) as errors involved in the approximations of the areas may cancel out when ratios are involved. The assumption here is that the (AS/A) values estimated from ADCP measurements (and wavelet analysis) are representative of the entire cross section including the areas that could not be reached using an ADCP. For the data reported in this paper, we were able to make measurements close to the banks, therefore this assumption is unlikely to affect our results and conclusions but it is not hard to imagine situations where a significant fraction of storage zones are in shallow areas and remain inaccessible to an ADCP. The success of studies aimed at understanding functional relationships between nutrient uptake and storage area depends critically on our ability to separate surface storage from hyporheic exchange [*Runkel et al.*, 2003; *Salehin et al.*, 2003]. This paper presents one approach for achieving this separation. We demonstrated that in one of our test reaches (reach A), the tracer data and the ADCP estimates of (AS/A) were in good agreement for both high and low discharge values. Additional data sets and analyses are required to test the strengths and weaknesses of this approach. As noted by [*Shields et al.*, 2003] and [*Dinehart and Burau*, 2005], the study of river reaches using ADCPs is hampered by the lack of custom software for data analysis. For this study, we created software for extracting ADCP data, smoothing, correction, visualization, wavelet analysis, image processing etc. The availability of standardized software will make it easier to study river reaches using ADCPs on a routine basis.

4.4.7. Estimating Longitudinal Dispersion in Rivers

Before computing the dispersion coefficient using Eq. (4.3), we examined the ADCP datasets to identify potential issues that could lead to a violation of the assumptions involved in Eq. (4.3). We examined the depth-averaged velocity profiles in the transverse direction to identify recirculating or secondary flow regions and the number of bad ensembles as a per cent of the total number of ensembles. Datasets with a large percent of bad ensembles are not used for estimating the dispersion coefficient as explained later. Results from typical ADCP surveys are shown in Figure 2 in which the variable plotted is the mean velocity $\left|\vec{V}\right| = \sqrt{u^2 + v^2 + w^2}$ as a function of the depth and width of the channel. Channel characteristics in all stream reaches are such that the width-to-depth ratio $\left(B / H\right) > 10$, therefore equation (4.16) can be expected to provide a reasonable approximation of the transverse dispersion coefficient [*Deng et al.*, 2001]; however, unless otherwise stated, we have used equation (4.17) to compute D_y. The raw velocity data obtained from the ADCPs is conditioned by orienting (rotating) the velocities and smoothing the vertical profiles to obtain a consistent mean velocity field. Figure 3 shows the vertical velocity profiles in two large rivers (Ohio and St. Clair). Power law and logarithmic profiles are fitted to the raw ADCP data. We notice that the logarithmic relation describes the data better in large rivers such as the St. Clair River (especially closer to the bottom boundary layer) and that smoothing produces much better conditioned data for further analysis. For evaluating the dispersion coefficient, however, we find that the two profiles make little difference. The reason is that during the evaluation of Eq. (4.3) only the mean velocity in the water column is used (which is relatively insensitive to the smoothing technique used). As described earlier, we have examined two approaches for orienting the velocities (rotating the velocities in the streamwise direction and projecting them along a direction that is normal to the transect track). These two methods together with the three profile smoothing methods (power law, log law and no smoothing) yield a total of six cases. Assuming that the tracer estimate of the dispersion coefficient in a given reach represents a reasonable averaged measure of dispersion in that reach, the relative "error" in the ADCP estimate for individual transects within the same river reach is calculated for all six cases for the Grand River datasets and the results are displayed as box plots in Figure 4.10. Results indicate that the method of orienting the velocities has a relatively larger influence on the dispersion results than the method used for profile smoothing. For the Grand River datasets used to generate Figure , log law smoothing with velocity projection in the streamwise direction gave the best (closest to the tracer) results. For the relatively shallow rivers such as Burns Ditch (not shown in Figure 4.10) the power law smoothing gave slightly better results.

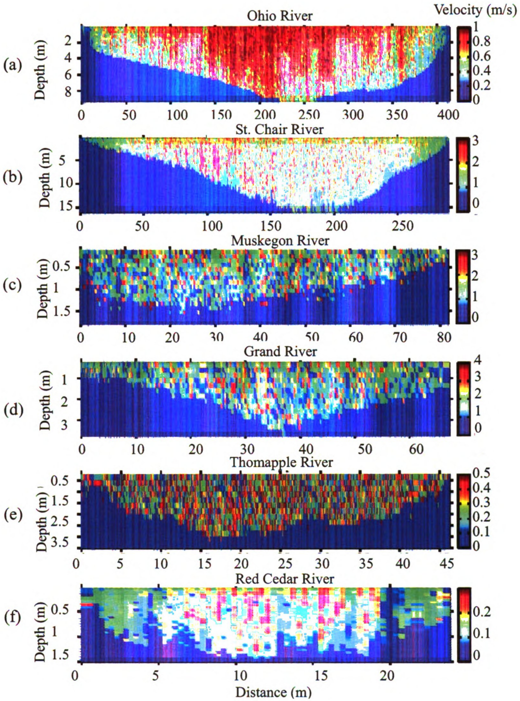


Figure 4.8. Sample ADCP transect data used for computing the longitudinal dispersion coefficients.

Since one of the objectives of this paper is to quantify the uncertainty in the dispersion estimates from the ADCP method by collecting repeated transect data at the same location (repeating this procedure for several different locations in a given reach), we plotted all the ADCP results against tracer data in Figure 5(a). The variability in the ADCP estimates within a given reach is shown using box plots and colors indicate different rivers. For the Red Cedar River, ADCP and tracer data are shown for reach A, a 1.4 km reach as described in [*Phanikumar et al.*, 2007]. Tracer data was collected for flow rates 19.06, 16.82, 14.41 and 2.49 m^3/s and tracer values of dispersion are marked in Figure 4.5 for all the four flow rates. Several ADCP datasets were collected within the same reach for flow rates 19.98, 3.6 and 4.7 m^3/s and the dispersion values are shown using box plots (The X-axis labels mark the locations of the box plots). For the Grand River, tracer and ADCP data are shown for reaches 2 and 3 as described in [*Shen et al.*, 2008]. The combined reach is approximately 23.8 km long which explains the larger variability in the ADCP estimates shown in Figure 4.5. Tracer values for Grand River plotted in the figure represent average values for reaches 2 and 3. For Ohio and St. Clair Rivers all the transect data reported in [*Holtschlag and Koschik*, 2003; *Koltun et al.*, 2006] are included to generate the box plots.

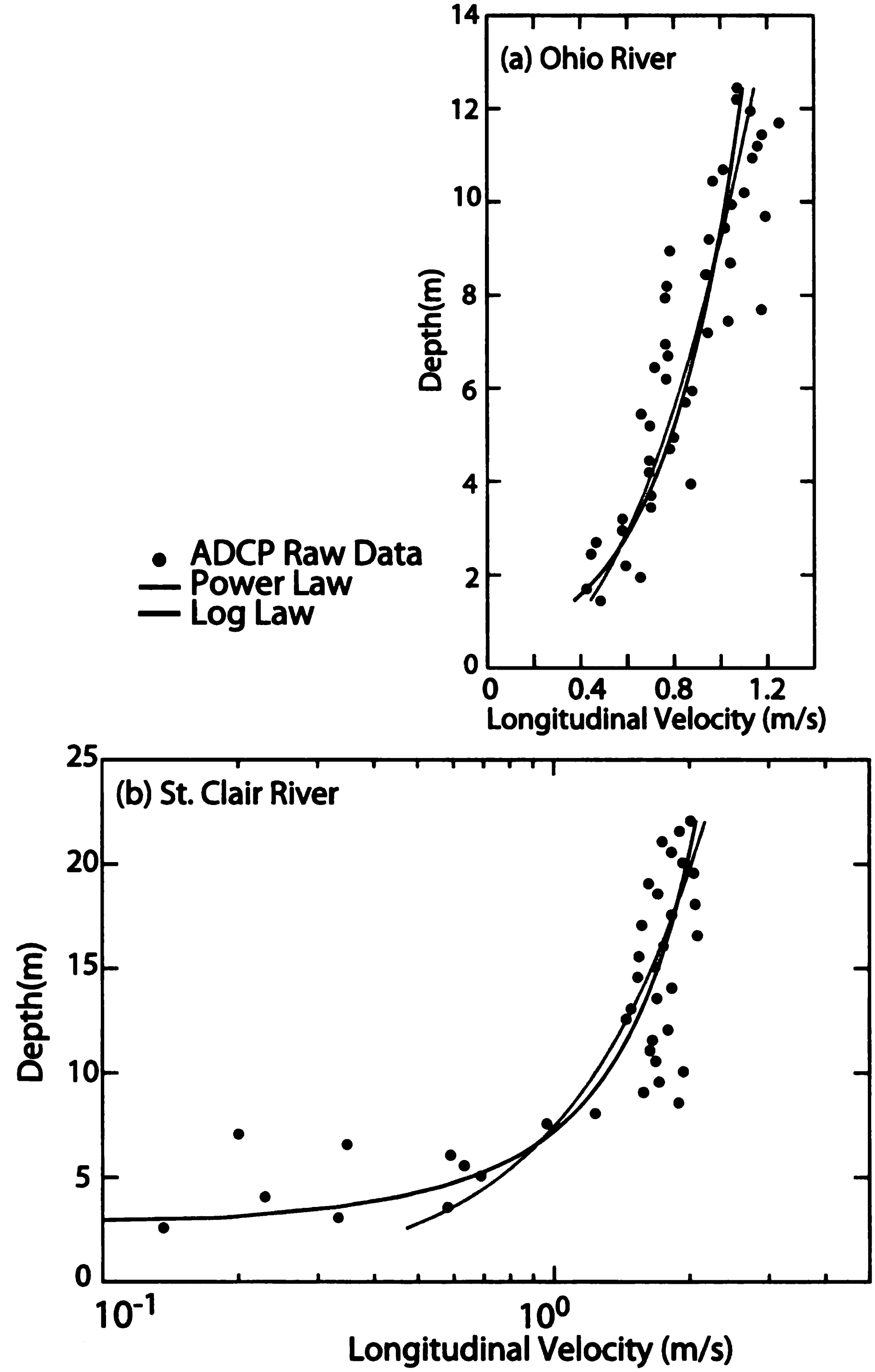


Figure 4.9. Vertical velocity profiles in the Ohio and St. Clair Rivers showing the effect of fitting a power law (black lines) and logarithmic profiles (red lines). The raw data from the ADCP is shown using symbols

