DEVELOPMENT OF SEDIMENTATION/SILTATION IMPAIRMENT THRESHOLDS IN NEW MEXICO

NEW MEXICO ENVIRONMENT DEPARTMENT
SURFACE WATER QUALITY BUREAU

MARCH 15, 2011

Overview

This document provides a summary of the seven step process the New Mexico Environment Department (NMED) Surface Water Quality Bureau (SWQB) initiated and helped complete in order to determine thresholds for bedded sediments in perennial, wadeable streams in New Mexico. The results of this process were used to revise the sedimentation/siltation assessment process for development of the 2012-2014 Integrated List (NMED/SWQB 2011). This effort was necessary to support an interpretation of the *State of New Mexico Standards for Interstate and Intrastate Surface Waters* narrative standard for bottom deposits found at NMAC 20.6.4.13 (NMWQCC 2011):

> *A. Bottom Deposits and Suspended or Settleable Solids:*
> *(1) Surface waters of the state shall be free of water contaminants including fine sediment particles (less than two millimeters in diameter), precipitates or organic or inorganic solids from other than natural causes that have settled to form layers on or fill the interstices of the natural or dominant substrate in quantities that damage or impair the normal growth, function, or reproduction of aquatic life or significantly alter the physical or chemical properties of the bottom.*

In 2008, the SWQB Sediment Workgroup was formed to review the previous sedimentation/siltation assessment protocol and recommend an approach for revision. As a result of workgroup discussions, SWQB and USEPA Region 6 contracted with Tetra Tech, Inc., to develop sediment translators or thresholds. The contractor generally followed the steps provided in USEPA's Framework for developing suspended and bedded sediment (SABS) water quality criteria (USEPA 2006). Several staff from Tetra Tech, Inc., USEPA Region 6, and SWQB worked as a team to complete this effort.

This effort included the identification of sediment characteristics that are expected under the range of environmental settings in New Mexico, especially in undisturbed or best available reference streams. The goal of this characterization was to enable SWQB to identify situations where bedded and suspended sediment expectations are not met, using sediment indicators that show responsiveness to disturbance. Examining the relationships between biological measures and sediment indicators helped to identify where disturbance caused sediment imbalance and biologically-relevant habitat degradation.

The approach to identifying potential sediment-related benchmarks in New Mexico followed seven basic steps, based on the USEPA Framework for developing SABS water quality criteria (USEPA 2006):

1. Review background information
2. Assemble datasets with potential sediment indicators
3. Establish reference sites
4. Classify sites
5. Characterize sediments
6. Describe stressor-response relationships
7. Recommend benchmarks or thresholds

This document provides a summary of this effort with respect to bedded sedimentation. The information in each step below is excerpted from the final report from the contractor (Jessup et al. 2010). The entire 100+ page final report details the results of each bedded sediment analysis, and includes additional information on suspended sediment indicators as well as transformed data exploration (i.e., residualization). This report is available on the SWQB web site at: http://www.nmenv.state.nm.us/SWQB/. Abbreviations are defined in Appendix A. An abbreviated description of the previous sedimentation assessment protocol utilized essentially from listing cycle 1998 through 2010 is also included in Appendix B for historical purposes.

Table of Contents

Step 1: Review background information 5
Step 2: Assemble datasets with potential sediment indicators 6
Step 3: Establish reference sites 9
Step 4: Classify sites 11
Step 5: Characterize sediments 16
Step 6: Describe stressor-response relationships 19
Step 7: Recommend benchmarks or thresholds 28
Conclusion and Research Needs 32
References Cited 35
- APPENDIX A: Abbreviations 38
- APPENDIX B: SWQB's Previous Sedimentation Assessment Protocol 39

Step 1: Review background information

In this step, the effects of excessive sedimentation/siltation on aquatic life are described, as well as the challenges of determining specific numeric thresholds because sedimentation can have both natural and anthropogenic components.

The biological effects of excess fine suspended and bedded sediments in streams and rivers are well established in the scientific literature (e.g., Waters 1995, Wood and Armitage 1997, Suttle et al. 2004, Opperman et al. 2005). These effects include displacement of interstitial habitat space, clogging of water movement through sediments, disruption of normal predator-prey relationships through visual impairment, decreased primary productivity, increased macroinvertebrate drift, abrasion or smothering of gills and other organs, and increased uptake of sediment-bound toxicants (Figure 1). While these effects are well known, the process for establishing benchmarks or thresholds of sediment effects is still new and evolving (USEPA 2006; Paul et al. 2008; Cormier et al. 2008; Jessup 2009a; Bryce et al. 2008, 2010).

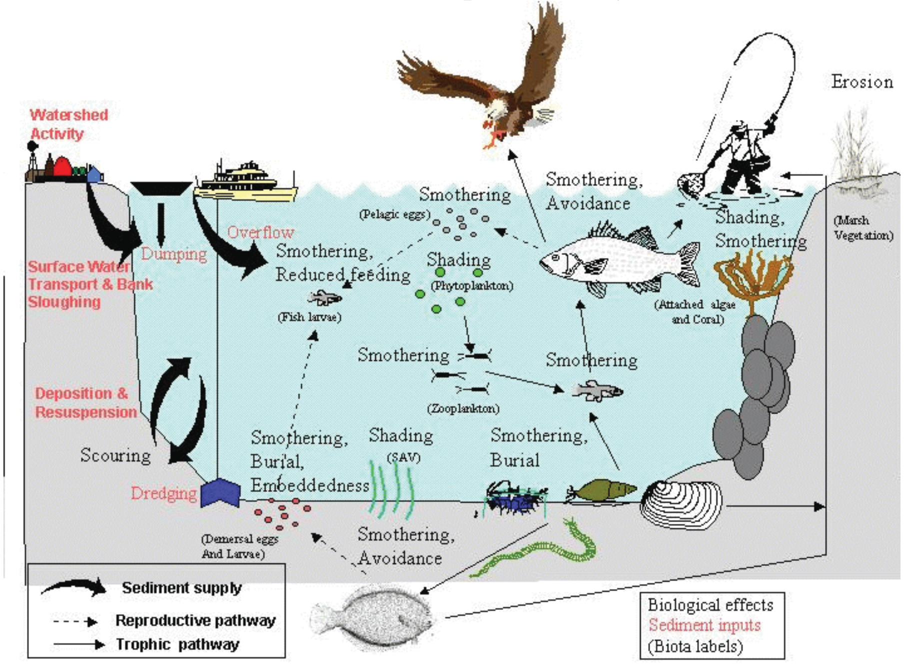

Figure 1. Conceptual diagram of bedded sedimentation effects (taken from USEPA 2006 – graphic courtesy of W. Munns, USEPA)

The condition of substrates in streams and rivers forms the foundation of habitat suitability for benthic organisms and their predators. The stability, size, shape, density, and porosity of the substrates directly affect the behavior of organisms and their ability to find food resources, hide from predation or other threats, and reproduce (Waters 1995, Wood and Armitage 1997). When substrate characteristics are out of balance in the channel or watershed, detrimental changes in the biological community structure and function can occur. Typically, increased disturbance leads to increased fine and mobile sediments, which leads to decreased stream habitat suitability.

Sediments cannot be treated as introduced pollutants such as pesticides because they are not uniquely generated through human input or disturbance. Rather, sediments are components of natural systems that are present even in pristine settings and to which stream organisms have evolved and adapted. Therefore, the detection of a sediment imbalance is more difficult than detecting an absolute concentration or percentage that represents a clear biological impact.

Substrate and suspended sediment characteristics can be considered impacted at a site under two circumstances. First, they can be considered impacted if they are not similar to expectations for undisturbed sites in the same environmental setting. A second case for impact can be made when the substrate characteristics are detectably affecting the biota. In the first case, substrates may be more fine, more coarse, more unstable, or more stable than they are expected to be under broadly-recognized, undisturbed conditions (reference or best available conditions) for that particular environmental setting. This, in itself, can be an indication that streambed substrates are impacted by human disturbance. Biotic responses to disturbed substrates can be variable, but sub-optimal biotic conditions are often associated with unbalanced sediments. These relationships can be demonstrated through analyses. Both of these assessment strategies can be used to develop numeric benchmarks.

The purpose of this analysis is to identify sediment characteristics that are expected under the range of environmental settings in New Mexico, especially in undisturbed reference streams. Through this characterization, it will be possible to identify situations where the expectations are not met, using sediment indicators that show responsiveness to disturbance. Associating biological measures with sediment indicators will further indicate situations where the disturbance causes sediment imbalance and biologically-relevant habitat degradation. The results of these analyses include a set of recommendations regarding the application of quantitative sediment benchmarks on New Mexico perennial streams. This includes a range of possible benchmarks, as well as the rationale and the strengths and weaknesses of each.

Step 2: Assemble datasets with potential sediment indicators

In this step, datasets are considered, selected, and collated in order to generate one comprehensive dataset of site characteristic, biological, and associated sediment habitat data sufficient for statistical analyses.

Data sources considered for analysis of sediment effects in New Mexico included the Ecological Data Application System (EDAS) maintained by SWQB and multiple sediment datasets compiled by SWQB, USEPA, and neighboring states (Table 1). Data from states neighboring New Mexico were used if sites were in an ecoregion that was contiguous with those in New Mexico and the site was within 50 -150 miles of the state border (generally within 50 but extending to 150 in the east and northeast or as needed to include sites in the less densely sampled plains and xeric ecoregions).

Table 1. Datasets used in analyses

Data set	Sampling Years	Source
NMED EDAS benthic macroinvertebrate data	1980 – 2007	NMED
NMED targeted riffle sediment data (taken during EMAP[b] sampling)	2006 – 2007	NMED
EMAP West	1999 – 2004	EPA; SWIMS web site[a]
EMAP Wadeable Streams Assessment	2000 – 2005	EPA; SWIMS web site[a]
EMAP Arizona Streams	2007	EPA; SWIMS web site[a]
EMAP New Mexico	1999 – 2001, 2006 – 2007	EPA; SWIMS web site[a]
EMAP Region 8 Colorado Streams	1994 – 1995	EPA; SWIMS web site[a]
NMED suspended sediment data	2001 – 2009	NMED
Site GIS characterization	2000 – 2010	EPA Region 6 analysis

NOTES:

a: The Surface Water Information Management System (SWIMS) was maintained by the USEPA National Health and Environmental Effects Research Laboratory-Western Ecology Division

b: EMAP = EPA's Environmental Monitoring and Assessment Program

GIS analysis was carried out by USEPA Region 6 to characterize land use, geology, and climatic conditions at each site and in the site catchments. Human disturbance variables analyzed at each site included land use, land cover, dams, road density, and road – stream intersections. The natural characteristics of sites and catchments included catchment area, National Hydrography Dataset (NHD) stream reach slope, land slope in the catchment and at the site, ecoregion designation, stream order, site elevation, soil permeability, precipitation, and geologic type for erodibility ratings.

Geologic types identified in the GIS analysis were categorized for erodibility (or fine sediment production potential). Catchment characteristics in relation to geologic erodibility were determined by the percentage of the site catchment with geological types that are highly or moderately erodible. Stream power, the product of discharge and channel slope, is a measure of the capacity of a waterway to carry bedload. An index of stream power was calculated as the product of catchment area, stream slope, and precipitation.

Data were manipulated to refine, transform, average, or calculate metrics on sediment indicators, biological indicators, and site/catchment characteristics. Biological metrics and indices were calculated from the taxa lists. These included the New Mexico Macroinvertebrate Stream Condition Index (NMMSCI - Jacobi et al. 2006) and other metrics that were components of the NMMSCI or otherwise believed to be generally responsive to sediment stresses in New Mexico streams. The NMMSCI was standardized as a proportion of the reported impairment benchmark in each of the biological site classes. If benthic macroinvertebrate data existed from several sampling events, the metrics from the sampling event that coincided with the sediment habitat sampling were used.

Sediment indicators calculated from the datasets that were of primary interest included relative bed stability (RBS), % fines, and percent sand & fines for bedded sediment measures (Table 2). Other derivations of these sediment indicators were calculated from these basic measures, and are discussed in detail in the full report (Jessup et al. 2010). Such derivations included statistical transformations or standardization of the % fines and percent sand & fines to natural variables (i.e., determination of residuals). The RBS index was calculated from variables in the EMAP habitat files. If sediment habitat data existed from several sampling events, these data were summarized for each site by averaging multiple sampling dates or replicates.

Table 2. Bedded sediment indicators

Sediment Indicator	**Description**
Relative Bed Stability	A measure of the relationship of the median particle size in a stream reach compared to the critical particle size calculated to be mobilized by standardized fluvial stresses in the reach. Median particle size is determined using a reach-wide pebble count (Peck et al. 2006). Critical particle size is calculated from channel dimensions, flow characteristics, and channel roughness factors (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**".
Percent Fines	The percentage of systematically selected (with random start) streambed substrate particles that are **≤0.06 mm** (silt, clay, or muck; not gritty when rubbed between fingers). Random particles were selected using either a reach-wide or targeted riffle pebble count. Abbreviated **"%Fines**".
Percent Sand & Fines	The percentage of systematically selected (with random start) streambed substrate particles that are **≤2.0 mm** in diameter. Random particles were selected using either a reach-wide or targeted riffle pebble count. Abbreviated "**%SaFn**".

