Watershed Data Assessment Framework
Final Draft

Version 1

Prepared For:

San Diego Storm Water Copermittees

June 1, 2004

MEC ANALYTICAL SYSTEMS

WESTON SOLUTIONS

Watershed Data Assessment Framework

Prepared For:

San Diego Storm Water Copermittees

Prepared By:

MEC Analytical Systems, Inc. / Weston Solutions, Inc.
2433 Impala Drive
Carlsbad, California 92008

June 2004

TABLE OF CONTENTS

1.0 INTRODUCTION 1

2.0 DEFINING CORE MANAGEMENT QUESTIONS 3

3.0 DATA QUALITY REQUIREMENTS 5
- 3.1 Chemistry Data 5
  - 3.1.1 Third Party Chemistry Data 9
- 3.2 Biological Data 10
  - 3.2.1 Third Party Benthic Data 11
- 3.3 Toxicity Data 12
  - 3.3.1 Third Party Toxicity Data 12
- 3.4 Ambient Bay and Lagoon Monitoring Data 12
  - 3.4.1 Sediment Chemistry Assessment 14
  - 3.4.2 Sediment Toxicity Assessment 14
  - 3.4.3 Benthic Infauna Assessment 15
  - 3.4.4 Third Party Data 15

4.0 WATERSHED DATA ASSESSMENT 15
- 4.1 Chemistry Data 18
  - 4.1.1 How to Determine Frequency of Occurrence for COCs 20
- 4.2 Biological Data 22
- 4.3 Toxicity Data 23
  - 4.3.1 Determining Persistent Toxicity for the Triad Decision Matrix 24
  - 4.3.2 Statistical Methods for Toxicity and COC 24
- 4.4 How to Use the Triad Decision Matrix 24
- 4.5 Long-Term Trend Data Analysis 28
- 4.6 ABLM Data Assessment 28
  - 4.6.1 Chemistry Data 28
  - 4.6.2 Biological Data 30
  - 4.6.3 Toxicity Data 30
  - 4.6.4 Data Integration 30
- 4.7 Impacts on Beneficial Uses 32
- 4.8 Cross-Watershed Comparisons 32
  - 4.8.1 Statistical Methods 33
    - 4.8.1.1 Scatterplots 33
    - 4.8.1.2 Analysis of Variance 33
    - 4.8.1.3 Cluster analysis 33
  - 4.8.2 Computer Modeling 34
    - 4.8.2.1 Event Mean Concentration Model 34
    - 4.8.2.2 Dynamic Storm Water Modeling 35
- 4.9 Core Management Questions – Gaining Understanding of Watershed Conditions 40

5.0 DATA ASSESSMENT TOOLS 41
- 5.1 Quantitative – Statistical Tools 41
  - 5.1.1 Data transformations 41
  - 5.1.2 T-Test 41
  - 5.1.3 ANOVA with Multiple Range Test 42
  - 5.1.4 Regression Analysis (and Outlier Test) 43
  - 5.1.5 Multiple Regression Analysis 44

5.1.6 Confidence Interval Test ... 45
5.1.7 Predictive Interval Test ... 46
5.1.8 Power Analysis ... 46
5.1.9 Outlier test ... 47
5.2 Semi-Quantitative ... 47
5.2.1 Cluster Analysis ... 47
5.2.2 Threshold comparisons ... 49
5.2.3 Ranking ... 51
5.2.4 Maps and graphics ... 51
5.2.5 Weight-of-evidence ... 53
5.3 Qualitative ... 53
5.3.1 Visual Observations ... 53
5.3.2 Photographic Documentation ... 53
5.3.3 Third Party Monitoring ... 53

6.0 REFERENCES ... 53

APPENDICES
A – Surface Water Ambient Monitoring Program (SWAMP) Quality Assurance and Quality Control
B – Summary of Special Sampling or Handling Requirements
C – QA/QC Data Evaluation Checklist for Chemical Analysis
D – QA/QC Data Evaluation Checklist for Benthic Analysis

ACRONYMS AND ABBREVIATIONS

ABLM	Ambient Bay and Lagoon Monitoring Program
AB411	California Assembly Bill 411
ACSW	Assessing California Surface Waters
AHC	Agua Hedionda Creek
ANOVA	analysis of variance
BL	Batiquitos Lagoon
BMI	benthic macroinvertebrates
BMP	best management practices
BOD	biochemical oxygen demand
BPJ	best professional judgment
CC	Chollas Creek
CDFG	California Department of Fish and Game
COC	constituents of concern
COD	chemical oxygen demand
CSBP	California Stream Bioassessment Procedure
CWA	U.S. Clean Water Act
D	day
DO	dissolved oxygen
DWS	dry weather station
EC	Escondido Creek
ELAP	California Environmental Accreditation Program
ELISA	enzyme-linked immunosorbant assay
ER-M	Effects Range-Median
ER-L	Effects Range-Low
EMC	event mean concentration
EPA	U.S. Environmental Protection Agency
GC/MS	gas chromatography/mass spectrometry
GIS	geographic information system
IBI	index of biotic integrity
JURMPs	Jurisdictional Urban Runoff Management Plans
L	Liter
LC25	concentration that causes 25 percent mortality
LC50	concentration that causes 50 percent mortality
M	month
MB	Mission Bay
MBAS	Methylene Blue Active Substances
MDL	method detection limit
MEC	MEC Analytical Systems
MLS	mass loading stations
MPN	most probable number
MS4	municipal separate storm sewers system
NELAP	National Environmental Laboratory Accreditation Program
NOAA	National Oceanic and Atmospheric Administration
NOEC	No Observed Effect Concentration
NPDES	National Pollutant Discharge Elimination System
NTU	nephelometric turbidity units
NURP	Nationwide Urban Runoff Program
OR	Otay River

PAH	polycyclic aromatic hydrocarbons
PC	Peñasquitos Creek
PCB	polychlorinated biphenyls
QA/QC	Quality Assurance/Quality Control
QAPP	Quality Assurance Project Plan
Rec-1	Water contact recreational beneficial use
Rec-2	Water non-contact recreational beneficial use
RSB	Rapid Stream Bioassessment
RWQCB	Regional Water Quality Control Board
SAP	Sampling and Analysis Plan
SANDAG	San Diego Association of Governments
SDC	San Dieguito Creek
SDR	San Diego River
SLR	San Luis Rey River
SMC	Stormwater Monitoring Coalition
SMR	Santa Margarita River
SR	Sweetwater River
SWRCB	State Water Resources Control Board
TC	Tecolote Creek
TDS	total dissolved solids
TIE	Toxicity Identification Evaluation
TJR	Tijuana River
TKN	Total Kjeldahl Nitrogen
TMDL	Total Maximum Daily Loads
TOC	total organic carbon
TRE	Toxicity Reduction Evaluation
TSS	total suspended solids
TU	toxic units
TUa	acute toxic units
TUc	chronic toxic units
TV	tolerance values
URMP	Urban Runoff Management Plans
USEPA	U.S. Environmental Protection Agency
USGS	U.S. Geological Survey
W	week
WURMPs	Watershed Urban Runoff Management Plans

TERMINOLOGY

The term Constituent is used to define an analyte or parameter measured in the monitoring program. It includes physical, chemical and microbiological measurements.

A Constituent of Concern (COC) is an analyte or parameter measured in the monitoring program that has exceeded a water quality objective or other action level. Action levels do not necessarily have to be established compliance levels or standards, but may include trigger levels set periodically by the Copermittees for program management. An example is the dry weather program action levels established in 2002 (See Section 3).

The frequency of occurrence of each COC has been defined as high, medium, or low using interim criteria defined by the Copermittees. Determining the frequency of occurrence of a COC was developed as a management tool in 2003 that is subject to improvements or modifications as more data and information is generated by the monitoring program. The interim criteria was established to consider or “flag” COCs that need to be watched and evaluated based on other available information to determine if they are or have a potential to become water quality problems. Watershed Copermittees should evaluate prioritized constituents of concern together with other available data from the watershed to determine if the COC(s) are providing an indication of a water quality problem.

Water quality problems are determined through the assessment of water quality data. One approach of identifying water quality problems is applying the interim criteria developed by the Copermittees which defines the problems by the frequency of occurrence of the COC as high, medium or low.

Prioritizing water quality problems is accomplished by taking into account the frequency of occurrence of COCs and other factors in the watershed, such as 303(d) listings, beneficial uses, existing projects and studies, and supplemental information.

1.0 INTRODUCTION

This Watershed Assessment Framework was developed for the San Diego Municipal Storm Water Copermittees to provide guidance for watershed data assessments so that analysis and management planning is more likely to be applied in a uniform and consistent manner across watersheds in the county. One of the goals of the document is to provide guidance as to how outside data can be incorporated into future watershed data assessments. The framework is designed to be adaptable to different circumstances in each watershed by describing the assessment strategy and providing an overview of the statistical tools available to conduct watershed data analysis. This Watershed Assessment Framework has been developed to be consistent with other San Diego region and statewide guidance documents.

This framework is anticipated to evolve with the scientific developments in watershed water quality assessment and in response to changing regulatory requirements. This document includes the following main sections: Section 2 describes the importance of defining core management questions; Section 3 describes data quality standards for watershed data collection and testing; Section 4 addresses watershed data assessment methodologies; and Section 5 lists the most commonly used statistical tools for watershed data assessment.

This guidance document may be used by individual watershed management groups to develop core management questions specific to the goals and needs of the stakeholders in the watershed. The data and information assessment guidance presented in Section 4 – Watershed Data Assessment will assist Copermittees in developing their action plans and watershed activities on a continuous basis. The assessment process has been summarized in a easy to follow flowchart that starts with data selection and ends with action planning.

Resource managers may find the Data to Wisdom models, developed by various systems theorists including Ackoff, R.L. (1989), provide a good foundation for the long-term approach to watershed assessments. The Data-Information-Knowledge-Wisdom (DIKW) hierarchy describes how the human mind is set in building blocks to process input and formulate conclusions. The DIKW model can be expanded to include an intermediate “understanding” step between knowledge and wisdom. The definitions of the terms data, information, knowledge, understanding and wisdom (DIKUW) used by Ackoff are relevant to watershed assessments in the following way:

Data are symbols that exist by themselves. Data is on a time scale and by itself cannot be classified as valid or not. Data is the first cornerstone of the DIKUW model presented in Figure 1-1.

Information is obtained by giving data meaning. In watershed assessment this may be achieved by applying several of the tools described in this document.

- Validation of the data through quality control checks
- Comparing data to water quality objectives
- Applying the various models or statistical tools to form relationships

Information is said to answer questions that start with “who”, “what”, “where”, and “when”. The San Diego watershed assessment framework is based on five core management questions presented in this document that when answered can lead to information about watershed conditions. Acquiring information is the second cornerstone of the DIKUW model as shown in Figure 1-1.

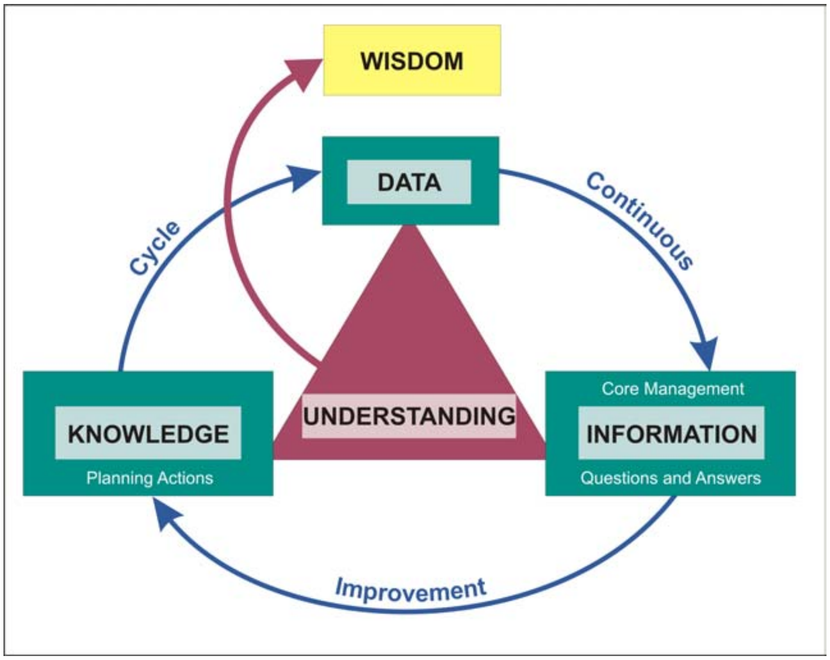


Figure 1-1. Five components of the data to wisdom model for watershed assessment

Knowledge is gained from combining data and information to provide answers to "how" type questions. In the context of this framework, one sample question is *How do we control or reduce bacteria sources to the receiving water body?* Gaining knowledge of the conditions in the watershed through data and information results in the ability to develop management actions to maintain (reduce or minimize) or improve conditions in the watershed. Knowledge is the third and last cornerstone of the DIKUW model and defines the continuous improvement cycle (Figure 1-1).

Understanding is described by Gene Bellinger et al. (2004) as an interpolative and probabilistic process that is cognitive and analytical. Understanding is gained by building knowledge on top of knowledge and synthesizing new knowledge or at least new information (new who, what, where and when questions). Therefore, creating a continuous improvement cycle ultimately leads to enhanced evaluation of the conditions in the watershed. Understanding in Figure 1-1 is at the center of the data-information-knowledge triangle and the continuous improvement circle.

Wisdom in contrast to understanding is noted by Bellinger as being extrapolative and non-deterministic which leads to a non-probabilistic process. The difficulty with obtaining wisdom is that it goes beyond the factual understanding and adds the dimension of values and ethics. At this level one makes a value judgment as to whether the end result of the effort is a benefit to the community and to what extent previously held questions have been answered. In other words, the continuous improvement cycle has come to an end (at least for the issue at hand) and our understanding cannot be expanded by more data, information, and knowledge. Wisdom is represented in Figure 1-1 as the end results of the effort.

This model is applied in this document to guide the systematic evaluation of data using scientific and statistically valid methods that will generate information to answer core management questions about the conditions in the watershed and be the foundation of planning actions to resolve issues of concern. The cycle of data-information-knowledge leads to an understanding of whether the actions taken have given the expected results. New understanding of how the watershed is connected and the cause and effect of actions promotes the generation of new questions and a new continuous cycle to the next level of understanding of the changing conditions in the watershed. Wisdom is achieved from interlacing values and ethics with our understanding of the conditions in the watershed and placing a judgment on the significance (i.e. social, economical, and political) of the achievements.

2.0 DEFINING CORE MANAGEMENT QUESTIONS

Core management questions utilize data collected from water quality assessments to determine the effectiveness of a monitoring program. The San Diego Municipal Stormwater Copermitees developed core questions to assess the effectiveness of the Jurisdictional Urban Runoff Management Plans (JURMPs) and Watershed Urban Runoff Management Plans (WURMPs). As noted in “A Framework for Assessing the Effectiveness of Jurisdictional Urban Runoff Management Programs” (2003) they recognized water quality as one component of the overall framework in this ‘effectiveness assessment’ by stating, “water quality assessments serve a number of important purposes, including the following: discharge and receiving waters characterization (baseline and trends); identification and characterization of water quality problems and the constituents/stressors of concern causing them; and, ongoing feedback for incremental program refinement and revision of program priorities”.

At this time, the Copermittees’ Assessing Effectiveness document lists these core questions:

- What are the high priority water quality problems and where do they occur?
- What is causing them (constituents/stressors of concern)?

Long-term questions may include:

- What are the changes to discharge and receiving water quality?

The answers to these and similar questions regarding water quality are combined with program planning questions to guide management decisions at the watershed and jurisdictional level. The Copermittees’ Assessing Effectiveness document is consistent with the DIKUW model.

The State of California has developed a draft document, “A Process for Addressing Impaired Waters in California” (2003; also known as the California Impaired Waters Guidance), to support the development of plans for addressing impaired waters in the State. This process will focus on TMDL issues and it requires that project needs be defined by a scientific approach that is determined by asking a series of technical questions that must be satisfied to understand the water quality problems and/or implement programs to improve water quality.

The basic technical questions discussed in that document include:

- ♦ Why is the water listed?
- ♦ When and under what conditions does the impairment occur?
- ♦ Does examination of the data show any obvious relationships?
- ♦ What are the data quality considerations?
- ♦ What are the potential sources contributing to the impairment?
- ♦ How is the loading delivered?
- ♦ What types of management measures might be considered to restore the impaired water?

To augment the assessment of water quality, existing monitoring programs should be identified that collect data (and perform analysis) that are relevant to answering core management questions. In general, more specific data collection provides for better characterization of problems and understanding of spatial and temporal scope. Data are not always collected with a specific question in mind and may not meet the necessary quality control criteria. The more specific the questions are, the more specialized the data collection and analysis will be and therefore, should be planned and managed accordingly. As the program moves forward data management will likely entail added resources, coordination and planning at a regional level.

The initial steps towards standards for data collection and management have been taken at several levels, including the development of the "Data and Information Management Plan" (San Diego Stormwater Copermittees and Project Clean Water, 2002) which outlines the major data management issues and data management frameworks for databases.

The watershed data assessment strategy presented in this document represents the current approach in San Diego County. These assessments are part of the WURMP and closely resemble the "Model Storm Water Monitoring Program for Municipal Separate Storm Sewer Systems in Southern California" developed by the Stormwater Monitoring Coalition's (SMC) Model Monitoring Technical Committee (2004). The SMC Model presents a flowchart that guides management questions that drive the design of the monitoring program based on the answers obtained. The core management questions defining the current monitoring program for San Diego Stormwater Copermittees can be summarized as follows:

1. What are the water quality conditions in the watershed?
2. Are water quality conditions in the watershed getting better or worse?
3. Are beneficial uses being impacted based on existing external information?
4. What is the relative contribution of urban runoff to the conditions in the watershed?
5. What are the sources to urban runoff that contribute to water quality conditions?

These five core management questions are the backbone of the watershed assessment framework and correspond to the data to information phase of the data to wisdom hierarchy discussed in the previous section. Guidance to evaluate data is provided in Section 3 – Data Quality Requirements and the five core management questions that lead to the actual assessment are discussed in detail in Section 4 – Watershed Data Assessment.

3.0 DATA QUALITY REQUIREMENTS

Determining what types of data are available for analysis is one of the first steps in watershed assessment aimed at answering project questions. In general, three types of data can be defined (Table 3-1):

Table 3-1. Data Type Descriptions.

Data Type	Description	Example
Type 1 or T1	Quantitative data	Wet weather MLS data
Type 2 or T2	Semi-quantitative data	Dry weather field test kit data
Type 3 or T3	Qualitative Data	Visual observations/physical descriptions

Quantitative data (T1) are best used for statistical analysis and are characterized by having been analyzed using consistent or comparable analytical methods, consistent reporting limits and collected with good temporal or spatial representation (as needed in response to project questions). These datasets can generally be subjected to several types of statistical analysis or tests and yield conclusive results with defined confidence.

Semi-quantitative (T2) data are defined as a dataset that may not have the desired quality (i.e. different reporting limits) or there are not enough data points to perform statistical analysis to answer the project question. In most cases, these datasets can provide partial information and define data gaps that need to be filled. In some cases, semi-quantitative data analysis reveals other questions that may need to be posed.

Qualitative data (T3) are generally used to corroborate other datasets (quantitative and semi-quantitative). Using qualitative data by itself to answer project questions is more likely to lead to erroneous solutions or unsupportable management decisions. Review and assessment of qualitative data are a valuable first step to defining questions or problems for which quantitative data should be sought to confirm the hypothesis or assumption.

Data quality should be assessed in a tiered approach that starts with selection of quantitative data (T1 dataset) to perform statistical analysis, semi-quantitative data (T2 dataset) assessment to corroborate the finding of the T1 dataset and/or identify data gaps; and lastly reviewing the qualitative data (T3 dataset) to define any additional questions or find supporting evidence.

The following sections outline the minimum requirements for chemistry, benthic and toxicity data so that they will satisfy Type 1 level quality (quantitative).

3.1 Chemistry Data

The term “chemistry data” is used here as a general term that includes general physical measurements (i.e., pH, conductivity), bacteriological testing (i.e., total coliform), wet chemistry (i.e., ammonia, turbidity), pesticides, and total and dissolved metals. Individual analytes or measurements are referred to as constituents.

