Physically Based Modelling of the Impacts of Climate Change on Streamflow Regime

Nazmus Shams Sazib
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd
Part of the Civil and Environmental Engineering Commons

Recommended Citation
Sazib, Nazmus Shams, "Physically Based Modelling of the Impacts of Climate Change on Streamflow Regime" (2016). All Graduate Theses and Dissertations. 5067.
https://digitalcommons.usu.edu/etd/5067

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact dylan.burns@usu.edu.
PHYSICALLY BASED MODELING OF THE IMPACTS OF CLIMATE CHANGE
ON STREAMFLOW REGIME

by

Nazmus Shams Sazib

A thesis submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Civil and Environmental Engineering

Approved:

David G. Tarboton
Major Professor

Jagath J. Kaluarachchi
Committee Member

David Stevens
Committee Member

Jeffery S. Horsburgh
Committee Member

Jiming Jin
Committee Member

Mark McLellan
Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2016
ABSTRACT

Physically Based Modeling of the Impacts of Climate Change on Streamflow Regime

by

Nazmus Shams Sazib, Doctor of Philosophy

Utah State University, 2016

Major Professor: David G. Tarboton
Department: Civil and Environmental Engineering

Understanding the implications of climate change on streamflow regime is complex as changes in climate vary over space and time. However, a better understanding of the impact of climate change is required for identifying how stream ecosystems vulnerable to these changes, and ultimately to guide the development of robust strategies for reducing risk in the face of changing climatic conditions. Here I used physically based hydrologic modeling to improve understanding of how climate change may impact streamflow regimes and advance some of the cyberinfrastructure and GIS methodologies that support physically based hydrologic modeling by: (1) using a physically based model to examine the potential effects of climate change on ecologically relevant aspects of streamflow regime, (2) developing data services in support of input data preparation for physically based distributed hydrologic models, and (3) enhancing terrain analysis algorithms to support rapid watershed delineation over large area.

TOPNET, a physically based hydrologic model was applied over eight watersheds across the U.S to assess the sensitivity and changes of the streamflow regime due to climate
change. Distributed hydrologic models require diverse geospatial and time series inputs, the acquisition and preparation of which are labor intensive and difficult to reproduce. I developed web services to automate the input data preparation steps for a physically based distributed hydrological model to enable water scientist to spend less time processing input data. This input includes terrain analysis and watershed delineation over a large area. However, limitations of current terrain analysis tools are (1) some support only a limited set of specific raster and vector data formats, and (2) all that we know of require data to be in a projected coordinate system. I enhanced terrain analysis algorithms to extend their generality and support rapid, web-based watershed delineation services. Climate change studies help to improve the scientific foundation for conducting climate change impacts assessments, thus building the capacity of the water management community to understand and respond to climate change. Web-based data services and enhancements to terrain analysis algorithms to support rapid watershed delineation will impact a diverse community of researchers involved terrain analysis, hydrologic and environmental modeling.

(160 pages)
PUBLIC ABSTRACT

Physically Based Modeling of the Impacts of Climate Change on Streamflow Regime

Natzmus Shams Sazib

Climate change has significant impact on streamflow regime, but general understanding of how climate change will affect the characteristics of streamflow regime is lacking. The need to evaluate the potential impacts of climate change on stream ecosystems makes it important to study how streamflow regime may change. This dissertation used a physically based hydrologic model to quantify the impact of climate change on streamflow regime. Distributed hydrologic models require diverse geospatial and time series data as inputs - e.g., topography, geology, soil, land use, and land cover. The heterogeneity in those data makes it difficult to acquire, manipulate, manage and reproduce. Therefore, physically based models, like the model used currently require an inordinate amount of time to process input data. This work addressed this problem by developing web services that automate the preparation of model input data and use standard formats for these data to facilitate analysis. Distributed hydrological models also require that the drainage structure of the watershed and spatial variability of subwatershed physical properties be quantified. Terrain analysis using a digital elevation model supports the delineation of a stream network and subwatersheds draining to each stream network reach. However, current terrain analysis tools have limitations in terms of
data type, size and inability to use a geographic coordinate system. This study reports on enhancements to terrain analysis algorithms developed to extend their generality and support rapid, web-based watershed delineation services. The physically based hydrologic model used in this study projected changes in streamflow regime under a changed climate that provide a basis for considering the implications of these changes for stream ecosystem biodiversity and formulating approaches to protect ecosystems that may be subject to change. Enhancements of terrain analysis algorithms to support rapid watershed delineation enable the sort of physically based modeling performed here and will impact a diverse community of researchers involved with terrain analysis, hydrologic and environmental modeling.
ACKNOWLEDGMENTS

At first, I am very grateful to Allah for giving me strength, patience and time in order to accomplish this achievement. I would like to express my greatest appreciation to Dr. David Tarboton for his support, assistance, and encouragement, throughout the period of my study. I am very happy to work with him for his continuous guidance, patience with me, time and energy invested in making me a good researcher. I also appreciate committee members Dr. Jagath J. Kaluarachchi, Dr. David Stevens, Dr. Jeffery S. Horsburgh, and Dr. Jiming Jim for their suggestions and guidance.

I would like to thank, my mother, sisters, brother, brother-in-law who gave me constant love and support to keep me concentrated and motivated on my research work and study. I want to give special thanks to my wife as I could not have completed without her love, patience and relentless support. Sincere thanks to Avirup Sen Gupta, Sonia Morshed, Tian Gan, Tseganeh Gichamo, Madeline Merck, Pabitra Dash and all of my lab and Bangladeshi friends for their company and help.

This work was supported by grant #R-828637-01 from the US EPA – Science to Achieve Results (STAR) program, William Penn Foundation grant number 103-14 to Stroud Water Research Center for the development of Model My Watershed functionality, and the US Army Corps of Engineers contract number W91238-15-P-0033. I would also thank the Utah Water Research Laboratory for providing me with financial aid for my degree and attending conferences.

Nazmus Shams Sazib
# CONTENTS

ABSTRACT ............................................................................................................................ iii
PUBLIC ABSTRACT ................................................................................................................ v
ACKNOWLEDGMENTS ............................................................................................................. vii
LIST OF TABLES .................................................................................................................. x
LIST OF FIGURES ................................................................................................................ xi

## CHAPTER

1 INTRODUCTION .................................................................................................................. 1

1.1 Problem Statement ........................................................................................................... 1
1.2 Objectives ......................................................................................................................... 4
1.3 Literature Review ............................................................................................................. 6
1.4 Contribution ...................................................................................................................... 13
References ........................................................................................................................... 14

2 USING A PHYSICALLY BASED MODEL TO EXAMINE THE POTENTIAL EFFECTS OF CLIMATE CHANGE ON STREAMFLOW REGIME ........................................... 18

Abstract ............................................................................................................................... 18
2.1 Introduction ....................................................................................................................... 19
2.2 Study Sites ......................................................................................................................... 23
2.3 TOPNET Model ............................................................................................................... 25
2.4 Weather Data ................................................................................................................... 29
2.5 Model Setup and Calibration ........................................................................................... 30
2.6 Results ............................................................................................................................... 33

   2.6.1 Calibration and Validation Results ........................................................................... 33
   2.6.2 Precipitation Sensitivity ......................................................................................... 35
   2.6.3 Temperature Sensitivity ........................................................................................... 35
   2.6.4 GCM Temperature Change for 2076-2095 ............................................................... 36
   2.6.5 GCM Precipitation Change for 2076-2095 ............................................................... 36
   2.6.6 Streamflow Regime Variable Changes for 2076–2095 ........................................... 37

2.7 Discussion ......................................................................................................................... 38
2.8 Conclusion ......................................................................................................................... 41
Tables and Figures ................................................................................................................ 43
References ........................................................................................................................... 56

3 DATA SERVICES IN SUPPORT OF INPUT DATA PREPARATION FOR PHYSICALLY BASED HYDROLOGIC MODELING ........................................................................... 62

Abstract ............................................................................................................................... 62
3.1 Introduction ....................................................................................................................... 63
3.2 TOPNET Model ............................................................................................................... 66
3.3 Why Automation of Input Data Preparation is Required ............................................... 67
3.4 Methods.......................................................................................................................... 68
  3.4.1 Functionality Design.................................................................................................. 68
  3.4.2 System Design ......................................................................................................... 69
3.5 Implementation of Hydro-DS Data Services ......................................................... 71
  3.5.1 EHV Services........................................................................................................... 71
  3.5.2 TOPNET Model Specific Services ........................................................................ 73
3.6 Data Service Implementation and Example Application ................................... 75
  3.6.1 Study Site ................................................................................................................ 76
  3.6.2 Input Data Pre-processing Using Data Services ..................................................... 77
3.7 Discussion and Conclusion ...................................................................................... 80
Tables and Figures .......................................................................................................... 83
References .......................................................................................................................... 92

4 ENHANCEMENTS TO HYDROLOGIC DIGITAL ELEVATION MODEL
ANALYSIS SOFTWARE TO SUPPORT RAPID WATERSHED DELINEATION..... 96

  Abstract ............................................................................................................................. 96
  4.1 Introduction .................................................................................................................. 98
  4.2 Enhancement to TauDEM and Rapid Watershed Delineation Methodology . 101
    4.2.1 Enhancements to TauDEM .................................................................................. 101
    4.2.2 Rapid Watershed Delineation (RWD) ................................................................. 105
    4.2.3 RWD Algorithm .................................................................................................. 109
  4.3 Software Implementation ........................................................................................... 111
    4.3.1 Implementation of TauDEM Enhancements ....................................................... 111
    4.3.2 RWD Implementation on Delaware River Basin............................................. 113
    4.3.3 Web based Rapid Watershed Delineation ....................................................... 114
  4.4 Discussion and Conclusion ...................................................................................... 114
Tables and Figures .......................................................................................................... 117
References .......................................................................................................................... 125

5 SUMMARY, CONCLUSIONS, BROADER IMPACTS AND
RECOMMENDATIONS ........................................................................................................ 127
  5.1 Summary and Conclusions ....................................................................................... 128
  5.2 Broader Impacts ......................................................................................................... 131
  5.3 Recommendations for Future Research .................................................................. 133

APPENDICES .................................................................................................................... 137

  Appendix A ....................................................................................................................... 138
  Appendix B ....................................................................................................................... 142
  Curriculum Vitae ............................................................................................................ 145
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Variables used to characterize streamflow regime.</td>
<td>43</td>
</tr>
<tr>
<td>2.2 Physical attributes of study watersheds selected to represent each streamflow regime class.</td>
<td>44</td>
</tr>
<tr>
<td>2.3 Details of the climate models and emission scenarios used for future climate projections.</td>
<td>45</td>
</tr>
<tr>
<td>2.4 TOPNET model parameters.</td>
<td>45</td>
</tr>
<tr>
<td>2.5 NSE and Bias for the calibration and validation period for eight watersheds.</td>
<td>46</td>
</tr>
<tr>
<td>3.1 HydroDS Essential Hydrologic variable data service functions</td>
<td>83</td>
</tr>
<tr>
<td>3.2 HydroDS TOPNET model specific service functions</td>
<td>84</td>
</tr>
<tr>
<td>3.3 TOPNET model parameters</td>
<td>85</td>
</tr>
<tr>
<td>4.1 Watershed attributes for rapid watershed delineation.</td>
<td>117</td>
</tr>
<tr>
<td>4.2 Inputs required for rapid watershed delineation.</td>
<td>118</td>
</tr>
<tr>
<td>4.3 Comparison of computation time required for rapid watershed delineation using RWD and traditional method.</td>
<td>118</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Spatial distribution of eight streamflow regime classes across the US.</td>
<td>47</td>
</tr>
<tr>
<td>2.2 Daily flow pattern for the representative stream in each class (i.e., the stream closest to)</td>
<td>48</td>
</tr>
<tr>
<td>2.3 TOPNET model schematic.</td>
<td>49</td>
</tr>
<tr>
<td>2.4 Comparison of observed and modeled (Daymet driven) streamflow variables for the climatological period of 1986-2005 for the eight selected watersheds.</td>
<td>50</td>
</tr>
<tr>
<td>2.5 Percentage changes of streamflow regime variables due to precipitation change only.</td>
<td>51</td>
</tr>
<tr>
<td>2.6 Percentage changes of streamflow regime variables due to temperature change only.</td>
<td>52</td>
</tr>
<tr>
<td>2.7 Box plots showing the annual and seasonal temperature changes for the lowest RCP 2.6 (white) and highest RCP 8.5 (gray) emission scenario for the period of 2076-95 compared to 1986-2005.</td>
<td>53</td>
</tr>
<tr>
<td>2.8 Box plots showing the annual and seasonal precipitation changes for the lowest RCP 2.6 (white) and highest RCP 8.5 (gray) emission scenario for the period of 2076-95 compared to 1986-2005.</td>
<td>54</td>
</tr>
<tr>
<td>2.9 Changes in streamflow regime variables due to climate changes at the end of the century for the lowest (RCP 2.6) and highest (RCP 8.5) emission scenarios.</td>
<td>55</td>
</tr>
<tr>
<td>3.1 Input data pre-processing steps for TOPNET model.</td>
<td>86</td>
</tr>
<tr>
<td>3.2 Architecture of HydroDS data services.</td>
<td>87</td>
</tr>
<tr>
<td>3.3 Watershed delineation processing using TauDEM.</td>
<td>87</td>
</tr>
<tr>
<td>3.4 Steps required for retrieving soil data from the SSURGO soil data base.</td>
<td>88</td>
</tr>
<tr>
<td>3.5 Daymet climate data retrieving and processing steps.</td>
<td>88</td>
</tr>
<tr>
<td>3.6 Streamflow data downloading steps.</td>
<td>88</td>
</tr>
<tr>
<td>3.7 Study site location and DEM of the Logan River Watershed.</td>
<td>89</td>
</tr>
<tr>
<td>3.8 TOPNET data preparation steps using HyrdoDS data services.</td>
<td>89</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>3.9 Morris sensitivity analysis of the TOPNET model parameters.</td>
<td>90</td>
</tr>
<tr>
<td>3.10 Comparison of simulated and observed streamflow for the Logan River Watershed.</td>
<td>91</td>
</tr>
<tr>
<td>4.1 Ellipsoid geometry of the Earth.</td>
<td>119</td>
</tr>
<tr>
<td>4.2 Data preprocessing steps for rapid watershed delineation.</td>
<td>120</td>
</tr>
<tr>
<td>4.3 Data Structure for Rapid Watershed Delineation.</td>
<td>121</td>
</tr>
<tr>
<td>4.4 Algorithm of Rapid Watershed Delineation.</td>
<td>121</td>
</tr>
<tr>
<td>4.5 Landing Page for RWD</td>
<td>122</td>
</tr>
<tr>
<td>4.6 User selected point for RWD</td>
<td>123</td>
</tr>
<tr>
<td>4.7 Watershed delineated for the clicked point</td>
<td>124</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Problem Statement

Streamflow regime variables are important to stream ecosystems, which provide many services to the environment and mankind. In addition to providing clean water for consumption and agriculture, stream ecosystems sustain biodiversity and provide support for basic ecological processes as well as important economic activities, including fisheries and recreation (Chen et al., 2013). Increased air temperature and changes in precipitation patterns are significant components of climate change that have the potential to alter the timing, pattern, and magnitude of streamflow that constitute the streamflow regime. Since streams comprise an ecosystem environment, there is the potential for ecosystem changes as streamflow regimes are altered. Potential impacts of climate change in the U.S. include decreases in snow depth, shifts in the timing of streamflow and more frequent droughts (Hayhoe et al., 2007; Fritze et al., 2011) Other possible impacts (Zhang et al., 2009) include changes in extreme high- and low-flow events and periods, increases in flooding and shifts in aquatic system process and functions (Johnson et al., 2012; Tohver et al., 2014). Assessment of changing streamflow regime variables due to climate change is motivated by the fact that water quality, quantity, and the ecological integrity of rivers are influenced not only by the magnitude but also timing, duration, frequency, and rate of change of streamflow. This assessment could be done by using statistical or physically based modeling approaches. Statistical modeling approaches use historical data to establish relationships and forecast the potential impact...
of change on the quantity of interest. These approaches are limited by the statistical relationships they depend on, which lose validity beyond the range of data where they have been fit. On the other hand, physically based modeling approaches use physical principles that are part of a model to assess change. Physically based models can thus provide a better understanding of climate impacts where potential future conditions may be outside the ranges experienced historically. In this study, physically based modeling was used to assess the changes in streamflow variables associated with climate change.

Physically based models require different types of geospatial data (e.g., geology, soil, land use, and land cover) and time series data (precipitation, temperature, etc.) to setup and run. Hydrologists obtain data from different sources such as United States Department of Agriculture (USDA) for geospatial data and the National Oceanic and Atmospheric Administration (NOAA) for climate data. However, the number and variety of available data sources make it difficult to quickly locate the most appropriate datasets for a specific model. Additionally, different data sources store data in different formats such as plain text, database, and markup language encoded files. Therefore, users need to learn specific and potentially different procedures for obtaining data from each source. Moreover, obtained data typically require additional preparation steps (e.g., regridding, clipping, reprojection, filling missing values, etc.) to convert to the form required by the model. As a result, researchers and practitioners spend a significant amount of time on basic data gathering and transformation, instead of scientific analysis and decision making. This work also addressed these issues by demonstrating how input data
preparation can be automated using web services to facilitate the setup of a physically based hydrologic model. Watersheds are widely recognized as the basic functional unit for hydrologic modeling. The emphasis on watershed approaches to answer water resource related questions has led to increased demand for watershed delineations and information derived from them. Furthermore, many watershed studies are now done at regional scales, requiring quick derivation of stream networks, watershed boundaries, and terrain characteristics at a large number of locations, spread across large areas (Chinnayakanahalli, 2010). Nevertheless, delineating a watershed spread across a large region is still cumbersome due to the processing burden of working with large Digital Elevation Models (DEMs). Terrain analysis tools, such as Terrain Analysis Using Digital Elevation Models (TauDEM) (Tarboton, 2002) support the delineation of watersheds and stream networks from within desktop Geographic Information Systems (GIS). However, common limitations of terrain analysis tools are: (1) they support only specific types of raster data; (2) they are time consuming for large watersheds and analyses may be limited due to memory limitations of the computer being used; and (3) input data have to be in a projected coordinate system. It is common practice in ArcGIS to project data from datasets such as the National Elevation dataset into a local projected coordinate system prior to hydrologic analysis. This has limitations as it alters (smooths, and sometimes introduces artificial stripes) the data and results in distortions to the data when large areas are processed (e.g., continents). To address these shortcomings, terrain analysis
capabilities were advanced through enhancements to the TauDEM algorithms and software package.

1.2 Objectives

Understanding the implications of climate change on streamflow regime is complex as (1) changes in climate will vary over space and time and (2) the hydrologic response to climate change will be further influenced by the attributes of specific watersheds, including land use, pollutant sources, and human use and management of water. However, better understanding impacts of climate change is required for identifying how we are vulnerable to these changes, and ultimately to guide the development of robust strategies for reducing risk in the face of changing climatic conditions. The first objective of this study was to improve our understanding of the impact of climate change on streamflow regime. We addressed the following questions by applying a physically based hydrologic model at different watershed across the U.S.

Questions:

1. How sensitive is streamflow regime to changes in climate inputs at the scale typical of global change projections?

2. For specific climate change projections over the next century, how are streamflow regime variables expected to change?

Physically based models used for quantifying the impact of climate change on streamflow regime require land use, streamflow, terrain, water quality, soils, meteorology and climate projection information. The volume of this information increases as the spatial resolution and extent of study areas increases. In addition to the diversity and
abundance of these data, they also need to be pre-processed and converted into forms that models can use directly. At the present time, most researchers perform such manipulation manually, which is time consuming, error prone, and difficult to reproduce. Therefore, the second objective of this study was to develop web based data services for accessing and preprocessing hydrologic model input data to set up a hydrologic model. This addressed following question:

Questions:

1. How can an instance of a physically based model be setup and populated with input data to address scientific problems without extensive computational and algorithmic knowledge and experience?

Physically based, distributed models require data that define the detailed drainage structure of the watershed to account for the spatial variability of physical properties. Terrain analysis algorithms support the delineation of watersheds and detailed drainage information from DEMs. To overcome current limitations in terrain analysis algorithms, the third objective of this study was to enhance TauDEM to increase its usability in terms of data type, size, and coordinate systems and to develop capability to organize pre-processed terrain data into a subwatershed based data structure to support rapid delineation of watersheds and derivation of watershed attributes over large areas. We addressed the following question for the third objective.

