JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

Vol. 48, No. 3 AMERICAN WATER RESOURCES ASSOCIATION June 2012

COMPARISON OF STREAM INVERTEBRATE RESPONSE MODELS FOR BIOASSESSMENT METRICS[1]

Ian R. Waite, Jonathan G. Kennen, Jason T. May, Larry R. Brown, Thomas F. Cuffney, Kimberly A. Jones, and James L. Orlando[2]

ABSTRACT: We aggregated invertebrate data from various sources to assemble data for modeling in two ecoregions in Oregon and one in California. Our goal was to compare the performance of models developed using multiple linear regression (MLR) techniques with models developed using three relatively new techniques: classification and regression trees (CART), random forest (RF), and boosted regression trees (BRT). We used tolerance of taxa based on richness (RICHTOL) and ratio of observed to expected taxa (O/E) as response variables and land use/land cover as explanatory variables. Responses were generally linear; therefore, there was little improvement to the MLR models when compared to models using CART and RF. In general, the four modeling techniques (MLR, CART, RF, and BRT) consistently selected the same primary explanatory variables for each region. However, results from the BRT models showed significant improvement over the MLR models for each region; increases in R^2 from 0.09 to 0.20. The O/E metric that was derived from models specifically calibrated for Oregon consistently had lower R^2 values than RICHTOL for the two regions tested. Modeled O/E R^2 values were between 0.06 and 0.10 lower for each of the four modeling methods applied in the Willamette Valley and were between 0.19 and 0.36 points lower for the Blue Mountains. As a result, BRT models may indeed represent a good alternative to MLR for modeling species distribution relative to environmental variables.

(KEY TERMS: modeling; macroinvertebrates; watershed disturbance; land use; prediction; statistical assessment.)

Waite, Ian R., Jonathan G. Kennen, Jason T. May, Larry R. Brown, Thomas F. Cuffney, Kimberly A. Jones, and James L. Orlando, 2012. Comparison of Stream Invertebrate Response Models for Bioassessment Metrics. *Journal of the American Water Resources Association* (JAWRA) 48(3): 570-583. DOI: 10.1111/j.1752-1688.2011.00632.x

INTRODUCTION

Modeling has increased markedly in the past decade in all areas of ecology, and major advances have been made in conceptual models and statistical techniques (Leathwick *et al.*, 2005; Austin, 2007; Cabecinha *et al.*, 2007; Turak *et al.*, 2011), which, in turn, help practitioners derive response models that better support the needs of bioassessment programs. A

[1]Paper No. JAWRA-11-0093-P of the *Journal of the American Water Resources Association* (JAWRA). Received July 26, 2011; accepted December 2, 2011. **Discussions are open until six months from print publication**.

[2]Respectively, Biologist (Waite), U.S. Geological Survey, Oregon Water Science Center, 2130 SW 5th Avenue, Portland, Oregon 97201; Biologist (Kennen), U.S. Geological Survey, New Jersey Water Science Center, Trenton, New Jersey 08628; Biologist (May and Brown) and Hydrologist (Orlando), U.S. Geological Survey, California Water Science Center, Sacramento, California 95819; Biologist (Cuffney), U.S. Geological Survey, North Carolina Water Science Center, Raleigh, North Carolina 27607; and Physical Scientist (Jones), U.S. Geological Survey, Utah Water Science Center, West Valley, Utah 84119 (E-Mail/Waite: iwaite@usgs.gov).

fundamental goal of bioassessment in stream ecology is a better understanding of the effects of human land use on stream biota and the processes at various scales that cause these effects. However, streams are complex spatial and temporal habitat mosaics that are directly and indirectly influenced by a combination of natural geology, climate, and human disturbance (Stanford *et al.*, 2005). Stream ecologists are trying to understand the spatial scales and processes associated with human and natural disturbances that are affecting the biota. Models provide a useful framework for testing hypotheses, determining potential direct and indirect linkages, and directing where further research is needed. The expansion and application of multivariate models in stream ecology are helping to address these issues and hopefully will lead to a broader understanding of ecological and anthropogenic pathways and responses (Oberdorff *et al.*, 2001; Cabecinha *et al.*, 2007; Turak *et al.*, 2011; Waite *et al.*, 2010).

Much of the research documenting the effects of land-use change on stream biota indicates that as the total watershed area in agricultural and/or urban land use increases, individual biological metrics and multimetric indices (MMIs) (such as an Index of Biotic Integrity, IBI) that reflect compositional changes in sensitive species generally decrease (Paul and Meyer, 2001; Allan, 2004; Van Sickle *et al.*, 2004; Cuffney *et al.*, 2005; Ode *et al.*, 2008; Waite *et al.*, 2010). Though some researchers have found a threshold response (i.e., a nonlinear or step function) of individual or multimetric biological indices to land-use indicators (e.g., Davis and Simon, 1995; Wang *et al.*, 2001; Walsh *et al.*, 2005; Hilderbrand *et al.*, 2010; King and Baker, 2010) much of the literature indicates that the response more often is a simple monotonic response with no initial resistance (Booth, 2005; Cuffney *et al.*, 2005, 2010; Kennen *et al.*, 2005; Morgan and Cushman, 2005; Roy *et al.*, 2005; Stanford *et al.*, 2005; Waite *et al.*, 2008, 2010). The debate about possible threshold responses continues not only because of the interest in determining, from a management perspective, where a threshold might occur along a land-use gradient, but also because of the effect thresholds and the resultant nonlinear responses have on the application of various modeling techniques. If biological responses to landscape measures are indeed complex and nonlinear, then newer modeling techniques such as classification regression trees (CART), random forest (RF) and boosted regression trees (BRT), multilevel hierarchical modeling, structural equation models, or artificial neural networks may be necessary to model these responses (Grace, 2006). However, if various biological responses to human disturbance are commonly simple and linear, then they should be more easily modeled via standard regression techniques, which are typically easier to develop and interpret.

There are three commonly used bioassessment variable types including individual biological metrics (e.g., Ephemeroptera, Plecoptera, and Trichoptera richness or EPT), combining individual metrics into a multimetric index (e.g., IBI) and development of the observed/expected ratio metric (O/E). Each method has its advantages and disadvantages, yet sometimes they can give different results in differ environmental settings (Herbst and Silldorff, 2006; Chessman *et al.*, 2010; Hawkins *et al.*, 2010). It is possible that individual metrics may be more stressor gradient specific and multimetric indices better at more general disturbance gradients, however, detailed comparison of these three methods is beyond the scope of this paper. We focus on two common individual biological metrics, the general tolerance of invertebrates to a multitude of stressors including sediment, temperature, dissolved oxygen, hydrological and habitat changes, nutrients, and contaminants following Barbour *et al.* (1999) and the ratio of the observed/expected taxa based on the RIVSPAC method (River Invertebrate Prediction and Classification System) (Clarke, 2000; Moss, 2000). The number of tolerant taxa is expected to increase while the O/E value is expected to decrease as the amount of disturbance to the stream increases.

Using the same dataset used in this paper, Waite *et al.* (2010) developed macroinvertebrate response models for three regions in the western United States (U.S.) and the best multiple linear regression (MLR) models based on Akaike Information Criterion (AIC) and R^2 from each individual region required only two or three explanatory variables to model macroinvertebrate metrics to explain 41-74% of the variation. In each region, their best model contained some measure of urban and/or agricultural land use, yet often the model was improved by including a natural explanatory variable such as mean annual precipitation or mean watershed slope (for the MLR equations, see Waite *et al.*, 2010). Two macroinvertebrate metrics, the richness of tolerant macroinvertebrates (RICHTOL) and some form of EPT richness, were common response variables in models developed among the three regions (Waite *et al.*, 2010). Models were developed for the same two invertebrate metrics even though the geographic regions they modeled reflect distinct differences in precipitation, geology, elevation, slope, population density, and land use. L. R. Brown, J. T. May, A. C. Rehn, P. R. Ode, I. R. Waite, and J. K. Kennen (personal communication) were also able to develop strong models using linear modeling techniques (MLR), they modeled an invertebrate index of biotic integrity (BIBI) across a gradient of urbanized streams in southern California and were able to explain approximately 48% of the variation based on MLR models including classification accuracy of 69 and 87% for impaired and unimpaired sites, respectively.