Minimum requirements for quantitative analysis of chemistry data are:

- A Quality Assurance Project Plan, Sampling and Analysis Plan, or
- Comparable documentation (analytical methods, QA/QC procedures) and
- Accredited laboratory.

Ideally, sampling should be conducted under a Quality Assurance Project Plan (QAPP) or Sampling and Analysis Plan (SAP) designed to define the sampling design and analytical methods to be used. It also provides consistency in the quality of data and identifies the number of samples that may be needed to conduct statistical analysis of the analytical data. The QAPP can also define the appropriate temporal and spatial sampling coverage based on the question(s) one seeks to answer. Detailed information about the QAPP can be found on the Surface Water Ambient Monitoring Program's (SWAMP) website (CSWRCB 2002). Sample collection and documentation prepared by trained field personnel is essential to the acquisition of representative samples.

Supporting documentation such as field logs and chain-of-custody forms may need to be reviewed to verify the sampling location, visual observations, time, date, weather conditions, etc. These data may provide information that supports the decision to incorporate third party data into an existing dataset.

Data from sampling events collected to characterize urban runoff or wet weather discharges should meet the reporting limits and holding times for the methods outlined in Table 3-2 to qualify for consideration with data collected by the Copermittee monitoring program. It is preferable that data collected be reported in the units listed in Table 3-2. Information documented at the time the sample is collected should include the sample type (grab, 24 hour composite, etc.) and preservation technique (entered in the chain-of-custody) to document that the sample complies with customary protocols and continues to be representative of actual conditions in the environment. Preservation requirements as listed in Table 1060:I (Cesceri et al 1998) are provided in Appendix B.

Data reported using other methods may not be considered with the same "weight" as data that satisfy the requirements in Table 3-2 and in Standard Methods. To incorporate analytical results obtained using other methods it is important to first verify if they are comparable. This can be accomplished by analyzing split samples using both methods or carefully reviewing the quality control reports submitted with the original laboratory data reports to determine if the data should be merged. The two methods may yield different results for split or comparable samples and best professional judgment must be used to determine if the difference is significant or not. Depending on the detection limits and the analyte it may be acceptable to have differences up to 50% between methods, yet in some other cases only 5% may be acceptable. The exercise of determining if the methods are comparable can prove to be economical if merger of the two datasets is deemed valuable based on the number of data points, temporal or spatial coverage, etc.

Environmental samples should be analyzed by a laboratory with accreditation for the constituent(s) of interest. In California, independent accreditation is issued under the Environmental Laboratory Accreditation Program (ELAP) or the National Environmental Laboratory Accreditation Program (NELAP) by the California Department of Health Services.

Table 3-2. Recommended Methods for Analytical Data.

Constituent	Volume Required	Recommended Method	Reporting Limit	Units
General Physical and Inorganic Non-Metals				
Total Dissolved Solids (TDS)	100 mL	SM 2540C	20	mg/L
Total Suspended Solids (TSS)	100 mL	SM2540D	40	mg/L
Turbidity	100 mL	SM 2130A-B	0.1	NTU
Total Hardness	150 mL	EPA 200.7	10	mg $CaCO_3$/L
pH	In field	EPA 150.1	0.1	S.U.
Specific Conductance	In field	SM 2510B	1.0	umhos/cm
Temperature	In field	SM 2550B	1.0	Deg C
Dissolved Phosphorus (Ortho-P)	250 mL	SM 4500B,E	0.05	mg/L
Total Phosphorus	250 mL	SM 4500B,E	0.10	mg/L
Nitrate and Nitrite	200 mL	SM4500NO2-NO3	0.1/0.05	mg/L
Total Kjeldahl Nitrogen (TKN)	500 mL	SM4500N C	0.5	mg/L
Ammonia as N	250 mL	SM 4500NH3B,C	0.1	mg/L
Biological Oxygen Demand, 5-day (BOD)	1000 mL	SM5210B	2	mg/L
Chemical Oxygen Demand (COD)	25 mL	EPA 410.4	25	mg/L
Dissolved Organic Carbon (DOC)	100 ml	EPA 415.1	1.0	mg/L
Total Organic Carbon (TOC)	100 ml	EPA 415.1	1.0	mg/L
Methylene Blue Active Substances (MBAS)	250 mL	SM 5540C	0.5	mg/L
Cyanide	500 mL	EPA 335.2	3	µg/L
Organics				
Oil and Grease (O&G)	500 mL	EPA 413.2	1.0	mg/L
Total Petroleum Hydrocarbon	125-1000ml	EPA 8015B	50	µg /L
OP Pesticides (Diazinon, Chlorpyrifos, Malathion)	1 L	ELISA/8141/GCMS	0.05	µg/L
OC Pesticides	1 L	EPA 8081	0.05	µg/L
Carbamate Pesticides	1 L	EPA 8321	0.05	µg/L
Base/Neutral and Acid Extractables	1 L	EPA 625SIM	0.05	µg/L
PAHs	1 L	EPA 8310	0.05	µg/L
Purgeables	1 L	EPA 624	0.05	µg/L
Purgeable Halocarbons	1 L	EPA 601	0.05	µg/L
Purgeables Aromatics	1 L	EPA 602	0.05	µg/L
Metals, Dissolved and/or Total				
Antimony (Sb)	75 mL	EPA 6020	0.006	mg/L
Arsenic (As)	75 mL	EPA 6020	0.002	mg/L
Cadmium (Cd)	75 mL	EPA 6020	0.001	mg/L
Chromium (Cr)	75 mL	EPA 6020	0.005	mg/L
Copper (Cu)	75 mL	EPA 6020	0.010	mg/L
Iron (Fe)	75 mL	EPA 200.9/SM 3500-FeB	0.025	mg/L
Lead (Pb)	75 mL	EPA 6020	0.005	mg/L
Mercury (Hg)	75 mL	EPA 1631	0.5	ng/L
Nickel (Ni)	75 mL	EPA 6020	0.005	mg/L
Selenium (Se)	75 mL	EPA 6020	0.005	mg/L
Zinc (Zn)	75 mL	EPA 6020	0.02	mg/L
Bacteriological*				
Total Coliform	200 mL	SM 9221B	20-1.6E6	MPN/100 mL
Fecal Coliform	200 mL	SM 9221E	20-1.6E6	MPN/100 mL
Enterococcus	200 mL	SM 9230	20-1.6E6	MPN/100 mL

* To full dilution series, no > value should be reported or 1.6 million

Water samples collected to represent urban runoff (dry weather), wet weather discharges or receiving water quality can provide Copermittees with information for various levels of interpretation and can answer either general or specific questions depending on the sampling plan or design. The nature of the questions can include temporal or spatial coverage considerations, as described below. Samples may be collected using grab or composite methods that are weighted (flow, time, etc.) based on the anticipated statistical needs and resource considerations (labor hours, analytical costs, etc.).

The data can then be used in the majority of situations to answer questions if it is collected following the recommended methods outlined in Table 3-2. It is also important that the questions to be answered be drafted clearly and translated into a specific sampling design. These datasets are generally quantitative in nature and can be subjected to the majority of the statistical tests and tools described in Section 5.

The reporting limit for the methods described in Table 3-2 may in some instances not be achievable because of matrix interferences, instrument sensitivity, etc. This can result in a dataset with non-detectable amounts reported with varying reporting limits. Datasets with varying reporting limits may also be encountered when historical datasets (over 5 years old) are reviewed and when varying methods or laboratories perform the sample analysis. The following paragraphs describe the appropriate methods to manage datasets with non-detect results.

Dataset With Various Reporting Limits

If a dataset contains multiple reporting limits such as 0.5, 0.1 and 0.01 mg/L, the highest non-detectable result is reported as <0.5 mg/L. One-half of this reporting limit (0.25 mg/L) should be substituted for all non-detect values as well as detected values (from the datasets with lower reporting limits) less than 0.5 mg/L. In this case, information is lost from the data known to be below 0.5 mg/L (actual values reported at or above 0.1 and 0.01, but below 0.5 mg/L). This approach is used when equal numbers of data points have the various reporting limits and the total dataset is small. In some cases, these datasets with varying reporting limits are not conducive to extensive statistical calculations. Best Professional Judgment should be used when applying this technique.

Dataset With a Few Data Points With Higher Reporting Limits

If a dataset contains multiple reporting limits such as 0.5 and 0.2 mg/L, and the number of data points reported as non-detect (<0.5 mg/L) in the dataset with a higher reporting limit are few compared to the those values reported as non-detect (<0.2 mg/L) in the dataset with a lower reporting limit, a decision can be made to exclude all the data reported with the high reporting limit from the assessment. This may result in the loss of some data points, but more information is gained with less uncertainty and more conservative estimates. The non-detect values in the dataset with the lower report limit are set to one-half the reporting limit (0.1 mg/L, in this example) prior to performing statistical analyses.

Dataset With Uniform Reporting Limit

Non-detect results from datasets with only one reporting limit, for example, 0.1 mg/L, should be substituted with one-half the reporting limit (0.05 mg/L in this example) prior to conducting the data analyses.

Data With Varying Units of Measurement

If data are reported using different units (mg/L and µg/L or Deg C and Deg F) the data can be converted to a uniform unit while preserving the same number of significant units prior to conducting statistical analysis.

3.1.1 Third Party Chemistry Data

Third party data should conform, as closely as possible, to the same quality control standards, sampling and testing methods, and documentation that data collected under the Copermittee Regional Monitoring Program in order to be assessed quantitatively. Otherwise, data collected by third parties (universities, citizen groups, etc.) may only be used semi-quantitatively or qualitatively.

General quality assurance and quality control guidelines are intended to provide assistance in third party data evaluation. Specific information can be found in several documents listed in Section 6.

To adequately meet data quality acceptability and facilitate data analysis (including statistical analysis) the quality assurance and quality control (QA/QC) checks should be included in the QAPP or SAP, and should be applied during all steps of sample analysis: sample collection and handling, analysis, data management and reporting. In general, the three components of a QA/QC program are:

- Accuracy
- Precision
- Contamination

Accuracy is determined from measuring the percent recovery of a known quality of a constituent that is added as a matrix spike or laboratory control spike to a sample. The amount detected is compared to the acceptability limits for the test or method. Precision is assessed using duplicate samples and by comparing the relative percent difference (RPD) to the acceptability limits.

Excellent data quality is achieved during the sampling phase by including:

- Blank Samples
- Duplicate Samples
- Matrix Spikes and Matrix Spike Duplicate Samples
- Quality Control Sample Schedule

Field and laboratory blank samples are analyzed to determine sources of potential contamination introduced during the sampling and analysis procedures. Therefore, blanks are used at each step to identify contamination. Trip blanks, field blanks, equipment blanks and method blanks should be included to various degrees based on the sampling design and budget. Duplicate samples measure the analytical precision or reproducibility of the data which is assessed by analyzing two identical samples. Duplicate samples can include field duplicates to represent any sampling technique bias and/or laboratory duplicates that represent analytical method bias. The precision is calculated as the relative percent difference (RPD) and is method dependent. The RPD is described for each method in the quality assurance section and can vary from as low as 15% for Total Organic Carbon (EPA 415.1) to as much as 50% for some organic compounds (EPA 625). To determine the accuracy and precision of the analytical method in the particular sample a matrix spike (MS) and matrix duplicate spike (MDS) are used. These measurements give an indication of any potential interference (solids, salts, etc.) produced by the sample itself and are method and sample specific. They also provide an indication of any other analyst or instrument errors that may exist. A quality control schedule is project specific and outlines the minimum number of quality control samples that will be collected and analyzed for the project. (EPA Guidance, 1995). An adequate quality control sample schedule should be a major consideration of the project's budget.

Quality Assurance and Quality Control Data Evaluation
Data reported by the analytical laboratory must be thoroughly reviewed and assessed to determine if it meets the project's data quality objectives described in the QAPP or SAP prior to incorporating into a database or being used (for water quality assessments or statistical analysis). The data evaluation can be performed in two steps:

- Initial Data Quality Screening
- Data Quality Evaluation

The initial data quality screening is a basic review of the laboratory data report against the information contained in the chain-of-custody form and the QAPP or SAP. The three components of the initial data quality screening are included in the QA/QC Data Evaluation Checklist (Appendix C). Any anomalies identified during the initial data quality screening should be reported to the laboratory for clarification and any changes must be reported in writing through an amended laboratory analysis report.

The data quality evaluation is an in-depth, structured process with systematic checks to ensure that the data reported is representative of the samples. This evaluation is intended to identify sources of contamination in the sampling and analytical processes, any deficiencies in the laboratory analyses, and errors in the data reporting. The main components of the data quality evaluation are shown in the QA/QC Data Evaluation Checklist (Appendix C). Any discrepancies found during the data quality evaluation may lead to data that is rejected (labeled with "R").

3.2 Bioassessment Data

The sampling protocol of the Rapid Stream Bioassessment Program includes the collection of stream benthic macroinvertebrates and also assesses the quality and condition of the physical habitat following the protocols of the California Stream Bioassessment Procedure (CSBP; Harrington 1999). Utilizing species specific tolerance values and community species composition, numerical biometric indices are calculated, allowing for comparison of relative habitat health among streams in a region. Over time, this information is used to identify ecological trends and aid analyses of the appropriateness of water quality management programs (Yoder and Rankin 1998).

The Copermittee's Rapid Stream Bioassessment Program's primary goal is to monitor two separate stream reaches (monitor reaches) in each of the ten watershed management areas that have stormwater (or wet weather) mass loading stations. The assessments are typically conducted in May and October. Invertebrates reside in streams for periods ranging from a month to several years, and have varying sensitivities to the multiple stressors associated with urban runoff. By assessing the invertebrate community structure of a stream, a cumulative measure of stream habitat health and ecological response is obtained. Of the two monitoring reaches, one is typically located as far downstream in the watershed as practicable, and the other is located farther upstream in the watershed, but where it is still affected to some degree by urban development. Where possible, sites are located in the same stream reach that California Department of Fish and Game (CDFG) has previously sampled. Ongoing reconnaissance of the streams, with the goal of finding riffles with the highest quality in-stream habitats, has resulted in re-location of some of the monitoring reaches.

Reference sites have been designated by CDFG and the Regional Water Quality Control Board (RWQCB) based on upstream land use characteristics as determined by Geographic Information Systems (GIS) datasets. When selecting reference monitoring sites for comparison with urban affected sites, statistical analyses by CDFG has indicated that in southern California, elevation is not a significant factor affecting benthic community quality (Ode et al. 2002).

The Rapid Stream Bioassessment data collected by the Copemittees includes a taxonomic listing of all benthic macroinvertebrates identified in the surveys, and calculation of the biological metrics listed in the CSBP. All organisms are identified to the standard taxonomic level described in the CAMLnet List of Californian Macroinvertebrate Taxa and Standard Taxonomic Effort, using standard taxonomic keys. Quality assurance of sample sorting is performed on a minimum of 10 percent of the samples to ensure at least a 90% removal rate of organisms. Taxonomic quality assurance is performed on 10% of the samples by taxonomists at the CDFG Aquatic Bioassay Laboratory in Rancho Cordova, CA.

Additionally, the Index of Biotic Integrity (IBI) for all monitoring reaches is calculated, following the most recent protocols developed by the CDFG Aquatic Bioassessment Laboratory for coastal southern California (Ode, Rehn, and May, unpublished data). In 2003, the IBI replaced the Benthic Macroinvertebrate Index (BMI) Ranking Score used to analyze the data and is a significant improvement because it gives an absolute value to the benthic community quality based on the range of reference conditions in the region. The IBI score is calculated using a formula with raw data (taxa or species listing) as the input. Any future revisions of the formula will necessitate that the raw data be available to perform the recalculation. Therefore, it is important to preserve the original data sets.

For each monitoring reach sampled, the physical habitat of the stream and its adjacent banks are assessed using U.S. EPA Rapid Bioassessment Protocols. Habitat quality parameters are assessed to provide a record of the overall physical condition of the reach. Parameters such as substrate complexity, channel alteration, frequency of riffles, width of riparian zones, and vegetative cover help to provide a more comprehensive understanding of the condition of the stream. Additionally, specific characteristics of the sampled riffles are recorded, including riffle length, depth, gradient, velocity, and substrate composition.

Water quality measurements are taken at each of the monitoring sites using an environmental monitoring instrument such as the YSI model 6600. Measurements include water temperature, specific conductance, pH, dissolved oxygen, and chlorophyll. *In situ* relative chlorophyll measurements were recently added to the water quality assessment (May 2003) to provide information on phytoplankton productivity. Stream flow velocity is measured with a portable flow meter such as the Marsh-McBirney Model 2000 instrument, or it may be visually estimated.

The final IBI score ranges give a relative rating of Very Good, Good, Fair, Poor or Very Poor to each reach surveyed.

3.2.1 Third Party Benthic Data

Third party benthic data should conform to the quality control standards, sampling and testing methods, and documentation described in the CSBP in order to be incorporated into a watershed assessment in a quantitative manner. This enables IBI scores to be developed and incorporated into the existing Copermittee dataset. Otherwise, data collected by third parties (universities, citizen groups, etc.) may only be used semi-quantitatively or qualitatively. The QA/QC Checklist provided in Appendix D summarizes the basic requirements of the data and may assist in evaluating third party data sets.

3.3 Toxicity Data

The Copermittee Regional Monitoring Program includes toxicity monitoring at the mass loading stations (MLS) during the wet weather period. Samples collected are evaluated for chronic or acute toxicity using the test species *Ceriodaphnia dubia* a freshwater cladoceran, *Hyalella azteca* a freshwater amphipod, and *Selenastrum capricornutum* a freshwater algae as noted in Table 3-3. As part of the Ambient Bay and Lagoon Monitoring Program, estuary sediments are tested for toxicity using *Eohaustorius estuaries* (refer to Section 3.4.2 for details).

Table 3-3. Species and Methods for Toxicity.

Constituent	Volume Required	Method	Reporting Limit	Units
Toxicity	20 L			
Ceriodaphnia dubia. 7-day chronic test		EPA-821-R-02-013	100%	NOEC
Ceriodaphnia dubia. 96-hour acute test (survival and reproduction)		EPA-821-R-02-012		NOEC
Hyalella azteca. 96-hour acute survival test		EPA-821-R-02-012[a]	100%	NOEC
Selenastrum capricornutum. 7-day chronic test		EPA-821-R-02-013	100%	NOEC
Sediment Toxicity				
Eohaustorius estuarius		EPA-823-B-98-004		NOEC

(a) Modified method is used

Data from other sources must at a minimum follow the EPA methods listed in Table 3-3 and be performed by a qualified laboratory accredited under the Environmental Laboratory Accreditation Program (ELAP) or the National Environmental Laboratory Accreditation Program (NELAP) by the California Department of Health Services.

3.3.1 Third Party Toxicity Data

Toxicity data that is reported using other methods or performed by laboratories accredited in other states or without accreditation should be evaluated thoroughly to determine if proper laboratory practices and protocols are followed, analysts are trained and follow written protocols/methods, and that a written quality assurance manual is available and implemented. All toxicity results should be independently verified for quality control by a professional aquatic toxicologist prior to incorporating into the watershed assessment. Seeking expert advice is highly recommended for toxicity data. Toxicity data that meet all the requirements are appropriate for use with Copermittee data. If incomplete or faulty quality control measures or other inconsistencies are found, the data should not be used quantitatively. Data may be used semi-quantitatively or qualitatively if deemed appropriate by the marine toxicologist. For example, toxicity data reported by a non-accredited laboratory for a species such as Fathead Minnow (*P. promelas*), which is not part of the Copermittee Monitoring Program, could be assessed to determine if it provides supporting evidence to an existing hypothesis regarding the constituents of concern in the watershed. It is important to note that these data may not be given equal "weight" under the scenario presented above.

3.4 Ambient Bay and Lagoon Monitoring Data

The Ambient Bay and Lagoon monitoring (ABLM) program is intended to provide an indication of how marine life in bays and lagoons is affected by urban runoff and to assess the overall health of the receiving waters. The program monitors coastal lagoons and estuaries, including

Mission Bay and Oceanside Harbor. Components of the monitoring program are based on the triad of sediment chemistry, toxicity, and ecological community (benthic infauna) in the coastal embayments (lagoons and bays) of San Diego County.