The LRBS indicator was developed by USEPA researchers as a tool to predict the expected substrate size distribution for streams (Peck et al. 2006). It has proven to be a sensitive and meaningful indicator in other studies (Jessup 2009a, Kaufmann et al. 2009). Because fluvial site conditions are major determinants of the substrate conditions in stream channels, the critical particle size calculated from fluvial characteristics is a predictor of dominant and stable substrate conditions. This indicator incorporates stream channel, shape, slope, flow, and sediment supply. Sediment conditions relative to the fluvial potential are better estimates of system stability and imbalance than absolute measures of fine sediment concentration because they intrinsically account for site-specific natural settings.

In essence, the LRBS calculation is used to determine the stream power based on channel measurements to predict the expected sediment particle size that would be moved during a bankfull flow event. This expected or "critical" particle size is calculated from channel dimensions, roughness factors, and shear stresses (Kaufmann et al. 2008). The logarithm ratio of the measured particle size to the expected particle size is a measure of the relative stability of the stream bed. In minimally disturbed streams, the measured geometric mean particle size should trend towards the expected particle size (i.e., the size the stream is capable of moving as bedload at bankfull). Thus, LRBS values near zero indicate a stable stream bed, whereas increasingly negative values indicate excess fine sediment. For example, a LRBS value of -1 means that the measured geometric mean bedded sediment particle size is ten times (10X) finer than the expected particle size moving during bankfull flow events. A calculated LRBS of -2 means that the observed particle size is 100X finer than the expected particle size moving during bankfull flow events, whereas values less than -3 indicate that the bed substrate may be moving even during low flow events.

Step 3: Establish reference sites

In this step, reference sites are determined. This step provides information on the expected, reference, or best available condition, and allows for comparison to other sites.

Indicators of the sediment and biological condition at a site were expected to correlate to the intensity of disturbance at the site and in the contributing watershed. Sites with minimal evidence of disturbance at the reach scale and in the catchments are expected to exemplify our best expectations for sediment and biological conditions (Stoddard et al. 2006). Conditions at these sites are the reference conditions. They are standards for comparison to other sites. Reference sites are also typically examined for patterns of natural variability that can be used to define site classes.

For bedded sediment sites, preliminary indicators of site reference status included existing designations or procedures used to designate reference sites in previous studies. These studies include EMAP (Stoddard et al. 2005), NMED benthic multi-metric 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 preliminary reference designations were further scrutinized by selective reapplication of existing site criteria, application of new criteria for land use and road density, checking agreement of multiple reference designations, and evaluation of aerial imagery. Where designations were contradictory (one approach indicates reference and another indicates degradation), sites were left with no designation and termed “other”. Placing sites in this “other” category reduced the number of reference or stressed sites for subsequent analyses. However, this approach gave more credibility to the reference and stressed conditions and the benchmarks that were based on them.

The screening and criteria process resulted in identification of 99 reference sites of the 229 sites used for the bedded sediment analysis (Figure 2). Most of the reference sites were in mountainous ecoregions which include the Southern Rockies (N = 37) and the Arizona/New Mexico Mountains (N = 41). Far fewer (25) stressed sites were identified, which were distributed relatively evenly among the ecoregions. Though there were numeric reference criteria for some variables, designations were made with subjective input (weight-of-evidence) because of the combination of multiple indications (numeric criteria, previous designations, screening examination of aerial imagery, and professional judgment of familiar sites).

Because sites may have been evaluated with multiple previous analyses, the following statistics (Table 3) overlap among programs, giving the appearance that more than 99 sites were designated as reference. The EMAP criteria that were adjusted to remove sediment variables resulted in 44 sites that met four of the six reference criteria and 56 that met three criteria. After comparing to other reference designations and examining maps, 82 of those sites were designated as reference. Twenty-three sites failed one or more of the six EMAP stressed criteria, resulting in 18 stressed site designations after additional screening.

The RIVPACS analysis identified 24 reference sites in New Mexico (Paul 2008), 21 of which were confirmed. Likewise, NMED designations in the EDAS database (used for biological indicator development) identified 24 reference sites, 20 of which were confirmed. Only one of five stressed sites identified in EDAS were confirmed. In Colorado, 15 of the 17 sites identified by CDPHE were confirmed as reference for the current analysis. Only one stressed site in Colorado was identified and confirmed. Five stressed sites were identified based on GIS information alone.

Table 3. Reference and stressed sites

Criteria Set	Potential Reference	Final Reference	Potential Stressed	Final Stressed
EMAP	100	82	23	18
RIVPACS	24	21	n/a	n/a
NMED EDAS	24	20	5	1
CDPHE	17	15	1	1
GIS	n/a	n/a	5	5

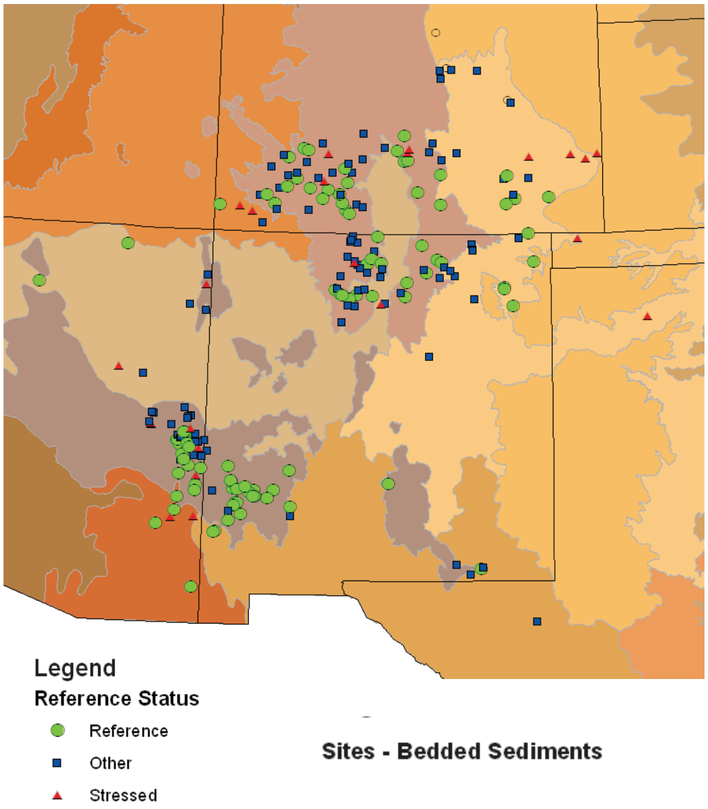

Figure 2. Site locations and reference status of bedded sediment sites in New Mexico and surrounding areas (includes Level 3 ecoregions boundaries)

Step 4: Classify sites

In this step, potential classes are explored in order to group similar sites into categories or site classes for benchmark development. This enables reference conditions to be defined for each site class so it is not necessary to identify individual reference sites for each study site in order to determine impairment.

Site classification is the process by which natural gradients among sites are examined to identify appropriate classes or “bins” of sites with similar sediment characteristics. The purpose of stream classification is to minimize within-class natural variability of sediment indicators so that anthropogenic disturbance can be recognized with less background noise. Potential site classification variables, sediment indicators, and biological variables were analyzed simultaneously to identify patterns of covariance. The analysis included several analytical methods (e.g., principal components analysis, correlation, examination of bi-plots, and distributions). Some additional techniques were exploratory and were not carried beyond preliminary stages, though they were sufficient to guide final analyses.

Ideally, reference sites alone would be used for site classification. However, the sediment indicators may respond differently to stress in the landscape depending on some natural characteristics, so it made sense to review patterns in all sites, not just reference sites. For instance, reference sites in erodible and resistant lithologies may have similar sediment signatures. However, with equal disturbance, the sites with erodible characteristics may have a much more profound signal of sediment stress, as was reported for Pacific Northwest streams (Kaufmann et al., 2009). It was also hypothesized that ecoregions and stream size (or power) may be important determinants of natural sediment conditions. Aggregate ecoregions used in the EMAP-West study — Mountains, Plains, and Xeric — were considered as a starting point for stream classification, but were not accepted without scrutiny.

Principal components analysis (PCA) was used as a primary tool for classifying sites. Correlation analysis was used to describe basic 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 analysis. The analysis was limited to reference sites when comparing individual pairs of variables to diminish some of the co-variability with stressors. Correlation results thus offer a slightly different perspective and new information compared to the PCA.

In the PCA, the first three factors explained 53% of the variability in the primary variables (Table 4). Sediment indicators were most strongly related to the first axis, and somewhat related to the second axis. The first principal axis was related to the natural variables, catchment area, elevation, stream slope, and precipitation and to the density of road crossings in the site catchment. Benthic macroinvertebrate metrics were also most strongly related to the first axis, and were therefore related to the sediment measures. The second axis was related to stream size (power and width). The third and fourth axes were related to site location (latitude and longitude) and land uses. The third and fourth axes were not strongly related to sediment or biological measures.

Table 4. PCA factor scores on the most important variables in all sites

Variable code	**Variable description**	**F. 1 30%**	**F. 2 12%**	**F. 3 11%**	**F. 4 6%**
LRBS_fin	Relative Bed Stability index (log10)	0.56	0.45	-0.17	-0.10
LRBS_NOR	RBS without bedrock or hardpan (log10)	0.58	0.44	-0.18	0.07
asPCT_SAFN	% sand & fine sediments at the site (arcsin(sqrt())	**-0.72**	-0.44	0.20	0.21
asPCT_FN	% fine sediments at the site (arcsin(sqrt())	-0.58	-0.48	0.20	0.12
Resid_pSAFN	% sand & fines (residual of critical diameter)	**-0.61**	-0.39	0.10	-0.13
Resid_pFN	% fines (residual of critical diameter)	-0.43	-0.42	0.10	-0.03
Res2pSAFN	% sand & fine (residual as per Stoddard)	-0.30	-0.38	0.30	-0.26
Res2pFines	% fines (residual as per Stoddard)	-0.15	-0.39	0.26	-0.12
LRdX_km2	Road crossings per km2 in the catchment	**-0.86**	0.37	0.10	-0.04
LArea_km2	Catchment area (log10(km2))	**-0.84**	0.47	-0.04	0.04
LSTREAMSLOP	Stream slope (log10(%), NHD data)	**0.84**	-0.08	-0.09	-0.06
ELEV_m	Site elevation (m)	**0.78**	-0.15	-0.20	0.22
STREAMORDE	Stream Order (Strahler)	**-0.75**	0.50	0.12	0.00
Precip	Precipitation (cm)	**0.72**	0.14	-0.24	0.08
LXSLOPE	Stream slope (log10(%), field data)	**0.66**	0.07	-0.33	-0.16
LPower	Stream power (log10(Precip*Area_km2*Xslope))	-0.50	**0.65**	-0.28	-0.03
LXWIDTH	Average site width (log10)	-0.46	**0.63**	-0.29	-0.14
Point_X	Latitude of sample	-0.36	-0.26	**-0.63**	0.10
Point_Y	Longitude of sample	0.00	-0.36	**-0.73**	0.10
as_Nindx	Natural land uses (%)	0.35	0.45	0.21	**0.72**
U_INDEX	Developed (urban) land uses (%)	-0.37	-0.45	-0.21	**-0.71**
*TotalTax	Total taxa (count)	0.42	0.20	0.04	-0.05
*BeckBI	Beck's Biotic Index (weighted count of sensitive taxa)	**0.63**	0.24	-0.19	-0.01
*IntolTax	Number of taxa intolerant of pollution (count)	**0.65**	0.17	-0.23	0.00
*EPTTax	EPT taxa (count)	0.46	0.39	-0.20	-0.09

NOTE: Scores with magnitude greater than 0.60 are shown in bold-type and considered to be strong relationships. See Appendix A for Variable Code definitions, and see Jessup et al. 2010 for descriptions of the residual calculations.

The relationships that were suggested by PCA and correlations were examined in box plots and bi-plots. Box and whisker plots were compared among expected site classes, such as ecoregions, to aid assessment of sediment variability. Ecoregions were examined using box plots because they are categorical and are one of the *a priori* variables of interest. Such illustrations of the distributions of values are simple to generate and easy to interpret for one variable at a time. They are not so useful for detecting interactive effects. Bi-plots were used to show patterns of relationships between two variables and to highlight tertiary attributes of the relationships such as reference status, ecoregion, or other covariants. LOWESS regressions were used to show trends at local portions of the gradients and to identify possible change-points.

Sites were classified so that sediment indicator values are calibrated to the many environmental settings in New Mexico. Preliminary analyses indicated that sediment indicators (LRBS, % fines, and % sand & fines) were most strongly related to environmental variables related to stream size, stream slope, precipitation, and riparian vegetation. The workgroup had an *a priori* assumption that sediment conditions would vary by ecoregion. This assumption was validated by the PCA results, which identified potential classification variables that are integral to the ecoregional classification scheme. Sediment conditions among ecoregions were investigated using box plots because ecoregional categories could not be interpreted in the PCA and correlation analyses.

