RIVER RESEARCH AND APPLICATIONS
River. Res. Applic. **26**: 118–136 (2010)
Published online 3 March 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/rra.1247

PREDICTING THE NATURAL FLOW REGIME: MODELS FOR ASSESSING HYDROLOGICAL ALTERATION IN STREAMS†

DAREN M. CARLISLE,[a]* JAMES FALCONE,[a] DAVID M. WOLOCK,[b]
MICHAEL R. MEADOR[a] and RICHARD H. NORRIS[c]

[a] *National Water-Quality Assessment Program, US Geological Survey, 413 National Center, Reston, VA 20192, USA*
[b] *US Geological Survey, Kansas Science Center, 4821 Quail Crest Place, Lawrence, KS 66049, USA*
[c] *Water Cooperative Research Centre and Institute for Applied Ecology, University of Canberra, Canberra, Australian Capital Territory 2601, Australia*

ABSTRACT

Understanding the extent to which natural streamflow characteristics have been altered is an important consideration for ecological assessments of streams. Assessing hydrologic condition requires that we quantify the attributes of the flow regime that would be expected in the absence of anthropogenic modifications. The objective of this study was to evaluate whether selected streamflow characteristics could be predicted at regional and national scales using geospatial data. Long-term, gaged river basins distributed throughout the contiguous US that had streamflow characteristics representing least disturbed or near pristine conditions were identified. Thirteen metrics of the magnitude, frequency, duration, timing and rate of change of streamflow were calculated using a 20–50 year period of record for each site. We used random forests (RF), a robust statistical modelling approach, to develop models that predicted the value for each streamflow metric using natural watershed characteristics. We compared the performance (i.e. bias and precision) of national- and regional-scale predictive models to that of models based on landscape classifications, including major river basins, ecoregions and hydrologic landscape regions (HLR). For all hydrologic metrics, landscape stratification models produced estimates that were less biased and more precise than a null model that accounted for no natural variability. Predictive models at the national and regional scale performed equally well, and substantially improved predictions of all hydrologic metrics relative to landscape stratification models. Prediction error rates ranged from 15 to 40%, but were $\leq$25% for most metrics. We selected three gaged, non-reference sites to illustrate how predictive models could be used to assess hydrologic condition. These examples show how the models accurately estimate pre-disturbance conditions and are sensitive to changes in streamflow variability associated with long-term land-use change. We also demonstrate how the models can be applied to predict expected natural flow characteristics at ungaged sites. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS: natural flow regime; predictive models; random forests; hydrologic modification

Received 25 August 2008; Revised 24 November 2008; Accepted 15 January 2009

INTRODUCTION

The importance of natural hydrologic processes to the ecological integrity of rivers is well known (Poff *et al*., 1997; Richter *et al*., 2003; Lytle and Poff, 2004), as is the notion that humans have dramatically altered river hydrology worldwide (Pringle *et al*., 2000; Jackson *et al*., 2001; Nilsson *et al*., 2005; Poff *et al*., 2006). Early recognition that certain minimum flows were required to maintain river biota (Bunn and Arthington, 2002; review in Tharme, 2003) have evolved to a more comprehensive view of the dynamic nature of ecological flow requirements. The paradigm of the natural flow regime (Poff *et al*., 1997) identified five major ecologically relevant components of streamflow: magnitude, frequency, duration, timing and rate of change. In a variety of ways, these hydrologic components maintain the chemical, physical and biological conditions that sustain biological diversity and ecosystem services of rivers (Poff *et al*., 1997; Bunn and Arthington, 2002; Lytle and Poff, 2004). Evidence of the importance of these hydrologic characteristics for river ecosystems is plentiful (Dynesius and Nilsson, 1994; Baxter, 1977; Jansson

*Correspondence to: Daren M. Carlisle, National Water-Quality Assessment Program, US Geological Survey, 413 National Center, Reston, VA 20192, USA. E-mail: dcarlisle@usgs.gov
†The contributions of Daren N. Carlisle, James Falcone, David M. Wolock and Michael R. Meador were prepared as part of their official duties as US Government employees.

et al., 2000; Bunn and Arthington, 2002). There is also mounting evidence that alteration of flow regimes is widespread (Nilsson *et al*., 2005; Poff *et al*., 2006, 2007), and potentially a major cause of ecological impairment (Wilcove *et al*., 1998; Pringle *et al*., 2000).

The natural flow regime must be quantified to assess hydrological alteration, but doing so has been a difficult challenge (Arthington *et al*., 2006). Attributes of the natural flow regime have been quantified for specific river systems using a variety of methods (Richter *et al*., 1997b; Acreman and Dunbar, 2004; Arthington *et al*., 2006), which provided benchmarks for the restoration of rivers known to be affected by hydrologic modification. Such predictions are often based on historic (e.g. pre-disturbance) flow records (Richter *et al*., 1997a; Henriksen *et al*., 2006) or modelled using data from nearby streamgages. Such pre-disturbance periods of record are undefined, however, for most gaged streams and non-existent for ungaged streams. Approaches to extrapolate estimates of hydrologic indices beyond known undisturbed basins have been suggested (Richter *et al*., 1997b), but we still have a limited ability to quantify expected flow regimes for most types of streams, especially at regional or national scales (Arthington *et al*., 2006).

The conceptual framework successfully employed for many years in biological assessments may also be applicable to hydrological assessments. Biological assessments require estimates of attributes (e.g. taxa richness and composition) expected in the absence of anthropogenic disturbances (Wright *et al*., 2000). The reference condition approach (Bailey *et al*., 2004) is widely used in bioassessments of freshwater ecosystems and is framed around two general concepts. First, the expected condition (hereafter '*E*') is site-specific, and derived from a collection of reference (i.e. under the least amount of human influence for a given region) sites that are environmentally similar to the site being assessed. This approach acknowledges the natural spatial and temporal variability associated with reference ecosystems. The second concept is that natural environmental characteristics can be used to explain variation in, and therefore make predictions for, ecological characteristics of reference sites. These models are used to make predictions of *E* at sites being assessed (Bailey *et al*., 2004). Because natural streamflow regimes are influenced by climate, soils, geology and other environmental characteristics (Puckridge *et al*., 1998; Poff *et al*., 2006), the rationale for empirically based estimates of *E* is compelling. The assessment is accomplished by quantifying the degree to which observed (*O*) conditions deviate from expected natural (*E*) conditions (Wright *et al*., 2000), which in concept as also been proposed and applied in hydrological assessments (Richter *et al*., 1997b; Arthington *et al*., 2006).

There are many published empirical models for predicting various attributes of the hydrologic regime (e.g. Sanborn and Bledsoe, 2006), but most have been applied to a limited number of hydrologic attributes and geographic regions. Further, the criteria for selecting reference sites, when adequately described, have been applied inconsistently among studies, thus limiting the ability to generalize. To our knowledge, most published empirical models were intended to predict observed (*O*) hydrologic attributes of ungaged basins, but none have applied this concept to establish an expectation (*E*) that is subsequently compared with *O* to assess alteration. Finally, no models have been applicable at scales sufficient for regionally or nationally consistent assessments of hydrologic alteration.

This paper describes the application and evaluation of the reference condition approach (sensu Bailey *et al*., 2004) to the assessment of hydrologic alteration. We first describe an exhaustive screening and reference-site selection process for streamgages across the conterminous US. We then evaluate the performance (i.e. precision and bias) of alternative modelling approaches for predicting metrics of the natural flow regime for these relatively undisturbed basins. We conclude with three case studies to illustrate how modelled estimates of *E* can be compared with observed (*O*) values at gaged sites to obtain a quantitative measure of hydrologic alteration.

METHODS

Selecting reference sites

Fundamental to the reference condition approach is identifying a collection of sites that are minimally influenced by human activity and that encompass the full range of environmental conditions (e.g. stream size) of sites for which assessments are desired (Bailey *et al*., 2004). One challenge is that the absolute quality of reference sites varies among different regions. For example, many basins in the western US remain in relatively pristine condition and can be considered representative of ecosystems largely undisturbed by human activity. In regions (e.g. U.S.

DOI: 10.1002/rra

Midwest) with long histories of intense landscape alteration, however, reference sites are ‘least disturbed’ or in ‘best available’ condition given current socioeconomic realities. We followed the recommendations of others (Bailey *et al*., 2004; Stoddard *et al*., 2006; Whittier *et al*., 2007) and applied numerous quantitative and qualitative criteria to identify gaged streams throughout the conterminous US that were, with respect to hydrology, minimally (i.e. pristine) or least disturbed (i.e. best available).

About 22 000 streamgages have been operated by the US Geological Survey (USGS) during the period 1950–2005. This study was limited to data from sites on perennial, small to medium-sized rivers ($<50\,000\,\mathrm{km}^2$) for which environmentally similar reference sites would likely be available and for which at least 20 years of flow records exist. Richter *et al.* (1997b) and Henriksen *et al.* (2006) found that 20 years was the minimum period of record for obtaining reliable estimates of streamflow characteristics. About 50 sites on highly intermittent streams (i.e. exhibited no flow for >6 mos year^{-1}) were excluded because they are poorly represented in the streamgage network and likely require a unique approach for understanding the natural streamflow characteristics (e.g. Fritz and Dodds, 2005). This preliminary screening process resulted in 5271 gages retained for further consideration (Figure 1).

