ECOHYDROLOGY
Ecohydrol. **7**, 903–926 (2014)
Published online 22 July 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/eco.1410

Updating the US hydrologic classification: an approach to clustering and stratifying ecohydrologic data

Ryan A. McManamay,* Mark S. Bevelhimer and Shih-Chieh Kao
Environmental Sciences Division, Oak Ridge National Lab, Oak Ridge, TN 37831, USA

ABSTRACT

Hydrologic classifications unveil the structure of relationships among groups of streams with differing streamflows and provide a foundation for drawing inferences about the principles that govern those relationships. Hydrologic classes provide a template to generalize hydrologic responses to disturbance and stratify research and management needs applicable to ecohydrology. We used a mixed-modelling approach to create hydrologic classifications for the continental USA using three streamflow datasets, a reference dataset compiled under more strict traditional standards and two additional datasets compiled under more relaxed assumptions. A variety of models were applied to each dataset, and Bayes criteria were used to identify optimal models and numbers of clusters. Using only reference-quality gauges, we classified 1715 stream gauges into 12 classes across the USA. By including more streamflow gauges ($n = 2402$ and 2618) of lesser reference quality in subsequent classifications, we observed minimal increases in dimensionality (i.e. multivariate space) at the expense of increasing uncertainty and outliers. Part of the utility of classification systems rests in their ability to classify new objects and stratify data by common properties. We constructed separate random forest models to predict hydrologic class membership on the basis of hydrologic indices or landscape variables. In addition, we provide an approach to assessing potential outliers due to hydrologic alteration based on class assignment. Departures from class membership due to disturbance take into account multiple hydrologic indices simultaneously; thus, classes can be used to determine if disturbed streams are functioning within the natural range of hydrologic variability. Published 2013. This article is a U.S. Government work and is in the public domain in the USA.

Supporting information may be found in the online version of this article.

KEY WORDS environmental flow; streams; water policy; aquatic conservation; dams

Received 5 November 2012; Revised 19 June 2013; Accepted 19 June 2013

INTRODUCTION

Classifications depict our current state of knowledge about a subject area (Melles *et al.*, 2012) and provide the structure and relationships within and among groups of objects (Sokal, 1974). These relationships provide a foundation for drawing inferences about the principles that govern relationships among different classes and how to interpret unclassified objects (Sokal, 1974). With regard to river systems, stream classifications and their use in management have a fairly long history (Horton, 1945; Strahler, 1957; Pennak, 1971; Rosgen, 1994). However, Melles *et al.* (2012) suggested that the advent of the 'computer age' dramatically enhanced opportunities for developing data-intensive classification systems across larger spatial scales and at higher resolutions (e.g. Bailey, 1983; Omernik, 1987; Omernik and Bailey, 1997; Snelder and Biggs, 2002; Snelder *et al.*, 2007). Likewise, discussions regarding novel approaches, evaluation/testing, and appropriate scales for river classifications systems have continued to increase in recent years (Snelder and Biggs, 2002; Snelder *et al.*, 2007; Leathwick *et al.*, 2011; Melles *et al.*, 2012; Olden *et al.*, 2012).

Hydrology varies extensively across continents and globally (Kennard *et al.*, 2010b; Haines *et al.*, 1988), yet streams display reoccurring patterns in the magnitude, duration, frequency, timing, and rate of change of flow events within regions (Acreman and Sinclair, 1986; Burn and Arnell, 1993; Poff *et al.*, 1997). These repeatable patterns naturally predispose streams to hydrologic classification. One of the primary justifications for developing hydrologic classifications is to provide a means for developing environmental flow standards to support the preservation of freshwater biodiversity and ecosystem services (Arthington *et al.*, 2006; Poff *et al.*, 2010). For example, streams that behave similarly hydrologically should share similar patterns in ecology (Arthington *et al.*, 2006) and respond similarly to a

*Correspondence to: Ryan A. McManamay, Environmental Sciences Division, Oak Ridge National Lab, Oak Ridge, TN 37922, USA.
E-mail: mcmanamayra@ornl.gov

given anthropogenic stressor (Arthington *et al*., 2006; Poff *et al*., 2010). Classifications alleviate some of the complexity of environmental flow management by consolidating hydrologic variation into stream types and managing for groups of streams rather than for the uniqueness of individual water bodies. The development of hydrologic classifications for use in environmental flow management has greatly expanded in recent years. Approaches have ranged from deductive techniques using regional boundaries (e.g. ecoregions) and environmental variables for inferring areas of similar hydrologic regimes to inductive techniques where hydrologic data directly inform classifications (Olden *et al*., 2012). In situations where hydrologic information is lacking, deductive approaches may be advantageous; however, these approaches assume that the actual number of hydrologic classes (i.e. represented hydrologic variability) is already known or assume that the structure of environmental variables is important in predicting hydrology (Olden *et al*., 2012). In addition, deductive approaches often only include best professional judgment as criteria and may not accurately represent or predict streamflow patterns (McManamay *et al*., 2012c). Inductive approaches, by comparison, utilize the available hydrologic information (i.e. stream gauges) and classification techniques that group streams according to similarities in hydrologic metrics (Olden *et al*., 2012). Inductive approaches to hydrologic classifications have been created at multiple scales including states (Kennen *et al*., 2007, 2009; Turton *et al*., 2008; Henriksen and Heasley, 2010; Liermann *et al*., 2012), regions (Monk *et al*., 2006; Sanborn and Bledsoe, 2006; Chinnayakanahalli *et al*., 2011; McManamay *et al*., 2012b), continents (Kennard *et al*., 2010b), and the world (Haines *et al*., 1988). Despite this intense growth, comprehensive testing of hydrologic classifications in generalizing patterns of disturbance and establishing environmental flow standards, one of the central precepts behind creating streamflow-based classes (Arthington *et al*., 2006; Poff *et al*., 2010), has not been fully addressed. Furthermore, with regard to ecological patterns, the predictive capacity of hydrologic classifications has received little attention (but see Monk *et al*., 2006; Chinnayakanahalli *et al*., 2011).

The latest hydrologic classification for streamflows within the conterminous USA was produced more than 15 years ago by Poff (1996), who documented ten dominant streamflow types of varying intermittency, perennial flows, and timing (Figure 1). Over two decades of US Geological Survey (USGS) streamflow gauge information have become available since Poff (1996) produced a hydrologic classification for 806 reference stream gauges in the conterminous USA (latest gauges used were from 1986). Thus, it becomes important to understand how increases in sample size may influence the representative hydrologic variation across the USA via changes in the number of classes and class membership. However, updating hydrologic classifications requires establishing data-quality standards for the inclusion of new information. One common approach in hydrologic classification is screening gauges for inclusion in a final 'reference' dataset (Olden *et al*., 2012). The screening process typically includes evaluating landscape distur-

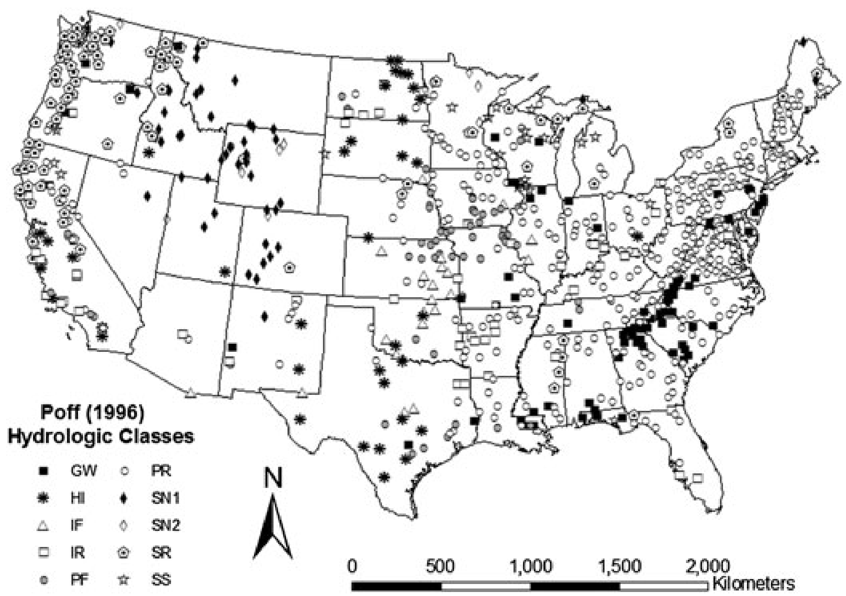

Figure 1. US hydrologic classification of 806 stream gauges into ten classes taken from Poff (1996). GW = groundwater, HI = harsh intermittent, IF = intermittent flashy, IR = intermittent runoff, PF = perennial flashy, PR = perennial runoff, SN1 = snowmelt 1, SN2 = snowmelt 2, SR = snow and rain, SS = super-stable.

bances upstream of each gauge, the hydrologic record length, and the extent of overlap among hydrologic records (Olden *et al*., 2012). Because most hydrologic classifications are constructed from natural streamflow patterns, the standards for inclusion can be quite strict and exclusive (Poff, 1996; Kennard *et al*., 2010a; Olden *et al*., 2012), which may limit the sample size and variation represented in the final dataset. Thus, high-data-quality standards often come at the expense of losses in hydrologic information.

The purpose of this study was to use an inductive approach to develop an updated hydrologic classification for the continental USA using three streamflow datasets of varying reference quality. We compiled a reference dataset under more strict traditional standards and two additional datasets under more relaxed assumptions. We then determined how the datasets varied in terms of dimensionality, predictive capacity, uncertainty (including outliers), and cluster stability. We also compared our hydrologic classes with those of Poff (1996) to describe how representative variation has changed. The utility of any classification system lies, in part, on its ability to stratify analyses and generalize patterns in disturbance. Thus, we provide an example of a multivariate method for determining the degree of hydrologic modification based on class membership.

METHODS

Identifying gauges for hydrologic classification

We used selected USGS gauges that fell into one of three categories: (1) reference, (2) nonreference with minimal hydrologic disturbance, and (3) pre-dam regulation. Although we define each category more thoroughly later, we provide a brief overview here. The categories represent a gradient from strict reference standards to progressively more relaxed assumptions, with relaxed assumptions indicative of lower reference quality, not high levels of disturbance. Reference gauges represent steams with the lowest amounts of anthropogenic disturbances based on landscape assessments and expert judgment (Falcone *et al*., 2010b). Nonreference streams may have slight disturbances in their upstream watershed that justify exclusion as reference-quality gauges. However, we presume that some disturbance is acceptable for inclusion in hydrologic classifications as long as streams function within the bounds of natural hydrologic variation. Lastly, gauges with discharge records that extend prior to dam regulation can potentially provide additional hydrologic information uninfluenced by disturbance. However, discharge records preceding dam construction also predate the availability of geospatial datasets and landscape information used to accurately assess disturbance and, thus, are prone to uncertainty. In addition, pre-dam records may represent different climatic regimes, which could influence classification results.

We define reference gauges as streams with the least amount of physiochemical and biological habitat disturbances within the current state of the landscape (Stoddard *et al*., 2006; Falcone *et al*., 2010b). Reference conditions and watershed disturbances may vary slightly among regions (Falcone *et al*., 2010b); thus, in order to represent continental-wide hydrologic variability, assessing levels of disturbance relative to each region is important. Reference gauges for the USA were provided in the Geospatial Attributes of Gages for Evaluating Streamflow, version II (GAGES II) database developed by (Falcone, 2011). The GAGES I and II databases were developed as part of a national effort to compile information on USGS stream gauges and their upstream watersheds in the conterminous USA, Alaska, Hawaii, and Puerto Rico (Falcone *et al*., 2010b, 2011). Geospatial information system (GIS) approaches were used to delineate watersheds and summarize natural (e.g. climate and soils) and anthropogenic (e.g. dams and land use) geospatial information for 9322 gauges (Falcone *et al*., 2010b, Falcone, 2011). Reference status was determined using three sources of information, including a GIS-based hydrologic disturbance index (HDI) (Falcone *et al*., 2010b), qualitative assessments of hydrologic alterations using 7·5′ topographic maps, and USGS Annual Water Data Reports (ADRs) (Falcone *et al*., 2010b). The HDI represents the cumulative disturbance of selected anthropogenic stressors summarized within each gauge's watershed. HDI stressors include major dam density, change in dam storage (1950–2006), percentage of canals and artificial paths along the mainstem of each gauge, distance to National Pollutant Discharge Elimination Sites, freshwater withdrawal estimates, and landscape fragmentation (Falcone *et al*., 2010a). Topographic maps provided visual assessments of dams, intense urbanization, and poor watershed practices that may be unaccounted for in HDI calculations. ADRs provided local expert judgment on the extent of hydrologic alterations above each USGS gauge including dam regulation, diversion, and withdrawals. Typically, at least 15 years of record is suitable for estimating hydrologic variables that are used to detect differences in the spatial variation, such as flow classifications (Kennard *et al*., 2010a). We selected reference gauges having at least a 15-year record.

Reference streams represented only a subset of the gauges available for the USA and left large areas of the landscape void of hydrologic information. On the basis of our judgment, we observed many streams with slight disturbance levels classified as nonreference in the GAGES II database. Furthermore, many nonreference streams had HDI values lower than those of reference streams, potentially indicating low hydrologic alteration. Thus, we attempted to identify additional nonreference gauges with minimal hydrologic disturbances to include in the classification. We define minimal hydrologic disturbance gauges as streams with no major upstream impoundments on the mainstem (Falcone *et al.*, 2010b), low total dam storage from tributaries, little to no diurnal fluctuations, and low urbanization (<15% area) and channelization in the watershed. Because we use only daily averages in hydrologic records, we suspect that minor subdaily fluctuations have little effect on hydrologic variables. We also presume that small diversions (defined later) would be insignificant to hydrologic classification because we standardize all magnitude variables by mean daily flow (see Hydrologic Classification section). Minimal hydrologic alteration was determined by a two-step screening procedure. First, we recalculated HDI indices using seven anthropogenic disturbance variables summarized for each watershed, which included 2009 dam storage, major dam density, freshwater withdrawals, percentage of stream distance as canals, percentage of watershed area comprised of agriculture irrigation (%AGIR), watershed fragmentation, and road density. Similar to Falcone *et al.* (2010a), we calculated thresholds for each disturbance variable based on percentiles (1st, 20th, 40th, 60th, 80th, 90th, 95th, and 98th percentiles). We then assigned scores of 1–8 for each, and the sum of the scores was used to calculate an HDI for each stream. We used percentiles (20% increments) to categorize HDI values into low, low-to-moderate, moderate, moderate-to-high, and high categories. We selected only gauges with low and low-to-moderate HDIs with at least a 15-year record (HDI < 14). As a secondary measure, we reviewed ADRs and screening comments for each gauge. We excluded gauges with reports and comments mentioning regulation by a dam upstream, major diversions (irrigation, hydropower, and municipalities), channelization, and agriculture. We included gauges where comments suggested slight diurnal fluctuations, small diversions for municipalities (≤10% of 7-day low flow), and some irrigation for agriculture. Some comments indicated diversions for irrigation for a given amount of agriculture acreage, which provided little information on the extent of hydrologic alteration. We evaluated values of %AGIR for each gauge where irrigation was mentioned to ensure their values fell within an acceptable range on the basis of values for reference and selected nonreference gauges in the same region. As a final measure for gauges with uncertain watershed disturbances, we examined plots of cumulative annual variation in flow versus time to identify apparent changes in streamflow patterns attributable to different anthropogenic stressors similar to methods provided by Vogl and Lopes (2009). Because of limited watershed information, HDIs could not be calculated for Alaska, Hawaii, and Puerto Rico gauges; thus, we had to rely on ADRs and evaluating plots of hydrographs.

