APPENDIX B

METHODS

TABLE OF CONTENTS

Page

1.0 METHODS B-1
- 1.1 STUDY AREA SELECTION B-1
  - 1.1.1 Selection of Arid and Semi-Arid Study Areas B-1
  - 1.1.2 Selection of Non-Arid Study Areas B-3
- 1.2 HISTORIC DATA SOURCES B-4
- 1.3 SITE RECONNAISSANCE B-4
  - 1.3.1 Aquatic Biology B-4
  - 1.3.2 Terrestrial Biology B-9
  - 1.3.3 Geomorphology B-9
- 1.4 PHYSICAL DATA ANALYSIS METHODS B-10
  - 1.4.1 Climate Assessment B-10
  - 1.4.2 Flow Assessment B-10
  - 1.4.3 Depth and Velocity B-11
- 1.5 CHEMICAL DATA ANALYSIS METHODS B-11
  - 1.5.1 Water Quality Summary Tables B-11
  - 1.5.2 Data Quality Assessment B-12
  - 1.5.3 Piper Diagrams B-13
  - 1.5.4 Box Plots B-14
  - 1.5.5 Principle Components Analysis B-14
- 1.6 DATA MANAGEMENT B-16
- 1.7 QUALITY ASSURANCE B-18

LIST OF FIGURES

B-1 Number of Families from Macroinvertebrate Studies Conducted across Western United States

B-2 Schematic of Box Plot Components

LIST OF TABLES

B-1 Table Structure for Physical and Chemical Data

APPENDIX B

1.0 METHODS

1.1 STUDY AREA SELECTION

This section discusses the selection of the 10 arid and semi-arid study areas and the non-arid study areas that were analyzed in the Habitat Characterization Study.

1.1.1 Selection of Arid and Semi-Arid Study Areas

The scope of work for the Habitat Characterization Study called for the selection of a representative set of effluent-dependent and effluent-dominated (effluent discharge equals a majority of the total flow) waters distributed across the arid West (**Figure B-1**). To assist with study area selection, the team relied on the information gathered from the Discharger's Survey of arid and semi-arid west streams (Arid West Water Quality Research Project [WQRP] 2000).

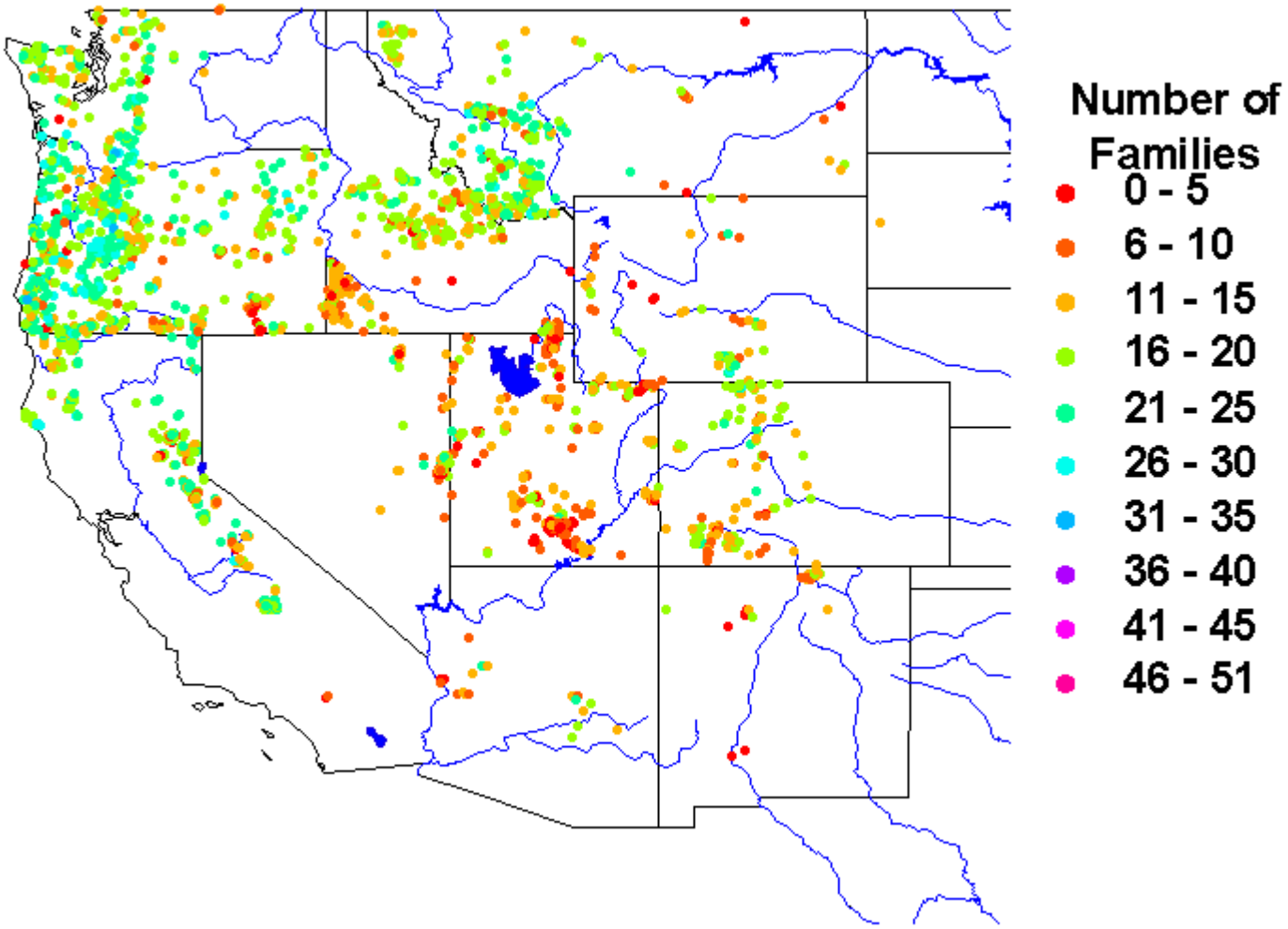
The Discharger's Survey used the U.S. Environmental Protection Agency's (EPA's) database to identify existing National Pollutant Discharge Elimination System (NPDES) permits within the 17 arid and semi-arid western states. This process identified 4,515 NPDES permits. Of this permit pool, 1,001 permits were issued to major municipal dischargers and of this pool, 251 permits had been issued in the arid and semi-arid portions of the 17 western states. Through the use of questionnaires and telephone interviews, it was determined that 71 of the 251 permit holders discharge to ephemeral, intermittent, or effluent-dependent streams. This set of dischargers represented the core of the survey and was used for additional analysis for the purposes of that survey.