We established several quantitative and qualitative measures to select reference sites from the candidate set of 5271 gages. Measured streamflow data were generally not used as screening criteria (one exception below) to minimize circularity in the assessment process (Bailey *et al*., 2004; Stoddard *et al*., 2006). We calculated several quantitative indicators of potential hydrologic modification using widely available geospatial databases (Table I), including the presence of water conveyance systems (e.g. canals and pipelines, Horizon Systems Corporation 2006), the locations and characteristics (e.g. storage capacity) of dams (USACE, 2006), locations of major pollutant dischargers (e.g. wastewater facilities, USEPA, 2006a), the extent of impervious land cover in the basin (Vogelman *et al*., 2001) and water withdrawals (USGS, 2007a). For each geospatial indicator, each site was scored (0–8) based on its value relative to various percentiles of the distribution of the metric among all sites nationally. Scores were then summed across all indicators producing an index of relative hydrologic impact for each site. Sites with index scores in the lowest quartile for each region (see below) were given highest consideration as possible reference, but additional criteria were also applied. Qualitative site-specific characterizations of hydrologic modification, including site descriptions published in USGS annual data reports (USGS, 2007b) and our own examinations of 7.5-min topographic maps and imagery were also used because of the inherent uncertainty in the databases used to

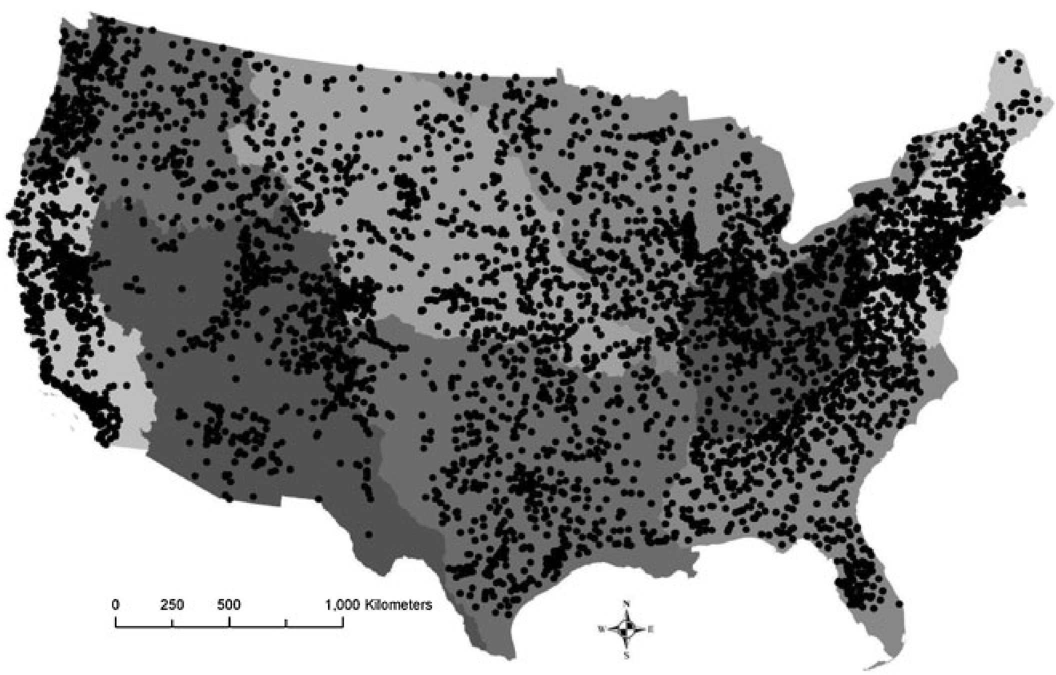

Figure 1. Streamgages in the conterminous US for which ≥20 years of continuous flow records exist since 1950, and that were screened for designation as reference quality for hydrologic analysis. The major hydrologic regions (HUC2) used in this study are also shown

DOI: 10.1002/rra

Table I. Geospatial indicators of potential hydrological modification used to screen reference-quality gaged river basins in the conterminous USA

General description	Indicators	Data source
Land cover	Urban land cover in basin	Vogelman *et al*., 2001
	Urban land cover in mainstem 600 m buffer	
	Agricultural land cover in basin	
	Agricultural land cover in mainstem 600 m buffer	
	Mining and transitional land cover in mainstem 600 m buffer	
Dams	Density of major dams in basin	USACE, 2006
	Linear distance of gage to nearest dam	
	Increase in reservoir storage 1950–2005	
	Dams/mainstem km	
Road density	Road density in basin	GeoLytics, 2001
Water conveyances	Per cent of stream length coded as canal/ditch/pipeline	Horizon Systems, 2006
Major dischargers	Linear distance of gage to nearest major discharger	USEPA, 2006a
Water withdrawal	Freshwater withdrawal	USGS, 2007a

generate geospatial variables. These qualitative data sources often revealed small diversions or other flow-regulating structures in the contributing basin or near streamgages. For sites in areas of irrigated agriculture, monthly discharge values were estimated using a water balance model (Wolock and McCabe, 1999) and compared to the observed monthly discharge to determine if irrigation withdrawal was significantly influencing streamflow during the growing season. After considering all of these criteria for each of the 5271 gaged sites, we classified each site as reference or non-reference.

Hydrologic data and metric selection

We selected 13 hydrologic metrics with the goal of evaluating how well each could be predicted using various modelling approaches. It was beyond the scope of this study to evaluate all metrics currently in use (see Olden and Poff, 2003). Rather, we selected metrics that represent ecologically relevant components of the flow regime (Poff *et al*., 1997) (Table II) and have been used by others. Twelve metrics were calculated using the algorithms of Henriksen *et al*. (2006), and one (skewness) was calculated using the unbiased method of Yevjevich (1972). All calculations were based on the entire period of record for each reference site (20–50 years). Many hydrologic metrics are calculated annually, and can therefore be represented as a distribution of yearly values for a given site. For convenience we built models for the median value of such metrics. Our final dataset therefore contained a single value of each of 13 hydrologic metrics for each reference site.

Evaluation of model performance

We used average (mean) and variability (SD) of the distribution of O/E values among reference sites as a measure of model performance. O for each site was the calculated value of each hydrologic metric, but E was derived in several ways (see below). Although in theory O/E should be unity at reference sites, measurement error (of hydrologic metrics and geospatial predictors) and variation left unexplained by the model produce a distribution of O/E values among reference sites (Bailey *et al*., 2004). This variation among reference site O/E values represents, in theory, the uncertainty with which the model would be applied to predict E at test sites. Minimizing this uncertainty would therefore improve our ability to statistically detect more subtle deviations from E at test sites. We used the root mean square error (RMSE) as a measure of the relative performance of each model (Van Sickle *et al*., 2006) because it accounts for both bias and precision of model predictions. Differences in the RMSE of reference site O/E values can be attributed to the success with which each model accounted for natural variation in E. Hence, models producing the lowest RMSE were deemed superior (see Hawkins *et al*., 2000b; Ostermiller and Hawkins, 2004; Linke *et al*., 2005; Van Sickle *et al*., 2005).

DOI: 10.1002/rra

Table II. Hydrologic metrics examined in this study and their general class of hydrological attribute

Description	Attribute	Units	Source
Daily variability	Average	Per cent of mean	Henriksen *et al*., 2006
Skewness*	Average	Dimensionless	Yevjevich, 1972
Annual runoff*	Average	$m^3\,sec^{-1}\,km^{-2}$	Henriksen *et al*., 2006
Base flow index	Low flow	Dimensionless	Henriksen *et al*., 2006
Annual maximum flow*	High flow	$m^3\,sec^{-1}\,km^{-2}$	Henriksen *et al*., 2006
Low flow pulses*	Low flow	events $year^{-1}$	Henriksen *et al*., 2006
High flow pulses*	High flow	events $year^{-1}$	Henriksen *et al*., 2006
Low flow duration*	Low flow	days $year^{-1}$	Henriksen *et al*., 2006
High flow duration*	High flow	days $year^{-1}$	Henriksen *et al*., 2006
Flood interval*	High flow	days $year^{-1}$	Henriksen *et al*., 2006
Flood-free days*	High flow	days $year^{-1}$	Henriksen *et al*., 2006
Predictability	Average	Dimensionless	Henriksen *et al*., 2006
Reversals*	Average	days $year^{-1}$	Henriksen *et al*., 2006

Descriptions coded with '*' are metrics for which the median value across years in the period of record was modelled.

Alternative models for predicting E

Our primary objective was to evaluate how well various modelling approaches predict *E* for each hydrologic metric. Classification of sites into discrete regions is a common approach in ecological assessments (Hawkins and Norris, 2000), and is based on the assumption that homogeneity of environmental conditions within regions will produce more accurate and precise estimates of expected ecological attributes (i.e. *E*) than estimates that do not 'control' for natural variation. We first developed a null model (hereafter null) that did not account for natural variability in *E*. For each metric, we calculated the median observed value across all reference sites nationwide and applied this value as *E* for all sites. We then calculated *O*/*E* for each reference site using this constant value of *E*. Although it is clearly unrealistic to assume that *E* would be constant across the nation for any hydrologic metric, the performance of this model is a useful baseline from which to compare how well the other approaches account for natural variation (Van Sickle *et al*., 2005). We then considered three landscape stratification/classification 'models' to derive estimates of *E*.

Our first landscape stratification model was based on 2-digit Hydrologic Unit Codes (hereafter HUC2s), which are aggregations of large river basins (Figure 1). We identified the HUC2 for each site, then aggregated some adjacent HUC2s to obtain at least 100 reference sites within each unit. We calculated *O*/*E* for all reference sites using a constant value of *E* for each HUC2, which was the median observed value of each metric across all reference sites within that unit.

