Chapter 7

Robust Data Analysis Systems for Urban Watershed Management Decision Support

Colleen L. Hughes and Ralph H. Kummeler

Recent advances in water quality and quantity monitoring technologies provide cost effective means of collecting large amounts of information for complex water quality problems. Technological advances in water quality data analysis, however, have lagged, especially for converting raw data into information which can support decisions, and robust software for managing and analyzing the information available.

As part of a Rouge River National Wet Weather demonstration grant, a suite of software called Watershed Analysis Tools for Reporting, Statistics, Hypothesis-testing and Display (WATRSHD) was designed, developed, and implemented. The WATRSHD data analysis system was based on prescribed analysis techniques. To ensure WATRSHD's suitability for large scale water quality management projects, it was developed and demonstrated using data collected by the Rouge River National Wet Weather Demonstration Project (a database containing over six million records of water quality and quantity information). Data analysis techniques, software applications developed, and case studies are presented in this chapter.

Hughes, C. and R. Kummeler. 1998. "Robust Data Analysis Systems for Urban Watershed Management Decision Support." *Journal of Water Management Modeling* R200-07. doi: 10.14796/JWMM.R200-07. © CHI 1998 www.chijournal.org ISSN: 2292-6062 (Formerly in *Advances in Modeling the Management of Stormwater Impacts*. ISBN: 0-9697422-8-2)

7.1 Introduction

Water pollution control legislation places a large emphasis on water quality monitoring as a means of supporting remediation and management planning as well as regulatory functions. The implementation of such policies over the past few decades coupled with improved data collection techniques such as electronic sensors and automatic samplers has created an explosion of available data characterizing conditions in our national waters (Hooper and Aulenbach, 1993). This data explosion has created an immediate need for more robust water quality information systems for managing, analyzing, and utilizing the information available (Barrows, 1989; Hooper and Aulenbach, 1993; Moss, 1989; Rood et al, 1996; Ward et al, 1990). As data collection becomes more cost effective, analysis objectives become more complex, and analysts realize the value of including data from supplemental sources, water quality databases expand rapidly in volume and complexity. The growing wealth of information offers analysts a great advantage in terms of combining data sets and sequencing analyses to maximize the usefulness of available data. As a result, more detailed and specialized data analysis techniques and applications are being developed for addressing water quality management objectives (Crisp, 1989; Hirsch et al, 1982; Hirsch and Slack, 1984; Hooper and Aulenbach, 1993; Hughes, 1996; Loftis and Ward, 1989; Mueller, 1989; Rood et al, 1996; Walker and Graczyk, 1993; Walker, 1994; Ward et al, 1990; Zou and Whittemore, 1993). Traditional (off the shelf) statistical packages and spreadsheet programs which have typically served as data analysis tools for water quality information systems, however, were not designed to effectively support the data management and analysis tasks that are required to conduct such in-depth analyses of large and structurally complex data sets (Barrows, 1989; Lorenz et al, 1989; Rood et al, 1996; Ward et al, 1990). Therefore, large scale water quality management programs are being forced to employ more robust systems that offer much greater power and flexibility and sacrifice the user friendliness of traditional data analysis software packages.

Effective water quality information systems, that facilitate sophisticated analyses of large and complex databases, must be designed and developed based on a thorough assessment of data analysis objectives and corresponding protocols (Messer, 1989; Moss, 1989). However, since adequate data for performing such detailed analyses have only recently become available, comprehensive guidelines for utilizing the information to its greatest potential and transforming it into information, that can be used in decision making, are still lacking. This situation was characterized by Ward et al (1986) as 'the data rich but information poor syndrome in water quality monitoring', which still holds today on a relative scale. Since data collection is the first component, initial research efforts focused on optimizing monitoring system performance and network design (Ward et al,

1990). As a result several references are now available that address 'standard' methods for data collection, laboratory analysis, and data handling, as well as network design and optimization (Gilbert, 1987; Ward et al, 1990). The ability to quantify the data collection end of a water quality monitoring system is much more advanced than the ability to quantify the information end of the system (Ward, 1989a; and 1989b Ward et al, 1990). The importance of closing the loop of the data collection, management, analysis, reporting, and utilization cycle cannot be over emphasized, if the primary objective of monitoring is to support management decisions. Therefore, the development of a data analysis strategy that outlines how the data will be used to support water quality management decisions is a critical prerequisite to the design and development of more robust systems for performing such analyses (Bell, 1989; Ward, 1989a).

7.2 Water Quality Data Analysis for Urban Watershed Management

Water Quality Information Systems that include database management systems (DBMS), geographical information systems (GIS), statistical analysis systems, and environmental simulation or modeling programs serve as effective tools for utilizing the information collected to support watershed management planning. Because the expense of collecting data, especially environmental monitoring data, is so high, it is important that data management, analysis, and reporting systems be made available for using the data collection investment to its greatest potential.

