JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION

Vol. 50, No. 6 AMERICAN WATER RESOURCES ASSOCIATION December 2014

BEDDED SEDIMENT CONDITIONS AND MACROINVERTEBRATE RESPONSES IN NEW MEXICO STREAMS: A FIRST STEP IN ESTABLISHING SEDIMENT CRITERIA[1]

Benjamin K. Jessup, Philip R. Kaufmann, Forrest John, Lynette S. Guevara, and Seva Joseph[2]

ABSTRACT: Excess fine sediments in streambeds are among the most pervasive causes of degradation in streams of the United States. Simple criteria for acceptable streambed fines are elusive because streambed fines and biotic tolerances vary widely in the absence of human disturbances. In response to the need for sediment benchmarks that are protective of minimum aquatic life uses under the Clean Water Act, we undertook a case study using surveys of sediment, physical habitat, and macroinvertebrates from New Mexico streams. Our approach uses weight of evidence to develop suggested benchmarks for protective levels of surficial bedded sediments <0.06 mm (silt and finer) and <2.0 mm (sand and finer). We grouped streams into three ecoregions that were expected to produce similar naturally occurring streambed textures and patterns of response to human disturbances. Within ecoregions, we employed stressor response models to estimate fine sediment percentages and bed stability that are tolerated by resident macroinvertebrates. We then compared individual stream sediment data with distributions among least-disturbed reference sites to determine deviation from natural conditions, accounting for natural variability across ecoregion, gradient, and drainage area. This approach for developing benchmark values could be applied more widely to provide a solid basis for developing bedded sediment criteria and other protective management strategies in other regions.

(KEY TERMS: environmental impacts; Relative Bed Stability; channel morphology; sediment; invertebrates.)

Jessup, Benjamin K., Philip R. Kaufmann, Forrest John, Lynette S. Guevara, and Seva Joseph, 2014. Bedded Sediment Conditions and Macroinvertebrate Responses in New Mexico Streams: A First Step in Establishing Sediment Criteria. *Journal of the American Water Resources Association* (JAWRA) 50(6): 1558-1574. DOI: 10.1111/jawr.12224

INTRODUCTION

Excessive or inadequate amounts of fine sediments can be stressful to aquatic life in streams. The mechanisms by which biota are affected by excessive fine sediments include displacement of interstitial habitat space, clogging of water movement through sediments, decreased primary productivity, increased macroinvertebrate drift, abrasion or smothering of gills and other organs, and increased uptake of sediment-bound toxicants (Waters, 1995; Wood and Armitage, 1997; USEPA, 2006a). As human disturbance and development increases in stream catchments, increasing amounts of fine sediments are delivered to the stream, generated through bank

[1]Paper No. JAWRA-13-0151-P of the *Journal of the American Water Resources Association* (JAWRA). Received June 26, 2013; accepted April 3, 2014. **Discussions are open until six months from print publication**.

[2]Senior Ecologist (Jessup), Center for Ecological Sciences, Tetra Tech, Inc., 73 Main Street, Suite 38, Montpelier, Vermont 05602; Research Physical Scientist (Kaufmann), Western Ecology Division, U.S. EPA National Health and Environmental Effects Research Lab/ORD, Corvallis, Oregon 97333; Life Scientist (John), Region 6, U.S. EPA, Dallas, Texas 75202; and Assessment Coordinator (Guevara) and Environmental Specialist (Joseph), New Mexico Environment Department, Surface Water Quality Bureau, Santa Fe, New Mexico 87505 (E-Mail/Jessup: benjamin.jessup@tetratech.com).

erosion, or resuspended from existing bed sediments (Southerland *et al.*, 2002; Zaimes *et al.*, 2004; Burcher *et al.*, 2007). While these effects are well known, the process for establishing thresholds of sediment effects is rather new and evolving (Relyea *et al.*, 2000, 2012; USEPA, 2006a; Bryce *et al.*, 2008, 2010; Cormier *et al.*, 2008; Paul *et al.*, 2008; Jessup, 2009a).

In New Mexico, excessive fine sediments in streams and the threats they pose to aquatic life uses are a growing concern. The State of New Mexico Environment Department (NMED) protects against degradation of aquatic life uses in streams below levels that are observed in the majority of relatively undisturbed systems. NMED enforces narrative criteria that require suspended or settleable solids from other than natural causes shall not be present in surface waters of the state in quantities that damage or impair the normal growth, function, or reproduction of aquatic life (New Mexico Administrative Code 20.6.4.13). The degrees to which given quantities of sediment are unnatural or detrimental are as yet loosely defined by the criteria. Aquatic life protection was the impetus for translating the narrative criteria into numeric sediment benchmarks that discerned natural and tolerable conditions from disturbed and intolerable conditions. The benchmarks will be considered as NMED progresses toward numeric criteria for sediments.

Fine sediments (<2 mm diameter) are natural components of streams that are present even in pristine settings to which stream organisms have evolved and adapted. Therefore, the detection of alterations in the amount of sediment is more difficult than detecting an absolute concentration or percentage that represents a clear biological impact (as happens with introduced pollutants such as pesticides). When an absolute sediment measure is assessed, the benchmark of protection or impairment should be established in the context of sites with similar sediment regimes, or of the same site class. Because sediments are known to accumulate and mobilize according to physical stream properties (Leopold *et al.*, 1964; Buffington and Montgomery, 1999), comparisons of existing substrates to the morphology and hydraulic characteristics of the stream can yield a sediment measure adjusted to the specific site (Kaufmann *et al.*, 2008). To describe bedded sediment conditions in New Mexico streams, we looked at background sediment conditions in the context of both site classes and specific channel effects on stream sediments in relatively undisturbed streams. Relatively undisturbed streams were defined using criteria for land uses, road densities, dams in the catchment, water quality, and riparian conditions. By describing the sediment conditions in undisturbed stream sites and the biological responses as conditions began to depart from the reference, we were able to suggest numeric benchmarks that NMED could consider as translators for their narrative sediment criteria. An established multimetric index of benthic macroinvertebrate condition (Jacobi *et al.*, 2006) and other individual metrics that are responsive to sediment conditions were used to assess aquatic life use attainment in New Mexico streams.

We focused on three bedded sediment indicators to characterize the sediment conditions in New Mexico streams. Surveys of streambed particle sizes and channel morphology were used to estimate percentages of fine particles and the diameter of the bed sediments that are mobilized during bankfull storm flows. The sediment indicators included in our analysis were the Relative Bed Stability (RBS) index (Kaufmann *et al.*, 2008) and the areal percentages of fines (particles < 0.06 mm) and sand and fines (<2 mm) on the surface of the wetted portion of the streambed.

The objectives of this study were to describe a process for developing benchmarks for nonconventional stressors that are also natural stream variables. It provides a means to develop stressor response models, to evaluate them using a comparative weight of evidence, and illustrates the process in three distinctly different geographic areas. The case study describes bedded sediment indicators in New Mexico streams in relation to degrees of human disturbance, to natural variability in landscape settings and stream channel morphology, and to biological conditions. The relationships were used to establish expectations for sediment indicator values that would reflect sediment conditions in relatively undisturbed streams, account for natural variability across stream types and settings, and be protective of minimum aquatic life use requirements. Our suggested sediment benchmarks based on evidence from multiple analytical approaches provide NMED a solid basis for selecting final sediment benchmarks and other protective management strategies. The assessment approach described here follows the United States Environmental Protection Agency (USEPA) framework for establishing sediment criteria (USEPA, 2006a), and is intended as an illustration of an approach that could be applied more broadly.

METHODS

Analytical Approach

The general approach to identifying numeric translator values (benchmarks) for narrative sediment criteria included characterizing sediment conditions in undisturbed (reference) streams, classifying sites by environmental and sediment characteristics, and relating sediment qualities to biological condition in a stressor response analysis. Benchmarks were recommended based on the weight of evidence from three analyses: reference distributions (USEPA, 2000), quantile regression (Cade *et al.*, 1999), and changepoint analysis (Qian *et al.*, 2003). Reference distributions describe expectations in least-disturbed sites, which were classified by natural site types to reduce variability. Quantile regression and changepoint analysis compare sediment conditions with biological conditions, using the biological conditions to indicate the degree to which aquatic life uses are supported. The points at which sediment indicators became clearly and consistently associated with poor biological conditions were identified as potential numeric benchmarks.

Data Sources

Multiple sediment and biological datasets were used in the analysis of bedded sediment conditions and effects (Table 1, Figure 1). Data were collected from 229 sites by NMED or neighboring state agencies using methods of the USEPA Environmental Monitoring and Assessment Program (EMAP) described by Peck *et al.* (2006), with data reduction as described by Kaufmann *et al.* (1999, 2008). Data from adjacent states were used when sites were in a level 3 ecoregion (Griffith *et al.*, 2006) that was contiguous with those in New Mexico and the site was within 80 kilometers of the state border.

Relative Bed Stability (Kaufmann *et al.*, 2008) is a measure of the stability of surficial bed particles during defined flow conditions. RBS is calculated as the ratio of the mean or median surficial bed particle diameter in a stream reach to the critical diameter, the calculated maximum diameter of bed particles that could be mobilized by bed shear stress experienced during a defined flood condition. In our case, we assumed the standard flood condition occurred on average every 1.5 years and was represented by the

TABLE 1. Sediment Datasets Used in Analyses.

Dataset	Sampling Years
EMAP West	1999-2004
EPA Wadeable Streams Assessment	2004
Regional EMAP Arizona Streams	2007
Regional EMAP New Mexico	1999-2001, 2006-2007
Regional EMAP EPA Region 8 Colorado	1994-1995

Note: EMAP, Environmental Monitoring and Assessment Program.

