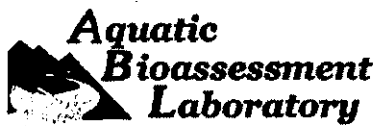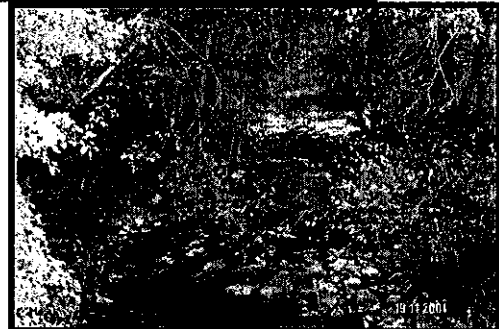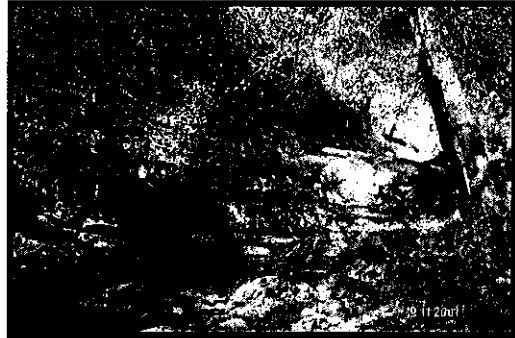124

Sacramento River Watershed Program
A Quantitative Framework for Reference Site Selection:
Case Study from the Sierra Nevada Foothills Ecoregion
December 2002

**Aquatic
Bioassessment
Laboratory**

California Department of Fish and Game
Office of Spill Prevention and Response
Water Pollution Control Laboratory
Aquatic Bioassessment Laboratory
2005 Nimbus Road
Rancho Cordova, CA. 95670

CITATION

Ode, Peter R. 2002. A Quantitative Framework for Reference Site Selection: Case study from the Sierra Nevada Foothills Ecoregion. Sacramento River Watershed Program. California Department of Fish and Game, Water Pollution Control Laboratory. Rancho Cordova, CA.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Previous Bioassessment Efforts in the Sacramento River Watershed

As part of its program to evaluate water quality in the Sacramento River watershed, the multi-agency Sacramento River Watershed Program (SRWP) has collected benthic macroinvertebrate samples for four seasons between 1997 and 2000. The California Stream Bioassessment Procedure (CSBP, Harrington 1999) was used to evaluate the benthic macroinvertebrate communities at 37 sites in watersheds draining the eastern and western sides of the basin. The first three years of data were collected by the California Department of Fish and Game (DFG-ABL), the California Department of Water Resources (DWR) and the United States Geological Survey (USGS), while the 2000 sampling event was performed only by DFG. The results of previous sampling events were reported in separate documents and are available in the administrative reports of the Sacramento River Watershed Program (www.sacriver.org).

Like many ambient bioassessment studies in California, the SRWP bioassessment projects were designed to assess baseline biotic conditions and to supplement existing water chemistry/toxicity data. While the initial sampling events provided good taxonomic surveys for the region, these studies called attention to the need for a context for interpreting the biological data. This context is provided by reference sites, which represent the desirable state or least disturbed state of biotic condition (Hughes *et al.* 1986, Karr and Chu 1999). This report presents a working example of a quantitative method for selecting reference sites that could serve to define reference condition for any region.

Establishing Reference Conditions

Reference sites are sections of streams that represent the desired state of stream health (*sensu* Meyer 1997) for a region of interest. Since natural stream communities vary both spatially and temporally, measures of biotic integrity should also be expected to vary. Once candidate reference reaches have been identified, these are used to characterize the range of biotic conditions expected for minimally disturbed sites. Deviation from this range is used as evidence that test sites are impaired.

Although natural biological variation poses challenges to the interpretation of stream condition, sound scientific approaches for interpreting impairment in the context of natural variation have been and continue to be developed for both multivariate and multimetric analytical techniques (Wright *et al.* 1993, Gibson *et al.* 1996, Bailey *et al.* 1998, Barbour *et al.* 1999, Karr and Chu 1999, Hawkins *et al.* 2000). Reference site selection is critical to the success of both of these approaches (Reynoldson *et al.* 1997, 2000).

Techniques for the selection of reference sites have been discussed extensively (Hughes and Larsen 1988, Hughes 1995, Rosenberg *et al.* 1999, Stoddard unpublished ms). There are many definitions of the term "reference condition" ranging from the pristine, undisturbed state of a stream, to merely the "best available" or "best attainable" conditions in a region. Since practical considerations limit our ability to find minimally disturbed sites, many reference condition approaches seek to identify a compromise, the "least disturbed condition" in a region (*sensu* Stoddard). In some regions, particularly those that have been severely impacted by human activity, it is necessary to select sites that represent the "best attainable" condition given best management practices in a manipulated ecosystem.

To date, most of the reference or "control" sites used in Californian bioassessment studies were selected for comparison to local conditions and have not been selected using common criteria that would permit comparison among projects. Almost all of these programs have used the subjective technique of "best professional judgement (BPJ)" in which reference sites are "hand-picked" by people familiar with the ecology of the region. While this is a legitimate approach that has been used widely in bioassessment programs (Hughes *et al.* 1986, Hughes 1995, Reynoldson *et al.* 2001), a problem with BPJ is that it is rarely quantified and is therefore not repeatable. This makes comparison with other projects difficult. In addition, several recent EPA analyses suggest that reference sites chosen by BPJ often are do not have significantly different biological signatures from sites chosen randomly (Paul Ringold, personal communication). These factors make BPJ less acceptable for a long term approach to using community data for water resource management and illustrate the need for a standardized and objective approach to selecting reference sites.

