5-2010

Characterizing Ecologically Relevant Variations in Streamflow Regimes

Kiran J. Chinnayakanahalli
Utah State University

CHARACTERIZING ECOLOGICALLY RELEVANT

VARIATIONS IN STREAMFLOW REGIMES

by

Kiran J. Chinnayakanahalli

A dissertation 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	Charles P. Hawkins Committee Member
Mac McKee Committee Member	Jack C. Schmidt Committee Member
Thomas B. Hardy Committee Member	Byron R. Burnham Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2010

ABSTRACT

Characterizing Ecologically Relevant

Variations in Streamflow Regimes

by

Kiran J. Chinnayakanahalli, Doctor of Philosophy

Utah State University, 2010

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

Maintaining the ecological health of streams is vital for sustainable water resources management. Streamflow is a primary factor influencing the structure and function of ecological communities. A quantitative understanding of how stream biota respond to variation in streamflow is required for stream bioassessment. This dissertation focuses on quantifying relationships between streamflow regime and stream macroinvertebrate richness and composition. The contribution comprises statistical models that predict stream macroinvertebrate class from streamflow regime and predict streamflow regime from watershed attributes, and a tool that helps derive watershed attribute variables used in these models.

The dissertation is a collection of three papers. In the first paper 12 variables were used to represent streamflow regime at 543 sites in the western US. Principal component analysis (PCA) and K-means clustering were used to obtain statistically independent factors and streamflow regime classes. We examined the relationship between these characterizations of streamflow and macroinvertebrate richness and composition at 63 of the 543 sites where there was also biological data. This analysis identified specific aspects of the streamflow regime that were useful in predicting macroinvertebrate richness and composition and that have potential application in classification-based bioassessment and management.

A regional-scale study such as this requires tools for efficiently delineating watersheds and deriving their attributes. Paper two presents a multiple watershed delineation tool that addresses issues such as a) incorrectly positioned outlets and b) large Digital Elevation Models. This tool has capabilities to delineate stream networks with the threshold that determines drainage density being objectively determined so that the resulting networks adhere to geomorphological stream network laws. It also derives a suite of geomorphological watershed attributes that were used in prediction models in paper three.

In paper three, we developed statistical models to predict streamflow regime class from watershed attributes. Four popular statistical methods were used and the uncertainty associated with class predictions for each method was quantified. Paper three also identified the watershed attributes that were most important for discriminating streamflow regime classes.

(226 pages)

ACKNOWLEDGMENTS

This dissertation is dedicated to my grandparents, *Anna* and *Akka*. They never understood what I was doing but in their world I am one of the best things to have happened. This is for them for making me feel all special.

I will be indebted forever to Dr. David Tarboton, whose guidance and wisdom have helped me develop and execute my dissertation ideas. His dedication to high academic standards, astute sense for the details, and challenging courses have greatly influenced and honed my research skills. I am grateful to him for providing me with his time and energy whenever I needed it. His loyalty to his students is out of the ordinary. He has been a true guru.

If persistence is a virtue in science, I am obliged to Dr. Charles Hawkins for cultivating some of it in me. His insightfulness and methods have been truly inspirational and his encouraging words have kept me going more times than I care to remember.

P. B. Medawar has said that a sincere critic of your research is your best friend because he prevents you from erring. Then I have two of my best friends in Drs. Jack Schmidt and Thom Hardy. I express my sincere gratitude to them for making me aware of research directions often lateral to mine. I am thankful to Dr. Mac McKee for his kind support and forever reminding me that I am an engineer at core.

I thank Ryan Hill and John Olson for their help at various stages of my dissertation. My sincere thanks to rock star Christina for her friendship and advice on life, to Ibrahim and Teklu for their company and above all for tolerating my views, to

Vinod for momos, to Shyam for his extremely honest opinions, and to Manjunath for his never flinching support.

Special thanks to Deepti—your presence in my life has been enriching. I am grateful to my family—my parents, Jayakumar and Vijaya Lakshmi; my sisters, Archana and Sahana; and my nephews, Sreejith and Pragyan—for their love and support and for silently bearing my long graduate life.

I extend my thanks to the Utah Water Research Lab for providing me with financial aid that made completion of this work possible. I would also like to thank the Water Initiative program at USU for helping me with funds for buying books and attending conferences.

This work was supported, in part, by grants #R-828637-01 and #R-830594-01 from the National Center for Environmental Research (NCER) STAR Program of the U.S. Environmental Protection Agency.

Kiran J. Chinnayakanahalli

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

River systems touch all spheres of human endeavors and have been subjected to human actions throughout the world; directly by dams, reservoirs and channelization, and indirectly by land-use developments [*Naiman et al.*, 2002]. Human influence has greatly affected riverine ecosystems - chemical contamination has occurred; physical habitat has deteriorated; some species of both flora and fauna have disappeared; non-native species have invaded and the functional characteristics of riverine ecosystems have been disrupted [*Petts*, 1994].

The initial concern for river health was mainly a response to pollution problems related to human health [e.g., *Karr*, 1991; *Petts*, 1994]. It was soon recognized that chemical control approaches aimed at reducing pollution were inadequate to maintain the overall health of the river. Additionally, the awareness that environmental systems that cannot sustain themselves cannot support human life [see *Karr and Chu*, 1997] contributed to the inclusion of concepts of ecological sustainability while defining the health of river systems [*Richter et al.*, 2003, 2005] Accordingly, the objective of the U.S. Clean Water Act (PL 92-500, Sec. 101, 33 U.S.C. 1251) has been "… restoring and maintaining the chemical, physical, and biological integrity of the Nation's waters."

One of the challenges facing river scientists to achieve the above objective is to be able to define the ecosystem needs clearly enough to help the formulation of policy and management actions to balance the competing demands [*Poff et al.*, 2003]. This dissertation strives to answer questions directly related to hydrologic needs of macroinvertebrates, an important group of aquatic biota.

Figure 1.1 illustrates how environmental factors interact and influence the structure and function of riverine ecosystems (Figure 1.1) [*Allan*, 1999; *Naiman et al.*, 2002]. These include sediment and flow that determine the physical habitat and are dictated by climate and watershed attributes. Flow habitat and watershed properties drive stream temperature. The quality of water in the stream is characterized by its chemistry which depends on watershed attributes including sources of contamination within the watershed. The composition, diversity and function of stream ecosystems depends on habitat, temperature and chemistry. These all need to be considered in river ecosystem management.

Nevertheless, among different environmental factors, the characteristics related to amount and variability of discharge are considered to be the most fundamental variables defining the stream ecosystem [see *Poff and Ward*, 1989; *Bunn and Arthington*, 2002] and the alteration of flow regimes is often claimed as the most serious threat to the ecological sustainability of rivers [e.g., *Richter et al.*, 1996]. To highlight the importance of hydrology for ecosystem sustainability, *Bunn and Arthington* [2002] describe four key mechanisms that link hydrology and aquatic biodiversity: a) flow is a major determinant of the habitat, a key driver of the aquatic composition, b) aquatic species have evolved life-history strategies in response to the natural flow regime, c) the natural pattern of the longitudinal and lateral connectivity in the river system is important for supporting populations of aquatic species and d) the invasion and success of non-native species is facilitated by alterations to streamflow.

Ecologists have identified 5 aspects of the streamflow regime that are thought to influence ecological processes in rivers (Figure 1.2): flow magnitude, duration, frequency, timing, and rate of change [*Poff*, 1996; *Poff et al.*, 1997; *Puckridge et al.*, 1998]. However, the relative effects of specific aspects of flow variation on the ecological structure and function of streams are still a source of significant uncertainty [*Bunn and Arthington*, 2002; *Snelder and Biggs*, 2002; *Monk et al.*, 2006]. A quantitative understanding of how stream biota respond to variation in streamflow regimes is a necessary precursor for developing strategies for effective assessment, conservation, and restoration of stream biota. The central theme of this dissertation is to provide a quantitative understanding of the interaction between hydrology and macroinvertebrate composition and richness over the scale of the western US. Towards achieving this objective, this dissertation offers solutions to two related questions;

a) how do we efficiently derive watershed boundaries and related watershed attributes from digital elevation models for multiple watersheds spread over large geographical regions?
b) how do we quantify the ecologically relevant streamflow characterizations at watersheds without streamflow data?

This dissertation is made up of five chapters including this introduction and a summary chapter (Chapter 5). The middle three chapters forming the core of this dissertation are written in the format of papers intended for publication as separate journal articles. Each is outlined in the following paragraphs.

Chapter 2 focuses on the main objective of this dissertation- characterizing ecologically relevant variations in the streamflow regime. The requirement of regional scale characterization of streamflow regime relevant to stream biota is in demand for use in bioassessment, monitoring and management of lotic ecosystems [*Wiken*, 1986;

Omernik, 1987; *Snelder and Biggs*, 2002]. Many studies have looked into the characterization of streamflow regimes relevant to stream ecology at regional scale, but they have not directly quantified the effects of streamflow regime on the biology of the stream [e.g., *Poff and Ward*, 1989; *Poff*, 1996; *Sanz and del Jalon*, 2005; *Sanborn and Bledsoe*, 2006]. Only a few studies have tested specific hypotheses on the interaction between hydrology and ecology at a regional scale [*Poff and Allan*, 1995; *Clausen and Biggs*, 1997; *Extence et al.*, 1999; *Riis and Biggs*, 2003; *Sheldon and Thoms*, 2006; *Suren and Jowett*, 2006; *Monk et al.*, 2007, 2008; *Konrad et al.*, 2008].

In Chapter 2, we characterized the flow regimes at 543 minimally impacted gauged streams in 13 western US states and tested whether invertebrate assemblage structure (taxa richness and composition) at 63 sites was associated with variation in flow regime. We first identified 12 streamflow variables deemed to be sufficient to quantify the five aspects of streamflow regime thought to influence ecological processes mentioned earlier (Figure 1.2). These were evaluated long-term flow records for each gauged stream. We then used Principal Component Analysis to condense the 12 dimensional flow data to 7 factors that characterized statistically independent properties of streamflow: 1) zero flow day factor, 2) flow magnitude, 3) predictability, 4) flood duration, 5) seasonality, 6) flashiness, and 7) base flow. These seven factors which quantify 98% of the variability from the original twelve variables are still deemed sufficient to quantify the five aspects of streamflow regime important to ecological processes.

We next used K-means cluster analysis to classify streams into 4 to 8 hydrologically different groups based on these 7 factors. We also used empirical models to estimate three aspects of stream temperature (mean annual, mean summer and mean winter) at each site. We classified the 63 sites with invertebrate data into 6 biotic groups defined by their taxonomic composition and developed Random Forests [*Breiman*, 2001] statistical models to predict both taxa richness at a site and the probability of taxonomic class membership from both streamflow and stream temperature variables.

From this study we were able to identify specific aspects of streamflow regime that were relatively more important in explaining the variation in the macroinvertebrate composition. We also tested continuous (7 factors) versus categorical characterization (from the K-means classification) of streamflow regime for their use in explaining the variation in the invertebrate assemblage composition. Based on observed to expected species ratio and Bray-Curtis measure [*Van Sickle*, 2008] we found that Random Forest models predicting macroinvertebrate composition from streamflow regime factors and temperature variables performed significantly better than null models. These models performed the best when both streamflow regime factors and temperature variables were used as predictors. We found that for the data used in this research, the base flow factor was most directly associated with invertebrate composition. Seasonality appeared to be another important streamflow regime factor influencing the invertebrate composition, but the effect of seasonality was hard to separate from the effect of temperature so this finding may be due to confounding between these two variables. We also evaluated the probability of each biotic group conditioned on streamflow regime class from counting their joint occurrence across the study sites. We found that the prediction of invertebrate composition based directly on conditional probability of the biotic groups with respect to the streamflow regime classes was as good as for the Random Forest models that used continuous streamflow variables as predictors. This means that there is little loss in fidelity involved in using streamflow regime classes as opposed to continuous streamflow regime variables or PCA factors. This is important because management approaches that use classification are easier to formulate than management approaches based on continuous variables.

Chapter 3 presents a GIS tool developed for deriving multiple watershed attribute data. Watersheds have been widely accepted as basic functional units for various water resources management purposes. The emphasis on watershed approaches to answer water resource related questions has increased the demand for information about watersheds of interest. Furthermore most of these studies are done at regional scales requiring quick derivation of watershed boundaries, stream network structure and characteristics at a large number of locations. Increased computational power, and GIS capabilities coupled with abundant spatial data have made it possible to derive watersheds and their characteristics digitally. Nevertheless, when delineating a large number of watersheds spread across large regions there are still some limitations.

For proper delineation of watersheds from a Digital Elevation Model (DEM) the outlet for which the watershed is being delineated should exactly be positioned on the digital representation of the stream. When the number of watersheds being delineated are small, this is a simple matter of manually shifting the coordinates of the outlets to match the digital streams. However, when delineating large number of watersheds, this can become very laborious. Further, most of the currently available GIS based watershed delineation tools have difficulty handling very large DEMs, such as a single DEM extending across half of State Utah.

The Multi-Watershed-Delineation (MWD) tool we developed addresses the above mentioned limitations in delineating multiple watersheds from large digital elevation models. This tool also derives a suite of stream network and watershed attributes relevant for prediction of streamflow regime. The MWD tool has two versions: 1) a standalone Graphical User Interface (GUI) program and 2) a command line executable. For one of the analyses in this dissertation, we ran the command line MWD tool in a batch process for nearly two days to create 441 watersheds spread across the western US. The drainage area ranged between 15 km^2 to 12416 km^2 and we used a DEM with approximately 30 m grid cell resolution for this run.

Within the context of this dissertation, this tool was important in quickly deriving watershed attributes for multiple watersheds that were then used in the prediction models for estimating the streamflow regime at ungauged watersheds. MWD derived the following attributes for each watershed: a) drainage area, b) elevation statistics, c) elevation –relief ratios, b) 15 hypsometric curve indices, c) two types of watershed shape factors, d) drainage density based on objectively estimated threshold for stream delineation, and e) stream network geomorphology.

Traditionally, predicting streamflow statistics at ungauged watersheds has been considered important for estimating flood magnitude [*Thomas et al.*, 2001; *Ries and Crouse*, 2002] and for quantifying the adequacy of the stream to waste disposal, irrigation requirements, and maintenance of suitable conditions for fish [*Riggs*, 1972; *Ries*, 1997; *Ries and Friesz*, 2000]. In the US, considerable work related to estimation of the flow statistics at ungauged basins has been carried out by the USGS under their National Flood Frequency program [*Jennings et al.*, 1993; *Ries and Crouse*, 2002] and more recently by their Streamstats programs [*Ries and Gray*, 2002]. As a result of this work, many statistical models have been developed to estimate either the low or high flow statistics which represent the magnitude component of the flow regime discussed in the previous section. Recently, focus has shifted to describing the biological quality of streams which requires that other ecologically relevant components of flow regime, namely frequency, timing, duration, and rate of change be estimated at ungauged sites to help understand the biological-hydrological interactions across broad geographic regions. This recent focus on the biological quality of streams has led to interest in estimating a broader range of hydrologic indices relevant to stream ecology at ungauged sites [*Sanborn and Bledsoe*, 2006].

One of the characterizations of streamflow regime considered in Chapter 2 is a series of classifications consisting of 4-8 streamflow regime classes (categorical response variables). Chapter 4 focuses on predicting these ecologically relevant streamflow regime classes at ungauged sites from watershed attributes (geomorphology, climate and soils/geology related).

We developed the following statistical models for predicting the streamflow regime classes at ungauged sites in the western US: 1) Linear discriminant methods, LDA [see *Rao*, 1965], 2) Classification and Regression Trees, CART [*Breiman et al.*, 1984], 3) Random Forests, RF [*Breiman*, 2001], and 4) Support Vector Machines, SVM [see *Vapnik*, 1998]. The uncertainty in the prediction was quantified using bootstrapping and the best model in each case was identified based on its classification error. We also identified the watershed attributes that most discriminated the streamflow regime classes.

For predicting classifications that had 4 to 8 streamflow regime classes, LDA, CART, RF and SVM had median prediction error ranging between 28-40, 30-47, 25-32 and 27-37% respectively suggesting that predictions of class for ungauged basins is possible with about 70% accuracy, and that the RF model was slightly better than the other models for predicting the streamflow regime classes used in this study.

This dissertation provides a knowledge base for: a) characterizing streamflow regimes relevant to stream ecology in general; b) quantifying the relationship between streamflow regime and stream macroinvertebrates (composition and richness) and identifying the streamflow regime variables that are most important in explaining the observed variation in the stream biota; c) delineating multiple watersheds and their stream network using large DEMs and deriving geomorphic attributes for these watersheds; and d) predicting streamflow regime classes at ungauged sites from watershed attributes.

References

Allan, D. J. (1999), *Stream Ecology: Structure and Function of Running Waters*, Kluwer Academic Publishers, Boston, Mass.

Breiman, L. (2001), Random forests, *Machine Learning*, *45*(1), 5-32.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (1984), *Classification and Regression Trees*, Wadsworth International Group, Belmont, Calif.

Bunn, S. E., and A. H. Arthington (2002), Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity, *Environ. Manage.*, *30*(4), 492-507.

Clausen, B., and B. J. F. Biggs (1997), Relationships between benthic biota and hydrological indices in New Zealand streams, *Freshwater Biol.*, *38*(2), 327-342.

Extence, C. A., D. M. Balbi, and R. P. Chadd (1999), River flow indexing using British benthic macroinvertebrates: A framework for setting hydroecological objectives, *Regulated Rivers-Research & Management*, *15*(6), 543-574.

Jennings, M. E., W. O. J. Thomas, and H. C. Riggs (1993), Nationwide summary of U.S.Geological survey regional regression equations for estimating magnitude and frequency of floods for ungaged sites, USGS Open-file.

Karr, J. R. (1991), Biological integrity - a long-neglected aspect of water-resource management, *Ecological Applications*, *1*(1), 66-84.

Karr, J. R., and E. W. Chu (1997), Biological monitoring: Essential foundation for ecological risk assessment, *Hum.Ecol. Risk Assess.*, *3*(6), 993-1004.

Konrad, C. P., A. M. D. Brasher, and J. T. May (2008), Assessing streamflow characteristics as limiting factors on benthic invertebrate assemblages in streams across the western united states, *Freshwater Biol.*, *53*(10), 1983-1998.

Monk, W. A., P. J. Wood, D. M. Hannah, D. A. Wilson, C. A. Extence, and R. P. Chadd (2006), Flow variability and macroinvertebrate community response within riverine systems, *River Research and Applications*, *22*(5), 595-615.

Monk, W. A., P. J. Wood, D. M. Hannah, and D. A. Wilson (2007), Selection of river flow indices for the assessment of hydroecological change, *River Research and Applications*, *23*(1), 113-122.

Monk, W. A., P. J. Wood, D. M. Hannah, and D. A. Wilson (2008), Macroinvertebrate community response to inter-annual and regional river flow regime dynamics, *River Research and Applications*, *24*(7), 988-1001.

Naiman, R. J., S. E. Bunn, C. Nilsson, G. E. Petts, G. Pinay, and L. C. Thompson (2002), Legitimizing fluvial ecosystems as users of water: An overview, *Environmental Management*, *30*(4), 455-467.

Omernik, J. M. (1987), Ecoregions of the conterminous united states, *Annals of the Association of American Geographers*, *77*(1), 118-125.

Petts, G. E. (1994), Rivers: Dynamic components of catchment ecosystems, in *The rivers handbook*, edited by P. Calow and G. E. Petts, pp. 3-22, Blackwell Scientific Publications, Boston, Mass.

Poff, N. L. (1996), A hydrogeography of unregulated streams in the united states and an examination of scale-dependence in some hydrological descriptors, *Freshwater Biol.*, *36*(1), 71-91.

Poff, N. L., and J. D. Allan (1995), Functional organization of stream fish assemblages in relation to hydrological variability, *Ecology*, *76*, 606-627.

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Larr, K. L. Prestegaard, B. D. Richter, R. E. Sparks, and J. C. Stromberg (1997), The natural flow regime: A paradigm for conservation and restoration of river ecosystems, *BioScience*, *47*, 769-784.

Poff, N. L., J. D. Allan, M. A. Palmer, D. D. Hart, B. D. Richter, A. H. Arthigton, K. H. Rogers, J. L. Meyer, and J. A. Stanford (2003), River flows and water wars: Emerging science for environmental decision making, *Frontiers in Ecology and the Environment*, *1*(6), 298-306.

Poff, N. L., and J. V. Ward (1989), Implications of streamflow variability and predictability for lotic community structure - a regional-analysis of streamflow patterns, *Can. J Fish. Aquat. Sci.*, *46*(10), 1805-1818.

Puckridge, J. T., F. Sheldon, K. F. Walker, and A. J. Boulton (1998), Flow variability and the ecology of large rivers, *Mar. Freshwater Res.*, *49*, 55-72.

Rao, C. R. (1965), *Linear Statistical Inference and Its Applications*, John Wiley & Sons, Inc., New York.

Richter, B. D., J. V. Baumgartner, J. Powell, and D. P. Braun (1996), A method for assessing hydrologic alteration within ecosystems, *Conservation Biology*, *10*(4), 1163-1174.

Richter, B. D., R. Mathews, and R. Wigington (2003), Ecologically sustainable water management: Managing river flows for ecological integrity, *Ecological Applications*, *13*(1), 206-224.

Richter, B. D., R. Roos-Collins, and A. C. Fahlund (2005), A framework for ecological sustainable water management, *Hydro Review*, 24(4), 84-91.

Ries III, K. G. (1997), August median steamflows in Massachusetts, USGS, Marlborough, Mass.

Ries III, K. G., and M. Y. Crouse (2002), The national flood frequency program, version 3: A computer program for estimating magnitude and frequency of floods for ungaged sites, 42 pp.

Ries III, K. G., and P. J. Friesz (2000), Methods for estimating low-flow statistics for Massachusetts streams, 81 pp, USGS Water Resources Investigation.

Ries III, K. G., and J. R. Gray (2002), Streamstats: A U.S. Geological survey web site for stream information, paper presented at Symposium on terrain analysis for water resources application, Univ. of Tex., Austin, Tex., 18 Dec.

Riggs, H. C. (1972), Low-flow investigations, Techniques of Water-Resources Investigations, Book 4, Chap. B1, 18 p, U.S. Geological Survey, Yakima, Wash.

Riis, T., and B. J. F. Biggs (2003), Hydrologic and hydraulic control of macrophyte establishment and performance in streams, *Limnol. Oceanogr.*, *48*(4), 1488-1497.

Sanborn, S. C., and B. P. Bledsoe (2006), Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon, *J. Hydrol.*, *325*, 241-261.

Sanz, D. B., and D. G. del Jalon (2005), Characterisation of streamflow regimes in central Spain, based on relevant hydrobiological parameters, *J. Hydrol.*, *310*(1-4), 266-279.

Sheldon, F., and M. C. Thoms (2006), Relationships between flow variability and macroinvertebrate assemblage composition: Data from four Australian dryland rivers, *River Research and Applications*, *22*(2), 219-238.

Snelder, T. H., and B. J. F. Biggs (2002), Multiscale river environment classification for water resources management, *J. Am. Water Resour. Assoc.*, *38*(5), 1225-1239.

Suren, A. M., and I. G. Jowett (2006), Effects of floods versus low flows on invertebrates in a New Zealand gravel-bed river, *Freshwater Biol.*, *51*(12), 2207-2227.

Thomas, J., Wilbert O., M. M. Grimm, and R. H. McCuen (2001), An approach for evaluating flood frequency estimates for ungaged watersheds, paper presented at World Water Congress 2001, ASCE, Orlando, Fla., 20 May.

Van Sickle, J. (2008), An index of compositional dissimilarity between observed and expected assemblages, *Journal of the North American Benthological Society*, *27*(2), 227-235.

Vapnik, V. (1998), *Statistical Learning Theory*, John Wiley Sons, Inc., New York.

Wiken, E. (1986), Terrestrial ecozones of Canada: Ottawa, *Ecological Land Classification Series no. 19. Hull*, Environment Canada, 26 pp.

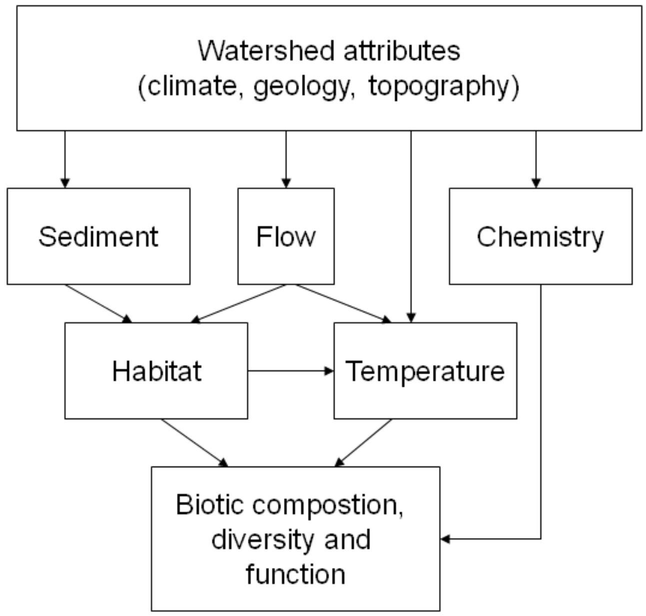


Figure 1.1. Conceptual diagram showing the causal relationships between watershed attributes; the flux of water, sediment, and chemicals from a watershed; structural and thermal habitat; and aquatic biota.

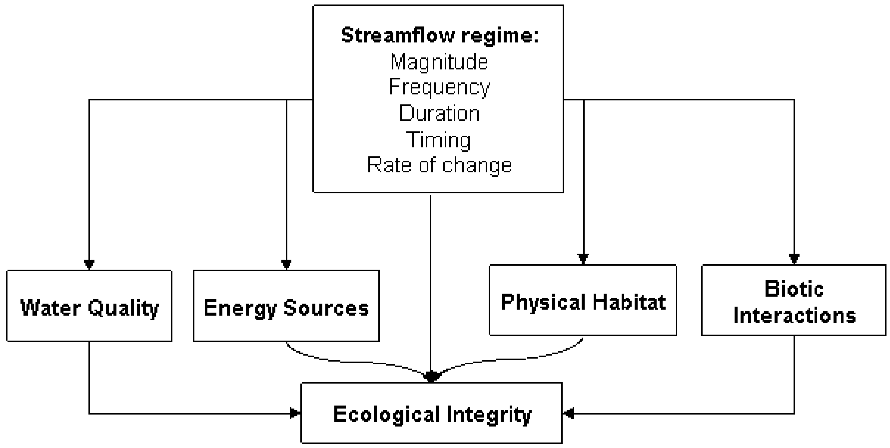


Figure 1.2. Conceptual depiction of how the five aspects of streamflow regime are related to ecological integrity of river systems. Source: *Poff et al. [1997]*.

CHAPTER 2

RELATIVE EFFECTS OF FLOW REGIME AND TEMPERATURE ON INVERTEBRATE TAXONOMIC RICHNESS AND COMPOSITION IN STREAMS OF THE WESTERN UNITED STATES[1]

Abstract

In this study we tested how strongly aquatic macroinvertebrate taxa richness and composition were associated with variation in flow regime and stream temperatures across streams of the western United States. We first used long-term flow records from 543 minimally impacted gauged streams to quantify 12 streamflow variables thought to be ecologically important. We then used Principal Component Analysis to reduce the dimensionality of the data from 12 variables to 7 principal component (PC) factors that characterized statistically independent aspects of streamflow: 1) zero flow day factor, 2) flow magnitude, 3) predictability, 4) flood duration, 5) seasonality, 6) flashiness, and 7) base flow. We used K-means to group streams into 4 to 8 hydrologically different classes based on these 7 factors. We also used empirical models to estimate mean annual, mean summer, and mean winter stream temperatures at each site. We then used invertebrate data from 63 sites to determine how well flow and temperature predicted both taxa richness and taxon-specific probabilities of capture at a site. We used Random Forests models for both predictions. We used the predicted taxon-specific probabilities of

---

[1] Coauthored by Kiran J. Chinnayakanahalli, David G. Tarboton, Ryan A. Hill, and Charles P. Hawkins.

capture to estimate how well predicted assemblages matched observed assemblages as measured by RIVPACS-type observed/expected (O/E) indices and Bray-Curtis dissimilarities. Based on observed to expected species ratio and Bray-Curtis measures, stream temperature and flow predicted assemblage composition better than a null model. Predictions were most precise when both temperature and streamflow PC factors were used, although predictions based on streamflow PC factors alone were also better than null model predictions. We were also able to predict assemblage composition from the conditional probabilities of hydrologic class membership nearly as well as Random Forests models that were based directly on the continuous principal component factors. Of the 7 factors of the streamflow regime we examined, variation in the factor describing the baseflow index, appeared to be most directly associated with invertebrate composition.

2.1. Introduction

A goal of stream ecology is to understand the environmental factors that structure natural communities. Natural flows are thought to be critical to the maintenance of healthy stream ecosystems [*Poff et al.*, 1997; *Bunn and Arthington*, 2002] , but we currently know less about the effects of flow on the distribution of stream invertebrates than that of temperature [e.g., *Sweeney and Vannote*, 1981; *Hawkins et al.*, 1997; *Poff and Zimmerman*, 2009]. Furthermore, we know little about the relative or interactive effects of these two factors on stream invertebrates.

Ecologists have identified 5 aspects of the streamflow regime that are thought to influence ecological processes in rivers: flow magnitude, duration, frequency, timing, and rate of change [*Poff*, 1996; *Poff et al.*, 1997; *Puckridge et al.*, 1998]. To quantify these five aspects, *Poff* [1996] focused on flow variables that represented the variability and predictability of low, average and high flow conditions. Other investigators have suggested additional streamflow variables to characterize streamflow regime [*Richter et al.*, 1996; *Puckridge et al.*, 1998; *Snelder and Biggs*, 2002; *Sanz and del Jalon*, 2005; *Sanborn and Bledsoe*, 2006; *Snelder et al.*, 2009]. *Olden and Poff* [2003] compiled a comprehensive list of 171 flow variables and noted that most variables generally described aspects of 1 of the 5 aspects of the streamflow regime listed above.

Classification has played an important role in efforts to synthesize and understand the variability of streamflow regimes and other stream properties, with a number of different classifications being developed for a variety of purposes [*Rosgen*, 1994; *Montgomery and Buffington*, 1998; *Snelder and Biggs*, 2002; *McDonnell and Woods*, 2004; *Wagener et al.*, 2007; *Snelder et al.*, 2009]. One approach to study the relationship between the biota and streamflow regime is to group watersheds into those with different streamflow regime classes and then assess if the composition and richness of the stream biota are significantly different across these streamflow regime classes. *Poff* [1996] used 10 streamflow variables to classify 806 relatively undisturbed gauged streams in the continental U.S. Others have also pursued watershed classifications relevant to stream biota [*Wiken*, 1986; *Omernik*, 1987; *Lipscomb*, 1998; *Snelder and Biggs*, 2002; *Snelder et al.*, 2004, 2005; *Snelder and Hughey*, 2005; *Sanborn and Bledsoe*, 2006] which is in demand for practical applications like bioassessment, monitoring and management of lotic ecosystems. .

Although our understanding of how streams differ in terms of their flow regimes has greatly increased over the last 15 years, there is still uncertainty regarding either how biota differ among streams with different flow regimes or flow regime classes, and how specific aspects of flow variation influence the ecological structure and function of streams [*Bunn and Arthington*, 2002; *Snelder and Biggs*, 2002; *Monk et al.*, 2006] particularly at regional scales. Several studies have considered the general relevance of regional variation in streamflow regimes for stream ecology, but they have not directly quantified relationships between flow regime and biotic assemblages [e.g., *Poff and Ward*, 1989; *Poff*, 1996; *Sanz and del Jalon*, 2005; *Sanborn and Bledsoe*, 2006]. A number of studies have tested specific hypotheses regarding the effects of regional differences in hydrologic regimes on the ecological properties of streams [*Poff and Allan*, 1995; *Clausen and Biggs*, 1997; *Extence et al.*, 1999; *Riis and Biggs*, 2003; *Sheldon and Thoms*, 2006; *Suren and Jowett*, 2006; *Monk et al.*, 2007, 2008; *Konrad et al.*, 2008]. However, such studies have often relied on aggregate biological measures, such as LIFE scores [*Extence et al.*, 1999], to summarize biotic responses to differences in flow regime rather than more direct measures of biodiversity such as taxonomic composition and richness. Furthermore, most of these studies have focused on the short-term response of stream biota to specific flow disturbances [e.g., *Stehr and Branson*, 1938; *Fisher et al.*, 1982; *Rae*, 1987; *Scrimgeour and Winterbourn*, 1989; *Boulton and Lake*, 1992; *Schlosser*, 1992; *Bickerton*, 1995; *Feminella*, 1996; *Miller and Golladay*, 1996; *Wood et al.*, 2000, 2001; *Cortes et al.*, 2002; *Wright et al.*, 2004; *Jackson et al.*, 2007; *Ilg et al.*, 2009]. These single-site or single-hydrological-event studies have established the general importance of different aspects of streamflow regime for stream biota, but it is difficult to draw inferences from this collective work regarding how biotic assemblages as a whole vary across large landscapes as a function of hydrologic regime. Studies that cover large, heterogeneous regions are needed.

In this study we focus directly on how aquatic macroinvertebrate taxonomic composition and richness vary across a spatially extensive set of streams that differ markedly in their flow regimes. The main goal of our study was to quantify the relationships between the richness and composition of stream invertebrate assemblages and both the long-term characteristics of streamflow and aspects of the temperature regimes that exist at the subcontinental scale of the western United States. Our specific objectives were to quantify the variation in flow regime that exists among streams in the western U.S. using a small number of carefully chosen variables, further condense these variables into independent factors, classify streams based on these factors into homogeneous classes, and determine how well flow regime factors or classes predict the composition and richness of stream invertebrate assemblages in the context of the thermal regime that also exists among streams. The results of this study have implications for both our understanding of natural stream invertebrate communities as well as our ability to conduct bioassessments.

2.2. Methods

2.2.1. General Approach to Characterizing Flow Regimes

We characterized flow regimes across the western United States by analyzing daily flow data collected at 543 relatively unimpaired streams in thirteen western U.S. states (Figure 2.1). The watershed areas for these sites ranged between 15-114,793 km^2. In past work there is limited consistency among the specific choice of ecologically relevant streamflow regime variables [*Richter et al.*, 1996; *Puckridge et al.*, 1998; *Snelder and Biggs*, 2002; *Sanz and del Jalon*, 2005; *Sanborn and Bledsoe*, 2006; *Snelder et al.*, 2009]. Based on *Olden and Poff's* [2003] suggestion and our own personal judgment, we selected 12 flow variables that we deem sufficient to characterize those flow regime properties important to stream biota. These 12 variables were: 1) base flow index (*BFI*), 2) daily coefficient of variation (*DAYCV*), 3) average daily flow (*QMEAN*), 4) average number of zero flow days (*ZERODAYS*), 5) bank full flow ($Q_{1.67}$), 6) Colwell's index of predictability (*P*), 7) Colwell's index of constancy (*M*), 8) Colwell's index of contingency (*C*), 9) average 7 day minimum flow ($\overline{7Q}_{\min}$), 10) average 7 day maximum flow ($\overline{7Q}_{\max}$), 11) average number of flow reversals ($\overline{R}$), and 12) flood duration (*FLDDUR*).

We then used Principal Component Analysis (PCA) with varimax rotation to identify a set of derived variables (factors) that were statistically independent of one another. Factor scores from the PCA were used in a K-means cluster analysis to classify the gauged streams into 4, 5, 6, 7, and 8 streamflow regime classes.

2.2.2. Flow Data

The Hydro Climatic Data Network (HCDN) is a national dataset of streamflow that is relatively free from anthropogenic influences and has been developed for studying natural variations in surface-water conditions [*Slack and Landwehr*, 1992]. The HCDN data cumulatively span the period between 1874 and 1988, but the periods of record differ between sites and not all sites are considered unaffected for this entire period. We only used data for the period where a site was considered by the USGS [*Slack and Landwehr*, 1992] to be not significantly impacted by flow regulation (listed in Appendix A). We refer to this as unimpaired streamflow. The period from 1940 to 1988 had the highest number of sites with unimpaired flow data (Figure 2.2). Fifty-one HCDN sites within our study area were excluded for one or more of the following reasons: a) closer examination revealed that they drained reservoirs, b) flows were unimpaired for less than 10 water years, or c) the HCDN database comments indicated that only monthly streamflow was considered free of human influence. We included flow data from an additional 32 gauged sites at which benthic invertebrate samples were collected and that *Carlisle et al.* [2009] indicated also had periods of unimpaired streamflow. For each site, we used daily streamflow records for only the period identified as having unimpaired flows to calculate values of the following 12 flow variables:

- The baseflow index (*BFI)* is the average across all years of the ratios of the annual lowest daily flow to the annual average flow expressed as a percentage. According to *Poff* [1996] *BFI* represents flow stability.

- The coefficient of variation of daily flows (*DAYCV)* is the ratio of the standard deviation of daily flows to the average of daily flows. *DAYCV* represents the overall variability of the streamflow regime [*Poff*, 1996].
- Mean daily discharge over all years of record *(QMEAN*, m^3/s) represents the magnitude of the flow.
- The mean number of zero flow days per year (*ZERODAYS*) quantifies low flow disturbances and intermittence in streamflow [*Poff*, 1996].
- The daily flow with a 1.67 year recurrence interval, $Q_{1.67,}$ is determined by fitting a log-normal probability distribution to the annual maximum daily flow series [*Dunne and Leopold*, 1978], then selecting the value that has a probability of exceedance of 1/1.67. Note that we base calculations on streamflow aggregated at the daily time scale, not the instantaneous peak values as is sometimes done. $Q_{1.67}$ is considered by some geomorphologists [e.g. *Dunne*, 1978] to be a measure of bank full or channel forming discharge, but the recurrence interval may vary regionally and with climate, and is generally between 1 and 2 years [*Poff*, 1996; *Wilkerson*, 2008].
- *Colwell's* [1974] indices of predictability (*P*), constancy (*C*), and contingency (*M*) quantify the temporal patterns of persistence and temporal organization of a seasonal process. A process is maximally predictable if it is constant or follows the same seasonal pattern from year to year. Predictability (*P*) is thus comprised of 2 separate components, constancy (*C*) and contingency (*M*) which are quantified based on entropy measures of uncertainty from Shannon's information

theory [see *Jelineck*, 1968]. A process has high constancy when the uncertainty is small regardless of season. A process has high contingency when the uncertainty contingent (i.e. conditional) upon season is small. Predictability combines constancy and contingency through $P = C + M$. P, C and M are scaled to range from 0 to 1.

- Calculation of Colwell's indices P, C and M is based on Shannon's entropy and requires that values be binned into discrete groups. As with all information measures absolute values are dependent on this binning, but a consistent binning allows relative comparisons. Seasonal organization can be quantified by dividing the year into periods (e.g. months or days) and constructing a table that gives the number of times a value is in group i and period j and is denoted as N_{ij}. This discretization provides t×s states, where t is the number of periods and s the number of value groups (bins) into which flow can be categorized. Following *Gordon et al*. [2004], with μ equal to the mean of daily streamflow values, we used the binning (<0.5μ, 1μ, 1.5μ, 2μ, 2.5μ, 3.0μ, >3.0μ) that defines groups scaled according to the mean of the daily streamflow values. We used months (i.e. t=12) to represent the seasonal cycle and counted the number of occurrences of daily streamflow values in states defined by groups (bins) and periods (months). *Colwell* [1974] can be referred for further details regarding the calculation of Colewell's index. P is scaled to be between 0 and 1, with the value 0 representing maximum uncertainty and the value 1 representing complete certainty as to which value group the streamflow is in each period. Constancy (C)

also scaled between 0 and 1, is a measure of temporal invariance and Contingency (*M*) (scaled between 0 and 1) is the degree to which time period and value group are dependent on each other and is a measure of seasonality.

- $\overline{7Q}_{\min}$ is the average annual minima of 7 day means of daily mean streamflow. For each year in the period of record, 7-day means are calculated from the daily mean streamflows and the minimum among them is the 7-day minimum flow for that year, $7Qmin_i$. $\overline{7Q}_{\min}$ is the average of those yearly 7-day minimum values and should characterize the average magnitude of low flow disturbances.
- $\overline{7Q}_{\max}$ is the average annual maxima of 7 day means of daily mean streamflow. For each year in the period of record, 7-day means are calculated from the daily mean streamflows and the maximum among them is the 7-day maximum flow for that year $7Qmax_i$. $\overline{7Q}_{\max}$ is the average of those yearly 7-day maximum values. $\overline{7Q}_{\max}$ should characterize the average magnitude of high flow disturbances.
- $\overline{R}$ is the average number of daily flow reversals per year. Flow-reversals are defined from the daily mean streamflow as days when the trend (increasing or decreasing) from the previous day is reversed. $\overline{R}$ represents a measure of daily flow stability.
- *FLDDUR* is flood duration calculated as the average number of days per year when flow equals or exceeds $Q_{1.67}$. *FLDDUR* is derived from the daily flows in excess of $Q_{1.67}$. Consequently *FLDDUR* is generally > 1/1.67 and quantifies the

duration of flooding in terms of the average number of days per year that flow is above the threshold.

The record lengths for 540 of these sites ranged between 10-103 years (Figure 2.3). Three sites for which we had invertebrate samples had less than 10 years of data (6, 7 and 8 years). Because we were concerned that records < 10 years in length would not adequately characterize long-term flow patterns (and hence biological associations), we conducted preliminary analyses both with and without these 3 sites. There were no significant differences between these two data sets in terms of model performance (described later), and we therefore present results based on the full data set. To maximize the data available for analysis, we did not constrain the period of record to be of either similar length or to cover a specific period of the overall record. As a consequence, the records at some sites were not continuous (i.e., there were missing years) and some sites had flow records available for different years than other sites. Because we were characterizing streams that were unimpaired, the intermittency (missing water-years) in the data should not affect the characterization of long-term flow regime. Differences in the period of record could potentially influence analyses because of natural variation in climate across years, but the fact that the most of the sites had unimpaired flow records between 1940 and 1988, should minimize such influences.

At many sites with biological data, the sample was collected later than the compilation of the HCDN database (that ended in 1988). There was a possibility that the streamflow regime had changed since 1988 at these sampling sites. We checked for evidence of potential streamflow regime change at these sites by computing streamflow regime variables for the periods where streamflow was designated by the USGS as unimpaired as well as for the period following this up to the most recent data available. This recent data has the highest potential for impact due to recent climate change. The correlation coefficient between HCDN unimpaired and post HCDN periods for 11 of the streamflow variables varied between 0.85 and 0.99, while for Zerodays the correlation coefficient was 0.64. This allayed concern that potential changes in the streamflow regime may bias the analysis.

2.2.3. Quantifying and Classifying Flow Regimes

We used Principal Component Analysis (PCA) [*Jackson*, 1991] with varimax rotation [*Kaiser*, 1958] based on the correlation between flow variables to identify the major statistically independent axes of hydrologic variation across stream gauge sites. Because PCA assumes that variables are normally distributed, we normalized each of the 12 flow variables using the *Box and Cox* [1964] transformation with parameter chosen to maximize the W-statistic in a Shapiro-Wilks normality test [*Royston*, 1982]. The transformed variables were then scaled by subtracting their mean and dividing by their standard deviation to obtain transformed standardized variables with mean of zero and standard deviation of one. Scaling removes the undue influence of a few variables on principle components (PCs) [*Jackson*, 1991].

PCA produces *NV* PCs where *NV* is the number of original variables. However, generally a relatively small number of the *NV* possible PCs are associated with most of the variation exhibited by the raw variables. Selection of a subset of the PCs for further analysis can focus on either selecting those first $nv < NV$ PCs associated with most of the variability in the original raw variables or identifying those PCs that provide unique information. Traditionally, choice of the subset of PCs used in analyses has followed the first approach [Kaiser's rule: *Lattin et al.*, 2003]. However, *Monk et al*. [2007] cautions that such traditional methods for variable selection may not represent all of the important aspects of the streamflow regime. In this work, we selected PCs based on how well they identified independent and unique aspects of the flow regime that we considered to be ecologically important.

We first chose a minimum number of PCs to work with by selecting those PCs with eigenvalues > 1 [based on Kaiser's rule: *Lattin et al.*, 2003]. We then used varimax rotation [*Kaiser*, 1958] on the PCs to obtain factors such that each variable is maximally aligned with a single factor. We inspected the resulting factors for the degree to which they represented each of the 12 variables as quantified by the variable factor loadings. If a variable was not represented in the set of factors based on its maximum loading we selected the PC with the next highest eigenvalue and repeated the varimax rotation. This process was continued until the selected PCs, when rotated into factors, had a loading of at least 0.6 from each of the 12 variables in at least one rotated factor. The final outputs from this process was table of loadings of the flow variables on varimax rotated PC factors and a matrix of factor scores, F, of dimension $543 \times k$, where k was the selected number of PCs.

We used the PC factor scores, F, in a K-means clustering analysis [*Gordon*, 1999] to identify streamflow regime classes. We used the kmeans() function available with the

R statistical software [*R Development Core Team*, 2007] to perform the K-means cluster analyses. K-means classification requires that the number of clusters be input. Because we had no a priori sense of how many classes would be optimal in terms of partitioning flow variability relevant to stream invertebrates, we examined a range of K values (K = 4 to 8). The number of classes we could examine was constrained by both resolution of flow information and sample size. $K < 4$ would not provide enough classes to discriminate all the streamflow regime characteristics of interest, whereas too few observations occurred per class when K was > 8.

2.2.4. Temperature Data

Temperature is an important environmental factor that impacts the distribution of stream invertebrates [e.g. *Sweeney and Vannote*, 1981; *Hawkins et al.*, 1997]. We included 3 measures of water temperature to help evaluate the potential importance of streamflow relative to another factor that can vary strongly over spatial and temporal scales. We used estimates of mean annual temperature (MAT), mean summer (June, July, August) temperature (MST), and mean winter (December, January, February) temperature (MWT) derived from stream temperature models (RMSE = 0.86, 2.2, and 1.7 ^{o}C, respectively) developed for the western United States (unpublished models, R. Hill and C. P. Hawkins, Utah State University, see Appendix B).

Because aspects of flow and temperature may co-vary among streams, we assessed the relationships between these 3 temperature measures and each of the different continuous measures of flow (PCs). We used backward stepwise multiple linear regression to determine which continuous flow factors were most strongly associated with temperature (response variable). We also used ANOVA to assess how much variation in stream temperature was associated with the flow classes.

2.2.5. Invertebrate Data

Between 1992 and 2003, USGS National Water-Quality Assessment Program (NAWQA) personnel collected benthic invertebrate samples at 63 of the 543 gauged sites (Figure 2.1) we used to characterize flow regimes. Samples for 59 of these sites were collected in one of the four months between June and September. For four of the sites the samples were collected in October, December or January. *Carlisle et al*. [2009] indicated that streamflow at all these sites was unimpaired, but that 30 of them had watersheds in non reference condition, meaning that the watersheds had alteration that may impact macroinvertebrates through other effects. Because our focus was on the effect of streamflow and limiting the invertebrate samples to only the 33 that Carlisle indicated were in reference condition would have resulted in a very small sample we used all 63 sites with invertebrate samples in our analyses.