With the set of 71 core dischargers identified, the project team worked with the WQRP to identify a subset of dischargers to be included in the Habitat Characterization Study. To the extent possible, study areas were selected that (1) were widely distributed across the arid and semi-arid West; (2) had a variety of historical physical, chemical, and biological data; and (3) had dischargers that were willing to share their own historical data and assist with providing access to the study stream. Ultimately, the following 10 study areas were selected using this process:

1. Salt and Gila Rivers Near Phoenix, Arizona

2. Santa Cruz River Near Nogales, Arizona

3. Santa Cruz River Near Tucson, Arizona



Number of Families from Macroinvertebrate Studies
Conducted across Western United States
Figure B-1

4. Santa Ana River Near San Bernardino, California

5. Fountain Creek Near Colorado Springs, Colorado

6. South Platte River Near Denver, Colorado

7. Las Vegas Wash Near Las Vegas, Nevada

8. Santa Fe River Near Santa Fe, New Mexico

9. Carrizo Creek Near Carrizo Springs, Texas

10. Crow Creek Near Cheyenne, Wyoming

Of these 10 study areas, all but one, Carrizo Creek, had a variety of historical data; however, Carrizo Creek was still included as a study area since it provided an example from the far eastern portion of the arid and semi-arid West.

1.1.2 Selection of Non-Arid Study Areas

After completion of the preliminary data analysis for the 10 study areas, the project Scientific Advisory Group recommended that additional data be gathered from a few non-arid areas. This effort was not conducted to universally characterize non-arid areas, but rather to provide preliminary data to identify physical, chemical, and biological differences between arid and non-arid streams. Also, no reconnaissance visit to these sites was possible within the scope of this project. For all of these reasons, the information obtained for these non-arid sites is not as complete as for the arid West sites. No systematic selection method was used to pick the non-arid sites. Because of limited time and budget, non-arid study areas were selected where it was known that at a minimum hydrologic, chemical, and aquatic biological data would be available. Initially, the project team selected two areas for analysis: North Carolina and the Kansas River near Kansas City, Kansas. North Carolina was selected because this state has established extensive biological monitoring as part of their water quality monitoring programs. The selection of these two locations also provided a gradient from east to west across the United States.

The North Carolina Department of Environmental and Natural Resources was contacted and agency staff worked with the project team to identify three locations where data were available upstream and downstream of wastewater treatment facilities and that represented a range of watershed sizes in that region. Once areas were selected, agency staff provided water quality wastewater treatment facility data from their files and databases. In addition, the project team gathered hydrologic data from nearby U.S. Geological Survey (USGS) gaging stations. The Kansas River data were obtained from the Johnson County Wastewater Department in Overland Park, Kansas and were used by permission.

1.2 HISTORIC DATA SOURCES

The databases of federal (e.g., EPA) and state agencies (e.g., departments of environmental quality) served as the starting points for the historical research. These databases were easily accessible, either through the Internet or through publications. The next step was to contact local colleges and universities. Wastewater treatment plants (WWTPs) within the 10 study areas were contacted for historical effluent sampling data and other relevant information. Agencies were contacted first by telephone to identify appropriate contact persons and type of data available. Written data requests were submitted when contacts were not responsive to inquiries by telephone. Records of conversations, including agency/organization, contact name, date, and issues discussed, were maintained by all individuals involved in collection of historical data.