Streams with pre-dam regulation hydrologic information were selected using the GAGES II database and the National Inventory of Dams (NID) (USACE, 2012). The NID has information for over 85 000 dams in the USA including their purpose, dimensions, storage capacity, contributing drainage area, river name, and year built. First, we chose previously unselected gauges from GAGES II with ≤10 major dams in their upstream watershed and with the mention of USGS station name or ADR: (1) dam regulation as the only source of hydrologic modification and (2) the specific name(s) of the dam contributing to regulation (e.g. Pound River below Flannagan Dam near Haysi, VA, gauge no. 03209000). Major dams are considered dams with ≥15-m heights or ≥6167-Ml storage capacity (5000 acre feet) (Falcone *et al.*, 2010b). ADRs typically mention the year in which dam regulation was initiated. However, for cases without information on initial year of regulation, we searched the NID database to find specific dams and their year of completion. We then generated a reduced list of regulated gauges with at least a 15-year record prior to dam construction (accounting for at least 2 years of pre-dam completion).

Although ADRs provide specific information on the dominant regulatory dam(s) upstream of each gauge, they may not mention cumulative regulation by other smaller, earlier constructed dams in the basin. We used spatial data to account for upstream regulatory dams and their year of construction. Within the NID database, we selected dams with heights >6 m to avoid including off-channel farm/holding ponds and minor impoundments. We imported the latitudes and longitudes of all selected gauges and dams in the NID into ARC GIS 9·3. Falcone, 2011 delineated watershed boundaries for each USGS gauge within the GAGES II database. We used a spatial join procedure to link all upstream dams in the NID to each gauge's watershed in ARC Map 9·3. We assessed each gauge individually to determine whether each record preceded the construction of other dams besides those mentioned in the ADRs. As a secondary measure for gauges with uncertain watershed disturbance, we used historical trends in urbanization (Brown *et al.*, 2005) and water use estimates (Kennen *et al.*, 2009) to estimate levels of disturbance for periods prior to dam construction (see McManamay *et al.*, 2012a, for methods). In addition, we reviewed plots of hydrologic records pre-regulation and post-regulation to ensure that apparent changes were associated with periods following our unregulated record.

Hydrologic classifications

Mean daily streamflow data for all stream gauges were downloaded from the USGS National Water Information System. Daily flow data for each gauge (entire record or entire pre-dam regulation record) were imported into the Hydrologic Index Tool (HIT) software (Henriksen *et al.*, 2006), which calculates 171 hydrologic indices reported by Olden and Poff (2003). Indices are grouped into five categories of flow including magnitude, frequency, duration, timing, and rate of change. Olden and Poff (2003) found that hydrologic variables tend to be correlated, leading to a redundancy of information; however, they also reported that the indicators of hydrologic alteration (IHA) variables explained the majority of information in 171 hydrologic indices. We reduced the variables to 110 indices that included the 32 (IHA) (Richter *et al.*, 1996), 32 range-of-variability variables associated with the IHA metrics (Richter *et al.*, 1997), and variables used in previous classifications (Poff, 1996; McManamay *et al.*, 2012b). Hydrologic variables and their descriptions are available in Table S1.

We divided all variables related to magnitude (e.g. mean January flow) by the mean daily flow to standardize all streams by river size (Kennen *et al.*, 2007; Kennard *et al.*, 2010b). We use mean daily flow rather than median daily flow because some streams were highly intermittent and had zero-flow values (Kennard *et al.*, 2010b). Zero flow in highly intermittent streams resulted in missing values for low-flow-related hydrologic variables (e.g. baseflow index and variation in short-duration low flows). In the case of highly intermittent streams, the dataset was scanned for missing values to determine if streams needed to be removed or if values could be estimated or calculated from other hydrologic variables. For example, the baseflow index is calculated as the 7-day minimum flow divided by the mean daily flow. Other variables could be estimated by linear regression using one or multiple hydrologic variables. For example, variation in low flows displayed inverse relationships to low-flow magnitudes. For missing values from perennial streams, we used the imputeData function (based on mix and mclust packages) in the R programming environment to estimate missing values. The mix package uses a multivariate normal regression probability model on the complete data (observed and missing values) (Schafer, 2012). Multiple imputation is based on Monte Carlo procedures in which missing values are replaced with the average of simulated values (Schafer, 1997). Following imputation, we rescanned the dataset for outliers.

In order to create a gradient from more traditional reference standards to progressively relaxed assumptions, we created three separate datasets that included only reference gauges (reference), reference and nonreference gauges (intermediate), and all reference, nonreference, and pre-dam regulation gauges (expanded). We performed a $\log(x+1)$ transformation on all variables and then performed a principal components analysis (PCA) based on correlations to reduce the dimensionality of each dataset. All variables were scaled by maximum values and centred from 0 to 1 prior to PCA. Following PCA, we reduced the reference, intermediate, and expanded datasets to 13 principal components (PCs) that explained 85%, 89%, and 89% of the overall variation, respectively. We used scores from the 13 PCs for classifications. Statistical significance of all PCs was determined using the broken-stick rule (Jackson, 1993). We evaluated the loadings of hydrologic variables on the first four PCs to examine which variables explain the majority of variation in each dataset. The ten hydrologic variables with the highest absolute loadings for each PC were selected (40 total variables). Results of the PCA analysis, including variable loadings, are provided in Table S2.

We conducted classifications separately for the reference, intermediate, and expanded datasets using PC scores. We used the mclust package in the R programming environment to cluster gauges on the basis of component scores (Fraley and Raftery, 2012). Because the underlying distributions and the appropriate clustering algorithm are typically unknown in classification situations, the Mclust function is advantageous in that it assumes a variety of model-based approaches and uses the Bayes criteria to identify the most likely model and number of clusters (Fraley *et al.*, 2012). In addition, the Mclust function uses hierarchical clustering to estimate initial structure and values followed by Gaussian mixture modelling with expectation–maximization parameter estimation. The Bayes information criterion (BIC) is then used to determine the best model (out of ten models with varying covariance structures) and number of clusters on the basis of the largest BIC value (Fraley *et al.*, 2012). The classification process is initialized by hierarchical modelling; thus, it does not require but allows the flexibility of the user to specify a conjugate prior on the number of mixture components (clusters), parameters, or model. Because of the large sample size and unknown mixture components for the current scale (conterminous USA and Alaska, Hawaii, and Puerto Rico), we determined that the initial hierarchical clustering procedure from 1 to a maximum of 20 clusters as an exploratory technique should have precedence over any arbitrary controls.

Cluster diagnostics

Comparing the gain in hydrologic information associated with adding gauges with the costs of relaxing reference data standards can be accomplished using different quantitative measures. For example, within the intermediate and expanded hydrologic classifications, including gauges with slight disturbances or pre-dam hydrologic information could have influenced the underlying multivariate distribution by (1) increasing variability (i.e. dimensionality) due to demographic changes in the representative sample of streams within a class, (2) shifting the dominant hydroclimatic period of record representative of gauges (in the case of pre-dam regulation), and (3) increasing unnatural correlations among variables due to hydrologic disturbance. An obvious first measure is determining whether additional gauges increase the overall dimensionality of the dataset. However, increases in dimensionality may result from including outliers from disturbed sites. Because classifications should consolidate variation, improvements in classifications can be quantified by increases in predictive capacity and decreases in variation due to error. From a management perspective, however, clusters should be stable, and observations should have a high amount of certainty of class membership to support decisions. We provide five quantitative diagnostics to compare the dimensionality, predictive capacity, uncertainty, presence of outliers, and cluster stability among the three datasets.

Dimensionality

We used scores from the 13 components in the PCA analysis for the expanded dataset to explore the dimensionality represented by each subset of gauges: reference, nonreference, and pre-dam-regulated gauges. We calculated PC ranges as the difference in minimum and maximum values for each of the 13 components for each of the three subsets. We then compared whether the PC ranges for nonreference and pre-dam gauges exceeded those for the reference gauges.

Predictive capacity

One approach to compare the predictive capacity of different hydrologic classifications is to determine how well classes predict hydrologic variables. We used a multivariate analysis of variance in R to determine the predictive capacity of hydrologic classes on the 32 IHA variables for the reference, intermediate, and expanded datasets. We used R^2-adjusted values and mean square error to assess the amount of variation explained in the hydrologic data by varying cluster solutions (2–20 classes) for each of the three datasets. All variables were $\log(x+1)$ transformed prior to analysis.

Uncertainty

The Mclust function provides a measure of uncertainty of class membership for each gauge for a given cluster solution (probability that sample is not a member of its assigned cluster). We explored differences in uncertainty among the reference, intermediate, and expanded datasets for varying cluster solutions (2–20 classes). Using only the optimal number of classes for each dataset, we also compared the mean uncertainty for all gauges and only reference gauges among the three datasets using a Kruskal–Wallis test.

Outlier analysis

Determining the multivariate distances of individual stream gauges within classes can also provide an assessment of the robustness of different gauge datasets within hydrologic classifications. We used squared Mahalanobis distances (D^2) to determine how the addition of gauges may have influenced the multivariate distribution and determine outliers (Mahalanobis, 1936). We hypothesized that including gauges with minimal hydrologic disturbances may create unnatural variation, thereby producing more outliers. Secondly, including pre-dam regulation information may increase variation by influencing the distribution of basin sizes represented within hydrologic classes or changing the predominant period of record (e.g. primarily pre-1950s).

Within the expanded hydrologic classification, we calculated D^2 for all stream gauges as the multivariate distance between each gauge and the centroid of the reference gauges within each class. Because D^2 takes into account the covariation among variables, it differs from Euclidean distance in that it is scale invariant (Mahalanobis, 1936). We calculated D^2 using the IHA variables and on the basis of the covariance matrix for all observations within a class. In order to detect outliers, we visually inspected plots of ordered D^2 values against quantiles of the chi-squared distribution within each class to determine any breaks in the distribution (Filzmoser and Gschwandtner, 2012). Outliers were detected as stream gauges deviating from a straight-line relationship (i.e. creating a break). Visually inspecting distributions and using break points can be a robust technique for detecting outliers in comparison with more automated or mathematical solutions (Garrett, 1989; Filzmoser, 2005).

Cluster stability

Assessing cluster stability provides some level of confidence that classifications are robust, i.e. classes will not change if a small number of streams are removed. We used the clusterboot function in R (fpc package) to repeat clustering procedures for the reference, intermediate, and expanded datasets after randomly removing observations from each dataset (Hennig, 2013). We repeated the bootstrapping procedure for 20 simulations for the optimal number of clusters for each dataset under two scenarios: removing 5% of gauges and removing two gauges. For each simulation, a Jaccard similarity index is calculated as a measure of each new cluster and its most similar respective original cluster (Hennig, 2013). We calculated a cluster stability index (CSI) as the average proportion of gauges reassigned to original clusters across the 20 simulations. CSI values <0·5 represent dissolved clusters whereas clusters with values >0·6 indicate true patterns (Hennig, 2008). Stable and very stable clusters have CSI values of 0·75 and 0·85, respectively.

Class descriptions, geographic affiliation, and class comparison

For simplicity, from here on, we only focus on the reference and expanded classifications. In addition, the reference and expanded classifications represent opposing endpoints of strict reference quality versus relaxed assumptions. In order to provide ecologically relevant names for each class, we evaluated patterns in 15 hydrologic metrics (box-and-whisker plots) among classes in the reference and expanded classification. The 15 variables included a subset of IHA indices (12) and were selected because of their ease of interpretation and documented ecological significance. We plotted class effect sizes simultaneously in order to visually isolate classes displaying the most divergence and hydrologic variables explaining the greatest amount of variation (effect sizes calculated as difference of each class mean from grand mean divided by pooled standard deviation, SD). We mapped all gauges by their latitude and longitude according to class membership to assess spatial patterns, which further aided in class depiction.

We compared both reference and expanded classifications with those of Poff (1996) to determine the degree of association among groups given increased sample size and dimensionality. We obtained the 1996 US flow classification dataset directly from N.L. Poff after request; however, it is also publicly available through Poff and Allan (1993). We used a chi-squared analysis to test whether the two classifications were statistically independent (i.e. not associated) and Cramér's V to determine the degree of association among the classifications (Cramér, 1999). In order to provide some relevance for comparison, we calculated statistics for association among the updated classification. We then evaluated individual class membership and association among classes.

Hydrologic and landscape predictive models

Many studies have utilized hydrologic models (Poff, 1996; Kennard *et al.*, 2010b; McManamay *et al.*, 2012c) and landscape models (Sanborn and Bledsoe, 2006; Kennard, 2010b; McManamay *et al.*, 2012c) to predict class membership. Developing tree-based classification models using hydrologic and landscape information can provide (1) an assessment of predictor importance in discriminating among classes (Kennard *et al.*, 2010b), (2) a mechanistic understanding of the structure or hierarchy of variable importance at different scales (McManamay *et al.*, 2012c), and (3) tools to predict class membership for ungauged streams or for hydrologic information not included in the classification. Hydrologic models can predict class memberships as hydrologic data become available, such as simulated hydrographs or gauge records excluded from the original classification that mature to a suitable length. In addition, hydrologic models can predict changes in class membership induced by changing hydrologic regimes due to disturbance or climate change (e.g. Liermann *et al.*, 2012). In contrast, landscape models can predict class membership in the absence of hydrologic information, thereby providing an approach to extrapolate membership to the landscape (Snelder *et al.*, 2009).