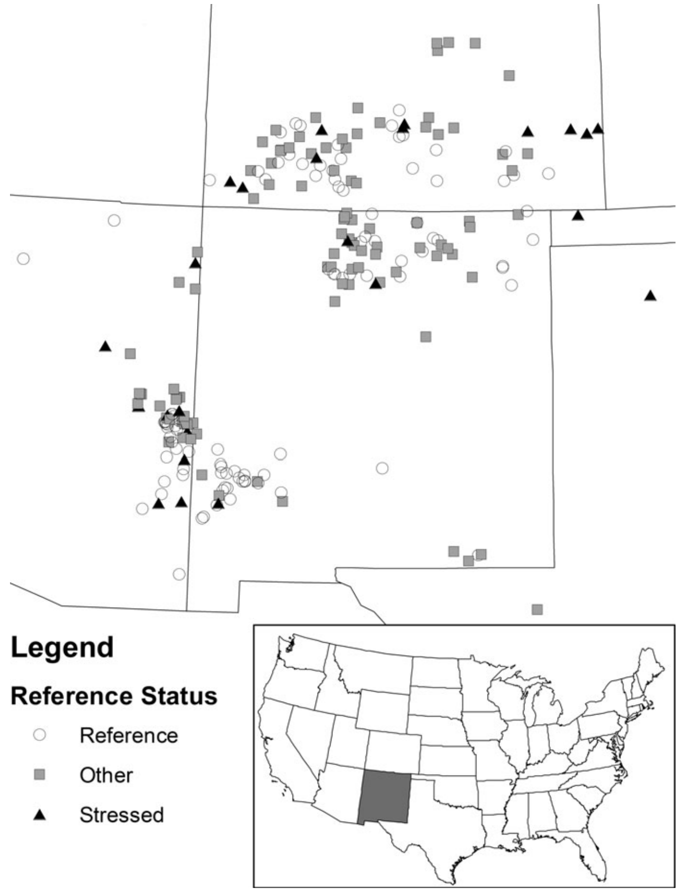

FIGURE 1. Site Locations and Reference Status of Bedded Sediment Sites in New Mexico and Surrounding Areas.

bankfull flow stage estimated from field indicators (Dunne and Leopold, 1978; Faustini *et al.*, 2009). Geometric mean particle diameter was determined using a reach-wide pebble count (Peck *et al.*, 2006) within the wetted portion of the channel and calculated from those data as described by Faustini and Kaufmann (2007). Bankfull critical particle diameter was calculated from slope, bankfull channel dimensions, and hydraulic roughness represented by residual pools and woody debris volumes (Kaufmann *et al.*, 2008). The measure is expressed as a logarithm of the ratio of geometric mean to critical particle size and is abbreviated as "LRBS". Because bedrock is sometimes a substantial portion of the streambed and will never be mobilized, it was removed from the percentage calculations to yield an LRBS excluding bedrock (LRBS_NOR). Percentages of fines (PCT_FN) and sand & fines (PCT_SAFN) were measured by systematically sampling 105 surface particles for a streambed pebble count (Peck *et al.*, 2006) and calculating the percentage of particles in each size category. The smaller particles were estimated based on either a visually distinct particle smaller than

2 mm (sand) or a smooth texture when rubbed between the fingers (fines).

Benthic macroinvertebrate samples were systematically collected from all habitat types in the stream reach in proportion to the areal proportion of those habitats (Peck *et al.*, 2006) using a rectangular net positioned downstream of stream substrates to capture organisms dislodged through agitation. Benthic samples were the basis for calculation of the New Mexico Macroinvertebrate Stream Condition Index (M-SCI) (Jacobi *et al.*, 2006) and other metrics that were components of the M-SCI or otherwise believed to be responsive to sediment stresses in New Mexico streams.

Geographic Information Systems (GIS) analyses were conducted to characterize land use, geology, and climatic conditions at each site and in the site catchments. Human disturbance variables analyzed at each site included field observations of near-stream human activities as well as basin land use, land cover, dams, road density, and road-stream intersections. The natural characteristics of sites and catchments included catchment area, stream reach slope, land slope in the catchment and at the site, level 3 and 4 ecoregion designations, stream order, site elevation, soil permeability, precipitation, and geologic type. For more information on data sources and GIS analysis techniques, see Jessup *et al.* (2010).

Reference Sites

Indicators of the sediment and biological conditions at sites were expected to be correlated with the intensities of disturbance. Sediment and ecological conditions at sites with minimal evidence of catchment or reach-scale disturbance represented our best approximation of the potential conditions in other sites of the same type (Stoddard *et al.*, 2006). Sediment and biological conditions at these sites defined the reference conditions for comparisons with other sites for both site classification and detection of impairment.

Site reference status was determined by combining two approaches, similar to methods used by Herlihy *et al.* (2008). First, we developed and applied reference criteria from the site data and second, we adopted reference designations already associated with each dataset. The criteria were modified from those used in Western-EMAP (Stoddard *et al.*, 2005a) or developed from distributions of the GIS data. The Western-EMAP reference and stressed site criteria included screening levels for chloride, total phosphorus, total nitrogen, and three measures of riparian disturbance: all disturbances, agricultural disturbances, and crop-related disturbances (Table 2). These criteria were region-specific, where regions were defined by aggregated ecoregions (Griffith *et al.*, 2006). Criteria related to suspended or bedded sediment concentrations were not used to define site reference status to avoid circularity when assessing sediment endpoints. Criteria were also developed from GIS data on land uses, road densities in the site catchments, and dams in the site catchment (Table 3). The distributions of values for GIS measures were used to guide selection of reference and stressed site criteria, as described by Stoddard *et al.* (2005a). In our application, we tallied a positive point for passing each reference criterion and a negative point for each stressed site indication. We summed the points for all the criteria and classified potential reference or stressed sites if the total points were ≥ 3 or <0, respectively. Sites were not considered of reference quality if they failed any of the stressed criteria.

Existing designations of reference sites from three projects included the NMED benthic multi-metric

TABLE 2. Anthropogenic Disturbance Screening Criteria Used in the National Wadeable Streams Assessment to Characterize Least-Disturbed (reference) and Most-Disturbed (stressed) Stream Reach Sample Sites (Stoddard *et al.*, 2005a). Observed values less than the first number (before) indicate reference conditions for that variable. Values greater than the second value indicate stress. Regions are as defined for Environmental Monitoring and Assessment Program.

Region	Ecoregion[1]	Chloride (mg/l)	Total Phosphorus (μg/l)	Total Nitrogen (μg/l)	Riparian Disturb. (W1_HALL)[2]	Riparian Disturb. (W1_HAG)[2]	Riparian Disturb. (W1H_CROP)[2]
MT-SW	23	300/1,000	50/200	750/1,000	1.25/2.5	0.25/0.6	0.2/0.5
MT-So. Rockies	21	200/1,000	25/200	750/1,000	1.25/2.5	1.0/1.4	0.2/0.5
PL-No. cultivated	25	1,000/2,750	200/900	2,000/4,000	—	0.6/1.4	0.15/0.25
PL-range	26	1,000/3,000	200/900	1,000/3,000	—	0.6/1.4	0.15/0.25
Xeric	20, 22, 24, 79	1,000/2,500	50/300	1,000/4,000	1.25/2.5	0.6/1.4	0.15/0.25

[1]Level 3 ecoregions are as follows: The Colorado Plateaus (#20), Southern Rockies (#21), Arizona/New Mexico Plateau (#22), Arizona/New Mexico Mountains (#23), Chihuahuan Deserts (#24), High Plains (#25), Southwestern Tablelands (#26), and Madrean Archipelago (#79).

[2]Proximity-weighted tally (Kaufmann *et al.*, 1999) of all riparian and near-stream human disturbances (W1_HALL), agricultural disturbances only (W1_HAG), or row-crop agricultural disturbances only (W1H_CROP).

TABLE 3. Reference and Stressed Criteria for Geographic Information Systems Variables.

Variable	Reference Threshold	Stressed Threshold
Natural land uses	>99%	<90%
Road density	<0.2 km/km^2	>0.7 km/km^2
Road crossing density	<0.1/km^2	>1.5/km^2
Dam density	<0.05 dams/km^2	>0.05 dams/km^2

index development (Jacobi *et al.*, 2006), NMED benthic predictive model development (Paul 2008), and Colorado Department of Environment and Public Health (CDPHE) multimetric index development (Jessup, 2009b). The existing designations were scrutinized for consistency among studies and with the numeric criteria in Tables 2 and 3. We also reviewed all potential reference and stressed sites with aerial imagery (GoogleEarth, available at: http://earth.google.com/) to identify any major misclassifications that were evident from land use patterns. Sites that showed ambivalent or contradictory reference characteristics and uncertain imagery were called "Other". The "Other" category also includes sites with an intermediate level of disturbance or insufficient data for categorization.

Site Classification

Site classification is the process by which natural gradients among sites are examined to identify sites with similar sediment supply, transport competence, and response to disturbance. The purpose of classification is to minimize within-class natural variability in indicators so that anthropogenic disturbance can be recognized with less background noise (Hughes, 1995; Barbour *et al.*, 1999). For sediment indicators, the stream characteristics that were expected to determine basic substrate conditions included channel dimensions, stream slope, catchment size, basin lithology, precipitation, and potential natural vegetation. Most of these variables are also related to categorical level 3 and 4 ecoregions (Griffith *et al.*, 2006). Potential site classification variables, sediment indicators, and biological variables were analyzed simultaneously to identify patterns of covariance that would suggest how sediment conditions could be classified according to environmental and biological characteristics.

Even though our aim was to classify sites based on nonanthropogenic characteristics, we examined patterns of response to human activities in the full range of sites. The sensitivity of ecological response to human disturbance is an inherent, or "natural" quality of streams that is detectable only by examining that response. For instance, Kaufmann *et al.* (2009) reported that reference sites in erodible and resistant lithologies in Pacific Northwest United States streams had similar RBS and amounts of fine sediments. However, with equal amounts of human activity, the sites draining erodible lithology experienced much greater sediment instability and excess fine sedimentation. This ecological response to disturbance should be recognized so that criteria are correctly set in areas of differing sensitivity. Our classification analysis included several methods, including principal components analysis (PCA), correlation analysis, and examination of biplots and distributions.

PCA was used as a primary tool for selecting site classification variables. We included all sites in the analysis and used natural, stressor, and indicator variables with continuous value distributions as the primary determinants. The ecoregion designations were not used in PCA because the values are categorical, not continuous. After establishing the PCA axes based solely on environmental variables, we examined the associations between biological metrics and those axes. We examined PCA axes that were correlated with sediment and biological variables to gain insight into potential scaling or classification variables that could be used to minimize natural variability in the sediment and biological endpoints. Variables were transformed as needed to approximate normal distributions using logarithmic and arcsine squareroot transformations.

