R. Pitt
April 5, 2007

Module 5: Statistical Analyses

Introduction

Statistical analyses are a critical component of research. The analyses that are to be conducted for a specific research activity must be carefully thought out in advance of any data collection and be an integral component of the experimental design activities. This module reviews a number of statistical tests that have been useful for a variety of water quality projects conducted by the author. The field of statistical analyses is very large and offers a great variety of tools. It is always worthwhile to consult an expert in environmental statistical analyses to help identify the most helpful and powerful tests for a specific set of objectives, experimental capabilities, and budget.

General Steps in the Analysis of Data

The analysis of data requires at least three elements, quality control/quality assurance of the reported data, an evaluation of the sampling effort and methods (and associated expected errors), and finally, the statistical analysis of the information. Quality control and quality assurance basically involves the identification and proper handling of questionable data. When reviewing previously collected data, it is common to find obvious errors that are associated with improper units or sampling locations. Other potential errors are more difficult to identify and correct. In some cases, the identification and rejection of "outliers" may result in the dismissal of rare data observations.

Experimental design efforts are usually associated with activities conducted prior to sample collection. However, many attributes of experimental design can also be used when evaluating previously collected data. This is especially useful when organizing data into relevant groupings for more efficient analyses. In addition, adequate sampling efforts are needed to characterize the information to the desired levels of confidence and power.

A general strategy in data analyses should include several phases and layers of analyses. Graphical presentations of the data (using exploratory data analyses) should be conducted initially. Simple to complex relationships between variables may be more easily identified through visual data presentations for most people, compared to only relying on descriptive statistical summaries. Of course, graphical presentations should be supplemented with statistical test data to quantify the significance of any patterns observed. The comparison of data from multiple situations (upstream and downstream of an outfall, summer vs. winter observations, etc.) is a very common experimental objective. Similarly, the use of regression analyses is also a very commonly used statistical tool. Trend investigations of water quality conditions with time are also commonly conducted.

Experimental Design

All sampling plans attempt to obtain certain information (usually average values, totals, ranges, etc.) of a large population by sampling and analyzing a much smaller sample. The first step in this process is to select the sampling plan and then to determine the appropriate number of samples needed. When evaluation previously collected data, it is often desirable and effective to organize the data according to a specific sampling plan (shown later).

Many sampling plans have been well described in the environmental literature. Gilbert (1987) has defined the following four main categories, plus subcategories, of sampling plans:

- Haphazard sampling. Samples are taken in a haphazard (not random) manner, usually at the convenience of the sampler when time permits. Especially common when the weather is pleasant. This is only possible with a very homogeneous condition over time and space, otherwise biases are introduced in the measured population parameters. It is therefore not recommended because of the difficulty of verifying the homogeneous assumption. This is the most common sampling strategy used when volunteers are used for sampling, unless the grateful agency is able to spend sufficient time to educate the volunteer samplers to the problems of this type of sampling and to specify a more appropriate sampling strategy.

- Judgment sampling. This strategy is used when only a specific subset of the total population is to be evaluated, with no desire to obtain "universal" characteristics. The target population must be clearly defined (such as during wet weather conditions only) and sampling is conducted appropriately. This could be the first stage of later, more comprehensive, sampling of other target population groups (multistage sampling).

- Probability sampling. Several subcategories of probability sampling have been described:

    - simple random sampling. Samples are taken randomly from the complete population. This usually results in total population information, but it is usually inefficient as a greater sampling effort may be required than if the population was sub-divided into distinct groups. Simple random sampling doesn't allow information to be obtained for trends or patterns in the population. This method is used when there is no reason to believe that the sample variation is dependent on any known or measurable factor.

    - stratified random sampling. This may the most appropriate sampling strategy for most receiving water studies, especially if combined with an initial limited field effort as part of a multistage sampling effort. The goal is to define strata that results in little variation within any one strata, and great variation between different strata. Samples are randomly obtained from several population groups that are assumed to be internally more homogeneous than the population as a whole, such as separating an annual sampling effort by season, lake depth, site location, habitat category, rainfall depth, land use, etc. This results in the individual groups having smaller variations in the characteristics of interest than in the population as a whole. Therefore, sample efforts within each group will vary, depending on the variability of characteristics for each group, and the total sum of the sampling effort may be less than if the complete population was sampled as a whole. In addition, much additional useful information is likely if the groups are shown to actually be different.

- multistage sampling. One type of multistage sampling commonly used is associated with the required subsampling of samples obtained in the field and brought to the laboratory for subsequent splitting for several different analyses. Another type of multistage sampling is when an initial sampling effort is used to examine major categories of the population that may be divided into separate clusters during later sampling activities. This is especially useful when reasonable estimates of variability within a potential cluster is needed for the determination of the sampling effort for composite sampling. These variability measurements may need to be periodically re-verified during the monitoring program.

- cluster sampling. Gilbert (1987) illustrates this sampling plan by specifically targeting specific population units that cluster together, such as a school of fish or clump of plants. Every unit in each randomly selected cluster can then be monitored.

- systematic sampling. This approach is most useful for basic trend analyses, where evenly spaced samples are collected for an extended time. Evenly spaced sampling is also most efficient when trying to find localized hot spots that randomly occur over an area. Gilbert (1987) present guidelines for spacing of sampling locations for specific project objectives relating to the size of the hot spot to be found. Spatial gradient sampling is a systematic sampling strategy that may be worthy of consideration when historical information implies a aerial variation of conditions in a river or other receiving water. One example would be to examine the effects of a point source discharge on receiving sediment quality. A grid would be described in the receiving water in the discharge vicinity whose spacing would be determined by preliminary investigations.

• Search sampling. This sampling plan is used to find specific conditions where prior knowledge is available, such as the location of a historical (but now absence) waste discharger affecting a receiving water. Therefore, the sampling pattern is not systematic or random over an area, but stresses areas thought to have a greater probability of success.

Box, *et al.* (1978) contains much information concerning sampling strategies, specifically addressing problems associated with randomizing the experiments and blocking the sampling experiments. Blocking (such as in paired analyses to determine the effectiveness of a control device, or to compare upstream and downstream locations) eliminates unwanted sources of variability. Another way of blocking is to conduct repeated analyses (such for different seasons) at the same locations. Most of the above probability sampling strategies should include randomization and blocking within the final sampling plans (as demonstrated in the following example and in the use of factorial experiments).

Sample size

An important aspect of any research is the assurance that the samples collected represent the conditions to be tested and that the number of samples to be collected are sufficient to provide statistically relevant conclusions. Unfortunately, sample numbers are most often not based on a statistically-based process and follow traditional “best professional judgment,” or are resource driven. The sample numbers should be equal between sampling locations if comparing station data (EPA 1983) and paired sampling should be conducted, if at all possible (the samples at the two comparison sites should be collected at the “same” time, for example), allowing for much more powerful paired statistical comparison tests. In addition, replicate subsamples should also be collected and then combined to provide a single sample for analysis for many types of ecosystem sampling. Various experimental design processes can be used that estimates the number of needed samples based on the allowable error, the variance of the observations, and the degree of confidence and power needed for each parameter (Burton and Pitt 2002).

Determination of Outliers
Outliers in data collection can be recognized in the tails of the probability distributions. Observations that do not perfectly fit the probability distributions in the tails are commonly considered outliers. They can be either very low or very high values. These values always attract considerable attention because they don't fit the mathematical probability distributions exactly and are usually assumed to be flawed and are then discarded. Certainly, these values (like any other suspect values) require additional evaluation to confirm that simple correctable errors (transcription, math, etc.) are not responsible. If no errors are found, then these values should be included in the data analyses as they represent rare conditions that may be very informative.

Analytical results less than the practical quantification limit (PQL) or the method detection limit (MDL) need to be flagged, but the result (if greater than the instrument detection limit, or IDL) should still be used in most of the statistical calculations. In some cases, the statistical test procedures can handle some undetected values with minimal modifications. In most cases, however, commonly used statistical procedures behave badly with undetected values. In these cases, results less than the IDL should be treated according to Berthouex and Brown (1994). Generally, the statistical procedures should be used twice, once with the less than detection values (LDV) equal to zero, and again with the LDV equal to the IDL. This procedure will determine if a significant difference in conclusions would occur with handling the data in a specific manner. In all cases of substituting a single value for LDV, the variability is artificially reduced which can significantly affect comparison tests. It may therefore be best to use the actual instrument reported value for many statistical tests, even if it is below the IDL or MDL. This value may be considered a random value, but it is probably closer to the true value than a zero or other arbitrary value, plus it retains some aspects of the variability of the data sets. Of course, these values should not be "reported" in the project report, or to a regulatory agency, as they obviously do not meet the project QA/QC requirements.

It is difficult to reject wet weather constituent observations solely because they are unusually high, as wet weather flows can easily have wide ranging constituent observations. High values should not automatically be considered as outliers and therefore worthy of rejection, but as rare and unusual observations that may shed some light on the problem.

Selection of Statistical Procedures
Most of the objectives of receiving water studies can be examined through the use of relatively few statistical evaluation tools. The following briefly outlines some simple experimental objectives and a selected number of statistical tests (and their data requirements) that can be used for data evaluation (Burton and Pitt 2002).

Statistical Power
Errors in decision making are usually divided into type 1 (α: alpha) and type 2 (β: beta) errors:

α (alpha) (type 1 error) - a false positive, or assuming something is true when it is actually false. An example would be concluding that a tested water was adversely contaminated, when it actually was clean. The most common value of α is 0.05 (accepting a 5% risk of having a type 1 error). Confidence is 1-α, or the confidence of not having a false positive.

β (beta) (type 2 error) - a false negative, or assuming something is false when it is actually true. An example would be concluding that a tested water was clean when it actually was contaminated. If this was an effluent, it would therefore be an illegal discharge with the possible imposition of severe penalties from the regulatory agency. In most statistical tests, β is usually ignored (if ignored, β is 0.5). If it is considered, a typical value is 0.2, implying accepting a 20% risk of having a type 2 error. Power is 1-β, or the certainty of not having a false negative. When evaluating data using a statistical test, power is the sensitivity of the test for rejecting the hypothesis. For an ANOVA test, it is the probability that the test will detect a difference amongst the groups if a difference really exists.

Comparison Tests
Probably the most common situation is to compare data collected from different locations, or seasons. Comparison of test with reference sites, of influent with effluent, of upstream to downstream locations, for different seasons of

sample collection, of different methods of sample collection, can all be made with comparison tests. If only two groups are to be compared (above/below; in/out; test/reference), then the two group tests can be effectively used, such as the simple Student's *t*-test or nonparametric equivalent. If the data are collected in "pairs," such as concurrent influent and effluent samples, or concurrent above and below samples, then the more powerful and preferred paired tests can be used. If the samples cannot be collected to represent similar conditions (such as large physical separation in sampling location, or different time frames), then the independent tests must be used.

If multiple groupings are used, such as from numerous locations along a stream, but with several observations from each location; or at one location; or from one location, but for each season, then a one-way ANOVA is needed. If one has seasonal data from each of the several stream locations for multiple seasons, the a two-way ANOVA can be used to investigate the effects of location, season, and the interaction of location and season together. Three-way ANOVA tests can be used to investigate another dimension of the data (such as contrasting sampling methods or weather for the different seasons at each of the sampling locations), but that would obviously require substantially more data to represent each condition.

There are various data characteristics that influence which specific statistical test can be used for comparison evaluations. The parametric tests require the data to be normally distributed and that the different data groupings have the same variance, or standard deviation (checked with probability plots and appropriate test statistics for normality, such as the Kolmogorov-Smirnov one-sample test, the chi-square goodness of fit test, or the Lilliefors test). If the data do not meet the requirements for the parametric tests, the data may be transformed to better meet the test conditions (such as taking the $\log_{10}$ of each observation and conducting the test on the transformed values). The non-parametric tests are less restrictive, but are not free of certain requirements. Even though the parametric tests have more statistical power than the associated non-parametric tests, they lose any advantage if inappropriately applied. If uncertain, then non-parametric tests should be used.

A few example statistical tests (as available in SigmaStat, SPSS, Inc.) are indicated below for different comparison test situations:

- Two groups
  - Paired observations
    - Parametric tests (data require normality and equal variance)
      - Paired Student's *t*-test (more power than non-parametric tests)
    - Non-parametric tests
      - Sign test (no data distribution requirements, some missing data accommodated)
      - Fiedman's test (can accommodate a moderate number of "non-detectable" values, but no missing values are allowed
      - Wilcoxon signed rank test (more power than sign test, but requires symmetrical data distributions)
  - Independent observations
    - Parametric tests (data require normality and equal variance)
      - Independent Student's *t*-test (more power than non-parametric tests)
    - Non-parametric tests
      - Mann-Whitney rank sum test (probability distributions of the two data sets must be the same and have the same variances, but do not have to be symmetrical; a moderate number of "non-detectable" values can be accommodated)
- Many groups (use multiple comparison tests, such as the Bonferroni *t*-test, to identify which groups are different from the others if the group test results are significant).
  - Parametric tests (data require normality and equal variance)
    - One-way ANOVA for single factor, but for >2 "locations" (if 2 "locations, use

Student's *t*-test)
- Two-way ANOVA for two factors simultaneously at multiple "locations"
- Three-way ANOVA for three factors simultaneously at multiple "locations"
- One factor repeated measures ANOVA (same as paired *t* test, except that there can be multiple treatments on the same group)
- Two factor repeated measures ANOVA (can be multiple treatments on two groups)

Non-parametric test
- Kurskal-Wallis ANOVA on ranks (use when samples are from non-normal populations or the samples do not have equal variances).
- Friedman repeated measures ANOVA on ranks (use when paired observations are available in many groups).

Nominal observations of frequencies (used when counts are recorded in contingency tables)
- Chi-square (X^2) test (use if more than two groups or categories, or if the number of observations per cell in a 2X2 table are > 5).
- Fisher Exact test (use when the expected number of observations is <5 in any cell of a 2X2 table).
- McNamar's test (use for a "paired" contingency table, such as when the same individual or site is examined both before and after treatment)

Data Associations and Model Building

These activities are an important component of the "weight-of-evidence" approach used to identify likely cause and effect relationships. The following list illustrates some of the statistical tools (as available in SigmaStat and/or SYSTAT, SPSS, Inc.) that can be used for evaluating data associations and subsequent model building:

- Data Associations

  Simple
  - Pearson Correlation (residuals, the distances of the data points from the regression line, must be normally distributed. Calculates correlation coefficients between all possible data variables. Must be supplemented with scatterplots, or scatter plot matrix, to illustrate these correlations. Also identifies redundant independent variables for simplifying models).
  - Spearman Rank Order Correlation (a non-parametric equivalent to the Pearson test).

  Complex (typically only available in advanced software packages)
  - Hierarchical Cluster Analyses (graphical presentation of simple and complex inter-relationships. Data should be standardized to reduce scaling influence. Supplements simple correlation analyses).
  - Principal Component Analyses (identifies groupings of parameters by factors so that variables within each factor are more highly correlated with variables in that factor than with variables in other factors. Useful to identify similar sites or parameters).

- Model building/equation fitting (these are parametric tests and the data must satisfy various assumptions regarding behavior of the residuals)

  Linear equation fitting (statistically-based models)
  - Simple linear regression ($y=b_0+b_1x$, with a single independent variable, the slope term, and an intercept. It is possible to simplify even further if the intercept term is not significant).
  - Multiple linear regression ($y=b_0+b_1x_1+b_2x_2+b_3x_3+\ldots+b_kx_k$, having k independent variables. The equation is a multi-dimensional plane describing the data).
  - Stepwise regression (a method generally used with multiple linear regression to assist in

identifying the significant terms to use in the model.)
- Polynomial regression ($y=b_0+b_1x^1+b_2x^2+b_3x^3+\ldots+b_kx^k$, having one independent variable
describing a curve through the data).

Non-linear equation fitting (generally developed from theoretical considerations)
- Nonlinear regression (a nonlinear equation in the form: $y=b^x$, where x is the independent variable. Solved by iteration to minimize the residual sum of squares).

- Data Trends
  - Graphical methods (simple plots of concentrations versus time of data collection).
  - Regression methods (perform a least-squares linear regression on the above data plot and examine ANOVA for the regression to determine if the slope term is significant. Can be misleading due to cyclic data, correlated data, and data that are not normally distributed).
  - Mann-Kendall test (a nonparametric test that can handle missing data and trends at multiple stations. Short-term cycles and other data relationships affect this test and must be corrected).
  - Sen's estimator of slope (a nonparametric test based on ranks closely related to the Mann-Kendall test. It is not sensitive to extreme values and can tolerate missing data).
  - Seasonal Kendall test (preferred over regression methods if the data are skewed, serially correlated, or cyclic. Can be used for data sets having missing values, tied values, censored values, or single or multiple data observations in each time period. Data correlations and dependence also affect this test and must be considered in the analysis).

Exploratory Data Analyses

Exploratory data analyses (EDA) is an important tool to quickly review available data before a specific data collection effort is initiated. It is also an important first step in summarizing collected data to supplement the specific data analyses associated with the selected experimental designs. A summary of the data's variation is most important and can be presented using several simple graphical tools. *The Visual Display of Quantitative Information* (Tufte 1983) is a beautiful book with many examples of how to and how not to present graphical information. *Envisioning Information*, also by Tufte (1990) supplements his earlier book. Another important reference for basic analyses is *Exploratory Data Analysis* (Tukey 1977) which is the classic book on this subject and presents many simple ways to examine data to find patterns and relationships. Cleveland (1993 and 1994) has also published two books related to exploratory data analyses: *Visualizing Data*, and *The Elements of Graphing Data*. The basic plots described below can obviously be supplemented by many others presented in these books. Besides plotting of the data, exploratory data analyses should always include corresponding statistical test results, if available.

Basic Data Plots

There are several basic data plots that need to be prepared as data is being collected and when all of the data is available. These plots are basically for QA/QC purposes and to demonstrate basic data behavior. These basic plots include: time series plots (data observations as a function of time), control plots (generally the same as time series plots, but using control samples and with standard deviation bands), probability plots (described below), scatter plots (described below), and residual plots (needed for any model building activity, especially for regression analyses).

Probability Plots

The most basic exploratory data analysis method is to prepare a probability plot of the available data. The plots indicate the possible range of the values expected, their likely probability distribution type, and the data variation. It is difficult to recommend another method that results in so much information using the data available. Histograms, for example, cannot accurately indicate the probability distribution type very accurately, but they more clearly indicate multi-modal distributions.

The values and corresponding probability positions are plotted on special normal-probability paper. This paper has a y-axis whose values are spread out for the extreme small and large probability values. When plotted on this paper, the values form a straight line if they are Normally distributed (Gaussian). If the points do not form an acceptably straight line, they can then be plotted on log-normal probability paper (or the data observations can be log transformed and plotted on normal probability paper). If they form a straight line on the log-normal plot, then the data is log-normally distributed. Other data transformations are also possible for plotting on normal-probability paper, but these two (normal and log-normal) usually are sufficient for most receiving water analyses.

Figures 1 and 2 are probability plots of stormwater data from the National Stormwater Quality Database (NSQD) (Maestre and Pitt 2005). These plots are for all conditions combined and represent several thousand observations. In most cases, it is obvious that normal probability plots do not indicate normal distributions, except for pH (which is already log-transformed). However, Figure 2 plots are log-normal probability plots and generally show much better normal distributions, as is common for stormwater data. However, some extreme values are still obviously not represented by log-normal probability distributions.

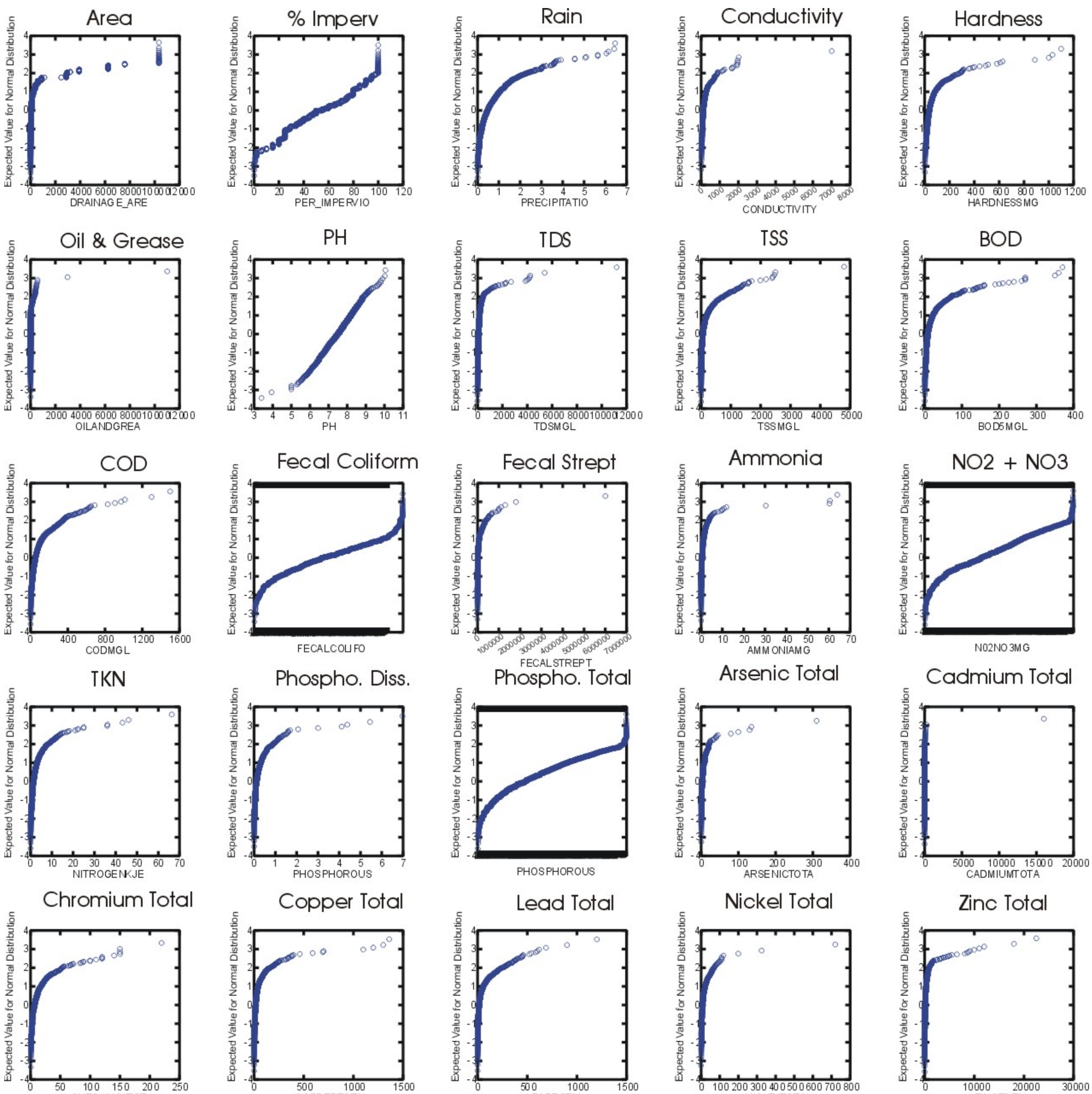


Figure 1. Probability plots of NSQD data (Maestre and Pitt 2005).

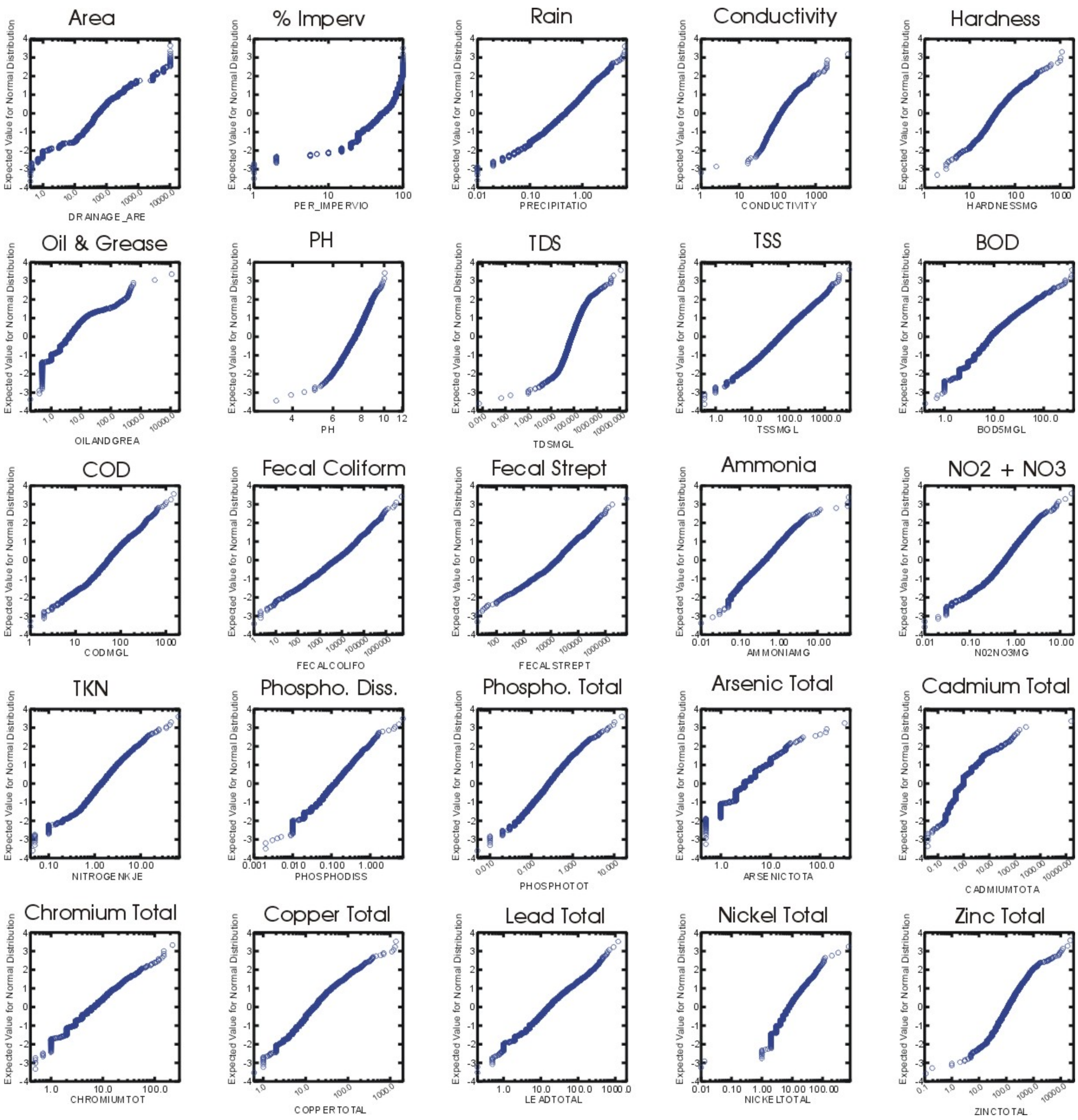


Figure 2. Log-probability plots of NSQD data (Maestre and Pitt 2005).

Figure 3 shows three types of results that can be observed when plotting pollutant reduction observations on probability plots, using data collected at the Monroe St. wet detention pond in Madison, WI, by the USGS and the WI DNR. Figure 3a for suspended solids (particulate residue) shows that SS are highly removed over a wide range of influent concentrations, ranging from 20 to over 1,000 mg/L. A simple calculation of percentage reduction would not show this consistent removal over the wide range. In contrast, Figure 3b for total dissolved solids (filtered residue) shows poor removal of TDS for all concentration conditions, as expected for this wet detention pond. The percentage removal for TDS would be close to zero and no additional surprises are indicated on this plot. Figure 3c, however, shows a wealth of information that would not be available from simple statistical numerical summaries. In this plot, filtered COD is seen to be poorly removed for low concentrations (less than about 20 mg/L, but the removal increases substantially for higher concentrations. Although not indicated on these plots, the rank order of concentrations were similar for both influent and effluent distributions for all three pollutants.

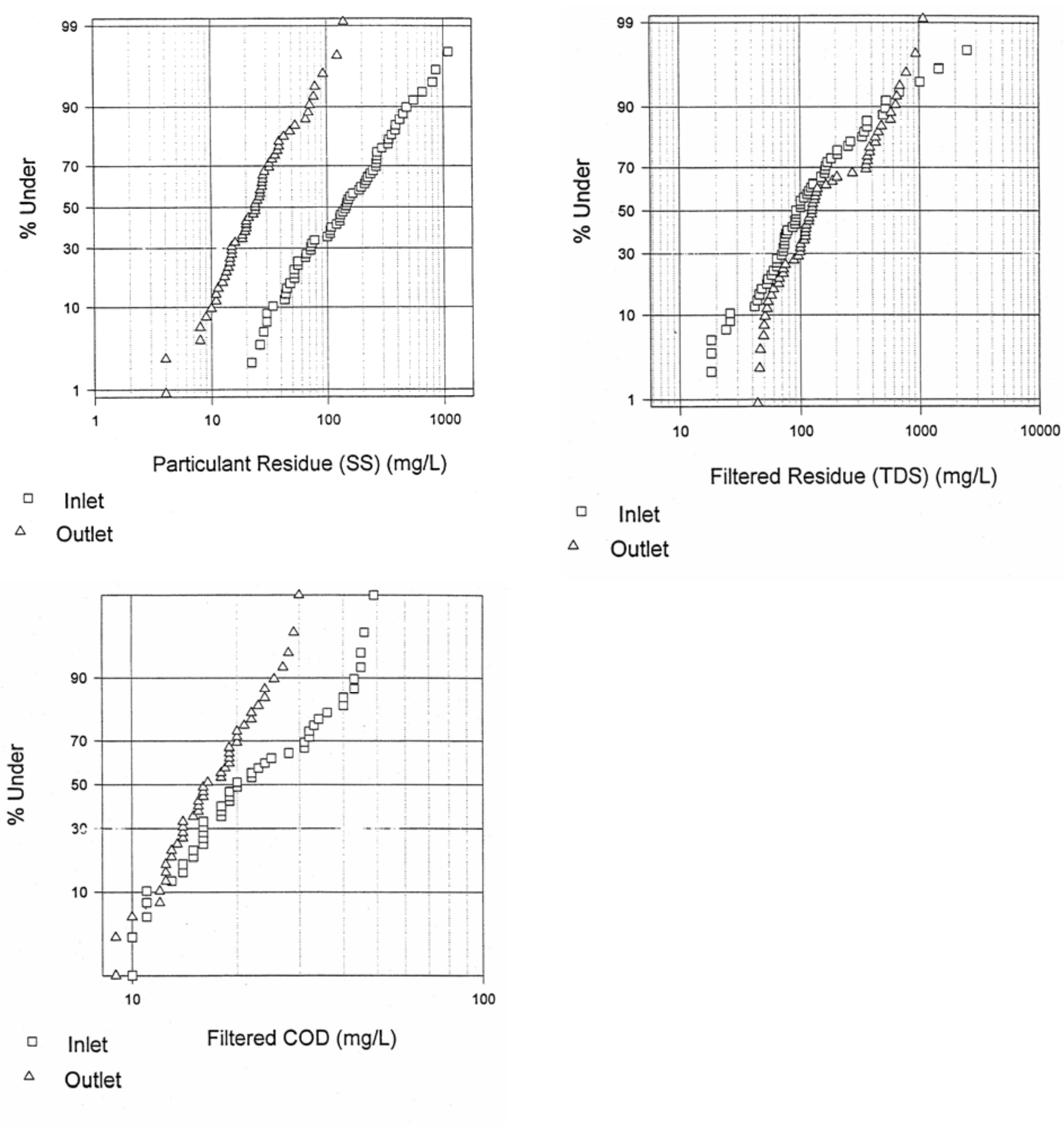


Figure 3. Influent and effluent observations for suspended solids, dissolved solids, and filtered COD at the Monroe St., Madison, WI, stormwater detention pond.

Generally, water quality observations do not form a straight line on normal probability paper, but do (at least from about the 10 to 90 percentile points) on log-normal probability paper. This indicates that the samples generally have a log-normal distribution and many parametric statistical tests can probably be used, but only after the data is log-transformed. These plots indicate the central tendency (median) of the data, along with their possible distribution type and variance (the steeper the plot, the smaller the COV and the flatter the slope of the plot, the larger the COV for the data). Multiple data sets can also be plotted on the same plot (such as for different sites, different seasons, different habitats, etc.) to indicate obvious similarities (or differences) in the data sets. Most statistical methods used to compare different data sets require that the sets have the same variances, and many require normal distributions. Similar variances would be indicated by generally parallel plots of the data on the probability paper, while normal distributions would be reflected by the data plotted in a straight line of normal probability paper.

Probability plots should be supplemented with standard statistical tests that determine if the data is normally distributed. These tests, at least some available in most software packages, include the Kolmogorov-Smirnov one-sample test, the chi-square goodness of fit test, and the Lilliefors variation of the Kolmogorov-Smironov test. They basically are paired tests comparing data points from the best-fitted normal curve to the observed data. The statistical tests may be visualized by imagining the best-fitted normal curve data and the observed data plotted on normal probability paper. If the observed data crosses the fitted curve data numerous times, it is much likely to be normally distributed than if it only crossed the fitted curve a few times.

Digidot Plot

Berthouex and Brown (1994) point out that since the best way to display data is with a plot, it makes little sense to present the data in a table. They highly recommend a digidot plot, developed by Hunter (1988) based on Tukey (1977), as a basic presentation of characterization data. This plot indicates the basic distribution of the data, shows changes with time, and presents the actual values, all in one plot. A data table is therefore not needed in addition to the digidot plot. A stem and leaf plot of the data is presented as the y-axis and the data are presented in a time series (in the order of collection) along the x-axis. Figure 4 is an example of a digidot plot, as presented by Berthouex and Brown (1994). The stem and leaf plot is constructed by placing the last digit of the value on the y-axis between the appropriate tic marks. In this example, the value 47 is represented with a 7 placed in the division between 45 and 50. Similarly, 33 is represented with a 3 placed in the division between 30 and 35. Values from 30 to 34 are placed between the 30 and 35 tic marks, while values from 35 to 39 are placed between the 35 and 40 tic marks. Simultaneously, the values are plotted in a time series in the order of collection. This plot can therefore be constructed in real time as the data is collected and obvious trends with time can be noted. This plot also presents the actual numerical data that can also be used in later statistical analyses.

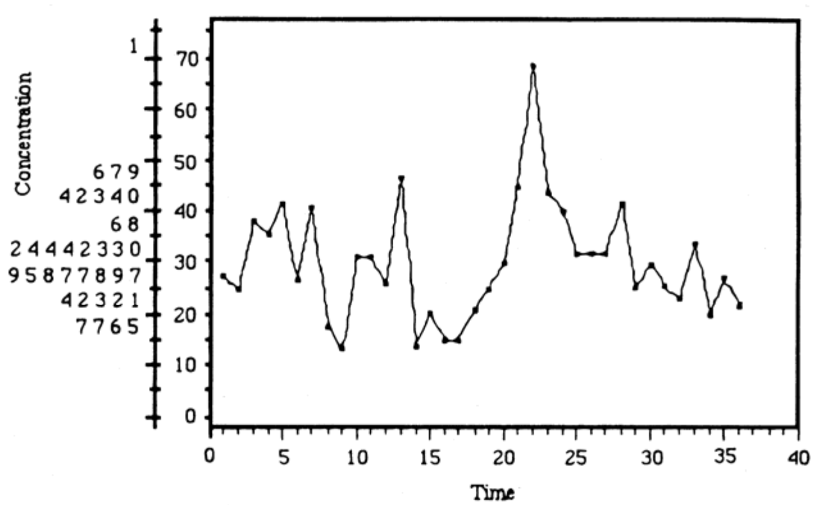


Figure 4. Digidot Plot (Berthouex and Brown 1994).

Scatterplots

According to Berthouex and Brown (1994), the majority of the graphs used in science are scatterplots. They stated that these plots should be made before any other analyses of the data is performed. Scatterplots are typically made by plotting the primary variable (such as a water quality constituent) against a factor that may influence its value (such as time, season, flow, another constituent like suspended solids, etc.). Figure 5 is a scatterplot showing COD values plotted against rain depth to investigate the possibility of a "first-flush," where higher concentrations are assumed to be associated with small runoff events (Pitt 1985). In this example, the smallest rains appear to have the highest COD concentrations associated with them, but the distribution of values is very wide. This may simply be associated with the much greater number of events observed having small rains and an increased likelihood of events having unusual observations to occur when more observations are made. When many data are observed for many sites, generally smaller rains do seem to be associated with the highest concentrations observed, but it is not a consistent pattern.

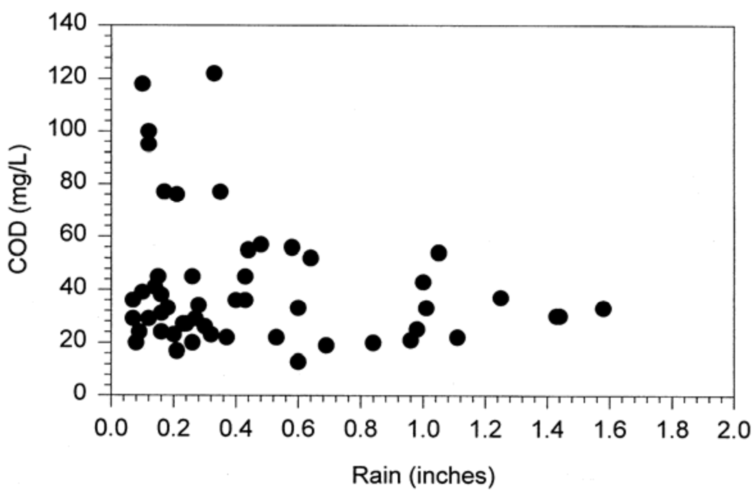


Figure 5. Scatterplot for Bellevue, Washington, COD stormwater concentrations, by rain depth (Pitt 1985).

Grouped scatterplots (miniatures) of all possible combinations of constituents can be organized as in a correlation matrix (Figure 6, Cleveland 1994). This arrangement allows obvious relationships to be easily seen, and even indicates if the relationships are straight-lined, or are curvilinear. In this example, the highest ozone values occur on days having the highest temperatures, and the lowest ozone concentrations occur on days having brisk winds and low temperatures. Figure 7 contains several scatterplots of NSQD data showing poor correlation of residential area stormwater concentration with rain depth (Maestre and Pitt 2005). Figure 8 are scatterplots used in QA/QC analyses of NSQD data showing reasonable relationships between constituents. In these cases, most of the dissolved copper and zinc concentrations are less than the concurrent total concentrations, as expected. Similarly, BOD5 is smaller than COD and ammonia is less than total Kjeldahl nitrogen values. Initially, several data sets were plotted with unreasonable relationships and review of the data indicated transcription errors that were corrected, for example.

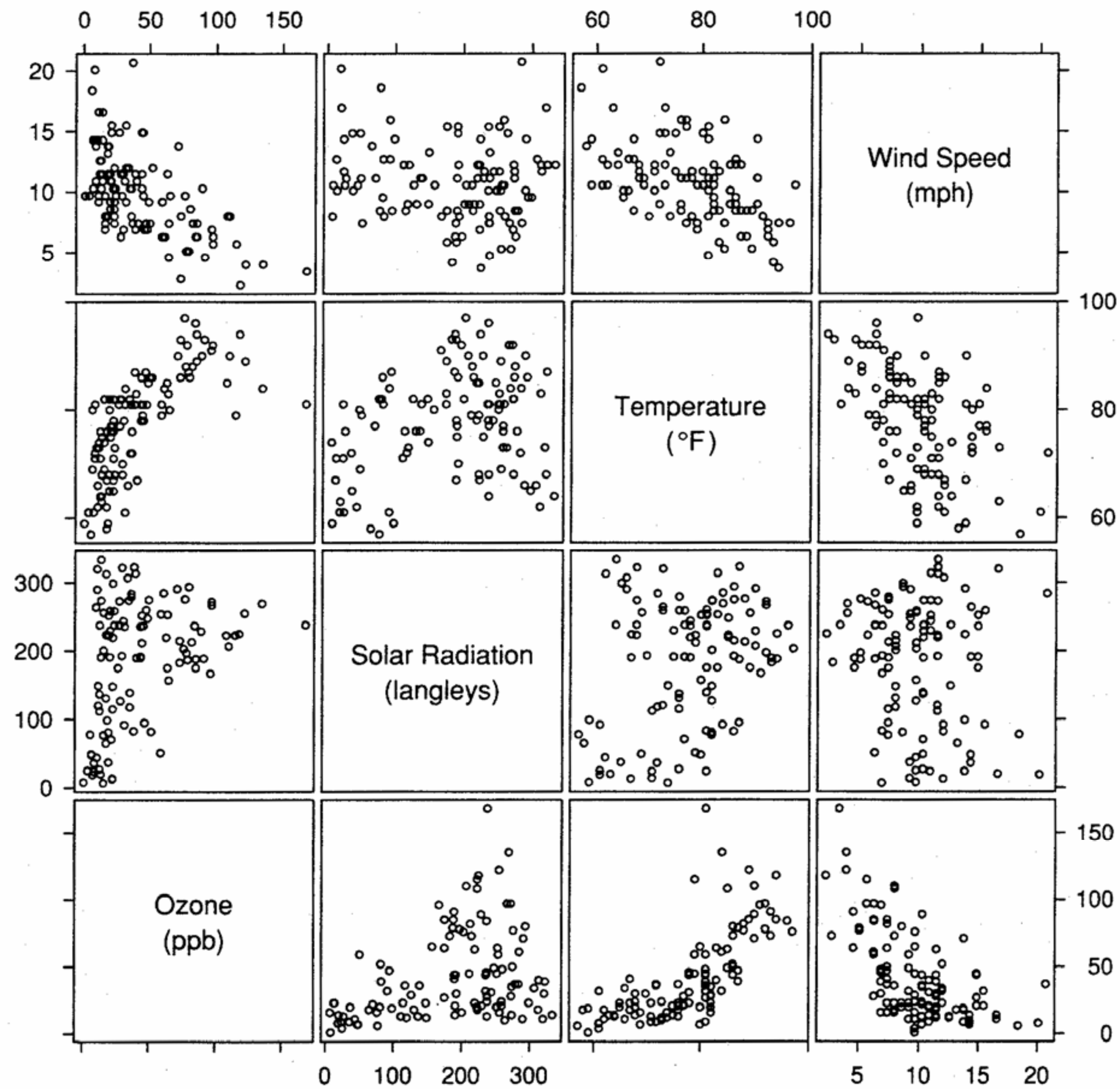


Figure 6. Grouped scatterplot for ozone, solar radiation, temperature, and wind speed (Cleveland 1994).

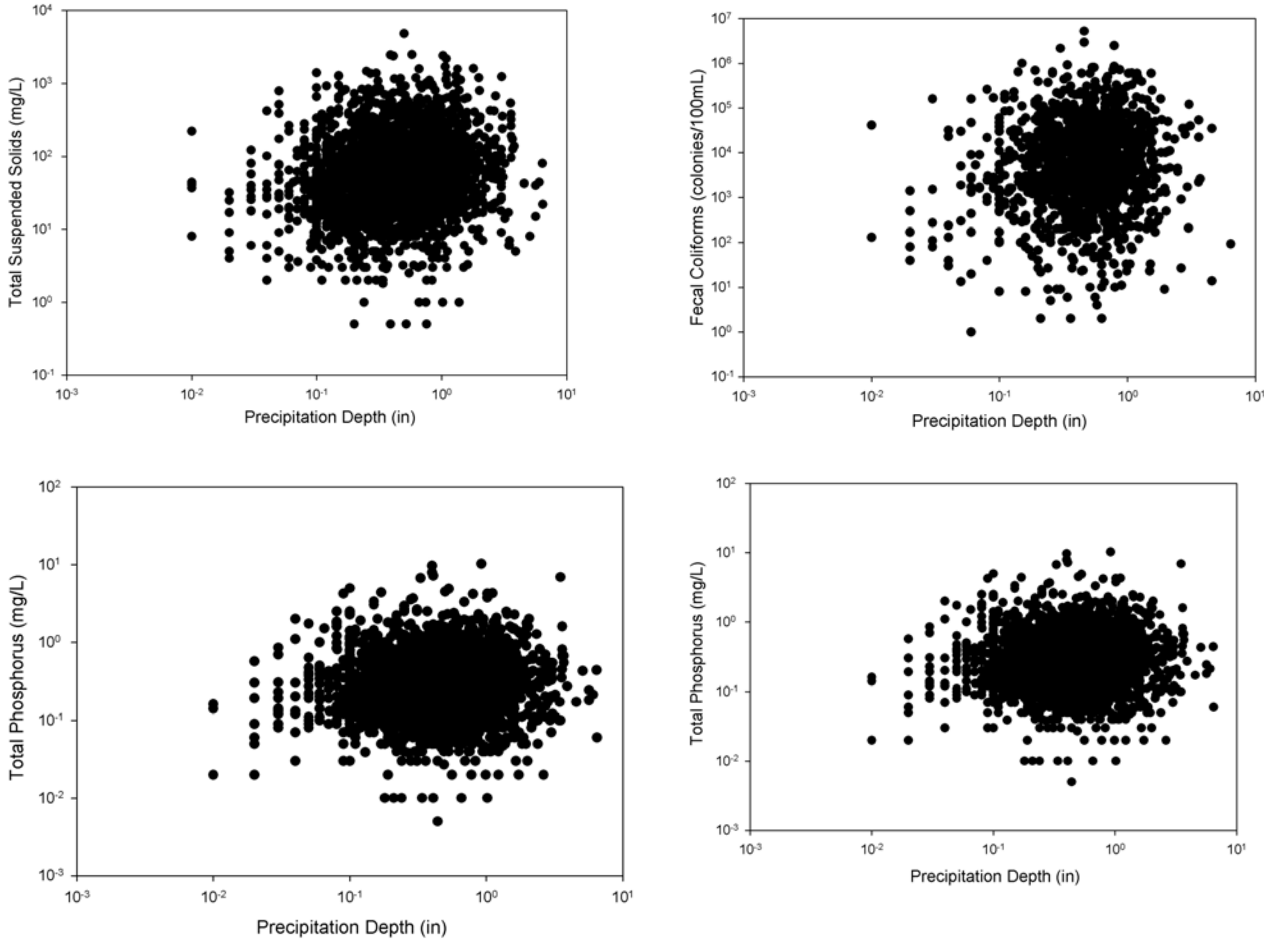


Figure 7. Scatterplots of NSQD data showing poor correlation of residential area stormwater concentration with rain depth (Maestre and Pitt 2005).

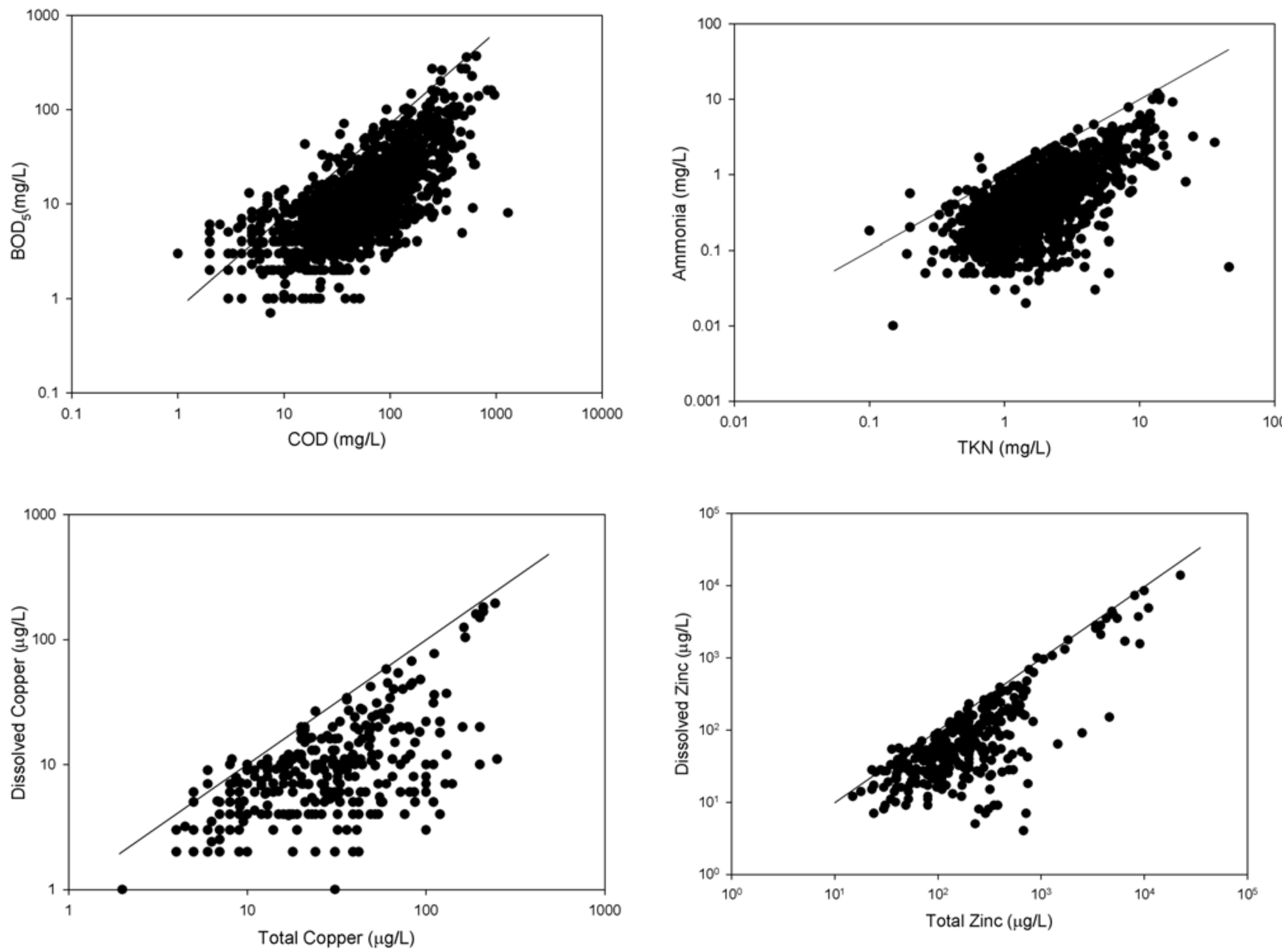


Figure 8. Scatterplots used in QA/QC analyses of NSQD data showing reasonable relationships between constituents (Maestre and Pitt 2005).

Grouped Box and Whisker Plots

Another primary exploratory data analysis tool, especially when differences between sample groups are of interest, is the use of grouped box and whisker plots. Examples of their use include examining different sampling locations (such as above and below a discharge), influent and effluent of a treatment process, different seasons, etc. These plots indicate the range and major percentile locations of the data, as shown on Figure 9 (Pitt 1985). In this example, seasonal groupings of stormwater quality observations for COD (Chemical Oxygen Demand) from Bellevue, Washington, were plotted to indicate obvious differences in the values. If the 75 and 25 percentile lines of the boxes do not overlap on different box and whisker plots, then the data groupings are likely significantly different (at least at the 95% level). When large numbers of data sets are plotted using box and whisker plots, the relative overlapping (or separation) of the plots can be used to identify possible groupings of the separate sets. In this case, there are no clear significant differences, but the summer season appears to have most of the highest concentrations observed.

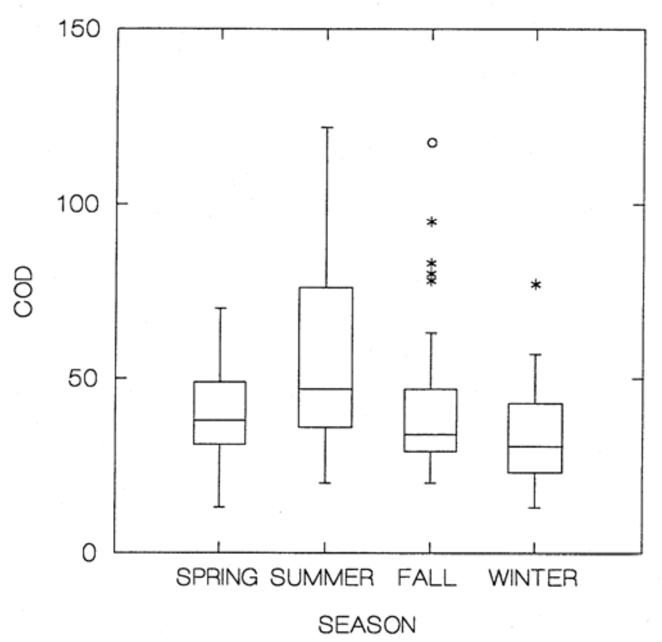


Figure 9. Grouped box and whisker plot for Bellevue, Washington, COD stormwater concentrations, by season (Pitt 1985).

To supplement the visual presentation with the grouped box and whisker plots, a one-way ANOVA test (or the Kurskal-Wallis ANOVA on ranks test) should be conducted to determine if there is any statistically significant difference between the different boxes on the plot. ANOVA doesn't specifically identify which sets of data are different from any other, however. A multiple comparison procedure (such as the Bonferroni *t*-test) can be used to identify significant differences between all cells if the ANOVA finds that a significance difference exists. Both of these tests (ANOVA and Bonferroni *t*-test) are parametric tests and require that the data be normally distributed. It may therefore be necessary to perform a log-transformation on the raw data. These tests will identify differences in sample groupings, but similarities (to combine data) are probably also important to know.

Figure 10 is a grouped box and whisker plot that shows significant differences in fluorescence values for groups of source waters. This was used in the inappropriate discharge study conducted by the Center for Watershed Protection and Pitt (2004) to distinguish groups of contaminated waters from clean water sources.

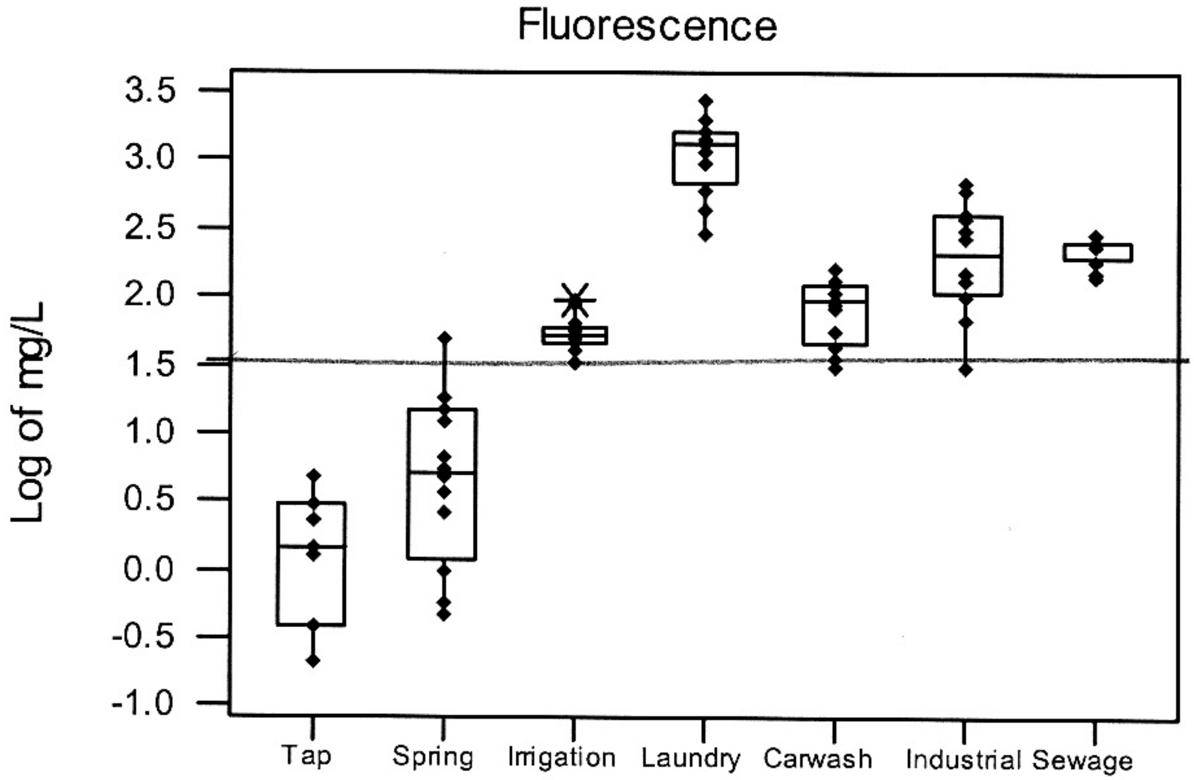


Figure 10. Grouped box and whisker plot indicating significant differences in fluorescence values for groups of source waters (CWP and Pitt 2004).