The program is based on long term trend data assessment (see Sections 4.5 and 4.6) that will allow for prioritization of coastal embayments. It is anticipated that if high priority coastal embayments are identified in the future, special studies or investigations will be conducted to answer specific questions defined by the data collected over the next few years.

There are wide variations in sediment characteristics within coastal embayments because of temporal variations in deposition patterns, the influence of stream and tidal channels, sequestering of contaminants by marshes and grasses, and connectivity with the ocean. Sediments that accumulate in coastal embayments as a result of urban runoff are dispersed according to the different energy conditions that are encountered at stream outfalls and in the embayment. Fine-grained sediments tend to accumulate in lower energy conditions between active stream and tidal channels; whereas, coarser sediments accumulate in stream and tidal channels as point bars. This variability complicates measuring and assessing the concentration and distribution of contaminants and requires that care be taken to specify the frequency and locations of field samples. Site assessments are further complicated by seasonal effects, which can be regular, or atypical, caused by drought that can reduce sediment outflow or high-energy storms that can displace large amounts of sediments and significantly alter the distribution and availability of contaminants.

Accounting for this inherent variability in monitoring coastal embayments requires comprehensive site assessments that reflect the possible range of variability of both long-term, periodic variations and infrequent, but often high-energy, episodic events. Such comprehensive assessments can be extremely labor intensive and expensive. Thus, rather than trying to directly measure contaminant loading in the water, the approach that was used in the ABLM Program focuses on the receiving water sediments where contaminants are most likely to be found. Fine-grained sediment particles in the size range typical of silts and clays (<64 microns in diameter) preferentially adsorb most contaminants found in the waters of coastal wetlands (Gibbs 1973, Moore et al. 1989, Kennish 1998). Fine-grained sediments tend to have large surface areas with unsatisfied surface charges that promote adsorption of ionic complexes of metals, PCBs, PAHs, and pesticides. This association is particularly strong where fine-grained sediments are associated with high levels of total organic carbon (TOC). Additionally, fine-grained organic sediment in overabundance can overwhelm the endemic flora and fauna of lagoons and estuaries. Fine-grained sediments with high TOC content pose the greatest risk to the biological communities in the coastal lagoons and estuaries.

The ABLM Program was initiated during the 2002-2003 monitoring year and utilized the association between fine-grain size and high TOC levels to spatially target areas in each embayment where contaminants were most likely to be found. During each year, the program will be conducted in two phases:

- **Phase I – Contaminant Targeting:** three areas in each embayment with the finest grain size and highest TOC concentration will be identified using a stratified random design.

- **Phase II – Sediment Assessment:** the areas identified in Phase I will be assessed using the same “triad” approach that is being utilized for the storm water runoff program: chemistry, toxicity, and biology of the sediments.

3.4.1 Water Quality and Sediment Chemistry Assessment

At each station, a portable multi-probe instrument is used to measure the following parameters at the sediment/water interface: depth, temperature, DO, pH, specific conductance, and turbidity (Table 3-4). In addition to water quality measurements, the following visual observations are recorded at each site: percent cover of algae or grasses (as an indicator of potential eutrophication), sediment type, color, and odor (such as hydrogen sulfide). Chemistry analyses are performed for organics (PAHs, PCBs, chlorpyrifos and diazinon) and total metals (Table 3-4).

Table 3-4. Analytical methods for the 2002-2003 Ambient Bay and Lagoon Monitoring Program.

Constituent	Volume Required	Method	MDL	Units	Holding Time
General Physical and Inorganic Non-Metals					
Temperature	In field	probe	na	°C	na
Dissolved Oxygen (DO)	In field	probe	na	mg/L	na
pH	In field	probe	0.1	S.U.	na
Specific Conductance	In field	probe	1	umhos/cm	na
Turbidity	In field	probe			na
Total Organic Carbon (TOC)	125 g	EPA 415.1	1.0	mg/L	28D
Grain Size	125 g	Plumb 1981	1.0	% dry wt.	14D
Organics					
PAHs (Total)	100 g	GC/MS SIMS	10	μg/Kg	14 D
PCBs (Total)	100 g	EPA 8082	4.68	μg/Kg	14 D
Chlorpyrifos	100 g	EPA 8141	2	μg/Kg	14 D
Diazinon	100 g	EPA 8141	2	μg/Kg	14 D
Metals, Total					
Antimony (Sb)	200 g	EPA 6020	0.6	mg/Kg dry wt	14 D
Arsenic (As)	200 g	EPA 6020	0.2	mg/Kg dry wt	14 D
Cadmium (Cd)	200 g	EPA 6020	0.1	mg/Kg dry wt	14 D
Chromium (Cr)	200 g	EPA 6020	0.4	mg/Kg dry wt	14 D
Copper (Cu)	200 g	EPA 6020	0.4	mg/Kg dry wt	14 D
Lead (Pb)	200 g	EPA 6020	0.1	mg/Kg dry wt	14 D
Nickel (Ni)	200 g	EPA 6020	0.2	mg/Kg dry wt	14 D
Selenium (Se)	200 g	EPA 6020	0.6	mg/Kg dry wt	14 D
Zinc (Zn)	200 g	EPA 6020	2.2	mg/Kg dry wt	14 D
Toxicity					
10 day acute with *Eohaustorius estuarius*	5 L	EPA 1994			6W

3.4.2 Sediment Toxicity Assessment

The locations for sediment chemistry are also used for sediment toxicity testing. A single composite of the three locations identified in Phase I is analyzed for toxicity. In the laboratory, toxicity testing is performed using a standard 10-day acute test using the estuarine amphipod *Eohaustorius estuarius* using U.S. EPA (1994) guidelines (Table 3-4). This test consists of a 10-day exposure of *E. estuarius* to sediment under static conditions. Amphipods are placed in glass chambers containing seawater and a 2 cm layer of test sediment. The number of surviving amphipods is measured at the end of the test and is used to calculate the percentage survival. Individuals are visually inspected to confirm proper size and healthy condition prior to use in sediment testing. All tests are initiated within 14 days of sediment collection and include native controls for quality assurance.

3.4.3 Benthic Infauna Assessment

For the benthic infauna assessment, a separate sample is collected at each station with a 0.1 m^2 grab sampler (either a box core or a Van Veen). The sample is sorted through a 1.0 mm sieve. Retained organisms and sediments are fixed in a buffered formalin solution and delivered to the laboratory for processing and long-term preservation. The infaunal samples are taken from the same station as samples for sediment chemistry and toxicity; however, the infaunal samples are not composited. Thus, there are three samples per embayment retained for infaunal analyses.

In the laboratory, infaunal samples are transferred from formalin solution to alcohol for processing. Organisms are separated from the sediments by trained technicians using dissecting microscopes into five major taxonomic groups: arthropoda (insects and crustaceans), annelida (worms), mollusca, echinodermata, and miscellaneous minor phyla. Upon completion of the sorting, the taxonomic groups are distributed to professional taxonomic experts in each of the categories for counting and identification of the organisms to the lowest practicable category (usually species and developmental stage).

3.4.4 Third Party Data

Third party data collected from San Diego County embayments may only be merged with data from the current ABLM Program if it is obtained under the same project design described above because the program is targeting monitoring sites that have very specific characteristics and are representative of worse-case conditions. Data collected from the embayments under alternative monitoring schemes may be used semi-qualitatively or qualitatively.

4.0 WATERSHED DATA ASSESSMENT

The watershed data assessments conducted in San Diego county closely resembles the “Model Storm Water Monitoring Program for Municipal Separate Storm Sewer Systems in Southern California” developed by the Stormwater Monitoring Coalition’s (SMC) Model Monitoring Technical Committee. The SMC Model Document presents a flowchart that guides management questions that drive the design of the monitoring program based on the answers obtained. The questions defining the current monitoring program for San Diego Stormwater Copermittees can be summarized as follows:

1. What are the water quality conditions in the watershed?
2. Are water quality conditions in the watershed getting better or worse?
3. Are beneficial uses being impacted based on existing external information?
4. What is the relative contribution of urban runoff to the conditions in the watershed?
5. What are the sources to urban runoff that contribute to water quality conditions?

These five core management questions are presented in a logical flow diagram (Figure 4-1).

As data are collected and analyzed over time it is anticipated that the questions will evolve and possibly expand with the addition of more refined questions requiring additional data analysis and in some cases collection of different types of data or special studies.

Watershed data assessments utilize a triad approach. The triad approach is a management tool used to determine the short and long term actions to address potential or actual water quality problems in the watershed. It is a management tool that can guide watershed decision making by stakeholders. During the annual water quality assessment, the high, medium or low frequency of COC(s) is evaluated for each watershed using the latest data collected and potential water quality issues are determined. In some cases confirmation of water quality problems will require that additional data be collected or assessed to understand the extent of the problem. Additional information to assess if a water quality problem exists may be available from third party data or a special study that can be used to answer questions relating to sources of the COC(s). In some instances, data from third parties may be used to further define the problem both spatially and temporally.

The three components of the triad are physical and chemical water quality data, toxicity data and benthic data. Sections 4.1 through 4.3 describe how each component of the triad is assessed independently and then these components are combined in Section 4.4 to determine a set of actions to address water quality and habitat problems. These sections answer the first core management question, *"What are the water quality conditions in the watershed?".* This question is currently addressed in this framework document by comparing physical and chemical water quality data from various programs (wet weather at mass loading stations, dry weather from jurisdictional programs, etc.) to specific benchmark values (Section 4.1). Other data, such as toxicity (Section 4.2) are reviewed to determine if pollutants (identified or not) have the potential to, or, are currently impacting test species selected to represent watershed habitat. Benthic data are reviewed and analyzed (Section 4.3) to determine actual habitat conditions in the watershed.

The second core management question, *"Are water quality conditions in the watershed getting better or worse?",* is discussed in two sections. In Section 4.5 (Long-Term Trend Data Analysis), multi-year data are used to identify trends in concentrations of COCs for which identifying the source(s) may be of interest and, in Section 4.6 (Ambient Bay and Lagoon Monitoring Data Assessment), are used to give an indication of how worst-case conditions may be changing over time. The third core management question, *"Are beneficial uses being impacted based on existing external information?",* is addressed in Section 4.7. The tools available to assess the fourth core management question, *"What is the relative contribution of urban runoff to the conditions in the watershed?",* and the fifth core management question, *"What are the sources to urban runoff that contribute to water quality conditions?",* are addressed through a Cross-Watershed Comparison that uses statistical tools and modeling found in Section 4.8.

Application of this watershed assessment process leads to a prioritization of water quality issues that can assist in short and long term planning efforts, and developing activities directed at maintaining or improving water quality. In Section 4.9, a method to periodically update core management questions and modify the watershed assessment strategy is described.

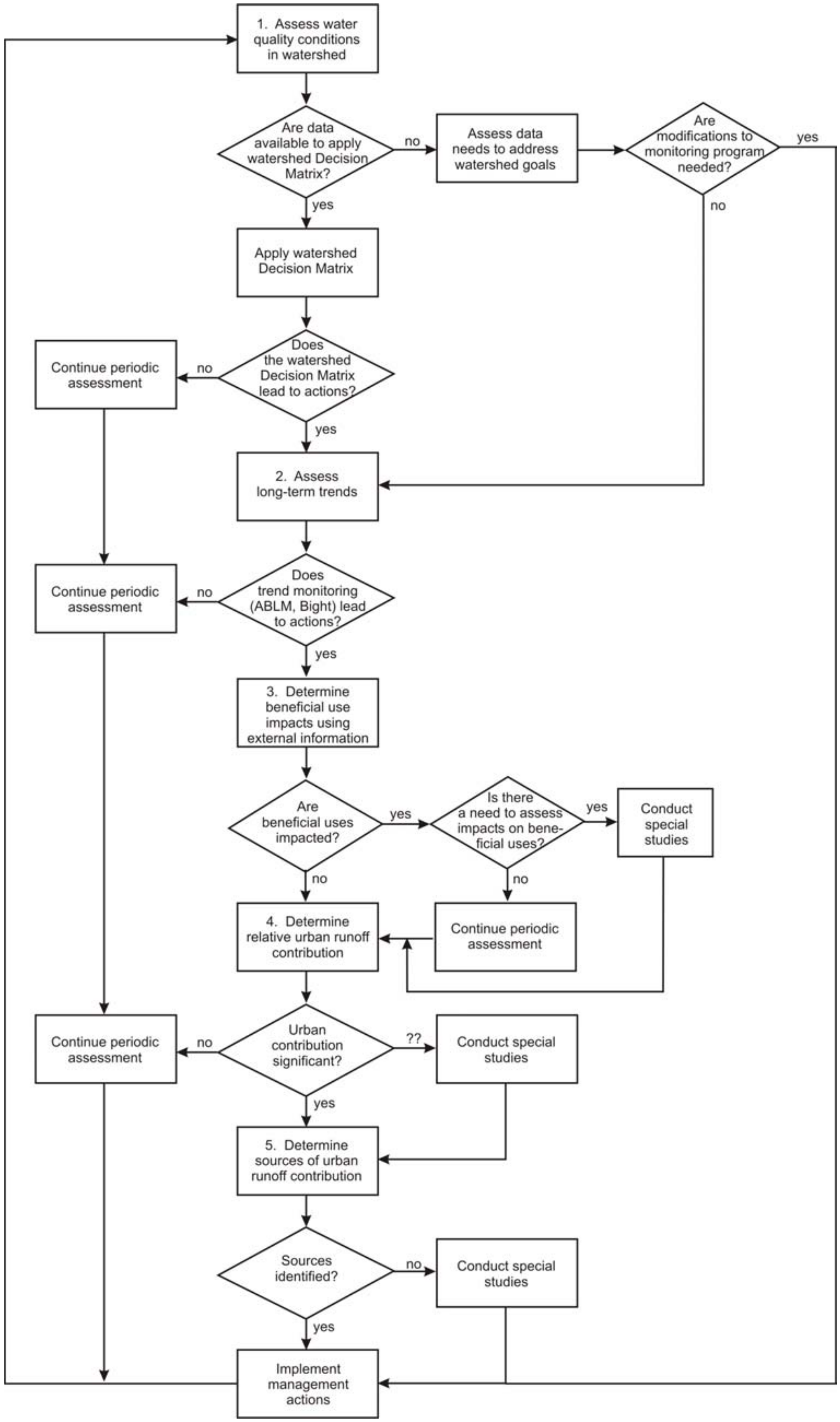
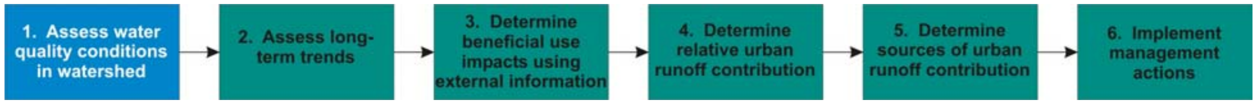


Figure 4-1. Watershed Assessment logical flow diagram for the five core management questions (Adapted from Figure 1 - SMC Model Storm Water Monitoring Program, 2004).

4.1 Chemistry Data



Wet weather chemistry data (physical, chemical, and bacteriological measurements) from the mass loading stations (MLS) are compared to the Water Quality Objectives shown in Table 4-1 and dry weather station data are compared to the Action Levels in Table 4-2 to determine the constituents that are exceeded most often in the watershed. The tables are not inclusive of all analytical measurements that can be conducted, but represent the constituents that are most common to water quality monitoring. If other chemistry data are available, the appropriate standards or water quality objectives should be identified while considering some of the following:

- What is the refined or follow-up question one seeks to answer?
- How was the sample collected – what does it represent?
- Does it meet basic quality control criteria (preservation, analytical methods, etc.)?
- Is the reporting limit low enough to allow for comparison to the objective, action level, or benchmark?
- If the data are to be merged with an existing dataset, are the data compatible (same reporting limits, analytical methods, etc.)?

Water quality objectives are defined in the San Diego County Copermittee program as benchmarks for comparison to monitoring results and do not necessarily reflect regulatory compliance for municipal stormwater discharges.

MLS wet weather results are compared to water quality objectives found in the following sources:

- San Diego Basin Plan (September 8, 1994)
- California Toxics Rule (CTR) 40 CFR 131 – 65FR 31682, May 18, 2000
- USEPA Multi-Sector General Permit (65FR 64746, October 30, 2002)
- California Department of Fish and Game

The water quality objectives utilized are the same across all watersheds in San Diego County except for total dissolved solids and fecal coliform. Total dissolved solids objectives are applied by hydrologic area or hydrologic sub-area as noted the 1994 Basin Plan (Table 4-1). Fecal coliform REC-2 standards apply at Tecolote Creek, Chollas Creek and Tijuana River while REC-1 standards are used for all other watersheds as shown in Table 4-1 below.

<u>Dry Weather Action Levels</u>
Dry weather action levels are established by the Copermittees to trigger investigations upstream of the sampling location and to eliminate illicit connections and illegal discharges (ICID). Dry weather action levels were initially established in 2002 and are updated on a yearly basis, as necessary. In order to allow for comparison with exceedances at the MLS, for which different water quality objectives are used, modifications are made that allow for comparison of MLS data and dry weather station (DWS) data. For example, in the 2003 watershed assessments the dry weather action levels for bacterial indicators were applied to the wet weather data instead of the Basin Plan REC-1 or REC-2 criteria in order to identify potential links between dry and wet weather constituents of concern. Similarly, turbidity action levels in dry weather samples are

evaluated using Best Professional Judgment (BPJ) while in wet weather (at the MLS) the Basin Plan water quality objective of 20 NTU is used. Therefore, when assessing dry and wet weather samples for turbidity at a watershed level the Basin Plan objective was used. See Table 4-2 for a summary of the dry weather action levels used to perform the data analysis.

Table 4-1. Water Quality Objectives for Wet Weather Monitoring at Mass Loading Stations.

Constituent	Units	WQO	Source
General / Physical / Organic			
Electrical Conductivity	umhos/cm		
Oil And Grease	mg/L	15	USEPA Multi-Sector General Permit
pH	pH Units	6.5-8.5	Basin Plan
Bacteriological			
Enterococci	MPN/100 mL		
Fecal Coliform	MPN/100 mL	400/4,000	Basin Plan REC-1/REC-2
Total Coliform	MPN/100 mL		
Wet Chemistry			
Ammonia As N	mg/L		
Un-ionized Ammonia as N	µg/L	25 (a)	Basin Plan
BOD	mg/L	30	USEPA Multi-Sector General Permit
Chemical Oxygen Demand	mg/L	120	USEPA Multi-Sector General Permit
Dissolved Phosphorus	mg/L	2	USEPA Multi-Sector General Permit
Nitrate As N	mg/L	10	Basin Plan
Nitrite As N	mg/L	1	Basin Plan
Surfactants (MBAS)	mg/L	0.5	Basin Plan
Total Dissolved Solids	mg/L	750	Basin Plan by watershed
Total Kjeldahl Nitrogen	mg/L		
Total Phosphorus	mg/L	2	USEPA Multi-Sector General Permit
Total Suspended Solids	mg/L	100	USEPA Multi-Sector General Permit
Turbidity	NTU	20	Basin Plan
Pesticides			
Chlorpyrifos	µg/L	0.02	CA Dept. of Fish & Game
Diazinon	µg/L	0.08	CA Dept. of Fish & Game
Malathion	µg/L	0.43	CA Dept. of Fish & Game
Hardness			
Total Hardness	mg $CaCO_3$/L		
Total Metals			
Antimony	mg/L	0.006	Basin Plan
Arsenic	mg/L	0.34/0.05	40 CFR 131/ Basin Plan
Cadmium	mg/L	0.0046	40 CFR 131
Calcium	mg/L		
Chromium	mg/L	0.016	CTR (Cr VI)
Copper	mg/L	0.0135	40 CFR 131
Lead	mg/L	0.082	40 CFR 131
Magnesium	mg/L		
Nickel	mg/L	0.47/0.1	40 CFR 131/ Basin Plan
Selenium	mg/L	0.02	40 CFR 131
Zinc	mg/L	0.122	40 CFR 131
Dissolved Metals			
Antimony	mg/L	(e)	40 CFR 131
Arsenic	mg/L	0.34 (c)	40 CFR 131
Cadmium	mg/L	(b)	40 CFR 131
Chromium	mg/L	(b)	40 CFR 131

Constituent	Units	WQO	Source
Copper	mg/L	(b)	40 CFR 131
Lead	mg/L	(b)	40 CFR 131
Nickel	mg/L	(b)	40 CFR 131
Selenium	mg/L	0.2 (d)	40 CFR 131
Zinc	mg/L	(b)	40 CFR 131

Blank spaces have been verified and no data is available.