Questions:
1. How can existing terrain analysis algorithms be enhanced to accommodate geographic coordinate input data, support rapid watershed delineation, and support deployment via web interfaces to enhance accessibility?

1.3 Literature Review

Climate change can occur naturally or as a result of human activities (Houghton et al., 1990). The IPCC (Houghton et al., 2001) provided strong evidence that most of the observed warming in climate is attributable to human activities and is not just natural climate variability. Climate change has the potential to alter the streamflow regime which is important to stream ecosystems. Alteration of the streamflow regime due either to human activities or climate change, may affect the aquatic organisms present due to change in flow of energy and sediments in streams (Gibson et al. 2005; Poff et al. 1997). The evidence of ecological responses to climate change, although at an early stage, is visible, and changes have been observed in the distribution of species as well as compositions of and interactions with communities (Walther et al., 2002). Hence, it is necessary to understand the response of streamflow regime to climate change.

Many global and regional studies have examined the impact of projected future climatic change on hydrologic variables. Milly (2008) focus on streamflow and water availability trends and find that an ensemble of twelve current climate models accurately accounts for twentieth-century changes. The same models project potentially crucial regional effects on streamflow in the future that could threaten the availability of freshwater in many regions of the world by the year 2050. Lehner et al. (2006) outlined a continental, integrated approach to analyzing the impacts of global change on future
flood and drought frequencies on a pan-European scale. Ficklin et al. (2013) evaluated the climate change impact on water resources within agriculture systems of the San Joaquin watershed, California (USA) using a semi-distributed hydrological model. Results of the study implied that changes in carbon dioxide (CO2) and climatic parameters (temperature, precipitation) significantly affect the water yield, evapotranspiration and other components of hydrological cycle. Arnell and Gosling (2013) assessed the impacts of climate change on a series of indicators of hydrological regimes across the global domain. They found that most climate models project increase in runoff in Canada and high latitude eastern Europe and Siberia and decrease in runoff in central Europe, around the Mediterranean, in central America and Brazil.

Tohver et al. (2014) used a physically based hydrologic model to examine the nature of changing hydrologic extremes (floods and low flows) for about 300 river locations in the Pacific Northwest. They found that the combination of warming and shifts in seasonal precipitation regimes resulted in projections of increased flooding and more intense low flows for most of the basins. Hydrological impacts on the Upper Grande River Basin (UGRB) under A1B climate change scenario were assessed using the Larvras Simulation of Hydrology (LASH) model (Viola et al., 2015). Their results showed a reduction of the streamflow between 2011-2040 but after 2041 until the end of the century results showed a considerable increase in the annual mean streamflow. The projected extension of the dry period and consequence of the declining streamflow may have caused a reduction in the water availability in the watershed of the region. Most of the above mentioned studies have focused on changes to daily streamflow, floods and
drought frequencies and not on changes to stream flow regime which refers to the pattern of streamflow variability and is important to stream ecology.

There are a number of studies that have looked at the impact of climate change on streamflow regime. Gibson et al. (Gibson et al., 2005) applied the Precipitation-Runoff Modeling System (PRMS) to the Cle Elum River, Washington, and the Chattahoochee–Apalachicola River in Georgia, Alabama and Florida to examine the impacts of future climate scenarios on flow regimes and how predicted changes might affect river ecosystems. Statistically downscaled temperatures and precipitations from one climate model were used as input to the two hydrological models. Cherkauer and Sinha (2010) quantified the impact of climate change on nine stream flow regime variables in the Lake Michigan Region. They found that the impact of regional climate change projections for early (water years 2010-2013) and midcentury (the water year 2040-2069) streamflow was highly variable. Dhungel et al. (2016) classified 601 reference condition streams in the contiguous U.S. using principal component factors from 16 streamflow regime variables selected for their relevance to stream ecology. This study predicted the changes in these streams by the end of the 21st century by using a statistical model. Downscaled climate projections of precipitation and temperature were used to predict the changes in these stream classes by 2100 using the Random Forest (RF) model (Breiman, 2001).

The above mentioned studies pre-dated the most recent ensemble of climate model projections referred to as the Coupled Model Intercomparison Project phase 5 (CMIP5) now available. CMIP5 provides a standard set of model simulations and projections of future climate change for both near term (through 2035) and long term (out
to 2100 and beyond) scales. Compared to earlier studies, the updated CMIP5 models produce higher resolution projections and use an updated set of greenhouse gas emission scenarios. There is therefore the need to evaluate the impact of these projections on streamflow regime.

Physically based, distributed hydrologic models are useful for answering science and management questions related to hydrological processes, the impact of climate and land use changes, as well as managing and supporting water resources. Distributed hydrologic models require different types of geospatial data (e.g., geology, soil, land use, and land cover) and time series data (precipitation, temperature, etc.) to setup and run.

For a small watershed composed of few model elements, this may not be too difficult, but for a large watershed, structuring and providing data for every model element can become a daunting task. Because of the inherent heterogeneity in geospatial data sets, the process of assigning parameters to a large number of model elements is an error prone and time intensive step (Bhatt et al., 2008). There have been a number of studies that deal with the input data preparation steps available for desktop model applications. PHIM GIS (Bhatt et al., 2008; Kumar et al., 2009; Bhatt et al., 2014) is one example where a GIS framework was tightly coupled to the Penn State Integrated Hydrologic Modelling System (PIHM, http://www.pihm.psu.edu/), and where model input pre-processing and input and output visualization are carried out. EcohydroLib (Miles, 2014) is another example of a software library that provides extensible tools for acquiring geospatial data directly from web services or local data sources for the preparation of model inputs for any watershed. It facilitates the rapid development of model inputs, easing comparisons
of study sites and model structures. Other examples include the Water Erosion Prediction Project (WEPP) interface on the Geographic Resources Analysis Support System (GRASS) (Engel et al., 1993); the interface between ARC/INFO and the Hydrologic Engineering Center (HEC) modelling system (Hellweger and Maidment, 1999); and ArcGIS-SWAT as an extension to ArcMap (Olivera et al., 2006). However, most of these software systems are desktop based and require installation, sometimes licenses, and are platform dependent and not be available for alternative platforms (Mac versus Windows versus Linux).

Current research is focused on extending these data analyses tool to web based services. Cyber GIS (Wang, 2013) is a project to advance approaches for performing spatial analysis over the Internet without the complexity of managing and installing the underlying software, hardware, and web server capabilities. HydroDesktop, the Consortium of Universities for the Advancement of Hydrologic Science Inc (CUAHSI) Hydrologic Information System (HIS) data access client provides a platform for time series data download, extraction and analysis by enabling map based discovery and retrieval of hydrologic time series with a specified domain of interest (Ames, 2012). However, HydroDesktop does not provide any spatial data, such as, for example, soil and land cover data required by many hydrologic models. HydroTerre (Leonard and Duffy, 2013) provides both spatial and time series data, which can be downloaded and used in hydrologic models, primarily targeting, but not limited to PIHM. Yet, downloaded data need to be pre-processed to use directly in the model, which requires significant time and effort. Tarboton et al. (2014) developed a web based system called HydroShare for data
sharing, analysis, and modelling in order to understand hydrology. It provided a community collaboration site that enables users to easily discover and access data and models, retrieve them to a desktop computer, or perform analyses in a distributed computing environment that includes grid, cloud, or high performance computing model instances as necessary.

Watersheds are widely accepted as the basic functional unit for water resources management studies and play an important role in defining the scope of the modeling domain, thereby impacting all further analysis and modeling steps. Watersheds are also used to identify the relationship among watershed attributes and streamflow regime variables (Kroll et al., 2004). Ecological studies often use watersheds as the basic unit for quantifying the effects of geomorphological, geological, and hydrological characteristics on structure and function of aquatic ecosystems (Sanz and Del Jalón, 2005).

Digital elevation models (DEM) represent the terrain from which watersheds can be derived automatically using GIS technology. The techniques for automated watershed delineation are becoming more prevalent due to increases in desktop computing power and GIS capabilities that allow complex operations involved in watershed delineation to be performed locally and fast.

There are several desktop tools available for automated watershed delineation from DEMs such as ArcGIS, BASINS (EPA, 2015), and TauDEM. ArcGIS has a large following within the water resources community. However, LIDAR data and stream networks can be too large to work with within the automated watershed delineation tool in ArcGIS for such a large area because processing is time consuming. BASINS,
developed by the EPA, provides an easy to use, automated watershed delineation tool. The automated watershed delineation tool allows simplified watershed delineation with limited user input (BASINS 4.0 User’s Manual). TauDEM is the terrain processing tool embedded in BASINS to prepare the DEM for watershed delineation. This process includes pit removal, computation of flow direction and slopes, calculating contributing area, channel network delineation, and subwatershed delineation with stream segment attributes (Tarboton, 2003).

Web-based, real-time GIS watershed delineation is the next step in the progression. The U.S. Geological Survey has developed a nationwide program called ‘Streamstats’ (Ries et al., 2009) for providing researchers with streamflow, physical and chemical characteristics at regional scale. Streamstats is a web based program that can provide commonly used streamflow measures at gauged and ungauged sites; it can also delineate watersheds and provide other useful watershed attributes at a user specified location (http://water.usgs.gov/osw/streamstats/). Environmental Systems Research Institute (ESRI) developed a web service to delineate watersheds by identifying a location users are interested in finding a watershed for, and whether users want to snap to the location of the nearest stream (http://resources.arcgis.com/en/help/). This web service used the 30m National Elevation Dataset (NED) for the continental United States (USGS, 2010-2011), and the 90m HydroSHEDS for the world. The CyberGIS-based TauDEM application is another example of a web based service that provides an online environment for computing- and data-intensive watershed delineation and hydrological analyses on advanced cyberinfrastructure (Fan, 2014). However, common limitations of
TauDEM are that it does not support multiple types of raster files (e.g., ESRI grid, img, etc.) and it does not support data that use a geographic coordinate system.

1.4 Contribution

This work was driven by the need for better understanding of the impact of climate change on stream flow regime. This dissertation also contributes towards advancing some of the cyberinfrastructure and GIS methodologies that support physically based hydrologic modeling. Our work included: (1) using a physically based model to quantify the impact of climate change on streamflow variables, (2) developing web based data services for automation of input data preparation of hydrologic model, and (3) enhancements of terrain analysis algorithm to support rapid watershed delineation. Each of these contributions is detailed in a chapter (paper) that follows. The impact of climate change on streamflow variables examined using a distributed, physically based hydrologic model is described in Chapter 2. Chapter 3 describes the web-based data services developed to set up a physically based hydrologic model entirely using server (cloud) functionality without the use of data or software on the user's desktop computer. Enhancements to terrain analysis algorithms and methods for web based rapid watershed delineation is presented in Chapter 4.
References


Chinnayakanahalli, K.J., 2010. Characterizing Ecologically Relevant Variations in Streamflow Regimes, Utah State University, Loga, UT.


Ficklin, D.L., Stewart, I.T., Maurer, E.P., 2013. Climate change impacts on streamflow and subbasin-scale hydrology in the Upper Colorado River Basin. PloS one, 8(8).


CHAPTER 2

USING A PHYSICALLY BASED MODEL TO EXAMINE THE POTENTIAL EFFECTS OF CLIMATE CHANGE ON STREAMFLOW REGIME

Abstract

Climate change will alter streamflow, but how climate change will affect specific aspects of streamflow is not well understood. Changes in streamflow regime variables due to climate change can be assessed through statistical or physically based models. The validity of statistical models is limited to the data for which they were developed, whereas physically based models have some ability to accommodate the non-stationary dynamics associated with climate change. We used TOPNET, a physically based, semi-distributed hydrologic model, to model 16 variables that quantify the streamflow regime for eight watersheds that represented the range of streamflow regimes occurring within the continental USA. TOPNET inputs were derived from Daymet climate data, Soil Survey Geographic database (SSURGO) data, and National Land Cover (NLCD) data. United States Geological Survey (USGS) streamflow data were used to calibrate and validate each model for a separate period of record. The models were then driven by downscaled future climate data to project future (2076–95) changes in the 16 streamflow regime variables. For most of the watersheds, streamflow regime variables derived from calibrated flow compared well with those from the observed flow. Magnitude-associated flow variables tended to increase with future climate change. For example, in watersheds...
in which precipitation was predicted to increase, the number of high-flow events also increased. Models also predicted change in the timing of flows. The hydrographs for streams that are currently snow-fed were predicted to shift to earlier in the water year, while streams that are currently rainfall-driven were not predicted to change much. These predictions of changes in streamflow regime variables due to climate change improve our understanding of how climate change may alter streamflow and hence stream ecosystems. Such considerations are important for water management and mitigation of the impacts of climate change on ecosystems.

2.1 Introduction

It is generally accepted that the increasing atmospheric concentration of greenhouse gases will result in climate change. Regionally, variable changes in the amount and intensity of precipitation have been observed in much of the United States (Groisman et al., 2012). Climate modeling experiments suggest these trends will continue throughout the 21st century, with continued warming accompanied by a general intensification of the global hydrologic cycle (Karl, 2009; Kharin et al., 2013). Increased temperature and changes in magnitude, timing, and form of precipitation are significant components of climate change that have the potential to alter the timing and magnitude of streamflow. Potential impacts of climate change in the U.S. include decreases in snow depth, shifts in the timing of streamflow, and more frequent droughts (Hayhoe et al., 2007; Fritze et al., 2011). Other possible alterations include changes in extreme high- and low-flow events and periods, increases in flooding, and shifts in aquatic ecosystem
processes and functions (Johnson et al., 2012; Petra and Hannes Müller, 2012; Tohver et al., 2014).

Understanding the implications of climate change on streamflow regime is complex as it varies over space and time, which challenges the stationarity assumption (Milly et al., 2008) on which standard water management practice has been based. The hydrologic response to climate change will also be influenced by the attributes of specific watersheds, including soils, land use, and management. A better understanding of the impact of climate change is required for identifying how aquatic ecosystems are vulnerable to these changes, and ultimately to guide the development of robust strategies for reducing risk in the face of changing climatic conditions. Such assessments could be approached using either empirical or physically based modeling. The empirical approach analyzes long-term, observed historical data and develops statistical models relating climate and watershed attributes to streamflow regimes. These models can then be used to estimate streamflow regime changes in response to climate change (Sanborn and Bledsoe, 2006; Dhungel et al., 2016). However, statistical models are fit using historical conditions, which may limit model predictive capability if future climate conditions differ from those observed in the historical record. In contrast, physically based models use equations that represent known physical processes to quantify hydrologic behavior and are, in theory, better suited to predictions where conditions differ from the range of historical observation.

Numerous studies have examined the potential impact of future climate change on streamflow (Lettenmaier et al., 1999; Salathé Jr, 2003; Jha et al., 2004; Barnett et al.,
Most of these studies focused on either the changes in water balance or the frequency of hydrological extremes. However, changes in the spatial distribution of flow regimes, which are potentially significant to aquatic ecology, are not typically estimated from large-scale hydrology or land surface models (Cherkauer and Sinha, 2010). Other studies (Melack et al., 1997; Mulholland et al., 1997; Gibson et al., 2005; Konrad and Booth, 2005; Dhungel et al., 2016) have examined how ecologically important hydrological and flow metrics may change for specific regions and species. For example, Dhungel et al. (2016) used a statistical model to predict changes in stream classification (derived from sixteen streamflow regime variables selected as relevant to stream ecology) by the end of the 21st century. They found thirty-three percent of the 601 sites were predicted to change to a different flow regime class by the year 2100. Snow-fed streams in the western United States (U.S.) were predicted to be less likely to change regimes, whereas both small, perennial, rain-fed streams and intermittent streams in the central and eastern U.S. were predicted to be most likely to change regime.

These prior studies have used statistical approaches based primarily on observed correlations. This makes them vulnerable to error in predicting changes due to climate change where conditions may be different from observations used in fitting the model. The assumption of a relationship of watershed and climate attributes with stream class or regime may not persist for changed climate conditions. Further, predictors used for developing the model may not be completely adequate for climate change applications. For instance, rainfall intensity, distribution, and temperature were excluded by Dhungel
et al. (2016) in developing their statistical model. This might result in a critical loss of information about projected changes in streamflow variables, which ultimately affects changes in flow class. These considerations motivate the use of physically based models to quantitatively assess the potential effects of climate change on streamflow regime.

In this study, we used a physically based model to examine projected changes in streamflow regime due to climate change. We specifically examined predictions from physically based models that we developed for variables used by Dhungel et al. (2016) at sites selected to be representative of the streamflow regime classes in the national scale classification developed by Dhungel et al. (2016). Our specific objectives were to (1) evaluate the performance of a physically based model for quantifying streamflow regime variables using historical data, (2) explore the sensitivity of different streamflow regime variables to changes in precipitation and temperature, and (3) quantify the changes in streamflow regime variables expected in response to climate change.

TOPNET, a physically based distributed hydrological model, was used to achieve those objectives. First, the model was set up for a set of study watersheds using climate, soil, land cover, and land use data. Observed streamflow data were used to calibrate the model at each site, and then validate it using a separate part of the period of record. The calibrated models were then run for a twenty-year baseline period (1986-2005) to estimate sixteen streamflow regime variables within each study watershed. These were compared with streamflow regime variables computed directly from the observed streamflow records to quantify the fidelity with which each of the models reproduced these streamflow regime variables. We quantified the sensitivity of the model to changes
in precipitation and temperature by separately imposing changes to the historic record of these inputs. However, climate changes are not expected to be so simple and uniform so we used bias-corrected and spatially downscaled future climate projections from five Global Climate Models (GCMs) to drive the calibrated TOPNET models for each site to assess the potential impact of climate change on streamflow regime variables ahead to the end of the 21st century compared to the baseline period.

2.2 Study Sites

Dhungel et al. (2016) assessed how climate change might affect eight classes of streamflow regimes that differed in terms of sixteen streamflow variables. Their classes were derived from streamflow data observed at 601 USGS stream gauging stations selected from Geospatial Attributes of Gages for Evaluating Streamflow (GAGES) (Falcone et al., 2010). These stations were selected because they were least affected by human influence on streamflow regime, which helps isolate the potential impacts of climate on natural flow patterns.

Dhungel et al. (2016) used long term daily streamflow data to characterize spatial variation in sixteen streamflow regime variables selected for their ecological importance. Dhungel et al. (2016) then performed principal component analysis with Varimax rotation (Jackson, 1991) to reduce the dimensionality to five uncorrelated streamflow factors. Examination of these factors revealed that they quantify low-flow, magnitude, flashiness, timing, and constancy aspects of the streamflow regime, respectively. Dhungel et al. (2016) then applied Ward’s hierarchical clustering (Ward JH, 1963) to the five streamflow factors to classify the 601 streams into eight streamflow classes. The eight
streamflow classes were characterized as (A1) Small Steady Perennial, (A2) Large Steady Perennial, (B1) Steady Intermittent, (B21) Early Intermittent, (B22) Late Intermittent, (C1) Early Flashy Perennial, (C21) Small Flashy Perennial, and (C22) Large Flashy Perennial streams (Figure 2.1). Daily flow statistics illustrate how the flow regime classes differed in terms of magnitude, timing, the number of reversals, and flashiness of daily flows (Figure 2.2). Class A1 and A2 watersheds have a generally sharp hydrograph peak towards the latter half of the water year and differ mostly by magnitude. Class B1, in comparison to classes B21 and B22, has a generally smoother hydrograph indicative of fewer flow reversals. Classes C1, C21, and C22 streams have smaller seasonal variability and generally flatter seasonal hydrograph peaks, but they have a degree of flashiness associated with perennial rain-fed streams. The early seasonal pattern of C1, relative to C21 and C22 is visible, as is the difference in magnitudes of flow between C21 and C22.

In this study, eight watersheds were selected (Figure 2.1), each one being representative of one of Dhungel's classes. The approach was to select the stream nearest the centroid of each stream class in principal component factor space. The total sum of squares (TSS) of the difference between individual stream and class mean for each factor was calculated for each stream in each class using Equation 1:

\[ TSS = \sum_{i=1}^{5} (X_i - \overline{X}_i)^2 \]  

where \( X_i \) represent the value of streamflow factors for each stream and \( \overline{X}_i \) represents the mean value of streamflow factor \( i \) for each class. Then, the stream with the minimum value of TSS is closest to the class centroid. This stream was taken as representative of
the class. Soil data for the stream closest to the centroid of the B21 class was not available; therefore, the next closest stream in this class was selected as representative. Table 2.2 lists the representative watersheds selected and their key watershed and climatologic properties for the baseline period of 1986-2005.

