Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Hydrology 325 (2006) 241–261

Journal of Hydrology
www.elsevier.com/locate/jhydrol

Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon

Stephen C. Sanborn [1], Brian P. Bledsoe *

Department of Civil Engineering, Colorado State University, Fort Collins, CO 80523, USA

Received 16 August 2004; revised 24 June 2005; accepted 6 October 2005

Abstract

Streamflow prediction in ungauged basins provides essential information for water resources planning and management and ecohydrological studies yet remains a fundamental challenge to the hydrological sciences. A methodology is presented for stratifying streamflow regimes of gauged locations, classifying the regimes of ungauged streams, and developing models for predicting a suite of ecologically pertinent streamflow metrics for these streams. Eighty-four streamflow metrics characterizing various flow regime attributes were computed along with physical and climatic drainage basin characteristics for 150 streams with little or no streamflow modification in Colorado, Washington, and Oregon. The diverse hydroclimatology of the study area necessitates flow regime stratification and geographically independent clusters were identified and used to develop separate predictive models for each flow regime type. Multiple regression models for flow magnitude, timing, and rate of change metrics were quite accurate with many adjusted R^2 values exceeding 0.80, while models describing streamflow variability did not perform as well. Separate stratification schemes for high, low, and average flows did not considerably improve models for metrics describing those particular aspects of the regime over a scheme based on the entire flow regime. Models for streams identified as 'snowmelt' type were improved if sites in Colorado and the Pacific Northwest were separated to better stratify the processes driving streamflow in these regions thus revealing limitations of geographically independent streamflow clusters. This study demonstrates that a broad suite of ecologically relevant streamflow characteristics can be accurately modeled across large heterogeneous regions using this framework. Applications of the resulting models include stratifying biomonitoring sites and quantifying linkages between specific aspects of flow regimes and aquatic community structure. In particular, the results bode well for modeling ecological processes related to high-flow magnitude, timing, and rate of change such as the recruitment of fish and riparian vegetation across large regions.

Keywords: Flow regime; Prediction; Stratification; Regression analysis; Aquatic environment; Cluster analysis

1. Introduction

Predicting streamflow behavior in ungauged (i.e. no streamflow data available) basins is a significant challenge facing the hydrologic sciences. Accurate estimates of hydrologic variables at ungauged sites

* Corresponding author. Tel.: +1 970 491 8410; fax: +1 970 491 8671.
E-mail addresses: ssanborn.02@engr.colostate.edu (S.C. Sanborn), bbledsoe@engr.colostate.edu (B.P. Bledsoe).
[1] Tel.: +1 970 491 8260; fax: +1 970 491 8671.

0022-1694/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2005.10.018

are not only important for water resources planning and management issues related to yield, storage, and extreme events, but are increasingly germane to ecological studies across a wide range of spatial and temporal scales. Streamflow regime encompasses the magnitude, timing, duration, and frequency of high and low flows, the rate of change of streamflow, and inter-annual variation and is increasingly cited as a 'master variable' that structures aquatic ecosystems and habitats (Poff and Ward, 1989; Richter et al., 1996; Poff et al., 1997; Baron et al., 2002). The specific hydrologic characteristics of greatest ecological relevance, however, remain largely unknown for most biotic communities. Moreover, a recent movement towards standardized biological monitoring at probabilistically selected stream and river sites (e.g. U.S. Environmental Protection Agency (USEPA), 2002) is providing unprecedented opportunities to examine biological-physical associations across broad geographic areas and hydrologic gradients. Because most biomonitoring sites are ungauged and there is substantial uncertainty regarding which hydrologic characteristics explain the observed biological variation, there is a burgeoning interest in the prediction of ecologically relevant hydrologic metrics for these sites. The need to generate a broad suite of hydrologic metrics at biological monitoring sites spanning large regions in the western United States (U.S.) provided the primary impetus for the work we present herein.

The literature is replete with studies focused on predicting selected streamflow attributes for ungauged streams (e.g. Thomas and Benson, 1970; Vogel et al., 1999; Chiang et al., 2002b). The typical approach to streamflow prediction in ungauged basins across large regions is to delineate geographic areas of similar streamflow pattern, use regression to relate watershed characteristics to streamflow variables, and assess model reliability (e.g. Jennings et al., 1994). In the western U.S., various processes control streamflow across multiple scales, hydroclimatic regions, and marked elevational gradients. The identification of distinct streamflow regime types at ungauged sites is critical for stratifying key processes and developing robust predictive models for each flow regime type. Selecting geographically contiguous regions of flow regime types from previously delineated regions (e.g. Thomas and Benson, 1970; Vogel et al., 1999) or examining residuals of regression Eqs. (Jennings et al., 1994; Tucci et al., 1995) facilitates stratification of ungauged streams. Cluster analysis can be used to objectively define geographically independent streamflow groups across regions (Hawley and McCuen, 1982; Haines et al., 1988; Burn, 1989; Hughes and James, 1989; Poff and Ward, 1989; Burn and Boorman, 1993; Kresch, 1993; Poff, 1996; Harris et al., 2000) and has been coupled with discriminant analysis to identify streamflow cluster membership of ungauged streams using catchment characteristics (Chiang et al., 2002a,b). Regression has been used within defined regions or strata to predict the mean and variance of flows, autocorrelation, flood durations and volumes (Thomas and Benson, 1970), parameters of statistical distributions to reproduce flow duration curves (e.g. Smakhtin et al., 1997; Fennessey and Vogel, 1990; Sugiyama et al., 2003), direct estimates of flood quantiles (Thomas and Benson, 1970; Surian and Andrews, 1999; Pitlick, 1994; Jennings et al., 1994), low flows (Ries and Friesz, 2000), and parameters of time series models to generate synthetic streamflows (Chiang et al., 2002a,b). Flow duration curves and flood quantile estimates provide valuable information about flow magnitudes and frequencies but cannot provide information about timing, a critical variable in many ecological studies. Synthetic monthly streamflows generated from a time series model include timing but omit information contained in metrics which characterize flows at the event or daily time scale.

In this paper, our purpose is to present: (1) a validated stratification scheme for identification and subsequent classification of streamflow regime types for ungauged sites, and (2) models developed for providing accurate estimates of a broad range of streamflow metrics of ecological interest at ungauged biological monitoring sites in 18 ecoregions (Omernik, 1995) of Colorado (CO), Oregon (OR), and Washington (WA). The disconnectedness and acute hydroclimatic variability of the study region and the prediction of unique streamflow metrics characterizing the entire flow regime provide practical insights regarding the efficacy of streamflow prediction for ecological studies across heterogeneous areas. In addition, we examine the utility of using different clustering schemes for high, low, and average flows and the effects of grouping streams with similar regimes but from different geographic regions into the same flow regime cluster. The conceptual framework and resulting models are useful for a variety of water resources and watershed management applications, large-scale biomonitoring studies, and understanding the linkages between streamflow, climate, and physical catchment characteristics.

2. Background

2.1. Study area

The USEPA Environment Monitoring and Assessment Program (EMAP) has collected physical and biological data from several hundred streams across the western U.S. (USEPA, 2002). A total of 286 EMAP sites in CO, WA, and OR were selected to develop and test a hierarchical, process-based river classification to improve stratification procedures for modeling biological variation. Nearly all of these sites occur on ungauged streams, leaving an incomplete picture of the hydrologic processes structuring biotic communities. To develop statistical models for streamflow at these ungauged sites, 162 U.S. Geological Survey (USGS) high quality streamflow gauges free of significant deviations from the natural streamflow pattern and with record lengths greater than 20 years were identified in CO, WA, and OR. Most of these gauges are part of the Hydroclimatic Data Network (HCDN; Slack et al., 1993) with additional reference gauges identified in CO by Surian and Andrews (1999) and in WA by Kresch (1993). The resulting set of gauged sites includes watersheds of many different sizes, elevations, climatic characteristics, and geologic types. Fig. 1 shows the major geographic features of the study area. Catchment area ranges from 4 to 19,632 km^2 and average precipitation is highly variable with 4500 mm per year falling in the North Fork of the Quinault basin in the Olympic Mountains to 335 mm per year falling in the Crab Creek basin in eastern WA. Gauge elevations range from near sea level to 3180 m for the Michigan River in CO.

The three states span several climatic zones and 18 ecoregions (Omernik, 1995). Along the Pacific coast, large cyclonic frontal storms frequently occur in winter generating 75–80% of the yearly precipitation (Redmond and Koch, 1991), and high streamflows with lower flows from April to October. These storms often produce snow at higher elevations in the Cascade Range. Generally, the rain-on-snow zone exists between 400 and 1500 m (Tague and Grant, 2004), where accumulated snow can be completely melted by a subsequent storm, generating very high flows. Above 1500 m, snow usually persists through

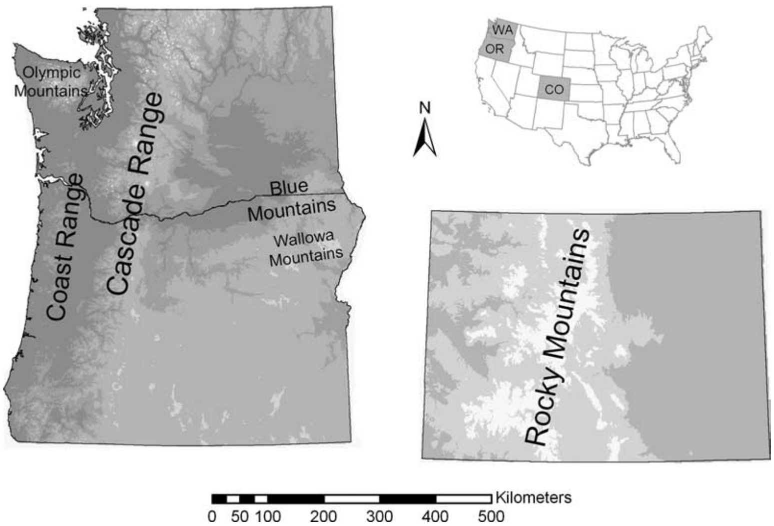

Fig. 1. Elevation throughout the study area is shown with lighter shades corresponding to higher elevations. Significant geographic features are labeled.

the winter season with the east side of the Cascades receiving less snow than the west due to orographic effects (Rasmussen and Tangborn, 1976). Low-flow volumes and recession characteristics on the western slope of the OR Cascades are dominated by geology (Tague and Grant, 2004). Eastern WA and OR lie in the rain shadow of the Cascades and receive much less precipitation. The higher elevations of the Blue and Wallawa Mountains receive greater amounts of precipitation relative to the semi-arid eastern uplands and build a winter snowpack (Western Regional Climate Center (WRCC) a,b) leading to elevated spring flows. Higher monthly flows for eastern OR and WA typically occur in the spring and summer due to precipitation and snowmelt. Peak flows there can be generated from a variety of processes including winter rainstorms, rain-on-snow events, snowmelt, or cloud-burst thunderstorms (Watershed Professionals Network, LLC, 2001). The mountains of CO accumulate a winter snowpack with runoff usually commencing in May and lasting into July with low flows from August to April. Summer convective thunderstorms and monsoonal events infrequently produce very large floods. The diversity of processes generating streamflow in these areas makes regionalization difficult yet essential to the development of models for predicting ecologically relevant streamflow parameters.