DFG-ABL/ SNARL Reference Site Selection Approach

In May 2000, the DFG-ABL and Sierra Nevada Aquatic Research Laboratory (SNARL) collaborated to develop a quantitative approach to selecting reference sites in California. The work was encouraged by the State and Regional Water Quality Control Boards as a way to standardize methodologies among regions. The approach combines use of landscape analysis tools (based on geographic information system data, GIS) with ground-truthing to identify a pool of reference sites that can be sub-sampled to define the range of variability in benthic communities in relatively undisturbed portions of a region of interest. The procedure consists of 5 main steps:

Step 1. Define region of interest and a set of stream types to be evaluated
Step 2. Identify major disturbances to the region and use GIS techniques to quantify potential impacts to different areas within the region
Step 3. Use these GIS-based impact estimates to identify least-disturbed candidate areas in a region
Step 4. Field reconnaissance of candidate areas to select final set of reference stream sites (contingent upon obtaining permission to sample)
Step 5. Quantitative local condition assessment to confirm high quality environments

The body of this report is divided into 5 sections based on the 5 steps in the methodology. Each step is further divided into three subsections: 1) "Theory and Methods" describes the general rationale for each step and the basic methodology for each step; 2) "Case Study" describes the application of each of these steps using an example performed in the Sierra Nevada Foothills ecoregion (SNFE) of the Sacramento River; and 3) "Discussion" summarizes some of the lessons learned in applying the method and discusses difficulties with the methodology, research needs and alternative approaches that could be tried.

STEP 1. CLASSIFICATION OF STREAM TYPES/ DEFINING THE STUDY REGION
Theory and Methods

Why Classify Streams?

Classification of streams can help partition natural variation and thereby improve our ability to detect deviation from reference condition (see Hughes 1995 for a review of the history of stream classifications). Ecoregions are commonly used for grouping reference streams (Hughes 1986, 1995, Whittier *et al.* 1988, Omernik 1995), but many other partitions may be equally useful including stream order, discharge, stream gradient, watershed area, elevation zones, prior beneficial use designations, etc. Combinations of classification schemes might be used to partition natural variation, and numerous clustering can be used to inform these decisions (Jongman *et al.* 1995, Legendre and Legendre 1998).

Defining the region of interest is another way to manage large amounts of biotic variation. The region of interest is determined by the scale of the questions and the entity asking the questions; it might be defined by regulatory or other political boundaries, zoographic regions or ecoregions, watersheds or other groupings.

It is important to consider a tradeoff intrinsic to creating multiple stream classes: the benefits of adding classes are countered by the costs of having too many reference pools to maintain, resulting in undesirable logistic and financial burdens. The objective, therefore, is to use the fewest classes necessary to effectively partition natural variability in the region. A benefit of this method's structure is that it allows decisions to be made on the basis of practical constraints (e.g. sample fewer classes or prioritize work in certain regions).

Identify Areas to be Used as Units of Analysis (Candidate Areas)

Because they integrate all upstream landuse activities, watersheds are a logical choice for analysis areas. Ideally, true watersheds (nested by stream order) would serve as the basic unit of analysis, but accurate watershed areas are not currently available for GIS analysis at the state scale. A reasonable alternative is to use existing CalWater Planning Watersheds (PWS) as the basic analysis unit. However, because they often do not match true watershed boundaries at the smallest scales, their use limits analysis to larger scale watershed boundaries. This is acceptable for coarse screening of target areas, but will need to be resolved before GIS can be used at a finer scale (see discussion for examples of alternate approaches).

Case Study

We focused on the eastern foothills of the Sacramento River watershed for this case study. The Sierra Nevada Foothills Ecoregion Section (SNFE) circumscribes an area bounded by the Cow Creek watershed at the north and the American River watershed at the south and between approximately 150 and 700 meters in elevation (Bailey 1994). Although it is not part of the Sacramento River drainage, we included the Cosumnes River watershed at the southern end of the range because of its potential for having good reference sites (see citations in SNEP report 1996).

This region is small enough to be manageable for this case study but is also subject to a wide range of landuse activities in the upstream watersheds, making it a good candidate for methods testing. An additional reason for choosing this region was that the SNFE is currently undergoing a great deal of development pressure and is a subject of ongoing and future bioassessments.

CalWater version 2.2 (1999) Planning Watersheds (PWS) were used as the basic unit of analysis. There are 236 PWS in the SNFE (Figure 3a).

Discussion and Research Priorities

Stream Classification

Efficient partitioning of natural variation through stream classification is a central issue in IBI based bioassessments. The selection of reference sites is dependent upon this classification step since it defines the range of stream classes to which the reference sites can be compared. There is a great deal of debate about this issue in the bioassessment literature (Whittier *et al.* 1988, Omernik 1995, see papers in JNABS Landscape Classification special issue, September 2000). Regardless of the factors used in stream classification, having a well-defined structure to the selection process makes it easier to compare across classes and projects. For example, scoring reference sites with a standardized method may allow classes to be combined later.

Defining Units of Analysis

One of the biggest current impediments to using this methodology is the lack of appropriate watershed based GIS data layers defining candidate areas. The method described here was developed within the constraints of the CalWater 2.2 dataset. The layers defined by the CalWater 2.2 watershed areas work well at larger scales, but break down below 3rd order streams, such that the smallest units (PWS) often contain many smaller watersheds. This makes analyses based at the PWS scale susceptible to missing landuse patterns occurring at the smaller scales (watersheds with either better or worse condition than the main watershed would be overlooked). We are currently exploring approaches to improve this:

One procedure that we are researching uses confluence points of streams to define the lower point in a watershed. This point would be used with elevation data to circumscribe a watershed. Watersheds would be nested such that 1st order streams would be contained within 2nd order streams and 2nd order streams would be contained within 3rd order streams. This would resolve the "true" watershed issue.

