Benthic Macroinvertebrates as Water Quality Indicators in Highly Urbanized Streams in the San Francisco Bay Region, California.

Final Report

August 15, 2005

Andree Breaux[1], Monique Born[2], Lynn Suer[1], Steve Cochran[3], Richard Looker[1]

[1]California Regional Water Quality Control Board, San Francisco Bay Region;
[2]Sustainable Land Stewardship Institute.
[3]Friends of the Estuary

With Assistance From: Roger Brewer, Naomi Feger, Jeff Kapellas, Revital Katznelson, Peter Krottje, Matt Cover, Steve Moore, Karen Taberski, Nelia White, and Farhad Ghodrati.

TABLES and FIGURES

ACKNOWLEDGMENTS

The macroinvertebrate portion of this study was funded by a grant from the California Coastal Conservancy to the Watershed Assessment Resource Center in order to test the validity of the creek assessment method described in this report and to assist in the development of protocols to be employed by volunteer creek monitors. Over the four year study, benthic macroinvertebrate collection and analysis was conducted by Monique Born of the Sustainable Land Stewardship Institute. In addition to the authors and those listed as providing special assistance for this project, many other Regional Board staff and watershed volunteers assisted with sampling over the four year study.

Conclusions and recommendations are those of the authors and do not necessarily reflect the opinions or policies of the San Francisco Bay Water Board or the State Water Resources Control Board.

Some Acronyms Used:
ANOVA: Analysis of Variance
BMI: Benthic Macroinvertebrate
CSBP: California Stream Bioassessment Protocol
CV: Coefficient of Variation
IBI: Index of Biological Integrity
SFB Water Board: San Francisco Bay, California Regional Water Quality Control Board (Region 2)
SLC: San Leandro Creek
SLSI: Sustainable Land Stewardship Institute
SWAMP: State Water Ambient Monitoring Program
WC: Wildcat Creek

ABSTRACT

We collected and analyzed benthic macroinvertebrate (BMI) communities from two creeks in east San Francisco Bay, California over a four-year period to test the hypothesis that these communities are reliable indicators of the water quality in urban creeks. In San Leandro Creek the preliminary results of the four-year study supported the hypothesis that highly impacted urban areas may have superior BMI communities in their upper watersheds. Wildcat Creek, however, revealed an impacted BMI community along with some toxicity even at the upper station. The relatively poor biological community at the highest station sampled during 2000 and 2001 may have been the result of excessive dog and human use. The addition of a sixth station in 2002 and 2003 at the upper limits of Wildcat Creek above the high use area provided good water quality and a high-quality BMI community, indicating that the creek still has the potential for high diversity.

The results indicate that BMIs can be a useful tool for assessing water quality even in highly impacted urban creeks. The most abundant groups in both creeks over the four years were chironomids, oligochaetes, simulid black flies, and baetids. Analysis of the variance of nineteen BMI community metrics indicates consistent differences between high and low quality stations over the four annual sampling events and substantial stability between creek stations from year to year. Some stations, within themselves, still had significant variation even with four years of data, but these differences would probably decrease with continued sampling. The most reliable metrics (those with the least significant differences at individual stations after four years) were taxa richness (the number of taxa) and tolerance value (the ability of the benthic macroinvertebrate to tolerate disturbance or pollution). Other useful metrics with consistently low coefficients of variation on both creeks included percent dominant taxa, Shannon diversity index, and percent collectors (although some of these metrics were among those that still had significant variation after four years when evaluated by analysis of variance).

In addition to providing data revealing consistent trends of the BMI communities in the two creeks over the four-year study, we hope to provide a framework for incorporating data collected by citizen volunteer monitors. Assessing the annual variability of the chemical, physical, and biological parameters of these two creeks should assist in assessing the overall health of other creeks being sampled as part of the Surface Water Ambient Monitoring Program (SWAMP), which is expected to rotate sampling between all creeks in the San Francisco Bay Region and the State over the coming years.

INTRODUCTION

California Regional Water Quality Control Boards (Water Boards) throughout the state are facing increasing pressure to certify the ecological health of creeks and other water bodies within their respective regions. Mercury, selenium, DDT, PCBs, dieldrin, diazinon, chlordane, siltation, pathogens, nutrients, and invasive species are some of the concerns listed on the San Francisco Bay Water Board's 303(d) and TMDL Priority List of impaired water bodies for 2000, 2001, 2002, and 2003. In 2000, we undertook a four-year study to measure some of these substances and others in two urban East Bay creeks – San Leandro Creek and Wildcat Creek -- to determine how they might change over an elevational and increasingly urbanized gradient from the upper watershed to the bay in both creeks. In 2001, the State Water Resources Control Board instituted a Surface Water Ambient Montioring Program (SWAMP), which is seeking to develop a state-wide comprehensive environmental monitoring program, into which we have fed our results. However, a further purpose of the our study was to test the effectiveness of the use of benthic macroinvertebrates (BMIs) for the assessment of water quality by determining if the benthic communities remained stable at the same testing stations over a four-year-period and if variations between stations could be explained against the expectation that higher quality communities of BMIs will appear at the upper stations where higher water quality is to be expected. The focus of this paper will be on the analysis of our BMI results with reference to the overarching problem of monitoring and regulating the water quality of urban creeks.

The paper covers the years from 2000 through 2003. Each year, water and BMIs were sampled over two to three days from each creek and sediment was collected once from each station in 2001 (and from one additional station in 2002). Due to budget constraints, only BMIs and limited water quality measurements were collected in 2003.

Both creeks are highly urbanized. Large urban parks are located at the top of each watershed, and extensive development is present along the mid and lower reaches, although parts of both watersheds are protected by East Bay Municipal Utilities District. Stations were selected in order to represent both the upper watershed – expected to show healthy BMI communities and good water and sediment quality – and the lower watershed – expected to show varying degrees of adverse impacts due to urbanization. Stations were also selected based on previous sampling of some of the sites by other groups or individuals, and on both physical and legal access. The two creeks in the context of the East San Francisco Bay are depicted on Figure 1 and the station locations are provided in Table 1.

The Beneficial Uses of the two urban streams include cold and warm freshwater habitat, sport fishing, estuarine habitat, groundwater recharge, fish migration, preservation of rare and endangered species, water contact recreation, noncontact water recreation, fish spawning, and wildlife habitat. The standards for protecting these beneficial uses are based on aquatic life protection; on human health via fish consumption, and contact recreation; and on wildlife protection via bioaccumulation and direct contact.