2.3 TOPNET Model

TOPNET (Bandaragoda et al., 2004; Ibbitt and Woods, 2004) is a distributed hydrologic model with basic model elements being topographically delineated subwatersheds that discharge into the stream network, which is then used to route flow to the outlet. It was selected to use for this study because of our familiarity with the code and because it represents the key physical processes involved in streamflow generation from rainfall and weather inputs at a level of detail sufficient to capture the physical sensitivities to changes in climate represented in daily weather inputs, while being simple enough to be calibrated and run sufficiently quickly for multiple study watersheds across the U.S. Other similar scale models (e.g., VIC (Xu et al., 1994) and SWAT (Arnold and Fohrer, 2005)) could have been used for this study and we expect would have produced similar results, though a comparison of multiple models was beyond our scope. Familiarity with the TOPNET code enabled us to customize data preparation and calibration workflows as needed. These capabilities were more important than the specific model used. Here we provide a detailed summary of TOPNET to give the reader information on the model representations of the physical processes involved.

Geographic Information System (GIS) tools can be used to estimate TOPNET model parameters from soil, vegetation, and topography information. The model is
calibrated using a multiplier approach that scales the overall magnitude of parameters while retaining the geographically derived spatial pattern of these input variables. Bandaragoda et al. (2004) applied TOPNET to the Distributed Model Intercomparison Project (DMIP) watersheds in South Central USA (Smith et al., 2004). Their study demonstrated that TOPNET was suitable for hydrologic prediction across the range of conditions in this study. Tarboton et al. (2007) applied TOPNET to the Nooksack River Watershed in Washington State. This work added the Utah Energy Balance snowmelt model to TOPNET, extending its capability to snowmelt dominated regions.

The TOPNET model was developed by coupling TOPMODEL (Beven and Kirkby, 1979; Beven, 1995a), which simulates flow in relatively small subwatersheds, with channel routing. This approach can model large watersheds by using smaller subwatersheds within the large watershed as model elements. The TOPNET model includes many enhancements beyond the original Beven and Kirkby TOPMODEL including: (1) calculation of reference evapotranspiration using the American Society of Civil Engineers (ASCE) standardized Penman-Monteith method (Walter and Brown, 2000; Allen, 2005) and (2) calculation of snowmelt using the Utah Energy Balance Snowmelt model (Tarboton, 1995a).

The physical processes (Figure 2.3) in the TOPNET model are represented by three major components: (1) Rainfall-Runoff Transformation, (2) Potential Evapotranspiration, and (3) Snow. The Rainfall-Runoff Transformation component of TOPNET uses TOPMODEL concepts for the representation of subsurface storage controlling the dynamics of the saturated contributing area and baseflow recession. A key
contribution of TOPMODEL is the parameterization of the soil moisture deficit (depth to water table) using a topographic index to model the dynamics of variable source areas contributing to saturation excess runoff. Within the Rainfall-Runoff Transformation component there are four sub-components: canopy interception storage, upper soil zone storage, groundwater saturated zone, and channel flow. Surface water input to the canopy interception store is comprised of rainfall and snowmelt. Potential evapotranspiration is first taken from the canopy interception storage, with any unsatisfied amount taken to act on the upper soil zone storage, with actual evapotranspiration based on soil moisture and vegetation factors. Throughfall is computed based upon canopy interception capacity surface water input, and water in canopy storage and is used as input to the upper soil zone storage. Based on the input and storage in the upper soil zone, recharge to groundwater and surface runoff is calculated. The upper soil zone calculation also accounts for potential upwelling from groundwater where the water table is shallow. The ground water saturated zone component calculates the local depth to water table, and the occurrence of saturation from below where the water table has risen to the surface using wetness index classes. Surface water input becomes saturation excess runoff in these saturated locations. Saturated zone calculations account for recharge, upwelling, and groundwater pumping and produce baseflow as an output. Baseflow and surface runoff from the upper soil zone storage are combined to calculate channel flow. A detailed description of the model can be found in (Bandaragoda et al., 2004; Tarboton, 2007).

Note that in the TOPMODEL/TOPNET saturated zone, discharge is simulated as an exponential function of saturated zone storage deficit. This parameterization has the
effect that baseflow is never zero. Two of the streamflow regime variables used from Dhungel et al. (2016) are based on days with zero flow. These are the extended low flow index (ELFI) and number of zero flow events (ZFE). We could have changed the model by introducing a new parameter representing a threshold below which model flow is set to zero, but this would have introduced another parameter requiring estimation or calibration, so was not done. Instead we have just retained the very small baseflow values that TOPNET outputs when saturated zone deficit is large. As a result of this, we do not expect the model to be able to reproduce ELFI and ZFE well. In particular, ZFE was always simulated as zero as there were no modeled zero flow events. Our interpretations of the results acknowledge this and focus on the other streamflow variables. ELFI and ZFE are included in the results for completeness.

Setting up a TOPNET model requires diverse spatial and temporal input data. The major spatial inputs include watershed boundaries, stream network, soils, land use, and land cover. Time series weather data inputs include daily precipitation, daily minimum and maximum air temperature, wind speed, and humidity. In addition to streamflow, TOPNET diagnostic output consists of time series of model state variables for each sub-basin. These variables include mean water table depth, soil zone storage, and canopy storage. Diagnostic output also includes information for each sub-basin on infiltration excess runoff, saturation excess runoff, base flow, drainage from the soil to the saturated zone (recharge), percent saturated area, potential evapotranspiration, and actual evapotranspiration.
2.4 Weather Data

The meteorological forcing that drives the TOPNET model are precipitation, temperature, and wind speed. Historical observed Daymet (Thornton, 2014) precipitation and minimum and maximum temperature data were used for the TOPNET model calibration and validation. In this study, weather sensitivity scenarios were developed from twenty years (1986-2005) of historical Daymet data by uniformly adjusting the historic record as follows: (1) -40% to 40% daily precipitation change with an interval of 10% and (2) -2°C to 6°C temperature change with an interval of 1°C.

This study used GCM output from the World Climate Research Program Coupled Model Inter-comparison Project Phase 5 (CMIP5) multi-model dataset for streamflow regime projection. The current study applied outputs from a subset of five participating models (Table 2.3), which were selected based on model performance, resolution, and data availability. The Intergovernmental Panel on Climate Change, Fifth Assessment Report, gave a range of plausible radiative forcing scenarios, called representative concentration pathways (RCP). These scenarios include a mitigation scenario leading to very low forcing level (RCP2.6), two medium stabilization scenarios (RCP4.5 and RCP6), and one very high emission scenario (RCP8.5). In this study, we used the lowest (RCP2.6) and highest (RCP8.5) emission scenarios for climate change analysis.

Climate model projection data were obtained from the ‘Downscaled CMIP3 and CMIP5 Climate and Hydrology Projections’ site (ftp://gdo-dcp.ucar.edu/archive/cmip5/bcca) hosted by Lawrence Livermore National Laboratory (LLNL) Program for Climate Model Diagnosis and Intercomparison.
PCMDI (Maurer et al., 2007). This dataset consists of a library of 112 fine-resolution climate projections based on sixteen climate models and different RCP scenarios, which were bias corrected and statistically downscaled to daily values with a spatial resolution of 1/8 degree and covers the whole U.S. Downscaling was achieved by using the daily bias-corrected and constructed analogues (BCCA) method outlined in Reclamation (Reclamation, 2013). Numerous studies have used this dataset to assess the potential impact of climate change on different aspects of hydrology (Sinha et al., 2010; Towler et al., 2010; Das et al., 2011; Todd et al., 2011). However, we found bias in precipitation and maximum and minimum temperature when compared with observed data for the baseline period (1986–2005). We therefore applied the quantile-based mapping method to correct biases based on observed data (Piani et al., 2010). This method maps the distribution of historical global climate data onto gridded observed Daymet data. The method is a relatively simple approach that has been successfully used in hydrology and many other climate studies (Piani et al., 2010; Maraun, 2013; Fang et al., 2015).

2.5 Model Setup and Calibration

The parameters of TOPNET are related to physical properties of the sub-watershed, including soils, topography, and land cover (Table 2.4). We applied the Terrain Analysis Using Digital Elevation Model (TauDEM) software (Tarboton, 2002) to a 30-m DEM obtained from the National Elevation Dataset (NED) to delineate stream networks and sub-watersheds for each of the eight representative study watersheds. TauDEM was also used to obtain the slope and specific catchment area for each grid cell in the DEM. These values were used to compute the topographic wetness index
distribution for each sub-watershed. Some of the sub-watershed properties (e.g., saturated hydraulic conductivity) were directly obtained from the Soil Survey Geographic database (SSURGO), and others (e.g., plant available porosity) were derived from soil texture information. Linear regression of logarithmic saturated hydraulic conductivity versus depth was used to fit the exponential function describing the decrease of hydraulic conductivity with depth that TOPMODEL assumes. It was also used to estimate saturated hydraulic conductivity at the surface, $K_0$ and sensitivity parameter, $f$. A linear regression of Daymet annual temperature versus station elevation was used to determine lapse rate to adjust daily temperature data for each study site. Canopy capacity, $CC$, and intercepted evaporation enhancement, $Cr$, were estimated from land cover and land use data.

In this study, only climate change impact was analyzed, and we assumed that TOPNET model parameters will not change with time, but do require calibration as the parameter estimates directly from soil and land cover result in poor simulations. Calibration does detract from the physical basis of model parameters, but is common in such hydrologic models. Differences between direct physical estimates of parameter and calibrated parameters are due to many factors, such as scale dependencies of the effective properties represented by the parameters. Here some of the physical basis of direct parameter estimates was retained using the multiplier approach that maintains the same relative differences of parameters across subwatershed and limits the degrees of freedom. The elitist multiobjective genetic algorithm (Deb, 2001) was used to calibrate sensitive TOPNET model parameters for each of the watersheds considering Nash Sutcliffe Efficiency ($NSE$) and $Bias$ as objective functions:
\[ \text{NSE} = 1 - \frac{\sum_{i=1}^{n} (Q_{\text{sim},i} - Q_{\text{obs},i})^2}{\sum_{i=1}^{n} (Q_{\text{obs},i} - \bar{Q}_{\text{obs}})^2} \]  

(2-2)

where \( Q_{\text{sim}} \) = simulated flow, \( Q_{\text{obs}} \) = observed flow, \( \overline{Q}_{\text{obs}} \) = mean of observed flow, \( i = \) time step, \( n = \) total number of time steps.

\[ \text{Bias} = \frac{\sum_{i=1}^{n} Q_{\text{sim},i}}{\sum_{i=1}^{n} Q_{\text{obs},i}} \]  

(2-3)

\( \text{Bias} \) can range between 0 and \(+\infty\), but indicates an excellent performing model when a value of 1 is generated. \( \text{NSE} \) can range between \(-\infty\) and 1, where the value of 1 indicates a perfect fit. \( \text{NSE} \) is similar to the commonly evaluated regression coefficient of determination (\( R^2 \)), but not exactly the same as it does not assume a linear model. \( \text{NSE} \) does suffer from similar shortcomings to \( R^2 \) in terms of fitting data when observations are not well distributed or representative of the full process variability. These potential problems were mitigated here by using multiple years of daily flow that capture the streamflow variability well, and the use of bias as a metric as well as \( \text{NSE} \) in a multi-objective calibration approach.

Our calibration procedure used multipliers, rather than individual sub-basin parameters as calibration variables. One multiplier value for each parameter applied uniformly to all of the sub-basins within the entire watershed is a parsimonious way to calibrate the model and maintain spatial variation between sub-watersheds based on GIS-derived parameter values. After setting all inputs, the model for each watershed was run
for the 1997–2005-time period. The first two years of simulations were used as a spin-up period and were excluded from the analysis. The four-years from 1999–2002 were used for calibration, and the three years from 2003–2005 were used for independent validation. Each calibrated model was run using Daymet meteorological forcing for the twenty-year baseline period (1986–2005) to estimate streamflow regime variables and then compared with the same variables derived from observed streamflow. Each model was then rerun for the hypothetical climate change scenarios to estimate the sensitivity of streamflow variables with climate change. After sensitivity analysis, each model was run again using input data from five GCMs for the baseline period. Finally, each model was run using input from the five GCMs for the period 2076–2095 to assess the impact of climate change on streamflow regime variables for the lowest (RCP2.6) and highest (RCP8.5) emission scenarios.

2.6 Results

2.6.1 Calibration and Validation Results

Overall, model performance statistics indicated satisfactory calibration (Table 2.5). The term “satisfactory” was based on the somewhat subjective model efficiency interpretations suggested by (Moriasi et al., 2007; Ficklin et al., 2013), where a calibration with NSE value >0.5 was considered to be a satisfactory calibration. The average NSE value for eight watersheds for the calibration and validation periods was 0.65 and 0.6 respectively, which together with visual comparisons of modeled and observed hydrographs corroborated that the TOPNET model was able to satisfactorily represent the daily streamflow in each watershed. During calibration, the lowest (0.44)
and highest (0.76) NSE values were found in the perennial flashy large (C22) and intermittent late (B22) watersheds respectively. The average Bias for eight watersheds for the calibration and validation was 1.2 and 1.18, indicating that model overestimated daily streamflow compared to observation. There was good agreement among streamflow regime variables derived from Daymet driven model simulation results with those of from observed data (Figure 2.4). Streamflow regime variables that refer to general flow conditions (e.g., mean discharge (Q mean), T50, coefficient of variation (CV), seven days maximum flow (Q7max), seven days minimum flow (Q7 min), and bank full flow (BFF)) produced by the TOPNET model had NSE value >0.75 and compared well with observed values (Figure 2.4). The Peak time, ELFI, Flood Duration (FD), High Flow Event (HFE), and Low Flow Event (LFE) values derived from the model simulation results compared well with observed values in most of the river basins, yet distinct differences occurred between modeled and observed values for some variables in some watersheds, indicating problematic performance of the model for those variables. Poor performance for ELFI and ZFE was expected for watersheds with significant periods of zero flow, as zero flow is never simulated in TOPNET due to its exponential storage and recession function. These discrepancies result in some bias in the estimation of streamflow regime variables. Because of this bias, changes in streamflow regime variables were quantified by comparing streamflow regime variables derived from the model driven by historical GCM data and future data rather than a direct comparison with historical observations.
2.6.2 Precipitation Sensitivity

The response of the streamflow regime variables varied significantly due to change in precipitation (Figure 2.5). The Qmean, Q7max, Q7min, and BFF increased with increasing precipitation for all of the watersheds and vice versa. However, the degree of change in those variables varied considerably across watersheds. The percentage change of a specific variable in response to precipitation change also varied among the watersheds. The sensitivity of Qmean to precipitation was highest in the intermittent late stream (B22) watershed, followed by the perennial steady small (A1), intermittent early (B21), intermittent steady (B1), perennial flashy large (C22), perennial flashy early (C1), perennial steady large (A2), and perennial flashy small (C21) watersheds. The CV of discharge decreased in most of the watersheds except perennial flashy small and large (C22 and C21) watersheds with increasing precipitation. Increasing precipitation resulted in decreasing ELFI in most of the watersheds except intermittent early (B21) and perennial steady large (A2) watersheds and vice versa. The changes of timing (day of peak flow and T50) of streamflow due to precipitation change were found to be negligible in all of the watersheds, excluding the intermittent early (B21) stream.

2.6.3 Temperature Sensitivity

Sensitivity of streamflow regime variables to changes in temperature, holding the precipitation fixed were also different among the eight streamflow class (Figure 2.6). The Qmean, Q7max, and BFF decreased in most of the watersheds in response to increasing temperature, but increased in the intermittent steady (B1) and perennial flashy early (C1)
watersheds. The Q7min decreased with increasing temperature in all of the watersheds. The amount of decrease of these variables in a specific watershed also differed. The T50 decreased in the perennial steady (A1), intermittent (B21 and B22) and perennial flashy small (C21) watersheds with increasing temperature. Increasing temperature resulted in earlier peak flows in the snow-fed watershed (A2). The sensitivity of day of peak flow to temperature for the rain-fed watershed was negligible. The ELFI decreased with increasing temperature in most of the watersheds except perennial flashy small (C21).

2.6.4 **GCM Temperature Change for 2076-2095**

The downscaled GCM projections indicate a robust signal of increasing temperature in all watersheds for both emission scenarios (Figure 2.7). Toward the end of the century, annual temperature is expected to increase by 3°C and 5°C under the lowest and highest emission scenarios, respectively, as compared to baseline period. The rate of temperature increase under the highest emission scenario is projected to be two times that of the lowest emission scenario by the end of the century. Under the lowest emission scenario, temperature increases were similar for all seasons; however, the temperature increase in summer was projected to be larger than in winter under the highest emission scenario. The highest increase in annual temperature was projected in the perennial flashy small (C21) watershed. Annual temperature was projected to increase by an average of 5° C in the intermittent streams (B1, B21, and B22).

2.6.5 **GCM Precipitation Change for 2076-2095**

Downscaled projections of precipitation varied significantly among GCMs. In the input data for our study watersheds, yearly precipitation is projected to increase in all of
the watersheds, with larger increases under the highest emission scenario (Figure 2.8). Annual precipitation was predicted to increase by 13% under the highest emission scenario in the perennial steady small (A1) watershed by the end of the century. On average, precipitation was projected to increase by 10% in all scenarios in the perennial steady large (A2 watershed). In the intermittent steady (B1) watershed, precipitation was projected to increase significantly (150%) during the summer season under the high emission scenario for the years 2076–95, although there was little change in projected annual precipitation. Among all of the watersheds, the intermittent early (B21) watershed had the highest increased precipitation under the highest emission scenario. The perennial flashy early (C1) watershed projection showed a minimal increase in precipitation for both emission scenarios, with a drop-off in summer precipitation. Yearly precipitation is likely to increase somewhat more than 10% under the highest emission scenario in the perennial flashy small and large watersheds (C21 and C22).

2.6.6 Streamflow Regime Variable Changes for 2076–2095

The GCM projected changes in precipitation and temperature drive changes in streamflow regime variables. Simulation results indicate that in many watersheds future streamflow regime is likely to be different from past experience (Figure 2.9). Projections of Qmean increased in most of the watersheds (e.g., perennial steady and flashy small (A1, and C21), intermittent early and late (B21 and B22)), but decreased in the intermittent steady (B1) and perennial flashy small (C21) watersheds. The highest increase was in the perennial steady small (A1) watershed for both climate change scenarios. Projected Q7max increased in all of the watersheds except perennial flashy
small (C21) for both emission scenarios with the highest increase in perennial steady small (A1) watershed. Projections of BFF also increased in all of the watersheds except intermittent steady (B1) and perennial flashy small (C21) for both RCPs. The projected Q7min decreased in most of the watersheds (e.g., perennial steady, perennial flashy small, and intermittent steady (A1, C21 and B1)).

Projections indicated an earlier day of peak flow in the snow-fed watersheds (A2 and C21) for both climate change scenarios. For example, day of peak flow advanced by 21 and 65 days in the perennial steady large (A2) watershed under the lowest and highest emission scenarios, respectively. The day of peak flow was projected to occur later in the intermittent early and late (B21 and B22) watersheds. T50 was projected to decrease in the perennial steady (A1 and A2) and flashy (C1 and C21) watersheds but increase in the intermittent early (B21) watershed for both emission scenarios, with the highest decreases of 27% (61 days) in the perennial steady large (A2) watershed under the highest emission scenario.

2.7 Discussion

The utility of our analyses for improving understanding and predicting streamflow regime due to climate change depends on several assumptions and conditions. First, we assumed that a physically based model (i.e., TOPNET) provides a representation of the physical processes involved in quantifying streamflow and can be used for estimating the impact of climate change on streamflow regime. Second, we assumed that the climate models we used produced a reasonably plausible representation of future climate change conditions in eight different watersheds across the U.S. Finally, we assumed that the
sensitivity analysis and prediction using TOPNET were informative, considering the inevitable uncertainties in the predictions. We address each of these issues below.

The usefulness of our analysis for quantifying the impact of climate change depends on how well we predict streamflow regime. Our physically based modeling results indicated that the simulated streamflows are generally consistent with observations during the calibration and validation periods. Our results indicated that the Qmean, Q7max, Q7min, CV, BFF, T50, P, C, and M were predicted well by the models, while FR, HFE, LFE, Peaktime, ELFI, and FLDD were problematic to the model. This might be associated with uncertainty in the input data, estimation of model parameters, and model assumption for describing physical process.