3. Methods

3.1. Streamflow and watershed metrics

A total of 84 streamflow metrics were computed to characterize the range of flow regimes in the study region. The Indicators of Hydrologic Alteration© (IHA; Richter et al., 1996) software was used to compute 33 streamflow metrics and their coefficients of variation using daily streamflow data obtained from the USGS. Olden and Poff (2003), in a study analyzing 171 published flow metrics, suggested that IHA sufficiently characterizes most flow regimes but nevertheless identify additional metrics that may add significant information. A subset of these additional metrics was selected for the flow regime types within the study region based on the recommendations of Olden and Poff (2003), physical interpretability, and economy of computation. These metrics were classified as describing high, low, or average streamflow magnitudes for use in the cluster analysis. A complete description of the final set of 84 streamflow metrics is provided in Table 1.

A Geographic Information System (GIS) was used to compute the monthly averages of climatological variables including precipitation, evapotranspiration, snowfall, temperature, and solar radiation. Current PRISM (Daly et al., 1994) data sets were used for precipitation and snowfall. Evapotranspiration data were acquired from Hobbins et al. (2001a,b) along with the spatially distributed temperature and solar radiation data sets they used. Physical properties of drainage basins are also important for predicting certain aspects of the streamflow regime and a review of streamflow estimation studies revealed many relevant physiographic basin characteristics. For instance, flood magnitudes and durations in larger basins are controlled by the slope and density of the drainage network (Pitlick, 1994). Variables utilized in this study include basin elevation, slope, relief, and soil properties. Geologic differences among sites were accounted for using a simple classification of unconsolidated, sedimentary, volcanic, and crystalline rock types. A complete list of the climatic and physiographic metrics is provided in Table 2.

3.2. Regionalization procedure

Such a hydroclimatically variable study area makes stratification crucial to developing accurate and interpretable models for streamflow metrics. Based on the hypothesis that optimal groupings of sites vary with different types of response variables, separate clustering schemes were developed using only high, low, and monthly flows, as well as the complete set of flow metrics. The high- and low-flow clustering schemes used the average monthly flow variables in addition to the respective high- or low-flow metrics (i.e. for low flows the Mn 1,2,7,30,90d metrics, NLoPl, Ml22, Fl3, etc.). The average, or monthly, scheme used only average monthly flows and the ‘all flow’ scheme encompassed all metrics. Three general statistical tools were used to create and subsequently assign ungauged sites to high-, low-, monthly-, and all-flow groupings. First, principal components analysis was used to reduce the streamflow variable set to variables explaining the most

Table 1
Streamflow regime metrics

Flow metric	Description	Units	Source	Pertubation
MAR	Mean annual runoff	m^3/s	Sanborn and Bledsoe	0
Flash	Avg annual 1-day maximum/average flow over all years	–	Sanborn and Bledsoe	0
Skew	Skewness of daily flows	–	Sanborn and Bledsoe	0
Avg_Oct[a]	Average October flow	m^3/s	IHA	0
Avg_Nov[a]	Average November flow	m^3/s	IHA	0
Avg_Dec[a]	Average December flow	m^3/s	IHA	0
Avg_Jan[a]	Average January flow	m^3/s	IHA	0
Avg_Feb[a]	Average February flow	m^3/s	IHA	0
Avg_Mar[a]	Average March flow	m^3/s	IHA	0
Avg_Apr[a]	Average April flow	m^3/s	IHA	0
Avg_May[a]	Average May flow	m^3/s	IHA	0
Avg_Jun[a]	Average June flow	m^3/s	IHA	0
Avg_Jul[a]	Average July flow	m^3/s	IHA	0
Avg_Aug[a]	Average August flow	m^3/s	IHA	0
Avg_Sep[a]	Average September flow	m^3/s	IHA	0
Mn1d[a]	Average annual 1-day minimum flow	m^3/s	IHA	0
Mn3d[a]	Average annual 3-day minimum flow	m^3/s	IHA	0
Mn7d[a]	Average annual 7-day minimum flow	m^3/s	IHA	0
Mn30d[a]	Average annual 30-day minimum flow	m^3/s	IHA	0
Mn90d[a]	Average annual 90-day minimum flow	m^3/s	IHA	0
Mx1d[a]	Average annual 1-day maximum flow	m^3/s	IHA	0
Mx3d[a]	Average annual 3-day maximum flow	m^3/s	IHA	0
Mx7d[a]	Average annual 7-day maximum flow	m^3/s	IHA	0
Mx30d[a]	Average annual 30-day maximum flow	m^3/s	IHA	0
Mx90d[a]	Average annual 90-day maximum flow	m^3/s	IHA	0
ZeroD[a]	Number of days per year with zero flow	–	IHA	0
BaseQ[a]	7-day minimum flow divided by mean flow for that year	–	IHA	0
DatMn[b]	Julian date of the minimum flow	–	IHA	0
DatMx[b]	Julian date of the maximum flow	–	IHA	0
NLoPl[a]	Average number of low pulses, low pulse defined as 1 standard deviation above the mean	–	IHA	1
DLoPl[a]	Average duration of low pulses	–	IHA	1
NHiPl[a]	Average number of low pulses, low pulse defined as 1 standard deviation below the mean	–	IHA	0
DHiPl[a]	Average duration of high pulses	–	IHA	0
RiseR[a]	Rise rate − mean of all positive differences	$m^3/s/day$	IHA	0
FallR[a]	Fall rate − mean of all negative differences	$m^3/s/day$	IHA	2300, 2[c]
Revs[a]	Number of flow reversals	–	IHA	0
Ma3	Coefficient of variation of daily flows	–	Olden and Poff	0
Ma40	(Mean monthly flow − median monthly flow)/median monthly flow	–	Olden and Poff	0.5
Ma41	Mean annual runoff divided by catchment area	cm	Olden and Poff	0
Ma44bs	Average variability in daily flows divided by median daily flows for each year, where variability is calculated as 90th–10th percentile	–	Sanborn and Bledsoe	0
Ml13	CV in minimum monthly flows	–	Olden and Poff	0
Ml14	Mean of lowest annual daily flow divided by median annual daily flow averaged across all years	–	Olden and Poff	0
Ml22	Mean annual minimum flows divided by catchment area	$m^3/s/km^2$	Olden and Poff	0
Mh1	Max monthly flow for Oct	m^3/s	Olden and Poff	0
Mh8	Max monthly flow for May	m^3/s	Olden and Poff	0
Mh17	Mean of 25th percentile from the flow duration curve divided by median daily flow across all years	–	Olden and Poff	0
Fl3	Total number of low flow spells (threshold equal to 5% of mean daily flow) divided by record length in years	–	Olden and Poff	0.01
Fh11	Mean number of discrete flood events per year	–	Olden and Poff	0

(*continued on next page*)

Table 1 (*continued*)

Flow metric	Description	Units	Source	Pertubation
Dl13	Mean annual 30-day minimum divided by median discharge	–	Olden and Poff	0
Dh12	Mean annual 7-day maximum divided by median discharge	–	Olden and Poff	0
Dh13	Mean annual 30-day maximum divided by median discharge	–	Olden and Poff	0
Th3	Max proportion of the year (num days/365) during which no floods have ever occurred over the period of record	–	Olden and Poff	0.5
BS1	(Sum(Abs($Q_{t+1}-Q_t$))/#days in record)/ Average flow over all years	–	Sanborn and Bledsoe	0

[a] Indicates CV of that metric was computed.
[b] Indicates standard deviation of that metric was computed.
[c] The two numbers are the pertubations for the average and CV, respectively.

variance in the data. Cluster analysis was then used to identify distinct flow regime types. Once flow regime clusters were established, discriminant analysis was used to predict the flow regime type of an ungauged site so that the appropriate set of cluster-specific regression models could be applied. A brief description of these statistical methods follows.

Principal components analysis is an ordination technique that creates linear combinations of variables to construct uncorrelated vectors (principal components, or PCs) that describe the dominant patterns of variance in the data (Joliffe, 1986). The PCs are mutually orthogonal, and therefore, uncorrelated with the first few ideally containing the most information. The variables that exhibit the highest loadings on a PC best explain that dimension of the data. This method is effective for identifying which variables explain the most variance, and therefore, contain the most information especially for data sets with a large number of variables.

To remove scale dependence from the streamflow metrics, monthly flows were expressed as a percent of mean annual runoff (Haines et al., 1988; Riggs and Harvey, 1990) and metrics describing flow magnitudes were normalized by catchment area. Data were standardized for use in the principal components analysis and cluster analysis to ensure each variable had equal weight. Principal components analysis was performed using the correlation matrix for the high-, low-, and all-flow variable sets. The number of variables chosen from each PC was proportional to the variance explained by that PC. This yielded a set of significant, non-redundant streamflow metrics for each flow variable set to use in the cluster analysis. It is important to choose a sound set of input variables for the cluster analysis because redundant variables can collectively bias the solution towards over-represented aspects of the flow regime.

Cluster analysis is used to find inherent groupings within data sets. Distances between data points and cluster centroids are evaluated and the observation assigned to the appropriate cluster. Hierarchical cluster analysis produces a dendrogram, or tree-like solution that depicts smaller clusters branching from the larger group to which they belong (Jobson, 1992). Partitioning methods such as k-means cluster analysis (Jobson, 1992) allow the number of clusters to be set and then finds the most effective clustering solution. Seeds are used to define the cluster centers initially and sites can be assigned to the nearest centroid; the seeds are then updated and the process is iteratively repeated. We utilized PROC FASTCLUS in SAS© (2001) to execute a k-means cluster analysis for each flow variable set. First, a very coarse solution of 20 clusters was computed and outliers were identified. Clusters with more than five members were then used as seeds for the next solution. The solution was refined to the specified number of clusters with the constraint that each cluster contained at least five observations. Canonical discriminant analysis was then performed and cases, identified by cluster, were plotted on the first two canonical axes to visually interpret the statistical quality of the solutions. A good cluster solution shows respondents in the same cluster closely spaced and far from other clusters (Hair and Black, 2000). The optimal number of clusters can be chosen (Milligan and Cooper, 1985) by comparing statistics