7.2.1. Analysis Goals and Objectives

The objectives of most water quality management projects can generally be condensed into three primary categories:

1. defining existing conditions;
2. evaluating trends; and
3. evaluating cause and effect.

The first includes identifying spatial, temporal, chemical, and biological water quality problems, answering the questions 'where, when, and what is the problem?' Evaluating trends is usually the next step. How are conditions changing? Is water quality improving or getting worse, or is it seasonal? Quantifying trends that demonstrate the success of a project is also a typical objective (Messer, 1989). When water quality problems are discovered or verified, evaluating cause and effect usually becomes an important objective in terms of defining the most appropriate management strategies.

7.2.2 Data Usability

A data usability review evaluates the practical utility of all data collected. This should include a comprehensive assessment of data quality and completeness with the intent of generating final data sets that will facilitate complex structured analyses and ensure the integrity of results. The process requires an intensive review of the data which includes: detailed flagging of raw data records, screening unacceptable data, correcting improperly reported data, adjusting data to compensate for equipment deficiencies, normalizing data to account for inconsistent reporting, enhancing data sets with pertinent analysis information, computing critical analysis parameters, and categorizing data for analysis purposes.

7.2.3. Statistical Methods

Water quality data analysis programs can take advantage of a broad range of statistical techniques for evaluating data and providing support for watershed management decisions. Since data analysis is often a very interactive and exploratory process where the next step is dependent on the results of the previous step, a tiered approach is often effective.

Estimation of Statistical Parameters

The estimation of statistical parameters such as the mean, maximum, minimum, standard deviation, and variance are an essential first step towards summarizing large amounts of data and providing guidance for continued analysis. For water quality parameters with regulatory defined water quality criteria or standards, the proportion of exceedance (number of exceedances divided by the number of observations) can provide a direct screening mechanism for identifying problem areas requiring more detailed analyses, that might include the number of exceedance events, the mean duration of exceedance events, and the maximum exceedance event duration. Utilizing these techniques for comparing different analysis scenarios can prove to be quite powerful. For example, comparing these statistics for different stations, seasons, weather conditions, land uses, and any other variable of interest can reveal a great deal of spatial and temporal information about water quality and governing mechanisms.

Water quality data analysis may require more sophisticated statistical techniques that allow the computation of a mean based on a weighting function. This technique is of particular use in computing flow weighted mean concentrations, usually referred to as event mean concentrations.

Correlation and Regression

Correlation matrices can be used to screen for relationships between variables. If an adequately predictive regression equation can be fitted, this analysis can be useful in optimizing sampling networks, supporting a larger scale

water quality model, or understanding cause and effect relationships. Regression and curve fitting techniques have also proved valuable in handling data below detection limits (Gleit 1985; Helsel 1990; Helsel 1995). Scatter plots are usually used in conjunction with correlation and regression results to provide a visual illustration.

Hypothesis Testing

Hypothesis testing attempts to qualify statistical inference. Hypothesis tests provide a means for reaching statistically-valid conclusions regarding differences between comparison groups for a given significance or confidence level.

Graphical Methods

Graphical methods provide a means of displaying data and statistical results in a manner which is often more conceptually interpretable than tabular formats. These techniques are essential in conveying the results of detailed statistical analyses to watershed managers. Time series plots such as hydrographs, pollutographs, and hyetographs are commonly used for analyzing temporal trends such as diurnal or seasonal behavior. Combining and overlaying these graphs using multiple parameter and stacked plotting techniques facilitates the evaluation of cause and effect relationships and correlations. Scatter plots are generally used when evaluating correlation and regression results. Histograms and probability plots are used when testing distributions. Cumulative distribution or frequency curves are useful for evaluating variability and threshold exceedances. Box and whisker plots serve as an excellent tool for evaluating differences between groups and thus providing a visual illustration and validation of hypothesis testing results.

Statistical Issues Pertinent to Water Quality Data

Statistical characteristics can significantly influence the results of certain statistical methods (Ward et al, 1990). The authors divide these attributes into two groups -- data limitations and statistical characteristics -- both of which can have an impact on the choice of statistical methods employed. Common data limitations and important statistical characteristics identified by Ward are:

- data limitations:
 1. missing values,
 2. sampling frequencies that change over the period of record,
 3. multiple observations within one sampling period,
 4. uncertainty in the measurement procedures,
 5. censoring the measurement signals,
 6. small sample size, and
 7. outliers.

- statistical characteristics:
 1. distribution shape,
 2. homogeneity of the variance,
 3. seasonality, and
 4. serial correlation.