We constructed hydrologic and landscape models to predict class membership for both the reference and expanded classifications using random forests in R. Random forests are an improved form of exploratory learning over traditional tree-based approaches in that they have increased classification accuracy and robust approaches to estimating variable importance (Breiman, 2001; Cutler *et al.*, 2007). Random forests improve accuracy by generating a large number of classification trees (typically 500) and then combining the predictions from all trees. Each tree is generated from a random subset of variables and a bootstrap subsample of the data (63% of observations). The remaining samples [out-of-bag (OOB) observations] are used in a cross-validation procedure to calculate misclassification rates (OOB error rate) and variable importance. To calculate variable importance, values of each variable for OOB observations are randomly permuted and then predicted using each tree. Variable importance (i.e. mean decrease in accuracy) is calculated as the difference in misclassification rates between randomly permuted OOB data and the original OOB divided by the standard error of all misclassification rates. For more information, refer to Breiman (2001) and Cutler *et al.* (2007).

For the hydrologic classification trees, we used two datasets: (1) the 32 IHA variables and (2) all 110 hydrologic variables (Table S1). Because the IHA variables are widely used and have been shown to explain the majority of variation in available hydrologic metrics (Olden and Poff, 2003), this provided an opportunity to determine the predictive capacity of IHA relative to other hydrologic metrics. Magnitude-related hydrologic variables were standardized by mean daily flow. We accessed information on climate, basin topography, and soils for all gauges using the GAGES II dataset (Falcone, 2011). We originally identified 77 landscape predictors, which were reduced to 52 after removing correlated variables. Variables that were more interpretable or were hypothesized to provide a better mechanistic understanding of how landscape factors structure hydrologic were retained over their correlated counterparts. We constructed random forests to predict reference and expanded class membership using only hydrologic or landscape variables. We used the OOB error rate to compare the predictive capacity among models. In addition, we used variable importance to compare variables with the highest explanatory power.

Utility of the hydrologic classification framework

We developed an approach to determine the degree of hydrologic alteration in gauges on the basis of class membership. In short, our approach consists of (1) assigning disturbed gauges to appropriate hydrologic classes on the basis of landscape predictive models and (2) using multivariate measures to assess deviation from class centroids and determining outliers. We assigned dam-regulated gauges (identified from the Methods section) to one of the expanded hydrologic classes using the landscape predictive model (random forest) developed in the previous section. We obtained post-dam construction hydrologic data for each regulated gauge and calculated 171 hydrologic indices using the Hydrologic Index Tool. We used two approaches to determine outliers. First, we applied the hydrologic predictive model (random forest from the previous section) to assign dams to classes on the basis of their current hydrologic regime. The second approach was similar to that mentioned in the Uncertainty and Outlier Analysis sections and consisted of calculating D^2 and determining outliers. D^2 was calculated as the distance between each regulated gauge and the centroid of the expanded gauges within each class. Because of class size, we based D^2 only on the 36 hydrologic variables (from the Uncertainty and Outlier Analysis sections) and the covariance matrix for all observations within a class (regulated and unregulated gauges). Outliers were detected as stream gauges deviating from a straight-line relationship in plots of ordered D^2 values versus chi-squared quantiles.

Falcone (2011) provided total dam storage per unit area (i.e. Ml km^{-2}) for each stream gauge and the straight-line distance from each gauge to the nearest major dam. Falcone (2011) also digitized mainstem stream reaches for each gauge and calculated stream sinuosity ratios as the curvilinear distance of each reach divided by the straight-line distance. We estimated the distance downstream in river kilometres from the nearest major dam to each gauge by multiplying the straight-line gauge-to-dam distance by stream sinuosity. We used generalized linear models to determine the effects of total dam storage and distance from dam on D values (i.e. normalized D^2 values) using Gaussian distribution. We performed a $\log(x+1)$ transformation on both explanatory variables.

RESULTS

Identifying gauges for hydrologic classification

Overall, there were 1715 gauges identified as reference streams by Falcone, 2011 with at least 15 years of record in the continental USA (Figure 2, Table S1). HDI scores for 9067 gauges (reference and nonreference) in the conterminous USA ranged from 1 to 38 (Figure 2). Of the 7120 nonreference streams in the conterminous USA, we identified 1923 with HDI values <14 (low and low-to-moderate disturbances). We evaluated ADRs for 145 nonreference Alaska, Hawaii, and Puerto Rico gauges to determine their inclusion in the classification. We selected a total of 674 nonreference gauges to include in the classification, 11 of which were from Alaska, Hawaii, or Puerto Rico (Table S1). We isolated 1180 gauges currently regulated by dams. Of these, we selected 229 with at least 15 years of hydrologic information pre-dam construction (Figure 2 and Table S1). Hence, the reference, intermediate, and expanded datasets were composed of 1715, 2402, and 2618 stream gauges, respectively.

Reference and nonreference streams showed consistent distributions in their hydrologic records; however, streams with pre-dam regulation information had considerably shorter periods of record (Figure 3). Reference and nonreference streams also displayed considerable temporal overlap in their records. Pre-dam-regulated streams displayed a shift in their records that only overlapped reference and nonreference gauges with long-term records.

Hydrologic classifications

For the reference dataset, eigenvalues for the first nine PCs exceeded eigenvalues from random data, suggesting statistical significance (Table S2). Only the first six PCs were significant for the intermediate and expanded datasets. Hydrologic variables with the highest loadings were similar among the three datasets. Baseflow indices, low and high flows of various duration, and monthly flow indices (averages and maxima) dominated the variables with highest loadings (Table S2). At least 15 IHA variables were represented among the 40 variables with the highest loadings within in each dataset.

For the reference classification, the best model (maximum BIC) occurred at 19 components with an ellipsoidal, equal shape (VEV) model; however, the VEV model displayed insignificant increases in BIC over the ellipsoidal, varying volume, shape, and orientation (VVV) model (Figure 4A). For the intermediate classification, the maximum BIC occurred at 15 components with VEV and VVV models whereas for the expanded classification, VEV or VVV models with 20 components had the maximum BIC values (Figure 4A). With the exception of the intermediate classification, we observed that BIC plots displayed increasing values with increasing numbers of components until each reached a plateau. Following the plateau, BIC values became unstable (increasing/decreasing) prior to reaching maximum values. In favour of a more parsimonious model, we truncated the number of classes to the number of components in which the local BIC maxima occurred following the initiation of the plateau. Following the plateau, we determined that 12, 15, and 15 components emerged as a local stable maximum for the reference, intermediate, and expanded classifications, respectively. Although the VEV model retained the highest local BIC value for the reference and intermediate classifications, the expanded classification displayed a local BIC maximum for the ellipsoidal, varying volume, shape, and orientation (VVV) model.

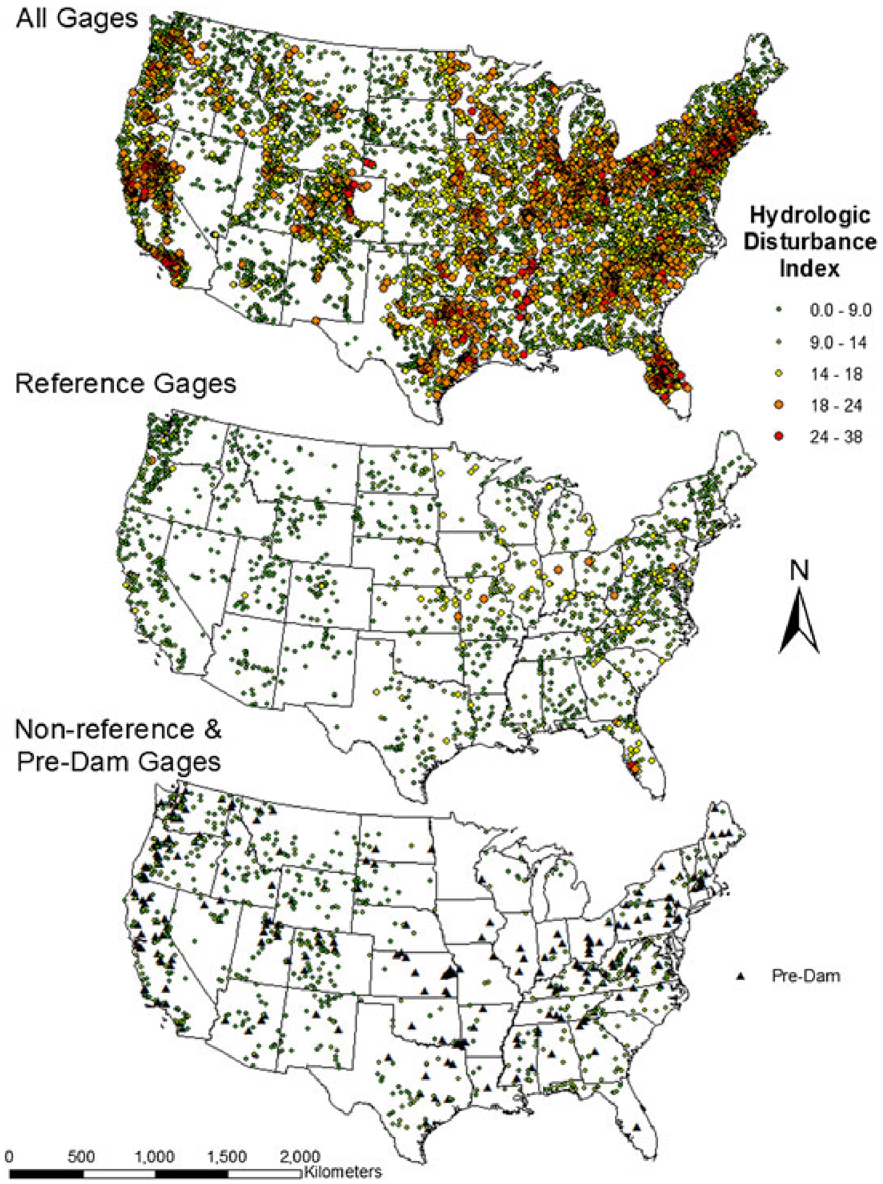

Figure 2. Hydrologic disturbance indices (HDIs) for all gauges, reference gauges, and selected nonreference gauges in our study. HDIs were not calculated for pre-dam regulation gauges because their record preceded the creation of geospatial landscape disturbance data. Likewise, Alaska, Hawaii, and Puerto Rico gauges lacked geospatial data needed to calculated HDIs.

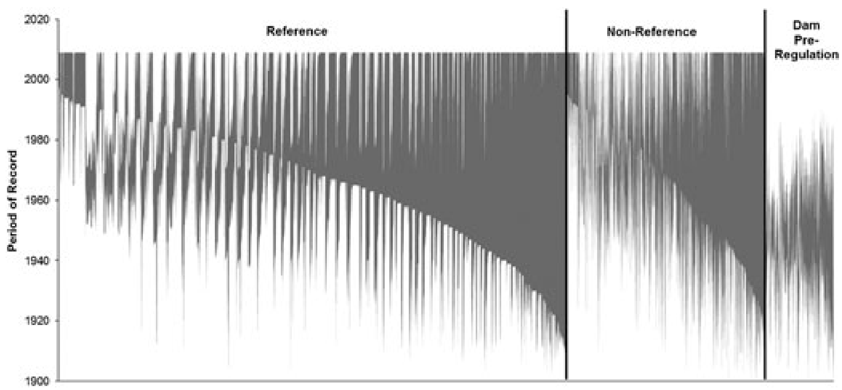

Figure 3. Period of record (total time span) of all reference ($n = 1715$), nonreference ($n = 674$), and pre-dam regulation gauges ($n = 229$) used in our study.

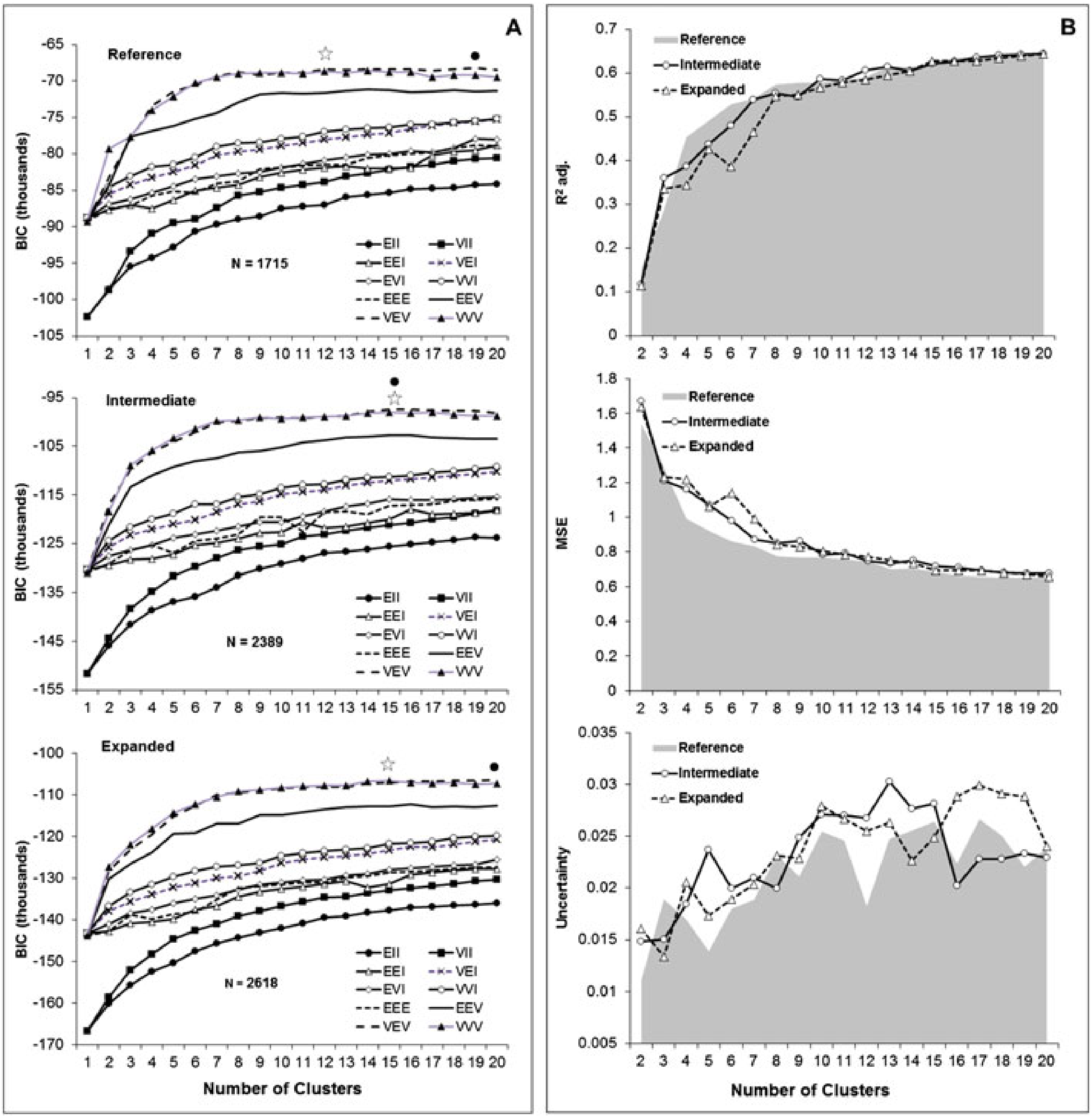

Figure 4. (A) Bayes information criteria (BIC) plots used to determine the best model and appropriate number of mixture components (clusters), which occur at the maximum BIC value. Stars indicate the models and associated clusters with a local BIC maximum value following the plateau, which suggests that additional components create some instability and do not explain considerable amounts of additional variation. Model names: spherical = EII (equal volume); VII (unequal volume); diagonal = EEI (equal volume, shape); VEI (varying volume, equal shape); EVI (equal volume, varying shape); VVI (varying volume, shape); ellipsoidal = EEE (equal volume, shape, orientation); EEV (equal volume, shape); VEV (equal shape); VVV (varying volume, shape, orientation). Stars and dots represent local peaks and maxima in BIC values, respectively. (B) Predictive capacity (R^2 adjusted), mean square error, and uncertainty associated with different cluster solutions for the reference, intermediate, and expanded datasets.

