Answers to questions posed by the Lahontan Regional Water Quality Contol Board regarding USFS Field, Laboratory and Data Analysis Protocols applied in the Lake Tahoe Basin Management Area

January 10, 2003

Prepared by Joseph Furnish, Regional Aquatic Ecologist, USFS, Regional Office, Pacific Southwest Region, Vallejo, CA

Sampling methodology

1. The LTBMU has not provided a proposal for the locations and frequencies that it proposes to sample over time to implement the bioassessment monitoring requirements of the Heavenly Valley Creek TMDL.

Answer: I have recommended to the LTBMU that samples be taken from paired stream reaches on both Heavenly Valley and Hidden Valley creeks utilizing the USFS field protocol (Hawkins et al. 2001, available at http://www.usu.edu/buglab/USUproto.pdf). This protocol has been applied throughout National Forest lands in California and the western United States, and is based upon extensive empirical data collected by the Environmental Protection Agency as part of their Environmental Monitoring and Assessment Program (EMAP). This protocol has also been adopted as part of the monitoring program by the California State Water Pollution Control Laboratory in Rancho Cordova (Jim Harrington, Biomonitoring Program Director, personal communication, 2003, see also http://www.dfg.ca.gov/cabw/emap/index.html).

Sample Locations: The test site for monitoring recovery of the biological community on Heavenly Creek is located below Patsy's and Groove lifts at an elevation of about 8,000 ft. Measurements taken there will provide water quality information reflecting cumulative effects of land management activities at the Heavenly Ski Resort. The paired site on Hidden Valley Creek is located at an elevation of about 9,000 and was chosen because it is relatively undisturbed and has geomorphological features similar to those of the site at Heavenly Creek. Comparison of conditions of the biological communities at these two sites will provide one source of information to determine whether trends are improving over time at Heavenly Valley.

Sample Frequencies: Benthic macroinvertebrate sampling has already been initiated at both of these sites during July of 2001 and 2002, which provide information on baseline conditions. Samples will be taken during the summer season, preferably during the month of July, every other year (i.e. 2004, 2006). Data from all of these samples will be available for analysis during 2007, at which time the Lahontan Board and LTBMU will make a determination regarding whether the objective of improving trends have been met. If a resetting event such as a major landslide or flood event occurs during this period, samples will be taken during the following summer season for at least two consecutive seasons to re-establish a baseline for comparison.

Analysis of data for determining trends in the benthic macroinvertebrate community: Two methods will be utilized to monitor trends: 1) a comparison of metric calculations for the Heavenly Valley and Hidden Valley sites, and 2) a comparison of the taxa

observed at the Heavenly Valley site to those predicted based upon a recently developed RIVPACS predictive model (Hawkins 2003) that was developed from a set of reference sites distributed throughout National Forest lands in California. Details for the development and effectiveness of this modeling technique for Sierran streams have been published in a peer-reviewed journal (Hawkins et at. 2000).

2. What criteria are used by the USFS to identify reference sites? (Reference sites should be of the same/similar stream class, with natural levels of sediment or at least documented/predicted lower levels of sediment exposure than the test sites.)

Answer: The set of reference sites sampled during 2000-01 to define our regional reference conditions are not based upon a paired (reference vs. "treated) basis. Instead, we have tried to define a network of reference sites that captures the variety of stream types that occur on National Forest lands throughout California based upon the same kinds of physical attributes listed above. They should also be relatively independent of land-use activities; they are the "independent variables" of a predictive model (see details that follow) and account for a significant amount of the variability between stream macroinvertebrate communities. Reference sites are also characterized by a relative absence of anthropogenic influences like timber harvest, road building, recreation, mining and grazing; the dependent variables.