The goals of the study were to:

- assess the health of the two creeks based on elevation and urbanization compared with each other and with western U.S. creeks in general;

- determine whether BMIs can be reliable water quality indicators in highly urbanized creeks;
- begin a set of baseline data that could be the foundation for future trend monitoring; and
- provide a framework for future data collection by citizen monitors.

It is important to emphasize that the data described in this report represent an average of only two days of sampling in each of the four years, and that data collected at other times may show different results. BMIs were selected as a tool of water quality analysis because they tend to represent stream health over longer time periods than discrete water or sediment samples. Consequently, the San Francisco Bay Water Board (or Region 2) along with other California Regional Water Boards are investigating the use of BMIs as biological indicators of water quality using the California Stream Bioassessment Procedure (CSBP) or a similar version, as described by Harrington and Born (2000). Our results, to be elaborated in the following report, although limited, indicate that the use of BMIs may be the most effective method of monitoring and regulating the water quality of creeks in a highly urbanized environment.

Results from this study will contribute to the overall assessment of the utility of analyzing BMIs in highly urbanized creeks in California and other western states with dense populations or highly impacted creeks. In 2001, Wildcat and San Leandro Creeks were part of the continuing pilot study for the San Francisco Bay Region's SWAMP which will tie ambient monitoring to the 305(b) report and the 303(d) list for the region. Although funding has been cut in the last two years, SWAMP hopes to examine over 50 planning watersheds and reservoirs in Region 2 over a fifteen-year rotational cycle. Since 2000, SWAMP has completed sampling in twelve planning watersheds with an additional five being sampled in 2004/05.

The first four years of bioassessment data have been analyzed for this report to determine whether the temporal variation between years and stations of the BMI populations are within an acceptable range of variation. A fifth and final year has been collected by the Sustainable Land Stewardship Institute (SLSI), but not yet analyzed. Further analysis of the BMI data from this study will be done by the SWAMP program as part of an overall regional and state assessment. Other agencies in the San Francisco Bay Area that are using the CSBP or variations of this procedure to assess creek health include the Alameda Countywide Clean Water Program, the Marin County Stormwater Pollution Prevention Program, and the Contra Costa Countywide Clean Water Program. For this study, volunteers assisted during some years with annual sampling, though no formal training sessions were provided. To the extent that training can be provided in the future, and citizen volunteers can produce data that passes the test of quality assurance, volunteers should prove useful for sampling, analyzing data, training other volunteers, and educating the public about the local creeks.

Physical Setting

The San Francisco Bay climate is predominantly Mediterranean with distinct annual dry and wet seasons which show large inter-annual and intra-annual variations. Average annual rainfall in the area ranges from 5 to 50 inches and averages about 23 inches (Collins et al. 2000). The highly seasonal flows generally bring more than 90% of the annual runoff during the wet season between November and April (San Francisco Bay Regional Water Quality Control Board, 1995).

Wildcat Creek is located in the East Bay area in Contra Costa County (Figure 2), and in Region 2's San Pablo Basin Hydrologic Unit. The watershed is approximately 8.7 square miles and includes Wildcat Creek and two impoundments (Lake Anza, and Jewel Lake), in addition to Harvey Creek which serves as a large tributary to Wildcat Creek (Collins et al. 2000). General land uses include protected park lands for recreation, education, and conservation; protected watershed lands of East Bay Municipal Utilities District; range land (for over 110 years in some areas); and urban uses. Management issues involve dams, water releases, flood control, recreational uses, urban runoff, rangeland, proposed developments, erosion and sedimentation, drinking water, and sensitive species (Moore 2000; San Francisco Bay Regional Water Quality Control Board 1995; Urban Creeks Council 1996). Activities or uses in Tilden Regional Park which might adversely impact water quality include pony rides, the “Little Farm”, a swimming beach at Lake Anza, a duck pond at Jewel Lake, horseback riding, a golf course, and a botanical garden. In Wildcat Canyon Regional Park cattle grazing has occurred since the early nineteenth century, but has been prohibited in Tilden Regional Park since 1936 (Urban Creeks Council 1996). Our original upper Wildcat Creek watershed station was located below one of the impoundments, the botanical garden, and the golf course, and showed indications of impacted biological communities. An additional station was added in order to provide a site less impacted by human uses.

San Leandro Creek is also located in the East Bay area in Contra Costa and Alameda Counties (Figure 3), and in Region 2's South Bay Basin Hydrologic Unit. The watershed is approximately 46.5 square miles and includes the following waterbodies: San Leandro Creek, Redwood Creek, Grass Valley Creek, Kaiser Creek, Buckhorn Creek, Moraga Creek, Indian Creek, Lake Chabot Reservoir, and Upper San Leandro Reservoir. General land uses include protected park lands for recreation, protected watershed lands of East Bay Municipal Utilities District, and urban uses. Specific management issues include dams, water releases, drinking water sources, water transfers, recreational uses, urban runoff, and sensitive species (East Bay Municipal Utilities District and C.R. James & Associates 2000; San Francisco Bay Regional Water Quality Control Board 1995 and Moore 2000).

Two major reservoirs contained by dams exist on the creek: the Upper San Leandro Reservoir, which is used for domestic water supply, and Lake Chabot, which is used for recreation and emergency supplies. Winter and spring flows are generally adequate to provide a variety of aquatic life, while summer conditions remain wet only on upper San Leandro Creek which is fed by water that leaks from the dam. The lower segment of the creek – about 5 miles -- has only isolated pools in the summer resulting from outfalls along the creek and possible discharge of groundwater as baseflow or springs. Stormwater in the winter and spring make up the flows in the urban segment (URSGWC 1999).

Although the downstream segment of the stream is thoroughly urbanized, it has been determined to have good fisheries habitat when water is flowing. Adequate flows that depend on precipitation and releases from the dam and Lake Chabot are key features for the future health of this creek.

Some of the fish species observed in both Wildcat and San Leandro Creeks include rainbow trout (*Oncorhynchus mykiss*) and stickleback (*Gasterosteus* sp.) (Sheppard et al. 2001; Collins et al 2000; Leidy 1999; Urban Creeks Council 1996). At the end of both creeks, marshes exists which harbor the endangered California clapper rail (*Rallus longirostris*) and, in Wildcat, the endangered salt marsh harvest mouse (*Reithrodontomys raviventris*) (Collins et al. 2000).

