FINAL REPORT: 2010 Wetland Program Development Grant
DEVELOPMENT OF STREAM METRICS FOR USE IN THE CONSERVATION ASSESSMENT AND PRIORITIZATION SYSTEM (CAPS)
December 23, 2019

Elizabeth Homa, Casey Brown, Kevin McGarigal, Brad Compton and Scott Jackson
University of Massachusetts Amherst

Introduction

In order to improve the ability of the Conservation Assessment and Prioritization System (CAPS) to assess the ecological integrity of rivers and streams, we developed new stream metrics to better address stressors affecting these riverine ecosystems. Five new metrics were considered, although we were not able to develop metrics for two of these stressors (temperature alteration and geomorphic alteration) because sufficient data were not yet available. Three new metrics were developed for three important stressors affecting rivers and streams: hydrologic alterations, nitrogen enrichment, and phosphorus enrichment. These metrics are based on empirical models that were developed to look at the relationship between human basin modifications and the resulting impacts on nutrient concentration and stream hydrology. Natural (unaltered) conditions were estimated by simulating basins with no alterations, allow us to calculate the difference between natural and altered conditions with respect to these stressors.

Following is a description of the models for nitrogen and phosphorus enrichment that served as basis for CAPS metrics. Attached as Appendix A is a journal publication that describes development of a flow alteration model that we used to create the hydrologic alteration metric.

Summary of the Nitrogen and Phosphorus Enrichment Modeling Approach

Empirical models were developed to look at the relationship between human basin modifications and the resulting impacts on nutrient concentrations. We developed multiple linear regression models, using principal component analysis to guide independent variable selection, to estimate current, altered nutrient concentrations from a full range of both natural and anthropogenic basin characteristics. Natural nutrient concentrations are then estimated by simulating basins with no alterations, allowing the difference between the natural and altered nutrient concentrations to be investigated. The model suggests that discharges, cropland and cranberry bogs are correlated with increases in phosphorus concentrations, and impervious surface, discharges and cropland area are correlated with increases in nitrogen concentrations.

Data

Data for the nutrient concentrations, total phosphorus (TP) and total nitrogen (TN), were taken from the Massachusetts Department of Environmental Protection (Mass DEP) Division of Watershed Management (DWM) WPP final water quality data for the 2005-2010 monitoring years. Selected sites had at least four measurements and were not excluded by the Mass DEP quality control process. For nitrogen, the TN concentration data were available from 621 sites with an average of 7 samples per site and range of 4 – 70 measurements. For phosphorus the TP concentration data were available from 569 sites with an average of 7 samples per site and range of 4 – 70 measurements. The dates of data collection for the data set ranged in time from 1/2005 to 11/2009. After removing basins smaller than 2 km^2, there were 434 sites, shown below in Figure 1.

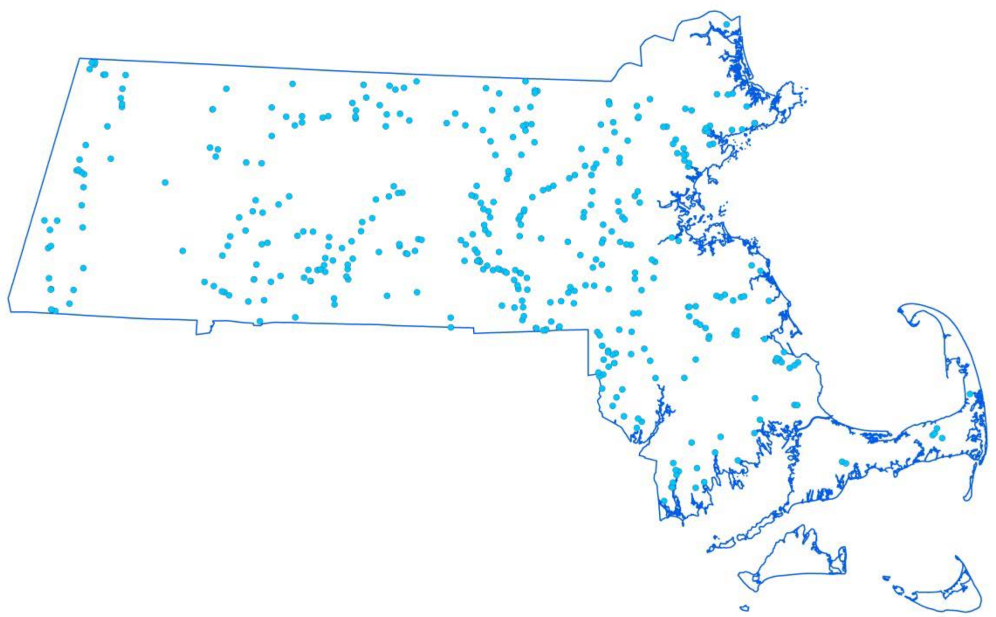

Figure 1 Locations of 434 sites for stream nutrient concentration data collected by Mass DEP 2005-2009.

Basin data

Data for the basin characteristics was collected from various sources as shown below in Table 1 and Table 2.

Table 1 Anthropogenic basin characteristics included in the model and sources of data. *Category for which details of the individual basin variables follow.

Variable	Source
Population	2010 census data
Discharge	NPDES discharges 2000-2005 in SYE wateruse db
Septic	1990 census data · % households on septic
Imperv	Mass GIS 2006
Land Use %*	Mass GIS 2005 Land Use

Land Uses Percentages: Alterations

- Cropland
- Cranberry_bog
- Nursery
- Orchard
- Pasture
- TOTAL PLANTED

- Commercial
- Industrial
- Urban_open
- Urban_public
- Transportation
- Mining
- Waste_disposal
- Junkyard
- TOTAL URBAN

- Multi_family_residential
- High_density_residential
- Medium_density_residential

- Low_density_residential
- TOTAL RESIDENTIAL

- Spectator_recreation
- Participatory_recreation
- Golf
- Water_based_recreation
- TOTAL RECREATION

Table 2 Natural basin characteristics included in the model and sources of data. *Category for which details of the individual basin variables follow.

Variable	Source
Basin Area	CAPS delineation
Climate*	PRISM
Bedrock Lithology*	Mass GIS 2004, Group A
Land Use %*	Mass GIS 2005 Land Use

Climate

- Max Temp
- Min Temp
- Mean Precipitation

The PRISM 800m climate data for 1981-2010:
http://www.prism.oregonstate.edu/products/matrix.phtml?vartype=tmin&view=maps

Bedrock Lithology

- Basin_Sedimentary
- Calcpelite
- Carbonate_Rocks
- Granite
- Mafic_Rocks
- Metamorphic_Rocks_Undivi

http://www.geo.umass.edu/stategeologist/frame_massgeo.htm

Land Uses Percentages: Natural

- Shrub_swamp
- Bog
- Shallow_marsh

- Deep_marsh
- Vernal_pool
- TOTAL LOWLAND
- Open_land
- Forest
- Forested_wetland
- Water_lentic
- Water_lotic

http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/datalayers/lus2005.html

Method

We performed natural log transformation of both the independent and dependent variables, consistent with previous regression models relating basin characteristics to lake or stream nutrient concentrations (Sorrano 2008, Dodds and Oakes 2004). Alternative transformations were evaluated both numerically, by considering univariate correlations between variables, as well as visually, by examining plots of independent against dependent variables. No improvement was found compared to the log-log model results.

$$\ln(c) = \beta_0 + \beta_1 ln(X_1 + 1) + \beta_2 ln(X_2 + 1) + \cdots \beta_{\mathrm{n}} ln(X_{\mathrm{n}} + 1)$$

where:

c	average nutrient concentration
X_i	basin characteristic i
β_i	model coefficient for basin characteristic X_i
n	the number of basin characteristics used as independent variables
i	1 to n

Adding one before taking the natural log of these terms allows a correct mapping between the value of the ln(X+1) term in the linear regression model and the value of X, so that when X = 0, ln(X+1) = 0. This solution allows for the correct representation of the ‘removal’ of anthropogenic modifications from the regression equation by setting the value of the corresponding terms to zero. The approach also avoids the problem of ln(0) being undefined.

Because the natural log is only defined for values above zero, any basin characteristic with negative values was shifted to be nonnegative by adding the minimum value plus a small increment to all values. Only one variable required shifting (min temp) as the remaining natural variables were all non-negative.

We used principal component analysis (PCA) to select a subset of the 51 highly inter-correlated basin characteristics considered as independent variables in each regression. Variables with the highest eigenvector loadings within inter-correlated sets of variables in the first set of components, determined using a scree test, were maintained in the set of candidate variables. The variable reduction process also involved determination of which characteristics had the highest univariate correlation coefficients with the dependent variable, and maintenance of variance inflation factors (VIFs) below 10, along with the analysis of the PCA loadings.

Based on the final sets of 7 and 8 variables, we conducted an “all subsets” regression algorithm, written in the statistical language R (R Development Core Team, 2006), which minimized the Akaike information criterion (AIC) to estimate the regression parameters and develop the final equations. One outlier was removed based on a very high Cook’s D (Kutner, 2005).

Weightings for the basin characteristics were calculated based on an ‘aquatic distance’ of each 30m square cell to the target point which, in this case, is the location of the sample taken for the nutrient concentration measurement. The ‘aquatic distance’ is calculated based on the land use of each cell traversed to reach the target point, whether the cell contains a stream channel and the slope of the land surface of the cell. The approach is based on a method presented by Randhir et al., 2001.

Models with weighted and non-weighted basin characteristics were compared. Variables weighted by aquatic distance had significantly higher univariate correlations with the dependent variables and the models using the weighted basin variables resulted in higher adjusted R^2 values, so only weighted variables were in the final models.

Results

Total Phosphorus

p	8
ADJR2	0.49
MSE	0.39
SE	0.70
BCF	0.19

Intercept	32.02
discharge	0.58
Cropland	0.26
Cranberry_bog	0.26
mintemp	0.54
meanprecip	-3.06
Forest	-0.44
Forested_wetland	0.11
Deep_marsh	-0.24

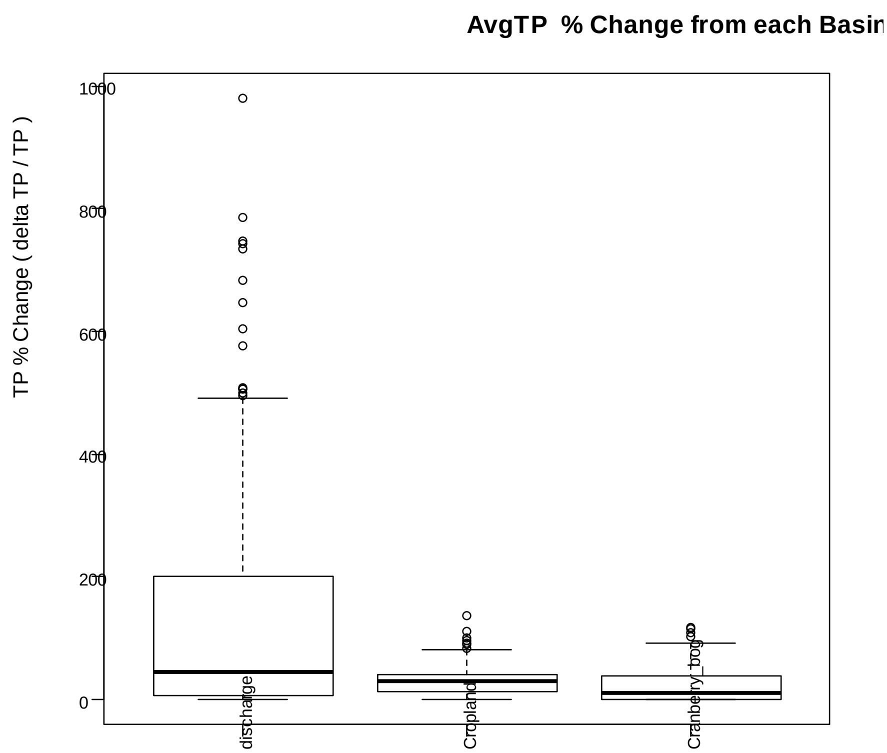

Figure 2 Percent change in Total Phosphorus due to each anthropogenic basin modification in the final model.