An even more promising variation on this method is to use stream segments rather than confluence points to define watershed boundaries. In this alternative, the upstream or downstream ends of stream segments (defined by points of confluence with tributary streams or receiving streams– a standard characteristic of GIS stream datasets) could be used to define the lowest point in a watershed. Then stream segments would serve as the targets for reconnaissance rather than watershed areas. The GIS analyses would still be made on the basis of watersheds, but the watershed could be more easily defined as the area upstream of a given point. This could increase the efficiency of the entire method since it would reduce or entirely eliminate the first reconnaissance stage, which is one of the more labor intensive steps.

STEP 2. QUANTIFY POTENTIAL IMPACTS TO CANDIDATE AREAS
Theory and Methods

Assemble a List of Land Use Disturbances of Interest

In a typical watershed, the list of potential impacts to stream condition is very long and

includes multiple point and non-point sources of disturbance. It is impractical to try to include all possible impacts when determining reference condition. A more practical approach is to include the most influential landuse impacts that have the potential to affect stream condition as the first selection criterion. Selection of impact types is clearly an important determinant of reference site selection, and this step provides considerable flexibility for deciding which factors to include or to emphasize. However, while this flexibility may often be desirable, it introduces a degree of subjectivity to the method. Depending on the set of impacts chosen and the manner in which they are integrated, different approaches could result in quite different pools of reference sites. It is, therefore, important to control the degree of subjectivity involved when selecting which landuse categories are measured and how they are measured.

Another criterion for selecting disturbances is that they need to be repeatably quantifiable. Ideally, impacts should be chosen based on *a priori* knowledge of effects on biota. Just as in IBI metrics selection, criteria like "signal to noise", "independence of measurement" or "orthogonality" can be applied to determine the best set of impacts to include and define the ways that they should be weighted (Hughes et al. 1998, McCormick *et al.* 2001).

A limitation to assembling a list of potential impacts is the availability of reliable data. Quantitative measures of many human activities (or human-related activities) are available in GIS formats from various state and federal agencies, but there is a very large amount of variation in the degree to which datasets are accurate, current and consistent across wide geographical ranges.

Defining the Area of Influence
A final consideration before analyzing impacts is to determine the area of influence that will be quantified for each candidate area. Possible approaches include analyzing the entire area upstream of the candidate area (the upstream watershed), analyzing the candidate area itself, and analyzing a riparian corridor around the streams in a candidate area (or the upstream watershed). Each is likely to give somewhat different answers and the appropriate scale depends on the degree to which the researcher needs the results to emphasize local vs. larger-scale conditions.

Use GIS tools to Summarize Potential Land Use Impacts for each Area
This is the quantitative step in the process. Using GIS tools like the EPA's Analytical Tools Interface for Landscape Analysis (ATtILA) extension, landuse data layers are analyzed to calculate impact scores for each impact identified in Step 2 (e.g. percent impervious surface in each target unit). At the end of the process, each target area is assigned scores for each land impact measure calculated in the model.

Although they provide a very useful set of tools, current ATtILA outputs are limited to analysis of land use and related measures (e.g. road density, percentages of different landuse activities, estimates of nitrogen and phosphorous loading). Additional criteria (e.g. presence of dams and other diversion structures, presence of mining activity, previous history of pollution events) can be scored the same way by hand (or other GIS tools could be built) using other datasets and maps to supplement the GIS data. Ultimately, important additional factors (like models for estimating sediment loading... AnnAGNPS, USDA 2000) can be integrated into ATtILA (or a similar software program) to provide a unified tool for analysis.

Case Study

For this case study, we focused on factors that could be analyzed with available GIS tools. All GIS analyses were performed with ArcView® 3.2 GIS database software, using the ArcView Spatial Analyst® extension (ESRI Products). For most landuse analyses, we used an ArcView extension developed by the EPA's Landscape Ecology Branch (LEB) in Las Vegas, NV. This extension (ATtILA) accepts a range of GIS data layers and produces summary statistics for areas defined by the user. For basic landuse data, we used the newly released National Landcover Database (NLCD, 2001), a joint product of the USGS and EPA. The NLCD is based on 1992 Landsat satellite imagery and is coded for 21 different types of landuse at the scale of 30 meter grids (Figure 3b). We added additional TEALE datasets available from the California Spatial Information Library (CaSIL, www.gis.ca.gov) for roads (by county) and population density (1996 census data by ZIP code). A version of the EPA's River Reach File 3 stream network layer with Strahler stream order assigned to each stream segment was provided by EPA in Corvalis, OR. Finally, a 30 Digital Elevation Model (DEM) grid was obtained from the USGS (same scale as the NLCD dataset). Figure 4 illustrates the major data layers used in this analysis.
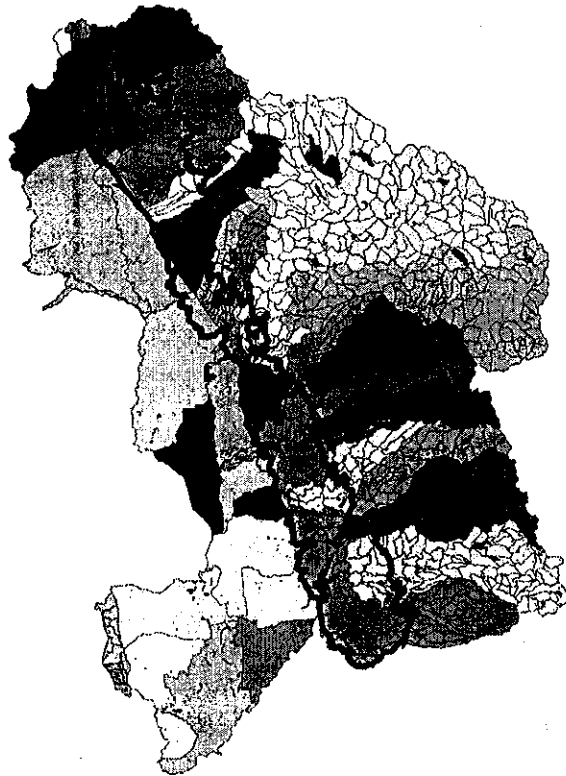
Figure 1. Map of CalWater 2.2 watershed areas showing main watersheds and planning watersheds (PWS). The Sierra Nevada Foothills Ecoregion is outlined in red.
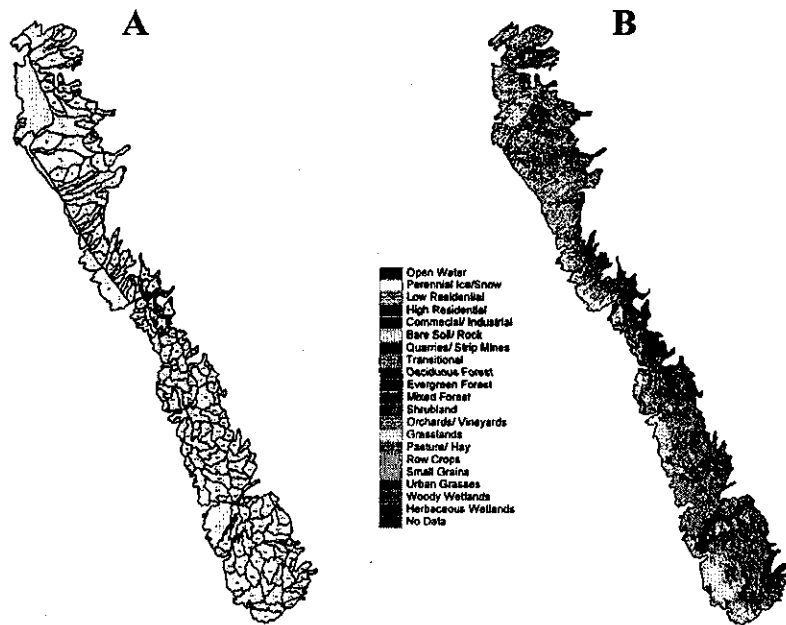
Figure 2. The Sierra Nevada Foothills Ecoregion (SNFE) of the Sacramento River watershed:
A) 236 candidate planning watersheds in the SNFE, B) NLDC landcover for the SNFE.