(a) Water Quality Objective is for unionized ammonia which may be calculated from ammonia as nitrogen using pH, temperature and salinity.

(b) Water Quality Objective for dissolved metal fractions are based on total hardness and are calculated as described by the USEPA Federal Register Doc. 40 CFR Part 131, May 18, 2000.

(c) Water Quality Objectives for dissolved metal fractions are based on water effects ratios (WER) and are calculated as described by the USEPA Federal Register Doc. 40 CFR Part 131, May 18, 2000.

(d) Water Quality Objective is based on the total recoverable form as described by the USEPA Federal Register Doc. 40 CFR Part 131, May 18, 2000.

(e) USEPA has not published an aquatic life criterion value.

Sources

USEPA National Pollutant Discharge Elimination System (NPDES) Storm Water Multi-Sector General Permit for Industrial Activities, 65 Federal Register (FR) 64746, Final Reissuance, October 30, 2000.

Siepmann and Finlayson 2000.

Basin Plan, September 8, 1994.

Assembly Bill 411 - Title 17 of the California Code of Regulations, Section 7958.

USEPA Federal Register Document 40 CFR Part 131, May 18, 2000.

Table 4-2. Dry Weather Action Levels for 2002.

Constituent	Action Level	Note
pH	<6.5 or >9.0	
Orthophosphate-P	2.0 mg/L	
Nitrate-N	10.0 mg/L	
Ammonia-N	1.0 mg/L	
Turbidity	20 NTU	Used Basin Plan WQO instead of BPJ when comparing with MLS data
Conductivity		Best professional judgment
MBAS	1.0 mg/L	
Oil and grease	15 mg/L	
Diazinon	0.5 ug/L	
Chlorpyrifos	0.5 ug/L	
Dissolved Cadmium	CTR	Used CTR table, 1-hour criteria. Action level is based on hardness. Where hardness data were not available, the average value for the watershed was substituted.
Dissolved Copper	CTR	
Dissolved Lead	CTR	
Dissolved Zinc	CTR	
Total Coliform	50,000 MPN/100 mL	2003 Action Levels defined by 95th percentile were applied at the MLS for comparison with DWS data. Basin Plan objectives are only available for Fecal coliform (REC-1 and REC-2) as shown in Table 4-1.
Fecal Coliform	20,000 MPN/100 mL	
enterococcus	10,000 MPN/100 mL	

4.1.1 How to Determine Frequency of Occurrence for COCs

The next step in the watershed assessment is to determine if any of the constituents of concern are exceeding benchmarks at a high frequency. Constituents may be evaluated yearly or on a longer timeframe to determine if management actions result in improvements (COC at a high frequency of occurrence changes to medium frequency of occurrence). Depending on the amount of data and the quality of the datasets, various sets of criteria can be developed to prioritize or rank constituents of concern in the watershed. Therefore, on a periodic basis watershed stakeholders will need to develop criteria tables appropriate to the available data. The 2002-03 interim criteria (Table 4-3) applied to watershed assessments incorporates data from mass loading stations, dry weather stations and coastal monitoring for bacteria.

Table 4-3. Interim Criteria for Evaluating Mass Loading and Dry Weather Station Data.

COC Frequency of Occurrence	Criterion No.	Definition
High ♦♦♦	1	Mass loading station tests results exceed WQO in greater or equal to 80% of samples.
	2	Six of the last consecutive storm samples at the MLS exceed WQO.
	3	Less than 80% and greater than or equal to 50% of the MLS samples exceed WQO and at least one DWS exceedance in the past year.
	4	Less than 80% and greater than or equal to 50% of the MLS samples exceed WQO and a significant increasing trend is found.
Medium ♦♦	5	Less than 80% and greater than or equal to 50% of the MLS samples exceed WQO and no exceedances or data available for DWS in the past year.
	6	Less than 80% and greater than or equal to 50% of the MLS samples exceed WQO and one or more exceedances found in last 2 years of monitoring at the MLS (generally applies to historical datasets).
	7	Greater than 50% of the DWS samples have exceedances in the past year.
Low ♦	8	DWS exceedances in 10 to 50% of the samples in the past year.
	9	MLS exceedances found in 25% to less than or equal to 50% of the samples and at least one exceedances found in last 2 years at the MLS (with or without DWS exceedances in the past year).
	10	Greater than 50% of the MLS samples have exceedances and no exceedances in the last 2 years at the MLS.
Coastal Program	11	Persistent exceedances (greater or equal to 80% of samples). Add one ♦ to bacteria determination (up to three ♦ maximum).

Note: Best professional judgment applies when unique situations arise (fewer samples at a site; sewage spills) and for toxicity once it is linked to a specific COC.

An example that applies the interim criteria presented in Table 4-3 is described below. The chemistry component of the triad decision matrix is determined to be persistently exceeding water quality objectives when the frequency of the COC is found to be high for the entire dataset being reviewed. For the San Diego County urban runoff monitoring program, this determination is made using the data collected for each watershed at the MLS. The span of the data varies from only a couple of years of wet weather for the majority of the watersheds to approximately ten years of monitoring for stations at Chollas Creek and Tecolote Creek.

The interim criteria take into account the exceedances at the MLS, DWS and coastal outfalls; and classify each COC as high, medium or low frequency of occurrence in the watershed. The classification of COC can change from year to year in response to the changes in the levels of the pollutants. The existing set of interim COC criteria is anticipated to evolve as the program matures and the data set expands. Periodic revisions or additions to the criterion definitions are anticipated in the future. Statistical tools that may be used to perform data analysis that support the development of future ranking criteria are found in Section 5.

DWS data were given less weight in the determination of watershed COC due to factors that include:

1. The dry weather monitoring program's main focus is to identify illicit connections and illegal discharges. Sample stations may not be representative of overall urban runoff quality since they include samples of ponded water.
2. Dry weather monitoring parameters are a subset of MLS monitoring parameters.
3. DWS may be located in the MS4 upstream of BMPs (detention basins, etc.) and samples may not be representative of urban runoff entering the receiving water.

For this evaluation, dry weather stations that only have field test kit results are not used in the assessment of COC. Only DWS monitored using laboratory analysis from grab samples including concurrent field test results are considered for comparison with MLS exceedances of water quality objectives. Only DWS located upstream of the MLS are taken into account when applying the interim COC criteria. Lastly, only DWS samples collected during routine monitoring and not as part of the ICID investigation phase of the program are used. The distinction between routine site visits and ICID investigations may not have been made in the 2002 dry weather database; therefore, minor deviations in the type of sampling data set used may have occurred.

If the number of DWS sampled was small, best professional judgment is used when applying the interim COC criteria. For example, if only three samples were collected and one exceedance was observed, then the 33% exceedance frequency may not be representative of watershed conditions.

Benchmarks for bacterial levels are judged differently in the MLS and DWS. The MLS water quality objective for fecal coliform was derived from the Basin Plan (REC-1 and REC-2) while DWS levels are compared to Copermittee defined action levels (Table 3-2) for all three bacterial indicators (total and fecal coliform and enterococcus). In order to compare the two datasets, the DWS action levels are applied to the MLS total coliform, fecal coliform, and enterococcus data. Otherwise, identification of bacterial indicators as potential COCs in the watershed between these two different data sets is not feasible.

Toxicity testing at the MLS does not measure a COC (for details see Section 4.3). Toxicity is a test to determine if an analyte (chemical or other) or group of analytes is present in concentrations capable of causing toxicity in the selected species. Once an analyte(s) is identified as the source of the toxicity through the TIE/TRE steps of the method, then it is possible to define toxicity as having a high frequency of occurrence because it has been positively linked to the actual constituent of concern identified to be causing the toxicity.

4.2 Bioassessment Data

The Rapid Stream Bioassessment data collected by the Copemittees includes a taxonomic listing of all benthic macroinvertebrates identified in the surveys, and calculation of the biological metrics listed in the CSBP. Additionally, calculation of the Index of Biotic Integrity (IBI) for all monitoring reaches is included, following the most recent version developed by the CDFG Aquatic Bioassessment Laboratory for coastal southern California (Ode, Rehn, and May, In Press).

For every monitoring reach, an Index of Biotic Integrity (IBI) was calculated utilizing the most recent method developed by CDFG (Ode, Rehn, and May, In Press). The IBI can also be used to evaluate community conditions over time to monitor the effects of habitat degradation or the success of restoration efforts.

Additional analysis of the data includes a comparison of IBI scores with habitat quality, and a cluster analysis (see Section 5.2.1 for a description of this statistical tool) to show the relationship between species associations and reach characteristics.

To gain better insight into benthic community structure and habitat associations, a cluster analysis is performed for each survey. This analysis illustrates relationships between species, and also associations between species and habitat types of the monitoring reaches. The analysis is based on a Bray-Curtis similarity index calculated on relative abundances of taxa by location. Monitoring sites with similar communities of taxa will cluster together; likewise, taxa that occur together at different locations will cluster together. The station clusters also show habitat type preferences by taxa.

For the Triad Decision Matrix, determining if benthic alterations have been found is necessary. The final IBI score ranges give a relative rating of Very Good, Good, Fair, Poor or Very Poor to each reach surveyed and it is the overall findings of multiple surveys that are applied to the Triad. Benthic community alteration is defined as conditions rated as Very Poor in San Diego County since some reference sites have been found to be rated as low as Fair and Poor.

4.3 Toxicity Data

The 2001-2002 Urban Runoff Monitoring Report included a recommendation that the Acute Toxicity Units (TUa) no longer be considered when assessing toxicity in this program for the following reasons. The California Ocean Plan states that it is only possible to have a TUa of 0 when the survival in the 100% concentration is 99% or better. It is not clear from this document whether the percent survival is absolute or relative to the control for the test. The endpoint is calculated relative to the control for this monitoring program. Even given this more liberal interpretation of the calculation of a TUa, in a test with forty animals per treatment, if one more animal dies in the 100% treatment than in the control, the sample exceeds its acute toxicity limit. The death of one animal in a toxicity test is more likely due to variability in response than to actual toxicity of a sample. Setting the limit at a TUa of 0 does not account for any of the variability inherent in toxicity tests. This is evidenced, for example, when the results for 2002-2003 *H. azteca* are reviewed. Out of 122 samples only 25 exceeded the acute toxicity limit with a range in TUa from 1.12 to 0.23, yet only 5 of those samples had mortality that was significant when compared to the control and all had a No Observed Effect Concentration (NOEC) of 50%. The TUa for the remaining samples ranged between 0.85 and 0.23. Therefore, it is more appropriate to set the limit based on an endpoint that statistically takes into account variability in its calculation.

The results reported for the Copermittee monitoring program focus on the acute toxicity limit as the NOEC of 100% for the test sample. This limit will take into account any inherent variability in the test, yet still be protective of the watershed. The seven-day chronic effects are estimated using the NOEC for both survival and reproduction. This is the highest concentration tested in which there was no statically significant effect on the survival or reproduction compared to the control response. Lower NOEC values equate to higher toxicity in the sample. The methods used in regional monitoring program are shown in Table 4-4.

Table 4-4. Toxicity Water Quality Objectives for wet weather monitoring at Mass Loading Stations.

Species/Test	Units	WQO	Source[1]
Toxicity			
Ceriodaphnia 96-hr	LC_{50} (%)	100	NPDES Order 2001-01; Appendix D-6
Ceriodaphnia 7-day survival	NOEC (%)	100	NPDES Order 2001; Appendix D-6
Ceriodaphnia 7-day reproduction	NOEC (%)	100	NPDES Order 2001; Appendix D-6
Hyalella 96-hr	NOEC (%)	100	NPDES Order 2001; Appendix D-6
Selenastrum 96-hr	NOEC (%)	100	NPDES Order 2001; Appendix D-6

(1) Modified from TUa to NOEC as noted in the text.

4.3.1 Determining Persistent Toxicity for the Triad Decision Matrix

Based on the explanation provided in the preceding section, toxicity test results are reported as the NOEC. The NOEC is the lowest concentration at which there is no statistical difference from the control. Therefore, a concentration of less than 100% is considered to have some degree of toxic effect.

Persistent toxicity is evident when more than 50% of the toxicity tests conducted to date for any given species at a specific site have a NOEC of less than 100%. The results of this determination are then combined with the high frequency constituents of concern (chemistry data) and benthic data in the Triad Decision Matrix to determine the actions to be taken.

4.3.2 Statistical Methods for Toxicity and COC

Toxicity can also be combined with the chemistry data to find relationships between the levels of particular constituents and the toxicity in a particular sample.

The relationship between toxicity and COC is generally evaluated by two methods. The first method uses a multiple regression model to correlate changes in toxicity to changes in COC levels in the water. This method groups data from all watersheds and is useful in providing general trends across the county and evaluating the effects of several COC at once. Sometimes thresholds of chemical concentrations are involved with toxicity whereby the organisms do not respond negatively until a certain chemical level is reached. Concentrations of COC above a specific threshold may no longer elicit a linear response in organism toxicity. Consequently, thresholds detract from the regression model. Therefore, a second method, threshold analysis, is used to clarify relationships following the regression analyses using the COC that were significant components of the final multiple regressions. The threshold analysis uses COC levels reported to be toxic in the literature where available and compares them to COC levels in the storm water samples.

4.4 How to Use the Triad Decision Matrix

The data from the three elements of monitoring (chemistry, toxicity, and benthic community assemblage) are assessed using the Triad Decision Matrix using the definitions in Table 4-5. This triad decision matrix was developed by MEC Analytical Systems, Inc. (MEC) and the Copermittee Monitoring Workgroup. This approach was presented to the RWQCB staff and approved as an appropriate technique for evaluating the health of watersheds and identifying possible actions to be taken. More recently, the Triad Decision Matrix was expanded by the

Stormwater Monitoring Coalition and is presented here with modifications that incorporate those new recommendations (Table 4-6).

Table 4-5. Triad Definitions for San Diego Storm Water Monitoring Program.

Triad Component	Definition
Persistent Exceedance of Water Quality Objectives	A constituent of concern with a high frequency of occurrence based on wet and dry weather data exceedances compared to established list of benchmarks or trigger levels
Evidence of Persistent Toxicity	More than 50% of the toxicity tests for any given species have a NOEC of less than 100%.
Indication of Benthic Alteration	IBI score indicates a substantially degraded community (very poor)

For each watershed, all three elements of the triad approach (chemistry, toxicity, and benthic community) are used. Chemistry data provide an indication of the pollutant load during a storm event and toxicity data an indication of the potential impacts to aquatic organisms during storm events. Dry weather chemistry data provides an indication of urban runoff pollutants. The benthic community data collected during stream bioassessment provides a more direct indication of the ecological health of the watershed in terms of insect/benthic community abundance and diversity. In addition, the stream bioassessment procedure provides information about the quality and condition of the physical habitat.

Results from chemistry, toxicity and relative benthic community health are assessed together using the triad approach to determine what short and/or long term actions are appropriate in a watershed. The general idea in the triad assessment is to look for persistence in the indicators. Persistence in several indicators provides an indication of an ecological concern that triggers the need to conduct short term actions, such as a TIE to identify the COCs in the watershed that may be responsible for storm water toxicity and/or benthic community degradation. Where long-term datasets are available, all the data are evaluated to identify persistent conditions. The majority of the mass loading stations are in their third year (2003-04) of monitoring and will soon have data from nine storm events available for the triad assessment.

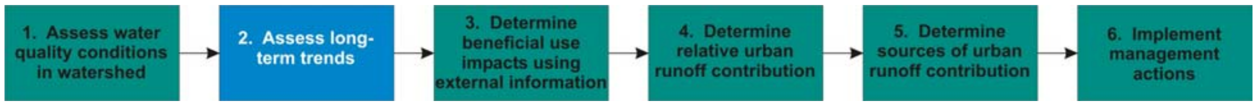
The triad approach does not consider fecal coliform and total dissolved solids for the purposes of triggering a decision or action. The bacteria parameters are not considered in the triad because they are not believed to influence toxicity responses in bioassay test organisms. Further, the REC-1 (water contact) and REC-2 (non-contact) WQOs for bacterial indicators are set for the protection of human health. Total dissolved solids are not considered since the water quality objectives for this COC as defined in the Basin Plan are set for municipal drinking water and do not necessarily reflect impacts to the ecology of the watersheds. However, fecal coliform and total dissolved solids data may be used to define high priority COC that lead to management actions even though they bypass the application of the triad decision matrix.

Table 4-6. Tabular Decision Matrix – chemical, toxicity, and benthic assemblage data available (adapted from SMC Model Storm Water Monitoring Program, 2004).

Chemistry	Toxicity	Benthic Alteration	Example Conclusions	Example Actions or Decisions
1. Persistent exceedance of water quality objectives (high frequency COC identified)	Evidence of persistent toxicity	Indications of alteration	Strong evidence of pollution-induced degradation	1)Toxicity tests at higher dilutions to better quantify toxicity; Use TIE to identify contaminants of concern, based on TIE metric. 2) Evaluate/identify upstream source as a high priority.
2. No persistent exceedances of water quality objectives	No evidence of persistent toxicity	No indications of alteration	No evidence of current pollution-induced degradation Potentially harmful pollutants not yet concentrated enough to cause visible impact	1) No immediate action necessary. 2) Conduct periodic broad scans for new and/or potentially harmful pollutants.
3. Persistent exceedance of water quality objectives (high frequency COC identified)	No evidence of persistent toxicity	No indications of alteration	Contaminants are not bioavailable Test organisms not sensitive to problem pollutants	1) TIE would not provide useful information with no evidence of toxicity. 2) Continue monitoring for toxic and benthic impacts. Consider whether different or additional test organisms should be evaluated. 3) Initiate upstream source identification as a low priority.
4. No persistent exceedances of water quality objectives	Evidence of persistent toxicity	No indications of alteration	Unmeasured contaminant(s) or conditions have the potential to cause degradation Pollutant causing toxicity at very low levels Synergistic effects of multiple chemicals at low levels causing toxicity	1) Recheck chemical analyses and evaluate detection limits relative to reported toxic levels. 2) Verify toxicity test results; Consider additional advanced chemical analyses. 3) Toxicity tests at higher dilutions to better quantify toxicity: Use TIE to identify contaminants of concern, based on TIE metric; Evaluate/investigate upstream source as a medium priority.
5. No persistent exceedances of water quality objectives	No evidence of persistent toxicity	Indications of alteration	Alteration may be due to physical impacts, not toxic contamination Test organisms not sensitive to problem pollutants Synergistic effects of multiple chemicals at low levels causing toxicity	1) No action necessary based on toxic chemicals. 2) Consider whether different or additional test organisms should be evaluated. 3) Consider potential role of physical habitat disturbance.

Chemistry	Toxicity	Benthic Alteration	Example Conclusions	Example Actions or Decisions
6. Persistent exceedance of water quality objectives high frequency COC identified)	Evidence of persistent toxicity	No indications of alteration	Toxic contaminants are bioavailable, but in situ effects are not demonstrable Benthic analysis not sensitive enough to detect impact Potentially harmful pollutants not yet concentrated enough to change community	1) Determine if chemical and toxicity tests indicate persistent degradation. 2) Recheck benthic analyses; consider additional data analyses. 3) Toxicity tests at higher dilutions to better quantify toxicity: • If recheck indicates benthic alteration, perform TIE to identify contaminants of concern, based on TIE metric. Evaluate/investigate upstream source as a high priority. • If recheck shows no effect, use TIE to identify contaminants of concern, based on TIE metric. Evaluate/investigate upstream source identification as a medium priority.
7. No persistent exceedances of water quality objectives	Evidence of persistent toxicity	Indications of alteration	Unmeasured toxic contaminants are causing degradation Pollutant causing toxicity at very low levels Synergistic effects of multiple chemicals at low levels causing toxicity Benthic impact due to habitat disturbance, not toxicity	1) Recheck chemical analyses and consider additional advanced analyses. 2) Toxicity tests at higher dilutions to better quantify toxicity. Use TIE to identify contaminants of concern, based on TIE metric. 3) Evaluate/investigate upstream source identification as a high priority. 4) Consider potential role of physical habitat disturbance.
8. Persistent exceedances of water quality objectives (high frequency COC identified)	No evidence of persistent toxicity	Indications of alteration	Test organisms not sensitive to problem pollutants Benthic impact due to habitat disturbance, not toxicity	1) TIE would not provide useful information with no evidence of toxicity. 2) Evaluate/investigate upstream source identification as a high priority. 3) Consider whether different or additional test organisms should be evaluated. 4) Consider potential role of physical habitat disturbance.