Total Nitrogen

p	7
ADJR2	0.63
MSE	0.18
SE	0.45
BCF	0.09

Intercept	4.79
imperv	0.47
discharge	0.43
septic	1.68
Urban_open	-0.19
Cropland	0.19
mintemp	0.38
Total_lowland	-0.11

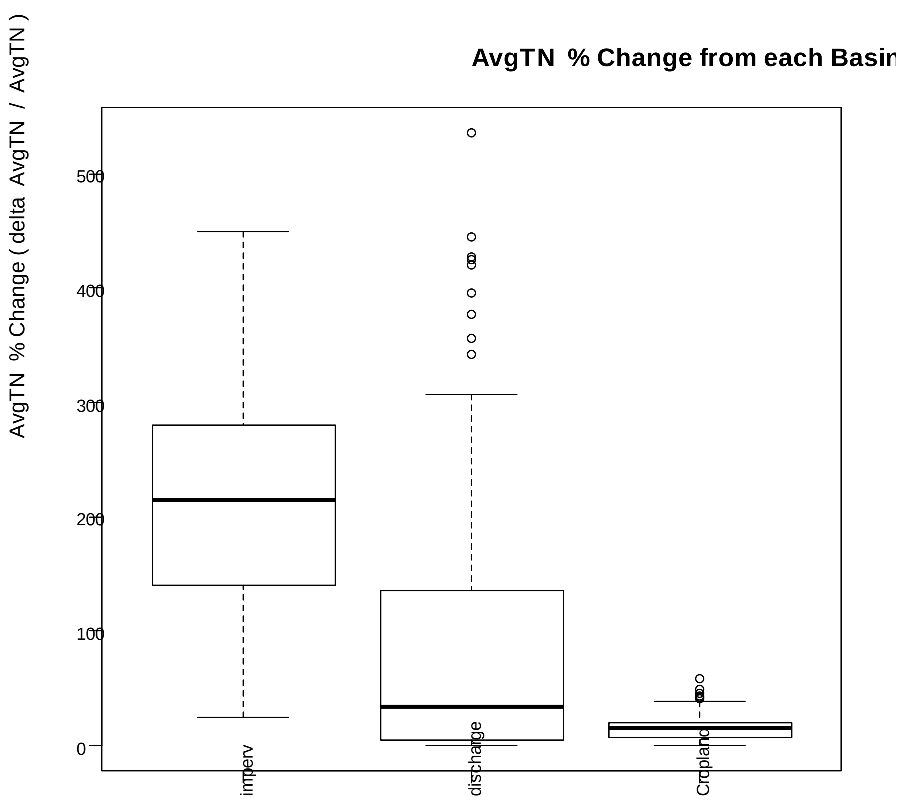

Figure 3 Percent change in Total Nitrogen due to each anthropogenic basin modification in the final model.

Results for total phosphorus, presented in Figure 2, show discharge as having the largest contribution to phosphorus concentration for the 434 stream locations in Massachusetts modeled in this study. The percent cropland and percent cranberry bogs are the other two basin modifications that the final regression model shows as having a significant impact on the phosphorus concentration.

For nitrogen the model results, presented in Figure 3, show that the percent impervious surface in a basin is correlated with a large percent increase in nitrogen. Discharge is also correlated with large increases in nitrogen, very high for some basins that receive large discharges, and percent cropland is also shown to be significant in the final model.

References:

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., & Smith, V. H. (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications, 8*(3), 559-568.

Dodds, W., & Oakes, R. (2004). A technique for establishing reference nutrient concentrations across watersheds affected by humans. *Limnology and Oceanography-Methods, 2*, 333-341.

Keller, S., T. Q. Zhang, S. Webb, R. Brugam, K. Johnson, and Z. -. Lin (2008), Effects of suburban land use on phosphorus fractions and speciation in the Upper Peruque Creek, Eastern Missouri, Water Environ. Res., 80, 316-323.

Kutner, M. H. (2005)*, Applied Linear Statistical Models* , 1396 pp., McGraw-Hill Irwin, Boston.

Lottig, N. R., E. H. Stanley, P. C. Hanson, and T. K. Kratz (2011), Comparison of regional stream and lake chemistry: Differences, similarities, and potential drivers, Limnol. Oceanogr., 56, 1551-1562.

Randhir, T., R. O'Connor, P. Penner, and D. Goodwin (2001), A watershed-based land prioritization model for water supply protection*, For. Ecol. Manage., 143*, 47-56.

Soranno, P. A., S. L. Hubler, S. R. Carpenter, and R. C. Lathrop (1996), Phosphorus loads to surface waters: A simple model to account for spatial pattern of land use, Ecol. Appl., 6, 865-878.

Soranno, P. A., Cheruvelil, K. S., Stevenson, R. J., Rollins, S. L., Holden, S. W., Heaton, S., et al. (2008). A framework for developing ecosystem-specific nutrient criteria: Integrating biological thresholds with predictive modeling. *Limnology and Oceanography, 53*(2), 773-787.

Appendix A

Estimating hydrologic alteration from basin characteristics in Massachusetts, Homa et al., 2013. Journal of Hydrology 503 (2013):196-208

Journal of Hydrology 503 (2013) 196–208

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Estimating hydrologic alteration from basin characteristics in Massachusetts

Elizabeth S. Homa [a,*], Casey Brown [a], Kevin McGarigal [b], Bradley W. Compton [b], Scott D. Jackson [b]

[a] Department of Civil and Environmental Engineering, University of Massachusetts, Amherst, MA 01003, United States
[b] Department of Environmental Conservation, Holdsworth Natural Resources Center, University of Massachusetts, Amherst, MA 01003, United States

ARTICLE INFO

Article history:
Received 28 February 2013
Received in revised form 7 September 2013
Accepted 8 September 2013
Available online 18 September 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Vazken Andréassian, Associate Editor

Keywords:
Hydrologic alteration
Ecodeficit
Ecochange
Flow duration curve
Multivariate regression models
Land use

SUMMARY

It is clear that humans are impacting the water cycle. There is interest in both determining where and how aquatic systems are most impacted by human development, and in determining the types and locations of basin modifications that are having the most impact. Instead of complex physical models of individual basins, we propose the use of a statistical approach to look at the relationship between human basin modifications and the resulting impacts on streamflow. We develop a set of multiple linear regression models, using principal component analysis to guide independent variable selection, to estimate current, altered streamflow from a full range of both natural and anthropogenic basin characteristics. Natural streamflow is then estimated by simulating basins with no alterations, and the difference between the natural and altered streamflow are summarized by use of the ecochange percent metric. The model suggests that dam storage, water withdrawals and discharges, and land use all impact stream flow and non-point source land use modifications such as impervious cover are potentially increasing low flows. The approach provides an opportunity to increase our understanding of the relation between human basin modifications and changes in streamflow. The model developed could potentially be used to estimate streamflow alteration at ungaged sites.

1. Introduction

Humans have been impacting the water cycle for thousands of years. In the past, a primary focus of human development was gaining control of water, securing a reliable supply of water and developing methods to increase agricultural production. The ability for humans to accomplish these goals has rapidly increased and significant alterations in the water cycle have occurred since industrialization (Vorosmarty and Sahagian, 2000; Jackson et al., 2001). Most commonly, the attainment of these goals outweighed any consideration for negative impacts on the functioning of the natural ecosystem. We are only now coming to appreciate the benefits of services provided to society by naturally functioning riverine systems (Poff et al., 1997; Richter et al., 1997; Jackson et al., 2001; Postel and Richter, 2003; Tharme, 2003).

Human development is clearly altering the natural characteristics of streamflow around the world (Jackson et al., 2001; Vorosmarty and Sahagian, 2000). The open question is: *how* are human alterations of the landscape changing streamflow? Many studies have focused on climate change (Risbey and Entekhabi, 1996; Chiew and McMahon, 2002; Milly et al., 2005), optimizing reservoir operations (Jager and Smith, 2008; Labadie, 2004; Suen and Eheart, 2006; Wardlaw and Sharif, 1999), or the total amount of water available to withdraw for human consumption (Weiskel et al., 2007; Archfield et al., 2010). However, changes to the land surface are also directly altering streamflows (Konrad and Booth, 2002; Foley et al., 2005; Poff et al., 2006). Thus, in addition to alterations to streamflow caused by "point sources", such as dams and water use, we can now benefit from making use of technology allowing us to consider "non-point source" alterations when evaluating human impacts on streamflows.

Previous assessments of distributed effects of human activities on water quality (Soranno et al., 2008) have not attempted to quantify the cumulative effect of various human activities on water quantity. The many studies addressing sources of hydrologic alteration have focused on the link to a specific type of basin alteration. For example, many studies have focused on analyzing the relationship between impervious cover or other land use changes to hydrologic alteration without considering other effects such as reservoir storage and water withdrawals (Jennings and Jarnagin, 2002; Poff et al., 2006; Roy and Shuster, 2009; Weiskel et al., 2010; Jacobson, 2011; Yang et al., 2011). It is now possible to take a more holistic approach with newly available data and analysis tools. Much geospatial data needed to quantify basin alterations is now

* Corresponding author. Tel.: +1 781 235 5576.
E-mail address: homa@ecs.umass.edu (E.S. Homa).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.09.008

available and the spatial extent and level of detail are increasing (Carlisle et al., 2010; Coles et al., 2010; Falcone et al., 2010; Weiskel et al., 2010). This study focused on developing a method of evaluating the relative impacts of both point and non-point source anthropogenic basin modifications on streamflow, and estimating the degree of hydrologic alteration at any site.

A major challenge in quantifying human alterations to streamflow is defining the "natural" or "unaltered" conditions. Human alterations to the land surface of a basin are more easily quantified because these types of alterations, such as percent impervious cover or percent agricultural land, can simply be considered nonexistent in an unaltered state. Quantification of the alterations to streamflow, on the other hand, requires a pre-alteration characterization of streamflow. This "natural" state of streamflow is typically represented by flow measurements before the basin modifications, or by estimating a natural flow for the basin (Pechlivanidis et al., 2011). Long records of stream flow data are required for the first approach and are often not available during the time periods needed before and after a modification such as construction of a dam or some other land alteration. Models for estimating natural flow range from the basic Drainage Area Ratio method to regional regression models developed from 'least altered' reference sites (Vogel et al., 1999; Armstrong et al., 2004; Kroll et al., 2004; Sanborn and Bledsoe, 2006; Armstrong et al., 2008; Archfield et al., 2010; Carlisle et al., 2010). These models make use of a range of geological, climate and topographic characteristics to estimate natural flows, and require the selection of reference sites with minimally altered drainage basins. Major limitations of these models include the requirement to somewhat arbitrarily identify what constitutes "minimally" altered, and the limited and decreasing number of minimally altered drainage basins.