The second stratification model was an increasingly more homogenous geographic region—level 3 ecoregions (USEPA, 2007). Previous studies (Richter *et al*., 1997b; Sanborn and Bledsoe, 2006) have suggested that ecoregions are a viable framework for predicting natural flow regimes because they are relatively homogenous in the environmental factors known to affect flow regimes (e.g. climate, geology, soils). We assigned each site to the ecoregion in which the majority of the basin fell, then aggregated enough adjacent ecoregions to obtain >10 reference sites within each region. We then calculated *O*/*E* for all reference sites using the median of observed values within each level 3 ecoregion for *E*.

We used hydrologic landscape regions (HLR) (Wolock *et al*., 2004) as our third stratification approach. HLR are aggregate river basins that share similar environmental factors (e.g. climate, soils, topography) known to influence streamflow. Unlike ecoregions or HUC2s, however, HLRs are not constrained to be contiguous, and therefore classify basins based on similarities believed to be relevant to hydrology irrespective of geographic location. We assigned each site to the HLR identified by Wolock *et al.* (2004) in which the majority of the basin fell. We then calculated *O*/*E* for all reference sites using the median of observed values within each HLR for *E*. Our estimates of precision for all stratification models were not calculated from independent (e.g. cross-validation) samples. We felt that sample sizes were too small in some regions to omit an adequate number of reference sites

DOI: 10.1002/rra

(e.g. 30–50) for an independent evaluation of bias and precision. As a consequence, performance for the stratification methods was probably overestimated.

In addition to the three landscape-stratification models, we developed models that made site-specific predictions of E using empirical relationships between variation in hydrologic metrics and environmental attributes among reference sites (sensu Bailey *et al*., 2004). We developed a single national model and HUC2-specific models for each hydrologic metric. We used a modelling method, random forests (RF) (Breiman, 2001), that is relatively new to ecology, but has proved to be more robust and accurate than traditional linear (e.g. multiple linear regression) or more complex (e.g. neural nets) methods (Lawler *et al*., 2006; Prasad *et al*., 2006; Cutler *et al*., 2007). RFs were developed as a method of improving the predictions of classification and regression trees (Liaw and Wiener, 2002; De'ath, 2007). A detailed description of the method is beyond the scope of this paper, but we provide a brief overview relevant to our results (also see Breiman, 2001). Classification and regression trees have been widely used for prediction and description in ecological studies because they have several desirable properties (De'ath and Fabricus, 2000), including the ability to model nonlinear relationships and interactions among predictors. RF were developed to alleviate one major concern of regression trees—that of overfitting (Hastie *et al*., 2001). In essence, RF produce thousands of regression trees, each with a bootstrapped sample of 70% of the observations where a randomly selected subset of the predictor variables is considered at each branch. The remaining 30% of observations are passed through each tree to evaluate predictive performance. On average, each observation is omitted from 1/3 of the trees, and the final prediction for each observation is obtained by averaging the predictions across all the trees where it was excluded. This cross-validation produces estimates of error that are close to what would be expected with independent data. The relative importance of predictors is also evaluated using the left-out observations. The predictive ability (RMSE in regression) of omitted observations is first calculated. Then, each predictor is in turn randomly permuted and the change in RMSE noted. The predictors are ranked by the per cent increase in RMSE obtained when each is permuted. The most important predictors will cause the largest decrease in RMSE when permuted. In sum, RF are a form of model averaging that has proved useful at 'learning' (sensu Hastie *et al*., 2001) from large datasets. We are not necessarily advocating the use of RF over more traditional methods, and a comparison of predictive ability among various approaches was beyond the scope of this paper. Rather, we selected a robust modelling approach that we felt would maximize our ability to make predictions of E. For all predictive models we used 80 predictor variables (Table III) representing climate, geology, soils, topography and geography. The RF models were performed with an implementation written for the R statistical system (R Development Core Team, 2006) by Liaw and Wiener (2002). Each RF model produced 2000 individual regression trees, and E for each site was obtained by averaging its predicted value across trees ($n \approx 500$) where it was withheld from model development. We calculated O/E for each metric using a predicted value of E from the single national RF model and a predicted value of E from each HUC2-specific RF model.

We hypothesized that, relative to the null model, increasingly refined landscape classifications (i.e. HUC2s, ecoregions, HLR) would provide progressively greater performance in estimates of E. We also expected that HUC2-predictive models would be more precise and less biased than a national model, but that both modelling approaches would produce better estimates than any of the stratification models. Because estimates of E from RF models are directly calculated from observations excluded from model development, we are confident that our estimates of model performance are similar to that encountered if applied to non-reference sites.

Example hydrologic assessments

In two ways, we illustrate the application of predictive modelling described in this paper. First, in the course of screening reference sites we excluded many that were influenced by proximal reservoirs constructed during the 1970s and 1980s and where a relatively unregulated gaged record exists from before dam construction. We reasoned that if the predictive models were capable of simulating the natural flow regime, predictions of E should agree with O from the pre-dam period. We randomly selected two such sites (USGS gages 05592000 and 01151500) that had recorded at least 20 years of flow before and after dam construction. To the best of our knowledge, there were no known hydrologic alterations prior to dam construction. We illustrate the assessment with a single hydrologic metric that we suspected would be sensitive to dams and other human activities in the watershed—daily variability. For each site, we generated a distribution of predictions by recording the individual

DOI: 10.1002/rra

Table III. Variables used in models to predict hydrological metrics among reference streams in the conterminous USA

Basin topography and location (USGS, 2006)
Latitude
Longitude
Drainage area
Standard deviation of elevation
Relief ratio (Pike and Wilson 1971)
Site (gage) elevation
Mean basin elevation
Climate (Daymet, 2006)
Mean annual precipitation
Mean monthly maximum precipitation
Mean monthly minimum precipitation
Mean annual number of days with precipitation
Mean monthly maximum number of days with precipitation
Mean monthly minimum number of days with precipitation
Mean monthly precipitation. (January—December)
Mean relative humidity
Potential evapo-transpiration
Mean annual temperature
Mean monthly maximum temperature
Mean monthly minimum temperature
Standard deviation of maximum monthly temperature
Standard deviation of minimum monthly temperature
Mean Julian date of first frost
Mean Julian date of last frost
Mean monthly temperature (January—December)
Soil properties (USDA, 2006)
Per cent basin soils in each of seven hydrologic groups
Mean permeability
Mean water capacity
Mean bulk density
Mean organic matter
Mean depth of water table
Mean soil thickness
Mean per cent clay, silt, sand
Mean per cent fine and coarse soils
Mean soil erodibility factor (from Universal Soil Loss Equation)
Mean runoff factor (from Universal Soil Loss Equation)
Geology (Reed and Bush 2005)
Per cent of basin each of nine geological classes
Dominant geologic class in basin

estimates of E from each tree ($n = 2000$) in the RF model. We then visually compared the observed (O) metric value to this distribution of modelled estimates. We also selected a gaged basin near Washington, DC (USGS gage 01654000) that has undergone substantial land-use change. We selected this site to demonstrate the modelling approach for a basin without a discrete hydrological alteration and, therefore, a pre-disturbance record. The streamgage has operated continuously since 1949, during which time the basin has been transformed from a mixed forest and pastoral landscape to an urban setting (Jennings and Jarnagin, 2002). We partitioned the flow record into 20-year blocks (1950–1970, 1960–1980, etc.) through 2000 and calculated daily variability of flow (O) for each time period. We then used the RF national model to generate a distribution of predictions of E for this metric (which was the same for each time period) and calculated O/E for each time period.

For these example assessments, we graphically compared the O/E value of the metric (i.e. for each time period) to the distribution of O/E values for all reference sites across the nation. A frequent aim of environmental

DOI: 10.1002/rra

assessment is to make an inference about whether observed conditions at a test site are different (with some known level of certainty) from expectations (i.e. reference), given the error associated with natural variability and the measurement and modelling processes. The distribution of *O*/*E* values among reference sites provides the basis for making decisions about the likelihood that a test site is significantly different (i.e. altered) from reference condition (Bailey *et al*., 2004; Stoddard *et al*., 2006). Thresholds defined by standard deviations or percentiles of the reference site distribution are frequently used to establish classes of severity of alteration (Wright *et al*., 2000; Bailey *et al*., 2004). Ultimately, assessment thresholds based on statistical significance, ecological relevance and societal values should be determined, but this was beyond the scope of our study.

A second application of the models we developed is the prediction of expected natural flows for ungaged segments of a river network. This application may be of interest to managers seeking the establishment of baseline conditions or setting goals for the protection of natural flows. We selected the Provo River basin in Utah to illustrate this application. The Provo River headwaters are located within a roadless, high elevation (up to 3000 m above sea level) basin. In its middle reaches, the river receives intrabasin water transfers from two separate drainages and is impounded in two large reservoirs. Abstractions for irrigation and municipal water supplies are made near the river terminus into the Great Salt Lake system. Five USGS streamgages with flow data from recent (1980–2000) years are located in the basin, none of which was used in model development. We calculated the GIS-based predictor variables for the upstream basin of every 30-m pixel located on the stream network, then used the RF model to predict *E* (days year^{-1}) for low flow duration for the entire network. We also calculated *O*/*E* for low flow duration (mean value from 1980 to 2000) for each of the five streamgages.