The level 3 ecoregions span considerable types of landforms in some cases. Therefore, level 4 ecoregions were considered for refinement of the classes. While a general division between mountains and plains was suggested, an additional transitional class seemed appropriate and necessary. Examination of sediment indicator values in level 4 ecoregions and consideration of ecoregion descriptions for regions that were poorly represented in this dataset lead to site classification similar, yet modified from the EMAP mountains, plains, and xeric regions. For this analysis, the regions are termed Mountains, Foothills, and Xeric. Classes were originally defined as in Table 5a. This scheme recognizes the differences between high elevation, steep sloped, lush vegetation mountain streams; lower and drier foothill streams; and flatter and still drier xeric streams. The first factor scores of the PCA were significantly different (ANOVA followed by Tukey's Honestly Significant Difference test, p<0.0001) among the site classes (Figure 3), indicating that the site classes have distinct environmental characteristics.

Table 5a. Original definition of bedded sediment site classes (Jessup et al. 2010)

Site Class	**Definition**
Mountains	Ecoregions 21 and 23, **except** 21d, 23a, 23b and 23e
Foothills	Ecoregions 21d, 23a, 23b, 23e and 79
Xeric	Ecoregions 20, 22, 24, 25, and 26

Ecoregion number	Level 3 Ecoregion Name
20	Colorado Plateaus
21	Southern Rockies
22	Arizona/New Mexico Plateau
23	Arizona/New Mexico Mountains
24	Chihuahuan Deserts
25	High Plains
26	Southwestern Tablelands
79	Madrean Archipelago

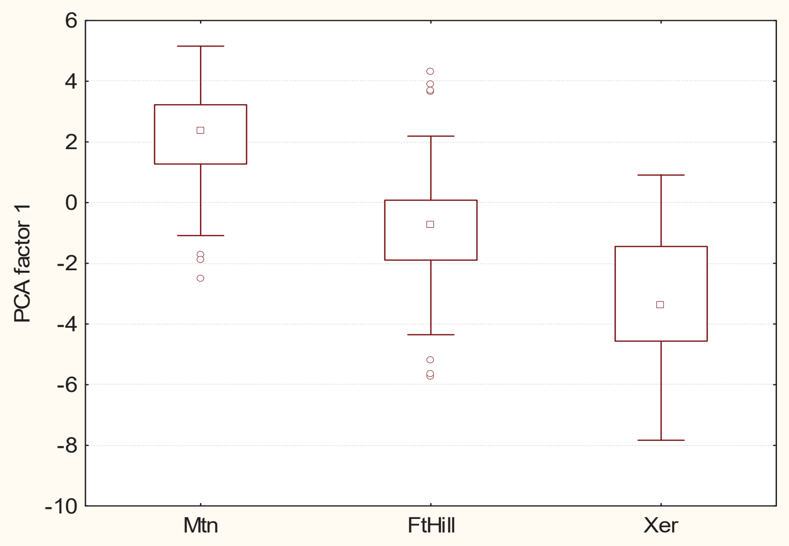

Figure 3. Three sediment site classes vs. PCA first factor

During development of New Mexico's revised sedimentation/siltation assessment protocol of the 2012-2014 listing cycle, further division of level 3 ecoregion 22 was investigated based on data for sites in level 4 ecoregions 22a, 22b, and 22f that were not available during the original sediment site class determination. These three ecoregions occur in the Upper Rio Grande valley upstream of the confluence with the Rio Chama. It was determined that sites characteristics related to the first PCA factor in these three level 4 ecoregions are generally more similar to sites characteristics in the foothills site class than the xeric site class, especially elevation, slope, and catchment area. Therefore, final sediment site classes for New Mexico were revised to divide level 3 ecoregion 22 into the foothills and xeric site class based on level 4 ecoregions (Table 5b) and are shown in Figure 4. This change has a very minor effect on the threshold determinations for the xeric sediment site class in Jessup et al. 2010 because only one of the seventeen xeric reference sites was in a level 4 ecoregion that has been moved to the foothills site class.

Table 5b. Final bedded sediment site classes

Site Class	**Definition**
Mountains	Ecoregions 21 and 23, **except** 21d, 23a, 23b and 23e
Foothills	Ecoregions 21d, 22a, 22b, 22f, 23a, 23b, 23e and 79
Xeric	Ecoregions 20, 24, 25, 26, and 22, **except** 22a, 22b, 22f

Ecoregion number	**Ecoregion Names***
20	Colorado Plateaus
21	Southern Rockies
21d	Foothill Woodlands and Shrublands
22a	San Luis Shrublands and Hills
22b	San Luis Alluvial Flats and Wetlands
22f	Taos Plateau
23	Arizona/New Mexico Mountains
23a	Chihuahuan Desert Slopes
23b	Madrean Lower Montane Woodlands
23e	Conifer Woodlands and Savannas

24	Chihuahuan Deserts
25	High Plains
26	Southwestern Tablelands
79	Madrean Archipelago

NOTES: * Additional written descriptions of level 4 ecoregions in New Mexico are available at: http://www.eoearth.org/article/Ecoregions_of_New_Mexico_(EPA).

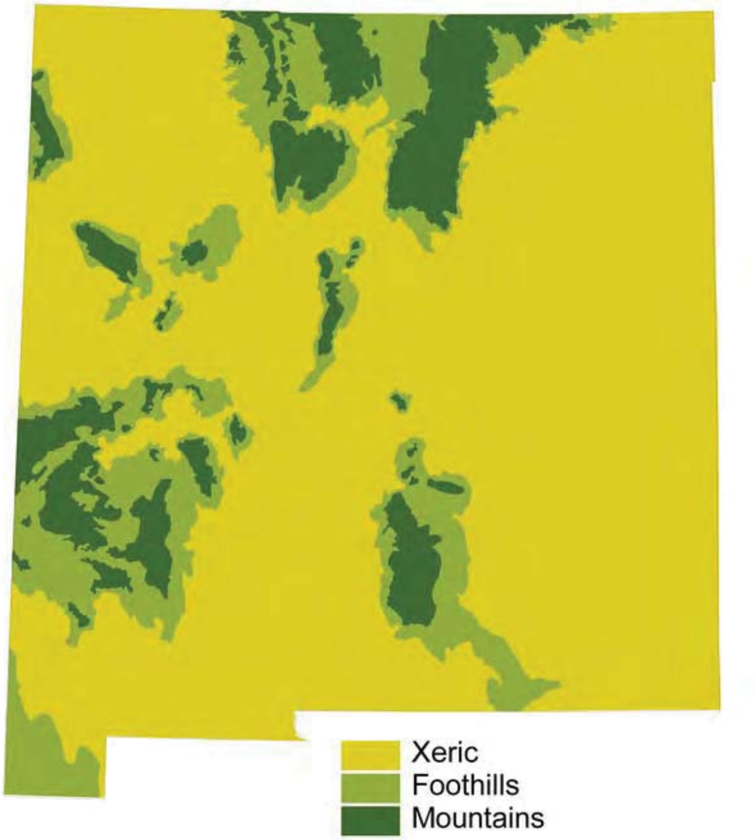

Figure 4. New Mexico Mountain, Foothills, and Xeric site class map

Lithologic erodibility was not found to be an important determinant of sediment classes in the PCA and correlation analyses. However, the following plot (Figure 5) suggests that in sites with resistant lithology, sediment conditions are less responsive to disturbance than in erodible sites. In both the mountainous and non-mountainous regions, reference sites have relatively consistent conditions across the erodibility scale. Stressed sites do not show excessively low LRBS values until the percentage of moderately and highly erodible rock cover in the catchment is greater than 40%. This pattern was also noted for the other sediment indicators (not shown). Excessive sediments in streams are less likely in areas with resistant rocks because the sources of fine particles are less plentiful and few particles are mobilized even in response to disturbance. This might be a consideration in applications of any benchmarks resulting from these analyses.

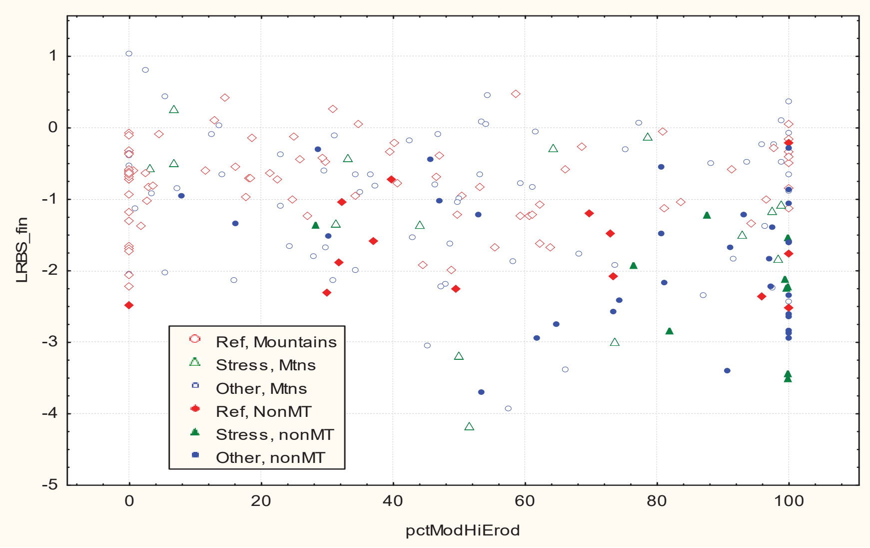

Figure 5. LRBS values in relation to the percentage of moderately and highly erodible lithology in the site catchments

Step 5: Characterize sediments

In this step, bedded sediment indicators at reference sites are analyzed to determine potential threshold levels for sedimentation/siltation impairment determination.

The sediment indicators currently in use and proposed for use in New Mexico streams were scrutinized to determine which were most appropriate for further analysis and application in the regulatory context. Sediment indicators should be selected based on: (a) association with designated uses; (b) availability and accessibility of data; (c) reliability of measurement characteristics; (d) appropriateness for the proposed analytical methods; and (e) applicability in the regulatory context.

Aquatic life uses are the designated uses that this report addresses in establishing potential sediment benchmarks. Because the scientific literature is replete with examples of sediment effects on stream biota (Waters 1995, Wood and Armitage 1997, Suttle et al. 2004, Opperman et al. 2005), the assumption is that each of the indicators is associated with the designated use, satisfying the first selection criterion (a) listed above. Each of the indicators being considered is a direct measure of either absolute sediment conditions or conditions relative to expectations for the environmental setting. This assumption of biological effects was confirmed in our examination of sediment-response relationships.

Sufficient data existed in and around New Mexico for analysis of all indicators listed in Table 1 and their derivatives based on environmental settings. The fewest samples were from targeted riffle sites, which may limit the certainty of some of those analyses. However, the targeted riffle sites were important in determining the sensitivity of the methods, so they were not eliminated (see Jessup et al. 2010 for targeted riffle analyses). Particle embeddedness was originally considered as a potential indicator, but it was abandoned because the workgroup had a general lack of confidence in the consistency of the methods used in its measurement.

All of the indicators were appropriate for the data analysis methods considered. Where data gaps existed in ancillary data, those data were removed from analysis or, if appropriate, mean values were substituted for missing values. For analyses that required normal distributions (e.g., PCA, Pearson correlations), variables were transformed using the logarithmic or Arc Sine – Square Root functions. Sediment indicators were averaged over time at individual sites when data from multiple sampling events were available.

Some indicators require more effort in the field for accurate measurement. Percent sand & fines, % fines, and turbidity are among the simplest indicators to sample. Relative bed stability (RBS) measures require more stream channel measurements or watershed characterization. For these reasons, much of the indicator characterization focused on comparisons among indicators. If simpler measures could substitute for more complex ones without sacrificing assessment accuracy, the simpler measure might be adequate and more efficient. If more complex measures were necessary for the most accurate assessments, then the simple measures may only suffice for screening applications. In contrast to LRBS, the percent fines and percent sand & fines measures are absolute quantities, which, except for natural variability captured by site classification, are more susceptible to natural variations.

Preliminary analyses suggested that log RBS (LRBS) calculated without bedrock or hardpan components (termed LRBS_NOR) might improve associations between the bedded sediment measure and biological responses. The LRBS_NOR was a measure that regarded only the potentially mobile streambed particles in determining the geometric mean particle size. Calculated values related to the mobile particle distribution were missing about 25% of the records. In those cases, LRBS_NOR was estimated from LRBS (code = LRBSfin) and the percentages of bedrock and hardpan. Only 10% of cases required any adjustment because 15% had no bedrock or hardpan.