An alternate modeling approach to estimate streamflow for a basin is to develop a physically-based hydrological model that is calibrated to represent the ungaged basin (Choi and Deal, 2008; Karvonen et al., 1999). Model parameter values are typically drawn from land cover attributes. Archfield et al. (2010) used regional regression to estimate parameter values for ungauged sites based on calibrated parameter values at nearby gauged sites. However, model structure uncertainty, input uncertainty and calibration uncertainty (Steinschneider et al., 2012) impede the application of these techniques for estimation of natural flow when prediction of alteration effects are likely to be dominated by uncertainty. These rainfall runoff models are typically costly and time consuming to build and complex to calibrate, so there is a high cost barrier and there can be a high degree of uncertainty in the results (O'Connell et al., 2007; Bulygina et al., 2011). However, new methods are being developed to auto-calibrate rainfall runoff models for ungaged catchments using regionalized data for soil hydrology and land use classifications (Bulygina et al., 2011; McIntyre and Marshall, 2010; Choi and Deal, 2008).

We chose to address the challenge presented by lack of definitively natural streamflow data by using all gaged sites in our regression model. We propose it is possible to learn about the relationship between basin alterations and streamflow by including all available sites in the model, not only pristine or 'least altered' sites, but sites representing a full range of basin alterations. Though there may be risk of additional model error and bias by using less homogenous data, the ability to consider a large number of characteristics for a large number of sites makes this approach appealing. Basin characteristics, both natural and anthropogenic, are included as potential independent variables to predict streamflow, and "natural" streamflow is estimated by setting the anthropogenic terms to zero. The difference between observed streamflow and estimated "natural" streamflow provides an estimate of streamflow alteration. Estimation of reference conditions for nutrient concentrations has been done in this way by Dodds and Oakes (2004) and Soranno et al. (2008). However, the only known study at this time to have modeled streamflow as a function of both natural and anthropogenic characteristics was done by Fitzhugh and Vogel (2011), who developed regional regression models of the median 1-day maximum flow to evaluate the impact of dams on flood flows in the US.

We suggest that there are multiple benefits to this approach. More sites are available if the regression model development is not limited to those with only minimally altered drainage basins, allowing the use of modeling techniques that are only possible with a larger data set. This approach does not require the rather arbitrary definition of what constitutes "minimally" altered, which is highly problematic in human-dominated landscapes such as the northeastern United States. Also, this approach presents the opportunity to estimate a 'completely' natural state, as compared to only estimating a 'least altered' state. Lastly, although not unique to this approach, this approach allows for estimation of hydrologic alteration at any ungaged site with characteristics within the range of the training data, with no streamflow record required.

2. Methods

2.1. Metric of alteration

In order to estimate the impact of human alterations on streamflow, we need to be able to quantify both the anthropogenic basin modifications and the resulting alterations of streamflow. The current conditions of anthropogenic modifications (such as percent impervious surface) can directly represent the basin alterations, since the natural state without modification would be, for example, no impervious surface at all. However, estimation of the alteration of streamflow requires metrics that represent the change from natural streamflow to altered streamflow.

Extensive research has been conducted on flow statistics to represent alterations to natural streamflow regimes. Many different hydrologic indices have been developed to characterize flow alteration (Richter et al., 1996; Olden and Poff, 2003; Gao et al., 2009; Poff et al., 2010; Poff and Zimmerman, 2010). The selection of a few simple statistics in the past has progressed, given new techniques and new computing power, to the development of many new approaches to quantify streamflow alteration. However, extensive sets of statistics such as The Nature Conservancy's 32 Indices of Hydrologic Alteration (IHA) statistics are often reduced to a smaller subset when actually applied. Gao et al. (2009) conducted an analysis of the IHA statistics using principal component analysis (PCA) and concluded that the annual ecodeficit and ecosurplus statistics (Homa et al., 2005; Vogel et al., 2007) best summarized the variability represented in the IHA. As shown in Fig. 1, the ecodeficit is defined as the area below the 'natural' flow duration curve (FDC) and the ecosurplus is defined as the area above the natural FDC, both normalized by the total area under the natural median annual FDC. The ecodeficit represents the amount of water that is not available due to the alterations and the ecosurplus represents extra water due to the alterations. A form of these two statistics are used in this study to summarize the effect of the alterations in each basin.

A benefit of using these particular statistics is that the ecodeficit and ecosurplus are calculated by comparing natural and altered FDCs (Vogel and Fennessey, 1994). Therefore, only FDCs need to be estimated for an ungaged site in order to calculate this metric of alteration; generating an estimate of a full daily time series of flow is not necessary. For this study an FDC is estimated by selecting a set of exceedance probabilities (0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95) and predicting each flow quantile

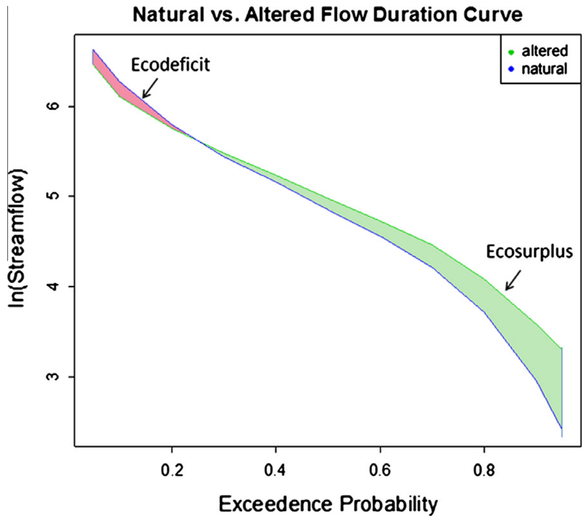

Fig. 1. An example of natural and altered Flow Duration Curves (FDCs) showing the definition of the ecodeficit and ecosurplus metrics.

with a separate regression equation. Each regression equation takes the general form of Eq. (1), described below.

$$\ln(Q_i) = \beta_0 + \beta_1 \ln(X_1) + \beta_2 \ln(X_2) + \cdots + \beta_n \ln(X_n) \tag{1}$$

where Q_i, stream flow at exceedance probability i; X_j, basin characteristic j; n number of variables; β_j, model coefficients for basin characteristic j.

The set of regression equations for the various streamflow quantiles provide estimates of actual flow for any basin for which the data are available to provide the independent variables. The estimate of natural flow is calculated from the same equations, but with the anthropogenic terms set to zero. The difference (or delta) in predicted versus natural streamflow as a percentage of the natural streamflow is measured at each modeled exceedence probability, as shown in Fig. 2. The average percent change in flow is calculated by normalizing the delta by the streamflow quantile and then averaging across the deficit or surplus. These percents are referred to as the ecodeficit percent and ecosurplus percent,

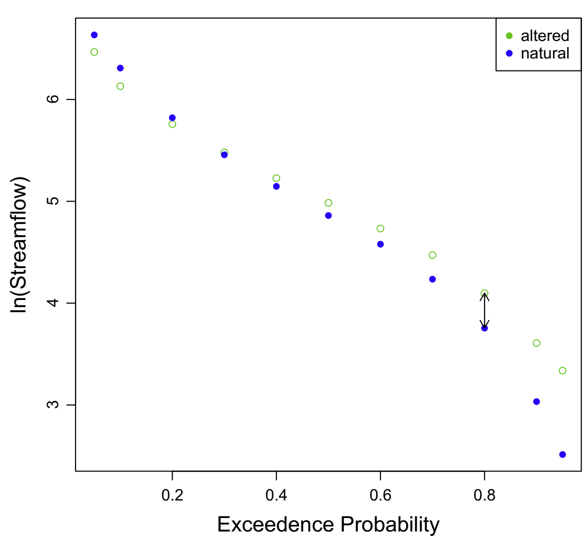

Fig. 2. An example of the estimation of natural and altered Flow Duration Curves (FDCs) for a sample basin. The delta between the points estimating the FDCs, shown at one exceedence probability by the two headed arrow, is used to estimate the ecodeficit and ecosurplus statistics.

and are a representation of %ΔQ. In order to rank sites by total alteration, these two percentages were summed to calculate a "total ecochange" metric.

2.2. Data

The USGS released the Geospatial Attributes of Gages for Evaluating Streamflow version II (GAGES II) in September, 2011 which provides delineated watershed boundaries for 9322 USGS stream gages across the nation. This dataset includes over 200 basin characteristics for each gage from a variety of primary sources. Daily streamflow data for at least 20 years is available from the USGS for each gage included in the dataset. The geospatial data includes both physical and climate data (soil, topology, temperature, precipitation, etc.) and anthropogenic basin characteristics (population, impervious surface, water use, dam density/storage). A reference to the original data source for each characteristic is provided (Falcone et al., 2010).

The data for the basin characteristics for our study come from the GAGES II dataset except for water use and discharges. These data were drawn from water withdrawal and discharge data available for basins within the state of Massachusetts (MA) as described below. The subset of sites selected for this study includes the six New England states: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island and Connecticut. There are 406 stream gages in these six states included in the GAGES II dataset. Of these 406, 190 gages have daily flow records available for 10/1/1996 to 9/30/2011 with fewer than 150 days of missing data over this fifteen year period. There are 42 out of these 190 drainage basins that are completely within MA, and thus 42 sites with point withdrawal and discharge data in addition to the basin characteristic data provided by Gages II.

Drainage basins for the 190 gages in the NE range in size from 10 to 25,000 km^2, and from 12 to 1785 km^2 for the subset of 42 basins within MA. Fig. 3 shows the locations of the 190 selected gages throughout New England.

The natural basin characteristics in GAGES II that were considered as candidate independent variables included drainage area, gage location (latitude and longitude), monthly basin temperatures and precipitation, geology (% soil types), elevation, stream density, and percent of land cover classes, resulting in a total of 86 variables. The anthropogenic basin characteristics that we considered as candidates included population, road density, impervious cover, number of dams, dam density, dam storage and percent of land use classes, resulting in a total of 12 variables. Data for the anthropogenic characteristics selected as independent variables in the final model (see below) will be described in more detail here. Fig. 4 below provides box plots of the four selected anthropogenic basin characteristics.

The data for watershed percent impervious surfaces was derived from the 30 m resolution U.S. Geological Survey (USGS) National Land Cover Database 2006 (NLCD06) (Falcone et al., 2010). The values for the 190 selected drainage basins ranged from 0.0025% for the Allagash River basin in northern Maine to 42.0% for the Aberjona River at Winchester, MA. The minimum percent impervious cover of all the basins within MA was 0.245% for a subbasin of the Swift River.