Cluster diagnostics

The dimensionality represented by reference streams increased slightly with the addition of nonreference gauges, but not with the addition of pre-dam regulation gauges (Table S2). PC ranges for nonreference streams exceeded those of the reference streams for the 1st, 7th, 8th, 9th, 10th, and 11th PCs, which cumulatively represented 41·6% of the overall variation. However, in most cases, the difference was minimal. Reference streams had the largest PC ranges in the remaining PCs, which represented 47% of total variation.

Predictive capacity (R^2 adjusted) gradually increased with the increasing number of clusters for the reference, intermediate, and expanded datasets (Figure 4B). However, adjusted R^2 was higher for the reference dataset than for the intermediate and expanded datasets for cluster solutions less than ten classes, after which there were no apparent differences. Similarly, mean square error values were lower for reference data than for the intermediate and expanded datasets for cluster solutions less than ten classes, after which there were no apparent differences (Figure 4B).

Typically, uncertainty showed increases with the increasing number of classes, with intermediate and expanded datasets displaying higher average values than the reference dataset (Figure 4B). Uncertainty was significantly different among the reference (mean = 0·023, SD = 0·077), intermediate (mean = 0·028, SD = 0·082), and the expanded classifications (mean = 0·025, SD = 0·069) for their respective optimal cluster solutions (Kruskal–Wallis, $\chi^2 = 7{\cdot}348$, $p = 0{\cdot}025$). Uncertainty in the intermediate classification was significantly higher than that in the reference classification (pairwise Wilcoxon, $p = 0{\cdot}029$), but not significantly different than that in the expanded classification (pairwise Wilcoxon, $p = 0{\cdot}822$). Uncertainty for only references gauges, however, was not significantly different among the reference, intermediate, or expanded classifications (Kruskal–Wallis, $\chi^2 = 3{\cdot}428$, $p = 0{\cdot}180$). Approximately, 11%, 12%, and 10% of gauges had uncertainties greater than 0·05 in the reference, intermediate, and expanded datasets, respectively. In addition, 7%, 7%, and 9% of gauges had uncertainties greater than 0·10 in the reference, intermediate, and expanded classifications, respectively.

Across the majority of classes in the expanded dataset, average D^2 values were typically higher for nonreference and pre-dam regulation gauges than for reference gauges (Figure 5). In most instances, average pre-dam regulation D^2 values were highest. The proportion of gauges classified as outliers followed a similar pattern to that of D^2 values with higher proportions of nonreference and pre-dam regulation gauges than reference gauges within each class (Figure 5).

Cluster stability was low for all datasets when 5% of observations were randomly removed (Figure 6). For the 5% removal scenario, average CSI values were 0·52 for both the reference and intermediate datasets and 0·60 for the expanded dataset. Cluster stability was dramatically higher when only two random observations were removed from each dataset (Figure 6). For the two-observation removal scenario, average CSI values were 0·78, 0·74, and 0·93 for the reference, intermediate, and expanded datasets, respectively. For the two-observation removal scenario, 10 of the 12 reference classes were indicative of true patterns (CSI > 0·6) whereas only 4 of the 12 reference classes were highly stable (CSI > 0·85). For the same scenario, 14 and 15 of the intermediate and expanded classes, respectively, were indicative of true patterns. For the same scenario, all 15 expanded classes were highly stable whereas only five intermediate classes were highly stable.

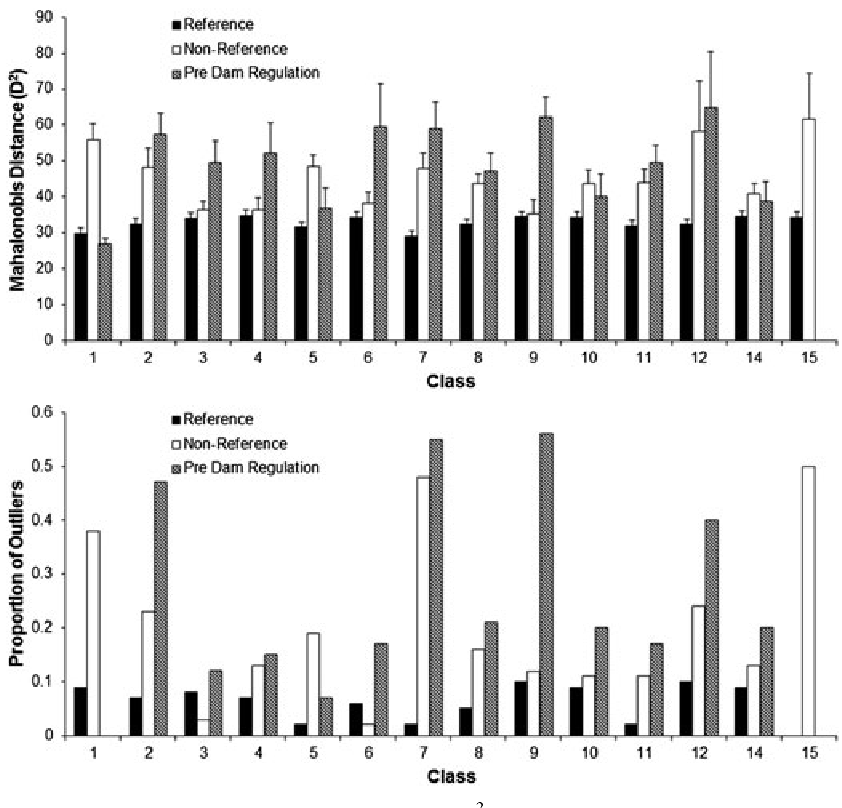

Figure 5. Levels of uncertainty represented by squared Mahalanobis distances (D^2) and the proportion of outliers for reference, nonreference, and pre-dam gauges within each class. D^2 values were calculated as distances from the multivariate centroid occupied by reference gauges in the expanded classification. Outliers were calculated on the basis of breaks in D^2 distributions (see Methods section).

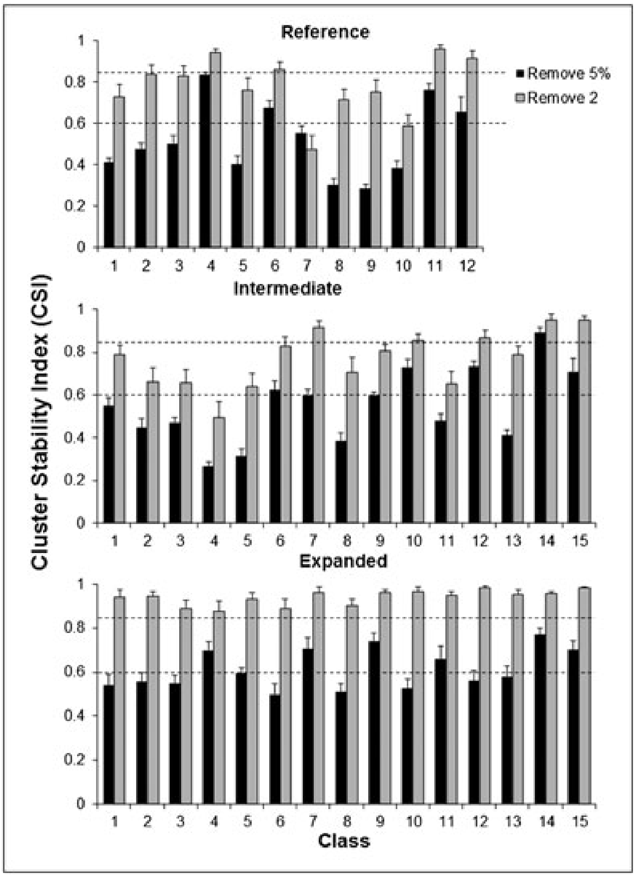

Figure 6. Cluster stability index (CSI) of reference, intermediate, and expanded classes for two simulations, randomly removing 5% of observation and two observations from each dataset. The lower dashed line (CSI = 0·6) is indicative of a threshold above which true patterns in the data emerge. The upper dashed line (CSI = 0·85) represents a threshold for highly stable clusters.

Class descriptions, geographic affiliation, and class comparison

Within the reference and expanded classifications, distinct classes emerged ranging from unstable intermittent streams to stable high-runoff streams (Figure 7 and 8, Tables I and II). The expanded classes were either expansions or merging of one or more existing reference classes. For example, the reference classification had four intermittent-type stream classes whereas five intermittent classes were represented in the expanded classification (Figure 7 and 8, Tables I and II). The hydrologic variables displaying the most distinction among classes in both classifications were daily variability, intermittency, maximum flows, and rise rates (Figure 9). One exception was that baseflows and low flows seemed to show more class separation in the expanded dataset (Figure 9).

For the reference classification, classes displayed some regional affiliation; however, there was substantial spatial overlap (Figure 7). Four classes showed varying degrees of duration of intermittency and high flows (Table I). The harsh intermittent class had extremely long intermittent periods punctuated by episodic flood events and was primarily situated in the south-west. The intermittent flashy (IF) class had extended periods of no flow punctuated by smaller, longer-duration flood events and very little geographic affiliation. Unpredictable intermittent (1–2) classes had low predictability and semi-flashiness (moderate frequency of high flows) but differed in geography and timing of flow events. Perennial runoff (PR) 1 and 2 streams had moderate stabilities and distinct seasonal extremes (high winter/spring and low summer/fall). PR1 streams had higher baseflows than PR2 and were found in eastern piedmont regions whereas PR2 streams had higher total runoff and were found in Appalachian regions. Stable high baseflow streams (SHBF) were situated in the south-eastern Blue Ridge Mountains and Pacific Northwest, both of which are characterized by high precipitation, sustained high baseflows, and moderately high runoff. Snowmelt (SNM) 1 and 2 streams had distinct periods of late-spring maximums during the initial SNM followed by receding flows through the summer. SNM1 streams were isolated to the eastern Rocky Mountains and had higher stability, higher baseflows, and early annual minimums (during winter) than SNM2 streams. Super-stable groundwater streams were broadly dispersed and characterized by high baseflows and high stability, both indications of spring-fed systems. Coastal high-runoff (CHR) streams had later annual maximums and slightly higher reversals (affected by tide) and were affiliated with coastal areas of the south-east, Alaska, and Hawaii. Western CHR (WCHR) streams were different than CHR streams in that WCHR had lower baseflows, late annual maximum flows, and very distinct seasonal flows (similar to PR streams).

Compared with the reference classification, the expanded classification tended to show more distinct regional affiliation (Figure 8). For example, PR1 and PR2 streams were more segregated to the south-east and north-east, respectively (Table II). Similarly, the SHBF streams in the reference classification split into two separate classes in the expanded classification: an SHBF class in the south-eastern Blue Ridge Mountains and a stable high-runoff (SHR) class in the Pacific Northwest (Figure 8). The SHBF class had a higher baseflow index whereas SHR streams had higher runoff. In the reference classification, CHR streams were split into late-timing runoff (LTR) and glacial high-runoff (GHR) streams after the inclusion of more gauges. LTR and GHR showed more regional affiliation and differed in the timing of flow events. GHR streams tended to have higher annual runoff and higher predictability than LTR streams. IF and unpredictable intermittent 2 streams in the reference classification were split into IF, IF south-west, and IF2 streams. In general, most classes in the expanded classification could be interpreted similarly to that of the reference classification with varying degrees of intermittency, stability, baseflows, runoff, and seasonality (Table II).