The following sources were contacted for information pertaining to each of the 10 study areas:

- EPA
- State Environmental Agencies
- USGS
- State/Local Universities
- Natural Resources Conservation Service
- The Nature Conservancy
- U.S. Fish and Wildlife Service (USFWS)
- Local Environmental Organizations
- National Oceanic and Atmospheric Administration
- Local Sources of Aerial Photography
- State Geological Surveys
- State Game and Fish Departments

1.3 SITE RECONNAISSANCE

This section discusses the methods used for conducting the site reconnaissance phase of the project for aquatic biology, terrestrial biology, and geomorphology.

1.3.1 Aquatic Biology

Site reconnaissance methods for aquatic biology included habitat assessment and macroinvertebrate sampling, as detailed below.

1.3.1.1 Aquatic Habitat Assessment

Physical aquatic habitat characteristics can significantly influence the aquatic community, such as the potential to colonize a given aquatic environment. This influence may be so strong as to override the potential limitations that may be attributed to water quality (Barbour et. al 1999; Karr 1991; Karr et al. 1986; Southwood 1977). Consequently, it is critical that habitat characterization occur at the same time that samples are collected to accurately characterize the aquatic community. Coupling habitat data analysis with biological community data analysis can yield results that allow a determination of whether habitat quality or water quality is the limiting factor in aquatic biological potential.

Aquatic habitat characterization methods incorporate a range of factors that influence both the structure and function of the aquatic community. Methods can range from detailed quantitative field methods to qualitative scoring methods that rate habitat on the basis of best professional judgment. Regardless of the method used, the outcome or interpretation of data analysis should lead to an evaluation of whether or not habitat is a limiting factor.

Federal agencies and biological research organizations use a variety of aquatic habitat assessment methods. Similarities can be identified among methods, but there are substantial differences as well. These differences are likely driven by the need to gather data for different purposes or agency missions. For the purposes of this study, two methods were used: (1) the method adopted by EPA, and (2) a habitat method developed specifically for this project. These methods incorporated two basic components: (1) qualitative information was recorded (e.g., aquatic vegetation density, water odor and clarity, substrate type, and the relative proportion of run, riffle, and pool habitat); and (b) specific characteristics associated with instream habitat, channel morphology, and the riparian zone were scored.

EPA Habitat Assessment Protocol for Low Gradient Streams. (Barbour et al. 1999). The EPA habitat assessment protocol was chosen because its use in bioassessments has gained general acceptance. Furthermore, while habitat assessment methods vary somewhat among federal agencies, most states use habitat assessments derived from the EPA protocols. The EPA habitat assessment protocols were originally published by Plafkin et al. in 1989, and were revised by Barbour et al. (1999). The EPA habitat protocol was selected because it represents a generally accepted approach to evaluating aquatic habitat in wadeable streams. Furthermore, this method or a variation of the method is used by many states in their bioassessment programs.

The EPA habitat assessment for low-gradient streams relies on the scoring of 10 habitat factors to rate a site. With a score ranging between 20 (best) and 0 (worst) each site is rated for the following factors:

- epifaunal substrate/available cover
- pool substrate characterization
- pool variability

- sediment deposition
- channel flow status
- channel alteration
- channel sinuosity
- bank stability
- vegetative protection
- riparian vegetative zone

The first five factors are rated for the specific sample location; the final five factors are rated for both the specific sample location and the larger stream reach within which the sample location is found.

Arid West Habitat Assessment Protocol. This alternative habitat assessment method, developed by project investigators specifically for effluent-dependent waters for use in this project, was used at each site to provide an opportunity to evaluate the applicability of the EPA protocol in effluent-dependent waters. Although the EPA habitat assessment approach has gained general acceptance over the years, concerns exist that habitat assessment in streams of the arid Southwest may not be accurately portrayed by direct application of the EPA habitat assessment procedures since the EPA method includes a presumption regarding what defines optimal, sub-optimal, marginal, and poor habitat.

The very nature of ephemeral, intermittent, and effluent-dependent southwestern streams may lead to poor ratings for an entire study area, even though this is a naturally occurring condition. Moreover, within this "poor" condition may lie a range of habitat quality not discernable with current Rapid Bioassessment Protocol (RBP) Habitat Assessment Method ratings. As an example, one of the RPB habitat parameters for low-gradient streams is designated "epifaunal substrate/available cover." The optimal condition is assumed to contain large areas of substrate suitable for benthic invertebrate colonization and fish cover, such as woody snags, submerged logs, and undercut banks. However, many low-gradient southwestern streams do not naturally have the appropriate riparian vegetation to contribute material to form woody snags and submerged logs. And, although undercut banks may have formed during high flow events, they do not always function as fish cover during the majority of the year when flows are low or absent. A different rating scale for this habitat parameter, scaled toward the more limited potential for many low-gradient southwestern streams, would lead to a more realistic picture of the optimum condition for the substrate/cover parameter.