Invertebrate samples were collected from 1.25 m^2 of stream bottom at each site following a standard protocol [*Moulton et al.*, 2000]. Samples were processed in the laboratory and a minimum target count of 300 (usually many more) randomly selected organisms were identified and enumerated. Invertebrate sample data included lists of taxa collected at each site and their counts. Taxa were generally identified to genus, but immature individuals of some genera cannot be distinguished from one another. Because the number of such problematic individuals varied across samples, we applied a standardized taxonomic resolution to all samples. This standardization involved either combining problematic taxa at a coarser level of taxonomic resolution (e.g., family) or excluding from analyses those individuals that could not be unambiguously assigned to a target level of taxonomic resolution [see *Carlisle et al.*, 2008]. The choice of combining or excluding individuals was based on both the frequencies of higher and lower resolution identifications for specific taxa and the best professional judgment by one of us (C. P. Hawkins) regarding the ecological differences between higher resolution taxa. In situations where the frequencies of higher and lower resolution identifications were similar, i.e., ~ 50% low resolution identifications, we proceeded as follows. If higher resolution taxa were ecologically similar, we tended to combine those taxa into a lower resolution taxon, whereas if higher resolution taxa were significantly different in their ecological preferences and tolerances, we tended to exclude the individuals identified to a coarser level.

Following taxonomic standardization and exclusion of ambiguous individuals, to limit bias from sample size differences we used the approach taken by the NAWQA program [*Moulton et al.*, 2000]. 500 individuals were randomly drawn from samples with more than 500 individuals. The entire sample was retained for samples with less than 500 individuals. From these samples, we extracted 2 assemblage-level measures: taxa richness (the number of unique types of organisms in a sample) and taxa composition (the list of specific taxa observed in a sample).

2.2.6. RIVPACS Approach and Macroinvertebrate-defined Groups of Sites

We used RIVPACS-type models [*Moss et al.*, 1987] to assess the associations between macroinvertebrate taxonomic composition and both continuous measures of flow variability (PC factors) and flow classes. Predictive models like RIVPACS are frequently used in bioassessment programs to evaluate the degree to which observed taxonomic composition matches the expected composition given specific environmental conditions [*Moss et al.*, 1987; *Wright et al.*, 1993; *Hawkins*, 2006].

The RIVPACS approach generally consists of the following steps [*Moss et al.*, 1987]: 1) classification of sites into groups based on their taxonomic composition (presence-absence data), 2) estimation of the frequencies of occurrence of different taxa within each group, 3) prediction of the probability of group membership for a site from environmental factors, and 4) estimation of probabilities of capture of specific taxa as the taxon occurrence frequency within each group combined with probabilities of group membership.

The classification of the 63 sites into macroinvertebrate groups required for the RIVPACS approach was based on their compositional similarity. We first used the Sørensen index to estimate compositional distance between all pairs of sites. We then used the flexible β hierarchical clustering method (β = -0.5) in the PC-ORD® software package [*McCune and Grace*, 2002] to construct a dendrogram that was used to identify different biologically-defined classes of sites. To facilitate interpretation of the dendrogram, compositional dissimilarity between sites and groups of sites was scaled by *Wishart's* [1969] objective function expressed as the percentage of information remaining. Wishart's objective function is a measure of information loss as clustering proceeds. It is calculated as the sum of squares of the distances between the centroids of each groups to the items in those groups. From the sample data in each group, we could estimate mean richness per group and the probability of occurrence of each taxon within each group. We also identified specific indicator taxa representative of each group following the method of *Dufrêne and Legendre* [1997] but applied to presence-absence data. These were used to illustrate the biological differences among the macroinvertebrate groups.

2.2.7. Invertebrate-Flow and Temperature Relationships

2.2.7.1. Null Models

We used null models to establish the values of model performance measures that would be expected from chance sampling alone. The null models predict the same richness and taxonomic composition at all sites within a population of sites [e.g., *Van Sickle et al.*, 2005]. Richness at each site was estimated as the average richness observed across all sites and taxonomic composition at each site was estimated as the frequencies of occurrence of different taxa among all sites. These null models ignore the effects of environmental variability among sites, and thus serve as a basis for evaluating models that include the effects of environmental variability.

2.2.7.2. Taxa Richness-Flow-Temperature Relationships

We used two type of modeling approaches for predicting taxa richness, Random Forests (RF), and direct prediction from contingency tables. For the first type, 8 different RF models were developed for predicting invertebrate taxa richness to help assess the relative performance of different combinations of predictors. These comprised a model using continuous streamflow factors (PCs) alone, categorical measures of streamflow alone (5 models, one for each streamflow classification), streamflow PCs plus measures of stream temperature, and temperature alone.

We used the random forest package [*Liaw and Wiener*, 2002] in the R software [*R Development Core Team*, 2007] to develop RF predictive models. RF models make no assumptions regarding the type of relationships (linear or non-linear) between predictor and response variables, can use both continuous and categorical predictors, and have been shown to perform well in a number of ecological settings [*Prasad et al.*, 2006; *Cutler et al.*, 2007]. When RF models are used in regression mode they predict the values of the response variable given different combinations of predictor variable values. In this case, the fit between observed and expected values can be expressed as R^2, which describes the fraction of variance in the response variable associated with the predictor variables. In regression mode, RF models quantify the importance of each predictor variable by the percentage increase in the mean square error (MSE) when the variable is left out.

The second type of model predicted taxa richness directly from the contingency table between macroinvertebrate groups and streamflow classes. Each site is associated with both a flow regime class and macroinvertebrate group. A contingency table gives the number of sites in each macroinvertebrate group occurring for a given streamflow regime class in each of the K=4-8 classifications. For a given streamflow regime classification, we computed the probability of a site to belong to different macroinvertebrate groups (P_b) directly from the contingency table. The taxa richness for a site was then estimated by averaging the mean richness per macroinvertebrate group weighted by P_b. This approach estimates the same taxa richness for all sites belonging to the same streamflow regime class.

2.2.7.3. Taxa Composition-Flow-Temperature Relationships

Similar to taxa richness, we also used RF and direct contingency tables for predicting taxa composition. The RF method was used in classification mode here, to predict the probability of macroinvertebrate group membership. Again there were 8 different RF models from 8 different sets of predictor variables. In classification mode, the importance of predictive variables is quantified by the Gini index score, a measure of the homogeneity at RF splits based on that variable [*Breiman et al.*, 1984]. For the direct contingency table approach, the number of sites in each macroinvertebrate group occurring for a given streamflow regime class was again used to derive the group membership probabilities. Once group membership probabilities had been estimated they were used in the RIVPACS approach to estimate the probabilities of capture of different taxa.

Agreement between observed and predicted assemblage composition can be measured as either the O/E ratio (where O is the number of taxa observed in the sample that were predicted to occur, and E is the number of predicted taxa, see Moss *et al*. 1987), or by a Bray-Curtis (BC) type measure of compositional dissimilarity between observed and expected taxa [*Van Sickle*, 2008]. We used both measures to assess the overall performances of the models.

The performance of O/E indices is typically assessed by the precision of model predictions. The 10th percentile of the distribution of O/E values across sample sites is a measure of model precision that is less affected by outliers than estimates of the standard deviation [*Van Sickle*, 2008]. A better model should have O/E 10th percentile values closer to 1. This was used to assess model performance relative to the null model and to evaluate between models with different predictor variables such as streamflow regime, or temperature or both. Because predictive models perform best on relatively common taxa, we restricted estimation of O/E statistics to just those taxa with predicted probabilities of capture > 0.5 [*Hawkins et al.*, 2000; *Van Sickle et al.*, 2007].

Van Sickle [2008] recommended comparing observed and expected assemblages based on the 90th percentile of BC values. 90th percentile BC values closer to 0 (greater similarity) indicate a better fit between observed and predicted assemblages. For consistency with O/E based assessments, BC estimates were also based on taxa with predicted probabilities of capture > 0.5.

2.2.8. Evaluation of Flow Regime Classification for Distinguishing Taxonomic Composition

To assess whether individual flow classes were associated with taxonomic composition, we also calculated classification strengths following *Van Sickle* [1997]. Classification strength is measured as the difference between the mean within-class similarity and the mean between-class similarity ($\overline{\mathrm{M}}$). We used the Sørensen index as the measure of between-site compositional similarity and constructed mean similarity dendrograms [*Van Sickle*, 1997] to visualize the relative strengths of association between individual classes and taxonomic composition.

2.3 Results

2.3.1. Independent Components of Flow Variability

The traditional approach to PC selection based on Kaiser's rule (eigenvalues greater than 1), would have retained only the first 3 components and 77% of the total variance in the flow data would have provided information on only the magnitude, flood duration and predictability aspects of the streamflow regime. By retaining seven PCs, selected following our iterative approach involving varimax rotation, the rotated PC factors (Table 2.1) explain 98% of the total variance in the flow data. The loadings reported in Table 2.1 define the rotated PC factors in terms of the original flow variables and can be used to physically interpreted the 7 factors to represent: 1) zero days, 2) magnitude, 3) predictability, 4) flood duration, 5) seasonality, 6) flashiness, and 7) baseflow. For example, *ZERODAYS* had a high positive loading with factor one meaning that streams with high values of factor 1 should be more susceptible to going dry than streams with low values of this factor.

2.3.2. Hydrologic Classification

When conducting the K-means analyses, we found that as *K* was incremented from 4 to 8, each subsequent classification resulted in the addition of a new class while retaining classes with attributes very similar to the previous ones. We therefore present the results for the first and the last classifications only (i.e. for *K*= 4 and 8) because they are representative of all the classifications.

The K-means clustering results in classes that are discriminated by differences in one or more flow factors (Figure 2.4 and 2.5). Examination of the distribution of flow factors for each of these classes identifies the dominant factors that characterize each class. In the $K = 4$ classification the classes are characterized by (1) seasonal streams, (2) smaller predictable intermittent streams with low baseflow, (3) mid-size perennial streams with low seasonality, and (4) big streams with low predictability and short flood duration (Figure 2.4). In the $K = 8$ classification the first four classes are characterized by the same factors as $K=4$ classes, with further classes characterized by (5) baseflow dominated streams, (6) big seasonal streams with high flood duration, (7) small unpredictable streams with high flood duration, and (8) small flashy streams with high susceptibility to drying (Figure 2.5).

Plots of the 5^{th}, 50^{th} and 95^{th} percentiles of average daily flows in each month for the different classes illustrate some of the major differences in seasonal pattern and magnitude among the $K = 8$ classes (Figure 2.6). The monthly mean values for the typical stream in each class (stream located closest to the centroid in the factor space of each class) illustrated similar, although not identical, differences among classes as did class 50th percentiles (Figure 2.6).

The *K* = 8 classes of streams represented a wide range of streamflow regimes. Streams belonging to the "seasonal stream" class (Class 1) were characterized by high seasonality (factor 5). Class 2 streams had high zero flow day factor and low baseflow. Streams in these watersheds can be intermittent. Class 3 streams had perennial flow (low zero flow day factor, factor 1) and low seasonality (factor 5). Class 4 consisted of big streams with unpredictable flows and short flood duration (factor 4). The fifth class included streams with high baseflow (factor 7) and high predictability (factor 3). In some cases these streams were intermittent (factor 1). Class 6 consisted of big (factor 2), seasonal (factor 5), perennial streams (factor 1) with long flood durations (factor 4). Class 7 streams were generally small (factor 2), unpredictable (factor 3), perennial (factor 1), and also had long flood duration (factor 4). Class 8 streams were small (factor 2), flashy (factor 6) streams with intermittent flow (factor 1).

Spatial structure was evident in some, but not all, streamflow classes (Figure 2.1and2.7). Sites in the first and the sixth classes occurred mostly in the Rocky Mountains and both were characterized by high seasonality. These 2 classes differed mostly in their size. The second streamflow class dominated the relatively dry landscape of North and South Dakota and the coastal regions of central and southern California. The third class occurred in Arizona, New Mexico, the plains east of the Rocky Mountains, and some arid parts of California. The fourth class occurred mostly in the

Washington, Oregon, and northern California coastal ranges. Streams belonging to the seventh class occurred most frequently in the interior plateaus of Utah and Nevada and the plains east of the Rocky Mountains. Classes 5, 6 and 8 did not have an obvious regional structure.

2.3.3. Relationships Between Flow Regime and Stream Temperature

The backward stepwise multiple linear regression of stream temperature on flow regime factors indicated that stream temperature co-varied with several aspects of flow. Mean annual stream temperature varied most strongly and negatively with flow seasonality (factor 5, standardized regression coefficient (SRC) = -0.66) and less strongly and negatively with baseflow (factor 7, SRC = -0.26) (adjusted R^2 = 0.40). Mean summer temperature varied negatively with 5 flow factors (zero days, predictability, seasonality, flashiness, and baseflow) with seasonality showing the strongest single association (SRC = -0.22, -0.26, -0.77, -0.25, and -0.31, respectively, adjusted R^2 = 0.56). Mean winter temperature was less strongly related to streamflow regime factors, but increased with increasing predictability (factor 3, SRC = 0.39) and decreased with both increasing flood duration (factor 4, SRC = -0.36) and seasonaility (factor 6, SRC = -0.39) (adjusted R^2 = 0.37). Flow regime classes showed similar associations with temperature. K = 4 to 8 classifications accounted for 42, 43, 44, 58, and 52% of the variation in mean annual temperature; 34, 39, 27, 52, and 44% of mean summer temperature; and 19, 27, 25, 40, and 38% of mean winter temperature respectively.

2.3.4. Invertebrate Assemblage Structure

Taxonomic richness varied from 13 to 44 taxa per sample across the 63 study streams (mean = 31). Taxa composition also showed considerable variability among sites as illustrated by the flexible β cluster diagram (Figure 2.8). For modeling purposes, we identified 6 groups (see dashed line in Figure 2.8) that represented a compromise between maximizing average within-group compositional similarity and the number of sites per group. The indicator species identified (Table 2.2) show that these groups are taxonomically and ecologically distinct from one another. For example, groups A and B consisted of taxa that require cool, fast-flowing water, whereas groups E and F included taxa more typical of warmer, slowing moving water.

2.3.5. Contingency Table Analyses

Invertebrate-defined classes (Figure 2.8) were non-randomly associated with flow regime classes for all classifications (Table 2.3, Chi-Square test $P < 0.00004$). The probabilities in this table are quite discriminating implying that in many cases, a streamflow regime class is associated with a single macroinvertebrate group (probability close to 100% for that macroinvertebrate group and 0 for the rest).

2.3.6. Associations Between Taxa Richness, Streamflow Regime, and Temperature

When predicted by Random Forests models, taxonomic richness was only weakly (11-15% of variation) associated with flow, temperature or streamflow class, although these values were statistically higher than that of the null model (Table 2.4). Seasonality of flow and the zero flow day factor were the most important flow regime predictors of invertebrate richness in the RF models when flow predictors were used alone, and mean summer temperature was the most important temperature predictor (Figure 2.9 a and b). When flow and temperature factors were used together, summer temperature was the most important predictor followed by seasonality of flow, zero flow day factor, flow predictability, mean annual temperature, mean winter temperature, flow flashiness, flow magnitude, baseflow, and flood duration. Conditional probability models for prediction of taxa richness based on 4, 5, and 7 classes accounted for slightly more (20 – 24%) of the variation. However, models based on 6 and 8 classes accounted for less (~11%) of the variation in taxa richness. Overall the predictability of taxa richness by this approach was found to be generally poor (R^2 values not > 0.24, Table 2.4).

2.3.7. Associations Between Taxa Composition, Streamflow Regime, and Temperature

All models predicting taxa composition performed substantially better than their respective null models as measured by both the 10^{th} percentile of O/E values and the 90^{th} percentiles of BC values (Table 2.4). Models incorporating both flow regime and temperature were best. Streamflow variables alone performed better than the 3 temperature variables alone in terms of O/E, but temperature alone performed better than flow variables alone in terms of BC. RF predictions based on flow classes were generally slightly worse than those based on continuous flow variables. The importance scores of flow factors in predicting taxa composition differed from those for predicting richness (Figure 2.7). Variation in baseflow was most useful in predicting composition followed by flow seasonality, flood duration, flow magnitude, flow predictability, zero flow day factor, and flashiness. When flow and temperature predictors were combined, summer temperature was the most important variable followed by baseflow and mean annual temperature (Figure 2.9 f). Mean winter temperature, flow magnitude, flood duration, flow seasonality, flow predictability, flow flashiness, zero flow dayfactor (in that decreasing order) were less important to RF predictions. Predictions based on the conditional probability models were generally better than RF models based on only flow variables for both O/E and BC measures of precision.

2.3.8. Evaluation of Flow Regime Classification for Distinguishing Taxonomic Composition

Each of the *K=4-8* streamflow regime classifications had similar, and weak, overall classification strengths with respect to invertebrate composition ($\overline{M}$ = 0.066 to 0.076, Figure 2.10) evaluated using the Sørensen index with presence-absence data [*Van Sickle*, 1997]. Different individual classes were more strongly associated with composition than other classes in all classifications. For example, class 1 (seasonal and predictable streams) accounted for more variation in composition than other classes. Compositional similarity of sites within classes 4 (large flows), 5 (predictable with high base flow), and 6 (large, flashy streams with low zero days) was moderately greater than mean between-class similarity. However, streamflow regime classes categorized by small streams (class 2, 7 and 8) and midsize -low seasonality streams (class 3) did not distinguish variation among sites in taxa composition.

2.4. Conclusions and Discussion

2.4.1. The Challenge of Identifying and Characterizing Ecologically Relevant Streamflow Variables

Although stream ecologists agree that flow regime is likely a primary determinant of the structure and function of natural stream ecosystems [*Resh et al.*, 1988; *Power et al.*, 1995], we still have difficulty quantifying which aspects of naturally occurring flow regimes most strongly affect stream ecosystems and predicting the biological consequences of altering these regimes, especially at a regional scale. To increase explanatory and predictive power, we must resolve several issues related to the characterization or classification of streamflow. A critical step is identifying the flow variables that are most useful in understanding ecological patterns and processes from the many variables available. A second issue is how to most effectively summarize the information contained in different flow variables into axes of hydrologic variation. This issue involves both the number and types of flow characteristics that are needed to adequately describe flow regimes, especially as perceived by biota. We also need to understand how finely we need to resolve classifications of flow regime or if classification into flow types is useful at all. Perhaps most importantly, we need to demonstrate that different aspects of flow variability are ecologically relevant and useful in either an explanatory or predictive capacity. Finally, to be useful in a management context, we need to know if we can predict the specific types of flow regime that characterize ungauged streams.

2.4.2. Streamflow Record

An ideal study design would have included long term streamflow records for all sites from the same period of record that extended to a common date of invertebrate sampling. Such a design would remove the uncertainties involved in comparing streamflow regimes across sites with a) unequal streamflow records and b) different periods of record. Comparing sites with records from different periods and different record lengths potentially confounds the effect of temporal variation in climate with spatial variation in climate. However such ideal data is not available so we worked with available data evaluating to the extent possible the potential impact of non stationarities on the results. We found that, the majority of sites we worked with had an unimpaired flow record between the years 1940 and 1988. This commonality of period for most sites should limit the potential for confounding by climate non stationarities.

2.4.3. Choice of Streamflow Variables and Their Effect on Classification

The first step in this work involved the selection of a set of streamflow variables that were thought to influence the macroinvertebrate richness and composition of streams. This choice influences the subsequent quantification and classification of flow regimes. But it was not straightforward from previous studies which variables to select to explore general relationships between streamflow variables and stream macroinvertebrates, especially at a regional scale. We selected variables based on insights from previous studies, discussions with colleagues, and our own experiences. Choice of the number of variables to use in a classification should also be based on our ability to interpret the resulting classification. For example, we could interpret differences among streams in terms of the 12 variables we chose, but we concluded it would be increasingly difficult to interpret and understand the physical characteristics of classifications based on more variables. The small number of variables used in this study and the use of PCA considerably reduced redundancy among the variables, which helped with both physical and potential ecological interpretations.

The use of varimax rotation in the PCA allowed us to associate PC factors with distinct aspects of the hydrologic regime and thereby enhance the physical interpretation of these factors. A traditional approach to PC selection would have led to the use of only 3 axes of streamflow variation and the use of only 77% of the information in the raw data. Using the enhanced capability for physical interpretation given by varimax rotation we continued adding PC factors until all key aspects of the flow regime that we had identified were included, resulting in a factorization that identified 7 physically distinguishable characteristics that captured 98% of the variance in the original streamflow regime data (Table 2.1).

2.4.4. Scaling Magnitude Related Streamflow Variables

We also treated magnitude related streamflow variables differently than previous researchers. Most prior work has characterized flow magnitude in terms of unit discharge by scaling discharge variables either by watershed area or mean flow [e.g., *Poff*, 1996; *Monk et al.*, 2007]. Because of this standardization, previous classifications would potentially group small and large streams together. Our use of unscaled magnitude related variables resulted in a magnitude factor (factor 2) that discriminated between small and large streams, which we showed was related to variation in invertebrate assemblage composition. For example, streams belonging to classes 1 or 6 are seasonal streams found along the Rocky Mountain range (Figure 2.7), but they differ in the quantity of water that is present in them. Stream size, often measured as watershed area, is well known to be strongly associated with both variation in taxonomic composition and ecosystem processes [e.g., *Vannote et al.*, 1980].

2.4.5. Flow Regime Classifications

At the scale of the western USA, climate has a major influence on the spatial structure of streamflow regime classes (Figure 2.1and 2.7). However, streams belonging to different classes were also often found in close proximity to each other. Such close proximity of different stream types arose, in part, because magnitude was a factor in the classification. This result implies that even though climate has a major influence on streamflow regimes, it will not be possible to identify geographically contiguous hydro-regions (comparable to ecoregions) that are spatially discrete from one another. Rather, stream segments will need to be individually characterized in recognition of the diversity of ecologically relevant flow regimes (or classes) that can occur within any climatic region.

While there have been many previous studies that have developed classification of streamflow, the classification presented here differs in the choice of underlying variables and the inclusion of flow magnitude as a factor in classification. The new classification was shown to have some degree of predictive capability for macroinvertebrate assemblage composition. This suggests that it has potential application in bioassessment and for identifying hydrologically similar watersheds to study, for example, the separate effects of hydrology and other factors on macroinvertebrate composition. Since the variables upon which this classification is based are relatively general descriptors of the stream environment this classification may have applicability beyond macroinvertebrate composition, a question that warrants further investigation.

2.4.6. Temperature and Streamflow Variables

Temperature is an important variable that regulates the local and regional composition of macroinvertebrates [e.g., *Sweeney and Vannote*, 1981; *Hawkins et al.*, 1997]. Because temperature variables co-varied with some of the streamflow variables, it was difficult to differentiate the biological effects of one set of variables from the other. For example, the strong association between temperature and seasonality is probably caused by the seasonal pulse of cold water that enters streams in the spring and early summer associated with the melting of snow. Although this co-variation confounds interpretation of the specific ecological importance of each variable, such co-variation implies that one type of variable might be used as a surrogate for the other type for predictive purposes (Figure 2.9). However, in our analysis use of both hydrologic and temperature variables resulted in the best predictions of taxonomic composition, which implies some degree of independent response to both types of variables (Table 2.4). In general, such joint consideration of streamflow and temperature regimes should provide a more robust characterization of the stream environment than either alone [e.g., *Harris et al.*, 2000] and thus allow more accurate predictions of the biological potential of different streams.

2.4.7. Relationships Between Flow Regimes and Biota

Primary goals of stream ecologists are to understand the independent and interactive affects of environmental factors on the structure and function of stream ecosystems [*Allan and Castillo*, 2007]. With such understanding, it should be possible to predict the biota expected to occur under different environmental conditions and hence assess the degree to which anthropogenic alteration in those environmental conditions will affect the ecological condition of streams [e.g., *Hawkins*, 2006]. Because the hydrologic regime is a fundamental component of stream habitat, it is imperative to understand how it affects both populations and communities of stream organisms. However, any analysis of biota-flow relationships assumes that we have adequately characterized real differences among streams in their hydrologic regimes.

Our analyses assumed that the hydrological characterizations we developed were relevant to each stream's biota at the time invertebrates were sampled. Given that considerable gaps sometimes existed between the period of record and when invertebrates were sampled, there is some concern that macroinvertebrate richness and composition may be more due to recent short term events, rather than the predictor variables we are using. For this reason we might expect relatively poor associations between our flow characterizations and the biota collected at a site. This is one of many sources of uncertainty in the analysis that may contribute to the remaining unexplained variability in the results. The fact that we detected relationships between streamflow regime and biota in spite of year-to-year variation within streams in some aspects of flow (Table 2.4 and Figure 2.10) supported our underlying hypothesis that long-term flow patterns are part of the hydrologic template that influences what specific organisms can establish and persist in a specific stream. However, further exploration of this issue may be warranted in the future as it has implications for the number and timing of macroinvertebrate sample collection in bioassessment.

Our modeling focused on two aspects of stream invertebrate assemblages: overall richness and taxonomic composition. Our results showed that overall taxa richness was not strongly associated with either flow regime or temperature in spite of the considerable variation that occurred in the number of invertebrate taxa found at our study streams (Taxa richness in Table 2.4). We conclude that other factors may have been more important in regulating overall taxa richness in our study streams [cf., *Vinson and Hawkins*, 1998]. Furthermore, total richness is strongly influenced by the number of rare taxa at a site. In open ecosystems like streams in which downstream drift can deliver many taxa to a site that cannot persist, estimates of total richness may tell us little regarding important ecological differences among sites and the factors regulating those differences. We observed reasonably strong relationships between the taxonomic composition at a site and both flow regime and temperature (90th percentile BC values in Table 2.4, Figure 2.9), perhaps because we focused on those taxa most common (estimated probability of detection > 0.5) at each site. This result was encouraging because it has clear implications for both our understanding of the factors that regulate taxonomic composition in streams and our ability to assess the effects of landscape and waterway alteration on stream ecosystems. Although associations do not necessarily imply causation, 2 factors (temperature and baseflow) stood out as being important in predicting taxonomic composition (90^{th} percentile BC values in Table 2.4, Figure 2.9). The role of temperature in structuring stream assemblages is well established, but we know less about which aspects of flow are critical in this regard. Several previous studies have focused on the role of flooding in structuring stream assemblages [e.g., *Boulton et al.*, 1992; *Robinson et al.*, 2004], but our results imply that future studies might profit by focusing on how the mechanisms associated with variation in baseflow affect assemblage composition.

Our modeling results are significant given that prediction of the taxa expected at a site is a critical component of bioassessment [e.g., *Hawkins*, 2006; *Stoddard et al.*, 2006; *Paulsen et al.*, 2008]. The accuracies of the models developed here are comparable with those in use in many bioassessment programs. For example, the 10^{th} percentile of reference site O/E values derived from a western USA-wide model that *Carlisle and Hawkins* [2008] used to assess the condition of invertebrate assemblages at NAWQA sampling sites was 0.84 (i.e., any site with a value < 0.84 would be considered impaired). That model was based on data collected from 729 reference sites and used 9 predictor variables, several of which were probably surrogates for flow variables. Use of direct estimates of 7 flow factors and 3 temperature variables produced a RF model of similar precision (10^{th} percentile of O/E values = 0.80), and use of flow variables or the best classification alone resulted in only slightly less precise RF models (with 10^{th} percentile of O/E values of 0.73 and 0.70, respectively). These values are also similar to the 10^{th} percentile values reported for several other O/E indices [*Hawkins*, 2006]. The use of direct measures of both flow and temperature should not only improve model accuracy and predictions, but allow a more direct interpretation of the likely causes of biological impairment when it is observed. For example, a low O/E value associated with a substantial difference between the expected and observed baseflow at a site implies that hydrologic alteration may be a cause of the observed biological impairment. Improvement of the models used for bioassessment will require that we be able to estimate both the hydrologic reference condition at ungauged sites [e.g., *Sanborn and Bledsoe*, 2006] in the same way that we estimated the expected thermal environment. The fact that use of both flow and temperature variables produced the best models of taxonomic composition is not surprising considering the frequent reference to these factors in the stream ecology literature [see *Allan and Castillo*, 2007]. It is unclear, however, that their separate effects can be cleanly distinguished from one another. Inferring that flow seasonality is important in structuring stream invertebrate assemblages from our results is especially suspect given its strong confounding with stream temperature. These issues notwithstanding, the associations between stream biota, flow regime, and temperature that we documented here provide a solid empirical basis for justifying future studies designed to refine the characterization of both flow and thermal regimes in streams.

2.4.8. Classes Versus Continuous Variables

In our analysis, we studied 2 types of streamflow regime characterizations: one based on the derived continuous streamflow regime factors and the other based on discrete hydrologic classifications. Random Forests models with continuous streamflow factors appeared to be slightly better than models based on categorical variables in predicting the taxonomic compositions across the 63 NAWQA stations (RF under taxa composition in Table 2.4). However, predictions based on direct conditional probabilities also performed relatively well (model type CP in Table 2.4) and the conditional probabilities were derived from classifications. The better performance of the continuous factors is probably due to the fact that some information is always lost when we collapse continuous factors into categorical classes. However, classifications are attractive to ecosystem managers because they are generally easier to communicate and implement. Our results showed that the use of flow regime classifications may not significantly compromise models when predicting taxonomic composition. The use of hydrologic classes in models may be especially attractive if predicting hydrologic class turns out to be easier than predicting continuous values of the different aspects of the flow regime.

References

Allan, J. D., and M. M. Castillo (2007), *Stream Ecology: Structure and Function of Running Waters*, Second ed., Springer, Dordrecht, The Netherlands.

Bickerton, M. A. (1995), Long-term changes of macroinvertebrate communities in relation to flow variations - the river Glen, Lincolnshire, England, *Regulated. Rivers: Research and Management*, 10(2-4), 81-92.

Boulton, A. J., and P. S. Lake (1992), The ecology of 2 intermittent streams in Victoria, Australia .2. Comparisons of faunal composition between habitats, rivers and years, *Freshwater Biol.*, *27*(1), 99-121.

Boulton, A. J., C. G. Peterson, N. B. Grimm, and S. G. Fisher (1992), Stability of an aquatic macroinvertebrate community in a multiyear hydrologic disturbance regime, *Ecology*, *73*(6), 2192-2207.

Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *Journal of the Royal Statistical Society*, *26*, 211-252.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (1984), *Classification and regression trees*, Wadsworth International Group, Belmont, Calif.

Bunn, S. E., and A. H. Arthington (2002), Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity, *Environ. Manage.*, *30*(4), 492-507.

Carlisle, D. M., and C. P. Hawkins (2008), Land use and the structure of western USA stream invertebrate assemblages: O/E and ecological traits, *Journal of the North American Benthological Society*, 27(4), 986-999.

Carlisle, D. M., C. P. Hawkins, M. R. Meador, M. Potapova, and J. Falcone (2008), Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages, *Journal of the North American Benthological Society*, *27*(1), 16-37.

Carlisle, D. M., J. Falcone, D. M. Wolock, M. R. Meador, and R. H. Norris (2009), Predicting the natural flow regime: Models for assessing hydrologic alterations in streams, *River Research and Applications,* 26(2), 118-136.

Clausen, B., and B. J. F. Biggs (1997), Relationships between benthic biota and hydrological indices in New Zealand streams, *Freshwater Biol.*, *38*(2), 327-342.

Colwell, R. K. (1974), Predictability, constancy, and contingency of periodic phenomena, *Ecology*, *55*, 1148-1153.

Cortes, R. M. V., M. T. Ferreira, S. V. Oliveira, and D. Oliveira (2002), Macroinvertebrate community structure in a regulated river segment with different flow conditions, *River Research and Applications*, *18*(4), 367-382.

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, and K. T. Hess (2007), Random forests for classification in ecology, *Ecology*, *88*(11), 2783-2792.

Dufrêne, M., and P. Legendre (1997), Species assemblages and indicator species: The need for a flexible asymmetrical approach, *Ecological Monographs*, *67*, 345-366.

Dunne, T. (1978), Field studies of hillslope flow processes, in *Hillslope Hydrology*, edited by M. I. Kirkby, pp. 227-293, Wiley-Interscience, New York.

Dunne, T., and L. B. Leopold (1978), *Water in Environmental Planning*, W H Freeman & Co, San Francisco, Calif.

Extence, C. A., D. M. Balbi, and R. P. Chadd (1999), River flow indexing using British benthic macroinvertebrates: A framework for setting hydroecological objectives, *Regulated Rivers-Research & Management*, *15*(6), 543-574.

Feminella, J. W. (1996), Comparison of benthic macroinvertebrate assemblages in small streams along a gradient of flow permanence, *Journal of the North American Benthological Society*, *15*(4), 651-669.

Fisher, S. G., L. J. Gray, N. B. Grimm, and D. E. Busch (1982), Temporal succession in a desert stream ecosystem following flash flooding, *Ecological Monographs*, *52*(1), 93-110.

Gordon, A. D. (1999), *Classification*, Second ed., Chapman & Hall/CRC, New York.

Gordon, N., T. McMahon, B. Finlayson, C. Gippel, and R. Nathan (2004), *Stream Hydrology: An Introduction for Ecologists*, Second ed., John Wiley & Sons Inc., Chichester.

Harris, N. M., A. M. Gurnell, D. M. Hannah, and G. E. Petts (2000), Classification of river regimes: A context for hydroecology, *Hydrological Processes*, *14*(16-17), 2831-2848.

Hawkins, C. P. (2006), Quantifying biological integrity by taxonomic completeness: Its utility in regional and global assessments, *Ecological Applications*, *16*(4), 1277-1294.

Hawkins, C. P., J. N. Hogue, L. M. Decker, and J. W. Feminella (1997), Channel morphology, water temperature, and assemblage structure of stream insects, *Journal of the North American Benthological Society*, *16*(4), 728-749.

Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella (2000), Development and evaluation of predictive models for measuring the biological integrity of streams, *Ecological Applications*, *10*(5), 1456-1477.

Ilg, C., F. Foeckler, O. Deichner, and K. Henle (2009), Extreme flood events favour floodplain mollusc diversity, *Hydrobiologia*, *621*, 63-73.

Jackson, E. J. (1991), *A Users Guide to Principal Components*, John Wiley & Sons, Inc., New York.

Jackson, H. M., C. N. Gibbins, and C. Soulsby (2007), Role of discharge and temperature variation in determining invertebrate community structure in a regulated river, *River Research and Applications*, *23*(6), 651-669.

Jelineck, F. (1968), *Probabilistic Informtaion Theory*, McGraw-Hill, New York.

Kaiser, H. F. (1958), The varimax criterion for analytic rotation in factor analysis, *Psychometrika*, *23*, 187-200.

Konrad, C. P., A. M. D. Brasher, and J. T. May (2008), Assessing streamflow characteristics as limiting factors on benthic invertebrate assemblages in streams across the western united states, *Freshwater Biol.*, *53*(10), 1983-1998.

Lattin, J., J. D. Carroll, and P. E. Gree (2003), *Analyzing Multivariate Data*, Brroks/Cole-Thomson Learning, Toronto, Ontario.

Liaw, A., and M. Wiener (2002), Classification and regression by randomforest, *R News*, *2*(3), 18-22.

Lipscomb, S. W. (1998), Hydrologic classification and estimation of basin and hydrologic characteristics of subbasins in central Idaho, 49 pp, USGS, Denver, CO.

McCune, B., and J. B. Grace (2002), *Analysis of Ecological Communities*, MjM Software, Gleneden Beach, Ore.

McDonnell, J. J., and R. Woods (2004), On the need for catchment classification, *J. Hydrol.* (299), 2-3.

Miller, A. M., and S. W. Golladay (1996), Effects of spates and drying on macroinvertebrate assemblages of an intermittent and a perennial prairie stream, *Journal of the North American Benthological Society*, *15*(4), 670-689.

Monk, W. A., P. J. Wood, D. M. Hannah, D. A. Wilson, C. A. Extence, and R. P. Chadd (2006), Flow variability and macroinvertebrate community response within riverine systems, *River Research and Applications*, *22*(5), 595-615.

Monk, W. A., P. J. Wood, D. M. Hannah, and D. A. Wilson (2007), Selection of river flow indices for the assessment of hydroecological change, *River Research and Applications*, *23*(1), 113-122.

Monk, W. A., P. J. Wood, D. M. Hannah, and D. A. Wilson (2008), Macroinvertebrate community response to inter-annual and regional river flow regime dynamics, *River Research and Applications*, *24*(7), 988-1001.

Montgomery, D. R., and J. M. Buffington (1998), Channel processes, classification, and response, in *River Ecology and Management: Lessons from the Pacific Coastal Ecoregion*, edited by R. Naiman and R. Bilby, pp. 13-42, Springer-Verlag, New York.

Moss, D., M. T. Furse, J. F. Wright, and P. D. Armitage (1987), The prediction of the macro-invertebrate fauna of unpolluted running-water sites in great Britain using environmental data, *Freshwater Biol.*, *17*, 41–52.

Moulton, S. R., II, J. L. Carter, S. A. Brother, T. F. Coffey, and T. M. Short (2000), Methods of analysis by the U.S. Geological survey national water quality laboratory: Processing, taxonomy, and quality control of benthic macroinvertebrates samples, U.S. Geological Survey Open-File Report, 49 pp.

Olden, J. D., and N. L. Poff (2003), Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, *River Research and Applications*, *19*(2), 101-121.

Omernik, J. M. (1987), Ecoregions of the conterminous united states, *Annals of the Association of American Geographers*, *77*(1), 118-125.

Paulsen, S. G., A. Mayio, D. Peck, J. Stoddard, E. Tarquinio, S. Holdsworth, J. Van Sickle, L. L. Yuan, C. P. Hawkins, A. T. Herlihy, P. R. Kaufmann, M. T. Barbour, D. P. Larsen, and A. R. Olsen (2008), Condition of stream ecosystems in the united states: An overview of the first national assessment, *Journal of the North American Benthological Society*, *27*, 812-821.

Poff, N. L. (1996), A hydrogeography of unregulated streams in the united states and an examination of scale-dependence in some hydrological descriptors, *Freshwater Biol.*, *36*(1), 71-91.

Poff, N. L., and J. D. Allan (1995), Functional organization of stream fish assemblages in relation to hydrological variability, *Ecology*, *76*, 606-627.

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Larr, K. L. Prestegaard, B. D. Richter, R. E. Sparks, and J. C. Stromberg (1997), The natural flow regime: A paradigm for conservation and restoration of river ecosystems, *BioScience*, *47*, 769-784.

Poff, N. L., and J. V. Ward (1989), Implications of streamflow variability and predictability for lotic community structure - a regional-analysis of streamflow patterns, *Can. J. Fish. Aquat. Sci.*, *46*(10), 1805-1818.

Poff, N. L., and J. K. H. Zimmerman (2009), Ecological responses to altered flow regimes: A literature review to inform the science and management of environmental flows, *Freshwater Biol.,* 55(1), 194-205.

Power, M. E., A. Sun, G. Parker, W. E. Dietrich, and J. T. Wootton (1995), Hydraulic food-chain models, *Bioscience*, *45*(3), 159-167.

Prasad, A. M., L. R. Iverson, and A. Liaw (2006), Newer classification and regression tree techniques: Bagging and random forests for ecological prediction, *Ecosystems*, *9*(2), 181-199.

Puckridge, J. T., F. Sheldon, K. F. Walker, and A. J. Boulton (1998), Flow variability and the ecology of large rivers, *Marine and Freshwater Research*, *49*, 55-72.

R Development Core Team (2007), R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria.

Rae, J. G. (1987), The effects of flooding and sediments on the structure of a stream midge assemblage, *Hydrobiologia*, *144*, 3-10.

Resh, V. H., A. V. Brown, A. P. Covich, M. E. Gurtz, H. W. Li, G. W. Minshall, S. R. Reice, A. L. Sheldon, J. B. Wallace, and R. C. Wissmar (1988), The role of disturbance in stream ecology, *Journal of the North American Benthological Society*, *7*(4), 433-455.

Richter, B. D., J. V. Baumgartner, J. Powell, and D. P. Braun (1996), A method for assessing hydrologic alteration within ecosystems, *Conservation Biology*, *10*(4), 1163-1174.

Riis, T., and B. J. F. Biggs (2003), Hydrologic and hydraulic control of macrophyte establishment and performance in streams, *Limnol. Oceanogr.*, *48*(4), 1488-1497.

Robinson, C. T., U. Uehlinger, and M. T. Monaghan (2004), Stream ecosystem response to multiple experimental floods from a reservoir, *River Research and Applications*, *20*(4), 359-377.

Rosgen, D. L. (1994), A classification of natural rivers, *Catena*, *22*, 169-199.

Royston, P. (1982), An extension of Shapiro and Wilk's W test for normality to large samples, *Applied Statistics*, *31*, 115-124.

Sanborn, S. C., and B. P. Bledsoe (2006), Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon, *Journal of Hydrology*, *325*, 241-261.

Sanz, D. B., and D. G. del Jalon (2005), Characterisation of streamflow regimes in central Spain, based on relevant hydrobiological parameters, *J. Hydrol.*, *310*(1-4), 266-279.

Schlosser, I. J. (1992), Effects of life-history attributes and stream discharge on filter-feeder colonization, *Journal of the North American Benthological Society*, *11*(4), 366-376.

Scrimgeour, G. J., and M. J. Winterbourn (1989), Effects of floods on epilithon and benthic macroinvertebrate populations in an unstable New Zealand river, *Hydrobiologia*, *171*(1), 33-44.

Sheldon, F., and M. C. Thoms (2006), Relationships between flow variability and macroinvertebrate assemblage composition: Data from four Australian dryland rivers, *River Research and Applications*, *22*(2), 219-238.

Slack, J. R., and J. M. Landwehr (1992), Hydro-climatic data network: A U.S. Geological survey streamflow data set for the United States for the study of climate variations, 1874 - 1988, 193 pp.

Snelder, T. H., and B. J. F. Biggs (2002), Multiscale river environment classification for water resources management, *Journal of the American Water Resources Association*, *38*(5), 1225-1239.

Snelder, T. H., B. J. E. Biggs, and R. Woods (2005), Improved eco-hydrological classification of rivers, *River Research and Applications* (21), 609-628.

Snelder, T. H., F. Cattaneo, A. M. Suren, and B. J. E. Biggs (2004), Is the river environment classification an improved landscape-scale classification of rivers?, *Journal of the North American Benthological Society*, *23*(3), 580-598.

Snelder, T. H., and K. F. D. Hughey (2005), The use of an ecologic classification to improve water resource planning in New Zealand, *Environ. Manage.*, *36*(5), 741-756.

Snelder, T. H., N. Lamouroux, J. R. Leathwick, H. Pella, E. Sauquet, and U. Shankar (2009), Predictive mapping of the natural flow regimes of France, *J. Hydrol.*, *373*(1-2), 57-67.

Stehr, W. C., and J. W. Branson (1938), An ecological study of an intermittent stream, *Ecology*, *19*, 294-310.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris (2006), Setting expectations for the ecological condition of streams: The concept of reference condition, *Ecological Applications*, *16*(4), 1267-1276.

Suren, A. M., and I. G. Jowett (2006), Effects of floods versus low flows on invertebrates in a New Zealand gravel-bed river, *Freshwater Biol.*, *51*(12), 2207-2227.

Sweeney, B. W., and R. L. Vannote (1981), Ephemerella mayflies of white clay creek: Bioenergetic and ecological relationships among six coexisting species, *Ecology*, *62*(5), 1353-1369.

Van Sickle, J. (1997), Using mean similarity dendrograms to evaluate classifications., *Journal of Agricultural, Biological and Environmental Statistics*, *2*, 370–388.

Van Sickle, J. (2008), An index of compositional dissimilarity between observed and expected assemblages, *Journal of the North American Benthological Society*, *27*(2), 227-235.

Van Sickle, J., C. P. Hawkins, D. P. Larsen, and A. T. Herlihy (2005), A null model for the expected macroinvertebrate assemblage in streams, *Journal of the North American Benthological Society*, *24*(1), 178-191.

Van Sickle, J., D. P. Larsen, and C. P. Hawkins (2007), Exclusion of rare taxa affects performance of the O/E index in bioassessments, *Journal of the North American Benthological Society*, *26*(2), 319-331.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing (1980), The river continuum concept, *Can. J. Fish. Aquat. Sci.*, *37*, 130-137.

Vinson, M. R., and C. P. Hawkins (1998), Biodiversity of stream insects: Variation at local, basin, and regional scales, *Ann. Review of Entomol.*, *43*, 271-293.

Wagener, T., M. Sivapalan, P. Troch, and R. Woods (2007), Catchment classification and hydrologic similarity, *Geography Compass*, *1*(4), 901-931.

Wiken, E. (1986), Terrestrial ecozones of Canada, in *Ecological Land Classification Series no. 19. Hull*, Environment Canada, QC.

Wilkerson, G. V. (2008), Improved bankfull discharge prediction using 2-year recurrence-period discharge, *Journal of the American Water Resources Association*, *44*(1), 243-258.

Wishart, D. (1969), An algorithm for hierarchical classifications, *Biometrics*, *25*, 165-170.

Wood, P. J., M. D. Agnew, and G. E. Petts (2000), Flow variations and macroinvertebrate community responses in a small groundwater-dominated stream in south-east England, *Hydrol. Processes*, *14*(16-17), 3133-3147.

Wood, P. J., D. M. Hannah, M. D. Agnew, and G. E. Petts (2001), Scales of hydroecological variability within a groundwater-dominated stream, *Regulated Rivers-Research & Management*, *17*(4-5), 347-367.

Wright, J. F., M. T. Furse, and P. D. Armitage (1993), RICPACS: A technique for evaluating the biological water quality of rivers in the UK, *Eur. Water Pollut. Control,* 3, 15-25.

Wright, J. F., R. T. Clarke, R. J. M. Gunn, N. T. Kneebone, and J. Davy-Bowker (2004), Impact of major changes in flow regime on the macroinvertebrate assemblages of four chalk stream sites, 1997-2001, *River Research and Applications*, *20*(7), 775-794.