Two different tracer estimates of dispersion are plotted in Figure 4.11a – in the first approach transient storage modeling (Eq. 4.2) is used to estimate the dispersion coefficient. In the second approach, the tracer data is fitted to an analytical solution of the ADE to obtain the dispersion coefficient as described in the earlier section. There are additional details associated with the tracer methods that are important to understand the comparison shown in Figure 4.11. Observed tracer data can be modeled in two different ways. In the first approach (method A), tracer mass is injected into the stream at $x = 0$ and parameters in the analytical or numerical solution are estimated by minimizing the deviation between simulated and observed concentrations at each of the downstream sampling stations. In this method parameters (e.g., D) estimated for reach 1 represent average conditions between the injection site and the first sampling location, however, parameters estimated for reach 2 represent conditions for *both* reaches 1 and 2 and so on. In the second approach (method B), upstream conditions in the form of concentration versus time data observed at the first sampling location are specified at the beginning of reach 2, therefore parameters estimated for reach 2 represent conditions in reach 2 only. For relatively small rivers and using the TS model Eq. (4.2), our experience indicates that the two approaches produce almost identical results [*Phanikumar et al.*, 2007]. However, this situation is different while using the ADE for large rivers. For this case, we found that methods A and B produce widely different results. This is not surprising since the ADE does not have a separate term for the dead zones. Therefore, the dispersion term in the ADE tends to capture the effects of the dead zones as well. As a result, if method A is used with the ADE in large rivers, the cumulative effects of dead zones as travel time increases could produce dispersion estimates that are unreasonably high compared to the local/point estimates from the ADCP. For the Grand River, for example, the TS estimates of the dispersion coefficient in the four reaches are 2.16, 1.6, 4.2 and 1.39 m^2/s respectively while the values based on the ADE using method A are 3.5, 27.72, 56.54 and 112.3 m^2/s respectively. Results obtained using method B in the same reaches (using the same initial mass) are 3.5, 13.05, 15.02, and 4.92 respectively. The ADE estimates shown in Figure 4.11 a are obtained using method B in which parameters estimated for a reach represent conditions only in that reach. We notice that the ADCP and tracer estimates are in good agreement as the flow rate changes over four orders of magnitude. In addition, the tracer estimate is closer to the median value of the dispersion coefficients obtained from the ADCP method indicating that it is beneficial to obtain multiple datasets at the same location. These results establish the ADCP method as a reliable alternative to the tracer method. From Figure 4.11a it appears that the difference between the TS and ADE estimates increases with flow indicating that dead zones play an important role at high flows. This was observed clearly for the Red Cedar River and the Grand River during our field studies (e.g., Figure 10 in [*Phanikumar et al.*, 2007]). At high flows, the low lying areas near the banks of the river are filled with stagnant water that contributed to additional storage. This additional storage is not available during low flow conditions. For low flows, the difference between the two models (TS and ADE) is not as high for the river sites considered in this paper and the dispersion coefficient from the TS model is in good agreement with the ADCP estimates. These comparisons indicate that the ADCP estimates of dispersion include the effects of dead zones as well. The ADCP measures velocities in both fast and slow moving regions of the river. There are no guidelines in the literature on what constitutes a dead zone (e.g., regions where velocities fall below a certain threshold value). In an earlier paper [*Phanikumar et al.*, 2007] we explored the idea of separating the flow field measured using an ADCP into relatively fast and slow moving regions using wavelet decomposition. We were successful in estimating the size of surface storage zones $\left(A_S / A\right)$ based on ADCP data and estimates compared favorably with results from a tracer-based method (TS modeling) for both high and low flows. These results support the fact that the dispersion coefficient estimated by the ADCP method includes contributions from dead zones.

The median dispersion values from repeated transect data shown in Figure 4.11a are plotted against the tracer values in Figure 4.11b. We also plotted values reported in the literature for comparison including data from Carr and Rehmann [*Carr and Rehmann*, 2007] and Fisher et al. [*Fischer*, 1979]. The lower end of the tracer values shown in Figure 4.11b come from either laboratory flume data reported by Fisher et al. [*Fischer*, 1979] or relatively smaller rivers such as the Burns Ditch or the Red Cedar River (present work). Results from individual datasets for all rivers are plotted in Figure

4.12(a) against some of the well known empirical relations available in the literature including the relations of Fisher et al. [*Fischer*, 1979], Seo and Cheong [*Seo and Cheong*, 1998] and Deng et al. [*Deng et al.*, 2001]. The ADCP method generally produces estimates that are comparable to the results from the empirical relations; however ADCP and tracer values are generally lower. The deviation $\left(\Delta D / D_{\mathrm{ADCP}}\right)$ where $\Delta D = \left(D_{\mathrm{ADCP}} - D_{\mathrm{Empirical}}\right)$ between the ADCP values and those from empirical relations is displayed using box plots in Figure 6(b) for all three empirical relations considered. We find that the Fisher et al. [*Fischer*, 1979] relation matches closely with our ADCP estimates. A Kruskal-Wallis one-way ANOVA test on ranks indicated that the three groups had statistically significant differences ($p \leq 0.001$) among their median values. Further analysis using Tukey's multiple pairwise comparison procedure indicated that the relation of Fischer et al. [*Fischer*, 1979] is responsible for the observed difference and that differences in the results based on the [*Seo and Cheong*, 1998] and Deng et al [*Deng et al.*, 2001] relations is not statistically significant.

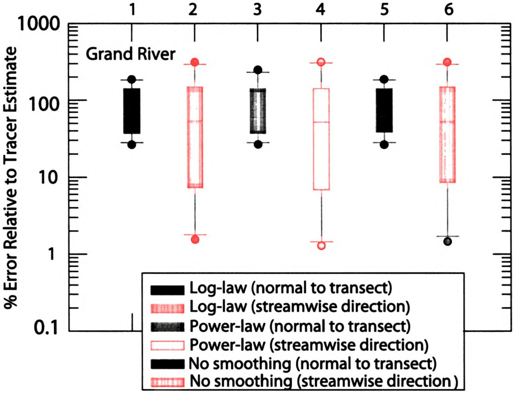


Figure 4.10. Effects of smoothing (logarithmic, power-law or no-smoothing) and velocity projection methods (rotating velocities in the streamwise direction or projecting them in a direction normal to the transect track) on the dispersion estimates from ADCP

Examination of our ADCP data indicated that there is a correlation between the quality of the dispersion estimates and the per cent bad ensembles in the transect data. Estimated dispersion numbers were found to be unrealistically high or low when the per cent bad ensembles exceeded about 12%. Bad ensembles can occur when communication is interrupted, when aquatic vegetation or large debris enters the field of the transducer beams or when a change in the ADCP operating conditions is warranted. A large number of bad ensembles could potentially influence the discharge measurement which can be a problem in itself. Figure 4.13 shows typical depth-averaged velocity profiles (Figures 4.13 a, b and c) and the computed discharge and the dispersion coefficient as a function of per cent bad ensembles within the same reach for Grand River. Symbols show the raw data and trend lines based on LOESS smoothing are also plotted (no attempt was made to fit the profile to satisfy the no-slip condition at the two banks). Examination of the velocity profiles can help isolate datasets that could potentially lead to a violation of the assumptions involved in Eq. (4.3).

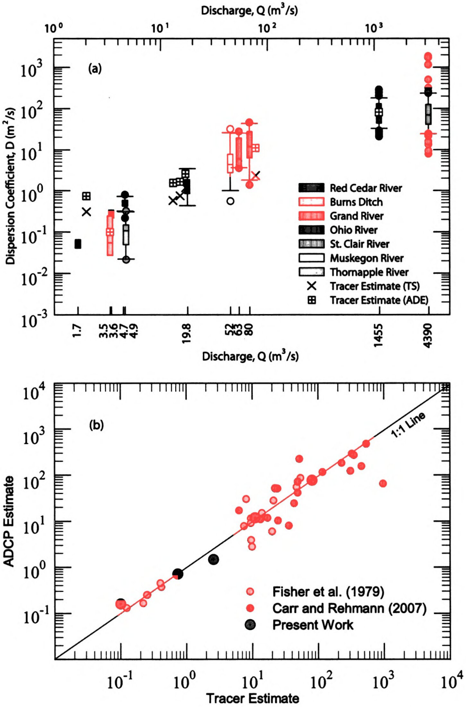


Figure 4.11. Comparisons between ADCP and tracer estimates of the dispersion coefficient: (a) Box plots denote the variability in D estimated using the ADCP method within a given river reach. Tracer estimates based on the ADE and the transient storage modeling are shown using different symbols. (b) Comparisons between ADCP and tracer estimates plotted on top of similar results reported in the literature.

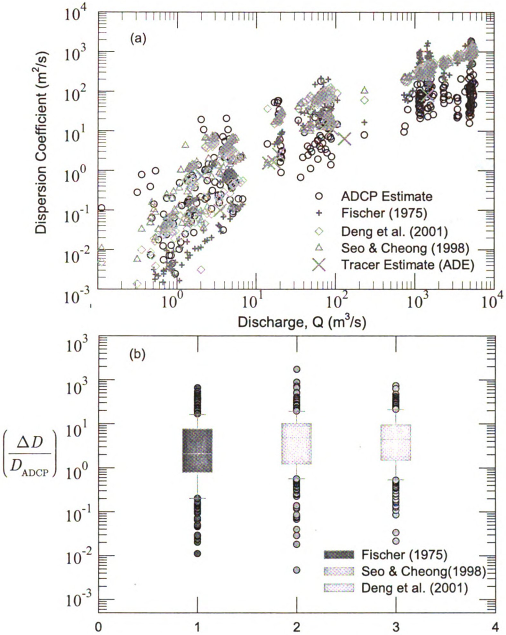


Figure 4.12. Comparison of ADCP estimates of the dispersion coefficient with tracer estimates and results from empirical relations. (b) Box plots showing the deviation between estimates of D based on the ADCP method and empirical relations

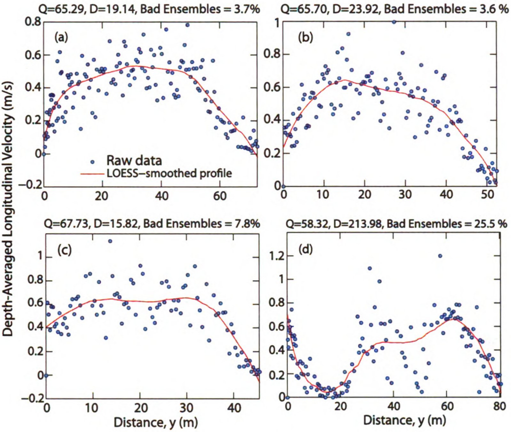


Figure 4.13. Depth-averaged longitudinal (u) velocity profiles plotted as a function of the transverse coordinate y .