Table 2
Drainage basin characteristics

Climatic					
Metric	Description	Units	Data	Pertubation	Period
prcp_mo	Mean total monthly precipitation	mm	PRISM	0	1971–2000
srad_mo	Mean monthly solar radiation	$KJ/m^2/day$	Hobbins et al.	0	1952–1994
avgt_mo	Average monthly temperature	°C	Hobbins et al.	30	1953–1994
mint_mo	Average minimum monthly temperature	°C	Hobbins et al.	30	1954–1994
snow_mo	Average total monthly snowfall	mm	PRISM	1	1971–2000
et_mo	Average monthly potential evapotranspiration	mm	Hobbins et al.	4.5	1962–1988
precipr1	Ratio of precipitation in wettest month to that of the driest	–	PRISM	0	
precipr2	Ratio of precipitation in the wettest 3 months to the driest 3 months	–	PRISM	0	
precip_t	Average annual precipitation	mm	PRISM	0	1971–2000
snow_t	Average annual snowfall	mm	PRISM	1	1971–2000
et_t	Average annual potential evapotranspiration	mm	Hobbins et al.	0	1962–1988
srad_t	Average annual solar radiation	$KJ/m^2/day$	Hobbins et al.	0	1952–1994
nt_ndjfm	Average minimum temperature during Nov, Dec, Jan, Feb, Mar	°C	Hobbins et al.	22.9	1953–1994
snow_ppt	Ratio of total annual snowfall to total annual precipitation	–	PRISM	0.1	

Physiographic			
Metric	Description	Units	Pertubation
avg_elev	Average basin elevation	m	0
min_elev	Minimum basin elevation	m	1
topo_wet	Average basin topographic wetness—ln(flow accumulation area) + ln[1/tan(slope)], where area is in square meters and slope is in radians	–	0
drainden	Drainage density	m^{-1}	0
urban_p	Percentage of basin that is urban	–	0.01
shed_slp	Average slope of the basin	m/m	0
da	Drainage area	km^2	0
relief_r	Basin relief divided by its length	–	0
slp_elon	Ratio of the slope of the basin to the elongation of the basin Elongation = diamater of a circle of the same area as the basin	m^{-1}	0
store_p	Percentage of basin that is lakes or other water storage	–	0.01
chan_slp	Average channel slope	m/m	0
aspect	Average aspect	Degrees	0
drain_cl	Area weighted soil drainage class	–	0
n_br_dep	Average area weighted minimum depth to bedrock	m	0
forest_p	Percentage of basin that is forested	–	0
for_urb	Ratio of forested percentage to urban percentage	–	0

(*continued on next page*)

Table 2 (*continued*)

Physiographic

Metric	Description	Units	Pertubation
uncons	Percentage of basin underlain by unconsolidated geologic type	–	0.01
sedim	Percentage of basin underlain by sedimentary geologic type	–	0.01
volcan	Percentage of basin underlain by volcanic geologic type	–	0.01
cryst	Percentage of basin underlain by crystalline geologic type	–	0.01

_mo indicates the metric is computed for each month.

such as the cubic clustering criterion (CCC; Sarle, 1983), pseudo F-statistic, and root-mean-square standard deviation (Chiang et al., 2002a), or by using a stopping rule (Wong and Schaack, 1982; Ratowsky, 1984). However, the interpretability of the solution was used as the primary decision factor (Poff, 1996; Haines et al., 1988), in combination with the CCC and pseudo F-statistic. Plots of the average monthly streamflows and clustering metrics for each cluster along with maps of the clusters were scrutinized to understand the solutions and make final judgments on the number of clusters.

With distinct flow regime classes determined, discriminant analysis was used as a classification tool to assign *ungauged sites* to the streamflow clusters derived using cluster analysis based on watershed characteristics. Twelve watersheds used in the cluster analysis to create the streamflow clusters extend into Canada, so these were not used in the discriminant analysis or regression analyses as the necessary watershed and climatic data were not readily available for these sites. Discriminant analysis constructs linear combinations of variables called discriminant functions and finds those that can most effectively partition the predefined groups (Hand, 1981). The watershed data were natural-log transformed to bring data into closer compliance with the assumption of normality and subjected to a principal components analysis to identify the most explanatory variables for the discriminant analysis. Merging similar variables having high loadings on the same PC created several new metrics; average minimum winter temperature (nt_ndjfm), for example. For discriminant analysis, there should be at least four observations per predictor per group (Hand, 1997) so a subset of variables thought to be significant for reflecting cluster differences was entered in SAS's PROC STEPDISC, a step-wise discriminant analysis procedure, to identify the best discriminators. Correlations between the best discriminators were checked and the best four with Pearson correlations less than 0.75 were used in the PROC DISCRIM procedure. Quadratic, rather than linear, discriminant functions were generated to allow for differences in group size. Cross-validation was used to compute classification error rates.

3.3. Multiple regression analysis

Multiple regression analysis (MRA) was used to derive, from gauged sites, equations relating streamflow metrics to physical and climatic basin characteristics within streamflow groups. A log-linear model of the following form is generally used in regional regression studies:

$$Q = kX_1^{\beta_1}X_2^{\beta_2}\ldots X_n^{\beta_n}\nu \tag{3.1}$$

where Q is the streamflow metric of interest; X_i, 1,…, n are basin characteristics; k, β_i, 1,…, n are model parameters; and ν are the lognormally distributed model errors. When the natural log of this model is taken, it becomes linear of the form:

$$\ln(Q) = \beta_0 + \beta_1\ln(X_1) + \beta_2\ln(X_2) + \cdots + \beta_n\ln(X_n) + \varepsilon \tag{3.2}$$

where $\beta_0 = \ln(k)$ and $\varepsilon = \ln(\nu)$ are normally distributed residuals with mean zero and variance σ_ε^2.

Error can be partitioned into (1) model error from parameter estimation and model selection, (2) measurement error, which includes the cross correlations between sites, and (3) sampling error, which deals with the varying record lengths of sites. Ordinary least squares (OLS) regression only accounts for model error, ignoring both other types. Tasker (1980) introduced weighted least squares (WLS), which accounts for model error and sampling error and Stedinger and Tasker (1985) presented generalized least squares (GLS) which accounts for all three types of error. We used WLS with the weighting scheme utilized by Vogel et al. (1999):

$$w_i = \frac{n_i}{\sum_{j=1}^{m} n_j} \tag{3.3}$$

where w_i is the weight for a site i; n_i is the record length in years for site i; and m is the number of sites in the cluster.

The PROC REG procedure in SAS© (2001) was used for the MRA. All data were perturbed so values were greater than zero and then natural-log transformed. Model selection algorithms in SAS limit the pool of independent variables to approximately the size of the number of observations before collinearity errors abound. The pool of predictors was narrowed to around 50, using the variables we figured would best stratify the flow regimes. Correlation matrices were then reexamined and variables eliminated until collinearity issues were resolved. Monthly precipitation and snow variables were usually reduced to a few months that adequately represent the seasonal patterns, and other variables were eliminated by understanding which variables represent dominant processes driving streamflow within the particular flow regime group. Variables eliminated due to high correlations with others are still largely represented by the variable retained, so minimal information is lost. Generally, it is desirable to have at least 10 observations per independent variable used in a multiple regression model (Hirsch et al., 1993). With the pool of possible independent variables in place, Mallows' C_p was used to select the best set of models given a constrained number of independent variables. This set was further examined and the choice of the final model was made from best judgment based on physical interpretability, stipulating primarily that the variance inflation factor (VIF), a measure of collinearity between independent variables, be less than 10 (Hirsch et al., 1993) and secondarily that the adjusted R^2 not be greater than 0.02 less than the maximum. Adjusted R^2 values provide a more conservative estimate of model fit because it accounts for the number of predictors relative to the number of observations.

Model robustness was evaluated using the PRESS (prediction residual sums of squares) statistic. In this procedure, one observation is removed from the data and the regression model refit. The difference between the predicted value from the model and actual value at that point is the prediction residual. This is done for all points and PRESS is the sum of squares of these prediction residuals. PRESS can also be used as a model selection criterion (Hirsch et al., 1993) and generally agrees with Mallows' C_p selections (Gilroy and Tasker, 1990). Model error can be assessed by examining the root-mean-square error (RMSE) of the model. Prediction intervals, the error associated with making a prediction of a single observation, can be computed as:

$$\mathrm{SE} = \sqrt{\exp[s_\mathrm{e}^2(1 + p/m)] - 1} \tag{3.4}$$

where s_e^2 is the variance of the residuals in the estimated model ($s_\mathrm{e}^2 = (\mathrm{RMSE})^2$) and is a measure of model error; p is the number of independent variables in the model; m is the number of sites in the cluster; and ps_e^2/m is the average sampling error (Vogel et al., 1999).

Model accuracy was assessed by examining both adjusted R^2 and SE values. All values of R^2 reported hereafter are measures of adjusted R^2.

4. Results

Regression models were selected for each cluster contingent on process-based interpretation and the statistical criteria described above. Many regression models performed well using any of the regionalization schemes. For the high-, low-, and average-clustering schemes 39, 36, and 39%, respectively, of the metrics for which models were derived had values of R^2 greater than 0.80 for all flow regime types. However, comparing the R^2 values for individual metrics using the appropriate high-, low- or monthly-flow classifications versus the 'all-flow' stratification indicated that these schemes did not provide significant overall improvement. The average R^2 value decreased by 0.01 for models in the snowmelt cluster and 0.03 for rain and snow cluster models. Rain cluster models improved by an average R^2 of 0.09 and the variable stream models showed a marked 0.12 decrease in R^2. These disparities can be attributed to differences in sample size that result in allowing more independent variables into the rain models and less into the models for variable streams. R^2 values for monthly flows were generally not affected by the clustering scheme used while model R^2 values describing flow magnitude variability were poorer. Accordingly, the 'all-flow' clustering scheme was used to develop models for all streamflow metrics and the following sections focus exclusively on these results.

4.1. Variable reduction and streamflow regime stratification

Principal components analysis was performed for all streamflow metrics to identify the variables describing the most variance in the data (Table 3). Each PC can be said to describe a specific 'dimension' of variability among flow regimes. The first PC (PC1), which accounts for 27.5% of the total variance, can be interpreted as a high-flow axis. Mean annual runoff per drainage area (Ma41) is biased towards high flows, average November flow (PMAR_NOV) corresponds to high flow for streams along the Pacific coast, and the number of high pulses (NHiPl) and maximum 90 day flow (Mx90d) are direct measures of high-flow events. For PC2, average March flow (PMAR_Mar) and the coefficient of variation (CV) of January average flows (CV_Jan) are positively loaded while average July flow (PMAR_Jul) has a negative loading. This PC represents flow variability, as flow regimes with high flows in March tend to be flashier with higher coefficients of variation of monthly flows, those with high flows in July have lower coefficients of variation, and CV_Jan is a direct measure of flow variability that is correlated with the overall variability. PC3 and PC4 are best interpreted as characterizing low-flow magnitudes and low-flow variability, respectively.

The metrics identified by the principal components analysis were used in the cluster analysis and a four-cluster solution was selected based primarily on the physical interpretability with attention to clustering statistics. Also, the discriminant analysis classification error rate for the five-cluster solution was considerably higher. Fig. 2 shows the result from a canonical discriminant analysis of the cluster solution. This procedure is a dimension reduction technique similar to principal components analysis, except it identifies vectors, which describe the most *between-group* variation (SAS©, 2001). The axes, canonical variables 1 and 2, have the same hydrologic interpretation as PC1 and PC2, respectively. The groups are quite distinct in this plot, with little overlap. Cluster 3 loads highest on the high-flow axis followed by clusters 2 and 4. Cluster 4, the most diffuse cluster, clearly has the highest loading on the variability axis with the other three clusters having similarly low values.