Table 2.1. Loadings for Varimax PC Factors from Box-Cox Transformed Streamflow Variables. High Loadings are in Bold Font

Factors	1	2	3	4	5	6	7
BFI	-0.299	0.006	-0.175	0.157	0.060	0.097	**0.895**
DAYCV	0.045	-0.215	0.336	-0.125	-0.316	-0.210	**-0.769**
QMEAN	-0.091	**0.928**	-0.204	-0.017	0.152	0.141	0.207
ZERODAYS	**0.813**	-0.235	0.174	0.012	-0.221	-0.162	-0.408
$Q_{1.67}$	-0.080	**0.951**	-0.120	-0.199	0.115	0.122	0.035
FLDDUR	0.002	-0.181	0.040	**0.967**	0.043	0.005	0.171
P	0.078	-0.144	**0.930**	0.014	0.181	-0.152	-0.203
C	0.127	-0.268	**0.822**	0.060	-0.350	-0.139	-0.272
M	-0.157	0.225	-0.004	0.047	**0.927**	0.101	0.209
$\overline{7Q}_{min}$	-0.200	**0.672**	-0.250	0.049	0.182	0.185	0.582
$\overline{7Q}_{max}$	-0.071	**0.981**	-0.084	-0.086	0.080	0.089	0.005
$\overline{R}$	-0.135	0.274	-0.238	0.005	0.119	**0.885**	0.226
% variance explained by each factor	7.3	29.4	15.8	8.6	10.5	8.3	18.2
Interpretation	Zero flow days	Magnitude	Predictability	Flood duration	Seasonality	Flashiness	Base flow

Table 2.2. Indicator Taxa for Each of the Macroinvertebrate-Defined Groups Identified from the Cluster Analysis (Figure 2.8). Taxa Within a Group are Ordered (Highest to Lowest) by their Indicator Values (Not Shown). Indicator Taxa Were Identified Following *Dufrêne And Legendre* [1997] we Indentified those Taxa that are over Represented in a Class Relative io ihe other Classes. Letters In Parentheses Identify the Taxonomic Order to which Each Taxon Belongs: C = Coleoptera (Beetles), D = Diptera (True Flies), E = Ephemeroptera (Mayflies), L = Lepidoptera (Butterflies/Moths), P = Plecoptera (Stoneflies), And T = Trichoptera (Caddisflies). Following the Standard Notation Genus are Italicized, but Family is not

Group					
A	B	C	D	E	F
Micropsectra (D)	*Lepidostoma* (T)	*Eukiefferiella* (D)	*Psephenus* (C)	Dryopidae (C)	*Paratanytarsus* (D)
Zapada (P)	*Claassenia* (P)		*Pteronarcys* (P)	*Hydropsyche* (T)	*Dubiraphia* (C)
Chloroperlidae (P)	Athericidae (D)		*Glossosoma* (T)	*Chimarra* (T)	*Caenis* (E)
Rhyacophila (T)	*Drunella* (E)		*Microcylloepus* (C)		*Dicrotendipes* (D)
Arctopsyche (T)	*Deuterophlebia* (D)		*Rheocricotopus* (D)		*Saetheria* (D)
Brillia (D)	*Acentrella* (E)		*Protoptila* (T)		*Thienemannimyia* (D)
Heterlimnius (C)	*Hexatoma* (D)		Pyralidae (L)		
Epeorus (E)	*Zaitzevia* (C)		*Rheotanytarsus* (D)		
Cleptelmis (C)			*Optioservus* (C)		
Perlodidae (P)			*Antocha* (D)		
Hesperoperla (P)					
Baetis (E)					
Simuliidae (D)					

Table 2.3. Probability That a Site Belongs to One of the Macroinvertebrate Group Given That Its Streamflow Regime Class Is Known. N Is the Number of Sites in Each Streamflow Regime Class. Probabilities of Macronivertebrate Group Membership > 0.5 Are Highlighted in Bold Font

Streamflow Class	Biotic classes						N
	A	B	C	D	E	F	
1	0.22	**0.61**	0.17	0.00	0.00	0.00	18
2	0.00	0.00	0.11	0.00	0.22	**0.67**	9
3	0.09	0.09	0.43	0.09	0.17	0.13	23
4	0.38	0.23	0.15	0.15	0.00	0.08	13
1	0.19	**0.69**	0.12	0.00	0.00	0.00	16
2	0.00	0.00	0.11	0.00	0.22	**0.67**	9
3	0.12	0.06	0.29	0.12	0.24	0.18	17
4	0.36	0.21	0.21	0.14	0.00	0.07	14
5	0.14	0.14	**0.71**	0.00	0.00	0.00	7
1	0.17	**0.75**	0.08	0.00	0.00	0.00	12
2	0.00	0.00	0.12	0.00	0.25	**0.62**	8
3	0.11	0.00	0.33	0.00	0.22	0.33	9
4	**0.50**	0.20	0.00	0.20	0.00	0.10	10
5	0.33	0.00	**0.67**	0.00	0.00	0.00	3
6	0.10	0.24	0.43	0.10	0.10	0.05	21
1	0.18	**0.73**	0.09	0.00	0.00	0.00	11
2	0.00	0.00	0.00	0.00	0.29	**0.71**	7
3	0.00	0.00	0.12	0.00	**0.50**	0.38	8
4	0.44	0.22	0.00	0.22	0.00	0.11	9
5	0.33	0.00	**0.67**	0.00	0.00	0.00	3
6	0.11	0.26	**0.53**	0.11	0.00	0.00	19
7	0.33	0.17	0.33	0.00	0.00	0.17	6
1	0.22	**0.67**	0.11	0.00	0.00	0.00	9
2	0.00	0.00	0.00	0.00	0.00	**1.00**	4
3	0.00	0.00	0.14	0.00	**0.57**	0.29	7
4	0.38	0.25	0.00	0.25	0.00	0.12	8
5	0.33	0.00	**0.67**	0.00	0.00	0.00	3
6	0.10	0.38	0.43	0.10	0.00	0.00	21
7	0.25	0.00	**0.50**	0.00	0.00	0.25	4
8	0.29	0.00	0.14	0.00	0.29	0.29	7

Table 2.4. Performance of the Random Forests (RF) and Conditional Probability (CP) Models in Predicting Taxa Richness and Taxonomic Composition. R^2 Measures the Strength of Relationships between Taxa Richness and Streamflow and Temperature Predictors. The 10th Quantile of O/E Values and the 90th Quantile of Bray-Curtis (BC) Values Measure How Well Streamflow and Temperature Predict Taxonomic Composition

Model type	Predictors	Taxa richness	Taxa composition	
		R^2	10th quantile of O/E	90th quantile of BC
Null	-	0.000	0.576	0.460
RF	7 flow factors	0.142	0.725	0.418
RF	7 flow factors + 3 temperature variables	0.148	0.795	0.344
RF	3 temperature variables	0.108	0.665	0.398
RF	4 flow classes	0.145	0.638	0.449
RF	5 flow classes	0.097	0.696	0.406
RF	6 flow classes	-0.044	0.669	0.460
RF	7 flow classes	0.017	0.675	0.427
RF	8 flow classes	-0.028	0.678	0.412
CP	4 flow classes	0.237	0.756	0.391
CP	5 flow classes	0.223	0.753	0.389
CP	6 flow classes	0.111	0.749	0.422
CP	7 flow classes	0.198	0.755	0.390
CP	8 flow classes	0.115	0.742	0.397

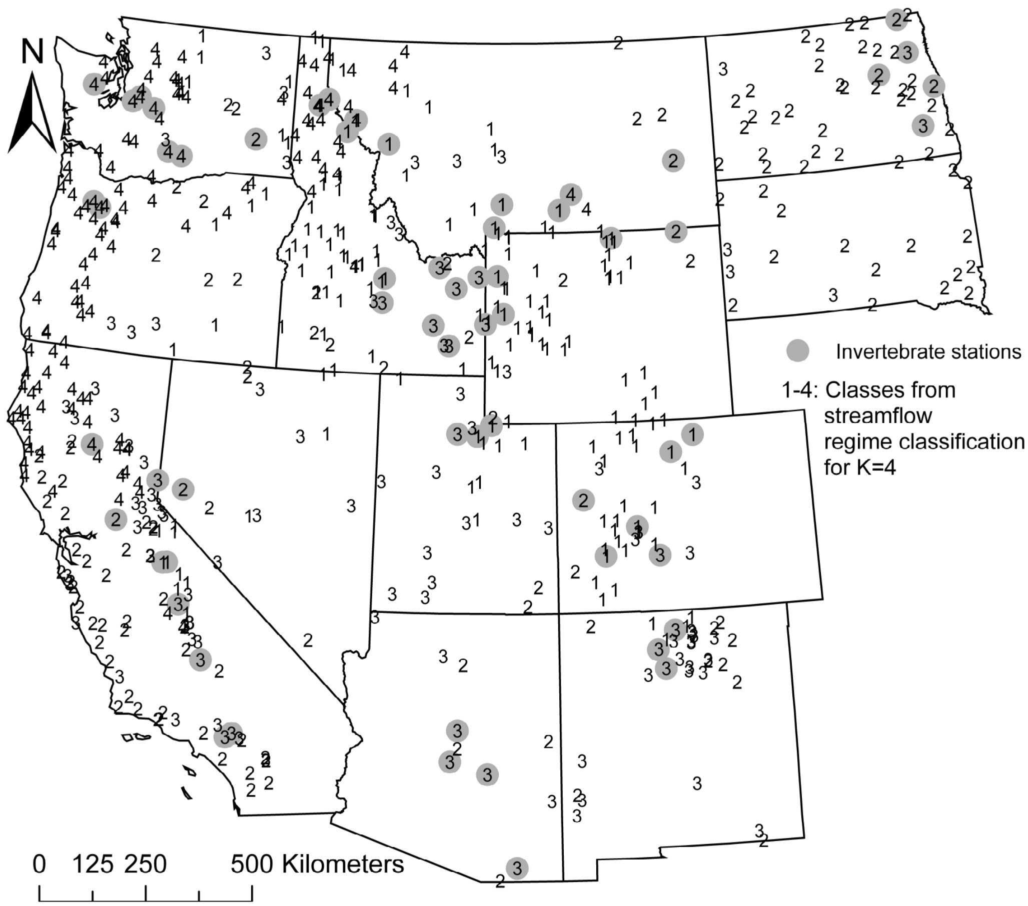
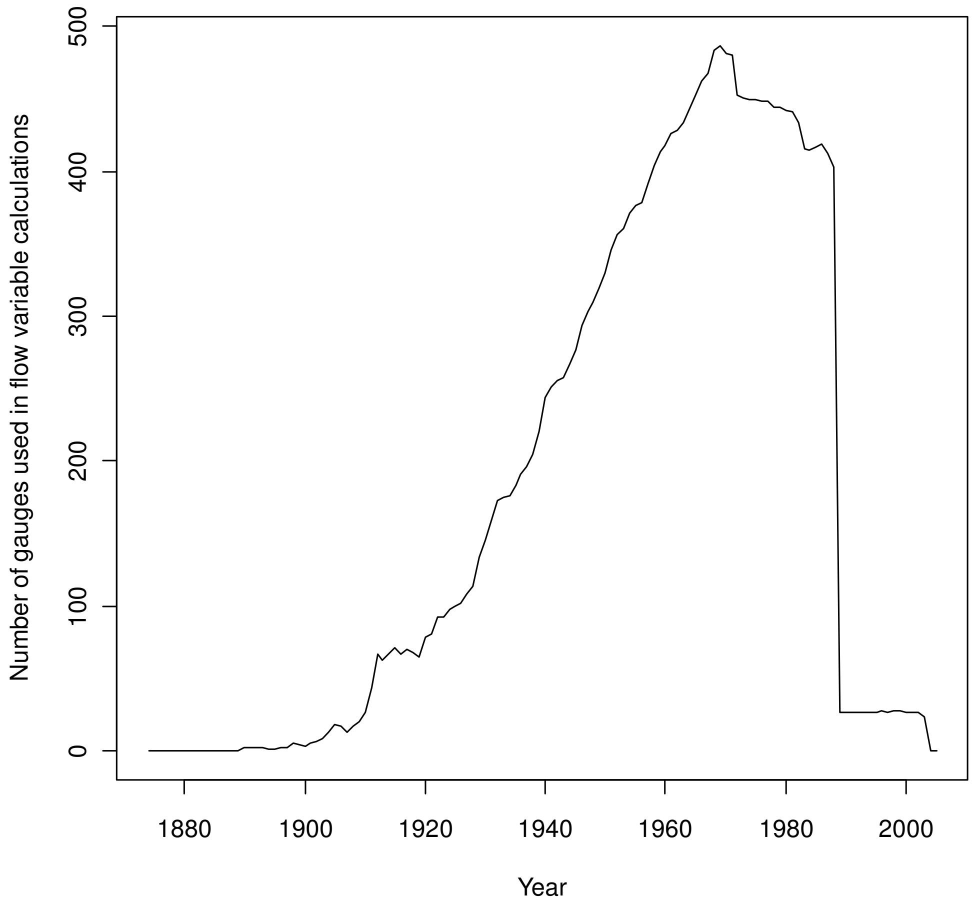


Figure 2.1. Location of 543 streamflow gauge sites used in this study. 511 sites are from the Hydro Climate Data Network (HCDN) with an additional 32 with benthic invertebrate data from *Carlisle et al.* [2009]. Numbers indicate regime class for K=4 classification. Sites with NAWQA benthic invertebrate samples are also indicated.



.
Figure 2.2. Distribution of number of gauges used in flow variable calculations by year.

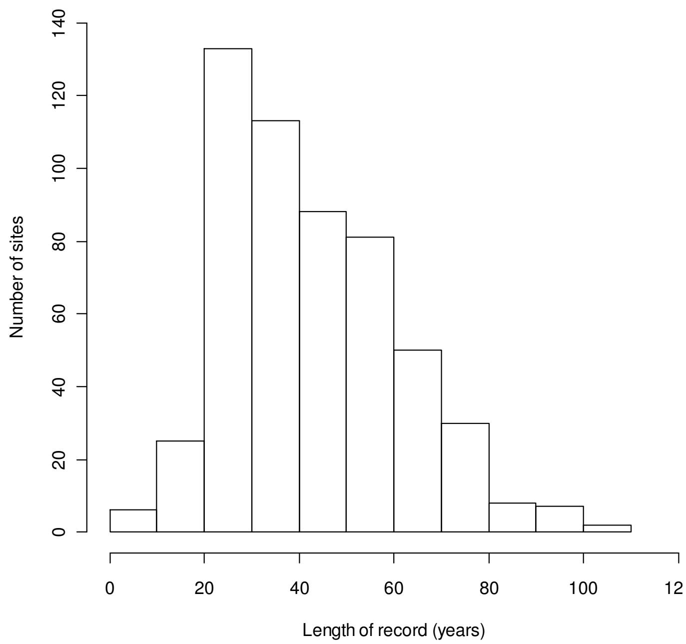


Figure 2.3. Distribution of length of records for 543 sites.

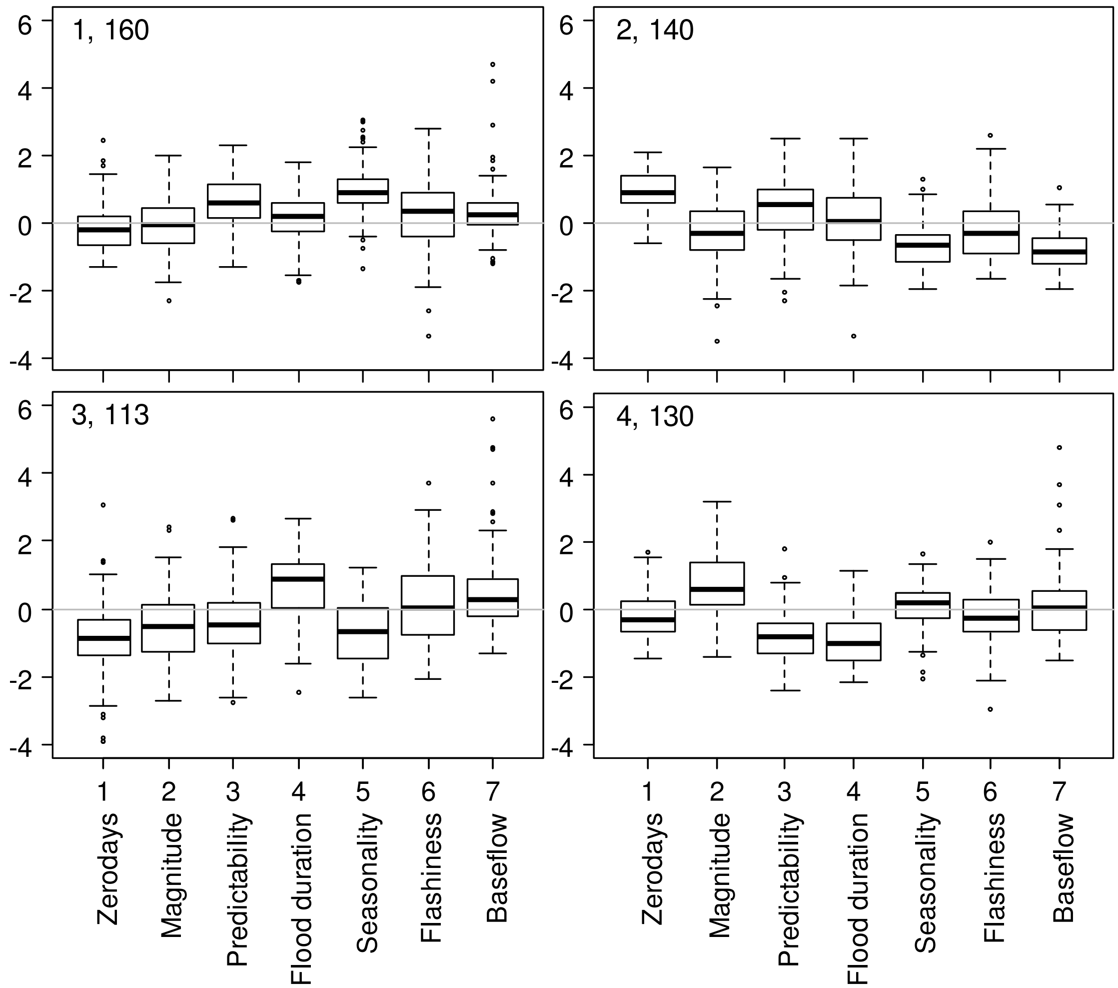


Figure 2.4. Box plots showing the distribution of varimax rotated PC factors across different flow regime classes for K=4 (the numbers on top of each plot represent the class number and the class size).

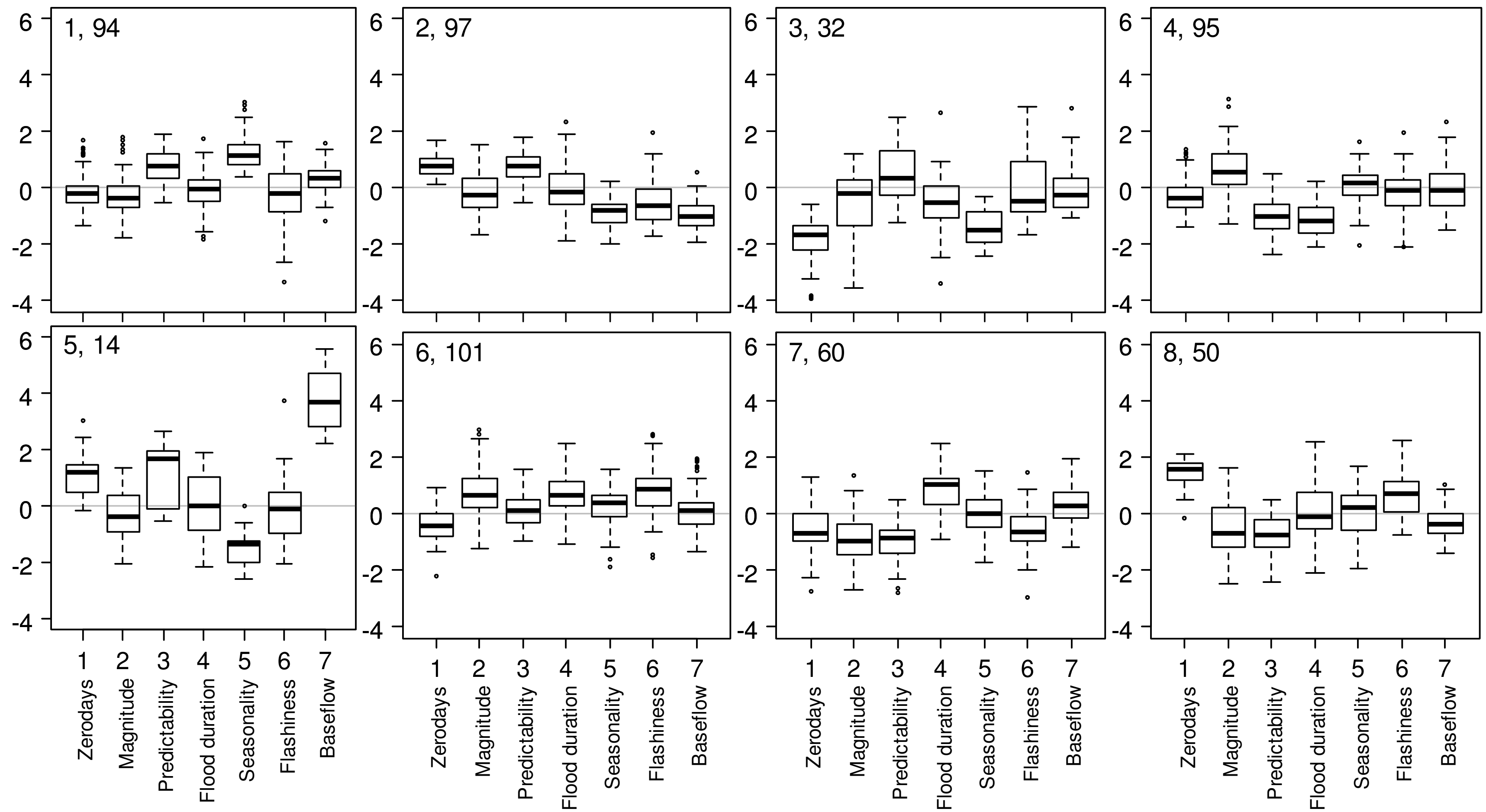


Figure 2.5. Box plots showing the distribution of varimax rotated PC factors across different flow regime classes for K=8 (the numbers on top of each plot represent the class number and the class size).

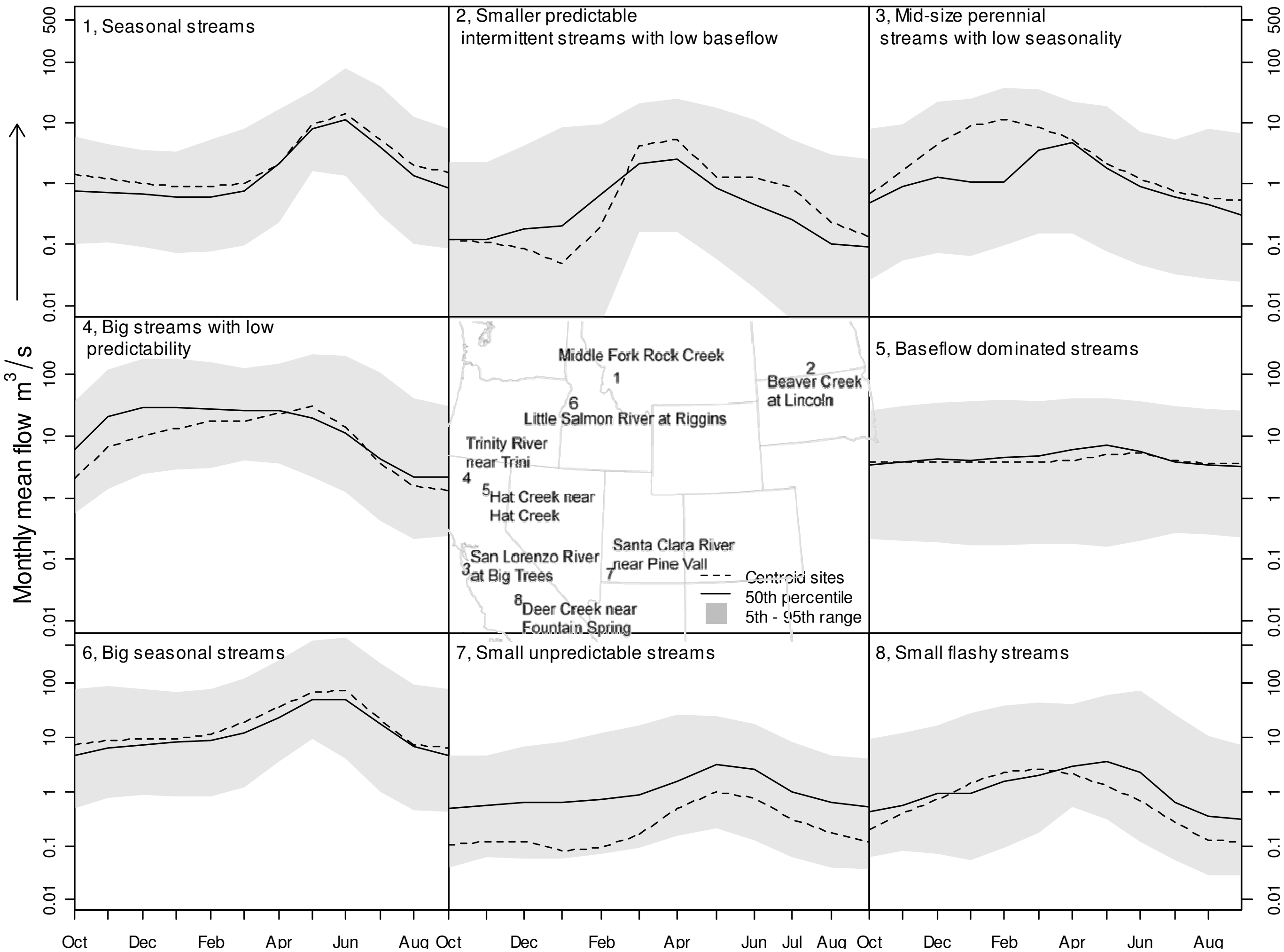


Figure 2.6. 5th, 50th, and 95th percentile of average monthly flows for each flow regime class and mean flow for streams closest to class centroids. Map at center indicates the site nearest to the centroid of each flow regime class.

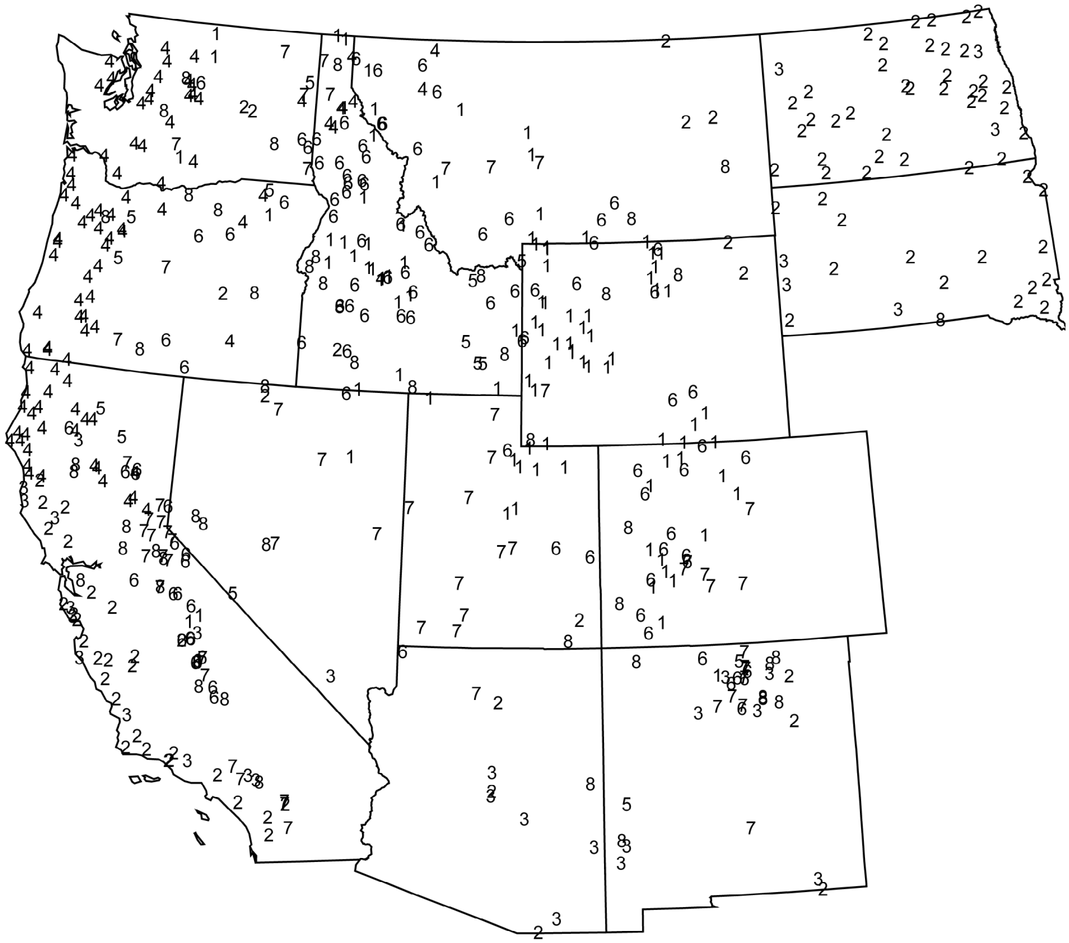


Figure 2.7. Spatial distribution of sites within each flow regime classes for K=8.

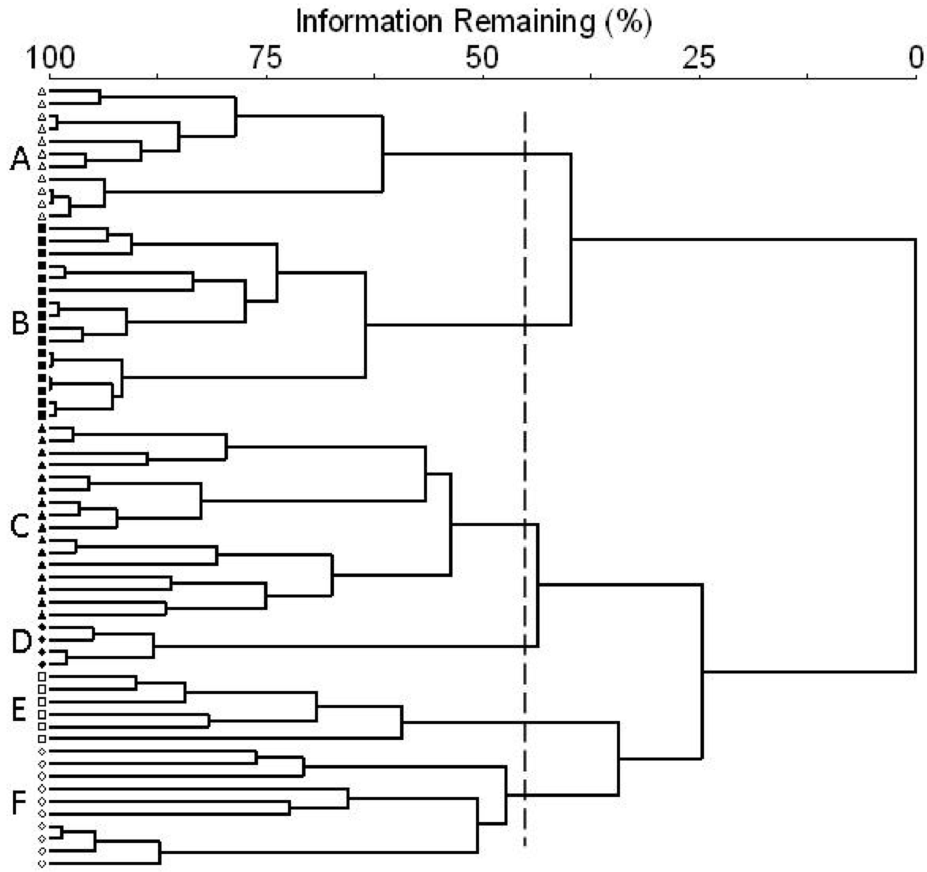


Figure 2.8. Dendrogram produced by the hierarchical clustering showing dissimilarities between individual sites and groups of sites in invertebrate taxonomic composition. The compositional distance between sites and groups of sites was scaled by *Wishart's* [1969] objective function expressed as the percentage of information remaining.

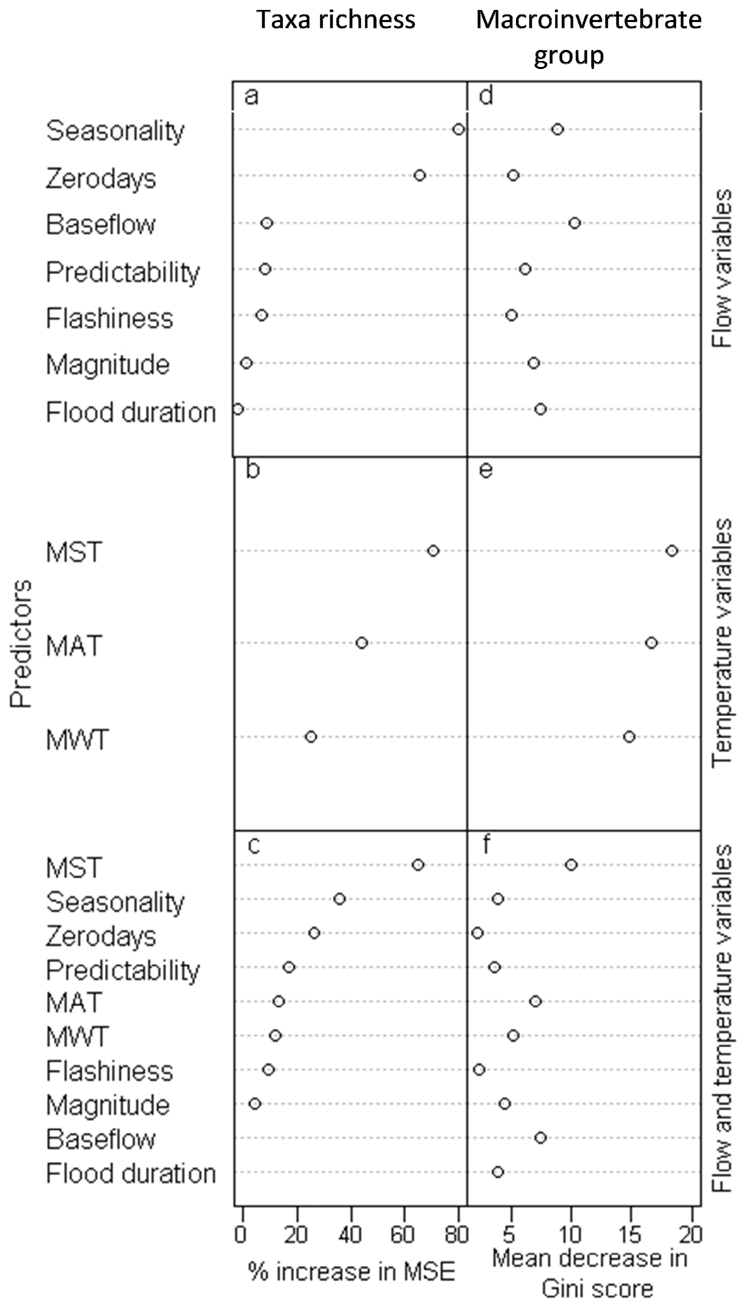


Figure 2.9. Variable-importance plots from Random Forests models for predicting taxa richness (a, b and c) and biotic class (d, e and f). Flow predictors only (a, d); temperature predictors only (b, e); and both flow and temperature predictors (c, f). Predictor variables are ordered in the same sequence for both taxa richness and macroinvertebrate group to facilitate comparisons. Mat = Mean Annual Temperature, MST = Mean Summer Temperature (Jun, Jul, Aug), MWT = Mean Winter Temperature (Dec, Jan, Feb)

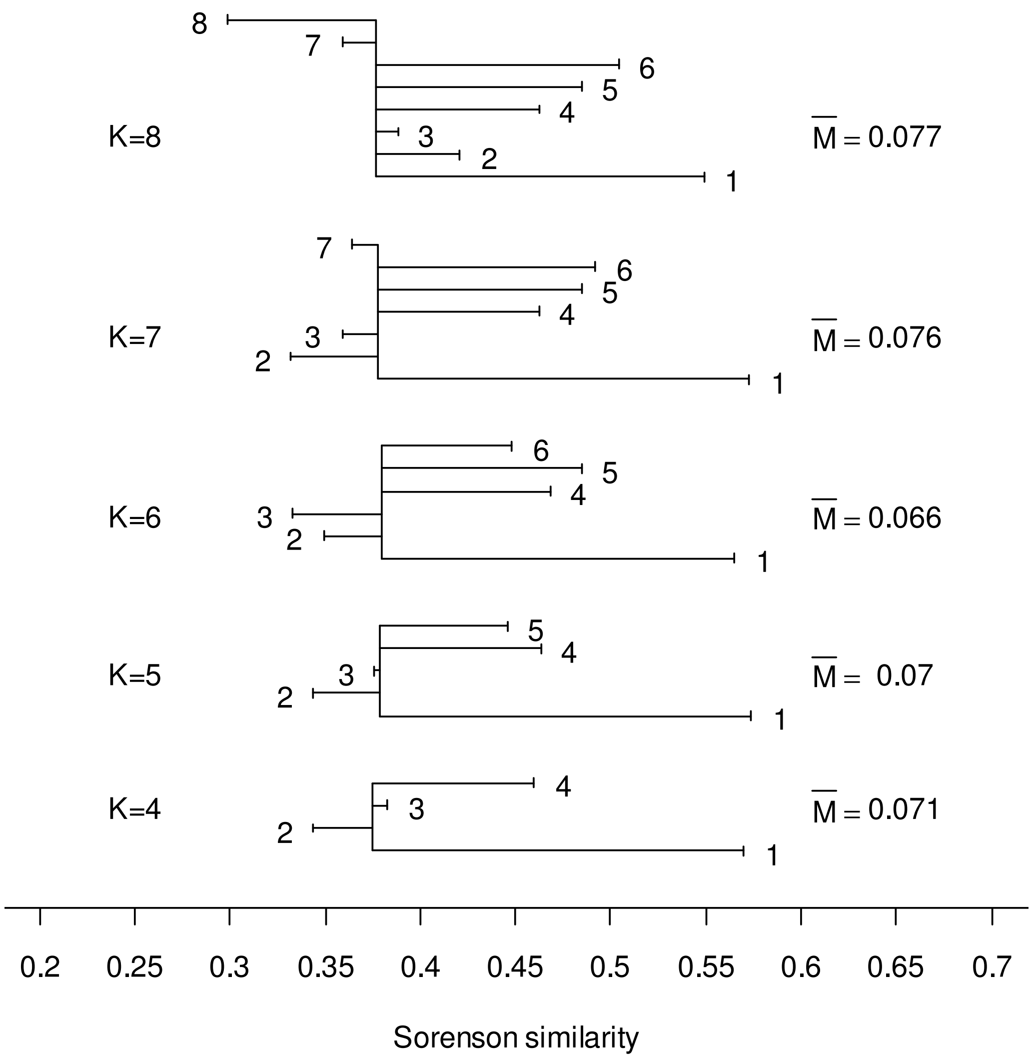


Figure 2.10. Mean similarity dendrograms showing strengths of the different flow regime classifications in accounting for variation among sites in invertebrate assemblage composition (presence-absence). $\overline{M}$ is the mean classification strength of the classes within each classification and the length of dendrogram branches illustrates the relative classification strength of individual classes.

CHAPTER 3

A TOOL FOR THE RAPID AUTOMATIC CHARACTERIZATION OF WATERSHEDS SPREAD OVER LARGE DIGITAL ELEVATION MODELS[1]

Abstract

Geographic Information System (GIS) methods for watershed and stream network delineation are based on the derivation of flow direction and flow accumulation from Digital Elevation Models (DEM), and the demarcation of watersheds upstream of specified outlet sites. This method can be laborious when the number of watersheds to be delineated is large, as one has to fill sinks in DEMs, process flow direction and flow accumulation and calculate watershed properties for each outlet site. Further, if the watershed outlet site location is not exactly on the digital representation of the stream, GIS based methods may result in the wrong watershed boundary. Additionally, when the sites are spread over relatively large geographical area, DEMs and other raster datasets necessary for watershed delineation can be large and may not be handled well by the currently available watershed delineation tools. This paper presents a tool developed from the functionality of ArcGIS and TauDEM that is specifically designed to delineate multiple watersheds spread over large raster data sets, and has capabilities to adjust the outlet site locations to the nearest streams based on the flow direction grid, if they are not already on the stream. This tool can be used in a batch process to delineate many

---

[1] Coauthored by Kiran J. Chinnayakanahalli, David G. Tarboton, Ryan A. Hill, John R. Olson, and Charles P. Hawkins.

watersheds in one run. This tool also provides options to delineate watersheds based on contributing area or a terrain curvature approach that better reflects variable geomorphology, and can objectively choose the appropriate threshold to delineate stream networks based on a stream drop test designed to identify the highest resolution stream network consistent with geomorphological scaling properties. Additional capabilities of this program include the computation of geomorphic variables such as hypsometric curve indices, shape factors, stream network geomorphology attributes, and average watershed properties from input grids. This tool is useful in deriving watersheds, stream networks and watershed attributes of importance to a variety of problems in hydrology, stream ecology and geomorphology.

3.1. Introduction

Watersheds are widely accepted as the basic functional unit for water resources management studies. For example, various state agencies along with the U. S. Environmental Protection Agency regulate Total maximum Daily Loads (TMDL) based on watersheds [*Tong and Chen*, 2002]. Numerous studies have used watershed information to develop relationships between watershed characteristics and streamflow variables of interest [*Vogel and Kroll*, 1992; *Vogel et al.*, 1999; *Kroll et al.*, 2004; *Sanborn and Bledsoe*, 2006]. Further, regional studies need information about watersheds to estimate parameters of various rainfall-runoff models [*Yadav et al.*, 2007; *Zhang et al.*, 2008]. Ecological studies use watersheds as the basic unit for quantifying the effects of geomorphological, geological and hydrological characteristics on structure and function of aquatic ecosystems [*Poff and Ward*, 1990; *Poff*, 1996; *Baeza Sanz and

Garcia del Jalon, 2005]. In many of these studies, watershed data is derived by applying the geographical boundary of the watershed to spatial data and then relating these watershed attributes to appropriate field measurements. Additionally, many studies also require a representation of the stream network. Auxiliary information such as elevation statistics, drainage density, etc. are then derived from or during the creation of the watershed boundary.

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 characteristics at a large number of locations, spread across large areas. Increases in computational power, GIS capabilities and availability of spatial data have made it possible to derive both watersheds and their characteristics digitally. The U.S. Geological Survey has developed a nationwide program called Streamstats [*Ries et al.*, 2008] 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 ungagued sites; it can also delineate watersheds and provide other useful watershed attributes at a user specified location. Nevertheless, delineating a large number of watersheds spread across large regions is still cumbersome due to the processing burden of working with large DEMs and due to steps in the process that require manual intervention such as precise positioning of outlets on digital streams.

Determining the relationships between field measured data and watershed attributes requires investigators to go through the cumbersome process of delineating watersheds upstream of each point where field data are collected. Generally, coordinates are recorded at a sample site using a GPS instrument. However, field site coordinates may not provide accurate watershed delineations because of human recording or instrument error, or differences between the actual and the modeled stream. Delineating watersheds from grid based digital elevation models (DEMs) requires the creation of a grid model of the stream network, and the position of the outlet for the watershed should coincide exactly with the stream model. When the two do not coincide due to error in the stream model or in the site's coordinates, the resulting delineation will be incorrect (Figure 3.1). Even when the outlets coincide with the modeled stream, they should be checked to see if they lie on the correct stream, since in cases where outlets are close to tributary junctions, outlets can be placed on the wrong stream, resulting in the wrong watershed being delineated due to the difference in surface flow paths [*Jensen*, 1991]. The outlet can be manually repositioned to solve this problem. For example, Streamstats requires that each outlet be manually selected from a GIS web based interface, but this can be laborious when number of such outlets is large [*Lindsay et al.*, 2008].

Another significant issue with delineating watersheds spread across broad geographical regions such as states or provinces is that the grid datasets may exceed the available computer memory, or may be too large for the available GIS algorithms. Although subsets of large grids can be created manually, this is a cumbersome approach to analyses at landscape scales.

This paper presents a tool called the Multi-Watershed Delineation (MWD) tool developed using ArcGIS and TauDEM (http://www.engineering.usu.edu/dtarb/taudem/) functionality to help quickly delineate watershed boundaries and derive watershed attributes for a large number of watersheds across broad geographical regions. The MWD tool can also correct outlets that are not positioned exactly on the streams derived from the DEM. It also derives watershed attributes that can aid in analyses involving watershed characteristics. The MWD tool comes in two versions: 1) a standalone windows program with a graphical user interface (GUI) and 2) a command line executable program. The first version can delineate watersheds within one large geographical region. The second version can be used in batch processing to extend MWD tool's capability to delineate watersheds within multiple regions. The MWD tool software and support materials are available for download at http://hydrology.neng.usu.edu/mwdtool.

3.2. Watershed Delineation from DEMs

Although watersheds are easy to conceptualize and delineate on a paper map, GIS delineations are less labor intensive, more reproducible, and less dependent on subjective judgment [*Djokic*, 2000]. GIS based watershed delineation processes construct a watershed boundary for each outlet by identifying all the locations connected to the outlet via overland flow paths [*Band*, 1986]. However, before delineating watershed boundaries from DEMs, several processing steps must be completed. These include pit filling, and the creation of flow direction and flow accumulation grids.

DEMs commonly contain local depressions called pits, which in most cases are artifacts from the DEM creation processes. Pits are grid cells that are completely surrounded by higher elevation grid cells making it impossible for flow to drain from the pit to any of the neighboring cells. This causes discontinuity in the routing of flow across the DEM. Therefore pits are removed from the DEM by increasing the elevation within the pits to elevations just sufficient to make them drain into one of their neighboring cells. This ensures proper surface flow routing across the DEM.

The pit filled DEM can then be used to derive a flow direction grid. A widely used flow direction algorithm, the D-8 method [*Marks et al.*, 1984; *O'Callaghan and Mark*, 1984; *Band*, 1986; *Jenson and Domingue*, 1988] assigns a number to each cell indicating the direction of the flow leaving the cell. This is in the direction of the steepest descent between the focal cell and its eight neighboring cells. Once the flow directions are assigned, flow paths can be traversed to identify all the cells that contribute flow to any grid cell. The total number of cells contributing flow to a focal cell multiplied by the grid cell area is the flow-contributing area for the focal cell. Flow-contributing areas are then assembled into a flow accumulation grid. A simple way to define a drainage network is to apply an area threshold to the flow accumulation grid. Channels are mapped as those cells with contributing areas equal to or exceeding the threshold. For a given location on this digitally mapped channel, a watershed can be mapped by simply assembling all of the upslope cells contributing flow at the location.

Several studies have shown shortcomings in stream networks derived from grid based DEMs in terms of the accuracy of the network structures [*Saunders and Maidment*, 1995; *Soille et al.*, 2003], stream length [*Callow et al.*, 2007; *Paz and Collischonn*, 2007] and watershed area [*Baker et al.*, 2006]. Many researchers have suggested a DEM reconditioning method commonly called "stream burning" [*Mizgalewicz and Maidment*, 1996; *Saunders and Maidment*, 1996; *Callow et al.*, 2007] to improve the accuracy of both stream network structures and watershed areas. Stream burning integrates a vector representation of the stream network with the DEM during the process of pit filling. Stream burning improves the grid representation of the stream by trenching the DEM at known stream locations indicated by vector stream network data. Stream burning should only be carried out when the vector stream network data is considered more accurate than the stream network obtained by the unconditioned DEM.

For a given outlet, general steps to delineate a watershed are a) obtain a DEM that encompasses the watershed being delineated and the vector stream network data for the region, b) remove pits from the DEM along with stream-burning, c) derive flow direction grid, d) derive flow accumulation grid, e) derive drainage network grid and f) make sure the outlet is exactly on the modeled stream of the drainage network grid, and g) delineate the watershed boundary for the given outlet.

3.3. Data Preprocessing and Setup

A regional scale DEM is typically used in the MWD tool to create watersheds (Figure 3.2). In the example presented here, the DEM encompasses a large part of the state of Utah, US. Note that a large geographical region does not necessarily mean a large DEM file size, because it depends on the dataset resolution. This particular example has 12187 × 17200 grid cells with cell size approximately 30 × 30 m. Within the region, there are 136 site locations that require watersheds to be delineated.

The MWD tool uses two types of hydrological boundaries called Large Hydrologic Unit (LHU) and Medium Hydrologic Unit (MHU). The role of MHU within MWD tool will be explained more elaborately in the following section. For now, it is sufficient to note that MHU consists of multiple polygonal regions, while LHU is a composite of these polygons (Figure 3.3). In this paper, we used USGS 4-digit Hydrological Unit Code (HUC) polygons as LHU and the corresponding 8-digit HUC polygons as MHU. Importantly, it was convenient and efficient for us to use 4-digit HUC regions for organizing our grid data across the western US. This was also useful to set up the MWD tool to run in a batch mode.