Comparing Multiple Sets of Data with Group Comparison Tests

Making comparisons of data sets are fundamental objectives of many receiving water investigations. Different habitats and seasons can produce significant affects on the observations. The presence of influencing factors, such as pollutant discharges or control practices, also affect the data observations. Berthouex and Brown (1994) and Gilbert (1987) present excellent summaries of the most common statistical tests that are used for these comparisons in environmental investigations. The significance of the test results (the α value, the confidence factor, along with the β value, the power factor) will indicate the level of confidence and power that the two sets of observations are the same. In most cases, an α level of less than 0.05 has been traditionally used to signify significant differences between two sets of observations, although this is an arbitrary criterion. In most cases, β is ignored (resulting in a default value of 1-β of 0.5), although some use a 1-β value of 0.8. An α value of 0.05 implies that the interpretation will be in error an average of 1 in 20 times. In some cases, this may be too conservative, while in others (such as where health and welfare implications are involved), it may be too liberal. The selection of the critical α value should be decided beforehand, while the calculated values for α should always be presented in the data evaluation (not simply stating that the results were significant or not significant at the 0.05 level, as is common). Even if the α level is significant, the magnitude of the difference, such as the pollutant reduction, may not be very important. The importance of the level of pollutant reductions should also be graphically presented using grouped box plots indicating the range and variations of the concentrations at each of the sampling locations, as described previously.

Comparison tests are divided into simple comparison tests between two groups (such as Student's t test) and tests that examine larger numbers of groups and interactions (such as Analysis of Variance Tests, or ANOVA).

Simple Comparison Tests with Two Groups

The main types of simple comparison tests are separated into independent and paired tests. These can be further separated into tests that require specific probability distribution characteristics (parametric tests) and tests that do not have as many restrictions based on probability distribution characteristics of the data (nonparametric data). If the parametric test requirements can be met, then they should be used as they have more statistical power. However, if information concerning the probability distributions is not available, or if the distributions do not behave correctly, then the somewhat less powerful nonparamteric tests should be used. Similarly, if the data gathering activity can allow for paired observations, then they should be used preferentially over independent tests.

In many cases, observations cannot be related to each other, such as a series of observations at two locations during all of the rains during a season. Unless the sites are very close together, the rains are likely to vary considerably at the two locations, disallowing a paired analysis. However, if data can be collected simultaneously, such as at influent and effluent locations for a (rapid) treatment process, paired tests can be used to control all factors that may influence the outcome, resulting in a more efficient statistical analysis. Paired experimental designs ensure that uncontrolled factors basically influence both sets of data observations equally (Berthouex and Brown 1994).

The parametric tests used for comparisons are the Student's *t*-tests (both independent and paired *t*-tests). All statistical analyses software and most spreadsheet programs contain both of these basic tests. These tests require that the variances of the sample sets be the same and are constant over the range of the values. These tests also require that the probability distributions be Gaussian (Normal). Transformations can be used to modify the data sets to these conditions. Log-transformations can be used to produce Gaussian distributions of most water quality data. Square root transformations are also commonly used to make the variance constant over the data range, especially for biological observations (Sokal and Rohlf 1969). In all cases, it is necessary to confirm these requirements before the standard *t*-tests are used.

Nonparametrics: Statistical Methods Based on Ranks by Lehman and D'Abrera (1975) is a comprehensive general reference on nonparametric statistical analyses. Gilbert (1987) presents an excellent review of nonparametric alternatives to the Student's *t*-tests, especially for environmental investigations from which the following discussion is summarized. Even though the nonparametric tests remove many of the restrictions associated with the *t*-tests, the *t*-tests should be used if justifiable. Unfortunately, seldom are the Student's *t*-test requirements easily met with environmental data and the slight loss of power associated with using the nonparametric tests is much more acceptable than misusing the Student's *t*-tests. Besides having few data distribution restrictions, many of the nonparametric tests can also accommodate a few missing data, or observations below the detection limits. The following paragraphs briefly describe the features of the nonparametric tests used to compare data sets.

Nonparametric Tests for Paired Data Observations. The sign test is the basic nonparametric test for paired data. It is simple to compute and has no requirements pertaining to data distributions. A few "not detected" observations can also be accommodated. Two sets of data are compared and the differences are used to assign a positive sign if the value in one data set is greater than the corresponding value in the other data set, or a negative sign is assigned if the one value is less than the corresponding value in the other data set. The number of positive signs are added and a statistical table (such as in Lehman and D'Abrera 1975, Table G shown below as Table 1) is used to determine if the number of positive signs found is unusual for the number of data pairs examined. This table shows that in order to have at least a 95% confidence that two sets of paired data are significantly different, only one out of eight pairs can have a larger data value in one set compared to the 7 larger ones in the other data set. As the number of pairs of observations increase, the allowable number of inconsistent values increases. With 40 pairs of observations, as many as 14 inconsistent values are allowed.

Table 1. Sign Test Statistical Tables (Lehman and D'Abrera 1975)

least # (of the pos, or -'s)

a \ N	2	3	4	5	6	7	8	9
0	.2500	.1250	.0625	.0313	.0156	.0078	.0039	.0020
1	.7500	.5000	.3125	.1875	.1094	.0625	.0352	.0195
2	1.0000	.8750	.6875	.5000	.3438	.2266	.1445	.0898
3		1.0000	.9375	.8125	.6562	.5000	.3633	.2539
4			1.0000	.9687	.8906	.7734	.6367	.5000

a \ N	10	11	12	13	14	15	16	17
0	.0010	.0005	.0002	.0001	.0001	.0000	.0000	.0000
1	.0107	.0059	.0032	.0017	.0009	.0005	.0003	.0001
2	.0547	.0327	.0193	.0112	.0065	.0037	.0021	.0012
3	.1719	.1133	.0730	.0461	.0287	.0176	.0106	.0064
4	.3770	.2744	.1938	.1334	.0898	.0592	.0384	.0245
5	.6230	.5000	.3872	.2905	.2120	.1509	.1051	.0717
6	.8281	.7256	.6128	.5000	.3953	.3036	.2272	.1662
7	.9453	.8867	.8062	.7095	.6047	.5000	.4018	.3145
8	.9893	.9673	.9270	.8666	.7880	.6964	.5982	.5000

a \ N	18	19	20	21	22	23	24	25
0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2	.0007	.0004	.0002	.0001	.0001	.0000	.0000	.0000
3	.0038	.0022	.0013	.0007	.0004	.0002	.0001	.0001
4	.0154	.0096	.0059	.0036	.0022	.0013	.0008	.0005
5	.0481	.0318	.0207	.0133	.0085	.0053	.0033	.0020
6	.1189	.0835	.0577	.0392	.0262	.0173	.0113	.0073
7	.2403	.1796	.1316	.0946	.0669	.0466	.0320	.0216
8	.4073	.3238	.2517	.1917	.1431	.1050	.0758	.0539
9	.5927	.5000	.4119	.3318	.2617	.2024	.1537	.1148
10	.7597	.6762	.5881	.5000	.4159	.3388	.2706	.2122
11	.8811	.8204	.7483	.6682	.5841	.5000	.4194	.3450
12	.9519	.9165	.8684	.8083	.7383	.6612	.5806	.5000

a \ N	26	27	28	29	30	31	32	33
0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
3	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
4	.0003	.0002	.0001	.0001	.0000	.0000	.0000	.0000
5	.0012	.0008	.0005	.0003	.0002	.0001	.0001	.0000
6	.0047	.0030	.0019	.0012	.0007	.0004	.0003	.0002
7	.0145	.0096	.0063	.0041	.0026	.0017	.0011	.0007
8	.0378	.0261	.0178	.0121	.0081	.0053	.0035	.0023
9	.0843	.0610	.0436	.0307	.0214	.0147	.0100	.0068
10	.1635	.1239	.0925	.0680	.0494	.0354	.0251	.0175
11	.2786	.2210	.1725	.1325	.1002	.0748	.0551	.0401
12	.4225	.3506	.2858	.2291	.1808	.1405	.1077	.0814
13	.5775	.5000	.4253	.3555	.2923	.2366	.1885	.1481
14	.7214	.6494	.5747	.5000	.4278	.3601	.2983	.2434
15	.8365	.7790	.7142	.6445	.5722	.5000	.4300	.3642
16	.9157	.8761	.8275	.7709	.7077	.6399	.5700	.5000

a \ N	34	35	36	37	38	39	40
0	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2	.0000	.0000	.0000	.0000	.0000	.0000	.0000
3	.0000	.0000	.0000	.0000	.0000	.0000	.0000
4	.0000	.0000	.0000	.0000	.0000	.0000	.0000
5	.0000	.0000	.0000	.0000	.0000	.0000	.0000
6	.0001	.0001	.0000	.0000	.0000	.0000	.0000
7	.0004	.0003	.0002	.0001	.0001	.0000	.0000
8	.0015	.0009	.0006	.0004	.0002	.0001	.0001
9	.0045	.0030	.0020	.0013	.0008	.0005	.0003
10	.0122	.0083	.0057	.0038	.0025	.0017	.0011
11	.0288	.0205	.0144	.0100	.0069	.0047	.0032
12	.0607	.0448	.0326	.0235	.0168	.0119	.0083
13	.1147	.0877	.0662	.0494	.0365	.0266	.0192
14	.1958	.1553	.1215	.0939	.0717	.0541	.0403
15	.3038	.2498	.2025	.1620	.1279	.0998	.0769
16	.4321	.3679	.3089	.2557	.2088	.1684	.1341
17	.5679	.5000	.4340	.3714	.3136	.2612	.2148
18	.6962	.6321	.5660	.5000	.4357	.3746	.3179
19	.8042	.7502	.6911	.6286	.5643	.5000	.4373
20	.8853	.8447	.7975	.7443	.6864	.6254	.5627

The Mann-Whitney signed rank test has more power than the sign test, but it requires that the data distributions be symmetrical (but with no specific distribution type). Without transformations, this requirement may be difficult to justify for water quality data. This test requires that the differences between the data pairs in the two data sets be calculated and ranked before checking with a special statistical table (as in Lehman and D'Abrera 1975). In the simplest case for monitoring the effectiveness of treatment alternatives, comparisons can be made of inlet and outlet conditions to determine the level of pollutant removal and the statistical significance of the concentration differences. StatXact-Turbo (CYTEL, Cambridge, MA) is a microcomputer program that computes exact nonparametric levels of significance, without resorting to normal approximations. This is especially important for the relatively small data sets that will typically be evaluated during most environmental research activities.

Friedman's test is an extension of the sign test for several related data groups. There are no data distribution requirements and the test can accommodate a moderate number of "non-detectable" values, but no missing values are allowed.

Nonparametric Tests for Independent Data Observations. As for the *t*-tests, paired test experimental designs are superior to independent designs for nonparametric tests because of their ability to cancel out confusing properties. However, paired experiments are not always possible, requiring the use of independent tests. The Wilcoxon rank sum test is the basic nonparametric test for independent observations. The test statistic is also easy to compute and compare to the appropriate statistical table (as in Lehman and D'Abrera 1975). The Wilcoxon rank sum test requires that the probability distributions of the two data sets be the same (and therefore have the same variances). There are no other restrictions on the data distributions (they do not have to be symmetrical, for example). A moderate number of "non-detectable" values can be accommodated by treating them as ties.

The Kruskal-Wallis test is an extension of the Mann-Whitney rank sum test and allows evaluations of several independent data sets, instead of just two. Again, the distributions of the data sets must all be the same, but they can have any shape. A moderate number of ties and non-detectable values can also be accommodated.

Comparisons of Many Groups

If there are more than two groups of data to be compared (such as in-stream concentrations at several locations along a river, each with multiple observations), one of the analysis of variance, or ANOVA, tests should be used. The commonly available one-way, two-way, and three-way ANOVA tests are parametric tests and require that the data in each grouping be normally distributed and that the variances be the same in each group. This can be visually examined by preparing a probability plot for the data in each group displayed on the same chart. The probability plots would need to be parallel and straight. Obviously, log transformations of the data can be used if assumptions are met when the data is plotted using log-normal probability axes. On Figure 3a, the influent and effluent probability plots for suspended solids at the Monroe St. wet detention pond site in Madison, WI, the probability plots are reasonably parallel and straight when plotted as log-normal plots. However, Figure 3c, a similar plot for dissolved COD, indicates that the plots are not parallel. Of course, these figures only contain two groupings of data (influent and effluent) and one of the previous two-group tests would be more efficient for this data.

If data from multiple stations along a river were collected during different seasons, it would be possible to use the two-way ANOVA test to examine the effects of different seasons and different locations, along with the interaction of these parameters. Three-way ANOVA tests can be used to evaluate the results of similar field sampling data (different locations, different seasons) and another factor, such as natural vs. artificial substrate samplers for benthic macroinvertebrates (or seining vs. electro-shocking for fish sampling). These tests would then indicate if the results from these different sampling procedures varied significantly by season, or sampling location. These analyses are more flexible than the factorial tests, as the factorial tests are most commonly only used for two levels (such as winter vs. summer; pools vs. riffles; and artificial substrate vs. natural substrate samplers). Factorial tests are more complicated when intermediate, or more than 2 levels, are being considered. However, the ANOVA tests are parametric tests and require multiple observations in each group, while the factorial tests are not and can be used with single observations per group (although that may not be a good idea considering the expected high variability in most environmental sampling).

A non-parametric test, usually included in statistical programs, for comparing many groups is the Kruskal-Wallis ANOVA on ranks test. This is only a one-way ANOVA test and would be only suitable for comparing data from different sampling sites alone, for example. This would be a good test to supplement grouped box and whisker plots.

Grouped comparison tests indicate only that at least one of the groups is significantly different from at least one other, they do not indicate which ones. For that reason, some statistical programs also conduct multiple comparison tests. SigmaStat, for example, offers: the Tukey test, Student-Newman-Keuls test, Bonferroni t-test, Fisher's LDS, Dunner's test, and Duncan's multiple range test. These tests basically conduct comparisons of each group against each other group and identify which are different.

Data Associations

Identifying patterns and associations in data may be considered a part of exploratory data analyses, but many of the tools (especially cluster, principal component, and factor analyses) may require specialized procedures having multiple data handling options that are not available in all statistical software packages, while some (such as correlation matrices discussed here) are commonly available.

Identifying data associations, and possible subsequent model building, is another area of interest to many investigators examining receiving water conditions. This is a critical component of the "weight-of-evidence" approach for identifying possible cause and effect relationships. The following are possible steps for investigating data associations:

1) re-examine the hypothesis of cause and effect (an original component of the experimental design previously conducted and was the basis for the selected sampling activities).
2) prepare preliminary examinations of the data, as described previously (most significantly, prepare scatter plots and grouped box/whisker plots).
3) conduct comparison tests to identify significant groupings of data. As an example, if seasonal factors are significant, then cause and effect may vary for different times of the year.
4) conduct correlation matrix analyses to identify simple relationships between parameters. Again, if significant groupings were identified, the data should be separated into these groupings for separate analyses, in addition to an overall analysis.
5) further examine complex inter-relationships between parameters by possibly using combinations of hierarchical cluster analyses, principal component analyses (PCA), and factor analyses.
6) compare the apparent relationships observed with the hypothesized relationships and with information from the literature. Potential theoretical relationships should be emphasized.
7) develop initial models containing the significant factors affecting the parameter outcomes. Simple apparent relationships between dependent and independent parameters should lead to reasonably simple models, while complex relationships will likely require further work and more complex models.

The following sections briefly describe these tools and present some interesting examples of their use.

Correlation Matrices

Knowledge of the correlations between data elements is very important in many environmental data analyses efforts. They are especially important when model building, such as with regression analysis. When constructing a model, it is important to include the important factors in the model, but the factors should be independent. Correlation analyses can assist by identifying the basic structure of the model.

Table 2 (Pitt 1987) is a standard correlation matrix that shows the relationships between measured rain and measured runoff parameters. This is a common Pearson correlation matrix, constructed using the microcomputer program SYSTAT (SPSS, Inc. Chicago, IL). It measures the strength of association between the variables. The Pearson correlation coefficients vary from -1 to +1. A coefficient of 0 indicates that neither of the two variables can be predicted from the other using a linear equation, while values of -1 or +1 indicate that perfect predictions can be made of one variable by only using the other variable. This example shows several very high correlations between pairs of parameters (>0.9). The paired parameters having high correlations are the same for both sites, indicating the same basic processes for rainfall-runoff. High correlations are seen between total runoff depth (RUNTOT) and rain depth (RAINTOT) and between runoff duration (RUNDUR) and rain duration (RAINDUR).

Table 2. Pearson Correlation Matrix (Pitt 1987)

Emery (Industrial)

	RAINTOT	RAINDUR	AVEINT	PEAKINT	DRYPER	RUNTOT	RUNDUR	AVEDIS	PEAKDIS	LAG
RAINTOT	1.000									
RAINDUR	0.533	1.000								
AVEINT	0.138	−0.387	1.000							
PEAKINT	0.512	−0.039	0.675	1.000						
DRYPER	0.169	0.273	−0.096	−0.132	1.000					
RUNTOT	0.906	0.562	0.007	0.405	0.075	1.000				
RUNDUR	0.501	0.965	−0.348	0.035	0.184	0.556	1.000			
AVEDIS	0.709	−0.013	0.480	0.654	−0.095	0.680	−0.026	1.000		
PEAKDIS	0.729	0.129	0.372	0.748	0.041	0.699	0.150	0.849	1.000	
LAG	0.135	0.220	−0.292	−0.217	0.052	0.205	0.134	0.098	0.107	1.000

Thistledowns (Residential/Commercial)

	RAINTOT	RAINDUR	AVEINT	PEAKINT	DRYPER	RUNTOT	RUNDUR	AVEDIS	PEAKDIS	LAG
RAINTOT	1.000									
RAINDUR	0.553	1.000								
AVEINT	0.321	−0.295	1.000							
PEAKINT	0.564	−0.104	0.827	1.000						
DRYPER	0.281	0.308	−0.190	−0.122	1.000					
RUNTOT	0.903	0.448	0.187	0.551	0.283	1.000				
RUNDUR	0.508	0.989	−0.322	−0.148	0.337	0.402	1.000			
AVEDIS	0.398	−0.178	0.593	0.817	−0.037	0.585	−0.227	1.000		
PEAKDIS	0.600	−0.051	0.659	0.917	0.009	0.702	−0.106	0.946	1.00	
LAG	−0.192	−0.037	−0.114	−0.202	−0.122	−0.184	−0.094	−0.138	−0.173	1.000

It is very important not to confuse correlation with causation. Box, *et al.* (1978) presents a historical example of a plot (Figure 11) of the population of Oldenburg, Germany, against the number of storks observed in each year. In this example, few would conclude that the high correlation between the increased number of storks observed and the simultaneous increase in population is a cause and effect relationship. The two variables observed are most likely related to another factor (such as time in this example, as both sets of populations increased over the years from 1930 to 1936). However, many investigators make similar improper assumptions of cause and effect from their observations, especially if high correlations are found. It is extremely important that theoretical knowledge of the system being modeled be considered. If this knowledge is meager, then specific tests to directly investigate cause and effect relationships must be conducted.

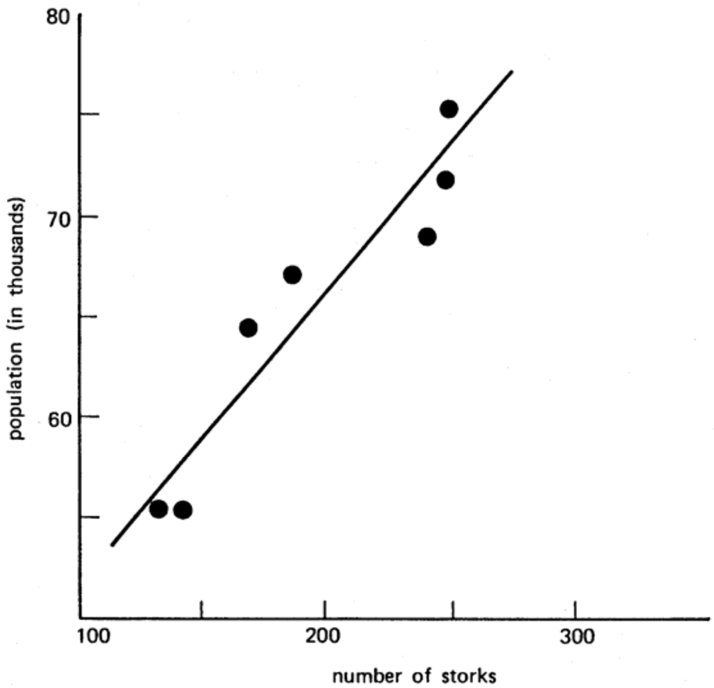


Figure 11. Possible cause and effect confusion from correlation tests (Box, *et al*. 1978).

Hierarchical Cluster Analyses

Another method to examine correlations between measured parameters is by using hierarchical cluster analyses. Figure 12 (Pitt 1987) is a tree diagram (dendogram) produced by SYSTAT using the same data as presented in the correlation matrix. A tree diagram illustrates both simple and complex correlations between parameters. Parameters having short branches linking them are more closely correlated than parameters linked by longer branches. In addition, the branches can encompass more than just two parameters. The length of the short branches linking only two parameters are indirectly comparable to the correlation coefficients (short branches signify correlation coefficients close to 1). The main advantage of a cluster analyses is the ability to identify complex correlations that cannot be observed using a simple correlation matrix. In this example, the rain total - runoff total and runoff duration - rain duration high correlation coefficients found previously are also seen to have simple relationships. In contrast, predicting peak runoff rates (PEAKDIS) requires more complex information. Therefore, the model used to predict peak runoff would have to be more complex, requiring additional information than required to just predict total runoff. Figure 13 is a cluster analysis from the National Stormwater Quality Database (NSQD) (Maestre and Pitt 2005) relating different stormwater constituent concentrations, rainfall, and site characteristics. Table 3 is an output from SYSTAT showing the distances of the joining branches. More detailed tables are available showing other joined constituents. Nitrogen compounds are closely related to rainfall conditions, but other constituents are more distantly related to each other. More detailed statistical analyses were conducted by Maestre and Pitt (2005) to examine other factors (such as geographical location, season, etc.).

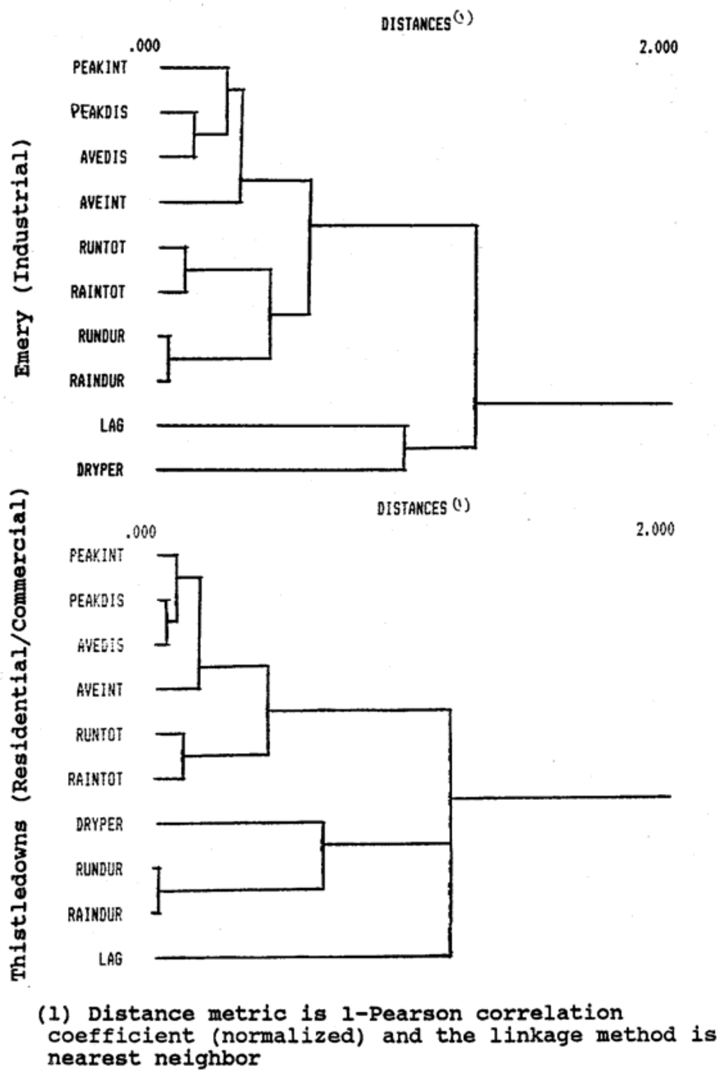


Figure 12. Tree diagram from cluster analyses of Toronto rainfall and runoff parameters (Pitt 1987).

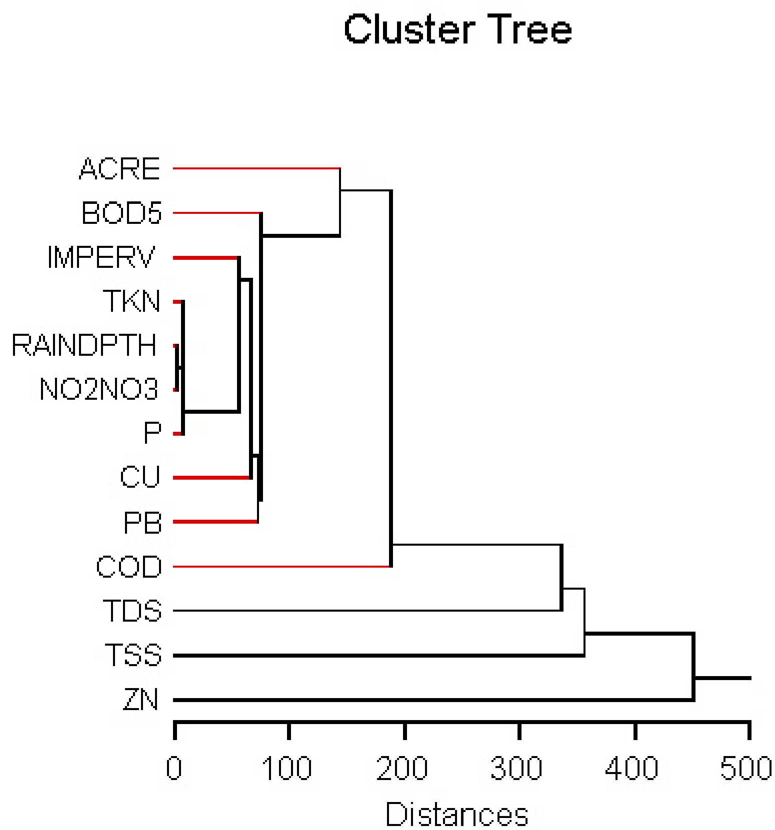


Figure 13. Cluster analysis for stormwater samples from the National Stormwater Quality Database (Maestre and Pitt 2005).

Table 3. SYSTAT Summary Table for Cluster Analysis

```
Distance metric is Euclidean distance
Single linkage method (nearest neighbor)
```

Cluster containing	and	Cluster containing	Were joined at distance	No. of members in new cluster
NO2NO3		RAINDPTH	1.960	2
TKN		NO2NO3	7.337	3
P		TKN	7.504	4
P		BOD5	76.164	5
P		COD	188.538	6
TDS		P	473.486	7
TDS		ZN	600.646	8
TSS		TDS	999.110	9

Principal Component Analyses (PCA) and Factor Analyses
Another important tool to identify relationships and natural groupings of samples or locations is with principal component analyses (PCA). Normally, data is autoscaled before PCA in order to remove the artificially large influence of constituents having large values compared to constituents having small values. PCA is a sophisticated procedure where information is sorted to determine the components (usually constituents) needed to explain the variance of the data. Typically, very large numbers of constituents are available for PCA analyses and a relatively small number of sample groups are to be identified. Salau, *et al.* (1997) used PCA (and then cluster analyses) to identify characteristics of sediment off Spain. Figure 14 shows the first two component loadings (collectively comprising most of the information) for about 60 constituents. The first principal component (PC1) is seen to be a near reversed image of the second principal component (PC2) (if a constituent is very important in one PC, it should be much less important in the other). Figure 15 shows a scatter plot of PC1 vs. PC2 values for different sample locations, showing how there are three main groups of samples, which generally corresponded to two sampling areas, plus a third group. The third group was then further analyzed using cluster analysis to examine more complex groupings and sampling subareas, as shown in the dendogram of Figure 16.

	%	cum		%	cum
PC1	75.4	75.4	PC3	5.2	89.4
PC2	8.8	84.2	PC4	3.8	93.2

%, percent of variance; cum, cumulative variance.

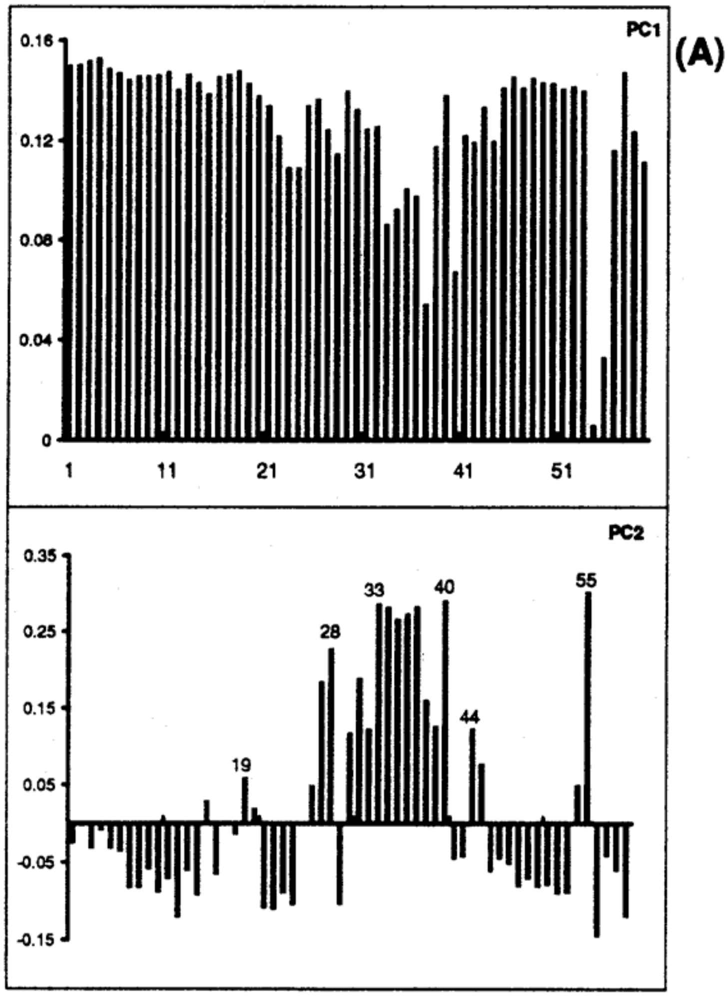


Figure 14. Loadings of principal components (Salau, *et al.* 1997).

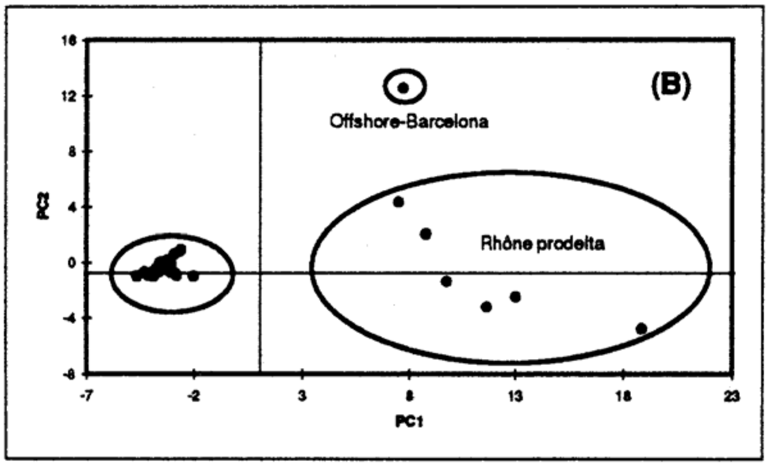


Figure 15. Score plots of principal components (Salau, *et al.* 1997).

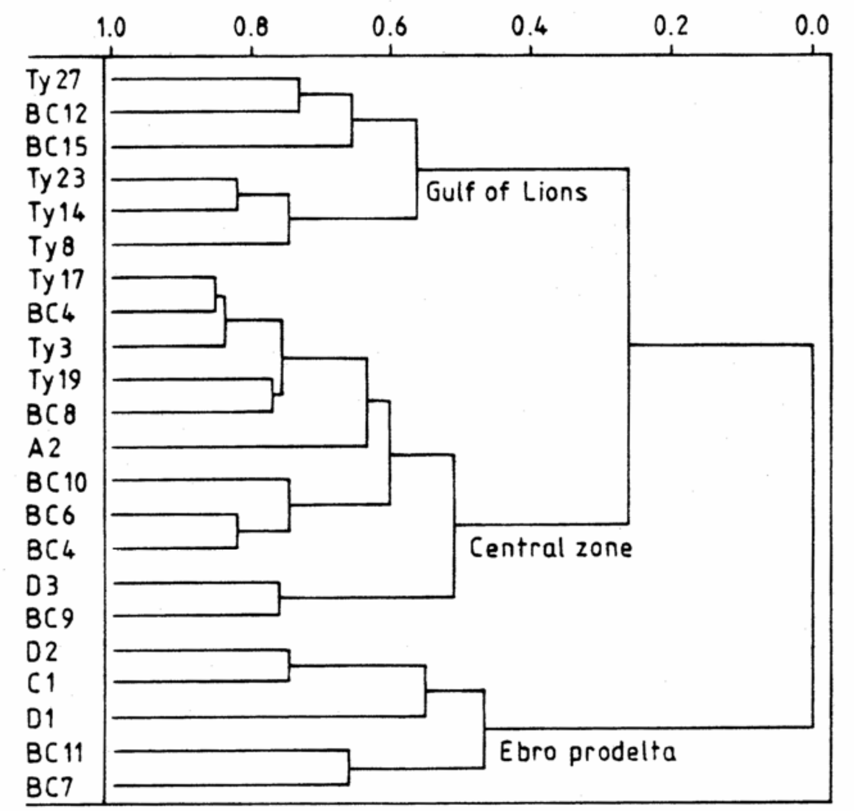


Figure 16. Dendogram of data, without two major groupings (Salau, *et al.* 1997).

Table 4 shows the latent roots (eigenvalues) and component loadings for a principal component analysis of the NSQD data. This shows that the first five components explained about 56% of the total variance of all the data. Hopefully, most of the variability would be explained with just the first few components. In this example, the first component (with 15% of the total variance explained) is mostly comprised of COD and BOD_5 values. TSS is spread out amongst at least three of the top five principle components.

Figure 17 is a scree plot produced by SYSTAT as part of the principle component analyses and shows the accumulative effect of additional factors in reducing variability for the NSQD data (Maestre and Pitt 2005). In this case, most of the components had similar benefits. It would be desirable to have a plot that was more concave, with much greater benefits associated with fewer initial components, and the accumulative effects tapering off for the later added factors.

Table 4. Principal Component SYSTAT Summary for NSQD Data (Maestre and Pitt 2005)

Latent Roots (Eigenvalues)

	1	2	3	4	5
	1.798	1.489	1.200	1.153	1.063
	6	7	8	9	10
	0.970	0.938	0.878	0.854	0.802
	11	12			
	0.496	0.357			

Component loadings

	1	2	3	4	5
COD	0.838	0.032	-0.303	-0.130	-0.045
BOD5	0.785	0.073	-0.402	-0.204	0.018
IMPERV	-0.130	-0.773	-0.250	0.046	0.133
ORDER	0.050	-0.762	0.103	-0.397	-0.008
ACRE	0.065	-0.029	0.459	-0.690	-0.329
NO2NO3	0.172	-0.151	0.227	0.220	-0.519
RAINDPTH	-0.168	0.272	0.093	-0.423	0.433
TDS	0.266	0.069	0.360	0.229	-0.382
P	0.350	0.055	0.427	0.190	0.331
TSS	0.280	0.134	0.396	-0.081	0.328
ZN	0.325	-0.338	0.258	0.369	0.276
TKN	0.147	-0.259	0.260	0.040	0.198

Variance Explained by Components

	1	2	3	4	5
	1.798	1.489	1.200	1.153	1.063

Percent of Total Variance Explained

	1	2	3	4	5
	14.987	12.407	9.997	9.612	8.859

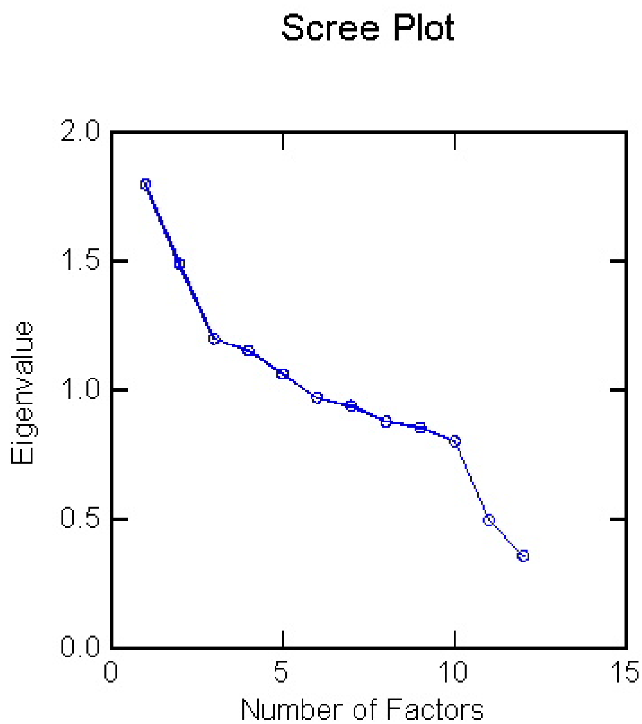


Figure 17. Scree plot showing accumulative effect of additional factors in reducing variability (Maestre and Pitt 2005).

Analysis of Trends in Receiving Water Investigations

The statistical identification of trends is very demanding. Several publications have excellent descriptions of statistical trend analyses for water quality data (as summarized by Pitt 1995). In addition to containing detailed descriptions and examples of experimental design methods to determine required sampling effort, Gilbert (1987) devotes a large portion of his book to detecting trends in environmental data and includes the code for a comprehensive computer program for trend analysis. Reckhow and Stow (1990) present a comprehensive assessment of the effectiveness of different water quality monitoring programs in detecting water quality trends using EPA STORET data for several rivers and lakes in North Carolina. They found that most of the data (monthly phosphorus, nitrogen, and specific conductance values were examined) exhibited seasonal trends and inverse relations with flow. In many cases, large numbers of samples would be needed to detect changes of 25 percent or less (typical for stormwater retro-fitting activities).

Spooner and Line (1993) present recommendations for monitoring requirements in order to detect trends in receiving water quality associated with nonpoint source pollution control programs, based on many years experience with the Rural Clean Water Program. These recommendations, even though derived from rural experience, should also be very applicable for urban receiving water trend analyses. The following is a general list (modified) of their recommended data needs for associating water quality trends with land use/treatment trends:

• Appropriate and sufficient control practices need to be implemented. A high level of participation/control implementation is needed in the watershed to result in a substantial and more easily observed water quality improvement. Controls need to be used in areas of greatest benefit (critical source areas, or in drainages below major sources) and most of the area must be treated.

• Control practice and land use monitoring is needed to separate and quantify the effects of changes in water quality due to the implemented controls by reducing the statistical confusion from other major factors. Monitor changes in land use and other activity on a frequent basis to observe temporal changes in the watershed. Seasonal variations in runoff quality can be great, along with seasonal variations in pollutant sources (monitor during all flow phases, such as during dry weather, wet weather, cold weather, warm weather, for example). Collect monitoring data and implement controls on a watershed basis.

• Monitor the pollutants affecting the beneficial uses of the receiving waters. Conduct the trend analyses for pollutants of concern, not just for easy, or convenient, parameters.

• Monitor for multiple years (at least 2 to 3 years for both pre- and post-control implementation) to account for year-to-year variability. Utilize a good experimental design, with preferable use of parallel watersheds (one must be a control and the other undergoing treatment).

Preliminary Evaluations before Trend Analyses are Used

Gilbert (1987) illustrates several sequences of water quality data that can confuse trend analyses. It is obviously easiest to detect a trend when the trend is large and the random variation is very small. Cyclic data (such as seasonal changes) often are confused as trends when no trends exist (type 1 error) or mask trends that do exist (type 2 error) (Reckhow and Stow 1990; Reckhow 1992). Three data characteristics need to be addressed before the data can be analyzed for trends because of confusing factors. These include:

• Measure data correlations, as most statistical tests require uncorrelated data. If data are taken close together (in time or in location), they are likely partially correlated. As an example, it is likely that a high value is closely surrounded by other relatively high values. Close data can therefore be influenced by each other and do not provide unique information. This is especially important when determining confidence limits of predicted values or when determining the number of data needed for a trend analyses (Reckhow and Stow 1990). Test statistics developed by Sen can use dependent data, but they may require several hundred data observations to be valid (Gilbert 1987).

• Remove any seasonal (or daily) effects, or select a data analysis procedure that is unaffected by data cycles. The nonparametric Sen test can be used when no cycles are present, or if cyclic effects are removed, while the seasonal Kendall test is not affected by cyclic data (Gilbert 1987).

• Identify any other likely predictable effects on concentrations and remove their influence. Normally occurring large variations in water quality data easily mask commonly occurring subtle trends. Typical relations between water quality and flow rate (for flowing water) can be detected by fitting a regression equation to a concentration vs. flow plot. The residuals from subtracting the regression from the data are then tested for trends using the seasonal Kendall test (Gilbert 1987).

Reckhow (1992) presents a chart listing specific steps that need to be taken to address the above problems. These steps are as follows:

(1) Check the data for deterministic patterns of variability (such as concentration versus flow by using graphical and statistical methods). If deterministic patterns exist, subtract the modeled pattern from the original data, leaving the residuals for subsequent seasonality analyses.

(2) Examine the remaining residuals (or data, if no deterministic patterns exist) for seasonal (can be short period, such as daily) variations. Again use graphical and statistical methods. If "seasonality" exists, subtract the modeled seasonality from the data (residuals from #1 above), leaving the remaining residuals for subsequent trend analyses.

(3) Conduct the trend analysis on the residuals from #2 above, using the standard seasonal Kendall test . If a trend exists, subtract the trend, leaving the remaining residuals for subsequent autocorrelation analyses.

(4) Test the remaining residuals from #3 above (or the raw data, if no deterministic or cyclic patterns or trends were found) for autocorrelation. If the autocorrelation is significant, re-evaluate the trends using an autocorrelated-corrected version of the seasonal Kendall (or regular Kendall) test. If no autocorrelation was found, use the standard seasonal Kendall test if seasonality was identified, or the standard Kendall test if no seasonality was identified. The final residual variation is then used (after correcting for autocorrelation) in calculating the required number of samples needed to detect trends for similar situations.

Statistical Methods Available for Detecting Trends

Graphical methods. Several sophisticated graphical methods are available for trend analyses that use special smoothing routines to reduce short-term variations so the long-term trends can be seen (Gilbert 1987). In all cases, simple plots of concentrations versus time of data collection should be made. This will enable obvious data gaps, potential short-term variations, and distinct long-term trends to be possibly seen.

Regression methods. A time-honored approach in trend analysis is to perform a least-squares linear regression on the quality versus time plot and to conduct a *t*-test to determine if the true slope is not different from zero (Gilbert 1987). However, Gilbert (1987) points out that the *t*-test can be misleading due to cyclic data, correlated data, and data that are not normally distributed.

Mann-Kendall test. This test is useful when missing data occur (due to gaps in monitoring, such as if frozen waters occur during the winters, equipment failures, or when data are reported as below the limit of detection). Besides missing data, this test can also consider multiple data observations per time period. This test also examines trends at multiple stations (such as surface waters and deep waters, etc.) and enables comparisons of any trends between the stations. This method also is not sensitive to the data distribution type. This test can be considered a nonparametric test for zero slope of water quality versus time of sample collection (Gilbert 1987). Short-term (such as seasonal changes) cycles and other data relationships (such as flow versus concentration) affect this test and must be corrected. If data are highly correlated, then this test can be applied to median values in each discrete time groupings.

Sen's nonparametric estimator of slope. Being a nonparametric test based on ranks, this method is not sensitive to extreme values (or gross data errors) when calculating slope (Gilbert 1987). This test can also be used when missing data occur in the set of observations. It is closely related to the Mann-Kendall test.

Seasonal Kendall test. This method is preferred to most regression methods if the data are skewed, serially correlated, or cyclic (Gilbert 1987). This test can be used for data sets having missing values, tied values, censored values (less than detection limits) or single or multiple data observations in each time period. The testing of homogeneity of trend direction enables one to determine if the slopes at different locations are the same, when seasonality is present. Data correlations (such as flow versus concentration) and dependence also affect this test and must be considered in the analysis.

The code for the computer program contained in Gilbert (1987) computes Sen's estimator of slope for each station-season combination, along with the seasonal Kendall test, Sen's aligned test for trends, the seasonal Kendall slope estimator for each station, the equivalent slope estimator for each season, and confidence limits on the slope.

Example of Long-Term Trend Analyses for Lake Rönningesjön, Sweden

An example showing the use of trend analyses for investigating receiving water effects of stormwater is presented here, using a Swedish lake example that has undergone stormwater treatment (Pitt 1995). The significant beneficial use impairment issue is decreasing transparency associated with eutrophication. The nutrient enrichment was thought to have been aggravated by stormwater discharges of phosphorus. Stormwater treatment was shown to decrease the phosphorus discharges in the lake, with an associated increase in transparency. The data available includes nutrient, chlorophyll *a*, transparency, and algal evaluations conducted over a 20 to 30 year period, plus treatment plant performance information for 10 years of operation. This trend evaluation was conducted by Pitt (1995) using data collected by Swedish researchers, especially Enell and Henriksson-Fejes (1989-1992).

A full-scale plant, using the Karl Dunkers' system for treatment of separate stormwater (the Flow Balancing Method, or FBM) and lake water, has been operating since 1981 in Lake Rönningesjön, Taby (near Stockholm), Sweden. The FBM and the associated treatment system significantly improved lake water quality through direct treatment of stormwater and by pumping lake water through the treatment system during dry weather. Figure 18 is an illustration of an idealized FBM system showing how inflowing stormwater is routed though a series of inter-connected compartments, before being discharged to the lake. A pump can also be used to withdraw water from the first compartment to a treatment facility. Figure 19 is a photograph of a FBM installation located at Lake Trehormingen, Sweden.

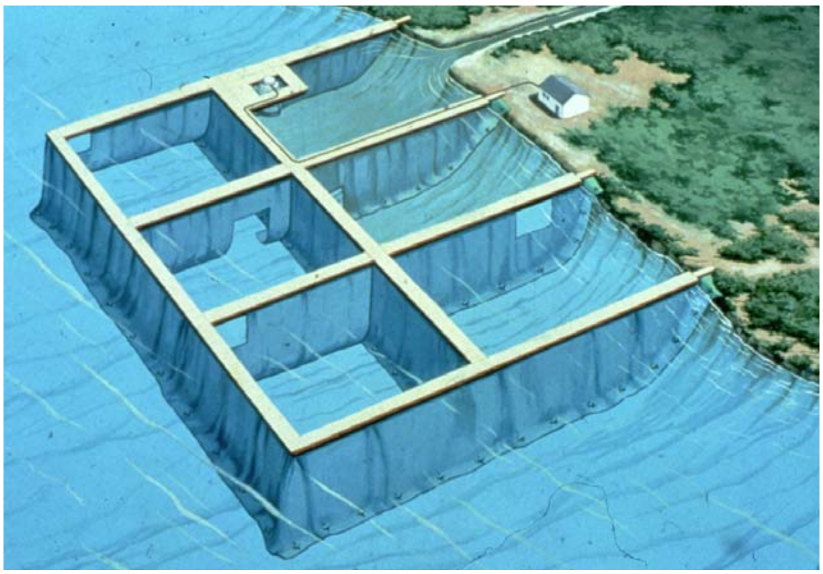


Figure 18. Drawing showing underwater features of an FBM facility (Karl Dunkers, Inc.).

Figure 19. FBM installation located at Lake Trehormingen, Sweden (Karl Dunkers, Inc.).

The annual average removals of phosphorus from stormwater and lake water by the ferric chloride precipitation and clarification treatment system were 66 percent, while the annual average total lake phosphorus concentration reductions averaged about 36 percent. Excess flows are temporarily stored before treatment. Stormwater is pumped to the treatment facility during rains, with excess flows stored inside in-lake flow balancing tanks. The treatment system consists of a chemical treatment system designed for the removal of phosphorus and uses ferric chloride precipitation and crossflow lamella clarifiers. The stormwater is pumped from the flow balancing storage tanks to the treatment facility. Lake water is also pumped to the treatment facility during dry periods, after any excess stormwater is treated.

The specific question to be addressed by this research was whether controlling phosphorus in stormwater discharges to a lake would result in improved lake water quality. Secondly, this evaluation was made to determine if the treatment system was designed and operated satisfactorily. The problem formulation employed for this project was a long-term trend analysis. Up to 30 years of data were available for some water quality parameters, including about 10 years of observations before the treatment system was implemented. Data was available for two sampling locations in the lake, plus at the stormwater discharge location. In addition, mass balance data was available for the treatment operation.

Monitored water quality in Lake Rönningesjön, near Stockholm Sweden, was evaluated to determine the changes in transparency and nutrient concentrations associated with retro-fitted stormwater controls. Statistical trend analyses were used to evaluate these changes. Several publications have excellent descriptions of statistical trend analyses for water quality data. In addition to containing detailed descriptions and examples of experimental design methods to determine required sampling effort, Gilbert (1987) devotes a large portion of his book to detecting trends in water quality data and includes the code for a comprehensive computer program for trend analysis.

Qualitative watershed and lake characterization

Lake Rönningesjön is located in Täby, Sweden, near Stockholm. Figure 20 shows the lake location, the watershed, and the surrounding urban areas. The watershed area is 650 ha, including Lake Rönningesjön itself (about 60 ha), and the urban area that has its stormwater drainage bypassing the lake (about 175 ha). The effective total drainage area (including the lake surface) is therefore about 475 ha. Table 5 summarizes the land use of the lake watershed area. About one-half of the drainage area (including the lake itself) is treated by the treatment and storage operation.

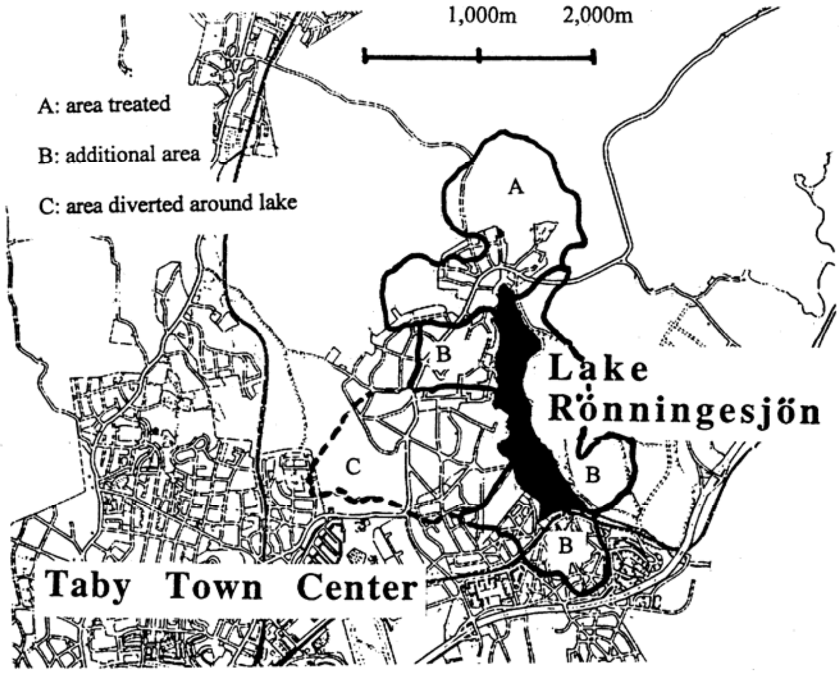


Figure 20. Lake Rönningesjön watershed in Taby, Sweden.

Table 5. Lake Rönningesjön Watershed Characteristics

	Area Treated	Additional Area	Total Area
urban	50 ha	100 ha	150 ha (32%)
forest	75 ha	80 ha	155 ha (32%)
agriculture	65 ha	45 ha	110 ha (23%)
lake surface	60 ha	0 ha	60 ha (13%)
total drainage	250 ha	225 ha	475 ha (100%)

The lake volume is about 2,000,000 m^3 and has an annual outflow of about 950,000 m^3. The estimated mean lake resident time is therefore slightly more than two years. The average lake depth is 3.3 m. It is estimated that the rain falling directly on the lake surface itself contributes about one-half of the total lake outflow.

The treatment process consists of an in-lake flow balancing storage tank system (the Flow Balancing Method, or FBM) to contain excess stormwater flows which are pumped to a treatment facility during dry weather. The treatment facility uses ferric chloride and polymer precipitation and crossflow lamella clarifiers. Figure 21 shows the cross-section of the FBM in the lake. It is make of plastic curtains forming the cell walls, supported by floating pontoons and anchored to the lake bottom with weights.

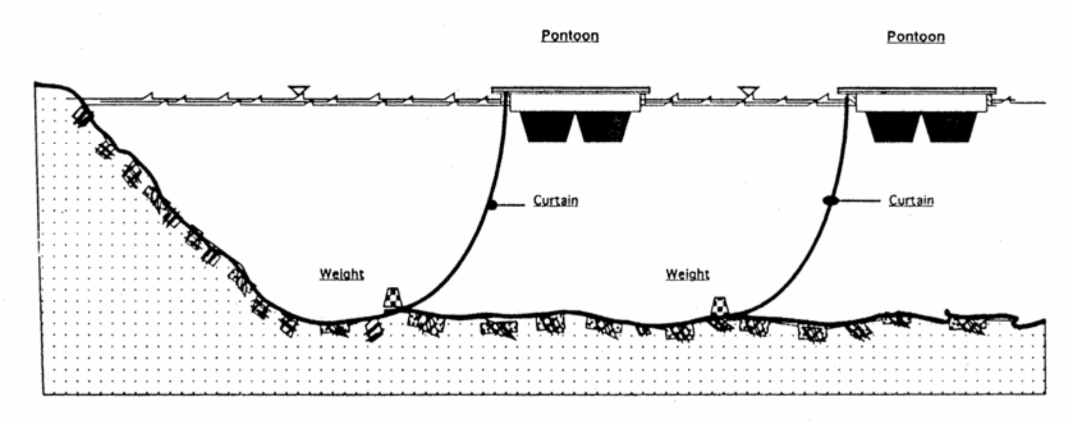


Figure 21. Cross-section of FBM in-lake tanks.

Figure 22 shows that the FBM provides storage of contaminated water by displacing clean lake water that enters the storage facility during dry weather as the FBM water is pumped to the treatment system. All stormwater enters the FBM directly (into cell A). The pump continuously pumps water from cell A to the chemical treatment area. If the stormwater enters cell A faster than the pump can remove it, the stormwater flows through curtain openings (as a slug flow) into cells B, C, D, and finally E, displacing lake water (hence the term flow balancing). As the pump continues to operate, stormwater is drawn back into cell A and then to the treatment facility. The FBM is designed to capture the entire runoff volume of most storms. The Lake Rönningesjön treatment system is designed to treat water at a higher rate than normal to enable lake water to be pumped through the treatment system after all the runoff is treated.