METHODS

The name, location, and description of all testing stations are listed in Table 1 along with the years for which data was collected. Sampling dates occurred in spring of each year and are provided in Table 2. BMIs and basic water quality measurements were collected in each of the four years; water quality samples were collected in 2000, 2001, and 2002; and sediment was collected in 2001 (at ten stations) and 2002 (at one station). Detailed methods and results are provided in previous annual reports for water quality and sediment analysis for this project [1], so only the more important points will be re-stated here. Since one goal was to evaluate the use of BMIs as a water quality indicator in highly urbanized creeks, the full four-year dataset for the results of the BMI creek community will be presented.

Procedures for collecting macroinvertebrate samples were followed to the extent practicable as described in Harrington and Born (2000) for the CSBP. Those procedures call for the random selection of three riffles from a stream reach with at least five riffles and the formation of one composite sample from three samples collected in a kick net from each riffle[2]. After samples are collected in a kick net and composited, they are sieved through a 0.5 mm mesh sieve, and placed in a jar with 95% ethanol. The number of composited samples from Wildcat Creek and San Leandro Creek totaled three per station per year (three composited samples from three different riffles), resulting in a total of 30 samples from the 10 stations for 2000 and 2001, 33 samples from the 11 stations in

[1]Breaux, A., M. Born, S. Cochrane, L. Suer, R. Brewer. A Watershed Assessment of Two East Bay Creeks, San Francisco Bay, California: third year report on bioassessment, water, and sediment quality sampled in April 2002 [Draft, June 2003]; Also see : second year report on bioassessment, water, and sediment quality sampled in May 2001 [Final Draft, December 2002]; and first year report on data collected in May 2000 [November 21, 2000], San Francisco Bay Regional Water Quality Control Board, Oakland, CA.

[2] Note these procedures have been slightly altered in the CSBP 2003 which now recommends that the stream reach being assessed be measured as a discreet length of 300 feet instead of as a five pool-riffle sequence. There is also the option of compositing the sample in the lab. See California Stream Bioassessment Procedure, Department of Fish and Game, Aquatic Bioassessment Laboratory, December 2003 (www.dfg.ca.gov/cabw/cabwhome.html.). For some studies with small data sets and limited funding, such as this one, replicates are good to keep in the analysis. If we move to a 500 count from a 900 count, the replicates will be lost because the samples will be composited (Tom King, pers. comm.) So, for this study which seeks to determine variation between years and between stations, we have chosen to keep the replicates and the 900 count, rather than to re-calculate for the 500 count. The SWAMP analysis can re-calculate for the 500- count in order to make this dataset comparable with the rest of the San Francisco Region and the State.

2002, and 34 samples in 2003 (one sample was added at the San Leandro Creek Tributary above station SL 5). The total number of samples for the four years was 127.

Those 127 samples were subsequently analyzed by staff and contractors from the Sustainable Land Stewardship Institute (SLSI), again using the CSBP described in Harrington and Born (2000). BMIs were sorted and identified in the laboratory to the species level. Sensitive orders were distinguished as Ephemeroptera, Plectoptera, and Trichoptera (EPT) taxa. Biological metric values were then calculated based on taxonomic richness (overall diversity and numbers of EPT taxa), composition measures (sensitive EPT Indices), tolerance/intolerance measures (percent tolerant and intolerant organisms, percent Hydropsychidae and Baetidae, percent dominant taxon*)*, and trophic measures (functional feeding groups), and overall abundance. This report is based on (1) two detailed reports for 2000 and 2001 written by M. Born (SLSI) (and included as appendices in the first two creek reports for this study [Breaux 2000 and 2001]), and (2) two additional data sets (without analysis) for 2002 and 2003.

RESULTS

BENTHIC MACROINVERTEBRATES (BMIs)

BMI communities in San Leandro Creek consistently followed the pattern of high quality at the upper stations (SL4 and 5) with decreasing quality at the lower stations (SL 1,2,3). Wildcat Creek had more variability throughout the creek and the BMI community at the original upper station at Lone Oak in Tilden Park (WC5) was almost as poor as the lower three stations (WC 1,2,3). The addition of WC6 at Big Springs for the last two years of the four year study (2002 and 2003), along with the restoration site at Alvarado Park (WC4), provided relatively high quality BMI communities for Wildcat Creek. These general results for both creeks are depicted in the tables and figures below.

Table 3 contains sixteen summary metrics for the eleven stations plus one single sample taken from a tributary (SL Trib) approximately 30 feet from the Upper San Leandro Creek (SL5) station collected in 2003. Five primary summary metrics averaged over the four years (only two years for WC6) for each creek are displayed in Figure 4. Taxonomic richness was generally higher in San Leandro Creek, and especially high in the upper creek stations (SL4, SL5, and SL Trib). On Wildcat, the four-year average was highest at WC4 followed by the two-year average at WC6 located at the highest station. The EPT index followed a similar and related pattern, though more of the lower creek stations on San Leandro lacked EPT taxa than on Wildcat Creek. Sensitive EPT taxa (those with tolerance values of less than four) were also found at the upper San Leandro Creek stations and, to a lesser extent, at the added upper creek station on Wildcat (WC6) and also at the restoration site on Wildcat (WC4). The final two metrics on Figure 4 showing percent intolerant and tolerant organisms are essentially mirror images of each other, with the organisms intolerant to pollution at the upper San Leandro Creek stations, and at WC6 and WC 4, and the tolerant organisms at the lower stations of both creeks as well as a substantial percentage at the highly used urban park (WC5).