7.3 A Robust Water Quality Data Analysis System: the Rouge Project Application

The Rouge Project serves as an excellent demonstration for this application because it is data intensive with complex water quality problems. A primary objective of the project is to demonstrate watershed management technologies with emphasis on transferability. The water quality information system for the Rouge Watershed is illustrated by providing a brief overview of known water quality problems and the Rouge Project's objectives, monitoring programs, DBMS, and decision support system. The development of a data analysis system is then described based on generalized design criteria provided in Section 7.2.

7.3.1 Rouge Watershed Background

The Rouge River Watershed serves as the primary drainage basin for over 438 square miles of Southeastern Michigan, including most of Wayne and Oakland counties and parts of Washtenaw county. The fan shaped basin consists of four major branches, numerous tributary streams, and more than 400 lakes, impoundments, and ponds. The Rouge River totals more than 127 miles in length with over 50 miles, (40%), flowing through public parklands. With approximately 1.5 million people living within the watershed, the Rouge is one of the most publicly accessible rivers in the State of Michigan.

7.3.2 The Rouge River National Wet Weather Demonstration Project

The pollution sources degrading the Rouge are common to many urban watersheds throughout the nation. In an effort to address this national problem, Congress appropriated nearly \$200 million to Wayne County's Department of the Environment to manage the Rouge Project. (*In this chapter \$ = US\$*). Funded through the U.S. Environmental Protection Agency (USEPA) as a demonstration project, its mission is to restore beneficial uses to the Rouge through implementing pollution control techniques that can be readily transferred to similar urban watersheds throughout the country. A successful watershed management plan must identify the most efficient and cost effective controls through a comprehensive evaluation of pollution sources and their impact on water quality. Providing

the necessary data to support these decisions is thus a critical component. The Rouge Project has initiated a highly intensive watershed-monitoring program. The program attempts to comprehensively address water quality issues across the entire watershed through an integrated approach, which includes a thorough sampling of both instream conditions and pollution sources.

Monitoring Programs

The 1993 and 1994 monitoring seasons focused primarily on defining baseline conditions with respect to instream water quality, sediment quality, and CSO water quality. The sampling network included 23 rain gages, 17 continuous water quality monitoring stations, 25 water level stations, and 45 periodically monitored water quality stations. Figures 7.1, 7.2, and 7.3 show the distribution of these stations throughout the watershed. The 23 rain gages, 25 water level monitors, and 17 continuous water quality monitors record rainfall volume, water level, dissolved oxygen, temperature, pH, and conductivity at 15 minute intervals throughout the monitoring seasons. The 1993 monitoring season included only the fall months, whereas the 1994 season lasted from April through November.

The 45 periodic monitoring stations were used to sample both wet and dry weather conditions. Each monitoring event, however, utilized only a subset of the total 45 sites available based on its specific objectives. Eight wet weather events were monitored for a series of eleven parameters (referred to as the “A” series)

Figure 7.1 Rouge project, rain gage monitoring stations.

Figure 7.2 Rouge project, level and flow monitoring stations.

which included: oxygen demand (BOD_5 , CBOD, and NBOD); solids (TSS and VSS); and nutrients (TKN, NH_3 , NO_2 , NO_3 , Total Phosphorus, and Ortho Phosphorus). A limited number of sites also included analyses for fecal coliform, fecal streptococcus, and *E.coli*. Samples were collected and analyzed at discrete intervals throughout the rainfall event so that variation in water quality could be examined over the course of the event. Most events included a pre-event sample for characterizing antecedent conditions. Two of the eight events also included analyses for eight different metals, where both total and dissolved concentrations were determined. Dry weather monitoring included at least two surveys for each branch, which included analyses for “A” series parameters, UBOD, metals, and coliform group bacteria.

The sediment quality monitoring program included two surveys. The first included a watershed reconnaissance of surface sediment quality. Nearly 200 surface grab samples were collected and analyzed for PCBs, PAHs, and metals. The second survey included approximately 20 sediment core samples collected from one of the impoundments, Newburgh Lake. Each core was divided into approximately ten cross sections and analyzed for the same parameters as the surface grabs.

The 1994 CSO monitoring program included a total of eleven sites, which were monitored for thirteen wet events. CSO samples included a first flush composite (generally the first 30 minutes of overflow), an event remainder composite, and a total event composite. In addition to recording flow volumes,

samples were analyzed for “A” series parameters, UBOD, and metals. TSS and BOD₅ samples were also collected at discrete intervals throughout the event in an effort to characterize the variability of CSO water quality.

The 1995 monitoring program expanded on the baseline programs with greater emphasis on obtaining more information on stormwater and additional non-point sources. The continuous rainfall, water quality, and level monitoring programs were maintained from May through November. Four additional dry weather events and five wet events were monitored. Stormwater monitoring included four sites and nine wet events. Non-point source monitoring included five sites and twenty events. Specialized programs were also conducted for collecting data regarding air deposition, landfill leachate, reaeration, time of travel, and sediment oxygen demand.