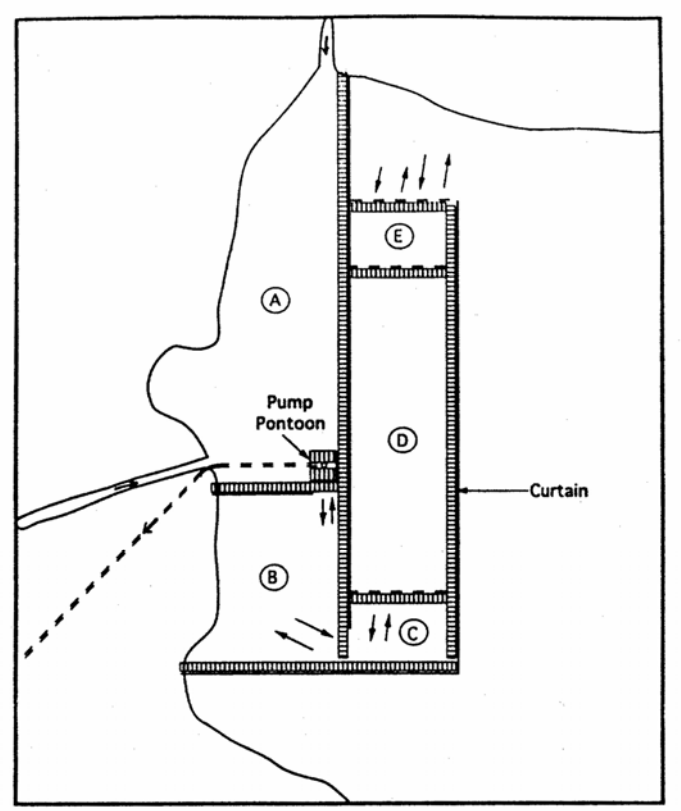


Figure 22. Flow pattern in FBM.

The FBM is mainly intended to be a storage device, but it also operates as a wet detention pond, resulting in sedimentation of particulate pollutants within the storage device. The first two cells of the FBM facility at Lake Rönningesjön were dredged in 1991, after 10 years of operation, to remove about one meter of polluted sediment.

The treatment flow rate is 60 m^3/hr (about 0.4 MGD). The ferric chloride feed rate is about 20 to 35 grams per cubic meter of water. About 30 m^3 of thickened sludge is produced per day for co-disposal with sludge produced at the regional sanitary wastewater treatment facility. The annual operating costs are about $28,000 per year (or about $0.03 per 100 gallons of water treated), divided as shown in
Table 6.

Table 6. Stormwater Treatment System Operating Cost Breakdown

chemicals	26%
electricity	8
sludge transport	3
labor	41
sampling and analyses	22

From 1981 through 1987, the FBM operated an average of about 5500 hours per year (about 7.6 months per year), treating an average of about 0.33 million m^3 per year. The treatment period ranged from 28 to 36 weeks (generally from April through November). The FBM treatment system treated stormwater about 40% of its operating time and lake water about 60% of its operating time. The FBM treatment system directly treated about one-half of the in-flowing waters to the lake (at a level of about 70% phosphorus removal).

Lake Rönningesjön and Treatment System Phosphorus Budgets

Two tributaries flow directly to the treatment facility. Excess flows (exceeding the treatment plant flow capacity) are directed to the FBM in the lake. As the flows in the tributaries fall below the treatment plant capacity, pumps in the FBM deliver stored stormwater runoff for treatment. When all of the stormwater is pumped from the FBM, the pumps deliver lake water for treatment. Tables 7 and 8 summarize the runoff and lake volumes treated and phosphorus removals during the period of treatment.

Table 7. Water Balance for Treatment System (m^3)

	From Trib. A	From Trib. B	Total Stormwater	From Lake	Total treated and discharged	Stormwater, % of total treated
1981	185,100	101,100	286,200	121,600	407,700	70
1982	112,700	41,000	153,700	238,700	391,900	39
1983	14,400	6,400	20,800	250,000	271,000	8
1984	122,000	53,000	175,000	95,000	270,000	65
1985	96,600	46,500	143,100	149,000	292,400	49
1986	216,000	86,000	302,000	48,000	350,000	86
1987	243,000	97,000	340,000	13,000	353,000	96
1988	26,200	19,300	45,500	186,300	231,800	20
1989	24,900	19,900	44,800	267,700	312,500	14
1990	12,160	8,330	20,490	201,270	221,760	9
1991	11,610	7,780	19,390	121,730	141,120	14

Table 8. Phosphorus Treatment Mass Balance (kg)

	From Trib. A	From Trib. B	From Lake	Total to treatment	P discharged to Lake	P removal	% removal
1981	20.3	16.8	10.2	47.3	13.6	33.7	71.2
1982	8.0	8.0	18.0	34.0	12.8	21.2	62.4
1983	1.5	2.5	20.0	24.0	11.0	13.0	54.2
1984	10.0	9.5	3.0	22.5	10.0	12.5	55.6
1985	7.1	5.9	2.1	15.1	4.3	10.8	71.5
1986	15.2	21.4	3.7	40.3	5.1	35.2	87.3
1987	18.6	7.5	1.7	27.8	4.3	23.5	84.5
1988	1.7	2.3	9.2	13.2	6.1	7.1	53.8
1989	1.7	1.4	14.1	17.2	7.6	9.6	55.8
1990	1.3	0.3	10.5	12.1	3.7	8.4	69.4
1991	7.7	9.8	5.6	23.1	8.9	14.2	61.5

There have been highly variable levels of phosphorus treatment from stormwater during the period of operation. The years from 1988 through 1990 had low phosphorus removals. These years had relatively mild winters with substantial stormwater runoff occurring during the winter months when the treatment system was not operating. Normally, substantial phosphorus removal occurred with spring snowmelt during the early weeks of the treatment plant operation each year. The greatest phosphorus improvements in the lake occurred during the years when the largest amounts of stormwater were treated.

The overall phosphorus removal rate for the 11 years from 1981 through 1991 was about 17 kg/year. About 40% of the phosphorus removal occurred in the FBM from sedimentation processes, while the remaining occurred in the chemical treatment facility. This phosphorus removal would theoretically cause a reduction in phosphorus concentrations of about 10 µg/L per year in the lake, or a total phosphorus reduction of about 100 µg/L during the data period since the treatment system began operation. About 70% of this phosphorus removal was associated with the treatment of stormwater, while about 30% was associated with the treatment of lake water.

Select Monitoring Parameters

Lake Rönningesjön water quality has been monitored since 1967 by the Institute for Water and Air Pollution Research (IVL); the University of Technology, Stockholm; the Limnological Institute at the University of Uppsala; and by Hydroconsult Corp. Surface and subsurface samples were obtained at one or two lake locations about five times per year. In addition, the tributaries being treated, incoming lake water, and discharged water, were all monitored on all weekdays of treatment plant operation. The creek tributary flow rates were also monitored using overflow weirs. Phosphorus, nitrogen, chlorophyll *a*, and Secchi disk transparency were all monitored at the lake stations.

Observed Long-Term Lake Rönningesjön Water Quality Trends

The FBM started operation in 1981. Based on the hydraulic detention time of the lake, several years would be required before a new water quality equilibrium condition would be established. A new water quality equilibrium will eventually be reached after existing pollutants are reduced from the lake water and sediments. The new water quality conditions would be dependent on the lake flushing rate (or detention time, estimated to be about 2.1 years), and the new (reduced) pollutant discharge levels to the lake. Without lake water treatment, the equilibrium water quality would be worse and would take longer to obtain.

Figure 23 is a plot of all chlorophyll *a* data collected at both the south and north sampling stations. Very little trend is obvious, but the wide swings in chlorophyll *a* values appeared to have been reduced after the start of stormwater treatment. Figure 24 is a three-dimensional plot of smoothed chlorophyll *a* data, indicating significant trends by season. The values started out relatively low each early spring and dramatically increased as the summer progressed. This was expected and was a function of algal growth. Homogeneity, seasonal Kendall and Mann-Kendall statistical tests (Gilbert 1987) were conducted using the chlorophyll *a* data. The homogeneity test was used to determine if any trends found at the north and south sampling stations were different. The probabilities that the trends at these two stations were the same were calculated as follows:

	χ^2	Probability
season	14.19	0.223
station	0.00001	1.000
station-season	0.458	1.000
Trend	21.64	0.000

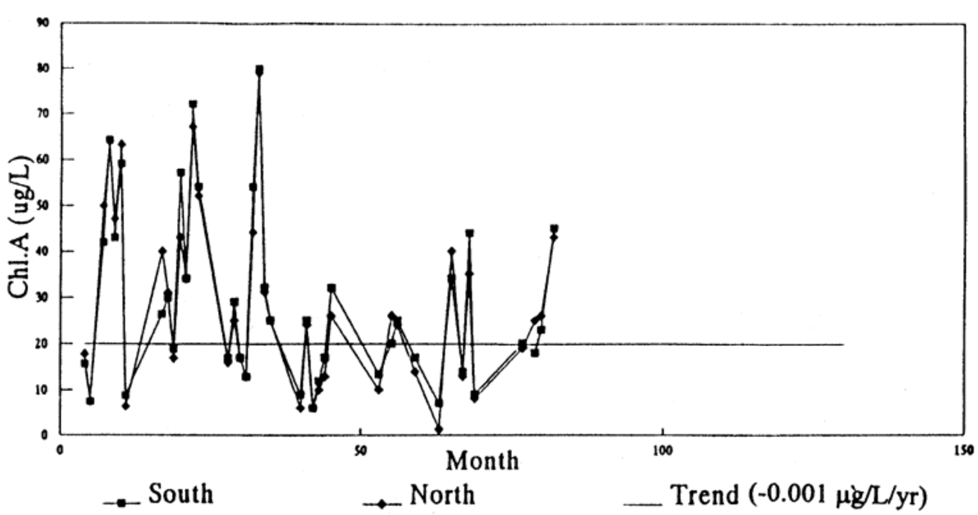


Figure 23. Chlorophyll a observations with time (µg/L).

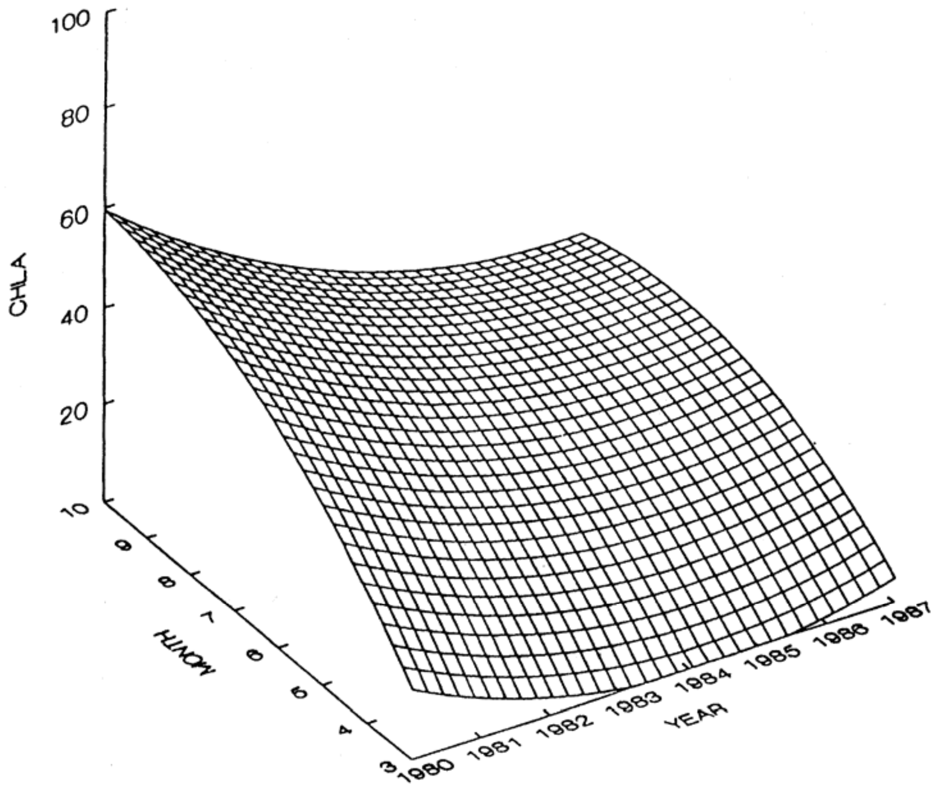


Figure 24. Chlorophyll a trends by season and year (µg/L).

This test shows that the trend was very significant (P<0.001) and was the same at both sampling stations (P=1.000). The seasonal trend tests only compared data obtained for each season, such as comparing trends for June observations alone. The station-season interaction term shows that the chlorophyll *a* concentration trends at the two stations were also very similar for all months (P=1.000). Therefore, the sampling data from both stations were combined for further analyses.

The seasonal Kendall test calculated the chlorophyll *a* concentration trends and determined the probabilities that they were not zero, for all months separately. This test and the Mann-Kendall tests found that both the north and south sampling locations had slight decreasing (but very significant) overall trends in concentrations with increasing years (P≤0.001). However, individual monthly trends were not very significant (P≥0.05). The trends do show an important decrease in the peak concentrations of chlorophyll *a* that occurred during the fall months during the years of the FBM operation. The 1980 peak values were about 60 µg/L, while the 1987 peak values were lower, at about 40 µg/L.

Swedish engineers (Söderlund 1981; and Lundkvist and Söderlund 1988) summarized major changes in the algal species present and in the algal biomass in Lake Rönningesjön, corroborating the chlorophyll *a* and phosphorus limiting nutrient observations. From 1977 through 1983, the lake was dominated by a stable population of thread-shaped blue-green algae species (especially *Oscillatoria sp. and Aphanizomenon flos aquae f. gracile*). Since 1985, the algae population was unstable, with only a small amount of varying blue green (*Gomphosphaeria*), silicon (*Melosira, Asterionella* and *Synedra*) and gold (*Chrysochromulina*) algae species. They also found a substantial decrease in the algal biomass in the lake. From 1978 through 1981, the biomass concentration was commonly greater than 10 mg/L. The observed maximum was about 20 mg/L, with common annual maximums of 15 mg/L in July and August of each year. From 1982 through 1986, the algal biomass was usually less than 10 mg/L. The observed maximum was 14 mg/L and the typical annual maximum was about 6 mg/L each late summer. The lake showed an improvement in its eutrophication level since the start of the stormwater treatment, going from hypotrophic to eutrophic.

Figure 25 is a plot of all Secchi disk transparency data obtained during the project period. A very large improvement in transparency is apparent from this plot, but large variations were observed in most years. A large improvement may have occurred in the first five years of stormwater treatment and then the trend may have decreased. The smoothed plot in Figure 26 shows significant improvement in Secchi disk transparency since 1980. This three-dimensional plot shows that the early years started off with clearer water (as high as 1 m transparency) in the spring and then degraded as the seasons progressed, with transparency levels falling to less than 0.5 m in the fall months. The later years indicated a significant improvement, especially in the later months of the year.

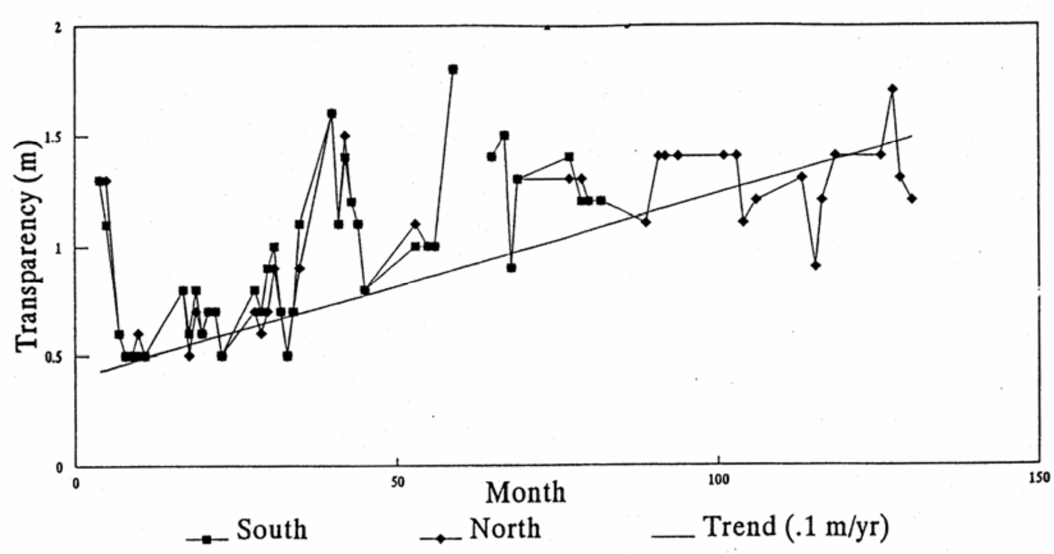


Figure 25. Secchi disk transparency observations with time (m).

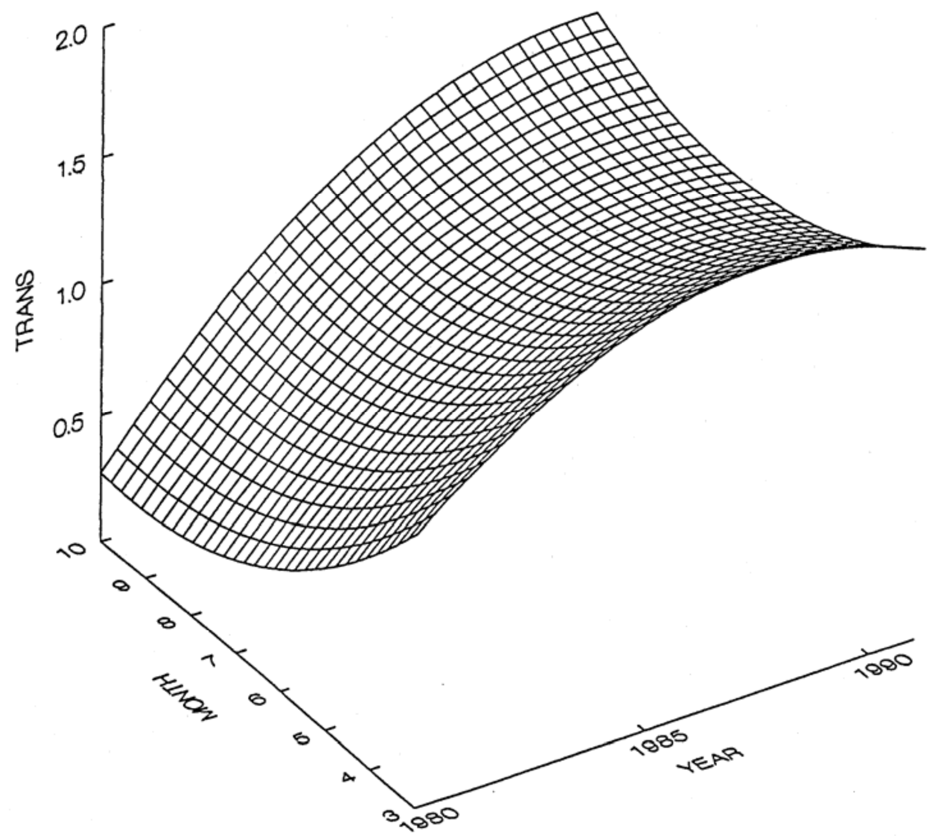


Figure 26. Secchi disk trends by season and year (m).

Homogeneity, seasonal Kendall and Mann-Kendall statistical tests (Gilbert 1987) were conducted using the Secchi disk transparency data. The homogeneity test was used to determine if any trends found at the north and south sampling stations were different. The probabilities that the trends at these two stations were the same were calculated as follows:

	χ^2	Probability
season	17.15	0.103
station	0.012	0.913
station-season	3.03	0.990
Trend	29.44	0.000

These statistics show that the observed trend was very significant ($P<0.001$) and was the same at both stations. The Seasonal Kendall and Mann-Kendall tests found that both the north and south sampling locations had increasing transparency values (the average trend was about 0.11 meter per year) with increasing years ($P<0.001$). The trend in later years was found to be less than in the early years. The transparency has remained relatively stable since about 1987 (ranging from about 1 to 1.5 m), with less seasonal variations.

Figure 27 plots observed phosphorus concentrations with time, while Figure 28 is a smoothed plot showing seasonal and annual variations together. The initial steep phosphorus concentration decreases in the early years of the FBM operation were followed by a sharp increase during later years. The increase was likely associated with the decreased levels of stormwater treatment during the mild winters of 1988 through 1990 when the treatment system was not operating; large amounts of untreated stormwater were discharged into the lake instead of being tied up as snow to be treated in the spring as snowmelt runoff.

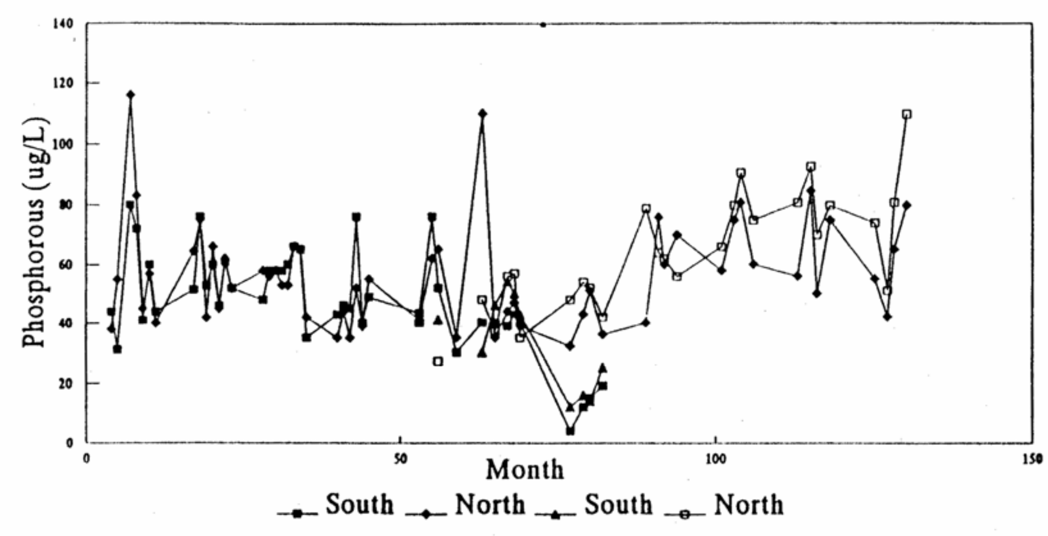


Figure 27. Total phosphorus observations with time (μg/L).

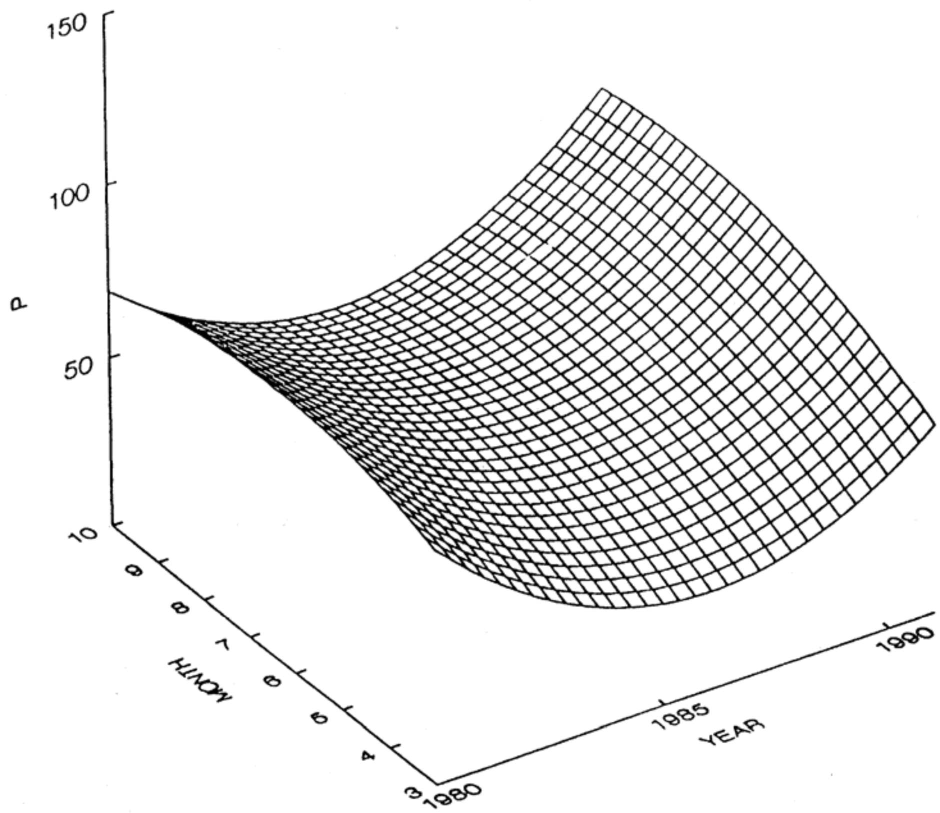


Figure 28. Total phosphorus trends by season and year (μg/L).

Individual year phosphorus concentrations leveled off in the summer (about July). These seasonal phosphorus trends were found to be very significant ($P \leq 0.002$), but were very small, using the seasonal Kendall test (Gilbert 1987). Homogeneity tests found no significant differences between lake sample phosphorus concentrations obtained at the different sampling locations, or depths, irrespective of season:

	χ^2	Probability
season	15.38	0.166
station	0.0033	0.954
station-season	1.64	0.999
Trend	12.43	0.000

The overall lake phosphorus concentrations ranged from about 15 to 130 µg/L, with an average of about 65 µg/L. The monitored stormwater, before treatment, had phosphorus concentrations ranging from 40 to >1,000 µg/L, with an average of about 200 µg/L.

An increase in nitrogen concentrations also occurred from the beginning of each year to the fall months. However, the overall annual trend decreased during the first few years of the FBM operation, but it then subsequently increased. These total nitrogen concentration variations were similar to the total phosphorus concentration variations. However, homogeneity, seasonal Kendall and Mann-Kendall statistical tests (Gilbert 1987) conducted using the nitrogen data found that neither the north or south sampling locations had significant concentration trends with increasing years ($P>0.2$). However, lake Kjeldahl nitrogen concentration reductions were found to occur during years when the FBM system was treating the largest amounts of stormwater.

Lake Water Quality Model

A simple water quality model was used with the Lake Rönningesjön data to determine the total annual net phosphorus discharges into the lake and to estimate the relative magnitude of various in-lake phosphorus controlling processes (associated with algal growth and sediment interactions, for example). These estimated total phosphorus discharges were compared to the phosphorus removed by the treatment system. The benefits of the treatment system on the lake water quality were then estimated by comparing the expected lake phosphorus concentrations as if the treatment system was not operating, to the observed lake phosphorus concentrations.

Thomann and Mueller (1987) presented the following equation to estimate the resulting water pollutant concentrations associated with varying input loadings for a well-mixed lake:

$$S_t = (M/V) \exp(-T/Td) \qquad \text{eq. 1}$$

where S_t = concentration associated with a step input at time t,
M = mass discharge per time-step interval (kg),
V = volume of lake (2,000,000 m^3),
T = time since input (years), and
Td = hydraulic residence time, or lake volume/lake outflow (2.1 years).

This equation was used to calculate the yearly total mass discharges of phosphorus to Lake Rönningesjön, based on observed lake concentrations and lake hydraulic flushing rates. It was assumed that the varying concentrations observed were mostly caused by varying mass discharges and much less by variations in the hydraulic flushing rate. The flushing rate was likely to vary, but by relatively small amounts. The lake volume was quite constant and the outflow rate was expected to vary by less than 20 percent because of the relatively constant rainfall that occurred during the years of observation (average rainfall of about 600 mm, with a coefficient of variation of about 0.15).

The total mass of phosphorus discharged into the lake each year from 1972 to 1991 was calculated using the following equation (an expansion of equation 1), solving for the M_{n-x} terms:

$$S_n = M_n [\exp(-T_n/Td)/V] + M_{n-1} [\exp(-T_{n-1}/Td)/V] + M_{n-2} [\exp(-T_{n-2}/Td)/V] +$$

$$M_{n-3}\,[\exp(-T_{n-3}/Td)/V] + \ldots \qquad \text{eq. 2}$$

where S_n is the annual average phosphorus concentration during the current year, M_n is the net phosphorus mass discharged into the lake during the current year, M_{n-1} is the phosphorus mass discharged during the previous year, M_{n-2} is the phosphorus mass that was discharged two years previous, etc.

The effects of discharges into the lake many years previous to a concentration observation have little effect on that year's observations. Similarly, more recent discharges have greater effects on the lake's concentrations. The magnitude of effect that each year's step discharge has on a more recent concentration observation is dependent on the $\exp(-T_n/Td)$ factors shown in equation 2. A current year's discharge affects that year's concentration observations by about 40 percent of the steady-state theoretical value (M/V), and a discharge from five years previous would only affect the current year's concentration observations by less than ten percent of the theoretical value for Lake Rönningesjön. Similarly, a new steady-state discharge would require about 4 years before 90 percent of its equilibrium concentration would be obtained. It would therefore require several years before the effects of a decrease in pollutant discharges would have a major effect on the lake pollutant concentrations.

The annual control of phosphorus ranged from about 10 to 50 percent, with an average lake-wide level of control of about 36 percent, during the years of treatment plant operation. It is estimated that there would have been about a 1.6 times increase in phosphorus discharges into Lake Rönningesjön if the treatment system was not operating. There was a substantial variation in the year to year phosphorus discharges, but several trends were evident. If there was no treatment, the phosphorus discharges would have increased over the 20 year period from about 50 to 75 kg per year associated with increasing amounts of contaminated stormwater associated with increasing urbanization in the watershed. With treatment, the discharges were held relatively constant at about 50 kg per year (as evidenced by the lack of any observed phosphorus concentration trend in the lake). During 1984 through 1987, the phosphorus discharges were quite low compared to other years, but increased substantially in 1988 and 1989 because of the lack of stormwater treatment during the unusually mild winters.

Figure 29 is a plot of the annual average lake phosphorus concentrations with time. If there had been no treatment, the phosphorus concentrations in the lake would have shown a relatively steady increase from about 50 to about 100 µg/L over the 20 year period. With treatment, the lake phosphorus concentrations were held within a relatively narrower range (from about 50 to 75 µg/L). The lake phosphorus concentration improvements averaged about 50 µg/L over this period of time, compared to an expected theoretical improvement of about 100 µg/L. Therefore, only about one-half of the theoretical improvement occurred, probably because of sediment-water interchange of phosphorus, or other unmeasured phosphorus sources.

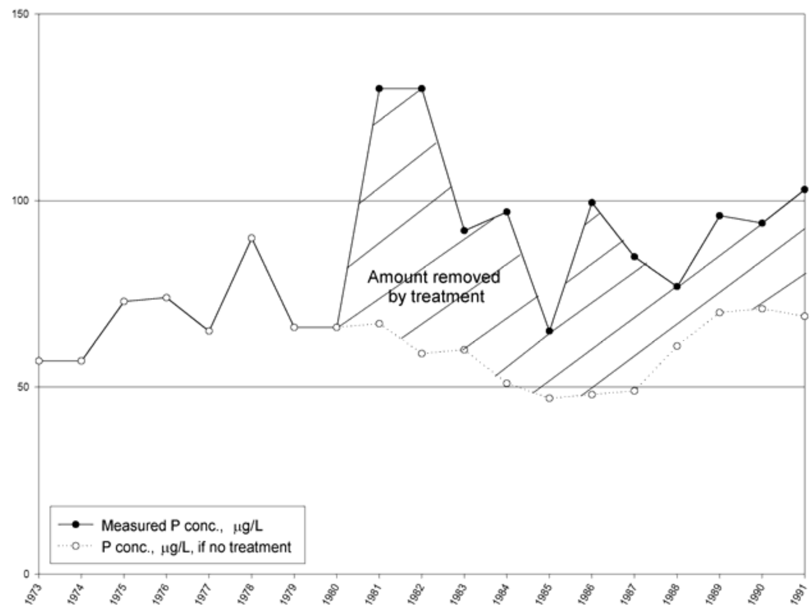


Figure 29. Effects of treatment on Lake Rönningesjön total phosphorus concentrations (μg/L).

Project Conclusions

The in-lake flow balancing method (FBM) for storage of excess stormwater during periods of high flows allowed for lower treatment flow rates, while still enabling a large fraction of the stormwater to be treated for phosphorus removal. The treatment system also enabled lake water to be treated during periods of low (or no) stormwater flow. The treatment of the stormwater before lake discharge accounted for about 70 percent of the total observed phosphorus discharge reductions, while the lake water treatment was responsible for the remaining 30 percent of the discharge reductions. The lake water was treated during 60 percent of the operating time, but resulted in less phosphorus removal, compared to stormwater treatment. The increased efficiency of phosphorus removal from stormwater compared to lake water was likely due to the more abundant particulate forms of phosphorus that were removed in the FBM by sedimentation and by the stormwater's higher dissolved phosphorus concentrations that were more efficiently removed during the chemical treatment process.

Lake transparency improved with treatment. Secchi disk transparencies were about 0.5 m before treatment began and improved to about 1 to 1.5 m after treatment. The total phosphorus concentrations ranged from about 65 to 90 μg/L during periods of low levels of stormwater treatment, to about 40 to 60 μg/L during periods of high levels of stormwater treatment.

The annual average removals of phosphorus by the ferric chloride precipitation and clarification treatment system were 66 percent, with a maximum of 87 percent. The observed phosphorus concentration improvements in the lake were strongly dependent on the fraction of the annual stormwater flow that was treated. The annual average total lake phosphorus discharge and concentration reductions averaged about 36 percent, or about one half of the maximum expected benefit.

The water sampling for this project was irregular. Only a relatively few samples were obtained in any one year, but up to 30 years of data were obtained. In addition, no winter data was available due to icing of the lake. In general, statistically-based trend analyses are more powerful with evenly spaced data over the entire period of time. However, this is typically unrealistic in environmental investigations because of an inability to control other important factors. If all samples were taken on the 15th of each month, for example, the samples would be taken under highly variable weather conditions. Weather is a significant factor in urban runoff studies, obviously, and this statistical methodology requirement would have severely confounded the results. The trend analyses presented by Gilbert (1987) enable a more reasonable sample collection effort, with some missing data. However, the procedure does require relatively complete data collected over an extended period of time. It would have been very difficult to conduct this analysis with only a few years of the data, for example. The seasonal patterns were very obvious when multiple years of before and after treatment were monitored. In addition, the many years of data enabled unusual weather conditions (such as the years with unusually mild winters) to stand out from the more typical weather conditions.

The analytical effort only focused on a few parameters. This is acceptable for a well designed and executed project, but prohibits further insights that a more expansive effort may obtain. Since this project was specifically investigating transparency associated eutrophication, the parameters evaluated enabled the basic project objectives to be effectively evaluated. However, the cost of labor for the sampling effort is a major component of an investigation like this one, and some additional supportive analyses may not have added much to the overall project cost while adding potentially valuable additional information.

In general, trend analyses require a large amount of data, typically obtained over a long period of time. These requirements cause potential problems. Experimental designs for a several year (or several decade) monitoring effort are difficult to carry out. Many uncontrolled changes may occur during a long period, such as changes in laboratory analyses methods. Laboratory method changes can affect the specific chemical species being measured, or at least have differing detection limit capabilities. This study examined basic measurements that have not undergone major historical changes, and very few "non-detectable" values were reported. In contrast, examining historical heavy metal data is very difficult because of changes in instrumentation and associated detection limits. The need for a typically long duration study also requires a long period before statistically relevant conclusions can be obtained. Budget reductions in the future always threaten long-term efforts. In addition, personnel changes lead to inconsistent sampling and may also possibly lead to other errors. Basically, adequate trend analyses require a large amount of resources (including time) to be successful. The use of historical data not collected for a specific trend analysis objective is obvious and should be investigated to supplement an anticipated project. However, great care must be expended to ensure the quality of the data. In most cases, incorrect sampling locations and dates, let alone obvious errors in reported concentrations, will be found in historical data files. These problems, in conjunction with problems associated with changing laboratory methods during the monitoring period, require special effort.

Example Stormwater Data Analysis

Sampling Effort and Basic Data Presentations

The following is an example of a large-scale stormwater data analysis effort recently conducted for the telecommunication industry. Table 9 lists the numbers of samples that were sent to our lab for analyses from the nine participating companies, by season.

Based on prior determinations. each strata needs about 10 separate samples in order to estimate the quality characteristics with an error level of about 25 percent. The goal of each participant is to obtain samples from four groups of locations (having 10 each) for each season:

1) old industrial/commercial (or central city) area
2) new industrial/commercial (or central city) area
3) old residential (or suburban) area
4) new residential (or suburban) area

The same areas were sampled during each season to minimize additional variation. The main seasons for sampling were winter and summer. Therefore, each participant was to collect a total of 40 samples per season, for at least these two seasons. The collection of additional samples for other seasons or land uses enabled further comparisons to be made.

Table 10 is an example partial listing of the cities sampled during this program, while Figure 30 shows their geographical distribution and associated EPA rainfall region. Thirty-two states, plus the District of Columbia were represented in this sampling effort. All EPA Rain Regions were also represented, although Regions 5, 8, and 9 had fewer samples. The sampled cities represent annual rainfalls ranging from about 7 inches (Phoenix) to about 65 inches (Pensacola).

Table 9. Samples Analyzed from Various Telecommunication Companies and Seasons

	Winter 1995/96	Spring 1996	Summer 1996	Fall 1996	Winter 1996/97	Spring/ Summer 1997	Fall/ Winter 1997	Winter 1997/98	Summer 1998	Winter 1998/99	Total
NYNEX	15	15	20	20	0	6	0	0	0	0	76
Bell Atlantic	14	0	26	22	24	20	0	0	0	0	106
BellSouth	3	36	36	32	31	0	0	0	0	0	138
SNET	0	0	0	20	20	0	0	0	0	0	40
Pacific Bell	0	0	0	0	0	21	23	0	0	0	44
GTE	0	0	0	0	0	24	23	16	0	0	63
U.S. West	0	0	0	0	0	0	0	35	35	0	70
Ameritech	0	0	0	0	0	0	0	40	40	0	80
AT&T	0	0	0	0	0	0	0	0	40	40	80
TOTAL	32	51	82	94	75	71	46	91	115	40	697

Table 10. Sampling cities and their locations.

SAMPLING LOCATIONS BY COMPANY AND STATE Note: (EPA Rainfall Zone Number, Average Annual Precipitation [in.]) after city name						
Ameritech	**Illinois** Frankfort (4, 35) Joliet (1, 33) Lemont (1, 33) New Lenox (1, 33)	**Indiana** Gary (1, 33) Indianapolis (2, 39) Merrillville (1, 33) St. John (1, 33)	**Michigan** Ann Arbor (1, 31) Dearborn (1, 31) Detroit (1, 31) Rockwood (1, 31) Southgate (1, 31)	**Ohio** Columbus (1/2, 37) Dublin (2, 37) Hilliard (2, 37) Graves City (2, 37)	**Wisconsin** Madison (1, 37)	
AT & T	**Missouri** St. Louis (4, 34)	**Montana** Billings (8, 15)	**Nebraska** Omaha/Lincoln (9, 30)	**Idaho** Boise (8,11)		
Bell Atlantic	**District of Columbia** (2, 39)	**Maryland** Baltimore (2, 42) Edgemere (2, 42) Prince Georges Cty. (2, 42)	**New Jersey** Fairfield (1, 42) Garfield (1, 42) Lincoln Park (1, 42)	**Pennsylvania** Monroeville (1, 41) Moon Township (1, 36) New Kensington (1, 36) Philadelphia (1, 41) Pittsburgh (1, 36) Trooper (1, 41) West Mifflin (1, 36)	**Virginia** Arlington (2, 44) Richmond (2, 44)	**West Virginia** Beckley (2, 40) Charleston (2, 42)
BellSouth	**Alabama** Birmingham (3, 55) Center Point (3, 55) Homewood (3, 55) Hoover (3, 55)	**Florida** Ft. Lauderdale (3, 58) Jacksonville (3, 53) Miami (3, 58) Orlando (3, 53) Pensacola (3, 65)	**Georgia** Conyers (3, 49) Decatur (3, 49) Fairington (3, 49)	**Louisiana** Baton Rouge (4, 60) Lafayette (4, 60) Lake Arthur (4, 60) Oakdale (4, 60)	**North Carolina** Asheville (2, 48)	
GTE	**Illinois** Bloomington (1, 35) Rantoul (1, 35) Dekalb (1, 33)	**Indiana** Lafayette (2, 39) Terre Haute (2, 39) Valparaiso (1, 33)	**Oregon** Beaverton (7, 37) Coos Bay (7, 46) Hillsboro (7, 37) Reedsport (7, 46) Tigard (7, 37)	**Washington** Anacortes (7, 39) Bothell (7, 39) Burlington (7, 39) Camas (7, 37) Everett (7, 39) Marysville (7, 39) Monroe (7, 39) Mount Vernon (7, 39) Mukilteo (7, 39)		

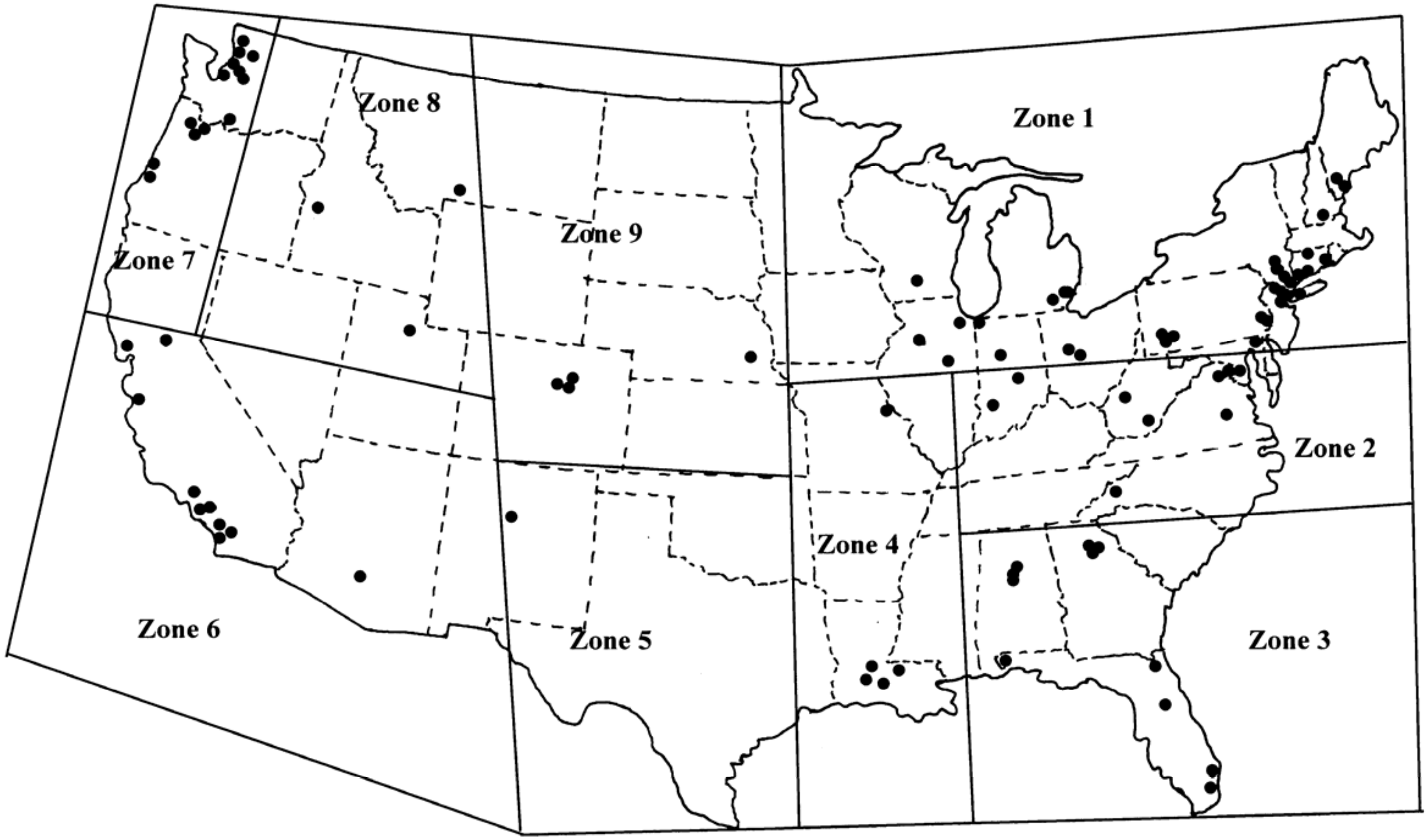


Figure 30. Map of US with sampling cities and EPA Rain Zones

BellSouth, U.S. West, Ameritech, and AT&T were close to having collected 40 samples for each of the two main seasons. BellSouth, NYNEX and Bell Atlantic also collected samples from all four seasons. SNET and Pacific Bell collected somewhat fewer samples. The total number of samples collected was close to the number as originally planned (at 80 per participant), but with half the number of locations sampled per some seasons, but twice as many seasons were represented for other areas. Very close to the total number of samples identified as our overall goal (720) was collected (697). About 390 sediment samples were also collected for concurrent analysis.

Summary of Data

Data Summaries

Most of the constituents have several hundred to almost 700 analyses available. Table 11 summarizes some of these data.

Table 11. Statistical Summary of Data

	Temp at Sampling (^oF)	Water depth (ft)	Sediment depth (ft)	Total Solids (mg/L)	Dissolved Solids (mg/L)	Suspended Solids (mg/L)	Volatile Total Solids (mg/L)
Number of Analyses	614	648	559	685	683	683	685
Number of Detectable	614	614	380	685	683	549	685
Percent Detectable	100	94.8	68	100	100	80.4	100
COV	3.24	1.31	0.68	0.58	0.54	0.32	0.57
1st Percentile	25.0	0	0	89	69	nd	14
5th Percentile	32.0	0	0	146	118	nd	22
10th Percentile	35.0	0.3	0	199	160	nd	31
15th Percentile	38.0	0.7	0	239	208	nd	36
20th Percentile	40.0	1.0	0	280	238	1	42
25th Percentile	45.0	1.3	0	314	278	4	48
30th Percentile	45.0	1.7	0	351	308	6	53
35th Percentile	50.0	2.0	0	385	342	9	59
40th Percentile	50.2	2.4	0	437	379	12	64
45th Percentile	55.0	2.8	0	487	424	15	71
50th Percentile	60.0	3.0	0.1	546	471	19	80
55th Percentile	64.2	3.5	0.1	625	544	25	92
60th Percentile	66.0	4.0	0.1	694	612	33	104
65th Percentile	70.0	4.0	0.2	770	712	41	119
70th Percentile	70.0	4.5	0.2	894	824	56	139
75th Percentile	75.0	5.0	0.3	1062	968	73	160
80th Percentile	76.0	5.6	0.3	1306	1144	98	188
85th Percentile	80.0	6.0	0.3	1665	1541	144	246
90th Percentile	85.0	7.0	0.3	2228	1935	208	337
95th Percentile	88.0	8.0	0.5	3597	3435	414	484
100th Percentile	105.0	13.0	1.5	32950	32756	3505	3264

Table 11. Statistical Summary of Data (continued)

	Volatile Dissolved Solids (mg/L)	Volatile Suspended Solids (mg/L)	Suspended Solids (mg/L)	% of Volatile Solids - Sediment	Turbidity Unfiltered (NTU)	Turbidity Filtered (NTU)	pH	Toxicity Unfiltered (I25% reduction)	Toxicity Filtered (I25% reduction)	COD Unfiltered (mg/L)
Number of Analyses	683	683	593	387	685	685	684	681	682	681
Number of Detectable	682	463	580	384	685	684	684	450	441	594
Percent Detectable	99.9	67.8	97.8	99.2	100	99.9	100	66.1	64.7	87.2
COV	0.66	0.23	0.35	1.16	0.25	0.53	13.32	0.84	0.82	0.69
1st Percentile	6	nd	nd	0.5	0.3	0.1	6.2	nd	nd	nd
5th Percentile	16	nd	2	1.1	0.6	0.2	6.6	nd	nd	nd
10th Percentile	24	nd	3	1.7	1.0	0.2	6.8	nd	nd	nd
15th Percentile	29	nd	4	2.3	1.4	0.3	7.0	nd	nd	1
20th Percentile	33	nd	5	3.0	1.8	0.4	7.1	nd	nd	4
25th Percentile	37	nd	6	3.6	2.2	0.4	7.2	nd	nd	6
30th Percentile	43	nd	7	4.0	2.7	0.5	7.3	nd	nd	8
35th Percentile	49	2	8	4.6	3.3	0.5	7.4	1	nd	10
40th Percentile	53	5	10	5.4	4.3	0.6	7.5	5	4	12
45th Percentile	60	7	11	5.8	5.3	0.7	7.5	9	8	13
50th Percentile	67	9	13	6.1	6.5	0.8	7.6	12	12	16
55th Percentile	75	12	14	6.7	8.2	0.9	7.7	17	16	19
60th Percentile	85	17	17	7.3	9.6	1.0	7.7	22	21	21
65th Percentile	99	20	21	8.7	11.9	1.1	7.8	29	27	24
70th Percentile	113	25	25	9.7	15.1	1.2	7.9	39	34	26
75th Percentile	132	32	34	11.0	17.8	1.5	7.9	51	45	32
80th Percentile	158	42	46	12.4	22.8	1.8	8.0	63	57	38
85th Percentile	192	51	66	14.3	32.8	2.1	8.1	76	74	47
90th Percentile	256	74	119	17.3	68.8	2.7	8.3	86	85	62
95th Percentile	371	126	308	21.4	127.4	3.9	8.4	97	95	118
100th Percentile	2093	3025	2123	67.8	2097.0	42.0	9.4	100	100	372

Exploratory Data Analysis of Rainfall and Runoff Characteristics for Urban Areas

Actual stormwater characteristics from the EPA's Nationwide Urban Runoff Program (EPA 1983), the EPA's Urban- Rainfall-Runoff-Quality Data Base (Heaney, *et al*. 1982), and from the Humber River portion of the Toronto Area Watershed Management Study (Pitt and McLean 1986) were examined by Pitt, *et al*. (2001). The Toronto area data were from two extensively monitored watersheds, a residential/commercial area and an industrial area. Most of the EPA's "Data Base" information used was from 2 locations in Broward County, FL; 1 site in Dade County, FL; 2 sites in Salt Lake City, UT; and 2 sites in Seattle, WA. Most of the data were obtained during the 1970s. These sites had the best representation of data of interest for these analyses and the sites were well described. Parameters examined included simultaneous rainfall and runoff depths, plus peak rain and flow rates. The following plots were prepared using this data:

- runoff depth versus rainfall,
- volumetric runoff coefficient (Rv) versus rainfall,
- NRCS curve number (CN) versus rainfall, and
- ratio of reported peak flow/peak rainfall versus rainfall.

In a similar manner, information from the EPA's NURP program (EPA 1983) was also investigated. A wider variety of information was collected during NURP, enabling additional relationships examining stormwater quality. Most of the data used here are from 5 sites in Champaign, IL; 2 sites in Austin, TX; 5 sites in Irondequoit Bay, NY; 1 site in Rapid City, SD; plus additional observations from Tampa, FL, Winston Salem, NC, and Eugene and Springfield, OR. Most of this data were obtained during the early 1980s and was subjected to rigorous quality control. Besides the four plots listed above, the following plots were also constructed examining potential water quality concentration relationships:

- total suspended solids concentration versus rainfall depth,
- COD concentration versus rainfall depth,
- phosphorous concentration versus rainfall depth,
- lead concentration versus rainfall depth,
- peak flow/peak rain versus rainfall depth, and
- peak flow rate versus peak rain intensity.

These plots were constructed to examine stormwater design methods using actual monitored data. These data can be used to examine many typical assumptions concerning stormwater drainage design and stormwater quality. Figures 31 through 39 show example plots for the John South Basin, a single family residential area, monitored during the EPA's NURP project in Champaign-Urbana, IL. The basic rainfall versus runoff plots (Figure 31) were made to indicate the smoothness of this basic relationship. A large scatter instead of a smooth curve may indicate measurement errors or uneven rainfalls over the catchment, or highly variable infiltration characteristics (due to changing soil moisture before the different rains). As shown on these plots, the runoff depth increases with increasing rain. However, several plots do show substantial scatter, mostly for sites having relatively small runoff yields. In addition, in some cases, more runoff was observed than could be accounted for by the rain. Errors in these measurements may be significant and would vary for the different sites. The following list shows possible measurement errors that may have affected this data:

- variable rainfall over a large test catchment that was not well represented by enough rain gages (Although several of the test catchments had multiple rain gages, most did not, and few were probably frequently re-calibrated in the field.),
- poorly calibrated monitoring equipment (Many flow monitoring equipment relied on using the Manning's equation in pipes, with assumed roughness coefficients, without independent calibration, while other monitoring locations used calibrated insert weirs.)
- transcription errors (Many of these older monitoring activities required manual transfer from field

equipment recorders to computers for analysis. In many cases, obvious “factor of ten” errors were made, for example.),
- newly developed equipment that has not been adequately tested, and
- difficult locations in the sewerage or streams that were monitored.

It is expected that the measurement errors were probably no less than about 25% during these monitoring activities. The effects of actual influencing factors can only be determined after the effects of these errors are considered.

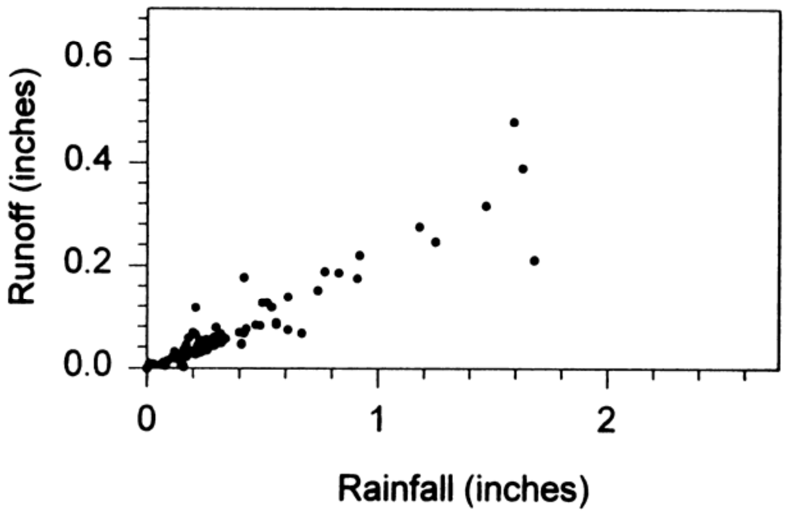


Figure 31. Runoff vs. rainfall.

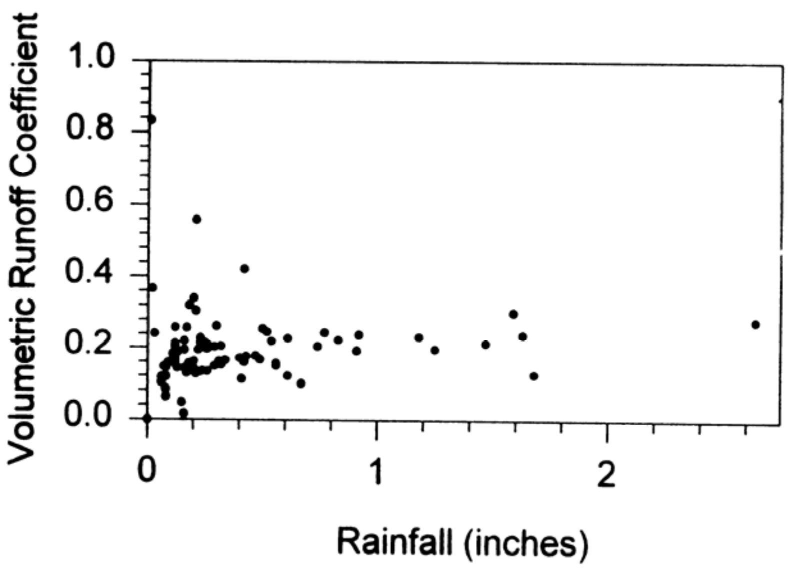


Figure 32. Rv vs. rainfall.