Correlation analysis was used to describe single factor relationships between sediment and environmental variables in reference sites. The Pearson product-moment correlation coefficient was calculated for a matrix of individual variables (sediment, natural, stressor, and benthic metrics) transformed as noted for the PCA. In contrast to the PCA, the correlation analysis was limited to reference sites to emphasize the effects of natural site conditions instead of disturbance levels.

The relationships that were suggested by PCA and correlations were examined in box plots and biplots. The distributions of variables in ecoregions were examined, especially for sediment variables and those environmental variables that were shown to be important determinants of sediment conditions with PCA or correlation analysis. Biplots were used to show patterns of relationships between variables and to highlight attributes of the relationships such as reference status, ecoregion, or other covariates.

Sediment Indicators

Quantiles of the reference distributions of sediment variables (LRBS_NOR, PCT_FN, and PCT_SAFN) were used to indicate the levels of sediment fining that are likely to occur without and with substantial anthropogenic disturbances. Below the median, indicator values (e.g., PCT_SAFN) are similar to the best half of the reference sites. Values above the 95th quantile are unlike most of the reference sites and clearly indicate excessive sedimentation. Typically, quartiles or other quantiles are used to define similarity to or separation from the reference condition (Barbour *et al.*, 1999; USEPA, 2000, 2006a). To identify potential sediment benchmarks, we used the 75th and 90th quantiles for the percentage indicators that increase with disturbance and the 10th and 25th quantiles for the LRBS_NOR indicator that decreases with increasing disturbance.

We also looked at the separation of reference and stressed sediment indicator values to determine whether the quantiles were valuable for discriminating stressed sites. The statistic used to determine separation of values was the discrimination efficiency (DE) (Flotemersch *et al.*, 2006), measured as the percentage of stressed sites with indicator values greater than the 75th quantile for fine sediment indicators or less than the 25th quantile for LRBS_NOR at reference sites.

Biological Responses to Sediment Conditions

Our recommendations for sediment indicator benchmarks in each site class were partially based on corroborated results from stressor response analyses. Biological responses to sediment conditions were interpreted from quantile regressions, changepoint analyses, and locally weighted sequential smoothing (LOWESS) regressions (Cleveland, 1979) comparing the three sediment indicators with selected benthic macroinvertebrate metrics and the M-SCI (Table 4). The selected metrics were components of the M-SCI, commonly used in biological assessments, or sensitive to substrate characteristics.

Quantile regression is a method for estimating relationships between variables for a selected quantile, often the upper or lower boundary of a distribution of stressor response data points (Cade *et al.*, 1999). The quantile regression line represents biological potential (plotted on the *y*-axis) in relation to the stressor of interest (plotted on the *x*-axis). When limiting factors such as sediments act as constraints on organisms, the potential maximum biological condition is observed as a sloping line on a wedge-shaped scatter plot of a biological metric against a sediment variable. Points that are not along the slope of the wedge represent sites where biological condition is diminished by factors not represented on the *x*-axis (Bryce *et al.*, 2010), which in this case is fine sediments. We used "R" software (R Development Core Team, 2010) and the "quantreg" package to estimate limiting relationships by quantile regression. Several upper quantile regression lines (75th, 85th, 90th, and 95th) were calculated and plotted. When the upper quantiles are relatively parallel, the biological potential is likely limited by the stressor variable (Cade *et al.*, 1999; Bryce *et al.*, 2008). The multiple upper quantiles were examined and indications of limiting effects were determined based on the consistency of the slopes. When the 90th quantile regression line was consistent with the other lines, it was plotted to illustrate the change in the biological resource for each increment of sediment disturbance.

In changepoint analysis, a potential threshold is identified as the point along an environmental gradient (sediment indicator) at which there is a high degree of change in the response variable (biological metric). For each biological metric, the data were successively divided into two groups, above and below a series of potential sediment benchmarks on the *x*-axis. With nonparametric deviance reduction (King and Richardson, 2003; Qian *et al.*, 2003), the sediment changepoint was identified as the point along the *x*-axis where the metric values above and below the benchmark had low variability within groups and high difference among groups. Changepoints were calculated using the "chngp.nonpar" package in R.

One caveat of the changepoint analysis is that a changepoint may be identified, and even determined to be statistically significant, when the changepoint value

TABLE 4. Benthic Macroinvertebrate Metrics Used for Biological Conditions.

Metric	Description	Source
M-SCI	Macroinvertebrate Stream Condition Index: a numeric combination of sensitive metrics	Jacobi *et al.* (2006)
Total taxa	All unique taxa in the sample	Barbour *et al.* (1999)
EPT taxa	Count of taxa in the Ephemeroptera, Plecoptera, and Trichoptera insect orders	Jacobi *et al.* (2006)
Ephemeroptera taxa	Count of Ephemeroptera taxa	Jacobi *et al.* (2006)
Clinger taxa	Count of taxa that maintain position in the stream by grasping stable substrates	Jacobi *et al.* (2006)
% Sensitive EPT	Percentage of EPT individuals, excluding the tolerant Hydropsychidae family	Jacobi *et al.* (2006)
Hilsenhoff's Biotic Index	The average tolerance value of all individuals in the sample	Hilsenhoff (1987)
Intolerant taxa	Count of taxa that are sensitive to pollution	Barbour *et al.* (1999)

is actually only an artifact of the available data and not an indication of a change in biotic response to sediment (Daily *et al.*, 2012; Qian and Cuffney, 2012). LOWESS regressions were used to show trends at local portions of the gradients and to identify possible changepoints as deflections in the regression line. In our analyses, we evaluated possible changepoints by examining the LOWESS regression line on biplots of biological metrics and sediment indicators. If the LOWESS fit did not coincide with a changepoint, then the value identified through changepoint analysis was disregarded.

RESULTS

Reference Site Identification

Our criteria and screening process identified 99 reference and 25 stressed sites from among the 229 sites used for the bedded sediment analysis (Figure 1). Most of the reference sites were in mountainous ecoregions ($N = 55$) and fewer were in the foothills ($N = 27$) and xeric areas ($N = 17$). Conversely, the stressed sites were mostly the xeric ($N = 10$) and foothill ($N = 9$) regions, with fewer in the mountains ($N = 6$).

Classification

In the PCA, the first three factors explained 53% of the variability in the environmental condition variables (Table 5). Sediment indicators were most strongly related to the first axis, which was also related to catchment area, elevation, stream slope, precipitation, and the density of road crossings in the site catchment. Benthic macroinvertebrate metrics were also most strongly related to the first axis. The second axis was related to stream size and the third axis was related to site location (latitude and longitude).

Correlation analysis identified relationships among variable pairs, including some that reinforce the

TABLE 5. Principal Components Analysis Factor Scores of the Most Important Variables on the First Three Factors, Showing the Variance Explained by Each Factor. Scores with magnitudes >0.60 are shown in bold type and considered to be strong relationships.

Variable Code	Variable Description	F. 130%	F. 212%	F. 311%
LRBS_NOR	RBS without bedrock or hardpan (log10)	0.58	0.44	−0.18
asPCT_SAFN	% Sand and fine sediments at the site (arcsin(sqrt))	**−0.72**	−0.44	0.20
asPCT_FN	% Fine sediments at the site (arcsin(sqrt))	−0.58	−0.48	0.20
LRdX_km2	Road crossings per km^2 in the catchment	**−0.86**	0.37	0.10
LArea_km2	Catchment area (log10(km2))	**−0.84**	0.47	−0.04
LSTREAMSLOP	Stream slope (log10(%), NHD-Plus data)	**0.84**	−0.08	−0.09
ELEV_m	Site elevation (m)	**0.78**	−0.15	−0.20
STREAMORDE	Stream order (Strahler, map scale 1:100,000)	**−0.75**	0.50	0.12
Precip	Precipitation (cm)	**0.72**	0.14	−0.24
LXSLOPE	Stream slope (log10(%), field data)	**0.66**	0.07	−0.33
LPower	Stream power index (log10(Precip*Area_km^2*Xslope))	−0.50	**0.65**	−0.28
LXWIDTH	Average site wetted width (log10)	−0.46	**0.63**	−0.29
Point_X	Latitude of sample	−0.36	−0.26	**−0.63**
Point_Y	Longitude of sample	0.00	−0.36	**−0.73**
TotalTax	Total taxa (count)	0.42	0.20	0.04
EPTTax	EPT taxa (count)	0.46	0.39	−0.20
IntolTax	Number of taxa sensitive to pollution (count)	**0.65**	0.17	−0.23

Note: RBS, Relative Bed Stability; EPT, Ephemeroptera, Plecoptera, and Trichoptera.

TABLE 6. Correlations (Pearson *r*) of the Environmental Variables Most Strongly Related to the Sediment Variables in Reference Sites.