4.5. Conclusions

As improved understanding of stream solute transport processes leads to better models and new approaches [*Boano et al.*, 2007; *Deng and Jung*, 2009], there is an imperative need for independent, field-based estimates of the dispersion coefficient to constrain models. The ADCP method of estimating the dispersion coefficient appears to be an excellent alternative to the tracer approach if care is taken to identify spurious data and repeated transects are used to estimate $\overline{D}$ (or another appropriate measure that represents average conditions within the stream reach). Our results indicate that a measure of D based on repeated transects is more reliable than individual estimates. For the river reaches in our work, the median value of the dispersion coefficients obtained from multiple datasets is found to be closer to the tracer estimate based on the ADE (using method B as described earlier). Our comparisons indicate that the ADCP method measures the influence of dead zones on the dispersion coefficient as does the estimate from the ADE. The ADCP method has its share of limitations including the inability to make measurements close to the banks and in shallow stream reaches. In addition, the method is not suitable for stream reaches dominated by meander bends, recirculating regions or secondary flows. Recent ADCP models (e.g., the Sontek s5 or M9) using smaller sensor heads, small (~ 5 cm) blanking distances and bin sizes and multiple acoustic frequencies (with features such as frequency hopping) have the potential to further improve our ability to estimate dispersion in streams and rivers. Additional datasets (including simultaneous tracer and ADCP data) and analyses are needed to further assess the relative strengths of this approach, especially for large rivers. The new ADCP models are also expected to improve estimates of surface storage zones by reducing the blanking distance.

In conclusion, we demonstrated how useful insights into stream solute transport processes can be obtained using high-resolution, three-dimensional hydrodynamic data obtained from ADCPs. Coupled with traditional tracer studies, these approaches have the potential to constrain TS modeling by providing independent estimates of the dispersion coefficient as well as surface storage in different reaches. The methods described in this chapter can be extended to obtain other types of useful information (e.g., residence times in different reaches, exchange rates with surface storage zones, lateral and vertical dispersion coefficients to name a few).

Appendix A . Supplemental Tables

Table A.1. List of input data to the watershed model and format

Data	Common Data Source	Format/Comments
DEM	Either DEM or NED from USGS	ASCII grid
LULC map	NLCD from US EPA or local agencies. Remotely sensed data	ASCII integer grid
LULC mapping table	Customized	A variable called lulc saved in .mat format that contains two fields: M for the transformation matrix as listed in Table A.2; G for the group information for each model RPT
Watershed Extent	Local agencies delineated watershed boundaries	ESRI shapefile: Polygon
River Hydrography Dataset	NHD from USGS	ESRI shapefile: Polyline Need to be processed. A river needs to be coded for its river ID. Currently, scattered observations of river width and type
Soils Type	SSURGO mapunit map	ASCII integer grid
SSURGO soils database	NRCS SSURGO database	A .mat file containing processed database information
Weather Station Locations	NCDC and other climatic data network	ESRI shapefile: Point
Climatic Data	NCDC and other climatic data network	NCDC downloaded format or MAWN downloaded format
Groundwater Hydraulic Conductivity	Local agencies or inferred from geological information	ASCII grid
Groundwater Aquifer Layer information	Local agencies or inferred from geological information	ASCII grid The thickness (or elevation) of each layer
Initial Groundwater Head	Local agencies or inferred from geological information	ASCII grid

Table A.2. An example transformation matrix from MDNR dataset to model classes

LU/ LC	Database definition	Water	Impervious	Pine	Oak	Shrub	Grass	Corn	Bare	Alfalfa
1	Low Intensity Urban		0.4		0.2		0.4			
2	High Intensity Urban		0.8		0.1		0.1			
3	Airport		1							
4	Road / Parking Lot		1							
5	Non-vegetated Farmland					0.2			0.8	
6	Row Crops							1		
7	Forage Crops									1
9	Orchards				1					
10	Herbaceous Openland						1			
11	(N/A)									
12	Upland Shrub									
13	Parks / Golf Courses				0.3		0.7			
14	Northern Hardwood Association				1					
15	Oak Association				1					
16	Aspen Association				1					
17	Other Upland Deciduous				1					

Table A.2. (cont'd)

18	Mixed Upland Deciduous				1						
19	Pines			1							
20	Other Upland Conifers			1							
21	Mixed Upland Conifers			1							
22	Upland Mixed Forest			0.5	0.5						
23	Water	1									
24	Lowland Deciduous Forest				1						
25	Lowland Coniferous Forest			1							
26	Lowland Mixed Forest			0.5	0.5						
27	Floating Aquatic	1									
28	Lowland Shrub										
29	Emergent Wetland	1									
30	Mixed Non-Forest Wetland	1									
31	Sand / Soil									1	
32	Exposed Rock		1								

Table A.2. (cont'd)

33	Mud Flats					0.2			0.8	
34	Other Bare / Sparsely Vegetated					0.2			0.8	

Table A.3. Summary of several watershed-scale hydrologic models

Model (Reference)	**GSSHA** [*Downer and Ogden*, 2004b]	**VIC** [Liang and Xie, 2001]	**MIKE-SHE** [DHI, 2001]	**MODHMS** [Panday and Huyakorn, 2004]
River Flow	1D diffusive wave and Preissman dynamic wave	Routing model of Lohmann for linearized St. Venant	1D diffusive wave	1D diffusive wave equation
Subsurface flow	Coupled Richards equation unsaturated/ saturated groundwater flow model	Variable infiltration capacity, Unit hydrograph routing	1D Richards equation unsaturated and 3D saturated flow	3D variably saturated flow equation
Surface flow	2D diffusive wave	Unit hydrograph routing	2D diffusive wave solved by implicit finite difference scheme	2D saint Venant equation
Components	Snow process, ET, infiltration, soil moisture, groundwater flow, overland flow, stream/groundwater interaction	ET, energy balance, Elevation bands, snow processes, frozen soil, lake model,	Interception, ET, overland and channel flow, unsaturated zone, saturated zone, snowmelt, groundwater-channel interaction, crop growth and nitrogen	Interception, ET, overland flow, channel routing, groundwater-channel interaction
Spatial Representation	2D structured grid, 1D unsaturated, 2D groundwater flow, 1D channel	Large structured grid (>1km)	2D structured grid surface flow, 3D saturated flow, 1D channel, 1D unsaturated	2D structured grid surface flow, 3D saturated flow, 1D channel

Table A.3. (cont'd)

Model (Reference)	**SWAT** [Arnold and Fohrer, 2005]	**tRIBs** [*Ivanov et al.*, 2004a]	**WEP** [Jia et al., 2001]
River Flow	Linear reservoir or Muskingum Kunge routing	Both Hydrologic routing and hydraulic routing	1D kinematic wave
Subsurface flow	Conceptual delayed recharge and baseflow recession	Quasi-3D cascade flow	1D unsaturated flow solved by Richards equation or Generalized Green and Ampt equation and quasi-3D groundwater flow
Surface flow	Conceptual surface lag method	2D cascade flow over TIN network	2D cascade flow
Components	Interception, ET, Curve number runoff, groundwater contribution, soil temperature, soil moisture, soil erosion, crop	Interception, energy balance, ET, infiltration, soil moisture redistribution, groundwater flow , surface runoff,	Interception, ET, infiltration, soil moisture, groundwater, overland flow, channel flow, surface energy balance, anthropogenic components, surface temperature
Spatial Representation	Subbasin delineated from DEM + HRU based on soil/land use combinations	TIN-based unstructured grids, 2D surface flow, 1D unsaturated, quasi 3D subsurface flow	2D structure grid cascade surface flow, 1D unsaturated, semi-3D saturated, 1D river

Appendix B. User's Manual for the model Graphical User Interface

B.1. Creating and Running the model in interactive mode

The proposed model is implemented in a mixed environment of Matlab and Fortran. Computationally intensive subroutines (mainly PDE solvers) are written in Fortran and linked to the Matlab main program via the Matlab mex interface. The main program is written in Matlab due to its efficiency in development and versatility in data handling. It is 'open-ended', meaning data is all conveniently available for inspection during run time. This is a big advantage for researchers interested in advancing the model further. The data interfacing capability of Matlab and efficiency is also important because it allows a researcher to spend less time writing auxiliary subroutines and can instead focus on the scientific part of modeling. The development of this model, along with the complete software package is impossible to be done by the author himself had the model not been written in this fashion.

The model is packaged in a folder that can be run under mainstream computational environments including Windows (32 and 64 bit), Linux (32 and 64) and Macintosh. The model have been compiled, run and tested on all of these platforms. Ensuring the compatibility and portability on all these platforms is not a trivial task. Such effort is spent mainly because the calibrated results shown in this dissertation is done on the

Linux environment at the High Performance Computing Center (HPCC) at MSU.

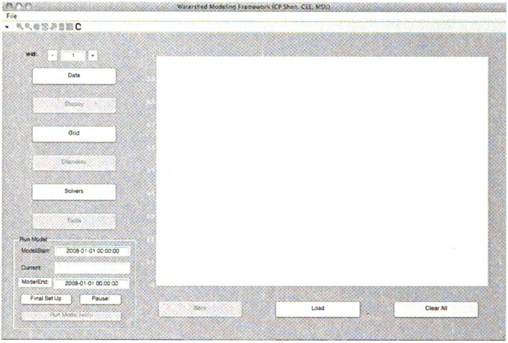
A Graphical User Interface has been developed to assist interfacing with data and setting up the model. The GUI has 7 core capabilities, including, loading data input, specifying grid parameters, discretizing data into model, specifying runtime parameters (Component solvers, Model Start Time, Model End Time, etc), running the model, saving/loading model and displaying the results. The GUI is mainly used during the setting-up stage. The model can be run with or without the GUI.