Our study explored the sensitivity of streamflow regime variables to precipitation and temperature. Our results show that sensitivity of stream flow regime variables varied from basin to basin and for the different variables. For example, Qmean was found to be more sensitive to precipitation than temperature. Additionally, streamflow regime variables that refer to streamflow magnitude (e.g., Q7max, BFF, Q7min) increased with increasing precipitation but decreased with increasing temperature and vice versa. The sensitivity of Qmean to precipitation across the watersheds appeared correlated with rainfall-runoff coefficient (Table 2.2), a finding that is in agreement with the conclusions of (Chiew, 2006; Johnson et al., 2012). The timing of streamflow (e.g., Peak time and T50) for snow-fed watersheds was more sensitive to temperature than precipitation; however, timing for rain-fed watersheds had little sensitivity to temperature and precipitation change.
This study projected a number of changes in stream flow regime associated with predicted climate change by the end of the century. For example, mean streamflow (Qmean) and high flow amount (Q7max and BFF) were predicted to increase in most of the watersheds by the end of the century due to increase in mean annual precipitation. However, low flow variables (Q7 min and ELFI) were predicted to decrease in most of the watersheds under the highest emission scenario due to the combined impact of increasing temperature and decreasing summer precipitation. These results are consistent with analyses conducted by Petra and Hannes Müller (2012), who used the WaterGap Global Hydrology Model (WGHM) to assess the impact of climate change on mean river discharge, low and high flows, and mean seasonal discharge. They found that mean annual discharge is projected to increase but low flows are projected to decrease.

Our study found that snow fed watersheds (A2) are projected to have an earlier peak time due to the compounding effects of increasing winter precipitation, more winter precipitation falling as rainfall instead of snow, and earlier snowmelt under higher temperature. These results are consistent with analyses conducted by (Huang et al., 2012), who predicted significant changes in streamflow timing and magnitude in extreme flows but they contrast with those of Dhungel et al. (2016) who predicted that snow-fed streams in the western mountain regions appear to be relatively insensitive to climate change. However, our results are not directly comparable with those of Dhungel et al. (2016) as they used a statistical model and different climate change scenarios. The ELFI was projected to decrease in most of the watersheds due to increase in mean discharge and decrease in low flow. The CV was increased in most of the watersheds, which might
be influenced by factors including frequency and magnitude of precipitation and land use and land cover characteristics (Nijssen et al., 2001; Chang et al., 2012).

The changes in streamflow regime variables mentioned in this study may affect ecological processes in streams. For example, reducing Q7min by 53% and increasing the LFE by 70% in the snow fed watershed (A2) might increase the stream temperature to the extent that detrimental impacts on fish occur. Additionally, shifting the timing of peak flow in the snow-fed watershed could alter the retention time of organic matter (Mulholland et al., 1997) and disrupt the recruitment of riparian species that rely on appropriately-timed high flows. However, a number of sources of uncertainty must be considered in interpreting results, including uncertainty in the emissions scenarios, uncertainty in the GCM simulations of future climate, uncertainty in the downscaling of these GCM outputs to the local scale, and uncertainty in the watershed models used to translate potential changes in local climate to watershed response. Quantification of the uncertainty in changes of streamflow regime due to climate change is beyond the scope of this study. However, further work involving quantifying the uncertainty is needed not only for purposes of detection and attribution, but also for strategic approaches to adaptation and mitigation.

2.8 Conclusion

We conducted these analyses to develop a better understanding of the impact of climate change on streamflow regime, which is required for identifying how stream ecosystems are vulnerable to these changes, and ultimately to guide the development of robust strategies for reducing risk in the face of changing climatic conditions.
was used to characterize the sensitivity and to assess the impact on streamflow regime variables due to climate change. This study shows that physically based modeling can reproduce streamflow regime variables and thus can be used to project their changes. However, TOPNET was not able to reproduce all variables, and poorly reproduced variables need to be recognized when interpreting the projections. Our results indicated that GCM projections were generally consistent in showing a continued warming trend over the end of the century but precipitation changes were more variable, with some models projecting increases while others project decreases.

We anticipated that simulated changes in streamflow regime variables due to climate changes would likely be significant and show a high degree of variability. We found the following patterns in projected streamflow regime changes. First, streamflow is likely to increase in amount and variability. The high flow amount (e.g., Q7max and BFF) is projected to increase. Second, the low flow condition (e.g., Q7 min and SI) is predicted to decrease under the highest emission (RCP8.5) scenario. Finally, the timing (e.g., Peaktime and T50) of stream flow is predicted to occur earlier for the snow-fed watersheds in both emission scenarios; however, the models reproduction of peak time was poor. Finally, our approach for predicting changes in streamflow regime due to climate change provides a method for estimating projected changes in streamflow regime variables. This approach contributes to a growing understanding of the complex relationship between climate change and streamflow regime variables and could be considered in the examination of climate change impacts on stream ecology.
Tables and Figures

Table 2.1 Variables used to characterize streamflow regime.

<table>
<thead>
<tr>
<th>Flow Variables</th>
<th>Flow Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Daily Discharge (Mean)</td>
<td>General</td>
</tr>
<tr>
<td>Coefficient of Variation of Daily Flows (COV)</td>
<td></td>
</tr>
<tr>
<td>Colwell’s Index of Predictability (P)</td>
<td></td>
</tr>
<tr>
<td>Colwell’s Index of Constancy (C)</td>
<td></td>
</tr>
<tr>
<td>Colwell’s Index of Contingency (M)</td>
<td></td>
</tr>
<tr>
<td>Flow Reversals Per Year (FR)</td>
<td></td>
</tr>
<tr>
<td>50% Flow Date (T50)</td>
<td>Dry</td>
</tr>
<tr>
<td>Number of Zero Flow Events (ZFE)</td>
<td></td>
</tr>
<tr>
<td>Extended Low Flow Index (ELFI)</td>
<td></td>
</tr>
<tr>
<td>Number of Low Flow Events (LFE)</td>
<td>Low</td>
</tr>
<tr>
<td>Average 7 Day Minimum Streamflow (Q7min)</td>
<td></td>
</tr>
<tr>
<td>Average 7 Day Maximum Streamflow (Q7max)</td>
<td></td>
</tr>
<tr>
<td>Bank Full Flow (Q167), Flood Duration (FLDD)</td>
<td>High</td>
</tr>
<tr>
<td>Time of Peak (Tp), Number of High Flow Events (HFE)</td>
<td></td>
</tr>
</tbody>
</table>

Note: These variables were selected by Dhungel et al. (2016) as being relevant for stream ecology. We used this selection for consistency with that study.
Table 2.2 Physical attributes of study watersheds selected to represent each streamflow regime class.

<table>
<thead>
<tr>
<th>Class</th>
<th>Name</th>
<th>Area (km²)</th>
<th>Mean Annual</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>T (°C)</td>
<td>P (mm)</td>
</tr>
<tr>
<td>A1</td>
<td>Martin Ck nr Paradise Valley, NV</td>
<td>454</td>
<td>7</td>
<td>486</td>
</tr>
<tr>
<td>A2</td>
<td>Lichsa River, ID</td>
<td>3054</td>
<td>5</td>
<td>1338</td>
</tr>
<tr>
<td>B1</td>
<td>Santa Cruz Ck nr Santa Ynez, CA</td>
<td>192</td>
<td>13</td>
<td>731</td>
</tr>
<tr>
<td>B21</td>
<td>Tar River, NC</td>
<td>429</td>
<td>15</td>
<td>1190</td>
</tr>
<tr>
<td>B22</td>
<td>Beaver Ck nr Mason, TX</td>
<td>558</td>
<td>18</td>
<td>782</td>
</tr>
<tr>
<td>C1</td>
<td>South Fork Coquille River, OG</td>
<td>443</td>
<td>10</td>
<td>2343</td>
</tr>
<tr>
<td>C21</td>
<td>Never Sink River nr Claryville, NY</td>
<td>173</td>
<td>6</td>
<td>1493</td>
</tr>
<tr>
<td>C22</td>
<td>Whiskey Chitto Ck nr Oberlin, LA</td>
<td>1305</td>
<td>19</td>
<td>1654</td>
</tr>
</tbody>
</table>

Note: Here T, P, Q and C stands for Temperature, Precipitation, Discharge and Runoff ratio respectively.
Table 2.3 Details of the climate models and emission scenarios use for future climate projections.

<table>
<thead>
<tr>
<th>Models Name</th>
<th>Research Center</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canadian Earth System Model version2 (CanESM2)</td>
<td>Canadian Centre for Climate Modeling and Analysis, Canada.</td>
</tr>
<tr>
<td>The Community Climate System Model version4 (CCSM4)</td>
<td>National Center for Atmospheric Research, USA.</td>
</tr>
<tr>
<td>Geophysical Fluid Dynamics Laboratory Climate Model version 3 (GFDL-CM3)</td>
<td>GFDL, National Oceanic and Atmospheric Administration, USA.</td>
</tr>
<tr>
<td>Model for Interdisciplinary Research on Climate-Earth System version 5</td>
<td>Studies (NIES), Japan Agency for Marine-Earth Science and Technology (JAMSTEC),</td>
</tr>
<tr>
<td>(MIROC5)</td>
<td>Japan.</td>
</tr>
<tr>
<td>Max-Planck Institute Earth System Model-Medium Resolution (MPI-ESM-MR)</td>
<td>MPI for Meteorology, Germany.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Emission Scenarios</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCP2.6</td>
<td>Peak in radiative forcing at ~ 3 W/m² before 2100 and decline</td>
</tr>
<tr>
<td>RCP8.5</td>
<td>Rising radiative forcing pathway leading to 8.5 W/m² in 2100</td>
</tr>
</tbody>
</table>

Table 2.4 TOPNET model parameters.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Estimated from</th>
</tr>
</thead>
<tbody>
<tr>
<td>f (m⁻¹)</td>
<td>Saturated store sensitivity</td>
<td>Soils (multiplier calibrated)</td>
</tr>
<tr>
<td>Kₒ (m/h)</td>
<td>Surface saturated hydraulic conductivity</td>
<td>Soils (multiplier calibrated)</td>
</tr>
<tr>
<td>Δθ₁</td>
<td>Drainable porosity</td>
<td>Soils (multiplier calibrated)</td>
</tr>
<tr>
<td>Δθ₂</td>
<td>Plant available porosity</td>
<td>Soils (multiplier calibrated)</td>
</tr>
<tr>
<td>d (m)</td>
<td>Depth of soil zone</td>
<td>Soils (multiplier calibrated)</td>
</tr>
<tr>
<td>C</td>
<td>Soil Zone drainage sensitivity</td>
<td>Soils (multiplier calibrated)</td>
</tr>
<tr>
<td>ψf (m)</td>
<td>Wetting front suction</td>
<td>Soils</td>
</tr>
<tr>
<td>V (m/h)</td>
<td>Overland flow velocity</td>
<td>360 (default)</td>
</tr>
<tr>
<td>CC (m)</td>
<td>Canopy capacity</td>
<td>Land cover and Land use (multiplier calibrated)</td>
</tr>
<tr>
<td>Cᵣ</td>
<td>Intercepted evaporation enhancement</td>
<td>Land cover and Land use (multiplier calibrated)</td>
</tr>
<tr>
<td>α</td>
<td>Albedo</td>
<td>Land cover and Land use</td>
</tr>
<tr>
<td>Lapse (°C/m)</td>
<td>Lapse Rate</td>
<td>Elevation- Average temperature</td>
</tr>
<tr>
<td>T</td>
<td>Transmissivity</td>
<td>Soils (multiplier calibrated)</td>
</tr>
</tbody>
</table>
Table 2.5 NSE and Bias for the calibration and validation period for eight watersheds.

<table>
<thead>
<tr>
<th>Watershed Class</th>
<th>Calibration</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NSE</td>
<td>Bias</td>
</tr>
<tr>
<td>A1</td>
<td>0.68</td>
<td>1</td>
</tr>
<tr>
<td>A2</td>
<td>0.73</td>
<td>1.3</td>
</tr>
<tr>
<td>B1</td>
<td>0.7</td>
<td>1.5</td>
</tr>
<tr>
<td>B21</td>
<td>0.58</td>
<td>1.1</td>
</tr>
<tr>
<td>B22</td>
<td>0.44</td>
<td>1.1</td>
</tr>
<tr>
<td>C1</td>
<td>0.75</td>
<td>1.2</td>
</tr>
<tr>
<td>C21</td>
<td>0.56</td>
<td>1.1</td>
</tr>
<tr>
<td>C22</td>
<td>0.76</td>
<td>1.2</td>
</tr>
<tr>
<td>Mean</td>
<td>0.65</td>
<td>1.18</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.76</td>
<td>1.5</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.44</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 2.1 Spatial distribution of eight streamflow regime classes across the US. The push pin symbol indicates the study site selected for each streamflow class as the stream with principal component factors closest to the median of the PC factors for that class.
Figure 2.2 Daily flow pattern for the representative stream in each class (i.e., the stream closest to the median of each of the PC factors). Gray shading gives the 5th to 95th percentile range of daily flows, fine lines give 25th and 75th percentile, and the bold (red) line gives the 50th percentile (median).
Figure 2.3 TOPNET model schematic.
Figure 2.4 Comparison of observed and modeled (Daymet driven) streamflow regime variables for the climatological period of 1986-2005 for the eight selected watersheds.
Figure 2.5 Percentage changes of streamflow regime variables due to precipitation change only.
Figure 2.6 Percentage changes of streamflow regime variables due to temperature change only.
Figure 2.7 Box plots showing the annual and seasonal temperature changes for the lowest RCP 2.6 (white) and highest RCP 8.5 (gray) emission scenario for the period of 2076-95 compared to 1986-2005. Individual plots show median changes with minimum to maximum range. The median for each GCM model was computed for the baseline and future period, then the difference between these quantities evaluated. Boxes show the range in these differences, annually and for each season across the five GCM models used.
Figure 2.8 Box plots showing the annual and seasonal precipitation changes for the lowest RCP 2.6 (white) and highest RCP 8.5 (gray) emission scenario for the period of 2076-95 compared to 1986-2005. Individual plots show median changes with minimum to maximum range. The median for each GCM model was computed for the baseline and future period, then the difference between these quantities evaluated. Boxes show the range in these differences, annually and for each season across the five GCM models used.
Figure 2.9 Changes in streamflow regime variables due to climate changes at the end of the century for the lowest (RCP 2.6) and highest (RCP 8.5) emission scenarios.
References


Ficklin, D.L., Stewart, I.T., Maurer, E.P., 2013. Climate change impacts on streamflow and subbasin-scale hydrology in the Upper Colorado River Basin. PloS one, 8(8).


CHAPTER 3

DATA SERVICES IN SUPPORT OF INPUT DATA PREPARATION FOR PHYSICALLY BASED HYDROLOGIC MODELING

Abstract

Physically based distributed hydrologic models are widely used for better understanding of hydrological processes, quantifying the response of the hydrologic systems to climate and land use changes, and managing water resources. Distributed hydrologic models require diverse geospatial (e.g., topography, geology, soil, land use, land cover, etc.) and time series data (e.g., temperature and precipitation) as inputs to set up and run. The volume of this information increases as the spatial resolution and extent of study areas increases. These data may be obtained from national spatial data available via the Internet and also from local sources. In addition to diversity and abundance, the inherent heterogeneity in these data makes it difficult to acquire, manipulate, and manage. At present, most researchers perform such manipulation manually, which is time-consuming, error-prone, and difficult to reproduce. In this paper, we describe data services we developed to automate the input data preparation steps for a physically based, distributed hydrologic model. The advantage of these web services in terms of ease in prototyping of the model is demonstrated using a case study application for a watershed in Utah, USA. These web services enable water scientists to spend less time extracting and formatting spatial and time series data to use as model inputs. Additionally, this

---

2 Coauthored by Nazmus Sazib, David G Tarboton and Pabitra Dash
approach alleviates the need for users to install and work with desktop GIS and other statistical software, thereby overcoming compatibility limitations and enhancing reproducibility.

3.1 Introduction

Physically based, distributed hydrologic models are useful for answering science and management questions related to hydrological processes, the impact of climate and land-use changes, as well as managing and supporting water resources. Distributed hydrologic models are data intensive and often require specific transformation of available data sets to convert to the form required by a model. Data may be available from a variety of different sources and this makes it difficult to quickly locate the most appropriate resource for a specific model. Additionally, the workflows used for data acquisition and transformation often require significant manual intervention which is labor intensive and difficult to reproduce. As a result, researchers and practitioners implementing hydrologic models spend a significant amount of time on basic data gathering and transformations, taking away from the time available for scientific analysis and decision making.

There have been a number of studies that automate the input data preparation steps which have the potential to overcome some of the limitations mentions above and are available for desktop application. PHIM GIS (Bhatt et al., 2008; Kumar et al., 2009; Bhatt et al., 2014) is one example in which a GIS framework was tightly coupled to the Penn State Integrated Hydrologic Modelling System (PIHM, http://www.pihm.psu.edu/) and where model input pre-processing and input and output visualization can be carried
out. EcohydroLib (Miles, 2014) is another example that provides extensible tools for acquiring geospatial data directly from web services or local data sources for the preparation of model inputs for any watershed. This facilitates rapid development of model inputs and eases comparisons of study sites and model structures, and is primarily focused on the RHESSys model. Other examples include the Water Erosion Prediction Project (WEPP) interface on Geographic Resources Analysis Support System (GRASS) (Engel et al., 1993); the interface between ARC/INFO and the Hydrologic Engineering Center (HEC) modeling system (Hellweger and Maidment, 1999); and ArcGIS-SWAT, an extension to ArcMap (Olivera et al., 2006) for preparing SWAT model inputs. However, most of these need to be installed on a particular operating system, some require proprietary software, and some have strict hardware and software requirements that must be met to ensure that they function properly.

Some of the issues listed above can be overcome through the creation of a web service based data preparation and modeling environment that provides access to existing models, their input/output datasets, and a mechanism to perform simultaneous simulations (Rajib et al., 2016). For example, HydroDesktop, the Consortium of Universities for the Advancement of Hydrologic Science Inc (CUAHSI) Hydrologic Information System (HIS) data access client provides a platform for time series data download, extraction, and analysis to hydrologists by enabling map based selection of watershed (Ames, 2012). However, HydroDesktop does not provide any spatial data, for example, soil and land cover data. HydroTerre (Leonard and Duffy, 2013) provides both spatial and time series data at the HUC-12 scale. Tarboton et al. (2014) developed a web
based collaboration environment called HydroShare for data sharing, analysis, and modelling in order to advance hydrology through integration of information from multiple sources. However, it does not provide tools for extracting and processing geospatial or time series information required to set up a hydrologic model.

Considering the background information presented in the previous paragraphs, there is a need to create cyberinfrastructure to set up and share physically based hydrologic models using server functionality to the maximal extent possible and minimizing the need for data or software on the user’s desktop computer. In this paper, we describe the design, architecture, implementation, and a case study involving a newly developed data service referred to as Hydro Data Services (HydroDS). HydroDS was designed to enable automating retrieval and processing of spatial and temporal data required by a physically based distributed hydrologic model such as TOPNET. These data services provide many capabilities needed by the hydrologic modelers, including: basic geoprocessing tasks (e.g., resampling, projecting, subsetting), watershed delineation, downloading of spatial data and time series climate data and sharing data through HydroShare. The advantage of these data services in terms of ease in configuring the TOPNET model is demonstrated using a case study application for a watershed in Logan, Utah, USA.

The paper is organized as follows: Section 3.2 introduces the TOPNET model. Section 3.3 motivates this work, outlining the problems involved in preparing TOPNET inputs from multiple data sources. Section 3.4 presents functionality and the system design of HydroDS services developed to address this problem. Section 3.5 gives the
implementation details of the HydroDS services. Section 3.6 gives an example application, and Section 3.7 gives discussion and conclusions.

3.2 TOPNET Model

TOPNET (Bandaragoda et al., 2004; Ibbitt and Woods, 2004) is a distributed hydrologic model with basic model elements being topographically delineated subwatersheds that discharge into the stream network, which is then used to route flow to the outlet. It was developed by combining TOPMODEL (Beven et al., 1979 and Beven et al., 1995a), which simulates flow in relatively small watersheds, with channel routing. This approach can be applied over a large watershed using subwatersheds within the larger watershed as model elements. A key contribution of TOPMODEL is the parameterization of the soil moisture deficit (depth to water table) using a topographic index to model the dynamics of variable source areas contributing to saturation excess runoff. The TOPNET model includes many enhancements beyond the original Beven and Kirkby TOPMODEL including (1) calculation of reference evapotranspiration using the ASCE standardized Penman-Monteith method (Walter and Brown, 2000; Allen, 2005) and (2) calculation of snowmelt using the Utah Energy Balance Snowmelt model (Tarboton, 1995a). Detailed descriptions of the model can be found in (Bandaragoda, 2012). Bandaragoda et al. (2004) applied TOPNET to the Distributed Model Intercomparison Project (DMIP) watersheds in South Central USA (Smith et al., 2004). Their study demonstrated that TOPNET was suitable for hydrologic prediction across the range of conditions in Bandaragoda study. Tarboton et al. (2007) applied TOPNET to the
Nooksack Watershed in Washington State. This work added the Utah Energy Balance snowmelt model to TOPNET, extending its capability to snowmelt dominated regions.