With these considerations in mind, it was believed that it may be appropriate to develop a habitat assessment method specifically designed for use in arid West streams. The Arid West Habitat Assessment form was developed on the basis of the best professional judgment of project biologists who have worked frequently in arid West streams. Although there certainly are

similarities between the arid West and EPA habitat assessment forms, the arid West form incorporated additional emphasis on instream habitat characteristics. The resulting form assessed the following habitat characteristics with a score ranging from 6 (best) to 0 (worst):

- current flow conditions
- bottom substrate type
- sediment deposition (riffles/runs)
- sediment deposition (pools)
- presence/absence of undercut banks, logs, roots, cobble, and boulders
- presence/absence of pools, riffles, and runs
- dominance of pools, riffles, or runs
- habitat diversity (riffles/runs)
- habitat diversity (pools)
- run/bend ratio
- channelization
- instream structures
- percent vegetative cover
- predominant vegetative bank cover
- predominant non-vegetative bank cover
- evidence of erosion

1.3.1.2 Macroinvertebrate Community Sampling

The project team conducted a screening-level field investigation of the macroinvertebrate community at each site. The primary purpose of this approach was to provide sufficient aquatic biological community data to identify macro-changes in the characteristics of the macroinvertebrate community along a longitudinal gradient from just upstream of the WWTP to some distance downstream of the WWTP discharge. In addition, sufficient data were gathered to allow a general comparison of macroinvertebrate community characteristics among the 10 study areas. The data collected were not intended to make definitive statements regarding water or habitat quality. Substantially more intensive sampling would be necessary to accomplish such a purpose.

The macroinvertebrate community screening-level procedure consisted of the following activities at each sample site:

1. Study biologists observed the aquatic habitat, then selected three sample locations that were representative of the variety of habitat available for colonization by benthic invertebrates. For example, if run, riffle, and pool habitats were available, then one sample was collected from each of these habitat types. Working from downstream to upstream, the three selected sample locations were sampled for one minute each. Data from each sample were recorded separately.

2. The sampling device was a standard D-Frame kick net with a 500-micron mesh bag. In runs and riffles, the sampler faced downstream, placed the net on the stream bottom, and kicked up the substrate in front of the net for one minute, thereby capturing dislodged benthic organisms in the net. In pools, the sampler kicked up the substrate to dislodge benthic organisms into the water column and then used the kick net to sweep the water column and capture the dislodged organisms. This kick and sweep activity was also conducted for a period of one minute. Upon completion of the sampling, the sample was deposited into a single sample tray. The kick net was checked for any remaining organisms attached to the net. Rather than removing these attached organisms and adding them to the sample tray, the organisms were identified, enumerated, and recorded.

3. Water was added to the sample tray to allow benthic organisms to swim or move freely among the debris and sediment in the sample. The water was then decanted into a white, gridded tray leaving the sediment behind in the first tray.

4. Benthic organisms were identified and counted in the white, gridded tray (grid contained 12 squares) to obtain relative abundance values for each taxon present. If the organisms of a specific taxon were particularly abundant, then the count was based on a subsample of the gridded squares in the tray. For example, if six (50 percent) of the squares were counted, then the actual organism count was the number counted in the six squares multiplied by two. If a specific taxon was not abundant, then the number of organisms was based on a count of all the squares within the tray.

5. Because considerable debris and sediment were often collected during the sampling process, it was often necessary to repeat Steps 3 and 4 in order to ensure that all benthic organisms were removed from the debris and sediment. After water was decanted from the original sample tray for the last time, a cursory check of any remaining debris and sediment was made to verify that no obvious benthic organisms remained.

6. Benthic organism identifications and counts were recorded on data sheets. The total count/taxon for each kick sample was based on the results of all counts obtained from repeating Steps 3 and 4 plus any counts obtained from the enumeration of organisms present on the net.

7. Because the sampling was only a screening-level effort, taxonomic identification was limited to what could be reliably identified in the field. Samples were not preserved, nor were they

sent to laboratories for confirmation or additional taxonomic identification. Field taxonomic identifications were generally limited to the family level for insects, the order level for crustaceans and snails, and the class level for any other taxa (e.g., worms and leeches).

1.3.2 Terrestrial Biology

Data included in the terrestrial ecology sections of this report were obtained from two basic sources: literature review and limited site reconnaissance. Literature review included contacts with local experts and personnel at the selected set of dischargers, from which the study team attempted to obtain unpublished reports and memos regarding terrestrial ecological resources. Literature sources reviewed included national and regional works on flora and fauna for each of the selected discharger sites, Internet data sources, scientific publications, and agency contacts. Data on threatened, endangered, or otherwise sensitive plant and animal species were obtained from USFWS websites and from state agencies having jurisdiction over such forms.