The ATtILA extension used information in the above data layers to calculate the percentages of various landuse activities (urban; forested; agricultural-row crops; agricultural-orchards/vineyards; agricultural-total), other correlated measures of human activity (population density; road length; road density; road crossings/stream mile; percent impervious surface), and estimated nitrogen and phosphorus loadings.

These measures were calculated for two local scale (candidate areas and riparian regions within each candidate area) analyses. We have also developed a method for analyzing all upstream areas, but this was not performed for this exercise. Results of the main analyses are reported in Appendix A.

30 m Slope DEMs   Census Tracts

NLCD Landcover GRID

Road Network   RF3 Stream Network
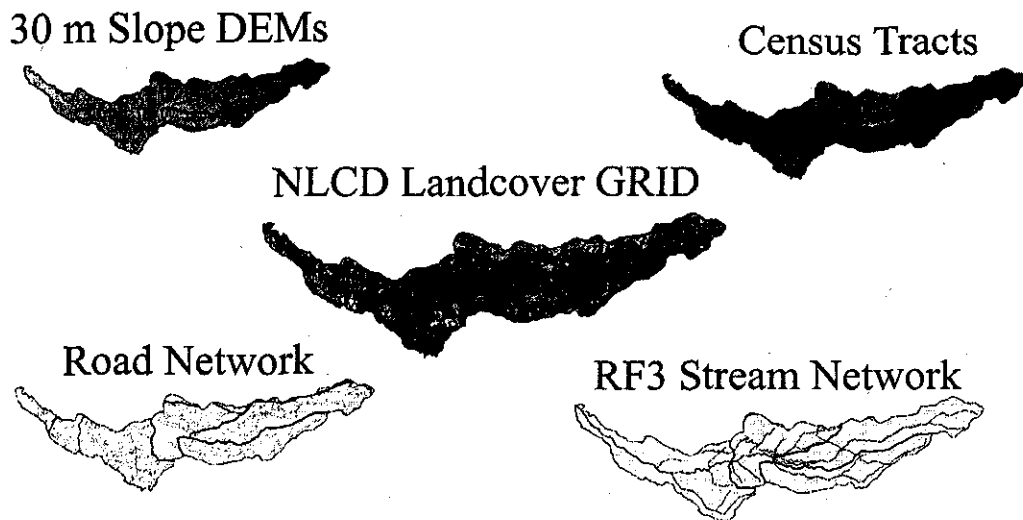
Figure 3. The major GIS data layers used for scoring landuse activity in target planning watersheds.

A great deal of refinement to this method could come from the inclusion of additional datasets and hand-calculated estimates of impacts not calculated by ATtILA such as the presence of dams and other diversion structures, other measures of flow manipulation, presence of mining activity, prior history of logging or grazing, prior fire history, prior history of pollution events, etc. The only additional factor that we considered was the presence of high densities of mining activity in upper watersheds. This was not quantified, but visual assessment of mining distribution was used to eliminate some candidate areas from the pool of target areas.

Note also that, for this exercise, we relied on the impact measures calculated by the ATtILA program and did not develop any procedures for screening which impacts to include. However, this is likely to be an important component in future iterations of this process.

Discussion and Research Priorities
Since this step introduces the most subjectivity to the method, it is an important area for further research. At this point we don't know how sensitive the method is to this subjectivity; its measurement should be one of our goals.

Two of the most promising research priorities are:
1. Improving the consistency of datasets (especially the landcover datasets)
2. Developing a quantitative methodology for deciding which impacts to measure and how and when to weight impacts differently. One approach would be to weight based on the degree of confidence in datasets. For example, we may have current and accurate timber harvest information but inaccurate information about mining impacts. We may have a good idea about mine locations, but not about which are active, and which are actively contributing contaminants to the stream? This is much harder to determine in a quantitative manner.