RESULTS

Reference sites

Our selection criteria yielded 1272 reference sites, which averaged 803 km^2 in basin size (range 2–2579) and were fairly uniformly distributed geographically across the conterminous US (Figure 2). Basins of reference sites were generally smaller and of slightly steeper topography than non-reference sites (Table IV), but were otherwise similar in natural characteristics.

Model performance

Regional stratification appeared to successfully account for some natural variation in *E*. Relative to null, estimates of *E* derived from any form of regional stratification markedly improved model performance (i.e. RMSE) for all metrics (Table V). With one exception (base flow index), ecoregion stratification produced better model performance than regionalization by HUC2 or HLR. Estimates of *E* derived from ecoregions improved performance by 12–60% relative to the null model.

RF predictive models accounted for substantially more natural variation in *E* than the best stratification (level 3 ecoregion) method (Table V). For all metrics, performance of the national-scale model was equal to that obtained by HUC2-specific models. Models using estimates of *E* derived from the national RF performed 9–61% better than models using *E* derived from ecoregion stratification.

Performance of RF predictive models varied substantially among hydrologic metrics (Table V). For most metrics RMSE ranged from 15 to 30%, but for models of low duration (RMSE = 0.34) and maximum flow (RMSE = 0.40) were somewhat higher. Models of metrics of average (Table II) flow conditions tended to perform (RMSE = 0.15–0.27) slightly better than models for high flow (RMSE = 0.15–0.40) and low flow (RMSE = 0.23–0.34).

The importance of predictor variables in RF models differed among metrics and spatial scales (Figure 3). Many of the 80 predictor variables were highly correlated (results not shown), and although this does not adversely affect RF predictions, measures of variable importance tend to be averaged among sets of highly correlated variables (Cutler *et al*., 2007). We therefore limit our interpretation of variable importance to broad groups of correlated variables. Variables describing patterns in precipitation were the most important predictors for most metrics at the national scale. The 10 most important predictors for each hydrologic metric at the national scale (Appendix) included combinations of climate, soils, topography and geographic location. The most important predictors varied

DOI: 10.1002/rra

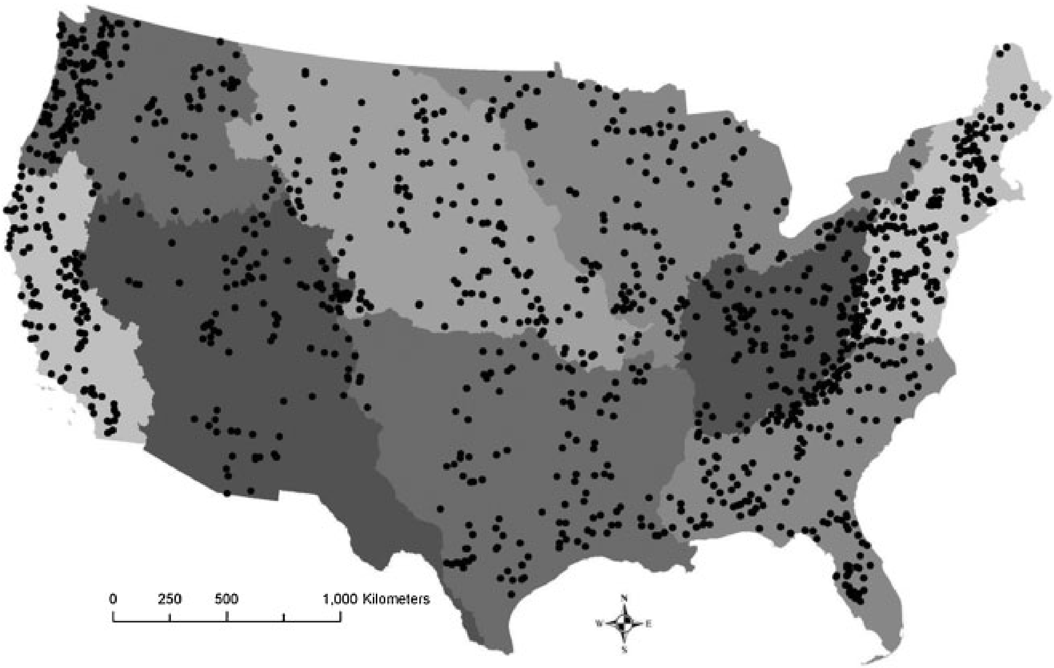

Figure 2. Reference-quality streamgages in the conterminous US relative to the major hydrologic regions (HUC2) used in this analysis

among metrics in the HUC2-specific models. Soil properties, precipitation and topography were the most important predictors of average flow metrics. Precipitation, soil properties and air temperature were the most important predictors of high flow metrics. Soil properties and air temperature were the most important predictors of low flow metrics.

Example assessments

Modelled estimates of *E* agreed well with observed data at sites with pre-dam flow records. Estimates of *E* derived from the national model were centred on observed values of daily variability from pre-dam flow records at both sites (Figure 4, Figure 5). Similarly, *O*/*E* values for daily variability from pre-dam periods were close to unity and well within the variability observed at reference sites across the nation. In contrast, observed values from post-dam flow records were well outside and much lower than model expectations for site 05592000, and the *O*/*E* value was well outside the reference-site distribution, indicating substantial hydrological alteration. For site

Table IV. Select environmental attributes of gaged river basins identified as reference-quality ($n = 1272$) and non-reference ($n = 3945$) across the conterminous US

Variable	Reference		Non-reference	
	10th, 90th	Median	10th, 90th	Median
Basin size (km^2)	31, 1820	274	73, 10 569	923
Mean basin slope (%)	1, 33	10	1, 26	6
Mean basin elevation (m)	97, 2321	497	119, 2350	424
Urban land cover (%)	0, 3	0	0, 16	1
Storage change ($10^6\,\mathrm{L\,km^{-2}}$)	0, 8	0	0, 273	8
Canal/ditch/pipeline (%)	0, 0	0	0, 10	0
Water withdrawal ($10^6\,\mathrm{L\,year^{-1}\,km^{-2}}$)	1, 31	4	1, 45	8

DOI: 10.1002/rra

Table V. Mean ($\overline{X}$), standard deviation (SD) and root mean square error (RMSE) of *O*/*E* for 13 hydrologic metrics across 1272 reference gages based on alternative approaches for estimating *E*

Metric	Null			HUC 2			Ecoregion 3			Hydrologic region			National model			HUC 2 model		
	$\overline{X}$	SD	RMSE	$\overline{X}$	SD	RMSE	$\overline{X}$	SD	RMSE	$\overline{X}$	SD	RMSE	$\overline{X}$	SD	RMSE	$\overline{X}$	SD	RMSE
Daily variability	1.06	0.42	0.42	1.01	0.34	0.34	1.01	0.31	0.31	1.01	0.34	0.34	0.99	0.24	0.24	0.99	0.24	0.24
Skewness	1.09	0.50	0.51	1.06	0.51	0.51	1.03	0.34	0.34	1.10	0.51	0.52	0.99	0.19	0.19	1.00	0.20	0.20
Annual runoff	1.31	1.37	1.40	1.28	1.31	1.34	1.09	0.68	0.69	1.22	1.25	1.27	0.98	0.27	0.27	0.97	0.27	0.27
Base flow index	1.01	0.43	0.43	1.03	0.42	0.42	1.04	0.40	0.40	1.00	0.39	0.39	0.99	0.25	0.25	0.99	0.25	0.25
Maximum flow	1.17	1.02	1.04	1.35	1.52	1.56	1.18	0.90	0.92	1.29	1.27	1.29	0.97	0.40	0.40	0.96	0.39	0.39
Low flow pulses	1.03	0.37	0.37	1.02	0.32	0.32	1.03	0.29	0.30	1.02	0.35	0.35	0.99	0.23	0.23	1.00	0.23	0.23
High pulses	1.00	0.53	0.52	1.06	0.50	0.51	1.04	0.34	0.34	1.07	0.49	0.49	0.98	0.24	0.24	0.99	0.23	0.23
Low duration	1.23	0.75	0.78	1.14	0.57	0.58	1.08	0.44	0.45	1.18	0.64	0.67	0.99	0.34	0.34	0.99	0.35	0.34
High duration	1.56	1.37	1.48	1.25	0.92	0.95	1.09	0.58	0.59	1.16	0.67	0.69	0.98	0.29	0.29	0.98	0.30	0.30
Flood interval	1.02	0.37	0.37	1.04	0.36	0.36	1.02	0.32	0.32	1.04	0.39	0.40	1.00	0.29	0.29	1.00	0.29	0.29
Flood-free days	1.01	0.28	0.28	1.06	0.32	0.32	1.01	0.19	0.19	0.97	0.22	0.23	1.00	0.15	0.15	1.00	0.16	0.16
Predictability	0.99	0.28	0.28	1.01	0.27	0.27	1.01	0.24	0.24	1.02	0.26	0.26	1.00	0.18	0.18	1.00	0.19	0.19
Reversals	0.98	0.21	0.21	0.99	0.19	0.19	1.00	0.17	0.17	0.99	0.20	0.20	1.00	0.15	0.15	1.00	0.15	0.15

01151500, however, post-dam observations were similar to pre-dam observations, and the *O*/*E* ratio was relatively unchanged from the pre-dam period.

Modelled estimates of *E* for daily variability at site 01654000 revealed substantial departures from the natural flow regime (Figure 6) through time. Observed values from the first 20 years of record were substantially larger than predictions of *E*, and values from subsequent 20-year periods continued to increase. *O*/*E* ratios were markedly outside the reference-site distribution for all time periods, indicating progressively more severe hydrological alteration through time.