Coefficients of Variation

For the summarized metrics on Table 3, coefficients of variation (CV) are included for fifteen of the metrics[3]; and CVs less than 25% are considered to have acceptable variation among the triplicate samples in the original data sets (Harrington and Born 2000). Using this rule of thumb, the number of CVs greater than 25% is listed on Table 1 for each annual metric. Low variation indicated by a low number (less than 5) of CVs greater than 25% is found in Wildcat Creek for taxonomic richness, percent dominant taxa, Shannon diversity index, tolerance value, and percent collectors, indicating that those metrics are fairly representative of the creek during the respective sampling events. Metrics from Wildcat with a high amount of variation indicated by ten or more instances of CVs greater than 25% are EPT index, sensitive EPT index, percent intolerant and tolerant taxa, and all the feeding types except collectors. Metrics with few occurrences (less than 5) of CVs greater than 25% on San Leandro Creek were taxonomic richness, percent dominant taxa, ETP taxa, sensitive EPT index, Shannon diversity, tolerance value, percent intolerant taxa, and percent collectors. The highest variation (greater than 10 times) on San Leandro Creek was found for percent tolerant taxa and all the feeding groups except collectors and shredders. Variation among feeding groups was found in the Russian River BMI study with noticeable differences in communities, though these differences revealed themselves in only slight differences in BMI metric values (California Department of Fish and Game 1999). The statistical analysis to be performed by SWAMP should provide further indication of the reliability of the metrics over the four annual collections. Note that the use of coefficients of variation is generally considered more useful for analyzing good dose-response metrics (e.g., sensitive EPT) from reference sites than from test sites (S. Moore & P. Krottje, pers. comm.).

Top Five Species

The top five species for each station and year are summarized in Table 4(a) and (b). Taxonomic groups include genus and species when provided, or higher taxonomic classifications when genus or species is not practical. For example, the Chironomid family is represented by four genera, the Oligochaete class is represented by several families, and the EPT taxa are distinguished from the other groups. Taxonomic Groups are ranked by tolerance value, with those more tolerant of pollution at the top of the table. Thus the lower Wildcat Creek stations (WC1, WC2, WC3) shown on Table 4(a) and the station in the highly used urban park (WC5) shows a substantial portion of the top five species from the high tolerance groups, dominated by oligocheate worms, chironomids, and species from the Baetis family which, while technically an Ephemeropteran (E), tends to be more tolerant of pollution than most of the other baetid species (Harrington and Born 2000). The restoration site (WC4) shows a cumulative total of eleven taxonomic groups over the four years which represent a total of only 74 percent, indicating high species diversity. The upper creek station on Wildcat (WC6) does not show as many species contributing to a relatively low percent total as WC4, but it does show a predominance of species that are intolerant to pollution. For the entire creek, the most abundant group in Wildcat Creek was the chrionomids, followed by oligocheaetes, baetids, and simulid black flies.

[3] Calculated by dividing the standard deviations by the sample means and expressing the results as a percent.

Table 4(b) provides the same information for San Leandro Creek revealing a predominance of highly tolerant groups at the lower stations (SL1, SL2, and SL3) but with greater diversity and less tolerant groups at the two upper stations (SL4, SL5). The lower stations are dominated by oligocheates, chironomids, copepods, cladocerans, and simulid blackfly larvae which can generally live in turbid environments with poor water quality (Harrington and Born 2000). The upper stations have relatively low cumulative percentages over the four years (SL4 = 66%; SL5 = 70%) representing a more even distribution of functional feeding groups and more pollution intolerant taxa. Like Wildcat Creek, San Leandro Creek's most abundant groups were chironomids and oligocheates with simulid black flies slightly more numerous than the baetids.

Tolerance Values

Figure 5(a) and (b) show the tolerance values (TV) for the 11 stations (plus the SL5 Tributary) averaged over the four years and grouped into categories of High (TV 8-10); Medium (TV 5-7); and Low (TV 0-4). The two lower Wildcat Creek Stations (WC1 and WC2) show both medium and high tolerant species with none in the low category. WC3 begins to show a few in the low category and higher upstream at the restoration site (WC4) a marked improvement in the BMI community is indicated. However, the community deteriorates again at WC5 which is impacted by a high use area with most of the station given to high and medium tolerant species. Finally, WC6 reverts to the more typical BMI community expected in the upper creek. Without the confounding circumstance of the restoration project station (WC4) and the highly used upper station (WC5) on Wildcat Creek, the tolerance values for the BMI community in San Leandro Creek show the typical pattern of improvement from lower stations (SL1, SL2, SL3) to the higher stations (SL4, SL5) with tolerant groups giving way to intolerant groups as the stations proceed upstream. The single one-time sample taken at the SLC Tributary to SL5 does include a small high tolerant group where none are seen in SL4 and SL5 but the medium and low tolerant groups predominate. While both Figure 5(a) and 5(b) show some variation for tolerance values between years at each station in both creeks, there is no major movement between years from one tolerance group to another at any station.

Functional Feeding Groups

Figure 6 displays the percent of functional feeding groups at each station averaged over four years (except WC6 which was averaged over two years and SLTr which was sampled only once and not in triplicate). Categories include shredders, filterer-collectors, collector-gatherers, scrapers or grazers, and predators. Predominant feeding types in small-order headwater streams are shredders and collectors; in larger streams with more open canopies and algae, scrapers are likely to predominate; and in large rivers with low gradients, fine sediment, and no canopy, collectors tend to predominate with some predators present (Harrington and Born 2000) . Generalist feeders which include collectors and filterers tend to be more tolerant of pollution, while specialists such as scrapers, piercers, and shredders are more sensitive to pollution and often reflect a healthy stream (U.S. EPA 1999).

The functional feeding groups in Wildcat Creek over the four year period were predominantly collectors represented by the oligocheate worms, cladocerans, and some of the chironomid groups. Even at the higher quality stations on Wildcat (WC4 and WC6), the collectors represent no less than half of the feeding groups. San Leandro Creek, on the other hand, had a predominance of collectors at the two lower stations (SL1 and SL2) and added a large portion of filterers at the middle station (SL3) where low flows and abundant algae tend to decrease water quality to the extent that filterers (mostly chironomids and simuliidae), not scrapers predominate. The upper two San Leandro Creek stations (SL4 and SL5) show the most balanced functional feeding group communities with all five groups well-represented. The SLTr sample also reflected a well-balanced community for the one time it was sampled.

Relative Ranking

Table 5 ranks the four-year averages for six of the major metrics (taxonomic richness, EPT index, sensitive EPT index, percent tolerant, percent intolerant, and Shannon diversity index) at each station against each other by first determining the statistical rank (the standard error) of each value and then providing a relative rank. Figure 7 shows the eleven stations ranked for each metric, with lower scores indicating the better BMI communities. SL5 and SL4 on San Leandro Creek have the best scores, followed by WC6 and WC4. The highly used urban park on Wildcat (WC5) ranks almost as poorly as the lower Wildcat Creek Stations (WC1 and WC2), and the lower San Leandro Creek stations (SL1 and SL2) score the worst for these metrics.