Reference distributions describe expectations in least disturbed sites, which are classified by natural site types. Percentiles of the reference distributions can be used as one line of evidence for establishing sediment benchmarks. Percentiles that are used to compare sites to reference conditions usually are from the median to the 90th percentile for values that increase with disturbance, such as % sand & fines. Below the median, conditions are similar to the best half of the reference sites. Values above the 90th percentile are unlike most of the reference sites and 10% of reference sites might have high values due to sampling and natural variability. Typically, quartiles are used to define similarity to or separation from the reference condition (Barbour et al. 1999, USEPA 2000, USEPA 2006).

Using the classification scheme to reduce natural variability of the indicators, the sediment conditions in reference sites were established. The reference sediment conditions are expected at all minimally disturbed sites of the same site class. Potential benchmarks were derived from the 10th and 25th percentiles for the LRBS indicator that decreases with increasing disturbance, and the 75th and 90th percentiles for the percentage indicators that increase with disturbance. Selection of benchmarks using these percentiles has precedence in state and federal biological and nutrient criteria programs (USEPA 2006).

Discrimination efficiency (DE) is a measure of the difference in values from reference sites to stressed sites. It is measured as the percentage of sites in stressed sites with indicator values worse than the 75th percentile of reference values. For the LRBS, stressed sites values below the 25th percentile of reference were used because these values are typically lower with greater stress.

From the statistics in Table 6, the % sand & fines indicator from reachwide sampling was among the best at discriminating stress in each region (see Jessup et al. 2010 comparison to the residuals of these values or indicators from target riffle sampling). Percent fines performed similarly, except that it performed poorly in the Xeric region. LRBS was not best at discriminating stress as defined here. LRBS 25th percentile of reference values are as expected and equal in the Mountains and Foothills.

Table 6. Bedded sediment indicator statistics for reference sites in three site classes.

Indicator	Ref N	Mean	Min	10%ile	25%ile	Median	75%ile	90%tile	Max	Std.Dev	DE[a]
Mountains											
PCT_SAFN	55	16	0	2.2	5.6	13.3	20.6	35.1	65.7	14.1	>50%
PCT_FN	55	9.7	0	0	2.8	5.7	12.9	24.6	61.9	11.7	>50%
LRBS_fin	55	-0.7	-2.2	-1.5	-1.1	-0.7	-0.4	-0.1	0.5	0.6	25-50%
LRBS_NOR	55	-0.8	-2.2	-1.5	-1.1	-0.8	-0.4	-0.1	0.4	0.5	25-50%
FootHills											
PCT_SAFN	27	27.7	0	5.1	18.6	27.1	36.9	45.2	61	15.5	<75%
PCT_FN	27	11.2	0	0.9	2.4	10.9	18.6	23.7	31.8	9.6	75%
LRBS_fin	27	-0.8	-2	-1.6	-1.1	-0.7	-0.4	0	0.3	0.6	>50%
LRBS_NOR	27	-0.9	-2.0	-1.7	-1.3	-1.0	-0.5	-0.1	0.2	0.6	50%
Xeric											
PCT_SAFN	17	57.2	19.5	29.8	43.8	53.3	74.3	84.4	99	22.8	>50%
PCT_FN	17	38.7	4.8	8.2	18.1	34.3	59	72.2	86.7	26	<25%
LRBS_fin	15	-1.8	-2.8	-2.5	-2.3	-1.9	-1.3	-0.9	-0.2	0.7	25-50%
LRBS_NOR	15	-2.0	-2.9	-2.7	-2.5	-1.9	-1.6	-1.1	-0.7	0.7	25-50%

NOTES: See Table 1 and Appendix A for indicator definitions. [a] DE = Discrimination Efficiency

Step 6: Describe stressor-response relationships

In this step, the biological responses to sediment conditions are explored using several statistical techniques. Three analytical techniques were use -- reference distributions, quantile regression, and change-point analysis.

Quantile regression and change-point analysis compare sediment conditions with biological conditions, using the biological conditions to indicate the degree to which aquatic life uses are supported. Quantile regression is a method for estimating functional relations between variables along the upper boundary of the conditional distribution of responses (Cade et al. 1999). If limiting factors such as sediments act as constraints on organisms, then the estimated effects for the measured factors are related to some upper limit. This is apparent when the biological measures tend to exhibit an upper limit that varies with the value of a disturbance variable — that is, the maximum biological condition generally falls beneath a sloping line in a scatter plot of biological condition against the disturbance variable. Points that are not along the slope (in the heel of the wedge) represent sites with worse biological conditions due to factors not represented on the *x*-axis. The slope represents biological potential (plotted on the *y*-axis) in relation to the disturbance of interest (plotted on the *x*-axis). Estimation of the limiting slope is accomplished through quantile regression, which was performed using R software (R Development Core Team 2010) and associated code (quantreg).

The quantile regression analysis was conducted such that several upper quantiles (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). This tells us there is a likely effect, but the point at which the effect becomes critical cannot be directly determined. The multiple upper quantiles were examined and parallelism was determined based on professional judgment regarding the consistency of the slopes and the meaningfulness of the 90th quantile regression line (good, flat, or inconsistent). When the 90th quantile regression line was good, it was plotted to illustrate the change in a biological resource for each increment of sediment disturbance.

The change-point is the point along an environmental gradient at which there is a high degree of change in the response variable. The data are divided into two groups, above and below a potential sediment benchmark, where each group is internally similar and the difference among groups is high. To determine the change-point, nonparametric deviance reduction was used (Qian et al. 2003, King and Richardson 2003) to identify benchmarks in biological responses to sediments. This technique is similar to regression tree models, which are used to generate predictive models of response variables for one or more predictors. Using this comparison, the change-point is the first split of a tree model with a single predictor variable (i.e., fine sediment percentage). Change-points and statistical significance were obtained using R software (R Development Core Team 2010) and associated code (chngp.nonpar).

One caveat of the change-point analysis is that a change-point may be identified, and even determined to be statistically significant, when the change-point value is actually only an artifact of the analysis and not an indication of a change in system properties. The method always finds a change-point, even in a dataset with a perfect straight line relationship between *X* and *Y*. It has been well established that sediment size affects macroinvertebrate assemblage characteristics (e.g., Waters 1995, Wood and Armitage 1997, Suttle et al. 2004, Opperman et al. 2005). Therefore, it is reasonable to believe an ecological benchmark does exist between certain biological metrics and sediment conditions. In our analyses, this relationship was evaluated by examining the locally-weighted regression line (LOWESS or loess) fit on biplots of biological metrics and sediment indicators. If the LOWESS fit did not show a change-point, then the value identified through change-point analysis was disregarded.

The LOWESS technique (Cleveland 1979) is designed to address nonlinear relationships, which may be important when investigating changing responses along a stressor gradient. LOWESS combines the simplicity of linear least squares regression with the flexibility of nonlinear regression. It achieves this by fitting simple models to localized subsets of the data to build up a function that describes the deterministic part of the variation in the data, point by point. LOWESS fits segments of the data to the model, essentially, at the central tendency of the data. This method does not require specification of a global function of any form to fit a model to the data but to simply fit segments of the data to the model. A bandwidth was used that considered 75% of the data for smoothing the slope at each data point. The LOWESS regression line can be used in combination with other indicators of sediment effects, primarily as a visual confirmation of changing biological measures at certain sediment indicator values.

The relationships between bedded sediment indicators and biological metrics were examined using bi-plots showing significant change-points, meaningful quantile regression lines, the LOWESS regression line, and reference points in comparison to non-reference points. From these plots, three things can be discerned:

1. When the sediment variable appears to be limiting the metric potential,
2. If the calculated change-point coincides with changes along the LOWESS regression line, and
3. If the change-point reasonably separates reference and non-reference points.

In Figures 6-12 below, significant change-points are shown as vertical lines, meaningful 90th quantile regression lines are shown as diagonal dash-dot lines, LOWESS regression lines are shown as non-linear solid lines, reference points are shown as solid circles, and non reference points are shown as crosses. For example, in Figure 6, the LOWESS regression line is variable but high until about 15% sand & fines, where it begins to drop. The significant change-point at 20% sand & fines is at the part of the LOWESS line that drops below a previous trough. The two indications are in general agreement. In addition, the quantile regression lines are relatively parallel, strengthening our case that the stressor is limiting the biological potential. Reference points are more common to the left of the change-point. In all, the 20% sand & fines change-point is a potential benchmark.

Final recommendation of a biological benchmark for an indicator in a site class would be based on corroborated results from all the metrics and the NMMSCI. In some cases, the lowest effect level may be appropriate to recommend as a benchmark. In other cases, the potential benchmarks indicated by significant change-points may be inappropriate and benchmarks may be recommended based on other indications. Metrics that do not show a significant change-point or do not appear to be limited by the stressor would not be used in determining a benchmark.

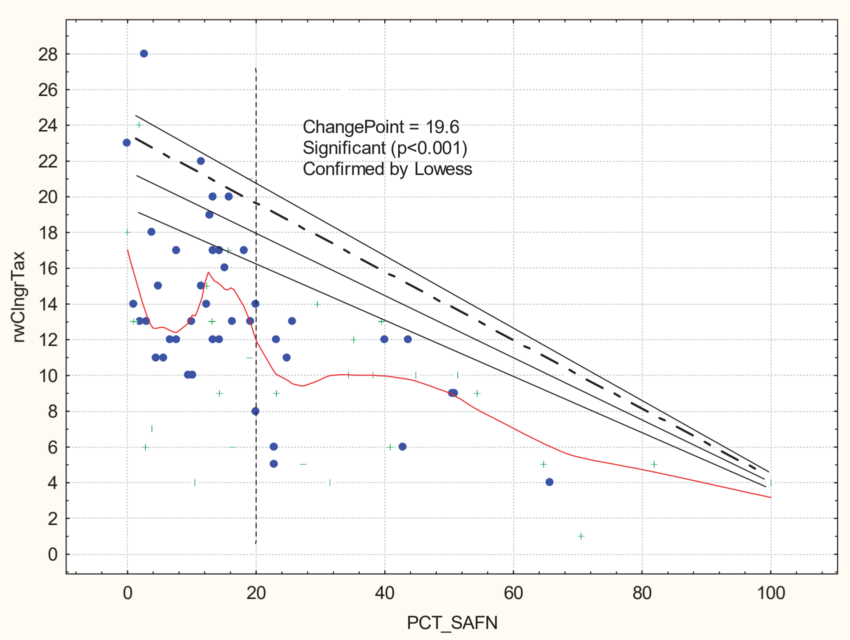

Figure 6. Example of significant change-point that is recognizable as a change in the y-axis value with respect to the x-axis. The LOWESS regression line shows an apparent shift near the change-point. The upper quantile regression lines are generally parallel, suggesting a true biological limitation. rwClngrTax = reachwide number of clinger taxa.

Only the % sand & fines, and LRBS_NOR plots by site class are included in this document. To see all change-point analyses graphs and full discussion of each, see Jessup et al. 2010.

Based on the quantile regression and change-point analyses of stressor-response in the three proposed site classes, possible effect levels to benthic macroinvertebrates are summarized in Table 7. For % sand & fines and % fines, the amount of fine sediments that appears to affect benthic macroinvertebrates is lowest in the mountains and highest in the xeric areas.