Reference site selection was initiated by requesting that Forest Service specialists nominate sites that would meet the conditions described above. Each forest submitted information on location, accessibility and incidence of land use activities. Our goal was to identify approximately 15 reference sites per forest and achieve a network consisting of about 250-300 sites for the entire region. Disturbed sites were also included so that the effectiveness of our ability to detect differences between reference and disturbed sites could be tested. By applying this procedure, about 250 streams were sampled during 2000-01. During the second year we attempted to fill any gaps in the representation of stream types and regions. During the first field season, many prospective reference sites were dry when crews arrived to sample, but we were able to successfully sample similar stream types in the same area the following year. No reference streams were sampled on the Lake Tahoe Basin Management Unit during 2000; 17 were sampled during 2001. Additional samples were taken from Hidden Valley Creek during 2002 and these data will also contribute to the reference site data set.

For sites sampled during 2000, the area of the watershed above each sample site was delineated and the size of the watershed calculated. Next an analysis of the watershed condition above each sample site was conducted using GIS data layers to calculate road density and near-stream (within 50 meters) road density for each watershed. We emphasized road density as a primary variable to distinguish reference from non-reference watersheds because it is highly correlated with anthropogenic disturbances like timber harvest and a primary avenue for delivery of sediments to streams. Geology of each watershed was characterized by calculating the area composed of each of 10 geologic formations or classes and the susceptibility of each to erosion. Consideration was also given to records of site condition and accessibility made by crews during their visit to each site. Both members of each crew took notes on the incidence of recreation, fire, grazing, bank erosion, timber harvest, water diversions and other forms

of disturbance. Based on these considerations, 140 of the first 174 streams sampled were designated as reference sites.

3. What criteria are used by the USFS to select stream reaches for sampling? (i.e., gradient, order, etc.) Is there a single class of stream type being sampled, or multiple classes? If multiple, how are the classes stratified?

Answer: Most of this question has already been addressed immediately above. However, there is the issue of selecting a stream reach within the valley segment or watershed of interest. As a general rule, once the sample area has been reached a stream reach of approximately 100-150 meters is selected based on the absence of disturbance in the immediate vicinity. The field crew first walks the stream reach to familiarize themselves with the site and find the reach that best represents the site. Once the stream reach has been selected, the objective is to identify four consecutive riffles within the reach where concentrated samples will be taken.

Pairs of sites were chosen on both Heavenly Valley and Hidden Valley creeks. The site below Patsys on Heavenly Creek was chosen because it is immediately below the Heavenly Ski Resort and represents the area where sediment entering the stream from the ski runs would be most concentrated and cumulative effects most apparent. The "property line" site was chosen to represent water quality where the Creek leaves National Forest lands. On Hidden Creek, both sites were chosen based on accessibility and because they had elevations and habitat types that were similar to their counterparts on Heavenly Creek.

4. The USFS sampling method includes no within-site replication, and I have seen no evidence that the precision of the method has been objectively quantified. What is the precision of the LTBMU sampling method?

Answer: The USFS field protocol for macroinvertebrate sampling does include collection of eight replicate 0.89 m^2 Surber samples. They are distributed as two randomly selected sites along a transect across each of four consecutive riffles. The resulting eight replicate samples are then composited as a single sample that represents a total area collected of 0.743 m^2. Combining sample replicates is a well-established convention in the collection of aquatic macroinvertebrates and is standard procedure for the EPA's Ecological Monitoring and Assessment Program (EMAP).

Field Protocol – Variability in Physical Habitat Parameters: During the field-collection portion of the bioassessment protocol, measurements of several physical habitat parameters are taken to characterize each stream reach sampled. Variability in the measurement of stream attributes can be partitioned into three sources: environmental heterogeneity, sampling variance and measurement error. Roper et al. (2002) evaluated variability among measurements from six streams in Central Idaho and 6-7 field crews for 13 stream habitat variables such as gradient, sinuosity, bank stability, width:depth ratios and substrate size. The sampling sessions involved all crews taking measurements in the same stream reach. They concluded that when crews are given thorough training

and standard protocols are applied, all but three substrate variables (i.e. percent fines, pool tail fines and D_{50}) were useful for discriminating between streams subject to different land management treatments. Differences among streams was by far the greatest source of variation and accounted for about 80% of the total variation. Variability, measured as coeffecients of variation (CV's), were much higher among streams (range of 5-72) than among individuals collecting the data (range of 5-38). According to literature reviewed by Roper et al. (2002), when sampling variance accounts for no more than 20% of the total it is likely to be a reliable attribute to monitor. According to this rule, all but three of 13 attributes were reliable for monitoring differences between sites.