To delineate watersheds, the MWD tool requires intermediate grids that represent hydrologic characteristics of the landscape. The creation of these grids is resource intensive but once created can be stored for delineating watersheds within the same region [*Djokic*, 2000], thus the MWD tool is time and resource efficient. In this paper we will call the permanent raster datasets "hydrologic grids". The geographic data needed to create the hydrologic grids to run the standalone MWD tool include two shapefiles and a raster (Table 3.1). The data other are obtained from various internet sources, primarily the USGS National Hydrologic Dataset (NHD) and National Elevation Dataset (NED).

An outlet file is created from points the user wants to delineate, based either on field GPS coordinates or points chosen from a map. For a given set of outlets requiring watershed delineation, we proceed by first identifying the 4-digit HUC they are contained within, and develop the necessary DEM data. For the example in Figure 3.2, we identified the 4-digit HUC (HUC 1602) which contains all the outlets and obtained the DEMs from National Elevation Dataset corresponding to this 4-digit HUC. The NED source provides DEMs in rectangular areas that are, in most cases, much smaller than the 4-digit HUCs. For each 4-digit HUC, we merged numerous NED DEMs to generate a large DEM representing the 4-digit HUC, with a sufficient buffer area around the 4-digit HUC to ensure that all hydrologic boundaries are captured within the polygon boundary. The resulting DEM will be called a regional DEM. From the regional DEM, we derive the required hydrologic grids from an ArcGIS Arc Macro Language (AML) script we developed. The AML script combines several sequential commands from the ArcInfo GIS software package into a single process to create the hydrologic grids from the input data listed in Table 3.1 (Figure 3.4). We created hydrologic grids for all 4-digit HUC regions within the western US (Figure 3.5). This organization of data enables us to rapidly delineate multiple watersheds anywhere in the western US. The GUI based MWD tool can handle only one LHU region at a time, but can be run in a batch mode that can delineate watersheds across multiple LHU regions in one processing step. Batch mode delineation is explained later.

3.4. How MWD Tool Works?

When delineating multiple watersheds across a large geographic area, two technical challenges may arise that most delineation tools are not equipped to handle. Firstly, the coordinate of the outlet may not coincide with the digital representation of the stream. This as mentioned earlier will lead to an error such as the incorrect watershed boundary being delineated. To solve this problem the MWD tool "rolls" an incorrectly placed outlet downhill until it contacts the stream channel with the help of the flow direction grid. The watershed is then delineated from this new location. The MWD tool first identifies if an outlet is not on the stream by comparing the outlet position with the stream network grid. When the outlet is found not to be on the stream, the MWD tool uses the flow direction grid to move the outlet to the down slope grid cell based on the flow direction and continues to do so until the outlet is placed on the stream-network.

In some cases, the user may not wish to move the outlets beyond a certain distance. For example, when delineating an ephemeral stream, it is commonly observed that due to the lack of digital representation of such streams, the outlet will be very far from the nearest modeled stream. In such cases, the outlet will be moved indefinitely until it encounters a stream which may often be the wrong stream. It is better not to move the outlet at all and examine them more carefully. This can be controlled in the MWD by inputting a maximum distance an outlet should be moved.

Secondly, delineating watersheds across broad geographic areas requires grid datasets too large for the memory of most computers. The MWD tool solves this problem by automatically clipping large hydrologic grids to a more manageable size with the use of MHU regions (Figure 3.6). The underlying assumption is that the MHU completely contains the watershed being delineated. This assumption is not always met and in such cases, the problem needs to be resolved by moving to coarser DEM and larger MHUs.

To ensure that the large hydrologic grids contain complete medium hydrologic units, we create the hydrologic grids from a series of medium hydrologic units merged to create a LHU. Although in this study we used the USGS 8-digit HUCs for medium hydrologic units and 4-digit HUCs for large hydrologic units, either hydrologic units may be smaller or larger depending on user's need. The MWD tool uses a MHU to clip only the area required for delineation from each of the hydrologic grids (Figure 3.6). During the clipping process, a buffer is added to the MHU to ensure that the polygon captures the hydrologic boundaries present in the landscape. Each site within a medium hydrologic unit is then delineated automatically with the routines of the TauDEM program [*Tarboton and Ames*, 2001]. The MWD tool then moves to the next MHU and repeats the process until it delineates all of the sites within the LHU. By sequentially processing the MHUs and the outlets within each MHU, the MWD tool provides an efficient way to derive multiple watersheds, their stream networks and associated topographic attributes.

3.5. Steps in MWD Tool

There are four executable steps in the GUI based MWD tool (Figure 3.7). The first step creates the association between the outlets and the MHUs. This is essential for the following steps and tells the grid cutting functions if a particular MHU has any outlets to be delineated, and if so, appropriate boundary coordinates are provided to do the clipping.

The second step checks if the outlets are positioned on the modeled streams. If not, this step repositions the outlet to lie on the modeled stream (see the previous section for details). This step also takes in as input the number of cells to move and if after moving the input number of cells, the outlet is still not on the modeled stream, the outlet is not moved at all. This is reflected in the output file created by this step.

There is no computation involved in the third step. It merely gives the user a choice to select a set of watershed variables to be calculated during the delineation process. Some of the watershed attributes like shape factor take considerable resources and time and if the user does not require them, they can be bypassed.

Watershed boundary and attributes are calculated in the fourth step. This step also requires the user to select between two network delineation algorithms (contributing area threshold or curvature threshold) and whether to perform drop analysis [*Tarboton and Ames*, 2001] to objectively estimate the threshold for delineating stream networks. For each outlet, the successful run of MWD will create two files in the ESRI shapefile format [*Environmental Systems Research Institute Inc.*, 1998]: a) watershed boundary and b) stream network.

3.6. Watershed Variables

The following watershed variables are calculated and saved in the polygon shapefile representing the boundary of the watershed.

1. Watershed area (Area): Area of the watershed in the horizontal units of the DEM (e.g., m^2).
2. Elevation statistics: Minimum, maximum, mean and standard deviation of elevation within the watershed boundary calculated from the DEM. Elevations are in the same vertical units as the DEM.

3. Elevation - relief ratio based on mean elevation (RR_{mean}): RR_{mean} is described in *Pike and Wilson* [1971] as,

$$RR_{mean} = \frac{(E_{mean} - E_{\min})}{(E_{\max} - E_{\min})} \tag{1}$$

4. Hypsometric curve indices: A watershed's hypsometric curve describes how elevation changes as one moves down through a watershed. The curve is a plot of the percentage of area greater than each elevation value (Figure 3.8). The MWD tool automatically determines the elevations for 15 increments of percent watershed area that can then be used to plot a hypsometric curve for each watershed. The output fields are prefixed with "Hypso" and a numeric suffix is attached that describes the area percentile reported for that field. For example, Hypso50 reports the elevation contour line above which 50% of the catchment area occurs. The 15 percentage increments are 1, 3, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, 97 and 99.

5. Elevation - relief ratio based on median elevation (RR_{median}): The hypsometric values are also used to calculate the median dimensionless elevation - relief ratio (RR_{median}):

$$RR_{median} = \frac{(Hypso50 - E_{\min})}{(E_{\max} - E_{\min})} \tag{2}$$

6. Drainage density: Drainage density, calculated as the length of channels per unit area, depends on the method used to derive the stream network. MWD tool can delineate the stream network based on two methods based on a) the support area

threshold (described earlier) and b) the DEM curvature based threshold. In the second method, upward curved grid cells are identified by following the procedure developed by *Peuker and Douglas* [1975]. The upward curved grid cells are then used as a weighting field in a weighted drainage area computation. A threshold on this weighted drainage area grid can then be used to define a drainage network.

Using an arbitrary threshold in the above methods create an arbitrary stream network which may not satisfy the empirical network scaling laws [*Horton*, 1945; *Schumm*, 1956]. Therefore MWD also provides options, whereby the stream initiation thresholds can be objectively selected using the method of *Tarboton and Ames* [2001]. The threshold is selected by examining the geomorphologic properties of the resulting stream network across a range of thresholds. Choosing too small of a threshold results in stream networks that violate empirical scaling laws for river networks. We test on the constant stream drop property [*Broscoe*, 1959] and pick the stream network corresponding to the smallest threshold for which a t-test for the difference between means of first order stream drops and higher order stream drops shows no significant difference. This selects the highest resolution stream network consistent with empirical network scaling laws. This approach also has a physical justification in terms of geomorphological landscape evolution [*Tarboton*, 1991, 1992]. This approach provides an automatic and objective approach to identify drainage density, a basic measure of the scale of the topography relevant for hydrology. *Horton* [1932, 1945] indicated that drainage density is inversely proportional to hillslope length. The use of upward curved grid cells in this method gives it a degree of adaptability to variable drainage density within the DEM domain based on terrain texture quantified in terms of curvature.

7. Shape factors: Two metrics related to the shape of the watershed are calculated.
    a. Shape1: Is the ratio of the watershed area to the square of the longest distance to the outlet on the flow path (dimensionless).
    b. Shape2: Is the ratio of the watershed area to the square of the mean distance to the outlet (dimensionless). Flow distance of each grid cell in the watershed to the outlet is calculated and then averaged to obtain the mean distance to the outlet.

   Small values for shape factors indicate elongated watersheds and larger values indicate circular watersheds.

The second output, a line shapefile, represents the stream network. This shapefile consists of a single line segment for each link of the delineated stream network. A link is defined as "an unbroken section of channel between successive nodes (sources, junctions, or outlet)" [see *Knighton*, 1998]. The information contained in this file is identical to the stream network information from TauDEM's functions. It can be broadly classified into the following two categories,

1. Network topology information: This can be used during network analysis to define connectivity relationships between the different links of the network.

2. Stream network variables: The following stream network characteristics are calculated for each link.
    a. Order: Strahler Stream Order.
    b. Magnitude: Shreve Magnitude of the link, defined as the number of upstream sources.
    c. Length: Length of the link.
    d. Drop: Drop in elevation from the start to the end of the link.
    e. Slope: Average slope of the link (computed as drop/length).
    f. Drainage area at the downstream end of the link.
    g. Drainage area at the upstream end of the link.
    h. Straight line distance from the start to the end of the link.
    i. Distance to the outlet from the downstream end of the link.
    j. Distance to the outlet from the upstream end of the link.
    k. Distance to the outlet from the midpoint of the link.

3.7. Batch Processing

The MWD tool is also built as a command line executable function that can be used in a batch mode to run multiple LHU regions at a time. This allows watersheds to be automatically delineated at any extent, up to entire continents. The batch-MWD program runs in steps similar to the GUI version, but has only three steps. The third and fourth steps of GUI-MWD are combined into the third step of batch-MWD. The batch-MWD reads its input from a text file instead of from a user interface. Figure 3.5 shows an example of the batch file setup to delineate watersheds (step 3) over all LHU regions in the western US. The points shown in Figure 3.5 are the streamflow gauges from the Hydro-Climatic Data Network database [*Slack and Landwehr*, 1992] for which we needed the watershed delineations. There are 562 streamflow gauges in this dataset, used here to illustrate the application of the MWD tool to delineate watersheds spread across large geographical regions in batch mode. It is also an example how USGS hydrological units can be used to organize data for efficient batch processing by the MWD tool. In the first attempt, using approximately 30 m grid cell resolution, the program ran for nearly two days (Step 3) and created 441 watersheds. The drainage area for the created watersheds ranged between 15 km^2 and 12416 km^2.

Note that in Figure 3.5 some of the watersheds are larger than 4-digit HUCs and could not be delineated during the first run of the batch process. To delineate these large watersheds, we ran the MWD tool on coarser DEMs (90m x 90m), utilizing 2-digit HUCs as LHUs and the corresponding 4-digit HUC polygons as MHUs.

3.8. Conclusions and Discussion

Both researchers and managers need to be able to measure watershed variables related to water chemistry, hydrology, geomorphology, and ecology across state or larger sized landscapes, to understand and predict how watersheds function at these large scales. However, users are faced with two major problems when trying to delineate multiple watersheds at these scales: 1) watershed outlets do not always coincide with the modeled stream and 2) data grids at these extents are too large for most computers. The MWD tool was developed using ArcGIS and TauDEM functionality to address both these problems, delineating watershed boundaries while simultaneously deriving stream networks and watershed attributes. The MWD tool can reposition outlets that do not coincide with the modeled stream to the nearest stream by rolling them down the hill following the flow paths of the flow direction grid. And, MWD tool resolves the problem of large DEMs by using hydrologic regions, to clip the DEMs and other grids to manageable size and automatically delineating watersheds for outlets within the clipped region.

The MWD tool incorporates watershed delineation functionality from TauDEM GIS and preprocessing functionality from ArcGIS to provide a methodology for rapid delineation of multiple watersheds and extraction of watershed properties over large areas. The MWD tool takes advantage of the fact that required processing such as pit filling, flow direction calculation etc, need not be repeated for watersheds within the same DEM. It also relies on a coarse large-scale partitioning of the domain into regional watersheds, the USGS HUC watersheds. Once the required inputs: DEM, pit filled DEM, flow direction grid, flow accumulation grid and stream network grid and shapefiles representing outlets and regional watersheds (MHU), are put together, MWD tool can easily delineate the watersheds contained within each MHU.

Another advantage of the MWD tool is its ability to use the drop-analysis algorithm at a regional scale to objectively derive the stream network and its properties. This characteristic of the MWD tool should result in delineating stream networks that are better at representing the natural texture of the topography and hence generate attributes that are relatively better descriptors of the stream networks.

The MWD tool has two modes: 1) GUI based and 2) command line executable. The GUI based mode can be used interactively to delineate watersheds within a single large DEM, while the command executable mode can be used in a batch process to delineate watersheds in multiple large DEMs and regional watersheds.

The MWD tool like any other software has some limitations. Our method for repositioning the outlets not coinciding with the modeled streams by “rolling” down the flow direction grid to the stream will reposition the outlet on the wrong stream if the original point is on the wrong side of the drainage divide. This may be a common occurrence for outlets near stream junctions. An advanced method like the one suggested by *Lindsay et al.* [2008] could solve this problem, but it requires names of the outlet and the stream be matched. This raises another issue, since many first order streams will not have stream names associated with them. The assumption that the watershed being delineated is contained within the MHU can also potentially lead to the MWD tool failing to delineate a watershed. Users interested in delineating larger rivers will find that even when using 8-digit HUCs with 1-2km buffers as MHUs, the MWD tool is unable to successfully delineate watersheds. Our solution is to use the corresponding 4-digit HUC as MHU, and the 2-digit HUC as the LHU. This will inevitably require that we use a coarser DEM to minimize the amount of memory needed for processing. Our experience has shown that delineating watersheds with coarser DEMs results in only minor differences between watershed boundaries. Large watersheds delineated this way have negligible (<5%) differences in watershed area when compared with watershed areas calculated with finer DEMs. However, other watershed attributes, such as stream delineation thresholds and associated drainage density might be more severely affected by using courser DEMs. When delineating a large number of watersheds, we do not expect any single tool to solve all the problems and MWD tool is no exception, but it considerably reduces the effort involved in such endeavors.

The MWD tool is specifically aimed at researchers who are working with regional scale issues and want information for hundreds of watersheds spread across large geographical regions. Any such efforts require lots of data processing and management. We have presented here one way of organizing and analyzing such a huge dataset. Such organization not only helps in efficient management of large grid datasets, but is also helpful in executing the MWD tool in a batch process. We have shown how this tool allows us to delineate multiple watersheds across the western US, so we can begin to assess how different watershed attributes effect processes at the landscape scales where natural resource management occurs.

References

Baeza Sanz, D., and D. Garcia del Jalon (2005), Characterisation of streamflow regimes in central Spain, based on relevant hydrobiological parameters, *J. Hydrol.*, *310*(1-4), 266-279.

Baker, M. E., D. E. Weller, and T. E. Jordan (2006), Comparison of automated watershed delineations: Effects on land cover areas, percentages, and relationships to nutrient discharge, *Photogramm. Eng.Remote Sens.*, *72*(2), 159-168.

Band, L. E. (1986), Topographic partition of watersheds with digital elevation models, *Water Resour. Res.*, *22*(1), 15-24.

Broscoe, A. J. (1959), Quantitative analysis of longitudinal stream profiles of small watersheds, Office of Naval Research, Project NR 389-042, Technical Report N0. 18, Department of Geology, Columbia University, New York.

Callow, J. N., K. P. Van Niel, and G. S. Boggs (2007), How does modifying a DEM to reflect known hydrology affect subsequent terrain analysis?, *J. Hydrol.*, *332*(1-2), 30-39.

Djokic, D. (2000), Dem preprocessing for efficient watershed delineation, in *Hydrologic and Hydraulic Engineering: Modeling and Support with Geographic Information Systems*, edited by D. Maidment and D. Djokic, Environmental Systems Research Institute, Inc., Redlands, Calif.

Environmental Systems Research Institute Inc. (1998), ESRI shapefile technical description.

Horton, R. E. (1932), Drainage basin characteristics, *Trans. American Geophysics Union*, *13*, 350-361.

Horton, R. E. (1945), Erosional development of streams and their drainage basins: Hydrophysical approach to quantitative morphology, *Geol. Soc. Amer. Bull.*, *56*, 275-370.

Jensen, S. K. (1991), Applications of hydrological information automatically extracted from digital elevation models, *Hydrol. Processes*, *5*, 31-44.

Jenson, S. K., and J. O. Domingue (1988), Extracting topographic structure from digital elevation data for geographic information system analysis, *Photogramm. Eng.Remote Sens.*, *54*(11), 1593-1600.

Knighton, D. (1998), *Fluvial Forms and Processes.*, Oxford University Press, New York.

Kroll, C., J. Luz, B. Allen, and R. M. Vogel (2004), Developing a watershed characteristics database to improve low streamflow prediction, *Journal of Hydrologic Engineering*, *9*(2), 116-125.

Lindsay, J. B., J. J. Rothwell, and H. Davies (2008), Mapping outlet points used for watershed delineation onto DEM-derived stream networks, *Water Resour. Res.*, *44*(8).

Marks, D., J. Dozier, and J. Frew (1984), Automated basin delineation from digital elevation data, *Geo Processing*, *2*, 299-311.

Mizgalewicz, P. J., and D. R. Maidment (1996), Modeling agrichemical transport in Midwest rivers using geographic information systems, 338 pp, Center for Research in Water Resources Online Report 96-6, University of Texas, Austin, Tex.

O'Callaghan, J. F., and D. M. Mark (1984), The extraction of drainage networks from digital elevation data, *Computer Vision, Graphics and Image Processing*, *28*, 328-344.

Paz, A. R., and W. Collischonn (2007), River reach length and slope estimates for large-scale hydrological models based on a relatively hill high-resolution digital elevation model, *J. of Hydrology*, *343*(3-4), 127-139.

Peuker, T. K., and D. H. Douglas (1975), Detection of surface-specific points by local parallel processing of discrete terrain elevation data, *Computer Graphics Image Procesing*, *4*, 375-387.

Pike, R. J., and S. E. Wilson (1971), Elevation -relief ratio, hypsometric integral, and geomorphic area-altitude analysis, *Geol. Soc. Am. Bull.*, *82*, 1079-1084.

Poff, N. L. (1996), A hydrogeography of unregulated streams in the united states and an examination of scale-dependence in some hydrological descriptors, *Freshwater Biol.*, *36*(1), 71-91.

Poff, N. L., and J. V. Ward (1990), Physical habitat template of lotic systems: Recovery in the context of historical pattern of spatiotemporal heterogeneity, *Environmental Management*, *14*(5), 629-645.

Ries, K. G., J. G. Guthrie, A. H. Rea, P. A. Steeves, and D. W. Stewart (2008), *Streamstats: A water resources web application*, 6 pp., U.S. Dept. of the Interior, U.S. Geological Survey, Reston, Va.

Sanborn, S. C., and b. P. Bledsoe (2006), Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon, *J. Hydrol.*, *325*, 241-261.

Saunders, W., and D. R. Maidment (1995), Grid-based watershed and stream network delineation for san antonio-nueces coastal basin., in *Proceedings of Texas Water '95: A Component Conference of the First International Conference of Water Resources Engineering*, edited, ASCE, San Antonio, Tex.

Saunders, W. K., and D. R. Maidment (1996), A GIS assessment of nonpoint source pollution in the san antonio—nueces coastal basin, 222 pp, Center for Research in Water Resources, University of Texas, Austin, Tex.

Schumm, S. A. (1956), The evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey, *Bulletin of the Geological Society of America*, *67*, 597-646.

Slack, J. R., and J. M. Landwehr (1992), Hydro-climatic data network: A U.S. Geological survey streamflow data set for the United States for the study of climate variations, 1874 - 1988, 193 pp.

Soille, P., J. Vogt, and R. Colombo (2003), Carving and adaptive drainage enforcement of grid digital elevation models, *Water Resour. Res.*, *39*(12).

Tarboton, D. G. (1991), Experimental studies of drainage basin composition and scale, Proposal submitted to U.S.G.S.

Tarboton, D. G. (1992), The interaction of diffusive and advective processes in the location of channel heads under local equilibrium conditions, *Eos Trans. AGU*, *73*(14), Spring Meeting Suppl., 188.

Tarboton, D. G., and D. P. Ames (2001), Advances in the mapping of flow networks from digital elevation data, paper presented at World Water and Environmental Resources Congress, ASCE, Orlando, Fla., May 20-24.

Tong, S. T. Y., and W. L. Chen (2002), Modeling the relationship between land use and surface water quality, *Journal of Environ. Manage.*, *66*(4), 377-393.

Vogel, R. M., and C. N. Kroll (1992), Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics, *Water Resour. Res.*, *28*(9), 2451-2458.

Vogel, R. M., I. Wilson, and C. Daly (1999), Regional regression models of annual streamflow for the United States, *J. Irrig. Drain. Div., Am. Soc. Civ. Eng.*, *125*(3), 148-157.

Yadav, M., T. Wagener, and H. Gupta (2007), Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, *Adv. in Water Resour.*, *30*(8), 1756-1774.

Zhang, Z. X., T. Wagener, P. Reed, and R. Bhushan (2008), Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization, *Water Resour. Res.*, *44*.

Table 3.1. Inputs Required to Create Hydrologic Grids for Use in the Multi-Watershed Delineation Tool

Inputs	Purpose	Type of File
Large Hydrologic Unit Files	Used to merge DEMs in the AML.	Polygon Shapefile
Stream Files (from NHD or other source)	Used in stream burning.	Line Shapefile
Raw DEM (from NED or other source)	Topographic data needed for creating other intermediate grids.	Grid data

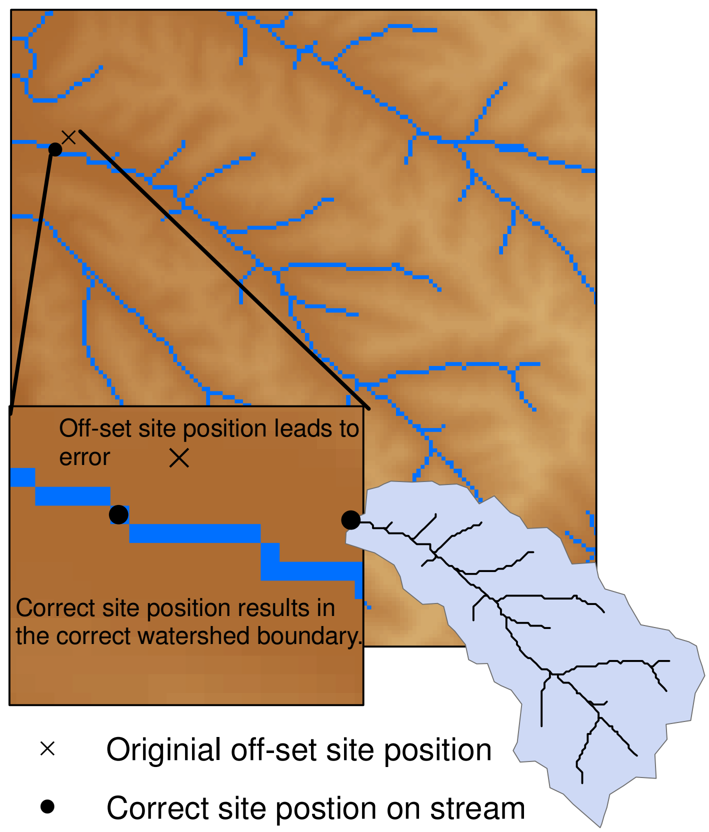


Figure 3.1. Effect of the site offset on watershed delineation.

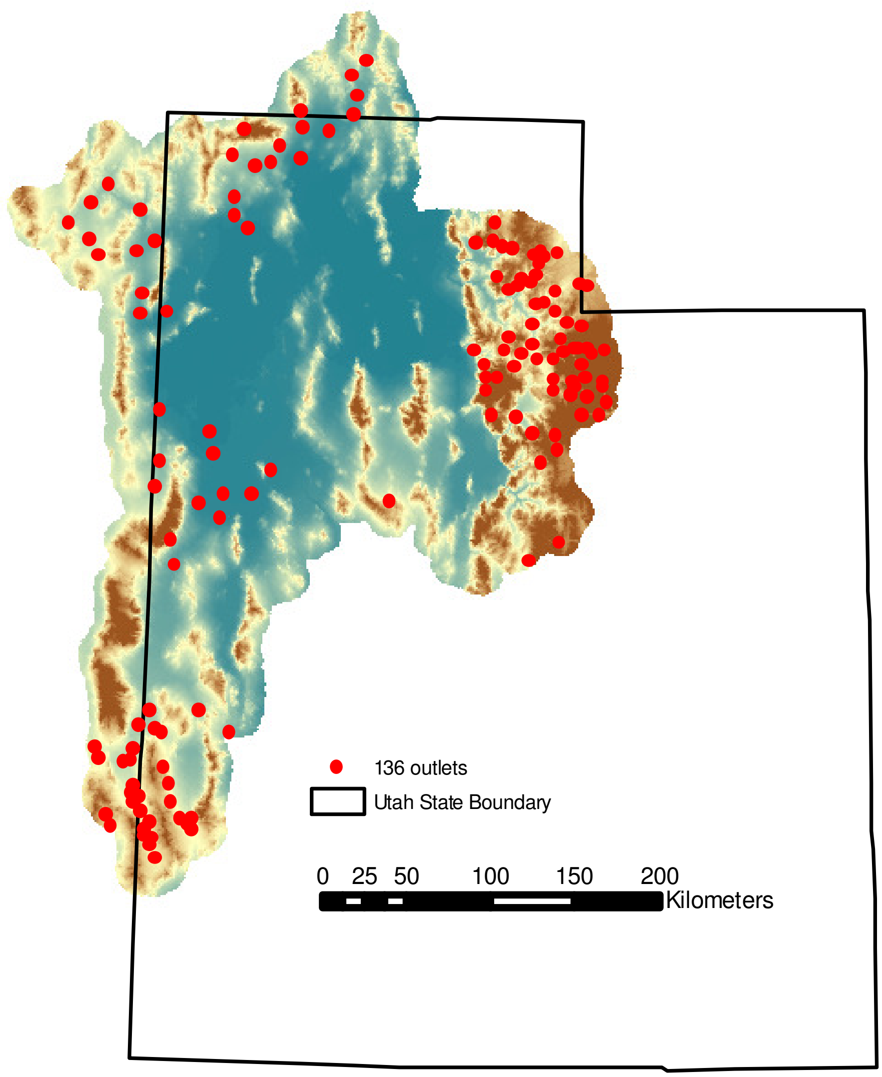


Figure 3.2. A LHU region representing a typical dataset for MWD tool. The region presented here is USGS HUC 1602 and covers sections of Idaho, Nevada and Utah.

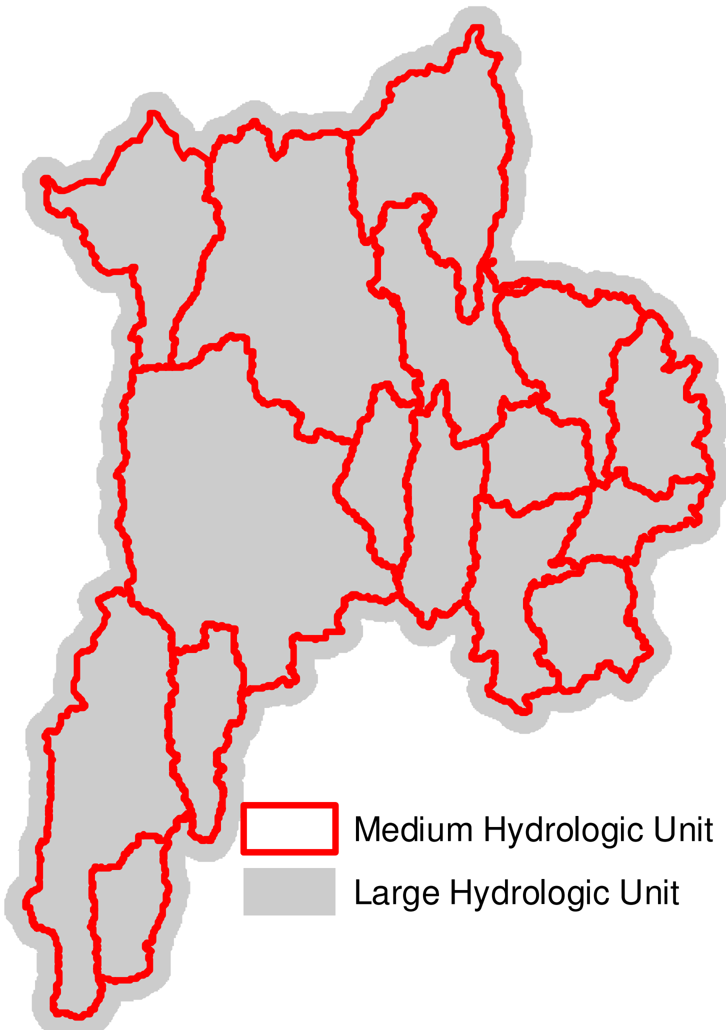


Figure 3.3. Example of LHU and MHU. In this example we have used 4-digit HUC numbered 1602 for representing LHU and the 8-digit HUCs within the LHU represent the MHU.

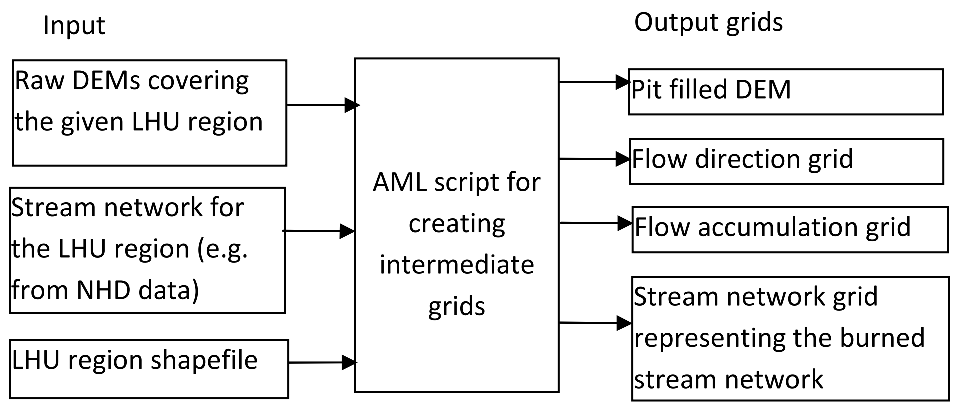


Figure 3.4. The input and output from ArcGIS AML. AML script is executed for each LHU region of interest.

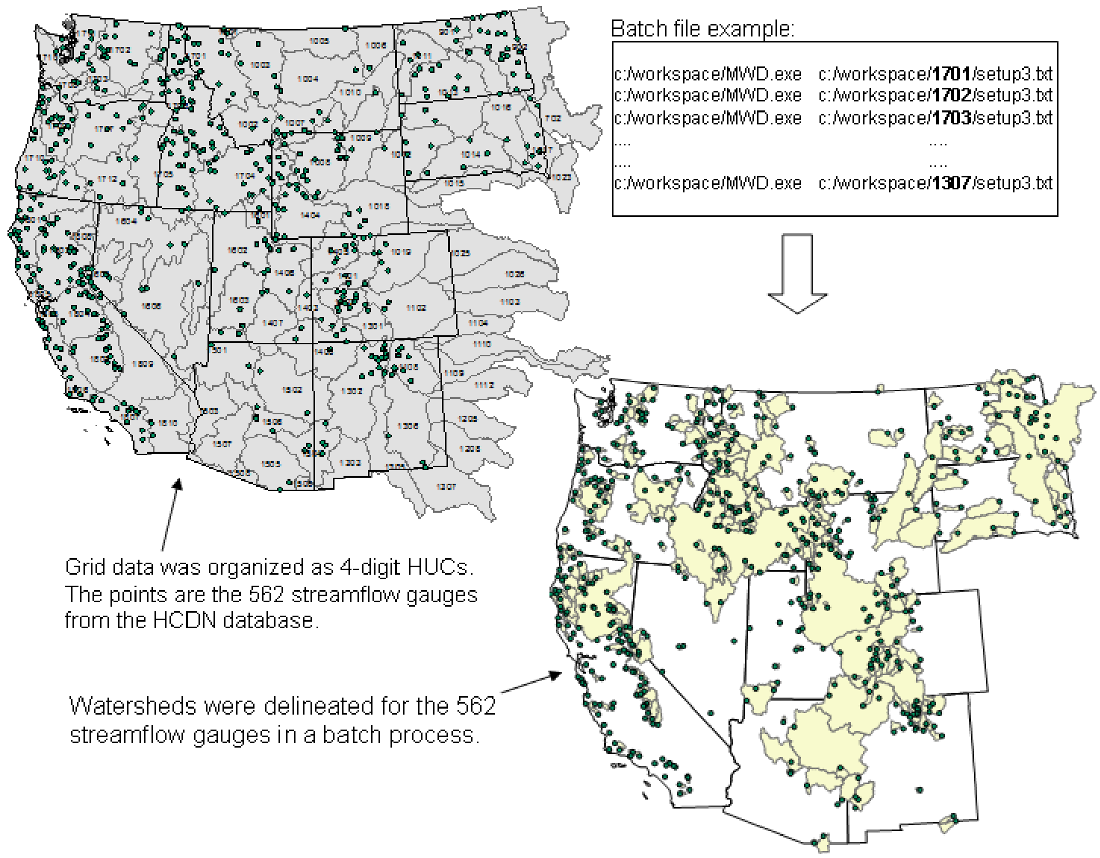


Figure 3.5. The 4 digit HUCs in the western US for which we created the grid data required to delineate watersheds. The bold number represents the folder for each LHU and setup3.txt is the input file for the corresponding LHU. The input file will also tell the MWD tool what step to run.

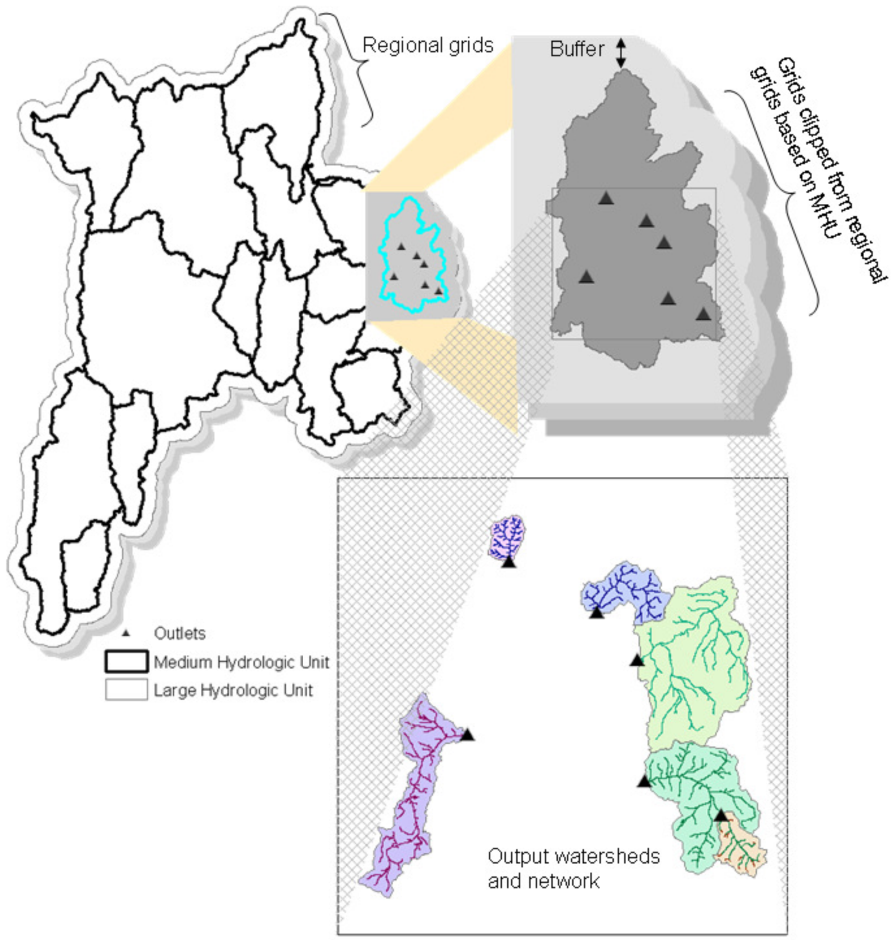


Figure 3.6. The MWD tool subdivides the landscape into smaller units with the use of polygon boundaries for efficient processing.

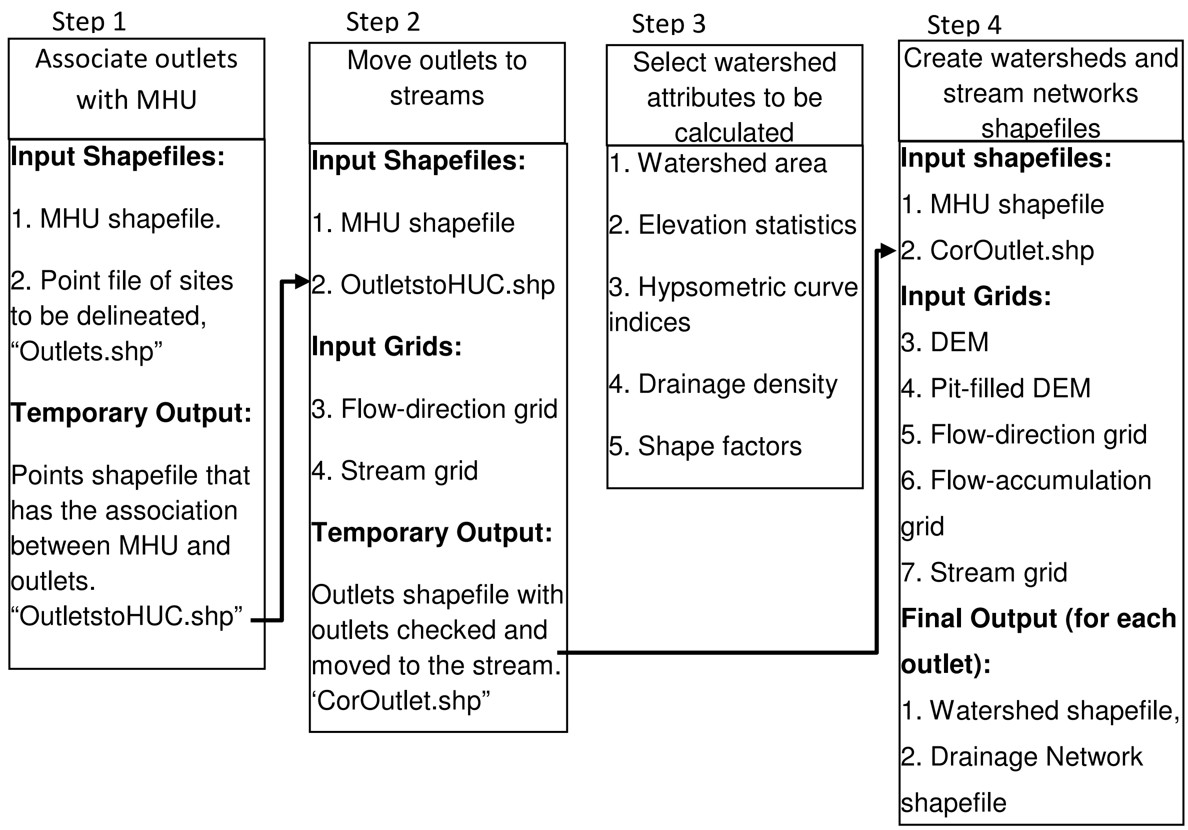


Figure 3.7. Different executable steps in the GUI-based MWD tool.

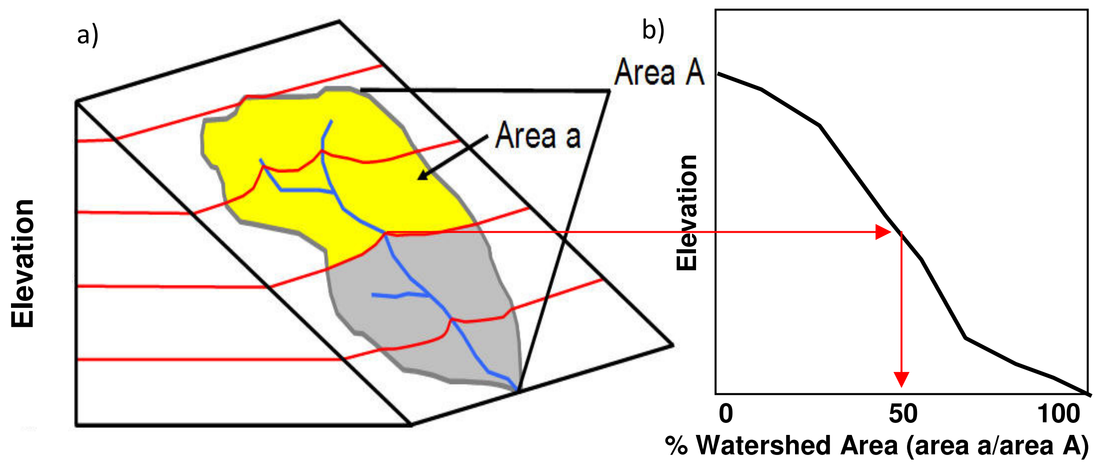


Figure 3.8. a) The elements of watershed hypsometry. Area A is total watershed area, Area A is 50% of total watershed area. b) A Hypsometric curve.

CHAPTER 4

PREDICTING NATURAL STREAMFLOW-REGIME CLASSES FROM WATERSHED ATTRIBUTES[1]

Abstract

Natural streamflow regime classifications are important for a variety of purposes, including bioassessment used in stream ecosystem management. A significant challenge is the extrapolation of natural streamflow regime classes to ungauged watersheds. In this paper, we used four popular statistical methods to develop models to predict streamflow regime class from watershed attributes. The predictions were into streamflow regime classifications based on variables chosen because of their ecological importance and developed using K-means clustering for 541 stream gauge stations in the western U.S. Five classifications with the number of classes ranging from 4 to 8 were used. Watershed attributes used as explanatory variables represented aspects of climate, geomorphology, geology and soil properties. The statistical methods used were Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART), Random Forests (RF) and Support Vector Machines (SVM). For LDA, CART and SVM a 10 fold cross validation method was used to estimate their respective parameters to optimize their performance. A bootstrapping analysis was then carried out to quantify the prediction error. The contingency table from the bootstrapping analysis between the actual and the predicted class was used to estimate the fraction of predictions that were correct, which is a measure of reliability of prediction of each class. For classifications, with number of

---

[1] Coauthored by Kiran J. Chinnayakanahalli, David G. Tarboton, and Charles P. Hawkins.

classes K from 4 to 8, LDA, CART, RF and SVM had median prediction error ranging between 28-40, 30-47, 25-32 and 27-37%, respectively. In terms of this measure of reliability, RF was best for predicting four of the classes, while LDA was best for two classes and CART and SVM were best for one class each. Given the efforts required to optimize the models and their relative performance, RF which requires the least or no optimization is more desirable as a predictive method. SVM with better predictor variable selection and model optimization could potentially perform as well as RF. This work is targeted towards classification based bioassessment where predictions from these models can be compared with observed streamflow regime and the differences used as indicators of hydrologic alteration. Further, biotic composition predicted based on streamflow regime class, and compared to the observed biotic composition may be used as a bioassessment tool for quantifying stream impairment.

4.1. Introduction

Streamflow and its patterns of variability have been considered important for maintenance of the ecological function, structure and composition of the riverine ecosystem [*Resh et al.*, 1988; *Power et al.*, 1995; *Clausen and Biggs*, 1997; *Wood and Armitage*, 2004; *Sanz and del Jalon*, 2005]. Estimation of a river's natural flow regime is frequently sought to serve as a reference point for management of river flows to sustain stream ecosystems [*Richter et al.*, 1996, 2003; *Poff*, 1997; *Bunn and Arthington*, 2002; *Snelder et al.*, 2009]. Natural flow regime classifications group streams into classes that are relatively homogeneous in terms of flow variability and such classifications are promoted as methods for defining units for management of river flows [*Snelder et al.*, 2009]. Stream classification that is relevant to the biota of the stream is in demand for its particular use in bioassessment, monitoring and management of lotic ecosystems [*Wiken*, 1986; *Omernik*, 1987; *Snelder and Biggs*, 2002].

The relationship between streamflow regime and the structure and functioning of the stream biota has not been sufficiently quantified [*Snelder and Biggs*, 2002; *Monk et al.*, 2006]. One way to examine the relationship between the biota and hydrology is to assess if the composition, structure and function of the stream biota are significantly different across natural streamflow regime classes. However, relating ecological measures to streamflow regime is difficult because streamflow is not gauged at many locations of interest where biological samples have been collected. Therefore there is a need to classify streamflow regime for ungauged watersheds, based upon watershed attributes.

One approach to the prediction of streamflow regime for ungauged watersheds is to group streams into homogeneous classes, either based on geographical or hydrological characteristics, and then use regression to predict streamflow variables from watershed attributes. A number of studies [*Riggs*, 1972, 1982; *Jennings et al.*, 1993; *Ries*, 1997; *Vogel et al.*, 1999; *Ries and Friesz*, 2000] have developed separate regressions for each region to predict streamflow variables. Once all the variables representing the streamflow regime are estimated, streamflow regime class can be determined from the classification rules. A second approach is to use statistical methods, like Linear Discriminant Analysis (LDA), to directly predict the streamflow regime class from watershed attributes without going through classification as an intermediate step [*Sanborn and Bledsoe*, 2006].

Although a number of statistical methods have been developed that are potentially promising for use in predicting streamflow regime class [*Breiman et al.*, 1984; *Cortes and Vapnik*, 1995; *Vapnik*, 1996; *Breiman*, 2001; *Hastie et al.*, 2001], there is still uncertainty regarding which method is best given the available watershed attributes and streamflow regime classes we are interested in predicting. There is a need to assess uncertainty and quantify the reliability of prediction models and identify watershed attributes that are effective discriminators for streamflow regime classes.

The objective of this work was to develop statistical models to be able to predict the streamflow regime class of a watershed from watershed attributes. In previous work we used 12 ecologically relevant streamflow variables, computed from the daily streamflow records at 541 sites in the western U.S., to categorize streamflow regime into from 4 to 8 streamflow regime classes (Chapter 2). In this paper these streamflow regime classifications will be represented by the letter K. For example K=4 refers to the streamflow regime classification that has 4 classes. This paper extends our previous work by exploring the capability to predict the streamflow regime class directly from watershed attributes. Four statistical methods of predicting streamflow regime class in ungauged watersheds were evaluated. We also quantified the uncertainty in each method. Among the watershed attributes used, we identified the watershed attributes that were most discriminating of the streamflow regime classes.