One important question that researchers are working to answer is whether the use of newer, more complex modeling techniques such as CART and regression trees improves our ability to predict biological metrics and potentially provide new insights into response patterns and mechanistic pathways. Generalized linear models (GLMs) and generalized additive models (GAMs) were introduced in the 1980s and 1990s as improved methods over MLR for data with non-normally distributed errors (e.g., presence–absence and count data) or nonlinear relations and usually outperform single regression trees (Elith *et al.*, 2008). Regression trees are one type of technique within the commonly used CART or decision tree family (e.g., Breiman *et al.*, 1984; De'ath and Fabricius, 2000; Prasad *et al.*, 2006). Trees attempt to explain variation in one categorical (classification) or continuous (regression) response variable by one or more explanatory variables, the resultant output being a dendogram or tree with varying numbers of branches or nodes. These techniques have a few properties that are highly desirable for ecological data analysis: (1) they can handle numeric, categorical, and censored response variables, (2) they are not affected by explanatory variables that follow non-normal distributions (i.e., skewed, Poisson, or bimodal), and (3) they can model complex interactions simply (De'ath, 2007). Maloney *et al.* (2009) found that CART models of watershed disturbance on BIBI values provided results that were intuitive and easy to interpret but they did not classify sites any better than logistic regression models; however, RF models showed minor improvements in performance over the other models. De'ath (2007) and Elith *et al.* (2008) show that BRTs outperform GLMs and GAMs in variable selection, predictive ability (higher R^2 and lower error), and can handle sharp discontinuities in data that are difficult for the other methods. Aertsena *et al.* (2010) also showed that BRT outperformed most modeling techniques (i.e., MLR, GLM, GAM, and CART), with the exception of artificial neural networks.

Over the past decade the estimate of O/E has become a common measure of biological condition for use in bioassessments (e.g., Hawkins, 2006; Carlisle *et al.*, 2008). The expected taxa for a site are commonly estimated by models (e.g., RIVPACS) (Clarke, 2000; Moss, 2000) of reference sites; this value is then compared to the actual taxa collected at a site. Models based on this approach have been developed in many international regions (e.g., Europe, New Zealand, and Australia) (Davies, 2000; Clarke and Murphy, 2006) and for separate regions within the U.S., including many states (Hubler, 2008). Recently, Hawkins *et al.* (2010) compared the response of three types of O/E models with five versions of MMIs for macroinvertebrates and found that in general, the O/E models were better able to distinguish managed or disturbed sites from reference sites than the MMIs. Due to these results and to its overall national and international popularity, we wanted to evaluate how models developed using O/E as the response variable would compare to models developed using single metrics, such as RICHTOL.

Our goal in this paper is to compare the overall performance (i.e., model fit, or R^2) of models developed using standard MLR techniques with more complex models developed using newer alternative techniques such as CART, RF, and boosted regression for the common macroinvertebrate metrics RICHTOL and O/E as the response variables. Also, we believe that the development of watershed disturbance predictive models such as those presented herein will build upon previous research to help the potential derivation of more complex models to better understand disturbance pathways in the landscape and ultimately the biocomplexity of aquatic systems.

METHODS

Data Aggregation and Landscape Analysis

For this comparative analysis we used the datasets (U.S. Geological Survey, U.S. Environmental Protection Agency, Oregon Department of Environmental Quality, and California Department of Fish and Game) previously aggregated for three regions in the western U.S. by Waite *et al.* (2010). A brief summary of the methods follows. Sites were evaluated based on the following criteria: invertebrate data sampled with comparable methods; upstream watershed area of between 13 and 259 km^2; and watersheds could not be nested (i.e., no spatial autocorrelation). Sites meeting these conservative criteria resulted in three study regions: Coastal Southern California ($n = 55$), the Blue Mountains ecoregion of eastern Oregon ($n = 148$), and the Willamette Valley ecoregion in north-central Oregon ($n = 96$) (Figure 1).

For consistency, watersheds were re-delineated for the selected sampling sites within the three study regions using USGS 7.5 min quadrangle digital raster graphics as base layers. The digital raster graphics were displayed on-screen along with National Hydrography Dataset (NHD) high resolution stream lines for each region (U.S. Geological Survey, 2007).

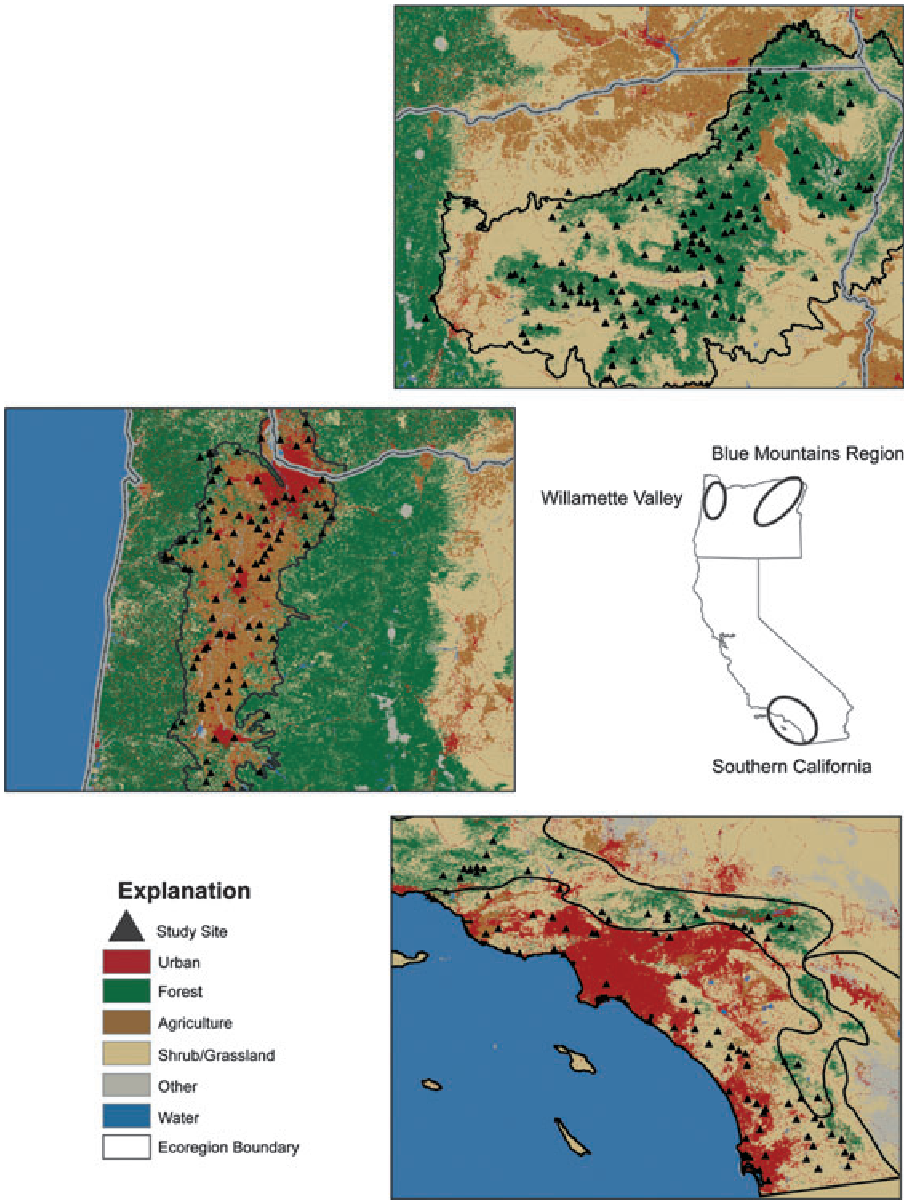

FIGURE 1. Map Showing Land Use and Land Cover for the Three Modeling Regions: Blue Mountains and Willamette Valley, Oregon, and Southern California.

Watershed boundaries were digitized on-screen at a scale of 1:10,000 or larger. Adjacent watershed polygons were edge matched to eliminate all overlaps and gaps. All work was conducted using ArcGIS, ArcMap 9.2 (Environmental Systems Research Institute, Redlands, CA; Table A1) GIS software.