The cluster solution has a strong geographic relation, as evidenced by a map of the gauges marked by streamflow cluster (Fig. 3), but is not entirely geographically dependent. Sites in clusters 2 and 3 are located predominately along and west of the Cascade crest with cluster 3 watersheds located mostly on the Olympic Peninsula and western slope of the North Cascades. Cluster 4 watersheds encompass most of eastern WA and OR with two additional sites in western CO. Cluster 1 watersheds cover nearly all of CO and much of the North Cascades. Plotting the average of the monthly flows for the streams of each cluster (Fig. 4) provides a general idea of the streamflow regime type associated with each cluster. Cluster 1 exhibits a strong snowmelt signal with high flows from April to July and low flows the rest of the year. These watersheds are located in the higher elevations of the CO Rockies and North Cascades, and one watershed is in the Wallowa Mountains. High winter flows define the cluster 2 signal, a result of the consistent cyclonic frontal storms bringing winter rains to the coastal regions. A bit of influence from snow can be detected from the persistence of the high flows into the late spring. Cluster 3 also shows evidence of these frontal rains with high winter flows,

Table 3
Variables with the highest loadings on the first four principal components, with the percent of variation explained by each PC in parentheses

PC 1 (27.5%)	PC 2 (21.1%)	PC 3 (11.8%)	PC 4 (6.9%)
Ma41	PMAR_Mar	BaseQ	CV_Mn30d
PMAR_Nov	PMAR_Jul	M113	
NHiPl	CV_Jan		
Mx90d			

Descriptions of the streamflow variables are provided in Table 1.

but has a smaller secondary peak in May due to snowmelt. Clusters 2 and 3 have similar monthly flow signals, but when individual streamflow metrics used in the cluster analysis are examined, their differences become more evident. Cluster 3 has higher magnitude long-term high flows (Mx90d), higher basin runoff production per drainage area (Ma41), more high-flow pulses (NHiPl), and somewhat less low-flow variation (CV_Mn30d). Cluster 4 has a general increase in flow beginning in November with a peak in May. This signal is the result of winter and spring rains, rain-on-snow events, and snowmelt affecting sites in eastern WA and OR. The 2 CO sites included in cluster 4 are a result of the aridity and, therefore, higher variability of these streams.

4.2. Determination of flow regime for ungauged streams

The step-wise discriminant analysis selected average annual precipitation (precip_t), average August precipitation (prcp_aug), nt_ndjfm, and average annual snowfall (snow_t) as the variables that best discriminate between the four streamflow regime clusters. The maximum correlation between these variables is −0.62, between snow_t and nt_ndjfm. The error rate for classifying the streamflow regime based on these watershed variables determined by leave-one-out cross validation is 13.29%. Six errors occurred when cluster 2 streams were classified as cluster 3. Four additional errors resulted from the misclassification of cluster 4 as snowmelt or rain streams. Box-and-whisker plots of the within-cluster distributions of watershed variables used in the discriminant analysis were created (Fig. 5) to facilitate interpretation of the discriminant analysis and the climatic properties of each cluster. As expected, clusters 2 and 3 receive much more precipitation than 1 and 4, with cluster 3 receiving the most. Cluster 1 has the highest snowfall, followed by cluster 3, then clusters 1 and 4. August precipitation effectively differentiates clusters 1 and

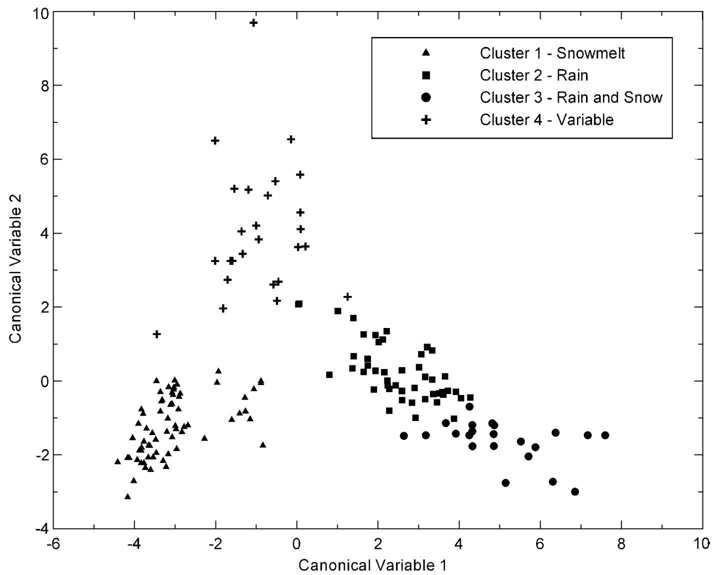

Fig. 2. Plot of the ‘all-flow’ clustering solution on the first two axes identified by canonical discriminant analysis.

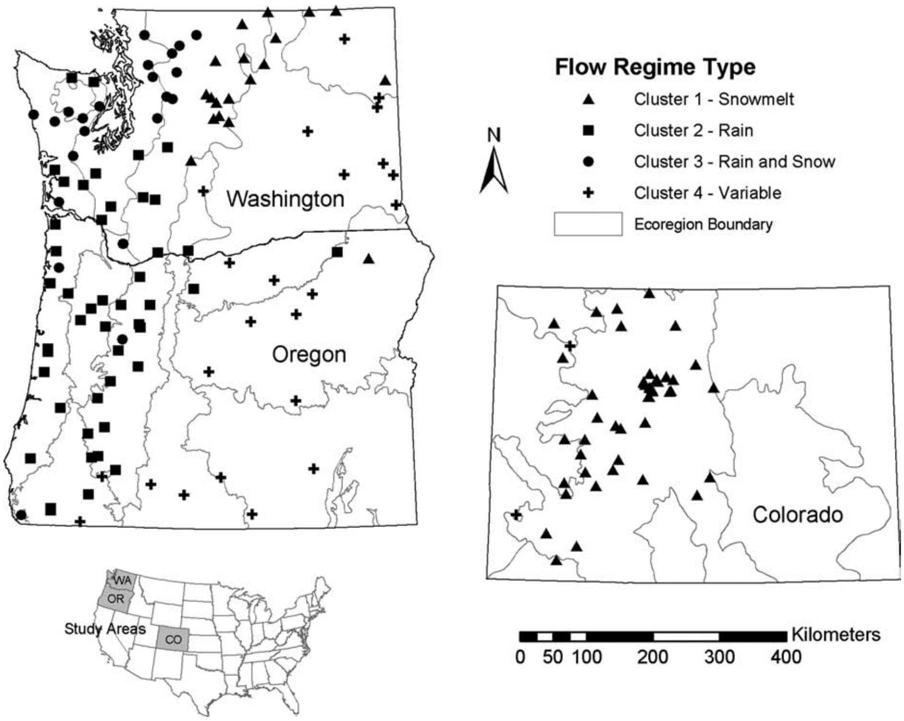

Fig. 3. USGS reference gauges classified by the 'all-flow' regime types superimposed on level 3 ecoregions.

4, and clusters 2 and 3. The nt_ndjfm metric provides further stratification for the snowmelt streams. The discriminant analysis essentially synthesizes the information contained in these plots so the flow regime type of ungauged streams can be accurately predicted. Equipped with a practical understanding of the clusters, we hereafter refer to cluster 1 ($m=61$) as a 'snowmelt' regime, cluster 2 ($m=47$) as a 'rain' regime, cluster 3 ($m=22$) as a 'rain and snow' regime, and cluster 4 ($m=20$) as a 'variable' regime.

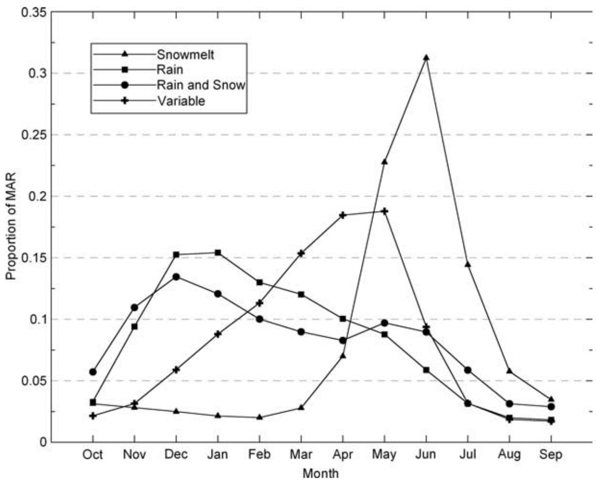

Fig. 4. Average monthly flows by flow regime type expressed as a proportion of mean annual runoff.

4.3. Cluster-specific regression models

Generally, the models for predicting various streamflow magnitudes were more accurate than models predicting high- and low-flow pulses and flow variability. In all, 26 streamflow metrics could be predicted well with R^2 greater than 0.65 for all streamflow regime types and seven more for every type except 'variable'. Nine metrics could be predicted with a value of R^2 greater than 0.89 for every type. The independent variables, parameter values, R^2, standard error, PRESS RMSE, and maximum VIF for these 33 models are shown in Tables 4a–4d. There were 19 metrics, which could not be predicted with an R^2 greater than 0.65 in any region, including CV_Jan, CV_Feb, CV_Mar, CV_Jul, CV_Aug, CV_Mn1, CV_Mn3, CV_Mn7, CV_Mn30, CV_Mn90d, NLoPl, CV_NLoPl, DHiPl, CV_DHiPl, skew, revs, Ml13, Fh11, and Th3.

To test the validity of clusters spanning such diverse regions, regression models for the 'snowmelt' cluster were generated separately for CO and Pacific Northwest (PNW) streams. Models for the entire cluster are biased towards CO streams, which account for 48 out of the 61 observations, so differences between models for the combined cluster and CO streams alone are small. However, 27 of the models constructed using only PNW streams outperform models for the entire cluster by a difference in R^2 of at least 0.20, despite a reduction in the number of independent variables from four to two.