The dam storage variable (STOR_NID_2009) in GAGES II is summed for each basin from a 2009 National Inventory of Dams (NID) database after being cross checked and corrected by the USGS (U.S. Army Corps of Engineers, 2010; Falcone et al., 2010). The database includes a total of 4083 dams in the six NE states with storage data, 1590 of those with storage data in MA. The values for dam storage for the NE basins ranges from 0 (17 basins) to 1093 megaliters/km^2 for the basin of the Chicopee River at

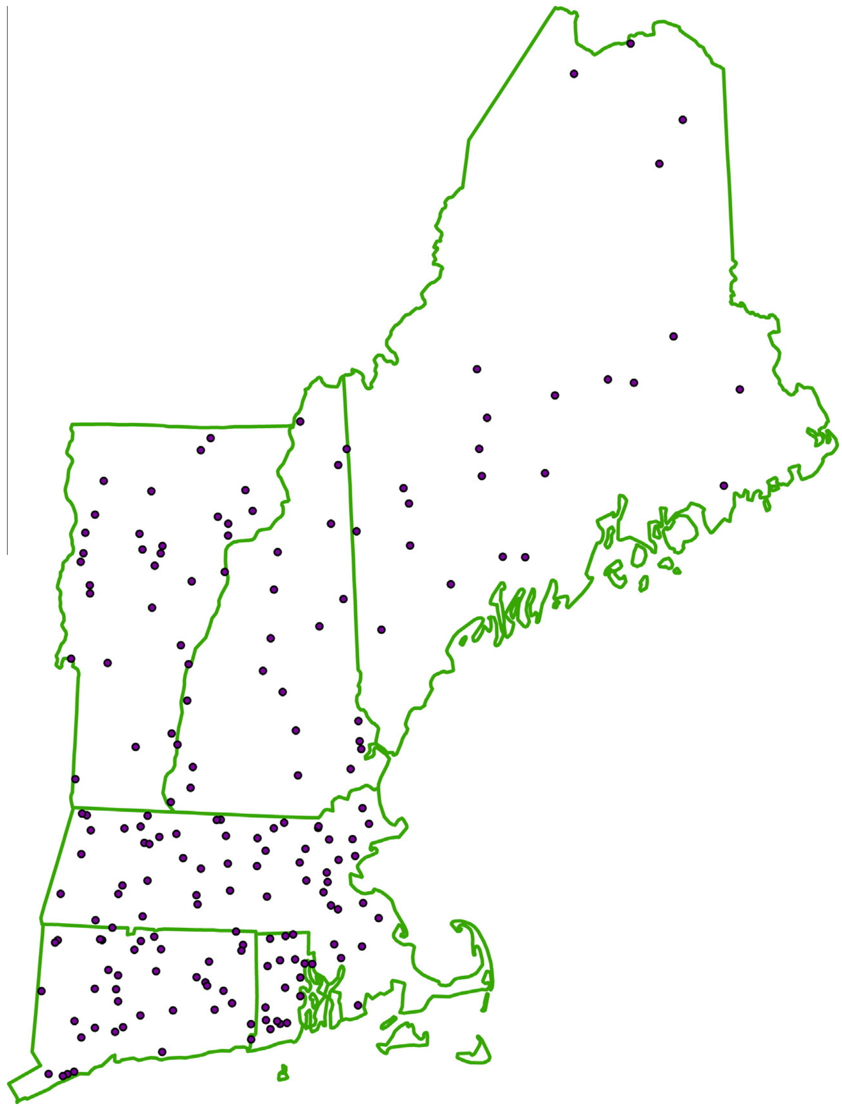

Fig. 3. Locations of the 190 USGS stream gages in the six Northeast states used in this study with available flow records for the period 10/1/96–9/30/10.

Indian Orchard, MA (Note that the Chicopee River basin includes the Quabbin Reservoir, the largest reservoir in New England).

The GAGES II water withdrawal variable was derived from county-level estimates (Hutson, 2007) which do not provide enough accuracy for our study given that basin sizes started at 10 km^2 and counties in this part of the country are much larger. The water use database developed for the Massachusetts Sustainable Yield Estimator provided the opportunity to utilize much more accurate estimates of average basin water use for the basins within the state of MA (Archfield et al., 2010). The database includes 6581 georeferenced points with ground water (GW) and surface water (SW) withdrawal and/or discharge rates regulated by the MA Department of Environmental Protection and the U.S. Environmental Protection Agency (EPA). The points covered by the regulations included GW and SW public water supply withdrawals greater than 100,000 gal/day, pollutant discharges greater than 10,000 gal/day, and National Pollutant Discharge Elimination System (NPDES) regulated SW discharges (Archfield et al., 2010). The values of water withdrawals for the 42 MA basins ranged from 0 (2 basins) to 285 megaliters/yr/km^2 for the basin of the Jones Rivers at Kingston. The values for the discharges ranged from 0 (8 basins) to 87.1 megaliters/yr/km^2 for the North Nashua River basin.

The best possible match was made between these dates for basin characteristics and the years of averaged flow data. The impervious surface data is from the NLCD06 which is 2006. The dam storage data is from 2009, and the water use and discharge data for Massachusetts is an average of data from 2000–2004.

We should point out that we are using an average of a 15 year period to represent a point in time (the most "current" point available), and comparing it to pre-settlement time. We propose that change during this 15 year time is not significant compared to change during the ~400 year period. The GagesII NLCD01_06_DEV basin characteristic representing the Watershed percent which changed to "Developed" (urban) land (NLCD classes 21–24) between NLCD 2001 and 2006 range from 0% to only 6.1%, whereas NLCD01_06_DEV representing the Watershed percent "developed" (urban) in 2006 ranged from 0.05% to 92.4% for the 42 study sites in MA.

2.3. Model development

We calculated streamflow quantiles from the USGS flow record for each stream gage ($n = 190$) at each of the exceedance probabilities for the period of record, 10/1/96 to 9/30/10. We used principal component analysis (PCA) to select a subset of the 98 highly intercorrelated basin characteristics in GAGES II (Appendix A) to be used as the independent variables in each regression equation. Variables with the highest eigenvector loadings within

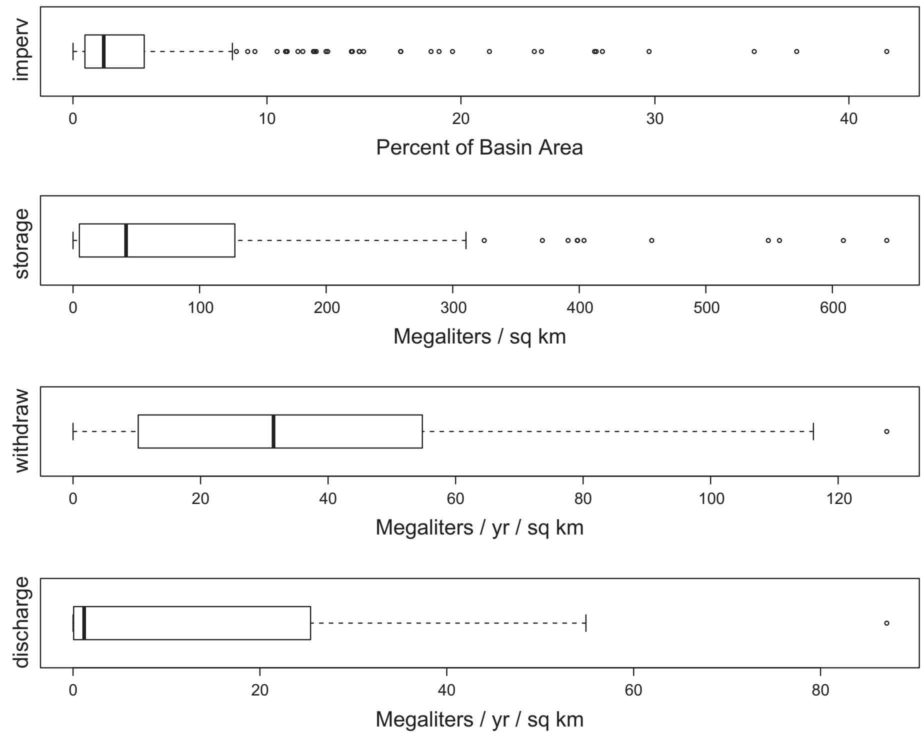

Fig. 4. Box plots showing the distributions of anthropogenic basin characteristics selected for the final model. Percent impervious cover (imperv) and storage volume (storage) include values for the 190 sites in New England. Water usage (withdraw) and water discharges (discharge) include values for the 42 sites in MA.

intercorrelated sets of variables in the first set of components, determined using a scree test, were maintained in the set of candidate variables. The variable reduction process also involved determination of which characteristics had the highest univariate correlation coefficients with the dependent variable, and maintenance of variance inflation factors (VIFs) below 5, along with the analysis of the PCA loadings.

The final set of variables were selected by determining the set with the minimum mean square error (MSE) using an "all subsets" regression algorithm, written in the statistical language R (R Development Core Team, 2006). Two outliers were removed when they exhibited very high Cook's D (Kutner, 2005). One of the two outliers removed had four times more storage than any other watershed due to the fact that almost one third of the basin was covered by the Quabbin Reservoir.

Use of a power law regression equation was consistent with previous streamflow regression models (Vogel and Kroll, 1992; Archfield et al., 2010; Vogel et al., 1999). Because the natural log is only defined for values above zero, any basin characteristics with negative values were shifted to be nonnegative by adding the minimum value plus a small increment to all values. For the set of eight natural variables selected in the final regression equations, only two required shifting (ASPECT_EASTNESS and LNG_GAGE) as the remaining natural variables were all above zero.

For the anthropogenic variables, however, modifications to the form of the terms in the regression equation were necessary. For the anthropogenic variables, a value of zero has a specific physical meaning in the model – the anthropogenic modification does not exist – and the model should indicate that the flow was unaltered by the characteristic. Three of the four anthropogenic variables in the final regression models had a significant proportion of zero values, and the remaining variable, impervious cover, had a minimum value of 0.0025%. Adding one before taking the natural log of these terms allowed a correct mapping between the value of the $\ln(X+1)$ term in the linear regression model and the value of X, so that when $X=0$, $\ln(X+1)=0$. This solution allows for the correct representation of the 'removal' of anthropogenic modifications from the regression equation by setting the value of the corresponding terms to zero. The approach also avoided the problem of ln(0) being undefined.

We did not take the same approach of adding one to the natural basin characteristics when modeling the relationship to flow. It makes physical sense for flow to approach zero as drainage area approaches zero. For the other natural characteristics the value of zero does not have any particular expected behavior on flow, and is outside the scope of the model because the values do not approach zero. The final regression equation is shown below in Eq. (2).

$$ln(Q_i) = \beta_0 + \beta_{n1} ln(X_{n1}) + \beta_{n2} ln(X_{n2}) + \cdots + \beta_{a1} ln(X_{a1}+1) + \beta_{a2} ln(X_{a2}+1) + \cdots \quad (2)$$

where Q_i, stream flow at exceedence probability i; X_{nj}, natural basin characteristic j; X_{ak}, anthropogenic basin characteristic k; β, model coefficients.The model translated into real space becomes the following:

$$Q_i = e^{\beta_0}(X_{n1})^{\beta_{n1}}(X_{n2})^{\beta_{n2}} \cdots (X_{a1}+1)^{\beta_{a1}}(X_{a2}+1)^{\beta_{a2}} \quad (3)$$

The first regression was run using data for the 190 stations in the six New England states against the full range of potential independent variables except water use and discharge because the georeferenced point data for water withdrawals and discharges were only available for Massachusetts. Therefore, a second regression was performed using the output of the first stage and the point withdrawals and discharges summed for the 42 basins that were completely within the state of Massachusetts. Consequently, the second stage regression focused on estimating the additional variance explained by withdrawals and discharges conditional on the variance accounted for by the other variables. The resulting set of equations for estimating the flow at each exceedence probability is shown below in Eq. (5).