Riparian buffer zone polygons were created within each watershed, extending 2 km upstream from the outlet of each watershed along the main stem and all tributaries and 90 m on either side of the stream centerlines. The buffers were created by selecting the appropriate NHD stream lines within each watershed and creating routes along each main stem and tributary flow path. The routes were then clipped to a distance of 2 km from the basin outlet and buffered. All abbreviations for riparian based explanatory variables begin with the letters "Rip"; otherwise, variables are watershed based (Table 1).

Spatial datasets representing landscape metrics of watershed disturbance were created for each watershed and riparian zone buffer from available national and regional datasets (Table A1) and included elevation, slope, land cover (1992 and 2001), population density, road networks, soil infiltration capacity, hydrography, pollution point sources, dams, and precipitation. Land-use summaries were based on either 1992 or 2001 spatial data (as described in Vogelmann *et al.*, 2001; Homer *et al.*,

TABLE 1. Description, Variable Code and Definition of Explanatory (landscape) and Predictor (invertebrate metrics) Variables Used for Response Model Development.

Explanatory Variables: Landscape		
Description	**Variable Code**	**Definition**
Watershed Scale Variables		
Percent urban land use	Urban	Percent watershed area in urban land use (NLCD 2000 categories 21, 22, 23, and 24)
Percent agricultural land use	Ag	Percent watershed area in agricultural land use (NLCD 2000 category 82)
Sum of percent Ag + Urban	Ag + Urb	Sum of percent watershed area in urban (NLCD 2000 categories 21, 22, 23, and 24) and agricultural (NLCD 82) land use
Percent forest	Forest	Percent watershed area in forest land use (NLCD 2000 categories 41, 42, 43)
Percent pasture	Pasture	Percent watershed area in pasture land use (NLCD 2000 category 81)
Percent shrub/scrub	Shrub	Percent watershed area in shrubland, shrub/scrub (NLCD 2000 category 52)
Road density	RdDens	Road density in watershed = Road length (km)/watershed area (km^2)
Mean population density	PopDen	Watershed mean population density based on 2000 census (persons/km^2)
Minimum elevation	Min-Elev	Elevation (m) at stream site, pour point of watershed
Mean slope percent	Slope	Mean percent watershed slope
Manmade stream density	MmStreams	Manmade stream density in watershed = manmade stream length (km)/ watershed area (km^2)
Mean annual precipitation	MnAnnPrecip	Mean annual precipitation (cm)
Soil infiltration rate	Soil_Mod-Infil	Hydrologic soil group B, moderate infiltration rate (min. infiltration rate 4-8 mm/h)
Riparian Scale Variables		
Percent urban land use	Rip_Urban	Percent buffer area in urban land use (NLCD 2000 categories 21, 22, 23, and 24)
Percent agricultural land use	Rip_Ag	Percent buffer area in agricultural land use (NLCD 2000 category 82)
Sum of percent Ag + Urban	Rip_Ag + Urb	Sum of percent buffer area in urban (NLCD 2000 categories 21, 22, 23, and 24) and agricultural (NLCD 82) land use
Percent forest	Rip_Forest	Percent buffer area in forest land use (NLCD 2000 categories 41, 42, 43)
Percent pasture	Rip_Pasture	Percent buffer area in pasture land use (NLCD 2000 category 81)
Percent shrub/scrub	Rip_Shrub	Percent buffer area in shrubland, shrub/scrub (NLCD 2000 category 52)
Road density	Rip_RdDens	Road density in buffer = Road length (km)/watershed area (km^2)
Mean population density	Rip_PopDens	Buffer area mean population density based on 2000 census (persons/km^2)
Mean slope percent	Rip_Slope	Mean percent buffer slope
Maximum elevation	Rip_Max-Elev	Maximum buffer elevation (m)
Response Variables: Invertebrate Metrics		
Observed/expected	O/E	Ratio of number of observed taxa at a site over the expected taxa based on modeled reference sites
Tolerant richness	RICHTOL	Average USEPA tolerance values for sample based on richness

2004), depending on which data source was closer to the macroinvertebrate sample date for that watershed. Watersheds and riparian zone buffers were used to define zones for analysis and calculate summary statistics. The 1992 and 2001 land cover datasets used slightly different classification schemes. Uniform codes based on the 2001 classification scheme were assigned to all land cover classes in the final summary statistics table (Fry *et al.*, 2009). We did not assess the distribution pattern of land use/land cover within the watershed though this can be important in some situations.

Description of Modeling Regions

The Coastal Southern California (SoCal; Southern and Central California Chaparral and Oak Woodlands Ecoregion) region has a Mediterranean climate of hot, dry summers and cool, moist winters (Ode *et al.*, 2005). Average precipitation at each site ranges from 25 to 50 cm/year. The geology of the ecoregion is dominated by recently uplifted and poorly consolidated marine sediments. Vegetative cover in this region consists mainly of chaparral and oak woodlands, though grasslands occur in some lower elevations and patches of pine are found at higher elevations (open low mountains or foothills). The landscape is currently dominated by urban development; the human population is approximately 19 million and is projected to exceed 28 million by 2025 (Ode *et al.*, 2005). Outside the urban centers, much of this region was historically grazed by domestic livestock or cultivated for fruits and vegetables, but most of this land has since been converted to urban uses.

The Blue Mountains (Blue_Mt) are the westernmost range of the Middle Rocky Mountains and, like the Cascade Range, are largely volcanic, with fertile plateaus and deeply fissured river valleys. Carved by two rivers (the John Day and Grande Ronde Rivers) the landscape has steep hillsides, bluffs and rimrock faces. Temperature and precipitation are highly correlated with elevation. Precipitation ranges from 22 to 45 cm/year along the river valleys and is >150 cm/year in the nearby mountains. This region is dominated by coniferous forests in mid to higher elevations and shrub and grassland in lower elevations, though much of the latter has been displaced by agriculture and grazing. The region has no large cities and urbanization is limited to scattered smaller cities and small towns.

The Willamette Valley (Will_V) ecoregion contains a mixture of rolling prairies, mixed forests, and extensive lowland valley wetlands. With temperate, dry summers and cool, wet winters, the Willamette River basin and surrounding area is characteristic of the Pacific Northwest climate. About 90% of the annual precipitation (100-130 cm/year) occurs during October through May (Uhrich and Wentz, 1999), falling as rain in the valley and snow in the mountains. The land use/land cover in the valley plains and foothills is primarily cultivated crops, pasture, and grasslands. Urbanization ranges from minimal to extensive (Waite *et al*., 2008). Centered on the confluence of the Columbia and Willamette Rivers, Portland is the most populous city in Oregon, with 539,000 people in city limits and nearly 3 million people in the Portland metropolitan area (U.S. Census Bureau, 2000). The population in the metropolitan area increased almost 30% from 1990 to 2000, with some suburban populations increasing more than 80% during the same period (U.S. Census Bureau, 2000). The drainage network in the Willamette Valley combines natural tributaries, complex networks of canals in agricultural areas, and stormwater canals and groundwater infiltration wells in cities.

The three geographic regions modeled in this study have differing natural settings and the extent and type of human disturbance in each respective region. SoCal has the driest climate, intermediate mean stream site elevation (Min-Elev) and percent agriculture, and the highest population density. Blue_Mt has the highest mean site elevation and mean watershed slope, intermediate mean precipitation, and the lowest population density, percent urban, and percent agriculture. Will_V has the greatest precipitation, lowest minimum site elevation, and the highest percent agriculture.