Figure 7.3 Rouge project, baseline monitoring stations.

The 1996 program includes the same rain gage network and scaled back continuous level and water quality networks with seventeen and eight sites, respectively. Biological, habitat, and stream bank erosion surveys were also conducted. Significant efforts for 1996 and beyond focus on collecting data from newly constructed best management practice sites which are now being implemented to control CSOs, stormwater, and non-point sources. (*written 1997*).

Database Management System

The Rouge Project uses an ORACLE relational database management system to store and manage the large volumes of data being collected. In addition to the data collected through the Rouge Project monitoring programs, Rouge Watershed data available from other sources have been included in the ORACLE database. Some of these sources include historical water quality and quantity data collected through the Michigan Department of Natural Resources (MDNR) and USGS, and recent MDNR biomonitoring data. By the end of the 1995 monitoring season, the ORACLE database contained over 6 million records. To provide efficient storage and retrieval, the database operates on a SunSparc 10/30 server with 8 gigabytes of disk space and 128 megabytes of RAM.

Water Quality Information System: WMSS

The Rouge Project Watershed Management Support System (WMSS) consists of a suite of decision support tools which include detailed mathematical models for simulating water hydraulics, quantity and quality; geographical information systems for managing and displaying spatial data; and database management systems for storage and retrieval of tabular data. The inclusion of a data analysis system for conducting statistical analyses supported by complex graphics provides analysts and decision makers with a versatile tool for evaluating not only monitoring data, but also GIS or model output data. Together, the suite of tools constitutes a system for conducting sophisticated analyses for answering many management or technical questions regarding the Rouge Watershed.

7.3.3 Data Analysis design System

The data usability and statistical procedures outlined in Section 7.2 provide the basis for the design of a data analysis system which should address common watershed analysis goals and objectives. Summarized below is a condensation of these design criteria and others necessary to ensure appropriate functionality and transferability.

Statistical Capabilities

1. Distribution testing: probability plots, histograms, and normality estimators such as Lilliefors Kolmogorov-Smirnov test, the Shapiro-Wilk test, skewness, and kurtosis.

2. Basic summary statistics: mean, median, geometric mean, maximum, minimum, quartiles, variance, standard deviation, coefficient of variance, proportion of exceedance, and percentiles.
3. Correlation statistics: correlation matrices, regression statistics, curve fitting, Pearsons correlation coefficient, Spearman's rho, Kendall's tau.
4. Hypothesis tests: parametric tests such as students t-test, paired t-test, and analysis of variance (ANOVA), nonparametric tests such as the Mann-Whitney test, Wilcoxon Signed Rank test, Kruskal-Wallis test, Levene's test, and the seasonal Kendall test, confidence limits.
5. Graphical displays: time series plots, box and whisker plots, scatter plots, probability plots, histograms, bar charts, frequency curves/cumulative distribution plots, stacked graphs, multi-parameter plots, and regression plots with confidence bands.

Data Management Capabilities

1. Input/Output: remote connection to DBMS with both read and write capabilities, flexible I/O external file formats, flexible printing and plotting options, flexible I/O graphics options.
2. Other formatting: flexible date and time formatting, dataset/table transposing capabilities, flexible data joining and appending capabilities.
3. Computations and data manipulation: powerful data management functionality for performing complex computations.

User Interface Capabilities

1. Programming language: high flexibility and control, application repeatability, project specific applications.
2. Graphical user interface: a user-friendly interface for commonly executed applications and standard reports.
3. Interactive and exploratory analysis capabilities: Dynamic spreadsheet, graphical, and tabular on screen displays with flexible editing and annotating capabilities.

System Performance Capabilities

1. Procedure execution: proper and complete execution of statistical and data management procedures using large data sets.
2. Quality Output: presentation quality graphics, flexible means for integrating results from different procedures.

Numerous software packages are available for performing statistical analyses supported by a wide range of graphical and data management capabilities. For analyzing large volumes of data using specific analysis procedures, a high-powered package was required. The SAS Institute offers a suite of data analysis software tools through their SAS System for Information Delivery, which features high-powered and flexible data access, management, analysis, and presentation capabilities in an applications development environment. The SAS System also offers a direct interface with several database management systems, including ORACLE, allowing efficient remote transfer of large data sets while maintaining the most up-to-date information. In addition, the SAS System is the EPA's standard software for statistical analysis and is widely used by water resource managers (Barrows, 1989; Crawford et al, 1983; Deb et al, 1989; Lorenz et al, 1989; Spooner et al, 1989; Young et al, 1989). Therefore, the SAS System not only satisfies the Rouge Project's data analysis needs, but can enhance the technology transfer and demonstration aspects of the project as well.