Combining Data from Different Sources

Quantitative measures of many human activities (or human-related activities) are available in GIS formats from various state and federal agencies, but there is a very large amount of variation in the degree to which datasets are accurate, current, and consistent across wide geographical ranges. We used the NLCD dataset because the data were consistently applied over the entire region that of interest and thus should provide comparable results across all candidate areas. However, the NLCL data is based on 10 year old Landsat imagery (note, however, that the data were only made available in summer 2001) and parts of the SNFE have undergone considerable development in the past 10 years. An alternative approach is to use the most recent data available for each region. Since this data would come from variety of sources, a map of the data for the entire region would resemble a patchwork quilt in which different areas have different data sources, qualities and origin dates. The consequences of combining datasets are unknown and represent another important research priority.

STEP 3. IDENTIFY LEAST-DISTURBED CANDIDATE AREAS WITH GIS TOOLS
Theory and Methods

Use Statistical Properties of Distributions to Screen Candidate Areas

GIS analysis results in impact scores for a number of measures for each candidate area. Frequency histograms are then created for each impact measure based on the scores for all candidate areas and these are used to set criteria for eliminating sites from consideration as having candidate reference reaches.

Criteria may be set visually by looking for "natural breaks" in distributions (which may indicate impact thresholds, Figure 1a) or by using statistical properties of the distributions to select cutoffs (e.g. eliminate all sites having road densities >1 or 2 standard deviation above the mean for the region, Figure 1b).
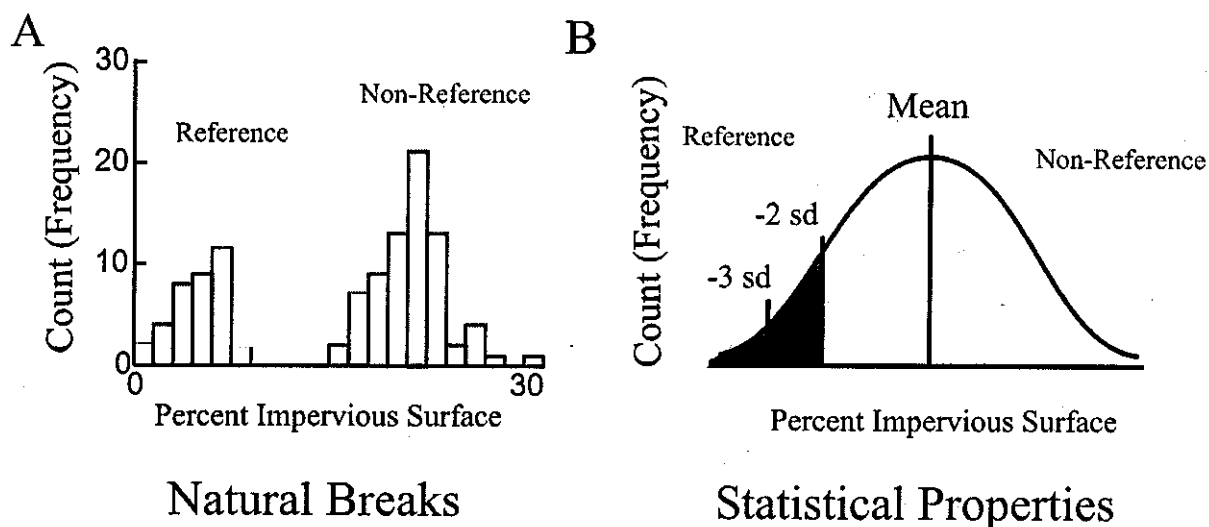

Figure 4. Alternate methods for establishing numeric cutoffs for impact scores using percent impervious surface as an example of an indicator of human activity: A) Natural breaks in the distributions may indicate real impact thresholds, B) Statistical properties of the

distribution can be used to define cutoffs.

Use Impact Scores to Identify Regions with Minimal Disturbance: Target Areas
Using the scoring criteria defined above, the next step is to progressively eliminate all candidate areas that do not meet all of the criteria established in Step 3a. This may require modification of the scoring criteria if too many or too few target areas are selected. This stage can be further modified to emphasize specific impact types based on *a priori* or *a posteriori* decisions about the relative importance of these factors (i.e. differential weighting of factors).

Case Study
After ATtILA was used to generate a series of impact scores, we used statistical properties of the score distributions of scores for each measure to eliminate a set of candidate areas from the target pool (Figure 1b). All candidate areas greater than two standard deviations worse than the mean (based on the *a priori* assumptions of the direction of change expected after impact) were eliminated. This was deliberately chosen to be a conservative screening tool; that is, we probably accepted a number of areas that had average impacts so that we didn't eliminate too many sites. In earlier iterations of this technique, we set the standards much more stringently (1 s.d.). This resulted in far too few target areas to find reference streams. To adjust for a high proportion of candidate areas with impact scores of 0, most metrics were log transformed (ln) to normalize the data before calculating standard deviation cutoffs (Figure 5). This was repeated for all the impact measures in a series of screening filters and at the end, sites that passed through all of the screens were placed in a pool of target areas. At the end of this process, 151 target areas were selected from the original 236 candidate areas for the reconnaissance stages of the method.
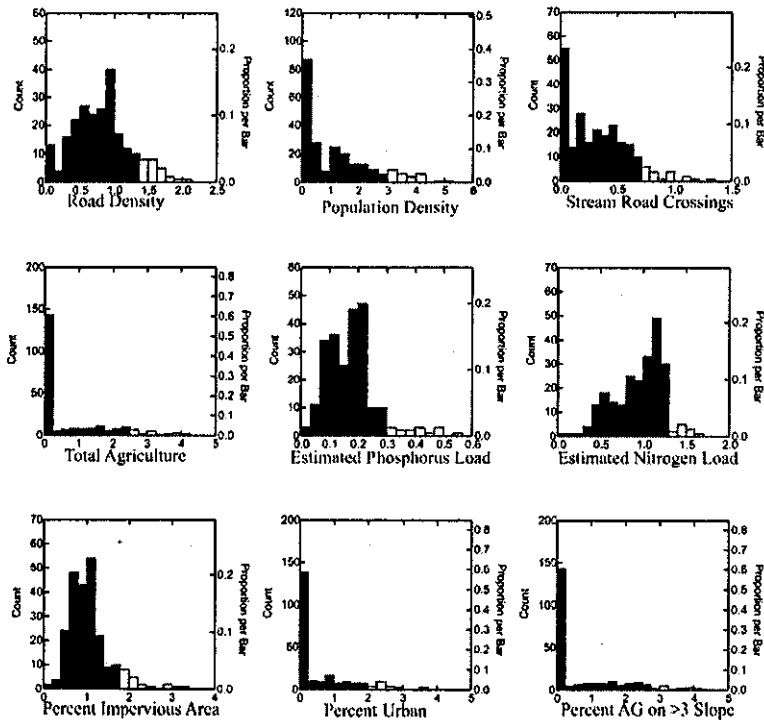