B.1.1. Installing and starting the model

The model can be run in the interactive mode (in Matlab) or the compiled mode (after compilation by the mcc compiler). Running the model in the interactive mode is the same as running any other Matlab programs. To create a model in the interactive mode with the GUI:

1. Start up Matlab and browse to the root directly of the model ($MROOT).
2. When Matlab path is under $MROOT, enter 'gpath'. This command adds the relevant directories into the Matlab paths and also set the values of some environmental variables (Env in matlab workspace)
3. Enter the command 'mygui'. This command brings up the model main GUI session. Fig B.1 shows the GUI window after it is started. (Depending on the Operating System and the Matlab version, the look of the windows may be

trivially different.)

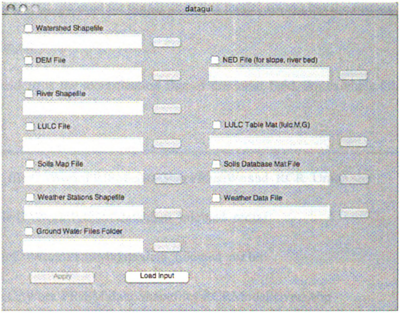


ModelStart: 2008-01-01 00:00:00
Current:
ModelEnd: 2008-01-01 00:00:00

Figure B.1. GUI after the model is started

B.1.2. Loading the data

To load the raw data for discretization, click on the 'Data' button on the main GUI. This brings up the data GUI (Figure B.2a). This window allows the user to specify the input files. The data items listed in this window is summarized in Appendix A. For each of these items, click on the item's checkbox and then use the file browser to load the file(s). For the 'Weather Data File Folder' and 'Ground Water Files Folder', a directory that contains relevant files should be loaded. The 'Soils Map Files' can accept multiple ASCII raster data files because normally SSURGO data is organized in county we may span several counties in our study domain. All other input boxes expect one input file. After a file is loaded, its path is shown in the Edit box below the checkbox (Figure B.2b).



☐ Watershed Shapefile

☐ DEM File

☐ NED File (for slope, river bed)

☐ River Shapefile

☐ LULC File

☐ LULC Table Mat (lulc.M,G)

☐ Soils Map File

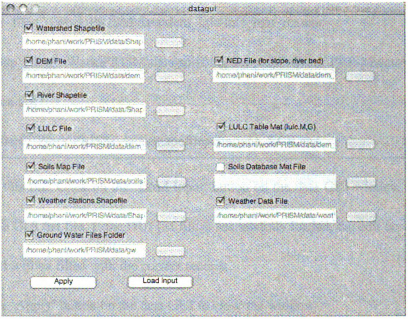
☐ Soils Database Mat File

☐ Weather Stations Shapefile

☐ Weather Data File

☐ Ground Water Files Folder

(a)



☑ Watershed Shapefile: /home/phani/work/PRISM/data/Shap

☑ DEM File: /home/phani/work/PRISM/data/dem

☑ NED File (for slope, river bed): /home/phani/work/PRISM/data/dem

☑ River Shapefile: /home/phani/work/PRISM/data/Shap

☑ LULC File: /home/phani/work/PRISM/data/dem

☑ LULC Table Mat (lulc.M,G): /home/phani/work/PRISM/data/dem

☑ Soils Map File: /home/phani/work/PRISM/data/soils

☐ Soils Database Mat File

☑ Weather Stations Shapefile: /home/phani/work/PRISM/data/Shap

☑ Weather Data File: /home/phani/work/PRISM/data/weat

☑ Ground Water Files Folder: /home/phani/work/PRISM/data/gw

(b)

Figure B.2. Data GUI. (a) before loading data (b) after loading data files

Another way to quickly load the data files and avoiding much of the human actions is to create a GUI list file. Using this GUI list file is the same as manually setting all the fields. A GUI list file used for the GUIs in this model always has the format:

Field_id:input

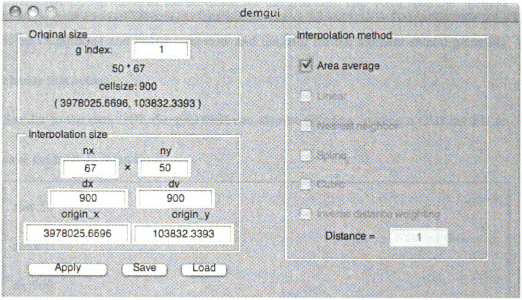
where Field_id is the identification flag of the input field. An example GUI list file looks like:

```
wtrshd_file:C:\Work\PRISM\data\Shapefiles\Wtrshd_RCR_Union.shp
dem_file:C:\Work\PRISM\data\dem_lulc\ned_rcr.txt
ned_file:C:\Work\PRISM\data\dem_lulc\ned_rcr.txt
riv_file:C:\Work\PRISM\data\Shapefiles\RCRModelRivers.shp
lulc_file:C:\Work\PRISM\data\dem_lulc\lulc_rcr_bigger.txt
lulcTB_file:C:\Work\PRISM\data\dem_lulc\lulc_mat_rcr.mat
soilsMap_file:C:\Work\PRISM\data\soils\ingham.txt;C:\Work\PRISM\data\soils\livingston.txt;
wea_file:C:\Work\PRISM\data\Shapefiles\Stations_RCR.shp
```

To load this file, Click on the 'Load Input' button on the data GUI window, and then select the GUI list file. If the GUI list file is successfully loaded, the edit boxes on the data GUI will be filled with the correct records.

Click the 'Apply' button on the data GUI to close the window.

B.1.3. Setting up the grid



g Index: 1
50 * 67
cellsize: 900
(3978025.6696, 103832.3393)

nx: 67 × ny: 50
dx: 900
dy: 900
origin_x: 3978025.6696
origin_y: 103832.3393

☑ Area average
☐ Linear
☐ Nearest neighbor
☐ Spline
☐ Cubic
☐ Inverse distance weighting
Distance = 1

Figure B.3. GUI after the model is started

Press the 'Grid' button on the main GUI will open up the grid GUI (Figure B.3). This window allows the user to specify discretization information. In the box 'nx' and 'ny', user needs to input the number of cells in x and y direction. In 'dx' and 'dy', spatial step size in meter should be input. The boxes origin_x and origin_y stand for the location of the lower left corner of the grid. Origin_x and origin_y information is automatically loaded for a given DEM raster grid. After filling in these fields, the user needs to click one of the options listed on the right to indicate what method should be used to aggregate DEM data in a grid cell. Because DEM maps usually have finer resolution than the computational grid, there can be many DEM point values inside each grid cell. 'Area average' means the elevation data inside one grid cell is averaged to obtain the grid cell elevation. 'Linear', 'Nearest Neighbor', 'Spline' and 'Inverse Distance Weighting are different methods of interpolation. These options will take an interpolated point value as the elevation for the grid cell. The 'Area Average' method is best supported and tested right now and the user for the moment should generally choose this option.

Similar to the data GUI, the grid GUI can also be filled by loading a GUI list file to save the manual input time:

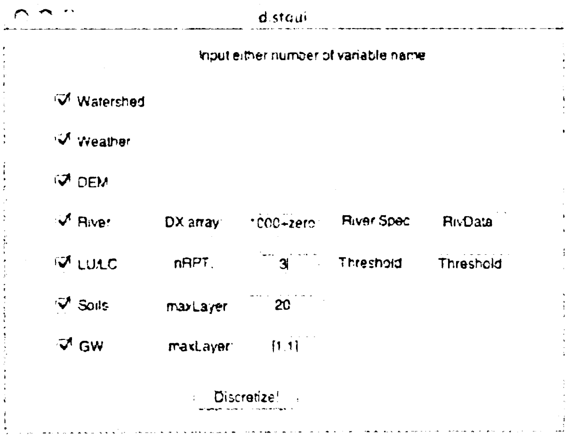
```
row:50

col:67

dx:900

dy:900

origin_x:3978025.6696

origin_y:103832.3393
```

This GUI list file can be written by first filling the information in the GUI and the using the 'Save' button. Then it can be loaded by using the 'Load' button. After the blanks are properly filled, hit the 'Apply' button to finish the grid set-up process. Note that the 'Apply' button is available only after one of the 'interpolation method' options on the right is selected.

B.1.4. Discretization

When grid information step is done, the 'Discretize' button on the main GUI is enabled (Figure B.5). Here we can discretize one, several, or all of the components of the watershed model. In this window the user should select the components that need to be discretized. If a new watershed model is being created, all the boxes should be checked. There are also a few edit boxes that needed to be filled out. The 'DX array' next to the 'River' check box states the spatial step-size for the discretization of the rivers. The program will evaluate the content of the box and to get an array (in Matlab), whose i-th element correspond to the i-th river dx. The nRPT next the 'LULC' checkbox is the number of RPT that are going to be modeled in the domain (see section *** for the meaning of nRPT). Other edit boxes should be left untouched at this moment.

Once the user clicks the 'Discretize!' button, the program will take a couple of minutes to go through the discretization steps for all the components. New data storages will be allocated in memory. Data will be discretized onto the computational grid previously specified. Any data in the memory will be cleaned if the box of that component is discretized.



Input either number or variable name

☑ Watershed				
☑ Weather				
☑ DEM				
☑ River	DX array	000~zero	River Spec	RivData
☑ LULC	nRPT	3	Threshold	Threshold
☑ Soils	maxLayer	20		
☑ GW	maxLayer	[1 1]		

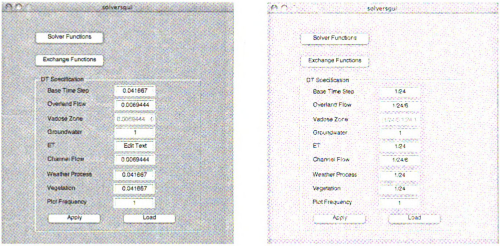
Figure B.4. Discretization GUI

B.1.5. Setting up solution schemes and time steps

The model is written in such a flexible way that different solution schemes to flow domains can replace the default value with great ease. After clicking on the 'Solvers' button the main the solvers GUI is bought up (Figure B.5). The solvers GUI contains two sub-buttons and the 'DT-specification' section where the user specifies the temporal time step that should be used for each component. First we should use the 'Solver Functions' sub-button to open the solver functions GUI. In this new window (Fig B.6a), the left column 'Field' shows the components. And the user is expected to fill in the middle column under 'Value' to indicate which solvers they want to choose for each component. These are expected to be Matlab function handles which accepts input in certain format. For the results published in this work, the settings can be directly loaded without any typing. On the new window, click on 'File→Open' and select the file '$MROOT/data/solvers.mat'. After clicking 'Yes' on the confirmation page (Figure B.6b), we see that some of the fields have been filled (Figure B.6c). The content in the filled fields are the Matlab function handles which correspond to a matlab .m file in the model package. Each of these files can be opened by typing 'edit (FILENAME)' under Matlab command prompt. For example, type the command 'edit GW_sol' in Matlab command prompt will open the file 'GW_sol.m'. Some fields are still 'void' because their solvers may have been combined in other functions. Hit 'OK' to close this window.