7.3.4 WATERSHD Development

In an effort to meet the needs of the Rouge Project and similar large scale water quality management projects, a suite of Watershed Analysis Tools for Reporting, Statistics, Hypothesis-testing and Display (WATRSHD) was developed, documented, and implemented based on the above outlined design guidelines using the SAS System as the software foundation. The ultimate goal was to create a versatile analysis system that will provide watershed management staff with all the data query, analysis, or reporting tools needed to support decisions regarding water quality monitoring networks, models, and management strategies. The focus was to maximize information utilization and minimize staff time requirements. Therefore, WATRSHD provides an adequate mix of standard reporting applications with more flexible tools for performing interactive exploratory analyses. Since the SAS System provides pre-established routines for accomplishing standard statistical procedures, the application development effort focused on integrating appropriate techniques to maximize information utilization as it relates to water quality data analysis. Several applications were designed to summarize, group, or format data to facilitate quick and easy analysis and reporting with minimal data management investment. The various analysis tools are available through the following application groups:

1. Time Series Plotter,
2. Time Series Analyzer, and
3. Modeling Results.

Time Series Plotter

Time series plots are an integral part of any water quality data analysis strategy that attempts to evaluate trends or cause and effect. Time Series Plotter consolidates the different types of data collected into easily interpretable displays to support both data QA/QC reviews and analysis.

Continuous Water Quality and Level/Flow QC Review Plots. This application generates stacked time series plots displaying rainfall, water level, dissolved oxygen, temperature, pH, and conductivity for all continuously monitored stations throughout the watershed (Hughes, 1996). The data for each site are displayed from upstream to downstream to facilitate analysis. The Rouge Project generates a set of stacked plots for each of the four primary branches of the watershed, with two for the Main branch, on a weekly basis. Options for generating similar Continuous Level/Flow QC Review Plots serve as the primary tool for QA/QC of monitored water level and calculated flow data.

Annual Water Quality Trend Comparison. As a tool for evaluating trends, an application for generating a time series display, that compares water quality from two different monitoring years, was developed. The application generates a set of five plots for each branch of the watershed based on a user selected date range and the selected years. An option for plotting hourly averages is also available.

Modeled Versus Observed Flow. Time series plotter also includes an application that generates stacked plots of modeled versus observed flow which are used for model calibration and verification, and for presentation of agreement between modeled results and actual conditions.

Time Series Analyzer

The Time Series Analyzer includes programs that require an analysis of temporal variations in water quality or quantity to generate the necessary data flags, adjustments, or summary information which serve as a basis for more detailed analyses. This application includes three tools for performing specific data usability and analysis functions as documented below.

Wet/Dry Event Definition. Since many urban water quality problems are related to wet weather sources such as stormwater runoff and combined sewer overflows, the differentiation of data collected during wet versus dry weather conditions can be an important analysis tool. As a result, an application was developed to identify data collected during wet weather based on a temporal analysis of continuously monitored water level or flow data. A flag designating wet weather, 'W', or dry, 'D', can then be assigned to water chemistry data records so that the impact of urban stormwater runoff, combined sewer overflows, and other wet weather sources can be investigated. Adjustable event initiation, termination, and antecedent dry period criteria are used to define wet

weather conditions and individual events. The beginning of an event is triggered by an assigned percent increase in flow. The event continues until the event termination criterion is met. The event termination criterion is based upon hydrograph separation techniques outlined by Linsley, et al. (1982). Linsley et al. (1982) outlined several options for hydrograph separation, but the approach which provided the most consistent results was based on the slope of a line connecting the event initiation and termination points. Event termination is determined when the slope of the line falls below a threshold provided by the user. This approach provides the user with sufficient flexibility to define events based on the sensitivity necessary for achieving the desired analysis objective.

Event Mean Concentrations and Loads. The event mean concentration (EMC) and load application uses the results of the wet/dry flagging application to support the computation of flow weighted averages and mass loadings for each monitored wet weather event based on an analysis of discrete analytical water quality samples strategically collected throughout the course of the event. The application computes a total EMC and load and an event EMC and load for each monitored station and event. The total load represents the total mass of a given constituent, which flows past the monitoring station over the course of each wet event. The total EMC represents the average concentration of the total volume of water, which flowed past the monitoring station during the event. In order to evaluate pollutant concentrations and loads added to the river by each rainfall event, antecedent samples collected prior to each wet event were used to extract existing pollutant loads from the total to calculate event loads and EMCs.