The expected duration of low flows varied from 53 to 88 days year^{-1} throughout the Provo River basin (Figure 7), and appeared to be significantly altered in some segments. Observed low flow duration (days year^{-1}) was close ($O = 59$) to expectations ($E = 53$–60) at the uppermost gage, but 4-50× less ($O = 1$–16) than expected ($E = 60$–70) at the next three downstream gages. In contrast, observed low flow duration ($O = 183$) at the most downstream gage was ∼3× greater than the expected ($E = 60$–70) natural condition.

DISCUSSION

Hydrological assessment is the process of determining whether human activity has altered the hydrological attributes of an ecosystem. Central to this process is a specification of the hydrological attributes that are expected in the absence of human modifications (i.e. pristine condition) or that are least modified given current socioeconomic and ecological limitations (i.e. best attainable, sensu Stoddard *et al*., 2006). A major distinction of our approach is that metrics of the expected natural flows (i.e. *E*) are predicted using climate, topography, soils and geology with a single national-scale model. A predictive model allowed us to quantify the natural flow regime for rivers that lack pre-disturbance flow records, which greatly increases the number of streams that can be assessed. Successful use of predictive models for hydrological assessments requires that models accurately predict the flow regime that should occur under undisturbed conditions. In addition, these predictions need to be made with sufficient precision that ecologically meaningful departures from the natural flow regime can be reliably detected.

Model performance and interpretability

Predictive models have promise for quantifying the reference conditions needed for assessing hydrological alteration, but our results suggest that more work may be needed to improve model performance and metric selection. Although the most precise estimates of *E* were obtained from predictive models, model error remained

DOI: 10.1002/rra

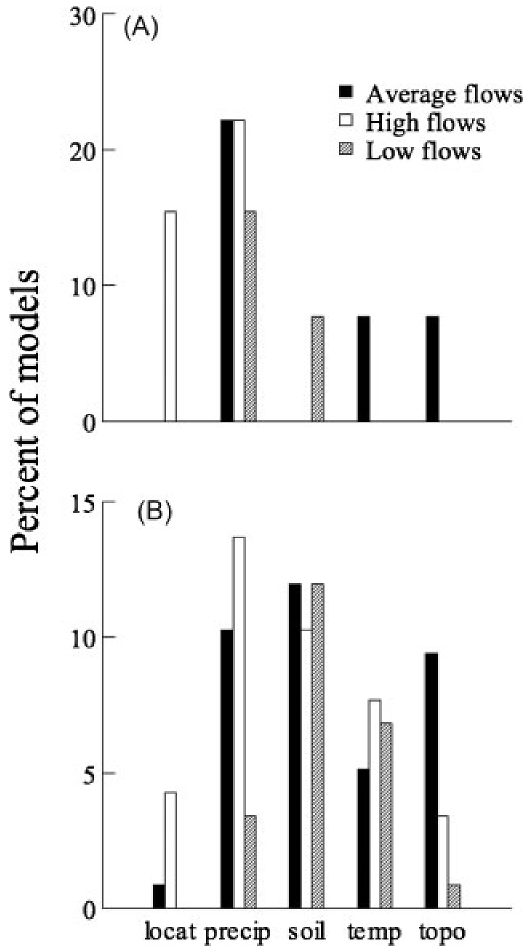

Figure 3. The frequency (% of total models) that an individual variable within each general category was ranked as the most important predictor by random forests models developed at a national scale (13 models, A), and at HUC 2 scales (117 models, B). General categories of predictors: locat = latitude & longitude, precip = precipitation, soil = soil properties, temp = air temperature, topo = topography and basin size

relatively high (e.g. >20%) for most metrics. Consequently, only rather substantial (e.g. >20%) departures from natural hydrologic conditions are reliably detectable using our models, depending on the statistical thresholds used. What remains unknown is the magnitude of hydrologic alteration that elicits biological responses (Arthington *et al*., 2006), and therefore what level of precision is required of methods used in hydrologic assessments. If, for example, biological communities are influenced by changes in a metric of <10%, none of the methods we examined would be sufficient to detect ecologically meaningful departures from the natural flow regime. By comparison, precision of biological metrics is frequently <20% (Wright *et al*., 2000; Ostermiller and Hawkins, 2004, Van Sickle *et al*., 2007), but national assessments have been performed using biological indicators with precision slightly >20% (USEPA, 2006b).

The utility of predictive models is enhanced if the underlying associations between predictor variables and hydrologic metrics are interpretable. Our nationwide models showed the dominant, continental-scale effects of precipitation patterns on hydrologic attributes, which have been reported in previous work (Wolock and McCabe, 1999). At regional scales, hydrologic attributes appeared to be influenced by local-scale factors such as soil properties and basin topography in addition to climatic factors (e.g. temperature and precipitation). An exhaustive analysis of the environmental factors that influence streamflow properties at various spatial scales is beyond the scope of this paper, but would likely identify predictors that would improve future predictive models.

DOI: 10.1002/rra

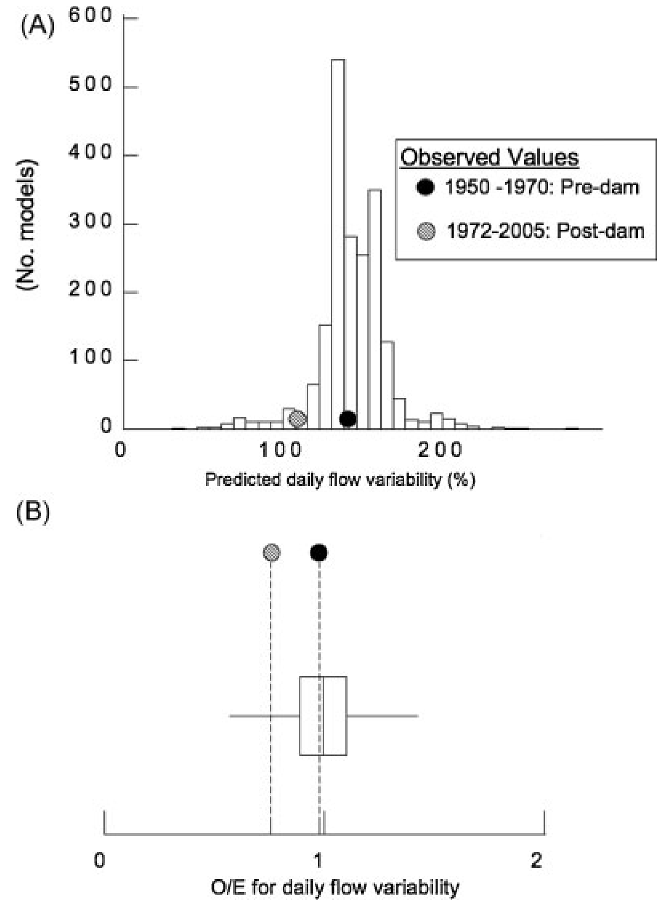

Figure 4. Observed daily flow variability of pre- and post-dam time periods compared to distribution of modelled expectations at gage # 05592000 (A). Ratios of observed and expected (=mean of expected distribution) flow variability for each time period compared to distribution of *O*/*E* values at reference sites nationwide (B)

Alternative approaches for predicting the natural flow regime

A major objective of landscape partitioning in environmental assessments is to improve precision of estimates of the expected condition by 'controlling' for natural variation (Hawkins and Norris, 2000). Regions are often a dominant framework for biological assessments, but there is significant evidence that environmental heterogeneity within ecoregions can be as large as that among regions (Hawkins *et al*., 2000a), and that variability in ecological conditions is best explained by a combination of regional and local (e.g. basin-scale) factors. There is also evidence that the ability of ecoregions to explain spatial variability in some aquatic ecosystem characteristics is no better than arbitrarily defined geometric regions (Wolock *et al*., 2004). Our results suggest that the environmental drivers of streamflow vary substantially even within relatively homogenous regions. The imprecision of estimates of *E* based on HUC2s indicates much heterogeneity in environmental factors that influence streamflow at that spatial scale. Precision of all metrics improved when *E* was estimated from more homogenous regions (ecoregion level 3), but was still less than modelled estimates for all metrics. Further, because estimates of *E* based on regions of similar hydrologic characteristics (e.g., HLR) were also less precise than predictive models, there appears to be strong evidence that local basin-scale factors (e.g. topography) in addition to (and interactive with) regional-scale factors (e.g. climate) must be considered when developing estimates of the natural flow regime. Such environmental heterogeneity within relatively small regions (e.g. level 3 ecoregions) also portends difficulty in estimating *E* for an assessed site by using data from reference gages within some proximity (e.g. 100 km radius). Such an approach is likely to fail in areas where topographic relief or geologic transitions produce sharp differences in climate, soils and