Variable Code	Variable Description	PCT_FN	PCT_SAFN	LRBS_NOR
ELEV_m	Site elevation (m)	−0.36	−0.51	0.36
Precip	Precipitation (cm)	−0.47	−0.64	0.46
LXSLOPE	Stream slope (log10(%), field data)	−0.58	−0.64	0.26
XCMGW	Riparian vegetation woody cover	−0.42	−0.41	0.44
LRdX_km2	Road crossings per km^2 in the catchment	0.37	0.45	−0.33
W1_HALL	Index of riparian disturbance	0.41	0.49	−0.40

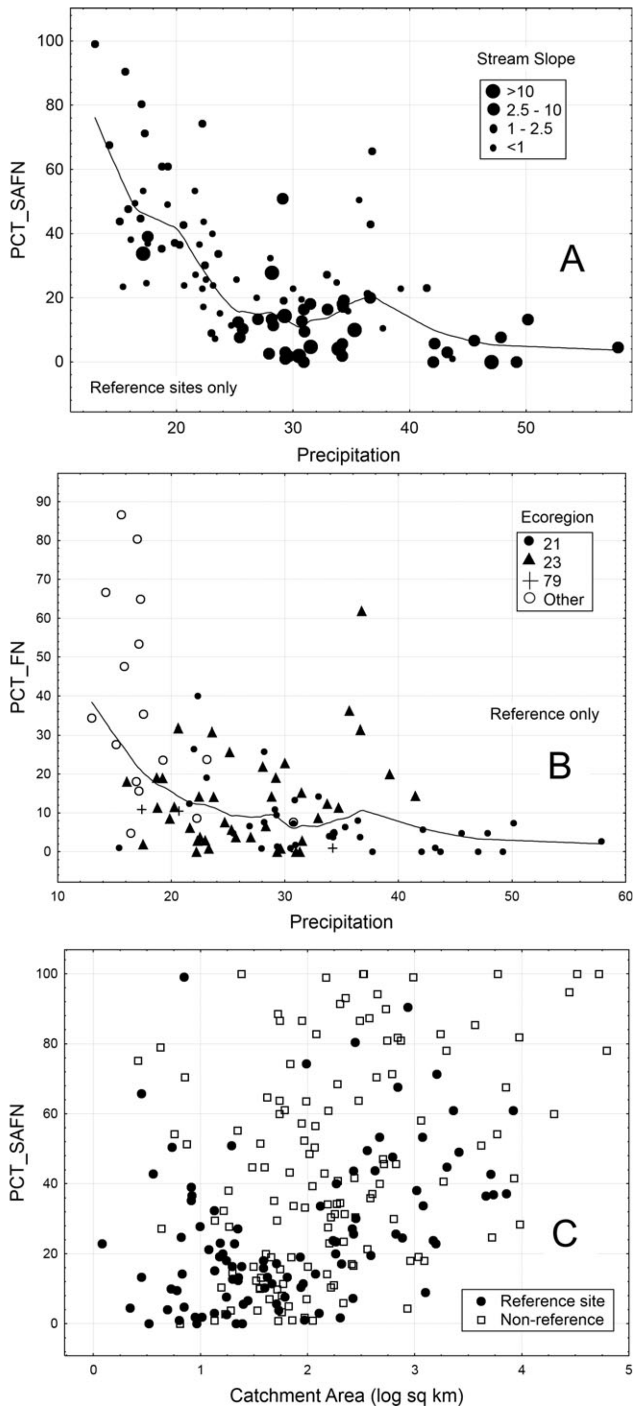

FIGURE 2. Relationships between Fine Particle Percentages and Influential Classification Variables: Precipitation, Slope, Ecoregion, and Catchment Size. The locally weighted sequential smoothing regression line is shown in relation to precipitation.

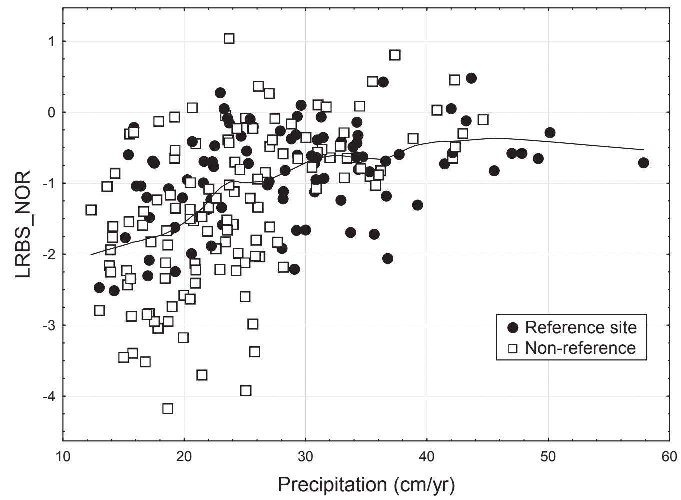

FIGURE 3. Relationships between LRBS_NOR and Precipitation, Showing the Locally Weighted Sequential Smoothing Regression Line.

results of the PCA. More variables were significantly related to PCT_SAFN than to PCT_FN or LRBS_NOR (Table 6). Among these were stream slope, precipitation, and elevation. The LRBS_NOR calculation accounts for stream slope and other channel morphology variables and is therefore less correlated with them. LRBS_NOR was more strongly related to precipitation, riparian vegetation, and riparian disturbance.

From the PCA and correlation analyses, it was clear that the primary classification variables were stream slope, precipitation, and elevation. Sites having lower precipitation (<25 cm/yr) and milder slopes (<2.5%) also have more fine sediments (>20 PCT_SAFN) (Figure 2). The Southern Rockies (#21), Arizona/New Mexico Mountains (#23), and Madrean Archipelago (#79) ecoregions had more precipitation and fewer fines compared to ecoregions in the plains, plateaus, and deserts. The relationship between PCT_SAFN and catchment size had a lot of scatter, though sites with catchments <100 km^2 had less fines than larger catchments. With LRBS_NOR, there was a difference in values above and below 20-25 cm/yr precipitation, as was observed with the other sediment indicators (Figure 3).

Values of all three sediment indicators were correlated with precipitation and woody vegetation, and the two fine sediment indicators were significantly correlated with slope (Table 6). Differences in these and other environmental characteristics are also captured by level 3 and level 4 ecoregions (Griffith *et al.*, 2006). Thus ecoregions were considered in developing a classification scheme that distinguished sediment expectations in streams and was relatively easy to conceptualize, communicate, and apply through mapping techniques. While a general difference between mountains and plains site types was obvious, an additional transitional class seemed appropriate and necessary. Distributions of site PCA factor scores in level 4 ecoregions were compared and when the ecoregions were grouped into Mountains, Foothills, and Xeric

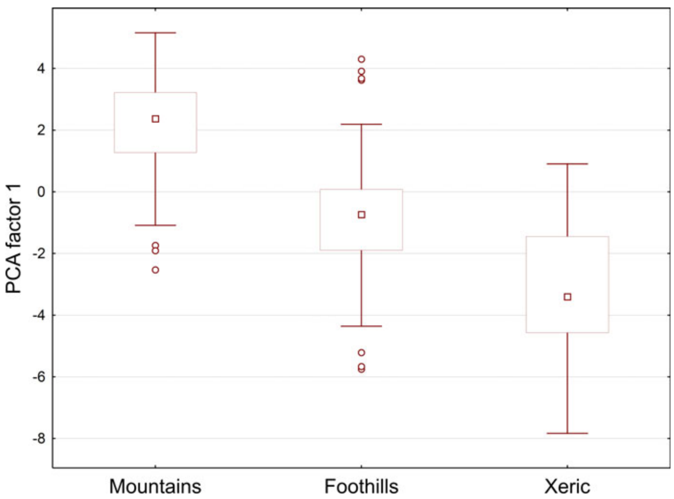

FIGURE 4. Principal Components Analysis (PCA) Factor 1 Scores in All Bedded Sediment Sites. The first axis was driven by sediment indicators, catchment area, elevation, stream slope, and precipitation.

areas, the PCA scores were homogenous within these classes (Figure 4). The nonoverlapping inter-quartile ranges suggest that the site classes have distinct environmental characteristics, especially regarding characteristics related to PCA factor 1; catchment area, stream slope, elevation, and precipitation. Level 4 ecoregions that were poorly represented in the data set were assigned to a class based on the ecoregional description. This classification recognizes the distinctions between high elevation, steep-sloped mountain streams; lower elevation streams in the drier foothills; and very low gradient streams in the still-drier plains and plateaus (Table 7 and Figure 5).

Sediment Reference Conditions

Sediment conditions in reference sites varied among site classes (Table 8). For the percentage measures, the 75th and 90th quantile values of reference were lowest in the Mountains, intermediate in the Foothills, and highest in the Xeric areas. The LRBS_NOR measure also showed patterns of increasing instability from the Mountains to the Foothills, and then the Xeric areas, as seen in the 25th quantile values. In the Xeric region, considerably more fine and unstable bed materials are evident when compared to the Mountains and Foothills.

Differences in indictor values among reference and stressed sites were greatest for the percentage indicators in the Foothills, where DE were at 75% or more (Table 8). Percent fines did not discriminate reference from stressed sites in the Xeric region. In each region, the LRBS_NOR was less discriminating than the PCT_SAFN indicator.

Biological Responses to Sediment Conditions

We looked for examples of strong relationships between the sediment indicators and the biological metrics using quantile regression, changepoint analysis, and LOWESS regression for all combinations of sediment indicators and benthic macroinvertebrate metrics. There were several cases where metrics appeared insensitive to sediment indicators, but no cases where the response of one metric contradicted response trends of other metrics. The most responsive metrics represented the most sensitive characteristics of the benthic macroinvertebrate assemblage that were worthy of protection through threshold setting. The unresponsive metrics were not used in setting thresholds. Metric responses are discussed in the following section, illustrated with selected biplots (Figures 6-8) and summarized in Table 9.

Mountains. Percent Sand and Fines. Seven of the eight biological metrics had consistent upper quantile regression slopes showing well-defined declines relating biota to progressive increases in sand and fine sediment. This suggested that the biological condition is limited by increasing percentages of sand & fine sediments in Mountain ecoregion streams. Based on the 90th quantile regression line, an increase in 20 PCT_SAFN results in a loss of five Ephemeroptera, Plecoptera, and Trichoptera (EPT) taxa. In four of the eight metrics, including EPT taxa (Figure 6A), the changepoint was identified near

TABLE 7. Definition of Sediment Site Classes, Including Names of the Level 3-4 Ecoregions Used to Define the Classes.