3.3 Why Automation of Input Data Preparation is Required

Setting up a TOPNET model requires several input datasets which are generated through a sequence of data processing steps (Figure 3.1). The figure illustrates the challenges and work involved in preparation of input data for the TOPNET model. These include the challenge that different datasets are required from different data sources such as digital elevation data from the National Elevation Dataset (NED), soil data from United States Department of Agriculture (USDA) State Soil Geographic Data Base (SSURGO), land use and land cover data from the National Land Cover Dataset (NLCD), and weather data from different sources such as Daymet (Thornton, 2014), PRISM (Daly, 2008), and NLDAS-2 (Xia, 2012). The Daymet data set provides daily gridded estimates of daily weather parameters for North America, with 1km x 1km spatial resolution. The PRISM dataset provides spatial resolution (4km x 4km) climate surfaces of temperature, precipitation and dew point temperature at monthly and daily time scales The NLDAS-2 dataset features meteorological variables at hourly time scales and 1/8th degree (approximately 12 km) resolution. Daymet was chosen for the TOPNET climate input as it has a fine spatial (1km x 1km) and temporal resolution (daily) which serves the purpose of hydrologic modeling.

At present, the acquisition of datasets is done manually, which is labor intensive and difficult to document in a reproducible way. Additionally, the formats of downloaded data have to be transformed to TOPNET input data required formats. This involves
generation of model elements, conversion of gridded spatial datasets into modeling parameters, creation of wetness index distribution for calculating saturation excess runoff, creation of distance to stream distribution for runoff routing, and the calculation of weights associated with point precipitation measurements. These transformations involve numerous data processing and time consuming steps as they do not simply convert one file format to another. Therefore, preparing input data for the model currently requires significant effort, especially for new users that must learn the steps and configure often complicated software to correctly execute the steps.

3.4 Methods

3.4.1 Functionality Design

Given the input data preparation challenges in running data-intensive hydrologic models, such as TOPNET described above, data services referred to as HydroDS (Hydro Data Services) were developed as part of the CI-WATER project (http://ci-water.org). The objective of CI-WATER was to acquire and develop hardware and software cyberinfrastructure (CI) to support the development and use of large-scale, high-resolution computational water resources models to enable comprehensive examination of integrated system behavior through physically based, data-driven simulation. It included servers and disk arrays connected to the four participating institutions (Utah State University, University of Utah, University of Wyoming, and Brigham Young University) via an ultra-high-speed network. Servers were deployed to host the data services, and regional model and computing clusters were used to access and run simulations (Jones, 2013).
The HydroDS data services (Table 3.1;Table 3.2) can be categorized into two major types, including (1) Essential Hydrologic Variables (EHV) services, and (2) TOPNET model specific services. The Essential Hydrologic Variables services are model independent and provide spatial and time series data and support four major functions: (1) Delineate watersheds and the stream network, (2) get soil data (3) get climate data, and (4) get streamflow data. The TOPNET specific services are model specific and transform downloaded data into TOPNET input file formats, which include (1) create node and reach link files (2) create wetness index and distance distributions (3) create model parameters and (4) create precipitation input weight files (referred to as rainweight). Detailed descriptions of these services are given in Section 3.5. In addition, there are utility services for common data processing tasks such as interpolation, resampling, projection of geospatial data and sharing output results through HydroShare. This study focuses on the design and implementation of EHV and TOPNET specific services as utility services had already been developed by others.

### 3.4.2 System Design

The application architecture (Figure 3.2) consists of two components: (1) a Python client library and analysis environment on a user’s Desktop and (2) a Server on which the HydroDS are deployed. The Python client library consists of multiple Python functions, where each function contains the code that can be invoked from a user’s desktop computer to make HydroDS REST web service calls. The data services are executed on the server side, with function calls and data transfer between client and server over the web. For each data service function on the server side, a corresponding
interface is implemented in the client tool. Each function has been designed so that it can be invoked stand-alone as well as being imported into other Python code, which was beneficial for unit testing.

Data underlying the services is either hosted on the server, or retrieved from a supporting data service on the fly. The choice for each dataset was made based on how often the data changes and how efficient its retrieval is. In general, our preference was to host geospatial data that is relatively static, but retrieve weather and streamflow data on the fly. Our selection was also based on comparative testing. For example, we found during development that HydroDS data service finished approximately 10 times faster than the manual method to retrieve spatial data for the 560 km² Logan River watershed. In the case of soil properties, we hosted the soil map unit key raster which is static information but obtained soil properties on the fly based on the map unit key from a SSURGO database soil property service.

A metadata repository is connected to these data, which stores information about the dataset properties, version, and technical attributes. Additionally, a variety of software such as R (R Development Core Team, 2010), TauDEM (Tarboton, 2002), and GDAL (GDAL, 2014) were installed on the server side to execute the data service functionality. R is a statistical software and scripting language initially developed for statistical analysis, such as hypothesis testing, time series analysis and plotting, and linear and nonlinear modeling (Carslaw and Ropkins, 2012). R is also extensively used in environmental data analysis, visualization, and modeling. In this study, we used several existing R packages such as rgdal (Bivand et al., 2016), SoilDB (Beaudette, 2014),
SSSOAP (Lang, 2012), dataRetrieval (Hirsch and Cicco, 2015), and DaymetR (Hufkens, 2010). TauDEM is a suite of Digital Elevation Model (DEM) tools used for delineation of streams and watersheds from topography as represented by a DEM. It was used because (1) it supports multiple formats of vector and raster data and projection systems, and (2), it has the advantage of parallel processing for handling larger DEMs. GDAL is an open source programming library and set of utilities for vector and raster data translation and processing. As a library, GDAL presents a single abstract data model to the calling application for all supported vector and raster formats.

The data produced by the services are accessible through a downloaded link that they return. There are also services that a user can use to have the results uploaded to HydroShare, where they can be stored and shared collaboratively. Internal service complexity is hidden from service clients, and backend processing is decoupled from client applications, making the core of the system independent of a specific platform. As a result, multiple users on different platforms can access the same service functionality.

3.5 Implementation of Hydro-DS Data Services

This section describes implementation details of HydroDS data services for retrieving geospatial and time series data and transforming these datasets to support TOPNET inputs.

3.5.1 EHV Services

The ‘Delineate Watershed’ function enables delineation of a watershed and stream network for a user selected spatial domain and outlet location. This function defines and delineates the stream network based on the TauDEM Peuker Douglas valley
identification and stream drop approach (Peucker and Douglas, 1975; Tarboton, 2001) that can objectively choose the appropriate threshold to delineate stream network consistent with geomorphological properties. This function automates the steps to obtain and preprocess a DEM for the area of interest and to delineate the stream network and subwatersheds draining to each stream network reach (Figure 3.3). The watershed delineation function is primarily based on functions from TauDEM and the GDAL geospatial library. This function produces watershed, wetness index and distance to stream rasters, streamnetwork shapefile and network topological connectivity information stored in tree and coordinate text files.

The ‘Get Soil Data’ function automates the steps (Figure 3.4) required for extraction of soil properties (e.g., soil hydraulic conductivity, transmissivity, and porosity) for the delineated watershed. This function obtains soil properties from the USDA SSURGO Data Base (http://websoilsurvey.nrcs.usda.gov/). SSURGO segmented the landscape into soil map-units (MU). The soil in each MU is sampled for soil properties in various horizons, and reported by grouping into soil components. The ‘Get Soil Data’ function first extracts the soil map unit raster for the watershed boundary and submits a query to Soil Data Access (SDA) requesting and processing horizontal level soil properties based on the values from contained map units. Then a two-step weighting process is done for deriving MU average values for soil properties. First, the horizontal level soil values are weighted by their thicknesses and then the component values are weighted by their percentage composition. The aggregate soil property values are converted into a standard raster object with cell values containing soil properties. This
function was developed using R code and used functionality from multiple existing R packages such as SoilDB, SSOAP, raster. The results of this function are soil properties (i.e., soil hydraulic conductivity, plant available porosity, transmissivity) rasters for the watershed.

The “Get Climate Data” function enables the user to download Daymet precipitation, temperature and vapor pressure data for a specific time period for a specific watershed. This function was developed using R code and used functionality from existing “DaymetR” and “raster” packages to automate the steps required for data downloading (Figure 3.5). This function first projects the watershed raster into a geographic coordinate system and converts raster grids into points. The converted points are used as input in the DaymetR’s `batchdownloaddaymet` function to download Daymet climate data. This `batchdownloaddaymet` function translates the geographic coordinates into projected Daymet (x, y) coordinates, which are used to extract daily weather from the Daymet database of daily-interpolated surface weather variables.

The “Get Streamflow data” function enables the user to download and process USGS daily streamflow data automatically (Figure 3.6) for a specific time period and USGS stream gage. Using a USGS gage number, start date, and end date as input, it uses the `readNWISdv()` function from the USGS’ R `dataRetrieval` package for downloading the data from NWIS.

### 3.5.2 TOPNET Model Specific Services

The TOPNET model is designed to simulate runoff originating from two dimensional area features referred to as “subwatersheds”. Runoff is then routed along a
linear channel network and accumulated at the point of interest referred to as a “node”. Each node has a direct contributing watershed, referred to as a ‘node catchment’. The association between nodes and node catchments is defined in the nodelinks table and is obtained by using the “Create Reach and Node Link” function. This function also provides information about reach linkage (e.g., reach identifier and upstream reaches) and properties (e.g., slope and length) for each of the reaches in the reachlink and reach properties files. It uses the stream network tree file and network coordinates file obtained from the ‘Watershed Delineation’ function for generating nodelink, reachlink and reach properties files. Details about the data structure of these files are given in Appendix A.

The topographic wetness index is used for estimating saturation excess runoff from excess precipitation on saturated areas in the TOPNET model. This runoff is delayed in reaching the outlet due to the time taken by within subwatershed travel, as well as travel in the stream network. A histogram of the downslope flow distances to the stream in each subwatershed is used to perform within subwatershed routing. The “Create Wetness Index and Distance to Stream Distribution” function uses topographic wetness index and distance to stream rasters as inputs and groups the values into bins for each subwatershed, tabulating the lower and upper bound of each bin and the proportion of the area within each bin, thereby providing a distribution of wetness index and distances to stream. The function is configured to have no more than 5% and 20% of the area in each bin, resulting on average in just over 20 and 5 wetness index and distance to stream classes correspondingly for each subwatershed. The data, which give the
distribution of each wetness index and distribution class for each subwatershed, are written to the input file ‘distribution.txt’ (Appendix A).

TOPNET model parameters (Table 3.2) are time invariant and describe the unchanging properties of the subwatersheds and are expressed at the spatial scale of a subwatershed. These parameters are derived by averaging over the grid cells within the subwatershed from soil and land use data. The “Create model parameters” function uses extracted soil, land use, and land cover data as inputs, and aggregate parameter values for each subwatershed are written into the model parameters text file (basispars.txt).

TOPNET is configured to derive aggregated subwatershed precipitation inputs as a weighted sum of point precipitation measurements. The weights associated with each gauge for each subwatershed are calculated as part of the preprocessing by the “Create Rainweight” function using linear interpolation based on Delaunay triangles, adjusted using an annual rainfall surface to account for topographic effects. The method for determining precipitation weights was used from Bandaragoda et al. (2004) and Tarboton et al. (2007). The “Create Rainweight” function evaluates these weights and writes them to the rainweight.txt (Appendix A) file for input to TOPNET. This procedure provides a way to estimate precipitation as a smooth surface based on nearby surrounding gauges, while at the same time adjusting point gauge values, which are often recorded at low elevation, for topographic effects that are represented by the annual precipitation surface.

3.6 Data Service Implementation and Example Application

The HydroDS data services described in the previous section were implemented on a server at Utah State University. To access the HydroDS data services, user
credentials (username and password) are required. We implemented simple security to require authenticated access to avoid common security vulnerabilities. Once credentials have been obtained, the data service’s functions can be called using a Python script. Users can call either a number of consecutive functions to automate the input data preparation steps, or each function can be called separately. The user needs to provide the bounding box around the watershed of interest, the approximate outlet location, specify a range of stream threshold values to get an optimum threshold value for defining the stream network and a range of time for the time series data in order to set up the model. Once complete, the user is provided with a link to the location where the processed data can be downloaded. An example is presented in the following sub section to illustrate the use of HydroDS data services for automating the input data preparation steps for the TOPNET model and then sharing the results in HydroShare.

3.6.1 Study Site

The Logan River watershed in northern Utah, USA was used as a study site for testing the HydroDS data services. It is located in the heart of the Bear River range with headwaters near the Utah-Idaho border. Figure 3.7 shows the location of the watershed together with its NED DEM. The elevation of the watershed varies between 1400 and 3000 m with an average elevation of 2306 m above mean sea level. The dominant land covers are deciduous and evergreen forest. The climate ranges from cold, snowy winters to hot, dry summers with average monthly temperatures ranging from -11°C in January to 25°C in July. Annual precipitation in the region is dominated by winter snowfall. Mean
annual flow is strongly correlated with annual precipitation, and peak discharges of the Logan River are associated with snowmelt from May through July.

3.6.2 Input Data Pre-processing Using Data Services

To create a TOPNET model input package using data services for the Logan River watershed, the following steps were followed (Figure 3.8):

1. Get approximate boundaries of watershed outlet location, in geographic coordinates, start and end time, stream threshold value. These inputs must be supplied by a user to start.

2. Get a subset of the NED DEM for the selected area using the “subset raster” HydroDS utility service.

3. Project and resample (re-grid) the DEM using the “project_resample_raster” HydroDS utility service.

4. Create an outlet shape file using the “create_outlet_shapefile” HydroDS utility service.

5. Project the outlet shape file using the “project_shapefile” HydroDS utility service.

6. Delineate the watershed using the DEM and outlet shapefile as inputs to the “Delineate Watershed” function.

7. Extract Soil properties data using the “Get soil data” function.

8. Download Daymet climate data using the “Get climate data” function.

9. Download stream flow data using the “Get streamflow data” function.

10. Calculate wetness index and distance to stream distribution using the “Create wetness index and distance to stream distribution” function.

11. Create node and reach link information using “Create node and reach link” function.
12. Extract land use data using the “project_clip_raster” HydroDS utility service.


14. Get a subset of PRISM annual rainfall grid for the selected area using the “subset raster” HydroDS utility service.

15. Project outlet shape file the “project_shapefile” HydroDS utility services


The TOPNET input package was also generated manually for the Logan River following Figure 3.1 described above. The manually derived TOPNET input files were then compared with those from Hydro-DS services and found to match well, validating the data services. However, minor differences (1%-5%) were found in subwatershed soil properties value due to use of gridded map unit key raster in HydroDS instead of map unit key shapefile for extracting and processing soil properties from SSURGO database. The gridded map unit key raster data have a 30 m cell size that approximates the vector polygon of the map unit key in an Albers Equal Area projection. This might result in mismatch of the map unit key value between raster and vector data set and consequently affect extracted soil properties.

The work to prepare a model input package was found to be much more efficient using the HydroDS data services. For the Logan River Watershed, a complete model input package can be prepared in 20 min, and the user does not need to remember the specific details of the sequence of steps to follow. Manually setting up a model for the Logan Watershed took 4 hours when done by the author, representing a knowledgeable
user familiar with the procedure. Learning and working out the procedure may take a new user week.

The TOPNET input data package created by the data service was then shared through HydroShare using HydroDS’s `create_hydroshare_resource` function which enables automatic sharing of TOPNET input package without manually uploading the data using HydroShare web page (https://www.hydroshare.org/). The purpose of sharing the TOPNET input package is to enable collaboration and sharing of information among a team working on this model and eventual publication of the complete data in support of research findings being published from the results, thereby enhancing research reproducibility and trust in the model results. Here the transfer to HydroShare occurred between servers, independent of the users' desktop system, a mode of working more amenable to large datasets.

The shared TOPNET input package was then downloaded from HydroShare to a local machine where parameters were calibrated and sensitivity analysis was performed. This demonstrated the suitability and usability of the Hydro-DS generated package in a typical hydrologic modeling exercise. Morris screening (Morris, 1992), a parameter screening method was performed to identify and rank parameters that have a significant impact on model outputs. This screening method was chosen as it is simple and particularly useful for computationally expensive models or multi-parameter models (e.g., TOPNET). It can screen out non-sensitive or unimportant parameters with a few runs of the model (Zhan et al., 2013). Figure 3.9 gives the results of the sensitivity analysis of the model parameters for two common metrics for the difference between
modeled and observed daily flow, namely Nash-Sutcliffe Efficiency (NSE) and Bias (Moriai et al., 2007). Three parameters (f, k, and dth1) were found to be most sensitive among twelve parameters for these measures. Then, these parameters were calibrated using the elitist multiobjective genetic algorithm (Deb, 2001) considering NSE and Bias as objective functions for the period of 2003 to 2006. Model performance was evaluated visually (Figure 3.10) and statistically (NSE, Bias). The TOPNET model was able to mimic the hydrological characteristics of the watershed reasonably well, with mean NSE and bias of daily runoff 0.8 and 0.92 respectively.

3.7 Discussion and Conclusion

Physically based models require diverse geospatial and time series data, which can be obtained from different sources. Due to the level of heterogeneity across different data sources, acquisition and processing of input datasets are labor intensive and difficult to automate. This paper describes the web-based data services developed to automate the retrieving and processing of geospatial and time series data to set up a physically based hydrologic model entirely using server functionality (or in other words in the "cloud" on servers remote from the user) without the use of data or software on the user’s desktop computer. The functionality developed was demonstrated using the TOPNET model and Logan River watershed as a case study. Data for the entire western U.S. were deployed on the server hosting our data services to evaluate the serving of this functionality across this region.

Automation of input data preparation steps enables water scientists to spend less time extracting and formatting data while making it easier for modeling workflows to be
shared and scientific results to be reproduced (Miles, 2014). The ‘Delineate Watershed’ function allows users to delineate a watershed and stream network without remembering all of the sequential steps required to complete the watershed delineation. The ‘Get Soil and Land use data’ tool allows users to extract and process soil and land cover data rapidly for any watershed in the area supported in a matter of minutes. The ‘Get Climate data’ function enables users to download climate data on the fly without going to the specific website and using specific tools.

Our approach for developing web services saves users from the complexity of installing and configuring software tools that may have complicated to set up dependencies on their computer. Instead a knowledgeable developer does this once for the server and enhances reproducibility of the workflows. One of the most attractive benefits of the web services is that users can access them from any computer connected to the Internet. It also allow users to get new functionality or software immediately without upgrading their PC as the the upgrades are installed on the host server. More generally, the outcome of this work is a methodology for creating server-side data processing services that could be applied for other data-intensive hydrologic, environmental, and Earth system models.

The tools described here were developed using open source, freely available scripting language and programs. The code is publicly available in a pubic GitHub repository (https://github.com/nazmussazib/TOPNET_PreProcessing) so that the user community outside the initial development team can participate in future improvements of the software by integrating new approaches and analytical techniques.
At present, these data services have some limitations. First, they do not allow users to execute the TOPNET model on the web. A next step is to develop new functionality for executing the model in the web. Secondly, at present the Hydro-DS server hosted NED DEM, PRISM annual rainfall grid, NLCD land use and land cover data for the CONUS but soil map unit raster for the western US only, which limits application of EHV services outside of western U.S. However, the EHV services can be easily expanded to any watershed in the continental U.S. (CONUS) by including soil map unit key raster for the CONUS. Similarly, the watershed delineation function could be extended to watersheds outside of the CONUS using the ASTER Global DEM (https://lpdaac.usgs.gov/products/aster_products_table/astgtm). However, the function for retrieving climate datasets cannot extend outside of the CONUS because the underlying dataset is only available for the CONUS.