Classification is the process of assigning an object to a class based on its attributes [*Hastie et al.*, 2001]. Statistical classification is closely related to pattern recognition, machine learning, and data mining concepts [*Zhao*, 2000]. A classification problem may be framed in terms of a training dataset, *T*, consisting of data points $(x_1,y_1), (x_2,y_2)..(x_n,y_n)$, where x_i is an input vector of length *p* and $y_i \in \{k;\ k=1,2...\ \mathrm{K}\}$ the response or target variable indicating a specific discrete class. Then, the objective of the classification method is to generate a decision rule which can predict the class labels *k* from a new input vector *x*. In the context of this paper, x_i represents the vector of watershed attributes or predictor variables for the i^{th} watershed and y_i the corresponding streamflow regime class.

In developing a classification model the prediction error defined as the percentage or fraction of cases the wrong class is obtained in a test data set is important for selecting among competing models [*Hastie et al.*, 2001]. This applies to each of the four models that are detailed below where it is generally possible to fit training data better by using increasing model complexity. Complexity here refers to the number of parameters or degrees of freedom defined by the structure of the model. To ensure that a model is not overfit, the complexity parameter should be selected by minimizing the prediction error on independent test data using a method such as *K*-fold cross validation [*Hastie et al.*, 2001].

In this paper four classification models that each take a different statistical approach to identifying the response class from input variables were evaluated for their ability to predict streamflow regime classes. Linear Discriminant Analysis assumes that the predictor variables are normally distributed within each class and then classification is by Bayes rule which selects the most probable class from the overlapping normal distributions [see *Hastie et al.*, 2001]. Classification and regression trees (CART) use tree structured classification rules based on a sequence of binary (yes or no) questions to determine class from the predictor variables [*Breiman et al.*, 1984]. Random Forests creates a number of classification trees by randomly sampling a fraction of the testing data and using CART, then assigning classes according to the class that receive the most votes among CART trees [*Breiman*, 2001]. Support Vector Machines (SVM) partition among classes using hyperplanes to serve as class boundaries within the space of the predictor variables. The support vectors are the specific vectors of predictor variables active in constraining any partitioning hyperplane [*Cortes and Vapnik*, 1995; *Vapnik*, 1996]. These methods span the range of statistical approaches available for classification. They have been implemented in a number of statistical packages. We used R implementations for each of them [*R Development Core Team*, 2007].

4.2. Data

Chapter 2 of this dissertation developed a set of streamflow classifications based on streamflow regime variables. These used daily streamflow data from 541 Hydro Climatic Data Network-HCDN [*Slack and Landwehr*, 1992] sites across thirteen states in the western U.S. to estimate 12 ecologically relevant streamflow regime variables (Table 4.1 and Figure 4.3). Chapter 2 then used Principal Component Analysis to reduce the dimensionality of the 12 variable flow data to 7 factors that characterized statistically independent aspects of streamflow. These factors were: 1) zero flow days, 2) magnitude,

3) predictability, 4) flood duration, 5) seasonality, 6) flashiness, and 7) baseflow. These factors are linear combinations of normalized (Box-Cox) streamflow variables (Table 4.2). Chapter 2 used the K-means clustering algorithm based on these 7 factors to classify streams into K classes with K ranging from 4 to 8, resulting in a total of five classifications that were used as response variables in the statistical models.

A wide ranging set of watershed attributes that might serve as predictor variables for streamflow regime class were identified. These included climate variables averaged over the watershed, topographic and geomorphologic variables and geologic and soil variables. These were assembled from nationally available data sources (Table 4.3).

Digital Elevation Model (DEM) Data from the National Elevation Dataset (NED) and stream network data from the National Hydrography Dataset (NHD) were used to derive watershed boundaries and watershed geomorphic attributes for the 541 HCDN sites. We used the Multi-Watershed Delineation tool-MWD (Chapter 3) which uses TauDEM [*Tarboton and Ames*, 2001] and ArcGIS to delineate watershed boundaries and derive watershed geomorphological attributes for sites spread across large geographical areas. Climate, soil and geologic parameters were then aggregated for each delineated watershed. Drainage density, the length of channels per unit area, is a basic measure of the scale of the topography relevant for hydrology that when based on a channel network extracted from a DEM is related to the method used to map stream initiation. An objective procedure [*Tarboton and Ames*, 2001] was used to determine the stream initiation threshold used in the estimation of drainage density.

The boundaries of the gauge watersheds were used to sample climate attributes from the Parameter-elevation Regression on Independent Slopes Model (PRISM) dataset [*Daly et al.*, 1994]. PRISM uses point measurements of climatic data and a digital elevation model to produce grid estimates of climatic variables like mean annual precipitation, mean monthly precipitation, mean monthly temperature etc, using regression based on elevation from nearby locations with similar slope and aspect.

The watershed boundaries were also used to sample the soils attributes from the State Soil Geographic -STATSGO (http://www.ncgc.nrcs.usda.gov/products/datasets/statsgo/) dataset. The STATSGO dataset is generalized from detailed soil survey data and is designed for regional analysis over broad geographic areas. The STATSGO data table links each location to corresponding soil attribute values. This information was used to create a raster data set for each attribute which was then averaged over the watershed boundary to obtain the watershed mean for that attribute. For some soil attributes such as available water capacity, bulk density etc, STATSGO provides high and low values. These were also rasterized and averaged over the watershed boundary to obtain watershed mean high/low values of those attributes (e.g. AWCH_AVE, BDH_AVE etc in Table 4.3).

The geologic attributes (Table 4.3) were sampled from a USGS-Generalized geologic map of the United States [*Reed and Bush*, 2001]. The USGS geological data was converted into grid format where each grid cell was categorized as one of the following geologic types: 1) Gneiss; 2) Granite; 3) Quarternary; 4) Sedimentary; and 5)

Volcanic. Watershed boundaries were then used to compute the percentage of the above geologic types in each watershed.

4.3. Classification Models

4.3.1. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) assumes normal distributions $N(\mu_k,\sum)$ for each class *k* in the space of its input variables, *x*. μ_k is the vector representing the centroid for each class *k,* and $\sum$ is the covariance matrix, assumed to be the same for all classes *k*. The LDA discriminant function is obtained by using Bayes theorem to evaluate the conditional probability of an input x belonging to class k. Under the assumption of normal distributions with equal covariance the following linear discriminant function can be derived [*Hastie et al.*, 2001],

$$L_k = \log(\pi_k) + \mu_k^T \Sigma^{-1}\left(x - \frac{1}{2}\mu_k\right) \tag{1}$$

where π_k is the prior probability for class *k*. Then, LDA predicts for input *x*, the class *k* which gives the maximum value for the discriminant function above. $\mu_k,\sum$ and π_k are estimated from training data as the mean of each class, pooled covariance across the classes and class membership proportions respectively.

LDA requires selection of the set of variables that best discriminate the different classes from among the competing predictor variables. Wilks-lambda [see *Mardia et al.*, 1979] was used in a forward-stepwise variable selection procedure to rank the predictor variables in order of their importance to discriminate the target classes. Wilk's lambda is the ratio of within group variance to the total variance and a small value for Wilk's lambda indicates good discrimination between groups.

We used the lda function available with the MASS [*Venables and Ripley*, 2002] package for R software to develop LDA prediction models. The predictor variables were first normalized by Box-Cox transformations. We implemented a 10 fold cross validation approach to determine the optimum complexity of the LDA model in terms of the number of predictor variables used, denoted as *pLDA*. K-fold cross validation involves splitting the data into K parts. One part is withheld for testing while the remaining K-1 parts are used as a training data set. The prediction error is estimated on the withheld data. The procedure is repeated K times, at each step withholding one part, fitting the model on the other parts and estimating prediction error from the withheld part. Error from the withheld parts accumulated across the K steps quantifies the overall prediction error. The cross validation was used to a) select the number of predictor variables, and b) select the most discriminative predictor variables.

For a given classification of the streamflow regime (one of K= 4 to 8) and for a given training dataset in the cross validation, we used the R function for Wilks-lambda, greedy.wilks from the kLar library [*Weihs et al.*, 2005] to rank the predictor variables in order of their discriminating importance. Then starting from the first two predictor variables from the rank list, LDA models were constructed on the training set, and the model prediction error estimated from the test data. For the same training dataset in the cross validation method, the process was repeated by including the next predictor variable from the ranked list and continued all the way up to 20 variables. Once the prediction error was estimated for varying LDA model complexity for a particular training dataset, the whole procedure was repeated for the remaining training and test datasets of the 10-fold cross validation. The 10-fold cross validation was repeated five times to obtain stable results. The number of predictor variables to be used for the final model, *pLDA*, was the one that gave on average the least prediction error across the five 10-fold cross validation repeats.

Since the ranking of the predictor variables can change with different training data sets used in the cross validation, it is necessary to select the set of variables that would on average be representative of the ranked lists. We considered all 50 ranked lists from the five repeats of 10-fold cross validation and identified the top *pLDA* variables in each list. We then selected the *pLDA* variables that appeared most frequently in these shortened ranked lists. By shortening the lists, we are looking at only the top *pLDA* variables and by choosing the most frequent ones, we have a set of variables that was on average representative of the ranked lists.

4.3.2. Classification and Regression Trees

Classification and regression trees (CART) use tree structured classification rules based on a sequence of binary (yes or no conditional) questions to determine class from the input vector (Figure 4.1). CART does not require the variables to be continuous or have any specific distribution. To start with, all data in T is considered to be at a node and the node is split into two nodes based on a threshold on one of the x predictor variables such that the resulting nodes are less impure. Impurity is a measure of how heterogeneous the nodes are and an ideal split would result in nodes that contain only one class (least impure). The Gini index given by

$$G(m) = \sum_{i \neq j} P_i P_j = \sum_{i=1}^{K} P_i (1 - P_i) \tag{2}$$

where P_i is the proportion of class i observations in node m is used to estimate the impurity at a node m.

The node that was split is called the parent node and the nodes resulting from the split are called child nodes. The first node that contains the entire x from all the classes k is the most impure node. At each node, CART considers all possible splits, n (n is the number of data points) in each predictor variable from the set of p-variables thus forming $n \times p$ possible splits. Each split is quantified by its goodness of split that is a measure of the decrease in impurity given by

$$\Delta G = G(m) - p_L G(m_L) - p_R G(m_R) \tag{3}$$

where $G(m)$, $G(m_L)$ *and* $G(m_R)$ are the impurity in the parent node, child left node and child right node respectively. p_L and p_R are the proportion of data points going into left and right child nodes. CART picks the split that result in the maximum ΔG. Each child node then acts as a parent node for subsequent splitting which is continued until partitioning of the child node no longer decreases the impurity significantly. Once splitting is terminated, the CART algorithm assigns each resulting terminal nodes to a class based on the majority class membership. A new data vector x can then be parsed through the tree based on the splitting rules to determine its classification according to the label of its terminal node. The complexity for CART is defined by the size of the tree and the optimal size of the tree can be determined by *K*-fold cross validation.

We used the Tree- package [*Ripley*, 2007] in the R software to develop CART models. We optimized the size of the tree by a 10-fold cross validation method using the function cv.tree within the Tree package. For each set of streamflow regime classes, the 10-fold cross validation exercise was carried out for different sizes of the tree and the optimum size of the tree for the final model was then determined as the one that gives the least average prediction error.

4.3.3. Random Forests

The Random Forests (RF) method [*Breiman*, 2001] creates a number of classification trees by randomly sampling a fraction of the testing data and using CART to develop a classification tree. The resulting ensemble of trees is called a random forest. A new input vector x is classified by each individual tree in the forest. The classification by each tree is taken as a vote for a class. The RF method then classifies the new input vector as belonging to the class that received the most votes.

Three important considerations in applying RF are, a) from the training set containing n objects, s objects are sampled with replacement to build each tree, b) among p-predictor variables, m $(<<p)$ variables are randomly sampled, and the best split among them is found as in CART and used to split the node and c) each tree is grown until the minimum specified size of the terminal node is reached. The reduction in dimensionality at each split from (b), enables the use of a large number of predictor variables overall, which can be problematic in some methods. The number of trees to be grown (*ntree*),

and number of predictor variables used at each split (*m*) are generally user specified. The minimum terminal node size is generally taken as one, growing each tree to its fullest extent [*Breiman*, 2001]. According to *Breiman* [2001], Random Forest testing error converges to a limit as the number of trees in the forest becomes large.

The RF model estimates prediction error based on the input vectors not used in tree construction, eliminating the need for *K*-fold cross validation. An advantage of RF is that it does not over fit the data as long as there are enough training data points and many trees can be grown without compromising the computational speed.

We used the randomForest package [*Liaw and Wiener*, 2002] in the R software to develop RF prediction models. Unlike in the above methods, RF does not have a complexity parameter and hence 10 –fold cross validation procedure was not used. To be consistent with the above methods, we had the RF model sample 90% of the data without replacement to grow each tree within the forest. This is different from the standard RF method which samples with replacement. The default values for parameters *ntree*, *m* and minimum terminal node size in the R randomForest package were used

In classification mode, RF models estimate the importance of predictor variables by the Gini index score, a measure of the impurity of nodes [*Breiman*, 2001]. The mean decrease in Gini index for a watershed attribute is a measure of reduction in impurity resulting from splits on the watershed attribute. It is summed over all splits and averaged over the number of trees in a RF model. The mean decrease in Gini index is used to assess their relative importance of watershed attributes in discriminating the streamflow regime classes. RF models were developed with the entire dataset for assessing the importance of variables.

4.3.4. Support Vector Machines

The Support Vector Machines (SVM) model developed by Vapnik and others [*Cortes and Vapnik*, 1995; *Vapnik*, 1996] was originally developed for binary classification (i.e $y_i \in \{-1, 1\}$) and is based on finding a hyperplane (Figure 4.2) that separates the two classes and maximizes the distance from the plane to the closest data point from either class [*Vapnik*, 1996]. The separating hyperplane is of the form $\{x : f(x) = x^T\beta + \beta_0 = 0\}$, where $\beta = \{\beta_1, \beta_2, .. \beta_p\}$ is a vector normal to the hyperplane. For two classes that can be linearly separated (Figure 4.2 a) we can find a hyperplane with the biggest margin between the training points by

$$\max_{\beta,\beta_0} C$$

subject to $$\frac{\left(x_i^T\beta + \beta_0\right)}{\|\beta\|} \cdot y_i \geq C, i = 1...n \qquad (4)$$

From linear algebra, it can be shown that $\left(x_i^T\beta + \beta_0\right)/\|\beta\|$ is the signed distance from the hyperplane. The condition in Equation 4 ensures that each data point is at least a distance C distance from the hyperplane and β and β_0 are chosen to maximize C. The constraint in (4) can be stated in an equivalent form as

$$\left(x_i^T\beta + \beta_0\right) y_i \geq C\|\beta\|, i = 1...n \qquad (5)$$

For any β and β_0 satisfying the above condition, any positively scaled multiple also fulfills the conditions, so we can arbitrarily set $\|\beta\| = 1/C$. With this, maximizing C

is equivalent to minimizing $\|\beta\|$ or since $\|\beta\|$ is positive to minimizing $\frac{1}{2}\|\beta\|^2$ for mathematical convenience. The optimization problem then becomes

$$\min_{\beta,\beta_0} \frac{1}{2}\|\beta\|^2$$

$$\text{subject to } \left(x_i^T \beta + \beta_0\right).\ y_i \geq 1, i = 1 \ldots n. \tag{6}$$

which is a quadratic programming (quadratic criterion with linear inequality constraints) problem that can be solved using the standard Lagrange multipliers approach [e.g. *Hastie et al.*, 2001] to obtain sample estimates $\hat{\beta}$ and $\hat{\beta}_0$ that optimally separates the input data.

The optimal hyperplane produced by the function $\hat{f}(x) = x^T\hat{\beta} + \hat{\beta}_0$ is then used for classifying new observations, *x*, as

$$\hat{G}(x) = sign\left\lfloor \hat{f}(x) \right\rfloor \tag{7}$$

For overlapping classes, SVM still maximizes *C*, but allows some points to be on the wrong side of the margin by introducing the slack variables $\xi = \{ \xi_1, \ldots.. \xi_n \}$. The constraint in Equation 4 can be modified to

$$\frac{\left(x_i^T \beta + \beta_0\right)}{\|\beta\|}.\ y_i \geq C(1 - \xi_i) \text{, with } \xi_i \geq 0 \ \ \forall i \tag{8}$$

The value ξ_i is the proportional amount of *C,* by which the prediction can be on the wrong side of the margin (Figure 4.2 b) and $\xi_i = 0$ for points on and on the right side of the margin. As in Equation 6 we recast the optimization problem as a quadratic problem involving the maximization of $\frac{1}{2}\|\beta\|^2$. Incorporating a penalty for data points on the incorrect side of the margin, the problem becomes

$$\max_{\beta,\beta_0} \frac{1}{2}\|\beta\|^2 + \gamma\sum_{i=1}^{n}\xi_i \tag{9}$$

subject to $\xi_i \geq 0, \left(x_i^T \beta + \beta_0\right) y_i \geq 1 - \xi_i, i = 1...n$

where γ is a parameter that controls the relative weight given to the penalty part of the objective. As before, this is solved using standard Lagrange multiplier methods.

Whereas the separable case resulted in a classification using no parameter, the result for the overlapping case depends on the parameter γ. This was optimized using *K*-fold cross validation. The classification rule for a new observation is given by Equation 7 as before.

This support vector classifier still defines linear boundaries between classes, but as with other linear methods, we can modify it to suit the non-linear case by using basis expansions. Once the basis functions $h_m(x)$, m=1,..M are decided, the procedure is the same as described above but now we would use the transformed features $h(x_i) = (h_1(x_i)....h_M(x_i))$, i = 1..n. to produce the nonlinear separating function $\hat{f}(x) = h(x)^T \hat{\beta} + \hat{\beta}_0$. The classifier is $\hat{G}(x) = sign\left[\hat{f}(x)\right]$ as before. It has been shown that when the Support Vector classifiers are modified to use basis functions, the solution involves $h(x)$ only through inner products [e.g. *Cristianini and Shawe-Taylor*, 2000; *Hastie et al.*, 2001]. That means we do not need to specify the actual transformation $h(x)$ at all, but only require kernel function $Kr(x,x') = \langle h(x), h(x')\rangle$ that gives the inner products in the transformed space. The optimal separating hyperplane is then given by

$$\hat{f}(x) = \sum_{i=1}^{n} \hat{\alpha}_i y_i Kr(x,x') + \hat{\beta}_0 \tag{10}$$

where α_i is the Lagrange multiplier associated with the data point i and corresponding observed class y_i. This equation only involves the data on the incorrect side and along the margin (support vectors) since the Lagrange multipliers, α_i, are zero for non-constraining data points.

One of the popular choices for kernel functions is the radial basis [*Hsu et al.*, 2009] given by

$$Kr(x, x') = \exp\left(-\kappa\|x - x'\|^2\right) \tag{11}$$

where κ is the kernel parameter. We used this kernel in our models because it is known to perform relatively better than other kernels in most cases [e.g. *Hastie et al.*, 2001; *Hsu et al.*, 2009]. Parameters κ and γ are tuning parameters that are optimized by K-fold cross validation.

SVM can be extended for multiple classes by what is called the "one-against-one" approach, in which k(k-1)/2 classifiers are constructed using each combination of pairs of classes and then using a voting strategy to select the ultimate class by counting votes from each binary classification [*Friedman*, 1996; *Kressel*, 1999]. An arbitrary rule is used to classify the rare cases where two classes have the same number of votes.

We used the svm function in the e1701 package [*Dimitriadou et al.*, 2008] within the R software for developing our SVM models. The radial basis function was used for the kernel option. For each streamflow regime categorization, we used all predictor variables. We first scaled the predictor variables according to *Hsu et al.* [2009] to range between -1 and 1. The tuning parameters, γ and κ were then determined by the grid search method based on 10 fold cross validation method [*Hsu et al.*, 2009]. The ten-fold cross validation process was carried out for pairs of (γ, κ) and the one with the best prediction error was picked. Like *Hsu et al.* [2009], we used exponentially growing sequence of $\gamma = 2^{-5}, 2^{-3}, \ldots, 2^{15}$ and $\kappa = 2^{-15}, 2^{-13}, \ldots, 2^{3}$ for our grid search. The best pairs of (γ, κ) determined for each set of streamflow regime classes define the final SVM models. *Hsu et al.* [2009] suggested a second grid search in the vicinity of the optimized parameters from the first to fine tune the parameters. When we tried this, we found that the results were not stable and hence retained the optimized values of the parameters from the first grid search.

Unlike the other methods used in this study, the R package used for SVM does not provide any measures or tools to interpret the affects of watershed attributes. This is also partly due to the structure of SVM models which does not render itself well for interpretations.

4.3.5. Uncertainty Estimation and Model Comparison

A bootstrapping analysis was carried out to compare the relative performance of the models. Bootstrapping also provided a common basis for model inter-comparison. Once the optimal model for each method was specified, in terms of complexity and input variables, the original data was randomly sampled without replacement to form training (90% of the data) and testing data. The bootstrapping analysis used 500 runs for each model (Figure 4.4). At each run, the data was randomly split into training and testing sets with 481 and 60 data points respectively. Note that this bootstrap estimation of uncertainty is additional to the 10-fold cross validation or bootstrapping included as part of the optimization of each model.

The construction of models on the training data consisted of estimating a) LDA parameters of the discriminant function, b) the splitting rules for CART and c) the SVM coefficients based on the support vectors. No parameters are required in application of the RF model. The models were then used on the test data to estimate a prediction error. Each bootstrap run results in a contingency table between the actual and the predicted classes. The average contingency table from all 500 runs was used to estimate the conditional probability of the observed class given the predicted class. This conditional probability was used as a measure of reliability of a model to predict a specific class.

Each of the bootstrap run contingency tables results can be used to obtain a estimate of the overall prediction error in terms of the fraction of sites misclassified. We thus have 500 estimates of this error from which it is also possible to obtain a distribution of prediction error for each model. The results report percentiles (5%, 50%, and 95%) of the overall prediction errors from each model estimated from these 500 runs. These measures of the distribution of prediction error provide a way to assess the relative performance of the models.

4.3.6. Relationship Between Watershed Attributes and Streamflow Regime Classes

For a given set of streamflow regime classes, K, the empirical distributions of watershed attributes were examined to estimate the separation between distributions among classes. The separation between distributions was quantified using the

Kolmogorov-Smirnov measure, D, where D is the maximum difference between the empirical watershed attribute distributions from two different streamflow regime classes. A high value of D is indicative of the discriminatory power of the attribute to distinguish between two classes. The most discriminatory watershed attributes for each pair of streamflow regime classes was then identified based on their D measure. The D measure helped identify which variables serve as discriminators between classes and where there are classes for which there are no strong discriminating variables, suggesting the need to search for additional variables that discriminate these classes.

4.4. Results

4.4.1. Linear Discriminant Analysis

The average prediction error from 10-fold cross validation was found to generally level off after a certain number of variables for most classes, indicating a point of diminishing returns with added complexity (Figure 4.5). From this analysis, we decided to use 5 predictor variables for K=4 and 10 predictor variables for classifications K=5 to 8.

The specific predictor variables for these LDA models were then identified based on their frequency of occurrence in the top *pLDA* positions in the 50 lists generated in the variable selection step (Table 4.4). The list is ranked according to decreasing frequency.

The results show that only watershed mean temperature (TMEAN_WS) appeared in all the five models, while ELEV_WS, XWD_WS and SQ_KM appeared in four, and precipjul, BDH_AVE and OMH_AVE appeared in three models. There is some indication of the particular classes these watershed attributes might be discriminating, but we cannot readily quantify such performance. For example, watershed area becomes relatively more important for K≥6 possibly because watershed area is a surrogate for flow magnitude and class 6 is dominated by high flow magnitude streams. Similarly, we suspect that ELEV_WS, XWD_WS are surrogates for class 1 streams (seasonal streams) and possibly for class 4 streams (big streams). We also suspect that the appearance of soils/geology attributes mostly for K≥5 is distinguishing class 5 streams (base flow dominated) from other classes.

4.4.2. Classification and Regression Trees

Similar to LDA we see that prediction error reaches a plateau of diminishing returns at between 4 and 10 terminal CART nodes (Figure 4.6). Ten terminal nodes were selected as the optimum size for each streamflow regime categorization. CART models were then developed using the entire dataset with the optimized tree size. The variables that CART identifies (Table 4.5) provide information on the quantities most able to discriminate streamflow classes. Fewer than 10 variables appears in each of these models because the same variable may be used for multiple splits in CART. XWD_WS and ELEV_WS appeared in all the five models, TMEAN_WS and Rdryness appeared in four, precipnov and SQ_KM appeared in three of the models. The variables precipjan and precipnov were highly correlated (correlation coefficient = 0.94) and one of them were used in each one of the classifications, suggesting that some measure of winter precipitation was used by CART in all the models. As for LDA, watershed area became relatively more discriminative for K= 6, 7, and 8 which we suspect is distinguishing class 6. We examined the tree structures for varying datasets (during the bootstrapping analysis) and observed that attributes towards the terminal nodes changed but most trees usually used the watershed attributes mentioned above.

4.4.3. Random Forests

Similar to the previous methods, XWD_WS, precipnov, ELEV_WS and TMEAN_WS were relatively more important than other variables in all models (Figure 4.7). As for the other methods, the area attribute, SQ_KM, is relatively more important for K=6 to 8. There are diminishing returns in terms of variable performance in discriminating between classes lower in the importance plot (Figure 4.7). Most of the soil/geology attributes are near the bottom of the lists, indicating that they were not used very frequently by the RF models.

4.4.4. Support Vector Machines

The optimal SVM tuning parameters γ estimated from the 10-fold cross validation method for K= 4 to 8 were 32, 128, 32, 32 and 32, while the corresponding κ value was estimated to be 2^{-10}, 2^{-10}, 2^{-10}, 2^{-4} and 2^{-8} respectively. SVM models use all the watershed attributes and these parameters do not assist in identifying discriminating attributes.

4.4.5. Uncertainty Estimation

Examining the fractions of predictions that were correct for each class from each model (Table 4.6) indicates that for K=4, the RF model was the most reliable predictor of classes 1, 3 and 4, while the LDA model was the most reliable predictor of class 2.

Percentiles (5%, 50%, and 95%) from the distribution of overall prediction error across the 500 bootstrap runs indicate for K=4 that overall RF has the least prediction error (50^{th} percentile or median error = 25%). For all models, the fraction of predictions that were correct was smallest for class 3. 56% of RF predictions for class 3 were correct. This is the highest correct prediction fraction for class 3 when K=4.

For K=8, the RF model was the most reliable predictor for classes 1, 3, 6 and 8. The LDA model was most reliable for classes 2 and 7; CART for class 4; and SVM for class 5. Again, RF had the least prediction error for all percentiles (median error, 32%) making it a slightly better prediction method.

For K=8, class 5 (BFI dominated streams), was not predicted well by any of the methods. For this class SVM was relatively more reliable (21% correct predictions), while the other models were ineffective in predicting this class. Further, in K=8, we found that 53% of LDA prediction as class 5 were actually from class 4; 98% of RF predictions were class 7 and 51% of SVM predictions were class 7.

Overall for K=8, classes 1, 2, and 4 were relatively well predicted by all the methods. Classes 3, 6, 7 and 8 contribute the most towards the overall prediction error with misclassification of these classes making up 61%, 66%, 68%, and 65% of total prediction error for LDA, CART, RF, and SVM, respectively.

Overall the results indicate that it is generally possible to predict these natural streamflow regime classes from geographically derived watershed attributes with about 70% accuracy. The median error for LDA, CART, RF and SVM classifications into K=

4 to 8 streamflow regime classes ranged between 28-40, 30-47, 25-31 and 27-37% respectively. The RF model was slightly better than the other models.

4.4.6. Identifying Watershed Attributes That Are Discriminators of Streamflow Regime Classes

The top 5 discriminators for each class pair was indentified based on their Kolmogorov-Smirnov statistic, D (Table 4.7). Classes 1 and 2, seasonal streams and small predictable intermittent streams respectively, have the highest difference (max D = 0.97), while classes 7 and 8, small predictable streams and small flashy streams respectively, are least distinguishable (max D= 0.42). The capability for variables to discriminate between classes is also reflected in the separation of, and overlap between their cumulative distributions (Figure 4.8). Aggregating the results in (Table 4.7) for K=8, the average D from the top 5 discriminators was 0.79, 0.84, 0.68, 0.83, 0.68, 0.71, 0.635 and 0.622 for classes 1 to 8 respectively. These average values support the prediction model results where classes 1, 2, 4 were relatively better predicted (have relatively higher D values and hence well discriminated) while classes 3, 5, 7 and 8 were more difficult to predict.

4.5. Conclusions and Discussion

The results indicate an improvement over previously reported values for predicting streamflow regime classes [*Jowett and Duncan*, 1990; *Detenbeck et al.*, 2005; *Beechie et al.*, 2006; *Snelder et al.*, 2009]. Improvement may be due to the classification being different and a more comprehensive suite of predictive variables and statistical methods being used. The fact that flow magnitude is a factor in the classifcation may make prediction easier.

The predictability of a specific class is a function of a) the model's capability to describe the relationship between streamflow regime classes and watershed-attributes, b) availability of a good surrogate measure for discriminating the hydrologic characteristics of classes and c) the proportion of data points in the class.

4.5.1. Statistical Models

We found that relatively good results were obtained from all four statistical methods we evaluated, despite them having varying capability to handle linear and non-linear relationships. LDA requires assumptions of normality of predictor variables and equivalence of class covariance matrices. The fact that LDA performed comparably to the other methods (see Table 4.6) indicates that transformations of predictor variables to normal were sufficient. LDA provided information about the discriminatory power of the watershed attributes used in the model through the variable selection process. However, the process of LDA implementation was the most tedious among the four models used in this work. CART is an attractive method because of its easy implementation and interpretation. It can also handle both categorical and continuous predictor variables. However our examination of the trees resulting from multiple runs during the bootstrapping to optimize complexity indicated structural variability in the lower level splits. Also it was the worst (albeit only slightly so) among the models compared in terms of prediction error (Table 4.6). RF extends the mechanism of CART, but by using an internal bagging procedure similar to bootstrapping is designed to automatically minimize the over-fitting. RF quantifies the importance of watershed attributes which gives it some capability to help understand the interactions between watershed attributes and flow regime classes. Implementing RF required the least amount of effort and overall RF had the best general performance (albeit only slightly so) in terms of prediction error (Table 4.6). SVM is still subject to ongoing research and lacks good variable selection tools. In our SVM implementation we used all watershed attributes and a simple grid based search for selecting the model parameters. We found that SVM comes close in performance to the RF model (Table 4.6) and was the only model that could predict class 5 with any reliability albeit low. This suggests that with better variable and parameter selection methods, SVM may potentially perform as well as or better than RF. However, with SVM, it is difficult to understand the interaction between watershed attributes and streamflow regime classes because of the complicated kernel transformations that occur within the SVM model.

Though RF was overall a better prediction method, we still suggest that other models be used in addition to RF because RF was not the most reliable predictor for all the classes (Table 4.6). We expect that combining information from different models can potentially improve overall prediction.

4.5.2. Class specific Predictions

It was evident in examining class specific prediction error (Table 4.6) that some classes are better predicted than others. The pattern was generally consistent across models indicating that this was in most cases due to the class rather that one of the models discriminating capability. This implies that some classes are poorly predicted mainly due to the absence of good watershed attributes that can discern hydrological differences among the classes, and not the method of prediction. This is also indicated by low Kolmogorov Smirnov (D) statistics and overlaps of the distributions of discriminating variables for the classes that are relatively poorly predicted (Table 4.7, Figure 4.8).

However in certain cases there were model specific differences. The most notable is for class 5 (in K=8) where all models except SVM failed completely in its prediction, while SVM performed better, albeit still poorly. Much of this is due to the small sample size of class 5. Other small sample size classes (class 3, 7, 8) also had lower correct predictions, but the better model in these cases was RF. The difficulties due to small sample size of class 5 suggests attempting to include more class 5 type streams in future analyses.

4.5.3. Watershed Attributes

The models and measures of watershed attribute discriminating capability tended to consistently identify a similar set of watershed attributes (TMEAN_WS, XWD_WS, ELEV_WS, SQ_KM, and monthly precipitation through precipnov, precpjan or precipjul). These attributes appeared in multiple models as well as having high D value for more than one pair of streamflow regime classes.

Classes 1, 2, and 4 have a relatively high measure of D, while class 5 made up of baseflow dominated streams has relatively small values of D (Table 4.7 and Figure 4.8) indicating the presence of attributes that can discern classes 1, 2, and 4, but a lack of attributes that discern class 5 and to some extent classes 7 and 8. Class 7 and class 8 were least separated from each other (smallest D value among the pair of classes). The main difference between these two classes is the flashiness represented by the number of flow reversals. This hydrologic characteristic does not appear to have a good surrogate among the watershed attributes used in this study and hence class 7 and 8 were often misclassified as one another. Class 3 had a similar problem and was not well distinguished from 7 or 8.

We suggest that overall and class specific prediction can be improved not so much by using another method but by developing or identifying better watershed attributes that can distinguish the hydrological characteristics between the streamflow regime classes.

Identifying better discriminators is especially significant for baseflow dominated streams (class 5) because of their importance for biota. Further, better discriminators for small unpredictable streams (class 7) and small flashy streams (class 8), should significantly increase the performances of the models.

References

Beechie, T., E. Buhle, M. Ruckelshaus, A. Fullerton, and L. Holsinger (2006), Hydrologic regime and the conservation of salmon life history diversity, *Biological Conservation*, *130*(4), 560-572.

Breiman, L. (2001), Random forests, *Machine Learning*, *45*(1), 5-32.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (1984), *Classification and Regression Trees*, Wadsworth International Group, Belmont, Calif.

Bunn, S. E., and A. H. Arthington (2002), Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity, *Environ. Manage.*, *30*(4), 492-507.

Clausen, B., and B. J. F. Biggs (1997), Relationships between benthic biota and hydrological indices in New Zealand streams, *Freshwater Biol.*, *38*(2), 327-342.

Colwell, R. K. (1974), Predictability, constancy, and contingency of periodic phenomena, *Ecology*, *55*, 1148-1153.

Cortes, C., and V. Vapnik (1995), Support-vector network, *Machine Learning*, *20*, 273-297.

Cristianini, N., and J. Shawe-Taylor (2000), *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge, UK.

Daly, C., R. P. Neilson, and D. L. Phillips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. App. Meteorol. 33*, 140-150.

Detenbeck, N. E., V. J. Brady, D. L. Taylor, V. M. Snarski, and S. L. Batterman (2005), Relationship of stream flow regime in the western lake superior basin to watershed type characteristics, *J. Hydrol.*, *309*(1-4), 258-276.

Dimitriadou, E., K. Hornik, F. Leisch, M. David, and A. Weingessel (2008), E1071: Misc functions of the department of statistics (e1071), Tu Wien.

Friedman, J. (1996), Another approach to polychotomous classification, 14pp, Online report, Department of Statistics, Stanford University.

Hastie, T., R. Tibshirani, and J. Friedman (2001), *The Elements of Statistical Learning*, Springer-Verlag, New York.

Hsu, C. W., C.-C. Chang, and C. J. Lin (2009), A practical guide to support vector classification, 15 pp, Online Report, National Taiwan University, Taipai, Taiwan.

Jennings, M. E., W. O. J. Thomas, and H. C. Riggs (1993), Nationwide summary of U.S.Geological survey regional regression equations for estimating magnitude and frequency of floods for ungaged sites, USGS Open-file.

Jowett, I. G., and M. J. Duncan (1990), Flow variability in New Zealand rivers and its relationship to in-stream habitat and biota, *N. Z. J. Mar. Freshwater Res.*, *24*(3), 305-317.

Knighton, D. (1998), *Fluvial Forms and Processes.*, Oxford University Press, New York.

Kressel, U. (1999), Pairwise classification and support vector machines, in *Advances in Kernel Methods*, edited by B. Scholkopf, et al., pp. 255-268, MIT Press, Cambridge, Mass.

Liaw, A., and M. Wiener (2002), Classification and regression by randomforest, *R News*, *2*(3), 18-22.

Mardia, K. V., J. T. Kent, and J. M. Bibby (1979), *Multivariate Analysis*, Academic Press, New York.

Monk, W. A., P. J. Wood, D. M. Hannah, D. A. Wilson, C. A. Extence, and R. P. Chadd (2006), Flow variability and macroinvertebrate community response within riverine systems, *River Research and Applications*, *22*(5), 595-615.

Omernik, J. M. (1987), Ecoregions of the conterminous united states, *Annals of the Association of American Geographers*, *77*(1), 118-125.

Pike, R. J., and S. E. Wilson (1971), Elevation -relief ratio, hypsometric integral, and geomorphic area-altitude analysis, *Geol. Soc. Am. Bull.*, *82*, 1079-1084.

Poff, N. L. (1996), A hydrogeography of unregulated streams in the united states and an examination of scale-dependence in some hydrological descriptors, *Freshwater Biol.*, *36*(1), 71-91.

Poff, N. L. (1997), Landscape filters and species traits: Towards mechanistic understanding and prediction in stream ecology, *Journal of the North American Benthological Society*, *16*(2), 391-409.

Power, M. E., A. Sun, G. Parker, W. E. Dietrich, and J. T. Wootton (1995), Hydraulic food-chain models, *Bioscience*, *45*(3), 159-167.

R Development Core Team (2007), R: A language and environment for statistical computing, edited, R Foundation for Statistical Computing, Vienna, Austria.

Reed, J. C., Jr., and C. A. Bush (2001), Generalized geologic map of the conterminous United States, USGS maps.

Resh, V. H., A. V. Brown, A. P. Covich, M. E. Gurtz, H. W. Li, G. W. Minshall, S. R. Reice, A. L. Sheldon, J. B. Wallace, and R. C. Wissmar (1988), The role of disturbance in stream ecology, *Journal of the North American Benthological Society*, *7*(4), 433-455.

Richter, B. D., J. V. Baumgartner, J. Powell, and D. P. Braun (1996), A method for assessing hydrologic alteration within ecosystems, *Conservation Biology*, *10*(4), 1163-1174.

Richter, B. D., R. Mathews, and R. Wigington (2003), Ecologically sustainable water management: Managing river flows for ecological integrity, *Ecological Applications*, *13*(1), 206-224.

Ries III, K. G. (1997), August median steamflows in Massachusetts, USGS, Marlborough, Mass.

Ries III, K. G., and P. J. Friesz (2000), Methods for estimating low-flow statistics for Massachusetts streams, 81 pp, USGS Water Resources Investigation.

Riggs, H. C. (1972), Low-flow investigations, Techniques of Water-Resources Investigations, Book 4, Chap. B1, 18 p, U.S. Geological Survey, Yakima, Wash

Riggs, H. C. (1973), Regional analyses of streamflow characteristics, Techniques of Water-Resources Investigations, Book 4, Chap. B3, 15 p, U.S. Geological Survey, Yakima, Wash.

Ripley, B. (2007), Tree: Classification and regression trees, R package version 1.0-26.

Sanborn, S. C., and B. P. Bledsoe (2006), Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon, *J. Hydrol.*, *325*, 241-261.

Sanz, D. B., and D. G. del Jalon (2005), Characterisation of streamflow regimes in central Spain, based on relevant hydrobiological parameters, *J. Hydrol.*, *310*(1-4), 266-279.

Slack, J. R., and J. M. Landwehr (1992), Hydro-climatic data network: A U.S. Geological survey streamflow data set for the United States for the study of climate variations, 1874 - 1988, 193 pp.

Snelder, T. H., and B. J. F. Biggs (2002), Multiscale river environment classification for water resources management, *Journal of the American Water Resources Association*, *38*(5), 1225-1239.

Snelder, T. H., N. Lamouroux, J. R. Leathwick, H. Pella, E. Sauquet, and U. Shankar (2009), Predictive mapping of the natural flow regimes of France, *J. Hydrol.*, *373*(1-2), 57-67.

Tarboton, D. G., and D. Ames (2001), Terrain analysis using digital elevation models, paper presented at Presentation at Forest Service 2001 Geospatial Conference, Salt Lake City, 16-19, April.

Vapnik, V. (1996), *The Nature of Statistical Learning Theory*, Springer-Verlag, New York.

Venables, W. N., and B. D. Ripley (2002), *Modern Applied Statistics with S*, Springer, New York.

Vogel, R. M., I. Wilson, and C. Daly (1999), Regional regression models of annual streamflow for the United States, *J. Irrig. Drain. Div., Am. Soc. Civ. Eng.*, *125*(3), 148-157.

Vörösmarty, C. J., C. A. Federer, and A. L. Schloss (1998), Potential evaporation functions compared on us watersheds: Possible implications for global-scale water balance and terrestrial ecosystem modeling, *J. Hydrol.*, *207*(3-4), 147-169.

Weihs, C., U. Ligges, K. Luebke, and N. Raabe (2005), Klar analyzing German business cycles., in *Data analysis and decision support*, edited by D. Baier, et al., pp. 335-343, Springer-Verlag, Berlin.

Wiken, E. (1986), Terrestrial ecozones of Canada: Ottawa, *Ecological Land Classification Series no. 19. Hull*, Environment Canada, 26 pp.

Wood, P. J., and P. D. Armitage (2004), The response of the macroinvertebrate community to low-flow variability and supra-seasonal drought within a groundwater dominated stream, *Archiv Fur Hydrobiologie*, *161*(1), 1-20.

Woods, R. (2003), The relative role of climate, soil, vegetation and topography in determining seasonal and long-term catchment dynamics, *Advances in Water Resources*, *26*, 295-309.

Zhao, G. (2000), A new perspective on classification, M.S. thesis, Dep. of Math., Utah State University, Logan.

Table 4.1. Variables Used for Streamflow Regime Classification from Daily Mean Streamflow Data

Streamflow variables	Description
BFI	Average across all years of the ratios of the annual lowest daily flow to the annual average flow expressed as a percentage.
DAYCV	Coefficient of variation of daily mean streamflow.
Qmean	Mean daily discharge.
Q1.67	Daily flow with a 1.67 year recurrence interval. See *Poff* [1996]
ZERODAYS	Average number of days each year with zero discharge.
FLDDUR	Flood duration calculated as the average number of days per year when flow equals or exceeds $Q_{1.67}$.
P	*Colwell's* [1974] predictability.
C	*Colwell's* [1974] constancy.
M	*Colwell's* [1974] contingency.
7Qmin	Average of annual minimum 7 day mean streamflow.
7Qmax	Average of annual maximum 7 day mean streamflow.
$\overline{R}$	Average number of flow reversals per year. Flow-reversals are defined from the daily mean streamflow as days when the trend (increasing or decreasing) from the previous day is reversed.

Table 4.2. Loadings for the Varimax Rotated PC Factors from Normalized (Box-Cox) Streamflow Variables. High Loadings are in Bold Font

Factors	1	2	3	4	5	6	7
BFI	-0.299	0.006	-0.175	0.157	0.060	0.097	**0.895**
DAYCV	0.045	-0.215	0.336	-0.125	-0.316	-0.210	**-0.769**
QMEAN	-0.091	**0.928**	-0.204	-0.017	0.152	0.141	0.207
ZERODAYS	**0.813**	-0.235	0.174	0.012	-0.221	-0.162	-0.408
$Q_{1.67}$	-0.080	**0.951**	-0.120	-0.199	0.115	0.122	0.035
FLDDUR	0.002	-0.181	0.040	**0.967**	0.043	0.005	0.171
P	0.078	-0.144	**0.930**	0.014	0.181	-0.152	-0.203
C	0.127	-0.268	**0.822**	0.060	-0.350	-0.139	-0.272
M	-0.157	0.225	-0.004	0.047	**0.927**	0.101	0.209
$\overline{7Q}_{min}$	-0.200	**0.672**	-0.250	0.049	0.182	0.185	0.582
$\overline{7Q}_{max}$	-0.071	**0.981**	-0.084	-0.086	0.080	0.089	0.005
$\overline{R}$	-0.135	0.274	-0.238	0.005	0.119	**0.885**	0.226
Descriptive characterization	Zero flow days	Magnitude	Predictability	Flood duration	Seasonality	Flashiness	Baseflow

Table 4.3. Watershed Attributes Used in the Statistical Models to Predict the Flow Regime Class

Metric	Description	Unit	Source
XWD_WS	Watershed average of the annual mean of the PRISM mean monthly number of days with measurable precipitation.	DAYS	PRISM
TMIN_WS	Watershed average of the coldest month's PRISM mean monthly air temperature	^{o}C	PRISM
TMAX_WS	Watershed average of the warmest month's PRISM mean monthly air temperature	^{o}C	PRISM
TMEAN_WS	Watershed average of the annual mean of the PRISM mean monthly air temperature	^{o}C	PRISM
MINP_WS	Watershed average of the driest month's PRISM mean monthly precipitation	mm	PRISM
MAXP_WS	Watershed average of the wettest month's PRISM mean monthly precipitation	mm	PRISM
MINWD_WS	Watershed average of the number of wet days in the month with fewest wet days from the PRISM mean monthly number of days with measurable precipitation	days	PRISM
MAXWD_WS	Watershed average of the number of wet days in the month with most wet days from the PRISM mean monthly number of days with measurable precipitation	days	PRISM
MEANP_WS	Watershed average of the annual mean of the PRISM mean monthly precipitation	mm	PRISM
RH_WS	Watershed average of the annual mean of the PRISM mean monthly relative humidity	%	PRISM
SD_TMIN_WS	Standard deviation across each watershed of the coldest month's PRISM mean monthly air temperature.. SD_MIN_WS measures thermal heterogeneity within a watershed during the coldest period of the year.	^{o}C	PRISM
SD_TMAX_WS	Standard deviation across each watershed of the warmest month's PRISM mean monthly air temperature . SD_MAX_WS measures thermal heterogeneity within a watershed during the hottest period of the year.	^{o}C	PRISM
LST32_AVE	Watershed average of the mean day of year (1-365) of the last freeze derived from the PRISM data.	day	
PETbar	Watershed average mean annual potential evapotranspiration	mm	[*Vörösmarty et al.*, 1998]
Rdryness	Climate dryness index [see *Woods*, 2003], the ratio of PETbar to MEANP_WS	-	
deltaE	Seasonal amplitude of potential evapotransipiration [see *Woods*, 2003]	-	Derived from [*Vörösmarty et al.*, 1998]
deltaP	Seasonal amplitude of rainfall [see *Woods*, 2003]	-	Derived from PRISM
Seasonality	Climate seasonality index, |deltaP-deltaE. Rdryness| [see *Woods*, 2003]		
precipJan	Watershed average mean January precipitation	mm	PRISM
precipmay	Watershed average mean May precipitation	mm	PRISM
precipjul	Watershed average mean July precipitation	mm	PRISM
precipsep	Watershed average mean September precipitation	mm	PRISM
precipnov	Watershed average mean November precipitation	mm	PRISM

Table 4.3. Continued

SQ_KM	Watershed area	Km^2	derived from NED
ELEV_WS	Mean watershed elevation	Meter	derived from NED
ELEV_MIN	Minimum elevation in the watershed derived from the National Elevation Dataset and watershed boundaries.	Meter	derived from NED
ELEV_MAX	Maximum elevation in the watershed derived from the National Elevation Dataset and watershed boundaries	Meter	derived from NED
ELEV_STD	Standard deviation of elevation (meters) across the watershed	Meter	derived from NED
SHAPE1	Ratio of the watershed area to the square of the longest distance to the outlet on the flow path	-	derived from NED
MeanSlp	Watershed average topographic slope		derived from NED
StdSlp	Watershed standard deviation of topographic slope		derived from NED
DDEN	Drainage density (see *Knighton* [1998]) in meters of stream per square meter of watershed determined from the stream network as created from drop analysis [*Tarboton and Ames*, 2001]	per Meter	derived from NED
RRMEDIAN	Dimensionless elevation - relief ratio (from *Pike and Wilson* [1971]), calculated as (ELEV_MED-ELEV_MIN)/(ELEV_MAX-ELEV_MIN) where ELEV_MED is the median elevation within a watershed	-	derived from NED
GNEISS	% of gneiss geology in the watershed derived from a simplified version of Reed & Bush (2001) - Generalized Geologic Map of the Conterminous United States.	%	
VOLCANIC	% of volcanic geology in the watershed derived from a simplified version of Reed & Bush (2001) - Generalized Geologic Map of the Conterminous United States.	%	
SDMNTRY	% of sedimentary geology in the watershed derived from a simplified version of Reed & Bush (2001) - Generalized Geologic Map of the Conterminous United States.	%	
GRANITIC	% of granite geology in the watershed derived from a simplified version of Reed & Bush (2001) - Generalized Geologic Map of the Conterminous United States.	%	
AWCH_AVE	Watershed mean high values of available water capacity of soils	fraction	STATSGO
BDH__AVE	Watershed mean high values of soil bulk density	g/cm3	STATSGO
KFCT_AVE	Watershed mean soil erodibility factor	-	STATSGO
OMH_AVE	Watershed mean high value of soil organic matter content	% by weight	STATSGO
PRMH_AVE	Watershed mean high values of soil permeability	inches/hr	STATSGO

Table 4.3. Continued

WTDH_AVE	Watershed mean high values of seasonally high water table. STATSGO reports the high and low values for "seasonally high water table". This is the watershed mean of high values.	Feet	STATSGO
RDH_AVE	Watershed mean high values of depth to bedrock	cm	STATSGO

Table 4.4. Watershed Attributes Selected from Five Repetition of 10-Fold Cross Validation for LDA Models and Ranked Based on Their Frequency in the Top *pLda* Positions.