Macroinvertebrate Data

Macroinvertebrate data from 1994 to 2005 assembled for this study were considered to be comparable in terms of sampling protocols (sampled habitat, number of composite samples, and total sampled area) and laboratory procedures, including sorting, subsample count level, and taxonomic resolution (personal communication state agency personnel, 2005; Waite *et al.*, 2010). In general, all macroinvertebrate samples were collected in similar habitats using kick-net techniques from five to eight separate areas and combined for a composite sample (Moulton *et al.*, 2002; Peck *et al.*, 2006; Hubler, 2008). Extensive review of the data was completed to make sure aggregated data from disparate sources included the same taxonomic groups, followed the same nomenclature, and had appropriate taxonomic resolution before data analysis was attempted. The Invertebrate Data Analysis System software (Cuffney, 2003) was used to resolve by region all taxonomic issues (taxonomic identification level and nomenclature), to remove ambiguous taxa (Cuffney *et al.*, 2007), and to randomly subsample raw counts to an equal 300 (Will_V) or 500 specimen count (the highest possible based on the data in each region) across all study regions. In general, data for dominant aquatic insect orders were resolved at genus level. Less common orders were often aggregated to family level. Rare organisms or those with difficult taxonomy were sometimes aggregated to order or higher. The dipteran family Chironomidae is considered an important bioindicator group, yet historically a difficult group to identify to genus or species. As a result, data for this group were assigned to six taxa levels (five subfamilies plus Chironomidae) from the various family to genus level identifications within the original data. Tolerance and functional group metrics were calculated using values from Barbour *et al.* (1999), supplemented with values from Wisseman's tolerances for the Pacific Northwest (Wisseman, 1996, unpublished data). Macroinvertebrate O/E values were estimated using two existing regional models (East and West of the Cascade Mountains) that were developed by Oregon Department of Environmental Quality (Hubler, 2008). O/E models were not ready for the SoCal region at the time of analysis so we were not able to test O/E values for this area.

MODEL DEVELOPMENT

Details of MLR model development procedures are outlined in Waite *et al.* (2010). In brief, model performance was assessed using a variety of statistics, including adjusted mean sum of squares (R^2), root mean squared error, AIC, predicted sum of squares, and regression coefficients in Waite *et al.* (2010). We adopted a model fitting approach for each response variable. We used a step-wise selection based on AIC for all models ranging from 1 to 5 environmental variables, as appropriate by region. If necessary, variables were transformed to improve their distributions to better adhere to assumptions of linearity. Models were developed for each geographic region separately due to the large spatial separation between each region and as described above, because the climatic and disturbance regimes were distinct. Model residuals, potential outliers, and interaction terms were evaluated. A description of variables used in model development is provided in Table 1. A MLR model was developed for the response variable RICHTOL for all three regions; it included two predictor variables (population density and riparian road density) for SoCal, three predictor variables for Blue_Mt (percent shrubs, percent agriculture, and mean annual precipitation in the watershed) and three predictor variables for the Will_V region (percent agriculture plus urban land use in the watershed, mean annual precipitation, and percent agriculture plus urban land use in the riparian zone) (Waite *et al.*, 2010). As a comparison to the MLR models developed by Waite *et al.* (2010) for RICHTOL, new models were developed for O/E for the Blue_Mt and Will_V regions.

To gain additional insight into these data and as a comparison against the MLR models, single regression trees, RF, and BRT models were developed for each region individually. Regression trees are one type of technique within the commonly used CART or decision tree family, and their use and technical details have been described extensively in the literature (e.g., Breiman *et al.*, 1984; De'ath and Fabricius, 2000; Prasad *et al.*, 2006); therefore, we will only provide a brief overview. Trees attempt to explain variation in one categorical (classification) or continuous (regression) response variable by one or more explanatory variables, the resultant output being a dendogram, or tree, with varying numbers of branches or nodes. Trees are developed following a hierarchical binary splitting procedure that attempts to find the best single explanatory variable that minimizes the within group and maximizes the among group dissimilarity in the response variable at each split. It does this for each explanatory variable entered into model development and can thus provide a list of the explanatory or predictive power of the variables. We used R statistics scripts and software (R Development Core Team, 2007, version 2.10.0) following the procedures outlined by Therneau and Atkinson (1997) to determine the proper single regression tree and the appropriate pruning of branches (De'ath and Fabricius, 2000; Prasad *et al.*, 2006). Trees have a few properties that are highly desirable for ecological data analysis: (1) they can handle numeric and categorical variables (2) they are not affected by explanatory variables that follow non-normal distributions (i.e., skewed, Poisson, or bi-modal), and (3) they can model complex interactions simply (De'ath, 2007).

Random forests and BRT are among a family of techniques used to advance single classification or regression trees by averaging the results for each binary split from numerous trees or forests thus reducing the predictive error and improving overall performance (De'ath, 2007; Elith *et al.*, 2008). In BRT, after the initial tree has been generated, successive trees are grown on reweighted versions of the data giving more weight to those cases that are incorrectly classified than those that are correctly classified within each growth sequence. Thus, as more and more trees are grown in BRT, the large number of trees increases the chance that cases that are difficult to classify initially are correctly classified, thus representing an improvement to the basic averaging algorithm used in RF (De'ath, 2007). Boosted trees and RF models retain the positive aspects of single trees seen in CART models, yet have improved predictive performance, nonlinearities and interactions are catered to or easily assessed, and they can provide an ordered list of the importance of the explanatory variables (Cutler *et al.*, 2007; De'ath, 2007). Though RF and BRT offers improved modeling performance over CART, the simple single tree obtained from CART is lost, making it more difficult to visualize the results. Partial dependency plots (PDP) are a way to visualize the effect of a specific explanatory variable on the response variable after accounting for the average effects of all other explanatory variables (De'ath, 2007; Elith *et al.*, 2008); these are presented in this paper for select models as examples (e.g., Figures 2 and 3). Random forest models were developed using the rpart library in R following methods outlined in Cutler *et al*. (2007) and BRT models were run using the gbm library in R and specific code from Elith *et al.* (2008). We used R^2 values for assessing the amount of variation explained among the four modeling techniques since it is a common and well understood measure that allowed us to put each model on the same measurement currency; other model performance measures such as confidence intervals and *p*-values are not included for simplicity.

RESULTS

In general, the four modeling techniques selected the same primary explanatory variables within each

TABLE 2. Explanatory Variables in Order of Importance in the Models for Four Modeling Methods for Two Macroinvertebrate Metrics for Each of Three Study Regions (SoCal, Southern California; Will_V, Willamette Valley; Blue_Mt, Blue Mountains, Oregon).

	MLR	CART	RF	BRT
SoCal				
RICHTOL	PopDen	PopDen	PopDen	PopDen
	Rip_RdDens	MmStreams	Min-Elev	Rip_Slope
		Min-Elev	Rip_Slope	Min-Elev
Will_V				
RICHTOL	Ag + Urb	Ag + Urb	Ag + Urb	Ag + Urb
	MnAnnPrecip	MnAnnPrecip	MnAnnPrecip	MnAnnPrecip
	Rip_Ag + Urb	Rip_Forest	Rip_Forest	Rip_Max-Elev
			Rip_Max-Elev	Rip_Forest
O/E	Ag + Urb	Forest	Forest	Ag + Urb
	MnAnnPrecip	Rip_Max-Elev	Rip_Max-Elev	Rip_Max-Elev
	Rip_Ag + Urb		Soil_Mod-Infil	MnAnnPrecip
			Rip_Forest	
Blue_Mt				
RICHTOL	Shrub	Shrub	Shrub	Shrub
	Ag	Slope	Slope	MnAnnPrecip
	MnAnnPrecip	MnAnnPrecip	MnAnnPrecip	Slope
O/E	MnAnnPrecip	Slope	Shrub	Slope
	Shrub	MnAnnPrecip	Slope	MnAnnPrecip
	Slope		MnAnnPrecip	Shrub

Notes: MLR, multiple linear regression; CART, classification and regression trees; RF, random forest; BRT, boosted regression trees; RICHTOL, average tolerance value for sample based on richness at a site; O/E, ratio of observed/expected taxa.

region with minor variation among model types (Table 2): (1) SoCal: population density, minimum elevation, and riparian slope, (2) Blue_Mt: percent shrub, mean annual precipitation (MnAnnPrecip), and watershed slope, and (3) Will_V: percent agriculture plus urban, MnAnnPrecip, riparian maximum elevation, and percent riparian forest (see Table 1 for definitions). Generally, the RICHTOL R^2 values for MLR were slightly higher than those for the CART and RF models for all three regions (Table 3); however, this was not the case for the O/E models for Blue_Mt. Nevertheless, these differences are probably not meaningful because the R^2 values for CART and RF models are determined by a cross-validation method that ensures no over-fitting and thus usually gives a lower, more conservative value than the MLR values. Interaction affects were tested for and found to not be significant in the models developed. Conversely, the BRT models showed considerable improvement in the R^2 values over all the other models for both response variables (i.e., RICHTOL and O/E). For example, the SoCal RICHTOL R^2 values for the MLR compared to the BRT model increased from 0.67 to 0.79, Blue_Mt showed an increase from 0.44 to 0.59 for RICHTOL and from 0.08 to 0.28 for O/E, and the Will_V R^2 values increased from 0.74 to 0.83 for RICHTOL and from 0.64 to 0.75 for O/E (Table 3).