Despite these limitations, our web services for automation of the input data preparation steps have significant potential to facilitate efficient hydrologic modeling. Additionally, the data services we developed not only benefit hydrologic modelers but may also benefit scientists from other disciplines who need to locate and analyze spatial and time series data.
Tables and Figures

Table 3.1 HydroDS Essential Hydrologic variable data service functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Delineate Watershed</strong></td>
<td>This function delineates watersheds and a stream network for a user selected spatial domain and outlet location. It also provides a stream network topology file (TauDEM tree file) that contains network topological connectivity information, and stream network coordinates file (TauDEM coord file) which stores coordinates and attributes from each grid cell along the network.</td>
</tr>
<tr>
<td><strong>Get soil and land use data</strong></td>
<td>This function performs automated extraction of soil properties (e.g., soil hydraulic conductivity, transmissivity, and porosity) and land cover data for the delineated watershed.</td>
</tr>
<tr>
<td><strong>Get climate data</strong></td>
<td>This function retrieves Daymet precipitation, temperature, and vapor pressure data from the Daymet website for a specific time period for the domain of interest.</td>
</tr>
<tr>
<td>Function</td>
<td>Description</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Create node and reach link</td>
<td>This function uses the stream network tree file and network coordinates file obtained from the ‘Delineate Watershed’ function for generating TOPNET nodelink, reachlink and reachproperties files.</td>
</tr>
<tr>
<td>Create wetness index and distance to stream distribution</td>
<td>This function is used to group the values of topographic wetness index and distance to stream into bins for each subwatershed, tabulating the lower and upper bound of each bin and the proportion of area within each bin.</td>
</tr>
<tr>
<td>Create model parameters</td>
<td>This function uses extracted soil, land use and land cover data for estimating the model parameters for each subwatershed.</td>
</tr>
<tr>
<td>Create rainweight</td>
<td>The weights associated with each gauge for each subwatershed are calculated by this function using linear interpolation based upon Delaunay triangles, adjusted using an annual rainfall surface to account for topographic effects.</td>
</tr>
<tr>
<td>Name</td>
<td>Description</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>$f (m^{-1})$</td>
<td>Saturated store sensitivity</td>
</tr>
<tr>
<td>$K_o (m/h)$</td>
<td>Surface saturated hydraulic conductivity</td>
</tr>
<tr>
<td>$\Delta \theta_1$</td>
<td>Drainable porosity</td>
</tr>
<tr>
<td>$\Delta \theta_2$</td>
<td>Plant available porosity</td>
</tr>
<tr>
<td>$d (m)$</td>
<td>Depth of soil zone</td>
</tr>
<tr>
<td>$\psi_f (m)$</td>
<td>Wetting front suction</td>
</tr>
<tr>
<td>$CC (m)$</td>
<td>Canopy capacity</td>
</tr>
<tr>
<td>$C_r$</td>
<td>Evaporation adjustment factor</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Albedo</td>
</tr>
<tr>
<td>$T$</td>
<td>Transmissivity</td>
</tr>
</tbody>
</table>
Figure 3.1 Input data pre-processing steps for TOPNET model.
Figure 3.2 Architecture of HydroDS data services.

Figure 3.3 Watershed delineation processing using TauDEM. The boundary identified by the dotted line represents the steps that are automated in the “Delineate Watershed” function.
Figure 3.4 Steps required for retrieving soil data from the SSURGO soil data base. The boundary identified by the dotted line represents the steps that are automated in the “Get soil and land use” function.

Figure 3.5 Daymet climate data retrieving and processing steps. The boundary identified by the dotted line represents the steps that are automated in the “Get climate data” function.

Figure 3.6 Streamflow data downloading steps. The boundary identified by the dotted line represents the steps that are automated in the “Get streamflow data” function.
Figure 3.7 Study site location and DEM of the Logan River Watershed.

Figure 3.8 TOPNET data preparation steps using HyrdoDS data services.
Figure 3.9 Morris sensitivity analysis of the TOPNET model parameters.
Figure 3.10 Comparison of simulated and observed streamflow for the Logan River Watershed.
References


Bivand, R., Keitt, T., Rowlingson, B., 2016. rgdal: Bindings for the Geospatial Data Abstraction Library.


Hufkens, K., 2010. DaymetR: Batch download single pixel or gridded DAYMET data.


Lang, D.T., 2012. SSOAP: Client-side SOAP access for S.


CHAPTER 4

ENHANCEMENTS TO HYDROLOGIC DIGITAL ELEVATION MODEL
ANALYSIS SOFTWARE TO SUPPORT RAPID WATERSHED
DELINÉATION

Abstract

Watersheds are widely recognized as the basic functional unit for water resource analysis. Terrain Analysis software support the delineation of watersheds within desktop Geographic Information Systems (GIS). However, traditional watershed delineation processing can be laborious and troublesome if the Digital Elevation Models (DEM) and other raster datasets necessary for watershed delineation are large due to high resolution or large geographic area. Further, this processing becomes more burdensome if the number of the watersheds to be delineated is large, as one has to fill sinks, process flow direction and flow accumulation for each outlet site. Additionally, current terrain analysis software often supports specific format of raster, vector and coordinate system data which limits their usability for other format data. This paper presents enhancements to terrain analysis software that have been developed to extend its generality and support web-based rapid watershed delineation services over a large area. The enhancements include (1) reading and writing raster data with the open-source Geospatial Data Abstraction Library (GDAL/OGR) to overcome file format limitation and (2) support for both geographic and projected coordinates. To support web services for rapid watershed

---

3 Coauthored by Nazmus Sazib and David G. Tarboton
delineation, a procedure was developed for sub setting the domain based on subwatersheds, with preprocessed data prepared for each subwatershed stored in a folder hierarchy. This allows the watershed delineation to function locally, while extending to the full extent of watersheds using preprocessed information. The functionality to support the web-based watershed delineation methodology was developed and tested for the Delaware River Basin. Additional capabilities of the rapid watershed delineation program include computation of average watershed properties and geomorphologic and channel network variables such as drainage density, basin relief and relief ratio. The advancements of terrain analysis software increase practical applicability in terms of raster and vector data type, size and coordinate system. The watershed delineation web service functionality is useful for web-based, software-as-a-service deployments that alleviate the need for users to install and work with desktop GIS software.

**Software availability:**

Name of Software: TauDEM 5.3.5, RWD

Developer: Utah State University

Hardware requirements: Windows PCs, Linux

Programming language used: C++, Python

Availability: TauDEM 5.3.5 and RWD are publicly available at [http://hydrology.usu.edu/taudem/taudem5/index.html](http://hydrology.usu.edu/taudem/taudem5/index.html) and [https://github.com/nazmussazib](https://github.com/nazmussazib) respectively.
4.1 Introduction

Watersheds are commonly used in water resource analysis to define the scope of a modeling domain and can be derived from Digital Elevation Models (DEM) using Geographic Information System (GIS) technology. Many studies have been performed that examine hydrologic terrain analysis and methods for watershed delineation from DEMs, and their application in terrain analysis software tools. The scientific literature includes many examples of both the background of the methods (Marks et al., 1984; Band, 1986; Jenson and Domingue, 1988; Fairfield and Leymarie, 1991) and application of the techniques such as ArcHydro (Maidment, 2002), TauDEM (Tarboton, 2002), Whitebox (Lindsay, 2014) and BASINS (EPA, 2015).

However, common limitations of existing terrain analysis software are: (1) they only support specific types of raster data, such as TauDEM support for the tiff format; (2) they are time-consuming to execute for large watersheds, and analyses may be limited due to memory limitations of the computer being used; (3) input data must be in a projected coordinate system; which has limitations as it alters (e.g., smooths and sometimes introduces artificial stripes) the data and results in distortions in the data when large areas are processed (e.g., continents), and (4) they support specific formats of vector data, such as TauDEM's support for only ESRI shapefile format vector data. Current approaches also rely heavily on GIS desktop software, which can have a steep learning curve for those unfamiliar with the software, and tedious data preparation steps to arrive at the desired watershed boundary dataset (Djokic, 1999; Kopp, 2013). Additionally, most of the terrain analysis software need to be installed on a particular
operating system and some require purchase of a software license. These factors limit their usability.

Web-based tools for watershed delineation have started to emerge. These have the potential to overcome some of the limitations associated with desktop computers such as alleviating the need for users to purchase, install, and work with desktop GIS software or download and work with large data files. For example, the U.S. Geological Survey has developed a nationwide program called ‘StreamStats’ (Ries et al., 2009) a web-based program that can provide commonly used streamflow measures at gauged and ungauged sites; it can also delineate watersheds and provide other useful watershed attributes at a user specified location (http://water.usgs.gov/osw/streamstats/). The Environmental Systems Research Institute (ESRI) developed a web service to delineate watersheds by identifying a location a user is interested in finding a watershed for, and whether the user wants to move (snap) input to the location of the nearest stream (http://resources.arcgis.com/en/help/). The CyberGIS-based TauDEM application is another example of a web-based service that provides an online environment for computing and data-intensive watershed delineation and hydrological analyzes on advanced cyberinfrastructure (Fan, 2014). However, a common drawback of those web based services is the use of low resolution DEM that does not represent the small-scale feature properly which can significantly affect the accuracy and reliability of the delineated watershed boundary and stream network.

Considering the background information presented in the previous paragraphs, there is a need to advance terrain analysis software to extend their usability to support
rapid watershed delineation over a large area. This paper describes the enhancement of terrain analysis software to support a rapid, web-based watershed delineation service. We chose to implement, demonstrate, and deploy this service using TauDEM, which is open source and commonly used for terrain analysis and watershed delineation. This study enhanced the TauDEM software and used it to develop a set of terrain analysis data for the entire Delaware River Basin at 10 m resolution to implement visualization capabilities in the Model My Watershed version 2 (MMW2) web user interface (https://app.wikiwatershed.org/). This study also developed new Rapid Watershed Delineation (RWD) functionality, based on TauDEM to derive watershed boundaries and attributes quickly for any watershed within the Delaware River Basin to support a web-based watershed delineation service for MMW2.

MMW2 provides a public web application to visualize maps of data and model outputs that are relevant to understand the effects of current and possible future land uses and watershed restorations on storm water runoff and stream water quality for any watershed within the entire Delaware River Basin. The need for a highly interactive website requires a capability to quickly delineate a watershed to an arbitrary (user input) outlet location anywhere in the area. The enhancements to terrain analysis software developed for MMW2 increase the practical applicability in terms of raster data type, size and coordinate system. The RWD functionality used in the MMW2 web platform reduces considerably the time required for watershed delineation and determination of basic watershed characteristics within a large area.
The paper is organized as follows: Section 4.2 presents the enhancements to TauDEM and rapid watershed delineation methodology. Section 4.3 describes the software implementation for TauDEM and RWD application to the Delaware River basin where TauDEM was used to produce terrain analysis products for display on the web, and the RWD code was deployed. Section 4.4 gives discussion and conclusions.

### 4.2 Enhancement to TauDEM and Rapid Watershed Delineation Methodology

#### 4.2.1 Enhancements to TauDEM

TauDEM was enhanced to support diverse geospatial raster and vector data to increase its applicability in terms of data type and size. The general purpose GDAL library (GDAL Development Team, 2015) was used to replace the custom TauDEM Tagged Image File Format (TiffIO) library to read and write raster data. GDAL is an open-source tool providing a single abstract data model to read and write a variety of geospatial raster data formats that has been widely used in raster-processing applications.

TauDEM uses a domain decomposition parallel implementation strategy. The processing domain defined by the extent of the input rasters is divided into equal parts based on the number of processes with any extra portion remaining being attached to the last partition (Wallis, 2009). With this strategy there is a need for each process to read in the data for the part of the domain it is responsible for, and to write the data for the part of the domain it computed at the end. Message passing between the processes is used where information needs to cross partitions. Wang (2012) have applied GDAL to parallel geospatial raster processing using a serial I/O mode for accessing raster data. In this approach, a master process takes charge of the entire I/O process between external and
internal memory, and other work processes access the data by communicating with the master process. The serial I/O in this approach can create a bottleneck when the size of the raster file exceeds the memory capacity of a single computer node. In enhancing TauDEM, we used a parallel read and sequential write method for reading and writing raster files. In the file read method, the master process first extracts the metadata (e.g., spatial extent and projection) from a raster file. Then, according to the domain partitioning, the master process sends the spatial extent information of each subdomain to the corresponding work process. Then, each process opens the shared input raster file and reads a partition from the domain assigned to it based on processor rank using the GDAL RasterIO function. This is done in parallel. To write results after computing, the master process (process 0) creates the output raster file with correct metadata information, and then each process writes back its local result into the corresponding part of the raster. The write IO operation is performed sequentially in a ring fashion because, at present, GDAL does not support parallel file writes.

The OGR library (GDAL Development Team, 2015) was used instead of shapelib (http://shapelib.maptools.org/) for reading and writing vector files. OGR supports access to a large number of vector data formats, including KML, ESRI shapefiles, GeoJSON, and SQLite. Objects in OGR include layers (OGRLayer), features (OGRFeature), and geometries (OGRGeometry). The OGRLayer contains the OGRFeature. The OGRFeature contains attribute values and a reference to the feature geometry. The OGRGeometry is an abstract base class, implemented by specific subclasses for the representation of point, multipoint, line, multiline, polygon, and multipolygon geometries.
To read a file type supported by OGR, the first step is to register all the format drivers and then open the data source and corresponding layer. An input argument is used to select the specific layer to use when an OGR dataset with multiple layers is provided as input. Next, the geometry type of the input layer is extracted to check its validity, as TauDEM only takes inputs from the point geometry type. Then, the spatial reference of the layer is extracted and compared with the input raster spatial reference to check for data consistency. The next step is identification of the field with name ‘Id’ used by TauDEM and reading information from it into a feature identifier array.

For writing an OGR file, all the drivers are first registered and then the driver name is fetched based on the user provided output file name. Next, data source and output layer are created based on the output file name, geometry type and spatial reference information. Attribute fields are created on the layer before any features are written on the layer. Field width and precision are then initialized in the output file. To write a feature to disk, an OGRFeature and a new geometry object is created and values of the attribute fields set. Finally, the data source is closed in order to ensure headers are written out properly.

TauDEM was also extended to support both projected and geographic coordinate system raster data. The raster data model holds data values on a grid with cell size in each direction specified (dx and dy). In projected coordinates, dx and dy are in length units. In geographic coordinates, cell sizes are in degrees, with the corresponding length on the surface of the earth implied by the Earth Ellipsoid of the projection (NIMA, 1997). The degree to length conversion is a function of latitude and earth ellipsoid parameters. A
function `geotolength` was developed to calculate cell lengths from latitude and ellipsoid parameters. The previous version of TauDEM used a single dx and dy for the entire domain, but here, since these vary with latitude we extended the internal raster class to use arrays to store them for each row of the raster. The quantities used in the evaluation of `geotolength` from ellipsoid parameters are illustrated in Figure 4.1 (Bekir, 2007). \( \phi \) is the latitude defined as the angle at the major axis from the tangent normal (equatorial plane).

The Ellipse equation is given by:

\[
\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1
\]  
(4-1)

where \( a \) and \( b \) are ellipse major and minor axis and \( x \) and \( y \) are coordinates of points on the ellipse, can be represented by:

\[
x = a \cos \beta
\]  
(4-2)

\[
y = b \sin \beta
\]  
(4-3)

where \( \beta \) is the angle from the center used to define the ellipse parametrically. For given \( \phi \),

\[
\tan \beta = \frac{b}{a} \tan \phi
\]  
(4-4)

\[
\beta = \arctan \left( \frac{b}{a} \tan \phi \right)
\]  
(4-5)

The N-S length along an arc may be evaluated by differentiating this with respect to latitude, expressing the result in terms of changes in \( x \) and \( y \) and evaluating arc-length using,

\[
ds^2 = dx^2 + dy^2
\]  
(4-6)
This gives the N-S length along an ellipsoid arc used for the north-south cell size array as:

\[ dy = \sqrt{(a \sin \beta)^2 + (b \cos \beta)^2} \, d\beta \]  

Similarly the E-W length is evaluated using the radius from the earth axis and longitude cell size.

\[ dx = rd\lambda = a \cos \beta \, d\lambda \]  

Equations (4-7) and (4-8) are used to calculate grid cell lengths from geographic latitude and longitude, where \( d\lambda \) represents longitude cell size increment.

The changes to the TauDEM functions also required changes to the TauDEM ArcGIS toolbox graphical user interface parameters, validation code, and help files. The parameter data type in the ArcGIS tool box for vector layers was changed from shapefile to feature layer to allow users to use multiple formats of vector data as input and output. The display name and validation code parameters were also changed to support multiple raster formats in the TauDEM functions. The command line for each function was included in the tool help file so that users can learn how to run functions through the command line window instead of through the ArcGIS tool box.

**4.2.2 Rapid Watershed Delineation (RWD)**

Rapid Watershed Delineation (RWD) functionality was developed to delineate a watershed boundary and to derive watershed attributes (Table 4.1) rapidly over a large area from a user selected point clicked on a map. For this, a scheme was developed to subdivide a large area into subwatersheds with preprocessed results, with connectivity information stored for each subwatershed. RWD then involves delineating a local
subwatershed to an arbitrary point and merging it with attached, preprocessed, adjoining watersheds.

The RWD methodology consists of data preprocessing steps for creating required preprocessed rasters (Table 4.2) and an algorithm for using preprocessed data to delineate a watershed from an arbitrary point over a large area. Prior to preprocessing, we adjusted the DEM to make it consistent with geographic shape information on stream hydrography, oceans, and estuaries. Grid cells within the ocean/estuary were lowered within a buffer distance of the shoreline, and beyond this were set to no data. This ensured that flow enters the ocean and then limits the algorithms from attempting to process elevation data within the ocean/estuary. This was done by using a number of ArcGIS geoprocessing tools (e.g., Feature to Raster and Map algebra) as well as Environment Settings to control raster cell size and extent. Grid cells along stream lines mapped at high resolution were lowered using a new TauDEM “FlowDirCond” function we developed, so that tracing along a stream line never results in an increase in elevation. The FlowDirCond function recursively traverses the subset of grid cells that have flow direction input, and lowers the elevation of any downstream grid cell that has elevation higher than an upstream grid cell, to the same value as the upstream grid cell. Input flow directions are constructed to be down input vector stream lines. This removed obstacles present in the DEM that sometimes occurred when it, for example, represented the road top elevation of a bridge crossing a river. These adjustments were not strictly necessary, and should only be used when this external information adds value. If the DEM is of high
fidelity and stream line hydrography is poor, then stream adjustments would be
detrimental.

The first preprocessing step was to run TauDEM functions (Pitremove, D8 flow
dir, D8Area, Stream definition by threshold, gridnet and Streamnet) to obtain terrain
property grids derived from the input DEM (i.e. hydrologically conditioned elevation,
slope, flow direction, flow accumulation, stream order, total length of streams draining to
each grid cell, and the stream network). The Pit Remove function identified all pits in the
DEM and raised their elevation to the level of the lowest pour point around their edge.
The D8 flow Dir function determined flow direction from each grid cell to one of its
adjacent or diagonal neighbors. The D8Area function calculated a grid of contributing
areas using the single direction of D8 flow model. The Stream definition by threshold
function was applied to the D8 contributing area to compute a preliminary stream grid
based on an input selected threshold value (5000 cells). Next, the StreamNet function
created a preliminary stream network. This stream network with this fixed area threshold
does not represent the full resolution drainage density, but was used for two purposes.
First it was used to define downstream outlets that drain to the ocean. The watersheds
upstream of these outlets served to define the domain over which the geomorphologically
derived stream network was computed, and for which RWD was implemented,
effectively excluding areas less than 0.5 km² draining directly to the ocean. Second, it
was used to decompose the domain into subwatersheds.

To derive a geomorphologically based stream network we used the TauDEM
implementation of the Peucker and Douglas algorithm (Peucker and Douglas, 1975), to
identify a stream network skeleton. These skeleton grid cells were used as input to a weighted contributing area calculation and thresholds of 20, 50 and 100 used to derive stream networks of different drainage density. Comparison of these with the topography and high resolution mapped streams suggested that the threshold of 50 was best, so this threshold was used to delineate the final geomorphologically based stream network. This stream network is geomorphologically based due to use of the Peuker and Douglas skeleton which makes it sensitive to the natural variability in drainage density with variations in terrain geomorphology. This final stream network served as a background layer for visualization in the MMW2 website and was used to calculate the distance to stream, stream order and total length of streams upslope of each point rasters by using \textit{D8 Distance to stream} and \textit{gridnet} functions. Additional terrain property rasters such as slope weighted contributing area and maximum upslope elevation were calculated by running \textit{D8 Area} and \textit{D8extreme upslope} functions, to support derivation of watershed properties such as average slope and basin relief.