K=4	5	6	7	8
ELEV_WS	ELEV_WS	ELEV_WS	SQ_KM	SQ_KM
TMEAN_WS	TMEAN_WS	GRANITIC	precipsep	TMEAN_WS
XWD_WS	ELEV_STD	SQ_KM	TMEAN_WS	precipsep
BDH_AVE	RDH_AVE	TMEAN_WS	precipjul	SD_TMAX_WS
precipjul	XWD_WS	precipsep	deltaE	deltaP
	PRMH_AVE	SD_TMAX_WS	SDMNTRY	precipnov
	StdSlp	SDMNTRY	OMH_AVE	XWD_WS
	SQ_KM	BDH_AVE	ELEV_WS	MAXP_WS
	BDH_AVE	XWD_WS	GRANITIC	GRANITIC
	precipjul	OMH_AVE	precipnov	OMH_AVE

Table 4.5. Watershed Attributes for Optimized CART Models

K=4	5	6	7	8
XWD_WS	XWD_WS	ELEV_WS	ELEV_WS	ELEV_WS
ELEV_WS	ELEV_WS	Rdryness	XWD_WS	XWD_WS
precipnov	Rdryness	TMEAN_WS	TMEAN_WS	SQ_KM
TMEAN_WS	TMEAN_WS	XWD_WS	SQ_KM	MEANP_WS
	TMIN_WS	SQ_KM	precipjan	Precipjan
	BDH_AVE	precipnov	Rdryness	Rdryness
	precipmay	deltaE		
	precipnov			

Table 4.6. Prediction Error and Model Reliability Quantified by the Fraction of Correct Predictions of a Class

K			Number of points in class	Fraction of predictions that were correct: LDA	CART	RF	SVM
4	Classes	1	160	0.701	0.756	0.771	0.759
		2	140	0.803	0.758	0.760	0.746
		3	113	0.504	0.470	0.562	0.553
		4	128	0.821	0.781	0.856	0.833
	Percentile Prediction Error	5		20.00	20.00	16.67	18.33
		50		28.33	30.00	25.00	26.67
		90		36.67	41.67	33.33	35.00
8	Classes	1	94	0.691	0.555	0.721	0.647
		2	97	0.829	0.730	0.789	0.815
		3	32	0.270	0.129	0.523	0.515
		4	94	0.787	0.860	0.778	0.809
		5	14	0.000	0.000	0.022	0.215
		6	101	0.605	0.428	0.627	0.529
		7	60	0.477	0.227	0.468	0.383
		8	49	0.354	0.183	0.403	0.386
	Percentile Prediction Error	5		30.00	36.67	21.67	26.67
		50		40.00	46.67	31.67	36.67
		90		48.33	58.33	40.00	46.67

Table 4.7. Kolmogorov-Smirnov Statistic (D) Based Top 5 Discriminators for Each Pair of Streamflow Regime Classes for K=8.

	2		3		4		5		6		7		8	
1	MINP_WS	0.97	XWD_WS	0.82	TMIN_WS	0.90	TMEAN_WS	0.68	SQ_KM	0.70	TMEAN_WS	0.64	TMEAN_WS	0.79
	TMAX_WS	0.97	LST32AVE	0.81	ELEV_WS	0.86	LST32AVE	0.64	TMEAN_WS	0.51	TMIN_WS	0.62	MINP_WS	0.77
	LST32AVE	0.96	deltaP	0.80	precipnov	0.86	TMIN_WS	0.62	TMAX_WS	0.44	MINP_WS	0.60	TMAX_WS	0.75
	XWD_WS	0.93	TMEAN_WS	0.80	LST32AVE	0.84	AWCH_AVE	0.60	DDEN	0.43	precipmay	0.59	LST32AVE	0.72
	deltaP	0.90	TMAX_WS	0.79	TMEAN_WS	0.84	MeanSlp	0.54	ELEV_STD	0.42	XWD_WS	0.58	TMIN_WS	0.65
2			AWCH_AVE	0.65	MAXWD_WS	0.94	LST32AVE	0.91	LST32AVE	0.86	ELEV_WS	0.80	ELEV_WS	0.72
			ELEV_STD	0.49	XWD_WS	0.94	SDMNTRY	0.81	ELEV_WS	0.84	AWCH_AVE	0.73	SD_TMAX_WS	0.67
			MeanSlp	0.49	Rdryness	0.93	ELEV_WS	0.77	TMAX_WS	0.81	TMAX_WS	0.68	SDMNTRY	0.61
			MAXWD_WS	0.49	precipnov	0.91	XWD_WS	0.75	SD_TMAX_WS	0.79	LST32AVE	0.68	MeanSlp	0.60
			deltaE	0.48	MEANP_WS	0.85	MINP_WS	0.74	MINP_WS	0.78	SD_TMAX_WS	0.66	ELEV_STD	0.59
3					XWD_WS	0.84	LST32AVE	0.75	LST32AVE	0.71	LST32AVE	0.53	deltaE	0.47
					MAXWD_WS	0.84	XWD_WS	0.66	TMAX_WS	0.63	TMAX_WS	0.52	MINP_WS	0.45
					Rdryness	0.82	SDMNTRY	0.63	XWD_WS	0.61	TMEAN_WS	0.49	TMEAN_WS	0.44
					MEANP_WS	0.77	TMEAN_WS	0.63	TMEAN_WS	0.61	MINP_WS	0.49	LST32AVE	0.40
					precipnov	0.76	MINP_WS	0.61	deltaE	0.59	precipmay	0.46	SD_TMAX_WS	0.40
4							LST32AVE	0.68	precipnov	0.82	precipnov	0.78	Rdryness	0.82
							TMIN_WS	0.67	ELEV_WS	0.80	Rdryness	0.77	MAXWD_WS	0.79
							precipnov	0.67	MEANP_WS	0.77	RH_WS	0.77	precipnov	0.77
							RH_WS	0.64	TMIN_WS	0.76	ELEV_WS	0.77	MEANP_WS	0.77
							MeanSlp	0.63	LST32AVE	0.74	MAXWD_WS	0.75	XWD_WS	0.71
5									SQ_KM	0.59	AWCH_AVE	0.58	RDH_AVE	0.55
									AWCH_AVE	0.53	MeanSlp	0.53	SDMNTRY	0.48
									SD_TMAX_WS	0.53	StdSlp	0.53	Sseasonality	0.47
									TMEAN_WS	0.50	VOLCANIC	0.47	Rdryness	0.47
									GRANITIC	0.49	SQ_KM	0.44	MINP_WS	0.45
6											SQ_KM	0.70	TMEAN_WS	0.59
											XWD_WS	0.43	LST32AVE	0.52
											WTDH_AVE	0.41	MINP_WS	0.49
											DDEN	0.41	TMAX_WS	0.46
											RH_WS	0.38	deltaP	0.46

Table 4.7. Continued

7		TMAX_WS	0.42
		AWCH_AVE	0.39
		SHAPE1	0.36
		LST32AVE	0.35
		SQ_KM	0.35

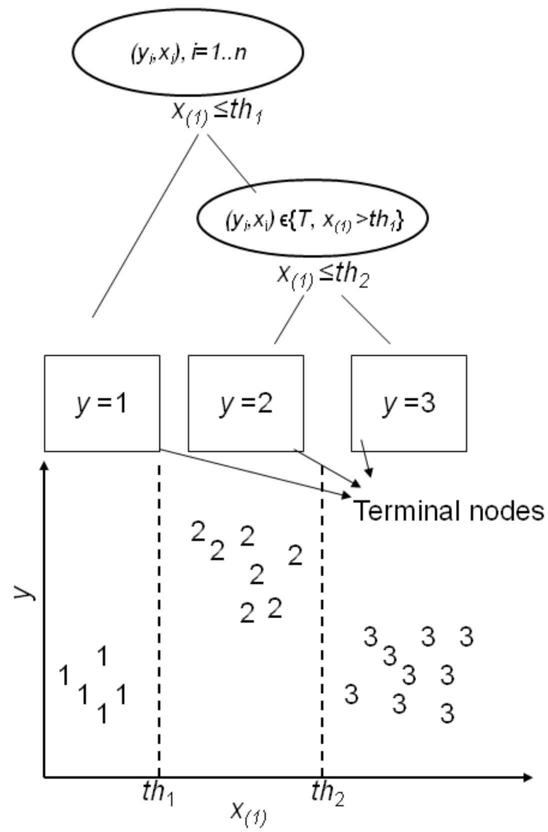


Figure 4.1. An example of CART's tree structured classification approach where a sequence of binary rules operating on the input vector split the data into classes. $x_{(j)}$ is the j^{th} attribute, where j=1..p.

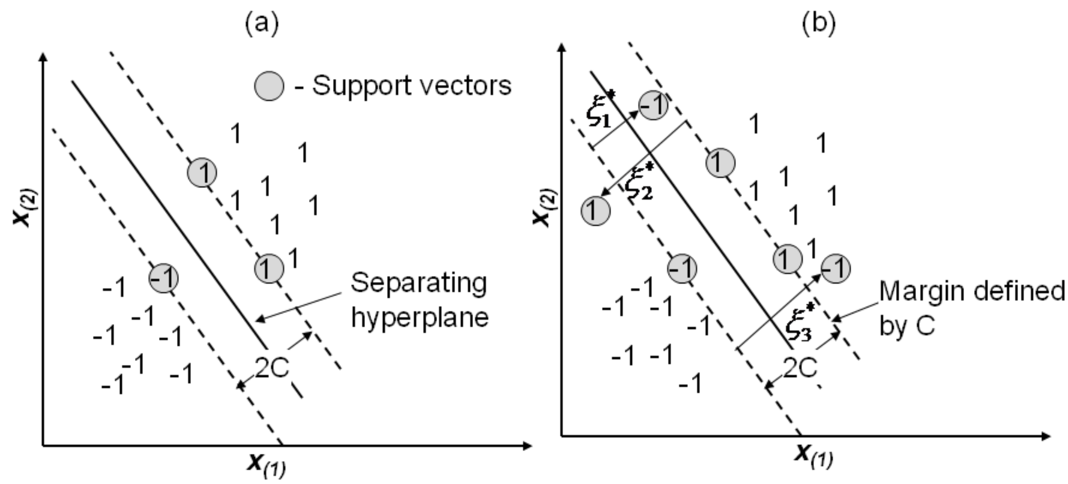


Figure 4.2. An example demonstrating SVM algorithm for two classes ($y \in \{-1,1\}$) and $x \in R^2$. $\xi^*_i = C\,\xi_i$ is the amount by which the points are on the wrong side of the margin. $x_{(j)}$ is the j^{th} attribute, where j=1..p.

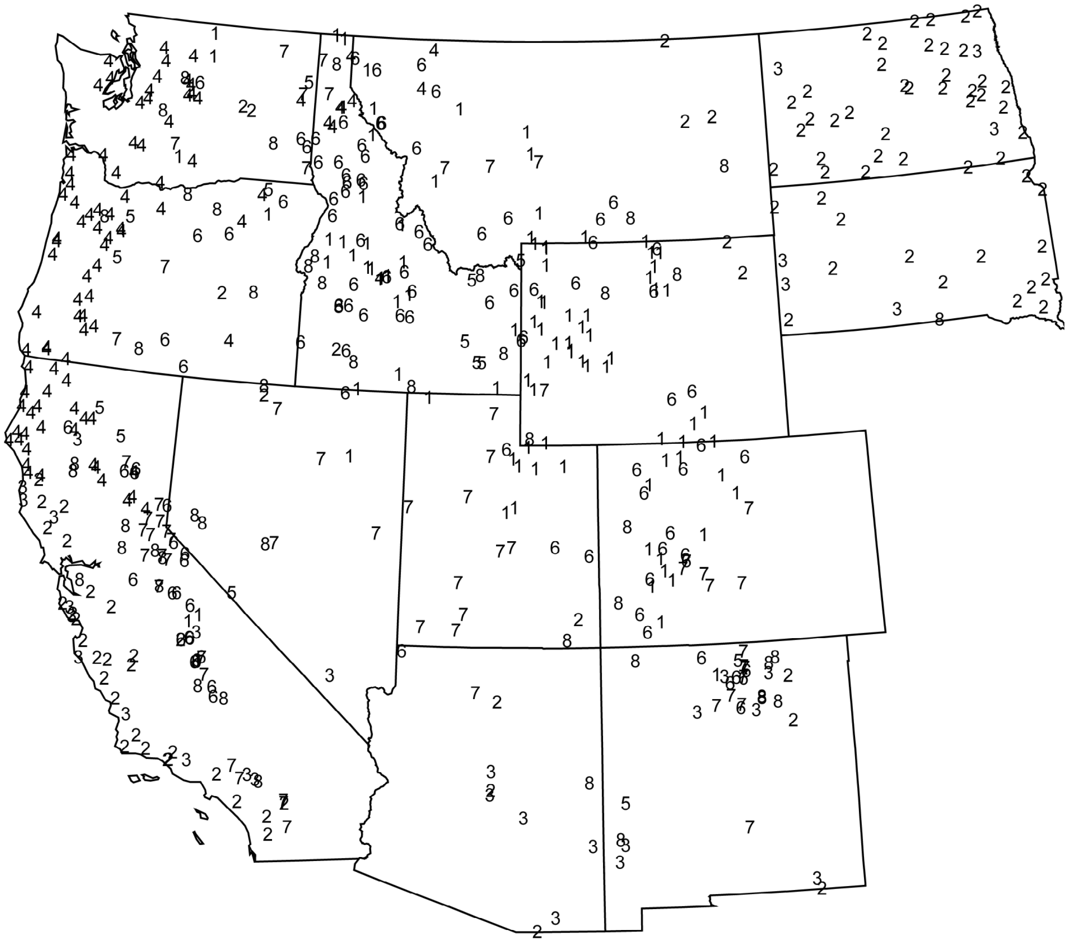


Figure 4.3. Stream gauge sites and their K=8 flow regime class. 1. Seasonal streams 2. Smaller predictable intermittent streams with low baseflow. 3. Mid-size perennial streams with low seasonality. 4. Big streams with low predictability. 5. Baseflow dominated streams. 6. Big seasonal streams. 7. Small unpredictable streams. 8. Small flashy streams.

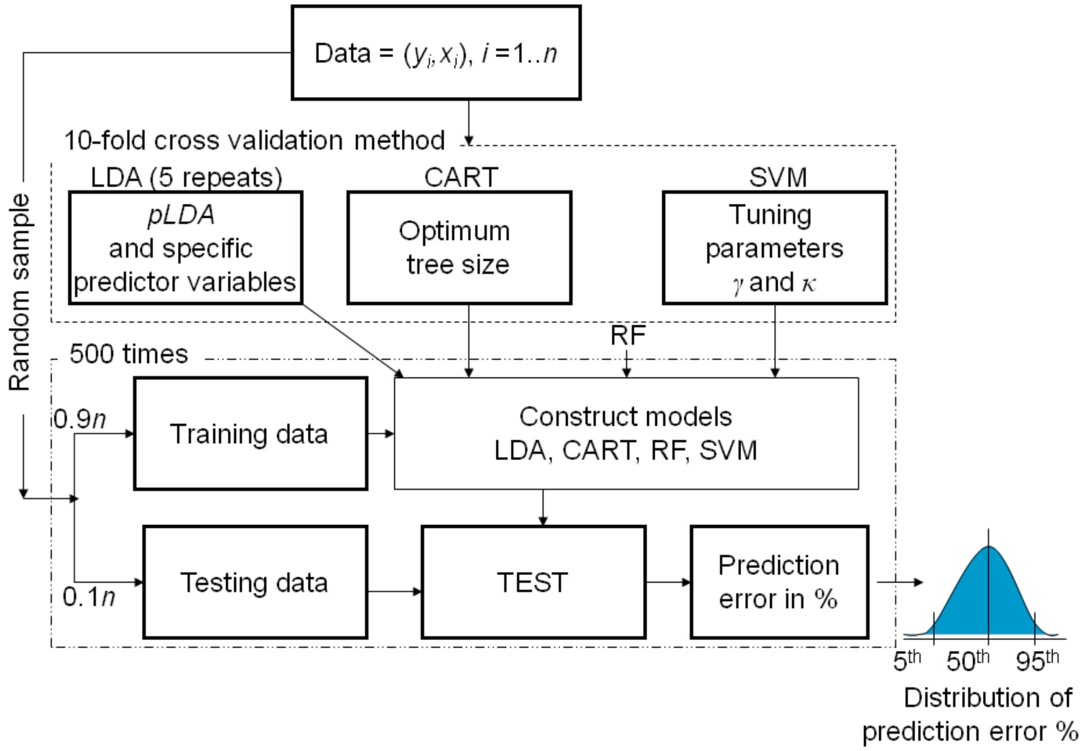


Figure 4.4. Flow chart describing the 10 fold cross validation and the uncertainty estimation from the four methods.

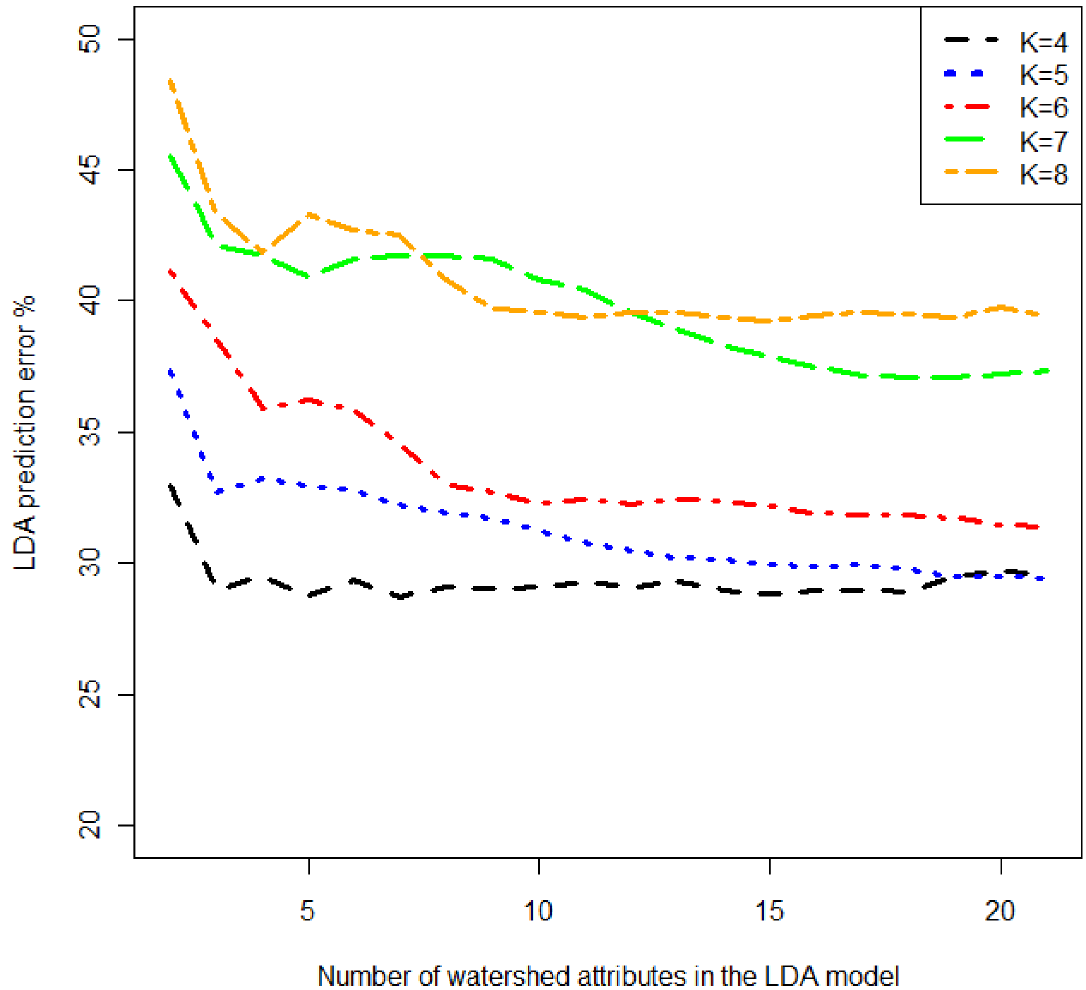


Figure 4.5. Average prediction error from Linear Discriminant Analysis K=4 to K=8 streamflow regime class predictions from five repeats of 10 -fold cross validation.

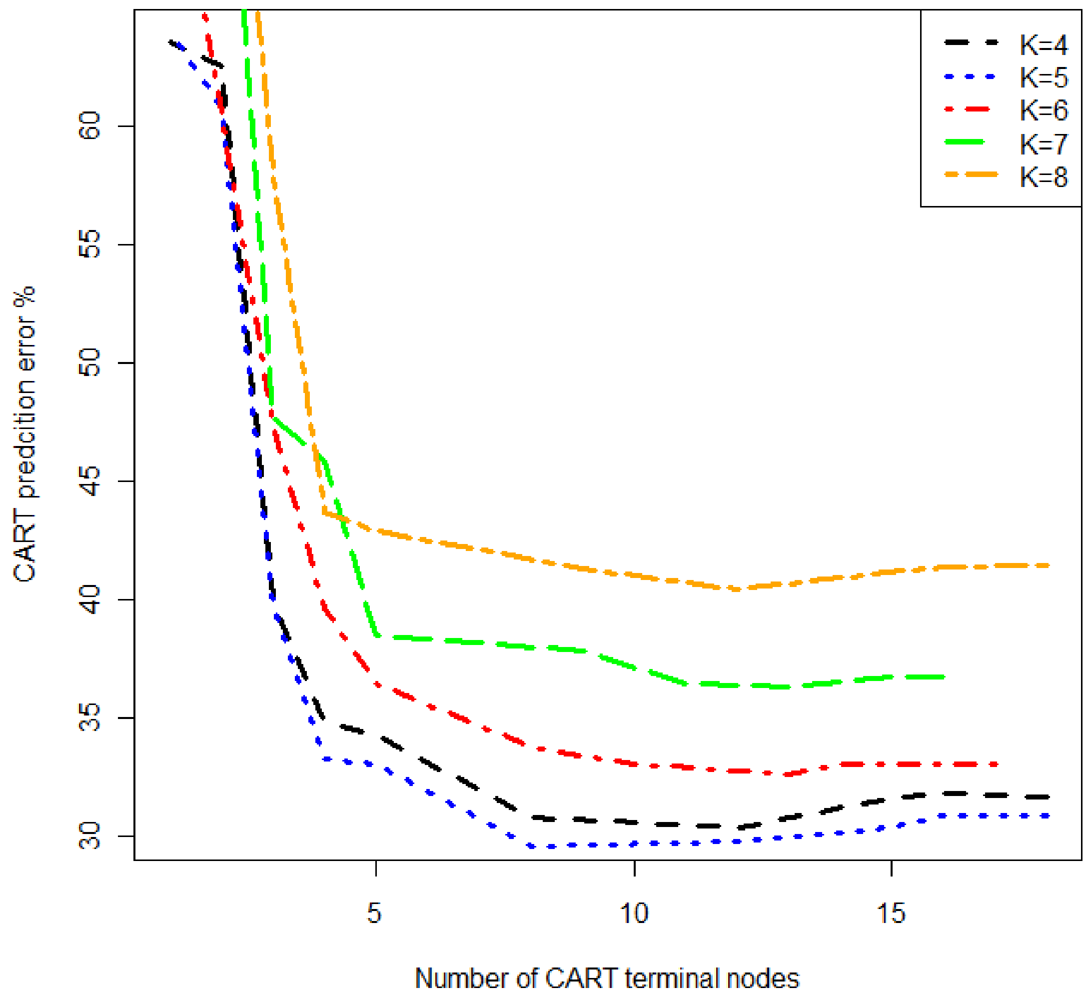


Figure 4.6. Average prediction error from CART K=4 to K=8 streamflow regime class predictions from 10 -fold cross validation.

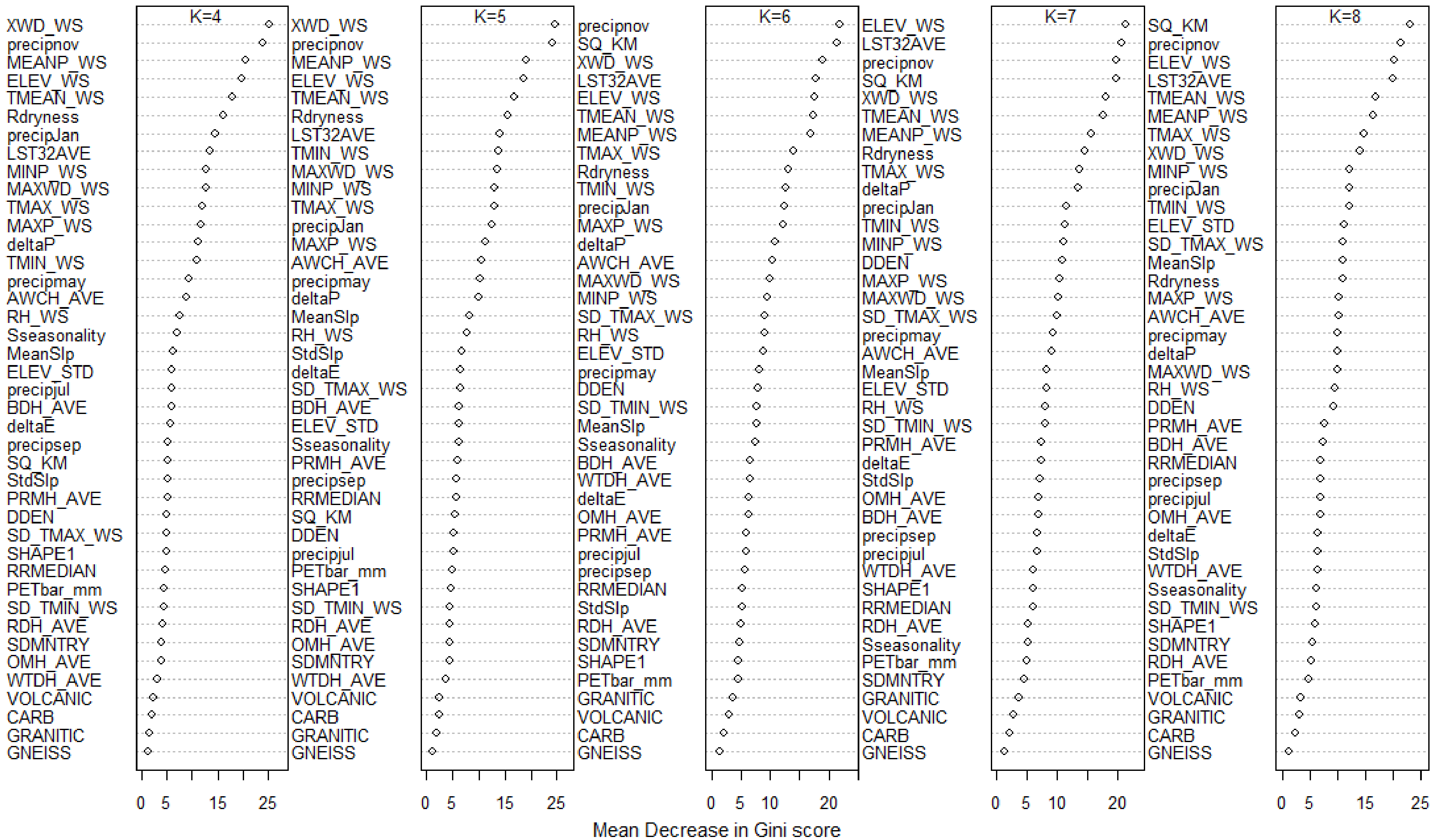


Figure 4.7. Variable importance plot from RF models.

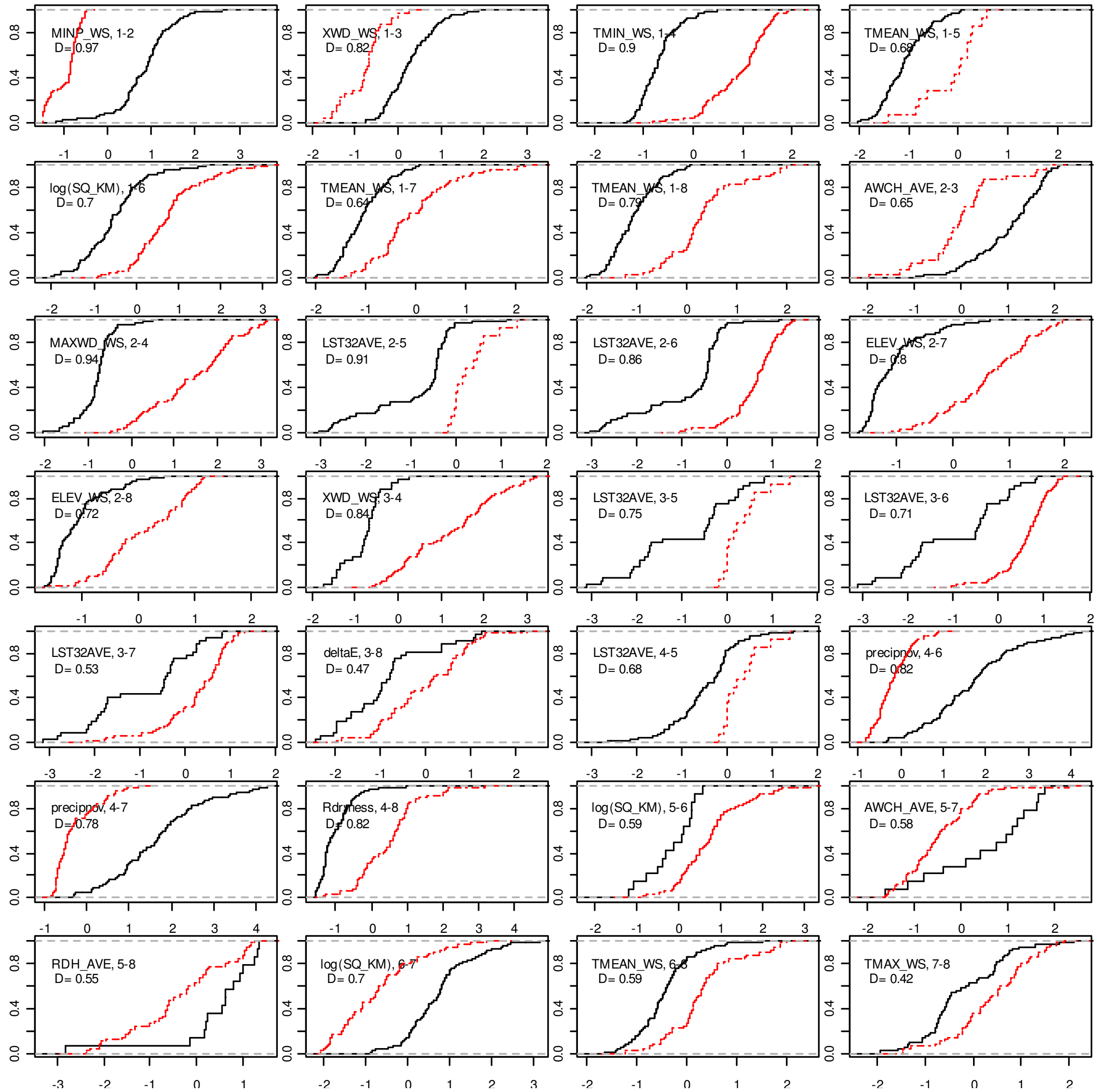


Figure 4.8. Distribution of the best discriminator for each pair of streamflow regime classes. The pair of numbers following the name of the watershed attribute refer to the classes and the second class is plotted in red. Watershed attributes are standardized.

CHAPTER 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This dissertation has considered the characterization of streamflow regimes for the purpose of understanding the relationship between hydrology and stream biota in terms of richness and composition. For a comprehensive treatment of such a purpose, it is essential to be able to predict streamflow regimes at ungauged sites and estimate the uncertainty associated with such predictions. It is also important to be able to efficiently derive watershed attributes used in this prediction. Chapters 2 through 4 present the main results of this dissertation. In this chapter I summarize important conclusions from these chapters and suggest directions and opportunities for future research.

5.1. Summary and Conclusions

The major objectives of this dissertation were to 1) classify streamflow regimes in the western US based on ecologically relevant streamflow variables, 2) quantify how strongly invertebrate taxa richness and composition were associated with variation in flow regime and stream temperatures, 3) develop a GIS based tool to delineate multiple watersheds and derive watershed attributes, 4) predict streamflow regime classes at ungauged sites from watershed attributes using different statistical approaches and assess their performances, 5) quantify the uncertainties associated with each of the prediction methods, and 6) identify key watershed attributes that are important in predicting the streamflow regime classes.

In paper 1 (Chapter 2), we identified and evaluated 12 ecologically relevant streamflow regime variables at 543 gauged watersheds from the USGS Hydro Climatic

Data Network dataset. Principal Component Analysis was used to reduce the dimension from 12 to seven statistically orthogonal streamflow regime factors that represent the following aspects of streamflow regime: 1) zero flow days, 2) magnitude, 3) predictability, 4) flood duration, 5) seasonality, 6) flashiness, and 7) baseflow. These factors were then used in K-means clustering to develop classifications consisting of 4 to 8 streamflow regime classes. We used invertebrate data from 63 gauged sites to test biota-flow regime and biota-temperature relationships. The test was carried out by first grouping the 63 sites into six biotic classes based on their taxonomic composition and taxa richness and then the probability of a biotic class was predicted by Random Forest models with flow characterizations and temperature as predictors.

From this analysis, we found that models predicting taxonomic composition from streamflow regime and temperature performed substantially better than null models. Models just using streamflow regime were still better than null models and the best prediction was achieved when both streamflow and temperature were used as predictors. However, we found only weak to moderate association between streamflow regime and taxa richness. For the data used, we identified baseflow index to be most directly associated with the invertebrate biotic composition.

Classification based approaches are preferred by ecosystem mangers because they are generally easier to communicate and implement. From the conditional probabilities based on streamflow regime classifications, we were able to predict biotic class membership as well as Random Forest models based on continuous variables. This result suggests that a simple biotic class prediction based on classification of streamflow regime groups is possible and can be used to predict taxonomic compositions from the streamflow regime classifications.

One of the main sources of uncertainty in comparing streamflow regimes of two or more watersheds arises from the streamflow record used in the characterization. In an ideal scenario, we would prefer to have natural streamflow records for all watersheds to be from the same period and extending up to the date of biological sampling. Due to imperfect data having varying periods of record, there is a possibility that climate variability influences the streamflow regime characterization. This needs to be considered when watershed management plans based on such characterizations are formulated.

As for many regional-scale studies examining the effects of environmental factors on stream ecosystems [*Poff and Ward*, 1990; *Poff*, 1996; *Baeza Sanz and Garcia del Jalon*, 2005], the basic spatial unit in this dissertation is a watershed. In this dissertation watershed data was obtained by first delineating a watershed boundary and related stream network from DEMs and then applying the geographical boundary of the watershed to spatial data that characterizes climate and soil/geology to obtain statistical measures such as mean, standard deviation etc for each watershed. This was done in Chapter 3. The resulting watershed attribute data was then used in Chapter 4 in models for predicting the streamflow regime classes at ungauged locations.

The emphasis on watershed approaches to address water resources related questions has led to increased demand for watershed delineations and information derived from them. Furthermore, many of these studies are done at regional scales, where quick derivation of stream networks, watershed boundaries, and characteristics at a large number of locations, spread across large areas is desired. Delineating a large number of watersheds spread across large regions is a challenge because, firstly, the coordinate of the outlet may not coincide with the digital representation of the stream. When only few watersheds are delineated, this is not a major problem, but when hundreds of watersheds are being delineated, this can become cumbersome. Secondly, delineating watersheds across broad geographic areas requires grid datasets that are may tax the memory of available computers.

Chapter 3 presents a Multi-watershed delineation (MWD) tool developed using ArcGIS and TauDEM functionality to enhance the capability for delineating multiple watersheds over large areas. The MWD tool is designed to quickly create watershed boundaries and derive other geomorphological attributes of the watershed for a large number of watersheds across broad geographical regions.

Delineating watersheds requires various DEM derived grids that represent the hydrologic characteristics of the landscape. They are: 1) pit filled elevation grid, 2) flow direction grid, 3) flow accumulation grid, and 4) digital stream grid (to identify the correctness of the outlet position). The creation of these grids is resource intensive but once created can be reused for delineating watersheds with the same region [*Djokic*, 2000]. The MWD tool uses this idea and preprocesses all the required grids before the actual delineation.

To solve the problem of the imprecise position of the outlet, the MWD tool moves an outlet that is not on a digitally delineated stream downhill by following the flow direction grid until it comes into contact with the stream. The watershed is then delineated for this new position of the outlet. The MWD tool solves the problem of memory required for large grids by automatically clipping large hydrologic grids to the size of regional watersheds called Medium Hydrologic Units (MHU) that are a more manageable size. The MHU may be created from regional scale USGS hydrologic units such as 8 digit HUCs. The MWD tool uses the MHU to clip only the area required for delineation from the necessary hydrologic grids. During the clipping process, a buffer is added to the MHU to ensure that the polygon captures the hydrologic boundary present in the landscape. Each site within a MHU is then delineated by using the common routines available within TauDEM program [*Tarboton and Ames*, 2001].

The MWD tool comes in two versions; 1) Graphical User Interface (GUI) standalone program and 2) command line executable. The GUI version is user friendly but can handle only one grid set at a time. In most of our cases, this grid set encompassed only a few 8 digit HUCs. The command line version can be used in a batch process to delineate multiple grid sets. In one of our runs, we ran the command line version in a batch process for almost two days to delineate 441 watersheds and their geomorphological attributes (drainage area: 15 to 12416 km^2). We used a DEM with approximately 30 m grid cell resolution and the resulting watersheds were spread across 13 states of the western US. The watershed boundaries created by this tool were later used to estimate different statistics from other spatial datasets (climate, soils and geology). These watershed attributes were then used in statistical models for predicting streamflow regime classes.

In paper 3 (Chapter 4), we used four popular statistical classification models; 1) Linear discriminant methods, LDA [see *Hastie et al.*, 2001], 2) Classification and Regression Trees, CART [*Breiman et al.*, 1984], 3) Random Forests, RF [*Breiman*, 2001], and 4) Support Vector Machines, SVM [see *Vapnik*, 1998] to predict the streamflow regime classes of chapter 2 at ungauged sites from watershed attributes. We used 541 of the sites used in paper 1 (Chapter 2) for this study. Two of the sites used in paper 1 were excluded as they were considered as outliers in one or more watershed attributes. Excluded watersheds had elevation related statistics that were high compared to the others and their removal resulted in better Box-Cox transformations for those attributes and satisfied the LDA assumptions better. These watersheds we suppose would not have affected other methods, as they are known to handle outliers better.

We used a 10 fold cross validation method with LDA, CART and SVM to optimize relevant model parameters. In the case of LDA, the 10 fold cross validation method was repeated 5 times to get a stable estimate of the number of parameters required to minimize the prediction error. The 10 fold cross validation step was also used in selecting the specific watershed attributes that were most important for LDA models. The 10 fold cross validation method was used to optimize the tree size in CART and, the cost and slack parameters in SVM. RF effectively represents a cross-validation extension of CART and does not require any additional cross-validation for optimization or selection of watershed attributes.

Optimized models were then used in a bootstrapping method for assessing the uncertainties in model predictions. For each of the 500 runs in the bootstrapping step, we randomly divided the data into a training set (481 data points) and testing set (60 data points) and developed each of the four models using the training set, then predicted the streamflow regime classes of the test dataset. The distribution of the classification error from 500 runs was used to compare the performance of the models. The average contingency table between the predicted and the actual classes from all 500 runs was used to estimate the conditional probability that the predicted class was correct. This was used as a measure of reliability of a model to predict a specific class.

For a given set of streamflow regime classes, K, the empirical distributions of watershed attributes were examined. The power of a watershed attribute to distinguish between two classes was quantified by the separation between distributions using the Kolmogorov-Smirnov measure, D.

We found that classes 1, 2, and 4 were relatively well predicted, class 6 was moderately well predicted and classes 3, 5, 7, and 8 were poorly predicted. This behavior consistent across the different models suggests that some classes are poorly predicted due to the absence of watershed attributes that could discern the hydrological differences between the classes. For example, classes 1, 2, and 4 have relatively high measure of D for the attributes that are most discriminating between them, while class 5 made up of baseflow dominated streams has relatively small values of D for its most discriminating attributes with other classes. Another aspect contributing to the poor prediction of class 5 is its size. In the 8 class categorization (K=8), class 5 had only 14 sites and this makes it difficult for the statistical models to predict it. Class 7 and class 8 were least separated from each other (smallest D value among the pair of classes). The main difference between these two classes is the flashiness represented by the number of reversals on the daily streamflow. This hydrologic characteristic is apparently not related to the watershed attributes used in this study and hence class 7 and 8 were often misclassified as one another. Class 3 had similar problem and was not well distinguished from classes 7 or 8.

The median error for LDA, CART, RF and SVM classifications into K=4 to 8 streamflow regime classes from watershed attributes ranged between 28-40, 30-47, 25-31, and 27-37%, respectively. This suggests that predictions of class for ungauged basins is possible with about 70% accuracy, and that the RF model was slightly more reliable than the other models. Scrutiny of the results revealed that RF was the most reliable in predicting four classes (1, 3, 6, and 8) while LDA was reliable for two (classes 2 and 7). CART and SVM were most reliable for one class each, classes 4 and 5, respectively.

When the underlying assumptions of LDA are met and a robust variable selection is employed, LDA generally can perform very well. But LDA is not very well suited for modeling non-linear relationships and the implementation was the most tedious among the methods used in this study. CART on the other hand can handle non-linear relationships very well and the interpretation is most intuitive, but it is very sensitive to the data used to develop the model. Hence CART had a somewhat larger prediction error. Distributions of prediction error support the finding based on the median that RF performs slightly better than the other models with a better overall accuracy. Even though with RF it is not as easy as CART to understand the relationship between streamflow regime classes and watershed attributes, some degree of interpretability is offered by the variable importance plots. Also, the fact that RF required the least amount of effort in terms of its implementation makes it an attractive method. Nevertheless, there still appears to be merit in using multiple methods because RF was not always the most reliable method when class specific prediction was considered.

SVM is still subject to ongoing research with relatively few studies in hydrology and ecology. SVM as implemented here performed almost as well as RF and was the only model that could predict class 5 with any reliability albeit low. It is possible that with good variable selection and better optimization methods, SVM can predict as well as or better than RF. However, with SVM it is very difficult to interpret the relationship between streamflow regime classes and the watershed attributes because of the complicated kernel transformations that occur within the SVM model.

5.2. Recommendations

From a broader perspective this dissertation has attempted to develop the foundation needed in regional stream ecological studies to understand the effects of environmental components on the structure and function of stream ecosystems. As a result there are a number of avenues that can be foreseen for carrying ahead this research.

Foremost among the needs to carry on this research would be to procure additional biological data at gauged sites. This would enable a better research design for studies relating ecology and streamflow regimes as it would avoid the uncertainties involved in predicting the streamflow regime.

The paradigm of natural streamflow regime [*Richter et al.*, 1996; *Poff et al.*, 1997] posits that the magnitude, frequency, duration, timing, and rate-of-change of streamflow are important components of flow regime that directly influence the ecological processes and patterns [*Poff et al.*, 2006]. However there is no uniformity in the use of specific streamflow regime variables to represent these components and the choice of streamflow variables affects the characterization and its ability to partition the naturally occurring biota. We argue here that the choice should be based on the ecological questions being addressed and may require iteration on variable selection to identify the best set of streamflow regime variables.

This dissertation provides a framework for testing different streamflow regime variables for their ability to describe biotic variation. For example, in our analysis we found that the average number of reversals had a weak relationship with both taxa richness and biotic class. For the next iteration, number of reversals can be replaced or excluded to refine the streamflow regime characterization before testing against taxonomic richness and composition. The process of iteratively selecting the variables can be repeated to arrive at a streamflow regime characterization with maximum power for describing biotic variation. The variables identified in such a process may also provide information on the specific streamflow variables that are most important for the biota.

Some of the aspects of streamflow regime that we are interested in examining further within the above framework are: (1) timing variables – Colwell's seasonality that was used as a timing variable used is only representative of the average timing of the seasonal cycle and it might be beneficial to use specific timing variables for peak and low flows, (2) low flow duration variables - similar to the flood duration, and (3) scaled high flow variables – this is to test variables that quantify the relative magnitude of high flows similar to the way base flow index quantifies the relative magnitude of low flows. These variables according to us could possibly increase the ecological relevance of the streamflow regime characterization.

In the previous section the possibility of climatic variations affecting the streamflow regime characterization was briefly mentioned. One approach to address this issue is to use hydrologic models [e.g. *Wigmosta et al.*, 1994; *Beven et al.*, 1995] to generate streamflow regimes for observed and forecast climate scenarios. Though challenging to implement at large scales for multiple sites, detailed hydrologic models are increasingly gaining prominence for studying the effects of climatic variations on regional hydrology and can also provide opportunities for studying the effects of climate change on stream ecology.

The MWD tool described in Chapter 3 is able to delineate multiple watersheds spread across large geographical regions based on the assumption that the watershed being delineated is contained within the MHU which is generally an 8 digit HUC. To meet this assumption, when delineating watersheds larger than an 8 digit HUC we have to use a 4 digit HUCs as MHUs and 2 digit HUC as LHU (larger regional watersheds). Moving from 4 digit HUCs to 2 digit HUCs also means a coarser DEM (commonly 90m DEM) for delineating watersheds and its attributes. The use of coarser DEM is not always desirable and may need a better solution. The underlying watershed delineation functions of the MWD tool are directly from TauDEM and the TauDEM capabilities have recently been enhanced to handle big DEMs via parallel programming [*Wallis et al.*, 2009a, 2009b]. Using the improved functionalities of TauDEM within the MWD concept could greatly improve the ability to handle very large DEMs without compromising the DEM resolution.