The next sub-button 'Exchange Functions' is the place to enter functions that calculate interactions among domains. For the current model structure, the user only need to input '@F_oc_dw' for the 'OC' field.

Then we are ready to specify the time steps. The unit of the input should be 'day'. A quick way to load the current setting is to click the 'Load' button on the solvers GUI and load the file '$MROOT/data/dt.txt'. This will automatically fill in the relevant fields. (Fig B.5b). Hit 'Apply' to save the settings.

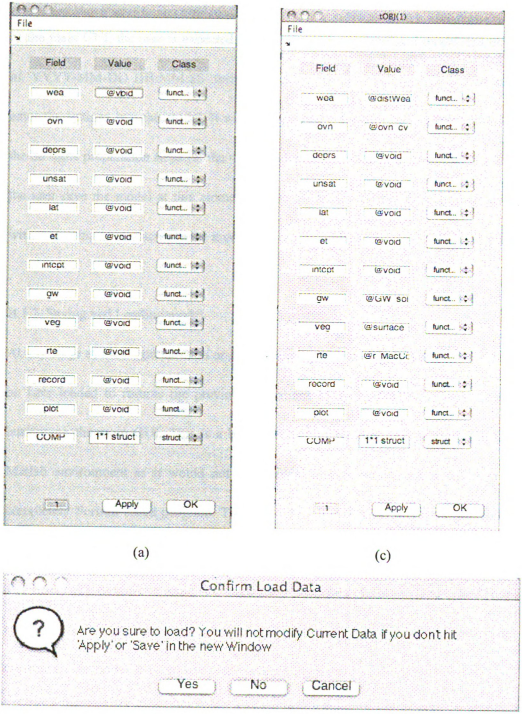


DT Specification
Base Time Step 0.041667
Overland Flow 0.0069444
Vadose Zone 0.0069444 (
Groundwater 1
ET Edit Text
Channel Flow 0.0069444
Weather Process 0.041667
Vegetation 0.041667
Plot Frequency 1

DT Specification
Base Time Step 1/24
Overland Flow 1/24/6
Vadose Zone 1/24/6 1/24 1
Groundwater 1
ET 1/24
Channel Flow 1/24/6
Weather Process 1/24
Vegetation 1/24
Plot Frequency 1

(a) (b)

Figure B.5. Solvers GUI. (a) after it is opened (b) after dt file is loaded



Field	Value	Class
wea	@void	funct...
ovn	@void	funct...
deprs	@void	funct...
unsat	@void	funct...
lat	@void	funct...
et	@void	funct...
intcpt	@void	funct...
gw	@void	funct...
veg	@void	funct...
rte	@void	funct...
record	@void	funct...
plot	@void	funct...
COMP	1*1 struct	struct

(a)

tOBJ(1)

Field	Value	Class
wea	@distWea	funct...
ovn	@ovn_cv	funct...
deprs	@void	funct...
unsat	@void	funct...
lat	@void	funct...
et	@void	funct...
intcpt	@void	funct...
gw	@GW_sol	funct...
veg	@surface	funct...
rte	@r_MacC	funct...
record	@void	funct...
plot	@void	funct...
COMP	1*1 struct	struct

(c)

Confirm Load Data

Are you sure to load? You will not modify Current Data if you don't hit 'Apply' or 'Save' in the new Window

(b)

Figure B.6. Loading solvers data into solvers functions GUI

B.1.6. Final model set up

Before model can be run, a few steps remain to be done. In the 'Run Model' section of the main GUI, the user is expected to set the model start time and model end time in 'YYYY-MM-DD HH:MM:SS' format. Clicking on the 'Final Set Up' button will enter the settings into the model. It will also execute a function called 'modelSetUp' the do some preparation steps for the model to be ready to run. It is recommended that the user save the model at this moment so the model can be run directly from here without having to re-discretize the model.

B.1.7. Saving and Loading model

At any stage of model preparation or model running, the model can be saved and can be later loaded to resume the previous operations by using the 'Save' and 'Load' buttons on the main GUI. This is a big advantage of building up the model in the Matlab environment as it would not be a trivial task to write such a utility in a completely Fortran-based program. The 'Save' and 'Load' functions have proven to be hugely convenient during the model development stage.

The 'Clear All' button can erase any model data from the current workspace and start a new main GUI. It is recommended to use 'Clear All' function before loading or creating a new model to avoid any potential memory issue.

B.1.8. Running the model

After the final model setup is done, the model is ready to be run and the 'Run Model' button is now enabled. Clicking on this button will start to run the model from the time specified in the 'ModelStart' box. The simulation can be paused by clicking on the 'Pause!' button and resumed by clicking on the 'Run Model' button again.

When the model is running, the Matlab command window will be displaying some time usage information (in seconds) after each day of simulation. The items are, sequentially:

Year, Julian Day, T-vadoze zone, T-overland flow, T-river flow, T-groundwater, T-total.

This will tell the user that the simulation is proceeding properly.

An example run-time information is given below:

```
>> run('nlST.mat',0,'example_Run')
Driver control file:
mat initalization    :nlST.mat
Model Project Name: example_Run
Start Running Model at wallclock: 23-Nov-2009 16:48:02
File Output to: example_Run.txt & example_Run_result.txt
Model Start Time: 01-Sep-2001
Model End Time: 31-Dec-2005
            2001           244        1.1549         1.6202          1.8675
0.35064        4.9932
            2001           245       0.83704         1.0274         0.89959
0.11235        2.8764
            2001           246        0.8243        0.98791         0.86801
0.11384        2.7941
            2001           247       0.76272        0.95858         0.86146
0.10863        2.6914
.................................................
```

B.2. Running the model in non-interactive mode

For model calibration or long term simulation, the model can be run in non-interactive mode (without the GUI). The user can either run the model in Matlab with a command or run a compiled version of the model without invoking Matlab at all.

To run the model in non-interactive mode, the .mat file saved after step B.1.6 must be accessible. At Matlab prompt, the model can be run using the following command:

Run(*matfile*, *par*, *prj*)

The *matfile* is the filename of the .mat that is saved after step B.1.6. It should be a string variable. *par* is a Matlab structure array for parameter adjustment information. If no parameter is to be changed, put number 0 at this argument location. *prj* is the project name for the simulation (a string). This name will be used to write output.

The model has also been compiled on into standalone executables that can be run on Windows/Linux/Macintosh platforms. However the Linux executable may not run on a random Linux environment due to the numerously different Linux systems and machine architectures. Normally a Linux program needs to be compiled from source. Assuming the compiled version can work properly. The command can be run at the command prompt of the operating system:

PRISM *driverfile*

driverfile is the filename of a text input driver file. This driver file contains control information for the model simulation. It has five lines, with explanations in the parenthesis:

project_name (Simulation name that is used to save output)

matfile (Mat file that contains the data)

ModelStartTime (YYYYMMDD, or 0 if ModelStartTime saved in the mat file is used)

ModelEndTime (YYYYMMDD, or 0 if ModelEndTime saved in the mat file is used)

Parameter_changer_file (a parameter changer control file, leave blank if none)

References

Adams, M. H. (1959), *Bacteriophage*, Interscience Publishers, Inc., New York.

Arnold, J. G., and P. M. Allen (1996), Estimating hydrologic budgets for three Illinois watersheds, *Journal of Hydrology*, *176*(1-4), 57-77.

Arnold, J. G., and N. Fohrer (2005), SWAT2000: current capabilities and research opportunities in applied watershed modelling, *Hydrological Processes*, *19*(3), 563-572.

Assouline, S. (2004), A model for soil relative hydraulic conductivity based on the water retention characteristic curve (vol 37, pg 265, 2001), *Water Resources Research*, *40*(2), -.

Atkinson, T. C., and P. M. Davis (2000), Longitudinal dispersion in natural channels: 1. Experimental results from the River Severn, U.K., *Hydrology and Earth System Sciences*, *4*(3), 345-353.

Barnett, T. P., and D. W. Pierce (2008), When will Lake Mead go dry?, *Water Resources Research*, *44*(3), -.

Barnett, T. P., and D. W. Pierce (2009), Reply to comment by J. J. Barsugli et al. on "When will Lake Mead go dry?", *Water Resources Research*, *45*, -.

Bencala, K. E., and R. A. Walters (1983), Simulation of Solute Transport in a Mountain Pool-and-Riffle Stream - a Transient Storage Model, *Water Resources Research*, *19*(3), 718-724.

Bencala, K. E., D. M. McKnight, and G. W. Zellweger (1990), Characterization of Transport in an Acidic and Metal-Rich Mountain Stream Based on a Lithium Tracer Injection and Simulations of Transient Storage, *Water Resources Research*, *26*(5), 989-1000.

Beven, K. (1978), The hydrological response of headwater and sideslope areas, *Hydrological Sciences Bulletin*, *23*, 419-483.

Beven, K., and M. Kirkby (1979), A physically based, variable contributing area model of basin hydrology., *Hydrological Sciences Bulletin*, *24*(11).

Beven, K. (1985), Distributed models, in *Hydrologic forecasting*, edited by M. J. Anderson and T. P. Burt, pp. 405-435, Wiley, New York.

Beven, K. (2002), Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system, *Hydrological Processes*, *16*(2), 189-206.

Beven, K. (2006), A manifesto for the equifinality thesis, *Journal of Hydrology*, *320*(1-2), 18-36.

Beven, K. J., and A. Binley (1992), The future of distributed models: model calibration and uncertainty prediction, *Hydrological Processes*, *6*, 279-298.

Bjerklie, D. M. (2007), Estimating the bankfull velocity and discharge for rivers using remotely sensed river morphology information, *Journal of Hydrology*, *341*(3-4), 144-155.

Boano, F., A. I. Packman, A. Cortis, et al. (2007), A continuous time random walk approach to the stream transport of solutes, *Water Resources Research*, *43*(10).

Borah, D. K., and M. Bera (2003), Watershed-scale hydrologic and nonpoint-source pollution models: Review of mathematical bases, *Transactions of the Asae*, *46*(6), 1553-1566.

Borah, D. K., and M. Bera (2004), Watershed-scale hydrologic and nonpoint-source pollution models: Review of applications, *Transactions of the Asae*, *47*(3), 789-803.