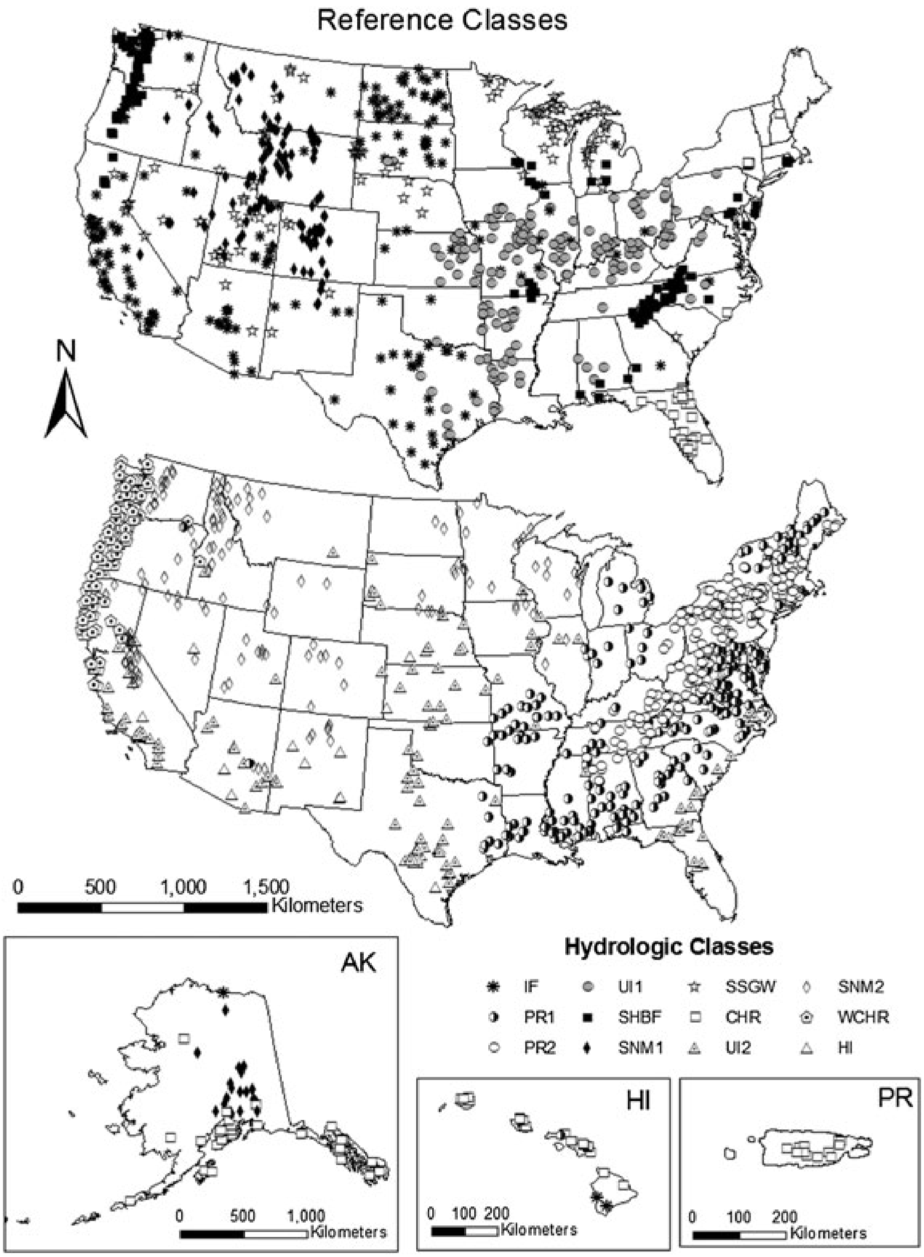

Figure 7. Geographic locations of 12 reference classes across the continental USA.

Reference and expanded classes were not statistically independent (Pearson's $\chi^2 = 10\,552$, $df = 154$, $p < 0{\cdot}0001$) and showed strong association (Cramér's $V = 0.748$). Only 428 of the 806 gauges from Poff (1996) were represented in our reference classification, and only 521 were represented in our expanded classification. Reference classes and Poff (1996) classes were not statistically independent (Pearson's $\chi^2 = 1244$, $df = 99$, $p < 0{\cdot}0001$) but only displayed moderate association (Cramér's $V = 0{\cdot}568$). Likewise, expanded classes and Poff (1996) classes were not statistically independent (Pearson's $\chi^2 = 1591$, $df = 126$, $p < 0{\cdot}0001$) and displayed slightly higher association (Cramér's $V = 0{\cdot}583$). Reference and expanded classes tended to be extensions of Poff (1996) classes showing some degree of affiliation; however, there was a large degree of overlap, with Poff (1996) classes being represented by multiple updated classes (Figure S3). For example, groundwater streams (GW) streams were predominately represented by SHBF streams in the reference classification; however, PR1, SNM2, and super-stable groundwater streams also made up a considerable proportion of the GW class (Figure S3). Likewise, PR streams

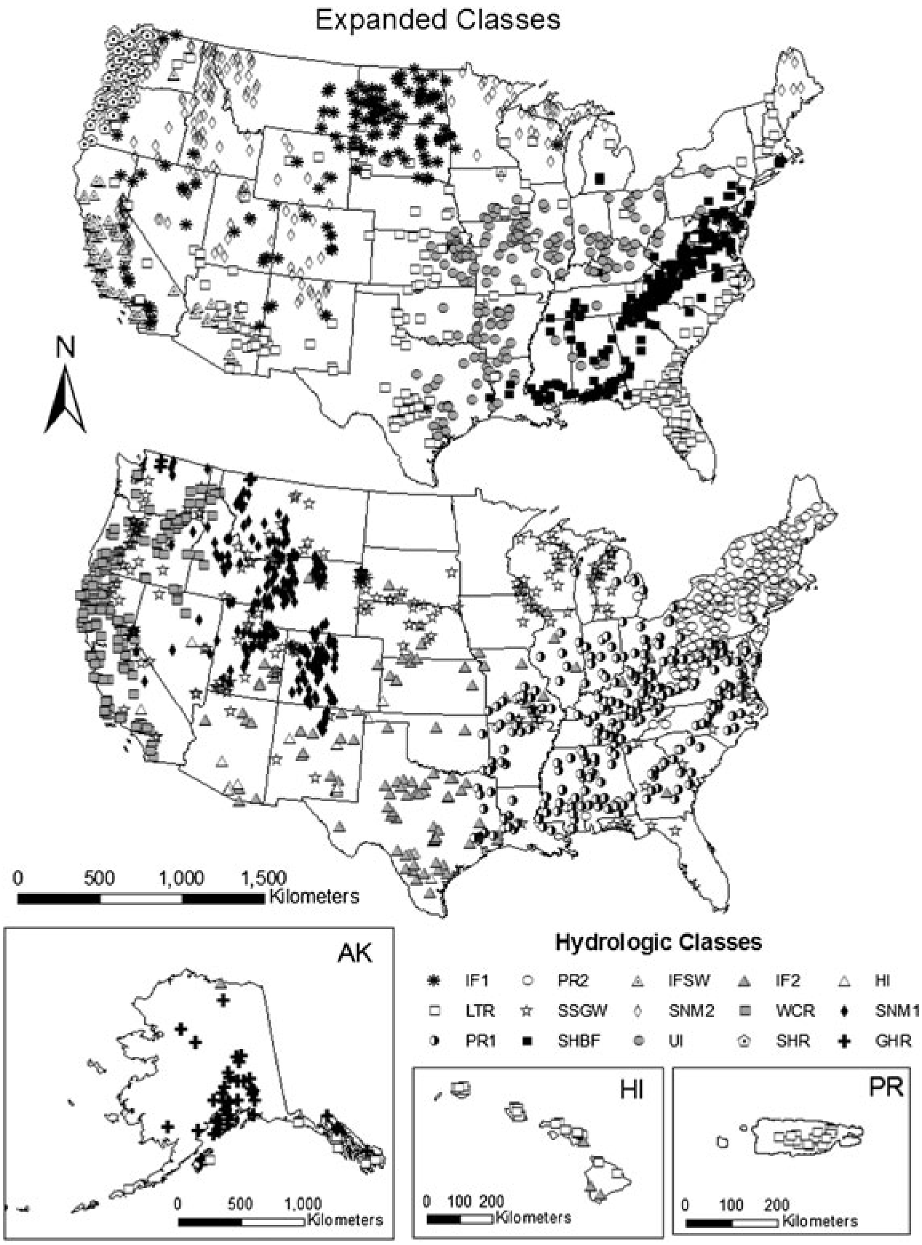

Figure 8. Geographic locations of 15 expanded classes across the continental USA.

were primarily represented by PR1 and PR2 streams in the reference class yet also shared membership with seven other reference classes.

Hydrologic and landscape predictive models

The predictive capacity of random forests for both the reference and expanded datasets was comparable. In addition, OOB error rates for hydrologic models built using only IHA variables were comparable with those built with all 110 variables. For the reference classification, the OOB error rate was 12·9% for the IHA variables compared with 9·95% for all 110 hydrologic variables. For the expanded classification, the OOB error rate was 11·1% for the IHA variables compared with 9·34% for all 110 hydrologic variables. IHA variables with higher mean decrease in accuracy values were also similar between the classifications and included high-flow frequency, average monthly flow indices, and low/high flows of various durations (Figure 10). The timing of high-flow events was more important than the timing of low flows in the expanded classification whereas the reference classification showed the opposite pattern. Low flows of various durations were more important than high flows in the reference classification whereas the opposite was true for the expanded classification (Figure 10).

Important landscape predictors were also very similar among the reference and expanded classifications (Figure 11).

Table I. Reference hydrologic class names, codes, descriptions, and geography.

Class	Name	Code	Characteristics	Geography
1	Intermittent flashy	IF	High intermittency, long high-flow duration	Variable
2	Perennial runoff 1	PR1	Similar to SHBF but lower baseflows, semi-stable	Eastern piedmont
3	Perennial runoff 2	PR2	Similar to PR2, but lower baseflows, higher runoff than PR1	Eastern Appalachians
4	Unpredictable intermittent 1	UI1	Moderate intermittency, low predictability, and semi-flashy flows	Eastern USA
5	Stable high baseflow	SHBF	High baseflows (smaller than SSGW), stable, and relatively high runoff	Blue Ridge Mountains and Pacific Northwest mountains
6	Snowmelt 1	SNM1	Distinct and consolidated periods of runoff, stable and relatively high baseflows, early annual minimum (winter freeze)	Eastern Rocky Mountains
7	Super-stable groundwater	SSGW	Very high baseflow, high stability, not necessarily high runoff	Variable (spring-fed systems)
8	Coastal high runoff	CHR	High runoff, very late annual max, and very early annual min; slightly higher reversals (potential tide effects)	Coastal areas (south-east, Alaska, Hawaii)
9	Unpredictable intermittent 2	UI2	Moderate intermittency, low predictability, and semi-flashy flows (different timing and runoff than UI1)	Variable
10	Snowmelt 2	SNM2	Less stable and lower baseflow than SNM1 otherwise similar	Western Rocky Mountains, Great Lakes
11	Western Coastal high runoff	WCHR	Distinct wet/dry seasons, lower baseflow than PR streams, but very high runoff, early annual maximum	Western coast
12	Harsh intermittent	HI	Very long periods of intermittency punctuated by episodic flows	South-west

Table II. Expanded hydrologic class names, codes, descriptions, and geography.

Class	Name	Code	Characteristics	Geography
1	Intermittent flashy 1	IF1	High intermittency, long high-flow duration	Variable but centralized to ND and SD
2	Late-timing runoff	LTR	Semi-stable, late annual maximum (different reasons for late timing in various geographical areas – e.g. hurricanes (FL) and rainfall (Western))	Coastal areas but mixed inland
3	Perennial runoff 1	PR1	Similar to PR2, but lower baseflows, higher variability and different timing than PR2, semi-stable	South-eastern piedmont and mountains
4	Perennial runoff 2	PR2	Similar to SHBF but lower baseflows, higher baseflow and different timing than PR1, semi-stable	North-east piedmont and mountains
5	Super-stable groundwater	SSGW	Very high baseflow, high stability, not necessarily high runoff	Variable
6	Stable high baseflow	SHBF	High baseflows (smaller than SSGW), stable, and relatively high runoff	South-eastern Blue Ridge Mountains
7	Intermittent flashy SW	IFSW	High intermittency, long high-flow duration, short low-flow duration	South-west California and Arizona
8	Snowmelt 2	SNM2	Similar to SNM1, but slightly higher baseflow, different timing of max flows and flow durations	Western Rocky Mountains, Great Lakes
9	Unpredictable Intermittent	UI	Moderate intermittency; low predictability, semi-flashy; later annual maximum than PR2	Eastern USA
10	Intermittent flashy 2	IF2	High intermittency, very late maximum	Variable
11	Western coastal runoff	WCR	Distinct wet/dry seasons, early maximum, high runoff	Western Coast (California)
12	Stable high runoff	SHR	Very high runoff, stable, moderate baseflows – lower than SHBF	Upper Western Coast (Washington and Oregon)
13	Harsh intermittent	HI	Extremely long periods of intermittency punctuated by episodic flows	South-west
14	Snowmelt 1	SNM1	Distinct and consolidated periods of runoff; stable, similar to SHBF; early annual minimum (winter freeze); slightly lower BF than SNM2, different timing max flows and f low durations	Eastern Rockies
15	Glacial high runoff	GHR	Highly predictable and high runoff, annual minimum occurs in winter (freeze), very long high-flow duration	High-elevation areas (Montana and Wyoming) and arctic (Alaska)

Landscape predictive models had OOB error rates of 25·48% for the reference classification and 24·15% for the expanded classification. Temperature and precipitation measures (primarily monthly averages) were the most important predictors for both classifications, whereas soils and soil hydrology variables were less important (Figure 11).

Utility of the hydrologic classification framework

Fourteen of the 15 expanded hydrologic classes were represented by dam-regulated gauges (Figure 12). Of the 1180 dam-regulated gauges, 229 gauges had pre-dam regulation information, all of which were correctly assigned to their original class using the landscape predictive model. A total of 581 gauges were classified to a different class (i.e. outlier) using the hydrologic predictive model (random forest) whereas only 329 were classified as outliers using breaks in D^2 values (Figure 12). Total dam storage had significant positive effects on D values but explained only 10% of the total variation (Figure 13). Distance from dam had significant negative effects of D values and explained 11% of the total variation (Figure 13).