In Chapter 4, four popular statistical methods were used to predict only classifications, the discrete characterization, even though seven continuous factors were also used to relate the streamflow regime to taxa richness and composition. Statistical models can be developed for predicting continuous streamflow regime variables to complete the overall characterization of the streamflow regime at ungauged sites. Three of the four methods (CART, RF and SVM) used in this study can be implemented for continuous variables without any major modifications to complete this task.

The SVM as mentioned earlier is a subject of active research within the machine learning and data mining community. The R library used to implement the SVM has a weak grid-search based optimization method to select model parameters and does not have any type of variable selection procedure. The SVM if updated with a better optimization and variable selection method can potentially improve the capability of predicting steamflow regime characterizations.

The pattern of prediction across classes was consistent across the different models. For example, class 5 was poorly predicted, while class 1, 2 and 4 were relatively better predicted by all the methods. This pattern implies that some classes are poorly predicted mainly due to the absence of good watershed attributes that can discern the hydrological differences among the classes and not the method of prediction. This view is further supported by the range of model capabilities used in this study. It is imperative to identifying better discriminators in order to increase the prediction performances of the model. Better discriminators for class 5 (base flow dominated streams), class 7 (small unpredictable streams) and class 8 (small flashy streams) are needed to increase the correct prediction percentage.

The key objective of paper 3 (Chapter 4) was to predict the streamflow regime classes at ungauged sites. Chapter 4 developed these capabilities using only gauged sites. The natural progression of this research then would be to predict the streamflow regime at the actual ungauged sites where we have the biology data and test the associations between biology and streamflow regime at these ungauged sites, similar to the methods in

Chapter 2. Such a study should assess the possible implications of uncertainties emanating from the streamflow regime prediction models.

This dissertation lays the ground work necessary to evaluate the ecological health of streams as it relates to hydrology. It provides the tools and knowledge base for assessing the overall affects of environmental factors on the structure and function of stream ecosystems. The streamflow characterizations created in this dissertation make up one piece of the puzzle needed to answer the questions asked in the original proposal funded by the EPA:

(1) can sequential application of classifications based on different types of watershed attributes provide insight regarding the stressors affecting aquatic ecosystem?

(2) can a watershed classification derived from a multivariate analysis of the joint variation in different types of watershed attributes achieve greater effectiveness in partitioning biotic variation among watersheds than classifications based on single factors?

Carrying forward this research to answer the above questions is vital to meet the goals of the Clean Water Act mentioned in Chapter 1 and is of great interest to the watershed science community involved in developing sustainable water management policies and practices.

References

Baeza Sanz, D., and D. Garcia del Jalon (2005), Characterisation of streamflow regimes in central Spain, based on relevant hydrobiological parameters, *J. Hydrol.*, *310*(1-4), 266-279.

Beven, K., R. Lamb, P. Quinn, R. Romanowicz, and J. Freer (1995), TOPMODEL, in *Computer models of watershed hydrology*, edited by V. P. Singh, pp. 627-668, Water Resources Publications, Highlands Ranch, Colo.

Breiman, L. (2001), Random forests, *Machine Learning*, *45*(1), 5-32.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (1984), *Classification and Regression Trees*, Wadsworth International Group, Belmont, Calif.

Djokic, D. (2000), Dem preprocessing for efficient watershed delineation, in *Hydrologic and hydraulic engineering: Modeling and support with geographic information systems*, edited by D. Maidment and D. Djokic, Environmental Systems Research Institute, Inc., Redlands, Calif.

Hastie, T., R. Tibshirani, and J. Friedman (2001), *The Elements of Statistical Learning*, Springer-Verlag, New York.

Poff, N. L. (1996), A hydrogeography of unregulated streams in the united states and an examination of scale-dependence in some hydrological descriptors, *Freshwater Biology*, *36*(1), 71-91.

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Larr, K. L. Prestegaard, B. D. Richter, R. E. Sparks, and J. C. Stromberg (1997), The natural flow regime: A paradigm for conservation and restoration of river ecosystems, *BioScience*, *47*, 769-784.

Poff, N. L., B. P. Bledsoe, and C. O. Cuhaciyan (2006), Hydrologic variation with land use across the contiguous united states: Geomorphic and ecological consequences for stream ecosystems, *Geomorphology*, *79*(3-4), 264-285.

Poff, N. L., and J. V. Ward (1990), Physical habitat template of lotic systems: Recovery in the context of historical pattern of spatiotemporal heterogeneity, *Environ. Manage.*, *14*(5), 629-645.

Richter, B. D., J. V. Baumgartner, J. Powell, and D. P. Braun (1996), A method for assessing hydrologic alteration within ecosystems, *Conservation Biology*, *10*(4), 1163-1174.

Tarboton, D. G., and D. P. Ames (2001), Advances in the mapping of flow networks from digital elevation data, paper presented at World Water and Environmental Resources Congress, ASCE, Orlando, Fla., May 20-24.

Vapnik, V. (1998), *Statistical Learning Theory*, John Wiley Sons, Inc., New York.

Wallis, C., R. Wallace, D. G. Tarboton, D. W. Watson, K. A. T. Schreuders, and T. K. Tesfa (2009a), Hydrologic terrain processing using parallel computing, in *The*

18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, edited by R. S. Anderssen, et al., pp. 2540-2545, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, Cairns, Australia.

Wallis, C., D. Watson, D. G. Tarboton, and R. Wallace (2009b), Parallel flow-direction and contributing area calculation for hydrology analysis in digital elevation models, in *PDPTA'09, The 2009 International Conference on Parallel and Distributed Processing Techniques and Applications*, Las Vegas, Nev., Jul 13-16.

Wigmosta, M. S., L. W. Vail, and D. P. Lettenmaier (1994), A distributed hydrology-vegetation model for complex terrain, *Water Resour. Res.*, *30*(6), 1665-1679.

APPENDICES

Appendix A

This appendix provides information about the USGS streamflow gauge sites and the period of record used to estimate the 12 streamflow regime variables in Chapter 2. It also provides a map of these sites along with their watershed boundaries. The sites in the map (Figure A.1) are labeled to refer the first column of Table A.1.

Table A.1. Streamflow Site Information. The Index Column is Used to Label the Sites in Figure A.1

Index	Site name	USGS code	Drainage Area, Sq. km	NAWQA Site	Periods used
1	HIDDEN ISLAND COULEE NR HANSBO	5098700	97		1962-1988
2	CYPRESS CREEK NR SARLES, ND	5098800	182		1962-1988
3	PEMBINA RIVER AT WALHALLA, ND	5099600	8576	Yes	1940-1988
4	PEMBINA RIVER AT NECHE, ND	5100000	8730		1932-1988
5	WINTERING RIVER NR KARLSRUHE,	5120500	1805		1938-1988
6	WILLOW CREEK NR WILLOW CITY, N	5123400	2970		1957-1988
7	BOUNDARY CREEK NR LANDA, ND	5123900	589		1958-1981
8	VERMEJO RIVER NEAR DAWSON, NM	7203000	771		1916-1917,1920-1920,1928-1988
9	PONIL CREEK NEAR CIMARRON, N.	7207500	438		1917-1918,1920-1924,1951-1988
10	RAYADO CREEK AT SAUBLE RANCH N	7208500	166		1917-1918,1931-1988
11	CANADIAN R NR TAYLOR SPRINGS,	7211500	7296		1940-1958,1965-1988
12	MORA RIVER NEAR GOLONDRINAS N.	7216500	684		1916-1920,1927-1986
13	COYOTE CREEK NEAR GOLONDRINAS,	7218000	550		1930-1988
14	MORA RIVER NR SHOEMAKER N MEX.	7221000	2811		1920-1924,1928-1988
15	CONCHAS RIVER AT VARIADERO, N.	7222500	1275		1937-1988
16	SWIFTCURRENT CREEK AT MANY GLA	5014500	80		1918-1919,1959-1988
17	WILD RICE RIVER NR RUTLAND, ND	5051600	1398		1960-1969,1971-1982

Table A.1. Continued

18	WILD RICE RIVER NR ABERCROMBIE	5053000	5325		1933-1988
19	SHEYENNE RIVER ABOVE HARVEY, N	5054500	1085		1956-1988
20	SHEYENNE RIVER NR WARWICK, ND	5056000	5299	Yes	1950-1988
21	MAUVAIS COULEE NR CANDO, ND	5056100	991		1957-1982
22	EDMORE COULEE NR EDMORE, ND	5056200	978		1958-1982
23	SHEYENNE RIVER NR COOPERSTOWN,	5057000	16563		1946-1988
24	BALDHILL CREEK NR DAZEY, ND	5057200	1769		1957-1988
25	MAPLE RIVER NR HOPE, ND	5059600	52		1965-1982
26	MAPLE RIVER NR ENDERLIN, ND	5059700	2158	Yes	1957-1988
27	RUSH RIVER AT AMENIA, ND	5060500	297		1947-1988
28	BEAVER CREEK NR FINLEY, ND	5064900	410		1965-1986
29	GOOSE RIVER AT HILLSBORO, ND	5066500	3080	Yes	1936-1988
30	MIDDLE BRANCH FOREST RIVER NR	5083600	122		1961-1988
31	FOREST RIVER NR FORDVILLE, ND	5084000	1167	Yes	1941-1988
32	LITTLE MINNESOTA RIVER NEAR PE	5290000	1121		1940-1981
33	WHETSTONE RIVER NEAR BIG STONE	5291000	1019		1932-1986
34	BEAR CREEK AB RESERVOIR NR IRW	13032000	197		1954-1971
35	BLACKFOOT RIVER AB RESERVOIR N	13063000	896		1915-1915,1968-1982
36	PORTNEUF RIVER AT TOPAZ ID	13073000	1459	Yes	1914-1915,1920-1988
37	MARSH CREEK NR MCCAMMON ID	13075000	904		1955-1988
38	GEORGE CREEK NEAR YOST UTAH	13077700	20		1960-1988
39	GOOSE CREEK AB TRAPPER CREEK N	13082500	1620		1912-1916,1920-1988
40	ROCK CREEK NR ROCK CREEK ID	13092000	205		1911-1912,1945-1974
41	BEAVER CREEK AT SPENCER ID	13113000	307		1941-1941,1969-1981,1987-1988
42	N FK BIG LOST RIVER AT WILD HO	13120000	292		1945-1988
43	BIG LOST RIVER AT HOWELL RANCH	13120500	1152	Yes	1905-1905,1949-1988
44	BIG WOOD RIVER NR KETCHUM ID	13135500	351		1949-1971

Table A.1. Continued

45	COMBINATION BIG WOOD R/SLOUGH	13139510	1638		1916-1987
46	LITTLE WOOD RIVER AB HIGH FIVE	13147900	635	Yes	1959-1974,1980-1981,1983-1988
47	BRUNEAU RIVER AT ROWLAND NV	13161500	978		1914-1918,1967-1988
48	EF JARBIDGE RIVER NR THREE CRE	13162500	217		1929-1932,1954-1971
49	EF BRUNEAU RIVER NR HOT SPRING	13167500	1587		1911-1914,1950-1971
50	BRUNEAU RIVER NR HOT SPRING ID	13168500	6733		1910-1914,1944-1988
51	BIG JACKS CREEK NR BRUNEAU ID	13169500	648		1940-1949,1966-1988
52	JORDON CREEK AB LONE TREE CR N	13178000	1126		1946-1952,1956-1971
53	BOISE RIVER NR TWIN SPRINGS ID	13185000	2125		1912-1988
54	SF BOISE RIVER NR FEATHERVILLE	13186000	1626		1946-1988
55	MORES CREEK AB ROBIE CREEK NR	13200000	1021		1951-1988
56	ROBIE CREEK NR ARROWROCK DAM I	13200500	40		1951-1971
57	MALHEUR RIVER NEAR DREWSEY,ORE	13214000	2330		1927-1988
58	SF PAYETTE RIVER AT LOWMAN ID	13235000	1167		1942-1988
59	LAKE FORK PAYETTE RIVER AB JUM	13240000	125		1946-1988
60	BIG WILLOW CREEK NR EMMETT ID	13250600	121		1963-1982
61	WEISER RIVER AT TAMARACK ID	13251500	93		1937-1971,1975-1975
62	PINE CREEK NR CAMBRIDGE ID	13260000	138		1939-1962
63	LITTLE WEISER RIVER NR INDIAN	13261000	210		1925-1927,1939-1971
64	MANN CREEK NR WEISER ID	13267000	143		1912-1913,1938-1961
65	VALLEY CREEK AT STANLEY ID	13295000	376		1912-1913,1922-1971
66	SALMON RIVER BL VALLEY CREEK A	13295500	1283		1926-1960
67	YANKEE FORK SALMON RIVER NR CL	13296000	499		1922-1948
68	SALMON RIVER BL YANKEE FORK NR	13296500	2053		1922-1971,1977-1988
69	SALMON RIVER NR CHALLIS ID	13298500	4608		1929-1971
70	CHALLIS CREEK NR CHALLIS ID	13299000	218		1944-1962
71	SALMON RIVER AT SALMON ID	13302500	9626		1913-1916,1920-1988

Table A.1. Continued

72	LEMHI RIVER NR LEMHI ID	13305000	2291	1956-1963,1968-1988
73	PANTHER CREEK NR SHOUP ID	13306500	1354	1945-1977
74	SALMON RIVER NR SHOUP ID	13307000	16051	1945-1981
75	MF SALMON RIVER NR CAPEHORN ID	13308500	353	1929-1971
76	BEAR VALLEY CREEK NR CAPE HORN	13309000	468	1929-1960
77	S FK SALMON RIVER NR KNOX ID	13310500	236	1929-1960
78	EF OF SF SALMON RIVER AT STIBN	13311000	50	1929-1941,1983-1988
79	JOHNSON CREEK AT YELLOW PINE I	13313000	543	1929-1988
80	LITTLE SALMON RIVER AT RIGGINS	13316500	1475	1952-1954,1957-1988
81	SALMON RIVER AT WHITE BIRD ID	13317000	34688	1911-1917,1920-1988
82	GRANDE RONDE R AT LA GRANDE, O	13319000	1736	1904-1912,1914-1915,1919-1923, 1926-1988
83	MINAM RIVER AT MINAM,OREG.	13331500	614	1913-1913,1966-1988
84	ASOTIN CR BLW KEARNEY GULCH NR	13334700	435	1960-1982
85	SELWAY RIVER NR LOWELL ID	13336500	4890	1930-1988
86	LOCHSA RIVER NR LOWELL ID	13337000	3021	1911-1912,1930-1988
87	S FK CLEARWATER RIVER NR ELK C	13337500	668	1945-1974
88	S FK CLEARWATER RIVER NR GRANG	13338000	2214	1912-1916,1924-1963
89	S FK CLEARWATER RIVER AT STITE	13338500	2944	1965-1988
90	CLEARWATER RIVER AT KAMIAH ID	13339000	12416	1911-1965
91	CLEARWATER RIVER AT OROFINO ID	13340000	14285	1931-1938,1965-1988
92	N FK CLEARWATER RIVER AT BUNGA	13340500	2550	1945-1969
93	N FK CLEARWATER RIVER NR CANYO	13340600	3482	1968-1988
94	CLEARWATER RIVER AT SPALDING,	13342500	24499	1911-1913,1926-1971
95	PALOUSE RIVER NR POTLATCH, ID.	13345000	812	1916-1919,1968-1988
96	SOUTH FORK PALOUSE RIVER AT PU	13348000	338	1935-1942,1961-1981
97	RUBY RIVER ABOVE RESERVOIR NEA	6019500	1377	1939-1988

Table A.1. Continued

98	MADISON RIVER NEAR WEST YELLOW	6037500	1075		1914-1917,1919-1921,1923-1973,1984-1986
99	GALLATIN RIVER NEAR GALLATIN G	6043500	2112		1890-1894,1931-1969,1972-1981,1985-1988
100	PRICKLY PEAR CREEK NEAR CLANCY	6061500	492		1909-1916,1922-1933,1946-1953,1955-1969,1979-1988
101	SHEEP CREEK NEAR WHITE SULPHUR	6077000	110		1942-1972
102	NORTH FORK SUN RIVER NEAR AUGU	6078500	660		1912-1912,1946-1968
103	BELT CREEK NEAR MONARCH, MT.	6090500	942		1952-1982
104	NORTH FORK MUSSELSHELL RIVER N	6115500	80		1941-1976
105	BIG DRY CREEK NEAR VAN NORMAN,	6131000	6538		1940-1947,1950-1968,1971-1972,1974-1988
106	ROCK CREEK BELOW HORSE CREEK,	6169500	840		1979-1988
107	REDWATER RIVER AT CIRCLE MT	6177500	1400		1932-1932,1936-1936,1938-1971,1975-1988
108	YELLOWSTONE RIVER AT YELLOWSTO	6186500	2575		1927-1982,1984-1986
109	TOWER CREEK AT TOWER FALLS,YNP	6187500	129		1924-1943
110	LAMAR RIVER NR TOWER FALLS RAN	6188000	1690		1924-1969
111	GARDNER RIVER NEAR MAMMOTH YNP	6191000	517		1939-1972,1985-1988
112	YELLOWSTONE RIVER AT CORWIN SP	6191500	6715	Yes	1890-1893,1911-1988
113	YELLOWSTONE RIVER NEAR LIVINGS	6192500	9091	Yes	1898-1905,1929-1932,1938-1988
114	CLARKS FORK YELLOWSTONE RIVER	6207500	2954		1922-1988
115	ROCK CREEK NEAR RED LODGE, MT.	6209500	317		1935-1982,1986-1986
116	YELLOWSTONE RIVER AT BILLINGS	6214500	30195	Yes	1929-1988
117	WIND RIVER NEAR DUBOIS, WYO.	6218500	594		1946-1988
118	DINWOODY CREEK ABOVE LAKES, NE	6221400	226		1958-1978
119	CROW C NR TIPPERARY WYO	6222700	77		1963-1988
120	BULL LAKE C AB BULL LAKE WYO	6224000	479		1942-1953,1967-1988

Table A.1. Continued

121	LITTLE POPO AGIE RIVER NEAR LA	6233000	320		1947-1971
122	GOOSEBERRY CREEK AT DICKIE, WY	6265800	243		1958-1978
123	NOWOOD R NR TENSLEEP, WY	6270000	2056		1939-1943,1951-1955,1973-1988
124	TENSLEEP CREEK NEAR TENSLEEP,	6271000	632		1911-1912,1915-1924,1944-1971
125	MEDICINE LODGE CREEK NEAR HYAT	6273000	222		1944-1971,1973
126	SHELL CREEK ABOVE SHELL CREEK	6278300	59		1957-1988
127	SOUTH FORK SHOSHONE RIVER NEAR	6280300	760		1957-1958,1960-1988
128	BEAUVAIS CREEK NEAR ST. XAVIER	6288200	256		1968-1977
129	LITTLE BIGHORN RIVER AT STATE	6289000	494		1940-1988
130	SOUTH TONGUE RIVER NEAR DAYTON	6297000	218		1946-1971
131	TONGUE RIVER NR DAYTON WYO	6298000	522	Yes	1920-1927,1929-1929,1941-1988
132	WOLF CREEK AT WOLF,WYO.	6299500	97		1946-1971
133	NORTH FORK POWDER RIVER NEAR H	6311000	63		1947-1988
134	CLEAR CREEK NEAR BUFFALO, WYO.	6318500	307		1918-1927,1939-1987
135	LITTLE MUDDY RIVER BL COW CREE	6331000	2240		1955-1983
136	BEAR DEN CREEK NR MANDAREE, ND	6332515	189		1967-1988
137	LITTLE MISSOURI R AT CAMP CROO	6334500	5043		1904-1906,1957-1988
138	LITTLE MISSOURI RIVER AT MARMA	6335500	11878		1939-1988
139	LITTLE MISSOURI RIVER NR WATFO	6337000	21274		1935-1988
140	KNIFE RIVER AT MANNING, ND	6339100	525		1968-1988
141	KNIFE RIVER NR GOLDEN VALLEY,	6339500	3149		1944-1988
142	KNIFE RIVER AT HAZEN, ND	6340500	5734		1938-1988
143	GREEN RIVER NR NEW HRADEC, ND	6344600	389		1965-1988
144	APPLE CREEK NR MENOKEN, ND	6349500	4301		1946-1988
145	CANNONBALL RIVER AT REGENT, ND	6350000	1485		1951-1988
146	CEDAR CREEK NR HAYNES, ND	6352000	1416		1951-1988
147	CEDAR CREEK NR RALEIGH, ND	6353000	4480		1963-1988
148	CANNONBALL RIVER AT BREIEN, ND	6354000	10496		1935-1988

Table A.1. Continued

149	BEAVER CREEK AT LINTON, ND	6354500	1836	1950-1988
150	SOUTH FORK GRAND R NEAR CASH S	6356500	3456	1947-1988
151	MOREAU R NEAR FAITH SD	6359500	6810	1944-1988
152	CHEYENNE R AT EDGEMONT SD	6395000	18286	1947-1988
153	CASTLE CR ABOVE DEERFIELD RES	6409000	203	1949-1988
154	ELK CR NEAR ELM SPRINGS SD	6425500	1382	1950-1988
155	BELLE FOURCHE RIVER BELOW MOOR	6426500	4206	1944-1970,1976-1983,1986-1987
156	SPEARFISH CR AT SPEARFISH SD	6431500	430	1947-1988
157	BAD R NEAR FORT PIERRE SD	6441500	7954	1929-1988
158	LITTLE WHITE R NEAR ROSEBUD SD	6449500	2611	1944-1988
159	WHITE R NEAR OACOMA SD	6452000	26112	1929-1988
160	KEYA PAHA R AT WEWELA SD	6464500	2739	1939-1940,1948-1988
161	JAMES RIVER NR MANFRED, ND	6467600	648	1958-1982,1986-1988
162	JAMES RIVER NR GRACE CITY, ND	6468170	2714	1969-1988
163	MAPLE R AT ND-SD STATE LINE	6471200	1833	1957-1988
164	SAND CR NEAR ALPENA SD	6476500	668	1951-1988
165	JAMES R NEAR SCOTLAND SD	6478500	52872	1929-1988
166	WEST FORK VERMILLION R NEAR PA	6478690	965	1962-1988
167	VERMILLION R NEAR WAKONDA SD	6479000	5555	1946-1983
168	BIG SIOUX RIVER NEAR BROOKINGS	6480000	9979	1954-1988
169	SKUNK CR AT SIOUX FALLS SD	6481500	1592	1949-1988
170	NORTH PLATTE RIVER NEAR NORTHG	6620000	3663	1916-1988
171	NORTH BRUSH CREEK NEAR SARATOG	6622700	96	1961-1988
172	ENCAMPMENT RIV AB HOG PARK CR	6623800	186	1965-1988
173	N PLATTE R AB SEMINOE RES NR S	6630000	10680	1940-1988
174	ROCK CR AB KING CANYON CANAL,	6632400	161	1966-1988
175	MEDICINE BOW R AB SEMINOE RESE	6635000	5942	1940-1988
176	ROCK CREEK ABOVE ROCK CREEK RE	6637750	24	1963-1988
177	LARAMIE RIVER NEAR JELM, WYO.	6658500	753	1905-1905,1912-1971
178	BEAR CREEK AT MORRISON, CO.	6710500	420	1901-1901,1920-1988

Table A.1. Continued

179	MIDDLE BOULDER CREEK AT NEDERL	6725500	93		1908-1910,1912-1988
180	HALFMOON CREEK NEAR MALTA, CO.	7083000	60		1947-1988
181	GRAPE CREEK NEAR WESTCLIFFE, C	7095000	819		1926-1927,1931-1961,1963-1988
182	SAN JOAQUIN R AT MILLER CROSSI	11226500	637		1922-1928,1952-1988
183	BEAR CR NR LAKE T.A.EDISON CAL	11230500	134		1922-1988
184	PITMAN C BL TAMARACK CREEK CAL	11237500	59		1929-1988
185	CANTUA CREEK NR CANTUA CREEK C	11253310	119		1967-1988
186	MERCED R AT HAPPY ISLES BRIDGE	11264500	463	Yes	1916-1988
187	MERCED RIVER AT POHONO BRIDGE	11266500	822	Yes	1917-1988
188	ORESTIMBA CREEK NR NEWMAN CALI	11274500	343		1933-1988
189	SF TUOLUMNE RIVER NR OAKLAND R	11281000	223		1924-1988
190	MIDDLE TUOLUMNE R AT OAKLAND R	11282000	188		1917-1988
191	CLAVEY RIVER NEAR BUCK MEADOWS	11283500	369		1960-1983,1987-1988
192	CLARK FORK STANISLAUS RIVER NE	11292500	173		1951-1988
193	NF STANISLAUS R BL SILVER CREE	11293500	71		1953-1987
194	HIGHLAND C BL SPICER MEADOWS R	11294000	116		1953-1988
195	COLE C NR SALT SPRINGS DAM CAL	11315000	54		1928-1942,1944-1988
196	FOREST CREEK NEAR WILSEYVILLE,	11316800	53		1961-1988
197	SACRAMENTO RIVER AT DELTA CALI	11342000	1088		1945-1988
198	HAT CREEK NEAR HAT CREEK CALIF	11355500	415		1927-1929,1931-1988
199	MCCLOUD RIVER NR MCCLOUD CALIF	11367500	916		1932-1988
200	MCCLOUD RIVER AB SHASTA LAKE C	11368000	1546		1946-1965
201	CLEAR CREEK AT FRENCH GULCH, C	11371000	294		1951-1988
202	CLEAR CREEK NR IGO CALIF	11372000	584		1941-1962
203	ELDER CREEK NEAR PASKENTA CALI	11379500	237		1949-1988
204	MILL C NR LOS MOLINOS CALIF	11381500	335		1929-1988
205	THOMES C AT PASKENTA CALIF	11382000	520		1921-1988

Table A.1. Continued

206	DEER CREEK NEAR VINA CALIF	11383500	532	Yes	1912-1915,1921-1988
207	BIG CHICO CREEK NEAR CHICO CAL	11384000	185		1931-1986
208	BUTT C AB ALM-BUT C TU NR PRAT	11400000	177		1937-1964
209	INDIAN CREEK NR CRESCENT MILLS	11401500	1892		1907-1909,1912-1917,1931-1988
210	SPANISH CREEK ABOVE BLACKHAWK	11402000	471		1934-1988
211	SPANISH C AT KEDDIE CALIF	11402500	497		1912-1933
212	EAST BRANCH OF NF FEATHER R NR	11403000	2624		1951-1961,1969-1982
213	OREGON CREEK AT CAMPTONVILLE,	11409300	59		1968-1988
214	NORTH YUBA RIVER BELOW GOODYEA	11413000	640		1931-1937,1939-1988
215	LATIR CREEK NEAR CERRO, N. MEX	8263000	27		1946-1970
216	RED RIVER AT MOUTH, NEAR QUEST	8267000	486		1952-1978
217	RIO HONDO NEAR VALDEZ, N. MEX.	8267500	93		1935-1988
218	RIO PUEBLO DE TAOS NEAR TAOS,	8269000	170		1915-1915,1941-1951,1964-1988
219	RIO LUCERO NEAR ARROYO SECO, N	8271000	42		1914-1915,1935-1951,1964-1988
220	RIO GRANDE DEL RANCHO NEAR TAL	8275500	212		1953-1982,1986-1988
221	RIO CHIQUITO NEAR TALPA, N. ME	8275600	95		1958-1980
222	RIO GRANDE BELOW TAOS JUNCTION	8276500	24832	Yes	1926-1988
223	EMBUDO CREEK AT DIXON, NM	8279000	781	Yes	1924-1925,1928-1929,1931-1955,1963-1988
224	RIO CHAMA AT PARK VIEW, N. MEX	8283500	1037		1914-1915,1931-1955
225	EL RITO NEAR EL RITO, N. MEX.	8288000	129		1932-1950
226	RIO OJO CALIENTE AT LA MADERA,	8289000	1073		1933-1988
227	SANTA CRUZ RIVER NEAR CUNDIYO,	8291000	220		1933-1988
228	JEMEZ RIVER NR JEMEZ,NM	8324000	1203		1937-1940,1950-1950,1954-1988
229	RIO MORA NEAR TERRERO, NM	8377900	136		1964-1988
230	PECOS R NR PECOS, NM	8378500	484		1920-1920,1924-1924,1931-1988
231	GALLINAS CREEK NEAR MONTEZUMA,	8380500	215		1927-1988

Table A.1. Continued

232	RIO RUIDOSO AT HOLLYWOOD, N. M	8387000	307		1954-1988
233	BLACK RIVER ABOVE MALAGA, N. M	8405500	878		1948-1988
234	DELAWARE RIVER NR RED BLUFF, N	8408500	1764		1938-1988
235	CRYSTAL RIVER AB AVALANCHE C,	9081600	428		1956-1988
236	TAYLOR RIVER AT ALMONT, CO.	9110000	1221		1911-1936
237	EAST RIVER AT ALMONT CO.	9112500	740		1911-1922,1935-1988
238	TOMICHI CREEK AT SARGENTS, CO.	9115500	381		1917-1922,1938-1972
239	TOMICHI CREEK AT GUNNISON, CO.	9119000	2716		1938-1988
240	LAKE FORK AT GATEVIEW, CO.	9124500	855		1938-1988
241	CURECANTI CREEK NEAR SAPINERO,	9125000	90		1946-1972
242	SMITH FORK NEAR CRAWFORD, CO.	9128500	110		1936-1988
243	NORTH FORK GUNNISON RIVER NEAR	9132500	1347		1934-1960
244	LEROUX CREEK NEAR CEDAREDGE, C	9134500	88		1937-1956,1961-1969
245	UNCOMPAHGRE RIVER NEAR RIDGWAY	9146200	381	Yes	1959-1988
246	UNCOMPAHGRE RIVER AT COLONA, C	9147500	1147		1913-1985
247	DOLORES RIVER BELOW RICO, CO.	9165000	269		1952-1988
248	DISAPPOINTMENT CREEK NEAR DOVE	9168100	376		1958-1986
249	COLORADO RIVER NEAR CISCO UTAH	9180500	61696		1914-1914,1916-1917,1923-1937
250	GREEN RIVER AT WARREN BRIDGE,	9188500	1198		1932-1988
251	PINE CREEK ABOVE FREMONT LAKE,	9196500	194		1955-1988
252	POLE CREEK BELOW LITTLE HALF M	9198500	224		1939-1971
253	FALL CREEK NEAR PINEDALE WYO	9199500	95		1939-1971
254	EAST FORK RIVER NEAR BIG SANDY	9203000	203		1939-1988
255	NORTH PINEY CREEK NEAR MASON,	9205500	148		1916-1916,1932-1971
256	FONTENELLE CR NR HERSCHLER RAN	9210500	389		1952-1988
257	BIG SANDY R AT LECKIE RANCH, N	9212500	241		1940-1971
258	EAST FORK OF SMITH FORK NR ROB	9220000	136		1940-1971

Table A.1. Continued

259	HAMS FORK BELOW POLE CREEK, NE	9223000	328	1953-1988
260	YAMPA RIVER AT STEAMBOAT SPRIN	9239500	1546	1905-1906,1911-1987
261	ELK RIVER AT CLARK, CO.	9241000	553	1911-1916,1918-1918,1920-1920,1932-1988
262	ELKHEAD CREEK NEAR ELKHEAD, CO	9245000	164	1954-1988
263	MILK CREEK NEAR THORNBURGH, CO	9250000	166	1953-1986
264	YAMPA RIVER NEAR MAYBELL, CO.	9251000	8730	1917-1987
265	SAVERY CREEK NEAR SAVERY, WY	9256000	845	1942-1946,1948-1971,1986-1988
266	ROCK CREEK NEAR HANNA, UTAH	9278500	312	1950-1969,1975-1988
267	WHITEROCKS RIVER NEAR WHITEROC	9299500	289	1910-1910,1919-1920,1930-1988
268	WHITE RIVER NEAR MEEKER, CO.	9304500	1933	1902-1906,1910-1988
269	FISH CREEK ABOVE RESERVOIR NEA	9310500	154	1939-1988
270	GREEN RIVER AT GREEN RIVER, UT	9315000	114793	1895-1899,1906-1962
271	MUDDY CREEK NEAR EMERY, UTAH	9330500	269	1911-1913,1950-1988
272	VALLECITO CREEK NEAR BAYFIELD,	9352900	185	1963-1988
273	ANIMAS RIVER AT DURANGO, CO.	9361500	1772	1898-1898,1900-1900,1913-1925,1928-1988
274	ANIMAS RIVER AT FARMINGTON, NM	9364500	3482	1914-1914,1920-1925,1931-1988
275	COTTONWOOD WASH NR BLANDING UT	9378700	525	1965-1987
276	SAN JUAN RIVER NEAR BLUFF, UTA	9379500	58880	1916-1916,1928-1928,1930-1940
277	LITTLE COLORADO R ABV LYMAN LA	9384000	1897	1941-1988
278	LITTLE COLORADO RIVER NEAR CAM	9402000	67835	1948-1988
279	BRIGHT ANGEL CREEK NEAR GRAND	9403000	259	1924-1973
280	EAST FORK VIRGIN RIVER NR GLEN	9404450	190	1967-1988
281	SANTA CLARA RIVER NR PINE VALL	9408400	48	1960-1988
282	VIRGIN RIVER AT LITTLEFIELD, A	9415000	13030	1930-1988
283	LEE CANYON NR CHARLESTON PARK,	9419610	24	1964-1988

Table A.1. Continued

284	GILA RIVER NEAR GILA, NM	9430500	4772		1929-1988
285	MOGOLLON CREEK NEAR CLIFF, NM	9430600	177		1968-1988
286	GILA RIVER NEAR REDROCK, NM	9431500	7242		1931-1955,1963-1988
287	TULAROSA RIVER ABOVE ARAGON, N	9442692	241		1967-1988
288	SAN FRANCISCO RIVER AT CLIFTON	9444500	7081		1914-1915,1917-1917,1928-1933,1936-1988
289	SAN PEDRO RIVER AT CHARLESTON,	9471000	3121	Yes	1905-1905,1913-1926,1929-1933,1936-1988
290	SANTA CRUZ RIVER NEAR LOCHIEL,	9480000	210		1950-1988
291	SALT RIVER NEAR ROOSEVELT, ARI	9498500	11023	Yes	1914-1988
292	WET BOTTOM CREEK NR CHILDS, AR	9508300	93		1968-1988
293	VERDE RIVER BLW TANGLE CR AB H	9508500	15017	Yes	1946-1988
294	SULPHUR CREEK ABV RESERVOIR NR	10015700	164		1958-1988
295	SMITHS FORK NEAR BORDER, WY	10032000	422		1943-1988
296	CUB RIVER NEAR PRESTON, IDAHO	10093000	81		1941-1952,1956-1986
297	BLACKSMITH FORK AB U.P.&L. CO,	10113500	673		1915-1917,1919-1988
298	WEBER RIVER NEAR OAKLEY, UTAH	10128500	415	Yes	1905-1988
299	CHALK CREEK AT COALVILLE UTAH	10131000	640		1928-1988
300	NORTH FORK PROVO RIVER NEAR KA	10153800	62		1964-1988
301	RED BUTTE CREEK AT FT. DOUGLAS	10172200	19	Yes	1964-1988
302	VERNON CREEK NEAR VERNON, UTAH	10172700	64		1959-1988
303	TROUT CR NR CALLAO UTAH	10172870	21		1960-1988
304	SEVIER RIVER AT HATCH UTAH	10174500	870		1915-1928,1940-1988
305	SALINA CREEK NEAR EMERY UTAH	10205030	133		1964-1988
306	OAK CREEK NR. FAIRVIEW, UTAH	10208500	30		1965-1988
307	BEAVER RIV NR BEAVER UTAH	10234500	233		1915-1988
308	STEPTOE C NR ELY, NV	10244950	28		1967-1988
309	S TWIN R NR ROUND MOUNTAIN, NV	10249300	51		1966-1988
310	CHIATOVICH C NR DYER, NV	10249900	95		1961-1982

Table A.1. Continued

311	BORREGO PALM C NR BORREGO SPRI	10255810	56		1951-1988
312	TAHQUITZ CR NR PALM SPRINGS CA	10258000	41		1948-1982,1984-1988
313	PALM CANYON CREEK NR PALM SPRI	10258500	238		1931-1941,1948-1988
314	ANDREAS CREEK NEAR PALM SPRING	10259000	22		1949-1988
315	BIG ROCK CREEK NEAR VALYERMO,C	10263500	59		1924-1988
316	W WALKER R BL L WALKER R NR CO	10296000	461		1939-1988
317	W WALKER R NR COLEVILLE, CA	10296500	640		1910-1910,1916-1937,1958-1988
318	WALKER R NR WABUSKA, NV	10301500	6656		1904-1904,1921-1923,1926-1935,1940-1941,1943-1943,1945-1988
319	E F CARSON R BL MARKLEEVILLE C	10308200	707		1961-1988
320	W F CARSON R AT WOODFORDS, CA	10310000	167		1939-1988
321	CARSON R NR FORT CHURCHILL, NV	10312000	3333	Yes	1913-1923,1925-1927,1929-1932,1934-1988
322	LAMOILLE C NR LAMOILLE, NV	10316500	64		1916-1922,1944-1988
323	HUMBOLDT R AT PALISADE, NV	10322500	12826		1903-1906,1912-1912,1914-1988
324	REESE R NR IONE, NV	10325500	136		1952-1980
325	MARTIN C NR PARADISE VALLEY, N	10329500	440		1922-1988
326	BLACKWOOD CREEK NR TAHOE CITY	10336660	29		1961-1988
327	TROUT CREEK NR TAHOE VALLEY CA	10336780	94		1961-1988
328	SAGEHEN CREEK NR TRUCKEE CALIF	10343500	27		1954-1988
329	MC DERMITT C NR MC DERMITT, NV	10352500	576		1949-1984,1986-1988
330	QUINN R NR MC DERMITT, NV	10353500	2816		1949-1982,1985-1985
331	DEEP CREEK ABOVE ADEL,OREG.	10371500	637		1923-1923,1933-1988
332	CHEWAUCAN RIVER NEAR PAISLEY,O	10384000	704		1925-1988
333	SILVIES RIVER NEAR BURNS,OREG.	10393500	2391		1904-1905,1910-1912,1918-1920,1923-1988

Table A.1. Continued

334	DONNER UND BLITZEN RIVER NR FR	10396000	512	1912-1913,1915-1916,1918-1921,1939-1988
335	SANTA YSABEL CREEK NEAR RAMONA	11025500	287	1913-1922,1944-1953
336	TEMECULA CREEK NEAR AGUANGA, C	11042400	335	1958-1988
337	CITY C NR HIGHLAND CA.+ CANALS	11055801	50	1925-1988
338	EAST TWIN CREEK NEAR ARROWHEAD	11058500	23	1921-1988
339	LONE PINE CREEK NR KEENBROOK C	11063500	39	1921-1938,1950-1988
340	SANTIAGO C A MODJESKA CA	11075800	33	1962-1988
341	ARROYO SECO NR PASADENA CALIF	11098000	41	1914-1915,1917-1988
342	SESPE CREEK NR WHEELER SPRINGS	11111500	127	1948-1988
343	SESPE C + FILLMORE IRR CO CA N	11113001	643	1940-1985
344	COYOTE CREEK NEAR OAK VIEW, CA	11117600	34	1959-1988
345	SANTA ANA CREEK NEAR OAK VIEW	11117800	23	1959-1988
346	SANTA CRUZ CR NR SANTA YNEZ CA	11124500	189	1942-1988
347	SALSIPUEDES CR NR LOMPOC CA	11132500	121	1942-1988
348	SISQUOC RIVER NEAR SISQUOC, CA	11138500	719	1944-1988
349	LOPEZ C NR ARROYO GRANDE CA	11141280	54	1968-1988
350	BIG SUR RIVER NR BIG SUR CALIF	11143000	119	1951-1988
351	SANTA RITA C NR TEMPLETON CALI	11147070	47	1962-1988
352	SAN ANTONIO RIVER NEAR LOCKWOO	11149900	556	1966-1988
353	SAN LORENZO C BL BITTERWATER C	11151300	596	1959-1988
354	ARROYO SECO NEAR SOLEDAD, CAL.	11152000	625	1902-1988
355	EL TORO CREEK NR SPRECKELS, CA	11152540	82	1962-1988
356	SOQUEL CR AT SOQUEL CALIF	11160000	103	1952-1988
357	SAN LORENZO RIVER NEAR BOULDER	11160020	16	1969-1988
358	ZAYANTE CREEK AT ZAYANTE CALIF	11160300	28	1958-1988
359	SAN LORENZO R AT BIG TREES CAL	11160500	271	1937-1988

Table A.1. Continued

360	PESCADERO CREEK NEAR PESCADERO	11162500	118		1952-1988
361	ARROYO VALLE BL LANG CN NR LIV	11176400	333		1964-1988
362	SAN RAMON CREEK AT SAN RAMON,	11182500	15		1953-1988
363	COMBINED FLOW OF KERN R AND KE	11186001	2166		1912-1988
364	KERN RIVER AT KERNVILLE CALIF	11187000	2583	Yes	1906-1912,1954-1988
365	SF KERN R NR ONYX CALIF	11189500	1357		1912-1913,1920-1925,1930-1942,1947-1988
366	DEER CREEK NEAR FOUNTAIN SPRIN	11200800	213		1969-1988
367	NF OF MF TULE R NR SPRINGVILLE	11202001	101		1941-1988
368	MF KAWEAH R NR POTWISHA CAMP C	11206501	261		1950-1988
369	MARBLE FK KAWEAH AT POTWISHA C	11208001	132		1951-1988
370	KAWEAH RIVER AT THREE RIVERS C	11209900	1070		1959-1988
371	SOUTH FORK KAWEAH RIVER AT THR	11210100	222		1959-1988
372	KAWEAH R NR THREE RIVERS CALIF	11210500	1329		1904-1961
373	KINGS RIVER AB NF NR TRIMMER C	11213500	2437		1927-1928,1932-1982
374	NF KINGS R NR CLIFF CAMP CALIF	11215000	463		1922-1957
375	MILL CREEK NEAR PIEDRA CALIF	11221700	325		1958-1988
376	KINGS R AT PIEDRA CALIF	11222000	4334		1896-1951
377	LOS GATOS CREEK AB NUNEZ CANYO	11224500	245		1946-1988
378	SOUTH YUBA RIVER NEAR CISCO, C	11414000	133		1943-1988
379	NF AMERICAN R AT NORTH FORK DA	11427000	876		1942-1988
380	DUNCAN CREEK NR FRENCH MEADOWS	11427700	25		1961-1988
381	PILOT CREEK ABOVE STUMPY MEADO	11431800	30		1961-1988
382	S.F. SILVER CREEK NEAR ICE HOU	11441500	70		1925-1959
383	KELSEY CREEK NEAR KELSEYVILLE,	11449500	94		1947-1988
384	NF CACHE C AT HOUGH SPRING NEA	11451100	154		1972-1988
385	NAPA RIVER NEAR ST. HELENA CAL	11456000	208		1930-1932,1940-1988

Table A.1. Continued

386	RUSSIAN RIVER NEAR UKIAH, CALI	11461000	256	1912-1913,1953-1988
387	DRY CREEK NR GEYSERVILLE CALIF	11465200	415	1960-1983
388	NAVARRO RIVER NEAR NAVARRO, CA	11468000	776	1951-1988
389	NOYO RIVER NR FORT BRAGG CALIF	11468500	271	1952-1988
390	MATTOLE RIVER NR PETROLIA CALI	11469000	614	1912-1913,1951-1988
391	OUTLET CREEK NR LONGVALE, CA.	11472200	412	1957-1988
392	MIDDLE FORK EEL R NR DOS RIOS	11473900	1907	1966-1988
393	ELDER CREEK NEAR BRANSCOMB CAL	11475560	17	1968-1988
394	SF EEL RIVER AT LEGGETT CALIF	11475800	635	1966-1988
395	SF EEL RIVER NR MIRANDA CALIF	11476500	1375	1940-1988
396	BULL CREEK NEAR WEOTT, CALIF.	11476600	72	1961-1988
397	EEL RIVER AT SCOTIA CALIF	11477000	7969	1911-1914,1917-1988
398	VAN DUZEN RIVER NR BRIDGEVILLE	11478500	568	1951-1988
399	LITTLE R NR TRINIDAD CALIF	11481200	104	1956-1988
400	REDWOOD C NR BLUE LAKE CALIF	11481500	173	1954-1958,1973-1988
401	REDWOOD CREEK AT ORICK CALIF	11482500	709	1912-1913,1954-1988
402	SPRAGUE RIVER NEAR BEATTY,OREG	11497500	1313	1954-1988
403	SPRAGUE RIVER NEAR CHILOQUIN,O	11501000	4045	1922-1988
404	SCOTT RIVER NEAR FORT JONES, C	11519500	1672	1942-1988
405	INDIAN CREEK NEAR HAPPY CAMP,	11521500	307	1958-1988
406	SALMON RIVER AT SOMES BAR CALI	11522500	1923	1912-1915,1928-1988
407	TRINITY R AB COFFEE C NR TRINI	11523200	381	1958-1988
408	TRINITY RIVER AT LEWISTON CALI	11525500	1841	1912-1960
409	S F TRINITY RIVER BL HYAMPOM,	11528700	1956	1966-1988
410	TRINITY R AT HOOPA CALIF	11530000	7304	1912-1913,1917-1918,1932-1960
411	SMITH RIVER NEAR CRESCENT CITY	11532500	1559	1932-1988

Table A.1. Continued

412	DUNGENESS RIVER NEAR SEQUIM, W	12048000	399		1924-1930,1938-1988
413	DUCKABUSH RIVER NEAR BRINNON,	12054000	170		1939-1988
414	NF SKOKOMISH R BLW STRCSE RPDS	12056500	146	Yes	1925-1988
415	SKYKOMISH RIVER NEAR GOLD BAR,	12134500	1370		1929-1988
416	SNOQUALMIE RIVER NEAR SNOQUALM	12144500	960		1903-1903,1927-1927,1959-1988
417	SAUK R ABV WHITECHUCK R NR DAR	12186000	389		1918-1920,1922-1922,1929-1988
418	SAUK RIVER NEAR SAUK, WASH.	12189500	1828		1929-1988
419	FISHER RIVER NEAR LIBBY, MT.	12302055	2145		1968-1988
420	FLOWER CREEK NEAR LIBBY, MT.	12303100	28		1961-1988
421	YAAK RIVER NEAR TROY, MT.	12304500	1961		1957-1988
422	BOULDER CREEK NR LEONIA ID	12305500	143		1929-1971,1974-1977
423	MISSION CREEK NEAR COPELAND, I	12316800	59		1959-1981
424	BOUNDARY CREEK NR PORTHILL ID	12321500	248		1931-1988
425	BOULDER CREEK AT MAXVILLE, MT.	12330000	183		1940-1988
426	MIDDLE FORK ROCK CREEK NEAR PH	12332000	315		1938-1988
427	CLARK FORK AT ST. REGIS, MT.	12354500	27415	Yes	1912-1923,1929-1988
428	MIDDLE FORK FLATHEAD RIVER NEA	12358500	2888		1940-1988
429	S F FLATHEAD R AB TWIN C, NR H	12359800	2970		1965-1982
430	SWAN RIVER NEAR BIGFORK, MT.	12370000	1718		1923-1988
431	PROSPECT CREEK AT THOMPSON FAL	12390700	466		1957-1988
432	PACK RIVER NR COLBURN ID	12392300	317		1959-1982
433	PRIEST R @ OUTLET OF PRIEST LK	12393500	1464		1914-1918,1920-1948
434	COLVILLE RIVER AT KETTLE FALLS	12409000	2578		1924-1931,1933-1988
435	COEUR D'ALENE R AB SHOSHONE CK	12411000	858	Yes	1951-1988
436	COEUR D'ALENE RIVER AT ENAVILL	12413000	2291	Yes	1940-1988
437	COEUR D'ALENE RIVER NR CATALDO	12413500	3123		1912-1912,1921-1972,1987-1988
438	ST. JOE RIVER AT CALDER, ID	12414500	2637		1912-1912,1921-1988