Field and Lab Protocols – Biological Variables: Mark Vinson, Director of the National Aquatic Monitoring Lab that processes USFS and BLM samples, conducted a major internal QA/QC evaluation of field sampling and laboratory protocols about one year ago. As part of this exercise several calculations were made including the Jaccard Coefficient similarity between aquatic invertebrate sample assemblage compositions and the coefficient of variation in taxa richness among six field crews, three sample sorting technicians, three taxonomists, and differences among identifications by the same taxonomist in different weeks. A full analysis of this effort has not yet been finished. However, based on initial results there was little variation in measures of taxonomic richness and considerable variation in abundance measures. Total variation in assemblage composition attributable to the three sources of variation in Table 1 is 8-10 percent and total CV's in taxa richness attributable to the four sources of variation ranged from 13-22 percent. Additional details about QA/QC procedures are available from the lab's website (http://www.usu.edu/buglab/).

Table 1. Summary of variability in data attributable to different activities associated with collection of field data, laboratory sample processing and taxonomic identifications.

Component of Variability Measured	Range in percent Similarity
Similarity in assemblage composition based on Jaccard Coefficient:	
Among field crews	92-96%
Among sample sorting technicians	Not yet determined
Among taxonomists	98%
Among identifications by the same taxonomist during different weeks	98%
Coefficients of Variation in taxa richness:	Range in Percent
Among field crews	4-7
Among sample sorting technicians	3-6
Among taxonomists	4-6
Among identifications by the same taxonomist during different weeks	2-3

5. It is my understanding that at least several (many?) different USFS crews are conducting bioassessment sampling throughout the Sierra Nevada. No QA/QC plan has been provided to us that documents the precision (i.e., error) expected when sampling is conducted by different personnel. A QA/QC plan must be an integral element of any such program. This is especially important if data collected by more than one crew will be

used in the analysis. What variability in the USFS reference data set is due to sampling error? Has USFS conducted any sampling at the same site(s) by different crews to quantify this variability?

Answer: You are correct that several crews (four during 2000 and 2001) collected samples to generate our reference site data. Prior to the 2000 field season, all crew members were trained together at Sagehen Creek Research Station near Truckee by myself, other Forest Service Specialists and Jeff Ostermiller of Utah State University, who is one of the coauthors of the field protocol. Prior to the 2001 field season, Charles Hawkins, also of Utah State University and senior author of the field protocol, conducted a training session for Forest Service personnel and the crew under contract from CDFG-Water Pollution Control Laboratory to collect samples from the Cleveland National Forest. The answer to the previous question provides details regarding precision of our field sampling protocols.

Laboratory identifications

1. We would like to see a QA/QC plan for the laboratory work.

Answer: Details of the laboratory QA/QC procedures are provided in the previous section and at the Laboratory website (provided above).

2. It is my understanding that the laboratory procedures used by the USFS's contractor(s) have varied over time. What method(s) have been used, and what methods are being used to ensure data comparability between "old" data and "new" data that will be used in the analysis for Heavenly Valley Cr TMDL? (This question is moot if only one lab method is used on all data for this study.)

Answer: All samples collected during 2000-01 from Heavenly Valley and Hidden Valley creeks, as well as all reference streams that were the basis of the RIVPACS predictive model, have been processed by the Logan lab according to their standard protocols.

Data analysis methods

1. Who will conduct the analysis of data to answer the monitoring question posed by the TMDL? LTBMU staff seems to imply that it will not be them. If not LTBMU staff, who?