It should be noted that other metrics could be selected and some of the six metrics used here are related variables (e.g., taxonomic richness and Shannon Diversity index; percent intolerant and percent tolerant) and these can be expected to move in similar directions at the same stations. With additional time and funding, a more complete analysis can be conducted in the future to further analyze the other major metrics, though it is expected that those selected here for analysis are representative of the overall quality of the BMI communities in both creeks. The final SLSI report for this project and the statistical analysis of this data provided by SWAMP, in addition to the analysis of all SWAMP data collected in the San Francisco Bay Region and in the entire state, should provide a more complete picture of the reliability of BMI data for assessing water quality.

Analysis of Variance

Analyses of Variance (ANOVAs) were conducted to test two hypotheses: first, that BMI communities would consistently differ between upper and lower stations (inter-station differences) and secondly, that the BMI metrics would remain constant within each station from year to year (intra-station variability). The ANOVAs used ninety-five percent confidence intervals for means based on pooled standard deviations, and significant differences were determined at an alpha value of 0.05, with further verification provided by F values and pair-wise comparisons.

Results testing the first hypothesis confirmed that there were differences between the stations on either creek over the four years (between station variation). The overall picture on both creeks shows consistent differences between stations for the nineteen major metrics provided by SLSI for this project. Given the large amount of pair-wise data analyzed for this test, no table is provided but Figure 8 shows the results of the four year analysis for each station on each creek (note that "SL Trib" is not a station for this project, but only a one-time sample taken above SL 5). Figure 8(a) shows taxa richness, EPT taxa, EPT index, sensitive EPT, Shannon diversity, tolerance value, percent intolerant, and percent tolerant. All of these metrics tend to show San Leandro Creek with the extremes of high quality BMI communities at the upper stations (SL 4 and SL 5) and the extremes of low quality BMI communities at the lower stations (SL1, 2, and 3). Wildcat Creek has more moderate values across the entire creek, with WC 4 and WC 6 showing the best BMI communities.

Figure 8(b) shows the dominant locations and abundance of the EPT taxa, with Ephemeroptera, Plecoptera, and Tricoptera prevalent at the upper San Leandro Creek stations. The percent Hydropsyhchidae (a relatively tolerant Tricoptera) is relatively high at the upper San Leandro and Wildcat Creek stations, but the percent Baetids (a relatively tolerant Ephemeroptera) is relatively high at most Wildcat Creek stations, even at the stations with the better BMI communities (WC4 and WC6).

Figure 8(c) shows the distribution, abundance, and variation of functional feeding groups at each station, with the collectors dominating the lower stations on both creeks, and present in substantial numbers throughout Wildcat, even at the higher quality stations (WC4 and WC6). Filterers are present in substantial, though varying, amounts at SL 3 and WC1, and present to a lesser degree at all other stations. Grazers are also present at all stations and most abundant at SL 4; the most variable stations are SL4, SL3, and WC4. Finally, predators and shredders are highest in the upper San Leandro Creek stations of SL4 and SL5 with fewer though still substantial populations of both at the higher quality Wildcat Creek stations of WC4 and WC6.

In sum, the nineteen metrics displayed on Figure 8 (a, b, c) reveal a consistent pattern of differences between stations over the four years of very high quality BMI communities on upper San Leandro Creek, very low quality ones on lower San Leandro Creek, and a range between those two extremes on Wildcat Creek. The two best stations on Wildcat are at the restoration project site (WC4) and at the highest station added in 2002 (WC6) where relatively higher flows in the former and low human or dog use in the latter provide situations for healthy BMI communities.

To test the second hypothesis that the individual creek stations were the same from year to year (intra-station or within station variation), ANOVAs were conducted for each station by year for seven major metrics including taxa richness, EPT taxa, sensitive EPT, Shannon diversity index, tolerance value, percent Baetids, and percent collectors[4]. The latter two metrics were selected arbitrarily only to check if there were differences between the groups over the years. Means and standard deviations for stations by year are presented in Table 6. Results of the ANOVAs are presented in Table 7(a) and Table 7(b) provides a summary of the number of measurements at each station showing significant differences over the four annual sampling events on each creek. While San Leandro Creek appears to have fewer significant differences than Wildcat Creek, the

[4] The SWAMP analysis for the San Francisco Bay Region is concentrating on the analysis of taxa richness, percent EPT taxa, and Shannon diversity. Two of these are analyzed in this report and a third, EPT taxa (the number of families from the EPT groups), is substituted for percent EPT (percent composition of EPT taxa). There is not likely to be a substantial difference between EPT taxa and percent EPT but the final SWAMP analysis may prove otherwise. Some of the SWAMP data analyzed for Region 2 found that sensitive EPT were more prevalent at moderately disturbed sites than at sites with little disturbance (Matt Cover, UC Berkeley, pers. comm., 3/30/04). This does not appear to be the case for the eleven stations sampled for this study, where SL 4 and 5 at upper San Leandro Creek had consistently higher sensitive EPT scores than the other, more disturbed sites on both creeks (See Figure 8a).

creeks are not comparable based on the ANOVAs for these selected metrics because three analyses on San Leandro Creek did not produce appropriate data for the statistical analysis and Wildcat Creek had a total of six stations (one sampled for only two of the four years) while San Leandro Creek had only five. Still, it is useful to point out which metrics had the fewest significant differences at each station given these qualifications.

Of the seven metrics, taxa richness had the least number of stations showing significant differences and was therefore the most reliable in terms of having the least variation, followed by tolerance value for those metrics with complete statistical analyses. While EPT taxa and sensitive EPT are not analyzed for all stations on San Leandro Creek, these two metrics show a fair amount of variation on Wildcat Creek with four of a total of six stations showing significant differences for EPT taxa and five of a total of six stations showing significance differences for sensitive EPT taxa over the four years. There were also significant differences at a substantial number of stations for the following metrics: Shannon diversity (four of five stations on San Leandro Creek and five of six stations on Wildcat Creek); percent Baetids (four of five on San Leandro Creek and four of six on Wildcat Creek); and percent collectors (four of five on San Leandro and five of six on Wildcat Creek). The 20 measurements at each station showing significant differences on San Leandro Creek represent 62% of the possible total of 32; and the 28 measurements at each station showing differences on Wildcat Creek represent 67% of the total of 42 measurements at each station analyzed statistically for variance. Figures for each of the stations with the seven metrics analyzed statistically are shown in Appendix 1.