First stage:

$$\ln(Q_i) = \beta_0 + \beta_n \ln(X_n) + \cdots + \beta_a \ln(X_a + 1) + \cdots \quad (4)$$

Second stage:

$$\ln(Q_i) = \gamma_0 + \gamma_1[\beta_0 + \beta_n \ln(X_n) + \cdots + \beta_a \ln(X_a + 1) + \cdots] + \gamma_2 \times \ln(X_{wth} + 1) + \gamma_3 \ln(X_{dis} + 1) \quad (5)$$

where Q_i, stream flow at exceedence probability i; X_n, natural basin characteristics; X_a, anthropogenic basin characteristics; β, model coefficients for the first regression; γ, model coefficients for the second regression; X_{wth}, average water withdrawal/km^2; X_{dis}, average water discharge/km^2.

3. Results

Eight natural basin variables and four alteration variables were selected for the final set of regression equations (Table 1). Not all the variables were selected in the final regression equations to predict each exceedence probability. The set of variables that minimized MSE were selected for each exceedence probability. The minimization of MSE is equivalent to maximization of the adjusted R^2 value since both are normalized by the difference between the number of observations and the number of parameters in each model.

The final regression equations are shown in Table 2. Each column represents the regression equation for estimating streamflow at a particular exceedence probability. Each row shows the values of the coefficients for the same variable in the different regression equations. Most coefficients tend to increase or decrease gradually across exceedance probabilities.

Model performance statistics are presented in Table 4. The Coefficient of Determination (R^2) ranges from 0.852 to 0.983 across exceedance probilities. The MSE is in units of the natural log of flow and is thus difficult to interpret in its absolute scale. Consequently, the number that is commonly reported for regional regression log-transformed models is the standard error (SE), which is shown in the table and reported as the coefficient of variation translated from log space to real space. The higher SE for lower flows (higher exceedance probabilities) is consistent with previous models estimating low flows (Maidment, 1993; Kroll et al., 2004). The *p*-values for each coefficient of each regression equation are shown in Table 3. Though some of the *p*-values are above 0.10 for some exceedence probabilities, each variable has p-values at multiple exceedence probabilities that are below 0.05.

A 'leave one out' cross validation was conducted by leaving out each of the 42 study sites in Massachusetts from both stages of regression and then estimating a value for this site from the resulting model. The Nash-Sutcliffe efficiency (NS) (Nash and Sutcliffe, 1970; Moriasi et al., 2007), shown in Table 5, was then calculated to evaluate the predictive ability of the model. The high NS values along with the fact that the MSE results for the cross validation are very close to the MSE values for the regression support the predictive ability of the model.

Boxplots in Figs. 5–8 display the estimated percent alteration of streamflow at each exceedance probability for the 42 stations in Massachusetts. In these figures, the estimated independent effect of each anthropogenic variable is viewed by setting one alteration variable to zero at a time and computing the percent alteration of streamflow at each exceedance probability.

Fig. 5 shows that on average dam storage in a basin decreases high flows (i.e., negative percent alteration at small exceedance probabilities), consistent with standard expectations for dammed flows and, in many cases, the purpose of the dam. The same figure shows storage increasing median flows and having no significant effect on low flows.

Fig. 6 depicts the effects of impervious cover on streamflow and reveals on average a decrease in high flows and an increase in low flows. These results contrast with the expectations often cited of increased impervious cover lowering base flows (Jacobson, 2011). However more and more studies now provide evidence of the opposite effect, consistent with our results, of impervious cover increasing low flows (Price, 2011). The decrease in high flows (peak flows are not represented in the model since the lowest exceedence probability modeled is 0.05) could be occurring due to stormwater storage systems which occur more frequently in areas with higher impervious cover.

Figs. 7 and 8 depict the estimates of the alterations caused by water withdrawals and discharges, respectively. Both figures show effects on streamflow in the expected manner – decreased by withdrawals or increased by discharges, with the strength of the effect greatest at lower flows. Note that the variability among basins increases with the mean percent alteration, such that the range for the 0.95 exceedance probability varies from just above zero to over 100% change in the natural streamflow.

Fig. 9 shows the percent alteration in streamflow when all of the anthropogenic modifications are considered together. For these 42 MA basins, the model shows that the cumulative effect of the anthropogenic basin modifications decreases high flows and increases low flows, and thus decrease the variability in daily streamflows. This will be discussed in more detail below.

Table 1
Variable names and descriptions for the set of basin characteristcs selected as independent variables in the final regression equations. Data for all except two (*) are from the USGS GAGES II database. Data for withdrawals and discharges are from the MA SYE wateruse database. Min, max and mean values shown of the 190 sites in 6 NE states except for MA only data (*).

Variable name	Description	Min	Mean	Max
DRAIN_SQKM	Watershed **drainage area**, sq km, as delineated in our basin boundary	9.4	1142.2	25049.5
LNG_GAGE	Longitude at gage, decimal degrees	−73.5	−71.7	−67.2
BAS_COMPACTNESS	Watershed compactness ratio, = area/perimeter^2 * 100; higher number = more compact shape	0.7	1.6	3.6
T_MAX_BASIN	Watershed average of maximum monthly air temperature (°C) from 800 m PRISM, derived from 30 years of record (1971–2000)	7.0	12.9	16.0
RH_SITE	Site average relative humidity (percent), from 2 km PRISM, derived from 30 years of record (1961–1990)	65	67.7	74
'DEC_PPT7100_CM	Mean December precip (cm) for the watershed, from 800 m PRISM data. 30 years period of record 1971–2000	6.2	9.9	15.6
SANDAVE	Average value of **sand** content (%) (STATSGO, 1997)	18.3	45.8	76.0
ASPECT_EASTNESS	Aspect "eastness". Ranges from −1 to 1. Value of 1 means watershed is facing/draining due east, value of −1 means watershed is facing/draining due west	−0.99	0.32	0.99
IMPNLCD06	Watershed percent impervious surfaces from 30-m resolution NLCD06 data (2006)	0.0026	4.6	51.3
STOR_NID_2009	Dam storage in watershed ("NID_STORAGE"); megaliters total storage per sq km (1 megaliters = 1,000,000 liters = 1,000 cubic meters) (2009)	0.0	136.3	4565.1
Withdrawals*	Water withdrawals in the basin; megaliters per year per sq km. (2000–2004)	0	38.5	127.6
Discharges*	Water discharges in the basin; megaliters per year per sq km. (2000–2004)	0	13.5	87.1

Table 2
Model coefficients for each regression equation estimating flow at the given exceedence probability.

Term	Exceedence probability										
	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
Intercept	10.1	2.5	−0.18	0.23	7.20	14.5	22.4	34.3	44.1	46.2	45.5
area	0.97	1.00	1.02	1.03	1.04	1.04	1.05	1.07	1.14	1.22	1.31
lng_gage	−2.24	−1.49	−1.78	−2.31	−3.98	−5.90	−7.79	−10.2	−12.6	−13.3	−11.9
compact	–	–	0.04	0.05	0.06	0.07	0.09	0.09	0.16	0.29	0.39
tmax_mean	−0.67	−0.48	−0.34	−0.34	−0.46	−0.57	−0.72	−1.21	−1.68	−2.10	−2.67
humidity	–	0.55	1.04	1.22	1.22	1.60	1.75	1.66	1.86	1.71	0
ppt_dec	0.70	0.81	0.98	1.16	1.30	1.31	1.36	1.44	1.63	2.24	2.68
sand	–	0.07	0.13	0.17	0.20	0.23	0.26	0.28	0.29	–	–
eastness	0.03	0.01	0	−0.02	−0.02	−0.03	−0.03	−0.04	−0.07	−0.12	−0.15
imperv	−0.02	−0.03	−0.03	−0.02	–	–	–	0.06	0.11	0.15	0.19
storage	−0.03	−0.02	−0.01	–	–	0.01	0.01	0.01	–	–	–
withdraw	–	–	–	–	−0.01	−0.02	−0.02	−0.04	−0.04	−0.06	−0.08
discharge	–	–	0.01	0.02	0.03	0.03	0.04	0.04	0.05	0.11	0.17

Table 3
P values for each coefficient of each regression equation estimating flow at the given exceedence probability.

Term	Exceedence probability										
	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
area	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
lng_gage	0.061		0.153	0.036	0.001	<0.001	<0.001	<0.001	<0.001	0.001	0.015
compact			0.143	0.069	0.105	0.076	0.063	0.117	0.035	0.022	0.027
tmax_mean	<0.001	<0.001	0.002	0.002	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
humidity			0.013	0.005	0.015	0.006	0.009	0.041	0.073	0.297	
ppt_dec	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
sand		0.164	0.006	0.001	0.001	0.001	0.002	0.006	0.049		
eastness	0.004	0.216	0.302	0.014	0.005	0.010	0.004	0.004	0.001	<0.001	0.001
imperv	0.193	0.028	0.014	0.133				0.019	0.001	0.004	0.010
storage	0.002	0.001	0.164			0.168	0.127	0.266			
withdraw						0.208	0.080	0.004	0.003	0.012	0.032
discharge			0.307	0.155	0.148	0.040	0.011	0.010	0.005	0.006	0.010

Table 4
Coefficient of determination (R^2), Mean Square Error (MSE) and real space Standard Error (SE) for the final regression equations for each exceedence probability.

Exceedence probability	R^2	MSE	SE (%)
0.05	0.983	0.017	13
0.10	0.987	0.014	12
0.20	0.992	0.010	10
0.30	0.992	0.010	10
0.40	0.991	0.011	11
0.50	0.990	0.013	11
0.60	0.989	0.015	12
0.70	0.988	0.017	13
0.80	0.981	0.029	17
0.90	0.926	0.141	39
0.95	0.852	0.375	67

Table 5
Nash Sutcliffe Efficiency (NS), Mean Square Error (MSE) and real space Standard Error (SE) for the final cross validation results for each exceedence probability.

Exceedence probability	NS	MSE	SE (%)
0.05	0.980	0.020	14
0.10	0.984	0.016	13
0.20	0.988	0.013	11
0.30	0.990	0.011	11
0.40	0.988	0.014	12
0.50	0.986	0.016	13
0.60	0.985	0.017	13
0.70	0.984	0.020	14
0.80	0.975	0.034	18
0.90	0.901	0.169	43
0.95	0.806	0.445	75

Fig. 10 depicts the ecochange metric for each of the 42 stations in MA and reveals the east-west development gradient across the state. Within this coarse-scale gradient in ecochange there is local variation driven by local differences in each basin.

4. Discussion

We conducted an empirical analysis of the relationship between anthropogenic modifications in drainage basins and the streamflow exiting these basins. A broad range of anthropogenic characteristics were considered given newly available GIS data allowing new and more accurate representations of anthropogenic modifications than in the past. These initial results provide insight about which alterations show clear signals to streamflow alteration and their relative impacts. In addition, the model can be used to estimate hydrologic alteration at ungaged sites.

There is little question that water withdrawals and discharges will have an impact on the amount of water in a basin. In fact, many models directly calculate water availability by subtracting withdrawals and/or adding discharges (Weiskel et al., 2010). It is reassuring that the empirical signal from the regression model confirms these expected results, and we see in Figs. 7 and 8 that these direct inputs and outputs of water have a larger percent impact on the lower flows. A benefit of modeling the water withdrawals and discharges in the same way as the other modifications is the ability to compare their relative impacts on streamflow.