Answer: I would suggest that we (USFS Regional Office, specialists from the LTBMU and the Lahontan Board) should reach mutual agreement on how to define a successful trajectory towards recovery. I believe that we should consider using both a multi-metric techniques based on the EPA's Rapid Bioassessment Protocol metrics and multi-variate techniques based on the RIVPACS model recently developed by Hawkins (2003). Using multimetric methodology, we might agree that the composite score of the suite of metrics selected will not decline below their present level (e.g. the composite score for metrics at Heavenly Valley should not be less than 60% of that for Hidden Valley), and that the metric scores for Heavenly Valley will be no less than 70% (e.g. indistinguishable from reference condition) in five years, barring any reset events such as a major flood or landslide. It would still remain for us to determine which metrics to use and how to

assign scores to each of them. The Squaw Creek TMDL scoring criteria might provide a basis to assign scores (i.e. 1, 3 or 5 depending on how close Heavenly scores are to reference condition) for each metric value.

For the RIPACS model, I would conclude that the aquatic macroinvertebrate community at the Heavenly Valley site was significantly impaired if its O/E score were less than two standard deviations from the mean O/E score for the reference sites. I plan to request that Chuck Hawkins calculate O/E scores for Heavenly Valley and Hidden Valley creeks to see if they have scores that place them in impacted and reference categories, respectively. This analysis will serve as a means to test the ability of the RIVPACS model to accurately classify these particular sites and provide another means of tracking recovery of the biological community.

2. It is my understanding that the USFS intends to rely primarily on multivariate analyses as packaged in the RIVPACS-type model for its regional bioassessment program. This model is proprietary, and not available in the public domain. How does USFS intend to make the model transparent to the public?

Answer: The USFS has a contract with Charles Hawkins at Utah State University to develop a RIVPACS predictive model based on the samples that were taken throughout the entire Pacific Southwest Region during 2000. A draft report providing details about the performance and precision of the model are provided in Hawkins (2003). This does not mean that we plan to abandon monitoring and analysis based on an Index of Biotic Integrity (IBI), multimetric approach. I consider these methods to be complementary and both may yield useful information. I consider both of these methods of analysis to be under development since IBIs have not yet been developed on a regional basis and are available for only a few sites (e.g. Squaw Creek TMDL). The credibility of certain metrics has not been established by publication in peer-reviewed journals. For example, the tolerance and intolerance values assigned to taxa in the western U.S. were originally developed by Hilsenhoff (1987) in Wisconsin, based on responses of macroinvertebrate communities to organic and nutrient enrichment. To my knowledge, they have not been validated for western taxa.

The initial evaluation of the RIVPACS predictive model has just been delivered by Charles Hawkins (see accompanying document), and it shows good potential as a monitoring tool for TMDLs or for other monitoring issues. A recent study by Dr. Hawkins (accompanying manuscript, Hawkins 2002) shows that in comparison with other methods (including IBI), RIVPACS assessments are equal to or better than other techniques in detecting biological impairment. The superiority of RIVPACS assessments is especially apparent under moderate stress as produced by forest practices and other land uses in western mountainous landscapes. Reynoldson et al. (2001) also found that the RIVPACS model performed better for accounting for variability in a data set from the Frazier River in British Columbia, and RIVPACS has been adopted by both the Oregon Department of Environmental Quality and the Washington Department of Ecology as a scientifically defensible means of assessing the biological condition of streams. Washington DOE is using the method to identify sites for 303d listing and to relate biological condition to TMDLs.

The precision, accuracy, and sensitivity of assessments based on RIVPACS models are well established in the literature (see Hawkins 2002 and bibliography in the draft evaluation report). According to my personal communications with Dr. Hawkins, he envisions that regional RIVPACS predictive models will eventually be available to users within 6 months on the Utah State University/USFS/BLM Monitoring Center website. Currently funding has been available from the EPA to begin development of the website, but it has not been sufficient to complete development and make it available to users.