DISCUSSION

We developed a series of hydrologic classifications at the scale of the continental USA using three datasets

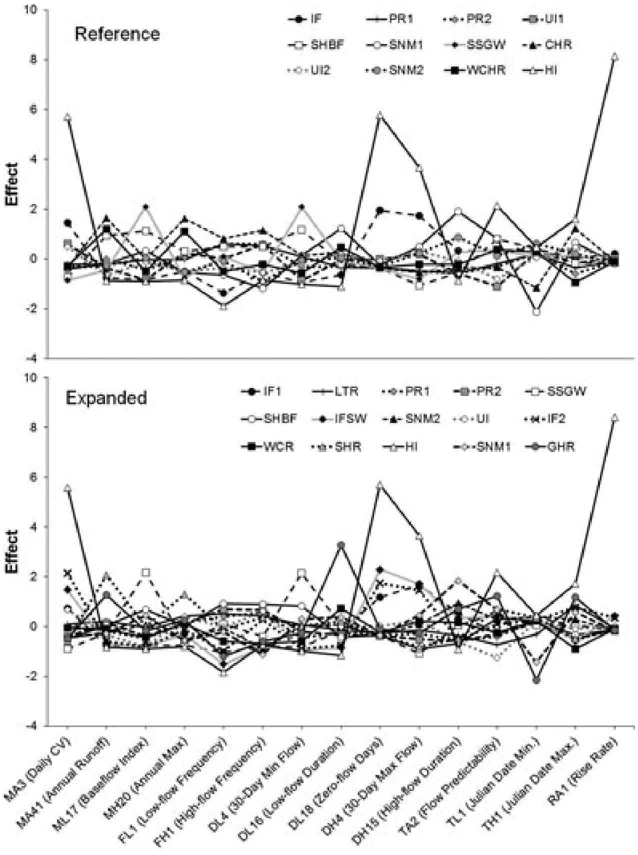

Figure 9. Effects of reference and expanded classes on 15 hydrologic variables. Effect sizes were calculated as (class mean – grand mean)/ pooled standard deviation across all classes.

representing a gradient of strict reference standards to progressively more relaxed assumptions. We presumed that the full variation in natural hydrology across the continental USA may be underrepresented by datasets composed of only strict reference standards. However, including additional streamflow gauges of lower reference quality in classifications led to minimal increases in dimensionality (i.e. hydrologic information) at the expense of increasing uncertainty in class membership. Although the expanded classification created more outliers, increases in streamflow information yielded more hydrologic classes with higher regional affiliation and, surprisingly, higher cluster stability. Thus, we failed to observe clear trade-offs between the information quality and quantity.

Levels of association between our updated classifications to that of the previous US hydrologic classification were weak, which indicated low structural similarity (i.e. poor nested or hierarchical structure) and considerable shared membership in divergent classes. Obviously, the level of hydrologic variation and resolution (gauges per unit area) represented in smaller sample sizes will highly influence classification solutions. However, understanding the influence of spatial extent (continent vs basin) and sample size on hydrologic clustering outcomes, and thus, management decisions, should be an area of future research.

Part of the utility of classification systems rests in their ability to stratify analyses (Wolock *et al.*, 2004). We provide two approaches of assigning gauges to classes and assessing potential outliers due to hydrologic alteration. Because of their multivariate nature, classification systems may be more robust in detecting hydrologically modified systems (i.e. outliers) than analyses evaluating individual hydrologic metrics or short-term patterns. Thus, in a multivariate sense, classes can be used to quantitatively determine if disturbed streams are functioning within the established range of natural streamflow variability (McManamay *et al.*, 2012a).

Hydrologic classifications

Our reference gauge dataset was compiled according to similar standards reported by Poff (1996) and included over two times the sample size (1715 stream gauges). Likewise, by including nonreference gauges and pre-dam regulation gauges in our expanded dataset, we increased the sample size from the reference dataset by over 50% (2618 stream gauges). Within any dataset representing natural patterns, increasing the sample size will increase variation, and hence dimensionality. Underrepresented streams of unique hydrologic character in smaller datasets may be manifested as new and separate classes in datasets of larger size; thus, we presumed that classifications created using smaller datasets may display nested structure within classifications created from larger datasets (McManamay *et al.*, 2012b). Although our classes showed some evidence of nested structure with those of Poff (1996), we observed a large degree of shared membership among divergent multiple clusters. Likewise, even though our reference and expanded classifications shared 66% of gauges and were constructed using the same methods, association values were still well below 1 (Cramér's $V = 0.78$). In summary, clustering outcomes utilizing updated information are uncertain and may not predictably nest within former classifications. Thus, updating classifications as new information becomes available is practically important.

Approaches to classifying the hydrology of streams have varied considerably in terms of the underlying data used in classification and the statistical methods (Olden *et al.*, 2012). Given the availability of hydrologic information, we determined that an inductive approach utilizing stream gauge information would be the most appropriate in accurately representing streamflow patterns across the USA. The appropriate applications of different clustering approaches vary depending on the situation, and the choice of one procedure over another can influence the outcome (Everitt *et al.*, 2001). The most commonly applied clustering procedures for inductive hydrologic classifications to date

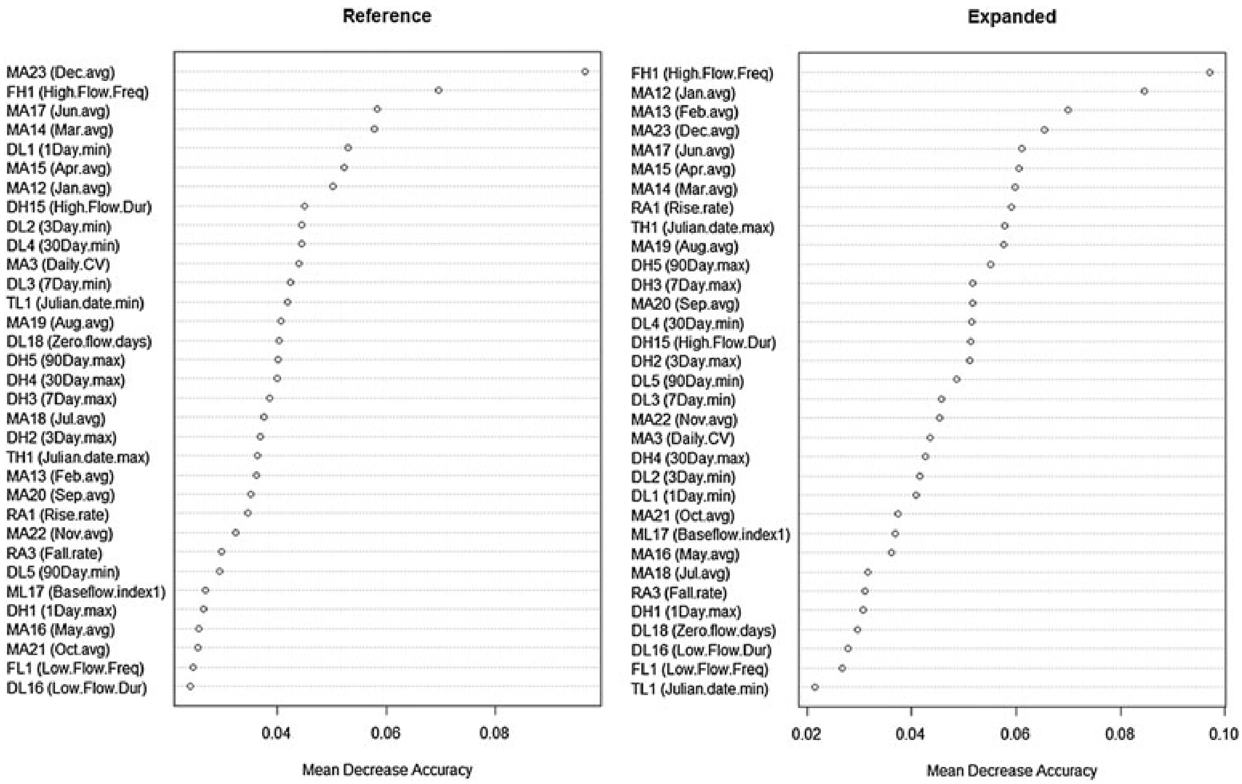

Figure 10. Variable importance of the 32 indicators of hydrologic alteration used in random forests to predict reference and expanded class membership. Variable importance was measured using mean decrease accuracy values (see methods).

have been hard-clustering methods, including hierarchical (e.g. Ward's algorithm) and partitioning algorithms (e.g. *k*-means) (Olden *et al*., 2012). These methods require users to determine *a priori* the number of clusters; however, there are methods to determine the most parsimonious portioning of variation (Everitt *et al*., 2001). Given that streams may share similar properties with multiple classes and the number of appropriate clusters may be unknown, we utilized a Bayesian mixture-modelling approach that applies multiple models and mixture components (clusters) to determine the best solution (Fraley and Raftery, 2007). Applying multiple algorithms is advantageous in situations where the most appropriate model is unknown, given the data. In addition, uncertainty and probability of class memberships can be explicitly reported.

When streams are classified according to patterns in natural streamflows, an obvious first step is screening gauges for inclusion in a final 'reference' dataset (Olden *et al*., 2012). Kennard *et al*. (2010a) concluded that gauges included in spatial analyses should have at least 15 years of record and at least 50% of overlap among gauge records in order to minimize uncertainty among and within hydrologic classifications. Reference and nonreference gauges had similar length and overlap in records; however, pre-dam regulation gauges, although similar in record length, did not share considerable temporal overlap with the majority of gauges. The most extensive difference in temporal overlap was between gauges with shorter records (typically occurring in last two decades) and pre-dam regulation gauges (occurring no later than the early 1980s) (Figure 3). After the reference dataset was augmented with nonreference and pre-dam regulation gauges, the underling variation increased only slightly. This suggests that the reference dataset represented the vast majority of hydrologic variation available and that adding more gauges is questionable unless predictive capacity increases. Although the inclusion of additional lower-reference-quality gauges increased outliers and increased uncertainty, the predictive capacities of cluster solutions were not different among datasets. Ultimately, this suggests that hydrologic classifications may be robust against small disturbances and slight changes in climatic regimes. However, this topic needs to be addressed more fully in future analyses.

Limitations of our approach

Although we attempted to be thorough in our classification approach and in our assessment of relaxing reference data standards, there are multiple limitations to both of these elements. First, the classification approach taken depends largely on objectives and also has management

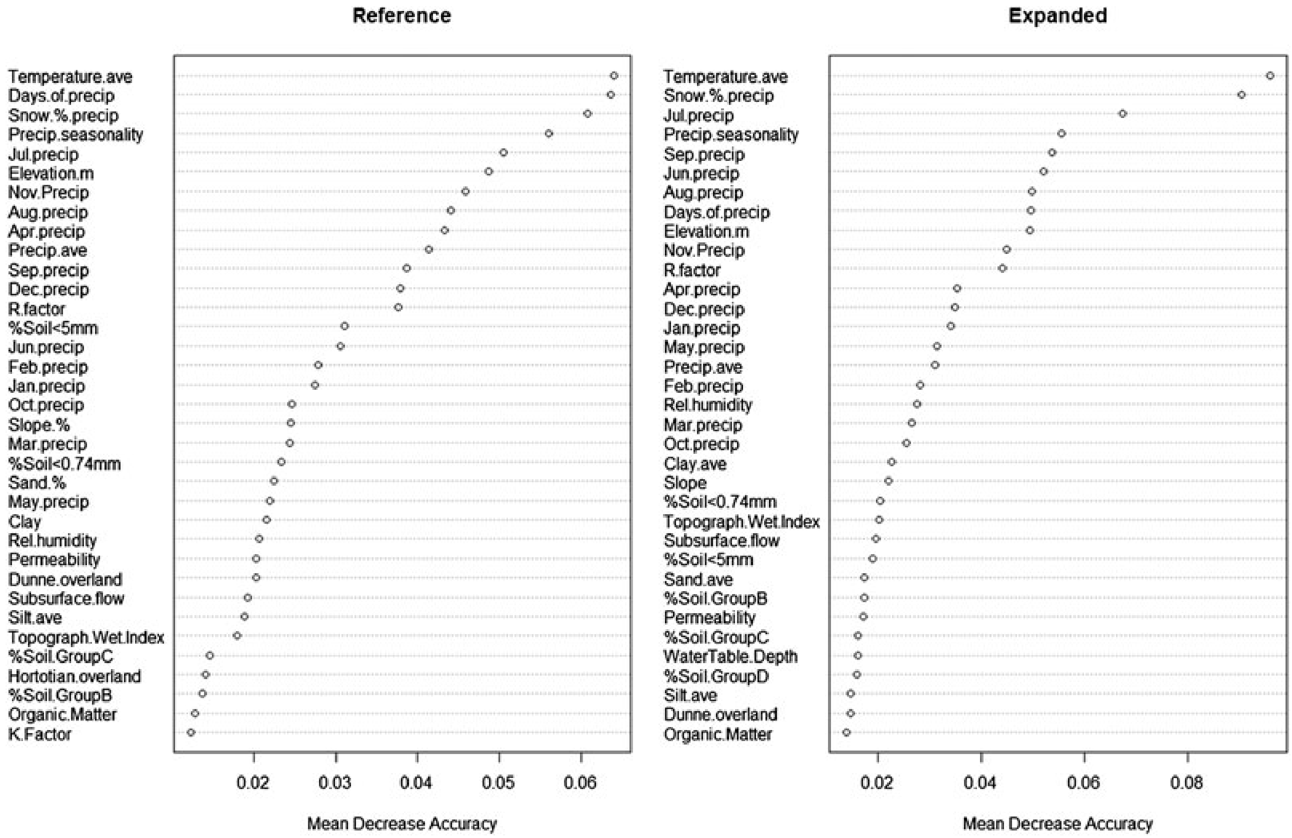

Figure 11. Variable importance of the top 35 landscape predictors (out of 52) used in random forests to predict reference and expanded class membership. Variable importance was measured using mean decrease accuracy values Methods section).

implications (Olden *et al*., 2012). To date, there are an immense number of potential clustering algorithms available (Nathan and McMahon, 1990); however, the choice of algorithm will have highly variable results on the same dataset (Olden *et al*., 2012). For example, partition-clustering approaches, such as *k*-means, are sensitive to the number of clusters and the order of the dataset (Everitt *et al*., 2001). In contrast, hierarchical algorithms, such as Ward's algorithm, are robust against cluster numbers and dataset order but will tend to produce classes of equal size (Everitt *et al*., 2001), which is highly unlikely given hydrologic data (Poff, 1996). The outcome is that different clustering solutions will vary in their stability on the basis of the algorithm applied. Furthermore, all cluster procedures are influenced by the underlying dissimilarity matrix and dimensionality, which can be influenced if observations are removed (Hennig, 2007). Although the Bayesian mixture-modelling approach we applied is a very flexible method, cluster solutions may be unstable because the optimal covariance matrix model selected by BIC may change with the removal of samples. In our case, the optimal covariance matrix models (VVV or VEV) remained consistent for all cluster solutions for the reference, intermediate, and expanded datasets (Figure 4A). With the removal of two observations from each dataset, average CSI values for clusters ranged from 0·78 in the reference dataset to 0·93 in the expanded dataset, which suggests fairly high stability. Comparisons of cluster stability for Bayesian mixture models and hard-based clustering methods have not been reported in literature; thus, it is difficult to ascertain whether mixed-model cluster approaches are more or less stable than other approaches.