Dams are built to provide hydroelectricity, reliable water supply, flood risk reduction, recreational uses and for many other reasons. Most of these dams are specifically meant to modify the magnitude and timing of natural flow in a river in order to control water storage in a reservoir. Studies that have focused on isolating

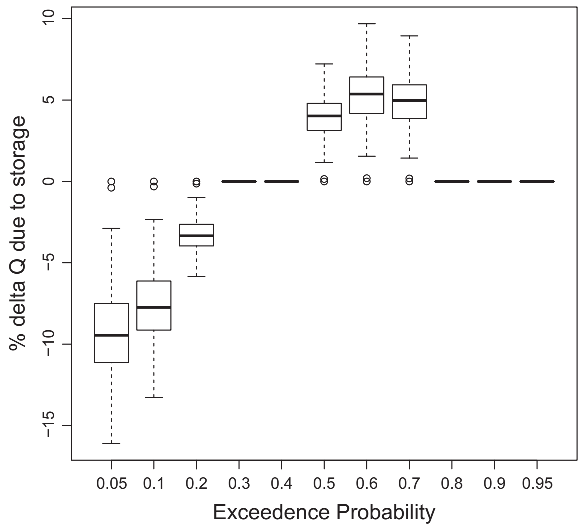

Fig. 5. Model-estimated percent alteration of streamflow due to reservoir storage in each of the 42 Massachusetts basins.

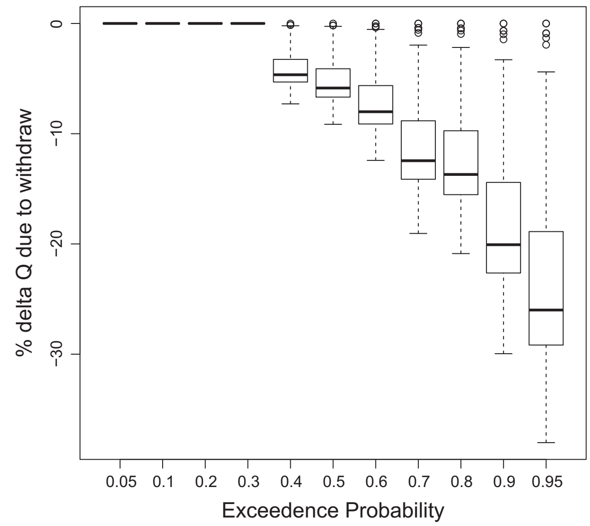

Fig. 7. Model-estimated percent alteration of streamflow due to water withdrawals in each of the 42 Massachusetts basins.

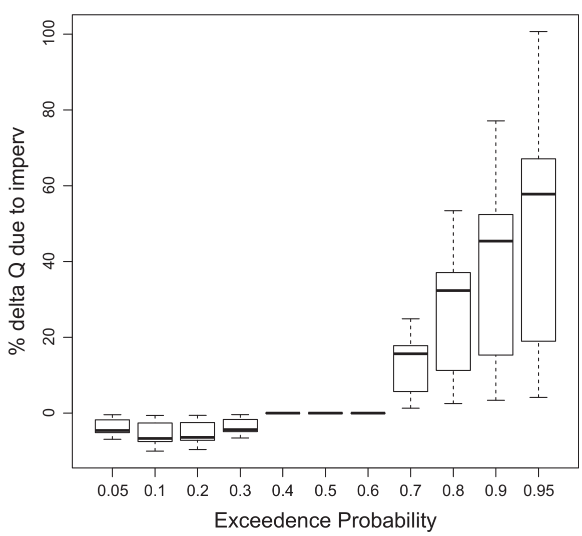

Fig. 6. Model-estimated percent alteration of streamflow due to impervious cover in each of the 42 Massachusetts basins.

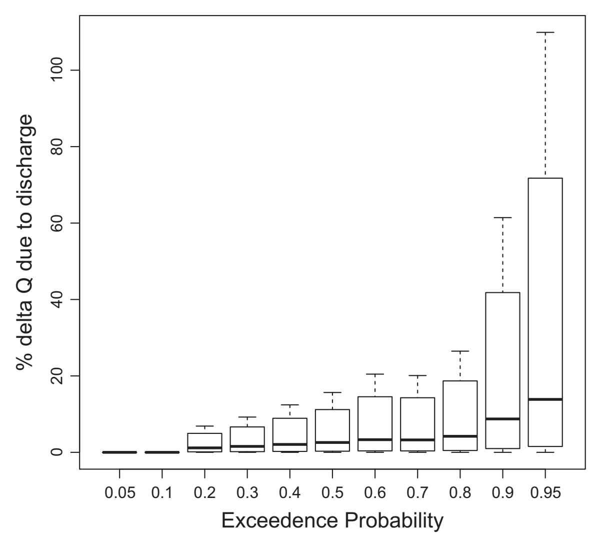

Fig. 8. Model-estimated percent alteration of streamflow due to water discharges in each of the 42 Massachusetts basins.

the impact of dams on streamflow have confirmed that dams typically decrease peak flows, increase minimum flows and decrease flow variability. These results are not surprising given the intent and operation of dams (Poff et al., 2006; Magilligan and Nislow, 2005; Fitzhugh and Vogel, 2011). Our results are consistent with this expectation and reveal a decrease in the highest streamflows (i.e., 0.05–0.20 exceedance probabilities) with dam storage (Fig. 5). The increase in streamflow at intermediate flows (i.e., exceedence probabilities 0.5–0.7) suggests that dam storage effects are not limited to just high and low flows. The signal strength at low flows was not strong enough to discern an effect.

The impact of impervious surface and other land use changes related to urbanization, for which impervious surface is often a surrogate, on streamflow has been a topic researched over many decades. Jacobson (2011) and Price (2011) provide reviews of impervious surface studies and propose that the general consensus from the first studies in the 1960s and 70s was that an increase in impervious cover resulted in an increase in high flows and a decrease in low flows due to less infiltration and recharge. Studies conducted in the more recent decades now provide evidence of complex interactions that produce various sets of possible results. Impervious surface may decrease recharge, but there is evidence of possible simultaneous decreases in evapotranspiration (ET) actually resulting in higher low flows (Jacobson, 2011; Price, 2011; Schueler et al., 2009). Others point to the possibility of areas of higher impervious cover having more leaks in water distribution pipes or heavily watered lawns which could contribute to higher low flows (Poff et al., 2006). Our results suggest an increase in low flows with a progressively higher percent change as the flows get lower (Fig. 6). These results are consistent with scenarios involving lower ET and possibly also explained by leaky pipes and heavily watered lawns.

4.1. Scope and limitations

Our modeling approach offers a couple of major advantages over other approaches. One of the most compelling features of our approach is that our estimate of natural streamflow is an

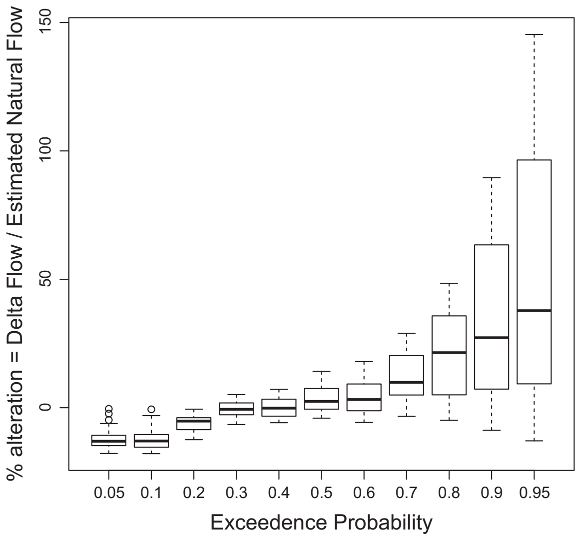

Fig. 9. Model-estimated percent alteration from the cumulative effects of all four basin alterations in the final regression equations: impervious cover, storage volume, withdrawals and discharges.

attempt to represent 'completely natural' conditions, and not just 'least altered" conditions. Studies using reference basins to represent natural conditions all need to compromise 'natural' in some way given the lack of available truly pristine sites. This has often resulted in the use of 'least altered' site as a surrogate for natural because truly natural no longer exists. For example, the MA Sustainable Yield Estimator (SYE) regressions (Archfield et al., 2010) were developed using a set of 'least altered' reference sites. A sample of half of these reference basins revealed sites with up to 27% impervious surface, up to 64.5 megaliters/yr/km^2 water withdrawls, up to 45.4 megaliters/yr/km^2 water discharges, and 151 megaliters/km^2 in storage volume. Similarly, the basins selected by Carlisle et al. (2010) as reference sites averaged significantly less storage and impervious cover than their non-reference sites but included sites with up to 75% of the top water withdrawals per area and the median water use was still 50% of the median non-reference water use.

Second, our model provides a means to estimate the degree of hydro-alteration at any ungaged site. The reference site approach taken by Carlisle et al. (2010) requires at least a few years of flow record to have the observed/altered flow with which to compare the natural flow estimates. The SYE estimates altered flow, but only altered by water withdrawals and discharges. Our analysis, on the other hand, included a wide range of anthropogenic characteristics and the final regression included dam storage volume and impervious surface in addition to water withdrawals and discharges.

Our results are not without important limitations. First, our results are limited by the fact that only the State of MA within the six New England states considered had water use data available with georeferenced withdrawals and discharges. Having these data in additional states would strengthen the regression results. Of course, our approach can be applied without accurate data on withdrawals and discharges, but given that these were two of the four retained anthropogenic variables in the final regression models, the reliability of the results may be suspect without these data or with coarse and inaccurate estimates. It is also worth noting that in other parts of the country county level data may be a more useful than in Massachusetts where counties are large and heterogeneous.

Second, the process of variable selection we used is complex and one among many alternative approaches that could be used for variable selection with unknown consequences. Additional work is warranted to investigate the model sensitivity to inclusion of different sets of predictors. We attempted to find a "good" and parsimonious set of predictors but cannot guarantee that we found the "best" set.

Third, the effects of highly correlated interdependent variables are necessarily difficult to distinguish. Not surprisingly, anthropogenic modifications to basins are typically highly confounded; the more developed a basin, typically the more impervious surface, number of dams, and magnitude of water withdrawals and discharges. It should be noted that this is a fundamental constraint facing all approaches. We attempted to minimize the multicollinearity among predictors by carefully selecting a largely uncorrelated set of

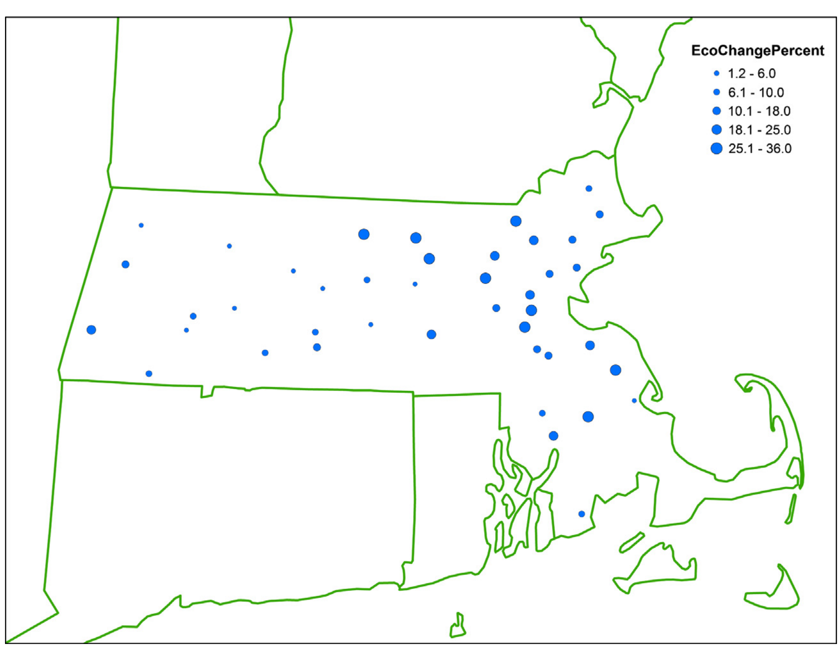

Fig. 10. Locations of 42 study stream gages in MA with the model estimated ecochange percent indicated by size of the point on the map.

variables. Despite this criterion, the final set of variables are not truly independent. Therefore, we are not able to truly distinguish the independent effect of each anthropogenic variable and the results depicted in Figs. 4–7 must be interpreted in this context.