Continuous Time Series Calibration Drift Adjustments. This application was developed as a QA/QC data usability tool for executing assigned magnitude adjustments to large intervals of time series data sets. Improvements in monitoring technology have provided major benefits by offering an inexpensive means of collecting continuous data records of water quality information, but not without at least minor drawbacks. The continuous nature of this method introduces added uncertainty with regards to the quality of data collected. The use of continuous monitors extends data QA/QC issues to include (i) Sensor Fouling, (ii) Calibration Drift, and (iii) Electronic Failures. To help quantify these uncertainties, site service visits are scheduled regularly to clean sensors, maintain electronics, and calibrate equipment. The calibration includes recording sensor readings before and after the cleaning and calibration adjustment process. These records can then be used with continuous time series plots to evaluate equipment performance and quantify data quality. In many cases, calibration discrepancy can be logically explained and quantified. This information provides an opportunity to improve data quality by compensating for the quantifiable error using magnitude adjustments.

In the case of the Rouge Project, dissolved oxygen is a primary water quality constituent of interest and is monitored at fifteen-minute intervals at 17 locations throughout the watershed. Equipment performance reviews revealed many

instances where the above-described QA/QC approach could be employed to significantly improve data quality. As a result, the assignment of calibration offset adjustments was incorporated into the data usability review process. Due to the nature of calibration error and calibration drift, adjustments are generally either constant or linearly interpolated offsets, respectively. In some circumstances both calibration error and drift are significant, requiring a constant and interpolated adjustment in series for the same range of data.

In order to provide a tool which can efficiently perform magnitude adjustments to large time series data sets based on interpolated values, an automated SAS application was developed to perform the adjustments. A DO Adjustment Table maintained by field monitoring staff as part of the QA/QC review serves as the only user input required. The table is structured to include the following information for each adjustment: (i) Station ID, (ii) Start Date, (iii) End Date, (iv) Adjustment Type, (v) Adjustment Factor and (vi) Adjustment Flag. The continuous time series data adjustment program reads the DO Adjustment Table, downloads the corresponding time series data from the ORACLE database, performs the assigned adjustments, and updates the final value and flag in the ORACLE database. Once final data values and flags are uploaded to the ORACLE database, the data adjustment procedure is verified using the Continuous Water Quality Final QC Review Plots generated by the Time Series Plotter application.

Modeling Results

Evaluating cause and effect was one of the three primary water quality data analysis objectives. Due to the numerous environmental factors involved and the inherent lack of control over these factors with regard to monitoring, it is difficult to evaluate cause and effect with monitoring data alone. Mathematical models, however, are specifically designed to support this type of analysis. Therefore, the analysis of data sets generated from numerous mathematical simulations becomes an important component of a decision support system. The interpretation of model results was supported by providing options for generating both graphical and tabular data analysis displays as listed below. The application was specifically designed to support WASP model output data sets. However, it will support data from other models providing output files are formatted similarly. The wet/dry options are available by integrating the Wet/Dry Time Series Analyzer algorithm into the application.

1. Time series plots comparing two alternatives for up to 12 parameters at up to 10 locations.
2. Frequency of exceedance plots comparing two alternatives for up to 12 parameters at up to 10 locations. An option for comparisons based on wet and dry weather conditions is also available.
3. Frequency of exceedance statistics.
4. Percentage of exceedance bar charts comparing alternatives and segments. An option for wet/dry comparisons is also available.

7.3.5 WATRSHD Implementation

Implementation includes integrating the SAS System and WATRSHD applications into the framework of the Rouge Project WMSS by ensuring proper configuration, documentation, training, and maintenance. This section outlines hardware and software requirements and system configuration.

Hardware and Software Requirements

The minimum hardware requirements include a 486 personal computer with a math coprocessor, 16 Megabytes of RAM, available hard drive space equal to three to four times greater than that of the largest data set to be processed, and a color plotter. Software requirements include the SAS System for Windows, version 6.10 or later, Windows version 3.1 or later, and a relational DBMS with corresponding connectivity software and drivers. If desired, a specialized DBMS can be excluded, since the SAS System can be configured to serve this role, also eliminating the need for connectivity software and drivers. In addition to the SAS System's Base package, the SAS Graph, Access, and Insight modules are required.

System Configuration

The configuration component of implementation involves setting up the SAS System and WATRSHD applications to work in conjunction with external software and hardware. In the case of the Rouge Project, this requires interfacing WATRSHD with the principal components of the WMSS -- the ORACLE DBMS, the ARC/INFO and ARC/VIEW GIS, and the SWMM and WASP mathematical models. In addition, system interfacing flexibility was expanded to include auxiliary decision support tools such as Quattro Pro spreadsheets, MS Access DBMS, and MS Publisher presentation software packages in an effort to maximize the resources available. Established interfacing techniques include open database connectivity (ODBC), file transfer protocols using PC TCP IP allowing cross-platform connectivity (SUN - PC), Windows Dynamic Data Exchange (DDE) and Object Linking and Embedding (OLE), standard ASCII import/export file formats, and multiple graphics file import/export file formats.

7.4 Case Studies

Rouge Project applications of WATRSHD have demonstrated the software's practical utility as a watershed management support tool. The following case studies provide examples of how the applications were used to support both technical and management decisions.