4.5 Long-Term Trend Data Analysis



The third core management question, "*Are water quality conditions in the watershed getting better or worse?",* is answered by looking at long term data sets both developed as part of the Copermittee Regional Monitoring efforts and other projects and monitoring programs.

A trend analysis of long term datasets is intended to track increases or decreases of the COC through time. Constituents of concern for MLS with historical data (Agua Hedionda Creek, Tecolote Creek, and Chollas Creek) are currently reviewed to identify trends in the concentration levels of these constituents. Other MLS will have larger data sets in the very near future that will allow for trend analysis spanning four to five years of monitoring during wet weather. Analyzing the concentration data to identify trends is only one of many ways to evaluate improving or deteriorating water quality. An increasing trend that shows concentrations of pollutants are nearing a WQO or continue to increase (when the water quality objective is already exceeded) is of particular concern and should be considered in the development of watershed actions that attempt to slow or reverse the trend.

Other sources of long term data available to assess overall water quality conditions are obtained through other monitoring like the Southern California Bight Projects. In the future, Copermittees may be able to determine links between watershed monitoring efforts and the results of Bight monitoring efforts.

The statistical tools used to determine trends are found in Section 5.

4.6 ABLM Data Assessment

As with the watershed assessments, the coastal embayments sampled in the ABLM program will be assessed using three data sources: 1) sediment chemistry; 2) sediment toxicity; and 3) benthic infauna community structure. Evaluations of the first year of the program are intended to provide an assessment of the worst-case conditions in each of the embayments, allow for comparisons across all of the embayments, and provide an indication of the impact on the embayment from its upstream watershed. In subsequent years, these long-term assessments can be completed along with an analysis of any temporal trends observed in the coastal embayments throughout San Diego County.

4.6.1 Chemistry Data

Currently, there are no universally accepted criteria for assessing contaminated sediments. However, the National Oceanographic and Atmospheric Administration (NOAA) has produced effects-based sediment quality values for evaluating the potential for constituents in sediment to cause adverse biological effects (Table 4-7). NOAA has defined two effects categories: Effects Range-Low (ER-L) and Effects Range-Median (ER-M). These parameters were developed from a large data set where results of both sediment toxicity bioassays (e.g., amphipod tests) and chemical analyses were available for individual samples. The ER-L concentrations are equivalent to the lower tenth percentile of available data screened by NOAA and indicate the low end of the range of concentrations at which adverse biological effects are observed or predicted in sensitive species and/or sensitive life stages. The ER-M values are concentrations based on the NOAA screened data at which effects are observed or predicted in 50% of the test organisms evaluated. In 1995, the ER-L and ER-M concentrations were updated to include an expanded data set (Long et al. 1995).

Table 4-7. Sediment Effects Guideline Values.

Parameter	Effects Range-Low (ER-L)	Effects Range-Median (ER-M)
Metals (mg/Kg)		
Antimony	2.0	2.5
Arsenic	8.2	70
Cadmium	1.2	9.6
Chromium	81	370
Copper	34	270
Lead	46.7	218
Nickel	20.9	51.6
Zinc	150	410
Organics (µg/Kg)		
Acenaphthene	16	500
Acenaphthylene	44	640
Anthracene	85.3	1,100
Florene	19	540
Naphthalene	160	2,100
Phenanthrene	240	1,500
Low-molecular weight PAH	552	3,160
Benz(a)anthracene	261	1,600
Benzo(a)pyrene	430	1,600
Chrysene	384	2,800
Dibenzo(a,h)anthracene	63.4	260
Fluoranthene	600	5,100
Pyrene	665	2,600
High molecular weight PAH	1,700	9,600
Total PAH	4,022	44,792
Total PCBs	22.7	180

Source: Long et al. 1995
ER-L = Concentration at lower tenth percentile at which adverse biological effects were observed or predicted.
ER-M = Concentration at which adverse biological effects were observed or predicted in 50% of test organisms.
mg/Kg = milligrams per kilogram.
µg /Kg = micrograms per kilogram.

Sediment chemistry data collected from each of the coastal embayments can be compared to the ER-L and or the ER-M data provided by NOAA. For each embayment, a comparison of the sediment chemistry data to the ER-L and ER-M values can be conducted. Because the ABLM program utilizes an approach that targets COC in each embayment (using TOC and grain size parameters), the individual assessments represent a worst-case scenario rather than a representative assessment of the embayment. The data should be interpreted to reflect this important distinction.

In addition, for each embayment ER-M values will be used to calculate a Total ER-M quotient. The concentration of each COC can be divided by its ERM to produce a quotient, or proportion of the ERM equivalent to the magnitude by which the ER-M value is exceeded or not exceeded. Quotients calculated for all contaminants in each embayment can then be summed to provide a Total ER-M Quotient. The Total ER-M Quotient thus represents an assessment for each embayment of the cumulative sediment chemistry relative to the NOAA threshold values. In this way, the cumulative risks of contamination to the benthic community can be assessed and compared between sites. This method has been used and evaluated by several researchers (Hyland et al. 1999, Carr et al. 1996, Chapman 1996, and Long et al. 1995) throughout the country and has proven to be a reliable basis for predicting anthropogenic-induced stress in benthic invertebrate communities.

4.6.2 Benthic Data

The benthic infauna data from each of the embayments can be assessed using a variety of indices common to ecological community structure evaluations. Some of the tools that are employed in the assessment include a species list, relative abundance, species diversity or richness, Shannon-Wiener Species Diversity Index, and an evaluation of the presence of sensitive and pollutant tolerant species. This information is incorporated into a two-way coincidence table that is used to perform a cluster analysis. The cluster analysis will show the relationship between the individual embayments and the various indicators used to describe the characteristics of the benthic infaunal community. Embayments with similar index or parameter scores will cluster together, providing a means by which the embayments can be ranked from best (least impacted community) to worst (most impacted community). The results of the cluster analyses can also be used to provide an individual assessment of each embayment.

4.6.3 Toxicity Data

Sediment toxicity results can be obtained from the exposure of the chosen test species (currently *Eohaustorius estuarius* an estuarine marine amphipod in a ten-day acute test) to sediments collected from each of the embayments. The percent survival of test organisms in sediments from the embayments is compared to percent survival in a control sample to assess benthic infaunal toxicity levels from each of the embayments sampled. A statistical evaluation is conducted for each of the embayments to determine if there is a statistically significant difference (using ANOVA) between toxicity in sediments from the embayment verses toxicity in the control.

In addition to the individual assessments of each embayment, the toxicity results are used to rank each of the embayments. The ranking is based on the percent survival *of E. estuarius* in the 10-day acute test, where the highest survival (lowest toxicity) receives a rank of one and the lowest survival (highest toxicity) receives a rank of 12.

4.6.4 Data Integration

Once all the ABLM data becomes available, a triad matrix or comparable assessment is developed so that the combination of benthic infauna data, sediment chemistry, and toxicity is used to develop a ranking of the embayments across the County (or in a watershed).

For each of the embayments, the three elements of the monitoring program are ranked individually for each site (1 to 12 for the 12 embayments assessed) as follows:

> **Sediment Chemistry –** The Total ER-M Quotient Value is used, where 1 represents low potential for toxicity and 12 represents high potential for toxicity;
>
> **Benthic Infauna –** The results of the Cluster Analysis are used to rank each site, where 1 represents the least-impacted site and 12 represents the most impacted; and

Sediment Toxicity – The results of the *Eohaustorius estuarius* percent survival value are used to rank each site, where 1 represents low toxicity and 12 represents high toxicity.

The individual ranks for each of the monitoring elements are added together and then these sums are ranked to develop an overall relative summed rank (where 1 represents the lowest summed score and 12 represents the highest summed score) of each embayment relative to the other embayments. A similar prioritization system is being developed for the wet and dry weather monitoring and bioassessment monitoring to rank the watersheds upstream of the coastal embayments, however, this process has not been finalized. The upstream watershed ranking system will utilize cluster analyses and triad results described in Sections 4.1 and 4.4. In this way, the relationship between the COC and impairments observed in the watersheds and corresponding embayments can be assessed.

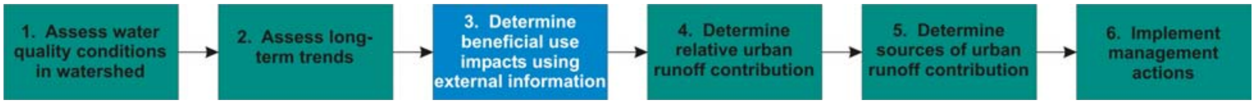
Table 4-8 presents a hypothetical example of how the results from the ABLM data assessment are ranked and used to develop a relative summed rank for the coastal embayments. It also presents a supposed ranking of the upstream watersheds. In this hypothetical example, Mission Bay (MB) had the lowest relative summed rank (1), meaning it was the least impacted of all the embayments monitored. In contrast, Batiquitos Lagoon (BL) had the highest summed rank (12), meaning it was the most impacted of the 12 embayments. Here, there are a total of 12 summed rankings but only 10 relative summed rankings because two sites had relative ranks of 7.5 and two sites had relative ranks of 9.5. After the rankings have been produced, a non-parametric statistical procedure, such as Kendall's Tau Test, can be used to determine the rank correlation among sites.

Table 4-8. Hypothetical Example of Ambient Bay and Lagoon Integrated Data Assessment.

Embayment	Sediment Chemistry Ranking (A)	Benthic Infauna Ranking (B)	Toxicity Ranking (C)	Sum of Ranked Values for Embayments (A+B+C)	Relative Summed Rank*	Upstream Watershed Rank
SMR	6	11	1	18	6	4
OH	9	8	4	21	7.5	7
SLR	12	5	6	23	9.5	9
BVL	11	3	9	23	9.5	10
AHL	5	1	11	17	5	6
BL	10	12	12	34	12	3
SEL	3	10	2	15	3	8
SDL	2	9	3	14	2	2
LPC	7	4	5	16	4	1
MB	1	2	7	10	1	12
SRE	8	7	10	25	11	5
TJR	4	6	11	21	7.5	11

* The lower the relative summed rank, the less-impacted the site is.

4.7 Impacts on Beneficial Uses



The third core management question: "*Are beneficial uses being impacted based on existing external information?*" can be answered in one of two ways. The first method is by reviewing and assessing existing information available from regulatory agencies. Sources of information to assess beneficial use impact for a watershed include:

- Beneficial Uses listed in the San Diego Basin Plan (1994).
- The State Water Resources Control Board 303(d) List of Water Quality Limited Water Bodies.
- The State Water Resources Control Board Monitoring List or other comparable list (pending final approval of the SWRCB Water Quality Control Policy for Developing the Section 303(d) List).
- The San Diego County Beach Closure Report and the yearly Beach Closure and Advisory Report.

These and supporting RWQCB strategies or workplans that address impacts to beneficial uses (TMDL implementation schedules, etc.) provide information that assists watershed stakeholders in the planning effort and development of management actions (data collection, source identification, etc.).

The second method comes from the need to assess impacts on beneficial uses because information from regulatory agencies does not exist or it is determined to be outdated or in some way inappropriate. In other words, new information (internally obtained) is needed to answer the core management question. In most of these cases, this can be accomplished by analyzing third party data or conducting special studies. The outcome of a special study, such as a Use Attainability Study, is to answer a core management question using new and current data and information.

4.8 Cross-Watershed Comparisons

This section addresses the fourth and fifth core management questions *"What is the relative contribution of urban runoff to the conditions in the watershed?"* and *"What are the sources to urban runoff that contribute to water quality conditions?"*

There are numerous methods available to assess the relative contribution and sources of urban runoff pollution. On a large scale, one method is examining constituent concentrations across watersheds and then grouping similar watersheds by constituent relationships. These types of analysis are performed using a variety of statistical tools of which the three most common are scatterplots, ANOVA and cluster analyses. Another method now available is computer modeling. Applying geographical, topographical, and land use data to analytical data can provide valuable information on the contribution and sources of pollutants. In Section 4.8.2 examples of the output of computer modeling are presented.

4.8.1 Statistical Methods

The statistical tools used for the cross watershed comparison include scatterplot analysis, analysis of variance (ANOVA), and multivariate cluster analysis. In this section, an overview of statistical methods applicable to cross-watershed comparisons is presented. The methods are discussed in detail in Section 5.

4.8.1.1 Scatterplots

Scatterplots provide a constituent based comparison among watersheds. The purpose of the scatterplots is to provide a visual representation to compare constituents between stations. Scatterplots are simple plots of concentrations of constituents plotted on the y-axis against the mass loading station identified on the x-axis. Each constituent and toxicity test is represented by its own scatterplot with all sampling dates plotted on a single graph. Where larger data sets are available for mass loading stations scatterplots include trend data plots. For scatterplots, where non-detect results are present, the detection limit is plotted as its value. Also, when a constituent concentration during separate storm events is the same, the scatterplot appears to have only one point at that concentration, because the points are co-located. All constituents are monitored at mass loading stations during three storms each year (with the exception of Santa Margarita) and all points are included in scatterplots.

4.8.1.2 Analysis of Variance

The ANOVA is used to determine statistical differences between the watersheds for the year as a whole (storms are used for replication). The term *analysis of variance* is sometimes a source of confusion. In spite of its name, ANOVA is concerned with differences between *means* of groups, not differences between *variances.* This analysis uses variances to detect whether the means are different. The ANOVA determines the variation (variance) *within* the groups that are being compared (e.g., monitoring stations), then compares that variation to the differences *between* the groups, taking into account how many subjects there are in the groups. If the observed differences between the means of groups are larger than those expected by chance relative to the underlying variance, a statistical significance is achieved. For watershed assessments, each constituent that was observed in any sample above the method detection limit was tested by ANOVA. Because this analysis needs to calculate a variance, the constituent with results below the detection limit at a station are handled in the following manner. If only one sample is below the detection limit, one-half the detection limit is used. If more than one sample is below the detection limit, each of the values is set so that the mean of all the values will be one-half the detection limit. For example, if the detection limit is 0.6 and there are two values below the detection limit, one will be set to 0.15 and the other will be set to 0.45 so that the mean of the two values is 0.3 (one-half the detection limit). The bacteriological measures are $\log_{10}$ transformed for this analysis.

4.8.1.3 Cluster Analysis

Cluster analysis is used to identify mass loading stations and sampling dates with similar constituent loadings. Multivariate cluster analysis is applied to the constituent and the toxicity endpoints (in terms of toxicity units) for each MLS and sampling time. This approach groups the station/times by the commonality of the constituent concentrations found at each one. Likewise, it groups the constituents according to similar loadings at stations. Prior to the analysis the bacteriological measures are $\log_{10}$ transformed and the data for each constituent is standardized by the overall mean value for each constituent.

4.8.2 Computer Modeling

Computer modeling is one valuable tool that can be used to assist in answering the last two core management questions:

1. What is the relative contribution of urban runoff to the conditions in the watershed?
2. What are the sources to urban runoff that contribute to water quality conditions?

This section describes two readily available models that Copermittees may use:

- Event Mean Concentration Model
- Dynamic Storm Water Model (Storm Water Management Model)

4.8.2.1 Event Mean Concentration Model

In general, comparisons of estimated event mean concentrations (EMCs) with measured EMCs at mass loading stations are performed to compare watershed areas using modeling methods. Currently the model input uses the land use EMCs calculated from results obtained from the first five monitoring years of the Regional Monitoring Program for residential, commercial, industrial and open land use categories for all parameters. EMCs for parks and open land are based upon Nationwide Urban Runoff Program (NURP) values with the exception of lead, where more recent data are used (Walker 1990), and BOD, total copper, and total cadmium, for which NURP results are not available. Results from calculated EMCs from Los Angeles County based on data collected from 1994-2000 routine monitoring of land use stations are used for these parameters. No parks/open space land-use sites have been monitored under the current program because the program focuses on urban land uses. Federal Highway Administration data (FHWA 1990) are used for roadways.

The San Diego Association of Governments (SANDAG) geographic information system (GIS) database and the SANDAG 2000 Generalized Land Use maps are used to determine land use within the catchments of the mass loading sites. SANDAG data are supplemented with information from San Diego State University and USGS. The model inputs include the percentage of impervious land coverage within each land use category, runoff coefficients for pervious and impervious areas, and average annual rainfall for each drainage basin. Percent impervious values are based upon literature values and URS Greiner Woodward Clyde data from studies throughout the country. For this analysis, pervious areas are assumed to have a runoff coefficient of 0.20 and impervious areas 0.95. Long-term isohyets are used to estimate average annual rainfall for each drainage basin.

The data for San Diego County land use, annual rainfall distribution, and drainage areas contributing to runoff to the mass loading stations are geographically linked together in an ArcView® GIS. The total loading values for each mass loading station are determined by multiplying the contributing runoff volume from each land use type by the various input constituent land use based EMCs. This total load is divided by the estimated total runoff volume. The runoff volume for each land use type is found by factoring in average annual rainfall, total acreage and percent of precipitation that would be expected to run off the land for each land use type.

<u>Model Uncertainties</u>
It should be noted that there is uncertainty in the EMCs estimated for a study area. Uncertainty arises from the following sources:

- The annual flow volumes are calculated as a product of annual precipitation depths within a given drainage basin (estimated from historical rainfall data) and the estimated area of that basin.
- The annual flow volume from a particular land use is estimated by broadly applying runoff coefficients based on general assumptions regarding ground surface imperviousness. Infiltration and evaporation are not considered explicitly.
- The percentage of each land use within a given basin is derived using the 2000 SANDAG GIS database, or more recent version. This database has some degree of uncertainty in terms of its currency and accuracy.
- The pollutant EMCs are calculated from a limited number of water quality samples. Statistically, the degree of uncertainty in the mean decreases with increased sample sizes.
- The pollutant EMCs are calculated from measured concentrations that are subject to laboratory and field errors, as well as inherent variability in storm water quality.
- Data collected during the land use monitoring period may not be representative of other years or long-term trends. There is considerable debate as to whether water quality varies significantly between wet and dry years.

The pollutant EMCs calculated for each land use and input into the model are computed from measured storm events assuming that the constituent concentrations are solely dependent on the land use characteristics of a given basin and that runoff from similar land uses throughout the study area have the same water quality.

4.8.2.2 Dynamic Storm Water Modeling

The previous model discussed is a static model that estimates the typical average loads and concentrations. Average annual rainfall, runoff coefficients, and typical concentrations from various land uses are used as inputs. While the static model estimates what values might be expected over the course of a year, it does not show variability between storms based on rain events or changes in water quantity and quality during a storm. Dynamic modeling addresses these issues.

The EPA's Storm Water Management Model (SWMM) is used to explore the potential of modeling the San Diego County watersheds to assist the Urban Runoff Monitoring Program. SWMM is a dynamic rainfall-runoff simulation model, primarily but not exclusively for urban areas, for single-event or long-term (continuous) simulation. The Rain and Runoff modules of SWMM are most appropriate to model the watersheds of San Diego County.

All twelve watersheds associated with the mass loading stations are modeled. Each watershed can be broken down into one or more sub-catchments. Sub-catchment data such as watershed area, width, impervious area, and average slope is calculated and input into the model with the aid of GIS. Rain gages at many mass loading stations provide five minute interval data available as input into the model.

The San Diego River watershed serves as a good example to explore the results, potential, and limitations of SWMM. A USGS gauging station exists near the mass loading station so the modeled flow can be compared to both the mass loading station monitored flow and the USGS calculated flow. For modeling purposes, the San Diego River watershed is divided into the upper and lower sub-catchments. Necessary data are input into the model, including impervious percent estimated to be higher in the more urban lower sub-catchment.

Figure 4-2 shows that there is a close correlation between USGS and mass loading station flow data as illustrated by the February 11, 2003 storm data. Therefore, the USGS data seem to provide an accurate account of the flow at the mass loading station for modeling purposes.