RWD is based on a partition of the domain into subwatersheds. The downstream end point of each stream reach in the initial (from 5000 cell threshold) stream network was identified as an outlet point that serves to seed the decomposition into subwatersheds. The initial stream network with uniform drainage density helps produce subwatersheds of roughly equivalent size. To avoid very small subwatersheds, outlet points that have another outlet point within a defined threshold distance (2.5 km) downstream were removed. The remaining outlets were used to subdivide the area into multiple (thousands of) subwatersheds draining to outlet points (Figure 4.2a and b).
Since each subwatershed is associated with one downstream subwatershed, this information was used to identify all upstream subwatersheds that contribute to a subwatershed. These were merged together to form the complete watershed associated with each subwatershed (Figure 4.2c). Then, a buffer around each subwatershed was created using a `create_buffer` function which was developed using functions in the OGR library. The purpose of creating a buffer is to avoid inaccurate evaluation of terrain properties due to truncation of the raster that results in omission of contributions from outside the computation area (edge contamination). Then, the terrain property grids for each of the buffered subwatersheds were extracted from existing, preprocessed TauDEM products (Figure 4.2d) and a directory was created for each subwatershed based on its ID number. All TauDEM preprocessed terrain property grid were kept in a `Main_Watershed` directory, and extracted terrain properties and complete watershed for each subwatershed were kept in the corresponding subwatershed directory (Figure 4.3). This organization of preprocessed data support the RWD algorithm that enables users to rapidly delineate a watershed in a large area described in the next section. All of the above preprocessing steps were automated by using a Python script `PreProcessorForRapidWatershedDelineation.py`.

### 4.2.3 RWD Algorithm

The RWD algorithm takes as input the coordinates of the point a user has clicked and a flag to indicate whether the watershed is to be delineated exactly from the point clicked, or from the nearest downslope point downslope on a stream. The following procedure is then performed (Figure 4.4):
Step 1: Given as input the user defined point of interest (point where the user clicks) and option for repositioning the point of interest to the nearest stream, identify the subwatershed ID in which this point resides and go to corresponding subwatershed directory for accessing necessary files (e.g., flow direction, stream and distance to stream rasters) required for local watershed delineation. These are accessed from the preprocessed data structure (Figure 4.3).

Step 2: Check if the point of interest is positioned on the stream (using subwatershed stream raster). If not, and if the input flag indicates to snap to the stream, run the TauDEM moveoutletstostream function to reposition the point to lie on the stream.

Step 3: Delineate the local watershed and identify any adjacent subwatersheds that drain to the outlet point using the TauDEM Gagewatershed function which uses subwatershed flow direction and subwatershed raster information as inputs. The logic for identifying the adjacent subwatershed ID is to trace upwards from the outlet point, marking all cells in the local watershed that have subwatershed grid value the same as subwatershed grid value of outlet point. If a point with a different subwatershed ID is encountered, the new subwatershed ID is added to the list of adjacent subwatersheds whose watershed is to be merged.

Step 4: Check the number of adjacent subwatersheds and merge with the local watershed. If the number of adjacent subwatersheds is greater than or equal one, then the local watershed is merged with adjacent watersheds using ogr2ogr ST_Union function to
form the complete watershed for the outlet point. If not, the local watershed delineated in step 3 is considered as the complete watershed.

Initially, merging a local subwatershed and adjacent subwatersheds was the most time-consuming operation in RWD. The initial development used GDAL and Python package library (e.g., Fiona and shapely) function for merging and dissolving watersheds. Following investigation we found that using the ogr2ogr ST_Union function instead of the function we developed, was able to speed up the merging task by a factor of 3.

Step5: Get the watershed attributes for the complete watershed. Watershed attributes (Table 4.1) are extracted for the complete watershed by

extract_value_from_raster function which was developed by using GDAL. For an input point of interest, this procedure creates two shapefiles: (1) the outlet point moved to the stream if move to stream was specified, and (2) the watershed boundary polygon with attributes. The RWD algorithm was implemented in “Rapid_Watershed_Delineation.py” and the source code is available on GitHub.

4.3 Software Implementation

4.3.1 Implementation of TauDEM Enhancements

The source code of TauDEM was changed to implement the methodology described in section 4.2.1 for enhancements to TauDEM. Reading and writing of raster data code resides in TauDEM’s TiffIO library, which was changed to support GDAL data types. Major changes in this library included: (1) a new constructor was created for getting metadata and spatial extent information and for calculating pixel width and depth
using the *geotolength* function, and (2) two separate functions were added for reading and writing raster data using GDALRasterIO functions.

TauDEM uses a striped partitioning approach where an input grid is divided horizontally into equal parts based on the number of parallel computation processes used to run the program. This data decomposition code resides in *linearpart.h* file. A new function *savedxyc* was added in *linearpart.h* file to store the array of pixel width and height for each partition. A number of TauDEM functions (e.g., D8 distance to stream and Dinfinity distance to stream) use pixel width and height locally and globally. The source code of those functions, where the pixel width and height were used locally, were replaced by the array of pixel width and height to support working with data in a geographic projection system.

TauDEM’s *readoutlet* function, which is used for reading vector data, was modified based on the methodology described in Section 4.2 to support multiple formats of vector data. The source code of the TauDEM functions (e.g., *moveoutletstostream* and *connectdown*) that write vector data was also changed to support OGR data types.

The updated TauDEM functions were tested to evaluate their applicability in terms of data type and coordinate system. To do this, a script was developed to run a series of TauDEM command line functions. The goal of running this script was to test each TauDEM function with each possible combination of multiple formats of raster and vector data as inputs and outputs. The output results were checked visually and were also compared with reference results for validation. The script, inputs, outputs and referenced results we used for testing are available on HydroShare (Sazib, 2016).
4.3.2 RWD Implementation on Delaware River Basin

RWD was developed to provide watershed delineation functionality for the Model My Watershed website (https://app.wikiwatershed.org/), supporting web based hydrologic and water quality analysis in the Delaware River Basin. The Delaware River Basin was chosen because it focused on the Model My Watershed website that drove the development of this functionality. The Delaware River drains to the Delaware Bay where it meets the Atlantic Ocean. In all, the basin contains 13,539 square miles, draining parts of Pennsylvania (6,422 square miles or 50.3 percent of the basin's total land area); New Jersey (2,969 square miles, or 23.3%); New York (2,362 square miles, 18.5%); and Delaware (1,004 square miles, 7.9%). The Delaware River basin DEM has 26514x48687 grid cells with cell size 10x10 m and requires more than a week to delineate a watershed due to processing burden of working with that large DEM, using TauDEM without RWD. We followed the procedure described in section 4.2.2 for preprocessing of required data to support rapid watershed delineation from an arbitrary point inside the Delaware River basin. Then, the RWD application was tested and compared with traditional watershed delineation method (deriving hydrologically condition DEM, calculating flow direction, flow accumulation, distance to stream etc) for multiple outlet points in the Delaware River Basin on a PC with 3.40 GHz CPUs with 32 GB random-access memory. The RWD method completed the largest delineation significantly faster than the traditional method. For instance, it finished approximately 623 time faster than the traditional method for the watershed with an area of 16922 km² as shown in Table 4.3.
4.3.3 Web based Rapid Watershed Delineation

The preprocessed data were used for web based watershed delineation services deployed at https://app.wikiwatershed.org/. When the website opens, the user is presented with a map showing political boundaries including terrain as a background data set (Figure 4.5). Once zoomed in sufficiently within the Delaware River Basin, the user activates the “Delineate Watershed” tab for watershed delineation. The user clicks “Stream Network” which triggers a cursor that can be clicked anywhere in the Delaware River basin for watershed delineation (Figure 4.6). Then the web service converts the clicked point to a geographic coordinate and passes that coordinate to the RWD function which returns a polygon that represents the watershed boundary (Figure 4.7) for the clicked point as output. The user can use the reset button to delineate a watershed for any location. Our work on this website was limited to development and testing of the RWD functions. The deployment was done by Azavea Inc.

4.4 Discussion and Conclusion

This paper presents enhancements to existing terrain analysis software to extend its generality and to support web-based watershed delineation services. The enhancements were implemented using the TauDEM software. Enhancements of TauDEM include (1) reading and writing raster data with GDAL, thus removing TauDEM’s limitation to use only the tiff data format for raster data, (2) reading and writing vector data with OGR, thus removing TauDEM’s limitation to using only the shapefile format for vector data, and (3) support for data stored in both geographic and projected coordinate systems. To support web-services for RWD, a procedure was
developed for sub setting the domain based on subwatersheds, with preprocessed data prepared and stored for each subwatershed. This allows a watershed to be delineated from an arbitrary point by processing a small local section of the data, then merging the result with preprocessed adjoining watersheds, thus producing a result very rapidly.

The implementation of GDAL and OGR in TauDEM increases its usability through support for multiple raster and vector formats. By relying on a third party library, this terrain analysis software was enhanced and made more sustainable as the GDAL/OGR library is maintained by others. Additionally, as raster and vector formats are added to GDAL/OGR, they will more easily become available to TauDEM, exemplifying the benefit to terrain analysis software of reliance on a general purpose library.

A new method, using latitude dependent grid cell dimensions was also developed to support using geographic coordinate system data in terrain analysis software. This methodology avoids the degradation in DEM quality that occurs when projecting data originally obtained in geographic coordinates. This feature is especially important when working with DEMs at the regional, continental or global scale.

RWD allows users to delineate a watershed and extract watershed properties across a large area rapidly. Starting from the DEM and using 5-8 TauDEM commands, it may take several hours to days, even using High Performance Computing (HPC) for delineating a watershed for any given point in an area as large as the Delaware River Basin. Sometimes it may not even be possible using a PC, as datasets may exceed the available computer memory, or may be too large for the available GIS algorithms. The
RWD resolves those problems by subdividing a large area into subwatersheds with preprocessed results, and connectivity information stored for each subwatershed. Storage of preprocessed results effectively trades storage for on demand computation and enables on demand results to be generated by processing only a small area (subwatershed scale) and then merging with preprocessed information.

The RWD tool takes advantage of the fact that required processing such as pit filling, flow direction calculation, etc., need only be run once for the large area. Once the required inputs (i.e., the outlet shapefile and maximum snapping distance to a stream) are put together, the RWD tool can easily delineate the watershed. Another advantage of this tool is that it can help users to delineate a watershed with a single click without needing to remember the sequence of steps to follow to complete the watershed delineation process.

The web-based watershed delineation service alleviates the need for users to install and work with desktop GIS software, thereby overcoming compatibility limitations and enhancing reproducibility. This also allows users to delineate watershed from any computer connected to internet therefore increases usability.

The enhancements to existing terrain analysis software for RWD make it easier and more efficient for researchers involved with terrain analysis and those working on hydrologic and environmental modeling to use watershed delineation and other DEM derived products in their research. The approaches and techniques developed in this study could be applied in other GIS and terrain analysis systems, or extended to other areas of terrain analysis, and hydrologic modeling.
### Tables and Figures

Table 4.1 Watershed attributes for rapid watershed delineation.

<table>
<thead>
<tr>
<th><strong>Watershed Attributes</strong></th>
<th><strong>Definition</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
<td>Area of the watershed in the horizontal units of DEM</td>
</tr>
<tr>
<td>Perimeter</td>
<td>Perimeter of the watershed in the horizontal units of DEM.</td>
</tr>
<tr>
<td>Basin Length</td>
<td>Flow length from the furthest point to the outlet.</td>
</tr>
<tr>
<td>Stream Order</td>
<td>The relative size of the streams, which is calculated based on Strahler order.</td>
</tr>
<tr>
<td>Basin Relief</td>
<td>Difference in the elevation of the highest point of a basin and the lowest point on the valley floor.</td>
</tr>
<tr>
<td>Relief Ratio</td>
<td>Ratio between the Basin Relief and Basin Length.</td>
</tr>
<tr>
<td>Total Stream Length</td>
<td>Total length of the stream of all orders.</td>
</tr>
<tr>
<td>Drainage density</td>
<td>The total length of the stream of all orders divided by drainage area.</td>
</tr>
<tr>
<td>Length of Overland flow</td>
<td>One-half the reciprocal of the drainage density.</td>
</tr>
<tr>
<td>Mean Slope</td>
<td>Average of slopes from all elevation grid cells.</td>
</tr>
</tbody>
</table>
Table 4.2 Inputs required for rapid watershed delineation.

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pit filled DEM raster</td>
<td>Calculate watershed attributes</td>
</tr>
<tr>
<td>Flow direction raster</td>
<td>Delineate watershed</td>
</tr>
<tr>
<td>D8 Contributing Area raster</td>
<td>Calculate watershed attributes</td>
</tr>
<tr>
<td>D8 Contributing Area with slope as weight raster</td>
<td>Calculate watershed attributes</td>
</tr>
<tr>
<td>Extreme upslope raster</td>
<td>Calculate watershed attributes</td>
</tr>
<tr>
<td>Subwatershed shapefile</td>
<td>Delineate watershed</td>
</tr>
<tr>
<td>Upstream Subwatershed shapefile</td>
<td>Delineate watershed</td>
</tr>
<tr>
<td>Stream raster</td>
<td>Delineate watershed</td>
</tr>
<tr>
<td>Distance to Stream raster</td>
<td>Delineate watershed</td>
</tr>
<tr>
<td>Stream Order raster</td>
<td>Calculate watershed attributes</td>
</tr>
<tr>
<td>Total upslope length raster</td>
<td>Calculate watershed attributes</td>
</tr>
<tr>
<td>Longest flow length raster</td>
<td>Calculate watershed attributes</td>
</tr>
<tr>
<td>Gage Watershed raster</td>
<td>Delineate watershed</td>
</tr>
<tr>
<td>Gage Watershed shapefile</td>
<td>Delineate watershed</td>
</tr>
</tbody>
</table>

Table 4.3 Comparison of computation time required for rapid watershed delineation using RWD and traditional method.

<table>
<thead>
<tr>
<th>Watershed Area (km²)</th>
<th>RWD (sec)</th>
<th>Traditional Method Sec (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>310.2</td>
<td>2.29</td>
<td>10 (0.16)</td>
</tr>
<tr>
<td>5126</td>
<td>2.63</td>
<td>14 (0.23)</td>
</tr>
<tr>
<td>7456</td>
<td>2.9</td>
<td>432 (7.5)</td>
</tr>
<tr>
<td>10992</td>
<td>3.41</td>
<td>1458 (24.3)</td>
</tr>
<tr>
<td>16922</td>
<td>3.94</td>
<td>2494 (42)</td>
</tr>
</tbody>
</table>
Figure 4.1 Ellipsoid geometry of the Earth.
Figure 4.2 Data preprocessing steps for rapid watershed delineation.

(a) Main Watershed

(C) Subwatershed 3 with upstream contributing subwatersheds

Flow Direction

Distance to Stream

Flow Direction for subwatershed 3

Distance to stream for subwatershed 3

(d) Hydrologic information extraction for Subwatershed 3
Figure 4.3 Data Structure for Rapid Watershed Delineation.

Figure 4.4 Algorithm of Rapid Watershed Delineation.
Figure 4.5 Landing Page for RWD
Figure 4.6 User selected point for RWD.
Figure 4.7 Watershed delineated for the clicked point.
References


CHAPTER 5

SUMMARY, CONCLUSIONS, BROADER IMPACTS AND RECOMMENDATIONS

Stream flow regime variables can be potentially impacted by climate change, which can be quantified by statistical or physically based models. The validity of statistical models is limited to the data for which they are developed, whereas physically based models have some ability to predict for non-stationary changes, such as may be the case with climate change. The research described in this dissertation aimed to improve understanding of how climate change may impact streamflow regimes using physically based hydrologic modeling. This required diverse geospatial and time series datasets, the acquisition and preparation of which are time-consuming, error-prone, and difficult to reproduce. This dissertation developed data services in support of input data preparation for physically based distributed hydrologic models. This input includes terrain analysis and watershed delineation over a large area. This dissertation also enhanced terrain analysis algorithms to support rapid watershed delineation over a large area.

Quantifying changes of streamflow regime variables in response to climate change could be useful for analysis of watershed ecosystems and future water resources planning and management. Web-based data services and enhancements to terrain analysis algorithms to support rapid watershed delineation will impact a diverse community of researchers involved with developing cyberinfrastructure for data access, terrain analysis, hydrologic and environmental modeling. Chapters 2 through 4 present the model
application, developed cyberinfrastructure and scientific results of this dissertation. In this chapter I summarize the contributions in each of these chapters.

5.1 Summary and Conclusions

The first paper (Chapter 2) focuses on the development and set up hydrological models to quantify the impact of climate change on stream flow regime. For this, a physically based, distributed hydrologic model was used (1) to explore the sensitivity of the streamflow regime variables and (2) to quantify the changes in streamflow regime variables due to climate change. To achieve those objectives, a physically based hydrologic model was set up with geospatial and time series climate data and was then calibrated and validated with observed streamflow for 8 watersheds across the U.S. Calibrated models were driven by historical meteorological forcing for the twenty-year baseline period (1986–2005) to estimate streamflow regime variables and then compared with those derived from observed streamflow. Annual precipitation and temperature changes were applied to the calibrated model separately in order to estimate the sensitivity of the streamflow regime variables. Bias-corrected and spatially-downscaled future GCMs projections were used to drive the calibrated model in order to assess the potential impact of climate change on streamflow regime variables for two emission scenarios.

Our study suggested that physically based modeling can reproduce certain streamflow regime variables and can thus be used to project their changes. However, the model was not able to reproduce all variables, and poorly reproduced variables need to be recognized when interpreting the projections. Our results indicated that GCM projections
were generally consistent in showing a continued warming trend over the end of the century but offer a much wider range of precipitation change. Model simulation of streamflow regime variable changes due to climate changes were found to be appreciable and to show a high degree of variability. We found the following patterns in projected streamflow regime changes. First, streamflow is likely to increase in amount and variability. Variables quantifying high flow (e.g., Q7max and BFF) were projected to increase. Variables quantifying low flow conditions (e.g., Q7 min and SI) were predicted to decrease under the highest emission scenario, while for the low emission scenario changes were small and hard to detect. Finally, the timing variables (e.g., Peaktime and T50) indicated that stream flow is predicted to occur earlier for snow-fed watersheds in both emission scenarios; however, the model’s reproduction of peak time is poor.

Physically based models used for quantifying the impact of climate change on streamflow regime require diverse geospatial and time series datasets. The acquisition and preparation of these datasets are time-consuming, error-prone, and difficult to reproduce. The second paper focused on developing data services in support of input data preparation for physically based distributed hydrologic models. We described the design, architecture, implementation, and a case study involving a newly developed data service referred to as Hydro Data Services (HydroDS). HydroDS was designed to enable automating retrieval and processing of spatial and temporal data required by a physically based, distributed hydrologic model.

These data services provide many capabilities needed by the hydrologic modelers, including: (1) Essential Hydrologic Variable services and (2) model specific services.
The Essential Hydrologic Variables services are somewhat model independent and allow users to delineate watersheds and extract soil, climate, and streamflow data rapidly for any area in the western U.S. The model specific data services are model dependent and use the EHV service results to produce model input data sets. The data services developed can be called from a user’s personal computer using a Python script that enables selection of target watershed boundaries, approximate outlet location, and start and end time for climate data, and can automatically call all the data processing steps to produce a model input package. A user can also run each individual service through a Python function call from the script.

The third paper (Chapter 4) describes the enhancement of terrain analysis algorithms to extend their usability in terms of size and coordinate system and to support rapid watershed delineation. The work was implemented, demonstrated, and deployed for the TauDEM software. The general purpose GDAL library was used instead of the project specific TIFF library for reading and writing raster data in TauDEM to support multiple raster data formats for terrain analysis. The OGR library was used instead of the Shapelib library for reading and writing vector files in TauDEM to increase its applicability in terms of vector data types.

To support data in a geographic coordinate system, single cell dimension values for pixel width (dx) and height (dy) were replaced by arrays of pixel width (dx) and height (dy) that vary with latitude. For data in a projected coordinate system, these retain a single value, while for data in a geographic coordinate system, an earth spheroid model is used to populate these arrays and allow the program to be used over large geographic
areas (e.g., continents) where these cell dimension change. After updating TauDEM, a rapid watershed delineation program was developed to support web based watershed delineation services. For this, an algorithm was developed to subdivide a large area into subwatersheds with preprocessed results, and connectivity information was stored for each subwatershed. Rapid watershed delineation was enabled by merging local subwatersheds delineated to an arbitrary point with connected upslope preprocessed full watersheds. This design was demonstrated over the Delaware River Basin with 10 m DEM from National Elevation Dataset.