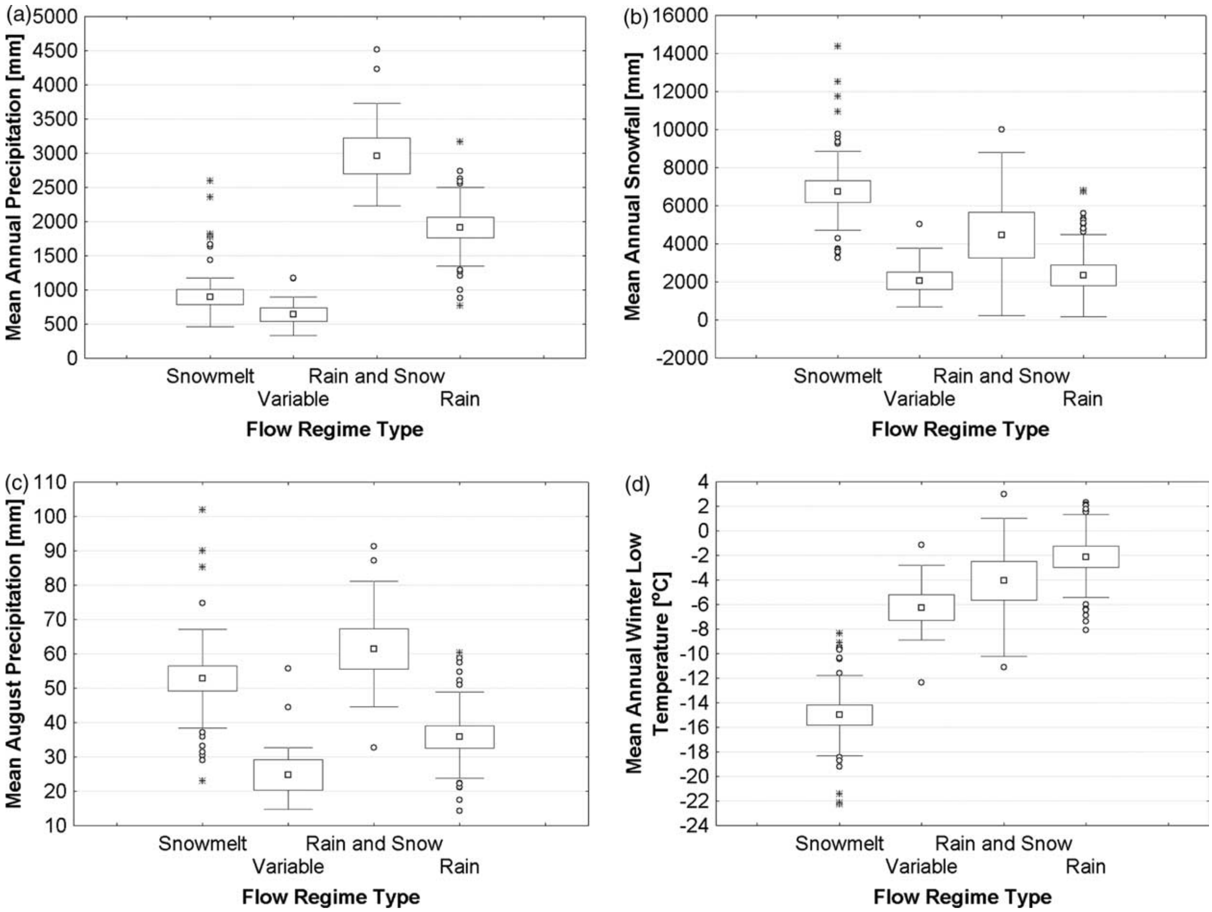

Fig. 5. Box-and-whisker plots depicting the distribution of watershed variables by flow regime class. The box represents data within two standard errors of the mean and the whiskers encompass all other points except the outliers and extremes shown.

Table 4a
Selected regression models and statistics for streams in the 'snowmelt' cluster

Metric	Adj. R^2	Var_1	Var_2	Var_3	Var_4	k	β_1	β_2	β_3	β_4	MaxVIF	PRESS RMSE	SE
MAR	0.985	prcp_mar	drainden	shed_slp	da	−8.5606	1.1799	0.5081	0.9078	0.8995	2.01	0.0291	0.0287
Avg_Oct	0.976	snow_-mar	shed_slp	da	store_p	−14.8247	0.9298	1.5085	1.0415	0.6134	1.58	0.0407	0.0391
Avg_Nov	0.981	shed_slp	da	precip_t	snow_jun	−19.6069	0.8701	1.0615	1.5318	0.0993	1.87	0.0386	0.0378
Avg_Dec	0.977	shed_slp	da	precip_t	snow_jun	−21.1856	0.8278	1.0865	1.7363	0.1013	1.87	0.0444	0.0434
Avg_Jan	0.973	shed_slp	da	precip_t	snow_jun	−20.9027	0.7826	1.0944	1.6939	0.0881	1.87	0.0480	0.0468
Avg_Feb	0.976	prcp_feb	snow_dec	shed_slp	da	−10.9531	1.6658	−0.6858	0.6635	1.0771	7.09	0.0460	0.0451
Avg_Mar	0.985	prcp_feb	topo_wet	snow_dec	da	−3.3398	1.8280	−2.6905	−0.8019	1.1043	6.18	0.0364	0.0361
Avg_Apr	0.977	da	forest_p	precip_t	snow_jun	−16.3934	0.9762	0.5625	1.9426	−0.2692	1.34	0.0467	0.0450
Avg_May	0.963	prcp_mar	drainden	da	forest_p	−6.4628	1.7131	0.4733	0.8624	0.6314	1.06	0.0489	0.0477
Avg_Jun	0.969	prcp_mar	shed_slp	da	precipr2	−10.5847	1.2313	0.9698	0.8595	−1.1762	1.85	0.0395	0.0388
Avg_Jul	0.969	shed_slp	da	store_p	snow_t	−15.1914	1.5886	0.9022	0.5971	0.9971	1.68	0.0412	0.0401
Avg_Aug	0.971	shed_slp	da	store_p	snow_jun	−6.7321	1.5718	0.9537	0.9715	0.2095	1.50	0.0411	0.0403
Avg_Sep	0.963	prcp_sep	shed_slp	da	snow_jun	−16.3489	1.5391	1.1352	1.0032	0.2592	1.63	0.0467	0.0459
Mn1d	0.961	prcp_mar	shed_slp	da	snow_jun	−15.6689	1.0723	0.9756	1.0871	0.1504	1.72	0.0552	0.0533
Mn3d	0.963	prcp_mar	shed_slp	da	snow_jun	−15.5586	1.0720	0.9642	1.0852	0.1470	1.72	0.0538	0.0520
Mn7d	0.964	prcp_mar	shed_slp	da	snow_jun	−15.4495	1.0699	0.9527	1.0863	0.1415	1.72	0.0527	0.0509
Mn30d	0.969	prcp_mar	shed_slp	da	snow_jun	−15.3019	1.0937	0.9239	1.0909	0.1226	1.72	0.0493	0.0479
Mn90d	0.973	prcp_mar	shed_slp	da	snow_jun	−15.4431	1.1688	0.8983	1.1034	0.1157	1.72	0.0462	0.0454
Mx1d	0.974	prcp_mar	drainden	shed_slp	da	−4.9384	1.2280	0.6231	0.7461	0.8129	2.01	0.0361	0.0350
Mx3d	0.973	prcp_mar	drainden	shed_slp	da	−5.1138	1.2299	0.6014	0.7292	0.8139	2.01	0.0369	0.0359
Mx7d	0.974	topo_wet	snow_jan	da	forest_p	1.7558	−5.9751	1.1355	0.8346	0.3622	1.38	0.0352	0.0352
Mx30d	0.973	topo_wet	snow_jan	da	forest_p	1.7423	−6.0901	1.1266	0.8392	0.3734	1.38	0.0359	0.0356
Mx90d	0.976	topo_wet	snow_jan	da	forest_p	2.0167	−6.6613	1.1445	0.8752	0.3488	1.38	0.0353	0.0352
CV_M-x90d	0.698	drainden	shed_slp	precipr2	snow_t	1.3337	−0.2917	−0.3668	0.2252	−0.3839	1.62	0.0212	0.0188
DatMn	0.753	avg_elev	shed_slp	volcan	snow_jun	2.7693	0.2901	0.1814	−0.0112	0.0382	1.29	0.0095	0.0092
DatMx	0.713	prcp_jan	snow_nov	shed_slp	precipr1	6.0407	−0.0361	0.0438	0.0247	−0.0189	5.80	0.0019	0.0018
RiseR	0.975	prcp_mar	drainden	shed_slp	da	−8.2033	1.1902	0.7508	0.9321	0.7963	2.01	0.0351	0.0342
FallR	0.769	topo_wet	da	store_p	chan_slp	3.8734	0.1681	−0.0082	−0.0178	0.0257	1.43	0.0017	0.0016
Ma3	0.647	prcp_sep	snow_oct	da	snow_jun	1.7476	−0.4002	0.1944	−0.0816	−0.1184	1.96	0.0164	0.0159
Ma41	0.934	prcp_mar	shed_slp	slp_elon	precipr2	−3.8122	1.1991	0.7590	−0.3091	−0.7663	1.96	0.0284	0.0269
Ml22	0.756	prcp_mar	prcp_may	shed_slp	snow_jun	−10.3171	1.7104	−1.0820	0.5618	0.1940	3.69	0.0539	0.0524
Mh1	0.975	prcp_oct	shed_slp	da	store_p	−7.6770	0.7715	0.9322	0.9585	0.6949	2.30	0.0427	0.0418
Mh8	0.975	prcp_mar	da	slp_elon	snow_jun	−6.5325	1.1159	0.9453	−0.3921	−0.1847	2.58	0.0400	0.0389

5. Discussion

Although we only present results from the 'all-flow' stratification scheme, each cluster solution produced many models with R^2 values exceeding 0.80. Many of the differences in model performance between the 'all-flow' and high-, low-, or average-flow clustering schemes could be attributed to differences in sample size. The cluster solutions were not markedly different among the high, low, and average schemes. Many sites were consistently classified together, which suggests that the flow regimes modeled in this study are fairly distinct in all three magnitude classes and have a high predictability component due to periodicity in the climatic processes driving streamflow. Thus, the 'all-flow' clustering solution collectively characterized the differences in each component of the flow regime. Analyses of data sets containing streams with flashier, less periodic regime types might garner additional benefits from a multi-faceted clustering scheme based on high, low, and average flows. For example, two streams might have low flows primarily driven by similar processes and characteristics such as geology and storage area, yet have high flows that behave quite differently due to climatological differences. High and low flows could then be predicted using models from separate clustering schemes, which more accurately represent processes driving that particular aspect of the flow regime. Technically, there is an optimal cluster solution for each metric, but it is not practical to develop a clustering solution for each streamflow metric.

In this study, the 'all-flow' scheme produced interpretable regression models with acceptable standard errors and R^2 values. Drainage area (da) was the best predictor for models estimating flow magnitudes. Eleven predictive models for monthly average flows normalized by drainage area had R^2 values greater than 0.70 for all flow regime types. The normalized maximum and minimum 1-, 3-, 7-, 30-, and 90-day flow models had R^2 values greater than 0.78 and 0.59, respectively, for all flow regime types. This demonstrates that streamflow drivers other than basin scale are well represented in the models. In the 'variable' regime grouping, models had higher standard errors than other flow regime types because of the variability associated with streams in more arid regions (Vogel et al., 1999) and typically had the lowest R^2 value for a particular model in any region. The 'rain' cluster produced models with R^2 values greater than 0.65 for 25 metrics in addition to those listed in Table 4b, the most of any flow regime type. This is likely due to the predictability of frontal winter rains driving streamflow. For these streams, snow is a principal source of flow variability as evidenced by the number of snow metrics appearing in the models of flow variation. The 'rain and snow' streams produced 14 additional robust models despite a small sample size that limited models to two independent variables. Again, this can be attributed to the regional climatic predictability and its direct link to streamflow. Models predicting monthly, and maximum and minimum 1-, 3-, 7-, 30-, and 90-day flow magnitudes for 'snowmelt' streams are exceptional ($R^2 > 0.95$), but the CVs of these metrics are not, with many R^2 values less than 0.65.