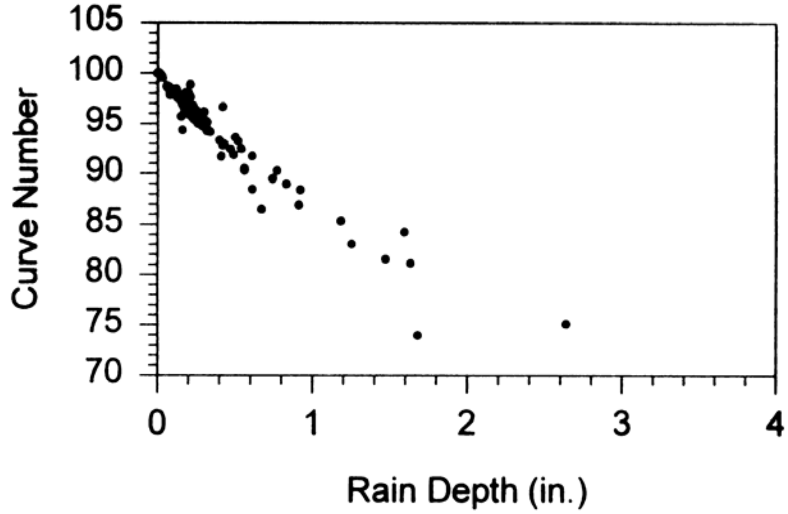


Figure 33. Curve number vs. rain depth.

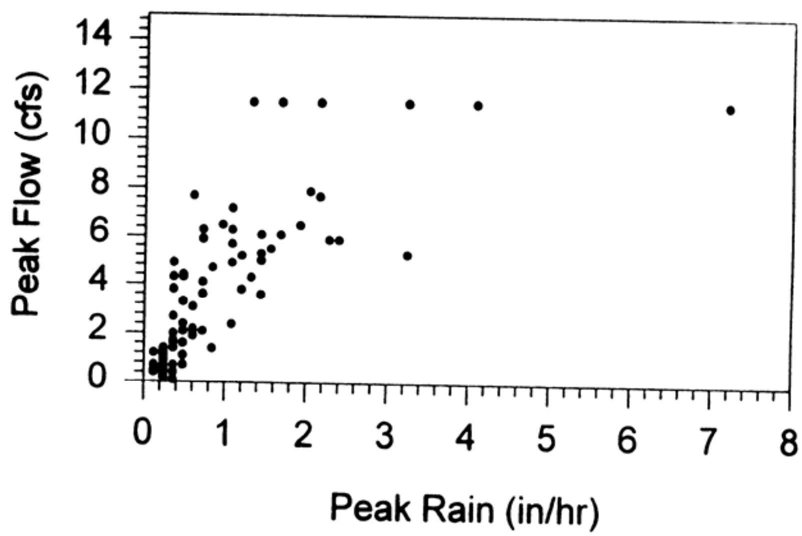


Figure 34. Peak flow vs. peak rain.

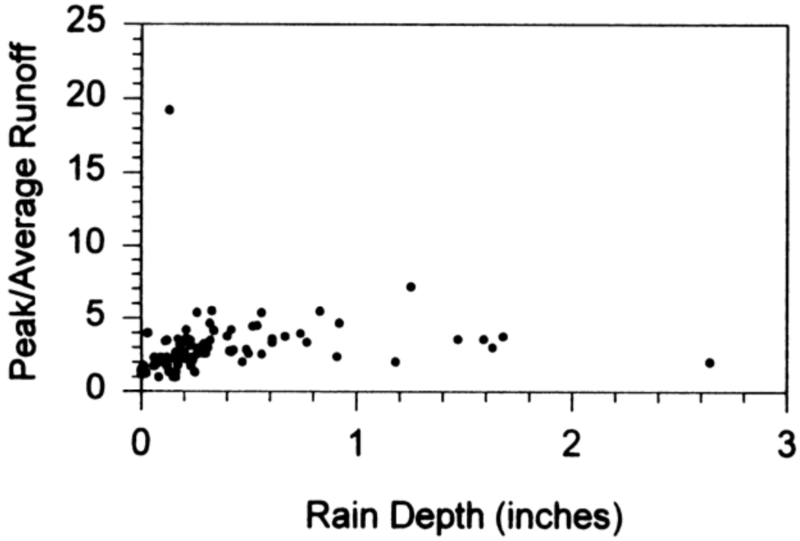


Figure 35. Peak/avg. runoff vs. rain depth.

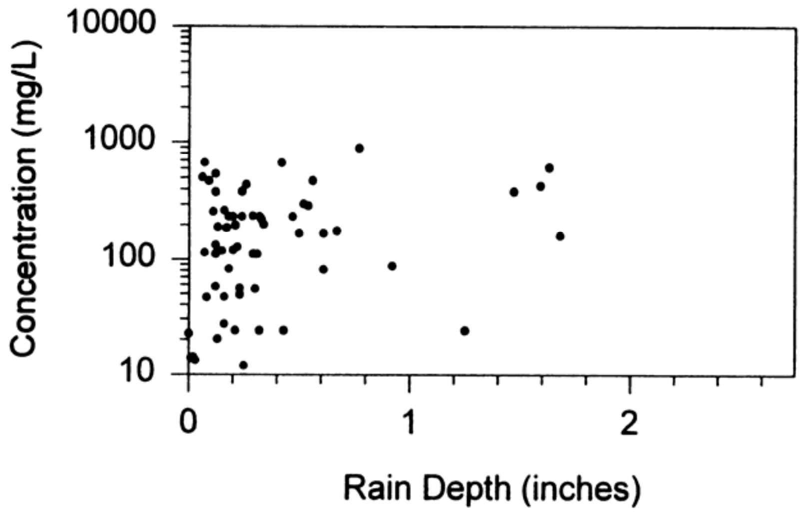


Figure 36. SS vs. rain depth.

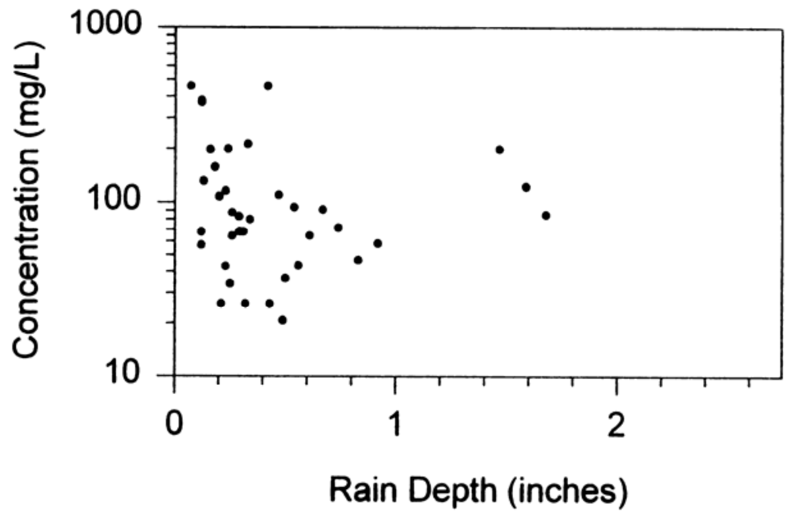


Figure 37. COD vs. rain depth.

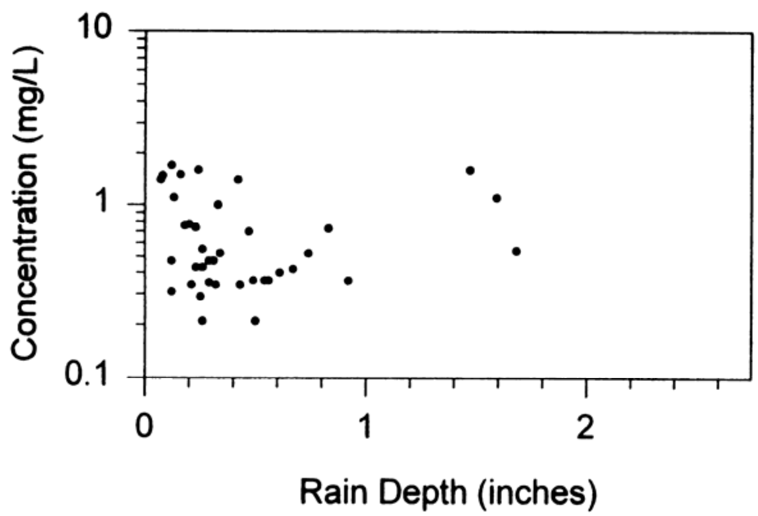


Figure 38. Phosphorus vs. rain depth.

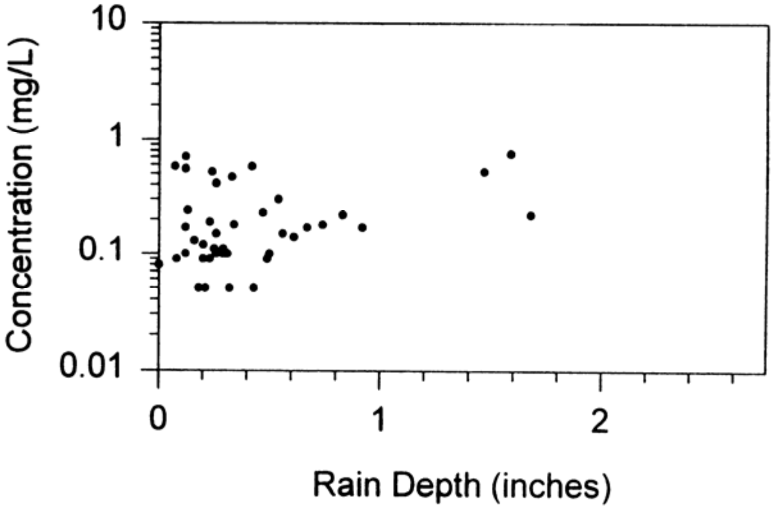


Figure 39. Lead vs. rain depth.

The plots of rainfall versus the volumetric runoff coefficient plot (Figure 32) shows the ratio of the runoff volume, expressed as depth for the watershed, to rain depth, or the Rv, for different rain depths. This is a related plot to the one described above. If the Rv ratio was constant for all events, the rainfall versus runoff depth plot described above, would indicate a straight diagonal line, with no scatter. It is typically assumed that the above described relationship would indicate increasing Rv values as the rain depth increased. Figure 31 shows a slight upwards curve with increasing rain depths. This is due to the rainfall losses making up smaller and smaller portions of the total rainfall as the rainfall increases, with a larger fraction of the rainfall occurring as runoff. The plot of Rv versus rainfall (Figure 32) would therefore show an increasing trend with increasing rain depth. In most cases, the plots of actual data indicate a large (random?) scatter, making the identification of a trend problematic. The use of a constant Rv for all rains may also be a problem because of the large scatter. In many cases, the long-term average Rv for a residential area may be close to the typically used value. In Figure 32, the values appear to center about 0.2 (somewhat smaller than the typically used value of about 0.3 for medium density residential areas), but the observed Rv values may range from lows of less than 0.04 to highs of greater than 0.5, especially for the smallest rains. The small rains probably have the greatest measurement errors, as the rainfall is much more variable for small rains than for larger rains, plus very low flows are difficult to accurately measure. Obviously, understanding what may be causing this scatter is of great interest, but is difficult because of measurement errors masking trends that may be present. In many cases, using a probability distribution to describe this variation may be the best approach.

The NRCS assumes that the CN is constant for all rain depths for a specific site. However, they specify several limitations, including:

- the CN method is less accurate when the runoff is less than 0.5 inch. It is suggested that an independent procedure be used for confirmation,
- the CN needs to be modified according to antecedent conditions, especially soil moisture before an event, and
- the effects of impervious modifications (especially if they are not directly connected to the drainage path) needs to be reflected in the CN.

Few of these warnings are considered by most storm drainage designers, or by users of NRCS CN procedures for stormwater quality analyses. Figure 33 shows the typical pattern obtained when plotting CN against rain depth. The CN for small rain depths is always very large (approaching 100), then it decreases as the rain depth increases. At some point, the observed CN values equal the NRCS published recommended CN. During rains smaller than this matching point, the actual CN is greater than the NRCS CN. Predicted runoff depths would therefore be much less than the observed depths during these rains. Very large differences in runoff depths are associated with small differences in CN values, making this variation very important.

Figure 34 shows the observed peak runoff flow rate versus the peak rain intensity. If the averaging period for the peak flows and peak rain intensities were close to the catchment time of concentration (t_c), the slope of this relationship would be comparable to the Rational coefficient (C). The averaging times for the peak values probably ranged from 5 minutes to 1 hour for the different projects. Unfortunately, this averaging time period was rarely specified in the data documentation. Most urban area t_c values probably range from about 5 to 15 minutes. As indicated in this figure, the relationship between these two parameters shows a general upward trend, but it would be difficult to fit a statistically valid straight line through the data. As noted above for the other two drainage design procedures, actual real-world variations (coupled to measurement errors) add a lot of variation to the predicted runoff flow and volume estimates. Most drainage designers do not consider the actual variations that may occur.

Figure 35 shows an example plot of the ratio of the peak runoff flow rate to the average runoff flow rate versus rain depth. These values can be used to help describe the shape of simple urban area hydrographs. If the hydrograph can be represented by a simple triangular hydrograph, then the peak flow to average flow ratio must be close to 2. As shown on these figures, this ratio is typically substantially larger than 2 (it can never be less than 1 obviously), indicating the need to use a somewhat more sophisticated hydrograph shape (such as a double triangular hydrograph that can consider greater flows). These plots indicate if this ratio can be predicted as a function of rain depth. In most cases, values close to 2 are seen for the smallest rains, but they ratio increases to 5, or more, fairly quickly, but with much variability.

Example plots for total suspended solids, COD, phosphorous, and lead are shown on Figures 36 through 39. It is commonly assumed that runoff pollutant concentrations are high for small rains (and at the beginning of all rains) and then taper off (the "first-flush" effect). As indicated on these plots, concentration has a generally random pattern. In many cases, the highest concentrations observed will occur for small events, but there is a large variation in observed concentrations at all rain depths. The upper limits of observed concentrations may show a declining curve with increasing rain depths, but the concentrations may best be described with random probability distributions. Analyses of concentrations versus antecedent dry periods can reduce some of this variability, as can analyses of runoff concentrations from isolated source areas.

Evaluation of Data Groupings and Associations

The telecommunication data was evaluated to identify correlations between various site characteristics and sediment and water quality. In addition, relationships between different parameters were also examined to find measurements that correlated with one another.

The most obvious correlation of the data with site conditions and with other parameters was for the very high winter dissolved solids and conductivity values in EPA Rain Region 1 compared to other seasons and areas. The snowmelt runoff during the winter seasons in the northeast dramatically affected the winter season quality of the sampled water collected for NYNEX and Bell Atlantic, especially for TDS and conductivity. In addition, increased dissolved solids and conductivity values were also found in some east coast locations that were tidally influenced by close-by brackish waters. Because of the very high chloride ion concentrations, several of the analytical methods were subjected to large interferences (especially the major ions by ion chromatography). These samples were re-analyzed using other methods less subject to interference to better determine the maximum concentrations, especially for nitrates.

The large amount of data collected during this project and the adherence to the original experimental design enabled a comprehensive statistical evaluation of the data. Several steps in data analysis were performed, including:

- exploratory data analyses (mainly probability plots and grouped box plots),
- simple correlation analyses (mainly Pearson correlation matrices and associated scatter plots),
- complex correlation analyses (mainly cluster and principal component analyses, plus Kurskal-Wallis comparison tests), and
- model building (based on complete 2^4 factorial analyses of the most important factors)

The following discussion presents the results of these analyses.

Exploratory Data Analyses

A series of plots were prepared that represented data relationships and groupings, arranged by parameter sets (solids, common parameters, bacteria, other sewage indicators, nutrients, heavy metals, and organics). Included for most parameters are the following plots:

- grouped box and whisker plots for all data, by season
- grouped box and whisker plots showing all residential and commercial/industrial data, separated by season and age,
- grouped box and whisker plots for all data by EPA rainfall zone and season
- grouped box and whisker plots separating data by company, season, age, and land use.
- overall probability plots
- probability plots separated by land use
- probability plots separated by age of development
- probability plots separated by season

The data indicated that the sampling effort needed as previously described was appropriate. Some of the parameters had high COV values, while others were more moderate, as expected. In almost all cases, the overall data for each constituent was best described using log-normal probability plots (the notable and obvious exception is for pH). This requires the use on nonparametric statistical methods, or transformations of the data using $\log_{10}$. The following discussion presents some of the obvious trends and relationships noted from these plots:

Solids Measurements in Water and Sediment Samples

The highest total solids observations were from older commercial and industrial areas. Winter water samples had the highest concentrations, followed by spring and summer observations, while the fall samples had the lowest concentrations. Almost all of the total solids were in the dissolved form (with a median TDS concentration of about

450 mg/L), with only small contributions from the suspended solids (median SS concentration of 20 mg/L). About 15% of the total and dissolved solids were in volatile forms, while about 50% of the suspended solids were in volatile forms.

The highest dissolved solids concentrations were observed during the winter sampling periods, with some TDS concentrations greater than 10,000 mg/L. The highest values were observed in samples from EPA rainfall zone 1 (specifically at NYNEX older residential sampling locations during the winter). Older commercial and industrial Bell Atlantic sites showed distinct trends in TDS by season, with the highest values observed during the winter, and then with steadily decreasing values through the year, with the lowest observed values during the fall season. The high TDS values associated with winter snowmelt inflow decreased by about ten-fold by the fall, likely by the less saline inflowing stormwater during the late spring, summer, and early fall seasons, or they may have been affected by local groundwaters that change in dissolved solids with time. A similar pattern was also observed at the SNET older residential, and the Ameritech mid-aged and older residential locations. Therefore, this pattern is very likely common to most areas using de-icing salts. Similar patterns were also observed for many of the conductivity measurements. Many of the AT&T sites in northern areas that were sampled in the summer of 1998 also had high TDS values, but the following winter samples were much lower in TDS, possibly because these winter samples may have been collected previous to the snowmelt season. Some of the coastal locations were noted to be directly affected by tidal conditions, with continuous high dissolved solids and conductivity conditions.

There were no apparent overall trends for turbidity by season, although the overall range observed was quite large (from <1 to about 2,000 NTU, with a median value of about 7 NTU). Filtration through 0.45 μm membrane filters reduced the turbidity values significantly (the maximum was reduced to about 45 NTU and the median to about 0.8 NTU). The largest turbidity values observed were from water samples collected from mid-aged and older residential areas located in EPA rainfall zones 1 and 3 (some samples from Bell Atlantic older residential areas approached 2,000 NTU). Samples from EPA rainfall zone 3 (especially newer residential area BellSouth samples) do indicate seasonal differences in turbidity, where the summer and (especially) fall samples averaged several times greater than the winter and spring samples. The BellSouth new residential area samples collected during the fall also had some of the highest turbidity values observed (several hundred NTU). A less distinct, but similar pattern, may also occur for EPA rainfall zone 2 samples.

The sediment had volatile contents ranging from <1 to about 70%, while the median volatile content was about 6%. There were no obvious relations of sediment volatile content for different seasons, land uses, or age of development.

Common Constituent Measurements in Water and Sediment Samples

A possible overall trend indicated lower pH values from spring water samples (median of about 7), higher pH values from winter and summer samples (medians of about 7.3), and the highest pH values (median of about 8) from fall samples. The fall samples from both residential and commercial/industrial areas were much higher than for the other three seasons. Only EPA rain regions 1, 2, and 3 had fall and spring samples, and all three of these areas experienced high fall samples. Rain regions 5, 6, and 9 showed lower summer pH values than for the winter samples.

There was also a wide range in color of the water samples, with no apparent overall relationships with season, age, or land use. In rain region 2, the summer and fall samples had higher colors than the winter and spring samples, especially for samples from older commercial/industrial areas. Many of the newer samples (from GTE, SNET, and PacBell sampling) also had much more color in the fall samples than in the winter samples. Residential area samples also had higher levels of color than samples from industrial and commercial areas.

COD did not vary greatly for different land uses, seasons, or age of development. About 20% of the samples did not have detectable COD, but maximum values approached 400 mg/L, and the median value was about 15 mg/L. Filtration reduced the overall COD values by about 30%, with the median filterable COD being about 10 mg/L, and the maximum filterable COD approaching 300 mg/L. The sediment COD values ranged from about 1,000 to 300,000 mg/kg, with the median about 85,000 mg/kg. These sediment COD values appear high, but about 75% of the volatile solids observations of the sediment had more than 10% volatile solids. The sediment samples from new areas had much lower COD values than sediment samples from older areas.

The hardness values of spring water samples were generally higher (harder), while the fall samples were generally lower (softer) than for the other seasons.

There was no overall pattern observed for ammonia measured in the water samples. The highest observations (up to 45 mg/L) were from samples collected from EPA rain region 1, especially during the winter and fall. Most of the ammonia observations were quite low, with very few exceptions. The highest nitrate observations (close to 200 mg/L) were from new commercial and industrial areas sampled in rain zones 1 and 3. The highest phosphate concentrations observed (about 20 mg/L) were from older residential areas, although water from older commercial and industrial areas also had relatively high phosphate concentrations (up to about 2 mg/L). EPA rain region 3 had the highest phosphate observations for each season.

About 300 water samples were analyzed for *E. coli* and enterococci from the samples collected during the later part of the project. Therefore, few samples were analyzed from the original project participants. Generally, bacteria was much reduced during colder winter periods in stormwater. However, when observing patterns for enterococci, the overall median values were quite similar for all seasons, while the median summer *E. coli* observations were substantially higher than for the other seasons. The bacteria values were highly variable, with similar ranges for the residential and the commercial areas. When examining the data for the different EPA rain regions, the winter samples from zone 1 (a colder area) had much lower bacteria counts less than the corresponding summer samples, while in zone 6 (a hot area) samples had reduced summer bacteria observations. Air temperatures during sampling ranged from about 15^{o}F to 100^{o}F. This implies that either extreme cold or hot weather conditions may reduce bacterial survival, as expected. Similar patterns were also found for enterococci bacteria observations.

Detergent, boron, fluoride, and potassium measurements were used as indicators of sanitary sewage contamination. Boron concentrations were higher in industrial and commercial areas compared to residential areas, fluoride concentrations were higher during the summer sampling periods, while potassium was highest in older areas. No other patterns were apparent for these constituents.

Heavy Metal and Organic Toxicant Measurements in Water and Sediment Samples

The toxicity screening tests (using the Azur Microtox® method) conducted on both unfiltered and filtered water samples indicated a wide range of toxicity, with no obvious trends for season, land use, or age. About 60% of the samples were not considered toxic (less than a I25 light reduction of 20%, the light reduction associated with the phosphorescent bacteria after a 25 minute exposure to undiluted samples), about 20% are considered moderately toxic, while about 10% are considered toxic (light reductions of greater than 40%), and 10% are considered highly toxic (light reductions of greater than 60%). Samples from residential areas generally had greater toxicities than samples from commercial and industrial areas. Samples from newer areas were also more toxic than from older areas. Further statistical tests of the data indicated that the high toxicity levels were likely associated with periodic high concentrations of salt (in areas using deicing salt), heavy metals (especially filterable zinc, with high values found in most areas) and pesticides (associated with newer residential areas).

Heavy metal concentrations have been evaluated in almost all of the water samples for copper, lead and zinc, and some filtered samples have been analyzed for chromium. From 564 to 674 samples (82 to 99% of all unfiltered samples analyzed) had detectable concentrations of these metals. Filterable lead concentrations in the water were as high as 173 μg/L, while total lead concentrations were as high as 810 μg/L. The winter Ameritech new residential areas had the highest zinc concentrations observed, with one value greater than 20,000 mg/L. The repeat samples from the following summer were much lower and more typical. The initially very high values may indicate increasing zinc concentrations as the water stands in the manholes for extended periods. Many of the zinc values were higher than 1,000 mg/L in both filtered and unfiltered samples. Some of the copper concentrations have also been high in both filtered and unfiltered samples (as high as 1,400μg/L). Chromium concentrations as high as 45 μg/L were also detected.

About 390 sediment samples were analyzed for heavy metals. An ICP/MS was used to obtain a broad range of metals with good detection limits. The following list shows the median observed concentrations for some parameters in the sediments (expressed as mg of the metal per kg of dry sediment):

Aluminum	14,000 mg/kg
Barium	50 mg/kg
Calcium	17,000 mg/kg
COD	85,000 mg/kg
Chromium	<10 mg/kg
Copper	100 mg/kg
Lead	200 mg/kg
Magnesium	5,000 mg/kg
Manganese	200 mg/kg
Nickel	<10 mg/kg
Strontium	35 mg/kg
Vanadium	<10 mg/kg
Zinc	1,290 mg/kg

The overall copper patterns indicate that the highest concentrations (over 1,000 μg/L) were found in samples obtained from older residential areas, especially in EPA rain zone 3, with almost as high copper values observed in some older commercial and industrial areas. Filtration did not significantly reduce the highest copper observations, but reduced most others by about 50%. Sediment from old areas had greater copper concentrations than sediment from new areas.

Lead concentrations were also highest (about 1,000 μg/L) in older residential area water samples, while samples from some older commercial and industrial areas also had high values. Rain zone 3 summer and fall lead observations were substantially larger than corresponding winter and spring observations. A similar, but smaller, difference was also noted for zone 1. This pattern was especially obvious for older commercial and industrial samples collected by BellSouth. Filtration significantly reduced the lead concentrations by about 75%. Filtered samples from zone 3 collected during the summer and fall were still greater than the samples collected during the winter and spring. Sediment from old areas also had greater lead concentrations than sediment from newer areas.

Residential area samples generally had larger zinc concentrations than the samples from commercial and industrial areas. Samples from the newest areas also had higher zinc concentrations compared to samples from older areas. Filtration reduced the highest zinc concentrations (about 3,600 μg/L) by about 20%, and most of the other values by about 35%. No overall patterns were observed for zinc concentrations in sediment samples.

Water samples from more than 600 locations were analyzed and verified for base neutral and acid extractable organic toxicants. About 120 of these samples were partitioned by filtering to identify the quantity of organics associated with the particulates and how much is soluble. Very few detectable organics were found, especially in the filterable fraction, even with the GC/MSD method detection limits ranging from 2 to 5 μg/L. The most common organic compounds found are listed below:

di-n-butyl phthalate: detected in 3.0% of the unfiltered water samples, maximum concentration of 4.7 μg/L
benzylbutyl phthalate: detected in 1.2% of the unfiltered water samples, maximum concentration of 21 μg/L
bis(2-ethylhexyl) phthalate: detected in 1.2% of the unfiltered water samples, maximum concentration of 15 μg/L
coprostanol: detected in 3.5% of the unfiltered water samples, maximum concentration of 80 μg/L

The phthalate ester compounds are probably associated with plastic components in the sampling areas. Coprostanol was also detected in many of the samples. This compound is used to help identify the presence of fecal contamination as high concentrations may imply sanitary sewage contamination of the water or pet wastes. Obviously, the median concentrations of these compounds were below the detection limits.

Water samples from about 580 manholes were analyzed for pesticides, with about 50 also filtered for partitioning pesticide analyses. Again, the pesticides were only detected in small fractions of the samples analyzed, as shown below:

delta BHC: detected in 10.4% of the unfiltered water samples, maximum concentration of 5.7 μg/L
heptachlor: detected in 1.6% of the unfiltered water samples, maximum concentration of 0.58 μg/L
aldrin: detected in 4.3% of the unfiltered water samples, maximum concentration of 0.30 μg/L
endosulfan I: detected in 1.6% of the unfiltered water samples, maximum concentration of 0.04 μg/L
alpha chlordane: detected in 4.2% of the unfiltered water samples, maximum concentration of 0.11 μg/L
4,4'-DDE: detected in 14% of the unfiltered water samples, maximum concentration of 0.36 μg/L
endosulfan sulfate: detected in 1.0% of the unfiltered water samples, maximum concentration of 0.58 μg/L
4,4'-DDT: detected in 1.9% of the unfiltered water samples, maximum concentration of 0.06 μg/L
endrin ketone: detected in 3.0% of the unfiltered water samples, maximum concentration of 0.96 μg/L
methoxychlor: detected in 4.0% of the unfiltered water samples, maximum concentration of 0.2 μg/L

Only two organic compounds were detected in more than 10% of the water samples (delta BHC and 4,4'-DDE). While only one pesticide had an observed concentration greater than 1 μg/L (delta BHC), some of these pesticide concentrations may be considered relatively high.

One of the most striking features of the sediment samples was their visibly wide range of physical characteristics such as texture, color, and odor. The sediments ranged in texture from grainy sand to an extremely fine silt or sludge. Color ranged from clear quartz to white sand to red clay to black sludge. Multi-colored sheens were observed on a few sediment samples. Odor of the sediment samples ranged from no detectable odor to a scent of nutrient rich potting soil to clearly discernible diesel or other petroleum compounds, to sulfur and sewage. It was thought that these characteristics would be related to the presence of organic toxicants.

An evaluation of the sediment collected from the telecommunication manholes revealed that most of the sediment was of silt to sand texture, and brown in color, indicating a relatively low level of organic contamination for most sediments analyzed. About 4% of the samples were clayey and black, indicating potentially high levels of organic contamination, while another 4% were clayey and red, also indicating the potential presence of high levels of organic contaminants. Another 25% are in a marginal category, being dark in color, but not of the finest texture.

Simple Correlation Analyses

Pearson correlations and other association analyses were conducted with the data to identify relationships between the different parameters. This was done to identify sets of parameters that could possibly be used as indicators of problematic conditions, especially by substituting simpler and less expensive analyses for more costly or time-consuming analyses. Tables 12 and 13 summarize the significant correlations identified through typical Pearson correlation matrix analyses using SYSTAT, version 8. Pearson normalization removed the effects associated with the range and absolute values of the observations. Correlation coefficients approaching 1.0 imply near perfect relationships between the data. These tables show all of the correlation coefficients larger than 0.5, with those greater than 0.75 highlighted in bold. The pair-wise deletion option was also used to remove data in the analysis if data for one observation of a pair of parameters being compared was absent, but keeping the parameter in the complete table for other possible correlations. Also shown on these tables are the highly significant regression slope terms relating the dependent variables to the independent variables.

Table 12 are correlation pairings that are also obvious, and possibly also useful as indicators. Most of the coefficients are relatively high (up to 0.98), indicating mostly strong correlations. These relationships are between obviously related parameters, such as between total solids (TS) and conductivity (Figure 40), which has a coefficient of 0.84. The "obvious" relationship between turbidity and suspended solids, however, is relatively poor, at only 0.53 (Figure 41). It is therefore possible to use conductivity as a good indicator of TDS for almost all conditions, but using turbidity as a indicator for SS is more problematic. There were also relatively high correlations between filtered and total forms of solids, toxicity, COD, and zinc. The correlations between total and filtered forms of copper and lead were less, but still likely useful. The regression slope terms indicate that the filtered form of toxicity is about 91% of the unfiltered form, implying that very little toxicity reduction is accomplished with filtration. Of course, correlations between unfiltered and filtered constituents should generally be high, as the unfiltered concentrations should always be greater than the filtered concentrations.

Table 12. Obvious and Useful Correlations

Independent and Dependent Variables	Pearson Coefficient	Regression slope term
TDS and total solids	**0.98**	1.03
conductivity (μS/cm)and total solids	**0.84**	0.59
conductivity (μS/cm) and TDS	**0.85**	0.57
suspended solids and volatile total solids	0.60	0.58
suspended solids and volatile suspended solids	0.70	0.45
turbidity (NTU) and suspended solids	0.53	1.3
volatile total solids and volatile TDS	0.65	0.49
volatile total solids and volatile SS	**0.86**	0.61
toxicity and filtered toxicity (both light decrease)	**0.79**	0.91
COD and filtered COD	**0.76**	0.58
zinc and filtered zinc (both μg/L)	**0.78**	0.69
copper and filtered copper (both μg/L)	0.69	0.4
lead and filtered lead (both μg/L)	0.69	0.2

Table 13 shows the parameter correlations of additional interest, as these are not as obvious as those listed above. These correlations are generally weaker than those shown on the previous tables (these range from 0.5 to 0.75), but deserve further investigation. Especially interesting are the frequent correlations between the unfiltered and filtered forms of zinc and the total and unfiltered forms of toxicity, for example. Another useful correlation shown is between copper and lead, indicating the relatively common joint occurrence of these two heavy metals.

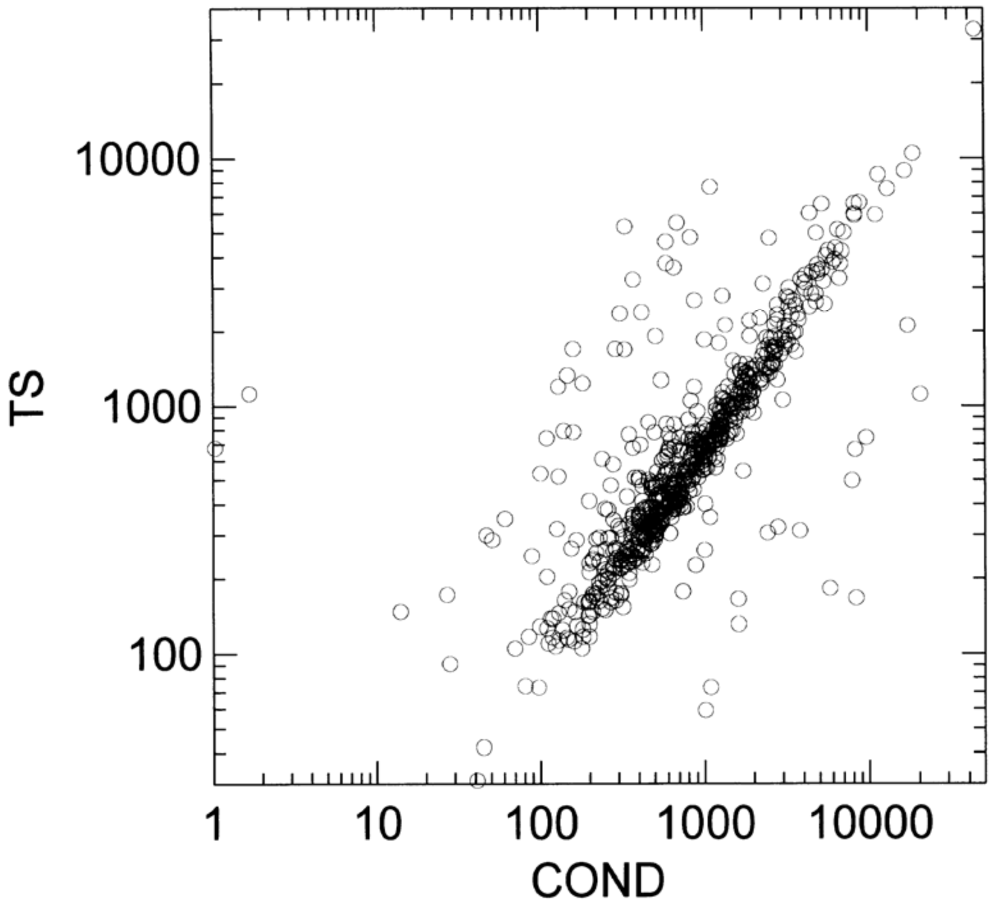


Figure 40. Strong correlation (0.84) between total solids and conductivity.

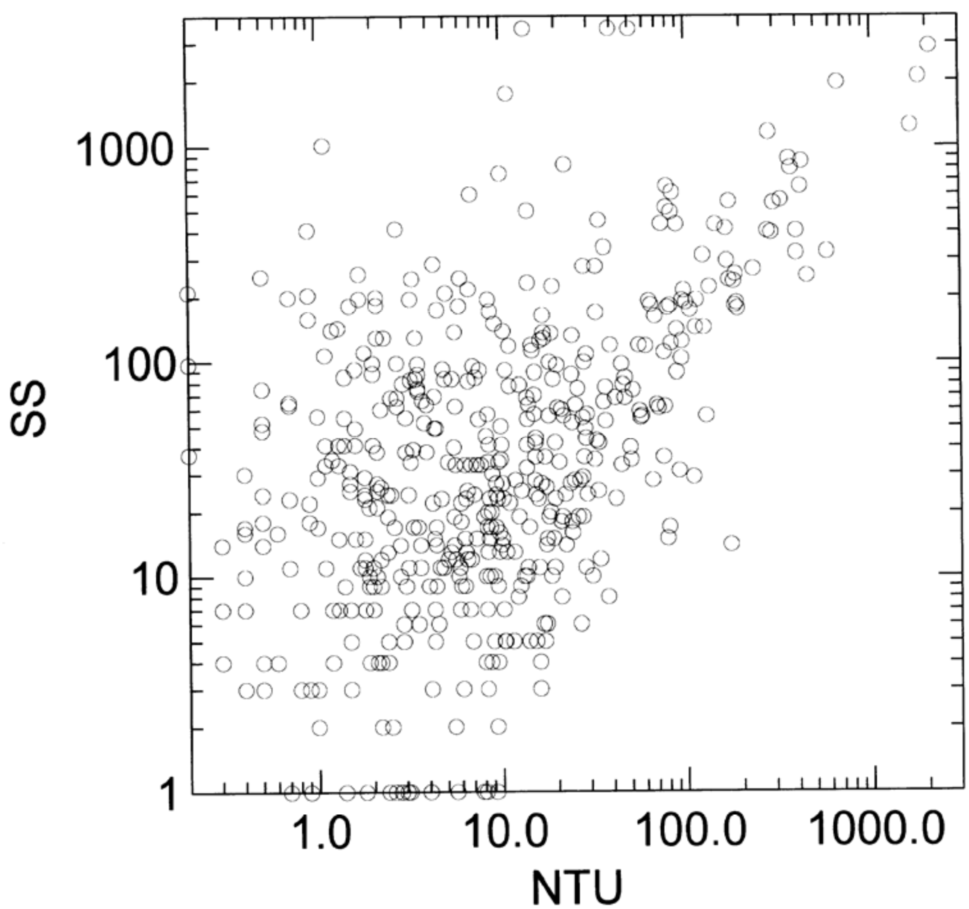


Figure 41. Weak correlation (0.53) between suspended solids and turbidity.

Table 13. Unexpected and Possibly Useful Correlations

Independent and Dependent Variables	Pearson Coefficient	Regression slope term
volatile TDS and hardness	0.66	1.3
filtered COD and phosphate	0.57	0.021
copper and lead (both μg/L)	0.52	0.32
zinc (μg/L) and toxicity (light decrease)	0.50	0.046
filtered zinc and toxicity (same as above)	0.55	0.058
zinc and filtered toxicity (same as above)	0.50	0.045
filtered zinc and filtered toxicity (same as above)	0.56	0.057
nitrate and ammonia	0.74	0.16

Complex Correlation Analyses

Additional analyses were conducted to identify more complex relationships between the measured parameters. These analyses do not prove any cause and effect relationship between parameters and conditions, but they do support a "weight-of-evidence" approach for reasonable hypotheses developed through different and supporting statistical methods. The complex correlation procedures used here examine inter-relationships between possible groups of parameters, compared to the pair-wise only comparisons presented earlier. Analyses between sub-groups of measurements, separated by expected important factors, are also presented.

One method to examine complex relationships between measured parameters is by using hierarchical cluster analyses. Figure 42 is a tree diagram (dendogram) produced by SYSTAT, version 8, using the water quality data for water samples collected from manholes. A tree diagram illustrates both simple and complex relationships between parameters. Parameters having short branches linking them are more closely related than parameters linked by longer branches. In addition, the branches can encompass more than just two parameters. The length of the short branches linking only two parameters are indirectly comparable to the correlation coefficients (very short branches signify correlation coefficients close to 1). The main advantage of a cluster analyses is the ability to identify complex relationships that cannot be observed using a simple correlation matrix.

In Figure 42, the shortest branches connect TDS and TS. As noted previously, almost all of the total solids are dissolved for these samples. Conductivity is also closely related to both TDS and TS. Other simple relationships are comparable to the higher correlation coefficients shown previously (Zn and filtered Zn, VTS and VSS, ammonia and nitrates, COD and filtered COD, etc.). There are relatively few complex relationships shown on this diagram: total toxicity is closely related to filtered toxicity and then to zinc and filtered zinc; phosphate is closely related to both copper and filtered copper; and hardness is related to the volatile solids.

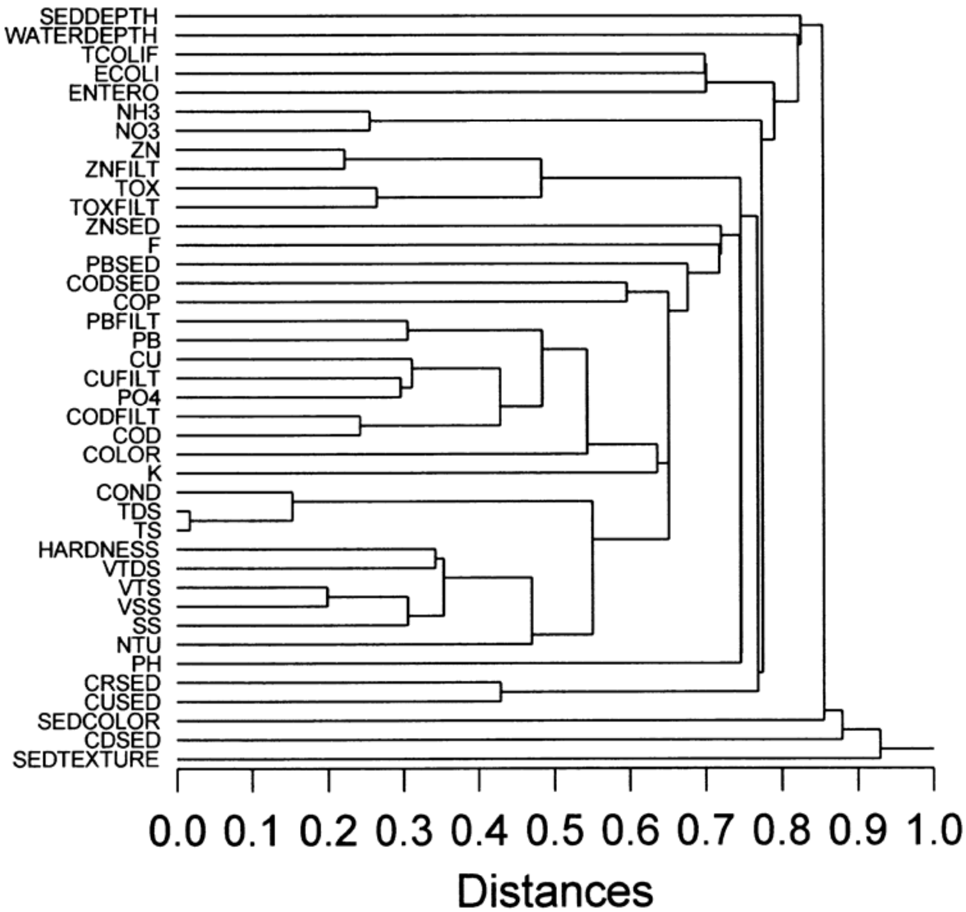


Figure 42. Dendogram showing complex relationships between monitored constituents

Another important tool to identify relationships and natural groupings of samples or locations is with principal component analyses (PCA). The data were auto-scaled before PCA in order to remove the artificially large influence of constituents having large values compared to constituents having small values. PCA is a sophisticated procedure where information is sorted to determine the components (usually constituents) needed to explain the variance of the data. Typically, very large numbers of constituents are available for PCA analyses with a relatively small number of sample groups desired to be identified. Component loadings for each principal component were calculated using SYSTAT, version 8, as shown in Table 14 (with the percent of the total variance explained for each component also shown).

Table 14. Loadings for Principal Components

Principal Component (% of total variance explained)	**1 (20.8%)**	**2 (14.2%)**	**3 (10.1%)**	**4 (9.4%)**	**5 (7.7%)**
Total solids	**0.771**	**-0.557**	0.011	0.190	0.104
TDS	**0.723**	**-0.629**	0.030	0.131	0.036
SS	0.424	0.322	-0.111	0.311	0.353
Turbidity	0.306	0.463	-0.110	0.381	0.381
pH	0.106	0.117	-0.338	-0.416	-0.206
Toxicity	0.269	0.173	0.339	0.154	**-0.674**
COD	**0.726**	0.304	0.057	-0.052	-0.037
Color	0.464	0.431	-0.059	-0.122	0.062
Conductivity	**0.649**	**-0.593**	0.041	0.193	0.058
Fluoride	0.280	-0.186	-0.177	-0.478	-0.045
Nitrate	0.170	0.183	**0.816**	-0.283	0.181
Phosphate	**0.571**	0.233	-0.154	-0.466	0.034
Hardness	0.385	-0.291	0.046	0.041	-0.278
Ammonia	0.107	0.088	**0.821**	-0.284	0.296
Potassium	0.344	0.031	-0.179	**-0.518**	-0.124
Zinc	0.206	0.355	0.265	0.370	**-0.613**
Copper	**0.521**	**0.523**	-0.211	-0.103	-0.056
Lead	0.298	0.488	-0.121	0.335	0.092

These first five components account for about 65% of the total variance of the data. The first two components are mostly dominated by total solids, TDS, COD, conductivity, phosphate, and copper. The third component is dominated mostly by nitrate and ammonia, the forth component is dominated by potassium, while the fifth component is dominated by toxicity and zinc.

Kurskal-Wallis nonparametric analyses were used like a one-way analysis of variance test to identify groupings of data that had significant differences between the groups, compared to within the groups. The groups examined were:

- Age
  - new (50 to 130 observations)
  - medium (65 to 150 observations)
  - old (100 to 300 observations)
- Season
  - winter (90 to 225 observations)
  - spring (50 to 100 observations)
  - summer (80 to 175 observations)
  - fall (50 to 115 observations)
- Land Use
  - commercial (75 to 200 observations)
  - industrial (30 to 65 observations)
  - residential (100 to 335 observations)

- Nearby traffic
  - light (85 to 160 observations)
  - medium (175 to 270 observations)
  - heavy (125 to 175 observations)

- EPA Rain Region
  - zone 1 (160 to 260 observations)
  - zone 2 (45 to 80 observations)
  - zone 3 (50 to 110 observations)
  - zone 4 (5 to 10 observations)
  - zone 5 (25 to 40 observations)
  - zone 6 (25 to 55 observations)
  - zone 7 (20 to 30 observations)
  - zone 8 (10 to 20 observations)

The number of data observations for each group component are also shown in the above list and has a significant effect on the probability of having a statistically significant difference between some of the group category components. The number of observations for some of the parameters are less than indicated, especially for those having low detection frequencies, or for screening parameters that were not evaluated for all samples. Most of the groupings had a large and relatively even number of observations in each subgroup. However, a few of the subgroups had small counts (such as for a couple of the rain zones). Table 15 lists the probabilities that the observed concentrations are the same amongst all of the categories. Probabilities smaller than 0.05 are traditionally considered significant and are indicated in bold.

Table 15. Kurskal-Wallis Probabilities that Concentrations are the same in each Category

mg/L, unless otherwise noted	**Total Number of Detectable Observations**	**Age**	**Season**	**Land Use**	**EPA Rain Region**
Total solids, mg/L	598	0.23	**<0.001**	0.53	**<0.001**
TDS	596	0.67	**<0.001**	0.4	**<0.001**
SS	483	**0.009**	0.25	0.36	0.21
VTS	598	0.1	**<0.001**	0.32	**<0.001**
VTDS	596	**0.028**	**<0.001**	0.13	**<0.001**
VSS	410	0.46	0.093	0.25	**<0.001**
Turbidity, NTU	598	0.67	**<0.001**	**0.002**	**<0.001**
pH	598	**0.03**	**<0.001**	**0.012**	**0.001**
Toxicity	394	**0.007**	0.086	0.14	**<0.001**
Toxicity, filtered	384	**0.001**	0.29	**0.024**	**0.001**
COD	596	**0.048**	**<0.001**	0.078	**<0.001**
COD, filtered	595	**0.001**	**<0.001**	**0.021**	**<0.001**
COD in sediment, mg/kg	320	**0.006**	0.79	**0.005**	**<0.001**
Color, color units	595	**0.026**	**0.032**	**0.035**	**<0.001**
conductivity, μS/cm	598	0.69	**<0.001**	0.53	**<0.001**
Total coliforms, #/100 mL	224	0.1	**<0.001**	0.5	**<0.001**
E. coli, #/100 mL	224	0.97	0.29	0.83	**<0.001**
Enterococci, #/100 mL	224	0.55	**0.001**	0.18	**0.018**
Fluoride, mg/L	594	0.57	**<0.001**	0.056	**<0.001**
NO3	595	0.26	0.064	0.78	**<0.001**
PO4	548	0.24	**<0.001**	0.36	**<0.001**
Hardness	598	**<0.001**	**<0.001**	**0.001**	**<0.001**
NH3	598	0.72	**0.039**	0.15	**<0.001**
K	593	**0.002**	0.11	**0.004**	**<0.001**
B	179	0.86	**0.009**	0.058	**0.031**
Zn, μg/L	541	**<0.001**	0.25	**<0.001**	**0.002**
Zn, filtered, μg/L	536	**<0.001**	0.19	**0.041**	**0.003**
Zn in sediment, mg/kg	275	0.78	0.45	0.56	**0.004**
Cu, μg/L	559	0.26	0.058	**0.009**	**0.008**
Cu, filtered, μg/L	554	0.079	**0.001**	0.063	**<0.001**
Cu in sediment, mg/kg	219	0.18	0.66	0.6	0.18
Pb, μg/L	555	**<0.001**	**0.013**	**0.029**	**0.005**
Pb, filtered, μg/L	552	**0.002**	**0.003**	**0.004**	0.084
Pb in sediment, mg/kg	237	**0.002**	0.58	0.76	0.33
Cr, μg/L	19	0.59	0.14	0.24	0.31
Cr in sediment, mg/kg	105	0.14	**0.001**	0.46	**<0.001**
Cd in sediment, mg/kg	45	0.055	**0.023**	0.091	0.11
di-n-butyl phthalate, μg/L	19	0.46	0.21	0.27	0.19
coprostanol, μg/L	23	0.061	0.62	0.71	0.15

The grouping that affected the most parameters was the EPA Rain Region, followed by the season, age, and lastly land use. The parameters affected by the most groupings were sediment accumulation, volatile total solids, filtered COD, hardness, potassium, and lead. Those affected by none of the groupings included chromium, and the organics (likely due to infrequent detections of these compounds). Zinc and copper sediment conditions were both affected by only one grouping each because of their relatively consistent concentrations found in all sediment samples.

Grouped box and whisker plots were prepared for selected parameters and for each grouping that was identified as having a significant difference during the Kurskal-Wallis analyses. Figure 43 shows high phosphate averaged concentrations associated with the southwest sampling locations, and with summer and winter seasons. Low averaged concentrations were noted in the southeast (although the largest phosphate concentration found was at a southeastern location).

Copper (Figure 44) had significant associations with different subcategories of region and land use. Copper (and lead) had very similar regional patterns, and copper, lead, and zinc all had higher average concentrations in residential areas.

Table 16 summarizes these associations.

Table 16. Significant Kurskal-Wallis Groupings

	Reasonable Associations	Opposite to Expected Associations
Total solids, mg/L	Geographical area	
TDS, mg/L	Geographical area, Season of sample collection	
Phosphate, mg/L	Geographical area, Season of sample collection	
Total coliforms, #/100 mL	Geographical area, Season of sample collection	
E. coli, #/100 mL	Geographical area	
Enterococci, #/100 mL	Geographical area, Season of sample collection	
Toxicity, I25, % light reduction		Age of surrounding area
Copper, μg/L	Geographical area, Land use	
Lead, μg/L	Geographical area, Land use Season of sample collection Age of surrounding area	
Zinc, μg/L	Geographical area, Land use	Age of surrounding area

Possible spurious correlations obviously occurred, although most of the associations appear reasonable and support the experimental design that directed the sampling effort. The age notation was periodically problematic for the field crews as it was sometimes difficult to obtain a reasonable estimate in areas that were very diverse.

Model Building

The most reasonable correlations (region, land use, age, and season) were used in these analyses to construct predictive models, based on the full-factorial sampling effort. The expanded geographical coverage, due to later-joining project participants from throughout the nation, allowed a geographical factor to also be considered in the final analyses. The sampling effort did not include a sufficient or representative number of areas to be sampled having other varying conditions of other potentially interesting factors. Therefore, the model building process was based solely on the full 2^4 factorial design using region, land use, age, and season, as the main factors, plus all possible interactions.

Since the experimental design was a full two-level factorial design, the following groupings were used to define the two levels used for each main factor, based on the number of observations in each grouping, the previous grouping evaluations, and the initial exploratory data analyses:

- age: old and medium combined (group A), vs. new (group B)
- season: winter and fall combined (group A), vs. summer and spring combined (group B)
- land use: commercial and industrial areas combined (group A), vs. residential areas (group B)
- region: EPA rain regions 1, 2, 8, and 9 (northern tier) (group A), vs. regions 3, 4, 5, 6, and 7 (milder) (group B)

The 597 sets of data observations used for this analysis were therefore divided into 16 categories corresponding to the complete factorial design, as shown in Table 17. Some samples did not have the necessary site information needed to correctly categorize the samples and were therefore not usable for these analyses. The "Group A" categories were assigned "+" values and the "Group B" categories were assigned "-" values in the experimental design matrix for the main factors. These 16 factorial groups account for all possible combinations of the four main factors. Twelve to more than 100 samples were represented in each factorial group and were used to calculate the means and standard errors.

Table 17. Factorial Design for Manhole Water and Sediment Characteristics

group	Number of observations in group	region	land use	age	season	region x land use	region x age	region x season	land use x age	land use x season	age x season	region x land use x age	region x land use x season	region x age x season	land use x age x season	region x land use x age x season
		R	L	A	S	RL	RA	RS	LA	LS	AS	RLA	RLS	RAS	LAS	RLAS
1	65	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	50	+	+	+	-	+	+	-	+	-	-	+	-	-	-	-
3	13	+	+	-	+	+	-	+	-	+	-	-	+	-	-	-
4	13	+	+	-	-	+	-	-	-	-	+	-	-	+	+	+
5	113	+	-	+	+	-	+	+	-	-	+	-	-	+	-	-
6	70	+	-	+	-	-	+	-	-	+	-	-	+	-	+	+
7	13	+	-	-	+	-	-	+	+	-	-	+	-	-	+	+
8	12	+	-	-	-	-	-	-	+	+	+	+	+	+	-	-
9	41	-	+	+	+	-	-	-	+	+	+	-	-	-	+	-
10	36	-	+	+	-	-	-	+	+	-	-	-	+	+	-	+
11	22	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+
12	19	-	+	-	-	-	+	+	-	-	+	+	+	-	+	-
13	42	-	-	+	+	+	-	-	-	-	+	+	+	-	-	+
14	47	-	-	+	-	+	-	+	-	+	-	+	-	+	+	-
15	21	-	-	-	+	+	+	-	+	-	-	-	+	+	+	-
16	20	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+

Table 18. Results of Full Factorial Statistical Tests on Characteristics of Water and Sediment Samples

		Total Solids (mg/L)	Dissolved Solids (mg/L)	Volatile Total Solids (mg/L)
Overall average:		957.84	884.96	157.67
Total number of observations:		590	588	590
Calculated polled standard error:		489.23	470.53	101.11
Standard error from high level interactions:		75.82	76.11	16.51
region	R	**700.34**	**678.59**	**92.78**
land use	L	127.88	119.74	19.09
age	A	90.01	63.86	-27.72
season	S	23.94	15.19	-17.96
region x land use	RL	195.61	216.54	9.55
region x age	RA	-8.82	-50.23	-16.27
region x season	RS	5.38	21.47	-26.46
land use x age	LA	115.83	112.09	38.23
land use x season	LS	-119.01	-125.23	-24.96
age x season	AS	44.41	25.81	5.36
region x land use x age	RLA	-69.60	-76.00	30.84
region x land use x season	RLS	23.94	15.19	-17.96
region x age x season	RAS	81.66	68.50	0.88
land use x age x season	LAS	-57.77	-60.19	8.43
region x age x land use x season	RALS	-115.40	-121.00	-4.12

Table 18. Results of Full Factorial Statistical Tests on Characteristics of Water and Sediment Samples (cont.)