DOI: 10.1002/rra

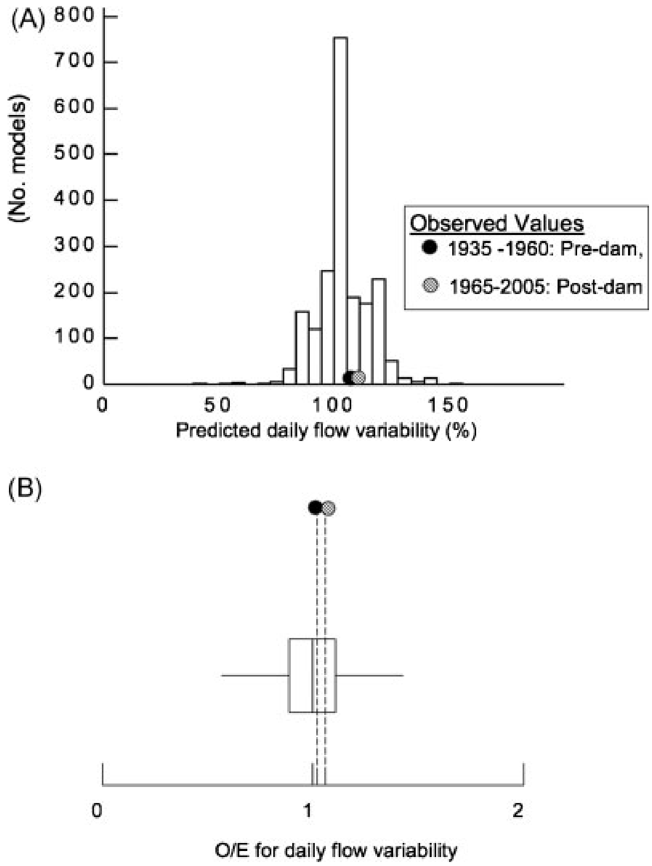

Figure 5. Observed daily flow variability of pre- and post-dam time periods compared to distribution of modelled expectations at gage # 01151500 (A). Ratios of observed and expected (=mean of expected distribution) flow variability for each time period compared to distribution of *O*/*E* values at reference sites nationwide (B)

the pathways of water through a watershed. An approach pioneered by Linke *et al.* (2005) uses the concept of proximity in multivariate space (e.g. multiple environmental gradients) to identify sets of reference sites that are relevant to a test site, which avoids the pitfalls of stratification.

We did not attempt to predict *E* from pre-defined stream classes as has been suggested by some (Arthington *et al*., 2006; Henriksen *et al*., 2006). The interest in establishing *E* from hydrologic classes defined *a priori* grew from the perceived need to tailor predictions for specific environmental conditions. Although we do not doubt the heuristic usefulness of stream hydrologic classes (e.g. perennial versus intermittent), we believe the concept holds some disadvantages for predicting *E*. The most notable difficulty with hydrologic classes is the inevitable error associated with accurately classifying non-reference sites. Since classes are based on hydrologic attributes from the discharge record, sites with known or suspected hydrologic modification cannot be directly classified. Discriminant function models can be developed that use watershed features to classify unknown sites (Henriksen *et al*., 2006; Sanborn and Bledsoe, 2006), but such models will almost certainly add an additional source of error to the ultimate predictions of *E* when sites are misclassified. Further, stream classifications based on hydrologic attributes often reflect differences in climate and topography—which are implicit in ecoregion classifications—when considered at broad geographic scales (Sanborn and Bledsoe, 2006). Admittedly, there are examples of significant natural differences in hydrologic attributes among streams within relatively small geographic settings (Poff, 1996), but we believe this is the reason RF models were so successful at predicting *E* for most hydrologic metrics.

One potentially counterintuitive result is that for all metrics, a single national model performed just as well as regional (HUC2) predictive models. We expected region-specific models would be better than a national model at extracting relations between environmental gradients and hydrologic metrics because some natural variation (e.g.

DOI: 10.1002/rra

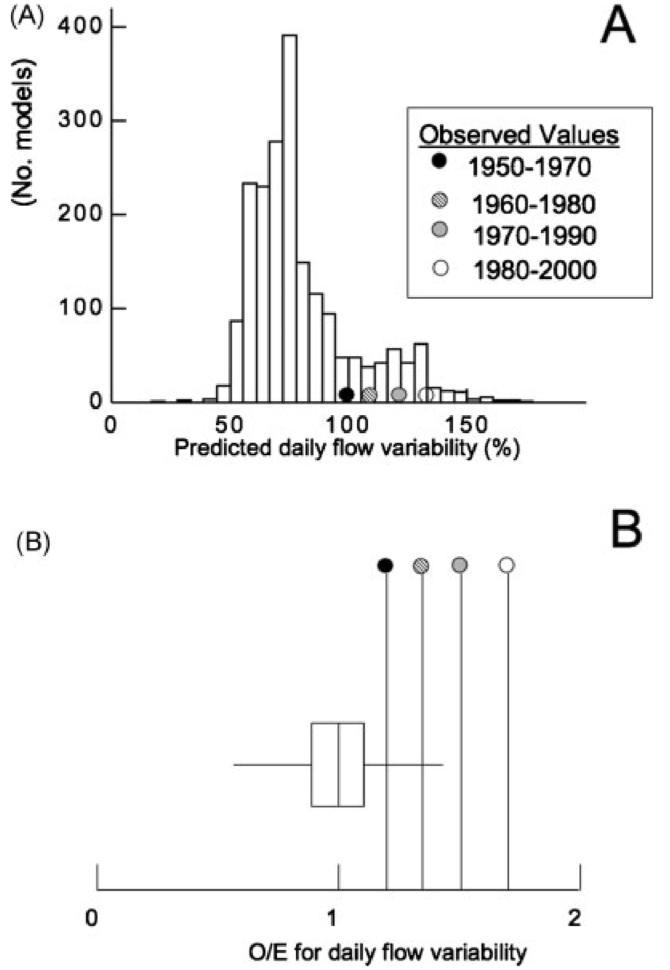

Figure 6. Observed daily flow variability of four time periods compared to distribution of modelled expectations at gage # 01654000 (A). Ratios of observed and expected (=mean of expected distribution) flow variability for each time period compared to the distribution of *O*/*E* values at reference sites nationwide (B)

climate) was 'controlled' through the regionalization process. It is apparent, however, that environmental heterogeneity within HUC2 basins was comparable to that manifested at a national scale. Alternatively, our statistical models may have been unable to improve predictions of *E* within HUC2 basins because important regional drivers of variation in flow regimes were inadequately represented by our predictor variables. Nevertheless, a single nationwide model is clearly appealing for conducting national assessments because there would be no need to aggregate multiple regional assessments based on models with potentially varying performance.

Example applications in hydrological assessment

Our application of predictive models was limited to a few illustrative examples, but demonstrated the approach has promise for estimating natural flow regimes and quantifying hydrologic alteration. The accuracy with which the predictive model estimated pre-dam flow characteristics at the two-dammed sites was not surprising given the demonstrated performance among reference sites. Our assessment of the dammed sites revealed that the dams did not alter hydrology equally, which is likely due to differences in dam operations and reservoir characteristics between the sites. A detailed analysis is beyond the scope of this study, but cursory examinations found that the reservoir above the hydrologically altered site had much larger storage capacity than that of the unaltered site. Our retrospective hydrologic assessment of an urbanized basin revealed a steadily increasing trend in flow variability.

DOI: 10.1002/rra

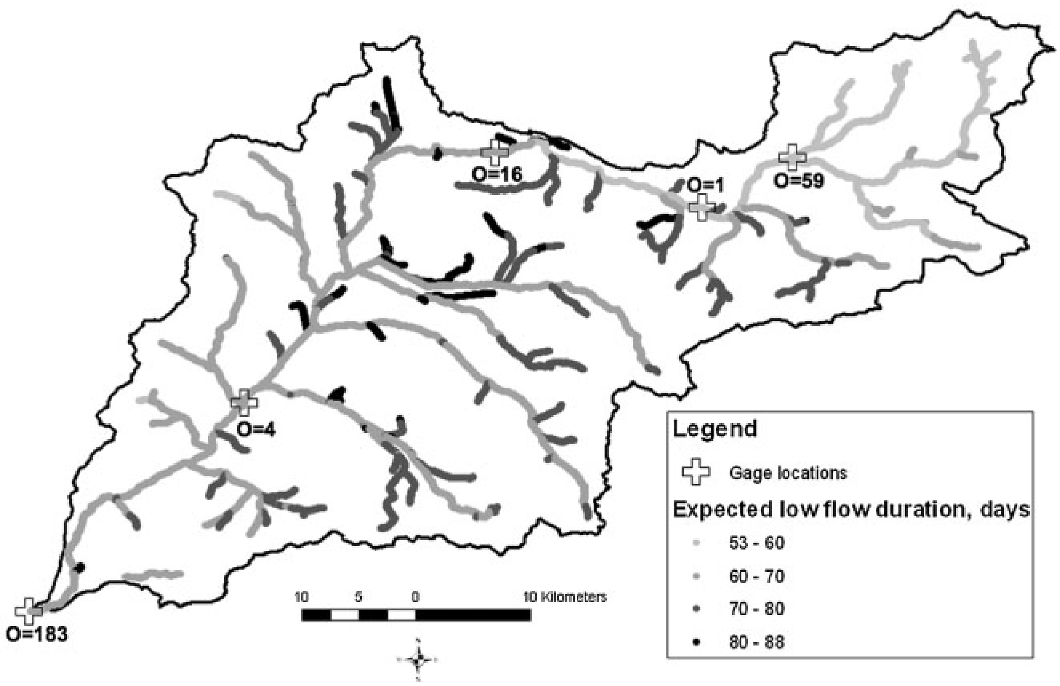

Figure 7. Model estimates for the expected low flow duration (days year^{-1}) for the stream network in the Provo River basin, Utah. Observed values for the period 1980–2000 are given at each gage

Previous trend analyses at this site (Jennings and Jarnagin, 2002) found that flashiness increased as the basin urbanized, even after correcting for trends in precipitation over the same time period. Our results suggest that significant hydrologic alteration ($O/E = 1.2$) occurred within the first 20 years of the flow record, over which time impervious cover in the basin increased from 3 to 21% (Jennings and Jarnagin, 2002). Impervious land cover increased to >33% through the remaining flow record and appears to have caused greater flow variability, a pattern well established in the literature (Paul and Meyer, 2001; Konrad and Booth, 2002).