Another difference in our approach is that most clustering approaches hard-assign streams to classes, i.e. streams do not share membership with other classes. However, realistically, hydrologic data do not always fall neatly into separate categories, and stream classes may overlap in multidimensional space (Olden *et al*., 2012). Because of the uncertainty associated with class membership and uncertainty in clustering algorithms, we chose to apply a fuzzy clustering procedure, which provides some probability that a given stream could, in fact, belong to all classes. Each observation was then assigned to the most probabilistic class, and uncertainty of misclassification was quite low. The limitation of this approach, however, is not in the probabilistic approach but in the interpretation of results. Typically, managers find uncertainty problematic when associated with classifications, especially hydrologic classes aimed to support environmental flow decisions. In

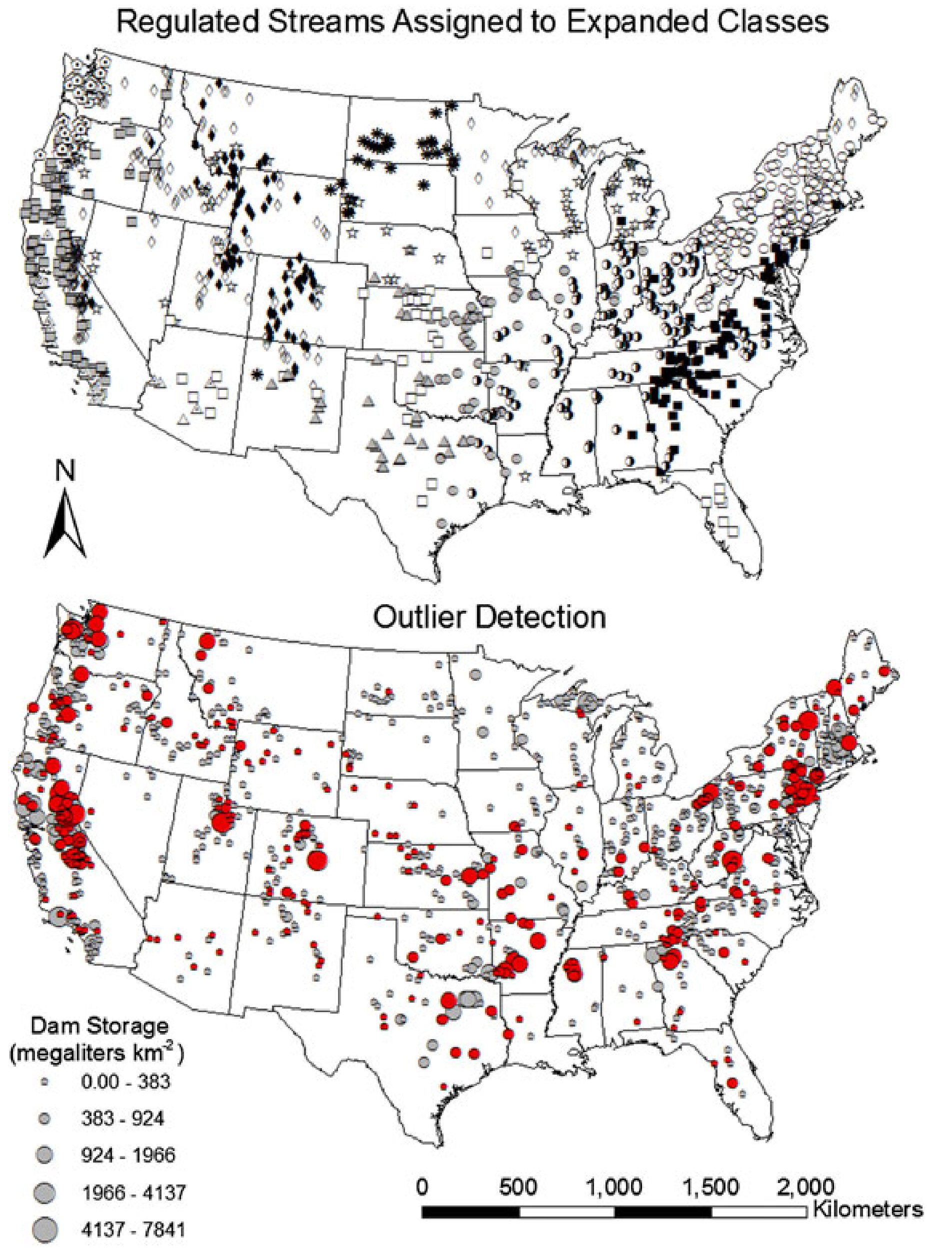

Figure 12. Dam-regulated streams assigned to expanded hydrologic classes using landscape random forest predictive model (top). Gauges designated as outliers (red) within each class in association with total dam storage (bottom).

addition, managers are likely to desire classifications uninfluenced by removing random subsets of the data. Although uncertainty in classification solutions will vary from the clustering algorithm and classification approach chosen, uncertainty will always exist no matter the approach; the key challenge is whether uncertainty is quantified, explicitly reported, and then accepted.

The other main focus of our paper was the balance between maximizing hydrologic information while ensuring reference data standards are met. While we attempted to construct datasets representing a gradient from strict standards to more relaxed assumptions, we admit that a more thorough analysis is required to fully address the hypothesis posed here. For example, a robust analysis aimed at quantifying the trade-offs between data quality and quantity might include developing a stronger gradient of data quality with highly disturbed streams being one endpoint. By using quantitative measures, an optimal solution could then be determined between two opposing endpoints: the highest-quality reference gauge information available versus gauges highly modified by anthropogenic disturbance. Within the current study, the nonreference and pre-dam regulation gauges did not truly represent a highly disturbed endpoint, but a semi-reference endpoint.

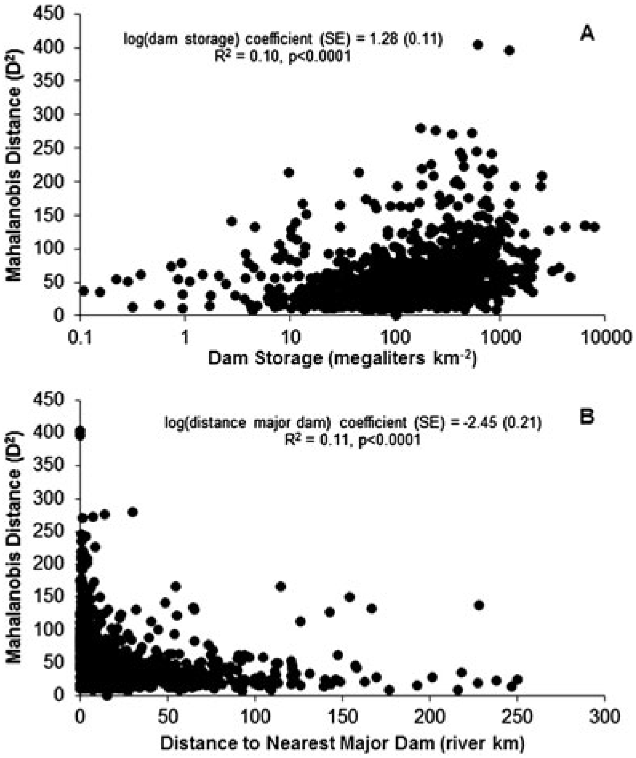

Figure 13. Squared Mahalanobis distances (D^2) plotted versus total dam storage (A) and versus downstream distance from the nearest major dam (B). Major dams are ≥15 m in height and/or ≥6167 Ml in storage capacity.

Hydrologic and landscape predictive models

Because classification systems organize information into similar groupings, classes represent the relationships within and among groups and the laws governing those relationships (Melles *et al.*, 2012). On the basis of class membership, general patterns among classes can be inferred and used to develop hypotheses and effectively communicate how systems behave (Melles *et al.*, 2012). In light of this reasoning, developing models to predict class membership can provide a framework to classify new observations and to determine the variables important in distinguishing classes.

We developed a hydrologic and landscape predictive model that can be used to assign streams to classes depending on the availability of information. Our hydrologic predictive models were highly accurate (87–90%) and exceeded the predictive capacities for hydrologic classification models reported elsewhere (e.g. 81%, Kennard *et al.*, 2010b; 85%, McManamay *et al.*, 2012c). Random forest models developed solely from 32 IHA variables performed similarly to models developed from all 110 hydrologic variables, which supports the assertion that IHA variables explain the majority of variation found in available hydrologic metrics (Olden and Poff, 2003). IHA metrics important in discriminating among classes were high-flow frequency, average monthly magnitudes, and, to a lesser extent, low and high flows of various durations. For all 110 variables, monthly flow indices (average, max, min, and cumulative annual variation in flow) dominated the most important hydrologic variables (Figure S1). Monthly flow indices were also the most important variables discriminating among classes within a south-eastern hydrologic classification (McManamay *et al.*, 2012b). The importance of monthly flows in discriminating classes may not be surprising because they comprise a considerable portion (12/32) of the IHA indices (Richter *et al.*, 1996) and a large portion (55/171) of all 171 hydrologic metrics (Olden and Poff, 2003). Monthly flows provide consolidated indices of various flow magnitudes and the seasonal timing in which they occur, and they have been widely used in developing environmental flow recommendations for river systems (Magilligan and Nislow, 2001; Pyron and Neumann, 2008; Gao *et al.*, 2009). In addition, monthly indices are interpretable and easily calculated in the absence of statistical software or input data types required for statistical software.

In the absence of sufficient natural hydrologic information, landscape variables can be used to assign streams to classes with fairly high accuracy. Within our analysis, temperature and precipitation variables had the highest explanatory power in discriminating among hydrologic classes compared with soils and soil-hydrology-related factors. Random forests produced models with accuracy rates of 74–76%, which is comparable with estimates from other studies. For example, Kennard *et al.* (2010b) reported an accuracy rate of 62% using environmental variables to discriminate among 12 Australian flow classes whereas Liermann *et al.* (2012) reported 75% accuracy for discriminating among seven flow classes in Washington (USA). In terms of predicting hydrologic class membership using landscape predictors, the highest reported accuracy we found was 87% in discriminating among four flow classes across Washington, Oregon, and Colorado (Sanborn and Bledsoe, 2006). Climate typically governs streamflow patterns at the continental scale (Carlisle *et al.*, 2010; Kennard *et al.*, 2010b; McManamay *et al.*, 2012c) whereas at basin-wide or regional scales the relative importance of localized factors, such as soils and topography, increases (Carlisle *et al.*, 2010; Knight *et al.*, 2011; McManamay *et al.*, 2012c). However, climate predictors explained the majority of variation in flow classifications developed for state or regional areas in the western USA (Sanborn and Bledsoe, 2006; Liermann *et al.*, 2012). Thus, the relative importance of landscape variables in predicting hydrology will likely depend on scale and location (Carlisle *et al.*, 2010).

The availability of hydrologic data across the landscape remains one of the largest challenges in determining relationships between flow and ecology (Knight *et al.*, 2008; Poff and Zimmerman, 2010). Thus, predicting the natural flow regime for ungauged sites has received considerable attention in recent years (Sanborn and Bledsoe, 2006; Zhu and Day, 2009; Carlisle *et al*., 2010; Knight *et al*., 2011; Murphy *et al*., 2012). The accuracy of models predicting individual hydrologic metrics can vary considerably with hydrologic metric, scale, and location (Sanborn and Bledsoe, 2006; Carlisle *et al*., 2010; Knight *et al*., 2011). However, Sanborn and Bledsoe (2006) suggested that models predicting hydrologic classes, rather than hydrologic metrics, may provide higher accuracies. Because streamflow regimes are characterized by multiple aspects of the hydrograph (magnitude, timing, frequency, duration, and rate of change) (Poff *et al*., 1997), it is reasonable to assume that multivariate groups may be more representative of stream hydrology than individual metrics.

Utility of the hydrologic classification framework

Hydrologic classes provide a contextual basis for developing generalities in hydrologic disturbances (i.e. commonalities in how streams response to disturbance). In addition, hydrologic classes provide a means to stratify hydrologic-mediated responses of ecological systems to disturbance (Arthington *et al*., 2006; Poff *et al*., 2010). Generalizing the hydrologic response of a stream to a given disturbance can provide information to guide future landscape and resource development, provide regional management strategies, and establish conservation priorities. More specifically, hydrologic classes provide a quantitative basis for establishing thresholds or sustainability boundaries for allocating water (Richter, 2010).

We provided a method for detecting outliers for streams according to their hydrologic class membership. Extrapolating hydrologic class membership to ungauged locations can be highly advantageous in determining the degree of hydrologic alteration in situations where natural flow information is missing, as in the case of regulated systems that lack pre-dam regulation data.

Determining the degree of hydrologic alteration on the basis of individual metrics has been applied in many settings and has been shown to successfully produce environmental flow recommendations (Richter *et al*., 1996; Mathews and Richter, 2007). However, in addition to univariate approaches, it may be appropriate to determine whether a regulated stream is functioning within the normal multivariate bounds of streamflow (i.e. hydrologic class membership). Our results suggest that multivariate distances from classes can be predicted; however, more in-depth model building will be required to increase predictive capacity and quantitatively generalize disturbances.

CONCLUSIONS

We developed updated hydrologic classifications at the scale of the continental USA that can be used as a framework to develop and test hypothetical relationships between flow alteration and ecology. Because classifications should be updated as new information and novel approaches become available, we envision that hydrologic classifications will continue to be reproduced at the spatial and temporal resolutions needed to suit the specific needs of managers. Our results indicate that multivariate approaches (i.e. classifications) may expand the sample size of hydrologic information potentially available to analyses evaluating patterns in natural flow. Limited availability of hydrologic information and the need for extrapolating hydrologic information to ungauged locations for ecological analyses have been on the forefront of ecohydrology (Carlisle *et al*., 2010; Eng *et al*., 2012). However, a common and necessary step in evaluating natural flow patterns is to screen gauges for reference quality and temporal resolution. The inclusion of gauges of lower reference quality increased cluster stability and only minimally increased dimensionality at the expense of increasing outliers and uncertainty. Including nonreference quality gauges may be inappropriate in univariate approaches; however, low-disturbance streams may be included in multivariate analyses assessing patterns of natural flow with fairly low uncertainty. Trade-offs among information quantity and quality in relation to hydrologic classifications should be explicitly tested under more rigorous approaches. For example, to fully test this hypothesis, datasets could be constructed that include stronger opposing endpoints than that found in our analysis (e.g. gauges representing the highest reference quality and those representing the most disturbed quality). Classification solutions that optimize the balance between dimensionality, predictive capacity, and uncertainty would be preferred.