Table 7. Summary of recommended benchmarks based on biological responses (i.e., quantile regression and change-point analysis)

Metric \ Site Class	**Mountains**	**Foothills**	**Xeric**
% sand & fines	20%	50-60%	74%
% fines	15%	22%	29%
LRBS	-1.25 units	-1.8 units	-1.0 units
LRBS_NOR	-1.1 units	-1.5 units	-1.25 units

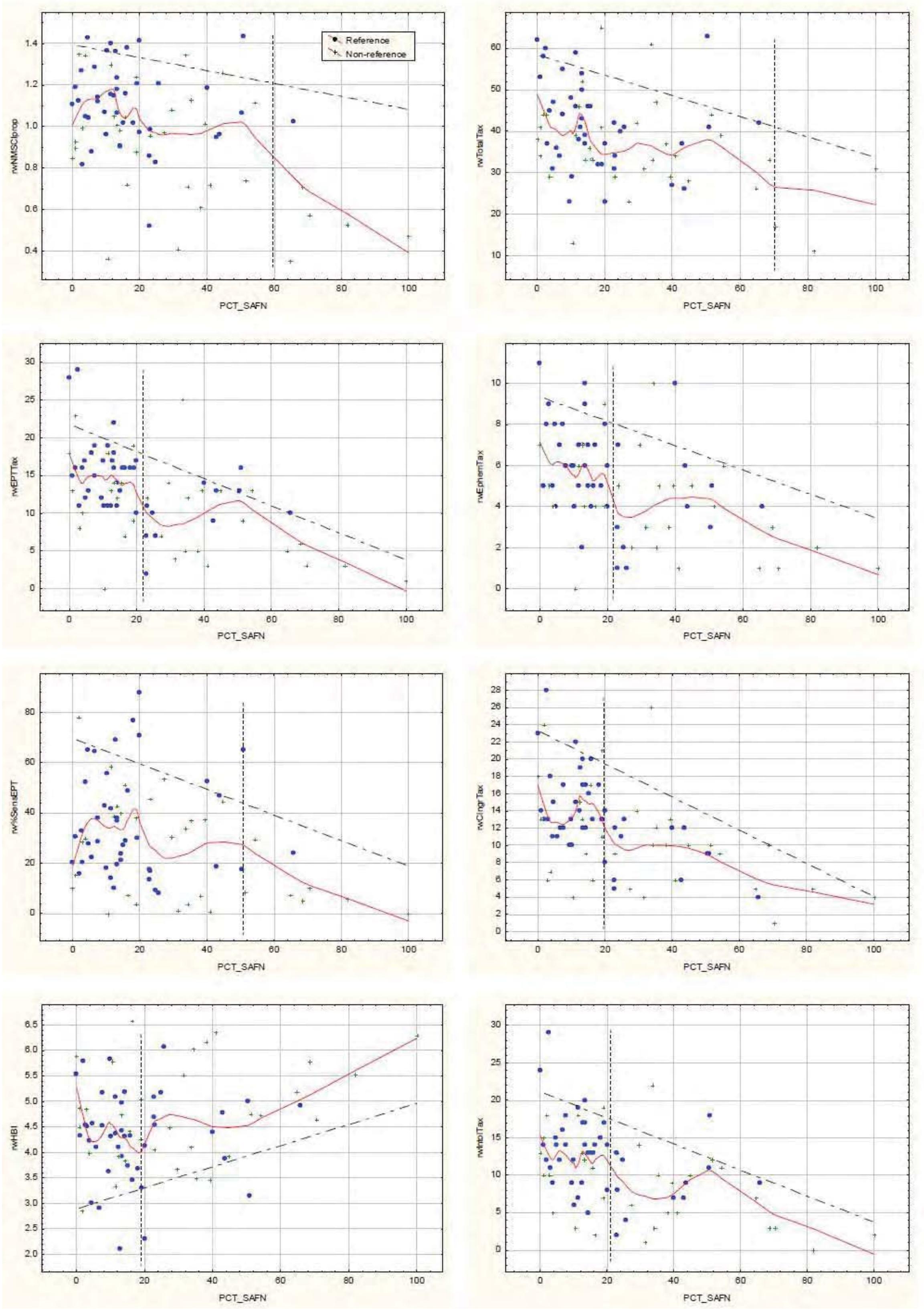

Figure 7. Biological responses to % sand & fines in the Mountain site class.

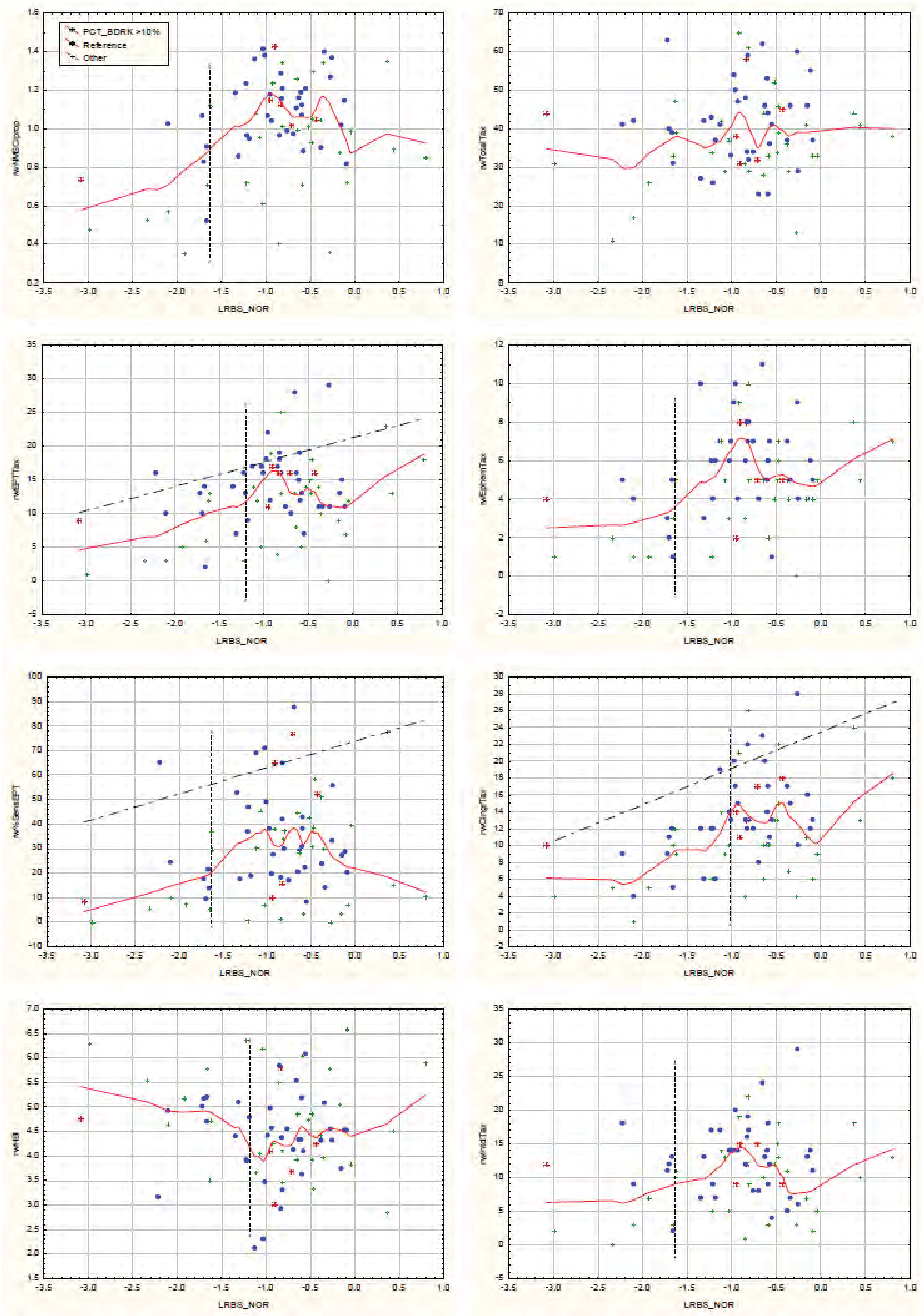

Figure 8. Biological responses to LRBS_NOR in the Mountain site class.

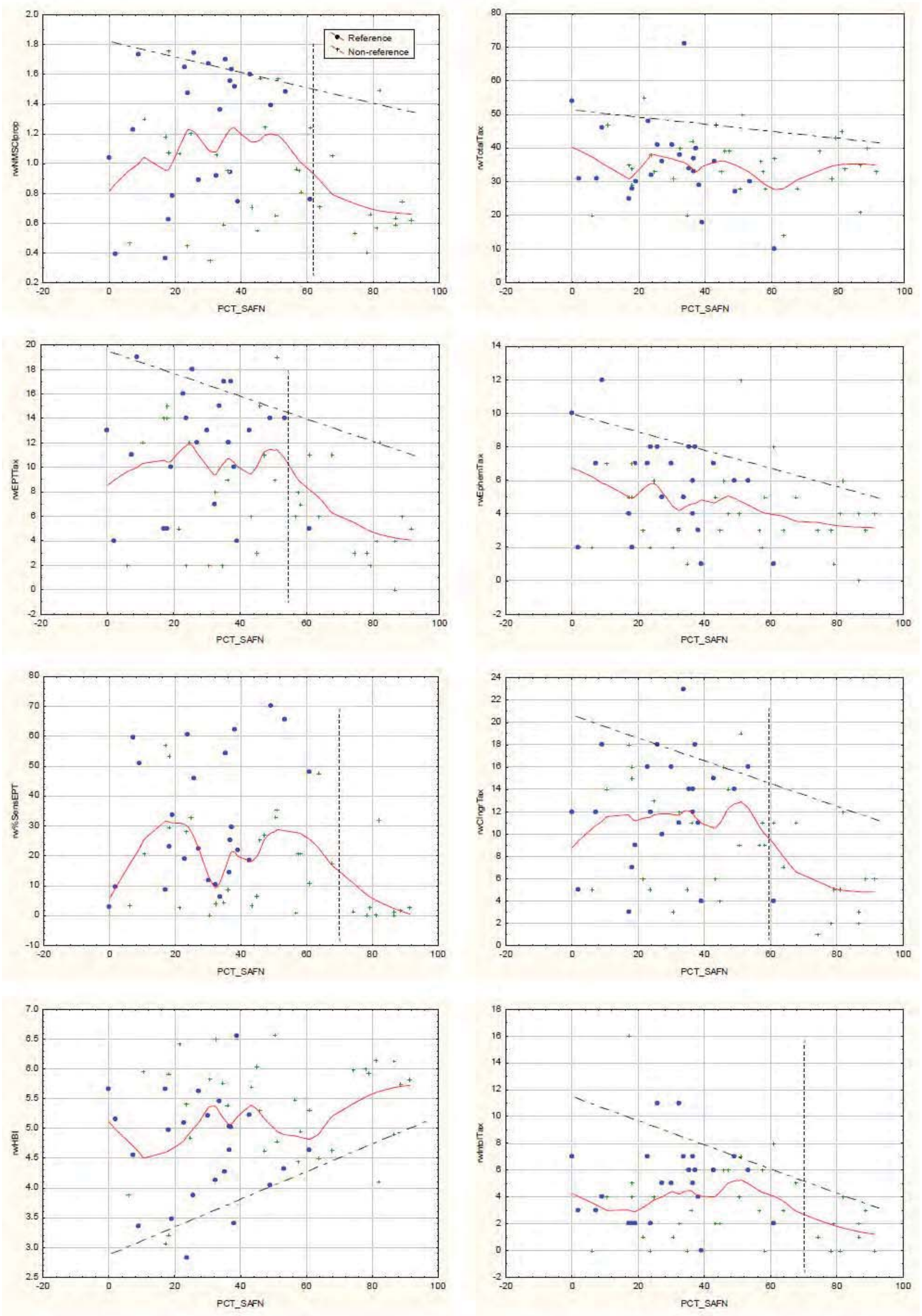

Figure 9. Biological responses to % sand & fines in the Foothills site class.

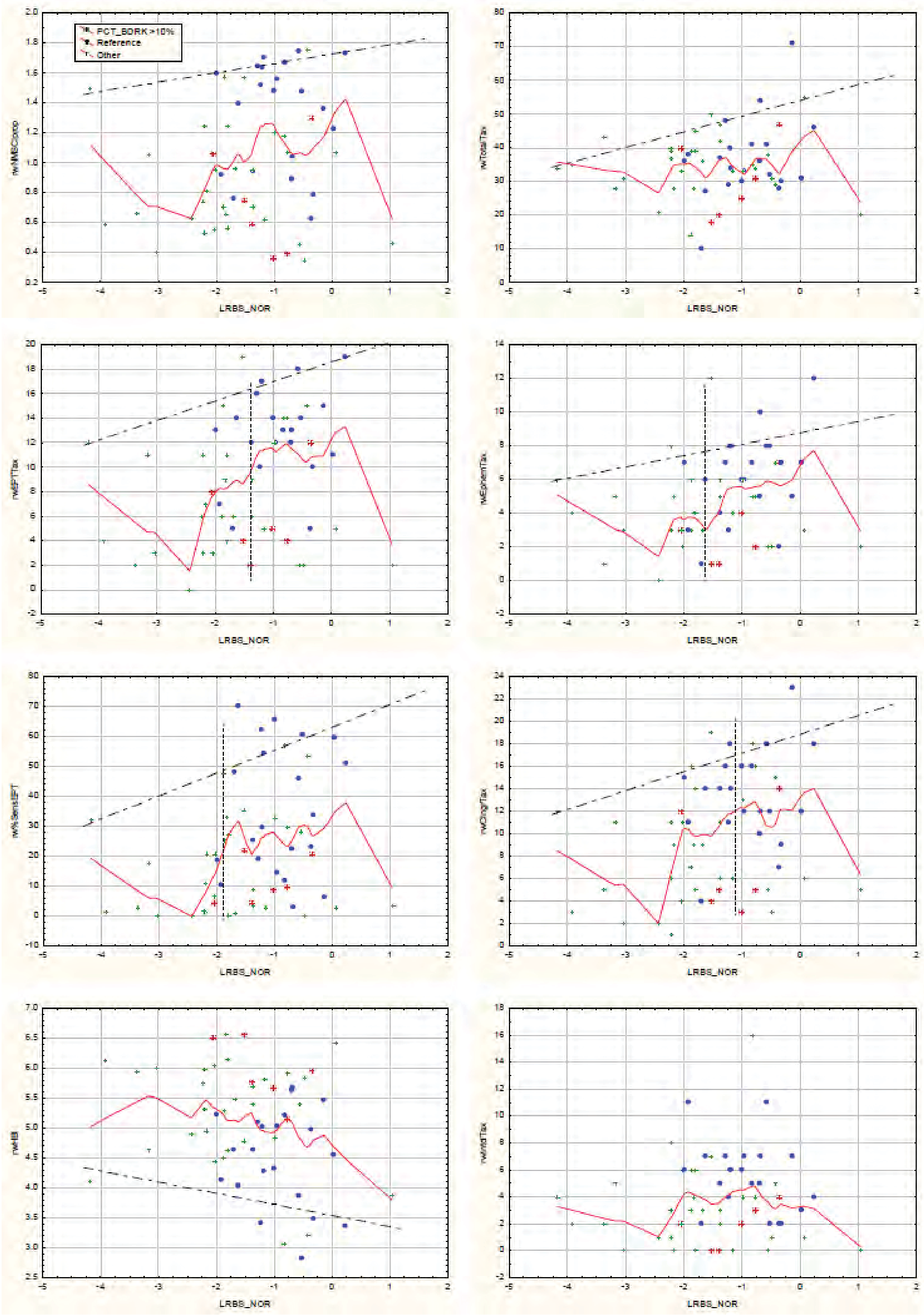

Figure 10. Biological responses to LRBS_NOR in the Foothills site class.