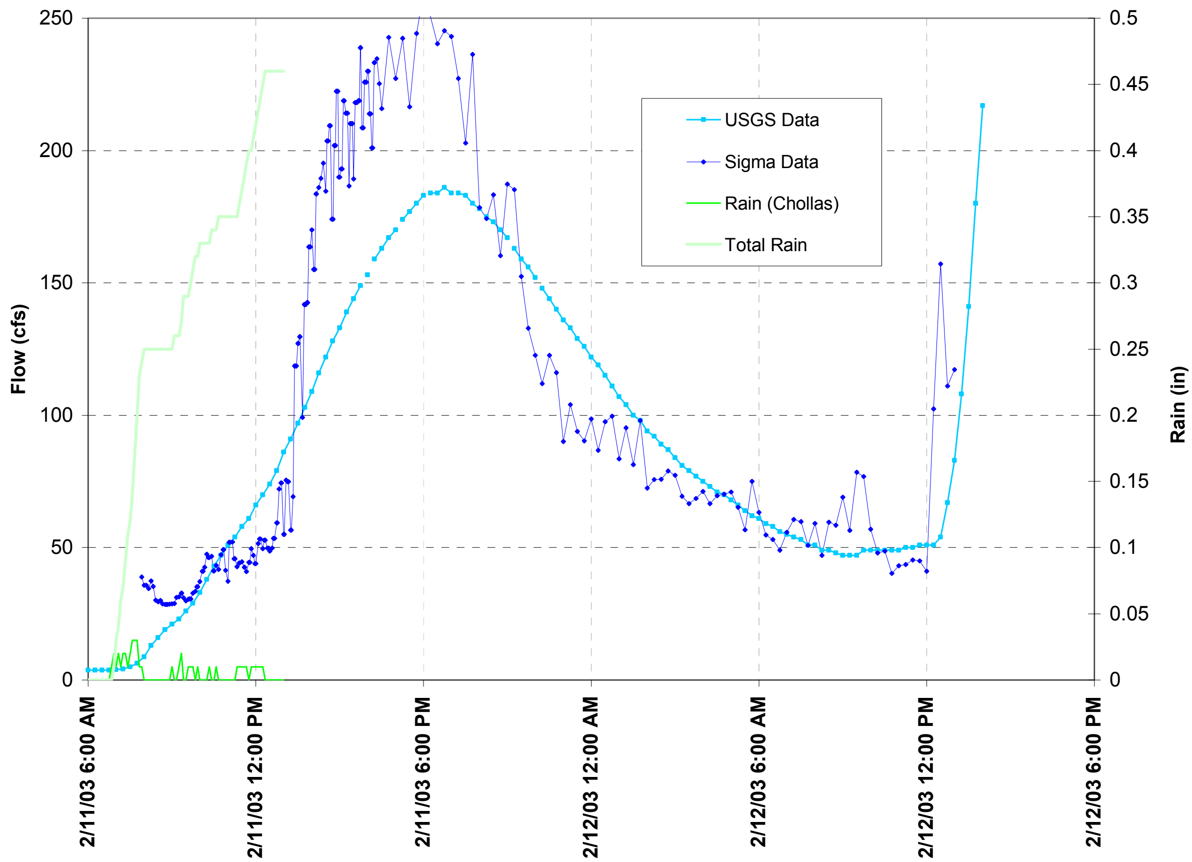


Figure 4-2. USGS and MEC (Sigma) monitored flow at San Diego River.

The USGS flow data are then used to calibrate and verify the modeled flow data. Sub-catchment parameters are adjusted to calibrate the model. The model uses the rain gage data to predict the resultant hydrograph. The modeled hydrograph in the flow section of Figure 4-3 compares favorably with the USGS hydrograph. The peak flow and time of peak flow are very similar.

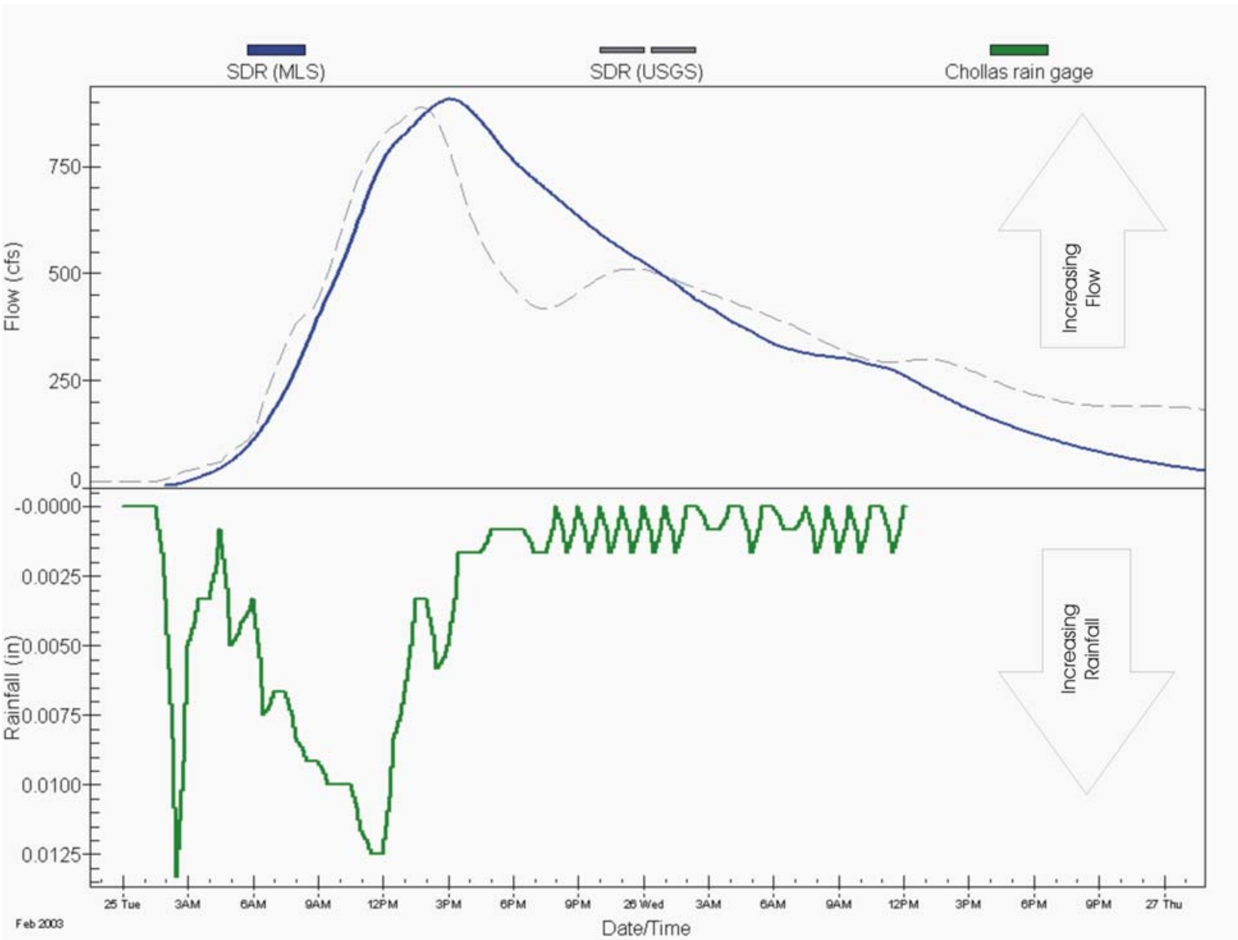


Figure 4-3. Actual rainfall and hydrograph (USGS) plotted with modeled hydrograph (MLS).

Modeling water quality is much more complex than modeling water quantity. There are large uncertainties in the physical, chemical, and biological processes that contribute to pollutant runoff. There are two main categories of water quality modeling available in SWMM, one physically based and the other empirically based. The physically based model involves gathering information on the “buildup” and “washoff” mechanisms of pollutants. By calibrating with historical data, the empirically based model attempts to predict contaminant concentrations based on their relation to the amount of flow. Some contaminant concentrations may typically be highest during low flows while other types of contaminants will tend to be higher during the high flow periods. Because of the complexity and lack of data in accurately modeling the physical processes, the empirical method was chosen to model San Diego River and for this framework document.

Water Quality Rating Curves

The empirical water quality method involves developing a rating curve for each constituent at a monitoring location. The curve shows the relation between concentration or load and flow. Six storm events over the past two storm seasons are used to fit rating curves based on the following equation:

$$C=K \cdot Q^X$$

Where C = constituent concentration at time, t.
Q= flow at time, t
K= coefficient that includes correct units for conversion
X=exponent

The EMC for each storm event is plotted against the average flow during the storm. The equations of the rating curves fit to historic data give estimates of K and X to be used in water quality modeling. A constant concentration would have a value of zero. Constituents that behave like particulates typically have X values greater than zero while dissolved constituents tend to have X values less than zero. The coefficient, K, can be used to calibrate the modeled concentration values to more accurately estimate real data. Figure 4-4 shows the San Diego River rating curves for ammonia (red), total suspended solids (blue), and total copper (green).

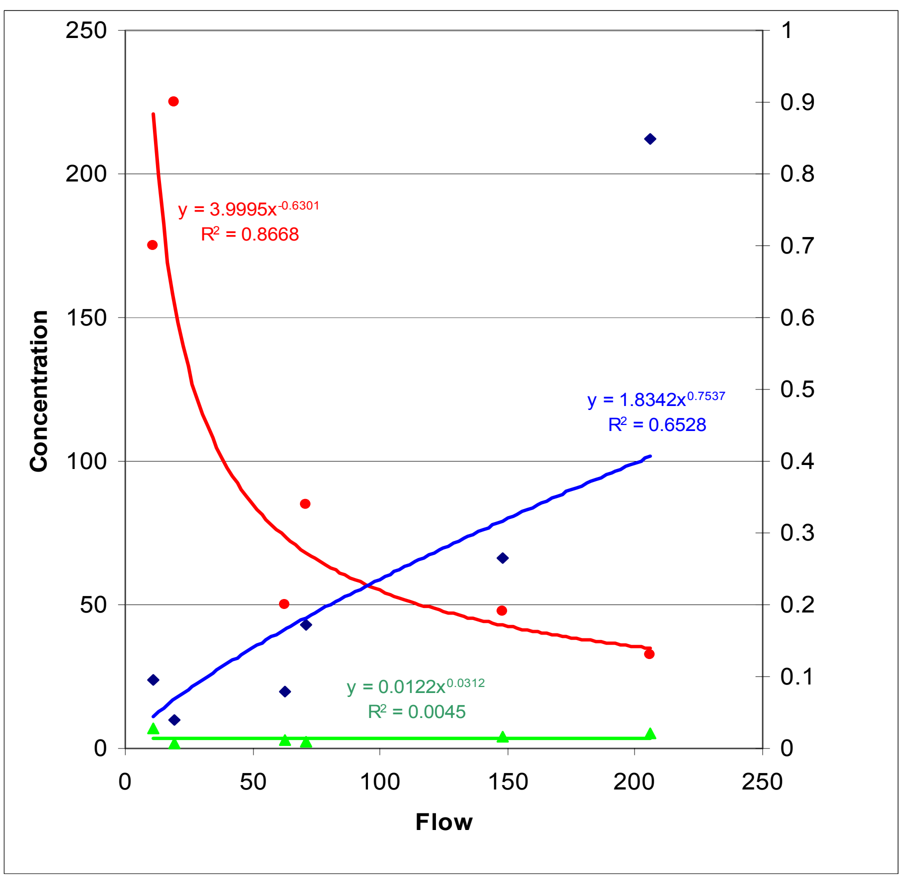


Figure 4-4. Water quality rating curves.

Water Quality Model Results
Using the rating curve approach, water quality of the San Diego River is presented as modeled for the February 25, 2003 storm event. Figure 4-5 shows how the concentrations of constituents may have varied during the storm. The modeled TSS values reached their highest values at peak flow while ammonia values were estimated to be lowest during peak flow. Total copper values remained relatively constant during the modeled storm. To more accurately verify the modeled water quality results, it would be necessary to analyze real water quality data at several times during a storm and compare this to modeled data.

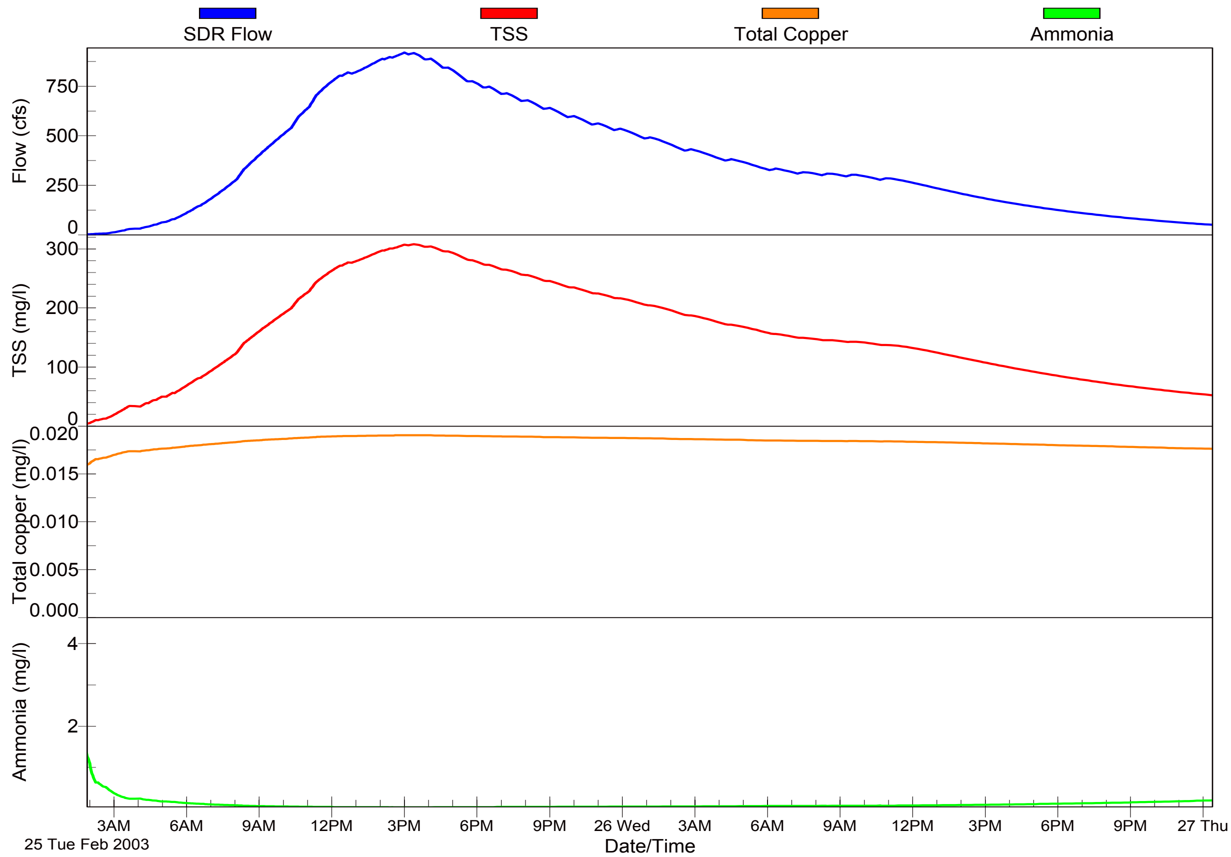
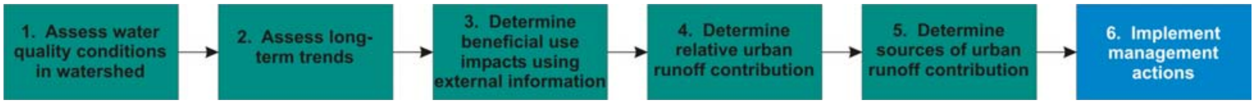


Figure 4-5. Modeled water quality results for San Diego River.

The SWMM model shows promise for estimating flow and water quality in the San Diego watersheds. More data are needed for calibration to reduce uncertainty, particularly in regards to water quality. A model calibration step can be performed by collecting and analyzing discrete water samples at various times during a storm. This would also provide valuable information to improve the model. If the model can be shown to accurately and consistently predict water quality at monitored stations, it is possible that the requirement for intensive monitoring could be reduced. The model may also be used to predict pollutant reductions when various changes to land use and other inputs are known.

4.9 Core Management Questions – Gaining Understanding of Watershed Conditions



The watershed assessment process described in the previous sections provides the ability to define activities or actions within the context of the urban runoff management plan targeted at improving water quality. To accomplish this, water quality issues are prioritized in order to develop short or long term actions. The goal of this watershed assessment is to support the Copermittees' efforts to plan and manage the urban runoff program in an effective manner. Defining actions to reduce or investigate the COC or water quality problems in the watershed is a collaborative process involving stakeholders.

The watershed program should periodically assess the core management questions to see if they are aligned with the framework. Any major discrepancies or deviations would indicate that an update to this framework is needed to keep it current. The flow diagram in Figure 4-6 shows the approach to period revisions presented in the SMC Model Document.

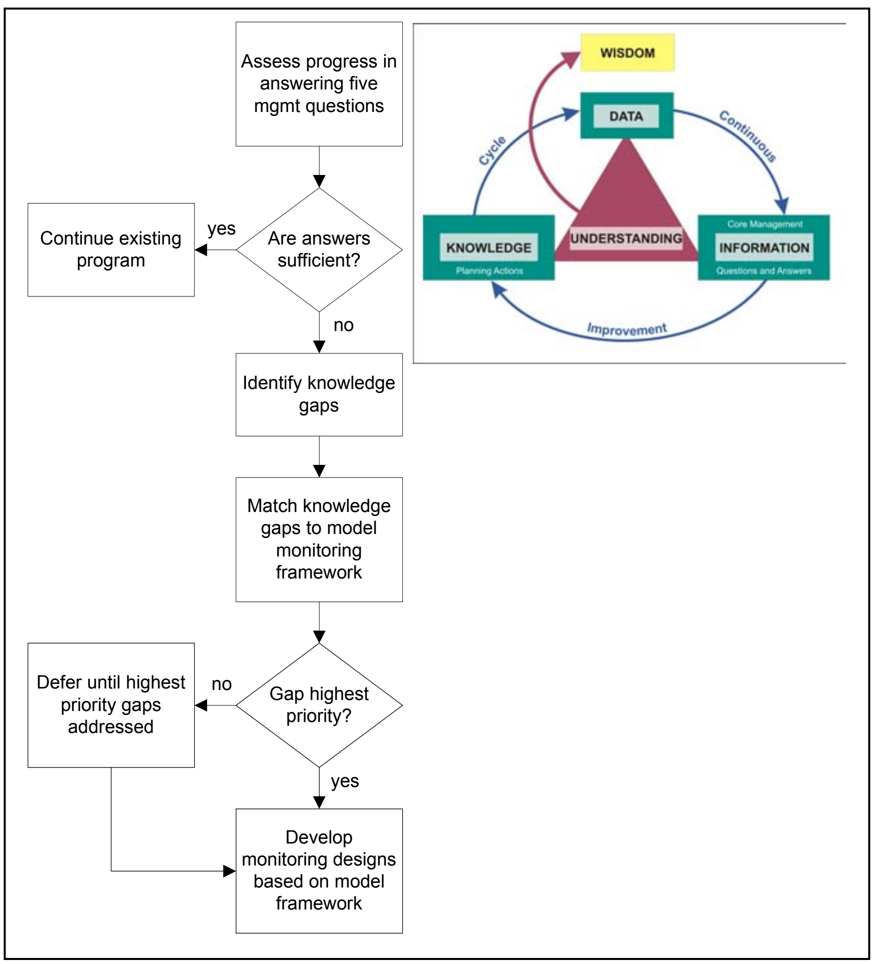


Figure 4-6. Long-term core management questions and strategic planning flow chart. (SMC Model Monitoring Committee - Figure 3)

As presented in Section 1.0 and Figure 1-1, the data to wisdom model presented (inset in Figure 4-6) has an “understanding” phase developed from the data-information-knowledge cornerstones and the continuous improvement cycle. The continuous improvement cycle is created by moving through those three phases repeatedly. In relation to the long-term core management question and strategic planning flow chart presented above the program should strive to identify new questions (knowledge gaps) and monitoring program adjustments to gain new understanding of the conditions as the program evolves.

5.0 DATA ASSESSMENT TOOLS

The first step in data assessment is to frame the questions you want to answer with the data. Once the questions are established and the data is collected then the correct statistical tool can be selected.