Site-specific data were obtained during visits to each of the 10 study areas. Within each of the 10 areas, a total of five sites were visited whenever possible. The five sites included one site upstream of the discharge point, the actual discharge point, and three sites downstream of the discharge point. There were more sites tested downstream of the discharge points since many areas had no flow upstream of the WWTP. Generally, the downstream sites were spread out in such a way that 90 percent of the watered reach of the effluent stream was covered. Covering 90 percent of the downstream watered reach was not possible at sites in Wyoming, California, and Colorado due to the perennial nature of the streams involved.

A field data sheet was completed at each of the five sites visited within each study area. The three-page data sheet, tailored from a published EPA form (EPA 1998), allowed the project team to record information on vegetation, wildlife, and human influences noted at each location. Vegetation data included species composition, relative abundance, and estimates of cover for canopy, understory, and ground cover. Human influences recorded included stream channel features (e.g., riprap, dikes, etc.), pavement, pipes, mining activity, recreation use, and other influences.

Plant species that were not readily identifiable in the field were collected for later identification and verification. Plants were collected by taping samples into a spiral notebook and recording location and habitat information on the page to which the sample was affixed. Specimens were identified in the office using standard plant taxonomy keys and verified by personnel at the University of Arizona herbarium.

In addition to plant samples and field data collected on standardized field data sheets, each site was also photographed and video-taped.

1.3.3 Geomorphology

Geomorphological observations and data were collected according to standard field procedures (for example, Leopold 1994; Leopold et al. 1964; Moody, et al. 2000). The procedure consisted of a single site assessment performed by a field geomorphologist. Morphometric data were

collected at each site and combined with general observations to evaluate the relevant landform and fluvial processes active at each site.

Sinuosity and gradients were measured from USGS 1:24,000 scale topographic sheets and/or aerial photography, as available. Fluvial morphometric data were measured in the field using a steel tape. Flow velocities were estimated by eye. Grain size textures and compositions were estimated in the field and no site samples were returned for laboratory testing. Photos were taken at the site to record general features. All notes were recorded in field notebooks retained by the investigator and photo-reproduced for the project files.

1.4 PHYSICAL DATA ANALYSIS METHODS

1.4.1 Climate Assessment

Surface and groundwater flows originate from precipitation. To quantify precipitation amounts and seasonal variability, the project team obtained historic daily precipitation data for all 10 sites from the National Climatic Data Center (NCDC). From these data, monthly precipitation amounts, average annual precipitation amounts, and extreme value precipitation (maximum) were calculated.

Historic temperature data also were collected from the NCDC for all 10 study areas to quantify seasonal variation and extreme value information. From these data, average monthly temperature, average annual temperature, and extreme value temperatures were calculated.

1.4.2 Flow Assessment

The project team collected average daily streamflow from USGS streamflow records for all study areas except Carrizo Springs, Texas. For each stream, data were collected from one gage upstream and one gage downstream of the WWTP inputs to the stream. If available, the project team obtained the average daily WWTP flows from the dischargers.

From the average daily streamflow and average daily WWTP flow, the project team performed the following analyses:

- hydrograph or average daily streamflow versus time for the period of record
- mean monthly flow to assess seasonal variation of streamflow
- Log Pearson Type III flood frequency analysis using annual series data
- Log Pearson Type III low flow frequency analysis using low flow duration of 7 days and 30 days (these low flow durations are typically used in establishing permit limits)
- flow duration curves to establish the percent of time that mean daily flows occur
- average annual yield at each gage (acre-feet)

- annual maximum flow versus annual mean flow to quantify "flashiness" of systems

Ratio of streamflow (inches) to rainfall (inches) through time was also used to assess if increase in urbanization had increased runoff amounts to stream systems.

1.4.3 Depth and Velocity

The project team obtained depth and velocity rating curves where available. The rating curve data were plotted on a log scale and fitted with a power regression to obtain the equation relating depth or velocity to flow. The power equations were used to classify the streams into four general categories of velocity and depth for benthic and fish communities (Novotny and Olem 1997): (1) slow and shallow (velocity $v < 0.3$ m/s; depth $H < 0.5$ m); (2) slow and deep ($v < 0.3$ m/s; $H > 0.5$ m); (3) fast and deep (> 0.3 m/s; $H > 0.5$ m); and (4) fast and shallow ($v > 0.3$ m/s; $H < 0.5$ m).

The project team also used the rating curve data to estimate at what flow the streambed would begin to move. Qualitative categorizations of substrate type were estimated during the site reconnaissance efforts. These substrate types were used to estimate mean particle size, which was then used to estimate critical shear stress. Using critical shear stress, the dimensions and corresponding flow at which the streambed moves were then determined.

1.5 CHEMICAL DATA ANALYSIS METHODS

Chemical data from the selected reconnaissance sites at each of the study areas were compiled and characterized using several methods, as described below.