TABLE 3. Comparison of R^2 Values for Four Modeling Methods for Two Macroinvertebrate Metrics for Each of Three Study Regions (SoCal, Southern California; Will_V, Willamette Valley; Blue_Mt, Blue Mountains, Oregon).

	MLR	CART	RF	BRT
SoCal				
RICHTOL	0.67 (2)	0.64 (3)	0.65 (3)	**0.79** (3)
Will_V				
RICHTOL	0.74 (3)	0.68 (3)	0.73 (4)	**0.83** (4)
O/E	0.64 (3)	0.62 (2)	0.61 (4)	**0.75** (3)
Blue_Mt				
RICHTOL	0.44 (3)	0.34 (3)	0.41 (3)	**0.59** (3)
O/E	0.08 (3)	0.15 (2)	0.07 (3)	**0.28** (3)

Notes: Number of variables in model in parentheses. MLR, multiple linear regression; CART, classification and regression trees; RF, random forest; BRT, boosted regression trees; RICHTOL, average tolerance value for sample based on richness at a site; O/E, ratio of observed/expected taxa. Highest R^2 value across all models is shown in bold.

The O/E metric derived from RIVPACS type models specifically calibrated for Oregon consistently had lower R^2 values than RICHTOL for the two regions tested (Table 3). Modeled O/E R^2 values were between 0.06 and 0.10 lower than RICHTOL values for each of the four modeling methods applied in the Will_V region and were between 0.19 and 0.36 points lower for the Blue_Mt region.

As mentioned above, all modeling procedures (i.e., MLR, CART, RF, and BRT) generally retained the same subset of explanatory variables. These variables, with some minor exceptions in the Blue_Mt study region, generally accounted for approximately a similar proportion of the variance in the

RICHTOL and O/E response models. R^2 values, however, do not provide a complete picture of the model response pattern, and the overall influence of a specific explanatory variable on the environmental system or process being modeled is typically lost when the model is fit to a linear or nonlinear form. Partial dependency plots, which are provided as a diagnostic tool in the BRT and RF model output, provide a way to more fully examine the relative influence of individual explanatory variables on the response variable given the modeled structure. As explained in De'ath (2007) and Elith *et al.* (2008), PDP provide a way to visualize the effect of a specific explanatory variable on the response variable after accounting for the average effects of all other explanatory variables. For example, PDPs for the four variables retained in the BRT model for Will_V are shown in Figure 2. In general, the plots show a near linear increase in RICHTOL as the amount of agriculture plus urban land use in the watershed increases (Figure 2A) and a decrease in RICHTOL as riparian maximum elevation increases (Figure 2D). However, the response in RICHTOL values flattens out at approximately 60% agriculture plus urban land use, then again increases rapidly from approximately 90 to 100%. Likewise, the PDP graph shows that there is rapid change in RICHTOL values from near 0 to 200 m in riparian maximum elevation followed by no response beyond 400 m. The pattern shown for mean annual precipitation (Figure 2B) follows the opposite pattern of the amount of agriculture plus urban land use in the watershed, RICHTOL values decrease rapidly from the lowest precipitation values until approximately 80 cm/year beyond which values show no response. As the amount of riparian forest cover declines (Figure 2C), RICHTOL values increase little until riparian forest values drop to about 30%, where there is a step-wise increase until the point when there is only about 5% riparian forest remaining, whereupon there is a rapid increase in tolerance values. The PDPs for O/E in the Will_V show remarkable similarity to that described above for RICHTOL except that, as one would expect due to the differences in the invertebrate metrics, the curves respond in opposite directions (Figure 3). There is a general linear decrease in O/E values as agriculture plus urban land use increases (Figure 3A), a sharp increase in O/E values as riparian maximum elevation increases to 200 m (Figure 3B) or when mean annual precipitation increases to about 70 cm/year (Figure 3C). As seen for RICHTOL, O/E showed an abrupt threshold-type response at low levels of riparian forest (Figure 3D) followed by a step

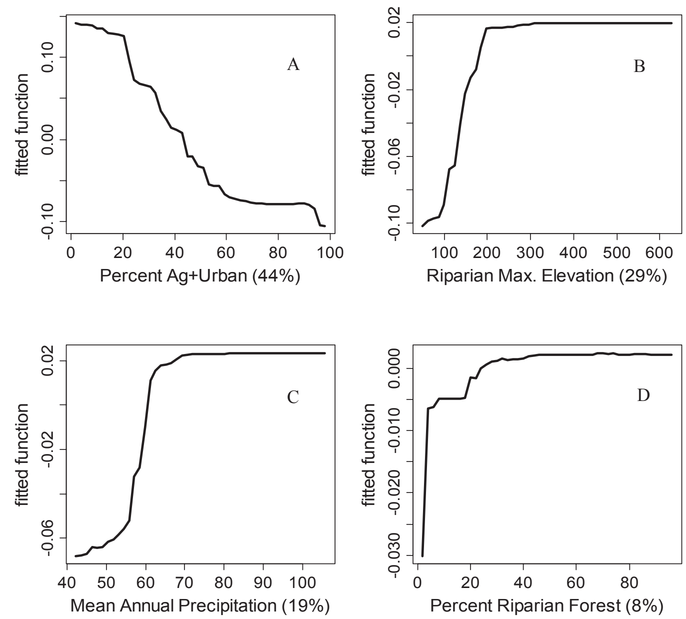

FIGURE 3. Partial Dependency Plots for Ag + Urb (A), Rip_Max-Elev (B), MnAnnPrecip (C), and Rip_Forest (D) in the Boosted Regression Model Developed for Observed/Expected (O/E) in Willamette Valley (Will_V). The *y*-axis represents the effect of the selected variable on the response variable O/E metric, the relative contribution of each explanatory variable is reported in parentheses. Refer to Table 1 for variable definitions.

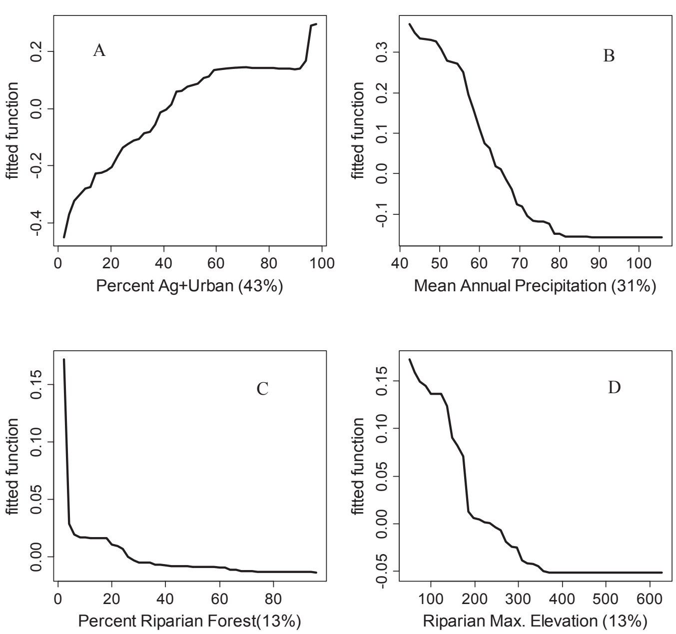

FIGURE 2. Partial Dependency Plots for Ag + Urb (A), MnAnnPrecip (B), Rip_Forest (C), and Rip_Max-Elev (D) in the Boosted Regression Model Developed for RICHTOL in Willamette Valley (Will_V). The *y*-axis fitted function represents the effect of the selected variable on the response variable RICHTOL; the relative contribution of each explanatory variable is reported in parentheses. Refer to Table 1 for variable definitions.

increase and a plateau above approximately 30% riparian forest cover.