Site Class	Level 3 and 4 Ecoregions
Mountains	The Southern Rockies (#21) and the Arizona/New Mexico Mountains (#23), excluding specific level 4 ecoregions in the Foothills (21d, 23a, 23b, and 23e)
Foothills	Foothill Woodlands and Shrublands (21d), Chihuahuan Desert Slopes (23a), Madrean Lower Montane Woodlands (23b), Conifer Woodlands and Savannas (23e), and the Madrean Archipelago (#79)
Xeric	Colorado Plateaus (#20), Arizona/New Mexico Plateau (#22), Chihuahuan Deserts (#24), High Plains (#25), and Southwestern Tablelands (#26)

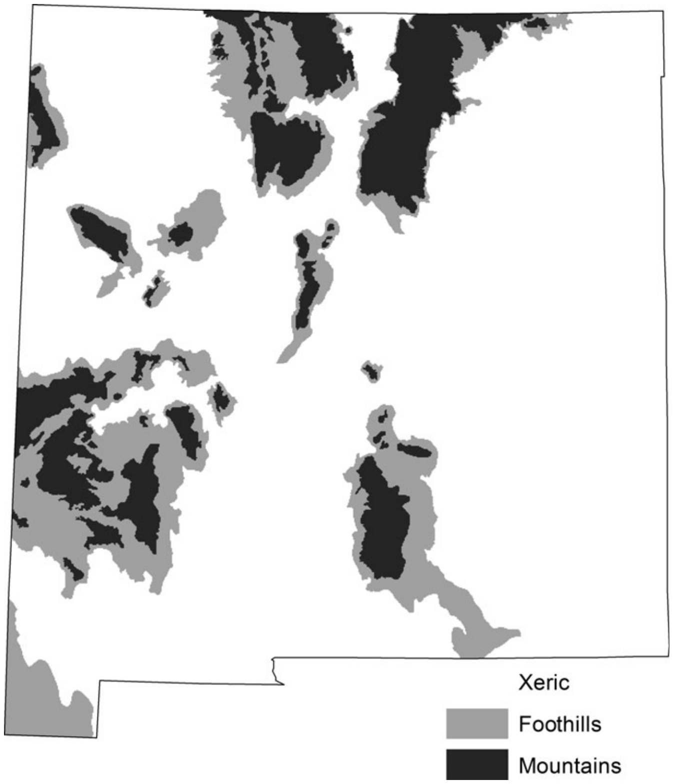

FIGURE 5. New Mexico Mountain, Foothills, and Xeric Site Classes.

20%. LOWESS regression lines also showed changes in slopes near 20 PCT_SAFN.

Percent Fines. Seven of the eight metrics had consistent upper quantile regression slopes, consistent with a negative response in the biological metrics with increasing percentages of fine sediments. On the basis of the 90th quantile regression line, an increase in 20 PCT_FN results in a loss of four EPT taxa. In all eight metrics, a significant changepoint was identified near 20%. This is greater than is indicated by the LOWESS regression shifts, which start at 10-15% (Figure 6B).

LRBS_NOR. Three of the eight metrics had parallel upper quantile regressions. Based on the 90th quantile regression line, a decrease of 0.5 LRBS_NOR units results in a loss of two EPT taxa. In four metrics, the changepoint was identified near −1.6 units. Changepoints identified for EPT taxa, clinger taxa, and the Hilsenhoff's Biotic Index (HBI) are higher, perhaps due to the more sensitive organisms represented in those metrics. The highest changepoint (near −1.0 units) was identified with the clinger taxa metric (Figure 6C). This coincided with changes indicated by the LOWESS regression lines for all metrics.

Foothills. Percent Sand and Fines. Six of the eight metrics had consistent upper quantile regression lines. Based on the 90th quantile regression line, an increase of 20 PCT_SAFN results in a loss of two macroinvertebrate taxa. In five of the eight metrics, a significant changepoint was identified in the range of 55-71%. The LOWESS regression lines showed changing trends at about 50 PCT_SAFN, as shown with EPT taxa (Figure 7A). At >50-60 PCT_SAFN, biological metrics show certain degradation. The quantile regression analysis and reference distributions suggest that effects could be occurring at lower levels.

Percent Fines. All of the metrics had consistent upper quantile regression lines. Based on the 90th quantile regression line, an increase of 20 PCT_FN results in a loss of one macroinvertebrate taxon or two EPT taxa. In five of the eight metrics, a significant changepoint was identified between 20 and 33 PCT_FN. This coincides with the LOWESS regression shifts, which start at about 20% (Figure 7B), even for those metrics with changepoints at higher levels. All reference sites had less than 32 PCT_FN.

TABLE 8. Bedded Sediment Indicator Statistics for Reference Sites in Three Site Classes.

Indicator	Ref N	Mean	Min	10%ile	25%ile	Median	75%ile	90%ile	Max	Std.Dev	DE (%)
PCT_SAFN											
Mountains	55	16	0	2.2	5.6	13.3	20.6	35.1	65.7	14.1	>50
Foothills	27	27.7	0	5.1	18.6	27.1	36.9	45.2	61	15.5	<75
Xeric	17	57.2	19.5	29.8	43.8	53.3	74.3	84.4	99	22.8	>50
PCT_FN											
Mountains	55	9.7	0	0	2.8	5.7	12.9	24.6	61.9	11.7	>50
Foothills	27	11.2	0	0.9	2.4	10.9	18.6	23.7	31.8	9.6	~75
Xeric	17	38.7	4.8	8.2	18.1	34.3	59	72.2	86.7	26	<25
LRBS_NOR											
Mountains	55	−0.8	−2.2	−1.5	−1.1	−0.8	−0.4	−0.1	0.4	0.5	25-50
Foothills	27	−0.9	−2	−1.7	−1.3	−1	−0.5	−0.1	0.2	0.6	~50
Xeric	15	−2	−2.9	−2.7	−2.5	−1.9	−1.6	−1.1	−0.7	0.7	25-50

Notes: DE, discrimination efficiency. Stressed site N = 6, 9, and 10 for the Mountains, Foothills, and Xeric areas, respectively.

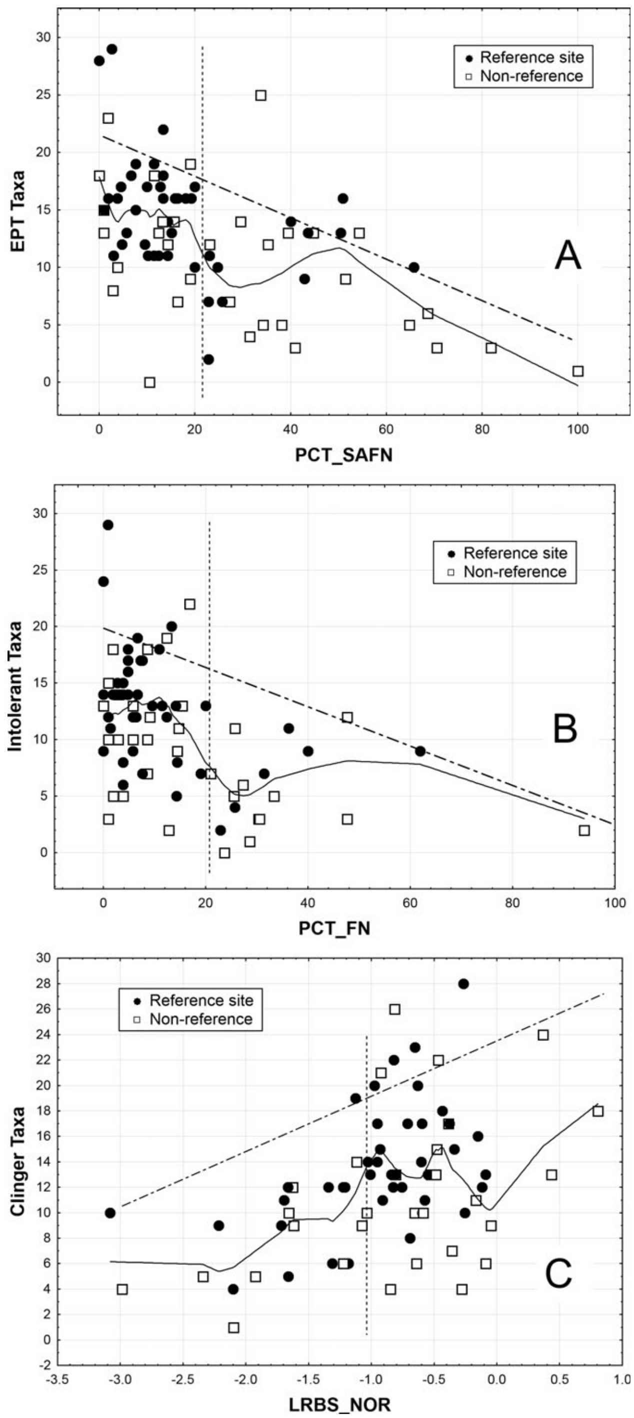

FIGURE 6. Benthic Macroinvertebrate Metrics in Relation to Sediment Indicators in the Mountain Site Class. Significant change points are shown as vertical lines, meaningful 90th quantile regression lines are shown as diagonal dash dot lines, locally weighted sequential smoothing regression lines are shown as nonlinear solid lines, reference points are shown as solid circles, and nonreference points are shown as open squares.