1.5.1 Water Quality Summary Tables

Water quality summary tables were generated for each site, including all parameters measured at the site. For each parameter, the following information was defined:

- parameter name and unit of measure
- period of record (POR) for the parameter at that site
- number of measurements during the POR
- frequency of detection (FOD)
- detection limit (if less than 100 percent FOD)
- basic statistics—minimum, median, maximum
- calculated statistics—mean, standard deviation, geometric mean, geometric standard deviation

The parameters were grouped alphabetically within the following categories:

- physical/field
- nutrients
- trace elements
- major ions
- organic compounds
- biological
- radiological
- sediments

A key was developed for the tables to detail the methods used for the calculated statistics. Calculated statistics were adjusted for the number of nondetects using Cohen's method. For purposes of this adjustment, the nondetects were treated as follows:

- Cohen's method was applied for parameters with FODs greater than or equal to 50 percent and less than 100 percent.
- Calculated statistics were not provided for cases where the FOD was less than 50 percent.
- In cases of multiple-valued detection limits, the highest listed detection limit for the parameter at the station was used for all nondetects.
- Values reported as zero were replaced with the highest reported detection limit.

1.5.2 Data Quality Assessment

The following four accuracy checks (Hem 1995) were used to assess of the quality of the chemical data for purposes of this study:

- major ion charge balance
- measured total dissolved solids (TDS) versus specific conductance
- calculated TDS versus measured TDS
- anions or cations versus specific conductance

The following data quality rating system was used:

Data Quality Rating	Major Ion Charge Balance (% Difference)	Measured TDS to Specific Conductance (Ratio)	Calculated TDS to Measured TDS (% Difference)	Anions or Cations to Specific Conductance (% Difference)
Good	<10	0.55-0.75	<10	<10
Marginal	10-20	0.50-0.55 or 0.75-0.90	10-20	10-20
Poor	>20	<0.50 or >0.90	>20	>20

Results were tabulated by study area and reconnaissance site including, for each site (for each of the four accuracy checks), the number of analyses with good, marginal, and poor ratings, and an overall rating for each site.

Two methods—Piper diagrams and box plots—were used to graphically display chemical data for this study. These methods are detailed below.

1.5.3 Piper Diagrams

Characterization of major cations (positively charged ions) and major anions (negatively charged ions) were accomplished using Piper diagrams (also known as trilinear diagrams). Piper diagrams are a simple, fundamental method of examining and evaluating differences and similarities of major ion composition among samples. In a water in which only one anion and one cation are present, the best way to graphically express the data is to plot the anion concentration versus the cation concentration for each sample. However, when more than one anion and cation are present, a third axis would be required that would necessitate a three-dimensional coordinate system. The Piper diagram allows all of the major chemical components of natural waters to be plotted on a single two-dimensional diagram by using a triangular coordinate system.

The Piper diagram consists of two triangles, one for major cations and one for major anions. The major cations shown on the cation triangle are calcium (Ca), magnesium (Mg), and sodium plus potassium (Na+K). The major anions shown on the anion triangle are sulfate (SO_4), chloride (Cl) and carbonate plus bicarbonate (CO_3+HCO_3). Ion concentrations are plotted on the triangles in units of percent milliequivalents per liter (% meq/L) for each of the cations and anions. The relative proportions of the cations and anions can be quickly determined on the triangular coordinate system. Points that plot near a particular corner of the triangle (e.g., Ca) indicate a higher proportion of that ion. The Piper diagram also contains a diamond representing the projections of the cation and anion triangles and associated data.

Piper diagrams were generated for each study area as follows:

- Sites containing either a complete set of cations or a complete set of anions were plotted on the associated cation or anion triangles.
- Sites containing complete sets of both cations (plotted on the cation triangle) and anions (plotted on the anion triangle) were projected onto the diamond portion of the diagram.

- The plotted points were keyed to site ID (USGS gaging station number) and location (upstream or downstream).

- Charge balances were examined to identify poor quality data that were not included on the diagrams.

Diagrams were also generated for the four non-arid sites and for the conditions of the WET tests.

In addition, one Piper diagram was generated containing data for all study areas. This diagram was used to characterize differences and similarities among the study areas. Each station was represented by a single point representing the average cation and anion concentrations. Points were plotted with calculated TDS circles.

1.5.4 Box Plots

Box plots provide a simple, fundamental method of examining data distributions and comparing the distributions among different groups (e.g., sampling sites). The components of the box plots are provided on **Figure B-2**.