		Volatile Dissolved Solids (mg/L)	Volatile Suspended Solids (mg/L)	Suspended Solids (mg/L) (direct)	% Volatile Solids of sediment	Turbidity Unfiltered (NTU)	Turbidity Filtered (NTU)	Toxicity Unfiltered (I25% Red)	Toxicity Filtered (I25% Red)	COD Unfiltered (mg/L)	COD Filtered (mg/L)
Overall average:		129.95	51.04	52.81	6.67	28.52	1.50	44.96	44.74	30.80	21.24
Total number of observations:		588	406	540	357	590	590	389	380	588	587
Calculated polled standard error:		88.98	70.46	50.58	3.02	30.07	1.09	17.02	16.95	18.27	13.94
Standard error from high level interactions:		10.62	15.30	11.27	0.94	6.11	0.25	5.58	4.31	4.40	2.60
region	R	**81.74**	15.77	11.15	2.65	18.72	0.26	-12.25	-5.74	-13.72	-7.71
land use	L	0.19	28.89	-19.23	-0.20	-19.27	-0.75	-9.23	**-14.91**	-0.21	-0.67
age	A	-40.60	17.26	11.20	2.00	5.52	-0.16	-8.14	**-13.54**	-4.29	-7.61
season	S	-21.48	-4.41	18.98	-1.81	7.06	-0.36	-6.17	-3.29	-0.84	-1.55
region x land use	RL	10.17	7.74	-25.70	1.46	-16.68	-0.19	6.91	-0.34	-0.90	0.73
region x age	RA	-42.08	41.10	31.68	-1.17	14.91	-0.11	7.13	1.11	4.57	2.64
region x season	RS	-17.99	-2.26	-14.79	-0.43	-7.70	-0.29	5.19	0.41	3.09	4.73
land use x age	LA	32.11	8.07	-17.18	-1.85	-8.14	0.25	0.99	6.23	-9.35	-4.49
land use x season	LS	-16.63	-6.12	-18.13	0.21	-5.79	0.46	2.88	6.45	-3.59	-2.13
age x season	AS	-1.55	9.76	-0.50	0.58	0.60	0.29	2.55	6.02	1.81	0.31
region x land use x age	RLA	6.65	32.73	-3.61	-0.01	-0.40	0.13	-6.26	-3.14	3.99	2.66
region x land use x season	RLS	-21.48	-4.41	18.98	-1.81	7.06	-0.36	-6.17	-3.29	-0.84	-1.55
region x age x season	RAS	-4.05	-1.62	6.25	0.95	3.85	0.25	-8.72	-8.05	-8.17	-4.12
land use x age x season	LAS	3.39	7.52	0.82	0.05	-7.96	-0.21	-0.88	-2.70	2.69	1.60
region x age x land use x season	RALS	-5.49	-4.59	-14.90	0.41	-7.62	-0.27	-1.29	-0.05	2.48	2.17

Table 18. Results of Full Factorial Statistical Tests on Characteristics of Water and Sediment Samples (cont.)

		COD mg/kg dry sediment	pH	Color Unfiltered	Color Filtered	Conductivity (μS/cm)	Total Coliform (MPN/100 mL)	*E. coli* (MPN/100 mL)	Enterococci (MPN/100 mL)	Fluoride (mg/L)
Overall average:		105200.92	8.59	49.19	27.66	1385.60	2056.96	171.56	398.13	0.39
Total number of observations:		333	590	590	590	590	225	225	225	586
Calculated polled standard error:		66053.07	7.95	49.48	24.69	742.01	1119.82	463.24	903.42	0.12
Standard error from high level interactions:		12760.55	1.97	12.98	5.60	129.75	621.91	129.53	326.53	0.05
region	R	**78579.17**	1.88	-4.32	-18.81	**1151.67**	458.14	198.71	-35.17	-0.05
land use	L	4532.72	2.11	-11.37	3.53	205.59	343.48	52.80	155.47	0.04
age	A	16182.47	1.95	-11.94	-12.61	30.29	-539.18	7.47	-399.15	0.00
season	S	-16815.29	-1.76	-5.17	3.27	244.08	-1103.80	-204.06	-363.19	-0.09
region x land use	RL	17137.33	2.10	-16.03	-5.64	450.21	-275.52	59.43	26.35	0.02
region x age	RA	-20079.08	1.98	3.98	7.11	45.29	137.95	-61.74	128.68	-0.05
region x season	RS	-11711.43	-2.06	-20.74	-4.58	-82.80	-1172.79	-204.99	-281.17	-0.02
land use x age	LA	-24373.31	2.00	-4.28	-10.80	86.87	-859.77	-104.25	-254.48	0.05
land use x season	LS	-568.16	-2.11	11.58	4.31	-12.40	-693.41	-118.71	-433.02	-0.02
age x season	AS	21520.74	-2.03	-1.50	-3.35	16.03	662.65	70.36	241.13	0.03
region x land use x age	RLA	-7907.59	2.07	18.68	8.21	-7.55	-101.03	-137.66	-204.62	-0.04
region x land use x season	RLS	-16815.29	-1.76	-5.17	3.27	244.08	-1103.80	-204.06	-363.19	-0.09
region x age x season	RAS	8863.49	-2.07	10.86	1.94	118.46	108.51	113.09	140.94	0.05
land use x age x season	LAS	-14278.79	-1.98	-18.07	-6.58	48.34	271.91	90.94	549.46	0.01
region x age x land use x season	RALS	13652.07	-1.97	-4.66	5.61	-90.35	-787.03	47.31	-193.72	0.03

Table 18. Results of Full Factorial Statistical Tests on Characteristics of Water and Sediment Samples (cont.)

		Nitrate (mg/L)	Phosphate (mg/L)	Hardness (mg/L as CaCO3)	Ammonia (mg/L)	Potassium (mg/L)	Boron (mg/L)	Zinc Unfiltered (μg/L)	Zinc Filtered (μg/L)
Overall average:		3.06	0.31	273.14	0.37	14.37	0.31	648.57	498.81
Total number of observations:		589	542	590	590	588	180	533	528
Calculated polled standard error:		8.09	0.31	107.02	1.74	13.99	0.52	269.34	247.20
Standard error from high level interactions:		2.02	0.07	21.68	0.43	2.07	0.13	89.21	91.39
region	R	0.28	-0.10	31.82	**0.49**	**-9.10**	0.14	-223.83	-122.58
land use	L	1.93	-0.15	-16.74	0.37	0.63	0.21	-246.26	-101.69
age	A	-2.80	0.19	-67.88	-0.39	2.42	-0.09	-252.09	-230.57
season	S	1.74	0.04	-32.70	0.45	-0.43	0.06	-124.14	-135.70
region x land use	RL	2.21	0.12	27.76	0.39	0.80	0.21	32.86	13.69
region x age	RA	-0.75	-0.14	-38.35	-0.39	-0.56	-0.12	119.95	34.16
region x season	RS	2.44	-0.08	4.28	0.48	1.40	0.12	34.49	28.42
land use x age	LA	-2.29	-0.21	**80.70**	-0.50	-1.92	-0.05	46.59	20.31
land use x season	LS	1.28	-0.07	-1.84	0.40	-3.49	0.15	33.52	-2.17
age x season	AS	-0.89	0.09	6.02	-0.40	-2.45	-0.11	81.38	114.03
region x land use x age	RLA	-1.77	0.13	-15.64	-0.39	-2.98	-0.14	-77.87	-79.89
region x land use x season	RLS	1.74	0.04	-32.70	**0.45**	-0.43	0.06	-124.14	-135.70
region x age x season	RAS	-2.64	0.01	-23.42	-0.37	3.01	-0.15	-127.26	-96.63
land use x age x season	LAS	-1.38	-0.05	17.47	-0.46	-1.77	-0.16	14.34	69.00
region x age x land use x season	RALS	-2.30	0.01	-13.45	-0.46	0.21	-0.11	43.75	53.51

Table 18. Results of Full Factorial Statistical Tests on Characteristics of Water and Sediment Samples (cont.)

		Zinc sediment (mg/kg)	Copper Unfiltered (μg/L)	Copper Filtered (μg/L)	Copper sediment (mg/kg)	Lead Unfiltered (μg/L)	Lead Filtered (μg/L)	Lead sediment (mg/kg)
Overall average:		3103.21	33.29	16.39	332.35	19.91	4.91	3178.74
Total number of observations:		271	552	546	215	547	544	233
Calculated polled standard error:		3347.84	33.60	20.36	na	17.99	4.77	na
Standard error from high level interactions:		841.43	4.02	3.81	142.50	4.99	1.00	4537.82
region	R	-80.81	-18.45	-16.63	-94.26	-4.57	-3.59	-4786.67
land use	L	-1410.25	-19.08	-9.93	23.15	-10.43	-2.74	-4718.76
age	A	-86.31	26.72	11.44	299.02	9.47	2.22	-3578.94
season	S	-806.70	2.65	5.54	-183.44	3.31	1.68	4510.31
region x land use	RL	5.26	17.28	9.14	64.67	1.92	0.68	4588.65
region x age	RA	-780.38	-9.25	-10.54	-135.48	3.76	-0.43	4451.73
region x season	RS	884.05	-9.42	-1.95	156.63	-4.55	-1.06	-4318.80
land use x age	LA	1021.87	-22.25	-8.17	-80.36	-5.67	-1.51	4490.85
land use x season	LS	357.50	-3.80	-4.51	-39.55	-2.59	-1.96	-4702.65
age x season	AS	469.26	0.93	2.05	-155.72	-2.02	-0.44	-4767.03
region x land use x age	RLA	128.72	7.27	6.13	-18.72	-7.09	0.55	-4459.42
region x land use x season	RLS	-806.70	2.65	5.54	-183.44	3.31	1.68	4510.31
region x age x season	RAS	-192.21	-0.25	-0.03	226.68	6.29	1.08	4874.38
land use x age x season	LAS	-725.52	0.75	0.07	40.99	3.15	0.10	4669.87
region x age x land use x season	RALS	1519.59	-4.49	2.09	120.25	-3.71	0.84	-4142.03

The factorial analyses were conducted using the group means. In addition, all parameters were also transformed by $\log_{10}$ to account for their correct log-normal data distributions. Table 18 shows the results of these analyses. Ten parameters were found to have significant models, with the most commonly occurring significant factor being the geographical region. Several parameters had significant interacting factors. All of the calculated effects for each parameters were plotted on probability plots (examples shown on Figures 45 through 47) to confirm the significant factors, which are indicated in bold type on Table 18.

Ten models were identified that had significant factors or combinations of factors. These models are listed below, along with the calculated values corresponding to the different levels for the significant factors:

Models with significant regional factors alone:

	R+ (northern tier states)	R- (milder climate)
Total solids (mg/L) = 958 + 350 R	1308 mg/L	608 mg/L
TDS (mg/L) = 885 + 339 R	1224 mg/L	546 mg/L
Volatile total solids (mg/L) = 158 + 46 R	204 mg/L	112 mg/L
Volatile dissolved solids (mg/L) = 130 + 82 R	172 mg/L	88 mg/L
Sediment COD (mg/kg) = 105,200 + 39,300 R	144,500 mg/L	65,900 mg/L
Conductivity (μS/cm) = 1390 + 576 R	1960 μS/cm	810 μS/cm
Potassium (mg/L) = 14.4 – 4.6 R	9.8 mg/L	18.9 mg/L

Model with significant land use and age effects alone:

	L+ and A+	L+ and A-	L- and A+	L- and A-
Filtered toxicity (I25%) = 44.7 – 7.5 L – 6.7 A	30.5 %	44.1 %	45.4 %	60.0 %

Models with significant land use and age interactions alone:

	LA+	LA-
Hardness (mg/L as $CaCO_3$) = 273 + 40 LA	313 mg/L	233 mg/L

Model with complex interactions with regional, land use, and season factors:

	RLS+	RLS-
Ammonia (mg/L) = 0.37 + 0.23 RLS	0.60 mg/L	0.14 mg/L

The effects and interactions are described below:

L+ and A+ (commercial or industrial and medium or old)
L+ and A- (commercial or industrial and new)
L- and A+ (residential and medium or old)
L- and A- (residential and new)

RLA+ (northern tier states and commercial or industrial and old; northern tier states and residential and new; milder climate and commercial or industrial and new; milder climate and residential and old)
RLA- (northern tier states and commercial or industrial and new; northern tier states and residential and old; milder climate and commercial or industrial and old; milder climate and residential and new)

RLS+ (northern tier states and commercial or industrial and winter; northern tier states and residential and summer; milder climate and commercial or industrial and summer; milder climate and residential and winter)

RLS- (northern tier states and commercial or industrial and summer; northern tier states and residential and winter; milder climate and commercial or industrial and winter; milder climate and residential and summer)

Obviously, the more complex interactions are more likely to be random, but the two-way interactions, and especially models having one or two main factors, are much more likely. The models containing only a single factor were mostly identified as being significant during the earlier described statistical tests.

Residual analyses were also conducted for each of these models, as shown on Figures 48 and 49. The predicted values were compared against all 597 data observations and their differences were plotted on probability plots. Legitimate models would produce residual probability distributions that are mostly random in nature (a straight line on a probability plot). These residual plots show that, in many cases, the upper 15 to 25 percent of the data are not adequately explained by the models. The models are therefore most useful to describe more typical conditions, from the lowest values to the 75^{th}, or possible higher, percentiles. The most extreme conditions that were observed in each category were more associated with factors other than those included in these models. As noted previously, much additional information was gathered and used in the simpler statistical tests previously presented that examined these other factors, but these other data were not adequately represented in each of the 16 major data grouping used in these factorial analyses. The following section examines the extreme conditions in more detail to attempt to identify patterns associated with the manholes that had the poorest water and sediment quality.

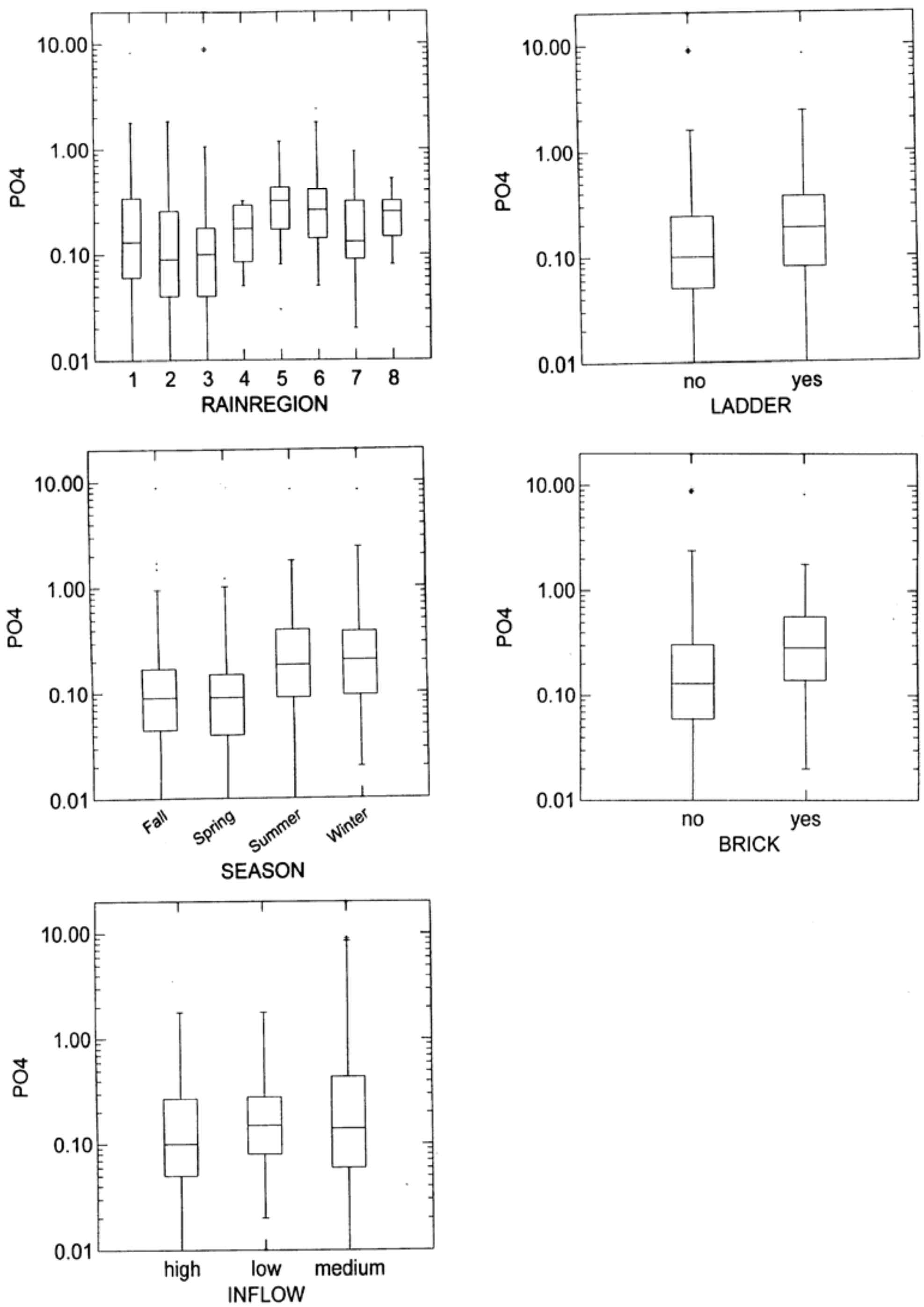
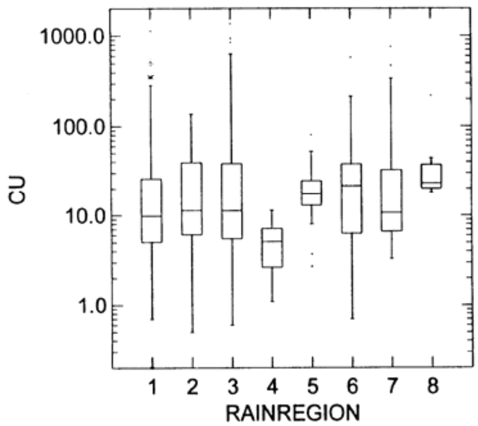
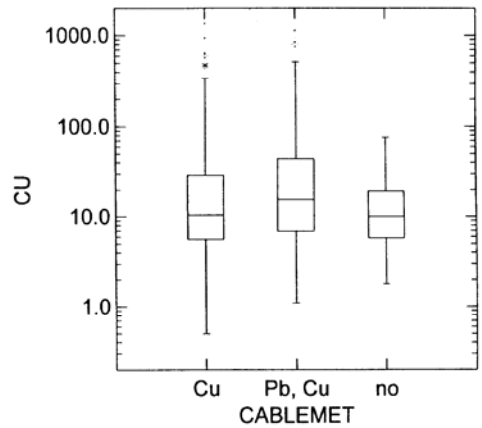
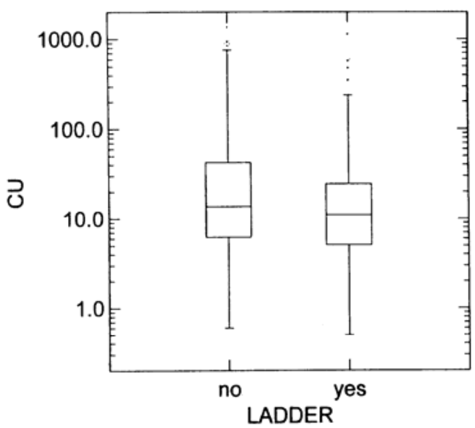
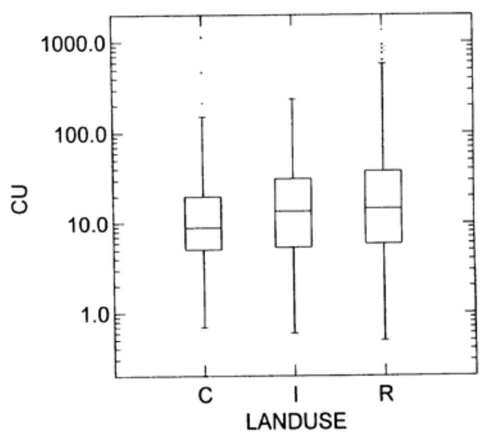


Figure 43. Statistically significant groupings for "phosphate" concentrations found in sampled water.


1000.0
100.0
10.0
1.0
CU
1
2
3
4
5
6
7
8
RAINREGION


1000.0
100.0
10.0
1.0
CU
Cu
Pb, Cu
no
CABLEMET


1000.0
100.0
10.0
1.0
CU
no
yes
LADDER


1000.0
100.0
10.0
1.0
CU
C
I
R
LANDUSE

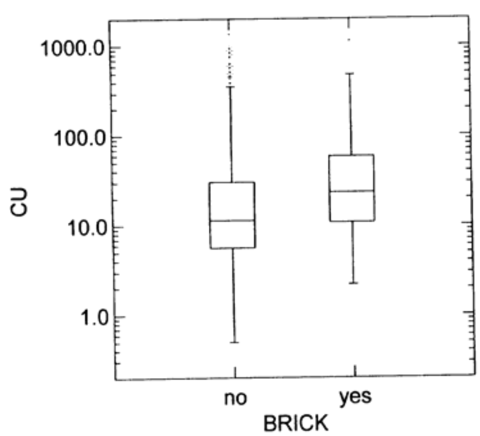

1000.0
100.0
10.0
1.0
CU
no
yes
BRICK

Figure 44. Statistically significant groupings for "copper" concentrations found in sampled water.

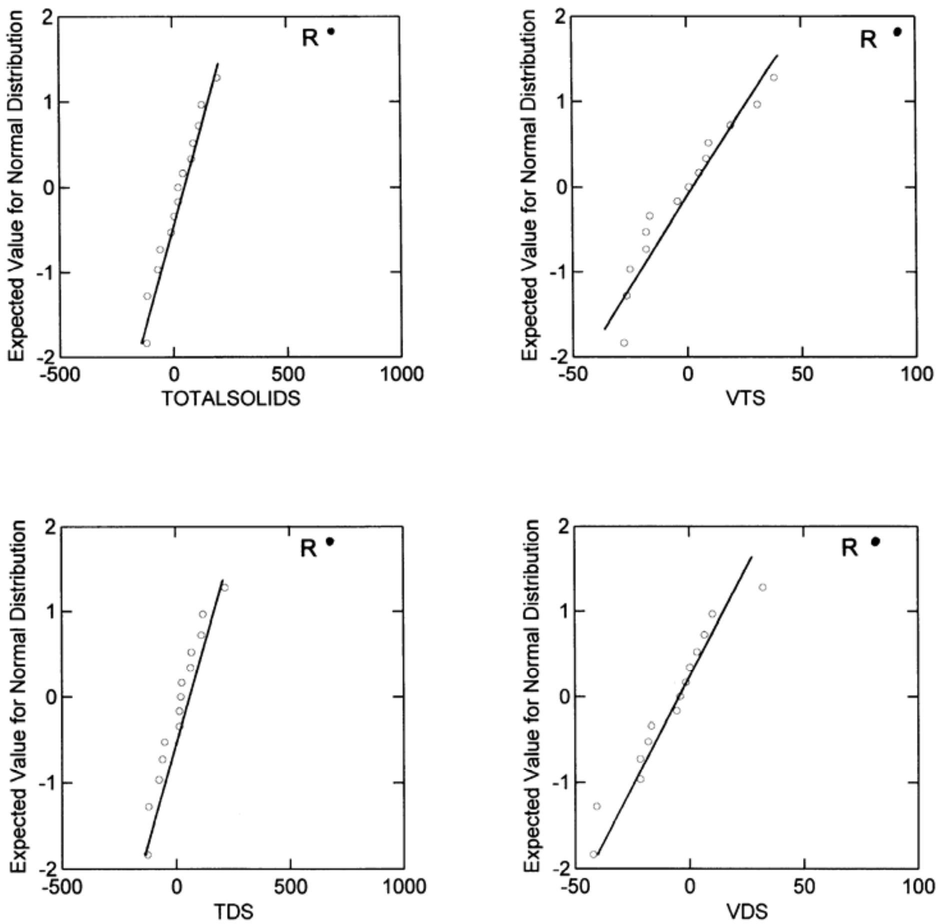


Figure 45. Significant main and interacting factors for solids concentrations in sampled water.

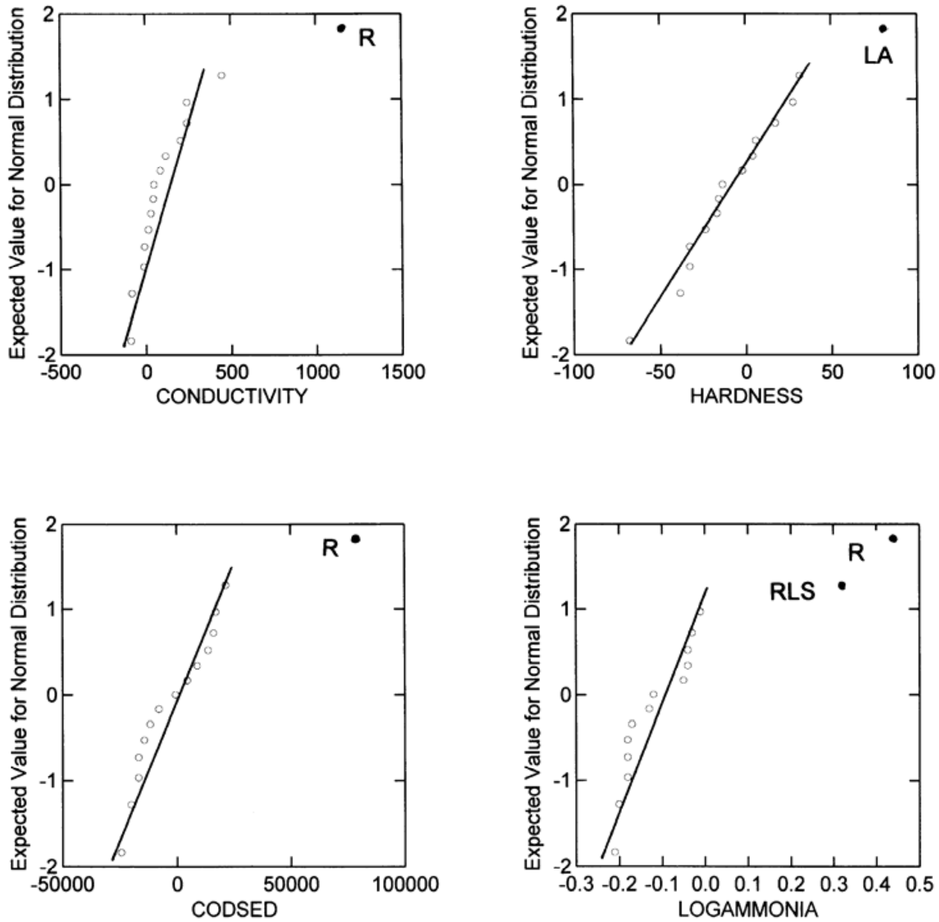


Figure 46. Significant main and interacting factors for common constituent concentrations in sampled water.

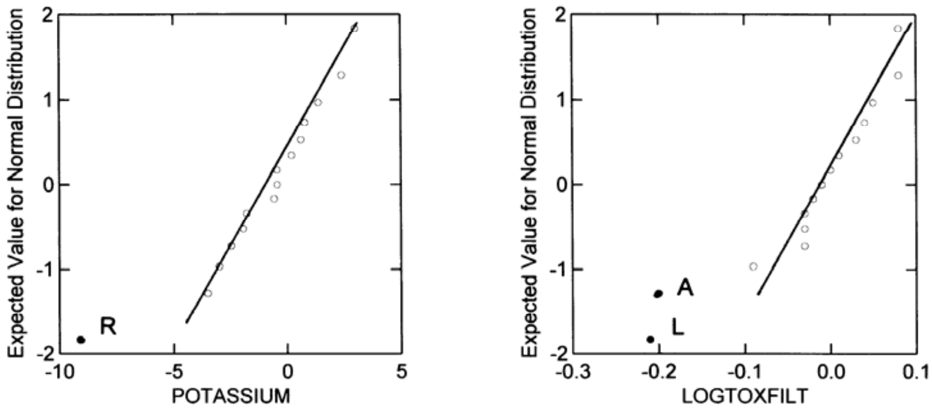


Figure 47. Significant main and interacting factors for potassium concentrations and toxicity in sampled water.

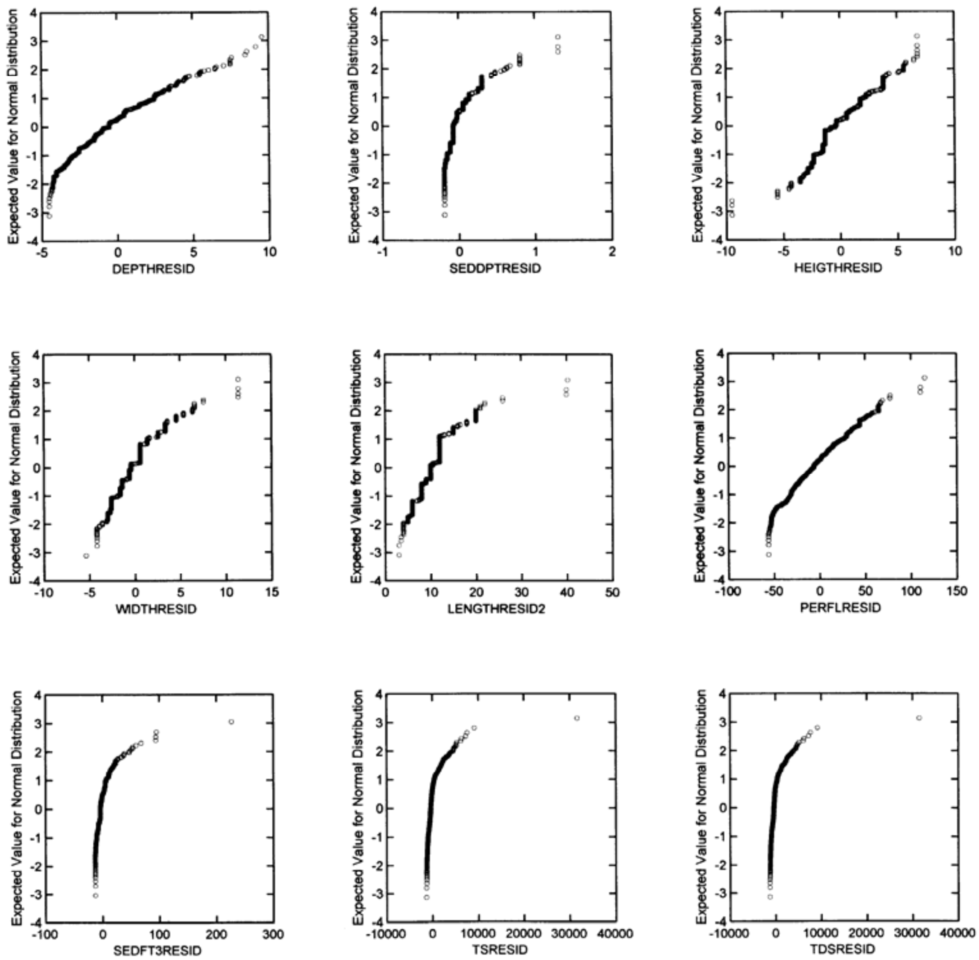


Figure 48. Residuals for significant factorial models.

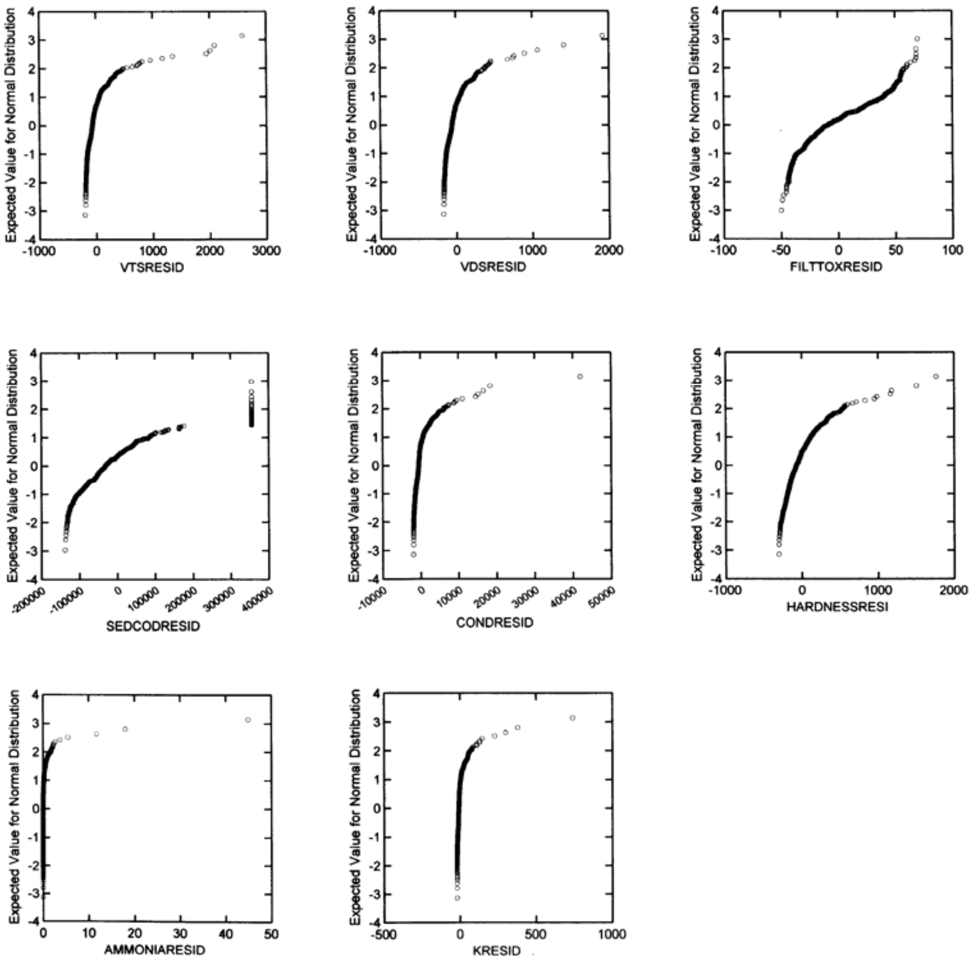


Figure 49. Residuals for significant factorial models (cont).

Figure 50 contains several very different plots that all have identical R^2 values. The use of the index of determination by itself can be misleading (data from Anscombe, in Draper and Smith 1981). The need for residual plots to confirm the regression assumptions and to visually examine the data, plus the use of ANOVA for evaluating the resulting regression equations is obviously critical.

As noted above, examination of the model residuals is a critical part of a model building exercise. When least-squares regression is used, residual analyses assist in confirming the requirements of the statistical test:

- the residuals are independent
- the residuals have zero mean
- the residuals have a constant variance (S2)
- the residuals have a normal distribution (required for making F-tests)

The residual analyses include several steps:

- Check for normality of the residuals (preferably by constructing a probability plot on normal probability paper and having the residuals form a straight line, or at least use an overall plot,
- plot the residuals against the predicted values,
- plot the residuals against the predictor variables, and

- plot the residuals against time in the order the measurements were made.

Figure 51 are example residual analysis plots, while Figure 52 shows several types of resulting patterns (Draper and Smith 1981). Only an even band is desired. Any curvature or tapering is undesirable and can likely be improved with data transformations.

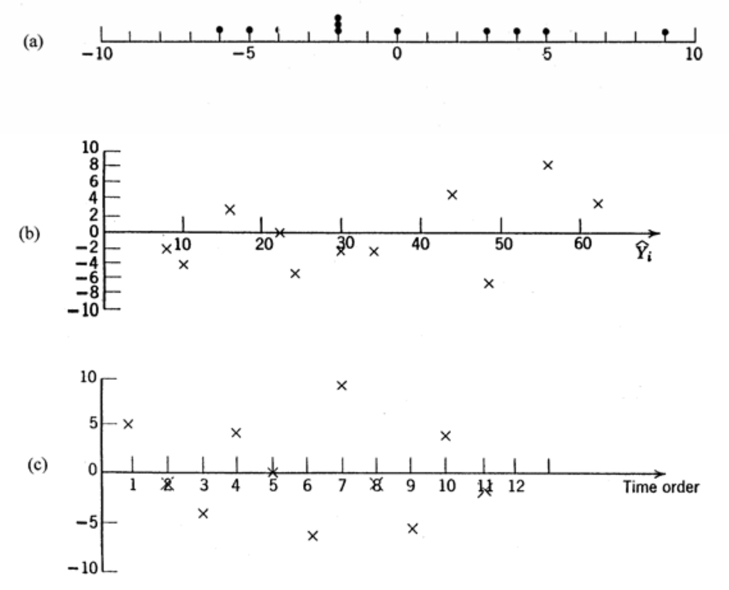


Figure 50. Example residual analysis plots (Draper and Smith 1981).

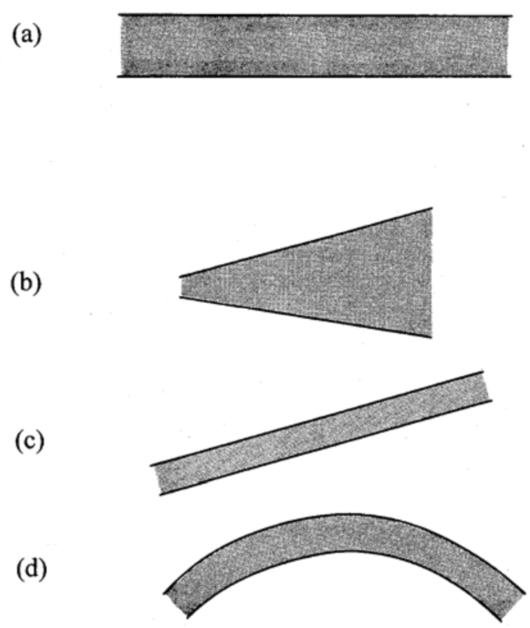


Figure 51. Possible residual patterns, only (a) is desired (Draper and Smith 1981).

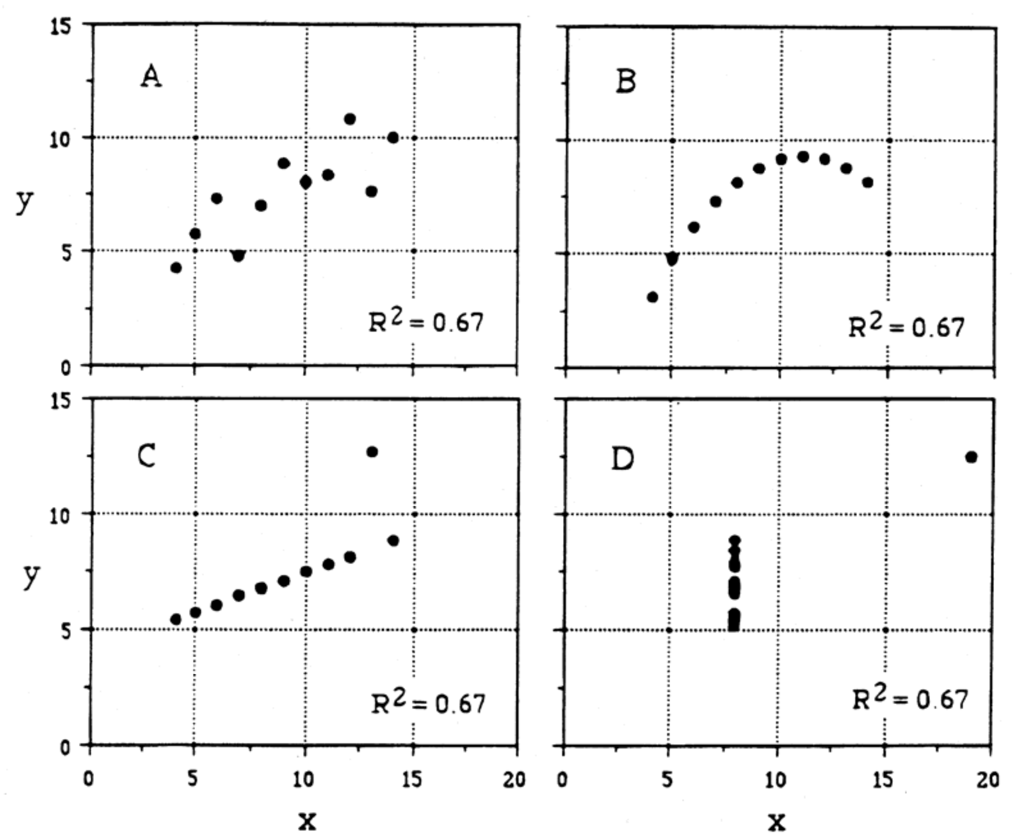


Figure 52. The use of the index of determination (R^2) can be misleading (data from Anscombe, in Draper and Smith 1981).

"Outliers" and Extreme Observations

Outliers are commonly detected using various statistical analyses and then eliminated from the data set to make analyses for straight forward and convenient. However, data should only be eliminated after much further examination, as extreme values may include highly valuable information. The following discussion presents an examination of the extreme values found during these monitoring activities.

As noted above, the factorial models developed for predicting the quality of water were not generally suited for the worst (extreme) cases. Since these situations are typically of high interest, further statistical analyses were conducted to identify patterns and conditions associated with these special locations. The most important water quality constituents (based on potential exceedences of criteria) were used to rank each location. The rankings were then averaged to identify the locations having the poorest quality water. The water quality constituents used for these rankings were as follows:

- Suspended solids
- Turbidity
- Conductivity
- Volatile total solids
- pH
- COD
- Phosphate
- Ammonia
- Nitrate
- Toxicity
- Copper
- Filtered copper
- Lead
- Filtered lead
- Zinc
- Filtered zinc

The observed water quality was ranked according to these constituents and the top ten percent where when compared to the other 90%. The locations selected in this group of high constituent values are shown on Table 19. Most EPA rain regions and all participating companies are represented in the list. In addition, about half of the samples were from locations during repeat samplings at other seasons. Since the areas were sampled during pumping operations, the repeated poor quality water found in these locations indicates that the sources of the poor quality water were relatively consistent for these areas and not the result of a single contaminating incident.

Table 19. Manholes Containing the Highest Water Quality Concentrations

Location	EPA Rain Region	Season	Age	Land Use
Ameritech				
4610 Tokay Blvd, Madison, WI	1	winter	old	resid
4610 Tokay Blvd, Madison, WI	1	summer	old	resid
402 Franklin St., Madison, WI	1	winter	old	resid
402 Franklin St., Madison, WI	1	summer	old	resid
5301 Cottage Grove Road, Madison, WI	1	winter	medium	resid
575 Science Dr., Madison, WI	1	winter	new	indus
Agriculture Drive, Madison, WI	1	summer	new	resid
1548 Carolina, Gary, IN	1	winter	old	resid
East 56th & Rosslyn, Indianapolis, IN	1	winter	old	resid
White & Edward Streets (NE corner), Frankfort, IL	1	winter	old	commer
Rte. 30 & School House Road (NE Corner), New Lenox, IL	1	summer	old	resid
Scovel between Grand Blvd & Vinewood, Detroit, MI	1	summer	old	resid
Grand River & Mackinaw, Detroit, MI	1	summer	old	commer
Old Fort & Woodruff, Rockwood, MI	1	summer	new	resid
Toledo-Dix South of Eureka, Southgate, MI	1	summer	old	commer
AT&T				
12th Avenue No. between 31st and 320 Streets, Billings, MT	8	summer	old	resid
MH #11672 - Highway 3, Billings, MT	8	summer	new	commer
Virginia Lane and Cotton Blvd., Billings, MT	8	summer	old	resid
19th & 20, Omaha, NE	9	winter	old	commer
6th Street & Willow, Omaha, NE	9	summer	old	resid
MH #21 27th & 20, Omaha, NE	9	summer	old	commer
MH #22 29th & 20, Omaha, NE	9	summer	old	commer
MH #112, Angelica, St Louis, MO	4	winter	old	commer
MH #322, St. Louis, MO	4	winter	new	resid
Highway 61/67, St. Louis, MO	4	summer	new	resid
MH #270, Vickers, St. Louis, MO	4	summer	old	commer
MH # 04, HW55 & Richardson Rd, St. Louis, MO	4	summer	new	commer
Bell Atlantic				
Rte. 123 N & Old Meadow Rd., McLean, VA	2	summer	medium	resid
Rte. 123 N & Old Meadow Rd., McLean, VA	2	winter	medium	resid
Marlboro Pike & Green Landing Rd., Prince Georges Cty., MD	2	summer	medium	resid
Marlboro Pike & Green Landing Rd., Prince Georges Cty., MD	2	spring	medium	resid
25 Plymouth St. (N of Rte. 46), Fairfield, NJ	1	spring	New	commer

Table 19. Manholes Containing the Highest Water Quality Concentrations (cont.)

Location	EPA Rain Region	Season	Age	Land Use
BellSouth				
8825 Jasper Rd., Jacksonville, FL	3	spring	old	resid
8825 Jasper Rd., Jacksonville, FL	3	summer	old	resid
8825 Jasper Rd., Jacksonville, FL	3	fall	old	resid
8825 Jasper Rd., Jacksonville, FL	3	winter	old	resid
NW 5th St. & 139th Av., Ft. Lauderdale, FL	3	spring	new	commer
NW 5th St. & 139th Av., Ft. Lauderdale, FL	3	summer	new	commer
NW 5th St. & 139th Av., Ft. Lauderdale, FL	3	fall	new	commer
Silver Palm Blvd. & NW 126th, Ft. Lauderdale, FL	3	summer	new	resid
Silver Palm Blvd. & NW 126th, Ft. Lauderdale, FL	3	fall	new	resid
Silver Palm Blvd. & NW 126th, Ft. Lauderdale, FL	3	winter	new	resid
Westward & Lenape Dr., Miami, FL	3	summer	old	resid
Westward & Lenape Dr., Miami, FL	3	fall	old	resid
4800 NW 102nd Av., Miami, FL	3	spring	new	resid
Coptek Rd., Pensacola, FL	3	summer	new	indus
GTE				
MH 0600056, Highway 45 S/LP#2 - Rantoul, IL	1	spring	medium	commer
MH 0600119, Rt 45 S, End of AF#1 - Rantoul, IL	1	fall	medium	resid
MH 1807, GE Rd - Bloomington, IL	1	fall	medium	resid
MH-1-DK-IL	1	winter		
47th & Rucker, Everett, WA	7	fall	old	resid
NE Dallas & NE 14th Avenue, Camas, WA	7	fall	medium	resid
NYNEX				
2011 Flatbush Av., Brooklyn, NY	1	winter	old	commer
2011 Flatbush Av., Brooklyn, NY	1	spring	old	commer
51st St. & 19th Av., Brooklyn, NY	1	summer	old	resid
51st St. & 19th Av., Brooklyn, NY	1	fall	old	resid
Dahill Rd. & 20th Av., Brooklyn, NY	1	winter	old	resid
North St. (across from St. Agnes Hosp.), White Plains, NY	1	winter	old	commer
Washington St. & Hudson St., Peekskill, NY	1	summer	old	resid
PacBell				
University Avenue & Lowell Street, La Mesa, CA	6	summer	old	commer
University Avenue & Lowell Street, La Mesa, CA	6	winter	old	commer
Green River Road & Crest Ridge Drive, Corona, CA	6	summer	new	resid
Green River Road & Crest Ridge Drive, Corona, CA	6	winter	new	resid
River Road & Archilbald Avenue, Norco, CA	6	summer	new	
Navajo Road & Park Ridge Street, San Diego, CA	6	winter	new	resid
SNET				
Norwalk Company Office, Washington St., Norwalk, CT	1	winter	old	commer
Wolcott Hill Rd. corner of Reed St., Weathersfield, CT	1	winter	old	resid

Table 19. Manholes Containing the Highest Water Quality Concentrations (cont.)

Location	EPA Rain Region	Season	Age	Land Use
U.S. West				
875 N. Beck Street (300 West), Salt Lake City, UT	8	winter	old	commer
875 N. Beck Street (300 West), Salt Lake City, UT	8	summer	old	commer
53 East Orpheum Ave (150 South), Salt Lake City, UT	8	winter	old	indus
53 East Orpheum Ave (150 South), Salt Lake City, UT	8	summer	old	indus
7th Street & Winged Foot, Phoenix AZ	6	summer	new	commer

Two-way cross-tabulations were used with SYSTAT, version 8, to identify groupings that were different for these top ten percent of the manholes compared to the other 90 percent of the data. The AT&T sites were not included in the analysis due to their being collected after the analyses were completed. The groupings examined were site characteristics noted on the field forms and included:

- EPA rainfall region
- Season of sample collection
- Age of surrounding area
- Land use of surrounding area
- Traffic in vicinity
- Site topography near manhole
- Road type

- Water odor
- Water clarity
- Water color
- Presence of surface sheen on water

- Sediment odor
- Sediment color
- Sediment texture

Pearson Chi-square statistics and the probabilities that the data subsets had the same distributions between the different groupings were calculated by SYSTAT, as shown on Table 20. The only groups that had significantly different groupings between the set of extreme observations and the rest of the observations (probabilities ≤ 0.05) were:

- Land use (more residential areas in the extreme group, and more commercial and industrial areas for the other 90% of the samples, opposite to what was originally expected)

- Water clarity (more cloudy and dark water in the extreme group and more clear water for the other 90% of the samples, as would be expected)

- Water color (more light, moderate, dark, and turbid water in the extreme group and more clear water for the other 90% of the samples, as would be expected)

- Sediment texture (more fine clay in the sediment for the extreme group and more coarser silt and sand in the sediment for the other 90% of the samples, as would be expected)

• Site topography (more moderate and steep slopes for the extreme group and more flat slopes for the other 90% of the samples, for unknown reasons)

These findings can be used to indicate a greater likelihood of high water quality constituent concentrations for water found in telecommunication manholes. It is recommended that areas having noticeable color and/or turbidity, along with sediments having a muddy texture (especially in residential areas) be given special attention.

Unfortunately, the use of these characteristics as the only screening tool results in substantial false negatives and false positives. As an example, combinations of these characteristics were compared to the complete set of samples, with the results summarized in Table 21. As the screening components increased, the number of hits was decreased, with increased "efficiency." The efficiency is calculated as the ratio of the rate of correct hits to total problem sites, compared to the total number of hits to the total number of sites. As an example, if 25% of the total sites were targeted (hits) and 50% of the problem sites were included in these hits, the efficiency would be 2.0. If the efficiency approaches 1.0, the number of problem sites identified is close to what would be expected with a random sampling, with no real benefit from using the screening criteria. As more criteria are included in the screening effort, the efficiency generally increases, but, unfortunately, so does the number of false negatives (ignores actual problems). The best plan may be to minimize the number of false negatives, while having a large efficiency factor. In this case, the use of color or land use may be best, if false negatives are to be reduced the most. If the largest number of correct hits of problem sites is desired for the least effort, then the combination of clarity, color, and texture is best (but with large numbers of false negatives because many problem sites will be missed).

As indicated, locations having colored and/or turbid water, especially with muddy sediments, should be examined more. Manholes located in residential areas (apparently especially newer areas) may also warrant additional attention, likely due to contaminated runoff water from landscaping maintenance operations.

Table 20. Cross-Tabulations of Sampling Area Characteristics Comparing Extreme Observations with Other Observations

EPA Rain Region	Other 90% of samples	Upper 10% of samples	Total %	Total number
1	42.599%	48.333%	43.160%	265
2	14.079	6.667	3.355	82
3 and 4	19.495	23.333	19.870	122
5 and 9	7.220	0.000	6.515	40
6	8.484	11.667	8.795	54
7	5.415	3.333	5.212	32
8	2.708	6.667	3.094	19
Total %	100.000%	100.000%	100.000%	
Total number	554	60	614	

Test statistic	Value	df	Probability that groups are the same
Pearson Chi-square	11.189	6.000	0.083

Season	Other 90% of samples	Upper 10% of samples	Total %	Total number
winter	37.184%	35.000%	36.971%	227
spring	15.704	11.667	15.309	94
summer	27.798	38.333	28.827	177
fall	19.314	15.000	18.893	116
Total %	100.000%	100.000%	100.000%	
Total number	554	60	614	

Test statistic	Value	df	Probability that groups are the same
Pearson Chi-square	3.264	3.000	0.353

Age of area	Other 90% of samples	Upper 10% of samples	Total %	Total number
new	21.561%	28.814%	22.278%	133
medium	26.580	15.254	25.461	152
old	51.859	55.932	52.261	312
Total %	100.000%	100.000%	100.000%	
Total number	538	59	597	

Test statistic	Value	df	Probability that groups are the same
Pearson Chi-square	4.103	2.000	0.129

Land use	Other 90% of samples	Upper 10% of samples	Total %	Total number
commercial	**44.853%**	29.310%	43.355%	261
industrial	**11.765**	6.897	11.296	68
residential	43.382	**63.793**	45.349	273
Total %	100.000%	100.000%	100.000%	
Total number	544	58	602	

Test statistic	Value	df	Probability that groups are the same
Pearson Chi-square	8.835	2.000	**0.012**

Table 20. Cross-Tabulations of Sampling Area Characteristics Comparing Extreme Observations with Other Observations (cont.)

Water odor	Other 90% of samples	Upper 10% of samples	Total %	Total number
none	93.721%	92.683%	93.631%	441
other	0.930	2.439	1.062	5
gasoline	1.163	2.439	1.274	6
sewage	4.186	2.439	4.034	19
Total %	100.000%	100.000%	100.000%	
Total number	430	41	471	

Test statistic	Value	df	Probability that groups are the same
Pearson Chi-square	1.569	3.000	0.666

Water Clarity	Other 90% of samples	Upper 10% of samples	Total %	Total number
clear	**77.979%**	46.341%	74.941%	320
cloudy	20.725	**41.463**	22.717	97
dark	1.295	**12.195**	2.342	10
Total %	100.000%	100.000%	100.000%	
Total number	386	41	427	

Test statistic	Value	df	Probability that groups are the same
Pearson Chi-square	30.769	2.000	**0.000**

Water Color	Other 90% of samples	Upper 10% of samples	Total %	Total number
clear	**55.764%**	27.660%	52.619%	221
light	18.231	**19.149**	18.333	77
moderate	13.941	**34.043**	16.190	68
dark	8.311	**8.511**	8.333	35
turbid	3.753	**10.638**	4.524	19
Total	100.000%	100.000%	100.000%	
Total number	373	47	420	

Test statistic	Value	df	Probability that groups are the same
Pearson Chi-square	21.078	4.000	**0.000**

Table 20. Cross-Tabulations of Sampling Area Characteristics Comparing Extreme Observations with Other Observations (cont.)

Surface sheen	Other 90% of samples	Upper 10% of samples	Total %	Total number
none	93.587%	90.909%	93.321%	517
partial	4.609	3.636	4.513	25
entire	1.804	5.455	2.166	12
Total	100.000%	100.000%	100.000%	
Total number	499	55	554	

Test statistic	Value	df	Probability that groups are the same
Pearson Chi-square	3.191	2.000	0.203

Sediment odor	Other 90% of samples	Upper 10% of samples	Total %	Total number
none	66.940%	48.649%	65.261%	263
other	2.186	2.703	2.233	9
gasoline	14.481	24.324	15.385	62
sewage	16.393	24.324	17.122	69
Total	100.000%	100.000%	100.000%	
Total number	366	37	403	

Test statistic	Value	df	Probability that groups are the same
Pearson Chi-square	5.114	3.000	0.164

Sediment color	Other 90% of samples	Upper 10% of samples	Total %	Total number
light	15.877%	16.216%	15.909%	63
medium	51.811	43.243	51.010	202
dark	32.312	40.541	33.081	131
Total	100.000%	100.000%	100.000%	
Total number	359	37	396	

Test statistic	Value	df	Probability that groups are the same
Pearson Chi-square	1.172	2.000	0.557

Sediment texture	Other 90% of samples	Upper 10% of samples	Total %	Total number
clay	13.774%	**37.838%**	16.000%	64
silt	**67.218**	45.946	65.250	261
sand	**19.008**	16.216	18.750	75
Total	100.000%	100.000%	100.000%	
Total number	363	37	400	

Test statistic	Value	df	Probability that groups are the same
Pearson Chi-square	14.620	2.000	**0.001**

Table 21. Examination of Screening Criteria to Identify Potentially Problematic Manholes

Characteristics	% of targeted samples correct	% of false positives (% of non-extreme sites included)	% of false negatives (% of total extreme sites missed)	Efficiency (rate of correct hits to total extremes to rate of hits to total observations)
Clarity x color x texture	62%	38%	87%	6.0
Color x land use x topography	24	76	83	2.5
Color x land use	26	74	62	2.5
clarity	20	80	63	2.0
color	17	83	43	1.7
texture	22	78	77	2.2
Land use	14	86	35	1.5
topography	11	89	52	1.1

Statistical Evaluation of a Water Treatment Control Device; the Upflow Filter

Controlled Experiments

Controlled sediment removal tests were also conducted for several media, different flow rates, and influent sediment concentrations. As shown in Figure 53, the percentage reductions for suspended solids for the mixed media tests and high influent concentrations (485 to 492 mg/L) were 84 to 94%, with effluent concentrations ranging from 31 to 79 mg/L for flows ranging from 15 to 30 gal/min. During the low concentration tests (54 to 76 mg/L), the reductions ranged from 68 to 86 mg/L, with effluent concentrations ranging from 11 to 19 mg/L. The coarser bone char and activated carbon media tests had slightly poorer solids removal rates (62 to 79% during the highest flow tests), but with much higher flow rates (46 to 50 gal/min). At flows similar to the mixed media (21 to 28 gal/min), these coarser materials provided similar removals (about 79 to 88% for suspended solids). The flow rates therefore seemed to be more important in determining particulate solids capture than the media type.