Various thresholds for assessing hydrologic alteration have been proposed (Richter *et al*., 1997b; Henriksen *et al*., 2006), and will ultimately be influenced by known ecological responses to hydrological alteration and wider societal values. Nevertheless, our results indicate that the daily streamflow variability observed at one dammed sited and in the urbanized basin was well outside the ranges of variability caused by natural variation and model error. We emphasize that the presence of significant hydrologic alteration at the sites we examined does not imply that ecological degradation has also occurred.

Application of RF models to estimate *E* for ungaged stream segments, as demonstrated for the Provo River, is possible but requires some caution. Model predictions agreed well with observed low flow duration at the gage in the undisturbed headwaters, which is remarkable given the fact that the model was developed at a national scale. However, we did not evaluate whether the environmental settings (i.e. the suite of predictor variables) of all stream pixels within the Provo basin were within the scope of the reference sites used in model development. Statistical evaluations of this sort are routinely performed when assessing biological condition with predictive models (e.g. Ostermiller and Hawkins, 2004), and could in theory be developed for RF models. Predictions of *E* for river segments that do not meet this test should be interpreted with caution.

Hydrological assessments and fixed values of E

Assessing hydrological alteration with a fixed *E* is not ideal in all situations. We suspect that fixed values of *E*, and *O* derived from several years of record, are suitable for regional or national assessments of average flow conditions for a defined time period. However, the natural flow regime is temporally variable because it is influenced by temporal variation in precipitation and runoff patterns. Fixed values of *E* are therefore not suitable for site-specific assessments

DOI: 10.1002/rra

or setting specific river management goals. We offer several possible ways to address this deficiency. First, as demonstrated in the example assessments, *E* can be expressed as a distribution of estimates created during cross-validation within the modelling process. When expressed as a distribution, predictions of *E* for a site represent uncertainty associated with error at all stages of the measurement and assessment process, including spatiotemporal variation in flow over the period of record. From such a distribution, bounds of acceptable variation in *E* could be used as thresholds for assessing a single site through time. Indeed, this general concept has been applied to site-specific assessments where *E* was derived from nearby or pre-disturbance gage records (e.g. Richter *et al*., 1997b).

An alternative approach is to estimate *E* for various percentiles of the hydrologic metric. For example, separate models could be developed for the 25th, median, and 75th percentiles of low flow duration, which when applied to an assessed site would provide bounds of 'acceptable' deviation for *O*. An additional approach is to construct separate models for estimating *E* for years of above-and below-average precipitation. For example, a 'dry year' model could be developed using reference site hydrologic and climatic data from years that were below the long-term average precipitation. A separate model could be developed using data for years with above-average precipitation. Such models would in theory provide upper and lower bounds of *E* for an assessed site that bracket annual variation in precipitation and runoff.

Lastly, we emphasize that *E* may not be stationary due to long-term changes in rainfall and runoff patterns (Milly *et al*., 2008). *E* is based entirely upon the period of record summarized by the hydrologic metrics at the reference sites, which was post-1950 in this study. The effects of long-term climate change on *E* cannot be ignored, and could be examined by developing models from different periods of the flow record with corresponding climate data.

CONCLUSIONS

We conclude with an important caveat and a call for further research. For the majority of the nation's rivers and streams the natural flow regime must be modelled. Our intent was to suggest a nationally consistent method of modelling expected natural characteristics of the flow regime in gaged and ungaged rivers and streams. The modelling approach we described successfully predicted average attributes of the natural flow regime at undisturbed sites across the nation and in diverse environmental settings. We believe the models can be applied to all stream segments for which basin-level predictors are available and that occur within the experience of the reference sites we identified. Further, quantification of hydrological alteration is now presumably possible at thousands of gaged streams across the contiguous US with sufficient periods of record for the calculation of *O*.

Although our ability to quantify the expected natural flow regime and deviation from it have improved, the ecological effects of hydrological alteration are still largely unknown. As a result, ecologists are still unable to specify what constitutes an ecologically benign deviation from natural streamflow conditions (Arthington *et al*., 2006). Until we better understand the relationships between ecological health and the magnitude of hydrological alteration, broad-brushed management prescriptions allowing specific magnitudes of deviation from natural flow characteristics (e.g. 25% removal of low-flow magnitude) are without scientific foundation. We join the chorus of others (Richter *et al*., 1997b; Arthington *et al*., 2006) in calling for studies that simultaneously assess ecological conditions and hydrological alteration in an effort to better understand how increasingly scarce water resources can be managed to balance the needs of aquatic life and human society.

ACKNOWLEDGEMENTS

This study would not be possible without the diligence and attention to detail by hydrologists too numerous to name, who monitor, maintain and report streamflow across the nation. Several people provided insight and review comments that improved the manuscript. Harry Lins, Tom Fitzugh, Bill Wilber and Sandy Cooper provided helpful comments on early versions of the manuscript. The thoughtful comments of two anonymous reviewers are also gratefully acknowledged.

DOI: 10.1002/rra

REFERENCES

Acreman M, Dunbar MJ. 2004. Defining environmental flow requirements—a review. *Hydrology and Earth System Sciences* **8**: 861–876.

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 RC, Norris RH, Reynoldson TB. 2004. *Bioassessment of Freshwater Ecosystems: Using the Reference Condition Approach*. Kluwer: Boston, Massachusetts, USA.

Baxter RM. 1977. Environmental effects of dams and impoundments. *Annual Review of Ecology and Systematics* **8**: 255–283.

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

Bunn SE, Arthington AH. 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. *Environmental Management* **30**: 492–507.

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

Daymet. 2006. Numerical Terradynamic Simulation Group. University of Montana. www.daymet.org accessed June 1, 2007.

De'ath G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* **88**: 243–251.

De'ath G, Fabricus KE. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* **81**: 3178–3192.

Dynesius M, Nilsson C. 1994. Fragmentation and flow regulation of river systems in the northern third of the world. *Science* **266**: 753–762.

Fritz KM, Dodds WK. 2005. Harshness: characterisation of intermittent stream habitat over space and time. *Marine and Freshwater Research* **56**: 13–23.

GeoLytics. 2001. *Census 2000 and Street 2000*. GeoLytics, Inc.: East Brunswick, NJ.

Hastie T, Tibshirani R, Friedman J. 2001. *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer.

Hawkins CP, Norris RH. 2000. Performance of different landscape classifications for aquatic bioassessments: introduction to the series. *Journal of the North American Benthological Society* **19**: 367–369.

Hawkins CP, Norris RH, Gerritsen J, Hughes RM, Jackson SK, Johnson RK, Stevenson RJ. 2000b. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. *Journal of the North American Benthological Society* **19**: 451–556.

Hawkins CP, Norris RH, Hogue JN, Feminella JW. 2000a. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* **10**: 1456–1477.

Henriksen JA, Heasley J, Kennen JG, Nieswand S. 2006. Users' manual for the hydroecological integrity assessment process software (including New Jersey assessment tools). *Open File Report 2006–1093*, U.S. Geological Survey.

Horizon Systems. 2006. National Hydrography Dataset Plus. Horizon Systems Corporation. Available at http://www.horizon-systems.com/nhdplus/ accessed June 1, 2007.

Jackson RB, Carpenter SR, Dahm CN, McKnight DM, Naiman RJ, Postel SL, Running SW. 2001. Water in a changing world. *Ecological Applications* **11**: 1027–1045.

Jansson R, Nillson C, Dynesius M, Andersson E. 2000. Effects of river regulation on river-margin vegetation: a comparison of eight boreal rivers. *Ecological Applications* **10**: 203–224.

Jennings DB, Jarnagin ST. 2002. Changes in anthropogenic impervious surfaces, precipitation, and daily streamflow discharge: a historical perspective in a mid-atlantic sub-watershed. *Landscape Ecology* **17**: 471–489.

Konrad CP, Booth DB. 2002. Hydrologic trends associated with urban development in western Washington streams. *U.S. Geological Survey Water Resources-Investigations Report 02-4040*.

Lawler JJ, White D, Neilson RP, Blaustein AR. 2006. Predicting climate-induced range shifts: model differences and model reliability. *Global Climate Change Biology* **12**: 1568–1584.

Liaw A, Wiener M. 2002. Classification and regression by random forest. *R News* **2/3**: 18–22.

Linke S, Norris RH, Faith DP, Stockwell D. 2005. ANNA: a new prediction method for bioassessment programs. *Freshwater Biology* **50**: 147–158.

Lytle DH, Poff NL. 2004. Adaptation to natural flow regimes. *Trends in Ecology and Evolution* **19**: 94–100.

Milly PCD, Betancourt J, Falkenmark M, Hirsch RM, Kundzewicz ZW, Lettenmaier DP, Stouffer RJ. 2008. Stationarity is dead: whither water management? *Science* **319**: 573–574.

Nilsson C, Reidy CA, Dynesius M, Revenga C. 2005. Fragmentation and flow regulation of the world's large river systems. *Science* **308**: 405–408.