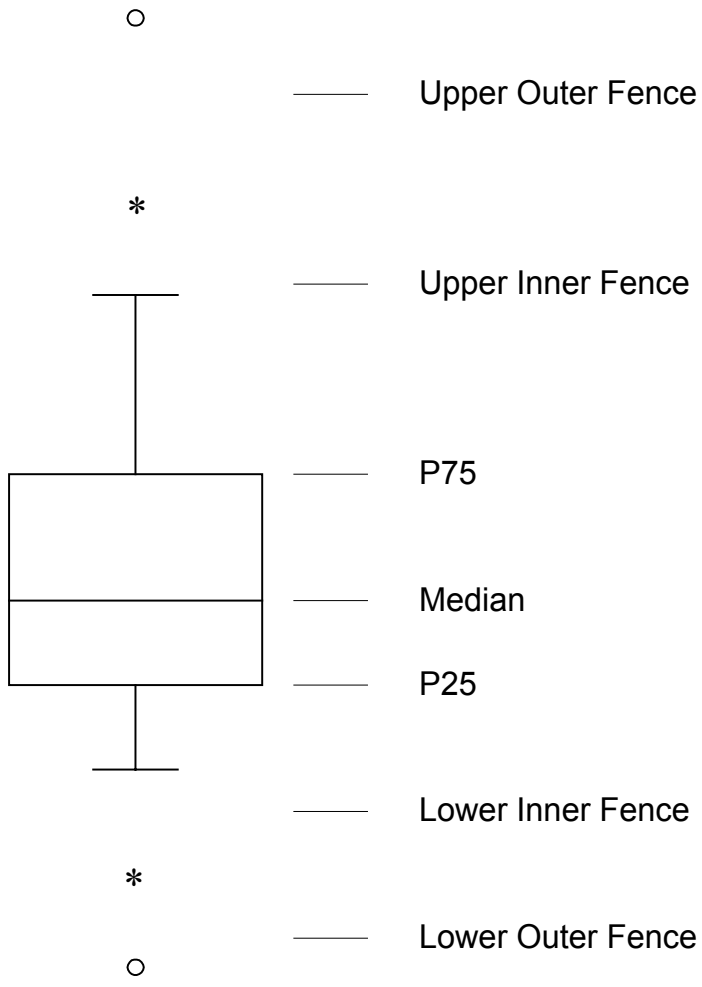
A series of box plots for all stations and study areas was generated, one for each parameter. The box plots within each study area were ordered from upstream to downstream along the stream, with locations upstream and downstream of WWTPs color-coded.

Key parameters represented by box plots were as follows:

- TDS
- hardness
- alkalinity
- pH
- ammonia
- dissolved oxygen

1.5.5 Principal Components Analysis

Principal components analysis (PCA) is a multivariate characterization method. The goal of the PCA is to represent the variation present in many variables using a small number of principal components (PCs), which are linear combinations of the original variables. Examination of the PCs enables identification of natural structures in the data that otherwise would be difficult to identify.



Key

P25	= 25th Percentile	Lower Outer Fence	= P25 - 3.0 (Median - P25)
P75	= 75th Percentile	Lower Inner Fence	= P25 - 1.5 (Median - P25)
*	= Outlier	Upper Outer Fence	= P75 + 3.0 (P75 - Median)
○	= Far Outlier	Upper Inner Fence	= P75 + 1.5 (P75 - Median)

Schematic of Box Plot Components
Figure B-2

PCA was conducted separately on three different sets of data as follows:

- major cations and anions only
- major cations and anions plus key water quality parameters (TDS, specific conductivity, hardness, pH, and temperature)
- all of the above plus parameters associated with WWTP effluent (dissolved oxygen, fluoride, nitrate plus nitrate, total Kjeldahl nitrogen, ammonia plus ammonium, and phosphate)

Evaluation of PCA results was limited to the three PCs that explained the highest amount of variation in the data sets. Each of these three PCs (one set for each of the three PCAs) was plotted as bar graphs that indicated the PC loadings for each of the original data variables included in the PCA. The PC loadings represented the correlations between each PC with each of the original variables. This method of presenting the PCA results allowed identification and characterization of the most important variable loadings for each PC.

In addition, PC scores for each sample included in the analysis were provided as scatter plots. The PC scores represented the values of the new PC variables. The PC scores were evaluated to characterize differences or similarities between individual stations or between study areas relative to the PCs.

1.6 DATA MANAGEMENT

The project team created databases for storing and manipulating information. Given the nature of the types of data collected, two databases were constructed. One database contained primarily numeric data, while the other contains primarily non-numeric descriptive data in the form of report titles, abstracts, and appendices.

The first database (numeric) contains data obtained directly from the EPA Storage and Retrieval (*STORET*) system and NCDC. Other data were retrieved indirectly through second-party sources: EarthInfo and Hydrosphere, both of Boulder, Colorado. USGS gaging station flow data were obtained from Hydrosphere. Additional streamflow data were obtained from *STORET* through EarthInfo. Climatic data for each study area, such as precipitation, were obtained for selected NCDC sites through Hydrosphere.

Surface water quality data were obtained from *STORET* and USGS. Historical *STORET* data were obtained from EarthInfo and supplemented with data obtained directly from EPA. Other water quality data were obtained directly from state USGS offices via telephone request.

Streamflow, climate, and water quality data were gathered based on the USGS Hydrologic Unit Code (HUC) in which each of the studied WWTPs is located. The numeric database was created in MS Access, a relational database. The database was normalized within data groupings; that is, the data from *STORET*, USGS, and NCDC were not combined but remained in separate table groupings.

The tables and their contents and number of records are summarized below in Table B-1. Prior to entry into a database, all data were preprocessed by restructuring and screening for suitable parameters before entry into the database.