Braud, I., N. Varado, and A. Olioso (2005), Comparison of root water uptake modules using either the surface energy balance or potential transpiration, *Journal of Hydrology*, *301*(1-4), 267-286.

Breuer, L., K. Eckhardt, and H. G. Frede (2003), Plant parameter values for models in temperate climates, *Ecological Modelling*, *169*(2-3), 237-293.

Burnash, R. J. C. (1995), The NWS river forecast system--Catchment modeling, in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, pp. 311-366, Water Resources Publications, Littleton, Colorado.

Campbell, G. S. (1985), *Soil physics with BASIC: transport models for soil-plant systems*, Elsevier, Amsterdam.

Campbell, G. S., and J. M. Norman (1998), *Introduction ot Environmental Biophysics*, Springer-Verlag, New York.

Carr, M. L., and C. R. Rehmann (2007), Measuring the dispersion coefficent with acoustic doppler current profilers, *Journal of Hydraulic Engineering-Asce*, *133*(8), 977-982.

Casulli, V. (1990), Semi-Implicit Finite-Difference Methods for the 2-Dimensional Shallow-Water Equations, *Journal of Computational Physics*, *86*(1), 56-74.

Casulli, V. (1999), A semi-implicit finite difference method for non-hydrostatic, free-surface flows, *International Journal for Numerical Methods in Fluids*, *30*(4), 425-440.

Celia, M. A., E. T. Bouloutas, and R. L. Zarba (1990), A General Mass-Conservative Numerical-Solution for the Unsaturated Flow Equation, *Water Resources Research*, *26*(7), 1483-1496.

Chakraborty, U. K. (2008), *Advances in Differential Evolution (Studies in Computational Intelligence)* Springer.

Chapra, S. C. (1997), Lecture 10, Distributed Systems, in: Surface water quality modeling, McCraw-Hill.

Chapra, S. C. (2008), *Surface Water-Quality Modeling*, 844 pp., Waveland Press Inc., Long Grove, IL.

Clement, T. P., W. R. Wise, and F. J. Molz (1994), A Physically-Based, 2-Dimensional, Finite-Difference Algorithm for Modeling Variably Saturated Flow, *Journal of Hydrology*, *161*(1-4), 71-90.

Corner, R. A., D. A. Albert, M. B. Austin, et al. (1999), Landtype Associations of Northern Michigan Section VII, available at http://www.mcgi.state.mi.us/mgdl/?action=thm, retrieved 2009-Nov-29, edited, Michigan Natural Features Inventory, Lansing, MI.

Coulibaly, P., and D. H. Burn (2004), Wavelet analysis of variability in annual Canadian streamflows, *Water Resources Research*, *40*(3), -.

Croley, T. E., and C. S. He (2005), Distributed-parameter large basin runoff model. I: Model development, *Journal of Hydrologic Engineering*, *10*(3), 173-181.

D'Angelo, D. J., J. R. Webster, S. V. Gregory, et al. (1993), Transient Storage in Appalachian and Cascade Mountain Streams as Related to Hydraulic Characteristics, *Journal of the North American Benthological Society*, *12*(3), 223-235.

Daubechies, I. (1988), Orthonormal Bases of Compactly Supported Wavelets, *Communications on Pure and Applied Mathematics*, *41*(7), 909-996.

Day, T. J. (1975), Longitudinal dispersion in natural streams, *Water Resources Research*, *11*(6), 909-918.

De Artigas, M. Z., A. G. Elias, and P. F. de Campra (2006), Discrete wavelet analysis to assess long-term trends in geomagnetic activity, *Physics and Chemistry of the Earth*, *31*(1-3), 77-80.

De Smedt, F., W. Brevis, and P. Debels (2005), Analytical solution for solute transport resulting from instantaneous injection in streams with transient storage, *Journal of Hydrology*, *315*(1-4), 25-39.

Demuren, A. O., R. V. Wilson, and M. Carpenter (2001), Higher-order compact schemes for numerical simulation of incompressible flows, part I: Theoretical development, *Numerical Heat Transfer Part B-Fundamentals*, *39*(3), 207-230.

Deng, Z. Q., V. P. Singh, and L. Bengtsson (2001), Longitudinal dispersion coefficient in straight rivers, *Journal of Hydraulic Engineering-Asce*, *127*(11), 919-927.

Deng, Z. Q., and H. S. Jung (2009), Scaling dispersion model for pollutant transport in rivers, *Environmental Modelling & Software*, *24*(5), 627-631.

DHI (2001), *MIKE SHE Code Verification and Validation (includes detailed analytical and numerical validation examples)*

DiGiammarco, P., E. Todini, and P. Lamberti (1996), A conservative finite elements approach to overland flow: The control volume finite element formulation, *Journal of Hydrology*, *175*(1-4), 267-291.

Dinehart, R. L., and J. R. Burau (2005), Repeated surveys by acoustic Doppler current profiler for flow and sediment dynamics in a tidal river, *Journal of Hydrology*, *314*(1-4), 1-21.

Dingman, S. L. (2002), *Physical Hydrology*, Prentice Hall.

Division, C. E. (1986), Urban Hydrology for Small Watersheds - Technical Release 55, edited by N. R. C. Service, U.S. Department of Agriculture.

Dogan, A., and L. H. Motz (2005), Saturated-unsaturated 3D groundwater model. II: Verification and application, *Journal of Hydrologic Engineering*, *10*(6), 505-515.

Downer, C. W., F. L. Ogden, W. D. Martin, et al. (2002), Theory, development, and applicability of the surface water hydrologic model CASC2D, *Hydrological Processes*, *16*(2), 255-275.

Downer, C. W., and F. L. Ogden (2004a), Appropriate vertical discretization of Richards' equation for two-dimensional watershed-scale modelling, *Hydrological Processes*, *18*(1), 1-22.

Downer, C. W., and F. L. Ogden (2004b), GSSHA: Model to simulate diverse stream flow producing processes, *Journal of Hydrologic Engineering*, *9*(3), 161-174.

Duan, Q. Y., S. Sorooshian, and V. Gupta (1992), Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models, *Water Resources Research*, *28*(4), 1015-1031.

Duan, Q. Y., S. Sorooshian, and V. K. Gupta (1994), Optimal Use of the SCE-UA Global Optimization Method for Calibrating Watershed Models, *Journal of Hydrology*, *158*(3-4), 265-284.

Dunne, T., and R. D. Black (1970), Partial Area Contributions to Storm Runoff in a Small New-England Watershed, *Water Resources Research*, *6*(5), 1296-&.

Dunne, T., T. R. Moore, and C. H. Yaylor (1975), Recognition and prediction of runoff-producing zones in humid regions, *Hydrological Sciences Bulletin*, *20*, 305-327.

Engelhardt, C., A. Kruger, A. Sukhodolov, et al. (2004), A study of phytoplankton spatial distributions, flow structure and characteristics of mixing in a river reach with groynes, *Journal of Plankton Research*, *26*(11), 1351-1366.

FAO (1998), Crop evaporation. Guidelines for computing crop water requirement--Irrigation paper no. 56, Available at http://www.fao.org/docrep/X0490E/X0490E00.htm, Rome.

Fernald, A. G., P. J. Wigington, and D. H. Landers (2001), Transient storage and hyporheic flow along the Willamette River, Oregon: Field measurements and model estimates, *Water Resources Research*, *37*(6), 1681-1694.

Fischer, H. B. (1979), *Mixing in inland and coastal waters*, xiv, 483 p. pp., Academic Press, New York.

Fisher, H. B., E. J. List, R. C. Y. Koh, et al. (1979), *Mixing in inland and coastal waters*, Academic Press, San Diego, California.

Freeze, A. R., and J. A. Cherry (1979), *Groundwater*, Prentice Hall.

Freeze, R. A., and R. L. Harlan (1969), Blueprint for a physically-based, digitally-simulated hydrologic response model, *Journal of Hydrology*, *9*, 237-258.

García, C. M., K. Oberg, and M. H. García (2007), ADCP Measurements of Gravity Currents in the Chicago River, Illinois, *Journal of Hydraulic Engineering (ASCE)*, *133*(12), 1356-1366.

Gooseff, M. N., J. K. Anderson, S. M. Wondzell, et al. (2005), A modelling study of hyporheic exchange pattern and the sequence, size, and spacing of stream bedforms in mountain stream networks, Oregon, USA (Retracted article. See vol 20, pg 2441, 2006), *Hydrological Processes*, *19*(15), 2915-2929.

Gottardi, G., and M. Venutelli (2008), An accurate time integration method for simplified overland flow models, *Advances in Water Resources*, *31*(1), 173-180.

Green, W. H., and G. A. Ampt (1911), Studies of soil physics. I: Flow of air and water through soils, *J. Agric. Sci.*, *4*, 1-24.

Greenberg, A. E., L. S. Clesceri, and A. D. Eaton (Eds.) (1992), *Standard methods for the examination of water and wastewater*, 18th ed., American Public Health Association, American Water Works Association, and Water Environment Federation, Washington, D.C.

Guan, S. G., C. H. Lai, and G. W. Wei (2003), Characterizing the spatiotemporal dynamics of turbulence, *Computer Physics Communications*, *155*(2), 77-91.

Gucker, B., and I. G. Boechat (2004), Stream morphology controls ammonium retention in tropical headwaters, *Ecology*, *85*(10), 2818-2827.

Gunduz, O., and M. M. Aral (2005), River networks and groundwater flow: a simultaneous solution of a coupled system, *Journal of Hydrology*, *301*(1-4), 216-234.

GWIM (2006), State of Michigan Public Act 148 Groundwater Inventory and Mapping Project (GWIM). Technical Report.

Harvey, J. W., B. J. Wagner, and K. E. Bencala (1996), Evaluating the reliability of the stream tracer approach to characterize stream-subsurface water exchange, *Water Resources Research*, *32*(8), 2441-2451.

Harvey, J. W., and B. J. Wagner (2000), *Quantifying Hydrologic Interactions Between Streams and Their Subsurface Hyporheic Zones*, Elsevier, New York.

Harvey, R. W., and J. N. Ryan (2004), Use of PRDI bacteriophage in groundwater viral transport, inactivation, and attachment studies, *FEMS Microbiol. Ecol.*, *49*(1), 3-16.

Haverkamp, R., M. Vauclin, J. Touma, et al. (1977), Comparison of Numerical-Simulation Models for One-Dimensional Infiltration, *Soil Sci Soc Am J*, *41*(2), 285-294.