11639

Figure 5. Selected frequency histograms of the main impact measures. All values were log transformed (ln) to normalize distributions. Blue bars represent candidate areas that passed each screening stage and white bars represent candidate areas that were eliminated from the candidate pool.

Discussion and Research Priorities

For this exercise, we deliberately chose conservative cutoffs so that we would not eliminate too many areas from consideration. Very likely, this resulted in our acceptance of sites that might not be ideal candidates for defining reference condition. The establishment of appropriate cutoff points is related to the general condition of the region of interest. Areas with no pristine conditions (e.g. Central Valley streams) should need less stringent cutoffs, while areas with large undisturbed regions (e.g. upper Sierra watersheds) should need more stringent cutoffs to define the range of reasonable expectations for "least disturbed" condition. The work reported here was intended as an exercise to work out the methodology and not as the final word on which sites should be selected to predict reference condition. We expect that we will have to adjust the sampling pool as our tool becomes more refined. These cutoffs can easily be adjusted and are expected to be less stringent in more impaired areas.

The method used here (using statistical properties of distributions to define score cutoffs) assumes a normal distribution of scores. Some distributions (e.g. highly skewed or bimodal) may be better interpreted by looking for natural breaks (Figure 1a).

STEP 4. FIELD RECONNAISSANCE OF CANDIDATE AREAS
Theory and Methods

Ground Truthing

Stage I– Rapid Reconnaissance
Once areas with potential for containing candidate reference sites are identified, field crews drive through the area to identify stream reaches that meet basic criteria for bioassessment sampling (e.g. adequate flow, practical access). Preliminary screening of streams within each target area identifies additional regions that need to be eliminated based on information not available through GIS tools (e.g. local landuse activity that wasn't recorded in the GIS data layers, flow status that didn't match that inferred from data layers). This step can (and ideally, should) be supplemented by coordinating with local watershed groups, landowner groups and other stakeholders to obtain information about potential stream reaches that are not close enough to roads to be picked-up by cursory reconnaissance.

Stage II–Identify Ownership and Obtain Access Permission
It is often desirable to select sampling locations that occur on publicly owned land or land with easy access. However, since it is important to sample streams from a truly representative set of sites within an area, it is often necessary to sample from reaches running through privately owned land. Reasonable efforts should be taken to obtain permission from landowners before rejecting candidate sites. This stage is very important and the quality of the final data set (and the ability to make inferences about reference conditions in the region of interest) will depend on the ability to obtain a representative set. The degree to which this stage is important

depends on the region of interest since some areas have more private ownership than others (e.g. upper watershed Sierra Nevada has much more publicly-owned lands than the SNFE).

Stage III–Intensive Habitat Scoring and Selection of Reference Sites for Sampling
Sites that make it through Stage I reconnaissance and for which legal access is obtained are evaluated using an intensive physical habitat scoring procedure that emphasizes quantitative physical measurements.

Refinement of the Reference Pool

The reference site pool is further refined based on chemical and physical habitat data collected at each site. Some candidate reference sites will be eliminated as land use changes occur, while others may be added if conditions improve. If sites are eliminated or added to the pool of reference sites, it is important that the biological data not be used as the basis for these changes to avoid circularity in defining the reference state.

Case Study
Field reconnaissance was performed in three stages:

Stage I Reconnaissance

In the first stage, field crews evaluated 154 sites from rapid surveys of streams in the target regions using field form presented in Appendix B. Of these, 112 were eliminated due to poor habitat, uncertain flow status and poor local landscape conditions. Summer 2001 and 2002 were relatively dry years in the Sacramento River valley and sites that had water in both summers are likely to have water in most water years. Data from Stage I Reconnaissance are presented in Appendix C.