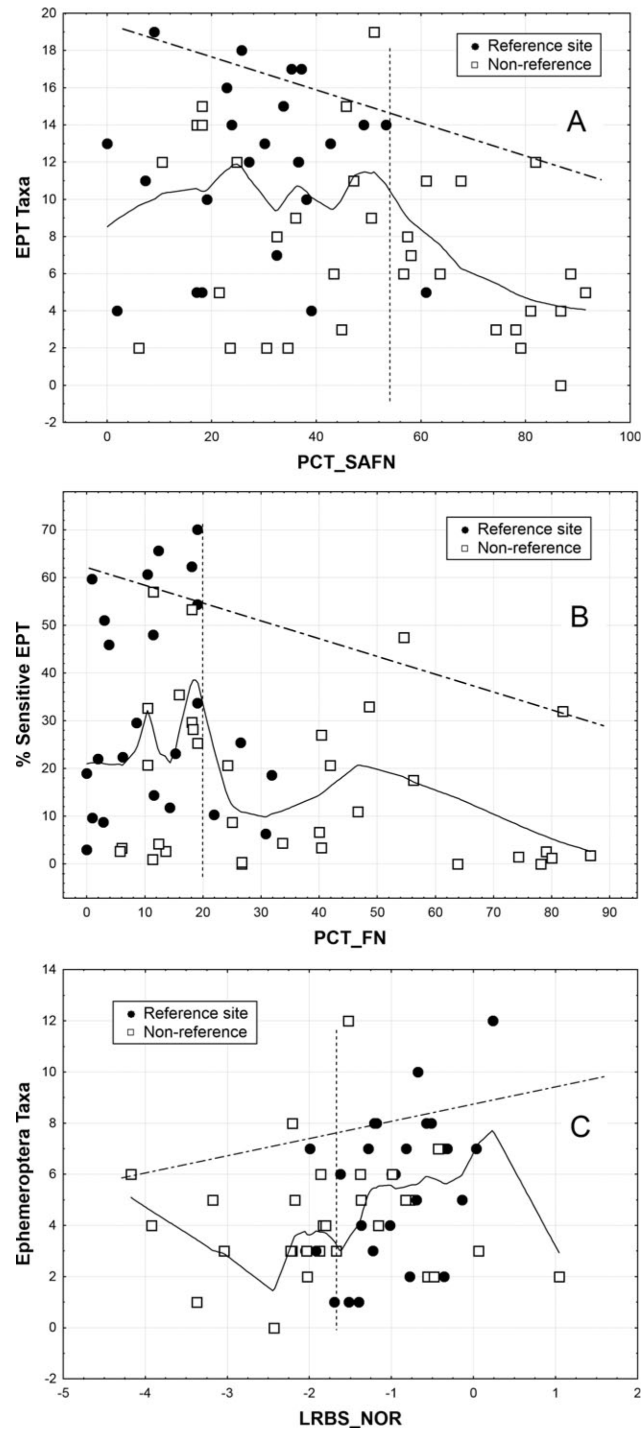

FIGURE 7. Benthic Macroinvertebrate Metrics in Relation to Sediment Indicators in the Foothill Site Class. Symbols are as described in Figure 6.

LRBS_NOR. Six of the eight metrics had consistent upper quantile regression lines, indicating that the biological condition is limited by decreasing bed stability. Based on the 90th quantile regression line, a decrease of 0.5 LRBS_NOR units results in a loss of two or three macroinvertebrate taxa. Changepoints were identified for four metrics, ranging from −2.2 to −1.3. The LOWESS regression lines were variable, though three showed a steeper descent starting at

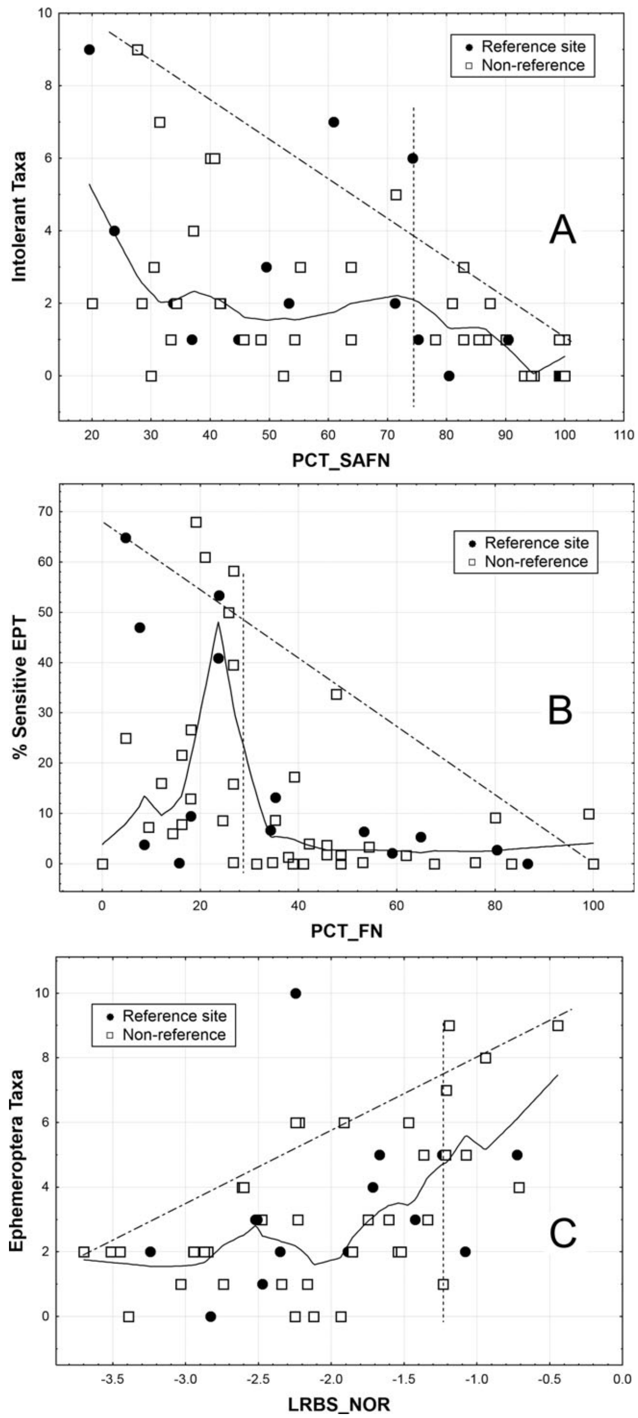

FIGURE 8. Benthic Macroinvertebrate Metrics in Relation to Sediment Indicators in the Xeric Site Class. Symbols are as described in Figure 6.

−1.3 units (Figure 7C). All reference sites have LRBS_NOR values greater than −2 units. The midpoint of the range of changepoints is −1.75 units.

Xeric Areas. Percent Sand and Fines. All of the metrics had consistent upper quantile regression lines. Based on the 90th quantile regression line, an increase of 20 PCT_SAFN results in a loss of five macroinvertebrate taxa in the Xeric areas. The range of changepoints was extremely broad, but five of the eight significant changepoint values were between 72 and 75%. The LOWESS regression lines did not show stark or consistent changes in trends. They were more or less gradual, except that EPT taxa and intolerant taxa had very level regression lines up until 74% (Figure 8A). The best indication from our analyses is that at >74 PCT_SAFN, biological potential is limited. The quantile regressions suggest that effects could be occurring at lower levels, but reference quantile values do not support lower benchmarks.

Percent Fines. Six of the eight metrics had consistent upper quantile regression slopes. On the basis of the 90th quantile regression line, an increase in 20 PCT_FN results in a loss of two EPT taxa. In five of the eight metrics, a significant changepoint was identified at 29%. This generally coincides with the LOWESS regression shifts, which have a peak at about 24% (Figure 8B).

LRBS_NOR. All of the metrics had consistent upper quantile regression slopes and significant changepoints. On the basis of the 90th quantile regression line, a decrease in 0.5 LRBS_NOR units results in a loss of three taxa. Most of the changepoints separated a few very good metric values from the remaining poorer values at LRBS_NOR values between −1.0 and −1.2 units. The lowest changepoint (for the HBI) was at −2.25. The changepoints generally coincide with the LOWESS regression descents (Figure 8C).

DISCUSSION

Potential Sediment Benchmarks

Benchmark values were recommended as translators of narrative sediment criteria for protecting stream resources and aquatic life from degradation. The benchmarks were selected to prevent certain degradation compared to most reference sites for the indicator values and obvious changes in at least two biological metric responses. They were not intended to protect against minimum effects, such as sediment indicator values that depart from median reference values or biological conditions that are slightly worse than the optimum observed metric values.

The sediment indicators that are most appropriate for assessments are PCT_SAFN and LRBS_NOR.

TABLE 9. Summary of Potential Benchmarks Based on Reference Distributions and Biological Changepoints. The reference quartiles show the 25th percentile for LRBS_NOR and the 75th percentile for PCT_SAFN and PCT_FN.

	Recommended Benchmark	Reference Quartile	NM-SCI	Total Taxa	EPT Taxa	Ephemeroptera Taxa	% Sensitive EPT	Clinger Taxa	HBI	Intolerant Taxa
Mountains										
PCT_SAFN	20	20.6	59.5	69.5	21.4	21.4	51.2	19.6		21.4
PCT_FN		12.9	19.5	17.9	19.5	19.5	19.5	19.5	19.5	20.5
LRBS_NOR	−1.1	−1.10	−1.64		−1.22	−1.64	−1.64	−1.15		−1.22
Foothills										
PCT_SAFN	37	36.9	62.3		55.0		71.0	59.5		71.0
PCT_FN		18.6	20.5		32.7		20.5	32.7	26.6	
LRBS_NOR	−1.3	−1.30			−1.37		−1.90	−2.21		
Xeric										
PCT_SAFN		74.3	72.8	99.5	74.8	72.8	28.0	87.0	74.8	74.8
PCT_FN	29	59.0			29.1	29.1	29.1		29.1	29.1
LRBS_NOR	−2.5	−2.50	−1.21	−1.01	−1.20	−1.22	−1.39	−1.20	−2.25	−1.01

Note: NM-SCI, New Mexico Stream Condition Index; EPT, Ephemeroptera, Plecoptera, and Trichoptera; HBI, Hilsenhoff's Biotic Index.

Less emphasis was placed on percent fines because it is already included in the measure of percent sand & fines and is more variable across sites, and because both sand & fines were relatively rare in streams with normal to powerful flows. However, in Xeric sites, the sand component was relatively common and the fines may take on more importance. From the two types of analyses, biological responses and reference distributions, we derived a set of summary statistics and recommended benchmarks (Table 9). In some cases, as in the Mountain site class, the agreement between the two approaches is remarkably consistent. In the Foothills and Xeric areas, there is a greater difference in benchmarks resulting from the two analytical approaches.

Mountains. The corroborating analyses in the Mountains suggested benchmarks of 20 PCT_SAFN and −1.1 LRBS_NOR units. At these benchmarks, biological conditions are generally good on one side and beginning to worsen on the other. In the analytical dataset, 67% of the Mountain sites had PCT_SAFN <20%, including 71% of reference sites and 50% of stressed sites. Seventy-six percent of sites had LRBS_NOR values greater than −1.1 units, including 75% of reference sites and 67% of stressed sites. From the stressor response analysis, it appears that biological impairment is nearly certain at observations above 35 PCT_SAFN or below −1.5 LRBS_NOR.