3. It is my understanding that the USFS's consultant (Chuck Hawkins, Utah State Univ) has made changes to the RIVPACS model, and the routines, algorithms, and assumptions in the model currently used by the USFS have not been disclosed. How does USFS intend to make the model transparent to the public (and to us)? When will this be done?

Draft Answer: Mostly covered in my answer to question 2. The development of the model for the Sierra Nevada was presented in Hawkins et al. (2000), but development of a website application is still in the future. There are, however, no fundamental differences between the mechanics of the original RIVPACS program, which is proprietary, and the software of Hawkins's team program. There are subtle differences in the methods used to account for biological variation among reference sites and hence the initial creation of predictive equations, but these modifications were made as refinements based on work completed by the original RIVPACS development team in Great Britain (Wright 1995), as well as work conducted by the Centre for Freshwater Ecology in Australia, and Hawkins's team at Utah State University.

4. The model produces an O/E ratio for each stream, but because we don't know the precision of the method, we do not know the significance of small changes in O/E. This raises the question: How will LTBMU will interpret the O/E results to answer the questions of whether: (a) the biological integrity of Heavenly Valley Creek is improving significantly over time, and (b) approaching that of Hidden Valley Creek?

Answer: The calculation of O/E scores for a set of reference sites produces a normal distribution of O/E scores with a mean score near 1.0. Generally, sites with scores < the 10^{th} percentile and > the 90^{th} percentile are considered to be in an impaired condition, although the specific threshold values are determined on a model by model basis. Therefore, if the O/E score for Heavenly Valley is within the central 80 percent of O/E scores, it would be considered to be in an unimpaired condition. RIVPACS bases its analysis on data generated from collections from throughout the region and does not rely on comparison to a single reference site. It is important to note, however, that assessments of individual sites are largely based on reference sites that share common environmental attributes with the test site. Often, the reference sites that are most influential in generating the list of expected taxa will be nearby sites. However, because some nearby sites may by physically different from the test site, those sites will be down-weighted, and sites further away that share naturally occurring physical characteristics will add statistical power to the assessment, which would otherwise be low because of the small number of local reference sites.

5. It is my understanding that the RIVPACS model relies on samples from numerous sites in order to obtain defensible O/E values. What zoogeographic region will be used for comparison to Heavenly Valley Creek, and what criteria were used to identify the comparability of this ecoregion?

Answer: You are correct that RIVPACS relies on samples taken throughout the region to generate a list of macroinvertebrate taxa that are expected to occur at a site in the absence of disturbance. This is based on the probability that a given taxon will occur in a stream based on the physical and chemical attributes of that stream as explained above. Model predictions are conceptually similar to a simple regression model in which values of the dependent variable (in this case the specific list of biota) vary in response to one or more independent variables (in this case latitude, elevation, stream size, biogeographic region, etc.). Our expectation of the specific fauna at a site will vary depending on where we are in the landscape. The RIVPACS models simply try to quantify what natural historians and ecologists have known to be true for decades.

6. Related to the previous question, at our meeting in May 2001, LTBMU staff indicated that they planned to sample approximately 10 "reference" sites in the Tahoe Basin for comparison to Heavenly Valley Cr. We asked if the RIVPACS model is robust enough to make conclusions based on only 10 reference sites. That question has not been answered and is still on the table.

Answer: See responses above. The RIVPACS model uses the entire set of samples taken throughout the region and does not rely on just those taken on the LTBMU, so the basis for the predictive model represents data collected from 134 reference sample sites distributed throughout California. The USFS plans to update and refine the model based on analysis of the samples taken during 2001. The critical issue here is whether all reference sites in the same basin are actually appropriate reference sites. In many cases, they will not be because of differences in elevation, stream size, stream gradient, geology, etc. Sites from outside the basin will often provide relevant information regarding the expected fauna and thus increase the robustness and defensibility of the assessments made.

7. If multimetric analyses are to be conducted on the USFS data set for the Heavenly Valley TMDL, how can defensible inferences be made in absence of objective quantification of method precision?