Hydrologic classifications provide a contextual framework for organizing hydrologic information, stratifying analyses, generating broadly applicable relationships, and providing building blocks to support environmental flow standard development. Although characterizing streams by their flow patterns has been an active area of research, determining the ability of hydrologic classes to predict ecological patterns has been poorly documented (but see Monk *et al*., 2006; Chinnayakanahalli *et al*., 2011). Furthermore, there is potentially a greater need to use ecological data to refine hydrologic classifications (i.e. supervised classifications) to ensure classifications are relevant to ecology and instream flow needs (Poff *et al*., 2010). Thus, to support assessments of ecological responses of hydrology-mediated disturbances, approaches that use ecological and hydrologic information simultaneously in classifications should be an area of active research.

ACKNOWLEDGEMENTS

This research was sponsored by the US Department of Energy's (DOE) Office of Energy Efficiency and Renewable Energy, Wind and Water Power Technologies Program. This paper has been authored by employees of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy. Accordingly, the publisher, by accepting the article for publication, acknowledges that the US Government retains a nonexclusive, paid-up, irrevocable, worldwide licence to publish or reproduce the published form of this manuscript or allow others to do so, for US Government purposes. We thank Henriette Jager, Emmanuel Frimpong, Jonathan Kennen, and an anonymous reviewer for providing editorial suggestions on earlier versions of this manuscript. Special thanks to Leroy Poff for providing his hydrologic classification for public use.

REFERENCES

Acreman MC, Sinclair CD. 1986. Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. *Journal of Hydrology* **84**: 365–380.

Arthington AH, Bunn SE, Poff NL, Naiman RJ. 2006. The challenge of providing environmental flow rules to sustain river ecosystems. *Ecological Applications* **16**: 1311–1318.

Bailey RG. 1983. Delineation of ecosystem regions. *Environmental Management* **7**: 365–373.

Breiman L. 2001. Random forests. *Machine Learning* **45**: 5–32.

Brown DG, Johnson KM, Loveland TR, Theobald DM. 2005. Rural land-use trends in the conterminous United States, 1950–2000. *Ecological Applications* **15**: 1851–1863.

Burn DH, Arnell NW. 1993. Synchronicity in global flood responses. *Journal of Hydrology* **144**: 381–404.

Carlisle DM, Falcone J, Wolock DM, Meador MR, Norris RH. 2010. Predicting the natural flow regime: models for assessing hydrological alteration in streams. *River Research and Application* **26**: 118–136.

Chinnayakanahalli KJ, Hawkins CP, Tarboton DG, Hill RA. 2011. Natural flow regime, temperature and the composition and richness of invertebrate assemblages in streams of the western United States. *Freshwater Biology* **56**: 1248–1265.

Cramér H. 1999. Mathematical Methods of Statistics. Princeton University Press: Princeton, NJ.

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. 2007. Random forests for classification in ecology. *Ecology* **88**: 2783–2792.

Eng K, Carlisle DM, Wolock DM, Falcone JA. 2012. Predicting the likelihood of altered streamflows at ungauged rivers across the conterminous United States. *River Research and Applications*, DOI: 10.1002/rra.2565

Everitt BS, Landau S, Leese M. 2001. Cluster Analysis, 4th edn. Arnold: London.

Falcone JA, Carlisle DM, Weber LC. 2010a. Quantifying human disturbance in watersheds: variable selection and performance of a GIS-based disturbance index for predicting the biological condition of perennial streams. *Ecological Indicators* **10**: 264–273.

Falcone JA, Carlisle DM, Wolock DM, Meador MR. 2010b. GAGES: a stream gage database for evaluating natural and altered flow conditions in the conterminous United States. Ecology 91: 621. Data Paper in Ecological Archives E091-045-D1; available online at: http://esapubs.org/Archive/ecol/E091/045/metadata.htm.

Falcone JA. 2011. GAGES-II: geospatial attributes of gages for evaluating streamflow. Accessed 9 November 2011 at: http://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml#Identification_Information

Filzmoser P. 2005. Identification of multivariate outliers: a performance study. *Austrian Journal of Statistics* **32**: 127–138.

Filzmoser P, Gschwandtner M. 2012. Package 'mvoutlier'. Multivariate outlier detection based on robust methods. Accessed 11 September 2012 at: http://cran.r-project.org/web/packages/mvoutlier/mvoutlier.pdf.

Fraley C, Raftery AE. 2007. Bayesian regularization for normal mixture estimation and model based clustering. *Journal of Classification* **24**: 155–181.

Fraley C, Raftery AE. 2012. Package 'mclust'. Model-based clustering/normal mixture modeling. Accessed July 22, 2012 at: http://cran.rproject.org/web/packages/mclust/mclust.pdf

Fraley C, Raftery AE, Murphy TB, Scrucca L. mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation technical report no. 597, Department of Statistics, University of Washington. June 2012.

Gao YX, Vogel RM, Kroll CN, Poff NL, Olden JD. 2009. Development of representative indicators of hydrologic alteration. *Journal of Hydrology* **374**: 136–147.

Garrett RG. 1989. The chi-square plot: a tool for multivariate outlier recognition. *Journal of Geochemical Exploration* **32**: 319–341

Haines AT, Findlayson BL, McMahon TA. 1988. A global classification of river regimes. *Applied Geography* **8**: 255–272.

Hennig C. 2007. Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis* **52**: 258–271.

Hennig C. 2008. Dissolution point and isolation robustness: robustness criteria for general cluster analysis methods. *Journal of Multivariate Analysis* **99**: 1154–1176.

Hennig C. 2013. Package 'fpc'. Flexible procedures for clustering. Accessed 10 June 2013 at: http://cran.r-project.org/web/packages/fpc/index.html

Henriksen, JA, Heasley J. 2010. Hydroecological stream classification and StreamFlow software development for North Carolina. Prepared For Environmental Defense Fund North Carolina Division of Water Resources. April 2010. Accessed September 25, 2012 at: http://www.ncwater.org/Data_and_Modeling/eflows/files/Final_NC_Report.pdf

Henriksen JA, Heasley J, Kennen JG, Niewsand S. 2006. Users' manual for the hydroecological integrity assessment process software (including the New Jersey Assessment Tools): U.S. Geological Survey, Biological Resources Discipline, Open File Report 2006-1093, 71.

Horton RE. 1945. Erosional development of streams and their drainage basins – hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin* **56**: 275–370.

Jackson DA. 1993. Stopping rules in principal components-analysis – a comparison of heuristic and statistical approaches. *Ecology* **74**: 2204–2214.

Kennard MJ, Pusey BJ, Mackay SJ, Olden JD,Marsh N. 2010a. Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies. *River Research and Applications* **26**: 137–156.

Kennard MJ, Pusey BJ, Olden JD, Mackay SJ, Stein JL, Marsh N. 2010b. Classification of natural flow regimes in Australia to support environmental flow management. *Freshwater Biology* **55**: 171–193.

Kennen JG, Hendriksen JA, Heasley J, Cade BS, Terrell JW. 2009. Application of the hydroecological integrity assessment process for Missouri streams. US Geological Survey Scientific Investigations Report 2009-1138.

Kennen JG, Henriksen JA, Nieswand SP. 2007. Development of the Hydroecological Integrity Assessment Process for determining environmental flows for New Jersey streams: U.S. Geological Survey, Scientific Investigations Report 2007–5206.

Knight RR, Gain WS, Wolfe WJ. 2011. Modelling ecological flow regime: an example from the Tennessee and Cumberland River Basins: *Ecohydrology*, in press. DOI: 10.1002/eco.246

Knight RR, Gregory MB, Wales AK. 2008. Relating streamflow characteristics to specialized insectivores in the Tennessee River Valley: a regional approach. *Ecohydrology* **1**: 394–407.

Leathwick JR, Snelder T, Chadderton WL, Elith J, Julian K, Ferrier S. 2011. Use of generalised dissimilarity modelling to improve the biological discrimination of river and stream classifications. *Freshwater Biology* **56**: 21–38.

Liermann CAR, Olden JD, Beechie TJ, Kennard MJ, Skidmore PB, Konrad CP, Imaki H. 2012. Hydrogeomorphic classification of Washington state rivers to support emerging environmental flow management strategies. *River Research and Applications* **28**: 1340–1358.

Magilligan FJ, Nislow KH. 2001. Long-term changes in regional hydrologic regime following impoundment in a humid-climate watershed. *Journal of the American Water Resources Association* **37**: 1551–1569.

Mahalanobis PC. 1936. On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India* **2**: 49–55.

Mathews R, Richter BD. 2007. Application of the indicators of hydrologic alteration software in environmental flow setting. *Journal of the American Water Resources Association* **436**: 1400–1413.

McManamay RA, Orth DJ, Dolloff CA. 2012a. Revisiting the homogenization of dammed rivers in the southeastern US. *Journal of Hydrology* **424–425**: 217–237.

McManamay RA, Orth DJ, Dolloff CA, Frimpong EA. 2012b. A regional classification of unregulated streamflows: spatial resolution and hierarchical frameworks. *River Research and Applications* **28**: 1019–1033.

McManamay RA, Orth DJ, Dolloff CA, Frimpong EA. 2012c. Regional frameworks applied to hydrology: can landscape-based frameworks capture the hydrologic variability?. *River Research and Applications* **28**: 1325–1339.

Melles SJ, Jones NE, Schmidt B. 2012. Review of theoretical developments in stream ecology and their influence on stream classification and conservation planning. *Freshwater Biology* **57**: 415–434.

Monk WA, Wood PJ, Hannah DM, Wilson DA, Extence CA, Chadd RP. 2006. Flow variability and macroinvertebrate community response within riverine systems. *River Research and Applications* **22**: 595–615.

Murphy JC, Knight RR, Wolfe WJ, Gain WS. 2012. Predicting ecological flow regime at ungaged sites. *River Research and Applications*, in press. DOI: 10.1002/rra.2570.

Nathan RJ, McMahon TA. 1990. Identification of homogeneous regions for the purposes of regionalization. *Journal of Hydrology* **121**: 217–238.

Olden JD, Poff NL. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications* **19**: 101–121.

Olden JD, Kennard MJ, Pusey BJ. 2012. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. *Ecohydrology* **5**: 503–518.

Omernik JM. 1987. Ecoregions of the conterminous United-States. *Annals of the Association of American Geographers* **77**: 118–125.

Omernik JM, Bailey RG. 1997. Distinguishing between watersheds and ecoregions. *Journal of the American Water Resources Association* **33**: 935–949.

Pennak RW. 1971. Toward a classification of lotic habitats. *Hydrobiologia* **38**: 321–334.

Poff NL. 1996. A hydrogeography of unregulated streams in the United States and an examination of scale-dependence in some hydrological descriptors. *Freshwater Biology* **36**: 71–91.

Poff NL, Allan JD. 1993. Streamflow variability, fish community structure, and implications for climatic change. U.S. Environmental Protection Agency Report ERL-D #2765. Duluth, MN.

Poff NL, Allan JD, Bain MB, Karr JR, Prestegaard KL, Richter BD, Sparks RE, Stromberg JC. 1997. The natural flow regime. *BioScience* **47**: 769–784.

Poff NL, Richter BD, Arthington AH, Bunn SE, Naiman RJ, Kendy E, Acreman M, Apse C, Bledsoe BP, Freeman MC, Henriksen J, Jacobson RB, Kennen JG, Merritt DM, O'Keeffe JH, Olden JD, Rogers K, Tharme RE, Warner A. 2010. The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. *Freshwater Biology* **55**: 147–170.

Poff NL, Zimmerman JZH. 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. *Freshwater Biology* **55**: 194–205.

Pyron M, Neumann K. 2008. Hydrologic alterations in the Wabash River watershed, USA. *River Research and Applications* **24**: 1175–1184.

Richter BD. 2010. Re-thinking environmental flows: from allocations and reserves to sustainability boundaries. *River Research and Applications* **26**: 1052–1063.

Richter BD, Baumgartner JV, Powell J, Braun DP. 1996. A method for assessing hydrologic alteration within ecosystems. *Conservation Biology* **10**: 1163–1174.

Richter BD, Baumgartner JV, Wigington R, Braun DP. 1997. How much water does a river need?. *Freshwater Biology* **37**: 231–249.

Rosgen DL. 1994. A classification of natural rivers. *Catena* **22**: 169–199.

Sanborn SC, Bledsoe, BP. 2006. Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon. *Journal of Hydrology* **325**: 241–261.

Schafer JL. 1997. Imputation of missing covariates under a general linear mixed model. Technical report, Dept. of Statistics, Penn State University. Accessed 25 September 2012 at: http://cran.r-project.org/web/packages/pan/vignettes/pan-tr.pdf

Schafer JL. 2012. Package 'mix'. Estimation/multiple imputation for mixed categorical and continuous data. Accessed 14 February 2012 at: http://cran.r-project.org/web/packages/mix/mix.pdf

Snelder TH, Biggs BJF. 2002. Multi-scale river environment classification for water resources management. *Journal of the American Water Resources Association* **38**: 1225–1240.

Snelder TH, Dey KL, Leathwick JR. 2007. A procedure for making optimal selection of input variables for multivariate environmental classifications. *Conservation Biology* **21**: 365–375.

Snelder TH, Lamouroux N, Leathwick JR, Pella H, Sauquet E, Shankar U. 2009. Predictive mapping of the natural flow regimes of France. *Journal of Hydrology* **373**: 56–67.

Sokal RR. 1974. Classification – purposes, principles, progress, prospects. *Science* **185**: 1115–1123.

Strahler AN. 1957. Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysical Union* **38**: 913–920.

Stoddard JL, Larsen DP, Hawkins CP, Johnson RK, Norris RH. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications* **16**: 1267–1276.

Turton D, Fisher B, Seilheimer TS, Esralew R. 2008. An assessment of environmental flows for Oklahoma. US Geological Survey Report 2008OK107B.

USACE. 2012. National Inventory of Dams. US Army Corps of Engineers. Accessed online June 24, 2011: https://nid.usace.army.mil.

Wolock DM, Winter TC, McMahon G. 2004. Delineation and evaluation of hydrologic landscape regions in the United States using geographic information system tools and multivariate statistical analyses. *Environmental Management* **34**: S71–S88.

Vogl AL, Lopes VL. 2009. Impacts of water resources development on flow regimes in the Brazos River. *Environmental Monitoring and Assessment* **157**: 331–345.

Zhu Y, Day RL. 2009. Regression modeling of streamflow, baseflow, and runoff using geographic information systems. *Journal of Environmental Management* **90**: 946–953.