The benchmarks we recommend are comparable to those suggested by other authors (Bryce *et al.*, 2010; Relyea *et al.*, 2012) for sediment-sensitive macroinvertebrates in western mountain streams. These others used data collected with similar reference site designations, identical field methods, and from a larger geographic area. Bryce *et al.* (2010) found that the 75th quantiles were 17 PCT_SAFN and 5 PCT_FN in a broad study of western mountain streams. When comparing sediment conditions to biota, these other authors emphasized protection of the most sediment-sensitive macroinvertebrates. Their analyses suggested that benchmarks could be set at 3 PCT_FN (Bryce *et al.*, 2010) and 10 PCT_SAFN (Bryce *et al.*, 2010; Relyea *et al.*, 2012). Relyea *et al.* (2012) found that all taxa could tolerate up to 10 PCT_SAFN. Taxa designated as "extremely sensitive" made up 5% of the 206 taxa analyzed and were impaired at 10-20 PCT_SAFN. For protection of aquatic vertebrates, Bryce *et al.* (2010) recommended 13 PCT_SAFN and 5 PCT_FN.

Reference quantiles suggest somewhat more stable substrates in reference streams for comparable studies in western mountain streams (Kaufmann *et al.*, 2012) and southwestern mountain streams (Stoddard *et al.*, 2005b). The 25th and 5th quantiles in those studies were −0.64 and −1.29, respectively (Stoddard *et al.*, 2005b) and −0.99 and −1.68, respectively (Kaufmann *et al.*, 2012). In the Pacific Northwest, assessment index values for fish and amphibians showed declining conditions at LRBS values less than −1.0 (Kaufmann and Hughes, 2006).

Foothills. In the Foothills, benchmarks of 37 PCT_SAFN and −1.3 LRBS_NOR units are recommended. These benchmarks are based on the reference quartiles, which are more protective than the benchmarks derived from biological responses (Table 9). In the analytical dataset, 52% of the Foothill sites had PCT_SAFN <37%, including 74% of reference sites and 44% of stressed sites. Fifty-one percent of sites had LRBS_NOR values greater than −1.3 units, including 78% of reference sites and 44% of stressed sites.

We did not recommend several less-protective potential benchmarks in the Foothills. Some of the metrics had higher changepoints for PCT_SAFN, with ranges from 55 to 71%, much higher than the 90th quantile of reference data (45%). Observations of PCT_SAFN above 45% or below −1.7 LRBS_NOR indicate certain degradation of sediment conditions in the Foothills.

Xeric Areas. Sand is a common substrate in the Xeric sites. The 75th quantile of reference and the biological responses suggest that up to 74 PCT_SAFN may be natural and tolerable. The 25th quantile of reference LRBS_NOR values was −2.5, which is lower than the value indicated by changepoints (−1.25 units). At −1.25 LRBS_NOR units, more than 71% of reference sites would fail the criterion. The lack of agreement between the reference distribution and biological effects for LRBS_NOR does not support selection of either for setting benchmarks in a weight of evidence approach. Other estimates of LRBS reference quantiles in southwestern xeric streams are higher in comparison to our observations, with 25th and 10th quantiles of −0.94 and −1.69, respectively (Stoddard *et al.*, 2005b).

A valid alternative indicator of sediment stress is PCT_FN, representing a particle fraction that is less common than sand in the Xeric stream substrates. A biological changepoint at 29 PCT_FN was consistent among metrics. However, the distributions of PCT_FN are similar for reference and stressed sites in the Xeric areas and the biological changepoint is not corroborated by the reference quantiles. In the analytical dataset, 47% of reference Xeric sites had <29 PCT_FN, as did 60% of the stressed sites. Site and catchment stressors used to define reference and stressed sites may not affect sediment supply and dynamics in Xeric streams due to the high background supply, especially of sand. The consistency of biological responses at 29 PCT_FN indicated a meaningful benchmark for protection of aquatic life.

Benchmark Applications

On the basis of our analyses, NMED now performs two levels of assessment in sequential order to determine whether there is excessive sedimentation/siltation in a stream reach. The PCT_SAFN sediment indicator is used in the Level One assessment because it is easily measured and strongly related with biological metrics. If PCT_SAFN indicate excessive fine sediment in comparison to the benchmarks, a Level Two survey is performed, including comparison of LRBS_NOR to benchmarks. The LRBS_NOR measure is appropriate as a second-tier indicator because it is scaled to hydro-geomorphic factors of the individual sites, as well as to the broader site classes. When used as a second-tier sediment indicator, LRBS_NOR helps explain whether high PCT_SAFN should be expected under specific site conditions or result from disturbed conditions.

SUMMARY AND CONCLUSIONS

The purpose of the preceding analyses was to illustrate a systematic process (using New Mexico as an example) for deriving numeric benchmarks of fine and unstable sediments that could be used by state environmental agencies like NMED to translate their narrative sediment standards and ultimately to protect their state's stream ecosystems. Benchmarks were suggested using multiple analytical methods that largely yielded congruent results. Therefore, the recommended benchmarks and supporting analyses provide a strong basis for NMED to make final selections of sediment thresholds and to establish procedures for their application (http://www.nmenv.state.nm.us/swqb/Sedimentation/). This approach may serve as a template for application in other states where similar data sets are available.

The analytical techniques used to identify potential benchmarks for sediments included reference distributions to describe expectations in least-disturbed sites, which were classified by natural site type. Quantile regression and changepoint analysis were used to compare sediment conditions with biological conditions, using the biological conditions to indicate the degree to which aquatic life uses were supported.

The RBS formulation after factoring out bedrock (LRBS_NOR) indicates site-specific hydraulic potential for moving bed sediments, so that the observed fine sediments are only considered in excess when the stream particles are more easily mobilized and transported than expected based on local channel characteristics. This allows evaluation of the potential of individual sites to retain or flush fine sediments. RBS has proven to be a sensitive and meaningful indicator in other studies (Kaufmann and Hughes, 2006; USEPA, 2006b; Jessup, 2009a; Kaufmann *et al.*, 2009).

Sediment can be defined as an absolute quantity (e.g., PCT_SAFN), without adjusting for shear stress as in the RBS. Percent sand and fines are easily measured and were strongly related to biological metrics. Biota seemed more responsive to PCT_SAFN than to LRBS_NOR. However, when used in conjunction with PCT_SAFN, LRBS_NOR can help explain whether high PCT_SAFN were expected for a given site or were a result of anthropogenic disturbance. LRBS_NOR can also be used to identify sites with deficient fine sediments, though we did not explore this effect in the current analysis.

The recommended benchmarks for protection of stream resources were heavily weighted by the sediment reference condition statistics. These statistics were the most direct measures of expected sediment conditions in undisturbed sites throughout New Mexico. The sediment reference conditions were characterized independently of the biological conditions and they are assumed to exemplify optimal habitat conditions for natural (relatively undisturbed) aquatic fauna.

Corroborating evidence for selection of benchmarks from reference conditions was found in the stressor response analysis of sediment conditions and macroinvertebrate metrics. Although the metrics are a direct measure of the protected resource, biological effects of sediments are undoubtedly confounded by environmental conditions other than sediments. However, through a combined interpretation of quantile regressions, changepoints, and LOWESS regression lines, we could discern some of the limiting and meaningful biological responses to sediments. For each indicator and site class, the most protective credible effect level was generally recommended as a benchmark.

Natural variability in sediment conditions that are associated with landscape-scale factors are accounted for by the site classes, which were determined by the level 4 ecoregions. The site classes (Mountains, Foothills, and Xeric areas) were identified through a PCA of environmental conditions and the sediment indicators. Expectations for the Mountains are that PCT_SAFN will be low and LRBS_NOR will be near 0 (higher than in the other ecoregions). In comparison, expectations in the Xeric areas are that highly mobile fine sediments are relatively common. Foothills sites should have intermediate sediment conditions. Because of these differences, assigning sites in proper classes is important for attaining accurate assessments.

While analyses converge to define the benchmarks presented, there are many benchmark-setting factors that cannot be addressed entirely in this article. For instance, reference and stressed sites are defined with rigor that varies by ecoregion to include sufficient samples for analyses. Mountain sites have relatively fewer stressors or stressor sources than Xeric sites. For this reason, the reference quantiles for Mountain sites may be more protective than those in the other regions.

In the Xeric sites, an alternative to the PCT_SAFN benchmark is to apply one based on PCT_FN, which exhibits a wider range in these sites. On the basis of biological responses, a benchmark of 29 PCT_FN is recommended; sites that exceed this benchmark are assessed as sediment stressed. However, because sediment indicator values at reference and stressed sites show poor to fair agreement with this benchmark, it should be used with caution. On the other hand, the reference sites in the Xeric region are more disturbed than in the other regions and we might see an improved correlation if some of the riparian disturbance could be accounted for in the reference distribution. With this approach, we could find the *y*-intercept or the minimum disturbance level and interpret regressions of reference conditions along one or more disturbance gradients (Kaufmann *et al.*, 2012).

In summary, this approach using field data and the weight of evidence from multiple analyses is credible for developing benchmark values because it corroborates findings using different methods and predictive variables. The case example was performed with a dataset from New Mexico. However, the approach could be applied more widely to provide a solid basis for developing bedded sediment criteria and protective management strategies in other regions.

ACKNOWLEDGMENTS

Funding for this project was provided by the U.S. Environmental Protection Agency through contract EP-C-08-004 to Tetra Tech, Inc. Field survey data were provided by the U.S. Environmental Protection Agency Wadeable Streams Assessment and Western Environmental Monitoring and Assessment Programs. James Hogan of the New Mexico Environment Department provided insightful suggestions and interpretations throughout the analyses. Angel Kosfisher of EPA Region 6 conducted GIS analyses. Erik Leppo of Tetra Tech helped in data management and analysis. Revisions suggested by Bob Ozretich, John Faustini, Lil Herger, and Larry Willis improved the manuscript, as did the comments provided by the associate editor and two anonymous reviewers. This paper was subjected to peer and administrative review at the National Health and Environmental Effects Research Laboratory's Western Ecology Division and approved for publication. Approval does not signify that the contents reflect the views of the USEPA, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

LITERATURE CITED

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). EPA 841-B-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Bryce, S.A., G.A. Lomnicky, and P.R. Kaufmann, 2010. Protecting Sediment Sensitive Aquatic Species in Mountain Streams through the Application of Biologically Based Streambed Sediment Criteria. Journal of the North American Benthological Society 29(2):657-672.