Fourth, low flows (0.9 and 0.95 exceedence probabilities) were the most difficult to predict accurately (Table 3), although our level of errors are comparable to other studies (Vogel and Kroll, 1992; Maidment, 1993). Price (2011) suggested that the influence of subsurface topography should be considered in order to properly model baseflow, and others have proposed measurements to estimate a baseflow regression constant that could be used to improve estimation accuracy of low flows (Vogel and Kroll, 1992). No doubt, our results would improve with the addition of information on subsurface topography, but this is not generally available.

Our regression models represent an average across the available input data set, and thus the estimates produced are a representation of average conditions; the actual individual circumstances in each basin will result in deviations from our predictions. Applying the model to locations where conditions are known to deviate should be avoided, such as largely groundwater dominated watersheds in southeastern MA.

Lastly, our aim was to estimate streamflow in the absence of any anthropogenic basin modifications. Unfortunately, there are no sites available with 0% impervious surface, so for this variable we had to extrapolate to zero in order to implement our estimate of 'pristine' natural conditions. Fortunately, the minimum percent impervious surface was very low at 0.025%, so the extrapolation was relatively minor. Zero values were present in the data for the other three anthropogenic variables: dam storage, water withdrawals, and discharges. Of course, having a few sites with the complete absence of each anthropogenic characteristic does not really provide a means to validate our natural streamflow predictions across the remaining modified sites. Indeed, this is the biggest limitation of our approach, but one that fundamentally constrains any approach. The only way to validate our results is to have sites with streamflow data prior to and after anthropogenic modifications, which unfortunately are rare or nonexistent.

5. Conclusions

We presented a method to evaluate the degree of hydrologic alteration for a basin given the availability of data to characterize a limited suite of anthropogenic basin characteristics. Impervious cover data are readily available and improving in accuracy given the use of satellite images and technology to process these images into useful formats for analysis. Dam data that are maintained nationally by the National Inventory of Dams includes point locations and storage volumes for the entire country. The critical data to implement our approach are accurate point data for water withdrawals and discharges such as we were able to obtain for the state of Massachusetts. Our results provide a first look at empirical evidence for the effects of basin alterations on streamflow that is typically only theorized.

The regional regressions we developed for each exceedance probability improved the percent error compared to existing regional regression for similar statistics that used only least altered sites. We have demonstrated that the ecochange metric is a convenient metric for summarizing the cumulative impact of anthropogenic basin modifications using the flow duration curve without the complexity and uncertainty involved with estimating a daily time series of flow at each site.

We propose that our regression approach could be an effective mechanism for estimating the degree of total streamflow alteration in Massachusetts basins. Our approach could also be applied to other locations if the necessary water withdrawal and discharge data becomes available. In addition, it can be applied to any location within the stream network, providing a tool for estimation of alteration with wide applicability. Note, it is critical to analyze streamflow alterations having considered these direct inputs and outputs; limiting the analysis to basins unaffected by withdrawals or discharges is too restrictive and biased towards the types of basins that have not been developed.

Models relating basin alterations to flow alteration are becoming valuable tools in conservation efforts to protect and sustain our water resources. In addition to creating a tool to estimate streamflow alteration at ungaged sites, our results have provided a step toward an increased understanding of how humans alterations to basins are affecting streamflow.

Acknowledgements

We thank Stacey Archfield and Sara Levin from the USGS for their support in procuring MA water use data; James Falcone from the USGS for his support for further information and guidance on the GAGES II database; and Richard Vogel for his review and insights.

Funding for this work was provided by the Wetland Program Development Grant Program administered by the U.S. Environmental Protection Agency under section 104 (b)(3) of the U.S. Clean Water Act. This project is part of an ongoing partnership with the Massachusetts Department of Environmental Protection and Office of Coastal Zone Management to develop and implement a comprehensive wetland assessment and monitoring program for Massachusetts. C. Brown's effort was funded by the National Science Foundation Grant Number CBET-1054762.

Appendix A

The set of 98 variables from USGS GAGES II database considered as independent variables.

Variable name	Description
DRAIN_SQKM	Watershed drainage area, sq km
BAS_COMPACTNESS	Watershed compactness ratio, = area/perimeter^2 ∗ 100
LAT_CENT	Latitude of centroid location of basin, decimal degrees
LONG_CENT	Longitude of centroid location of basin, decimal degrees
PPTAVG_BASIN	Mean annual precip (cm) for the watershed
PPTAVG_SITE	Mean annual precip (cm) at the gage location
T_AVG_BASIN	Average annual air temperature for the watershed (°C)
T_AVG_SITE	Average annual air temperature at the gage location (°C)
T_MAX_BASIN	Watershed average of maximum monthly air temperature (°C)
T_MAXSTD_BASIN	Standard deviation of maximum monthly air temperature (°C)
T_MAX_SITE	Gage location maximum monthly air temperature (°C)

(continued on next page)

Appendix A (*continued*)

Variable name	Description
T_MIN_BASIN	Watershed average of minimum monthly air temperature (°C)
T_MINSTD_BASIN	Standard deviation of minimum monthly air temperature (°C)
T_MIN_SITE	Gage location minimum monthly air temperature (°C)
RH_BASIN	Watershed average relative humidity (%)
RH_SITE	Site average relative humidity (%)
FST32F_BASIN	Watershed average of mean day of the year of first freeze
LST32F_BASIN	Watershed average of mean day of the year of last freeze
FST32SITE	Site average of mean day of the year of first freeze
LST32SITE	Site average of mean day of the year of last freeze
WD_BASIN	Watershed average of annual number of days of measurable precipitation
WD_SITE	Site average of annual number of days of measurable precipitation
WDMAX_BASIN	Watershed average of monthly maximum number of days of measurable precipitation
WDMIN_BASIN	Watershed average of monthly minimum number of days of measurable precipitation
WDMAX_SITE	Site average of monthly maximum number of days of measurable precipitation
WDMIN_SITE	Site average of monthly minimum number of days of measurable precipitation
PET	Mean-annual potential evapotranspiration (PET)
SNOW_PCT_PRECIP	Snow percent of total precipitation estimate, mean for period 1901–2000
PRECIP_SEAS_IND	Precipitation seasonality index
JAN_PPT7100_CM	Mean January precip (cm) for the watershed
FEB_PPT7100_CM	Mean February precip (cm) for the watershed
MAR_PPT7100_CM	Mean March precip (cm) for the watershed
APR_PPT7100_CM	Mean April precip (cm) for the watershed
MAY_PPT7100_CM	Mean May precip (cm) for the watershed
JUN_PPT7100_CM	Mean June precip (cm) for the watershed
JUL_PPT7100_CM	Mean July precip (cm) for the watershed
AUG_PPT7100_CM	Mean August precip (cm) for the watershed
SEP_PPT7100_CM	Mean September precip (cm) for the watershed
OCT_PPT7100_CM	Mean October precip (cm) for the watershed
NOV_PPT7100_CM	Mean November precip (cm) for

Appendix A (*continued*)

Variable name	Description
	the watershed
DEC_PPT7100_CM	Mean December precip (cm) for the watershed
JAN_TMP7100_DEGC	Average January air temperature for the watershed (°C)
FEB_TMP7100_DEGC	Average February air temperature for the watershed (°C)
MAR_TMP7100_DEGC	Average March air temperature for the watershed (°C)
APR_TMP7100_DEGC	Average April air temperature for the watershed (°C)
MAY_TMP7100_DEGC	Average May air temperature for the watershed (°C)
JUN_TMP7100_DEGC	Average June air temperature for the watershed (°C)
JUL_TMP7100_DEGC	Average July air temperature for the watershed (°C)
AUG_TMP7100_DEGC	Average August air temperature for the watershed (°C)
SEP_TMP7100_DEGC	Average September air temperature for the watershed (°C)
OCT_TMP7100_DEGC	Average October air temperature for the watershed (°C)
NOV_TMP7100_DEGC	Average November air temperature for the watershed (°C)
DEC_TMP7100_DEGC	Average December air temperature for the watershed (°C)
CLAYAVE	Average value of clay content (%)
SILTAVE	Average value of silt content (%)
SANDAVE	Average value of sand content (%)
ELEV_MEAN_M_BASIN	Mean watershed elevation (m)
ELEV_MAX_M_BASIN	Maximum watershed elevation (m)
ELEV_MIN_M_BASIN	Minimum watershed elevation (m)
ELEV_MEDIAN_M_BASIN	Median watershed elevation (m)
ELEV_STD_M_BASIN	Standard deviation of elevation (m) across the watershed
ELEV_SITE_M	Elevation at gage location (m)
RRMEAN	Dimensionless elevation – relief ratio, calculated as (ELEV_MEAN − ELEV_MIN)/(ELEV_MAX − ELEV_MIN)
RRMEDIAN	Dimensionless elevation – relief ratio, calculated as (ELEV_MEDIAN − ELEV_MIN)/(ELEV_MAX − ELEV_MIN)
SLOPE_PCT	Mean watershed slope, %
ASPECT_DEGREES	Mean watershed aspect, ° (degrees of the compass, 0–360)
ASPECT_NORTHNESS	Aspect "eastness". Ranges from −1 to 1
ASPECT_EASTNESS	Aspect "northness". Ranges from −1 to 1
STREAMS_KM_SQ_KM	Stream density, km of streams per watershed sq km
STRAHLER_MAX	Maximum Strahler stream order in watershed

Appendix A (*continued*)

Variable name	Description
HIRES_LENTIC_PCT	Percent of watershed surface area covered by "Lakes/ Ponds" + "Reservoirs"
PCT_1ST_ORDER	Percent of stream lengths in the watershed which are first-order streams
PCT_2ND_ORDER	Percent of stream lengths in the watershed which are second-order streams
HIRES_LENTIC_NUM	Number of Lakes/ Ponds + Reservoir water bodies
HIRES_LENTIC_DENS	Density (#/sq km) of Lakes/ Ponds + Reservoir water bodies
HIRES_LENTIC_MEANSIZ	Mean size (ha) of Lakes/ Ponds + Reservoir water bodies
FORESTNLCD06	Watershed percent "forest"
PLANTNLCD06	Watershed percent "planted/ cultivated" (agriculture)
WATERNLCD06	Watershed percent Open Water
DECIDNLCD06	Watershed percent Deciduous Forest
EVERGRNLCD06	Watershed percent Evergreen Forest
MIXEDFORNLCD06	Watershed percent Mixed Forest
SHRUBNLCD06	Watershed percent Shrubland
GRASSNLCD06	Watershed percent Herbaceous (grassland)
PASTURENLCD06	Watershed percent Pasture/Hay
CROPSNLCD06	Watershed percent Cultivated Crops
WOODYWETNLCD06	Watershed percent Woody Wetlands
EMERGWETNLCD06	Watershed percent Emergent Herbaceous Wetlands
PDEN_2000_BLOCK	Population density in the watershed, persons per sq km
PDEN_DAY_LANDSCAN_ 2007	Population density in the watershed during the day, persons per sq km
PDEN_NIGHT_LANDSCAN_ 2007	Population density in the watershed at night, persons per sq km
ROADS_KM_SQ_KM	Road density, km of roads per watershed sq km
RD_STR_INTERS	Number of road/stream intersections, per km of total basin stream length
IMPNLCD06	Watershed percent impervious surfaces
FRAGUN_BASIN	Fragmentation Index of "undeveloped" land in the watershed.
MINING92_PCT	Percent 1quarries-strip mines–gravel pits land cover in watershed,
PCT_IRRIG_AG	Percent of watershed in irrigated agriculture

References

Archfield, S.A., Vogel, R.M., Steeves, P.A., Brandt, S.L., Weiskel, P.K., Garabedian, S.P., 2010. The Massachusetts Sustainable-Yield Estimator: A Decision-support Tool to Assess Water Availability at Ungaged Stream Locations in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2009–5227, 41p.