5.1 Quantitative – Statistical Tools

5.1.1 Data transformations

To correctly apply statistics to data, the data need to meet the assumptions of each statistical tool. The assumptions vary according to the statistical tool, but most typical is the assumption of normally distributed data. If data are not normally distributed, transformations can be applied that may provide a normal distribution. Commonly used transformations are log, square root, and arcsine. Values that vary by orders of magnitude, such as bacterial concentrations, are typically log (base 10) transformed. Other biological data are usually log or square root transformed. Data values expressed as percents are transformed by the arcsine transformation.

5.1.2 T-Test

<u>When is the tool applicable?</u>
To test for differences between two sets of data, e.g. between two locations or two time periods.

<u>What does the tool tell you?</u>
Whether two sets of data are different (two-tailed test) or whether one set is lower than the other (one-tailed test).

<u>How many data points are needed for proper application?</u>
Approximately equal numbers of data points in each set. Ten observations or more in each data set is desired, but the tool can be performed with fewer.

<u>Important considerations and proper use of the tool</u>
Data need to be normally distributed or transformed to a normal distribution. If the variances in the two sets are not equal, the t-test should be performed using the pooled variance.

<u>Example</u>
In the example shown below, the effect of benz(a)anthracene on the test clam species at Site 4 was compared to the experiment control. The Pr > |t| was greater than 0.05. Therefore, we concluded that there is no statistically significant difference between site 4 and the control.

POLA 212-215 Review 09:50 Monday, February 23, 2004 58
T-test Results, This is a 2-tailed result

-------------------- SPECIES=CLAM ANALYTE=Benz(a)anthracene group=SITE04_ --------------------

The TTEST Procedure

Statistics

Variable	Station	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err
numres	REF	5	0.9479	2.5	4.0521	0.7489	1.25	3.5919	0.559
numres	SITE04_	5	-0.175	10.9	21.975	5.3437	8.9191	25.629	3.9887
numres	Diff (1-2)		-17.69	-8.4	0.8879	4.3016	6.3684	12.2	4.0277

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
numres	Pooled	Equal	8	-2.09	0.0705
numres	Satterthwaite	Unequal	4.16	-2.09	0.1027

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
numres	Folded F	4	4	50.91	0.0022

The p-value is the two-tailed probaility using the t distribution. It is the probability of a t-value of equal or greater absolute value for the null hypothesis. For a one tailed test, we would half this probability. If the p-value is <0.05(the alpha), we will conclude that the hypothesis is significantly different from zero. In the example, the p-value for the difference between the two samples is greater than 0.05, so we conclude that the difference in means is not statistically significantly different from 0.

This is the two-tailed signifance probablity
In the example, the probability is less than 0.05
So there is evidence that the variance of the two samples are different. Therefore, we may want to use the second method (Satterthwaite variance estimator) in the t-test.

5.1.3 ANOVA with Multiple Range Test

When is the tool applicable?
When one wants to test for differences between three or more sets of data, e.g. between locations or time periods.

What does the tool tell you?
The ANOVA tells you whether any of the datasets are different from any other dataset. If a significant difference is indicated, a multiple range test can then be run to determine which sets are different from each other, or which sets are different from a control set.

How many data points are needed for proper application?
Approximately equal numbers of data points in each set. Ten observations or more in each data set is desired, but can be performed with fewer.

Important considerations and proper use of the tool
Data should be normally distributed or transformed to a normal distribution. Always look at the results of the ANOVA to accept or reject the hypothesis that the datasets are not significantly different. If the hypothesis is rejected (i.e., the datasets are different), a multiple range test such as a Tukey or Dunnett test can be applied. The Tukey test is used to determine which of several groups is different from another of the groups and can handle moderate variation in the number of data points in each dataset. The Dunnett test is typically used when one desires to know which of several groups has a larger (or smaller) mean than a reference or control group.

Example
In the example below, ANOVA was performed on the total phosphorus results from ten copermittee urban runoff monitoring stations. Stations that have mean results that are not significantly different from one another are grouped by color ranges. The total phosphorus mean result from Station TJR *is* significantly different from all other stations while the mean from station TC *is not* significantly different from Stations AH, CC, SDR, EC, SMR, and PC.

Analyte	Prob > F	TJR	TC	AH	CC	SDR	EC	SMR	PC	SR	SLR	SDC
Total Phosphorus (mg/l)	0.0001	2.26	1.16	0.95	0.70	0.64	0.61	0.58	0.57	0.30	0.25	0.11

5.1.4 Regression Analysis (and Outlier Test)

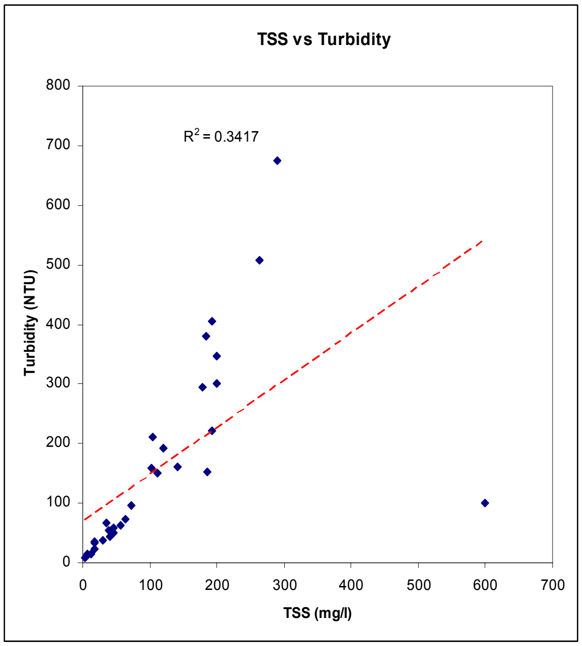
When is the tool applicable?
To look for relationships between two parameters (correlations) or for changes with time (trend analysis).

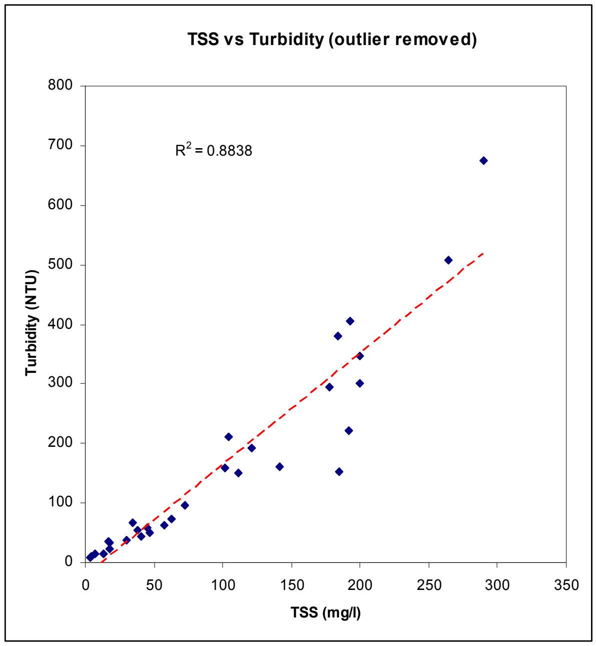
What does the tool tell you?
Whether the relationship between the two parameters has a slope that is not zero (horizontal line) and how good the fit is to the line (R^2).

How many data points are needed for proper application?
Minimum of three data points; greater than 5 are preferable.

Important considerations and proper use of the tool
Outliers in the data can be highly influential in regression analysis. Scatter plots of the data should always be used in conjunction with a regression analysis to determine if the regression equation is appropriate for the relationship and to check for potential influential values. Examine plots of the regression data and use regression diagnostic tools to evaluate the residuals (difference between actual value and that predicted by the regression); if outliers are reasonably suspected use an outlier test to determine if they are in fact outliers.

Example
The examples below show regression analysis that was used to look for a correlation between total suspended solid and turbidity results. The example on the left shows a scatter plot with a weak correlation (R^2=0.3417). Review of the scatter plot shows there is one point (TSS = 600 mg/l) which is out of line with all of the other data. The first thing to do is to review the data to make sure the value is not an error upon data entry. If there is no reason to correct the data point, use an outlier test to make sure that the point is actually an outlier and not just an extreme value. In this case the data point was an outlier, the scatter plot on the right has the outlier removed and shows a stronger correlation (R^2=0.8838).





5.1.5 Multiple Regression Analysis

<u>When is the tool applicable?</u>
To look for a relationship between a response variable and a combination of other parameters (predictors) that could be influencing a response.

<u>What does the tool tell you?</u>
Which parameters contribute to the observed concentrations of the response variable and to what extent each one contributes to the observed response.

<u>How many data points are needed for proper application?</u>
Minimum of three data points, number of data points must be at least two larger than the number of predictors used in the model.

<u>Important considerations and proper use of the tool</u>
Outliers in the data can be highly influential in regression analysis. Examine plots of the regression data for each predictor; if outliers are reasonably suspected use an outlier test to determine.

<u>Example</u>
If we suspected dissolved copper and diazinon affected amphipod mortality, we could use multiple regression analysis to determine the relationship. The multiple regression calculates the relationships between amphipod mortality and each of the predictor variables as well as the combined relationship with both predictors to determine which factors have a significant relationship with mortality. This multiple regression could result in one or both predictors showing a significant relationship. A linear multiple regression relationship combining both predictors would be of the form

$$Y_{am} = Int + a \bullet X_c + b \bullet X_d$$

Where

Y_{am}	amphipod mortality
Int	Intercept
a	contribution of copper to mortality
X_c	copper concentration
b	contribution of diazinon to mortality
X_d	diazinon concentration.

The output of the multiple regression shows the coefficients of the predictor variables, the contribution of each to the model R^2, and the significance of each variable as well as the overall model significance.

5.1.6 Confidence Interval Test

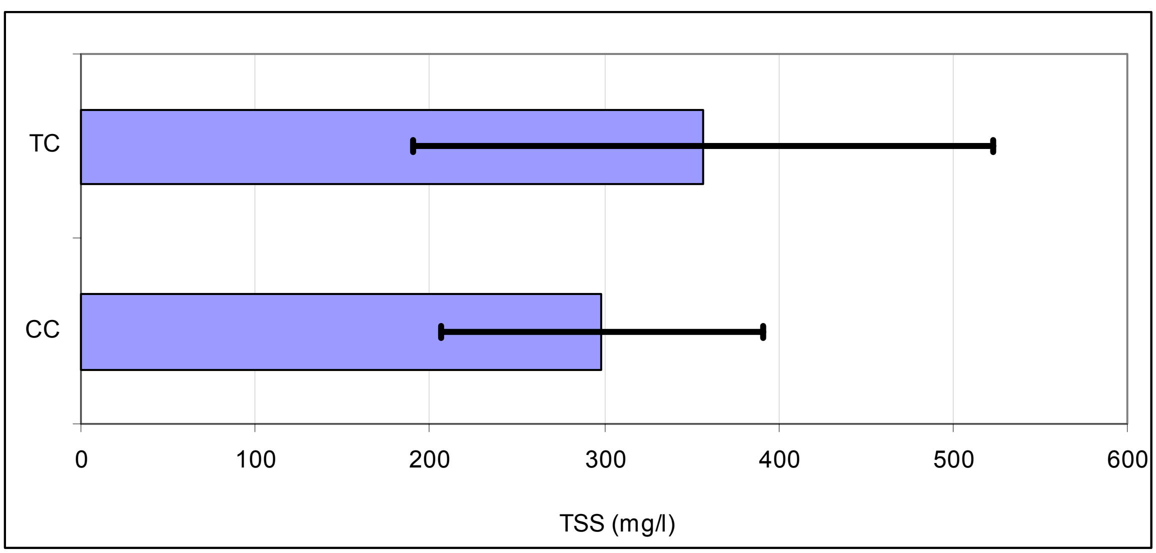
<u>When is the tool applicable?</u>
When one wants to determine the precision with which a new mean value falls within the confidence bounds of a known set of data.

<u>What does the tool tell you?</u>
If the mean of a new set of values is within a set confidence interval of another (e.g., reference) set of numbers.

<u>How many data points are needed for proper application?</u>
A minimum of two data points to compute a mean. The confidence interval on the reference set becomes smaller as the number of observations increases, a minimum of 10 observations for these dataset would be desirable.

<u>Important considerations and proper use of the tool</u>
Appropriate data transformations should be applied prior to calculating confidence intervals and means for comparison.

<u>Example</u>
In the example below, the mean of Chollas Creek (CC), represented by the blue bar, falls within the 95% confidence interval of the mean of Tecolote Creek (TC).



5.1.7 Predictive Interval Test

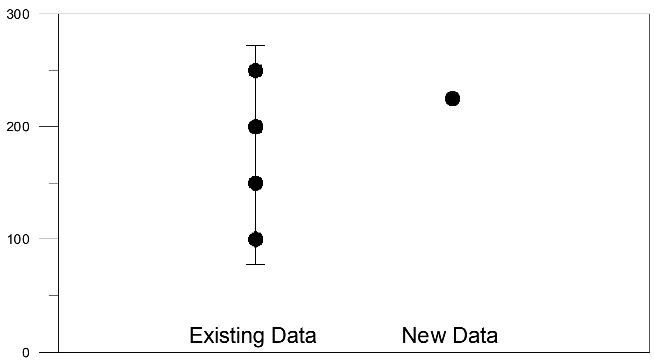
When is the tool applicable?
This tool is similar to the confidence interval, but this calculation provides the precision for a single new data point compared to a known set of data.

What does the tool tell you?
If the new value is within a set predictive interval of another (e.g., reference) set of numbers.

How many data points are needed for proper application?
One data point is needed for the new value. The confidence interval on the reference set becomes smaller as the number of observations increases, a minimum of 10 observations for these dataset would be desirable.

Important considerations and proper use of the tool
Appropriate data transformations should be applied prior to calculating confidence intervals comparing the new value to the interval.

Example



The predictive interval and individual data values are shown on the left of the example, the new data value is shown on the right. In this example the new data falls within the predictive interval of the existing data.

5.1.8 Power Analysis

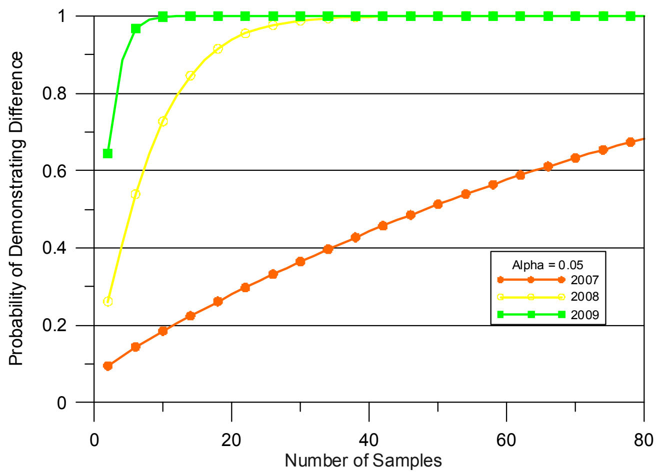
When is the tool applicable?
To determine the number of samples necessary to statistically see a pre-determined difference between two sets of data. Alternatively, to know what the power of a statistical test is based on the number of samples available.

What does the tool tell you?
Power estimates tell you what the chance is of failing to detect a false hypothesis (i.e., saying two sets of data are similar when they are not) and how many samples will be necessary to avoid this problem. Typically, power estimates of 80% or greater (at an alpha level [error] of 0.05) are desired to be able to make strong statements about statistical results.

How many data points are needed for proper application?
Ideally enough data points to provide a good estimate of the variation in the data. This test can also be used with hypothetical values to get estimates based on *a priori* assumptions about the data.

Important considerations and proper use of the tool
There are several ways to perform a power analysis. The proper test is determined by the statistical method that is to be used in the comparison of the datasets.

Example



This example shows the number of samples needed to demonstrate a difference between current concentrations and target concentrations in each of three years with the power for each shown on the y-axis. In this example, the target concentrations are lower each year, resulting in larger differences, therefore the power is stronger with fewer samples in the later years.

5.1.9 Outlier test

When is the tool applicable?
This tool is applicable when one is reasonably certain that one or more data points appear to be unduly influencing the results.

What does the tool tell you?
The results can suggest that a suspect data point is truly an outlier or only an extreme value.

How many data points are needed for proper application?
Enough data points to be reasonably certain a single data point is more than just an extreme value.

Important considerations and proper use of the tool
Suspected outliers should be investigated first to determine if an error occurred when processing the sample or a data entry error occurred. In environmental data values can vary widely because of natural causes and thus should be considered extreme values and not true outliers. If after careful consideration, an outlier is still suspected, apply a conservative outlier test and exclude the data point if the results suggest it is an outlier.

Example
As shown in Section 5.1.4.

5.2 Semi-Quantitative – Statistical Tools

5.2.1 Cluster Analysis

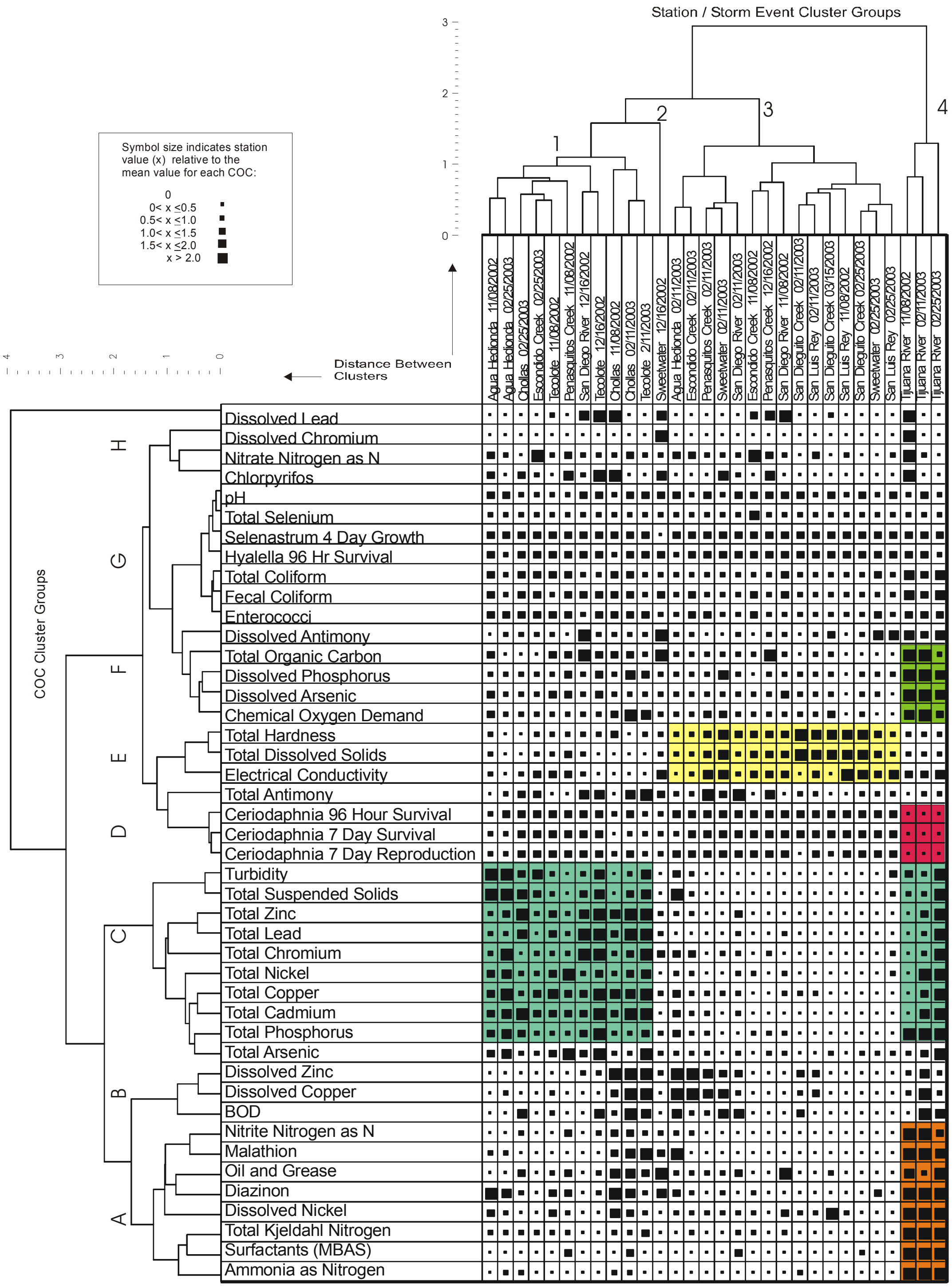
When is the tool applicable?
Cluster analysis is used when one wants to determine, for example, which of many sites and/or times have similar patterns of concentrations of species or contaminants.