Geographically separating the 'snowmelt' cluster improved models of flow variability for both CO and PNW streams but did not substantially improve monthly, maximum, or minimum flow magnitude models. The relatively poor performance of some models based on all snowmelt sites indicates that they do not adequately represent the processes controlling certain aspects of the regional snowmelt regimes. Measures of monthly variability are higher for CO than for the PNW streams from May to September but lower the rest of the year. Although the snowpack is generally more variable in the PNW than in CO (Serreze et al., 1999), the PNW streams selected for this study have consistently larger drainage areas that likely mask yearly micro-climatic variation, leading to less overall variability. We hypothesize that this combined with greater storage in deeper PNW snowpacks and more gradual spring warming produces greater consistency in monthly flows during the snowmelt season. During the remainder of the year, CO streams experience baseflow conditions while PNW streams receive considerable amounts of rain, resulting in more variable monthly flows. Although the monthly signals for these regimes are quite similar, examining the regimes on a daily time scale reveals that CO streams almost never deviate from baseflow conditions in the winter but PNW streams can have large events due to rain or rain-on-snow events. This explains why many metrics

Table 4b
Selected regression models and statistics for streams in the 'rain' cluster

Metric	Adj. R^2	Var_1	Var_2	Var_3	Var_4	Intercept	B1	B2	B3	B4	MaxVIF	PRESS RMSE	SE
MAR	0.984	prcp_sep	da	precip_t	et_t	−19.3589	0.4215	1.0026	1.1640	0.8154	6.31	0.0203	0.0191
Avg_Oct	0.959	prcp_sep	shed_slp	da	nt_ndjfm	−5.6412	1.8489	−0.5861	1.0740	−1.6135	1.47	0.0347	0.0337
Avg_Nov	0.961	prcp_jun	prcp_nov	da	nt_ndjfm	−16.9275	0.5293	1.4439	0.9987	1.1491	1.71	0.0317	0.0307
Avg_Dec	0.970	da	precip_t	et_t	nt_ndjfm	−26.2560	0.9838	1.8804	0.7342	1.4917	2.25	0.0298	0.0271
Avg_Jan	0.977	da	precip_t	et_t	nt_ndjfm	−25.3508	0.9933	1.6679	0.7330	1.7061	2.25	0.0272	0.0245
Avg_Feb	0.976	da	precip_t	et_t	nt_ndjfm	−24.1348	0.9896	1.5311	0.7997	1.4803	2.25	0.0278	0.0247
Avg_Mar	0.976	snow_feb	da	precip_t	et_t	−24.0077	−0.2172	0.9757	1.6264	1.5002	2.18	0.0261	0.0237
Avg_Apr	0.974	prcp_mar	prcp_sep	da	et_t	−19.6544	0.4999	0.8443	0.9496	1.5693	4.26	0.0257	0.0240
Avg_May	0.966	prcp_may	da	et_t	nt_ndjfm	−3.9309	1.0802	0.9999	0.4665	−2.4604	1.41	0.0334	0.0296
Avg_Jun	0.960	prcp_apr	da	store_p	nt_ndjfm	4.5331	0.9773	1.0072	0.2677	−3.9622	1.79	0.0405	0.0371
Avg_Jul	0.935	prcp_sep	topo_wet	da	nt_ndjfm	−0.7407	1.4643	2.1757	1.1491	−4.9952	1.54	0.0541	0.0533
Avg_Aug	0.909	prcp_sep	topo_wet	da	nt_ndjfm	−2.6973	1.4408	2.9956	1.1808	−5.0583	1.54	0.0678	0.0671
Avg_Sep	0.920	prcp_sep	shed_slp	da	nt_ndjfm	2.5823	1.7975	−0.9668	1.2119	−4.3769	1.47	0.0584	0.0585
Mn1d	0.902	prcp_sep	shed_slp	da	nt_ndjfm	7.5430	1.5530	−1.2514	1.2770	−5.6917	1.47	0.0736	0.0729
Mn3d	0.905	prcp_sep	shed_slp	da	nt_ndjfm	7.4702	1.5753	−1.2247	1.2811	−5.7306	7.19	0.0398	0.0719
Mn7d	0.906	prcp_sep	shed_slp	da	nt_ndjfm	7.3033	1.5903	−1.2004	1.2789	−5.7075	1.47	0.0724	0.0713
Mn30d	0.909	prcp_sep	shed_slp	da	nt_ndjfm	6.5248	1.6173	−1.1141	1.2605	−5.4993	7.19	0.0380	0.0681
Mn90d	0.922	prcp_sep	shed_slp	da	nt_ndjfm	4.5598	1.7106	−0.9579	1.2247	−4.9754	1.47	0.0716	0.0595
Mx1d	0.939	prcp_nov	shed_slp	da	nt_ndjfm	−14.4183	0.8943	0.6815	0.8949	2.2065	1.50	0.0373	0.0367
Mx3d	0.951	prcp_nov	shed_slp	da	nt_ndjfm	−14.4360	0.9520	0.5619	0.9196	2.1024	5.24	0.0174	0.0335
Mx7d	0.958	prcp_nov	shed_slp	da	nt_ndjfm	−14.3472	1.0076	0.4625	0.9360	1.9425	1.50	0.0344	0.0312
Mx30d	0.968	da	precip_t	et_t	snow_ppt	−18.8321	0.9671	1.4385	0.8906	−0.2071	5.46	0.0178	0.0267
Mx90d	0.973	da	precip_t	et_t	snow_ppt	−18.9973	0.9745	1.4478	0.8508	−0.1838	1.50	0.0327	0.0246
CV_M-x90d	0.651	prcp_feb	prcp_jun	shed_slp	nt_ndjfm	−1.7602	−0.2122	−0.2463	0.1811	0.6614	1.56	0.0166	0.0164
DatMn	0.687	drainden	slp_elon	volcan	snow_ppt	5.4138	−0.0248	−0.0365	−0.0261	0.0606	2.46	0.0070	0.0046
DatMx	0.708	prcp_oct	avg_elev	sedim	nt_ndjfm	6.9760	−0.0662	−0.0323	−0.0061	−0.1647	5.07	0.0029	0.0027
RiseR	0.930	prcp_nov	shed_slp	da	nt_ndjfm	−20.5627	1.0357	0.8901	0.8274	2.8064	1.50	0.0388	0.0381
FallR	0.699	da	store_p	drain_cl	snow_ppt	2.7120	−0.2320	−0.5433	0.4035	0.2006	1.43	0.0838	0.0483
Ma3	0.849	prcp_sep	shed_slp	store_p	nt_ndjfm	−7.2073	−0.2607	0.5818	−0.1086	1.9934	1.45	0.0201	0.0193
Ma41	0.925	prcp_sep	drain_cl	precip_t	et_t	−13.5765	0.3613	0.1432	1.2760	1.0411	6.89	0.0196	0.0182
Ml22	0.758	prcp_sep	shed_slp	da	nt_ndjfm	9.4463	1.5530	−1.2514	0.2770	−5.6917	1.47	0.0736	0.0729
Mh1	0.934	snow_apr	snow_dec	da	precip_t	−16.9771	0.3759	−0.4667	1.0702	1.9439	7.64	0.0428	0.0429
Mh8	0.955	prcp_apr	da	et_t	nt_ndjfm	−6.4825	0.9752	0.9664	0.8258	−2.0676	1.73	0.0368	0.0330

describing flow at a daily time scale (e.g. Ma3, Ma44, CV_Mn1d, CV_Mn3d, Dh12, NHiPl) had R^2 values that were improved by more than 0.20 when treating PNW streams separately.

This study suggests that geographically independent stratification within the confines of broad climatic regions is an appropriate strategy for modeling suites of streamflow metrics. Spatially contiguous regions are inherently deficient for regions of considerable climatic or topographic heterogeneity because they fail to reflect the patchiness of key processes driving streamflow including precipitation and temperature. Geographically independent approaches make physical sense for heterogeneous regions and can result in robust models as in Chiang et al. (2002a,b), but there appear to be limits to this approach. Caution should be exercised when mixing sites from different hydroclimatic and geologic regions as the benefits of an increased sample size may be offset by loss of model fidelity to processes for certain types of metrics. Selecting an appropriate level of stratification is critical to any study where groupings are necessary to facilitate accurate predictive modeling, but these groups cannot be 'unacceptably heterogeneous' (Gabriele and Arnell, 1991). A hierarchical clustering technique could facilitate the investigation of alternative clustering solutions for each streamflow type by selecting a different significance level in the dendrogram from which to select the solution. More robust descriptors of the driving processes, especially climatic and geologic heterogeneity in mountainous regions, may further reduce the need for geographically dependent regionalization schemes.

The numerous streamflow metrics modeled in this study provide information that is potentially useful in identifying which hydrologic characteristics best explain taxonomic and functional variation in stream biota. For example, maximum and minimum flow metrics coupled with information about numbers and durations of flow pulses provide a first-order approximation of relative flow disturbance in the ungauged streams, an important influence on benthic macroinvertebrate communities (Resh et al., 1988; Townsend et al., 1997; Wood et al., 2000; Matthaei and Townsend, 2000). Flow timing is also very important for many ecological studies due to interactions between streamflow and the specific needs of aquatic and riparian biota at different life stages. Predicted monthly streamflow signals combined with other metrics unique to this study, such as the average and standard deviation of the date of maximum flow, could be useful in guiding the introduction of spring spawning fish whose success depends upon survival of fry that are susceptible to high flows (Poff and Allan, 1995; Fausch et al., 2001). The resulting framework and metrics also have implications for restoration activities on ungauged or poorly gauged streams that could benefit from estimates of pre-disturbance flow regime metrics including channel-forming discharges for geomorphic design. In general, metrics thought to be directly linked to a specific ecological question can be carefully chosen, clustered on, and modeled using the framework developed in this study.

Cottonwood (*Populus* spp.) recruitment is another example of an ecological process that is strongly controlled by specific streamflow components. High flows create and transport seeds to nursery sites at an elevation 60–150 cm above the base flow stage. Seeds must be deposited on the falling limb of the hydrograph and the stage must not fall at a rate greater than approximately 2.5 cm/day so the roots can retain contact with the receding moisture zone. The subsequent summer and autumn flows must be high enough to provide water to the seedling (Mahoney and Rood, 1998). The magnitude, timing, duration, and the rate of change of flows must be known to determine if a specific site has suitable flows for recruitment. For snowmelt streams, models for monthly flows, the one-day annual maximum flow, and its average date of occurrence all have R^2 values greater than 0.76. Such information could be useful, for example, in restoring a severely disturbed ungauged stream. Relevant flow metrics could be estimated, and channel cross sections and floodplain elevations designed such that the criteria of the 'recruitment box' are met, thereby increasing the likelihood of cottonwood regeneration. Because many of the metrics related to riparian processes are dependent upon precipitation, the models could also be used as part of a broader framework for exploring the potential ramifications of climate change for riparian and aquatic communities across large regions.