DISCUSSION

It is encouraging that the MLR and the CART and RF (regression tree family) modeling techniques gave similar results selecting in general the same main explanatory variables (Table 2) and explaining similar amounts of variation (Table 3), which may indicate that the MLR methods used in this study are appropriate for these types of ecological data. The BRT models, however, did show notable improvement in model fit with increases in R^2 values ranging from 0.09 through 0.15 for RICHTOL to 0.11 through 0.20 for O/E compared to MLR models (Table 3). L. R. Brown, J. T. May, A. C. Rehn, P. R. Ode, I. R. Waite, and J. K. Kennen (personal communication), using a MMI for macroinvertebrates (i.e., BIBI) sampled across a strong urbanization gradient, also showed a notable improvement in model performance for BRT compared to MLR. De'ath and Fabricius (2000) suggest that for complex or messy data, even single regression trees will often outperform MLR and are preferred for determining variable selection and interaction effects due to the issue that MLR models with complex data are frequently difficult to interpret because they will often include too many variables with high order interactions. It was found that CART and RF models did not outperform the RICHTOL MLR models in this analysis which supports our overarching hypothesis that MLR will generally perform as well as many of the tree modeling techniques when data follows a general linear response or when, in the case of the three regions evaluated, there are few explanatory variables with no high order interactions. Maloney *et al.* (2009) found that CART models of land-use disturbance on macroinvertebrate IBI metrics provided results that were intuitive, but they did not classify sites any better than logistic regression models; however, unlike in this study, their RF models showed minor improvements in performance over CART and logistic regression models.

In general, regression trees allow the inclusion of more variables in the model building phase than MLR, allow for easier testing for interaction affects and produce a list of variables explaining the importance of variation in the response variable. In addition, the PDPs from BRT or RF can offer valuable insights into the pattern or form of the response variable based on select explanatory variables improving model interpretation. For example, the PDPs for Will_V (Figures 2 and 3) revealed that the response rate changed or flattened out and provided additional insight into potential thresholds along the range of the individual explanatory variables that are not easily depicted with MLR models.

The identification of thresholds (i.e., transition points in ecological condition) is of growing interest to the scientific and regulatory community, especially for forecasting the loss of biodiversity (Hilderbrand *et al.*, 2010) or for understanding system recovery (Clements *et al.*, 2010; Qian and Cuffney, 2012). More research is clearly needed to help better detect nonlinear and possible threshold responses (Dodds *et al.*, 2010) and new analytical tools are emerging (i.e., BRT results shown in this study) that can assist with identifying changes in taxa occurrence across an environmental gradient (Qian and Cuffney, 2012).

Even though we were able to successfully develop strong MLR models indicating that the primary responses were linear in nature (Waite *et al.*, 2010), the BRT PDPs reveal potential thresholds in the response variable in at least some of the regions (e.g., the Will_V PDPs shown for RICHTOL and O/E in Figures 2 and 3) that were not seen in the MLR models. It is possible that since MLR models assume linearity that they may sometimes miss nonlinear/thresholds in some explanatory variables. The response of RICHTOL and O/E for watershed agriculture plus urban (Ag + Urb) was primarily linear with a small step function at the end (Figures 2A and 3A). The two riparian variables, riparian maximum elevation (Rip_Max-Elev; Figures 2D and 3B) and riparian forest (Rip_Forest; Figures 2C and 3D) on the other hand showed potential thresholds. The response of the two invertebrate metrics to changes in Rip_Max-Elev showed no response from 600 to 400 m for RICHTOL and to 200 m for O/E, after which there was a steep increase or decrease to the lowest elevation (Figures 2D and 3B). It is likely that riparian elevation is acting as a surrogate for the natural climatic and geologic trend that occurs in the Willamette Valley, trending from the valley floor with low stream gradient and lower elevation and precipitation to higher values for these and other variables as one moves toward the foothills of the Coast or Cascade Ranges on either side of the valley. The response of RICHTOL and O/E to changes in Rip_Forest showed a slow but continuous linear increase or decrease as the amount of Rip_Forest decreased from 100% to approximately 5%, after which there appears to be a rapid change in either of the metric values, which may indicate a strong threshold at or near the 5% level. This suggests that as percent forest in the riparian zone along streams drops below approximately 5-10%

land cover, stream integrity degrades rapidly possible due to the reduction in natural buffering capacity seen in healthy riparian systems. L. R. Brown, J. T. May, A. C. Rehn, P. R. Ode, I. R. Waite, and J. K. Kennen (personal communication) found a similar response in the MMI they modeled (BIBI) against four explanatory variables across a strong urbanization gradient in some California streams. They showed that the amount of agriculture plus urban land use in the riparian zone and mean annual precipitation in the watershed showed approximate linear responses, though in opposite directions. They also found a threshold-type response in the BIBI to low values of population density (approximately 300 persons/km^2) in the watershed. Similar to the findings in this study, L. R. Brown, J. T. May, A. C. Rehn, P. R. Ode, I. R. Waite, and J. K. Kennen (personal communication) found that the BRT method appeared to be more sensitive for detecting nonlinear response patterns such as thresholds, for determining potential surrogate variables, and for model corroboration.

The overall poorer performance of the O/E metric compared to the single metric RICHTOL across all models was notable, yet the especially poor performance in the Blue_Mt region was particularly surprising (Table 3). When comparing the ability of O/E and a multimetric invertebrate IBI to differentiate between reference and degraded sites, Herbst and Silldorff (2006) found that the two methods were in close agreement for sites in eastern Sierra Nevada of California. Hawkins *et al.* (2010) compared the performance of a multimetric index and O/E for 225 sites from five ecoregions in the interior Columbia Basin, including many of the sites used in this study from the Blue_Mt ecoregions. They found that the O/E metric was better at distinguishing among the three disturbance classes, particularly between the intermediate and high disturbance classes than the multimetric index. The discrepancy between the poor performance of O/E in the Blue_Mt region in our study and the strong performance in their study may be due to a larger underlying disturbance gradient within their dataset, which resulted from the inclusion of data from multiple ecoregions. Models derived for the Will_V region, where there was a larger disturbance gradient than that found in the Blue_Mt region, showed relatively little difference in performance between the O/E and RICHTOL metrics. It is also possible that the lower R^2 for the O/E models may be because we are not able to model nor account for the error associated with estimation of the raw O/E metric values. Chessman *et al.* (2010) found that O/E values did not distinguish among site disturbance groups based on hydrologic alteration in Australia even though taxonomic richness and assemblage composition could. However, it is yet unclear why O/E performance would be inhibited in areas with a shorter disturbance gradient than that shown in Hawkins *et al.* (2010). One possibility is that because these O/E models are based on a subset of taxa that occur at 50% of the reference sites and therefore operate with a reduced taxa list, specifically with the relatively rare and arguably with the more sensitive portion of the taxa list removed, the resulting O/E values may be less able to distinguish the small more subtle differences among sites, such as that seen in the Blue_Mt study region. In contrast, the RICHTOL metric uses all the taxa that occur at a site and may be a more sensitive measure of changes in assemblage integrity in areas of low anthropogenic disturbance.

CONCLUSIONS

Waite *et al.* (2010) were able to successfully develop MLR models for the three distinct and separate regional datasets presented in this study for individual macroinvertebrate metrics (e.g., RICHTOL, EPT). This study developed alternate models, CART, RF, BRT, for the same datasets and compared them to the MLR models previously developed. The O/E metric performed nearly as well as RICHTOL in the Will_V region where there was a strong disturbance gradient but performed poorly in Blue_Mt, a region with a relatively weak gradient. Though the data modeled in this study were not particularly noisy or complex, the BRT models, in all cases, outperformed the MLR methods and provided specific information on the form of the response function for each variable giving important insight into potential thresholds in the data. As a result of this ecological modeling comparison, BRT models may indeed represent a good alternative to MLR for modeling species distribution relative to environmental variables. Modeling results indicate that even when the response pattern is simple and strongly linear, BRT models not only markedly improve model fit, but can also help to corroborate results from other methods, provide additional information on potential interactions among variables, and support greater insight into understanding the response profile of a given metric, whether it be a linear, step, or a threshold function, across environmental gradients that may not be easily seen with MLR. Models like these can be used to better understand potential causal linkages between environmental drivers and stream biological attributes or condition and predict expected values of macroinvertebrate metrics at unsampled sites.

APPENDIX

TABLE A1. Sources of Geographical Information System (GIS) and Digital Data Used in Model Development.