Answer: The draft document by Hawkins (2003) and references cited therein provide detailed information characterizing the precision of the RIVPACS model. The distribution of observed/expected values for this model has a mean of 1.01 and a standard deviation of 0.19. According to Hawkins (2003) "the most precise RIVPACS models produced to date have reference site O/E standard deviations of ~0.15, thus this initial model is slightly less precise that the best models." It is likely that greater precision will be achieved by inclusion of additional reference sites collected during 2001 and additional predictor variables (e.g. percent calcareous geology) for which data were not available when the model was developed.

Hilsenhoff, W.L. 1987. An improved biotic index of organic stream pollution. Great Lakes Entomologist 20:31-39.

Reynoldson, T.B., D.M. Rosenberg and V.H. Resh. 2001. Comparison of models predicting invertebrate assemblages for biomonitoring in the Fraser River catchment, British Columbia. Can. J. Fish Aquat. Sci. 58:1395-1410.

ROD. 2001. Record of Decision for the Sierra Nevada Forest Plan Amendment Final Environmental Impact Statement, U.S. Department of Agriculture, Forest Service, Pacific Southwest Region, Vallejo, CA. i + 55 pp., appendices.

Roper, B.B., J.L. Kershner, E. Archer, R. Henderson and N. Bouwes. 2002. An evaluation of physical stream habitat attributes used to monitor streams. Journal of the American Water Association. In press.

Wright, J.F. 1995. Development and use of a system for predicting the macroinvertebrate fauna of flowing waters. Australian Journal of Ecology 20:181-197.

Finally, I would like to provide an answer to your question about what the LTBMU Forest Supervisor meant when you recall she said "...we will be having biomonitoring standards imposed upon us when we adopt the Sierra Nevada Framework, which will be an amendment to our Forest Plan. So we will be doing biomonitoring under the Forest Service protocol...We will have to use the Forest Service protocol." Although I was not there and can not substantiate this statement, I want to point out that I believe it is accurate because the Sierra Nevada Framework Record of Decision does commit the Forest Service to the following standard under Riparian Conservation Objective #2, which in part states: "Prior to activities that could effect streams, determine if relevant geomorphologic characteristics...are within the range of natural variability for the reference stream type as described in the Pacific Southwest Region Stream Condition Inventory (SCI) protocol. If properties are outside the range of natural variability, implement restoration actions that will result in an improved trend" (ROD, Appendix A-55). Bioassessment is part of the SCI protocol, so she is technically correct in stating that we are required by the ROD to utilize the USFS bioassessment protocol although this is not explicitly stated.

In conclusion, thank you for providing the opportunity to shed some light on these monitoring issues. You have asked a number of basic questions that have allowed me to make a statement about the quality of our bioassessment program. I hope that my answers will serve the purpose of allowing us to reach agreement on a workable solution for proceeding with monitoring and recovery of the biological community at Heavenly Valley Creek.

References Cited

Hawkins, C.P. 2002. Bioassessment is a science not a religion: Or why the debate about methods is both healthy and needed. Unpublished transcript of a presentation made at the 2002 North American Benthological Society (NABS) Technical Issues Committee Workshop on Performance Characteristics and Accuracy of Statistical Approaches in Benthic Ecology, Pittsburg, PA.

Hawkins, C.P. 2003. Summay of initial development and testing of a RIVPACS-type model for assessing the biological condition of streams in Region 5 USDA Forest Service. Draft Report submitted to the USFS, Pacific Southwest Region, Vallejo, CA.

Hawkins, C.P. R.H. Norris, J.N. Hogue and J.W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10(5):1456-1477.

Hawkins, C., J. Ostermiller, M. Vinson and R.J. Stevenson 2001. Stream Invertebrate, Periphyton, and Environmental Sampling Associated with Biological Water Quality Assessments: Field Protocols. BLM National Aquatic Monitoring Center, Logan, Utah.