While there were still statistically significant differences at some stations for several of the seven metrics analyzed here, it is important to point out that some of the metrics analyzed are related to others; that the remaining twelve metrics might reveal less variation (though we doubt this), and that a longer data set and/or more intensive data collections at each site over a year is likely to decrease the amount of variation over the years (though completely different seasons might produce completely different communities).

In sum, over the study period, the BMI communities remained distinct at each station, revealing consistent differences between stations from year to year. While some stations had more variation than others, the overall picture was one of similarity. Upper San Leandro Creek stations had excellent BMI communities, lower San Leandro Creek had very poor BMI communities and Wildcat Creek had BMI communities in between the two extremes, with WC4 and WC 6 showing the best groups on that creek. While there was still enough within station variation from year to year to show statistical differences for some metrics, these differences would be expected to decrease with continued sampling over a longer period.

As noted above, the location of WC5 in the upper part of Wildcat Creek did not conform to the “river continuum” theory where the upper station should have a more sensitive and less pollution tolerant BMI community than most of the lower stations and, consequently, a higher station at Big Springs (WC6) was incorporated as a permanent station in 2002 for this study. Reasons why the lower station at WC4 may have a more pollution intolerant BMI community compared to WC5 include the following: WC4 has the advantage of stronger flows, is located at the site of a creek restoration project where weirs have been placed to direct flows and willow wattles have been planted to encourage vegetation growth, and is also extensively lined with stone walls that may decrease erosion despite high use by dogs and humans. WC5, on the other hand, is located in an area of the creek with lower flows that spill into a pool that is used extensively by dogs and people and has a high incidence of eroded banks. A study by East Bay Regional Parks District found WC5 to have the highest use at six creek stations surveyed with up to 109 human users, a mean of 28 dogs, and elevated turbidity levels for 80% of the 2-hour survey period. WC4 at Alvarado Park was impacted less, but still used substantially by dogs and humans. The study found that for all the six sites surveyed, 66 of the 154 dogs (43%) were observed actually disturbing the creek. Thus a likely explanation for the poor BMI community at WC5 is the high dog use which causes bank erosion and high turbidity (Sheppard et al. 2001).

WATER

[As noted above, sediment and water samples were discussed in earlier reports, so only the highlights are mentioned here.]

Flow rates for the two creeks are summarized in Table 8. Flows were generally low on both creeks and tended to be higher at the higher elevations on Wildcat Creek though WC4, the midway station containing the restoration project, had the highest flows in 2001. SLC3, the mid-way station on San Leandro Creek also tended to have slightly higher flows than the other stations for two of the three years.

Pathogens

Table 9 shows the high variability from single coliform samples collected with the BMI samples collected each year. The sampling conducted by SWAMP in 2001 (Table 10) at six of the eleven stations used for this study was based on the required five samples over 30 days for total, fecal, and E. coli pathogens. As indicated on Table 10, most sites sampled for the SWAMP study that are also part of this study exceeded some criteria used by SWAMP, with some of the stations showing considerable variability from week to week (e.g., Huckleberry Preserve and Root Park on San Leandro Creek). The high variability suggests the need for more coliform testing in both creeks carried out by the counties or cities. If persistent problems are revealed with further testing, then signs should be posted stating that the water is not safe for drinking or primary contact recreation. Looking at the bacterial levels of these two creeks in the context of all the SWAMP sites in Region 2 and in the state, will put them in a broader perspective.

Toxicity Tests

Table 11 summarizes the results of the three-species chronic toxicity tests conducted in 2001 and 2002 at the two upper and two lower creek stations. Species tested were a green algae (*Selanastrum capricornutum*), a crustacean (*Ceriodaphnia dubia*), and the fathead minnow (*Pimephales promelas*). Of twenty possible toxicity tests, two were invalid and another four were potentially invalid due to a pathogenic infection caused by the bacteria *Columnaris* which is evidenced by a fungal halo. For the reliable tests, the lowest station on San Leandro Creek (SL1) caused toxicity in three species, and Lone Oak at Tilden (WC5) was toxic to two species, although the latter station was only tested in one of the two years. In 2002 Big Springs (WC6) was substituted for WC5. Even

WC6, however, was toxic to minnows (though this may have been due to the *Columnaris* bacteria.) Given the general lack of consistent toxicity in the results, the problems with lab controls in some instances, and the presence of bacteria that may have confounded the results at the two upper watershed stations, these tests remain inconclusive and do not validate or invalidate the hypothesis that the higher stations will be less toxic than the lower ones. Toxicity tests are part of the overall SWAMP and other sites may show more conclusive results than those from Wildcat and San Leandro creeks. If not, toxicity tests can be reserved for sites with expected contamination where they can supplement data for other parameters, such as chemical concentrations.

Water samples for dissolved metals and some total metals (specifically mercury) was generally unreliable so that data is not included here.

SEDIMENT

Sediment was collected once from each station for pesticides, metals, and PAHs (Table 12). Stations with three or more pesticides included those with poor BMI communities namely WC5, SL3, SL2, and SL1 though WC2 and WC1 also had poor BMI communities with no pesticides detected. Sediment arsenic levels tended to be higher on Wildcat Creek though were generally within the background range for the San Francisco Region. Nickel exceeded some screening levels at most stations including those with the best BMI communities, indicating that the higher levels are not deleterious to aquatic life. TPH and PAC were only measured at stations WC3, SL2, and SL1 due to visible or aromatic signs of the substances. TPH and PAC were found at both SL2 and SL1 and TPH and oil and grease were found at WC3. Detection limits for PCBs were too high to be useful, but split samples analyzed by a different laboratory for two stations – SL2 and SL1 – showed elevated levels at both stations. These two lower San Leandro creek stations typically had extremely low or no flows during sampling.

Final tables included here are Table 13 which contains metals accumulation levels in mussels planted in San Leandro Bay at the base of San Leandro Creek in 1998, and Table 14 which contains water quality data collected with the BMI samples in 2000, 2001, 2002, and 2003 by staff from the SFB Water Board and volunteer creek monitors. These tables are not analyzed here but made available for comparison with future data and analysis.

DISCUSSION

Results of this study show that the nineteen BMI metrics displayed consistent trends with significant differences between stations (inter-station differences) for most of the BMI metrics analyzed over the four year period. The BMI analysis of 19 metrics revealed substantial stability between creek stations from year to year (inter-station variation) with stations following consistent patterns throughout the study. Comparing both creeks, San Leandro ranked highest for its upper stations and lowest for its lower stations while Wildcat Creek stations fell between the two extremes of San Leandro Creek's upper and lower stations.