Spatial Dataset	Data Source	Source Data Format	Processing Format	Resolution/ Scale	Reference
Hydrography	National Hydrography Dataset (NHD)	Vector	Vector	1:24,000	U.S. Geological Survey, National Hydrography Dataset, Digital data, *accessed* January 2007 at http://nhd.usgs.gov/data.html
Land Cover 1992	National Land Cover Dataset 1992 (NLCD)	Raster	Vector	30 m	U.S. Geological Survey, National Land Cover Dataset 1992, Digital data, *accessed* March 2003 at http://landcover.usgs.gov/natllandcover.php
Land Cover 2001	NLCD 2001	Raster	Vector	30 m	U.S. Geological Survey, National Land Cover Dataset 2001, Digital data, *accessed* January 2007 at http://www.mrlc.gov/
Elevation	National Elevation Dataset (NED)	Raster	Raster	10 m	U.S. Geological Survey, National Elevation Dataset, Digital data, *accessed* May 2007 at http://seamless.usgs.gov/
Slope	NED	Raster	Raster	10 m	U.S. Geological Survey, National Elevation Dataset, Digital data, *accessed* May 2007 at http://seamless.usgs.gov/
Road networks	U.S. Census Bureau Tiger	Vector	Vector	1:100,000	U.S. Census Bureau, TIGER line data, Digital data, *accessed* May 2007 at http://www.census.gov/geo/www/tiger/
	Ground Transportation Roads Publications Arc	Vector	Vector	1:24,000	Oregon BLM, Ground Transportation Roads Publication Arc, Digital data, *accessed* July 2007 at http://www.blm.gov/or/gis/
Soil infiltration capacity	USDA NRCS STATSGO	Vector	Vector	1:250,000	Natural Resource Conservation Service, STATSGO soils data, Digital data, *accessed* May 2007 at http://datagateway.nrcs.usda.gov/
Population density	U.S. Census Bureau Census 2000	Vector	Raster	30 m	U.S. Census Bureau, Census 2000, Digital data, *accessed* May 2007 at http://www.census.gov/main/www/cen2000.html
Precipitation	Oregon State University PRISM	Raster	Raster	30 arc-seconds	PRISM Group, Oregon State University, Precipitation data for the U.S., Digital data, *accessed* May 2007 at http://www.prismclimate.org
Dams	National Inventory of Dams	Vector	Vector	Various	U.S. Army Corps of Engineers, National Inventory of Dams, Digital data, Not publicly available

ACKNOWLEDGMENTS

We thank each of the individuals that provided the macroinvertebrate data and associated advice: Shannon Hubler, Oregon Department of Environmental Quality; Alan Herlihy, Oregon State University and Environmental Protection Agency; Pete Ode and Andy Rehn, California Department of Fish and Game. We thank the many dedicated USGS colleagues who acquired the biological samples used in this analysis. We also gratefully acknowledge Alan Herlihy, John Van Sickle, and Anthony Olsen (Environmental Protection Agency, Corvallis) for their insight and advice on issues related to modeling macroinvertebrates, and Dr. Yangdong Pan (Portland State University) for advice on R programming and statistics. Finally, we thank one USGS reviewer and one outside reviewer for their review of the manuscript prior to submission to the journal.

LITERATURE CITED

Aertsena, W., V. Kinta, J. Van Orshovena, K. Özkanb, and B. Muysa, 2010. Comparison and Ranking of Different Modelling Techniques for Prediction of Site Index in Mediterranean Mountain Forests. Ecological Modelling 221:1119-1130.

Allan, J.D., 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. Annual Review of Ecology, Evolution and Systematics 35:257-284.

Austin, M., 2007. Species Distribution Models and Ecological Theory: A Critical Assessment and Some Possible New Approaches. Ecological Modelling 200:1-19.

Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling, 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates, and Fish (Second Edition). U.S. Environmental Protection Agency, Office of Water, EPA Report 841-B-99-002, Washington, D.C.

Booth, D.B., 2005. Challenges and Prospects for Restoring Urban Streams: A Perspective From the Pacific Northwest of North America. Journal of the North American Benthological Society 24:724-737.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone, 1984. Classification and Regression Trees. Wadsworth and Brooks/Cole, Monterey, California.

Cabecinha, E., P. Silva-Santos, R. Cortes, and J. Cabral, 2007. Applying a Stochastic-Dynamic Methodology (StDM) to Facilitate Ecological Monitoring of Running Waters, Using Selected Trophic and Taxonomic Metrics as State Variables. Ecological Modelling 207:109-127.

Carlisle, D.M., J. Falcone, and M.R. Meador, 2008. Predicting the Biological Condition of Streams: Use of Geospatial Indicators of Natural and Anthropogenic Characteristics of Watersheds. Environmental Monitoring and Assessment 151:143-160.

Chessman, B.C., H.A. Jones, N.K. Searle, I.O. Growns, and M.R. Pearson, 2010. Assessing Effects of Flow Alteration on Macroinvertebrate Assemblages in Australian Dryland Rivers. Freshwater Biology 55:1780-1800.

Clarke, R.T., 2000. Uncertainty in Estimates of River Quality Based on RIVPACS. *In*: Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques, J.F. Wright, D.W. Sutcliffe, and M.T. Furse (Editors). Freshwater Biological Association, Ambleside, pp. 39-54.

Clarke, R.T. and J.F. Murphy, 2006. Effects of Locally Rare Taxa on the Precision and Sensitivity of RIVPACS Bioassessment of Freshwaters. Freshwater Biology 51:1924-1940.

Clements, W.H., N.K.M. Vieira, and D.L. Sonderegger, 2010. Use of Ecological Thresholds to Assess Recovery in Lotic Ecosystems. Journal of the North American Benthological Society 29:1017-1023.

Cuffney, T.F., 2003. User's Manual for the National Water-Quality Assessment Program Invertebrate Data Analysis System (IDAS) Software: Version 3. U.S. Geological Survey Open-File Report 03-172, 103 pp.

Cuffney, T.F., M.D. Bilger, and A.M. Haigler, 2007. Ambiguous Taxa: Effects on the Characterization and Interpretation of Invertebrate Assemblages. Journal of the North American Benthological Society 26:286-307.

Cuffney, T.F., R.A. Brightbill, J.T. May, and I.R. Waite, 2010. Responses of Benthic Macroinvertebrates to Environmental Changes Associated With Urbanization in Nine Metropolitan Areas. Ecological Applications 20:1384-1401.

Cuffney, T.F., H. Zappia, E.M.P. Giddings, and J.F. Coles, 2005. Urbanization Effects on Benthic Macroinvertebrate Assemblages in Contrasting Environmental Settings: Boston, Birmingham, and Salt Lake City. *In*: Effects of Urbanization on Stream Ecosystems, L.R. Brown, R.H. Gray, R.M. Hughes, and M.R. Meador (Editors). American Fisheries Society Symposium 47: 361-408.

Cutler, D.R., T.C. Edwards, K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, and J.J. Lawler, 2007. Random Forests for Classification in Ecology. Ecology 88:2783-2792.

Davies, P.E., 2000. Development of a National River Bioassessment System (AUSRIVAS) in Australia. *In*: Assessing the Biological Quality of Freshwaters: RIVPACS and Other Techniques, J.F. Wright, D.W. Sutcliffe, and M.T. Furse (Editors). Freshwater Biological Association, Cumbria, United Kingdom, pp. 113-124.

Davis, W.S. and T.P. Simon, Editors. 1995. Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making. Lewis Publishers, Boca Raton, Florida.

De'ath, G., 2007. Boosted Trees for Ecological Modeling and Prediction. Ecology 88:243-251.

De'ath, G. and K.E. Fabricius, 2000. Classification and Regression Trees: A Powerful Yet Simple Technique for Ecological Data Analysis. Ecology 81:3178-3192.

Dodds, W.K., W.H. Clements, K. Gido, R.H. Hilderbrand, and R.S. King, 2010. Thresholds, Breakpoints, and Nonlinearity in Freshwaters as Related to Management. Journal of the North American Benthological Society 29:988-997.

Elith, J., J.R. Leathwick, and T. Hastie, 2008. A Working Guide to Boosted Regression Trees. Journal of Animal Ecology 77:802-813.

Fry, J.A., M.J. Coan, C.G. Homer, D.K. Meyer, and J.D. Wickham, 2009. Completion of the National Land Cover Database (NLCD) 1992-2001 Land Cover Change Retrofit Product. U.S. Geological Survey Open-File Report 2009-1379.