7.4.1 Case Study 1: Best Management Practices

The two primary decision support tools available to Rouge Watershed managers for evaluating best management practices (BMPs) were water quality monitoring data and predictive mathematical models. The models were used to predict water quantity and quality for four different technical alternatives (potential BMPs). At the technical level, WATRSHD was used to support the development and calibration of the models. The modeled versus observed plotting component of Time Series Plotter was used to assist in calibration of SWMM. The water quality, level, and rainfall stacked time series plots were used to support the development of WASP, by providing a tool for evaluating cause and effect relationships. Runoff EMCs and Loads calculated with WATRSHD's Time Series Analyzer application served as estimates for input to the quality model and Total EMCs and Loads were used in model calibration and verification.

Once model development and calibration were complete, WATRSHD was used to analyze both monitoring and model output data so that comparisons of technical alternatives could be conducted. The Model Results application of WATRSHD was used to generate time series plots, frequency of exceedance curves, and summary statistics comparing technical alternatives and existing conditions. Numerous scenarios were evaluated by modeling staff. In an effort to convey results to watershed managers, graphics showing representative findings, were selected and included in summary displays using WATRSHD's Model Results option for creating CGM graphic files.

7.4.2 Case Study 2: Presentation to the Federal Court

A legal battle in the Federal Court regarding the Rouge Watershed's compliance with the Clean Water Act has continued for several years. Issues have changed from NPDES permits for industrial sources to municipal wastewater treatment plant permits to the current issue of CSO and stormwater permits. As the regulatory agency, the Michigan Department of Environmental Quality (MDEQ) is responsible for enforcing compliance and issuing permits. Communities in the Rouge Watershed are responsible for discharges of both CSOs and stormwater from their municipalities and have been requested by MDEQ to submit a permit proposal stating how they plan to control pollution from these sources. If a suitable plan is not provided, significant fines will be imposed. The Federal Court Judge Feikens recognizes the financial burden both compliance and/or fines may place on involved communities. He therefore summoned representatives from key communities in the Watershed to discuss this issue in court. Federal grant funding is available, requiring a minimal match by local

communities, to address this problem at this time. Judge Feikens is therefore urging community leaders to work together to devise a cost-effective plan and to take advantage of available federal funding. To support his argument, representatives from the Rouge Program Office were asked to present the findings of the Rouge River National Wet Weather Demonstration Project. Summary results of baseline monitoring program activities were presented to emphasize the severity of the problem. By interfacing WATRSHD with ArcView plots reporting the frequency of water quality standard violations were overlayed on a Rouge Watershed map to characterize the spatial extent of the problem (Hughes, 1996).

7.5 Summary and Conclusions

As water resource managers demand more and more information to support responsible restoration and management decisions and as environmental monitoring technology improves, water quality information systems must expand to accommodate the growing data volumes and information expectations. In an effort to provide data analysis systems that can keep pace with expanding technology and knowledge, a suite of watershed analysis tools for reporting, statistics, hypothesis-testing and display (WATRSHD) was designed, developed, and implemented based on generalized water quality data analysis guidelines.

Analysis guidelines are outlined based on defining three primary objectives common to most water quality management projects with the purpose of providing design criteria for a highly transferable analysis system. Using these criteria, WATRSHD was designed to provide the necessary software tools for performing prescribed analysis techniques and corresponding data QA/QC reviews in an environment that is both robust to large data sets and user-friendly. In an effort to ensure WATRSHD's suitability for large scale watershed management projects, the application was developed based on data collected under the Rouge Project which has compiled a database containing over six million records.

References

- Barrows, Edward H, 1989 "Analysis and Integration of Ecological Monitoring Data" Proceedings International Symposium of the Design of Water Quality Information Systems, Colorado Water Resources Research Institute, Information Series No. 61.
- Bell, H.F, 1989 "Water Quality Information Systems: Past, Present and Future An Industrial Perspective" Proceedings International Symposium of the Design of Water Quality Information Systems, Colorado Water Resources Research Institute, Information Series No. 61.