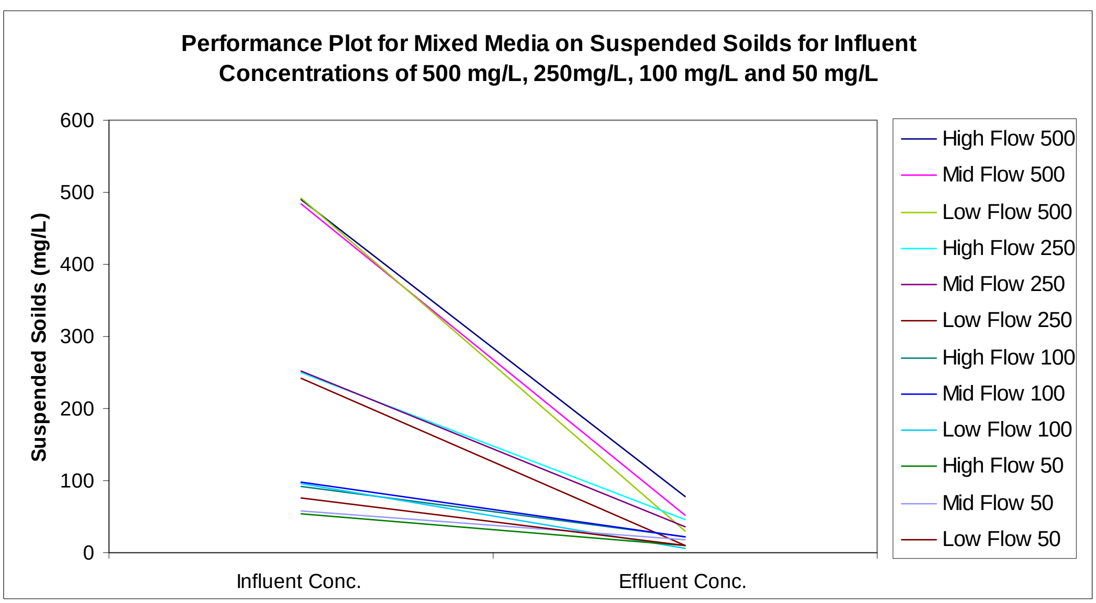


Figure 53. Performance plot for mixed media for suspended solids at influent concentrations of 500 mg/L, 250 mg/L, 100 mg/L and 50 mg/L.

Actual Storm Event Monitoring

Every storm evaluated had a hyetograph (rainfall pattern) and hydrograph (runoff pattern) prepared with the treatment flow capacity marked for that particular event. An example is shown in Figure 54.

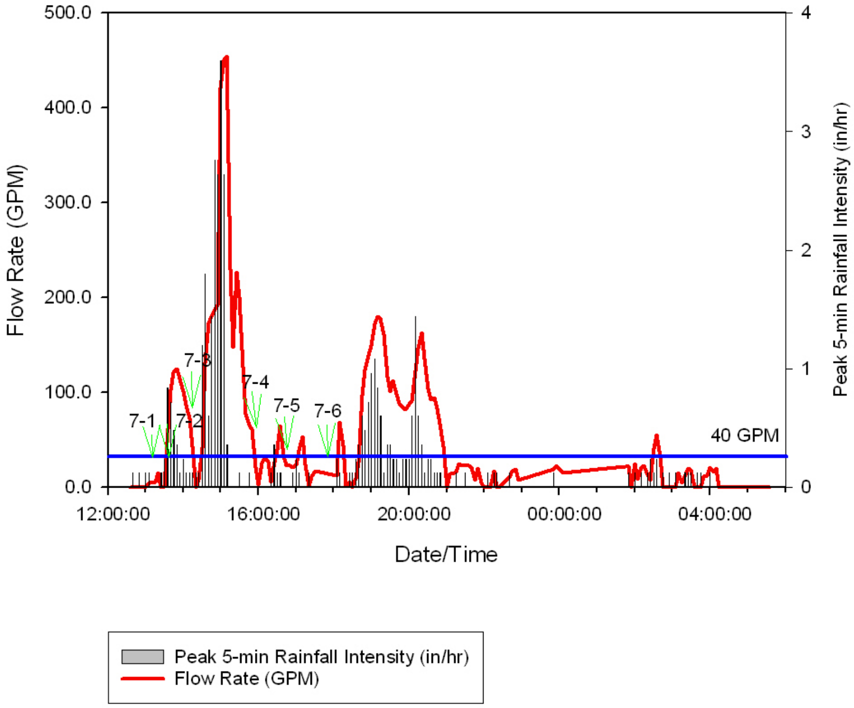


Figure 54. Hydrograph and hyetograph for Hurricane Katrina (August 29, 2005).

Thirty-one separate rains occurred during the 10 month monitoring period from February 2 to November 21, 2005. The monitoring period started off unusually dry in the late winter to early summer months. However, the mid summer was notable for severe thunderstorms having peak rain intensities (5-min) of up to 4 inches per hour. The late summer was also notable for several hurricanes, including Hurricane Katrina on August 29, 2005 that delivered about 3 inches of rain over a 15 hour period, having peak rain intensities as high as 1 in/hr in the Tuscaloosa area. During the monitoring period, the treatment flow rates were observed to decrease with time, as expected. Figure 55 relates the decreasing flow rate with rain depth. The filter was always greater than the specified 25 gpm treatment flow rate during the 10 month period. It is estimated that the 25 gpm treatment flow would be reached after about 30 inches of rainfall (in an area having 0.9 acre of impervious surfaces), or after about 45,000 ft^3 of runoff, or after about 160 lbs of suspended solids, was treated by the filter.

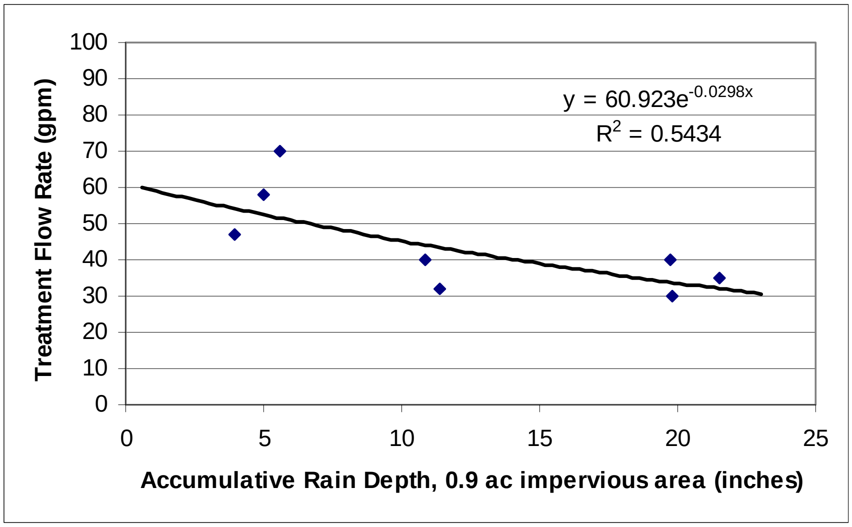


Figure 55. UpFlo™ filter treatment rate with rain depth.

These data indicate that the performance of the UpFlo™ filter is dependent on influent concentrations. As an example, the following figures show the analyses for suspended solids. Figure 56 is a scatterplot of the observed influent concentrations vs. the effluent concentrations, while Figure 57 is a line plot that connects paired influent and effluent concentrations. These plots show generally large reductions in TSS concentrations for most events.

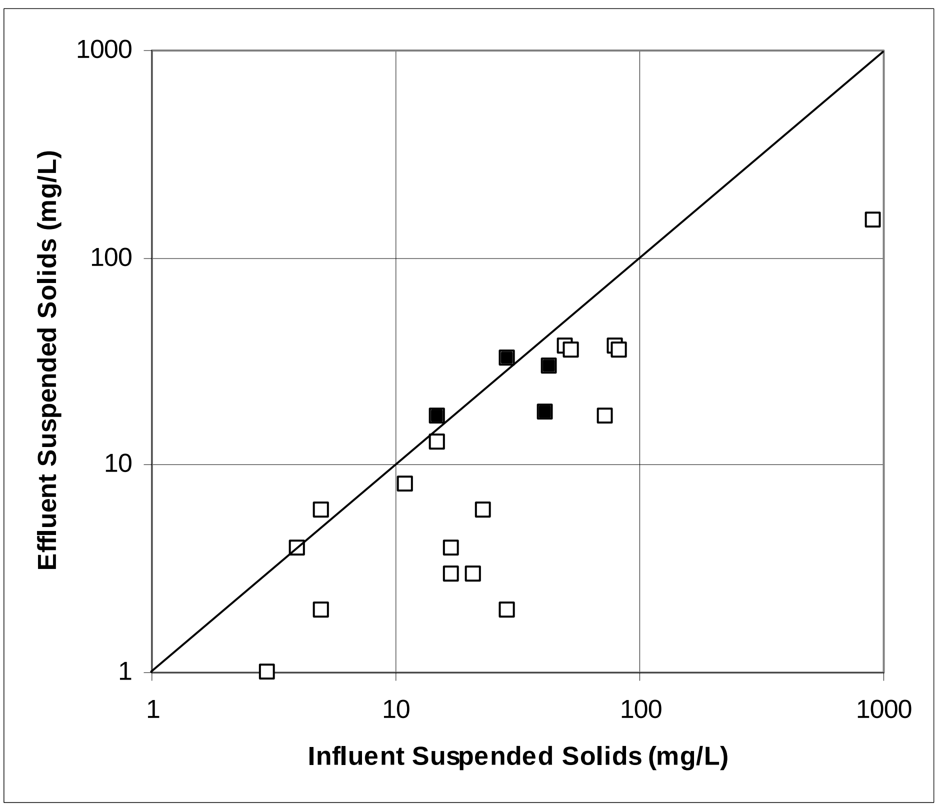


Figure 56. Scatterplot of observed influent and effluent suspended solids concentrations (filled symbols are events that had minor filter bypasses).

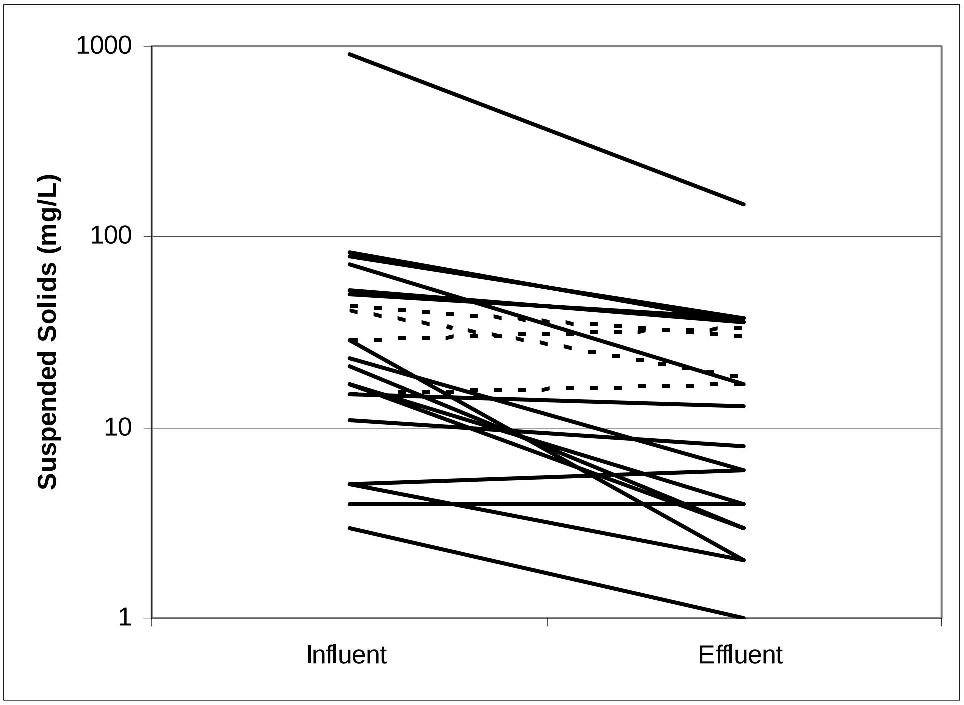


Figure 57. Paired influent and effluent suspended solids concentrations.

The nonparametric sign test was also used to calculate the probability that the influent equals the effluent concentrations. For the TSS data, P < 0.01, indicating with >99% confidence that the influent does not equal the effluent concentrations. Therefore, the test was statistically significant at least at the α 0.05 level.

These data were fitted to regression equations to predict the effluent concentrations from the influent conditions. In all cases, the data needed to be log-transformed in order to obtain proper residual behavior. For TSS, the following equation was found to be very significant, according to the ANOVA analyses:

Effluent Suspended Solids, log mg/L = 0.730 * (Influent Suspended Solids, log mg/L)

Regression Statistics on Observed Influent vs. Effluent Suspended Solids, log mg/L

Multiple R	0.94
R Square	0.89
Adjusted R Square	0.85
Standard Error	0.37
Observations	24

ANOVA

	df	SS	MS	F	Significance F
Regression	1	25.4	25.4	187	3.11E-12
Residual	23	3.12	0.136		
Total	24	28.55			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
X Variable 1*	0.730	0.053	13.7	1.56E-12	0.620	0.841

* the intercept term was determined to be not significant during the initial analyses and was therefore eliminated from the model and the regression and ANOVA reanalyzed.

As indicated on the ANOVA analyses above, the intercept term was not significant when included in the model, so that term was removed, and the statistical test conducted again. The overall significance of the model is very good ($F<<0.001$), and the adjusted R^2 term is 0.85. The P-value for the slope term of the equation is also highly significant ($P<<0.001$) and the 95% confidence limit of the calculated coefficient is relatively narrow (0.62 to 0.84). Figure 58 is a plot of the fitted equation along with the observed data, while Figure 59 contains the residual plots, all showing acceptable patterns.

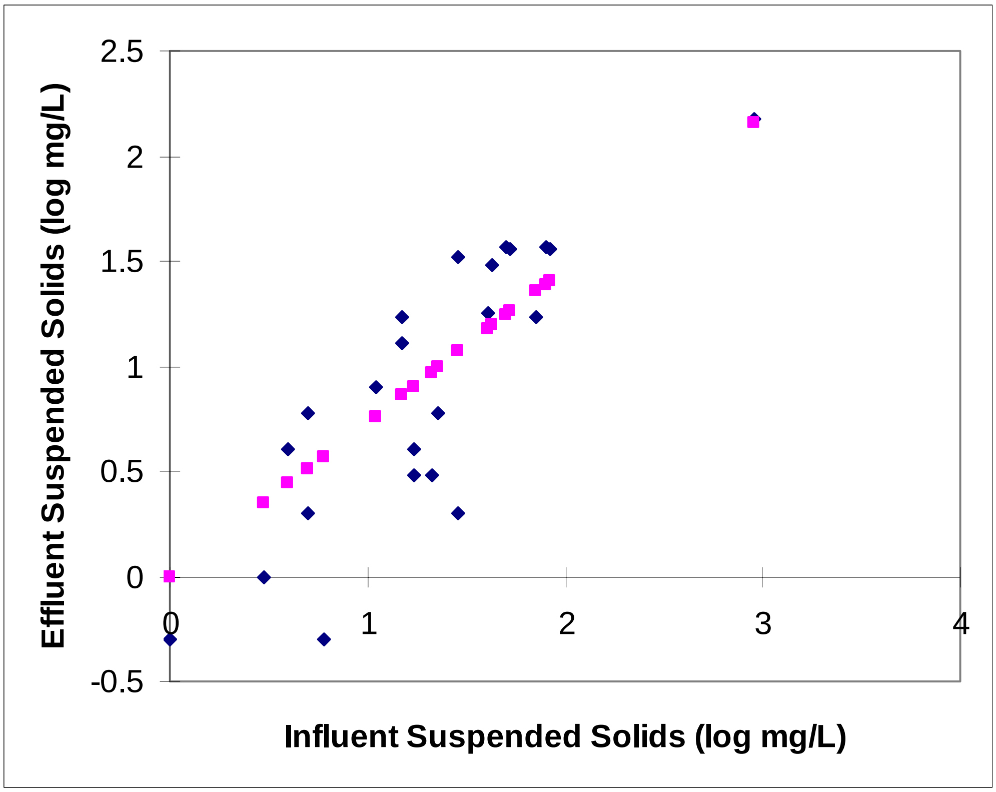


Figure 58. Fitted equation and data points for influent and effluent suspended solids.

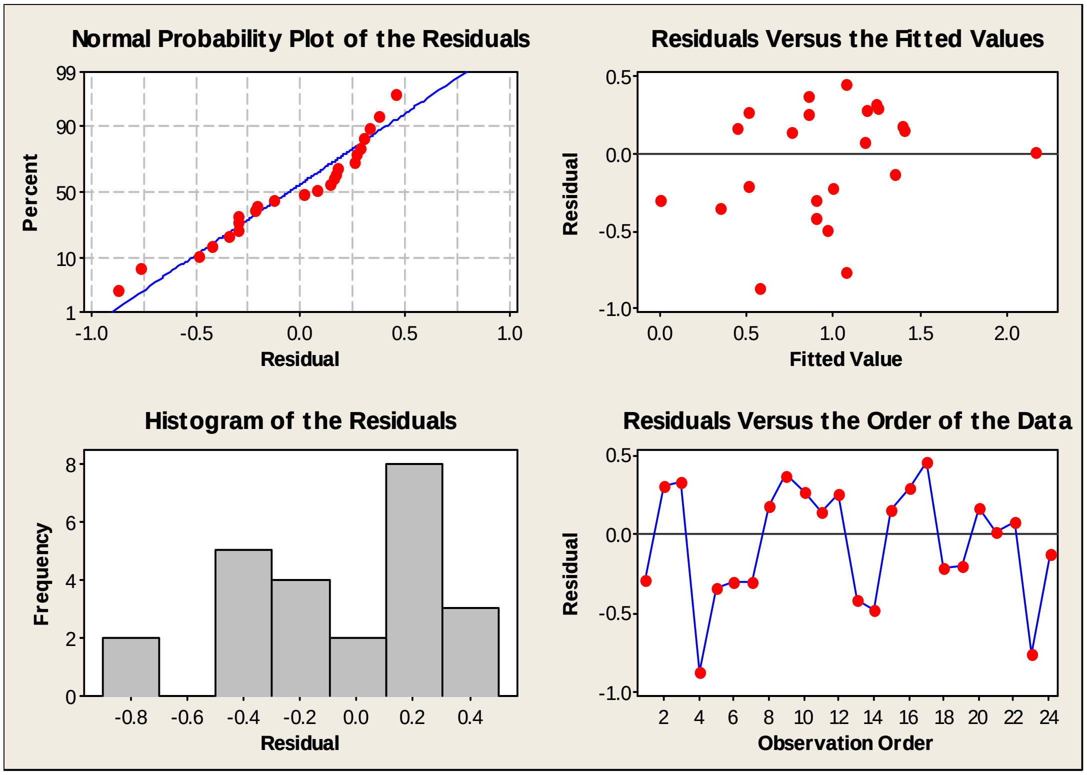


Figure 59. Residual analyses of fitted equation for suspended solids influent vs. effluent.

Confidence intervals of the influent vs. effluent plots are shown in Figure 60, while Figure 61 shows the confidence intervals for calculated percentage reduction values. As indicated in Figure 61, the TSS reductions would be >70% when influent concentrations exceeded about 80 mg/L, >80% when influent concentrations exceeded about 300 mg/L, and >90% when influent concentrations exceeded about 1000 mg/L.

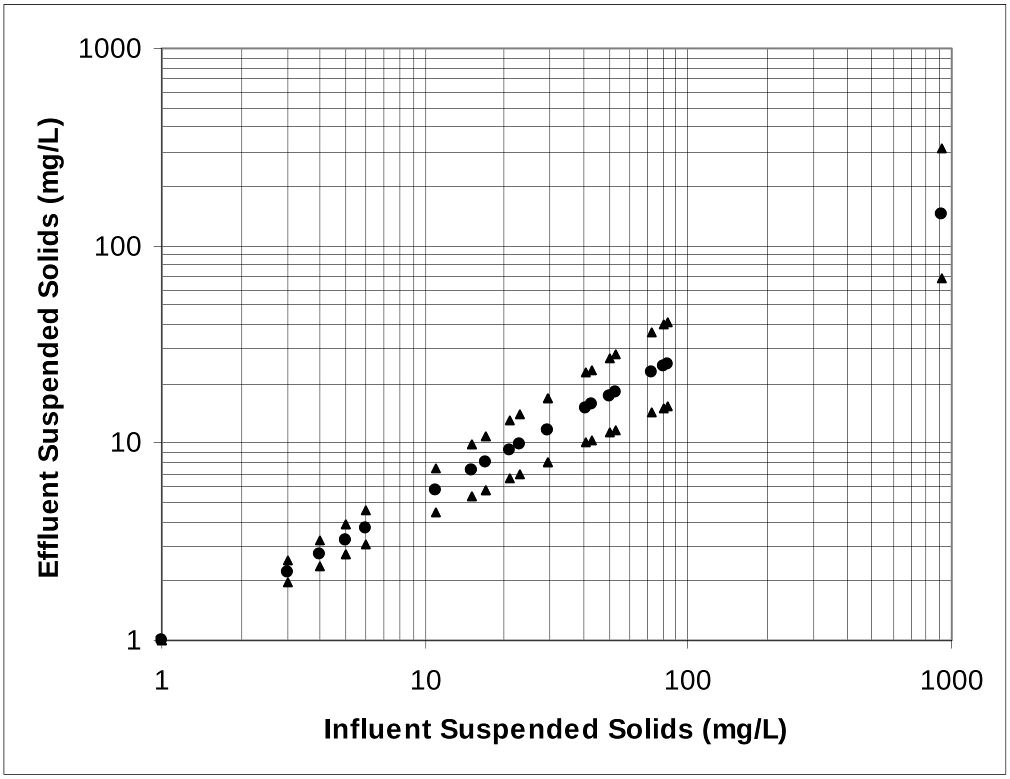


Figure 60. Predicted effluent concentrations for different influent concentrations, with 95% confidence limits.

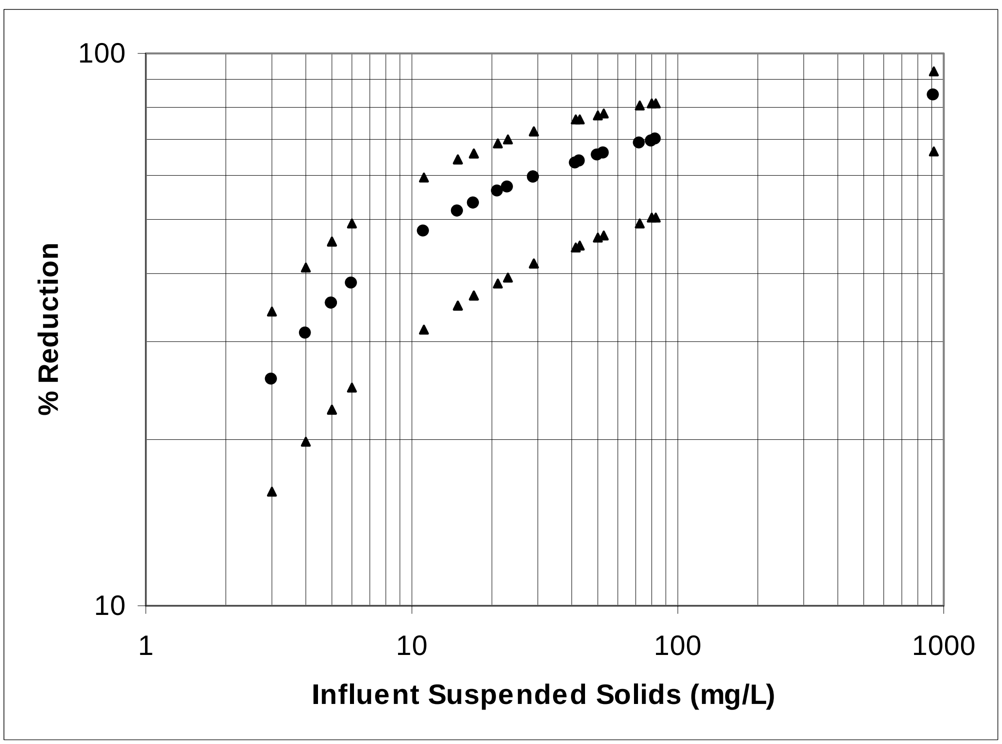


Figure 61. Percentage reductions as a function of influent concentrations, with 95% confidence limits.

Tables 22 summarizes the expected mass balance of particulate material removed by the UpFlow™ filter during the sampling period, considering both the measurements from the automatic samplers (for suspended material <150 µm in size) and the larger material retained in the sump, assuming all the runoff was treated by the filter, with no bypass, and all material greater than about 250 µm would be retained in the filter and sump. The suspended solids removal rate is expected to be about 80%, while the removal rates for the other monitored constituents are expected to be about 72 to 84%, depending on their associations with the different particle sizes.

Table 22. Calculated Mass Balance of Particulate Solids for Monitoring Period

particle size range (µm)	SS influent mass (kg)	SS effluent mass (kg)	SS removed (kg)	% reduction
0.45-3	9.3	2.8	6.6	70
3-12	18.7	6.4	12.3	66
12-30	22.4	7.7	14.7	66
30-60	26.7	6.8	19.9	74
60-120	4.6	1.8	2.9	61
120-250	19.8	4.3	15.5	78
250-425	11.5	0.0	11.5	100
425-850	17.1	0.0	17.1	100
850-2,000	10.5	0.0	10.5	100
2,000-4,750	4.8	0.0	4.8	100
>4,750	3.5	0.0	3.5	100
sum	148.9	29.8	119.2	80

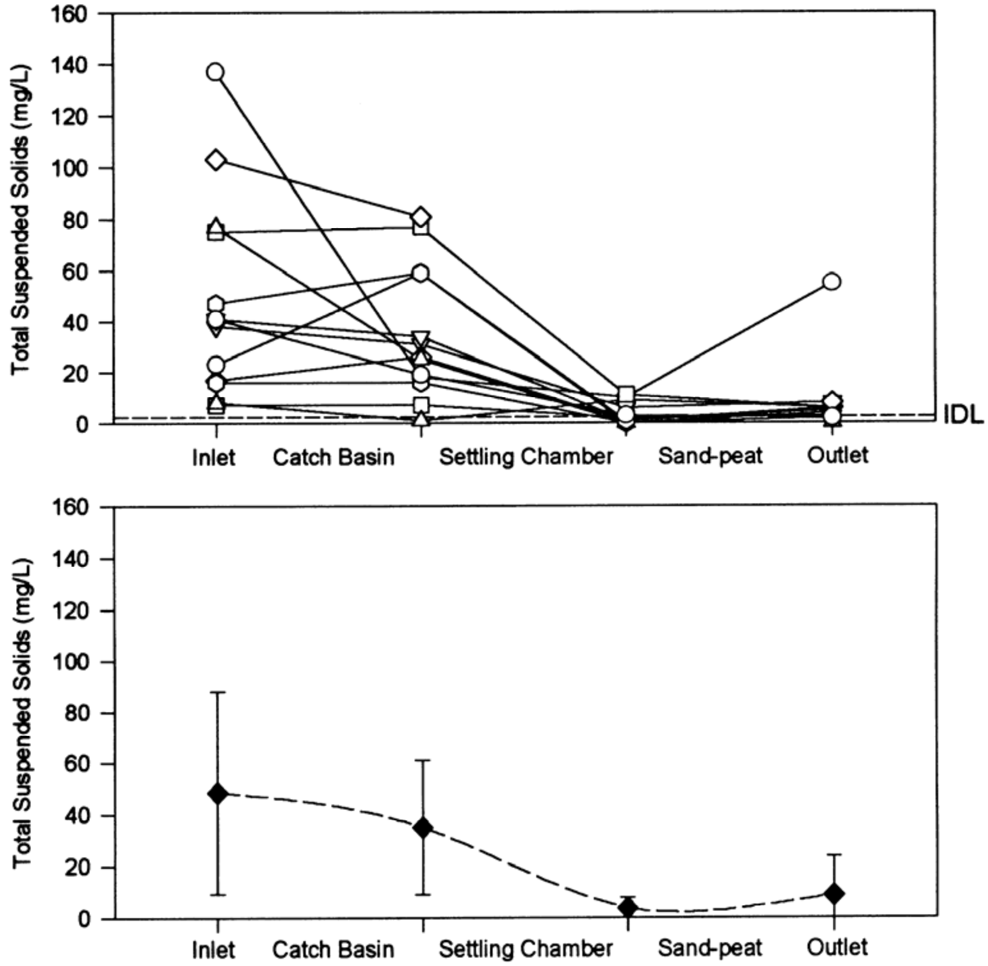
Other Exploratory Data Methods used to Evaluate Stormwater Controls

There are many other ways to present data from stormwater control practices. Several of these are shown in the following discussion.

Figure 62 is a plot showing the TSS concentrations of influent water and after several stages of treatment in the multi-chambered treatment train (MCTT) (Pitt, *et al.* 1999). Even though the influent quality was highly variable, the effluent was quite consistent. The first event, with a high effluent, was associated with rinsing fine media that hadn't been adequately cleaned. Table 24 is a listing of the TSS data for these MCTT tests (mg/L) for each of the 12 events. The following discussion outlines a simple analysis protocol that examined this data.

The first step in any analyses is to prepare several simple data plots. Figure 63 is a scatterplot of influent and effluent TSS observations. Except for the one high effluent observation, most of the effluent appears to be relatively constant and not affected by the influent conditions. If this was the case, a regression analysis with ANOVA would result in the slope term being insignificant and the intercept being significant. This would imply that there is no relationship between the influent and effluent TSS quality, and the effluent quality is constant for all conditions, a very favorable outcome. Figure 64 is the same plot, but with log transformations. In this case, there appears to be a positive trend between the influent and effluent, although slight. Figure 65 contains box and whisker plots of the influent and effluent TSS data, in actual and log space. Normal and log-normal probability plots of the influent and effluent MCTT TSS data are shown in Figure 66. These plots show reasonable parallel probability lines for the log-normal plot. Figure 67 shows a log-normal probability plot of the influent TSS data and Anderson-Darling test results, indicating a good fit (after the one large effluent data value was removed as that was an unusual observation associated with the first test and media that was not completely washed).

Figure 68 shows the data and the fitted regression line, with the 95% significance limits. The limits are very wide due to the few data observations (11 sets shown here). Table 23 shows the ANOVA results for the fitted regression line of this TSS MCTT data. This shows that the regression is not significant and that there is no significant relationship between the influent and effluent TSS observations. The effluent TSS can therefore best be described using a probability plot, as the little variability present cannot be adequately explained by the changing influent conditions. Far from being a problem with statistical analyses, this is the desired result from a control device: the effluent quality is consistent and not related to influent conditions. Of course, the excellent quality of the effluent is also very important!



	Catch Basin Chamber	Settling Chamber	Sand-peat Chamber	MCTT Overall
Concentration Difference 1-sided P Value	0.1543	0.0010	-0.1191	0.0002
Min. Percent Reduction	-157	-800	-500	25
Max. Percent Reduction	88	100	45	100
Median Percent Reduction	17	91	-400	83
Std. Dev. of Percent Reduction	65	257	240	22
COV of Percent Reduction	7.4	19	-1.5	0.28

Figure 62. Line plot and statistical summaries showing performance of MCTT for different treatment components (Pitt, *et al.* 1999).

Table 23. Total Suspended Solids Data for MCTT tests (mg/L) (Pitt, *et al.* 1999).

STORM	INLET	OUTLET
1	137	55
2	7	3
3	8	6
4	38	8
5	17	6
6	16	4
7	23	<2.5
8	75	6
9	77	<2.5
10	41	5
11	103	8
12	41	<2.5

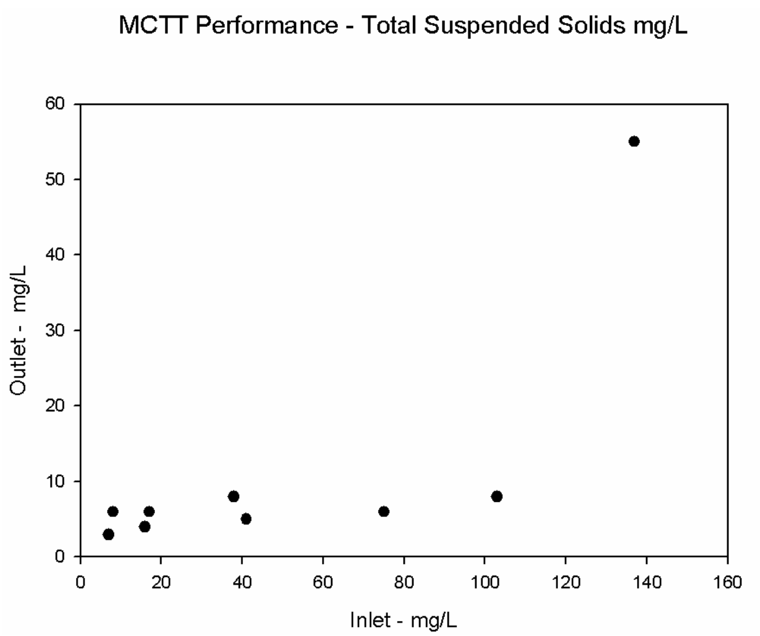


Figure 63. Plot of influent and effluent MCTT TSS data (Pitt, *et al.* 1999).

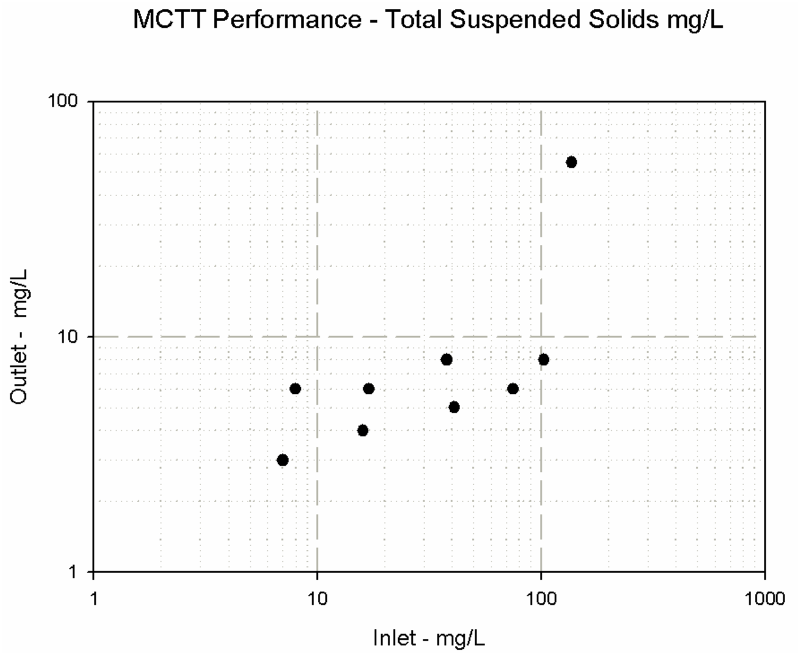


Figure 64. Plot of influent and effluent MCTT TSS data, log transformed data (Pitt, *et al.* 1999).

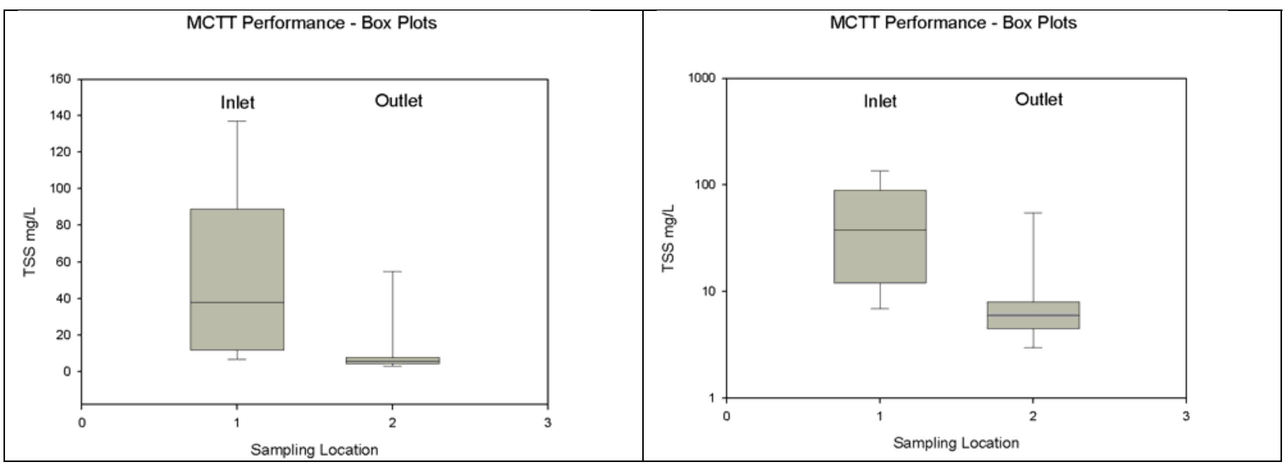


Figure 65. Box and whisker plots of influent and effluent TSS data for MCTT, in actual and log space.

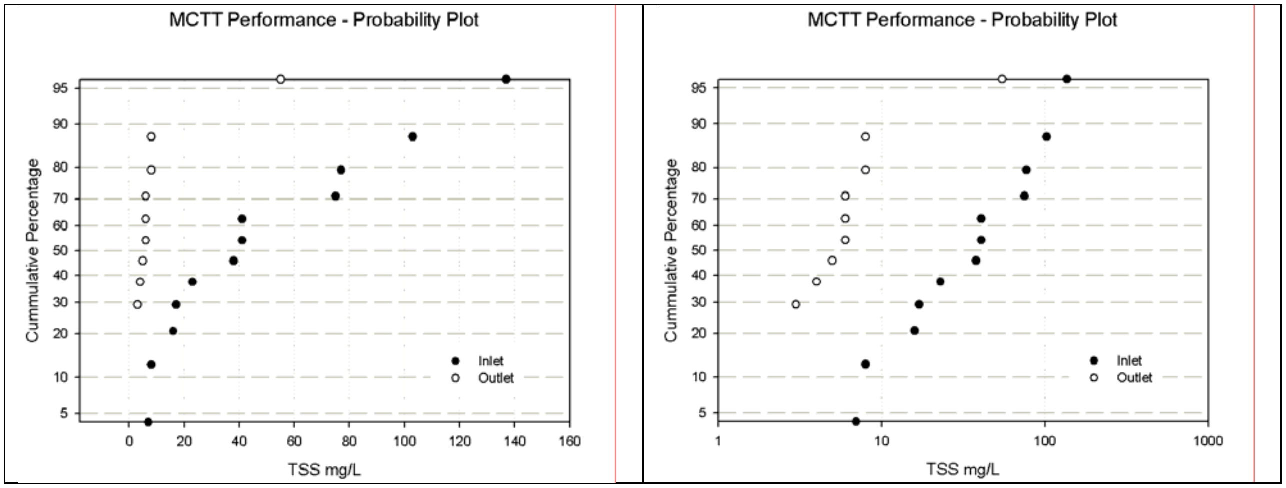


Figure 66. Normal and log-normal probability plots of influent and effluent MCTT TSS data.

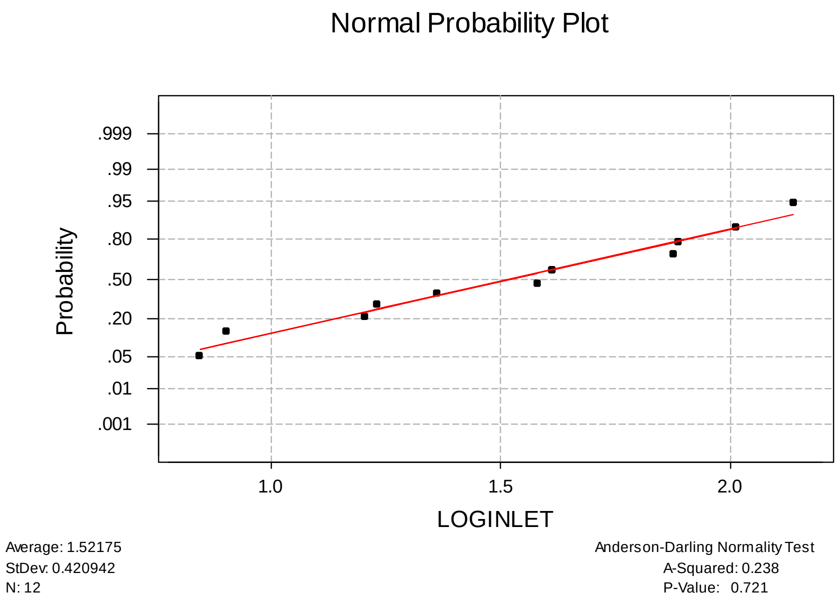


Figure 67. Log probability plot of influent TSS data and Anderson-Darling test results.

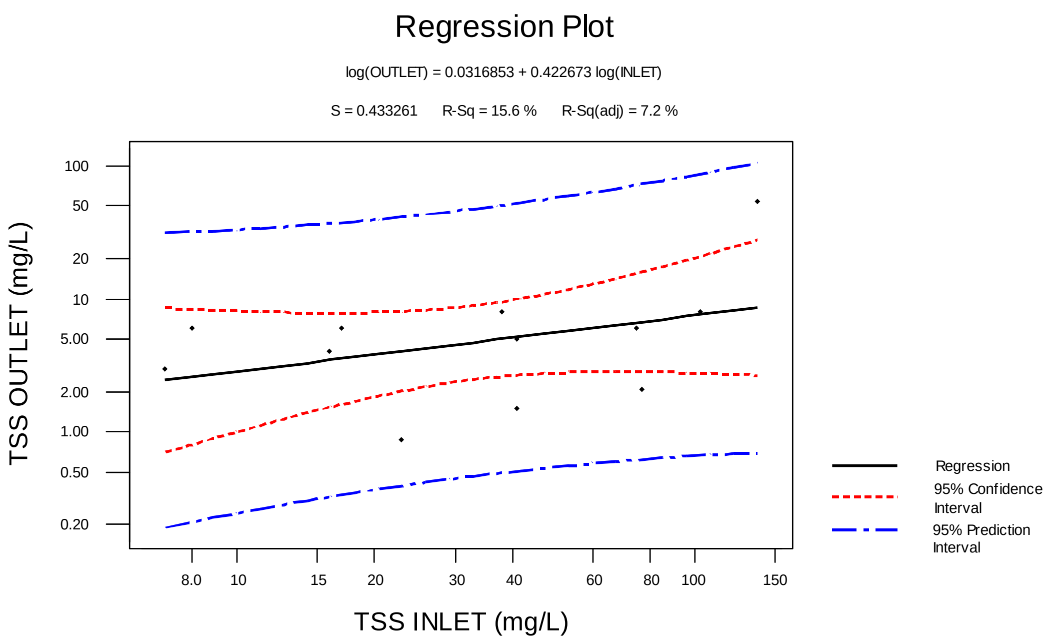


Figure 68. Data and regression line, with 95% significance limits.

Table 24. ANOVA Results for Regression Analysis of TSS MCTT Data.

Dependent variable is: **LOGOUTLET**
No Selector
R squared = 15.6% R squared (adjusted) = 7.2%
s = 0.4332 with 12 - 2 = 10 degrees of freedom

Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	0.347854	1	0.347854	1.85
Residual	1.87625	10	0.187625	

Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	0.0333252	0.4876	0.0683	0.9469
LOGINLET	0.421692	0.3097	1.36	0.2032

Figure 69 is a comparison of two alternative upflow treatment schemes, comparing the benefits of a suitable sump (Johnson, *et al*. 2003). The benefit of the sump was much more obvious for turbidity than for total solids, although it still provided a significant improvement for all constituents.

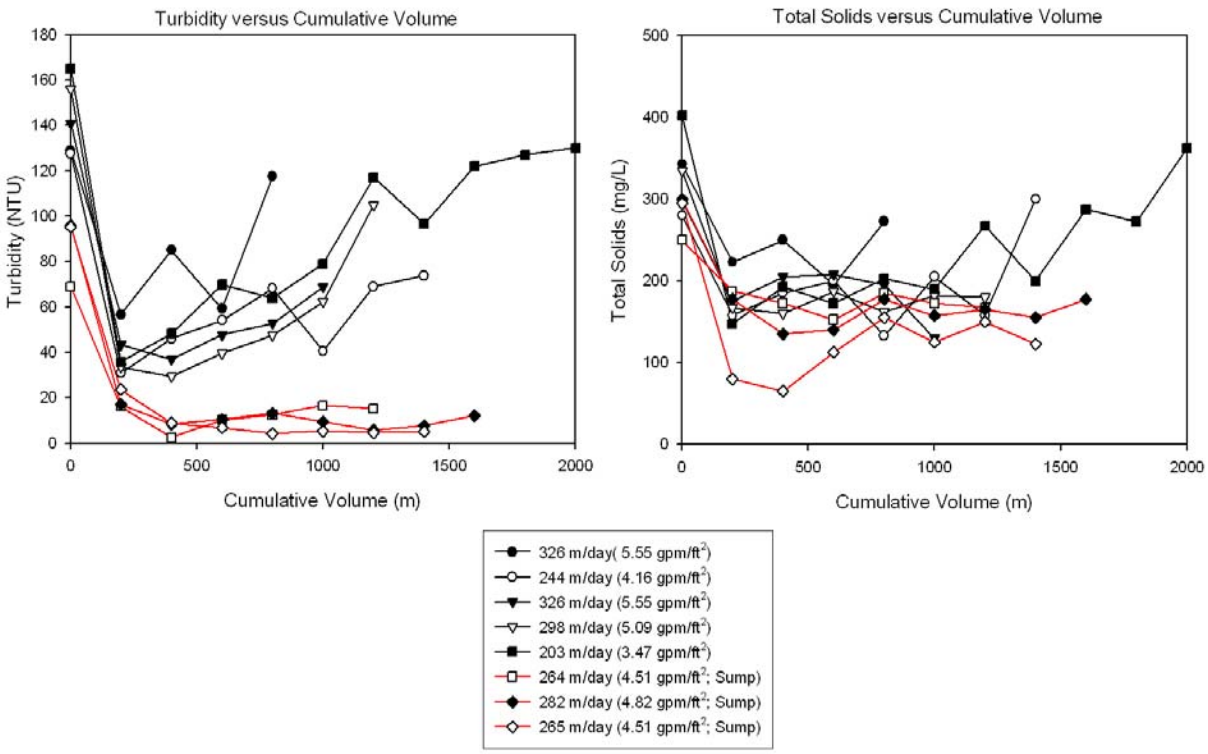


Figure 69. Comparisons of two alternative upflow treatment schemes (Johnson, *et al.* 2003).

Evaluation of Bacteria Decay Coefficients for Fate Analyses

A series of experiments were conducted to determine if sampling handling had a significant effect on measured microorganism values. Other tests were also conducted to identify and measure the fate mechanisms of there microorganisms. These example tests are summarized in the following discussion.

Fate Mechanisms for Microorganisms

Lake Tuscaloosa water samples containing total coliforms and *E. coli* were subjected to a series of simple laboratory tests to identify the effects of mixing and settling on the measured levels. Table 25 shows the results of the measured values for total coliforms over a several day period. One set of samples were rigorously mixed before 100mL was withdrawn for IDEXX total coliform analyses, while the other samples were left carefully undisturbed, and the 100mL of sample was pipetted without stirring the sample. There was an obvious downward trend in bacteria counts (#/100mL) with time for mixed and quiescent samples, but the reduction in values appeared to be greater for the quiescent sample set.

Table 25. Total Coliform Observations after Several Days

Time (day)	Quiescent (MPN)	Mixed (MPN)	Difference (MPN)
1	1413.6	1732.87	319.27
2	517.2	1299.65	782.45
3	727	727	0
5	116.2	691	574.8
6	54.6	517.2	462.6
7	12.2	410.6	398.4

The following analysis examined these differences to identify if they were significant. Figure 70 shows the probability plots for these two sets of data and Anderson-Darling test statistic (AD) indicates that they are not significantly different from normal probability plots (p values larger than 0.05, more samples would be needed to show that they are significantly different from a normal distribution). The standard deviations of both data sets are also similar. Figure 71 is a similar plot of the differences between the two data sets and also indicates a normal distribution.

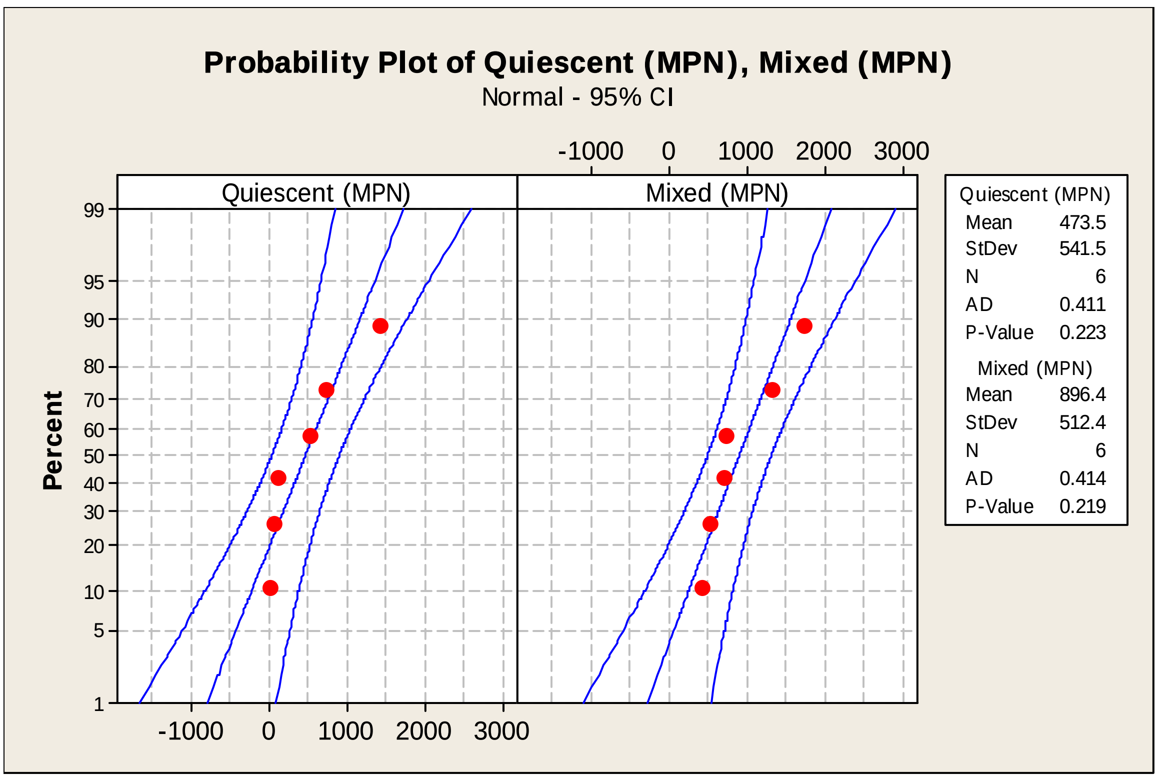


Figure 70. Probability plot for total coliforms (MPN) in quiescent and mixed samples.

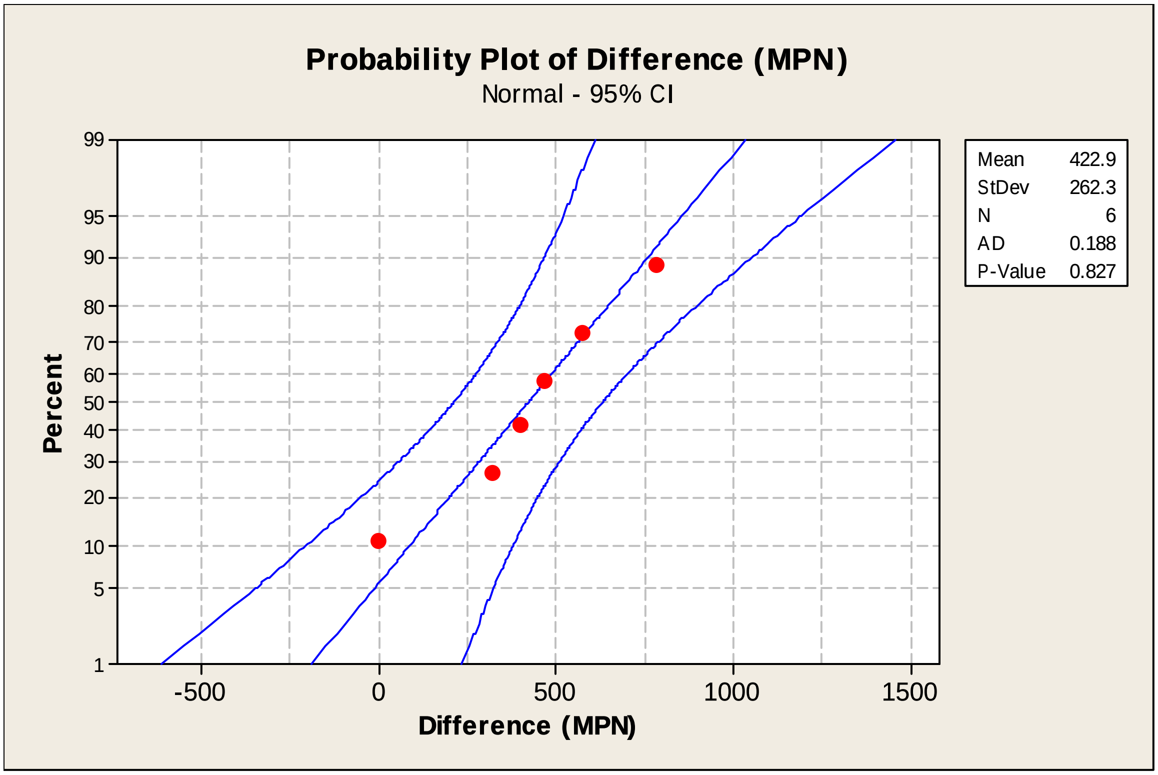


Figure 71. Probability plot of differences in total coliforms in mixed and quiescent samples (MPN).

Since these are paired samples and the difference between the mixed and quiescent samples is normal, it is possible to use the t-test:

Hypothesis: Let μ1 denote the mean MPN of total coliforms when the sample is in a mixed condition and let μ2 be the mean MPN of total coliforms when the sample is in a quiescent condition.

H_O : μ1= μ2
Ha : μ1> μ2

The test is performed at a significance level of 5% α = 0.05

The results of the paired t-Test (using Minitab) are:

Paired T-Test and CI: Mixed (MPN), Quiescent (MPN)

Paired T for Mixed (MPN) - Quiescent (MPN)

	N	Mean	StDev	SE Mean
Mixed (MPN)	6	896.387	512.440	209.203
Quiescent (MPN)	6	473.467	541.461	221.050
Difference	6	422.920	262.339	107.100

95% CI for mean difference: (147.612, 698.228)
T-Test of mean difference = 0 (vs not = 0): T-Value = 3.95 **P-Value = 0.011**

As the P-value is less than the specified significance level we can infer that at 5% significance level the data provides sufficient evidence to conclude that the mean MPN (#/100 mL) of total coliforms is greater in mixed samples than in quiescent samples.