Armstrong, D.S., Parker, G.W., Richards, T.A., 2004, Evaluation of Streamflow Requirements for Habitat Protection by Comparison to Streamflow Characteristics at Index Streamflow-Gaging Stations in Southern New England: U.S. Geological Survey Water-Resources Investigations Report 03-4332, 108p.

Armstrong, D.S., Parker, G.W., Richards, T.A., 2008, Characteristics and Classification of Least Altered Streamflows in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2007-5291, 113p.

Bulygina, N., McIntyre, N., Wheater, H., 2011. Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes. Water Resour. Res. 47, W02503.

Carlisle, D.M., Falcone, J., Wolock, D.M., Meador, M.R., Norris, R.H., 2010. Predicting the natural flow regime: models for assessing hydrological alteration in streams. River Research and Applications 26, 118–136.

Chiew, F., McMahon, T., 2002. Modelling the impacts of climate change on Australian streamflow. Hydrol. Process. 16, 1235–1245.

Choi, W., Deal, B.M., 2008. Assessing hydrological impact of potential land use change through hydrological and land use change modeling for the Kishwaukee River basin (USA). J. Environ. Manage. 88, 1119–1130.

Coles, J.F., Cuffney, T.F., McMahon, G., Rosiu, C.J., 2010. Judging a brook by its cover: the relation between ecological condition of a stream and urban land cover in New England. Northeast. Nat. 17, 29–48.

Dodds, W., Oakes, R., 2004. A technique for establishing reference nutrient concentrations across watersheds affected by humans. Limnol. Oceanogr. Methods 2, 333–341.

Falcone, J.A., Carlisle, D.M., Wolock, D.M., Meador, M.R., 2010. GAGES: a stream gage database for evaluating natural and altered flow conditions in the conterminous United States. Ecology 91 (2), 621, Data Paper in Ecological Archives E091-045-D1. <http://esapubs.org/Archive/ecol/E091/045/metadata.htm>.

FitzHugh, T.W., Vogel, R.M., 2011. The impact of dams on flood flows in the United States. River Res. Appl. 27, 1192–1215.

Foley, J. et al., 2005. Global consequences of land use. Science 309, 570–574.

Gao, Y., Vogel, R.M., Kroll, C.N., Poff, N.L., Olden, J.D., 2009. Development of representative indicators of hydrologic alteration. J. Hydrol. 374, 136–147.

Homa, E., Vogel, R.M., Smith, M.P., Apse, C.D., Huber-Lee, A., 2005. An Optimization Approach for Balancing Human and Ecological Flow Needs, EWRI 2005 World Water and Environmental Resources Congress. Anchorage, Alaska.

Hutson, S.S., compiler, 2007. Guidelines for Preparation of State Water-use Estimates for 2005: U.S. Geological Survey Techniques and Methods Book 4, 36p (Chapter E1). <http://pubs.usgs.gov/tm/2007/tm4e1>.

Jackson, R., Carpenter, S., Dahm, C., McKnight, D., Naiman, R., Postel, S., Running, S., 2001. Water in a changing world. Ecol. Appl. 11, 1027–1045.

Jacobson, C.R., 2011. Identification and quantification of the hydrological impacts of imperviousness in urban catchments: a review. J. Environ. Manage. 92, 1438–1448.

Jager, H.I., Smith, B.T., 2008. Sustainable reservoir operation: can we generate hydropower and preserve ecosystem values? River Res. Appl. 24, 340–352.

Jennings, D.B., Jarnagin, S.T., 2002. Changes in anthropogenic impervious surfaces, precipitation and daily streamflow discharge: a historical perspective in a mid-atlantic subwatershed. Landscape Ecol. 17, 471–489.

Karvonen, T., Koivusalo, H., Jauhiainen, M., Palko, J., Weppling, K., 1999. A hydrological model for predicting runoff from different land use areas. J. Hydrol. 217, 253–265.

Konrad, C.P., Booth, D.B., 2002. Hydrologic Trends Resulting from Urban Development in Western Washington Streams U.S. Geological Survey Water-Resources Investigation, Report 02e4040.

Kroll, C., Luz, J., Allen, B., Vogel, R., 2004. Developing a watershed characteristics database to improve low streamflow prediction. J. Hydrol. Eng. 9, 116–125.

Kutner, M.H., 2005. Applied Linear Statistical Models. McGraw-Hill Irwin, Boston, 1396pp.

Labadie, J., 2004. Optimal operation of multireservoir systems: state-of-the-art review. J Water Resour Plann Manage-ASCE 130, 93–111.

Magilligan, F.J., Nislow, K.H., 2005. Changes in hydrologic regime by dams. Geomorphology 71, 61–78.

Maidment, D.R., 1993. Handbook of Hydrology. McGraw-Hill, New York.

McIntyre, N., Marshall, M., 2010. Identification of rural land management signals in runoff response RID G-5410-2011. Hydrol. Process. 24, 3521–3534.

Milly, P., Dunne, K., Vecchia, A., 2005. Global pattern of trends in streamflow and water availability in a changing climate. Nature 438, 347–350.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3), 885–900.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—a discussion of principles. J. Hydrol. 10, 282–290.

O'Connell, E., Ewen, J., O'Donnell, G., Quinn, P., 2007. Is there a link between agricultural land-use management and flooding? Hydrol. Earth Syst. Sci. 11, 96–107.
Olden, J., Poff, N., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes RID A-8535-2010 RID C-1239-2009. River Res. Appl. 19, 101–121.
Pechlivanidis, I.G., Jackson, B.M., Mcintyre, N.R., Wheater, H.S., 2011. Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. Global Nest J. 13, 193–214.
Poff, N.L., Zimmerman, J.K.H., 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows RID C-1239-2009. Freshwat. Biol. 55, 194–205.
Poff, N., Allan, J., Bain, M., Karr, J., Prestegaard, K., Richter, B., Sparks, R., Stromberg, J., 1997. The natural flow regime RID C-1239-2009. Bioscience 47, 769–784.
Poff, N.L., Bledsoe, B.P., Cuhaciyan, C.O., 2006. Hydrologic variation with land use across the contiguous United States: geomorphic and ecological consequences for stream ecosystems. Geomorphology 79, 264–285.
Poff, N.L. et al., 2010. The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. Freshwat. Biol. 55, 147–170.
Postel, S., Richter, B., 2003. Rivers for Life – Managing Water for People and Nature. Island Press, Washington, DC, 253pp.
Price, K., 2011. Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: a review. Prog. Phys. Geogr. 35, 465–492.
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. <http://www.R-project.org>.
Richter, B., Baumgartner, J., Powell, J., Braun, D., 1996. A method for assessing hydrologic alteration within ecosystems. Conserv. Biol. 10, 1163–1174.
Richter, B., Baumgartner, J., Wigington, R., Braun, D., 1997. How much water does a river need? Freshwat. Biol. 37, 231–249.
Risbey, J., Entekhabi, D., 1996. Observed Sacramento Basin streamflow response to precipitation and temperature changes and its relevance to climate impact studies. J. Hydrol. 184, 209–223.
Roy, A.H., Shuster, W.D., 2009. Assessing impervious surface connectivity and applications for watershed management. J. Am. Water Resour. Assoc. 45, 198–209.
Sanborn, S., Bledsoe, B., 2006. Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon. J. Hydrol. 325, 241–261.
Schueler, T.R., Fraley-McNeal, L., Cappiella, K., 2009. Is impervious cover still important? Review of recent research. J. Hydrol. Eng. 14, 309–315.
Soranno, P.A., Cheruvelil, K.S., Stevenson, R.J., Rollins, S.L., Holden, S.W., Heaton, S., Torng, E., 2008. A framework for developing ecosystem-specific nutrient criteria: integrating biological thresholds with predictive modeling. Limnol. Oceanogr. 53, 773–787.
Steinschneider, S., Polebitski, A., Brown, C., Letcher, B.H., 2012. Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change. Water Resour. Res. 48, W11525. http://dx.doi.org/10.1029/2011WR011318.
Suen, J., Eheart, J., 2006. Reservoir management to balance ecosystem and human needs: incorporating the paradigm of the ecological flow regime. Water Resour. Res. 42, W03417.
Tharme, R., 2003. A global perspective on environmental flow assessment: emerging trends in the development and application of environmental flow methodologies for rivers. River Res. Appl. 19, 397–441.
U.S. Army Corps of Engineers, 2010. National Inventory of Dams. U.S. Army Corps of Engineers. <http://crunch.tec.army.mil/nidpublic/webpages/nid.cfm> (accessed August 2010).
Vogel, R., Fennessey, N., 1994. Flow-duration curves. 2. New interpretation and confidence-intervals. J. Water Resour. Plann. Manage.-ASCE 120, 485–504.
Vogel, R., Kroll, C., 1992. Regional geohydrologic–geomorphic relationships for the estimation of low-flow statistics. Water Resour. Res. 28, 2451–2458.
Vogel, R., Wilson, I., Daly, C., 1999. Regional regression models of annual streamflow for the United States. J. Irrig. Drain. Eng.-ASCE 125, 148–157.
Vogel, R.M., Sieber, J., Archfield, S.A., Smith, M.P., Apse, C.D., Huber-Lee, A., 2007. Relations among storage, yield, and instream flow. Water Resour. Res. 43, W05403.
Vorosmarty, C., Sahagian, D., 2000. Anthropogenic disturbance of the terrestrial water cycle. Bioscience 50, 753–765.
Wardlaw, R., Sharif, M., 1999. Evaluation of genetic algorithms for optimal reservoir system operation. J. Water Resour. Plann. Manage.-ASCE 125, 25–33.
Weiskel, P.K., Vogel, R.M., Steeves, P.A., Zarriello, P.J., DeSimone, L.A., Ries III, K.G., 2007. Water use regimes: characterizing direct human interaction with hydrologic systems. Water Resour. Res. 43, W04402.
Weiskel, P.K., Brandt, S.L., DeSimone, L.A., Ostiguy, L.J., Archfield, S.A., 2010, Indicators of Streamflow Alteration, Habitat Fragmentation, Impervious Cover, and Water Quality for Massachusetts Stream Basins: U.S. Geological Survey Scientific Investigations Report 2009-5272, 70p., Plus CD–ROM. <http://pubs.usgs.gov/sir/2009/5272/>.
Yang, G., Bowling, L.C., Cherkauer, K.A., Pijanowski, B.C., 2011. The impact of urban development on hydrologic regime from catchment to basin scales. Landscape Urban Plann. 103, 237–247.