Healy, R. W., and P. G. Cook (2002), Using groundwater levels to estimate recharge, *Hydrogeol J*, *10*(1), 91-109.

HEC (2000), *Hydrologic Modeling System: Technical Reference Manual*, US Army Corps of Engineers Hydrologic Engineering Center, Davis, CA.

Hilberts, A. G. J., P. A. Troch, and C. Paniconi (2005), Storage-dependent drainable porosity for complex hillslopes, *Water Resources Research*, *41*(6), -.

Hill-Rowley, R., T. McClain, and M. Malone (2003), Static water level mapping in east central Michigan, *Journal of the American Water Resources Association*, *39*(1), 99-111.

Holtschlag, D. J., and J. A. Koschik (2003), An acoustic Doppler current profiler survey of flow velocities in St. Clair River, a connecting channel of the Great Lakes, U.S. Geological Survey, Lansing, Michigan.

Hornberger, G. M., J. P. Raffensperger, P. L. Wiberg, et al. (1998), *Elements of Physical Hydrology*, The Johns Hopkins University Press, Baltimore.

Horton, R. E. (1933), The role of infiltration in the hydrologic cycle, *Transactions-American Geophysical Union*, *14*, 446-460.

HPCC (2009), High Performance Computing Center, Michigan State University, webpage at: http://www.hpcc.msu.edu/, retrived on 2009-Nov-29, edited, East Lansing, MI.

IPCC (2007), Climate change and water in *Climate Change 2007: Synthesis Report.* , in *Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.*, edited.

Irfan, A. (2002), Hydrodynamic and logitudinal dispersion in the Red Cedar River, Michigan State University, East Lansing.

Ivanov, V. Y., E. R. Vivoni, R. L. Bras, et al. (2004a), Catchment hydrologic response with a fully distributed triangulated irregular network model, *Water Resources Research*, *40*(11), -.

Ivanov, V. Y., E. R. Vivoni, R. L. Bras, et al. (2004b), Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: a fully-distributed physically-based approach, *Journal of Hydrology*, *298*(1-4), 80-111.

Ivanov, V. Y., R. L. Bras, and E. R. Vivoni (2008a), Vegetation-hydrology dynamics in complex terrain of semiarid areas: 2. Energy-water controls of vegetation spatiotemporal dynamics and topographic niches of favorability, *Water Resources Research*, *44*(3), -.

Ivanov, V. Y., R. L. Bras, and E. R. Vivoni (2008b), Vegetation-hydrology dynamics in complex terrain of semiarid areas: 1. A mechanistic approach to modeling dynamic feedbacks, *Water Resources Research*, *44*(3), -.

Jamieson, R., R. Gordon, D. Joy, et al. (2004), Assessing microbial pollution of rural surface waters - A review of current watershed scale modeling approaches, *Agric. Water Manage.*, *70*(1), 1-17.

Jenkins, T. M., T. M. Scott, M. R. Morgan, et al. (2005), Occurrence of alternative fecal indicators and enteric viruses in Michigan rivers, *J. Gt. Lakes Res.*, *31*(1), 22-31.

Jensen, M. E., R. D. Burman, and R. G. Allen (1990), *Evapotranspiration and irrigation water requirements. ASCE Manuals and Reports on Engineering Practice No. 70*, ASCE, N.Y.

Jia, Y. W., G. H. Ni, Y. Kawahara, et al. (2001), Development of WEP model and its application to an urban watershed, *Hydrological Processes*, *15*(11), 2175-2194.

Jia, Y. W., H. Wang, Z. H. Zhou, et al. (2006), Development of the WEP-L distributed hydrological model and dynamic assessment of water resources in the Yellow River basin, *Journal of Hydrology*, *331*(3-4), 606-629.

Jobson, H. E. (1997), Predicting travel time and dispersion in rivers and streams, *J. Hydraulic Eng (ASCE)*, *123*(11), 971-978.

Karvonen, T., H. Koivusalo, M. Jauhiainen, et al. (1999), A hydrological model for predicting runoff from different land use areas, *Journal of Hydrology*, *217*(3-4), 253-265.

King, F. H. (1899), Principles and conditions of the movements of groundwater, 86-91 pp.

Kirkham, D. (1967), Explanation of Paradoxes in Dupuit-Forchheimer Seepage Theory, *Water Resources Research*, *3*(2), 609-&.

Kolpin, D. W., E. T. Furlong, M. T. Meyer, et al. (2002), Pharmaceuticals, hormones, and other organic wastewater contaminants in US streams, 1999-2000: A national reconnaissance, *Environmental Science & Technology*, *36*(6), 1202-1211.

Koltun, G. F., C. J. Ostheimer, and M. S. Griffin (2006), Velocity, Bathymetry, and Transverse Mixing Characteristics of the Ohio River Upstream from Cincinnati, Ohio, October 2004—March 2006 U.S. Geological Survey, Reston, Virginia.

Kumar, P., and E. Foufoulageorgiou (1993), A Multicomponent Decomposition of Spatial Rainfall Fields .1. Segregation of Large-Scale and Small-Scale Features Using Wavelet Transforms, *Water Resources Research*, *29*(8), 2515-2532.

Lai, C. T., and G. Katul (2000), The dynamic role of root-water uptake in coupling potential to actual transpiration, *Advances in Water Resources*, *23*(4), 427-439.

Lange, J., N. Greenbaum, S. Husary, et al. (2003), Runoff generation from successive simulated rainfalls on a rocky, semi-arid, Mediterranean hillslope, *Hydrological Processes*, *17*(2), 279-296.

Leach, H. R., H. L. Cook, and R. E. Horton (1933), Storm-flow prediction, *Transactions-American Geophysical Union*, *14*, 435-446.

Lehmann, F., and P. H. Ackerer (1998), Comparison of iterative methods for improved solutions of the fluid flow equation in partially saturated porous media, *Transport in Porous Media*, *31*(3), 275-292.

Leveque, R. J. (2007), *Finite difference methods for ordinary and partial differential equations: Steady state and time-dependent problems*, SIAM, Society for Industrial and Applied Mathematics.

Leverett, F., and F. B. Taylor (1915), The Pleistocene of Indiana and Michigan and the history of the Great Lakes, *U. S. Geol. Surv. Monogr*, *53*, 209-211.

Li, S. G, and Q. Liu (2004), Interactive groundwater (IGW): An innovative digital laboratory for groundwater education and research, *Computer Applications in Engineering Education*, *11*(4), 179-202.

Li, S. G, Q. Liu, and S. Afshari (2006), An object-oriented hierarchical patch dynamics paradigm (HPDP) for modeling complex groundwater systems across multiple-scales, *Environmental Modelling & Software*, *21*(5), 744-749.

Liang, X., and Z. H. Xie (2001), A new surface runoff parameterization with subgrid-scale soil heterogeneity for land surface models, *Advances in Water Resources*, *24*(9-10), 1173-1193.

Lockwood, O. G, and H. Kanamori (2006), Wavelet analysis of the seismograms of the 2004 Sumatra-Andaman earthquake and its application to tsunami early warning, *Geochemistry Geophysics Geosystems*, *7*, -.

Luce, C. H., D. G Tarboton, and K. R. Cooley (1999), Sub-grid parameterization of snow distribution for an energy and mass balance snow cover model, *Hydrological Processes*, *13*(12-13), 1921-1933.

Luce, C. H., and D. G Tarboton (2004), The application of depletion curves for parameterization of subgrid variability of snow, *Hydrological Processes*, *18*(8), 1409-1422.

Mace, R. E. (1997), Determination of transmissivity from specific capacity tests in a karst aquifer, *Ground Water*, *35*(5), 738-742.

Mallat, S. G (1989), Multiresolution Approximations and Wavelet Orthonormal Bases of L2(R), *Transactions of the American Mathematical Society*, *315*(1), 69-87.

Martin, N., and S. M. Gorelick (2005), MOD_FreeSurf2D: A MATLAB surface fluid flow model for rivers and streams, *Computers & Geosciences*, *31*(7), 929-946.

MAWN (2010), Michigan Automated Weather Network, East Lansing, Michigan, 48823, Available at: http://www.agweather.geo.msu.edu/mawn/, edited.

Maxwell, R. M., and S. J. Kollet (2008), Quantifying the effects of three-dimensional subsurface heterogeneity on Hortonian runoff processes using a coupled numerical, stochastic approach, *Advances in Water Resources*, *31*(5), 807-817.

McCuen, R. H., Z. Knight, and A. G. Cutter (2006), Evaluation of the Nash-Sutcliffe efficiency index, *Journal of Hydrologic Engineering*, *11*(6), 597-602.

MDEQ (2010), 1987 Bedrock Geology, available at http://www.mcgi.state.mi.us/mgdl/?action=thm, retrieved 2009-11-29.

MDNR (2010a), 1982 Quaternary Geology, available at: http://www.mcgi.state.mi.us/mgdl/?action=thm, retrieved 2009-Nov-29.

MDNR (2010b), 2001 IFMAP/GAP Lower Peninsula Land Cover, available at http://www.mcgi.state.mi.us/mgdl/?rel=thext&action=thmname&cid=5&cat=Land+Cover+2001, retrieved 2009-Nov-28.

Morrice, J. A., H. M. Valett, C. N. Dahm, et al. (1997), Alluvial characteristics, groundwater-surface water exchange and hydrological retention in headwater streams, *Hydrological Processes*, *11*(3), 253-267.

Muste, M., K. Yu, and M. Spasojevic (2004), Practical aspects of ADCP data use for quantification of mean river flow characteristics; Part 1: moving-vessel measurements, *Flow Meas. Instrum.*, *15*(1), 1-16.

NCDC (2010), National Climatic Data Center, available at http://www.ncdc.noaa.gov/oa/climate/climatedata.html#daily, edited.

Neitsch, S. L., J. G. Arnold, J. R. Kiniry, et al. (2005), *Soil and Water Assessment Tool theoretical documentation version 2005*, US Department of Agriculture - Agricultural Research Service, Temple, Texas.

Neupauer, R. M., K. L. Powell, X. Qi, et al. (2006), Characterization of permeability anisotropy using wavelet analysis, *Water Resources Research*, *42*(7), -.

Noilhan, J., and S. Planton (1989), A Simple Parameterization of Land Surface Processes for Meteorological Models, *Monthly Weather Review*, *117*(3), 536-549.