Table B-1
Table Structure for Physical and Chemical Data

Table Name	Contents	Number of Records
TblHydrodataPRECIPITATION	Hourly and Daily precipitation up to Sept 98 for selected stations	370,249
TblHydrodataStreamFlow	Daily streamflow	312,176
TblSelectedStations	Stations listed in Pima_Study_matrix030600.xls	81
TblHydrodataStations	Station information associated with precipitation and streamflow data	661
TblSelectedWaterQualityStations	Water Quality stations listed in Pima_Study_matrix030600.xls	34
TblSTORETAgencies	Agencies of STORET data	40
TblSTORETData	STORET data from 1958 to 1998	243,556
TblSTORETParameters	STORET Parameters	1,035
TblSTORETStations	STORET stations that are surface water	868
TblUSGSParameterList	USGS WQ parameters	537
TblUSGSWQData	USGS WQ Data to 2/2000	91,697
TblUSGSWQStations	USGS Station Data	446

The second database (non-numeric) consisted of a Reference Table, Contacts Table, Keywords Table, and Abstracts Table. The Reference Table included bibliographic information for all sources collected by the project team. The following list include a brief description of each field included in the Reference Table:

- ***Stream Reach***. The name of the stream and its location.

- ***Author***. The full name of the author(s) to the extent the full name was known. For the sake of brevity in final documents, full names were reduced to initials if necessary.

- ***Date***. Date the source was published. If retrieved from a website with no publishing date, the last revision date for the website was used.

- ***Publishing Information***. All available publishing information. If the source was from the Internet and already published (e.g., the *Federal Register*), only publishing information was included. If the source was obtained from an Internet website and there was no publishing information, the full website address was entered into this field.

- ***Origin***. Original location of the source (e.g., library, agency gray literature, author's name, Internet, etc.). If the information was found at a website, the web address was included along with the date the site was accessed.

- ***Retrieved***. Indication of whether the document was retrieved or not by (i.e., "yes" or "no").

- ***Location***. Current location of the document (e.g., in the project files, in use by a project team member). If the document was not retrieved, this field was left blank.

- ***Notes***. Any pertinent notes or comments about the document (e.g., reason for non-retrieval or document status at a library).

- ***Description***. Brief description (as brief as two or three sentences) of the source as it related to this project.

The Contacts Table included the following information: name, affiliation, phone, e-mail address, date contacted, notes, initials of the investigator making the contact, and the location of the Record of Conversation form. Each conversation was briefly summarized in the "notes" field, with a more detailed account of the conversation in the Record of Conversation form.

The Keyword Table was used to record keywords used to search libraries, the Internet, and other databases. The purpose of the Keyword Table was to help avoid duplication of efforts, transfer ideas on the use of keywords between investigators, and allow for input from outside technical experts and other reviewers on searching techniques. This table included the following fields: stream reach, location searched, keywords searched, other keywords, and initials of the investigator.

The Abstracts Table was created to store abstracts of collected literature. This table was used only when an electronic copy of the abstract could be accessed and copied onto the table. It included the following fields: stream reach, author, date published, title, and abstract.

1.7 QUALITY ASSURANCE

Methods for examining chemical data quality are outlined in the Quality Integrated Work Plan (QIWP) (Pima County Wastewater Management Department 2000). For each site presented in **Appendix C**, a quality review of the electronic data was conducted via the procedures outlined in the QIWP.

In addition to gathering electronic chemical records, electronic flow records were assembled. USGS maintains large records of measured historical daily mean and daily peak flows for streams and rivers across the country. These data are obtained from stationary staff gages that continually monitor stream stage values. The staff gage levels are then converted to flow values using a stage-discharge relationship. Empirical equations define each stage-discharge relationship at each gage and are continually monitored for accuracy and modified if necessary. Detailed velocity/depth/width profiles are measured periodically at each gage site and translated into flow values to check the assumed stage-discharge relationship. For a stable stream (in terms of changing morphology), velocity/depth/width measurements are made about every six to eight weeks to verify the stage-discharge relationship. For an unstable stream, measurements are made about every week to define how the stage-discharge relationship is changing. Provisional flow

calculations are made continuously based on the most recent channel measurements and estimates of how the channel characteristics may be changing with time. Provisional flow values may be altered throughout a given year as new channel measurements better define how the channel morphology is changing with time. At the end of the year, all channel measurements are reviewed. Final decisions are made about the stage-discharge relationships throughout the year and the flow record is refined or recomputed as necessary. This record is then passed through a rigorous review process and, once approved, the data are considered final and are published (Wahl et al. 1995).

Specific Quality Assurance procedures related to surface water data are detailed in district quality assurance plans. The USGS has a policy that each District office is required to prepare a District Surface Water Quality Assurance Plan. The plan for each District describes the policies and procedures that ensure high quality in the installation of flow gages, the collection of surface water data (gage heights, velocity/depth/width profiles, etc.), the analysis of these data, and the storage and publication of the final data (Arvin 1995). Because all stream flow and stage-relationship data were collected from USGS records, it is assumed that the data are considered Level 1 data as outlined in the QIWP.