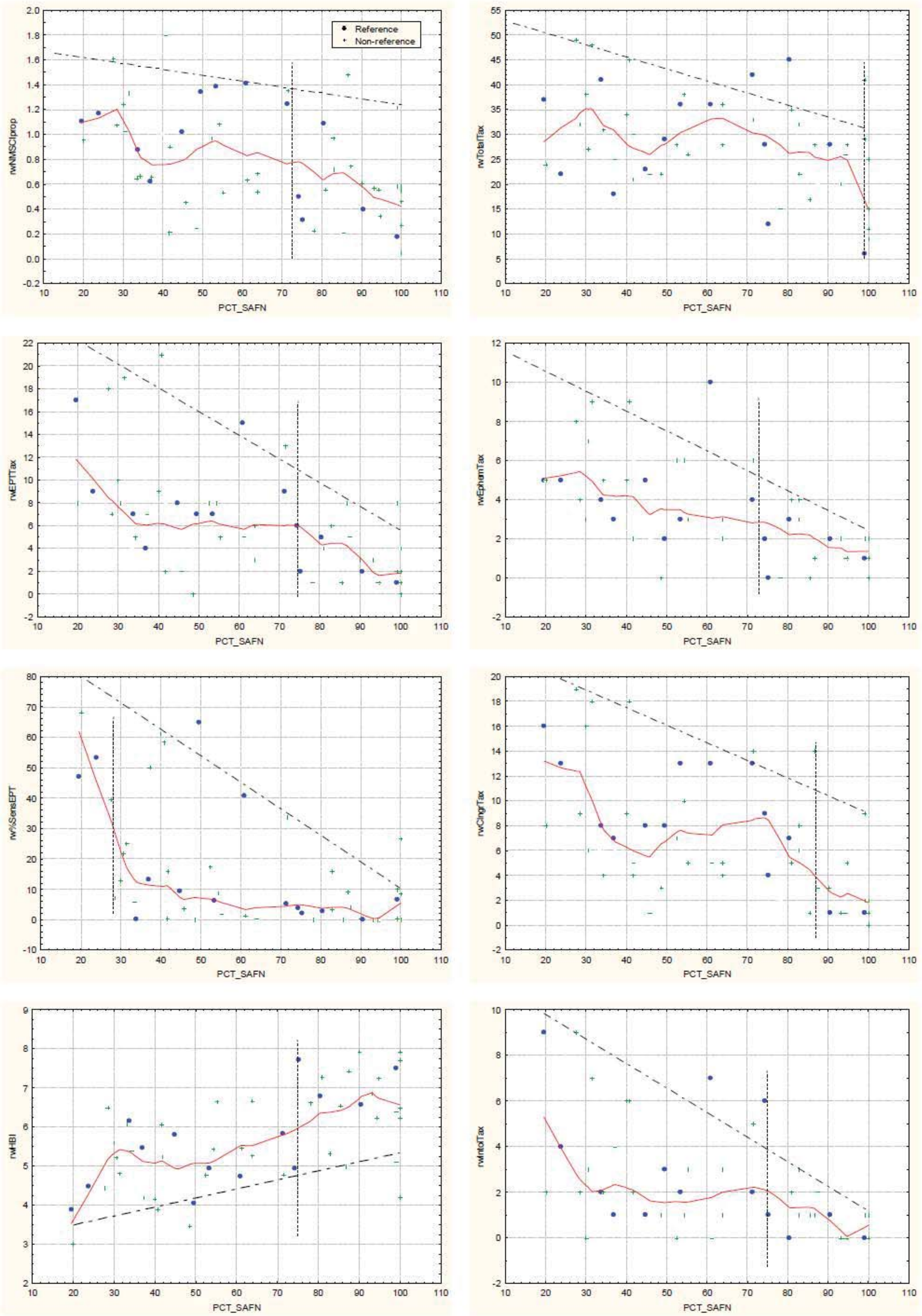

Figure 11. Biological responses to % sand & fines in the Xeric site class.

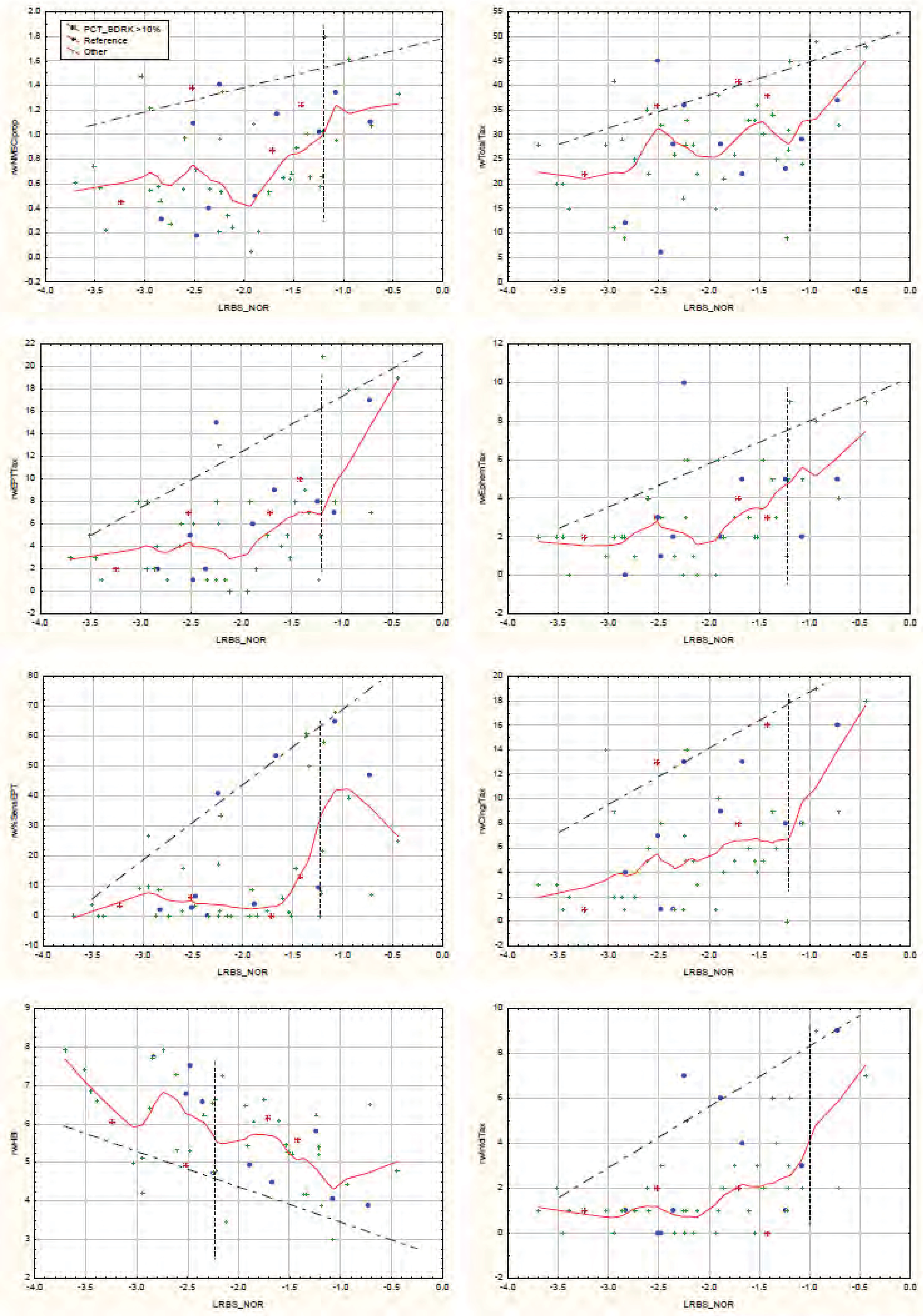

Figure 12. Biological responses to LRBS_NOR in the Xeric site class.

Step 7: Recommend benchmarks or thresholds

In this step, potential sedimentation/siltation benchmarks are recommended for each site class based on the results of step 6.

For each sediment indicator and site class, a benchmark value was recommended through a weight-of-evidence approach that considered the multiple analytical approaches presented above and the strength of each analysis. When formulating a recommendation, considerable weight was afforded the sediment reference condition statistics. These values are the most direct measures of expected sediment conditions in undisturbed sites throughout New Mexico. The sediment reference conditions are characterized independently of the biological conditions. However, 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 analysis of associations among sediment and biological indicators. Biological effects are less direct indicators of required sediment conditions because the biota are undoubtedly affected by other environmental conditions, not just sediments.

The purpose of the preceding analyses was to provide assessment tools by identifying benchmarks of sediment conditions that can be related to protection or impairment of natural ecological systems, including the biological components. To facilitate both broad, general assessments and detailed evaluations of sediment conditions, a two-tiered procedure is recommended. In the first tier, a measure that is relatively easy to collect and communicate will be compared to a set of known condition benchmarks. This would give a general indication of the sediment conditions, potential biological impairment, and indicate when the second tier is needed. The second tier would give more detailed and different information, allowing refined geomorphic interpretation of the first-tier results.

For bedded sediments, the sediment indicators that are most appropriate for first- and second-tier assessments are % sand & fines and Relative Bed Stability excluding bedrock and hardpan (LRBS_NOR). Percent sand & fines are appropriate because they have precedent in NMED standards, show expected relationships with biological measures, are relatively easy to measure consistently, and are straightforward to communicate. Percent sand & fines may be a better measure than % fines because the sand component has similar modes of biological effect as finer material and fines alone are more variable across sites (and are relatively rare in streams with normal to powerful flows). In Xeric sites, the sand component may be relatively common and the fines may take on more importance.

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. This allows evaluation of the potential of the specific site in terms of retaining or flushing fine sediments. When used as a second-tier assessment tool, LRBS_NOR could help explain whether high % sand & fines were expected for a given site or are a result of disturbed conditions. LRBS_NOR can also be used to identify sites with deficient fine sediments, though this condition was not fully explored in the current analysis.

The way that the two indicators can be used for a two-tiered assessment could be as follows for a given site. First, identify the site class and associated benchmarks for % sand & fines and LRBS_NOR. Second, compare the observed % sand & fines to the benchmark and determine whether there is a potential sediment impairment. Then, assuming LRBS_NOR data were collected, compare the observed value to the benchmark to help interpret site-specific sediment conditions. If there is no potential impairment indicated in the first tier, there may not be a need to use the second-tier assessment.

The benchmarks for first- and second-tier assessments can be guided by this report and analysis. However, the final decision on benchmarks is the responsibility of NMED, which may consider factors beyond the scope of this report. Therefore, the following suggestions for benchmark establishment and application are meant only to serve as a starting point.

From the two types of analyses, biological responses and reference distributions, summary statistics and benchmarks for the two recommended indicators are summarized (Table 8). 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.

Table 8. Summary of potential benchmarks based on biological responses and reference distributions. Recommended benchmarks are **bolded** and discussed further in the text below.

Metric \ Site Class	Mountains	Foothills	Xeric
Biological Effects			
% sand & fines	**20%**	50-60%	**74%**
% fines	15%	22%	**29%**
LRBS_NOR	**-1.1 units**	-1.5 units	-1.25 units
Reference Distributions			
% sand & fines (reference 75 %ile)	21%	**37%**	**74%**
% sand & fines (reference 90 %ile)	35%	45%	84%
% fines (reference 75 %ile)	13%	19%	59%
% fines (reference 90 %ile)	25%	24%	72%
LRBS_NOR (reference 25 %ile)	-1.1 units	**-1.3 units**	**-2.5 units**
LRBS_NOR (reference 10 %ile)	-1.5 units	-1.7 units	-2.7 units

In the following graphs (Figures 13 - 15), observations in the upper right quadrant show impairment using the Tier 1 indicator, but not using the Tier 2 indicator. In these sites, higher percentages of sand & fines may be natural and impairment may not be indicated. Observations in the upper left quadrant could be assessed as impaired with a high degree of confidence. Likewise, sites in the lower right quadrant could be called unimpaired with a high degree of confidence. Sites in the lower left quadrant have low % sand & fines (passing the Tier 1 benchmark) and low LRBS_NOR values (failing the Tier 2 benchmark). If a site was only assessed using Tier 1 measures, it would not show impairment and might not warrant a Tier 2 assessment. If a Tier 2 assessment was available, the site's assessment would be somewhat uncertain, as it would show low % sand & fines, but otherwise unstable substrates.