5.2 Broader Impacts

Understanding the implications of climate change on streamflow regime is complex. However, a better understanding of the impact of climate change is required for identifying how we are vulnerable to these changes, and ultimately to guide the development of robust strategies for reducing risk in the face of changing climatic conditions. This study explored possible scenarios for future stream flow regime changes. The results from this study illustrate some of the changes in streamflow that may occur with climate change, and thus provide information to the water management community in planning water supply policy and regulations that respond to climate change. The findings also have implication for restoring and recovering ecosystem processes, endangered species, and flood plain areas, as well as improving flood forecasting.

Data services developed in this study significantly reduce the time for acquiring and manipulating the geospatial and time series data required to drive a physically based, hydrologic model. This has the potential to save time and improve scientists’ ability to
perform reproducible hydrological modeling as well as to share input data and model parameterizations with colleagues using collaborative systems such as HydroShare. The approach avoids users having to install and configure models for their desktop platform, thereby overcoming compatibility limitations and enhancing reproducibility. One of the most attractive benefits of these data services is that users can access them from any computer connected to the Internet, which provides the ability to set up a hydrologic model using a PC or Mac and overcomes potential hardware, data storage, and bandwidth limitations of a user’s desktop client. More generally, the outcome of this work is a methodology for creating server-side data processing services that could be applied for other data-intensive hydrologic, environmental, and Earth system models.

Enhancements to TauDEM increase its practical applicability in terms of raster and vector data types, file size, and coordinate systems. The updated TauDEM supports geographic coordinate system data and avoids the degradation in DEM quality that occurs when projecting data originally obtained in geographic coordinates. We also developed the capability to organize pre-processed terrain data into a subwatershed based data structure to support rapid delineation of watersheds and derivation of watershed attributes over a large area. It enables a novice (e.g., middle school to college student) as well as expert user (e.g., conservation practitioner or municipal decision-maker) to perform rapid watershed delineation (RWD) by clicking near a desired outlet location without remembering the sequential steps to follow to complete the watershed delineation process. The RWD functionality also offers immediate access to watershed attribute information that would otherwise require expensive and time consuming computations.
Enhancements to TauDEM that support RWD will impact a diverse community of researchers involved with terrain analysis and hydrologic and environmental modeling. It is anticipated that the approaches and techniques adopted to make this a reality will be of significant interest to the community of researchers and students from many different scientific communities and from many different locales working to develop cyberinfrastructure for data access, terrain analysis, and hydrologic modeling.

5.3 Recommendations for Future Research

- Both climate and land use change have impacts on evapotranspiration (ET), subsurface flow, infiltration, and the streamflow regime. However, the effect of land use change was not considered while assessing the impact of climate change on the streamflow regime variables. Therefore, examining the combined impact of land use and climate change on streamflow is recommended. This can be evaluated by either integrating land use models or incorporating EPA’s future land use change data with the hydrologic model.

- This study used a physically based, hydrologic model to quantify the impact of climate change. However, there are multiple hydrologic models available, and the choice of hydrologic model can have an impact on climate change projection uncertainty. Therefore, it would be valuable to use multiple models to quantify the uncertainty in streamflow regime variable changes due to climate change.

- The climate projection data used here applied statistical methods for spatial downscaling, with the underlying assumption that climate projection preserves the occurrence frequency of daily precipitation events from the historic period. This may
not always be the case. Further studies can be done to explore the effect of different downscaling approaches such as dynamic downscaling on climate change analysis.

- In this study, one watershed was selected with principal component factors closest to the centroid of the stream in each stream class to quantify the impact of climate change. This single watershed might not represent the stream class properly. Thus, further work is recommended for selecting multiple watersheds from each streamflow class to quantify the impact of climate change on stream class. This will require (1) diverse spatial and historical time series data to set up, calibrate, and validate the hydrologic model for each watershed, (2) preprocessing Global Climate Model data, and (3) post processing the results for a large number of watersheds. Currently, the data services described in Chapter 3 have the functionality for automation of input data preparation steps for the hydrologic model. However, they do not support automated calibration, validation, and post processing of the output results. Future work is recommended to develop the data services for addressing these limitations and that will also facilitate the input and output data preparation steps to quantify the impact of climate change on streamflow class.

- The current data services allow the user to download automatically and preprocess Daymet climate data. However, there are multiple historical datasets available that can be used as a climate forcing in hydrological modeling. Extending the automated extraction to other rainfall products is recommended. This will enable the user to evaluate relative strength and weakness and take into account the uncertainty in the data sets.
- Web based data services do not provide formal descriptions of the inputs required to set up hydrologic model and output results. Additional work is needed to add such description, which will facilitate integration of this data service with other development environments such as HydroShare, which allow water scientist to discover and share hydrology data and models using a networking web portal paradigm (https://www.hydroshare.org/).

- These web based data services developed are accessible through Python scripts written by users. However, this requires knowledge of Python, which is a limitation. Integration with a graphical user interface that enables the user to learn the system quickly and use it efficiently would extend the accessibility of the data services.

- The RWD function was implemented in the Delaware River basin where the DEM with 10 m resolution was used as a base product. The DEM accuracy/scale has an effect on resulting watersheds and stream networks. Studies have shown that a more detailed DEM will produce more accurate hydrologic model results than DEMs of lower resolution. Thus extending this work using high resolution LIDAR data is recommended.

- RWD described in Chapter 4 can delineate a watershed rapidly across a large geographical region based on the assumption that the study area is subdivided based on the subwatershed. To meet this assumption, basic terrain analysis such as flow direction, flow accumulation, and stream network generation needs to be done in the study area. This may require extensive computational work. However, preprocessed products such as flow directions and the stream network are already available in the
NHDplus data set for the CONUS with 30 m spatial resolution. Thus, further work is recommended for using NHDplus data as preprocessing products for RWD implementation over the CONUS.

- Presently, RWD does not produce a detailed stream network as an output. A stream network is often needed for water quality and quantity modeling. Development of a new function that also produces the stream network with the delineated watershed is recommended. This can be achieved either by modifying the existing TauDEM StreamNet function used for watershed and stream network delineation or by using preprocessed stream network information.
APPENDICES
Appendix A

Appendix A 1: Data structure of nodelinks.txt

<table>
<thead>
<tr>
<th>File Name</th>
<th>nodelinks.txt</th>
</tr>
</thead>
<tbody>
<tr>
<td>File Function</td>
<td>Provides information on nodes-drainages connectivity</td>
</tr>
</tbody>
</table>
| File Format     | 1\textsuperscript{st} line – header descriptor of fields  
|                 | 2\textsuperscript{nd} line till the end of file – one value per each field, total number of fields = 10  
|                 | Total number of records = number of nodes |

<table>
<thead>
<tr>
<th>Fields Description</th>
<th>Field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NodeId</td>
<td>Numeric, integer</td>
<td>An internally defined node number in a sequence starting with 1</td>
</tr>
<tr>
<td></td>
<td>DownNodeId</td>
<td>Numeric, integer</td>
<td>The NodeId of the Downstream Node</td>
</tr>
<tr>
<td></td>
<td>DrainId</td>
<td>Numeric, integer</td>
<td>WRIA1 drainage ID containing the node</td>
</tr>
<tr>
<td></td>
<td>ProjNodeId</td>
<td>Numeric, integer</td>
<td>The point of interest identifier used in the project from the node point of interest file via the stream network tree</td>
</tr>
<tr>
<td></td>
<td>DOutFlag</td>
<td>Numeric, integer</td>
<td>A flag to indicate whether the node is the outlet of the drainage</td>
</tr>
<tr>
<td></td>
<td>ReachID</td>
<td>Numeric, integer</td>
<td>Identifier of the TOPNET reach that ends at this node.</td>
</tr>
<tr>
<td></td>
<td>Area</td>
<td>Numeric, float</td>
<td>The area in m\textsuperscript{2} draining directly to that node without flowing to another node first</td>
</tr>
<tr>
<td></td>
<td>AreaTotal</td>
<td>Numeric, float</td>
<td>The total area in m\textsuperscript{2} draining to each node</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>Numeric, float</td>
<td>Local coordinates for the node</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>Numeric, float</td>
<td>Local coordinates for the node</td>
</tr>
</tbody>
</table>
## Appendix A 2: Data structure of rchlink.txt

<table>
<thead>
<tr>
<th>File Name</th>
<th>rchlink.txt</th>
</tr>
</thead>
<tbody>
<tr>
<td>File Function</td>
<td>Provides information on contributing reach linages.</td>
</tr>
</tbody>
</table>
| File Format    | 1<sup>st</sup> line – header descriptor of fields  
                 2<sup>nd</sup> line till the end of file – one value per each field |

<table>
<thead>
<tr>
<th>Fields Description</th>
<th>Field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Column1</td>
<td>Numeric</td>
<td>Reach identifier</td>
</tr>
<tr>
<td></td>
<td>Column2</td>
<td>Numeric</td>
<td>Upstream reaches or sub basin number (0 means no inflows)</td>
</tr>
<tr>
<td></td>
<td>Column3</td>
<td>Numeric</td>
<td>Upstream reaches or sub basin number (0 means no inflows)</td>
</tr>
<tr>
<td></td>
<td>Column4</td>
<td>Numeric</td>
<td>Downstream X coordinate</td>
</tr>
<tr>
<td></td>
<td>Column5</td>
<td>Numeric</td>
<td>Downstream Y coordinate</td>
</tr>
<tr>
<td></td>
<td>Column6</td>
<td>Numeric</td>
<td>Monitoring point identifier</td>
</tr>
</tbody>
</table>

## Appendix A 3: Data structure of rchproperties.txt

<table>
<thead>
<tr>
<th>File Name</th>
<th>rchproperties.txt</th>
</tr>
</thead>
<tbody>
<tr>
<td>File Function</td>
<td>Provides information on contributing reach properties.</td>
</tr>
</tbody>
</table>
| File Format       | 1<sup>st</sup> line – header descriptor of fields  
                 2<sup>nd</sup> line till the end of file – one value per each field |

<table>
<thead>
<tr>
<th>Fields Description</th>
<th>Field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Column1</td>
<td>Numeric</td>
<td>Slope</td>
</tr>
<tr>
<td></td>
<td>Column2</td>
<td>Numeric</td>
<td>Manning’s n in mm units, i.e., factor 10 less than usual</td>
</tr>
<tr>
<td></td>
<td>Column3</td>
<td>Numeric</td>
<td>Width in mm</td>
</tr>
<tr>
<td></td>
<td>Column4</td>
<td>Numeric</td>
<td>Length in mm</td>
</tr>
</tbody>
</table>
Appendix A 4: Data structure of rainweights.txt

<table>
<thead>
<tr>
<th>File Name</th>
<th>rainweights.txt</th>
</tr>
</thead>
<tbody>
<tr>
<td>File Function</td>
<td>Provides information on weights used to interpolate precipitation information</td>
</tr>
</tbody>
</table>
| File Format | 1st line – header descriptor  
Subsequent rows – give information below |

<table>
<thead>
<tr>
<th>Field Description</th>
<th>Field</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subbasin no</td>
<td>Numeric, integer</td>
<td>Identifier for sub-basin number</td>
</tr>
<tr>
<td></td>
<td>Number of gauges</td>
<td>Numeric, integer</td>
<td>Number of gauges used in weighting precipitation for corresponding sub basin</td>
</tr>
<tr>
<td></td>
<td>StationID 1</td>
<td>Numeric, integer</td>
<td>Identifier for rain gauge station</td>
</tr>
<tr>
<td></td>
<td>Weight 1</td>
<td>Numeric, float</td>
<td>Weight of rain gauge station towards subbasin precipitation</td>
</tr>
<tr>
<td></td>
<td>StationID 2</td>
<td>Numeric, integer</td>
<td>Identifier for rain gauge station</td>
</tr>
<tr>
<td></td>
<td>Weight 2</td>
<td>Numeric, float</td>
<td>Weight of rain gauge station towards subbasin precipitation</td>
</tr>
</tbody>
</table>
Appendix A 5: Data structure of distribution.txt

<table>
<thead>
<tr>
<th>File Name</th>
<th>distribution.txt</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>File Function</strong></td>
<td>- Provides information wetness index distribution and hillslope flow distance distribution for each subbasin.</td>
</tr>
<tr>
<td><strong>File Format</strong></td>
<td>Block of information is repeated for each of the subbasin</td>
</tr>
<tr>
<td><strong>Fields Description</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Line</strong></td>
<td><strong>Type</strong></td>
</tr>
<tr>
<td>Line 1</td>
<td>text</td>
</tr>
<tr>
<td>Line 2</td>
<td>Numeric, integer</td>
</tr>
<tr>
<td>Line 3</td>
<td>text</td>
</tr>
<tr>
<td>Line 4 .... Line n</td>
<td>Two columns: Numeric, float, Numeric, float</td>
</tr>
<tr>
<td>Line n+1</td>
<td>text</td>
</tr>
<tr>
<td>Line n+2</td>
<td>Numeric, integer</td>
</tr>
<tr>
<td>Line n+3</td>
<td>text</td>
</tr>
<tr>
<td>Line n+m</td>
<td>Two columns: Numeric, float, Numeric, float</td>
</tr>
<tr>
<td>Line n+m+1</td>
<td>text</td>
</tr>
<tr>
<td>Line n+m+2</td>
<td>Numeric, float</td>
</tr>
</tbody>
</table>
Appendix B

Coauthor approval letter
September 15, 2016

Dr. Charles Hawkins
Professor
Watershed Science/Ecology Center, USU
Logan, UT, 84322
Email: chuck.hawkins@usu.edu

Dear Dr. Hawkins,

I am in the process of preparing my dissertation in the Civil and Environmental Engineering Department at Utah State University. I hope to complete my degree in September, 2016.

I am requesting your permission to include the attached paper, of which you are coauthor, as a chapter in my dissertation. I will include acknowledgements to your contributions as indicated. Please advise me of any changes you require.

Please indicate your approval of this request by signing in the space provided, attaching any other form or instruction necessary to confirm permission. If you have any questions, please contact me.

Thank you,
Nazmus Shams Sazib

I hereby give permission to Nazmus Shams Sazib to use and reprint all of the material that I have contributed to Chapter 2 of this dissertation.

Charles Hawkins
September 15, 2016

Pabitra Dash
Software Engineer
Utah Water Resource Laboratory
Logan, UT, 84321
Email: pabitra.dash@usu.edu

Dear Mr. Dash,

I am in the process of preparing my dissertation in the Civil and Environmental Engineering Department at Utah State University. I hope to complete my degree in September, 2016.

I am requesting your permission to include the attached paper, of which you are coauthor, as a chapter in my dissertation. I will include acknowledgements to your contributions as indicated. Please advise me of any changes you require.

Please indicate your approval of this request by signing in the space provided, attaching any other from or instruction necessary to confirm permission. If you have any questions, please contact me.

Thank you,
Nazmus Shams Sazib

I hereby give permission to Nazmus Shams Sazib to use and reprint all of the material that I have contributed to Chapter 3 of this dissertation.

Pabitra Dash
Curriculum Vitae

Nazmus Sazib  
nazmus.s.sazib@aggiemail.usu.edu
285 N 300 E apt 3  
sazibap25@gmail.com
Logan, UT, 84321  
435-374-9339

HIGHLIGHTS:
- More than five years’ experience as a researcher working both independently and in multidisciplinary teams
- Experience in developing and applying hydrologic models
- Experience in revamping physically based hydrologic model, terrain analysis and satellite rainfall algorithm and code
- Experience in terrain analysis, watershed delineation and flood mapping
- Experience in data analysis using ArcGIS, Matlab, R, Python
- Strong programming skills in Python, C++
- Experience in using High performance computing for analyzing big data

RESEARCH INTERESTS:
- Hydrologic and hydraulics modeling, Flood mapping, Computational hydrology, terrain analysis, Climate change
- High performance computing, GIS, CyberGIS

EDUCATION:
Doctor of Philosophy, Civil and Environmental Engineering, Fall 2012-present, Utah State University, Logan, UT.  
Dissertation: Physically based Modeling to quantify the Impact of Climate change on Stream flow regime.  
Advisor: Dr. David Tarboton

Master of Science, Civil and Environmental Engineering, June 2012 University of Louisiana, Lafayette, LA.  
Advisor: Dr. Emad Habib

Bachelor of Science, Civil and Environmental Engineering; June 2007  
Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh  
Dissertation: Developing pile foundation design software.  
Advisor: Dr. Md Saiful Alam Siddiquee

RESEARCH EXPERIENCE:
Research Assistant, Utah State University, August 2012-Present
- Revamped the algorithm and code of existing physically based hydrologic model to support climate change study
- Set up, calibrated and validated hydrologic model in different regions across US
- Quantified the impact of climate change on streamflow regime
- Developed data services to automate the extraction of spatial data required by the physically based model
- Automated getting and processing of time series climate data
- Developed web based services to automate the input data preparation steps of hydrologic model
- Used GDAL and OGR library in terrain analysis algorithm to increase its usability
- Modified terrain analysis algorithm to support geographic system
- Developed and designed a new methodology for rapid watershed delineation
- Generating flood map using terrain analysis software (TauDEM)
- Designed a new methodology for extracting river hydraulic properties from DEM.
- Evaluation of different sources of climate variables in hydrologic modeling.

Research Assistant, University of Louisiana, August 2010 - July 2012
- Developed and applied a new bias correction scheme to correct satellite rainfall data for applying in hydrologic model
- Revamped existing satellite rainfall extraction algorithm for using in the Nile Basin.
- Performed static and dynamic calibration of the satellite rainfall estimation algorithm

TEACHING EXPERIENCE:

Teaching Assistant, Utah State University, August 2012-Present
- Advised graduate/undergraduate students for hydrology, GIS in water resource engineering courses
- Helped students for doing GIS term project
- Graded quizzes and assignments

Teaching Assistant, University of Louisiana, August 2010- May 2011
- Helped undergraduate hydrology and Microstation students during office hours
- Graded quizzes and assignments
- Lecturer, Stamford University Bangladesh, January 2009- July 2010
- Taught undergraduate hydrology and environmental engineering courses
- Supervised undergraduate students for their thesis

**COMPUTER SKILLS:**

- Models: TOPNET, SWAT, HEC-HMS, HBV, WEAP
- Software: ArcGIS, Matlab, R, TauDEM, QGIS, GDAL
- Language and Scripts: Python, C++, FORTRAN
- Operating system: Windows, Linux

**PUBLICATION:**

- Sazib, N., D. G. Tarboton, 2015, Using a physically based distributed hydrologic model to quantify the impact of climate change on stream flow regime, in preparation for submission to *Journal of Hydrology*.
- Sazib, N., D. G. Tarboton, 2015, Data services for automation of input data preparation steps for a physically based distributed model, in preparation for submission to *Environmental Modeling and Software*.
- Sazib, N., D. G. Tarboton, 2015, Enhancements of TauDEM to support rapid watershed delineation, in preparation for submission to *Environmental Modeling and Software*.

**CONFERENCE PRESENTATION:**

- Sazib, N., D. G. Tarboton, 2015, Enhancement of TauDEM to support rapid watershed delineation, *AGU Meeting 2015*, San Francisco (CA), USA.
- Sazib, N., D. G. Tarboton, 2015, Development of Data services for automation of input data preparation steps for a physically based model, *CyberGIS Meeting 2015*, Reston (VA), USA.
- Sazib, N., D. G. Tarboton, 2015, Enhancement of TauDEM to support rapid watershed delineation, *CHAUSI meeting 2015*, Tuscaloosa (AL), USA.
- Sazib, N., D. G. Tarboton, 2015, TauDEM tools for hydrologic terrain analysis, *UWRL open house*, 2015, Logan (UT), USA.
- Sazib, N., D. G. Tarboton, 2014, Using a physically based distributed model to quantify the effect of climate change on
ecology aspects of streamflow regime, *AGU Meeting 2014*, San Francisco (CA), USA.

- Sazib, N, Habib, E., 2012 Bias correction of satellite Rainfall Estimation Enhancement of TauDEM to support rapid watershed delineation, *AMS Meeting*, 2012, New Orleans (LS), USA.

**AWARDS:**

- NSF travel fund for 2015 CyberGIS meeting
- USU Graduate Student Travel Award 2015 for AGU conference
- USU PhD Engineering Student Travel Award 2015 for AGU conference
- USU Graduate Student Travel Award 2014 for AGU conference
- USU PhD Engineering Student Travel Award 2014 for AGU conference