What does the tool tell you?
The results group the biological or chemical parameters in a hierarchical classification scheme on a dissimilarity scale. The larger the distance between two parameters, the more different they are in relative concentration and their distribution among sites and/or times. Likewise, the sites and times are grouped in their own hierarchical classification based on the co-occurrence of parameters found at them.

How many data points are needed for proper application?
A minimum of one data point is needed for each parameter at each site and/or time.

Important considerations and proper use of the tool
A full matrix of data is necessary to achieve an accurate end product. If a small number of data points are missing for a parameter that must be included, use an estimate for the values based on best professional judgment.

Example
The cluster represented here groups the COC into eight Cluster Groups (A-H) and the Station/Storm Events into four cluster groups (1-4) as shown by the letters next to the dendrograms. Combining the dendrograms with a table of relative concentration can explain why certain COC or sites/times group together. For instance, Station/Storm Event Cluster Group 4 is mainly differentiated from the other three groups by the high concentrations of the COC in Cluster Groups A, D, and F while Station/Storm Event Cluster Group 3 is set apart by COC Cluster Group E. The cluster analysis will give an indication of which COC are common to which Stations and/or Storm Events.



5.2.2 Threshold comparisons

When is the tool applicable?

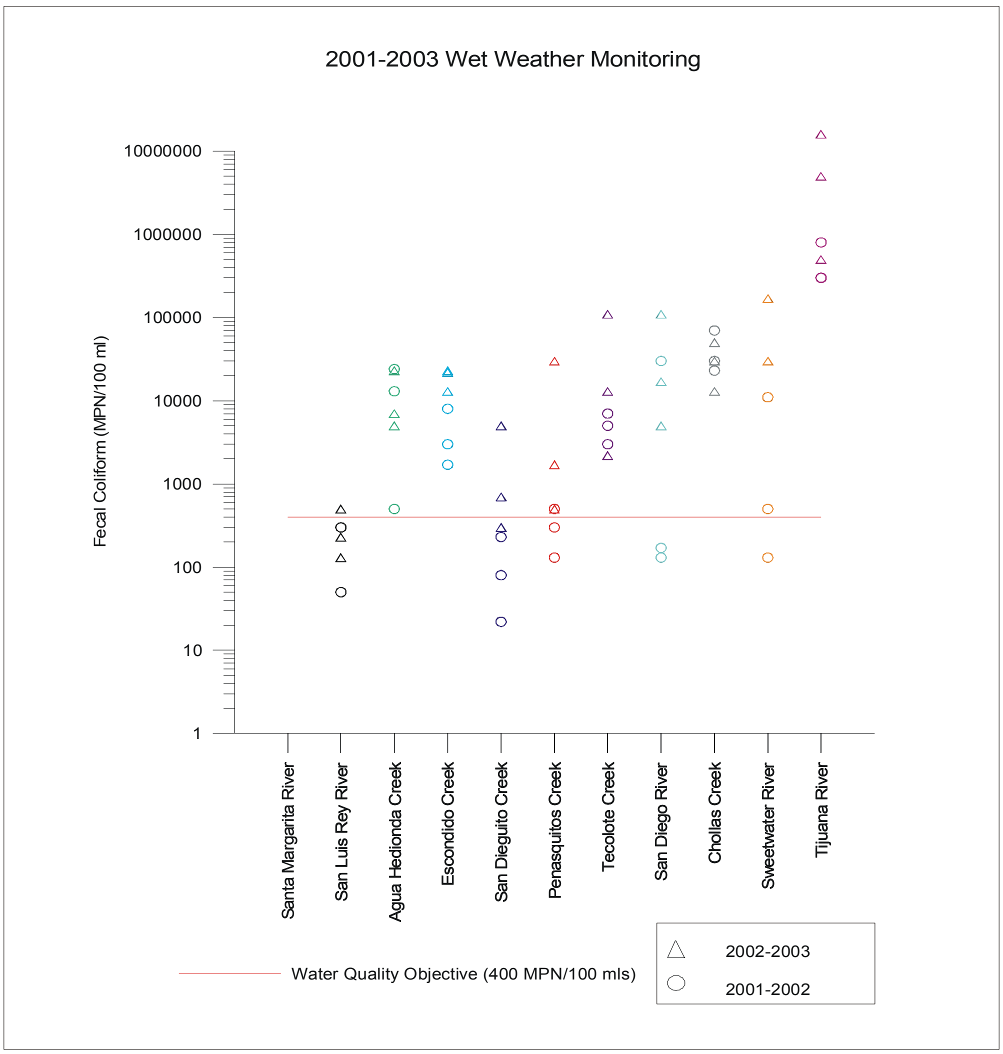
This tool is used when there are established threshold values for a particular parameter and the frequency of values above (or below) the threshold is of interest.

<u>What does the tool tell you?</u>
The frequency of data exceeding threshold values can indicate sites or times that are indicative of problems that may need further investigation.

<u>How many data points are needed for proper application?</u>
More is better in this case as environmental data can vary widely. A larger number of values will give more confidence in any conclusions made from the results.

<u>Important considerations and proper use of the tool</u>
Make sure you have enough data points to support your conclusions. Check data for detection limit problems prior to using threshold comparisons.

<u>Example</u>
The example below shows fecal coliform concentration results from the 2001-2003 wet weather monitoring season. The red line indicates the water quality objective with a threshold value of 400 MPN. All sample from Agua Hedionda, Escondido Creek, Tecolote Creek, Chollas Creek, and Tijuana River exceeded this threshold.



5.2.3 Ranking

When is the tool applicable?
Ranking the data is applicable when parametric statistical tests are not applicable because the data do not meet the assumptions of the test. Ranking is also used in non-parametrical statistical testing where no replicate observations are available. Ranking of parameter values by site and/or time can also quickly turn numerical data into integer numbers that are more rapidly assessed by visual observation. Ranking is also used assign numerical values to categorical data such as “3, 2, 1” for “high, medium, low”.

What does the tool tell you?
Ranking can show you which sites and/or times are consistently higher or lower when compared to all other sites and/or times. Ranked data can be used in non-parametric tests to statistically determine differences among groups of samples.

How many data points are needed for proper application?
As with the other statistical tools, the more data available will give more accurate results. Generally a minimum of ten observations per category are desirable.

Important considerations and proper use of the tool
Ranking data observations can give you an idea of patterns in the data. One must take care to compare like data with ranks and if applying ranks to categorical data, make sure that all categories found over the sites and/or times are ranked consistently.

Example
In the example below, several sediment sites were measured for fine grain percentage, TOC, and other parameters. The sites were given a ranking for fines and TOC. The highest combined rank was used to select three sites for further analysis.

Sampling Site	TOC and Grain Size Distribution in Phase I								Ranking for Phase II				
	Gravel (%)	Sand (%)	Silt (%)	Clay (%)	Median (µm)	Mean (µm)	Fines (%)	TOC (%)	Fines Rank	TOC Rank	Rank Sum	Highest Rank	Phase II
SLE 1L-1	0.01	79.7	11.7	8.57	365	105	20.3	0.64	3	5	8		
SLE 1M-1	0.29	88.4	6.76	4.55	166	185	11.3	0.33	1	2	3		
SLE 1R-1	0.00	58.3	27.9	13.8	95.2	47.0	41.7	1.39	7	7	14	*	Yes
SLE 2L-1	0.14	82.7	10.1	7.12	252	148	17.2	0.30	2	1	3		
SLE 2M-1	0.06	28.1	40.8	31.0	17.5	15.7	71.8	1.44	9	8	17	*	Yes
SLE 2R-1	0.08	65.7	19.2	15.0	176	45.8	34.3	1.04	4	6	10		
SLE 3L-1	0.60	60.1	25.5	13.8	104	54.6	39.3	0.39	5	4	9		
SLE 3M-1	0.47	57.9	30.9	10.8	104	77.0	41.7	0.37	6	3	9		
SLE 3R-1	1.65	45.5	34.1	18.8	53.4	30.9	52.8	1.67	8	9	17	*	Yes
Mean of all Sites	0.37	62.9	23.0	13.7	148	78.8	36.7	0.84					

5.2.4 Maps and graphics

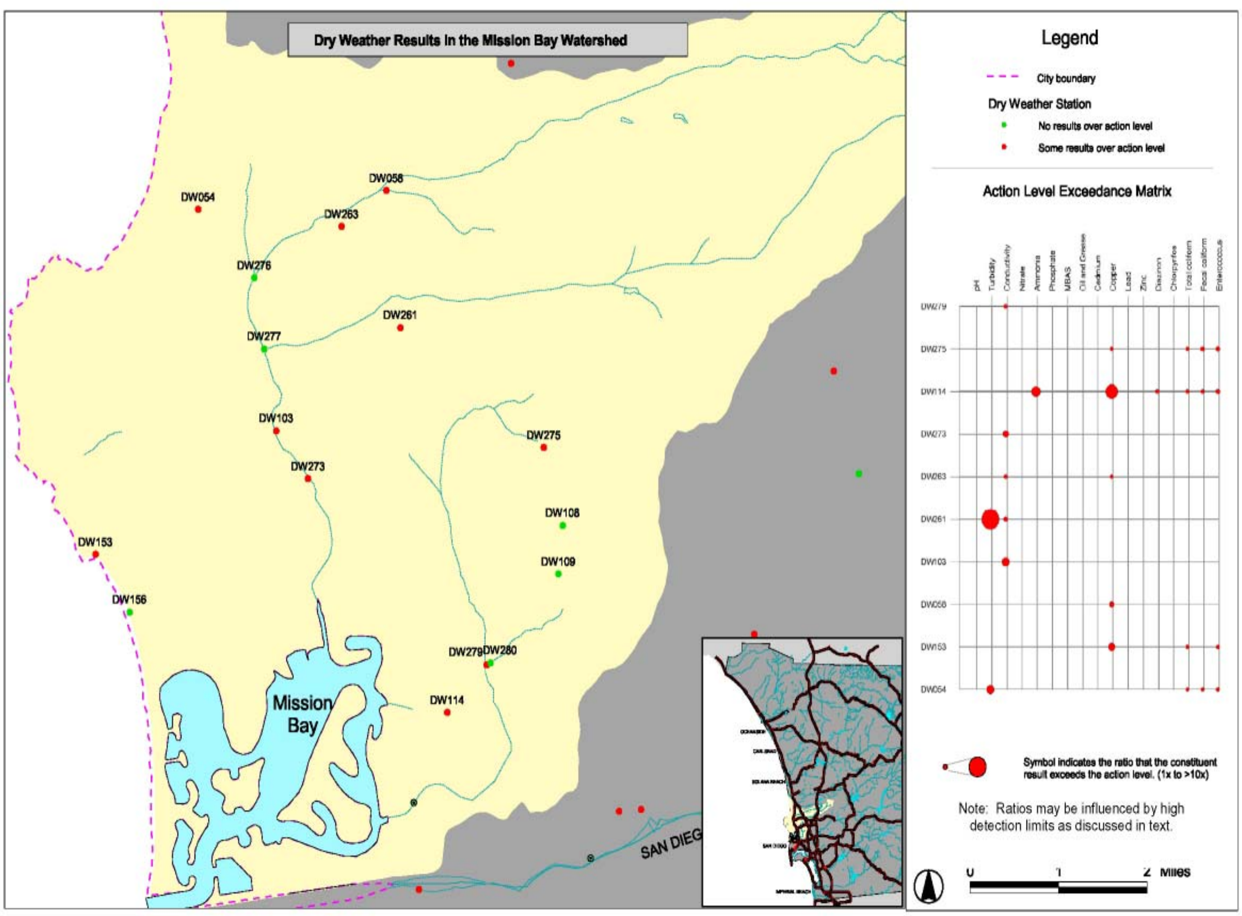
When is the tool applicable?
Graphics are useful tools in just about every data analysis. They provide a visual picture of the data and their inter-relationships that are hard to picture from a list of numbers. Mapping of data points is valuable when geographic patterns are part of the question.

<u>What does the tool tell you?</u>
Graphing data can show correlation between parameters, temporal trends, and differences or similarities between sites and/or times. Presentation of results from virtually every statistical tool is enhanced by the use of graphics or maps. With the use of GIS, maps can show patterns related to elevation, geographic features, land use, and other parameters.

<u>How many data points are needed for proper application?</u>
Any number of data points can be graphed or mapped, meaningful graphics are best produced with a minimum of ten data points.

<u>Important considerations and proper use of the tool</u>
Use of graphics to portray results of analyses with statistical tools should use axes scaled to the same transformation as used in the analyses.

<u>Example</u>
The map below shows where dry weather monitoring results have exceeded action levels. The geographic nature of this display can be useful in tracking down the source of pollutants.



5.2.5 Weight-of-evidence

When is the tool applicable?
Weight-of-evidence is useful when data are limited or inconclusive. Obtaining similar, but perhaps non-significant because of limited data, patterns for many parameters can indicate potential future significant results.

What does the tool tell you?
This approach can give one indications of potential trouble locations or indicate where further sampling may be best placed to provide additional data that may enable a quantitative data analysis.

How many data points are needed for proper application?
Because this approach can be applied when the number of data points are limited, it is advantageous to have the data cover as many parameters as possible.

Important considerations and proper use of the tool
Limited data can lead to misleading results. Environmental data are variable, if data used in the analysis are temporally limited; one may only be observing a temporal peak or valley and conclusions may not be valid.

5.3 Qualitative – Assessment Tools

5.3.1 Visual Observations

Visual observations whether documented as part of analytical samples or independent of other sample collection efforts can provide additional information that may strengthen a quantitative or semi-quantitative analysis.

5.3.2 Photographic Documentation

Photographic documentation can provide evidence of physical conditions at a specific time and place. To be most useful, photographs should be documented as to date and location with time stamps and recognizable landmarks.

5.3.3 Third Party Monitoring

Third party monitoring (citizen groups, universities, etc.) can provide additional data that may strengthen a quantitative or semi-quantitative analysis. Data from these sources must be screened as described previously for chemical and biological (or benthic) data to ensure that they meet minimum specifications and are comparable in quality prior to use.

6.0 REFERENCES

Quality Assurance and Quality Control.

CSWRCB/California State Water Resources Control Board. 2002. Quality Assurance Management Plan for the State of California's Surface Water Ambient Monitoring Program (SWAMP). Available Online: http://www.swrcb.ca.gov/swamp/qapp.html

CALTRANS/California Department of Transportation. 2000. Guidance Manual: Stormwater Monitoring Protocols (Second Edition). CTSW-RT-00-005. July, 2000.

USEPA/United States Environmental Protection Agency. 1999. USEPA Contract Laboratory Program: National Guidelines for Organic Data Review. USEPA Office of Emergency and Remedial Response. EPA540/R-99/008. October, 1999.

USEPA/United States Environmental Protection Agency. 1994. Guidance for the Data Quality Objectives Process. USEPA Office of Research and Development. EPA-600-R-96-055. September 1994.

USEPA/United States Environmental Protection Agency. 1995. Guidance on the Documentation and Evaluation of Trace Metals Data Collected for Clean Water Act Compliance Monitoring. USEPA Office of Water. EPA 821-B-95-002. April 1995.

USEPA/United States Environmental Protection Agency. 1999. Methods and Guidance for Analysis of Water on CD. USEPA Office of Water. EPA 821-C-99-004. June 1999.

Basic Statistical references.

Carr, R.S., Chapman, D.C., Howard, C.L., and J.M. Biedenbach. 1996. Sediment quality triad assessment survey of the Galveston Bay, Texas system. Ecotoxicology 5, 341-364.

Chapman, P.M. 1996. Presentation and interpretation of sediment quality triad data. Ecotoxicology 5, 327-339.

Dixon, W.J. and F.J. Massey. 1982. *Introduction to Statistical Analysis*. McGraw-Hill Science/Engineering/Math, 672 p.

Hyland, J.L., Van Dolah, R.F., and T.R. Snoots. 1999. Predicting stress in benthic communities of southestern U.S. estuaries in relation to chemical contamination of sediments. Environmental Toxicology and Chemistry 18(11), 2557-2564.

Long, E.R., Scott, G.I., Kucklick, J., Fulton, M., Thompson, B., Carr, R.S., Scott, K.J., Thursby, G.B., Chandler, G.T., Andreson, J.W., and G.M. Sloane. 1995. Final Report. Magnitude and extent of sediment toxicity in selected estuaries of South Carolina and Georgia. NOAA Technical Memorandum NOS ORCA: 178 p.

Sokal, R.R. and F.J. Rohlf. 1969. *Biometry*. W.H. Freeman and Company. 776 p.

StatSoft, Inc. 2004. Electronic Statistics Textbook. Tulsa, OK: StatSoft. Available Online: http://www.statsoft.com/textbook/stathome.html.

Zar, J.H. 1974. *Biostatistical Analysis*. Prentice-Hall, Inc. 620 p.

Other references

Ackoff, R.L. 1989. From Data to Wisdom. Journal of Applied Systems Analysis. Vol.; 16. pp.3-9.

Bellinger, G., Castro, D. and Mills, A. 2004. Data, Information, Knowledge and Wisdom. Available Online: http://www.systems-thinking.org/dikw/dikw.htm

CDFG/California Department of Fish and Game, Aquatic Bioassessment Laboratory. 2003. CAMLnet, List of Californian Macroinvertebrate Taxa and Standard Taxonomic Effort.

California Regional Water Quality Control Board, San Diego Region. 1994. Water Quality Control Plan for the San Diego Basin (9).

California State Assembly Bill – Title 17 of the California Code of Regulations, Section 7958.

Clesceri, L.S., A.E. Greenberg, and A.D. Eaton, eds. 2000. Standard Methods for the Examination of Water and Waste Water. 20th ed. New York: American Public Health Association (APHA), American Water Works Association (AWWA), and Water Environmental Federation (WEF).

Gibbs, R.J. 1973. Mechanisms of trace metal transport in rivers. Science 180:71-73.

Harrington, J.M. 1999. California Stream Bioassessment Procedures (CSBP). California Department of Fish and Game, Water Pollution Control Laboratory. Rancho Cordova, CA.

Kennish, M.J. 1998. *Pollution Impacts on Marine Benthic Communities*. CRC Press, New York. 310 p.

Model Monitoring Technical Committee – Stormwater Monitoring Coalition. 2004. Unpublished report. Model Storm Water Monitoring Program for Municipal Separate Storm Sewer Systems in Southern California.

Moore, J.N., E.J. Brooks, and C. Johns. 1989. Grain size partitioning of metals in contaminated coarse-grained river flood plain sediment, Clark Fork River, Montana, USA. Environmental Geology and Water Science 14:107-115.

Ode et al. 2002 Biological Assessment Report. Results of May 2001 Reference Site Study and Preliminary IBI. Submitted to San Diego Regional Water Quality Control Board. California Department of Fish and Game, Office of Spill Prevention and Response, Water Pollution Control Laboratory. Rancho Cordova, California.

Ode, P.R., A.C. Rehn, J.T. May. 2003. Unpublished Report. A Benthic Macroivertebrate Index of Biotic Integrity for Southern Coastal California.

San Diego Municipal Stormwater Copermittees. 2003. A Framework for Assessing the Effectiveness of Jurisdictional Urban Runoff Management Programs. October, 2003.

San Diego Municipal Stormwater Copermittees and Project Clean Water. 2002. Data and Information Management Plan. January, 2002.

Sharma, Nikhil. The Origin of the Data Information Knowledge Wisdom Hierarchy. Available Online: http://www.personal.si.umich.edu/~nsharma/dikw_origin.htm

Siepmann, S. and B. Finlayson. 2000. Water Quality Criteria for Diazinon and Chlorpyrifos. California Department of Fish and Game Administrative Report 00-3. 65 p.

USEPA/United States Environmental Protection Agency. 2000. National Pollutant Discharg Elimination System (NPDES) Storm Water Multi-Sector General Permit for Industrial Activities. 65 Federal Register (FR) 64746, Final Reissuance. October 2000.

USEPA/United States Environmental Protection Agency. 2000. Federal Register Document 40 CFR Part 131. May, 2000.