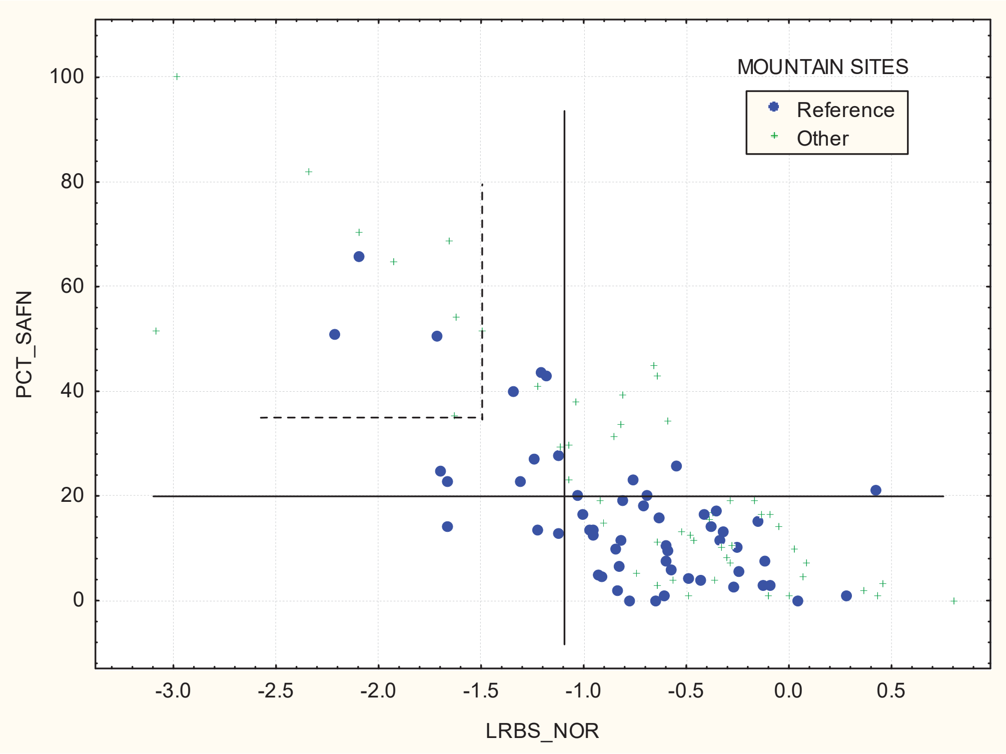

Figure 13. Sediment benchmarks suggested through analyses of biological effects (solid, recommended) and alternative benchmarks based on the 10th/90th reference percentiles (dashed) for the Mountain site class.

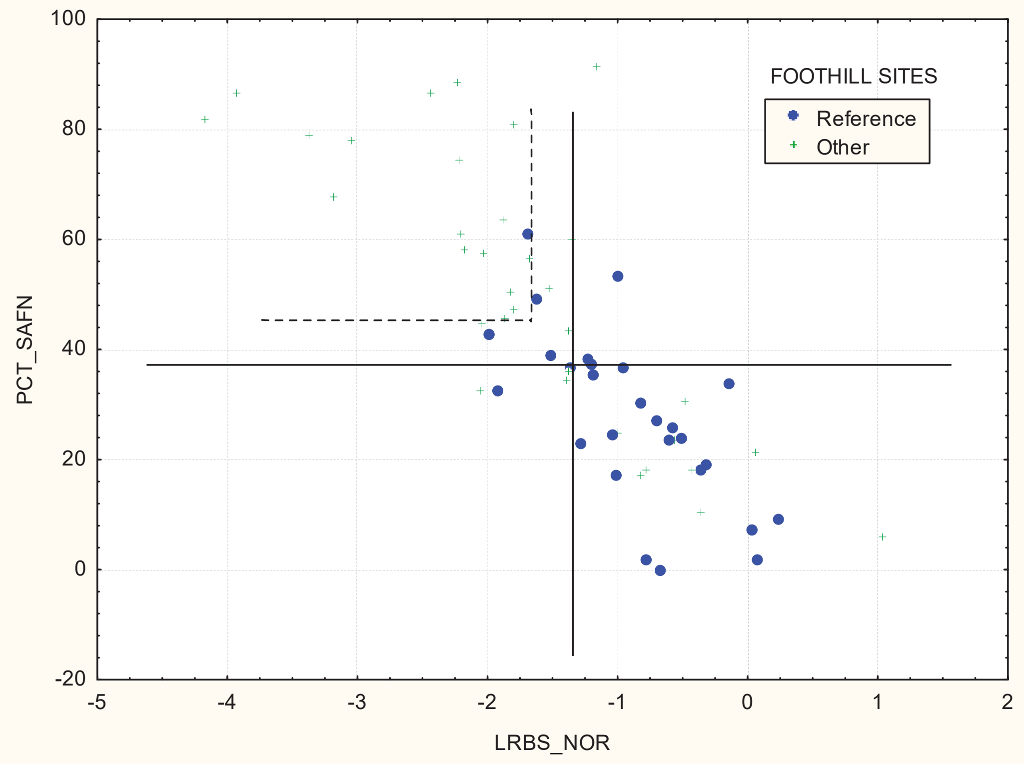

Figure 14. Sediment benchmarks suggested through analyses based on the 25th/75th reference percentiles (solid, recommended) and alternative benchmarks based on biological effects (dashed) for the Foothill site class.

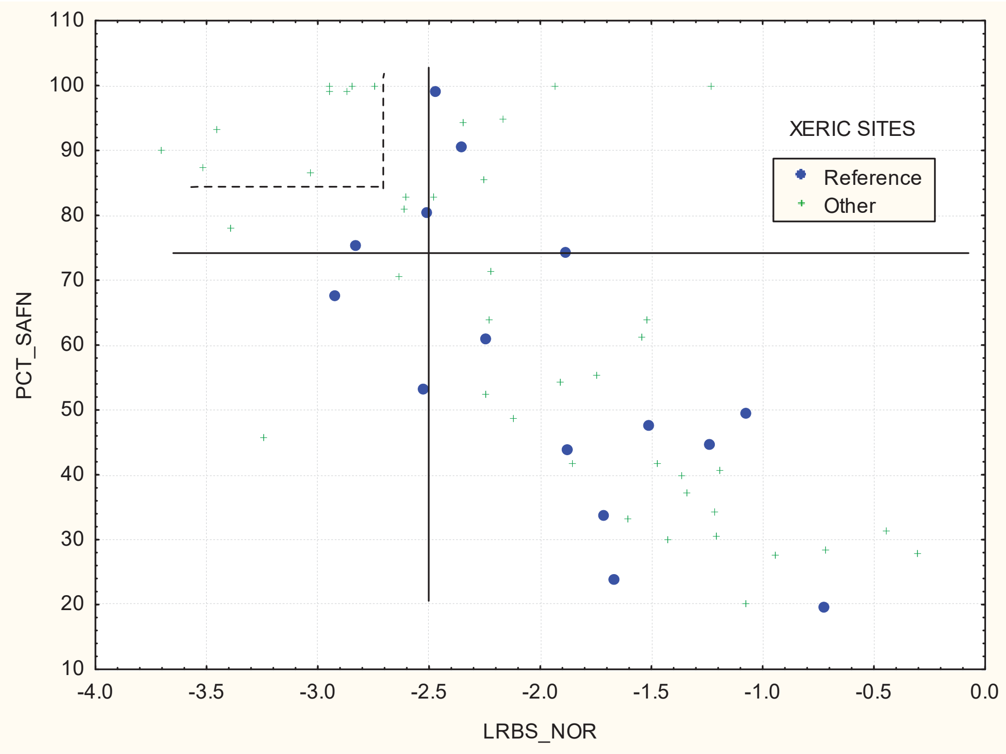

Figure 15. Sediment benchmarks suggested through combined analyses of biological effects and reference percentiles for the Xeric site class. Solid benchmarks are recommended.

Analysis of multiple sediment indicators, their responsiveness to site disturbance, and their effects on benthic macroinvertebrates resulted in identification of potential benchmarks for the bedded sediment indicators % sand & fines and Relative Bed Stability (LRBS_NOR) in three site classes, Mountains, Foothills, and Xeric areas. The site classes distinguish sediment expectations across the State and were identified through a principal components analysis (PCA) of environmental conditions and the sediment indicators. Percent sand & fines are easily measured and related strongly with biological metrics. LRBS_NOR is a formulation that considers site-specific hydraulic potential for moving bed sediments, so that the observed fine sediments are only considered imbalanced when the streambed is more easily mobilized and transported than expected due to unstable stream bed conditions. The two indicators can be applied in a two-tiered assessment that first considers the simpler indicator of biological impairment, and then refines the assessment with the second indicator of geomorphic impairment, as needed. Recommended benchmark values are as follows (Table 9):

Table 9. Recommended benchmarks

Site Class	**Tier 1 % sand & fines**	**Tier 2 LRBS_NOR units**
Mountains	< 20	> -1.1
Foothills	< 37	> -1.3
Xeric	< 74	> -2.5

Conclusion and Research Needs

Through multiple lines of evidence, bedded sediment conditions were determined that supported benthic macroinvertebrate assemblage integrity and represented least-disturbed reference conditions. The recommended benchmarks are specific to site classes that incorporate much of the natural variability in sediments within the reference data. They can be applied in a two-tier system that allows for rapid or detailed assessments.

Natural variability derived from landscape-scale factors is accounted for by the site classes, which are determined by the level IV ecoregions. It is noted that fine sediments were more likely to be high in non-reference sites with erodible lithologies and less likely in any sites with resistant lithologies. Erodibility is related to the landforms and geologies that contribute to ecoregion definition, and to some degree site classes account for erodibility effects. In the assessment approach to be defined by NMED, a qualitative statement about lithologic erodiblility and the likelihood of detecting excessive fine sediments may be appropriate for each site. For those sites with resistant lithologies, assessments that show no sediment impairment would not imply a lack of disturbance at the site or in the catchment.

The recommended benchmarks are quite different among site classes. Expectations for the Mountains are that % sand & fines will be low and LRBS_NOR will be high. In comparison, expectations in the Xeric areas are that fine and mobile 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.

In Table 10, "agreement" is the percentage of site assessments that agree with the reference status, or in other words, the percentage of reference sites passing the benchmark and the percentage of stressed sites failing the benchmark. Agreement percentages near 75% in reference sites reflects use of the 75th or 25th percentile of reference when selecting benchmarks. When applying the two-tiered system with impairment indicated only by failing both indicator benchmarks, the correct identification of reference sites improves relative to each single indicator, while the correct identification of stressed sites is equal or somewhat worse.

Table 10. Agreement of assessment results and reference status for individual indicators and the two-tiered assessment system.

Site Class	% sand & fines benchmark	Agreement % sand & fines Ref/Strs	LRBS_NOR benchmark	Agreement LRBS_NOR Ref/Strs	Agreement 2 tiers Ref/Strs
Mountains	< 20	71/50	> -1.1	75/33	80/33
Foothills	< 37	74/56	> -1.3	78/56	85/44
Xeric	< 74	71/60	> -2.5	76/30	87/30

Type I and Type II assessment errors should be considered. Type I errors, incorrectly assessing a reference site as impaired, occurs in 13 to 20% of reference sites using the two tiered system in this analytical data set. Type II errors, incorrectly assessing a stressed site as un-impaired, occurs in 56 to 70% of stressed sites. The error appears to be unbalanced, allowing impairment to go undetected in many sites. There are valid reasons for accepting higher Type II than Type I error. A prominent effect observed was that disturbance at the site or in the watershed often did not result in higher stream sediments in sites with resistant lithology. Since the stressed sites are not defined using measures that certainly cause sediment stress, greater Type II error in the sediment indicators is expected. A lack of disturbance, as measured using our reference site criteria, appears to reflect a lack of factors that cause sediment stress. In addition, there are more reference sites than stressed sites. Therefore, there is more confidence in the definition of reference sites and the agreement of the indicators with that definition than there is for stressed sites and agreement in that group.

While the presented analyses inform decisions regarding benchmark selection for sediments in New Mexico streams, they also generate additional questions that could not be thoroughly answered without further effort.

Precision analysis on the indicators would allow NMED to assess measurement error and temporal variability, which would enhance interpretation of sediment assessment certainty. Many of the data points used in this analysis were from single grab samples or observations. Multiple data points were available for each site in a small number of cases, and average values were used in analyses. Same-day replicate sampling was not evident in the data set, but would be necessary to assess measurement error while controlling for temporal changes.