Stage II Reconnaissance

Field crews then performed more intensive field surveys at the remaining 42 sites using the field form presented in Appendix D. In this stage, field crews delineated a CSBP style reach (a reach is defined as having at least 5 riffle/ pool sequences, Harrington 1999) and set up 9 transects (5 riffle transects and 4 inter-riffle transects) at which they collected a series of qualitative and quantitative physical habitat measurements. Physical characteristics consisted of both instream [(physical dimensions; riffle gradient; substrate composition (7 metrics); RBP visual scores (10 metrics)] and riparian measures (8 metrics). At the end of this stage, sites were assigned an overall habitat score based on 7 of the physical habitat metrics (canopy cover, minimum riparian buffer width, specific conductance, riffle complexity, bank erosion and percent fines) and the lowest 12 scores were eliminated, leaving 30 sites.
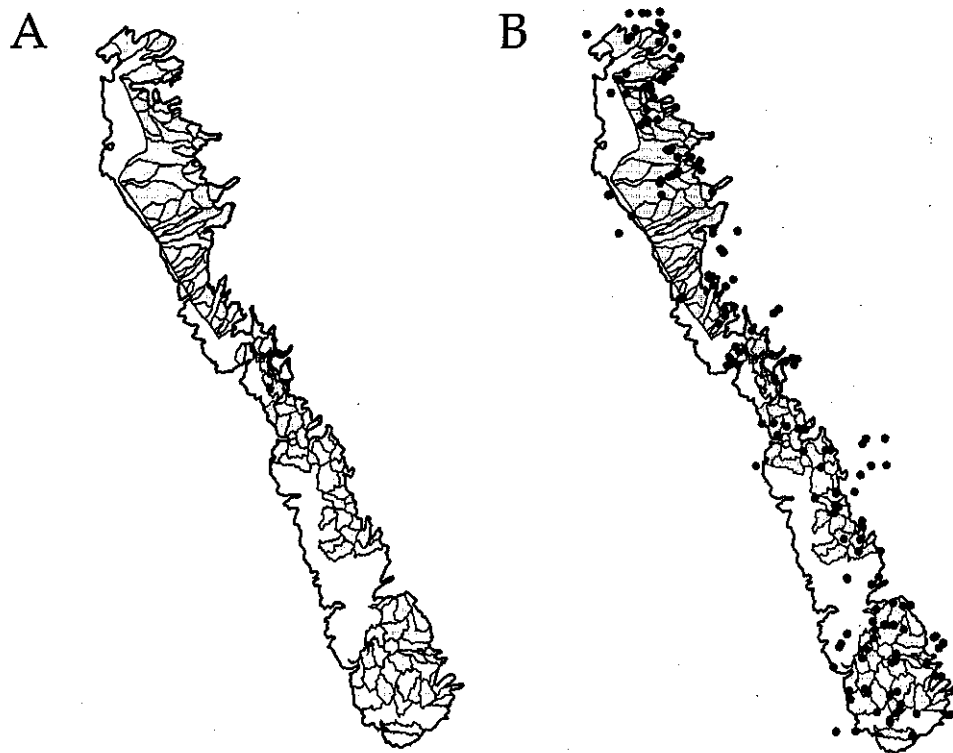
Figure 6. Map of the Sierra Nevada Foothills Ecoregion demonstrating A) the 151 target areas selected from all 236 candidate areas and B) sites sampled in Stage I reconnaissance.

Ownership and Access Permission

After going through considerable effort to obtain permission to sample sites on private lands, the number of potential sites was reduced to less than 20 sites by failure of owners to respond to our requests and by outright denial of access.

Stage III Reconnaissance

In summer and fall 2002, we made an effort to supplement the site list by repeating the above stages on an additional 28 sites. Note that this involved more intensive search for sites within the already selected target areas rather than any change in the target areas. At the end of this stage, we had permission to sample 32 sites that passed all screening stages.
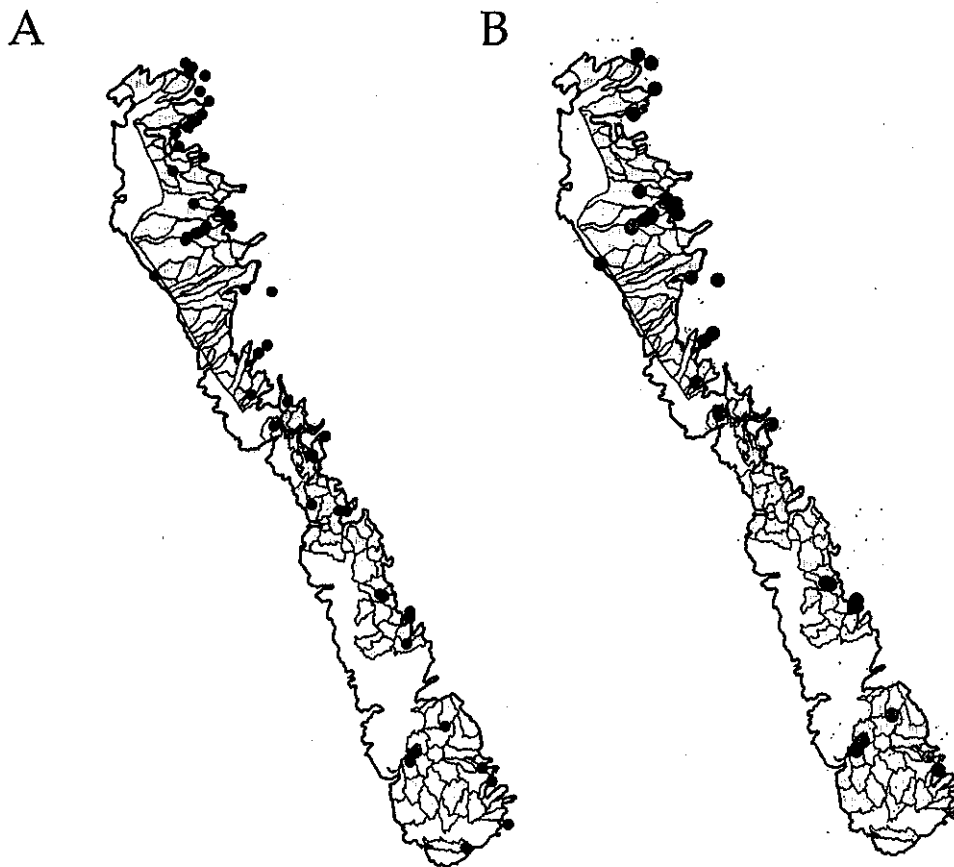
Figure 7. Map of the Sierra Nevada Foothills Ecoregion demonstrating sites remaining after A) Stage II reconnaissance and B) Stage III reconnaissance. Red dots in B represent sites that were sampled for macroinvertebrates in 2002.

Discussion and Research Priorities

The most important lesson we learned in completing this exercise was the importance of building relationships with local stakeholders. Although the short term nature of this exercise made it difficult to develop more than cursory relationships with stakeholders (regional boards, watershed groups, landowner group, tribal groups, etc.), this is clearly a critical part of making any reference site selection methodology work, especially in regions with a large degree of private ownership. Our original intent was to obtain a pool of at least 100 sites that passed all stages of the screening process. However, many factors, but primarily access and ownership issues caused us to reduce that number to less than 40 sites.