Table 4c
Selected regression models and statistics for streams in the ‘rain and snow’ cluster

Metric	Adj. R^2	Var_1	Var_2	Intercept	B1	B2	MaxVIF	PRESS RMSE	SE
MAR	0.980	da	precip_t	−9.3414	0.9832	0.8702	1.01	0.0244	0.0240
Avg_Oct	0.928	prcp_aug	da	−7.7405	1.4269	0.8311	1.10	0.0496	0.0483
Avg_Nov	0.901	da	precip_t	−14.8134	0.9608	1.6054	1.01	0.0602	0.0567
Avg_Dec	0.908	da	nt_ndjfm	−7.0897	0.8769	1.9647	1.00	0.0642	0.0572
Avg_Jan	0.923	da	nt_ndjfm	−7.8816	0.8775	2.1927	1.00	0.0615	0.0541
Avg_Feb	0.917	da	nt_ndjfm	−7.9171	0.8669	2.1940	1.00	0.0629	0.0558
Avg_Mar	0.934	da	nt_ndjfm	−7.6878	0.8928	1.9971	1.00	0.0557	0.0488
Avg_Apr	0.937	prcp_may	da	−7.5406	1.0907	0.9600	1.03	0.0497	0.0454
Avg_May	0.912	da	snow_t	−5.4565	1.0195	0.3556	1.02	0.0686	0.0620
Avg_Jun	0.857	da	snow_t	−7.6108	1.0360	0.5881	1.02	0.1137	0.0942
Avg_Jul	0.855	da	nt_ndjfm	5.3240	1.2034	−3.3414	1.00	0.1019	0.1040
Avg_Aug	0.844	da	nt_ndjfm	4.1836	1.1380	−3.0271	1.00	0.0994	0.1007
Avg_Sep	0.851	da	precipr1	0.3407	0.8722	−1.4835	1.11	0.0892	0.0842
Mn1d	0.912	da	precipr2	0.2086	1.0131	−6.3067	1.09	0.0711	0.0681
Mn3d	0.913	da	precipr2	0.2895	1.0151	−6.3934	1.07	0.0380	0.0679
Mn7d	0.913	da	precipr2	0.5091	1.0088	−6.5622	1.09	0.0709	0.0683
Mn30d	0.909	da	precipr2	1.1670	0.9944	−7.0076	1.04	0.0386	0.0705
Mn90d	0.900	da	precipr2	2.3028	0.9558	−7.5684	1.09	0.0714	0.0744
Mx1d	0.914	da	nt_ndjfm	−3.4346	0.8809	1.3066	1.00	0.0528	0.0507
Mx3d	0.925	da	nt_ndjfm	−3.7236	0.8882	1.2832	1.62	0.0241	0.0473
Mx7d	0.931	da	nt_ndjfm	−3.7403	0.8945	1.1600	1.00	0.0489	0.0449
Mx30d	0.953	da	precip_t	−9.3143	0.9517	1.0118	1.77	0.0244	0.0367
Mx90d	0.959	prcp_dec	da	−6.3569	0.7531	0.9803	1.00	0.0460	0.0337
CV_Mx90d	0.657	slp_elon	forest_p	−1.4203	0.1103	0.5641	1.23	0.0249	0.0195
DatMn	0.857	topo_wet	forest_p	6.0589	−0.2972	−0.6780	1.01	0.0103	0.0093
DatMx	0.872	store_p	sedim	6.0946	0.0458	−0.0169	1.41	0.0058	0.0046
RiseR	0.882	da	nt_ndjfm	−6.8858	0.8137	1.5863	1.00	0.0612	0.0588
FallR	0.680	da	sedim	4.8126	−0.1358	−0.0243	1.01	0.0158	0.0160
Ma3	0.818	precipr2	nt_ndjfm	−2.8543	1.3477	0.6534	1.15	0.0212	0.0198
Ma41	0.766	sedim	precip_t	−0.1296	0.0419	0.7222	1.23	0.0213	0.0202
Ml22	0.789	prcp_aug	nt_ndjfm	−6.2016	1.8594	−1.4248	1.01	0.0550	0.0528
Mh1	0.917	da	precip_t	−11.7462	0.8994	1.3137	1.01	0.0477	0.0475
Mh8	0.936	avg_elev	da	−4.9727	0.4609	1.0141	1.00	0.0510	0.0484

This study describes a generalized methodological approach for conducting future large-scale flow prediction studies. Due to the growing interest in efficiently characterizing the entire streamflow regime at sites spanning disparate regions, robust stratification schemes are critical. Numerous ecological studies have successfully distinguished flow regimes, but have not examined the feasibility of predicting a comprehensive suite of streamflow metrics for ungauged sites. This study represents a synthesis of ideas from various streamflow regionalization and prediction studies to provide a holistic approach to broad-scale flow regime modeling based on hydroclimatic processes. To improve this framework, more advanced statistical procedures such as artificial neural networks or genetic algorithms could be used to develop models for metrics, which could not be predicted accurately using MRA. CART analysis (De’ath and Fabricius, 2000) could potentially improve the clustering procedure and eliminate the need for discriminant analysis. These methods use complex algorithms that can better represent non-linearity and may provide more accurate models than standard multivariate statistical techniques for certain streamflow metrics.

6. Conclusions

The methodology developed in this study provides a means for classifying streamflow regime types and

Table 4d
Selected regression models and statistics for streams in the 'variable' cluster

Metric	Adj. R^2	Var_1	Var_2	Intercept	B1	B2	MaxVIF	PRESS RMSE	SE
MAR	0.934	da	precip_t	−21.6712	1.0483	2.4845	1.51	0.0780	0.0771
Avg_Oct	0.837	da	precip_t	−28.0866	1.1291	3.1619	1.51	0.1413	0.1370
Avg_Nov	0.886	da	precip_t	−32.8239	1.1832	3.9093	1.51	0.1184	0.1191
Avg_Dec	0.836	da	precip_t	−32.6315	1.2257	3.9102	1.51	0.1483	0.1511
Avg_Jan	0.837	da	precipr1	−8.3535	1.0292	1.4883	1.15	0.1492	0.1504
Avg_Feb	0.885	da	precipr1	−7.9582	1.0685	1.2783	1.15	0.1242	0.1238
Avg_Mar	0.904	da	precip_t	−20.6470	1.1577	2.2728	1.51	0.1035	0.1054
Avg_Apr	0.909	da	forest_p	−3.1044	0.8656	0.6402	1.02	0.0941	0.0966
Avg_May	0.835	da	forest_p	−2.2767	0.7592	0.7731	1.02	0.1348	0.1264
Avg_Jun	0.726	da	precip_t	−20.4987	0.9942	2.3775	1.51	0.1839	0.1669
Avg_Jul	0.774	da	precip_t	−22.1950	1.0601	2.3833	1.51	0.1693	0.1578
Avg_Aug	0.682	da	precip_t	−22.1830	1.0253	2.3098	1.51	0.2039	0.1916
Avg_Sep	0.702	da	precip_t	−24.2773	1.0708	2.5794	1.51	0.2037	0.1910
Mn1d	0.612	snow_jun	da	−7.8875	0.8126	0.8811	1.04	0.2528	0.2533
Mn3d	0.617	da	precip_t	−28.6421	1.1588	3.0548	1.51	0.2551	0.2482
Mn7d	0.623	da	precip_t	−28.3881	1.1427	3.0443	1.51	0.2498	0.2418
Mn30d	0.663	da	precip_t	−27.5595	1.1267	2.9658	1.51	0.2292	0.2191
Mn90d	0.733	da	precip_t	−27.4214	1.1542	2.9612	1.51	0.2007	0.1906
Mx1d	0.834	da	precipr1	−3.3689	0.7815	0.9598	1.15	0.1215	0.1121
Mx3d	0.863	da	precip_t	−16.9177	0.9341	2.1678	1.51	0.1068	0.1026
Mx7d	0.897	da	precip_t	−16.8853	0.9482	2.1171	1.51	0.0911	0.0890
Mx30d	0.917	da	precip_t	−17.0229	0.9623	2.0607	1.51	0.0817	0.0804
Mx90d	0.925	da	precip_t	−18.7829	1.0029	2.2326	1.51	0.0811	0.0794
CV_Mx90d	0.547	prcp_may	volcan	2.0539	−0.7398	−0.0832	1.03	0.0376	0.0377
DatMn	0.341	forest_p	precip_t	4.0159	−0.0974	0.2351	1.56	0.0308	0.0222
DatMx	0.795	precip_t	snow_t	5.9388	−0.1332	0.1360	1.13	0.0074	0.0072
RiseR	0.785	da	precipr1	−6.9021	0.7524	1.0142	1.15	0.1373	0.1280
FallR	0.792	topo_wet	da	3.9891	0.1581	−0.0201	1.10	0.0038	0.0034
Ma3	0.420	forest_p	sedim	0.5001	−0.2302	0.1105	1.36	0.0590	0.0479
Ma41	0.844	precip_t		−12.4572	2.3585		1.00	0.0757	0.0762
Ml22	0.602	snow_jun	precip_t	−19.9335	0.6694	2.0645	1.08	0.2059	0.1987
Mh1	0.823	Da	precip_t	−27.1792	1.0427	3.3037	1.51	0.1334	0.1345
Mh8	0.886	Da	forest_p	−1.6245	0.7906	0.6771	1.02	0.1085	0.1021

predicting several magnitude, frequency, duration, timing, and rate of change characteristics at ungauged stream biomonitoring sites. The procedure yielded several robust models for predicting a suite of streamflow regime metrics of ecological interest in ungauged basins across disparate hydroclimatic regions of CO, WA, and OR. Flow magnitude models were generally more accurate than models describing flow variability. Models for certain metrics, particularly in clusters containing multiple ecoregions separated by large distances (ca. 2000 km), perform better when geographically independent groups are further divided to better reflect gross differences in geologic and climatic factors. Stratification schemes based specifically on high-, moderate-, and low-flow aspects of the flow regimes all produce reliable models but yielded no appreciable improvement in model performance in this study. This somewhat surprising result should be tested further, especially in study areas with less predictable regimes to further explore the utility of clustering on specific aspects of the flow regime and the transferability of streamflow data across regions. The prediction of these ecologically significant streamflow metrics at ungauged sites quantifies a 'master variable' that profoundly influences aquatic communities, thereby facilitating interpretation of biological variation and stratification of biomonitoring sites for water-quality assessment.

This research also points to the possibility of combining readily predicted flow regime metrics and geomorphic attributes to explore and quantify linkages between specific aspects of the flow regime and aquatic community structure. In addition to presenting a framework for simultaneously modeling a comprehensive suite of ecological streamflow metrics, this research contributes new information and questions for the Predictions in Ungauged Basins initiative of the International Association of Hydrological Sciences.