Collection of sufficient data for calculation of LRBS is somewhat cumbersome or time consuming in the field. NMED has expressed interest in estimation of conditions in general categories as an alternative to extensive measurements. Conceptually, estimated channel roughness is feasible, but should be tested before implementation. Testing could include comparison of LRBS calculated from both quantitative measures and estimations, including estimations made by different field crew members for calibration.

Lithologic erodibility appears to control the degree to which disturbance at a site or in the catchment results in excess fine or mobile sediments. The workgroup hypothesized that minor changes in sediments in resistant lithologies might have comparable biological effects as major changes in erodible lithologies. The reasoning was that biota in resistant lithologies are more dependent on stable conditions than biota in streams with commonly shifting streambeds. Because reference streams in resistant and erodible lithologies were indistinguishable as site classes – they had similar sediment characteristics – the sediment sensitivities of biota within the two erodibility groups were not assessed. Information on minor sediment-biota relationships in resistant lithologies would help identification of sediment stress in resistant lithologies, but would require sensitive and precise sediment and biological measures. Lithologic erodibility may be important in suspended sediment analysis also, and such data would be worth developing.

In our analysis, it appeared that biota were more responsive to absolute measures of sediment composition than to measures of sediment relative to the stream potential, like LRBS or the residual measures of % fines and % sand & fines. However, differences in sediment composition occur not only in response to disturbance, but also in the full range of natural settings. The fauna that are naturally adapted to fine and mobile sediments occur with fine sediments, but may not indicate disturbance unless the presence of such sediments are unexpected. The measures used in the recommended two-tiered system include one absolute measure and one relative measure. The requirement that both indicators fail the benchmarks before impairment is assessed assures that the signal at the site is of both regionally abundant fine sediments and sediment amounts that are more than the stream can efficiently transport. Our understanding of the conditions to which biota respond would be enhanced with additional examination of the biotic responses to relative sediment supply along the natural gradient of fine sediment abundance. Obviously, the causes and mechanisms of effects would be difficult to tease apart.

Assessment of multiple biological assemblages would enhance interpretations of stressor-response relationships. Fish samples may be obtained as well as macroinvertebrate samples to show different sensitivities among the entire biotic system. Such data were not available for the current effort.

Unbalanced sediments undoubtedly occur in the presence of other stressors. A multiple stressor analysis focused on isolating sediment effects may not be possible without sediment chemistry, but other correlated factors could be assessed. In our analysis, sediment indicators and biological metrics were correlated, which implies a causative relationship that was never proven. Causative analyses may be infeasible, though more effort could be applied towards identifying sites in which multiple stressors are either prominent or lacking.

References Cited

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

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, 2010, 29(2):657–672.

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.

Jacobi, G.Z., M.D. Jacobi, M.T. Barbour, and E.W. Leppo. 2006. Benthic macroinvertebrate stream condition indices for New Mexico wadeable streams.

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

Jessup B. 2009b. Recalibration of the Macroinvertebrate Multi-Metric Index for Colorado. Prepared for the Colorado Department of Public Health and Environment, Water Quality Control Division – Monitoring Unit, Denver, CO 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. Prepared for the U.S. Enviromental Protection Agency, Region 6, Dallas, TX and the New Mexico Environment Department, Santa Fe, NM. Prepared by Tetra Tech, Inc., Montpelier, VT.

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

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 (JAWRA) 45(2):434-459.

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.

New Mexico Environment Department Surface Water Quality Bureau (NMED SWQB). 2011. Procedures for Assessing Use Attainment for the State of New Mexico Integrated Clean Water Act §303(d) List and §305(b) Water Quality Monitoring and Assessment Report. Santa Fe, New Mexico.

New Mexico Water Quality Control Commission (NMWQCC). 2011. *State of New Mexico Standards for Interstate and Intrastate Surface Waters.* 20.6.4 NMAC as amended through January 14, 2011. Available at: ftp://ftp.nmenv.state.nm.us/www/swqb/Standards/2010/20.6.4NMAC-Integrated2010-12-15.pdf

Opperman, J.J., K.A. Lohse, C. Brooks, N.M. Kelly, and A.M. Merenlender. 2005. Influence of land use on fine sediment in salmonid spawning gravels within the Russian River Basin, California. Canadian Journal of Fisheries and Aquatic Sciences 62: 2740–2751.

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):1-17.

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, MD.

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. Available at: http://www.epa.gov/wed/pages/publications/authored/EPA620R-06003EMAPSWFieldOperationsManualPeck.pdf

Qian, S.S., King, R.S., Richardson, C.J., 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. Available at: http://www.R-project.org

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. 2005. Environmental Monitoring and Assessment Program (EMAP): western streams and rivers statistical summary. EPA 620/R-05/006. Office of Research and Development, US Environmental Protection Agency, Washington, DC.

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.

Suttle, K.B., M.E. Power, J.M. Levine, C. McNeely. 2004. How fine sediment in riverbeds impairs growth and survival of juvenile salmonids. Ecological Applications 14(4): 969–974.

U.S. Environmental Protection Agency (USEPA). 2000. Nutrient Criteria Technical Guidance Manual; Rivers and Streams. Office of Water, office of Science and Technology. EPA-822-B-00-002.

USEPA. 2006. Framework for developing suspended and bedded sediment (SABS) water quality criteria. Office of Water, Office of Research and Development. EPA-822-R-06-001.

Waters T.F. 1995. Sediment in streams—Sources, biological effects and control. American Fisheries Society Monograph 7. Bethesda (MD): American Fisheries Society. 251 p.

Wood, P.J. and P.D. Armitage 1997. Biological effects of fine sediment in the lotic environment. Environmental Management 21:203–217.

APPENDIX A: Abbreviations

%Fines, PCT_FN	Percent Fines
%SaFn, PCT_SAFN	Percent Sand and Fines
ALU	Aquatic Life Use
ANOVA	Analysis of Variance
CDPHE	Colorado Department of Environment and Public Health
CW	cold water
DE	Discrimination efficiency
EDAS	Ecological Data Application System
EMAP	Environmental Monitoring and Assessment Program
EPA	Environmental Protection Agency
EPT	Ephemeroptera, Plecoptera, and Trichoptera
FtHl	Foothills
GIS	Geographic Information System
HBI	Hilsenhoff Biotic Index
HQCW	high-quality cold water
km	kilometer
LOWESS	locally-weighted regression line
LRBS	Log Relative Bed Stability
LRBS_fin	Log Relative Bed Stability final version (Kaufmann et al. 2008)
LRBS_NOR	LRBS without bedrock or hardpan
MCW	Marginal Cold-water
Mtn	Mountain
MWW	Marginal Warm-water
mg/L	milligrams per liter
n/a	Not applicable
NHD	National Hydrography Dataset
NMAC	New Mexico Administrative Code
NMED	New Mexico Environment Department
NMMSCI	New Mexico Macroinvertebrate Stream Condition Index
ntu	Nephalometric Turbidity Units
PCA	Principal components analysis
RBP	Rapid Bioassessment Protocol
RBS	Relative Bed Stability
Ref	Reference
RIVPACS	River Invertebrate Prediction and Classification System
RMSE	Root mean squared error
RW	Reach-wide
SABS	suspended and bedded sediment
SWIMS	Surface Water Information Management System
TMDL	Total Maximum Daily Load
TSS	Total Suspended Solids
USEPA	United States Environmental Protection Agency
WQS	Water Quality Standards
WSA	Wadeable Streams Assessment
WW	Warm-water
Xer	Xeric

APPENDIX B: SWQB's Previous Sedimentation Assessment Protocol

The below contains a brief description of SWQB's sedimentation/siltation assessment approach (NMED/SWQB 2009) prior to the 2012-2014 Integrated List. This approach was generally in effect from the 1998 to the 2010 listing cycle. For a full copy of the previous assessment approach, please contact SWQB at 505-827-2904:

> In order to properly assess a study site or stream reach for impairment due to excessive sedimentation, a specific reference site was selected, or a reference condition empirically defined, for comparison. Physical measurements of the stream bottom substrate were made alongside measurements of the biological component. Biological sampling and pebble counts always usually performed concurrently to capture an accurate picture of the stressor and response, as the amount of fine substrate present and the biological community changes with stream flow and season.
>
> Pebble counts were performed to determine the percent fines (<2 mm) at the study site and reference site. The intermediate axis of particles was measured within the wetted perimeter of the channel and tallied using standard Wentworth size classes (Bunte and Abt, 2001) from 10 equidistant transects (10 particles/transect as a minimum) selected along a longitudinal stream section of a representative riffle or run single habitat area being biologically sampled or evaluated.
>
> Since the narrative standard for bottom deposits is dependant on biological condition, the assessment of this physically-based narrative sedimentation criteria was determined using a biological response variable that will link excess settled sediment levels to designated use attainment. New Mexico chose the community composition of macroinvertebrates as the most informative biological response in determining sedimentation impacts to aquatic life. Benthic macroinvertebrates at the study site were collected in a representative riffle area and consisted of either three quantitative samples using a Hess sampler or three composited kick samples (semi-quantitative) covering an area of approximately one meter for one minute. For valid biological comparisons to an individual reference site, sampling procedures were identical between the reference and study site(s). Depending on the ecoregion of the study site, a benthic macroinvertebrate impairment determination utilizing either the Rapid Bioassessment Protocols (RBPs) or Mountain Stream Condition Index (M-SCI) as described in the main assessment protocol was performed (NMED/SWQB 2009).
>
> The final assessment was determined using the matrix in Table 1. This was accomplished by taking the increases between percent fines and matching it with the appropriate physical assessment use support category in the far left column. The physical assessment use category can then be matched with the biological assessment use category located on the top row to obtain a use support category for aquatic life use based on biological and physical indicators of increased stream bottom sediment.
>
> It is noteworthy that under certain situations, the physical indicators (i.e., percent fines) indicated full support, while the biological assessment indicated non support. In these cases, factors other than sediment alone, such as extremes in pH, low oxygen,

temperature, lack of stream flow, and toxicity, etc., may be responsible for the reduction in biological integrity at a particular site. In this case, the assessment unit was listed under Category 5C with an impairment of “Benthic-Macroinvertebrate Bioassessments (Streams)” on the Integrated Clean Water Act §303(d)/305(b) list until the exact cause of impairment could be determined. Potential causes of impairment such as those listed above will then be quantified by examining such things as chemical and physical data collected at or near the site in question.

Table 1. Final assessment matrix for determining aquatic life use support categories by combining physical and biological assessments as sediment indicators

Biological **Physical**	**Impaired** **(Non Support)** RBP Index < 79% of ref [1] M-SCI Score < 56.70 [2]	**Non-impaired** **(Full Support)** RBP Index > 84% of ref [1] M-SCI Score > 56.70 [2]
Non-Support Percent Fines >28% increase over reference	☐ Non-Support	☐ Full Support
Full Support Percent Fines <27% increase[3] over reference	☐ Full Support (Sedimentation/Siltation); ☐ Non-Support (Unidentified Biological Impairment)[4]	☐ Full Support

NOTES:

[1] RBP Index should be used in Ecoregions 22, 24, 25, and 26. RBP Index score based on Plafkin et al. (1989). The 4% gap allows for some best professional judgment.

[2] M-SCI should be used in Ecoregions 21 and 23. M-SCI and Score based on Jacobi et al. (2006).

[3] Raw percent values of ≤20% fines (pebble counts) at a study site should be evaluated as fully supporting regardless of the percent attained at the reference site.

[4] Reduction in the relative support level for the aquatic life use in this particular matrix cell is probably not due to sediment. It is most likely the result of some other impairment (temperature, D.O., pH, toxicity, etc.), alone or in combination with sediment. Label as Category 5C on the Integrated §303(d)/305(b) list as described in the text above to indicate that further study is needed.

References:

Bunte, K. and S. Abt. 2001. *Sampling Surface and Subsurface Particle-size Distributions in Wadable Gravel-and-Cobble-Bed Streams for Analysis in Sediment Transport, Hydraulics, and Streambed Monitoring.* Gen. Tech Report RMRS-GTR-74. United States Forest Service. Rocky Mountain Research Station. Fort Collins, Colorado.

Jacobi G.Z., M.D. Jacobi, M.T. Barbour, E.W. Leppo. 2006. Benthic Macroinvertebrate Stream Condition Indices for New Mexico Wadeable Streams. Jacobi and Associates and Tetra Tech, Inc. for New Mexico Environment Department, Surface Water Quality Bureau.

New Mexico Environment Department Surface Water Quality Bureau (NMED SWQB). 2009. *Procedures for Assessing Use Attainment for the State of New Mexico Integrated Clean Water Act §303(d) List and §305(b) Water Quality Monitoring and Assessment Report*. Santa Fe, New Mexico.

Plafkin, J.L., M.T. Barbour, K.D. Porter, S.K. Gross, and R.M. Hughs. 1989. U.S. Environmental Protection Agency. *Rapid Bioassessment Protocols for Use in Streams and Rivers*. Office of Water Regulations and Standards. Washington, D.C. EPA/444/4-89-001.