STEP 5. QUANTITATIVE LOCAL CONDITION ASSESSMENT
Theory and Methods

After both reconnaissance stages have been completed, the resulting pool of sites is available for biological sampling. A subsample of the pool of reference sites is sampled for benthic

invertebrates and these data are analyzed to define the range of biological metric values in the pool of reference sites. Reference sites may be sampled for other measures of stream or riparian health (e.g. fish or algal communities, water column chemistry, toxicity, etc.).

Case Study
In October 2002, we sampled 30 sites for benthic macroinvertebrates and physical habitat condition. The results of this sampling event will be reported in a separate document.

Discussion and Research Priorities
The data from the biological sampling will be analyzed in 2003. When this is complete, we will use the data to summarize the biological characteristics of reference sites. We expect that there will be considerable refinement of the reference pool over time as the methodology improves. We know, for example, that some sites in the American River watershed should be eliminated from the reference pool because of recent landuse changes that were not picked up in the GIS analysis. However, we can't overstate the importance of not using biological information to decide which sites are included in the reference pool because of the danger of circular definition of reference condition.

SUMMARY

The goal of this exercise was to describe a method for quantitative reference site selection that could be used to improve consistency and repeatability across bioassessment projects and to identify priority research needs to make it a practical tool. With some resources put into this development and working out some of the technological and information gaps, this will provide an excellent tool for standardizing reference site selection.

The most significant advantage of this method is that it provides a mechanism for standardizing the way that reference sites are chosen across a wide range of projects. Each of the steps involved in the method has aspects that can be modified to meet the needs of particular regions or project goals. However, even when different variations are used, this common structure provides a mechanism for communicating and comparing sites among projects.

We expect that we will need to add or subtract sites from the final list as we develop the methods, but the reference sites resulting from this exercise can be used for a wide range of applications including both multivariate (Reynoldson and Wright 2000) and multimetric bioassessments (Hughes 1995, Yoder and Rankin 1995). Potentially impaired sites (test sites) in the SNFE can be compared to this reference group to determine whether they are impaired.

In the interests of improving this methodology, we recommend that the SRWP coordinate efforts with other agencies and regions that are developing programs to and have the same needs. We recommend putting resources into the research needs identified in the discussion sections of each step. We also recommend putting resources into developing models that can be used to assess impact components that are not easily summarized by the ATtILA model. For example a model predicting sediment load (AnnAGNPS sedimentation model, USDA 2000) was developed by the University of Nevada, Reno. Other critical needs include

11644

estimating mining impacts, pesticide impacts and a means for summarizing the intensity of water manipulation within candidate areas.

Additional Research Needs

There are a number of related questions that should be considered while refining this methodology. The method can be used before this is done, but the quality of our inferences about stream condition will be greatly improved with better understanding of how our assessment of reference condition relates to these issues.

1. How important is inter-annual and seasonal variation?
2. How often should sites be re-sampled?
3. How does reference condition differ in non-perennial streams?
4. How many sites are needed to adequately define reference condition?

LITERATURE CITED

Bailey, R.C. M.G. Kennedy, M.Z. Dervish, and R.M. Taylor. 1998. Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. Freshwater Biology 39: 765-774.

Bailey, R. G., Avers, P. E., King, T., McNab, W. H., eds. 1994. Ecoregions and subregions of the United States (map). Washington, DC: USDA Forest Service. 1:7,500,000. With supplementary table of map unit descriptions, compiled and edited by W. H. McNab and R. G. Bailey.

Barbour, M.T., J. Gerritsen, B.D. Snyder and J.B. Stribling. 1999. Revision to rapid bioassessment protocols for use in stream and rivers: periphyton, BMIs and fish. EPA 841-D-97-002. U.S. Environmental Protection Agency. Washington DC.

Davis, W. S. and T.P. Simons, eds. 1995. Biological Assessment and Criteria: Tools for Resource Planning and Decision Making. Lewis Publishers. Boca Raton, FL.

Gibson, G.R., M.T. Barbour, 1996. Biological Criteria: Technical guidance for streams and small rivers. EPA 822-B-96-001. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

Harrington, J. M. 1999. California stream bioassessment procedure (CSBP). California Department of Fish and Game, Water Pollution Control Laboratory. Rancho Cordova, CA.

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

Hawkins, C.P., R.H. Norris, J. Gerritsen, R.M. Hughes, S.K. Jackson, R.K. Johnson, and R.J. Stevenson. 2000. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. Journal of the North American Benthological Society 19: 541-556.

Hughes, R.M. 1995. Defining acceptable biological status by comparing with reference conditions. Pages 31-47 in: W.S. Davies and T.P. Simon (eds.). Biological assessment and criteria: Tools for water resource planning and decision making. Lewis Publishers, Ann Arbor, Michigan.

Hughes, R.M., D.P. Larsen, and J.M. Omernik. 1986. Regional reference sites: a method for assessing stream potentials. Environmental Management 10: 629-635.

Hughes, R.M. and D.P. Larsen. 1988. Ecoregions: an approach to surface water protection. Journal of the Water Pollution Control Federation 60: 486-493.

Hughes, R.M., P.R. Kaufmann, A.T. Herlihy, T.M. Kincaid, L. Reynolds, and D.P. Larsen. 1998. A process for developing and evaluating indices of fish assemblage integrity. Canadian Journal of Fisheries and Aquatic Sciences 55: 1618-1631.

Jongman, R.H.G., C.J.F. ter Braak and O.F.R. van Tongeren [Eds.] 1995. Data Analysis in Community and Landscape Ecology. Cambridge University Press.

Karr, J.R. and E.W. Chu. 1999. Restoring Life in Running Waters - Better Biological Monitoring. Island Press, Covelo, CA.