Taxa richness and tolerance value were the most reliable metrics (i.e., those with the least variation) of the fifteen metrics with CVs analyzed for general variation and of the seven metrics analyzed for statistical variation between years at the same stations (intra-station variation). Other useful metrics with consistently low coefficients of variation on both creeks included percent dominant taxa, Shannon diversity index, and percent collectors. A preliminary analysis of SWAMP data for the San Francisco Region also recommended taxa richness and tolerance value, along with diptera taxa richness and percent EPT taxa, as reliable metrics based on low natural variability, ability to discriminate disturbance and low redundancy (Cover 2004).

While analysis of the variation between the four years at individual stations (intra-station variation) revealed acceptable variation for two of the seven metrics analyzed, namely taxonomic richness and tolerance value, the remaining five metrics had either inappropriate data for this type of analysis or showed statistically significant differences between years. The between year variation in the Wildcat/San Leandro Creeks study in some metrics is likely to decline with additional sampling. Unfortunately this four year study is concluded, but future sampling can take place during any future April or May to determine whether the statistical variation declines with stable communities or whether human disturbances or climatic variations alter the predominant BMI communities at any station. The number of years required to find little or no variation at stations depends on the type of statistical analysis performed, the sample size, and the number of years that data is collected. Thus, additional sampling over the years or different types of statistical analyses run by SWAMP, SLSI, or an academic or non-profit institution may help determine the appropriate length of the sampling period for adequate trend analysis in these two creeks and other creeks in the region or state. For the purpose of this paper, the overall BMI communities on Wildcat and San Leandro Creeks appeared stable and consistent over the four spring sampling events.

In the interest of trend monitoring, it would be useful to investigate the relationship between significance values, power, estimated variance, the number of samples, length of monitoring period, and precision criteria to determine the number of samples and years that would be required to detect population changes under various scenarios. Paige et al (2000) reported that most five-year surveys lack the funds to detect trend levels useful for conservation purposes, but that with every added year, the required sample size decreases substantially and levels off between 10 and 20 years. Thus it is not unreasonable that -- even with strained budgets -- cities, counties, the state, or academic or nonprofit institutions, might set up and independently fund annual BMI monitoring schedules that might continue indefinitely.

Attempts are being made throughout the state of California to develop reliable indices of biological integrity (IBI) for effective BMI assessments (e.g., Ode et al. 2005; CA. Department of Fish and Game, 1999 and 2003; Harrington and Born 2000; Puckett 2002). An updated IBI for the north coast should be available next year (J. Harrington, Department of Fish and Game, pers. comm.). A recent report on the progress of Southern California's IBI recommended the following metrics as reliable: percent collector-gatherer +collector-filterer individuals, percent non-insect taxa, percent tolerant taxa, Coleoptera richness, predator richness, percent intolerant individuals, and EPT richness (Ode et al. 2005). Some of these have been used in the San Francisco Region and others may be inappropriate for this region, though most of these summary metrics could be developed based on the taxonomic lists provided by the Aquatic Bioassessment Laboratory for SWAMP and by SLSI. The forthcoming analysis of the SWAMP data from the San Francisco Region and the other California Regional Water Quality Control Boards will provide recommendations for developing an IBI for the San Francisco Region and for selecting which BMI metrics are most reliable for describing these communities in wadeable streams of varying levels of disturbance (Matt Cover, pers. comm., SFB Water Board).

With continued data from SWAMP being added over the coming years, along with BMI data from the other programs using the CSBP, targeted riffle, or multi-habitat methods, the San Francisco Bay Area should be able to develop its own IBI in the future as a useful tool for ranking streams. Moreover, as SWAMP continues to describe BMI communities in San Francisco Bay region creeks, the Water Board is attempting to move from descriptions of populations to studies designed to detect differences between sites based on land uses or nonpoint sources that affect streams (Cover 2004). Consistent use of the CSBP or comparable assessment tool can develop a reliable reference database for the region and can also produce future hypotheses that can be tested with adequate replication, stratification, and statistical power. The success of the CSBP and any other assessment methods will "depend on its ability to detect environmental perturbation outside the range of natural variability" (Rosenberg and Resh 1996) which is the analytic approach currently being taken by the San Francisco Bay SWAMP.

RECOMMENDATIONS

The following recommendations should be considered to either standardize the SWAMP protocols throughout the Region and State or to simply improve future sampling, analysis, and understanding of San Leandro Creek and Wildcat Creek.

Recommendations for SWAMP:

1. Analyze the fifth year of data (collected in May 2004) which were not available for this four-year report. Emphasis should be on the following components to determine stream health (Jim Harrington, pers. comm.):
    a. convert all 5 years of data to new composite sample recommended by SWAMP;
    b. characterize the two watersheds by producing a human disturbance gradient;
    c. include physical/habitat data as well as hydrologic and watershed data for the two streams;
    d. determine if chemistry data (e.g., conductivity) can help discriminate between streams and the stream gradients;
    e. use multivariate statistical techniques to determine if there are relations with IBI values, physical habitat values, chemistry, watershed characteristics or human disturbance factors.

2. Sample these same eleven sites again when funds are available (spring is the preferred season since upper creeks are sometimes dry in the fall) to determine whether stations

retain the same BMI communities and whether statistical differences decrease at individual stations with more data. If possible, involve citizen monitors from Prop 13 grant for Contra Costa County.

3. Analyze the remaining twelve BMI metrics at individual stations to determine whether there are statistically significant differences (though no substantial differences are expected).

4. Continue the SWAMP sampling state-wide, as it is an excellent tool for gaining consistently collected BMI data.

5. Determine the sampling strategy and statistical analysis requirements necessary to determine BMI population trends in typical California creeks.

6. Determine the usefulness of toxicity testing by reviewing the toxicity data in this report and that collected by SWAMP. If results are confounding, as is the case for Wildcat and San Leandro Creeks, consider the costs versus benefits of toxicity testing. In addition, analyze the variability of pathogens in creeks statewide to determine the reliability of these measurements.

7. Maximize training opportunities to assess water quality through the analysis of BMI communities.

Recommendations for Sampling Wildcat and San Leandro Creeks:

8. As for #1 above, continue to collect future data at these same eleven stations when funding is available. Combine any new data with the data from this study and analyze it statistically to determine if the variation between stations and between years is within acceptable limits.