Acknowledgements

This research was supported by the USEPA Science to Achieve Results (STAR) program, grant number R82863601. We wish to thank Jose Salas and David Merritt for their suggestions and comments on an earlier draft. We also thank Phil Chapman, Sean Mahabir, and Xiaofan Cao in the Statistics Department at Colorado State University for their assistance with statistical analyses; and Dana Watson and Keith Olson for assistance with GIS analyses.

References

Baron, J.S., Poff, N.L., Angermeier, P.L., Dahm, C.N., Gleick, P.H., Hairston Jr., N.G., Jackson, R.B., Johnston, C.A., Richter, B.D., Steinman, A.D., 2002. Meeting ecological and societal needs for freshwater. Ecol. Appl. 12 (5), 1247–1260.

Burn, D.H., 1989. Cluster analysis as applied to regional flood frequency. J. Water Resour. Plan. Manage. 115 (5), 567–582.

Burn, D.H., Boorman, D.B., 1993. Estimation of hydrological parameters at ungauged catchments. J. Hydrol. 143, 429–454.

Chiang, S.M., Tsay, T.K., Nix, S.J., 2002a. Hydrologic regionalization of watersheds I: methodology development. J. Water Resour. Plan. Manage. 128, 3–11.

Chiang, S.M., Tsay, T.K., Nix, S.J., 2002b. Hydrologic regionalization of watersheds II: applications. J. Water Resour. Plan. Manage. 128, 12–20.

Daly, C., Neilson, R.P., Phillips, D.L., 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J. Appl. Meteorol. 33 (2), 140–158.

De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81 (11), 3178–3192.

Fausch, K.D., Taniguchi, Y., Nakano, S., Grossman, G.D., Townsend, C.R., 2001. Flood disturbance regimes influence rainbow trout invasion success among five Holarctic regions. Ecol. Appl. 11, 1438–1455.

Fennessey, N., Vogel, R.M., 1990. Regional flow-duration curves for ungauged sites in Massachusetts. J. Water Resour. Plan. Manage. 116 (4), 530–549.

Gabriele, S., Arnell, N., 1991. A hierarchical approach to regional flood frequency analysis. Water Resour. Res. 27 (6), 1281–1289.

Gilroy, E., Tasker, G., 1990. Multicollinearity and influential observations in hydrologic model selection. In: Berk, K., Malone, L. (Eds.), Computing Science and Statistics. Proceedings of the 21st Symposium of the Interface, pp. 350–354.

Haines, A.T., Finlayson, B.L., McMahon, T.A., 1988. A global classification of river regimes. Appl. Geogr. 8, 255–272.

Hair Jr., J.F., Black, W.C., 2000. Cluster analysis. In: Grimm, L.G., Yarnold, P.R. (Eds.), Reading and Understanding More Multivariate Statistics. American Psychological Association, Washington, DC, pp. 147–206.

Hand, D.J., 1981. Discrimination and Classification. Wiley, Chichester, England.

Hand, D.J., 1997. Construction and Assessment of Classification Rules. Wiley, Chichester, England.

Harris, N.M., Gurnell, A.M., Hannah, D.M., Petts, G.E., 2000. Classification of river regimes: a context for hydroecology. Hydrol. Process. 14, 2831–2848.

Hawley, M.E., McCuen, R.H., 1982. Water yield estimation in the western United States. J. Irrigat. Drain. Div. 108 (IR1), 25–34.

Hirsch, R.M., Helsel, D.R., Cohn, T.A., Gilroy, E.J., 1993. Statistical analysis of hydrologic data. In: Maidment, D.R. (Ed.), Handbook of Hydrology. McGraw-Hill, New York (Chapter 17).

Hobbins, M.T., Ramirez, J.A., Claessens, L.H.J.M., 2001a. The complementary relationship in estimation of regional evapotranspiration: the complementary relationship areal evapotranspiration and advection-aridity models. Water Resour. Res. 37 (5), 1367–1387.

Hobbins, M.T., Ramirez, J.A., Brown, T.C., 2001b. The complementary relationship in estimation of regional evapotranspiration: an enhanced advection-aridity model. Water Resour. Res. 37 (5), 1389–1403.

Hughes, J.M.R., James, B., 1989. A hydrological regionalization of streams in Victoria, Australia, with implications for stream ecology. Aust. J. Mar. Freshwater. Res. 40, 303–326.

Jennings, M.E., Thomas Jr., W.O., Riggs, H.C., 1994. Nationwide Summary of US Geological Survey Regional Regression Eqs. for Estimating Magnitude and Frequency of Floods for Ungauged Sites, 1993. USGS Water-Resources Investigations Report 94-4002, 196 pp.

Jobson, J.D., 1992. Applied Multivariate Data Analysis Volume II: Categorical and Multivariate Methods. Springer, New York.

Joliffe, I.T., 1986. Principal Component Analysis. Springer, New York.

Kresch, D.L., 1993. Variability of Streamflow and Precipitation in Washington. USGS Water-Resources Investigations Report 93-4132, 32 pp.

Mahoney, J.M., Rood, S.B., 1998. Streamflow requirements for cottonwood seedling recruitment—an integrative model. Wetlands 18 (4), 634–645.

Matthaei, C.D., Townsend, C.R., 2000. Long-term effects of local disturbance history on mobile stream invertebrates. Oecologia 125, 119–126.

Milligan, G.W., Cooper, M.C., 1985. An examination of procedures for determining the number of clusters in a data set. Psychometrika 50, 159–179.

Olden, J.D., Poff, N.L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. River Res. Appl. 19, 101–121.

Omernik, J.M., 1995. Ecoregions: a spatial framework or environmental management. In: Davis, W.S., Simon, T.P. (Eds.), Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. Lewis, Boca Raton, FL, pp. 49–62.

Pitlick, J., 1994. Relation between peak flows, precipitation, and physiography for five mountainous regions in the western USA. J. Hydrol. 158, 219–240.

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, 71–91.

Poff, N.L., Allan, J.D., 1995. Functional organization of stream fish assemblages in relation to hydrological variability. Ecology 76 (2), 606–627.

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

Poff, N.L., Allan, J.D., Bain, M.B., Karr, J.R., Prestegaard, K.L., Richter, B.D., Sparks, R.E., Stromberg, J.C., 1997. The natural flow regime: a paradigm for conservation and restoration of river ecosytems. Bioscience 47, 769–784.

Rasmussen, L.A., Tangborn, W.V., 1976. Hydrology of the North Cascades region, Washington 1: runoff, precipitation, and storage characteristics. Water Resour. Res. 12 (2), 187–202.

Ratowsky, D.A., 1984. A stopping rule and clustering method of wide applicability. Bot. Gazette. 145 (4), 518–523.

Redmond, K.T., Koch, R.W., 1991. Surface climate and streamflow variability in the Western United States and their relationship to large-scale circulation indices. Water Resour. Res. 27 (9), 2381–2399.

Resh, V.H., Brown, A.V., Covich, A.P., Gurtz, M.E., Li, H.W., Minshall, G.W., Reice, S.R., Sheldon, A.L., Wallace, J.B., Wissmar, R.C., 1988. The role of disturbance in stream ecology. J. N. Am. Benthol. Soc. 7, 433–455.

Richter, B.D., Baumgartner, J.V., Powell, J., Braun, D.P., 1996. A method for assessing hydrologic alteration within ecosystems. Conserv. Biol. 10, 1163–1174.

Ries III, K.G., Friesz, P.J., 2000. Methods for Estimating Low-flow Statistics for Massachusetts Streams. USGS Water-Resources Investigations Report 00-4135, 81 pp.

Riggs, H.C., Harvey, K.D., 1990. Temporal and spatial variability of streamflow. In: Wolman, M.G., Riggs, H.C. (Eds.), Surface Water Hydrology. Geological Society of America, Boulder, CO, pp. 81–96.

Sarle, W.S., 1983. Cubic Clustering Criterion. SAS© Technical Report A-108, The SAS Institute Inc., Cary, North Carolina.

SAS©, 2001. SAS Version 8.02. The SAS Institute, Inc., Cary, NC.

Serreze, M.C., Clark, M.P., Armstrong, R.L., McGinnis, D.A., Pulwarty, R.S., 1999. Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. Water Resour. Res. 35 (7), 2145–2160.

Slack, J.R., Lumb, A.M., Landwehr, J.M., 1993. Hydro-Climatic Data Network (HCDN): Streamflow Data Set, 1874–1988. USGS Water-Resources Investigations Report

Smakhtin, V.Y., Hughes, D.A., Creuse-Naudin, E., 1997. Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa. Hydrol. Sci. J. 42 (6), 919–936.

Stedinger, J.R., Tasker, G.D., 1985. Regional hydrologic analysis 1. Ordinary, weighted, and generalized least squares compared. Water Resour. Res. 9, 1421–1432.

Sugiyama, H., Vudhivanich, B., Whitaker, A.C., Lorsirirat, K., 2003. Stochastic flow duration curves for evaluation of flow regimes in rivers. J. Am. Water Resour. Assoc. 39 (1), 47–58.

Surian, N., Andrews, E.D., 1999. Estimation of geomorphically significant flows in alpine streams of the Rocky mountains. Regul. Rivers Res. Manage. 15, 273–288.

Tague, C., Grant, G.E., 2004. A geological framework for interpreting the low-flow regimes of Cascade streams, Willamette River Basin, Oregon. Water Resour. Res. 40. doi:10.1029/2003WR02629 W04303.

Tasker, G.D., 1980. Hydrologic regression with weighted least squares. Water Resour. Res. 6, 1107–1113.

Thomas, D.M., Benson, M.A., 1970. Generalization of Streamflow Characteristics from Drainage-basin Characteristics. USGS Water-Supply Paper 1975, 55 pp.

Townsend, C.R., Scarsbrook, M.R., Doledec, S., 1997. The intermediate disturbance hypothesis, refugia, and biodiversity in streams. Limnol. Oceanogr. 42, 938–949.

Tucci, C., Silveira, A., Sanchez, J., Albuquerque, F., 1995. Flow regionalization in the upper Paraguay basin. Brazil. Hydrol. Sci. 40, 485–497.

US Environmental Protection Agency, 2002. Research Strategy: Environmental Monitoring and Assessment Program. EPA 620/R-02/002. Office of Research and Development, US Environmental Protection Agency, Research Triangle Park, North Carolina.

Vogel, R.M., Wilson, I., Daly, C., 1999. Regional regression models of annual streamflow for the United States. J. Irrigat. Drain. Eng. 125, 148–157.

Watershed Professionals Network, LLC, 2001. Hydrologic Process Identification for Eastern Oregon, 34 pp.

Western Regional Climate Center. Climate of Oregon. Desert Research Institute, Reno, Nevada. http://www.wrcc.dri.edu/narratives/OREGON.htm.

Western Regional Climate Center. Climate of Washington. Desert Research Institute, Reno, Nevada. http://www.wrcc.dri.edu/narratives/WASHINGTON.htm.

Wong, M.A., Schaack, C., 1982. Using the kth nearest neighbor clustering procedure to determine the number of subpopulations. In: Proceedings of the Statistical Computing Section. American Statistical Association, pp. 40–48.

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