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

Ostermiller JD, Hawkins CP. 2004. Effects of sampling error on bioassessments of stream ecosystems: applications to RIVPACS-type models. *Journal of the North American Benthological Society* **23**: 363–382.

Paul MJ, Meyer JL. 2001. Streams in the urban landscape. *Annual Review of Ecology, Evolution, and Systematics* **32**: 333–365.

Pike RJ, Wilson SE. 1971. Elevation-relief ratio, hypsometric integral, and geomorphic area-attitude analysis. *Geological Society of America Bulletin* **82**: 1079–1084.

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.

DOI: 10.1002/rra

Poff NL, Allan JD, Bain MB, Karr JR, Prestegaard KL, Richter BD, Sparks RE, Stromberg JC. 1997. The natural flow regime: a paradigm for river conservation and restoration. *BioScience* **47**: 769–784.
Poff NL, Bledsoe BP, Cuhaciyan CO. 2006. Hydrologic variation with land use across the contiguous United States: geomorphic and ecological consequences for stream ecosystems. *Geomorphology* **79**: 264–285.
Poff NL, Olden JD, Merritt DM, Pepin DM. 2007. Homogenization of regional river dynamics by dams and global biodiversity implications. *Proceedings of the National Academy of Sciences* **104**: 5732–5737.
Prasad AM, Iverson LR, Liaw A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* **9**: 181–199.
Pringle CM, Freeman MC, Freeman BJ. 2000. Regional effects of hydrologic alterations on riverine macrobiota in the new world: tropical-temperate comparisons. *BioScience* **50**: 807–823.
Puckridge JT, Sheldon F, Walker KF, Boulton AJ. 1998. Flow variability and the ecology of large rivers. *Marine and Freshwater Research* **49**: 55–72.
R Development Core Team. 2006. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org
Reed JC, Bush CA. 2005. Generalized geologic map of the conterminous United States. In USGS National Atlas (available at http://pubs.usgs.gov/circ/1300/pdf/cir1300_508.pdf). US Geological Survey, Denver, Colorado.
Richter BD, Baumgartner JV, Braun DP, Powell J. 1997b. A spatial assessment of hydrologic alteration within a river network. *Regulated Rivers: Research and Management* **14**: 329–340.
Richter BD, Baumgartner JV, Wigington R, Braun DP. 1997a. How much water does a river need? *Freshwater Biology* **37**: 231–249.
Richter BD, Matthews R, Harrison DL, Wigington R. 2003. Ecologically sustainable river flows for river integrity. *Ecological Applications* **13**: 206–224.
Sanborn SC, Bledsoe BP. 2006. Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon. *Journal of Hydrology* **325**: 241–261.
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.
Tharme RE. 2003. A global perspective on environmental flow assessment: emerging trends in the development and application of environmental flow methodologies for rivers. *River Research and Applications* **19**: 397–441.
USACE [United States Army Corps of Engineers]. 2006. National inventory of dams. Available at http://crunch.tec.army.mil/nidpublic/webpages/nid.cfm accessed June 1, 2007.
USDA [United States Department of Agriculture]. 2006. U.S. general soil map. http://www.ncgc.nrcs.usda.gov/products/datasets/statsgo/ accessed June 1, 2007.
USEPA [United States Environmental Protection Agency]. 2006a. Water discharge permits compliance system. Available at http://www.epa.gov/enviro/html/pcs/index.html accessed June 1, 2007.
USEPA [United States Environmental Protection Agency]. 2006b. *Wadeable Streams Assessment: A Collaborative Survey of the Nation's Streams. EPA 841-B-06-002*. Office of Research and Development and Office of Water: Washington, DC.
USEPA [United States Environmental Protection Agency]. 2007. Ecoregions of North America. Available at http://www.epa.gov/wed/pages/ecoregions.htm
USGS [United States Geological Survey]. 2006. National elevation dataset. http://ned.usgs.gov accessed June 1, 2007.
USGS [United States Geological Survey]. 2007a. Water use in the United States. Available at http://water.usgs.gov/watuse/ accessed June 1, 2007.
USGS [United States Geological Survey]. 2007b. Water data reports. Available at http://pubs.usgs.gov/wdr/ accessed June 1, 2007.
Van-Sickle J, Hawkins CP, Larsen DP, Herlihy AT. 2005. A null model for the expected macroinvertebrate assemblage in streams. *Journal of the North American Benthological Society* **24**: 178–191.
Van-Sickle J, Huff DD, Hawkins CP. 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. *Freshwater Biology* **51**: 359–372.
Van Sickle J, Larsen DP, Hawkins CP. 2007. Exclusion of rare taxa effects performance of the O/E index in bioassessments. *Journal of the North American Benthological Society* **26**: 319–331.
Vogelman JE, Howard SM, Yang L, Larson CR, Wylie BK, Van Driel N. 2001. Completion of the 1990s national land cover data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing* **67**: 650–662.
Whittier TR, Stoddard JL, Larsen DP, Herlihy AT. 2007. Selecting reference sites for stream biological assessments: best professional judgment or objective criteria. *Journal of the North American Benthological Society* **26**: 349–360.
Wilcove DS, Rothstein D, Dubow J, Phillips A, Losos E. 1998. Quantifying threats to imperiled species in the United States. *BioScience* **48**: 607–615.
Wolock DM, McCabe GJ. 1999. Explaining spatial variability in mean annual runoff in the conterminous United States. *Journal of Climate Research* **11**: 149–159.
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**: 71–88.
Wright JF, Sutcliffe DW, Furse MT (eds). 2000. *Assessing the Biological Quality of Freshwaters: RIVPACS and Other Techniques*. Freshwater Biological Association: Cumbria, United Kingdom.
Yevjevich V. 1972. *Probability and Statistics in Hydrology.* Water Resources Publications: Littleton, Colorado, USA.

DOI: 10.1002/rra

APPENDIX

The 10 most important predictors in national-scale random forests models predicting hydrologic metrics

Metric	Top 10 predictors (in decreasing importance)
Daily variability	Mean number of days with measurable precipitation, per cent of soils with high runoff potential, per cent soils with moderately high runoff potential, mean soil erodibility factor, per cent fine (<0.074 mm) soils, per cent silt soils, per cent of soils with moderately low runoff potential, mean soil permeability, October mean precipitation, per cent sand soils
Skewness	May mean temperature, September mean temperature, August mean temperature, mean number of days with measurable precipitation, mean monthly maximum number of days with measurable precipitation, per cent sand soils, watershed area, mean soil permeability, potential evapo-transpiration, longitude
Annual runoff	Mean number of days with measurable precipitation, mean annual precipitation, November mean precipitation, mean monthly minimum number of days with measurable precipitation, mean monthly maximum number of days with measurable precipitation, mean monthly minimum precipitation, September mean precipitation, mean basin slope, per cent clay soils, April mean precipitation
Base flow index	Per cent soils with moderately high runoff potential, mean basin slope, per cent soils with moderately low runoff potential, mean soil erodibility factor, mean day of last frost, per cent soils with low runoff potential, July mean precipitation, per cent fine (<0.074 mm) soils, per cent soils with high runoff potential, mean soil permeability
Annual maximum flow	Mean annual precipitation, watershed area, May mean precipitation, January mean precipitation, mean rainfall/runoff factor, per cent soils with moderately high runoff potential, per cent sand soils, mean soil thickness, April mean precipitation, per cent fine (<0.074 mm) soils
Low flow pulses	Mean monthly minimum precipitation, watershed area, August mean precipitation, per cent soil organic matter, July mean precipitation, February mean precipitation, per cent soils with high runoff potential, mean soil bulk density, variability of minimum monthly air temperature, per cent soils with moderately low runoff potential
High flow pulses	September mean precipitation, mean soil erodibility factor, mean monthly minimum precipitation, watershed area, October mean air temperature, per cent soil organic matter, soil rainfall/runoff factor, November mean precipitation, October mean precipitation, May mean precipitation
Low flow duration	Mean monthly minimum precipitation, August mean precipitation, mean monthly minimum number of days with measurable precipitation, longitude, per cent soil organic matter, December mean air temperature, mean basin elevation, per cent clay soils, mean soil erodibility factor, watershed area
High flow duration	May mean precipitation, per cent soil organic matter, watershed area, mean relative humidity, October mean air temperature, August mean precipitation, soil rainfall/runoff factor, mean day of last frost, mean monthly maximum precipitation, mean basin elevation
Flood interval	Longitude, November mean air temperature, February mean air temperature, latitude, mean monthly maximum number of days with measurable precipitation, January mean temperature, mean number of days with measurable precipitation, per cent coarse (2 mm) soils, per cent coarse (5 mm) soils, variation of monthly minimum air temperature
Flood-free days	Longitude, December mean precipitation, October mean air temperature, mean number of days with measurable precipitation, variation in monthly minimum air temperature, mean basin elevation, August mean precipitation, basin mean slope, per cent soils with moderately low runoff potential, September mean air temperature
Predictability	Watershed area, soil permeability, August mean air temperature, per cent soils with moderately low runoff potential, mean number of days with measurable precipitation, variation of elevation in basin, per cent silt soils, per cent soils with low runoff potential, variation of monthly maximum air temperature, per cent soils with moderately high runoff potential
Reversals	Mean monthly minimum precipitation, longitude, watershed area, soil organic matter, May mean precipitation, soils with moderately low runoff potential, January mean air temperature, basin mean slope, September mean precipitation, mean relative humidity

DOI: 10.1002/rra