9. For water and sediment sample collection and analysis, follow the SWAMP Quality Assurance Management Plan (Puckett 2002 or later draft) to assure that field and laboratory standards are met. For additional guidance see the "Electronic Template for SWAMP-Compatible Quality Assurance Project Plans" (Nichol and Reyes, 2004). Do not measure total or dissolved metals unless these field and laboratory standards are followed.

10. Encourage volunteers to participate in water quality and quantity sampling and in the physical habitat assessment of BMI communities.

11. Analyze data from the following data sets: (a) SWAMP's continuous water quality monitors for the two creeks in this study and compare it to other creeks (b) A. Riley's historical bioassessment data from Wildcat Creek (c) water quality data (turbidity, dissolved oxygen, temperature, pH, and conductivity) has been collected monthly since 1995 at several stations on the San Leandro Creek; this unique data should be graphed and analyzed to look at trends over the years in the San Leandro Creek and to assess the

usefulness of data collected by volunteers; (d) hydrographs from the four years of BMI data from the two creeks sampled here.

12. For long term watershed monitoring add two estuarine stations to the freshwater stations, e.g., Martin Luther King restoration site on San Leandro Creek, and Wildcat Marsh on Wildcat Creek. Also assess the biological communities in San Leandro and San Pablo Bays. This would require additional funds but would allow comparisons of the health of estuarine versus freshwater communities using indicators appropriate to each community type in areas known to be impacted by pesticides, metals, and PCBs.

Management Recommendations:

13. Consider flushing San Leandro Creek stations with dam/reservoir water to clear out the stagnant algal mats. It may be better to have more water released from the reservoirs in pulses to improve stagnant conditions than to continue the small but steady trickle throughout the dry season. Currently the Upper San Leandro Dam releases water to San Leandro Creek which flows into the Chabot Reservoir which again enters San Leandro Creek when the Chabot Dam is opened (East Bay Watershed Sanitary Survey 2000). The feasibility of removing any elevated pesticides or PAHs in the three lower San Leandro Creek should be investigated before any flushing is attempted.

14. Consider fencing the pool and creek at Lone Oak in Tilden Park (WC5) to protect steelhead habitat and improve the BMI community.

REFERENCES:

Breaux, A., M. Born, S. Cochrane, L. Suer, R. Brewer 2003 [draft]. A watershed assessment of two east bay creeks, San Francisco Bay, CA.: third year report on bioassesment, water, and sediment quality sampled in April 2002 [Draft, June 2003]. Reports also available on 2000 (first year) and 2001 (second year) sampling. San Francisco Bay Regional Water Quality Control Board, Oakland, CA.

California Department of Fish and Game. 2003. California Stream Bioassessment Procedure. Aquatic Bioassessment Laboratory. www.dfg.ca.gov/cabw/cabwhome.html.

California Department of Fish and Game. 1999. Bioassessment Fact Sheet and Report on "An Index of Biological Integrity for First to Third Order Russian River Tributary Streams". Rancho Cordova, Ca. www.dfg.ca.gov/cabw/cabwhome.html.

Collins, J., L. Collins, R. Grossinger, L. McKee, and A. Riley. 2000. (Draft version). Wildcat Creek Landscape Change. A history of sediment-water relations and the role of people in the Wildcat Watershed, Contra Costa County, CA.

Cover, M. 2004. San Francisco Bay Reegional Water Quality Control Board’s Surface Water Ambient Monitoring Program Two-Year Aquatic Bioassessment Interpretive Report (Draft).

East Bay Municipal Utilities District and C.R. James and Associates. 2000. *East Bay Watershed Sanitary Survey Update 2000.* Oakland, CA.

Harrington, J. and M. Born. 2000. *Measuring the Health of California Streams and Rivers.* Sustainable Land Stewardship Institute, Sacramento, CA

Leidy, R. 1999. Bay Area EcoAtlas, San Francisco Estuary Institute, Richmond, CA., [www.sfei.org].

Moore, S. 2000. Memo on the Status of Implementation of the Regional Monitoring and Assessment Strategy (RMAS). 9/18/00, 8 pages + tables and figures. San Francisco Bay Regional Water Quality Control Board. Oakland, CA.

Nichol, G. and E. Reyes. 2004. Electronic Template for SWAMP-Compatible Quality Assurance Project Plans. State Water Resources Control Board. Sacramento, CA.

Ode, P. A. Rehn, and J. May. 2005. A quantitative tool for assessing the integrity of Southern Coastal California streams. *Journal of Environmental Management*, Vol. 35, No. 1, pp 1-13.

Pacific EcoRisk. 2001. Chronic toxicity evaluation of four ambient creek water samples (Samples collected May 11 and 14, 2001); and 2002. Chronic toxicity evaluation of ambient water samples (Samples collected April 29 & 30, 2002). Prepared for Sequoia Analytical. Martinez, CA.

Paige, C. J. Gibbs, and S. Droege. 2000(?). Power analysis of wildlife monitoring programs: exploring trade-offs between survey design variables and sample size requirements. USGS Patuxent Wildlife Research Center. http://www.pwrc.usgs.gov/resshow/droege3rs/salpower.htm

Puckett, M. 2002. Quality Assurance Management Plan for the State of California's Surface Water Ambient Monitoring Program ("SWAMP"). [draft] CA Department of Fish and Game, Monterey, CA. Prepared for the State Water Resources Control Board, Sacramento, CA

Rosenberg, D. and V. Resh. 1996. Use of Aquatic Insects in Biomonitoring. Pages 87-97 in *An Introduction to the Aquatic Insects of North America,* eds., R. Merritt and K. Cummins. Kendall/Hunt Publishing Co., Dubuque, Iowa.

Sheppard, J. A. Theile, and P. Alexander. 2001. Impacts of Disturbance on Habitat and Biota of East Bay Creeks. Report by Fisheries Program, East Bay Regional Parks District, Oakland, CA.

Urban Creeks Council. 1996. Wildcat Creek Monitoring Project Final Report. Berkeley, CA.

URS Greiner Woodward Clyde (URSGWC). 1999. Volunteer Monitoring in San Leandro Creek, Technical Summary Report 1995-1998. Report prepared for Alameda Countywide Clean Water Program, Hayward, CA.