Bryce, S.A., G.A. Lomnicky, P.R. Kaufmann, L.S. McAllister, and T.L. Ernst, 2008. Development of Biologically Based Sediment

Criteria in Mountain Streams of the Western United States. North American Journal of Fisheries Management 28:1714-1724.

Buffington, J.M. and D.R. Montgomery, 1999. Effects of Hydraulic Roughness on Surface Textures of Gravel-Bed Rivers. Water Resources Research 35(11):3507-3521.

Burcher, C.L., H.M. Valett, and E.F. Benfield, 2007. The Land-Cover Cascade: Relationships Coupling Land and Water. Ecology 88:228-242.

Cade, B.S., J.W. Terrell, and R.L. Schroeder, 1999. Estimating Effects of Limiting Factors with Regression Quantiles. Ecology 80(1):311-323.

Cleveland, W.S., 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of American Statistical Association 74:829-836.

Cormier, S.M., J.F. Paul, R.L. Spehar, P. Shaw-Allen, W.J. Berry, and G.W. Suter, II, 2008. Using Field Data and Weight of Evidence to Develop Water Quality Criteria. Integrated Environmental Assessment and Management 4(4):490-504.

Daily, J.P., N.P. Hitt, D.R. Smith, and C.D. Snyder, 2012. Experimental and Environmental Factors Affect Spurious Detection of Ecological Thresholds. Ecology 93(1):17-23.

Dunne, T. and L. Leopold, 1978. Water in Environmental Planning. W.H. Freeman and Co., San Francisco, California.

Faustini, J.M. and P.R. Kaufmann, 2007. Adequacy of Visually Classified Particle Count Statistics from Regional Stream Habitat Surveys. Journal of the American Water Resources Association 43(5):1293-1315, doi: 10.1111/j.1752-1688.2007.00114.x.

Faustini, J.M., P.R. Kaufmann, and A.T. Herlihy, 2009. Downstream Variation in Bankfull Width of Wadeable Streams across the Conterminous United States. Geomorphology 108:292-311.

Flotemersch, J.E., J.B. Stribling, and M.J. Paul, 2006. Concepts and Approaches for the Bioassessment of Non-Wadeable Streams and Rivers. EPA/600/R-06/127. U.S. Environmental Protection Agency, Cincinnati, Ohio.

Griffith, G.E., J.M. Omernik, M.M. McGraw, G.Z. Jacobi, C.M. Canavan, T.S. Schrader, D. Mercer, R. Hill, and B.C. Moran, 2006. Ecoregions of New Mexico (color poster with map, descriptive text, summary tables, and photographs). U.S. Geological Survey, Reston, Virginia (map scale 1:1,400,000).

Herlihy, A.T., S.G. Paulsen, J. Van Sickle, J.L. Stoddard, C.P. Hawkins, and L.L. Yuan, 2008. Striving for Consistency in a National Assessment: The Challenges of Applying a Reference Condition Approach at a Continental Scale. Journal of the North American Benthological Society 27:860-877.

Hilsenhoff, W.L., 1987. An Improved Biotic Index of Organic Stream Pollution. The Great Lakes Entomologist 20(1):31-40.

Hughes, R.M., 1995. Defining Acceptable Biological Status by Comparing with Reference Conditions. *In*: Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making for Rivers and Streams, W. Davis and T. Simon (Editors). Lewis, Boca Raton, Florida, pp. 31-47.

Jacobi, G.Z., M.D. Jacobi, M.T. Barbour, and E.W. Leppo, 2006. Benthic Macroinvertebrate Stream Condition Indices for New Mexico Wadeable Streams. New Mexico Environment Department, Santa Fe, New Mexico.

Jessup, B., 2009a. Development of Bedded Sediment Benchmarks for Oregon Streams. Oregon Department of Environmental Quality and U.S. EPA Region 10.

Jessup, B., 2009b. Recalibration of the Macroinvertebrate Multi-Metric Index for Colorado. Colorado Department of Public Health and Environment, Water Quality Control Division – Monitoring Unit, Denver, Colorado and EPA Region 8.

Jessup, B.K., D. Eib, L. Guevara, J. Hogan, F. John, S. Joseph, P. Kaufmann, and A. Kosfiszer, 2010. Sediment in New Mexico Streams: Existing Conditions and Potential Benchmarks. U.S. Enviromental Protection Agency, Region 6, Dallas, Texas and the New Mexico Environment Department, Santa Fe, New Mexico. Prepared by Tetra Tech, Inc., Montpelier, Vermont.

Kaufmann, P.R., J.M. Faustini, D.P. Larsen, and M. Shirazi, 2008. A Roughness-Corrected Index of Relative Bed Stability for Regional Stream Surveys. Geomorphology 99:150-170.

Kaufmann, P.R. and R.M. Hughes, 2006. Geomorphic and Anthropogenic Influences on Fish and Amphibians in Pacific Northwest Coastal Streams. American Fisheries Society Symposium 48:429-455.

Kaufmann, P.R., D.P. Larsen, and J.M. Faustini, 2009. Bed Stability and Sedimentation Associated with Human Disturbances in Pacific Northwest Streams. Journal of the American Water Resources Association 45(2):434-459.

Kaufmann, P.R., P. Levine, E.G. Robison, C. Seeliger, and D.V. Peck, 1999. Quantifying Physical Habitat in Wadeable Streams. EPA 620/R-99/003. Environmental Monitoring and Assessment Program, U.S. Environmental Protection Agency, Corvallis, Oregon.

Kaufmann, P.R., D.V. Peck, and S.G. Paulsen, 2012. NRSA Technical Support Documentation; Fluvial (River and Stream) Physical Habitat Condition Assessment (DRAFT 7/16/2012). U.S. EPA Office of Research and Development, Corvallis, Oregon.

King, R.S. and C.J. Richardson, 2003. Integrating Bioassessment and Ecological Risk Assessment: An Approach to Developing Numerical Water-Quality Criteria. Environmental Management 31:795-809.

Leopold, L.B., M.G. Wolman, and J.P. Miller, 1964. Fluvial Processes in Geomorphology. W.H. Freeman and Co., San Francisco, California.

Paul, J.F., S.M. Cormier, W.J. Berry, P.R. Kaufmann, R.L. Spehar, D.J. Norton, R.E. Cantilli, R. Stevens, W.F. Swietlik, and B.K. Jessup, 2008. Developing Water Quality Criteria for Suspended and Bedded Sediments—Illustrative Example Application. Water Practice 2(1):117.

Paul, M.J., 2008. Draft Predictive Bioassessment Models for New Mexico Streams. Prepared for New Mexico Environment Department, Santa Fe, NM. Prepared by Tetra Tech Inc., Owings Mills, Maryland.

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, P.L. 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. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C.

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

Qian, S.S., R.S. King, and C.J. Richardson, 2003. Two Statistical Methods for the Detection of Environmental Thresholds. Ecological Modeling 166:87-97.

R Development Core Team, 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org, *accessed* August 2010.

Relyea, C.D., G.W. Minshall, and R.J. Danehy, 2000. Stream Insects as Bioindicators of Fine Sediment. *In*: Proceedings of the Water Environment Federation, Watershed 2000. Water Environment Federation, Vancouver, British Columbia, pp. 663-686.

Relyea, C.D., G.W. Minshall, and R.J. Danehy, 2012. Development and Validation of an Aquatic Fine Sediment Biotic Index. Environmental Assessment 49:242-252.

Southerland, A.B., J.L. Meyer, and E.P. Gardiner, 2002. Effects of Land Cover on Sediment Regime and Fish Assemblage Structure in Four Southern Appalachian Streams. Freshwater Biology 47:1791-1805.

Stoddard, J.L., D.P. Larsen, C.P. Hawkins, R.K. Johnson, and R.H. Norris, 2006. Setting Expectations for the Ecological Condition

of Streams: The Concept of Reference Conditions. Ecological Applications 16:1267-1276.
Stoddard, J.L., D.V. Peck, A.R. Olsen, D.P. Larsen, J. Van Sickle, C.P. Hawkins, R.M. Hughes, T.R. Whittier, G. Lomnicky, A.T. Herlihy, P.R. Kaufmann, S.A. Peterson, P.L. Ringold, S.G. Paulsen, and R. Blair, 2005a. Environmental Monitoring and Assessment Program (EMAP): Western Streams and Rivers Statistical Summary. EPA 620/R-05/006. Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.
Stoddard, J.L., D.V. Peck, S.G. Paulsen, J. Van Sickle, C.P. Hawkins, A.T. Herlihy, R.M. Hughes, P.R. Kaufmann, D.P. Larsen, G. Lomnicky, A.R. Olsen, S.A. Peterson, P.L. Ringold, and T.R. Whittier, 2005b. An Ecological Assessment of Western Streams and Rivers. EPA 620/R-05/005. U.S. Environmental Protection Agency, Washington, D.C.
USEPA (United States Environmental Protection Agency), 2000. Nutrient Criteria Technical Guidance Manual, Rivers and Streams. Office of Water. EPA-822-B-00-002, Washington, D.C.
USEPA (United States Environmental Protection Agency), 2006a. Framework for Developing Suspended and Bedded Sediment (SABS) Water Quality Criteria. Office of Water, Office of Research and Development. EPA-822-R-06-001, Washington, D.C.
USEPA (United States Environmental Protection Agency), 2006b. Wadeable Streams Assessment: A Collaborative Survey of the Nation's Streams. EPA-841-B-06-002. United States Environmental Protection Agency, Office of Water, Washington, D.C.
Waters, T.F., 1995. Sediment in Streams. American Fisheries Society Bethesda, Bethesda, Maryland.
Wood, P.J. and P.D. Armitage, 1997. Biological Effects of Fine Sediment in the Lotic Environment. Environmental Management 21(2):203-217.
Zaimes, G.N., R.C. Schultz, and R.M. Isenhart, 2004. Stream Bank Erosion Adjacent to Riparian Forest Buffers, Row-Crop Fields, and Continuously Grazed Pastures along Bear Creek in Central Iowa. Journal of Soil and Water Conservation 59(1):19-27.