A similar set of analyses was used to determine if mixing or quiescent settling had any effect on *E. Coli* (MPN) values (also measured using the IDEXX method of analyses). The following presents similar analyses as were shown above for total coliforms. Visually, although the E. coli values decrease significantly with time, the difference between the mixed and quiescent sample results are much smaller than for the total coliforms.

Table 26. *E. Coli* Values for Mixed and Quiescent Conditions

Time (Day)	Quiescent (MPN)	Mixed (MPN)	Difference (MPN)
0	52.8	46.5	-6.3
0.125	51.2	48.7	-2.5
0.25	37.9	45.7	7.8
0.5	35.9	37.3	1.4
1	32.7	27.8	-4.9
2	10.9	12.1	1.2
3	15.8	11.9	-3.9
5	3.1	2	-1.1

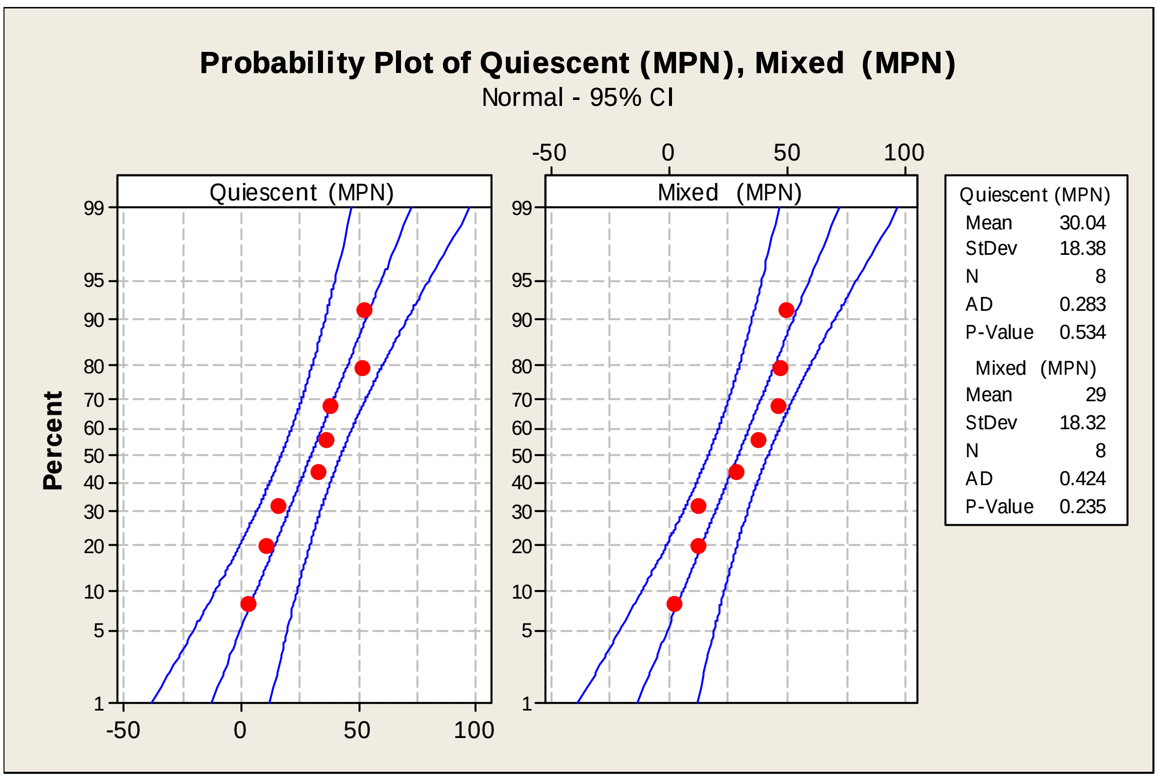


Figure 72. Probability plot for *E. Coli* (MPN) in quiescent and mixed samples.

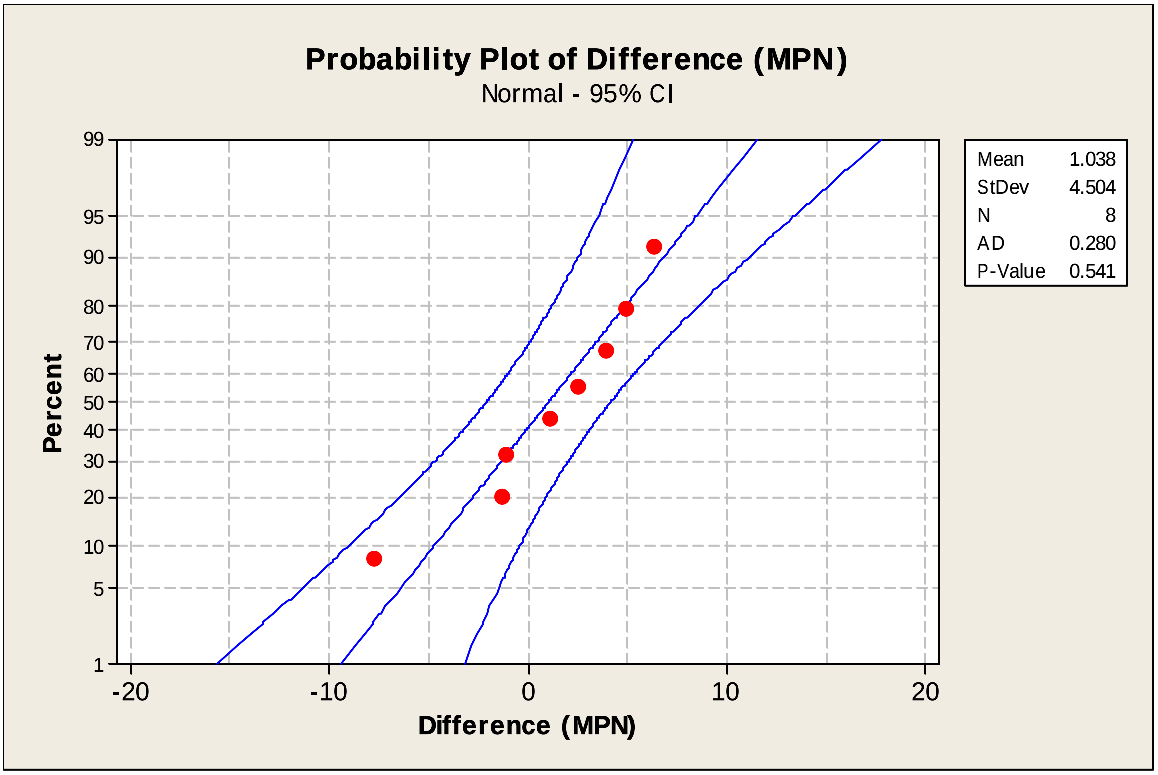


Figure 73. Probability plot of difference in mixed and quiescent sample *E. Coli* values (MPN).

Again, since these are paired samples and the difference between the mixed and quiescent samples is normal it is possible to use a paired t-test.

Hypothesis: Let µ1 denote the mean MPN of *E. Coli* when the sample is in mixed condition and let µ2 be the mean MPN of *E. Coli* when the sample is in quiescent system.

H_O : µ1= µ2
Ha : µ1> µ2

The test is performed at a significance level of 5% $\alpha = 0.05$

The results of the paired t-Test are

<u>Paired T-Test and CI: Mixed (MPN), Quiescent (MPN)</u>

Paired T for Mixed (MPN) - Quiescent (MPN)

	N	Mean	StDev	SE Mean
Mixed (MPN)	8	29.0000	18.3248	6.4788
Quiescent (MPN)	8	30.0375	18.3764	6.4970
Difference	8	-1.03750	4.50395	1.59239

95% CI for mean difference: (-4.80289, 2.72789)
T-Test of mean difference = 0 (vs not = 0): T-Value = -0.65 P-Value = 0.535
As the P-value is greater than the specified significance level, there are not enough samples to show that there is a significant difference between the two sample sets at the 0.05 level.

Decay Rate Curves of Lake Microorganisms

The above data allows calculations of the decay rates for the tested microorganisms to be directly calculated. Figures 74 through 76 are plots of the observed values for the different time periods. These can be used to determine the first order equation decay rates that are needed in bacteria fate modeling. Because of the difference in the decays from the mixed and quiescent samples, the effects of settling, separately from “dieoff” as a decay function can be quantified.

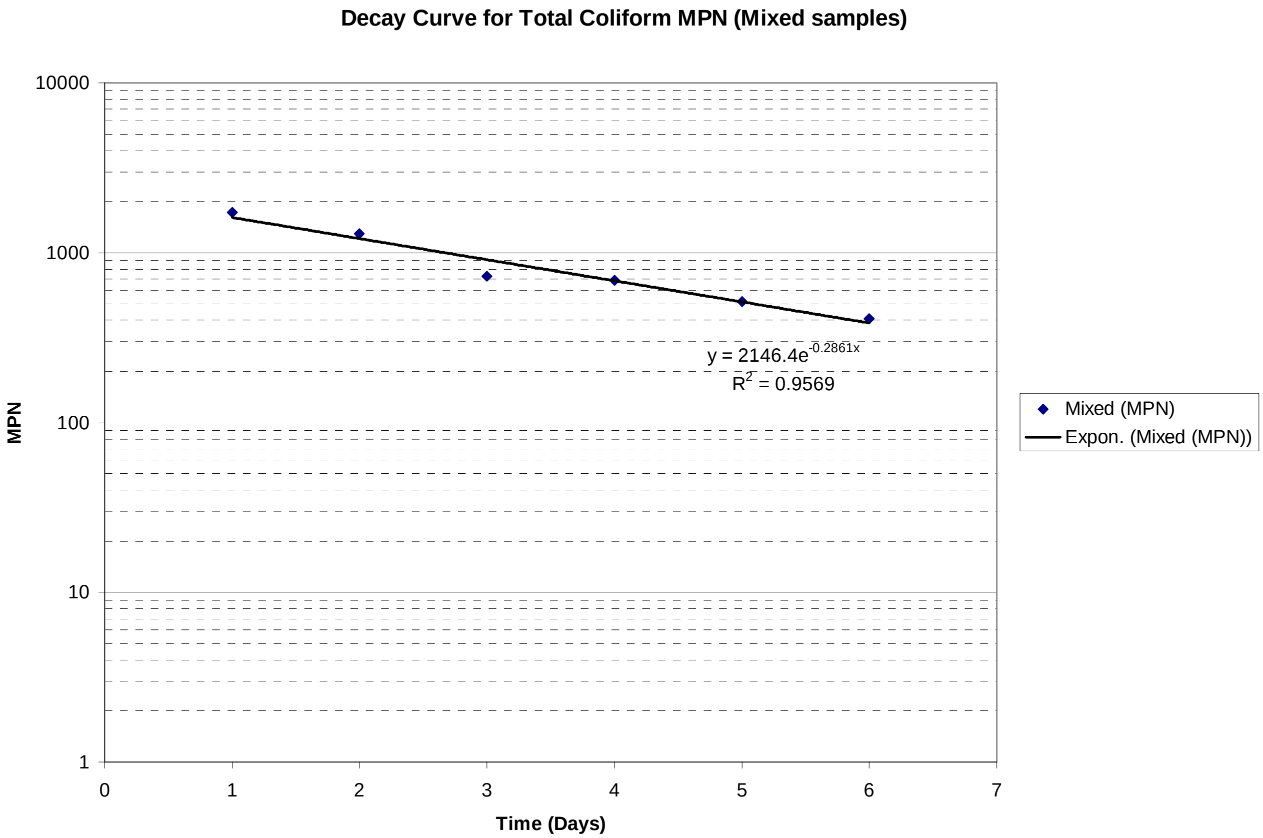


Figure 74. Decay rate for total coliforms in mixed samples.

From two points on the best fit line (900, 3day) and (500, 5day)

$$k = \frac{-\ln(S2/S1)}{(D2-D1)} \text{ per day}$$

$$k = \frac{-\ln(500/900)}{(5-3)} = 0.3 \text{ per day}$$

Therefore the decay rate for total coliforms in a mixed system for Lake Tuscaloosa is 0.3 per day, similar to the reported values in the literature.

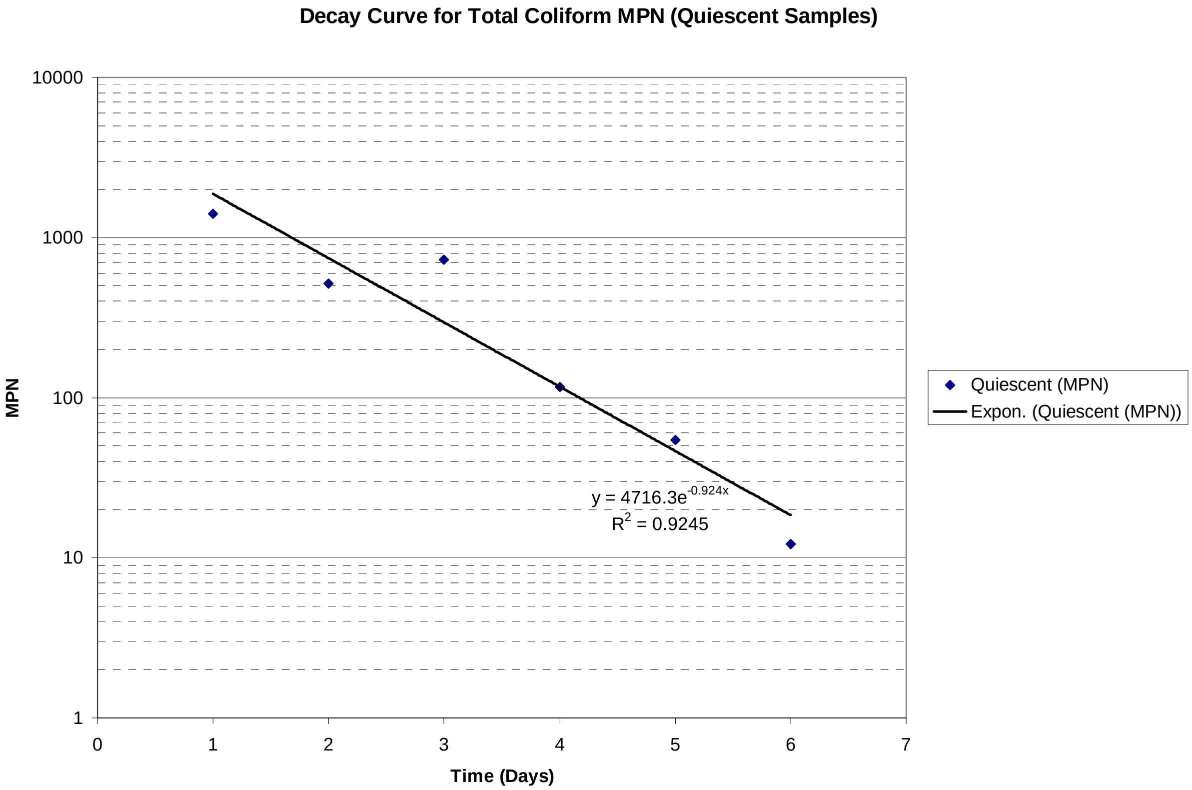


Figure 75. Decay rate for total coliforms in quiescent samples.

From two points on the best fit line (800, 2day) and (48, 5day)

$$k = \frac{-\ln(S2/S1)}{(D2-D1)} \text{ per day}$$

$$k = \frac{-\ln(48/800)}{(5-2)} = 0.93 \text{ per day}$$

Therefore the decay rate for total coliforms in a quiescent system for Lake Tuscaloosa is 0.93 per day, substantially greater than usually reported. The difference between these decay rates (0.63/day) can be attributed to gravitational settling, while the mixed decay rate (0.3/day) can be attributed to dieoff. Using the settling component, much more accurate fate predictions can be made concerning coliform bacteria in Lake Tuscaloosa.

The following plots are for *E. coli* decay rate calculations. Since there was no significant difference in the quiescent and mixed sample, settling was an important fate mechanism for *E. Coli*. and the total loss can be attributed to dieoff.

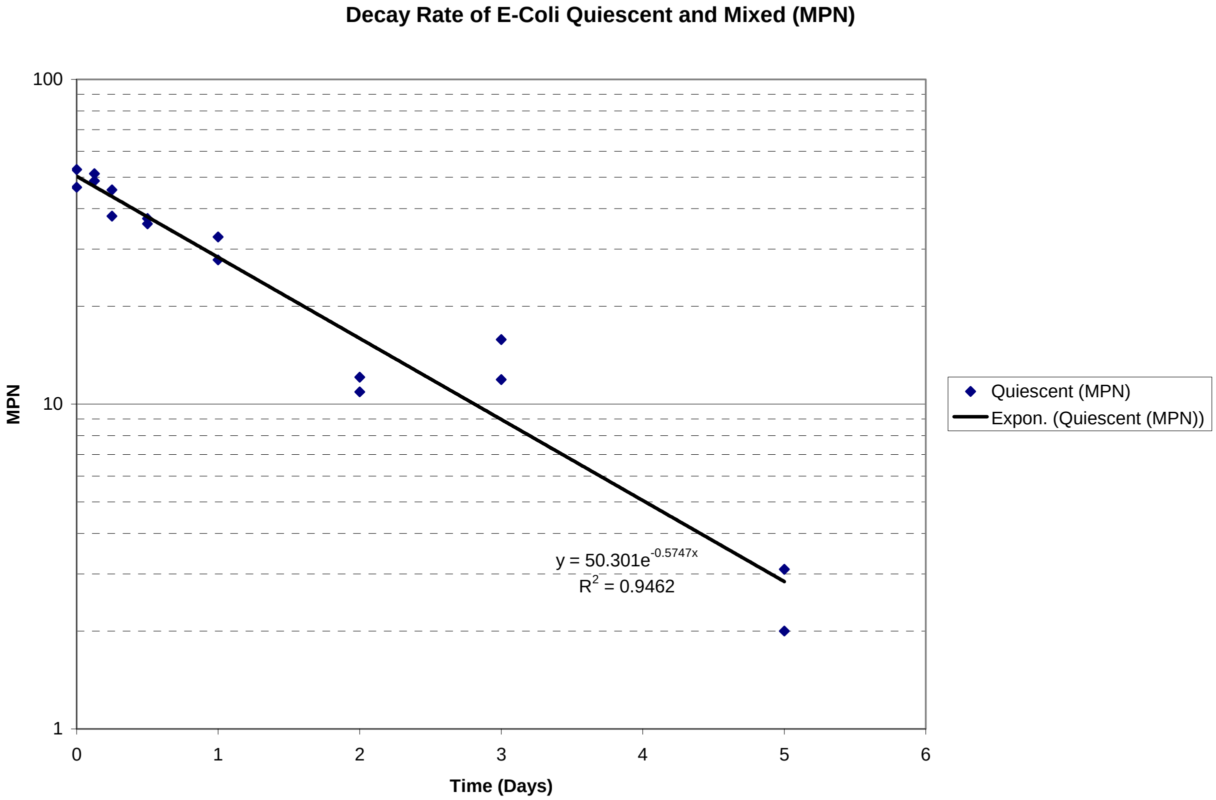


Figure 76. Decay rate for *E. coli* in mixed and quiescent samples.

From two points on the best fit line (900, 3day) and (500, 5day)

$$k = \frac{-\ln(S2/S1)}{(D2-D1)} \text{ per day}$$

$$k = \frac{-\ln(9/29)}{(3-1)} = 0.58 \text{ per day}$$

Therefore, the total decay rate for *E. Coli* in Lake Tuscaloosa is 0.58 per day, with very little attributed to gravitational settling.

References

Berthouex, P.M. and L.C. Brown. *Statistics for Environmental Engineers.* Lewis Publishers, Boca Raton, FL, 1994.

Box, G.E.P., W.G. Hunter, and J.S. Hunter. *Statistics for Experimenters*. John Wiley and Sons. New York. 1978.

Burton, G.A. Jr., and R. Pitt. *Stormwater Effects Handbook: A Tool Box for Watershed Managers, Scientists, and Engineers*. CRC Press, Inc., Boca Raton, FL. August 2001. 911 pgs.

Center for Watershed Protection and R. Pitt. *Illicit Discharge Detection and Elimination; A Guidance Manual for Program Development and Technical Assessments*. U.S. Environmental Protection Agency, Office of Water and Wastewater. EPA Cooperative Agreement X-82907801-0. Washington, D.C., 357 pgs. Oct. 2004.

Cleveland, W.S. *Visualizing Data*. Hobart Press, Summit, NJ. 1993.

Cleveland, W.S. *The Elements of Graphing Data*. Hobart Press, Summit, NJ. 1994.

Enell, M. and J. Henriksson-Fejes. *Dagvattenreningsverket vid Rönningesjön, Täby Kommun. Undersokningsresultat* (Investigation Results of Water Purification Works near Täby Municipality), in Swedish. Institutet for Vatten - och Luftvardsforskning (IVL). Stockholm, Sweden. 1989-1992.

EPA (U.S. Environmental Protection Agency). *Results of the Nationwide Urban Runoff Program*. Water Planning Division, PB 84-185552, Washington, D.C., December 1983.

FLUENT 6.2. Computational Fluid Dynamic (CFD) Software. User's Manual. http://www.fluent.com/

Gilbert, R. 0., *Statistical Methods for Environmental Pollution Monitoring*. New York, NY: Van Nostrand Reinhold, 1987.

Hunter, J.S. "The Digidot Plot." *American Statistician*. Vol. 42, No. 54. 1988.

Hwang,H.M., Foster,G.D. "Characterization of polycyclic aromatic hydrocarbons in urban stormwater runoff flowing into the tidal Anacostia River." *Environmental Pollution* 140, 416-426. 2005.

Kittegoda, N, Rosso, R. Statistics, Probability, and Reliability for Civil and Environmental Engineers, McGraw-Hill, 1997

Lehman, E.L. and H.J.M. D'Abrera. *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day and McGraw-Hill. 1975.

Lundkvist, S. and H. Söderlund. "Rönningesjöns Tillfrisknande. Resultat Efter Dag-och Sjövattenbehandling Åren 1981-1987" (Recovery of the Lake Rönningesjön in Täby, Sweden. Results of Storm and Lake Water Treatment over the Years 1981-1987), in Swedish. *Vatten*, Vol. 44, No. 4. 1988. pp. 305-312.

Mackay, Donald, Wan-Ying Shiu., Kuo-Ching Ma. *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals*. Volume II, Lewis Publishers. 1992.

Maestre, A. and R. Pitt. *The National Stormwater Quality Database, Version 1.1, A Compilation and Analysis of NPDES Stormwater Monitoring Information*. U.S. EPA, Office of Water, Washington, D.C. (final draft report) August 2005.

Mahler,B.J., Van Metre,P.C., Bashara,T.J., Johns,D.A., "Parking lot sealcoat: An unrecognized source of urban polycyclic aromatic hydrocarbons." *Environmental Science and Technology* 39, 5560-5566, 2005.

Navidi, W, *Statistics for Engineers and Scientists*, 1st Edition, McGraw-Hill, 2006,

Peter C. Van Metre, Barbara J. Mahler, Edward T. Furlong. "Urban Sprawl Leaves Its PAH Signature." *Environmental Science & Technology* 34, 4064-4070, 2000.

Pitt, R. *Characterizing and Controlling Urban Runoff through Street and Sewerage Cleaning*. U.S. Environmental Protection Agency, Storm and Combined Sewer Program, Risk Reduction Engineering Laboratory. EPA/600/S2-85/038. PB 85-186500. Cincinnati, Ohio. 467 pgs. June 1985.

Pitt, R., and J. McLean. *Toronto Area Watershed Management Strategy Study: Humber River Pilot Watershed Project*. Ontario Ministry of the Environment, Toronto, Ontario. 486 pgs. 1986.

Pitt, R. *Small Storm Urban Flow and Particulate Washoff Contributions to Outfall Discharges*. Ph.D. dissertation. Department of Civil and Environmental Engineering, University of Wisconsin, Madison. 1987.

Pitt, R. "Water quality trends from stormwater controls." In: *Stormwater NPDES Related Monitoring Needs* (Edited by H.C. Torno). Engineering Foundation and ASCE. pp. 413-434. 1995.

Pitt, R. and S. Clark, *Communication Manhole Water Study: Characteristics of Water Found in Communications Manholes*, prepared for Bellcore (Telcordia) and the Office of Water and Wastewater, EPA, July 1999.

Pitt,R., Roberson,B., Barron,P., Ayyoubi,A., and Clark,S. "*Stormwater treatment at critical areas: The multi-chambered treatment train (MCTT).* " U.S. Environmental Protection Agency, Water Supply and Water Resource Division. National Risk Management Research Laboratory. EPA 600/R-99/017.Cincinnati,OH. 1999.

Pitt, R., M. Lilburn, S. Nix, S.R. Durrans, S. Burian, J. Voorhees, and J. Martinson, *Guidance Manual for Integrated Wet Weather Flow (WWF) Collection and Treatment Systems for Newly Urbanized Areas (New WWF Systems),* EPA, 2001.

Reckhow, K.H. and C. Stow. "Monitoring Design and Data Analysis for Trend Detection." *Lake and Reservoir Management*. Vol. 6, No. 1. 1990. pp. 49-60.

Reckhow, K.H., K. Kepford, and W. Warren-Hicks. *Methods for the Analysis of Lake Water Quality Trends* School of the Environment, Duke University. Prepared for the U.S. Environmental Protection Agency. October 1992.

Salau, J.S., R. Tauler, J.M. Bayona and I. Tolosa. "Input characterization of sedimentary organic contaminants and molecular markers in the Northwestern Mediterranean Sea b exploratory data analysis." *Environmental Science & Technology*. Vol. 31, no. 12, pg. 3482. 1997.

Schmidt, S, Launsby, R. *Understanding Industrial Designed Experiments*, 4th Edition, Air Academy Press,1997.

SCS (now NRCS) (U.S. Soil Conservation Service). *Urban Hydrology for Small Watersheds*. Tech. Release No. 55, U.S. Dept. of Agriculture, June 1986.

Söderlund, H. "Dag-och Sjövattenbehandling med Utjämning i Flytbassänger Samt Kemisk Fällning med Tvärlamellsedimentering" (Treatment of Storm- and Lakewater with Compensation in Floating Basins and Chemical Precipitation with Crossflow Lamella Clarifier.), in Swedish. *Vatten*, Vol. 37, No. 2. 1981. pp. 166-175.

Sokal, R.R and F.J. Rohlf. *Biometry: The Principles and Practice of Statistics in Biological Research*. W.H. Freeman and Co. New York. 1969.

Spooner, J. and D.E. Line. "Effective Monitoring Strategies for Demonstrating Water Quality Changes from Nonpoint Source Controls on a Watershed Scale." *Water Science and Technology*. Vol. 28, No. 3-5. 1993. pp. 143-148.

Thomann, R.V. and J.A. Mueller. *Principles of Surface Water Quality Modeling and Control*. Harper and Row. New York. 1987.

Tufte, Edward R. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut 06410. 1983.

Tufte, Edward R. *Envisioning Information*. Graphics Press, Cheshire, Connecticut 06410. 1990.

Tukey, John W. *Exploratory Data Analysis*. Addison-Wesley Publishing Co. 1977.

Appendix A: Factorial Analyses Examples

Examples of an Experimental Design using Factorial Analyses: Sediment Scour

Introduction

This detailed example of using factorial analyses to design an experiment was prepared by Humberto Avila, a Ph.D. student in Water Resources Engineering at the University of Alabama.

Accumulation of sediment and potential subsequent scour is one of the sediment transport processes in a stormwater drainage system. Sediment can be captured in inlets and manholes during rainfall events. The accumulation rate, or sediment-retaining performance, depends on the size and geometry of the device, the flow rate, sediment size, and specific gravity of the sediment. In the same way, scour phenomenon includes all those parameters previously mentioned in addition to the water protection layer and the consolidation of the sediment bed due to the aging phenomenon. Once the runoff ceases, sediment consolidates in the settling chamber and two different phases of the new sediments are formed in the manhole: a new sediment layer on the top of the previously captured sediment and a water layer above the sediment to the elevation of the outlet. This scenario corresponds to the initial condition of the scour analysis, which is the subject of this experiment.

The purpose of this experiment is to evaluate the importance of the parameters and their interactions on the phenomenon of scour or migration of sediment out of a conventional inlet catchbasin, an experimental design was performed and analyzed with 4 parameters which are flow rate, sediment size, water protection depth, and specific gravity. Each factor was evaluated at 2 levels.

A 2-dimensional Computational Fluid Dynamic (CFD) model was implemented in Fluent 6.2, using the Eulerian multiphase model, with which is possible to include two phases: an upper layer of water and a submerged dense layer of sediment. The evaluation consists in determining the reduction of sediment mass from the chamber over the time under the effect of a submersible vertical water jet.

Parameters

Four (4) parameters were evaluated in this experiment: flow rate, sediment size, water protection depth, and specific gravity. Each factor was evaluated at 2 levels: flow rates at 1.6 L/s and 20.8 L/s (25 and 267 GPM), sediment diameter sizes at 50 μm and 500 μm, water protection depths at 0.2 m and 1.0 m above the sediment, and specific gravities at 1.5 and 2.5.

Model and Response

A 2-dimensional Computational Fluid Dynamic (CFD) model was implemented in Fluent 6.2 by using the Eulerian multiphase model, with which it is possible to include two phases: and upper layer of water and a submerged dense layer of sediment. The evaluation consists in determining the reduction of sediment mass into a chamber through time under the effect of a submersible vertical water jet. Figure A1 shows the general configuration of the 2-D CFD model implemented for this experiment. The figure shows the location of the inlet and outlet.

Normally, the response or responses are selected before performing an experiment. In this case, the loss of sediment through the time was selected as the measurable response. However, after a preliminary analysis of the results, an the necessity of having only one value for the response, the loss of sediment after 1,000 sec of continuous flow (16 min) was selected as the final response to be evaluated in the experimental analysis. Figure A2 shows the reduction of sediment mass over the time in the case ABC (flow rate at high, depth of water at high, diameter at high, and specific gravity at low).

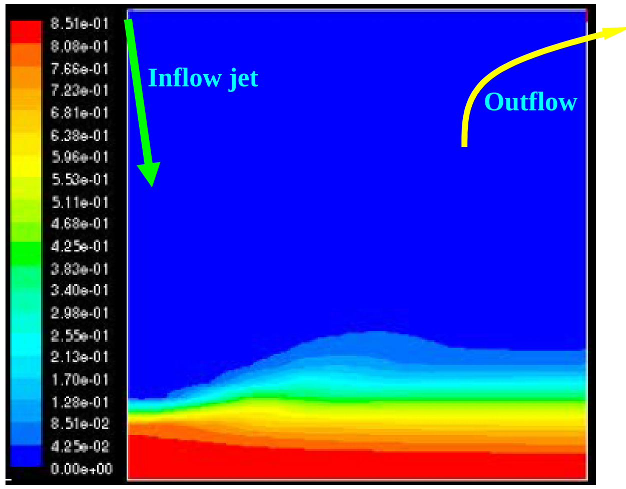


Figure A1. General representation of a simulation. Inflow, and outflow directions are indicated by arrows. Upper layer of water in blue, and sediment layer in color scale.

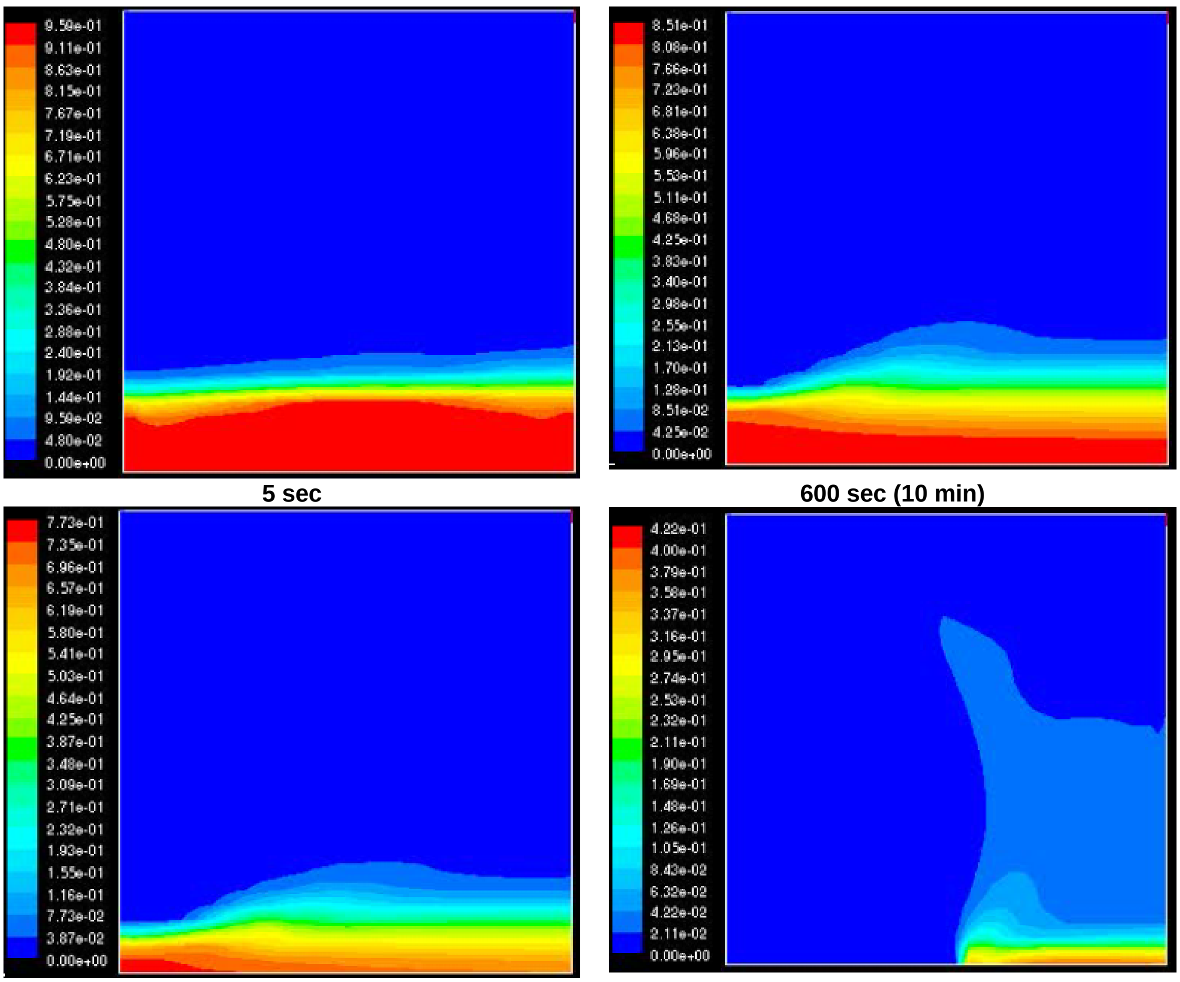


Figure A2. Simulation of case ABC (flow rate at high, depth of water at high, diameter at high, and specific gravity at low) – Colors represent Volume Fraction of Sediment.

Experimental Design
Considering that 4 factors at 2 levels each will be evaluated, a 2-level full factorial analysis for 4 factors is required; this is a 24 factorial analysis. The total number of runs is 24 = 16 runs, considering all four single factors and their interactions. Table A1 shows the experimental set up for a 24 factorial analysis.

Treatment	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
I	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+
a	+	-	-	-	-	-	-	+	+	+	+	+	+	-	-
b	-	+	-	-	-	+	+	-	-	+	+	+	-	+	-
ab	+	+	-	-	+	-	-	-	-	+	-	-	+	+	+
c	-	-	+	-	+	-	+	-	+	-	+	-	+	+	-
ac	+	-	+	-	-	+	-	-	+	-	-	+	-	+	+
bc	-	+	+	-	-	-	+	+	-	-	-	+	+	-	+
abc	+	+	-	-	+	-	-	-	-	+	-	-	+	+	+
d	-	-	-	+	+	+	-	+	-	-	-	+	+	+	-
ad	+	-	-	+	-	-	+	+	-	-	+	-	-	+	+
bd	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+
abd	+	+	-	+	+	-	+	-	+	-	-	+	-	-	-
cd	-	-	+	+	+	-	-	-	-	+	+	+	-	-	+
acd	+	-	+	+	-	+	+	-	-	+	-	-	+	-	-
bcd	-	+	+	+	-	-	-	+	+	+	-	-	-	+	-
abcd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table A1. Coded design matrix for a full factorial of 4 factors each at 2 levels (2^4 design)

A minimum of 3 replicates are required for this experiment to provide a 95% confidence in $\widehat{s}$ and 99.99% confidence in $\widehat{y}$. However, considering that the experiment was performed by using a computational model, only one replicate was performed, which provides about 95% confidence in $\widehat{y}$ and requires a residual analysis to evaluate which factors affect the variances $\widehat{s}$.

Results
After simulating all 16 scenarios for 3,600 sec, the reduction of sediment depth (sediment loss) was plotted as a function of the time. As previously mentioned, the analyzed response was the loss of sediment at 1,000 sec of continuous flow. The sediment depth is the inverse of the water protection depth; then if the water depth is 0.2 m, the sediment depth is 1.0 m; and if the water depth is 1.0 m, then the sediment depth is 0.2 m. Figure A3 shows the results obtained from the 2D-CFD model.

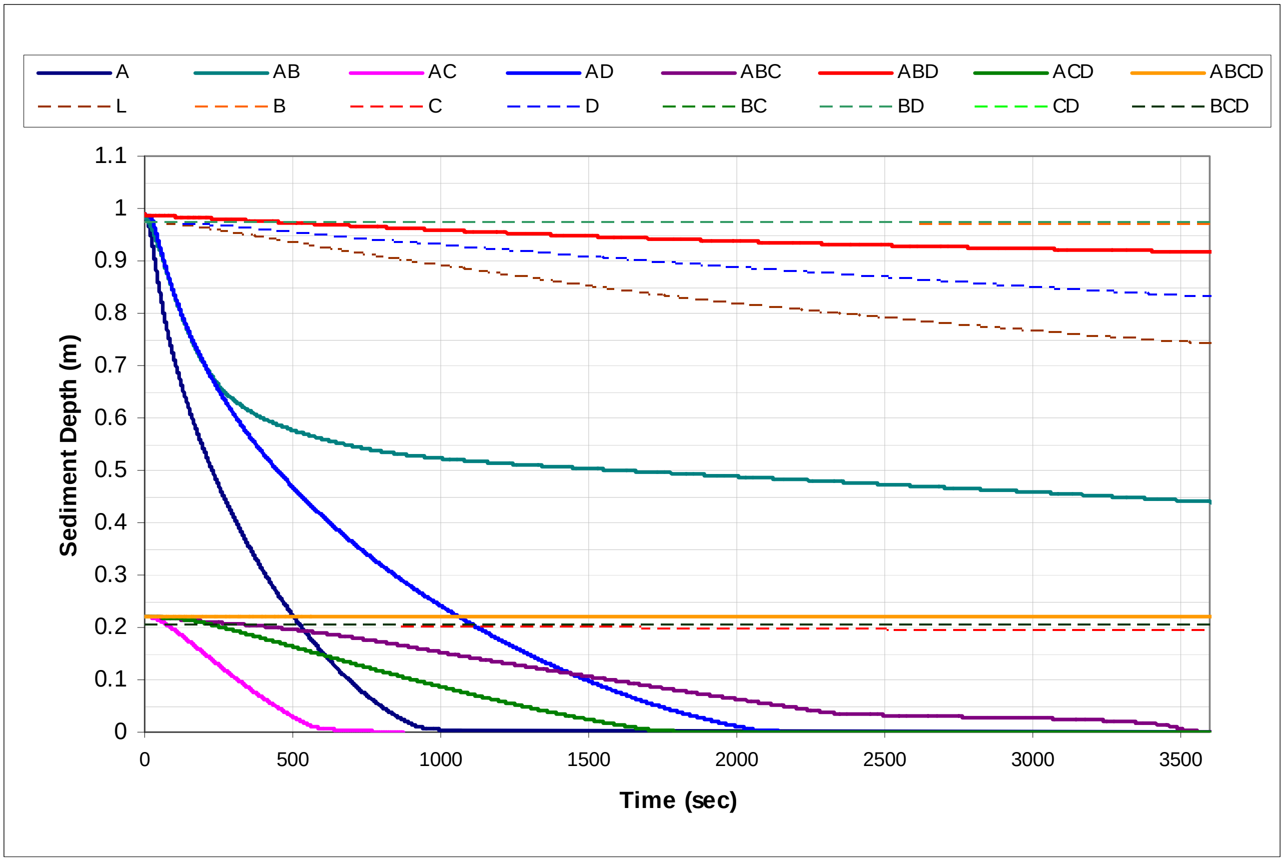


Figure A3. Experimental results from the 2D-CFD model – Sediment depth as a function of time

The analysis of the results of the experiment consists of determining the significant factors that affect the response. The significant factors that affect the response are called "*location factors*" and need to be considered in the prediction equation. A residual analysis is necessary to evaluate if the assumptions of the model are appropriate.

Location Factors and Prediction Equation

The location factors can be determined by three methods: eye ball with half-effects, normal probability plot of the effects, and a regression analysis to determine the *p-values*.

The first method is by eye ball, at which the half-effect of each factor is ranked and plotted to determine which factors have more effect than the others. This method is a first approach and may not be accurate when there is not substantial difference between half-effects; therefore, a more accurate method is required.

The following steps are required to determine the half-effects of factor A:

Determine the average of the low settings of factor A:

Avg Y @ -1 = (0.0831+0.0444+0.0018+0.0002+0.0011+0+0+0)/8 = 0.0163

Determine the average of the high settings of factor A:

Avg Y @ +1 = (0.9835+0.7465+0.2186+0.1324+0.4655+0.0293+0.0667+0)/8 = 0.3303

Determine the effect □:

Δ = (*Avg Y @ +1) – (Avg Y @ -1) = 0.3303-0.0163 = 0.3140*

Determine the half-effect D/2:

Δ/2 = 0.3140/2 = 0.1570

The same procedure is done for the other factors.

Additionally, the grand mean (Y_{grand} or $\bar{\bar{y}}$) is calculated as the average of all the responses:

Y(grand)=0.0831+0.0444+0.0018+0.0002+0.0011+0+0+0.9835+0.7465+0.2186+0.1324+0.4655+0.0293 +0.0667+0)/16 = 0.1733

Table A2 shows the results of the effects and half-effects.

A	B	C	D	AB	AC	AD	BC	CD	BD	ABC	ABD	ACD	BCD	ABCD	Y
-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	0.0831
-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1	1	-1	0.0444
-1	-1	1	-1	1	-1	1	-1	-1	1	1	-1	1	1	-1	0.0018
-1	-1	1	1	1	-1	-1	-1	1	-1	1	1	-1	-1	1	0.0002
-1	1	-1	-1	-1	1	1	-1	1	-1	1	1	-1	1	-1	0.0011
-1	1	-1	1	-1	1	-1	-1	-1	1	1	-1	1	-1	1	0.0000
-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1	0.0000
-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	-1	0.0000
1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	0.9835
1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	0.7465
1	-1	1	-1	-1	1	-1	-1	-1	1	-1	1	-1	1	1	0.2186
1	-1	1	1	-1	1	1	-1	1	-1	-1	-1	1	-1	-1	0.1324
1	1	-1	-1	1	-1	-1	-1	1	-1	-1	-1	1	1	1	0.4655
1	1	-1	1	1	-1	1	-1	-1	1	-1	1	-1	-1	-1	0.0293
1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1	0.0667
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.0000
														Y(grand)	*0.1733*

	A	B	C	D	AB	AC	AD	BC	CD	BD	ABC	ABD	ACD	BCD	ABCD
Avg Y @ -1	0.0163	0.2763	0.2942	0.2275	0.2603	0.2783	0.2224	0.1061	0.1384	0.1821	0.1217	0.1870	0.1432	0.1619	0.1574
Avg Y @ +1	0.3303	0.0703	0.0525	0.1191	0.0864	0.0683	0.1243	0.2405	0.2082	0.1645	0.2250	0.1596	0.2035	0.1847	0.1892
Δ	0.3140	-0.2060	-0.2417	-0.1084	-0.1739	-0.2101	-0.0981	0.1344	0.0698	-0.0175	0.1033	-0.0274	0.0603	0.0228	0.0318
Δ/2	*0.1570*	*-0.1030*	*-0.1209*	*-0.0542*	*-0.0869*	*-0.1050*	*-0.0491*	*0.0672*	*0.0349*	*-0.0088*	*0.0517*	*-0.0137*	*0.0301*	*0.0114*	*0.0159*

Table A2. Analysis of effects (Δ) for a 2^4 design

Figure A4 shows the half-effect of each factor ranked from the maximum to the minimum. In the figure it is possible to see that factors A (flow rate), However, it is not clear wheatear the other factors are significant or not.

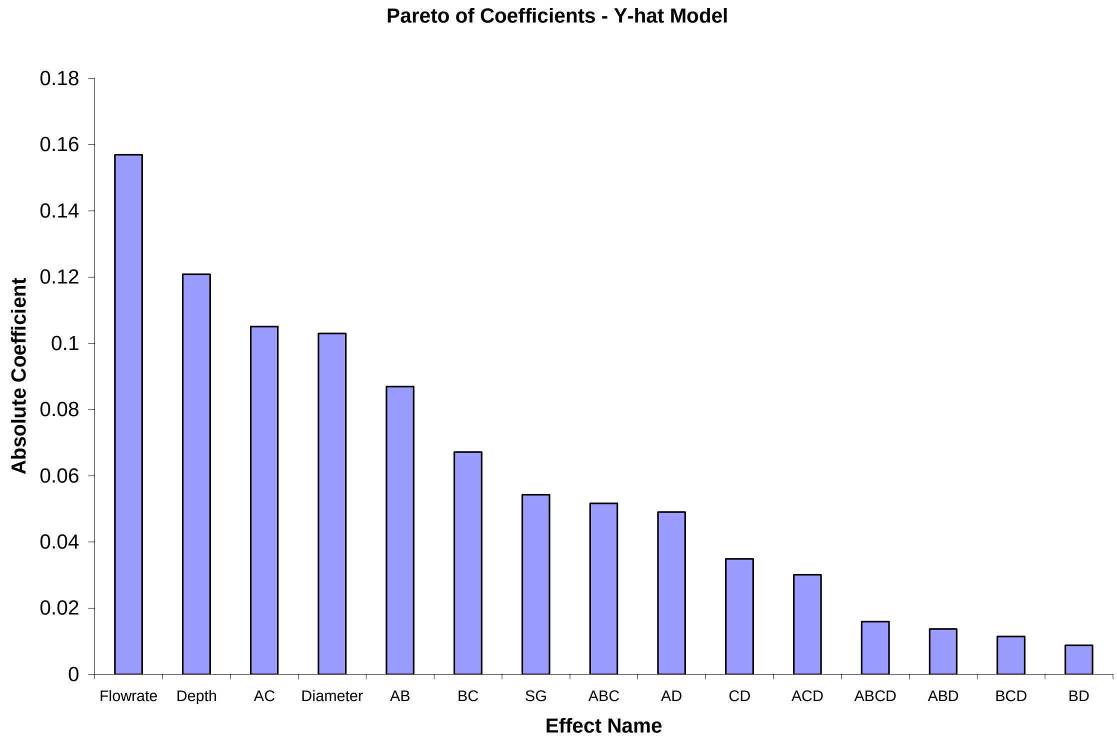


Figure A4. Pareto diagram of coefficients (half-effects) for the prediction equation

The second method is plotting a normal probability plot using the effects calculated in Table A2. Figure A5 shows that factors A, C, B, AC, and AB have a significant effect. The normal probability line should pass over the four, and three-way interactions which are more expected to be no significant.

To create the normal probability plot it is necessary to rank the effects from the smallest to the largest, and calculate the probability of each effect by using $p_i = \frac{(i - 0.5)}{n}$, where i = a specific rank, and n = maximum rank. Then, the Z-score with a distribution N (0, 1) is calculated for each effect using the probability p previously calculated; use NORMINV(pi, 0, 1) in Excel to calculate Z-score.

Visually, it is possible to detect that factor A, and interactions AC and AB are the farther from the normal probability line, and that factors B, and C do not look as far as AC and AB. However, if a higher order term is included in the model (an interaction), then all linear effects included in the higher order term need to be included in the model regardless of their significance; this is knows as the *hierarchy law*. For example, if the interaction AB is significant but not A and B, the prediction equation has to include A, B, and AB.

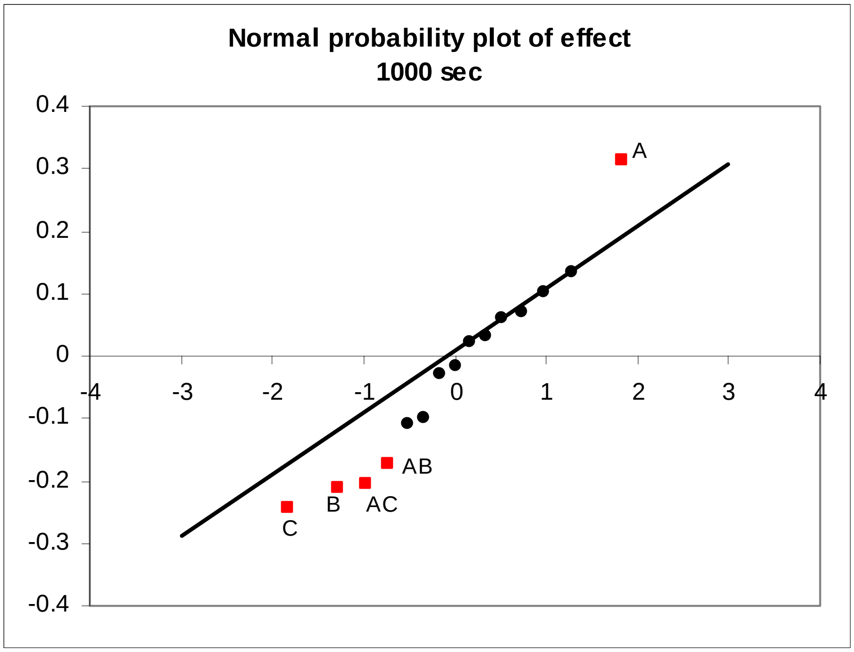


Figure A5. Normal probability plot of effects

The third method is the determination of the *p-value* for each factor using ANOVA. However, considering that this is a factorial experiment without replicates, it is not possible to calculate the error sum of squares (SSE) which is based on the standard deviation calculated from the replicates. Then, it is reasonable to assume that the higher-order interactions (four and three-way interactions) are no significant in the model, and then the sum of square of those interactions can be used like SSE. The methodology to calculate the p-values is the following:

1. Calculate the sum of square of each factor (each factor has 1 degree of freedom).

   Sum of square of each factor $MSB = \frac{N}{4}\Delta^2$ (only for 2-level designs)

   where

   MSB = The Mean Square Between for each factor
   N = total number of response values obtained in the entire experiment
   Δ = effect of each factor.

2. Calculate the Mean Square Error (***MSE***) adding the MSB of the four and three-way interactions (ABC, BCD, ABD, ACD, and ABCD) and dividing by the degrees of freedom which is 5 (one for each interaction).

3. Calculate *F* statistics for main effects and interactions by dividing the sum of square by the mean square error *F=MSB/MSE.*

4. Calculate the p-value of each factor using the F-statistic and the following degrees of freedom: df_{MSB}= 1 for 2-level design, and df_{MSE}= 5 (number of degrees of freedoms used to calculate MSE).

5. Identify the significant factors at significant levels α = 5% or α = 10%. A significant level of α = 5% was used for this example.

Table A3. ANOVA – calculation of p-values

	Effect	MSB	*F*	p-value
A	0.314	0.395	29.79	**0.0028**
B	-0.206	0.170	12.83	**0.0158**
C	-0.242	0.234	17.63	**0.0085**
D	-0.109	0.047	3.56	0.1179
AB	-0.174	0.121	9.14	**0.0293**
AC	-0.210	0.176	13.31	**0.0148**
AD	-0.098	0.039	2.91	0.1486
BC	0.134	0.072	5.45	0.0668
BD	-0.017	0.001	0.09	0.7738
CD	0.070	0.019	1.47	0.2799
ABC	0.103	*0.043*	3.22	0.1329
ABD	-0.027	*0.003*	0.22	0.6554
ACD	0.060	*0.014*	1.09	0.3436
BCD	0.023	*0.002*	0.16	0.7070
ABCD	0.032	*0.004*	0.31	0.6029
Sum. Three and four-way interactions		0.066		
MSE		0.0132		

According to Table A3, the significant factors and interactions that affect the response are A, B, C, AB, and AC. Those factors and interactions have to be in the prediction equation. The prediction equation can be written in terms of the grand mean and half-effects, excluding the no-significant factors.

$$\hat{y} = \bar{\bar{y}} + \left(\frac{\Delta_A}{2}\right)A + \left(\frac{\Delta_B}{2}\right)B + \left(\frac{\Delta_C}{2}\right)C + \left(\frac{\Delta_{AB}}{2}\right)AB + \left(\frac{\Delta_{AC}}{2}\right)AC$$

where,

$\hat{y}$ = predicted response (*Y pred*)

$\bar{\bar{y}}$ = grand mean (Y grand)

$\left(\frac{\Delta}{2}\right)$ = half-effects of each factor or interaction.

The prediction equation is given as

$$\hat{y} = 0.1733+0.157A-0.1030B-0.1209C-0.1050AC$$

Residual Analysis

In order to check the assumptions of the linear model presented previously, it is necessary to evaluate the trend, the homoscedastic, independence, and normality of the residuals.

Residual is defined as the difference between the observed and predicted values,

$$e = y_{obs} - y_{pred},$$

where

y_{obs} = observed responses
y_{preds} = predicted responses

Trend and homoscedastic of the residual is evaluated by plotting the residuals as a function of the fitted or predicted values. If the plot shows no substantial trend curve, and the vertical spread does not vary too much along the horizontal length of the plot, with the exception of the edges (homoscedastic), it is possible, but not certain, that the assumption of the linear model is appropriate (Navidi 2006). If the previous conditions do not apply to the plot, the linear model is not appropriate.

Figure A6 shows that there is not trend and the plot looks homoscedastic. However, considering that there are only few points, it is not possible to have a clear visual impression of homoscedastic or heteroscedastic. Therefore, the linear model should be considered as tentative (Navidi 2006).

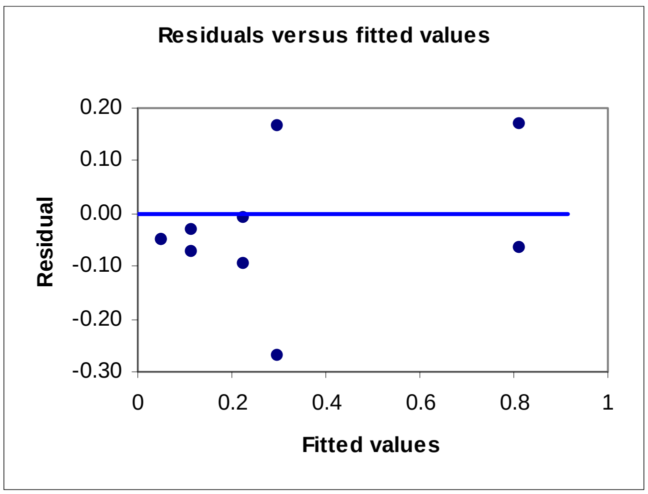


Figure A6. Residuals versus fitted values

Independence is evaluated by plotting the residuals as a function of the order of observation. This evaluation gives an idea about how the response varies over time, so it may be necessary to include the variable time into the model.

Considering that the response was based on computational results, it would not be necessary to evaluate this parameter. Additional analysis (not included in this experiment) performed with the variation of scour over the time shows that there is an evident dependency of the time and the results at 60 sec are different than the results at 1,000 sec for example. However, this analysis was focused only on the time 1,000 sec, so the dependency is not applicable.

Normality of the residuals is analyzed by plotting a normal probability plot of the residuals. If the plot looks straight, the residuals are normally distributed. Figure A7 shows that the residuals look pretty normal.