Table A.1. Continued

439	ST. MARIES RIVER NEAR SANTA ID	12414900	704		1966-1988
440	ST MARIES RIVER AT LOTUS ID	12415000	1119		1921-1966
441	HAYDEN CK BELOW N FK, NR HAYDE	12416000	56		1949-1953,1959-1959,1966-1988
442	HANGMAN CREEK AT SPOKANE, WASH	12424000	1764		1949-1977,1979-1988
443	LITTLE SPOKANE RIVER AT ELK, W	12427000	294		1949-1971
444	LITTLE SPOKANE RIVER AT DARTFO	12431000	1702		1930-1932,1948-1988
445	ANDREWS CREEK NEAR MAZAMA, WAS	12447390	57		1969-1988
446	METHOW RIVER AT TWISP, WA	12449500	3331		1920-1929,1934-1962
447	STEHEKIN RIVER AT STEHEKIN, WA	12451000	822		1912-1915,1928-1988
448	ENTIAT RIVER NEAR ARDENVOIR, W	12452800	520		1958-1988
449	WHITE RIVER NEAR PLAIN, WASH.	12454000	384		1955-1983
450	WENATCHEE RIVER BELOW WENATCHE	12455000	699		1933-1958
451	WENATCHEE RIVER AT PLAIN, WASH	12457000	1513		1911-1929,1932-1979
452	ICICLE CREEK ABV SNOW CR NR LE	12458000	494		1937-1971
453	WENATCHEE RIVER AT PESHASTIN,	12459000	2560		1930-1988
454	WENATCHEE RIVER AT MONITOR, WA	12462500	3331		1963-1988
455	CRAB CREEK AT IRBY, WASH.	12465000	2668		1943-1988
456	WILSON CREEK AT WILSON CREEK,	12465500	1093		1952-1957,1959-1971,1973-1973
457	AMERICAN RIVER NEAR NILE, WASH	12488500	202		1940-1988
458	NORTH FORK AHTANUM CREEK NEAR	12500500	176		1911-1915,1932-1978
459	PACIFIC CREEK AT MORAN, WY	13011500	433		1945-1975,1979-1988
460	BUFFALO FORK ABOVE LAVA CREEK	13011900	827		1966-1988
461	CACHE CREEK NEAR JACKSON WY	13018300	27		1963-1988
462	GREYS RIVER AB RESERVOIR, NR A	13023000	1147		1938-1938,1954-1988
463	PALOUSE RIVER BELOW SOUTH FORK	13349210	2038		1964-1972,1976-1988
464	PALOUSE RIVER AT HOOPER, WASH.	13351000	6400	Yes	1898-1899,1901-1906,1909-1911,1914-1915,1952-1988

Table A.1. Continued

465	S.F. WALLA WALLA RIVER NEAR MI	14010000	161		1908-1909,1911-1917,1932-1988
466	UMATILLA RIVER AB MEACHAM CR N	14020000	335		1934-1988
467	WILLOW CREEK AT HEPPNER, OREG.	14034500	248		1952-1982
468	CAMAS CREEK NEAR UKIAH, OREG.	14042500	310		1915-1917,1922-1923,1942-1988
469	M FK JOHN DAY R AT RITTER, ORE	14044000	1318		1930-1988
470	JOHN DAY RIVER AT SERVICE CREE	14046500	13030		1930-1988
471	JOHN DAY R AT MCDONALD FERRY,O	14048000	19405		1906-1988
472	CROOKED R NR PRINEVILLE, OREG.	14080500	6912		1942-1959
473	WHITE RIVER BELOW TYGH VALLEY,	14101500	1068		1918-1988
474	KLICKITAT RIVER NEAR PITT, WAS	14113000	3320		1910-1911,1929-1988
475	SANDY RIVER NEAR MARMOT, OREG.	14137000	671		1912-1915,1917-1918,1920-1988
476	FALL CR. NEAR LOWELL, OREG.	14150300	302		1964-1988
477	ROW RIVER ABOVE PITCHER CREEK	14154500	540		1936-1988
478	MCKENZIE R AT MCKENZIE BRIDGE,	14159000	891		1911-1962
479	NO SANTIAM R BL BOULDER CR NR	14178000	553		1908-1909,1929-1988
480	BREITENBUSH RIVER ABV FRENCH C	14179000	276		1933-1987
481	SOUTH SANTIAM RIVER BELOW CASC	14185000	445		1936-1988
482	QUARTZVILLE CREEK NEAR CASCADI	14185900	254		1964-1964,1966-1988
483	THOMAS CREEK NEAR SCIO,OREG.	14188800	279		1963-1987
484	LUCKIAMUTE RIVER NEAR SUVER, O	14190500	614		1906-1911,1941-1988
485	WILLAMETTE RIVER AT SALEM,OREG	14191000	18637		1910-1916,1924-1941
486	WILLAMINA CREEK NEAR WILLAMINA	14193000	166		1935-1988
487	MOLALLA R AB PC NR WILHOIT, OR	14198500	248		1936-1988
488	PUDDING RIVER NEAR MOUNT ANGEL	14201000	522	Yes	1940-1965
489	CLACKAMAS RIVER AT BIG BOTTOM,	14208000	348		1921-1970
490	EAST FORK LEWIS RIVER NEAR HEI	14222500	320		1930-1985,1987-1988

Table A.1. Continued

491	CISPUS RIVER NEAR RANDLE, WASH	14232500	822		1930-1988
492	COWLITZ RIVER NR RANDLE, WASH.	14233400	2637		1968-1988
493	COWEMAN RIVER NEAR KELSO, WASH	14245000	305		1951-1979
494	YOUNGS RIVER NEAR ASTORIA, ORE	14251500	103		1928-1958
495	NEHALEM RIVER NEAR FOSS, OREG.	14301000	1708		1940-1988
496	WILSON RIVER NEAR TILLAMOOK, O	14301500	412		1915-1915,1932-1988
497	NESTUCCA R NR BEAVER OREG	14303600	461		1965-1988
498	FIVE RIVERS NR FISHER, OREG.	14306400	292		1961-1963,1968-1988
499	ALSEA RIVER NEAR TIDEWATER, OR	14306500	855		1940-1988
500	SIUSLAW R NR MAPLETON, OREG.	14307620	1505		1968-1988
501	JACKSON CREEK NEAR TILLER, ORE	14307700	389		1956-1986
502	SOUTH UMPQUA RIVER AT TILLER,	14308000	1149		1911-1911,1940-1988
503	STEAMBOAT CREEK NEAR GLIDE,ORE	14316700	581		1957-1988
504	LITTLE RIVER AT PEEL, OREG.	14318000	453		1955-1988
505	SOUTH FORK COQUILLE RIVER AT P	14325000	433		1917-1926,1930-1988
506	ROGUE RIVER ABOVE PROSPECT, OR	14328000	799		1909-1910,1924-1988
507	ELK CREEK NEAR TRAIL, OREG.	14338000	330		1947-1988
508	APPLEGATE RIVER NEAR COPPER, O	14362000	576		1939-1979
509	ILLINOIS RIVER AT KERBY, OREG.	14377000	932		1927-1961
510	ILLINOIS RIVER NEAR KERBY, ORE	14377100	973		1962-1988
511	CHETCO RIVER NR BROOKINGS, ORE	14400000	694		1970-1988
512	Clarks Fork Yellowstone River	6208500	5176	Yes	1922-2003
513	L POWDER RIVER AB DRY C NR WES	6324970	3167	Yes	1973-2003
514	Powder River near Locate MT	6326500	33454	Yes	1939-2003
515	CACHE LA POUDRE R A MO OF CN,	6752000	2701	Yes	1901-2003
516	SAGUACHE CREEK NEAR SAGUACHE,	8227000	1523	Yes	1911-2003
517	RITO DE LOS FRIJOLES IN BANDEL	8313350	45	Yes	1984-1996
518	COLORADO R BELOW BAKER GULCH,	9010500	164	Yes	1954-2003

Table A.1. Continued

519	DRY FORK AT UPPER STATION, NEA	9095300	249	Yes	1996-2003
520	EAST RIVER BL CEMENT CREEK NR	9112200	609	Yes	1964-2003
521	WEST CLEAR CREEK NEAR CAMP VER	9505800	617	Yes	1965-2003
522	BEAR RIVER NEAR UTAH-WYOMING S	10011500	440	Yes	1943-2003
523	TRUCKEE R A FARAD CA	10346000	2386	Yes	1910-2003
524	ST. JOE RIVER AT RED IVES RANG	12413875	274	Yes	1998-2003
525	St. Regis River near St. Regis	12354000	776	Yes	1911-2003
526	Bitterroot River near Missoula	12352500	7204	Yes	1901-2003
527	TOPPENISH CREEK NEAR FORT SIMC	12506000	312	Yes	1910-2003
528	SATUS CR BELOW DRY CR NEAR TOP	12508500	1114	Yes	1914-2003
529	SNAKE RIVER AB JACKSON LAKE AT	13010065	1244	Yes	1984-2003
530	SPRING CREEK AT SHEEPSKIN RD N	13075983	46	Yes	1981-2003
531	MEDICINE LODGE CREEK NR SMALL	13116500	691	Yes	1922-2003
532	HENRYS FORK NR REXBURG ID	13056500	7475	Yes	1910-2003
533	FALLS RIVER NR SQUIRREL ID	13047500	855	Yes	1905-2003
534	SALT RIVER AB RESERVOIR NR ETN	13027500	2122	Yes	1954-2003
535	LITTLE GRANITE CREEK AT MOUTH	13019438	54	Yes	1982-1992
536	LITTLE ABIQUA CREEK NEAR SCOTT	14200400	25	Yes	1994-2003
537	TAYLOR CREEK NEAR SELLECK, WAS	12117000	44	Yes	1946-2002
538	NEWAUKUM CREEK NEAR BLACK DIAM	12108500	70	Yes	1945-2002
539	GREEN RIVER ABV TWIN CAMP CREE	12103380	42	Yes	1993-1999
540	KINGS R BL NF NR TRIMMER CA	11218500	3436	Yes	1963-1993
541	COSUMNES R A MICHIGAN BAR CA	11335000	1372	Yes	1908-2003
542	CAJON C BL LONE PINE C NR KEEN	11063510	145	Yes	1972-2003
543	CUCAMONGA C NR UPLAND CA	11073470	25	Yes	1930-1975

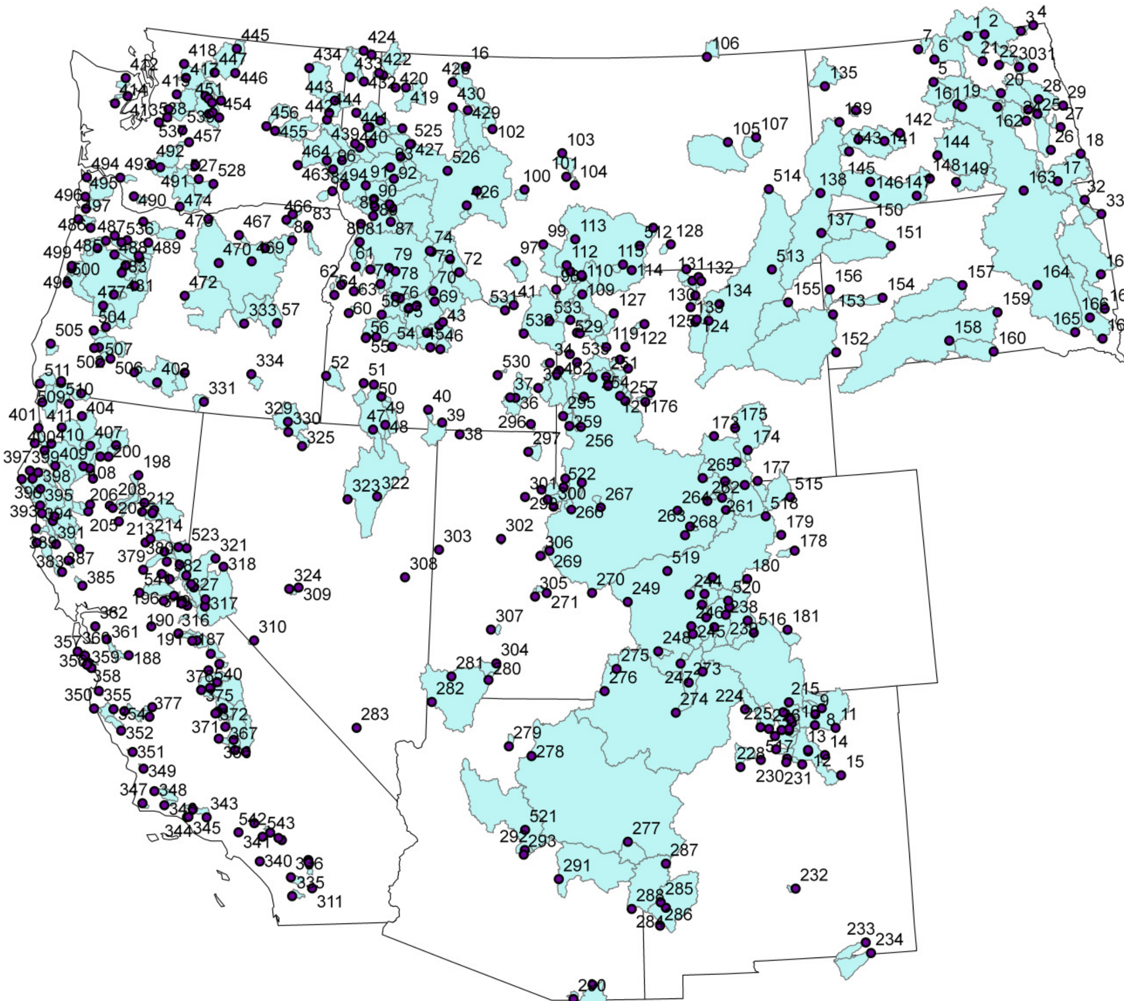


Figure A.1. Map of streamflow sites and their watersheds. The sites labels refer the first column in Table A.1.

Appendix B

This appendix describes the backward stepwise multiple regression models developed by Ryan A. Hill and Charles P. Hawkins (personal communication, 2008) for predicting mean annual temperature (MAT), mean winter temperature (MWT) and mean summer temperature (MST) of streams in western US. The predictor variables of these models are given in Table B.1 and Table B.2 gives the model statistics. The common predictor variables between the temperature models and the models predicting the streamflow regime class are described in Chapter 4. The rest of the variables are described in Table B.3. Table B.4 lists the sites used in developing these regressions. These are shown in the map in Figure B.1 labeled to refer the first column of Table B.4.

Table B.1. The Predictor Variables and their Coefficients Used in the Temperature Models

MAT		MWT		MST	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
CONSTANT	21.462	CONSTANT	3.220	CONSTANT	8.173
BDH_AVE	1.888	GRANITIC	-0.014	BDH_AVE	4.905
ELEV_MEAN	0.000	TMIN_PT	0.020	LATITUDE	-0.210
HYDR_AVE	8.585	TMEAN_WS	0.044	LOGRCHSLP	-0.748
LATITUDE	-0.267	PRMH_AVE	0.226	LOGSQ_KM	0.683
LOGSQ_KM	0.554	WTDH_WS	-0.466	LST32F_AVE	-0.028
LST32F_AVE	-0.015	LOGSQ_KM	0.162	PRMH_AVE	-0.049
LOGMAXP_PT	-0.994			RDH_AVE	0.029
OMH_AVE	-0.240			SHAPE1	-6.863
SHAPE1	-3.076			TMAX_WS	0.040
TMEAN_PT	0.031				

Table B.2. Model Statistics for Backward Stepwise Multiple Regression Models of Temperature Variables

	MAT	MWT	MST
RMSE (^{O}C)	0.97	1.73	2.14
Adj. R^2	0.86	0.74	0.73

Table B.3. Watershed Attributes That Were Used in the Temperature Model but Not Described in Chapter 4

Metric	Description	Unit	Source
TMEAN_PT	Annual mean of the PRISM mean monthly air temperature at the outlet.	mm	PRISM
TMIN_PT	Coldest month's PRISM mean monthly air temperature at the outlet.	mm	PRISM
LOGMAXP_PT	Log10 of the wettest month's PRISM mean monthly precipitation at the outlet.		PRISM
HYDR_AVE	Ratio of minimum of mean monthly flows on record to the mean of the maximum monthly flows, calculated as a watershed average.	-	Derived from USGS streamflow
LOGSQ_KM	Log10 of the watershed drainage area		
LOGRCHSLP	Log10 of channel reach slope as measured by		National Hydrography Dataset (NHDPlus)
LATITUDE	Latitude of the gauge	Deceimal degree	

Table B.4. Temperature Site Information. The Index Column is Used to Label the Sites in Figiure B.1

Index	Site name	USGS code	Drainage area, Sq. Km
1	BREITENBUSH	14179000	273
2	KALAMA RIVER	14223600	522
3	CALIFORNIA G	7081800	29
4	WINBERRY CRE	14150800	113
5	FALL CREEK B	14151000	481
6	BLUE CR ABV	12433542	16
7	PICEANCE CRE	9306007	460
8	PICEANCE CRE	9306042	653
9	PICEANCE C B	9306045	663
10	PICEANCE CRE	9306061	802
11	PICEANCE CRE	9306200	1312
12	PICEANCE CRE	9306222	1692
13	BLUE CR NR M	12433561	48
14	M F WILLAMET	14145500	1018
15	S FK ROGUE R	14334700	634
16	QUINN R NR M	10353500	2848
17	MARTIS C A H	10339250	13
18	BIG HOLE RIV	6024580	4153
19	CEDAR RIVER	12117500	331
20	CEDAR RIVER	12119000	440
21	LOST CREEK N	14158980	199
22	MIDDLE FORK	14361590	131
23	ELK CREEK NE	14337800	204
24	RALSTON CREE	6719725	96
25	NAPA RIVER N	11456000	213
26	ELWHA RIVER	12045500	695
27	ELK CREEK BE	14337830	268
28	MIDDLE FORK	12141300	401
29	RALSTON CREE	6719740	111
30	BIRCH CR NR	13116970	62
31	SALT CREEK N	9179200	82
32	SNAKE RIVER	13010200	1299
33	BUMPING RIVE	12488000	192
34	ELK CREEK NE	14338000	337
35	SAN LORENZO	11160500	276
36	CARBERRY CRE	14361700	180
37	YELLOWSTONE	6187550	3518
38	YELLOWSTONE	6186500	2606
39	YAKIMA RIVER	12474500	142
40	WILLOW CREEK	13058000	1659
41	MARTIS C NR	10339400	103
42	PAULINA CREE	14063300	45
43	KACHESS RIVE	12476000	164
44	HORSE CREEK	14159100	362

Table B.4. Continued

45	ROGUE RIVER	14335075	1784
46	SIMILKAMEEN	12442500	9100
47	YANKEE FORK	13296000	485
48	APPLEGATE RI	14362000	577
49	APPLEGATE RI	14366000	1252
50	APPLEGATE RI	14369500	1810
51	GARDNER RIVE	6191000	512
52	EAST BOULDER	6197800	101
53	E FK SALMON	13297453	487
54	BLACKS FORK	9224700	7721
55	CALAVERAS R	11308600	468
56	ARKANSAS RIV	7081200	252
57	LITTLE SNAKE	9260000	10478
58	BEAR CREEK A	13032000	203
59	SF SNOQUALMI	12143400	113
60	NORTH FORK Q	12039300	191
61	MALHEUR RIVE	13216350	6292
62	NO SANTIAM R	14178000	558
63	NORTH SANTIA	14181500	1171
64	NORTH SANTIA	14183000	1696
65	NORTH SANTIA	14184100	1892
66	SANTIAM R AT	14189000	4608
67	N F FLATHEAD	12355500	4026
68	ARKANSAS RIV	7083710	615
69	SPOKANE RIVE	12433000	16002
70	COWLITZ RIVE	14233400	2653
71	CHAMOKANE CR	12433200	447
72	S FK BOISE R	13186000	1660
73	MIDDLE FORK	7124050	130
74	MCKENZIE R A	14159000	903
75	MF WILLAMETT	14148000	2409
76	MIDDLE FORK	14150000	2605
77	MF WILLAMETT	14152000	3491
78	BIG BUTTE CR	14337500	641
79	TAYLOR ARROY	7126325	126
80	SKAGIT RIVER	12179000	3325
81	YAMPA RIVER	9260050	20461
82	SKAGIT RIVER	12181000	3604
83	STANISLAUS R	11299997	2536
84	STANISLAUS R	11303000	2862
85	FISHER RIVER	12302055	2173
86	CACHE LA POU	6752260	2968
87	CLARK FORK N	12323800	1285
88	S. UMPQUA RI	14312260	4567
89	MERCED R NR	11272500	3277
90	MCKENZIE RIV	14159800	1940
91	MCKENZIE RIV	14162400	2208

Table B.4. Continued

92	MCKENZIE R N	14162500	2401
93	MCKENZIE RIV	14163900	2793
94	ROGUE RIVER	14337600	2435
95	OKANOGAN RIV	12445000	11997
96	FOUNTAIN CR	7105530	1069
97	FOUNTAIN CRE	7105800	1314
98	FOUNTAIN CRE	7106000	1768
99	ROGUE RIVER	14338100	2819
100	ROGUE R AT D	14339000	3151
101	WEBER RIVER	10141000	5308
102	FOUNTAIN CRE	7106300	2198
103	FOUNTAIN CRE	7106500	2400
104	ARKANSAS RIV	7087200	1676
105	N UMPQUA RIV	14317500	2302
106	NORTH UMPQUA	14319500	3515
107	TRUCKEE R A	10346000	2417
108	TRUCKEE R AT	10348000	2746
109	TRUCKEE R NR	10348200	2785
110	TRUCKEE R AT	10350000	3700
111	TRUCKEE R AB	10350390	4044
112	TRUCKEE R BL	10350400	4120
113	TRUCKEE R RT	10350405	4120
114	TRUCKEE R AT	10350500	4146
115	TRUCKEE R NR	10351700	4692
116	TUOLUMNE R A	11290000	4950
117	WILLAMETTE R	14166000	8904
118	WILLAMETTE R	14171750	11604
119	JOHN DAY R A	14048000	19801
120	PURGATOIRE R	7124200	1308
121	SALMON RIVER	13293800	786
122	SALT RIVER A	13027500	2221
123	SKAGIT RIVER	12199000	7830
124	SKAGIT RIVER	12200500	8029
125	ARKANSAS RIV	7091200	2730
126	SATUS CR AT	12508621	1490
127	BOISE RIVER	13213000	10124
128	UMPQUA RIVER	14321000	9433
129	ROGUE RIVER	14359000	5312
130	VAN BREMER A	7126200	456
131	TONGUE RIVER	6308500	13979
132	S F TRINITY	11528700	1979
133	ROGUE R AT G	14361500	6364
134	BIRCH CR AT	13116980	672
135	WILLAMETTE R	14191000	18832
136	RUBY RIVER N	6023000	2520
137	GREEN RIVER	9234500	39153
138	YELLOWSTONE	6191500	6804

Table B.4. Continued

139	ROGUE RIVER	14370400	8561
140	ROGUE RIVER	14372250	9971
141	ROGUE RIVER	14372300	10198
142	FEATHER R A	11407000	9381
143	TETON RIVER	13055000	2271
144	CLARK FORK A	12324200	2595
145	WILLAMETTE R	14211720	28936
146	DOLORES RIVE	9169500	5257
147	OWYHEE RIVER	13184000	28693
148	WHITE RIVER	9304200	1662
149	WHITE RIVER	9304600	2096
150	DOLORES RIVE	9171070	5570
151	DOLORES RIVE	9171100	5573
152	WHITE RIVER	9304800	2655
153	WHITE RIVER	9306395	9219
154	WHITE RIVER	9306500	10109
155	WHITE RIVER	9306600	10170
156	WHITE R BLW	9306700	10453
157	WHITE RIVER	9306900	12933
158	OWENS R BL T	10277400	7773
159	BIG HOLE RIV	6025500	6409
160	SAN JUAN RIV	9368000	33271
161	YAKIMA RIVER	12508990	13887
162	YAKIMA RIVER	12510500	14544
163	PEND OREILLE	12398600	9590
164	CLARK FORK A	12324680	4600
165	YELLOWSTONE	6192500	9226
166	BIG HOLE RIV	6026400	7134
167	HORSE CREEK	7123675	3656
168	GUNNISON RIV	9152500	20452
169	BELT CREEK N	6090500	918
170	CLARKS FORK	6208800	5470
171	MADISON RIVE	6041000	5730
172	VIRGIN RIVER	9408135	3638
173	HUMBOLDT R N	10321000	11239
174	RIO GRANDE N	8251500	19393
175	ARKANSAS RIV	7094500	6373
176	JORDAN RIVER	10171000	9036
177	VIRGIN RIVER	9408150	3954
178	SOUTH PLATTE	6711565	8783
179	GALLATIN RIV	6052500	4630
180	BELT CREEK N	6090610	2069
181	SOUTH PLATTE	6720500	12347
182	SAN JUAN RIV	9379500	59578
183	S F FLATHEAD	12359800	3001
184	ARKANSAS RIV	7097000	10244
185	SAN JOAQUIN	11260815	53256

Table B.4. Continued

186	SAN JOAQUIN	11261500	54137
187	PURGATOIRE R	7126300	4998
188	FLATHEAD RIV	12363000	11524
189	FLATHEAD RIV	12372000	18332
190	SAN JOAQUIN	11274570	63059
191	COLORADO RIV	9070500	11378
192	CLARK FORK N	12331900	6864
193	SNAKE RIVER	13032500	13455
194	CLARK FORK A	12334550	9496
195	COLORADO RIV	9071100	11618
196	COLORADO RIV	9071750	11750
197	SAN JOAQUIN	11290500	69057
198	PURGATOIRE R	7126485	7184
199	ROARING FORK	9085000	3762
200	POWDER RIVER	6326500	33907
201	POWDER RIVER	6326520	34733
202	SAN JOAQUIN	11303500	72207
203	SEVIER RIVER	10224000	15371
204	SOUTH PLATTE	6764000	59143
205	PURGATOIRE R	7128500	8946
206	SNAKE RIVER	13037500	14818
207	CLARK FORK A	12340500	15587
208	ARKANSAS RIV	7099400	11948
209	CLARK FORK B	12353000	23346
210	ARKANSAS RIV	7099970	12254
211	CLARK FORK A	12353650	26447
212	CLARK FORK N	12354700	27997
213	GREEN RIVER	9315000	105289
214	YELLOWSTONE	6214500	30610
215	ARKANSAS RIV	7109500	16193
216	KLAMATH RIVE	11523000	31072
217	COLORADO RIV	9085100	15577
218	RIO GRANDE A	8313000	36145
219	BELLE FOURCH	6437000	15017
220	JEFFERSON RI	6036650	24759
221	COLORADO RIV	9095500	20686
222	SACRAMENTO R	11390500	39736
223	COLORADO RIV	9163530	46742
224	COLORADO RIV	9163500	46250
225	SACRAMENTO R	11447650	63854
226	DESCHUTES RI	14092500	20861
227	DESCHUTES RI	14103000	27783
228	MISSOURI RIV	6054500	38008
229	BIGHORN RIVE	6294700	59181
230	ARKANSAS RIV	7124000	37082
231	YELLOWSTONE	6309000	125021
232	MISSOURI RIV	6090800	63117

Table B.4. Continued

233	YELLOWSTONE	6329500	179003
234	MISSOURI RIV	6109500	87765
235	SNAKE RIVER	13077000	56656
236	SNAKE R NR M	13081500	62949
237	MISSOURI RIV	6177000	212274
238	ARKANSAS R N	7137500	64611
239	PECOS RIVER	8405200	54455
240	PECOS RIVER	8407000	59991
241	SNAKE RIVER	13154500	92931
242	SNAKE RIVER	13213100	152557
243	SNAKE RIVER	13269000	178362
244	CHANNEL A NE	5056410	5095
245	SHEYENNE RIV	5059000	24080
246	SHEYENNE RIV	5059400	24161
247	SOURIS RIVER	5114000	21198
248	SOURIS RIVER	5116000	22547
249	SOURIS RIVER	5120000	27680
250	GIBBON RIVER	6037000	296
251	MADISON RIVE	6037500	1126
252	HIGHWOOD CRE	6090720	315
253	EAST FORK PO	6179000	1834
254	BEAR DEN CRE	6332515	191
255	WHITE R NEAR	6452000	25852
256	JAMES RIVER	6470500	11008
257	JAMES RIVER	6470830	13296
258	MUD LAKE NR	6470985	14190
259	JAMES R AT C	6471000	18700
260	JAMES R AT A	6473000	25218
261	CANADIAN RIV	6619400	114
262	CANADIAN RIV	6619450	408
263	EF ARKANSAS	7079300	129
264	HALFMOON CRE	7083000	61
265	CHACUACO CRE	7126470	1086
266	COLORADO RIV	9034500	2135
267	MUDDY CREEK	9041500	749
268	EAST MIDDLE	9092850	57
269	EAST FORK PA	9092970	53
270	PARACHUTE CR	9093000	364
271	PARACHUTE CR	9093500	510
272	ROAN CREEK N	9095000	834
273	LEWIS WASH N	9106200	15
274	LEACH CREEK	9152650	38
275	ADOBE CREEK	9152900	40
276	BIG SALT WAS	9153270	367
277	REED WASH NE	9153300	74
278	WEST SALT CR	9153400	435
279	EAST SALT CR	9163310	509

Table B.4. Continued

280	MACK WASH NE	9163340	41
281	SALT CREEK N	9163490	1130
282	PINE CREEK A	9196500	196
283	BIG SANDY RI	9215550	2929
284	BIG SANDY RI	9216050	4491
285	SALT WELLS C	9216565	90
286	MIDDLE CREEK	9243700	61
287	FOIDEL CREEK	9243800	22
288	FOIDEL CREEK	9243900	45
289	SAGE CREEK A	9244415	11
290	WATERING TRO	9244460	11
291	HUBBERSON GU	9244464	21
292	GOOD SPRING	9250400	104
293	WILSON CREEK	9250507	52
294	TAYLOR CREEK	9250510	19
295	WILSON CREEK	9250600	71
296	JUBB CREEK N	9250610	20
297	MORGAN GULCH	9250700	69
298	YAMPA RIVER	9251000	8759
299	STEWART GULC	9306022	155
300	WILLOW CREEK	9306058	125
301	BLACK SULPHU	9306175	267
302	CORRAL GULCH	9306235	22
303	CORRAL GULCH	9306242	82
304	CORRAL GULCH	9306244	98
305	YELLOW CREEK	9306255	679
306	EVACUATION C	9306410	261
307	EVACUATION C	9306420	674
308	EVACUATION C	9306430	738
309	MANCOS RIVER	9370800	783
310	MANCOS RIVER	9370820	829
311	HARTMAN DRAW	9371400	87
312	MCELMO CREEK	9371500	598
313	MCELMO CREEK	9371520	606
314	RED BUTTE CR	10172200	19
315	BEAVER RIV A	10237000	792
316	SOUTH TWIN R	10249300	50
317	ALAMO R AT D	10254670	2978
318	NEW R AT INT	10254970	1471
319	MOJAVE R A L	10261500	1356
320	BIG ROCK CRE	10263500	59
321	LEVIATHAN C	10308783	11
322	LEVIATHAN C	10308790	21
323	BRYANT C BL	10308794	56
324	BRYANT C NR	10308800	83
325	MARYS RIVER	10313400	183
326	GENERAL C NR	10336645	20

Table B.4. Continued

327	WARD CREEK N	10336670	6
328	WARD CREEK T	10336672	4
329	WARD CREEK A	10336676	25
330	DONNER UND B	10396000	532
331	ARROYO TRABU	11047300	140
332	SAN DIEGO CR	11048500	105
333	LOS ANGELES	11103000	2144
334	LA RIV A WIL	11103010	2145
335	SANTA CLARA	11108500	1669
336	PIRU C NR PI	11110000	1112
337	SESPE CREEK	11111500	129
338	SESPE CREEK	11113000	650
339	SANTA PAULA	11113500	103
340	SATICOY DIV	11113900	1
341	JALAMA C NR	11120600	53
342	CANADA HONDA	11120900	30
343	SALSIPUEDES	11132500	122
344	ARROYO VALLE	11176600	577
345	BIG C BL HUN	11237000	185
346	FRESNO R NR	11257500	337
347	CHOWCHILLA R	11258980	521
348	SALT SLOUGH	11261100	800
349	MUD SLOUGH N	11262900	96
350	MERCED R A H	11264500	470
351	ORESTIMBA CR	11274538	393
352	MOKELUMNE R	11325500	1724
353	MF COTTONWOO	11374400	633
354	COTTONWOOD C	11375810	1025
355	STONY CREEK	11387000	1551
356	MF FEATHER R	11392500	1776
357	WB FEATHER R	11405300	291
358	YUBA RIVER N	11421000	3460
359	SF AMERICAN	11439500	499
360	SF AMERICAN	11445500	1743
361	AMERICAN R A	11446500	4922
362	EF RUSSIAN R	11461500	189
363	EF RUSSIAN R	11462000	271
364	BIG SULPHUR	11463200	221
365	RUSSIAN R NR	11464000	2050
366	NAVARRO R NR	11468000	786
367	EEL R AB DOS	11472500	1825
368	MF EEL R AB	11472800	528
369	SOUTH FORK E	11475500	114
370	REDWOOD C NR	11481500	175
371	REDWOOD C AT	11482200	479
372	REDWOOD C A	11482500	103
373	GRASS VALLEY	11525600	80

Table B.4. Continued

374	TRINITY RIVE	11525655	2100
375	SMITH R NR C	11532500	1588
376	WYNOOCHEE RI	12037400	465
377	NF SKOKOMISH	12056500	147
378	N.F. SKOKOMI	12059500	303
379	SOUTH FORK S	12060500	198
380	SKOKOMISH RI	12061500	593
381	NISQUALLY RI	12082500	359
382	NISQUALLY RI	12086500	773
383	GREEN RIVER	12113000	1073
384	N.F. SNOQUAL	12142000	166
385	TANK CR NR L	12197040	5
386	MINKLER CR N	12197110	14
387	BLACK CREEK	12197680	3
388	WISEMAN CR N	12197700	9
389	TOBACCO RIVE	12301300	1085
390	YAAK RIVER N	12304500	2043
391	ROCK CREEK N	12334510	2304
392	FISH CREEK B	12353450	628
393	ST. REGIS RI	12354000	788
394	STILLWATER R	12365000	1441
395	SWIFT CREEK	12365800	201
396	WHITEFISH RI	12366000	444
397	THOMPSON RIV	12389500	1648
398	BULL RIVER N	12391550	366
399	BIG CR AB E	12414350	101
400	OKANOGAN RIV	12439500	1469
401	SULPHUR CR W	12508850	435
402	FALLS RIVER	13047500	845
403	BOISE RIVER	13185000	2153
404	MORES CREEK	13200000	1029
405	PAYETTE RIVE	13238000	3079
406	PAYETTE RIVE	13247500	5744
407	MIDDLE FORK	13257000	221
408	REDFISH LAKE	13293900	111
409	MEADOW CREEK	13318050	97
410	MEADOW CREEK	13318060	128
411	MINAM RIVER	13331500	619
412	GEDNEY CREEK	13336300	125
413	WHITESAND CR	13336620	641
414	CROOKED FORK	13336630	438
415	N FK CLEARWA	13340600	3355
416	TWENTY ONE R	13342200	17
417	WHITE RIVER	14101500	1070
418	BULL RUN R N	14138850	125
419	FIR CREEK NE	14138870	13
420	NO FK BULL R	14138900	21

Table B.4. Continued

421	SOUTH FORK B	14139800	40
422	MIDDLE FORK	14144800	669
423	HILLS CR AB	14144900	137
424	FALL CR. NEA	14150300	19
425	COAST FORK W	14152500	188
426	S FK MCKENZI	14159200	407
427	SOUTH FORK M	14159500	538
428	BLUE R BL TI	14161100	118
429	LOOKOUT C NR	14161500	62
430	BLUE R AT BL	14162200	228
431	CALAPOOIA R	14172000	268
432	CALAPOOIA RI	14173500	958
433	LITTLE NORTH	14182500	287
434	SOUTH SANTIA	14185000	458
435	MIDDLE SANTI	14185800	268
436	QUARTZVILLE	14185900	256
437	SOUTH SANTIA	14187200	1443
438	SOUTH SANTIA	14187500	1645
439	THOMAS CREEK	14188850	25
440	SOUTH SANTIA	14188900	2693
441	TUALATIN RIV	14202500	126
442	TUALATIN RIV	14207500	1830
443	TILTON R ABV	14236200	360
444	GREEN R ABV	14240800	320
445	N.F. TOUTLE	14241100	736
446	NEHALEM RIVE	14301000	1744
447	NESTUCCA R N	14303600	471
448	BIG ROCK CRE	14304850	17
449	SILETZ RIVER	14305500	526
450	ALSEA RIVER	14306500	857
451	SIUSLAW R NR	14307620	1528
452	S. UMPQUA RI	14308600	1666
453	ROGUE RIVER	14330000	986
454	WEST BRANCH	14337870	40
455	ELLIOTT CREE	14361600	146

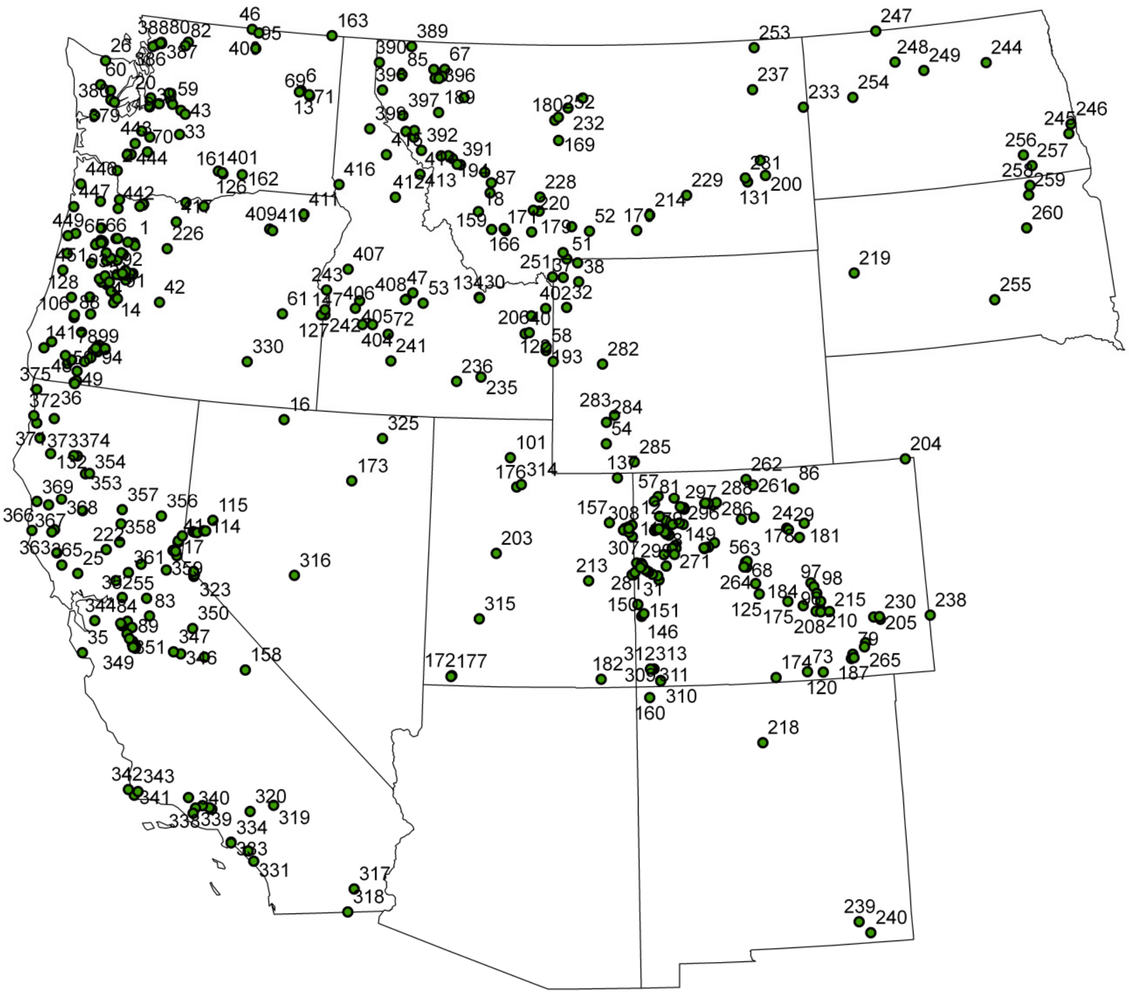


Figure B.1. Spatial distribution of temperature sites indexed according to the first column in Table B.4

Appendix C

Coauthor Approval Letters

UtahState
University
Department of Civil and Environmental Engineering
4110 Old Main Hill
Logan, UT 84322-4110
Telephone: (435) 797-2932
Fax: (435) 797-1185

1-11-2010

Ryan A. Hill
Western Center and Environmental Engineering Department,
Utah State University,
Logan, UT 84322

Dear Ryan,

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 January of 2010.

I am requesting your permission to include the attached paper, of which you are a coauthor, as a chapter in my dissertation. I will include acknowledgments 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,

Kiran J. Chinnayakanahalli

I hereby give permission to Kiran J. Chinnayakanahalli to use and reprint all of the material that I have contributed to Chapter 2, 3 and Appendix B of his dissertation.

Ryan A. Hill

UtahState
University
Department of Civil and Environmental Engineering
4110 Old Main Hill
Logan, UT 84322-4110
Telephone: (435) 797-2932
Fax: (435) 797-1185

1-11-2010

John R. Olson
Western Center and Environmental Engineering Department,
Utah State University,
Logan, UT 84322

Dear John,

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 January of 2010.

I am requesting your permission to include the attached paper, of which you are a coauthor, as a chapter in my dissertation. I will include acknowledgments 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,

Kiran J. Chinnayakanahalli

I hereby give permission to Kiran J. Chinnayakanahalli to use and reprint all of the material that I have contributed to Chapter 3 of his dissertation.

John R. Olson

CURRICULUM VITAE

Kiran J. Chinnayakanahalli
Utah Water Research Laboratory
Utah State University
8200 Old Main Hill
Logan, UT 84322-8200
(435) 797-2946
k.c@aggiemail.usu.edu

Education

Ph.D. Civil and Environmental Engineering, Utah State University, Logan, UT, Expected January 2010. Dissertation: Characterizing Ecologically Relevant Variance in Streamflow Regime. Advisor: David G. Tarboton.

M.S. Civil and Environmental Engineering, Utah State University, Logan, UT, 2004. Thesis: An objective method for the intercomparison of terrain stability models and incorporation of parameter uncertainty. Advisor: David G. Tarboton.

B.E. Environmental Engineering, University of Mysore, India. 1998.

Professional Experience

Research Assistant: 2000-Present, Utah Water Research Laboratory, Utah State University, Logan, UT.

Teaching Assistant: Probabilistic and Statistical Methods in Engineering, at Utah State University. Fall semester, 2005

Project Engineer: Department of environmental Engineering, SCJE, Mysore, India Apr - Nov, 1999.

Expertise

Hydrologic modeling
Digital Elevation Model Analysis.
Watershed classification.
Distributed and stochastic hydrologic modeling.
Fluvial geomorphology and terrain stability mapping.
Multivariate statistics for classification and prediction.
Stream ecology and environmental management.

Professional Society Memberships

American Geophysical Union (2003 – 2009).

Refereed Publications

Chinnayakanahalli, K. J, D. G. Tarboton, R. T. Pack. *An integral measure for evaluating the performance of spatial process predictor*. Targeted journal - Water Resources Research.

Chinnayakanahalli, K. J, D. G. Tarboton, C. P. Hawkins. *Characterizing ecologically relevant variation in streamflow regime.* Targeted journal - Freshwater Biology.

Chinnayakanahalli, K. J, D. G. Tarboton, R. Hill, J. Olson, C. P. Hawkins. *A tool to delineate multi watersheds spread across large Digital Elevation Models.* Targeted journal - Journal of Hydrology.

Chinnayakanahalli, K. J, D. G. Tarboton, C. P. Hawkins. *Prediction of hydrological classes of streams based on watershed attributes.* This is a manuscript in preparation. Targeted journal - Water Resources Research.

Other Publications, Reports, and Conference Proceedings

Chinnayakanahalli, K. J, C. Kroeber, R. Hill, J. Olson, D. G. Tarboton, and C. Hawkins. 2006. *Manual for regional watershed analysis using the multi-watershed delineation tool*. Utah State University. http://hydrology.neng.usu.edu/mwdtool/

Theses

Chinnayakanahalli, K. J., (2010), Characterizing Ecologically Relevant Variations in Streamflow Regime, PhD Dissertation, Utah State University, Logan, UT, 2004.

Chinnayakanahalli, K. J., (2004), An objective method for the intercomparison of terrain stability models and incorporation of parameter uncertainty, M.S. Thesis, Utah State University, Logan, UT, 144pp.

Conference Presentations, Posters, and Abstracts

Tarboton, D. G., K J Chinnayakanahalli, D P Ames, R Woods, 2001, "A Geographic Information System Tool for Hydrologic Model Setup", Presentation at AWRA/UCOWR specialty conference on "Decision Support Systems for Water Resources Management" - June 26 - 30, 2001 - Snowbird, Utah.

Chinnayakanahalli, K. J, D. G. Tarboton and R. T. Pack, 2003, "An Objective Method for the Intercomparison of Terrain Stability Models," Eos Trans. AGU, 84(47): Fall Meet. Suppl., Abstract H31C-0480.

Chinnayakanahalli, K. J, D. G. Tarboton and C. P. Hawkins, 2005, "Predicting Hydrologic Flow Regime for Biological Assessment at Ungauged Basins in the Western United States," Eos Trans. AGU, 86(52): Fall Meet. Suppl., Abstract H54C-04.

Chinnayakanahalli, K. J, D. G. Tarboton, J. Olson, R. Hill, and C. Kroeber. 2006. "A Tool to Delineate Watersheds and River Networks for Multiple Sites Spread over Large Digital Elevation Models". Hydrology Days, Colorado State University, Fort Collins.

Chinnayakanahalli, K. J, D. G. Tarboton, and C. Hawkins. 2007. "Predicting Streamflow Variables at Ungauged Sites Using Spatially-Referenced Regression Methods". Spring Runoff. April 5, Logan, UT.

Chinnayakanahalli, K. J, D. G. Tarboton, and C. Hawkins. 2007, "Classification of Watersheds for Bioassessment Based on Hydrological Variables", Eos Trans. AGU, 88(52): Fall Meet. Suppl., Abstract H53B-1233.

Chinnayakanahalli, K. J, D. G. Tarboton, and C. Hawkins. 2008, "Evaluating the use of streamflow regime for predicting taxonomic composition for use in bioassessment", Eos Trans. AGU, 89(53): Abstract H51A-0782.

Professional Activities

Reviewed for Journal of American Water Resources Associations.

Served on student committee for the Water Initiative program at Utah State University to organize a) the seminar series for 2007-2008 and b) the Spring Runoff Conference, 2008 and 2009.

Awards

J. N. Tata Endowment Scholarship for the year 2000-2001.

Awarded sixth rank in the Bachelor of Engineering, University of Mysore, India, 1998.

Best student oral presentation, Spring Runoff conference, 2008, Utah State University.