- Crawford, Slack, and Hirsch, 1983, "Nonparametric Tests for Trends in Water-Quality Data Using the Statistical Analysis System" U.S. Geological Survey, Reston, VA.
- Crisp, Norman H, 1989, "Regional Surface Water Quality Characteristics of Nebraska" Proceedings International Symposium of the Design of Water Quality Information Systems, Colorado Water Resources Research Institute, Information Series No. 61.
- Deb, Yeh, and Kreamer, 1989, "Water Quality Information Management System" Proceedings International Symposium on the Design of Water Quality Information Systems, Colorado Water Resources Research Institute, Information Series No. 61.
- Gilbert, R.O, 1987, *Statistical Methods for Environmental Pollution Monitoring*, Van Nostrand Reinhold, New York.
- Gleit, Alan, 1985, "Estimation for Small Normal Data Sets with Detection Limits" *Environmental Science and Technology*, Vol. 19.
- Helsel, D.R, 1990, "Less than Obvious: Statistical Treatment of Data Below the Detection Limit" *Environmental Science and Technology*, Vol. 24, No. 12.
- Helsel, D.R. and Hirsch, R.M, 1995, *Statistical Methods in Water Resources*, Elsevier.
- Hirsch, R.M., Slack, J.R. and Smith, R.A., 1982, "Techniques of Trend Analysis for Monthly Water Quality Data" *Water Resources Research*, Vol. 18, No. 1.
- Hirsch, R.M., and Slack, J.R., 1984, "A Nonparametric Trend Test for Seasonal Data with Serial Dependence" *Water Resources Research*, Vol. 20, No. 6.
- Hooper, R.P. and Aulenbach, B.T, 1993, "Managing the Data Explosion" *Civil Engineering*, May 1993.
- Hughes, C.L. "Design, Development, and Implementation of Robust Data Analysis Systems for Urban Watershed Management Decision Support: An Application of the Rouge River National Wet Weather Demonstration Project", Wayne State University, Ph.D. Dissertation, 1996.
- Linsley, Kohler, and Paulhus, 1982, "Hydrology for Engineers", Third Edition, McGraw-Hill.
- Loftis, J.C. and Ward, R.C, August 1979, "Regulatory Water Quality Monitoring Networks—Statistical and Economic Considerations" EPA-600/4-79-055.
- Lorenz, Sims, Killeen and Fay, 1989, "Managing Data from a Large-Scale Groundwater Monitoring Program" Proceedings International Symposium of the Design of Water Quality Information Systems, Colorado Water Resources Research Institute, Information Series No. 61, 1989.
- Messer, Jay J, 1989, "Why Monitor" Proceedings International Symposium of the Design of Water Quality Information Systems, Colorado Water Resources Research Institute, Information Series No. 61.
- Moss, Marshall E, 1989, "Water Quality Data in the Information Age" Proceedings International Symposium of the Design of Water Quality Information Systems, Colorado Water Resources Research Institute, Information Series No. 61.
- Mueller, David K, 1989, "Use of Box Plots and Trend Analyses to Evaluate Sampling Frequency at water Quality Monitoring Sites" Proceedings International Symposium of the Design of Water Quality Information Systems, Colorado Water Resources Research Institute, Information Series No. 61.
- Rood, S.G.; Bristol, C.R.; and Hughes, C.L. June 16-19, 1996, "Interacting with Large Data Sets", Presented at the Water Environment Federation Urban Wet Weather Pollution Conference, Quebec City, Quebec, Canada.

- SAS Institute Inc., 1990, "SAS Procedures Guide", Version 6, Third Edition, Cary, NC: SAS Institute Inc.
- Spooner, Dickey, and Gilliam, 1989, "Determining and Increasing the Statistical Sensitivity of Nonpoint Source Control Grab Sample Monitoring Programs" Proceedings International Symposium on the Design of Water Quality Information Systems, Colorado Water Resources Research Institute, Information Series No. 61.
- Walker, J.F. and Graczyk, D.J., 1993, "Preliminary Evaluation of Effects of Best Management Practices in the Black Earth Creek, Wisconsin, Priority Watershed" *Water Science Technology*, Vol. 28, No. 3-5.
- Walker, J.F., 1994, "Statistical Techniques for Assessing Water Quality Effects of BMPs" *Journal of Irrigation and Drainage Engineering*, Vol. 120, No. 2.
- Ward, Loftis, and McBride, 1986, "The 'Data-Rich but Information Poor' Syndrome in Water Quality Monitoring" *Environmental Management*, Vol. 10, No. 3.
- Ward, R.C., 1989a, "Opening Remarks" Proceedings International Symposium on the Design of Water Quality Information Systems, Colorado Water Resources Research Institute, Information Series No. 61.
- Ward, R.C., 1989b, "Water Quality Monitoring - A Systems Approach to Design" Proceedings International Symposium on the Design of Water Quality Information Systems, Colorado Water Resources Research Institute, Information Series No. 61.
- Ward, Loftis, and McBride, 1990, "Design of Water Quality Monitoring Systems", Van Nostrand Reinhold, New York.
- Young, Thomas and Ford, Jesse, 1989, "Site Selection Procedures for Temporally Integrated Monitoring of Ecosystems (TIME)" Proceedings International Symposium on the Design of Water Quality Information Systems, Colorado Water Resources Research Institute, Information Series No. 61.
- Zou, Shimin and Whittemore, Donald, 1993, "Identification of Surface Water Quality Patterns of the Arkansas River Basin" *Journal of Environmental Science and Health*, Vol. A28, No. 9.