Grace, J.B., 2006. Structural Equation Modeling and Natural Systems. Cambridge University Press, New York, 365 pp.

Hawkins, C.P., 2006. Quantifying Biological Integrity by Taxonomic Completeness: Its Utility in Regional and Global Assessments. Ecological Applications 16:1277-1294.

Hawkins, C.P., Y. Cao, and B. Roper, 2010. Method of Predicting Reference Condition Biota Affects the Performance and Interpretation of Ecological Indices. Freshwater Biology 55:1066-1085.

Herbst, D.B. and E.L. Silldorff, 2006. Comparison of the Performance of Different Bioassessment Methods: Similar Evaluations of Biological Integrity From Separate Programs and Procedures. Journal of the North American Benthological Society 25:513-530.

Hilderbrand, R.H., R.M. Utz, S.A. Stranko, and R.L. Raesly, 2010. Applying Thresholds to Forecast Potential Biodiversity Loss From Human Development. Journal of the North American Benthological Society 29:1009-1016.

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan, 2004. Development of a 2001 National Landcover Database for the United

States. Photogrammetric Engineering and Remote Sensing 70(7):829-840.

Hubler, S., 2008. Wadeable Stream Conditions in Oregon. Oregon Department of Environmental Quality Laboratory Division—Watershed Assessment Section, Portland, Oregon. DEQ07-LAB-0081-TR.

Kennen, J.G., M. Chang, and B.H. Tracy, 2005. Effects of Landscape Change on Fish Assemblage Structure in a Rapidly Growing Metropolitan Area in North Carolina, USA. *In*: Effects of Urbanization on Stream Ecosystems, L.R. Brown, R.H. Gray, R.M. Hughes, and M.R. Meador (Editors). American Fisheries Society Symposium 47: 39-52.

King, R.S. and M.E. Baker, 2010. Considerations for Analyzing Ecological Community Thresholds in Response to Anthropogenic Environmental Gradients. Journal of the North American Benthological Society 29:998-1008.

Leathwick, J.R., D. Rowe, J. Richardson, J. Elith, and T. Hastie, 2005. Using Multivariate Adaptive Regression Splines to Predict the Distributions of New Zealand's Freshwater Diadromous Fish. Freshwater Biology 50:2034-2052.

Maloney, K.O., D.E. Weller, M.J. Russell, and T. Hothorn, 2009. Classifying the Biological Condition of Small Streams: An Example Using Benthic Macroinvertebrates. Journal of the North American Benthological Society 28:869-884.

Morgan, R.P. and S.F. Cushman, 2005. Urbanization Effects on Stream Fish Assemblages in Maryland, USA. Journal of the North American Benthological Society 24:643-655.

Moss, D., 2000. Evolution of Statistical Methods in RIVPACS. *In*: Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques, J.F. Wright, D.W. Sutcliffe, and M.T. Furse (Editors). Freshwater Biological Association, Ambleside, pp. 25-37.

Moulton, II, S.R., J.G. Kennen, R.M. Goldstein, and J.A. Hambrook, 2002. Revised Protocols for Sampling Algal, Invertebrate and Fish Communities as Part of the National Water-Quality Assessment Program. U.S. Geological Survey Open-File Report 02-150, 75 pp.

Oberdorff, T., D. Pont, B. Hugueny, and D. Chessel, 2001. A Probabilistic Model Characterizing Fish Assemblages of French Rivers: A Framework for Environmental Assessment. Freshwater Biology 46:399-415.

Ode, P.R., C.P. Hawkins, and R.D. Mazor, 2008. Comparability of Biological Assessments Derived From Predictive Models and Multimetric Indices of Increasing Geographic Scope. Journal of the North American Benthological Society 27:967-985.

Ode, P.R., A.C. Rehn, and J.T. May, 2005. A Quantitative Tool for Assessing the Integrity of Southern Coastal California Streams. Environmental Management 35:493-504.

Paul, M.J. and J.L. Meyer, 2001. Streams in the Urban Landscape. Annual Review of Ecology and Systematics 32:333-365.

Peck, D.V., A.T. Herlihy, B.H. Hill, R.M. Hughes, P.R. Kaufmann, D.J. Klemm, J.M. Lazorchak, F.H. McCormick, S.A. Peterson, S.A. Ringold, T. Magee, and M. Cappaert, 2006. Environmental Monitoring and Assessment Program – Surface Waters Western Pilot Study: Field Operations Manual for Wadeable Streams. EPA/620/R-06/003. Office of Research and Development, US EPA, Corvallis, Oregon.

Prasad, A.M., L.R. Iverson, and A. Liaw, 2006. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. Ecosystems 9:181-199.

Qian, S.S. and T.F. Cuffney, 2012. To Threshold or Not to Threshold? That's the Question. Ecological Indicators 15:1-9.

R Development Core Team, 2007. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Roy, A.H., M.C. Freeman, B.J. Freeman, S.J. Wenger, W.E. Ensign, and J.L. Meyer, 2005. Investigating Hydrologic Alteration as a Mechanism of Fish Assemblage Shifts in Urbanizing Streams. Journal of the North American Benthological Society 24:656-678.

Stanford, J.A., M.S. Lorang, and F.R. Hauer, 2005. The Shifting Habitat Mosaic of River Ecosystems. Verhandlungen Internationaler Vereinigung fur Limnologie 29:123-136.

Therneau, T.M. and E.J. Atkinson, 1997. An Introduction to Recursive Partitioning Using the RPART Routines. Technical report no. 61. Mayo Clinic, Rochester, MN, 52 pp.

Turak, E., S. Ferrier, T. Barrett, E. Mesley, M. Drielsma, G. Manion, G. Doyle, J. Stein, and G. Gordon, 2011. Planning for the Persistence of River Biodiversity: Exploring Alternative Futures Using Process-Based Models. Freshwater Biology 56:39-56.

Uhrich, M.A. and D.A. Wentz, 1999. Environmental Setting of the Willamette Basin, Oregon. U.S. Geological Survey Water-Resources Investigations Report 97-4082-A, 20 pp.

U.S. Census Bureau, 2000. Census 2000 Redistricting Data Summary File: U.S. Census Bureau Technical Documentation Public Law, 94-171, 223 pp.

U.S. Geological Survey, 2007, National Hydrography Dataset (High Resolution), Digital Data, http://nhd.usgs.gov/data.html, *accessed* November 2007.

Van Sickle, J., J. Baker, A. Herlihy, P. Bayley, S. Gregory, and P. Haggerty, 2004. Projecting the Biological Condition of Streams Under Alternative Scenarios of Human Land Use. Ecological Applications 14:368-380.

Vogelmann, J.E., S.M. Howard, L. Yang, C.R. Larson, B.K. Wylie, and N. Van Driel, 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 67:650-652.

Waite, I.R., L.R. Brown, J.G. Kennen, J.T. May, T.F. Cuffney, J.L. Orlando, and K.A. Jones, 2010. Comparison of Watershed Disturbance Predictive Models for Stream Benthic Macroinvertebrates for Three Distinct Ecoregions in Western US. Ecological Indicators 10:1125-1136.

Waite, I.R., S. Sobieszczyk, K.D. Carpenter, A.J. Arnsberg, H.M. Johnson, C.A. Hughes, M.J. Sarantou, and F.A. Rinella, 2008. Effects of Urbanization on Stream Ecosystems in the Willamette River Basin and Surrounding Area, Oregon and Washington. Chapter D of "Effects of Urbanization on Stream Ecosystems in six Metropolitan Areas of the United States". Scientific Investigations Report 2006-5101-D, 62 pp.

Walsh, C.J., T.D. Fletcher, and A.R. Ladson, 2005. Stream Restoration in Urban Catchments Through Redesignin Stormwater Systems: Looking to the Catchment to Save the Stream. Journal of the North American Benthological Society 24:690-705.

Wang, L.J., J. Lyons, P. Kanehl, and R. Bannerman, 2001. Impacts of Urbanization on Stream Habitat and Fish Across Multiple Scales. Environmental Management 28:255-266.

Wisseman, B., 1996. Version 1.0. Unpublished Ecological Coding Attributes for Freshwater Invertebrate Taxa in Western North America. Available from Aquatic Biology Associates, Inc., Corvallis, Oregon.