NRCS (1986), Urban Hydrology for Small Watersheds -- TR55, edited by U. S. D. o.

Agriculture.

Olivera, F., M. Valenzuela, R. Srinivasan, et al. (2006), ArcGIS-SWAT: A geodata model and GIS interface for SWAT, *Journal of the American Water Resources Association*, *42*(2), 295-309.

Pairaud, I., and F. Auclair (2005), Combined wavelet and principal component analysis (WEof) of a scale-oriented model of coastal ocean gravity waves, *Dynamics of Atmospheres and Oceans*, *40*(4), 254-282.

Palancar, M. C., J. M. Aragon, F. Sanchez, et al. (2003), The. determination of longitudinal dispersion coefficients in rivers, *Water Environment Research*, *75*(4), 324-335.

Panday, S., and P. S. Huyakorn (2004), A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow, *Advances in Water Resources*, *27*(4), 361-382.

Perucca, E., C. Camporeale, and L. Ridolfi (2009), Estimation of the dispersion coefficient in rivers with riparian vegetation, *Advances in Water Resources*, *32*(1), 78-87.

Phanikumar, M. S., and J. T. McGuire (2004), A 3D partial-equilibrium model to simulate coupled hydrogeological, microbiological, and geochemical processes in subsurface systems, *Geophysical Research Letters*, *31*(11), -.

Phanikumar, M. S., I. Aslam, C. P. Shen, et al. (2007), Separating surface storage from hyporheic retention in natural streams using wavelet decomposition of acoustic Doppler current profiles, *Water Resources Research*, *43*(5), -.

Piedallu, C., and J. C. Gegout (2007), Multiscale computation of solar radiation for predictive vegetation modelling, *Annals of Forest Science*, *64*(8), 899-909.

Prasad, R., D. G. Tarboton, G. E. Liston, et al. (2001), Testing a blowing snow model against distributed snow measurements at Upper Sheep Creek, Idaho, United States of America, *Water Resources Research*, *37*(5), 1341-1356.

Press, W. H., B. P. Flannery, S. A. Teukolsky, et al. (1997), *Numerical Recipes in*

FORTRAN 77: The Art of Scientific Computing, Cambridge University Press.

Puzio, M. C., and G. J. Larson (1982), Surficial geology of Ingham County, map, Lansing.

Rawls, W. J., T. J. Gish, and D. L. Brakensiek (1991), Estimating soil water retention from soil physical properties and characteristics, in *Advances in Soil Science*, edited by B. A. Stewart, Springer, New York.

Reed, S., V. Koren, M. Smith, et al. (2004), Overall distributed model intercomparison project results, *Journal of Hydrology*, *298*(1-4), 27-60.

Reeves, H. W., K. V. Wright, and J. R. Nicholas (2003), Hydrolgeology and simulation of regional groundwater level declines in Monroe County, Michigan, U.S. Geological Survey.

Runkel, R. L. (1998), One-dimensional Transport with Inflow and Storage (OTIS): A Solute Transport Model for Streams and Rivers, 73 pp, US Dept. of the Interior, US Geological Survey.

Runkel, R. L., D. M. McKnight, and E. D. Andrews (1998), Analysis of transient storage subject to unsteady flow: diel flow variation in an Antarctic stream, *Journal of the North American Benthological Society*, *17*(2), 143-154.

Runkel, R. L. (2002), A new metric for determining the importance of transient storage, *Journal of the North American Benthological Society*, *21*(4), 529-543.

Runkel, R. L., D. M. McKnight, and H. Rajaram (2003), Modeling hyporheic zone processes - Preface, *Advances in Water Resources*, *26*(9), 901-905.

Rutherford, J. C. (1994), *River Mixing*, John Wiley & Sons, New York.

Salehin, M., A. I. Packman, and A. Worman (2003), Comparison of transient storage in vegetated and unvegetated reaches of a small agricultural stream in Sweden: seasonal variation and anthropogenic manipulation, *Advances in Water Resources*, *26*(9), 951-964.

Sangrey, D. A., K. O. Harropwilliams, and J. A. Klaiber (1984), Predicting Groundwater Response to Precipitation, *Journal of Geotechnical Engineering-Asce*, *110*(7), 957-975.

Satterlund, D. R. (1979), Improved Equation for Estimating Long-Wave-Radiation from the Atmosphere, *Water Resources Research*, *15*(6), 1649-1650.

Schmid, B. H. (2004), Simplification in longitudinal transport modeling: case of instantaneous slug releases, *Journal of Hydrologic Engineering*, *9*(4), 319-324.

Seo, I. W., and T. S. Cheong (1998), Predicting longitudinal dispersion coefficient in natural streams, *Journal of Hydraulic Engineering-Asce*, *124*(1), 25-32.

Shen, C. P., M. S. Phanikumar, T. T. Fong, et al. (2008), Evaluating bacteriophage P22 as a tracer in a complex surface water system: The Grand River, Michigan, *Environmental Science & Technology*, *42*(7), 2426-2431.

Shields, F. D., S. S. Knight, S. Testa, et al. (2003), Use of acoustic Doppler current profilers to describe velocity distributions at the reach scale, *Journal of the American Water Resources Association*, *39*(6), 1397-1408.

Shu, C. W., and S. Osher (1989), Efficient Implementation of Essentially Non-Oscillatory Shock-Capturing Schemes .2., *Journal of Computational Physics*, *83*(1), 32-78.

Shu, C. W. (1997), Essentially non-oscillatory and weighted non-oscillatory schemes for hyperbolic conservation laws, NASA.

Simard, A. (2007), Predicting groundwater flow and transport using Michigan's statewide wellogic database, 109 pp, Michigan State University, East Lansing.

Simpson, M. R. (2001), Discharge measurements using a broad-band acoustic Doppler current profiler, Sacramento, California.

Singh, V. P. (1996), *Kinematic wave modeling in water resources: Surface water hydrology*, Wiley-Interscience.

Spokas, K., and F. Forcella (2006), Estimating hourly incoming solar radiation from limited meteorological data, *Weed Science*, *54*(1), 182-189.

Stanley, D. J., and A. G. Warne (1993), Nile Delta - Recent Geological Evolution and Human Impact, *Science*, *260*(5108), 628-634.

Stephenson, D., and M. E. Meadows (1986), *Kinematics Hydrology and Modelling*, Elsevier, Amsterdam.

Stoppelenburg, F. J., K. Kavar, M. J. H. Pastoors, et al. (2005), Modeling the interactions between saturated and unsaturated groundwater. Off-line coupling of LGM and SWAP, RIVM Rep. 500026001/2005. RIVM, Bilthoven, the Netherlands.

Strzepek, K. M., G. W. Yohe, R. S. J. Tol, et al. (2008), The value of the high Aswan Dam to the Egyptian economy, *Ecological Economics*, *66*(1), 117-126.

Sukhodolov, A., C. Engelhardt, A. Kruger, et al. (2004), Case study: Turbulent flow and sediment distributions in a groyne field, *Journal of Hydraulic Engineering-Asce*, *130*(1), 1-9.

Taylor, G. I. (1954), The dispersion of matter in turbulent flow through a pipe, *Proc. R. Soc. London.*, *Ser. A*, *223*, 446-468.

Thackson, E. L., and K. B. Schnelle (1970), Predicting effects of dead zones on stream mixing, *Journal of Environmental Engineering*, *96*, 319-331.

Thompson, J. R., H. R. Sorenson, H. Gavin, et al. (2004), Application of the coupled MIKE SHE/MIKE 11 modelling system to a lowland wet grassland in southeast England, *Journal of Hydrology*, *293*(1-4), 151-179.

Tipping, E., C. Woof, and K. Clarke (1993), Deposition and Resuspension of Fine Particles in a Riverine Dead Zone, *Hydrological Processes*, *7*(3), 263-277.

Twarakavi, N. K. C., J. Simunek, and S. Seo (2008), Evaluating interactions between groundwater and vadose zone using the HYDRUS-based flow package for MODFLOW, *Vadose Zone J*, *7*(2), 757-768.

U.S.EPA (2000), National Primary Drinking Water Rgulations: Ground Water Rule; Proposed Rules. 40 CFR Parts 141 and 142, 30194-30274 pp.

Uzarski, D. G, C. A. Stricker, T. M. Burton, et al. (2004), The importance of hyporheic sediment respiration in several mid-order Michigan rivers: comparison between methods in estimates of lotic metabolism, *Hydrobiologia*, *518*(1-3), 47-57.

van Dam, J. C., and R. A. Feddes (2000), Numerical simulation of infiltration, evaporation and shallow groundwater levels with the Richards equation, *Journal of Hydrology*, *233*(1-4), 72-85.

van Dam, J. C., P. Groenendijk, R. F. A. Hendriks, et al. (2008), Advances of modeling water flow in variably saturated soils with SWAP, *Vadose Zone J*, *7*(2), 640-653.

van Genuchten, M. T. (1980), A Closed-Form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils, *Soil Sci Soc Am J*, *44*(5), 892-898.

van Griensven, A., and W. Bauwens (2003), Multiobjective autocalibration for semidistributed water quality models, *Water Resour. Res.*, *39*(12).

van Griensven, A., T. Meixner, S. Grunwald, et al. (2006), A global sensitivity analysis tool for the parameters of multi-variable catchment models, *Journal of Hydrology*, *324*, 10-23.

VanderKwaak, J. E., and K. Loague (2001), Hydrologic-response simulations for the R-5 catchment with a comprehensive physics-based model, *Water Resources Research*, *37*(4), 999-1013.

Vauclin, M., D. Khanji, and G Vachaud (1979), Experimental and Numerical Study of a Transient, 2-Dimensional Unsaturated-Saturated Water Table Recharge Problem, *Water Resources Research*, *15*(5), 1089-1101.

Venetis, C. (1969), A study of the recession of unconfined aquifers, *Bull. Int. Assoc. Sci. Hydrol.* , *14*(4), 119-125.

Vitousek, P. M., H. A. Mooney, J. Lubchenco, et al. (1997), Human domination of Earth's ecosystems, *Science*, *277*(5325), 494-499.

Vrugt, J. A., C. G. H. Diks, H. V. Gupta, et al. (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resources Research*, *41*(1), -.

Yang, J., P. Reichert, K. C. Abbaspour, et al. (2008), Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China, *Journal of Hydrology*, *358*(1-2), 1-23.

Zheng, C. M. (2002), *Applied Contaminant Transport Modeling*, Wiley-Interscience.