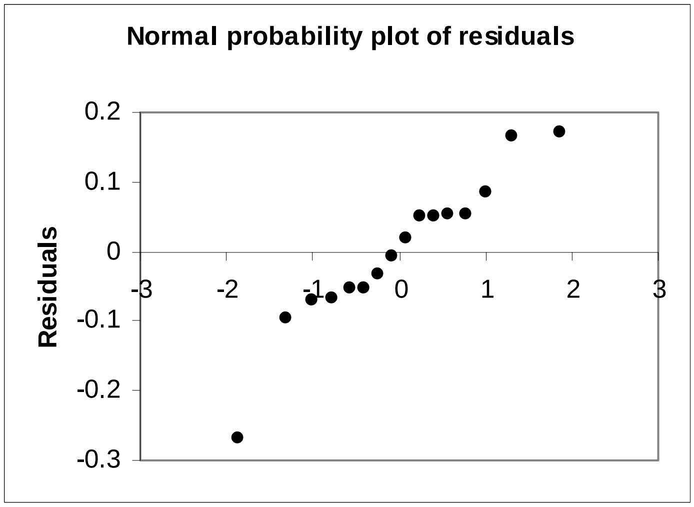


Figure A7. Normal probability plot of residuals

Finally, a comparison between the actual and the predicted response is showed in Figure A8. The figure shows that most of the predicted responses are close to the actual values with the exception of two values that are unpredicted by about 20%.

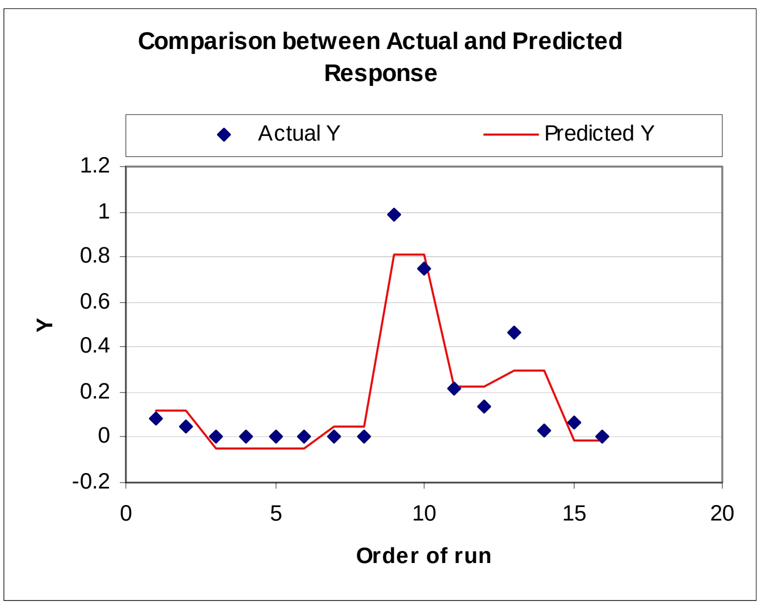


Figure A8. Comparison between Actual and Predicted Response

Example using Factorial Analyses to Evaluate Existing Data: Lake Tuscaloosa Water Quality

Introduction

This example was prepared by Tom Creech, a MS student in Biological Sciences at the University of Alabama as part of his thesis research. This analysis was conducted to better understand the processes which might be controlling the fate of wastewater and metals in the Lake Tuscaloosa reservoir, the main water supply to Tuscaloosa, AL, region. The data was obtained during December 2002 and January, 2003, and the basic relationships are summarized in the following discussion.

The data show that in the absence of extended periods of heavy rain, there is a clustering of sample locations by land cover (developed or undeveloped). After a period of heavy rain in February, the sample locations become less saline by dilution and generally enriched in dissolved iron to some degree. The clustering of sample locations by land cover is less distinct after periods of rain, probably based on the degree of surface water input.

Salinity is influenced by the water's source. Rainfall has extremely low Na. Groundwater has elevated Na due to rock weathering reactions. Wastewater has elevated Na due to detergents and human waste. Iron is a minor nutrient that participates in biological processes. It also precipitates from solution, depending on the concentration of total dissolved solids and oxidation-reduction conditions. Rainfall has extremely low Fe. Groundwater receives dissolved Fe from rock weathering reactions and decay of organic matter. Wastewater can have elevated Fe and it does contain elevated levels of other nutrients.

Therefore, there are several possible processes involved in determining the water quality in the lake. These are examined by grouping the undeveloped sample locations before heavy rainfall as a reference in Figure A9. The grouping of developed sites before rainfall exhibit elevated sodium and depleted iron, relative to the undeveloped sites. The high sodium values are suggestive of groundwater and wastewater sources. A variety of nutrients are associated with wastewater, and may be stimulating the biological activity. The iron depletion can be associated with increased biological activity, along with iron precipitation due to elevated TDS, and differences in the availability of an initial source of iron.

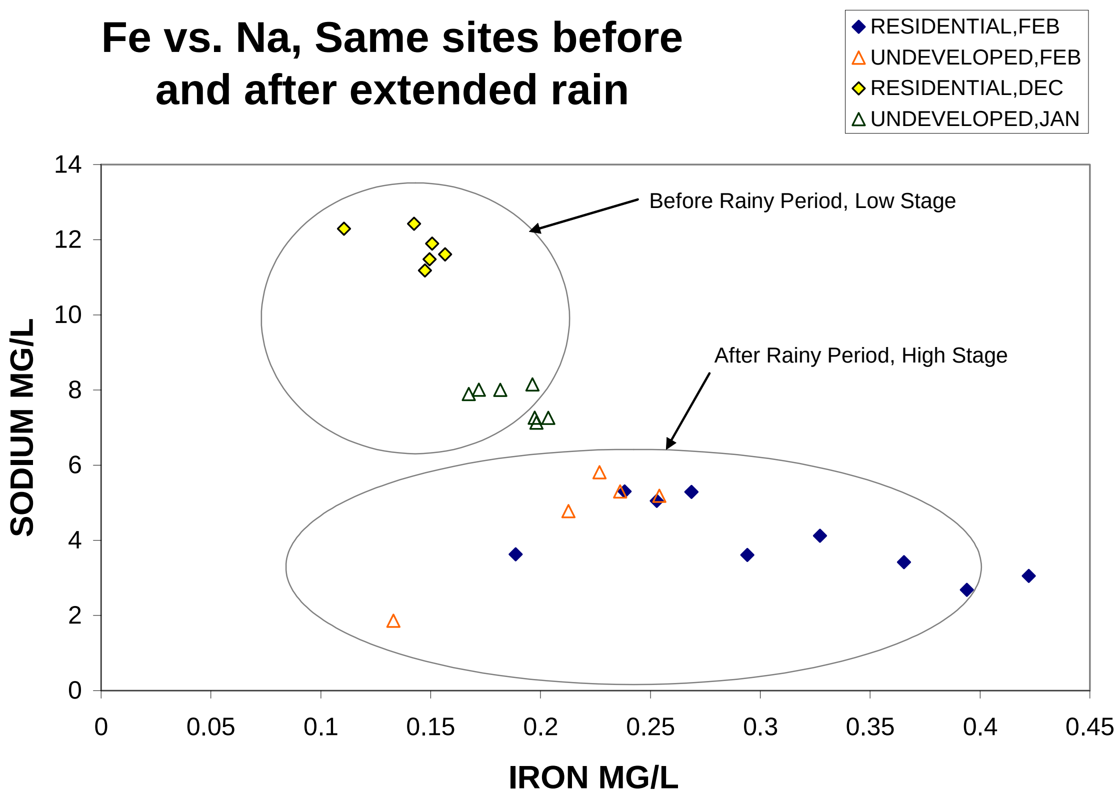


Figure A9. Clustering of sampling locations by sodium and iron concentrations during dry and wet weather.

The samples collected after rainfall are enriched in iron and depleted in sodium relative to the undeveloped, pre-rain samples. The depletion in sodium reflects dilution of the lake by rainwater. However, the relative enrichment of iron is more curious. It is possible that since lower TDS (and therefore, salinity) favors iron solubility instead of precipitation, more dissolved iron was observed despite changes in source (rainfall vs. groundwater +/-wastewater). The dilution and cloudy conditions may have also lowered the lake's productivity, and thus the demand for iron as a nutrient.

Experimental Design for Lake Tuscaloosa

To gain a better understanding of these and other relationships, a sample collection strategy was developed that will enable the data to be analyzed in a 2^3 full-factor statistical test. The factorial design will compare land cover, season, and lake stage (as an indication of incoming dilution flow, specifically from precipitation), as follows:

Land cover:	(+) developed	(-) undeveloped
Stage:	(+) high	(-) low
Season:	(+) summer	(-) winter

These three factors are likely to be primary controls over the water chemistry at a given location. Samples have been collected from 20 locations during the first winter, during both low stage and high stage conditions. To complete such a factorial analysis, summer samples from low stage and high stage conditions are also needed.

Experimental Design and Factorial Analysis for the North River site

Because the data set cannot yet produce a complete factorial analysis, related historical data was examined. Water quality data was selected for the North River USGS gage station because the river is the principle tributary to Lake Tuscaloosa. To better understand the potential geochemical interactions involving dissolved iron, the following 2^3 factorial test was conducted:

Discharge:	(+) less than median	(-) greater than median
Season:	(+) winter	(-) summer
Conductivity:	(+) less than median	(-) greater than median

Of the historical data available, these factors likely offer the best opportunity to investigate differences in the source and fate of dissolved iron in this river. Table A4 summarizes the data used in this analysis.

Table A4. North River Water Quality Data used in 2^3 Factorial Analysis

SAMPLE DATE	DISCHARGE, CUBIC FEET PER SECOND	SPECIFIC CONDUCTANCE MICROSIEMENS/ CM AT 25 DEG C	IRON, DISSOLVED (UG/L AS FE)
8/22/1989	12	485	180
6/1/1995	95	382	20
8/2/1995	10	371	50
7/8/1986	38	305	330
2/2/1982	83	183	110
7/14/1987	50	171	250
6/1/1989	79	150	390
1/3/1984	149	143	170
11/1/1978	9.8	136	610
8/3/1979	23	126	240
6/30/1983	29	123	530
7/1/1980	47	121	350
1/4/1983	120	117	80
11/19/1987	148	111	260
6/30/1981	32	108	260
7/30/1980	31	107	420
6/29/1982	15	101	380
7/7/1979	56	100	470
1/1/1982	202	95	700
8/15/1978	54	95	580
1/7/1987	330	94	60
1/1/1981	167	92	370
5/27/1977	69	85	140
1/19/1977	63	85	60
7/4/1984	43	85	220
10/31/1981	289	84	230
1/3/1986	812	81	100
5/30/1979	247	81	220
5/13/1977	30	80	660
8/9/1988	76	77	350
6/1/1982	85	75	10
5/3/1979	243	71	110
6/7/1977	92	70	370
6/3/1980	91	70	110
11/30/1979	352	68	120
1/2/1985	757	67	230
1/31/1980	265	67	110
1/6/1979	401	65	120
7/5/1985	623	63	50
12/21/1988	1240	62	50
2/23/1990	2100	59	110
12/9/1978	338	57	200
2/19/1977	337	57	100
1/3/1980	402	56	80
1/31/1981	3860	49	170
4/30/1980	1020	48	70
2/7/1979	1290	43	80

The results of the factorial analysis are summarized in Table A5, with the standard error calculations shown in Table A6.

Table A5. Results of 2^3 Factorial Analysis of North River Water Quality Data

MEAN DISCHARGE=	381.5409	MEDIAN DISCHARGE=	95	I used 95 as a (+/-) boundary
MEAN CONDUCTIVITY=	95.06818	MEDIAN CONDUCTIVITY=	85	I used 90 as a (+/-) boundary
SEASONS= SUMMER(-) AND WINTER (+)				

Table of Contrast Coefficients: (Means and Effects not rounded to significant digits)

		Main Effects			Two and Three factor interactions					
condition	mean	season	discharge	conductivity	sd	sc	dc	sdc	Fe microg/L (mean)	N samples
1	+	-	-	-	+	+	+	-	20.00	1
2	+	+	-	-	-	-	+	+	273.33	6
3	+	-	+	-	-	+	-	+	340.77	13
4	+	+	+	-	+	-	-	-	360.00	2
5	+	-	-	+	+	-	-	+	112.50	4
6	+	+	-	+	-	+	-	-	130.77	13
7	+	-	+	+	-	-	+	-	265.71	7
8	+	+	+	+	+	+	+	+	60.00	1
effect:	195.39	21.28	122.47	-106.28	-114.52	-115.00	-81.25	2.53		
First case error:	75	150	150	150	150	150	150	150	(see calculated effects table)	
Second case error:	38	75	75	75	75	75	75	75	(see calculated effects table)	
Third case error:	39	77	77	77	77	77	77	77	(see calculated effects table)	

Note: The second and third case error calculations were after the extreme values were eliminated.

There are several large effects; however none of the effects are necessarily significant because they are all within the range of standard error (see first case error). However, flow is shown to have the largest effect on the measured water quality. The number of samples in a given population varied from 1 to 13. Within each sample population, there was a significant standard deviation (summarized below). This wide natural variation obscures the effects of the factors and factor interactions. The interdependence of the selected variables (season, discharge, and conductivity) further diminishes the effectiveness of this approach. Ideally, factors should be more independent of one another. The parameter being measured should be the main dependent variable.

Table A6. Standard Error Calculations for North River Factorial Tests

The standard error of the mean (for each condition's mean) is the standard deviation of the sample group divided by the square root of the sample size
The standard error of the main effects and factor interactions is calculated differently, and is shown in another table.

Summary table of individual yields for each condition: (Not yet rounded to significant digits)

condition	1	2	3	4	5	6	7	8
Fe microg/L:	20	60	50	110	50	50	10	60
		80	180	610	70	80	110	
		170	240		110	80	140	
		260	250		220	100	220	
		370	260			100	354	
		700	330			110	370	
			350			110	660	
			380			120		
			390			120		
			420			170		
			470			200		
			530			230		
			580			230		
Standard Dev.	#DIV/0!	238.80	145.91	353.55	75.88	58.23	216.66	#DIV/0!
Square Root N	1.00	2.45	3.61	1.41	2.00	3.61	2.65	1.00
Standard Error	N/A	97.49	40.47	250.00	37.94	16.15	81.89	N/A
Average (microg/L)	20.00	273.33	340.77	360.00	112.50	130.77	265.71	60.00

Summary

The following shows the effects and standard errors. As noted above, none of the factors were clearly significant based on the large errors, but the discharge effects seem to be most important.

Effect	Estimate +/- Standard Error		Probability (%)
Average	195.39	+/- 75	
Main Effects			
Season	21.28	+/- 150	21.42
Discharge	122.47	+/- 150	7.40
Conductivity	-106.28	+/- 150	64.29
Two-Factor Interactions			
S x D	-114.52	+/- 150	78.57
S x C	-115.00	+/- 150	92.66
D x C	-81.25	+/- 150	50.00
Three-Factor Interaction			
S x D x C	2.53	+/- 150	35.71

A probability plot can also be used to identify significant factors and factor interactions. Outliers from the normal probability line indicate significant factors or factor interactions.

The results can also be graphically displayed in the x,y,z coordinate system. Figure A10 shows the concentration of iron at each corner of the cube formed by the three factors studied.

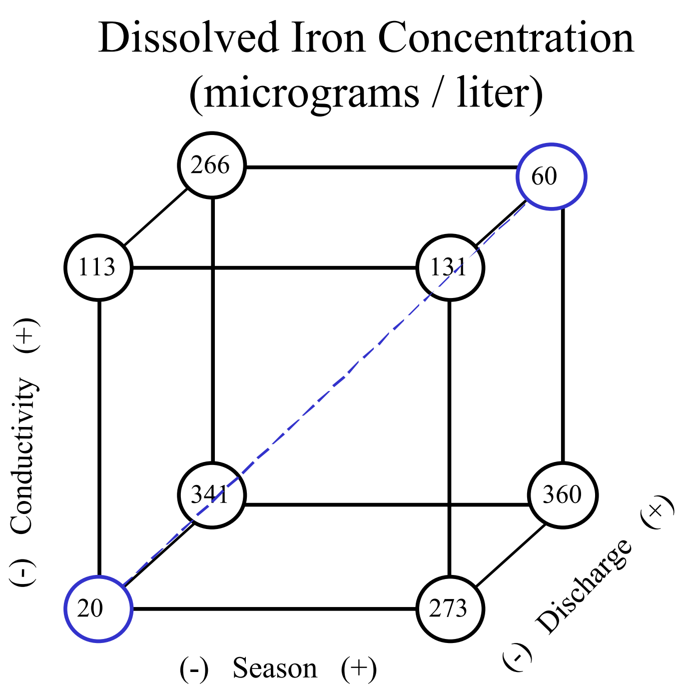


Figure A10. 2^3 Factorial diagram showing observed iron concentrations.

The result shows that the diagonal between the 60 microgram/L concentration and the 20 microgram/L concentration is significant because the two values are much lower than the rest. A diagonal indicates that there is a significant two-factor interaction.

Factorial Analysis used in Modeling the Fates of Polycyclic Aromatic Hydrocarbons (PAHs) Affecting Treatability of Stormwater

Abstract

This example of using a factorial analysis approach for fate modeling was prepared by Jejal Reddy, a Ph.D. student in the Department of Civil, Construction, and Environmental Engineering at the University of Alabama. The first part of this discussion examines the sensitivity of the different factors that affect the partitioning of PAHs commonly found in stormwater into different environmental phases using the fugacity calculation methods presented by Mackay, *et al* (1992). The predictions indicated that most of the PAHs are partitioned onto particulates than in the water or air phases. The second part of this discussion compares the predicted portioned values with actual stormwater PAH association values observed during prior research (Pitt, *et al*. 1999). Other than a few exceptions (Benzyl butyl phthalate, floranthene, and pyrene), the predicted percentages are in general agreement with the field measurements made by Pitt, *at al.* (1999). The third part of the discussion describes the effects of selected variables (temperature, PAH concentrations, suspended solids concentrations, and the organic fraction of the suspended solids) on partitioning of the PAHs by using a full 2^4 factorial experimental design (Box, *et al*. 1978). Concentrations of the PAHs and the concentrations of the suspended solids, and to a lesser extent the organic content of the suspended solids, were found to affect the partitioning of the PAHs into sediment matter.

Introduction

Polycyclic aromatic hydrocarbons (PAHs) are a major concern affecting public health and the natural environmental due to their carcinogenic and mutagenic properties. After the public drinking water act was implemented, PAH contributions from industrial sources were reduced, but expanding urbanization has increased the PAHs contribution from stormwater runoff. The increases in PAH concentrations in the environment is coincident with increases in reported automobile usage (Metre, *et al.* 2000). Stormwater runoff from impervious areas, along with wear and tear of vehicle tires and asphalt road surfaces, are responsible for much of the PAHs contributed to surface waters, especially associated with particulate matter. Due to their persistent organic pollutant (POP) nature, PAHs persist in the environment for long periods of time and accumulate to higher and higher concentrations with new discharges.

When PAHs are present in stormwater, they are partitioned into different phases which affect their treatability and how they should be analyzed. Sorption plays an important role in the fate of these organic contaminates. Due to their extremely low solubility and their hydrophobic nature, most PAHs are predominantly associated with particulate matter. PAHs in urban runoff can occur in both particulate and soluble forms, although studies have identified the particulate forms as being the most predominate (Pitt, *et al.* 1999). According to the Hwang and Foster study on urban stormwater runoff in Washington DC (2005), particulate-associated PAHs account for 68-97% of total PAHs in the runoff. Fortunately, the organic contaminates associated with particulate matter can be more readily removed by common sedimentation stormwater control practices compared to filterable PAHs. The particulate-bound PAHs also tend to settle and accumulate in receiving water sediments. The behavior of contaminants in the environment depends primarily on their physical and chemical properties and the reactivity of the compound. The important properties of compounds that affect their treatability and fate include their partition coefficients, Henry's law constant, and water solubility, amongst others. Examining the factors influencing the partitioning of the organic contaminants is very important in understanding the treatability of the organics and when conducting risk assessments associated with contaminated receiving waters.

The first part of this discussion examines the sensitivity of the different factors that affect the partitioning of PAHs commonly found in stormwater into air, water, suspended solids, and sediment phases using the fugacity calculation methods presented by Mackay, *et al* (1992). Typical stormwater and urban receiving water conditions are used in these calculations. The second part of this paper compares the predicted portioned values with actual stormwater PAH association values observed during prior research (Pitt, *et al.* 1999). The third part of the discussion describes the effects of selected variables on partitioning of PAHs using a full 2^4 factorial experimental design (Box, *et al*. 1978).

Methodology

The fugacity models described by Mackay, *et al.* (1992) are methods used to determine the partitioning of a chemical contaminant into solid, liquid, and gaseous phases once they are released into the environment. Fugacity is defined as the escaping tendency of a chemical substance from a phase. To study the partitioning behavior of PAHs in the environment, Mackay's level I calculations (which do not consider bioaccumulation rates or kinetics) were used as a preliminary assessment. The level I fugacity model describes the partitioning of the chemical contaminant into solid, liquid and gaseous phases once they are released into the environment, and assume equilibrium conditions. This model is based on the physical-chemical properties of the chemical contaminant and the media. These properties include temperature, flows and accumulations of air, water and solid matter. The composition of the media is also an important property of the media. The physical-chemical properties of the contaminant chemical include the partition coefficients, Henry's law constant, and solubility of the contaminant. Equations involved in the model calculations are shown below:

$$C = Z * f \quad \text{(or)} \quad f = \frac{M}{\sum (V_i * Z_i)}$$

Where, C = Concentration of contaminant, mol/m^3; Z = fugacity capacity constant, mol/m^3; f = fugacity, Pa; V_i = Volume of the corresponding phases; and Z_i = fugacity capacities of each phase for air, water, sediment, and suspended sediment for i =1, 2, 3, and 4 respectively and are defined as follows.

$$Z_1 = \frac{1}{RT};\ Z_2 = \frac{1}{H};\ Z_3 = Z_2 * \mathrm{P}_3 * \phi_3 * \frac{K_{OC}}{1000};\ \ Z_4 = Z_2 * \mathrm{P}_4 * \phi_4 * \frac{K_{OC}}{1000}$$

Where, R = gas constant (8.314 J/mol K); T = absolute temperature (K); H= Henry's law constant (Pa.m^3/mol); K_{OC} = Organic-water partition coefficient; P_3 = density of sediment (kg/m^3); P_4 = density of suspended sediment (kg/m^3); $Ø_3$= organic fraction of sediment; and $Ø_4$= organic fraction of suspended sediment.

Pitt, *et al.* (1999) conducted analytic research considering more than thirty organic contaminants commonly found in stormwater runoff. They analyzed more than 100 samples collected from different sources areas in and around Birmingham, AL. The source areas represented by the samples included roofs, parking areas, storage areas, streets, loading docks, and vehicle service areas, plus nearby urban creeks, in residential, commercial, industrial and mixed land use areas. Among all the organic contaminants analyzed, polycyclic aromatic hydrocarbons were detected most frequently. The concentrations of organics detected varied considerably among the different source areas. Roof runoff, vehicle servicing areas, and parking areas were found to have the largest concentrations of organic toxicants in collected runoff. The fugacity model predicted partition values were compared to actual monitored partition PAH values obtained by Pitt, *et al.* (1999). Table A10 shows the concentrations and percentage of selected PAHs portioned in water and suspended solids from this prior research.

The final part of this paper examines the effects of some selected environmental factors on the partitioning of the PAHs into different media using a full 2^4 factorial experimental design (Box, *et al.* 1978). The full factorial design experimental setup used is helpful in studying the effects of individual variables and also the effects of interactions of the variables. The design matrix used in this factorial study is shown in Table A7. The factors studied, and their low and high values used in the calculations, are shown in Table A8. The low and high values of the factors were chosen based on typical observations for stormwater and urban receiving waters.

Table A7. 2^4 Factorial Design (Box, *et al*. 1978)

Run	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	+	+	+	-	+	+	-	+	-	-	+	-	-	-	-
3	+	+	-	+	+	-	+	-	+	-	-	+	-	-	-
4	+	+	-	-	+	-	-	-	-	+	-	-	+	+	+
5	+	-	+	+	-	+	+	-	-	+	-	-	+	-	-
6	+	-	+	-	-	+	-	-	+	-	-	+	-	+	+
7	+	-	-	+	-	-	+	+	-	-	+	-	-	+	+
8	+	-	-	-	-	-	-	+	+	+	+	+	+	-	-
9	-	+	+	+	-	-	-	+	+	+	-	-	-	+	-
10	-	+	+	-	-	-	+	+	-	-	-	+	+	-	+
11	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+
12	-	+	-	-	-	+	+	-	-	+	+	+	-	+	-
13	-	-	+	+	+	-	-	-	-	+	+	+	-	-	+
14	-	-	+	-	+	-	+	-	+	-	+	-	+	+	-
15	-	-	-	+	+	+	-	+	-	-	-	+	+	+	-
16	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+

(A, B, C, D are factors to be studied, + High value, - Low value, combinations of A, B, C, D indicates factors interactions)

Table A8. Values used in Factorial Analysis.

Variable	Low value	High value
Temperature (A), ^{o}C	5	25
Concentration of Contaminant (B), µg/L	10	300
Concentration of Suspended Solids(C), mg/L	10	500
Organic Fraction of Suspended Solids (D)	0.05	0.2

Results

The predicted partition values, as percentages, are shown in Table A9. The values indicate, as expected, that most of the PAHs are partitioned more onto the sediment than in the other phases. The low molecular weight PAHs (having fewer carbon rings) are mostly partitioned into the water phase compared to those having higher molecular weights. Figures A11 and A12 show the relationships between the $\log K_{ow}$ and the $\log K_{oc}$ values of the PAHs and their partitioning into water and sediment phases, respectively. PAHs with $\log K_{ow}$ or $\log K_{oc}$ values greater than 4 are mostly partitioned onto sediment compared to other phases.

Table A9. Predicted PAH Partition Values

Contaminant	Log K_{ow}	K_{oc}	H (Pa-M^3/Mole)	% partitioning into different phases at equilibrium			
				Air	water	Sed-iment	Sus. Sed.
Bis(2-chloroethyl) ether	1.39	23	1.31	0	98	2	0
1,3-Dichlorobenzene	3.49	310	192.5	0	77	23	0
Bis(chloroisopropyl) ether	2.58	73	11.14	0	93	7	0
Hexachloroethane	4.28	1175000	253.3	0	0	98	2
Naphthalene	3.51	1300	46.61	0	44	55	1
Phenanthrene	4.52	23000	25.93	0	4	94	2
Anthracene	4.34	26000	1.79	0	4	94	2
Benzyl butyl phthalate	4.78	85687	0.132	0	1	97	2
Fluoranthene	5.22	59500	1712	0	2	96	2
Pyrene	5.32	73350	1.89	0	1	97	2
Benzo(a)anthracene	5.91	333261	0.067	0	0	98	2
Chrysene	5.71	210273	0	0	0	97	2
Benzo(b)fluoranthene	6.57	1200000	1.21	0	0	98	2
Benzo(k)fluoranthene	6.45	1100000	105.47	0	0	98	2
Benzo(a) pyrene	6.06	980000	0.24	0	0	98	2
Benzo(g,h,I) perylene	7.1	5161594	0.01	0	0	98	2

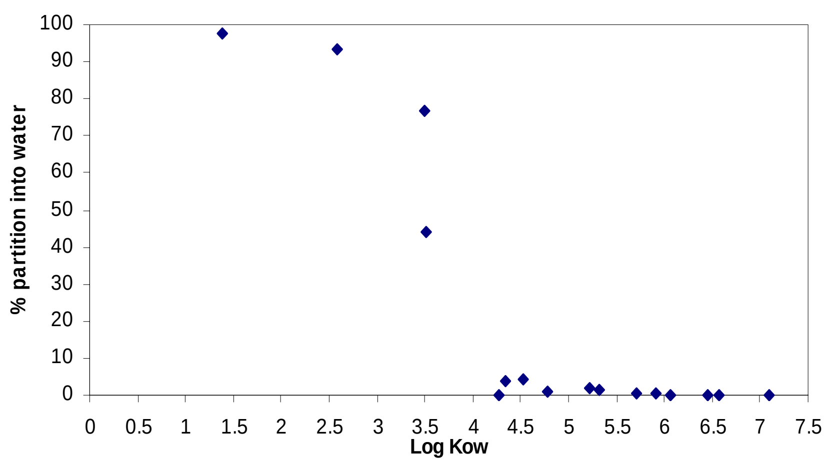


Figure A11. logK_{ow} versus % partition of PAHs into water phase.

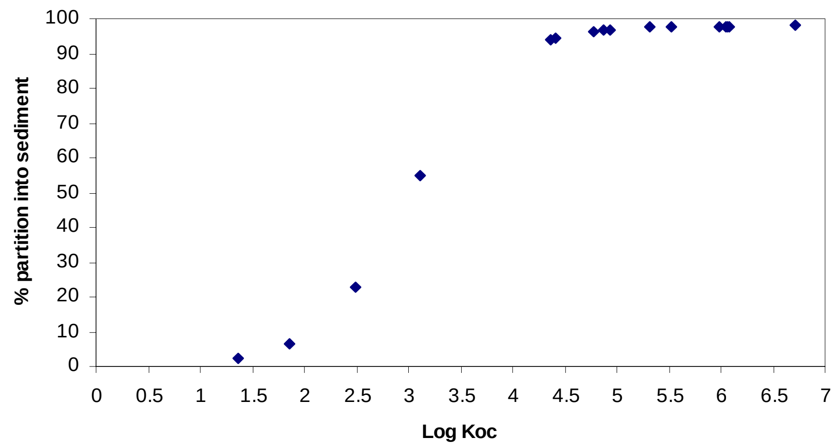


Figure A12. LogKoc versus % partition of PAHs into sediment phase.

Tables A10 and A11 indicate the percentage partitioning of the PAHs into the different phases, as observed by Pitt, *et al*. (1999), and the model-predicted values, respectively. Figure A13 is a plot showing the relationships between the observed and predicted partitioning. The comparison of predicted and observed values showed that the predicted percentages are in general agreement with the field measurements. Benzyl butyl phthalate, floranthene, and pyrene show somewhat higher observed percentages of partitioning onto suspended solids compared to the model-predicted values. Variations in concentrations of PAHs associated with particulate matter depend on the source areas, as shown by Mahler, *et al*. (2005). They found that particulate bond PAHs in runoff from coal-tar sealed parking areas was 65 times higher than found from un-sealed parking areas. Similarly, Pitt, *et al.* (1999) observed high concentrations of organics from vehicle servicing area compared to all other source areas monitored.

Table A10. Percentage partitioning of selected PAHs observed by Pitt, *et al*. 1999

Contaminant	Amount of contaminants (µg/L)			% Association	
	Non-filtered	Filtered (in water phase)	In Suspended Solids phase	Water	Sus – Solids
Benzyl butyl phthalate	73	2	71	3	97
Fluoranthene	28	7	21	25	75
Pyrene	31	2	29	8	92
Benzo(a)anthracene	32	0	32	0	100
Benzo(b)fluoranthene	61	0	61	0	100
Benzo(k)fluoranthene	47	0	47	0	100
Benzo(a)pyrene	70	0	70	0	100
Benzo(g,h,I) perylene	20	0	20	0	100

Table A11. Predicted % Partition of Selected PAHs

Contaminant	% Partitioning		
	Air	Water	Suspended Solids
Benzyl butyl phthalate	0	37	63
Fluoranthene	0	45	54
Pyrene	0	40	60
Benzo(a)anthracene	0	13	87
Chrysene	0	19	81
Benzo(b)fluoranthene	0	4	96
Benzo(k)fluoranthene	0	4	96
Benzo(a)pyrene	0	5	95
Benzo(g,h,I) perylene	0	1	99

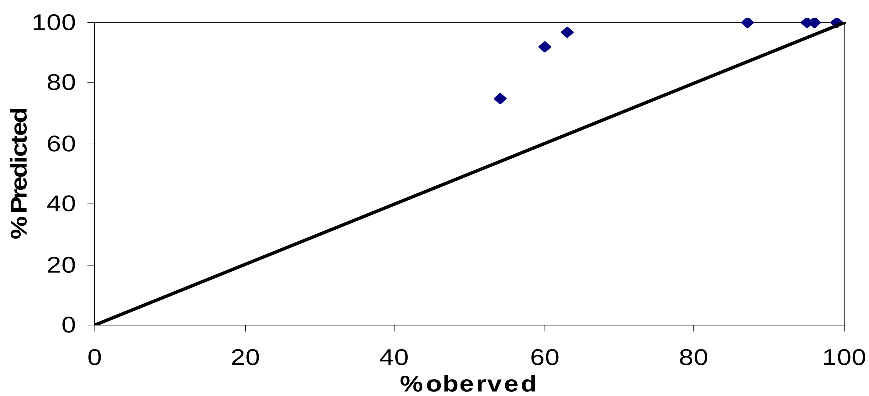


Figure A13. Comparison of predicted values versus observed values.

The analysis of the effects of environmental factors on partitioning of PAHs indicated that the main variables which affect PAH partitioning onto suspended sediment were the concentrations of the PAH compounds and the concentrations of the suspended solids. The organic content of the suspended solids also affected the partitioning of the PAHs into suspended solids, but to a lesser extent. In the case of partitioning into the water phase, the concentration of the PAHs was found to have the greatest positive effect, and the concentration of the suspended solids had a significant negative effect (the higher the SS concentration, more of the PAHs were associated with the sediment). Figures A14 and A15 are probability plots indicating the significant factors affecting anthracene partitioning into the water phase and suspended sediment phase, respectively. Indicated factors, B is concentration of contaminant, C is concentration of suspended solids and D is organic fraction of suspended Solids. The term BC indicates the interaction of factors B and C.

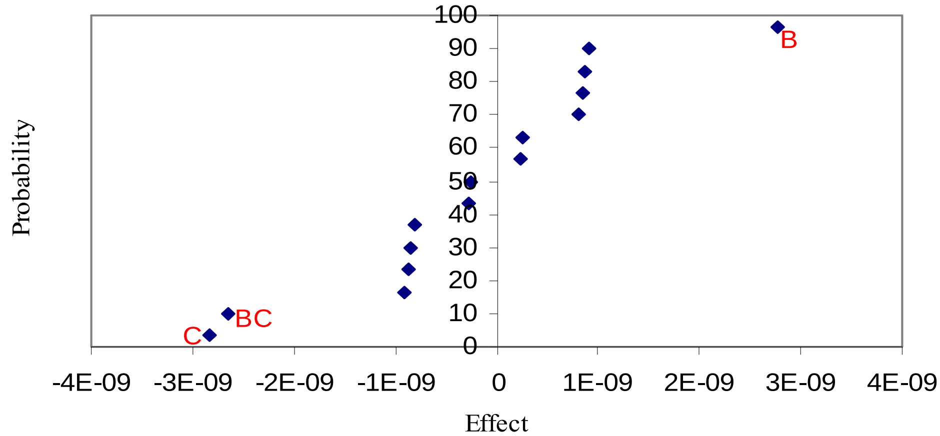


Figure A14. Probability plot to identify important factors affecting Anthracene partitioning into the water phase.

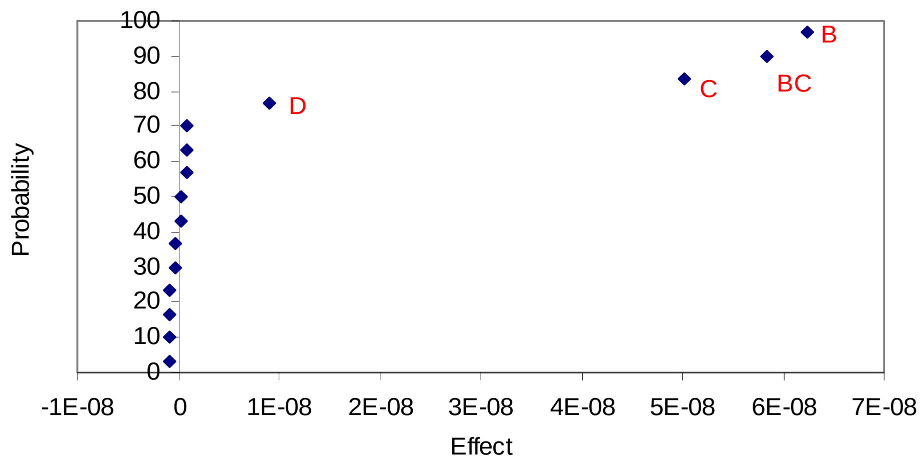


Figure A15. Probability plot to identify important factors affecting Anthracene partitioning into suspended sediment phase.

Conclusions

The fugacity level 1 calculations were performed for selected environmentally important PAH contaminants. The model-predicted values show that the contaminants are more likely to be associated with the solid phase (mostly with sediment) and less with other phases. There is a clear similarity between predicted and actual observations when compared to prior research (Pitt, *et al.* 1999) in identifying the most important media for PAH associations in the environment. The field measurements showed a greater percentage of PAHs associated with particulate matter than the percentage predicted by the fugacity model. This may be due to the variable properties of the suspended solids, or the conditions of the environment.

The factorial analysis identified concentrations of suspended solids, the concentration of the contaminant, and their interaction, as major factors affecting the PAH partitioning onto suspended matter. The identified behavior of PAHs association with suspended particulates helps in identifying better treatment options for the control of PAH contamination from stormwater. As modeling and field results shows, PAHs are mostly associated with particulate matter in water systems. The most common method currently used by analytical laboratories to analyze PAHs is solid phase extraction (SPE). This method is not reliable as the true recovery of PAHs from particulates using SPE procedures is very poor. The use of continuous extraction using separation funnels and multiple solvents has been shown by Pitt, *et al.* (1999) to be much more suitable for samples containing significant amounts of PAHs associated with particulates. Unfortunately, that is a tedious process. Current research at the University of Alabama is developing and testing a more reliable and quicker method for the analysis of PAHs associated with different particle sizes, using a sequential procedure focusing on thermal desoprtion for the particulate-bound PAHs, and SPE for the filterable PAH forms.

Acknowledgements
This material is based upon work supported by the National Science Foundation under Grant No. EPS-0447675. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Appendix B: Examples for Specific Statistical Tests

Probability Plot Preparation using Excel

Excel can be used to prepare probability plots, but it takes several steps, as shown in the following examples:

	C	D	E	F
	rank	value	p	Z
7		45.5878		
	1	6	0.02381	-1.98075
8		46.8340	0.07142	
	2	5	9	-1.46523
9		47.2413	0.11904	
	3	3	8	-1.17976
10		47.7358	0.16666	
	4	8	7	-0.96742
11		49.4814	0.21428	
	5	6	6	-0.79164
12		50.6825	0.26190	
	6	7	5	-0.63748
13		51.4490	0.30952	
	7	5	4	-0.4972
14		52.0580	0.35714	
	8	5	3	-0.36611
15		52.5684	0.40476	
	9	3	2	-0.24104
16		57.9105	0.45238	
	10	8	1	-0.11965
17		59.0166		
	11	4	0.5	-1.4E-16
18		59.5010	0.54761	0.11964
	12	2	9	8
19		59.6059	0.59523	
	13	1	8	0.24104
20		59.8476	0.64285	0.36610
	14	4	7	6
21			0.69047	0.49720
	15	60.6413	6	1
22		61.1086	0.73809	0.63748
	16	6	5	4
23		63.0009	0.78571	0.79163
	17	3	4	9
24		63.2458	0.83333	0.96742
	18	2	3	2
25			0.88095	1.17976
	19	64.2746	2	1
26		66.5883	0.92857	1.46523
	20	3	1	4
27		67.4906		1.98075
	21	2	0.97619	2

To calculate the p values (in column E for the example):

=+(C7-0.5)/$C$27

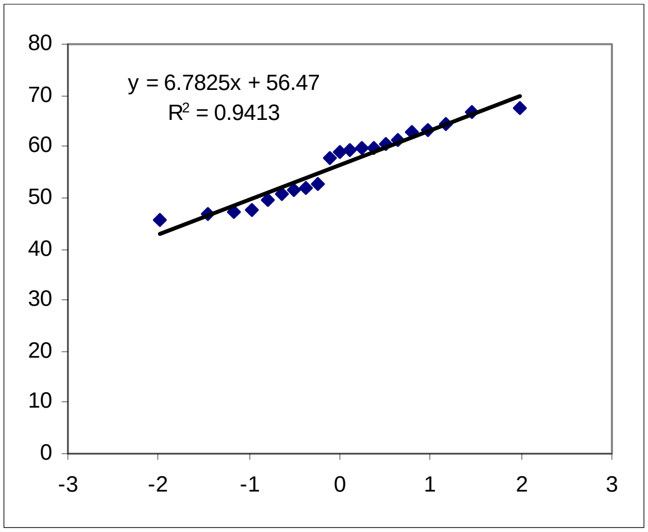
The total number of samples is in row C27, the largest rank.

This is for row 7 and the actual sorted values are in column D (not used) and the ranks are in column C. The Z scores are then calculated for each observation:

=+NORMINV(E7,0,1)

The mean of the distribution is 0 and the standard deviation is 1 for this example.

Again for row 7, with the Z scores in column F and the probability values in column E. The Z scores are plotted on the X-axis and the actual data values are plotted on the Y-axis:



In this case, a first-order polynomial regression was fitted to the probability plots. This enables "eye-balling" that data fit (the data should be on a "straight" line if normally distributed), but the regression information does not present any statistical significance for normality. Many statistical programs offer tests to verify if the data is normally distributed. The Anderson-Darling is one such test. In that case, the data are compared to corresponding points on the fitted normal probability line and a paired test indicates if they are from the same population. A values of <0.05 indicates that they are significantly different, and data is not normally distributed. These plots and tests can also be conducted on log-transformed data to check for log-normalcy. The following example compares regular and log-transformed data:

	A	B	C	D
1	rank	**Conductivity (uS/cm @25ºC)**	p	Z
2	1	7	0.00266	-2178.9
3	2	14	0.007979	-1820.87
4	3	24	0.013298	-1638.14

5	4	25	0.018617	-1510.73
6	5	27	0.023936	-1411.35
7	6	28.9	0.029255	-1329.1
8	7	32.8	0.034574	-1258.46
9	8	33.8	0.039894	-1196.26
10	9	35	0.045213	-1140.48
11	10	35.9	0.050532	-1089.75
12	11	36.9	0.055851	-1043.11
13	12	37.2	0.06117	-999.853
14	13	39	0.066489	-959.446
15	14	39.4	0.071809	-921.469
16	15	39.8	0.077128	-885.592
17	16	39.9	0.082447	-851.55
18	17	40.2	0.087766	-819.123
19	18	41.6	0.093085	-788.131
20	19	43	0.098404	-758.421
21	20	45	0.103723	-729.866
22	21	45	0.109043	-702.354
23	22	45.3	0.114362	-675.79
24	23	45.8	0.119681	-650.093
25	24	46	0.125	-625.188
26	25	47	0.130319	-601.014
27	26	47	0.135638	-577.513
28	27	47.2	0.140957	-554.635
29	28	49	0.146277	-532.335
30	29	50	0.151596	-510.574
31	30	50	0.156915	-489.314
32	31	50.4	0.162234	-468.524

33	32	51.4	0.167553	-448.172
34	33	52	0.172872	-428.233
35	34	53	0.178191	-408.681
36	35	54.3	0.183511	-389.494
37	36	55	0.18883	-370.65
38	37	56	0.194149	-352.131
39	38	58	0.199468	-333.918
40	39	59	0.204787	-315.995
41	40	60	0.210106	-298.347
42	41	60.1	0.215426	-280.96
43	42	60.8	0.220745	-263.82
44	43	64.2	0.226064	-246.914
45	44	65	0.231383	-230.233
46	45	65.5	0.236702	-213.763
47	46	66	0.242021	-197.496
48	47	67.7	0.24734	-181.422
49	48	68	0.25266	-165.532
50	49	69	0.257979	-149.817
51	50	69	0.263298	-134.269
52	51	69.2	0.268617	-118.881
53	52	70.4	0.273936	-103.645
54	53	70.5	0.279255	-88.5545
55	54	71	0.284574	-73.6032
56	55	72	0.289894	-58.7848
57	56	74.9	0.295213	-44.0934
58	57	75.6	0.300532	-29.5233
59	58	76.6	0.305851	-15.0693
60	59	77	0.31117	-0.7261
61	60	78	0.316489	13.51112

62	61	78.8	0.321809	27.64709
63	62	80	0.327128	41.68634
64	63	80	0.332447	55.6332
65	64	80	0.337766	69.49187
66	65	80.65	0.343085	83.26635
67	66	83	0.348404	96.96055
68	67	84.7	0.353723	110.5782
69	68	85	0.359043	124.1229
70	69	85	0.364362	137.5982
71	70	86	0.369681	151.0075
72	71	86	0.375	164.3541
73	72	87	0.380319	177.6412
74	73	88	0.385638	190.8718
75	74	88	0.390957	204.049
76	75	91	0.396277	217.1758
77	76	92.6	0.401596	230.255
78	77	96	0.406915	243.2895
79	78	99	0.412234	256.2819
80	79	99.6	0.417553	269.2351
81	80	100	0.422872	282.1515
82	81	103	0.428191	295.0338
83	82	110	0.433511	307.8846
84	83	110	0.43883	320.7063
85	84	110	0.444149	333.5014
86	85	112	0.449468	346.2723
87	86	114	0.454787	359.0215
88	87	115	0.460106	371.7512

89	88	117	0.465426	384.4639
90	89	118	0.470745	397.1618
91	90	123	0.476064	409.8472
92	91	140	0.481383	422.5225
93	92	144	0.486702	435.1898
94	93	145	0.492021	447.8516
95	94	146	0.49734	460.5099
96	95	150	0.50266	473.1672
97	96	150	0.507979	485.8255
98	97	157	0.513298	498.4873
99	98	160	0.518617	511.1547
100	99	160	0.523936	523.8299
101	100	160	0.529255	536.5154
102	101	166	0.534574	549.2133
103	102	167	0.539894	561.9259
104	103	170	0.545213	574.6556
105	104	172	0.550532	587.4048
106	105	178	0.555851	600.1757
107	106	190	0.56117	612.9708
108	107	196	0.566489	625.7925
109	108	197	0.571809	638.6433
110	109	200	0.577128	651.5256
111	110	200	0.582447	664.4421
112	111	201	0.587766	677.3952
113	112	210	0.593085	690.3877
114	113	210	0.598404	703.4221
115	114	213	0.603723	716.5013

116	115	217	0.609043	729.6281
117	116	220	0.614362	742.8054
118	117	222	0.619681	756.036
119	118	225	0.625	769.323
120	119	230	0.630319	782.6696
121	120	230	0.635638	796.0789
122	121	230	0.640957	809.5542
123	122	240	0.646277	823.0989
124	123	245	0.651596	836.7166
125	124	250	0.656915	850.4108
126	125	260	0.662234	864.1853
127	126	260	0.667553	878.0439
128	127	262	0.672872	891.9908
129	128	280	0.678191	906.03
130	129	300	0.683511	920.166
131	130	310	0.68883	934.4032
132	131	337	0.694149	948.7464
133	132	338	0.699468	963.2004
134	133	350	0.704787	977.7705
135	134	360	0.710106	992.4619
136	135	370	0.715426	1007.28
137	136	370	0.720745	1022.232
138	137	370	0.726064	1037.322
139	138	370	0.731383	1052.558
140	139	379	0.736702	1067.946
141	140	386	0.742021	1083.494
142	141	390	0.74734	1099.209
143	142	419	0.75266	1115.1

144	143	420	0.757979	1131.174
145	144	451	0.763298	1147.441
146	145	465	0.768617	1163.91
147	146	472	0.773936	1180.592
148	147	501	0.779255	1197.497
149	148	510	0.784574	1214.637
150	149	512	0.789894	1232.024
151	150	538	0.795213	1249.672
152	151	541	0.800532	1267.595
153	152	565	0.805851	1285.808
154	153	572	0.81117	1304.327
155	154	620	0.816489	1323.171
156	155	621	0.821809	1342.358
157	156	636	0.827128	1361.91
158	157	706	0.832447	1381.85
159	158	724	0.837766	1402.201
160	159	781	0.843085	1422.991
161	160	792	0.848404	1444.251
162	161	795	0.853723	1466.012
163	162	829	0.859043	1488.312
164	163	830	0.864362	1511.19
165	164	830	0.869681	1534.691
166	165	874	0.875	1558.866
167	166	990	0.880319	1583.77
168	167	1010	0.885638	1609.468
169	168	1010	0.890957	1636.031
170	169	1030	0.896277	1663.543

171	170	1250	0.901596	1692.098
172	171	1340	0.906915	1721.808
173	172	1490	0.912234	1752.8
174	173	1502	0.917553	1785.227
175	174	1502	0.922872	1819.269
176	175	1507	0.928191	1855.146
177	176	1576	0.933511	1893.123
178	177	1597	0.93883	1933.53
179	178	1635	0.944149	1976.785
180	179	1720	0.949468	2023.423
181	180	2180	0.954787	2074.153
182	181	2420	0.960106	2129.939
183	182	2580	0.965426	2192.14
184	183	2820	0.970745	2262.775
185	184	3200	0.976064	2345.029
186	185	4350	0.981383	2444.41
187	186	4440	0.986702	2571.817
188	187	5320	0.992021	2754.545
189	188	8150	0.99734	3112.579
190	mean	466.8385638		
191	stdev	949.3003352		

Calculation of probability values (cell C2):

=(A2-0.5)/$A$189

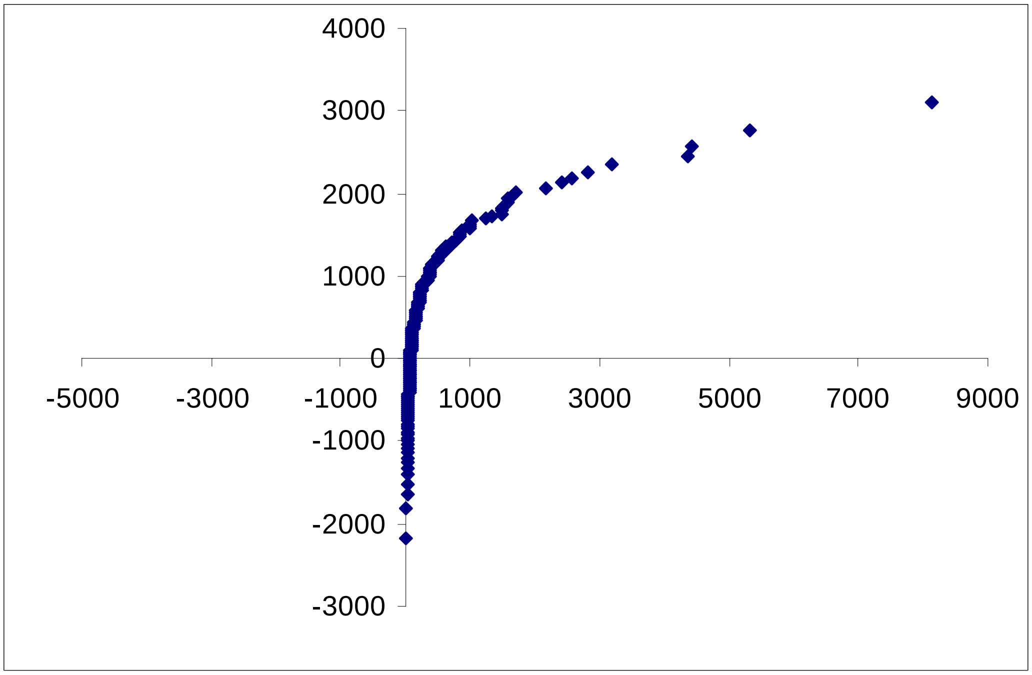
The total number of observations is in cell A189 (the highest rank)

The Z scores are calculated (cell D2):

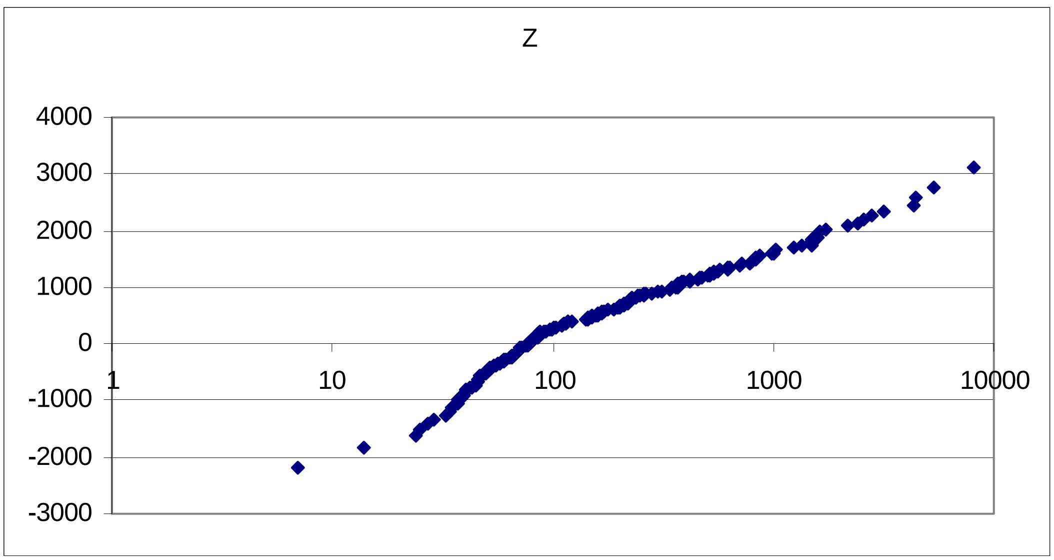
=NORMINV(C2,$B$190,$B$191)

The mean value is in B190 and the standard deviation is in cell B191.

Plot of data without transformations:



Plotting with the x-axis in log space, showing a much better bit to a straight line:



Comparisons of Two Sets of Data using Excel

One of the most common statistical tests is to compare two sets of data. Excel can be used for some basic tests, using t-tests.

Paired Tests:

If data is collected in pairs, many confounding factors are hopefully eliminated, as it is assumed that similar unmeasured factors are affecting both sets of data in a similar manner. Paired sampling is therefore recommended, if possible, although seldom can it be assumed that all confusion is eliminated! Paired sampling usually is associated with treatment units, where simultaneously influent and effluent samples are taken, for example. The following example shows how the basic paired (dependent) t-tests. For the t-test to be valid, the data must be normally distributed and the two data sets must have the same standard deviations. If not, either transformations can be used to obtain normal data (usually log transformations), or a non-parametric test should be used.

station A	station B
23	34
45	65
67	54
38	42
58	63
85	75
104	85
56	75
34	46
17	25

paired data above

t-Test: Paired Two Sample for Means

	Variable 1	*Variable 2*
Mean	52.7	56.4
Variance	744.4555556	381.8222222
Observations	10	10
Pearson Correlation	0.887012071	
Hypothesized Mean Difference	0	
df	9	
t Stat	-0.870996873	
P(T<=t) one-tail	0.203194421	
t Critical one-tail	1.833113856	
P(T<=t) two-tail	0.406388842	
t Critical two-tail	2.262158887	

Again, the summary table shows the column statistics for each of the sampling sites (mean, variance, and number of observations), and the statistical tests for differences. Normally, a p value of 0.05, or less, is usually used to signify significant differences in the data sets. If the p value is not (as in this example), there is not sufficient data to show that they are different (the means are close together for the variance observed and many more data may be needed to be confident that a difference exists). It is not proper to say that they are from the same population if the p is large. Excel also shows p values for one-tail and for two-tail tests. A one-tailed test is applied if one of the two sets of data is assumed to be larger than the other before the test is conducted. A two-tailed test is used if only a difference is to be examined, with no prior hypothesis that a specific set is larger than the other. In this example, neither case resulted in a significant difference, requiring additional data. If a one-tail test is used ("easier" to prove a significant difference, as the resulting p value is smaller than the calculated two-tailed p value), this must be clearly stated as part of the experimental design. This would be an obvious hypothesis for a treatment system when the effluent is hypothesized to have lower concentrations than the untreated influent.

Independent Tests:

The following example is for an independent t-test, where the data was not collected as pairs. This would occur for seasonal samples for example, where different times are associated with the samples. Again, in this example, not enough samples have been collected to say they are from different populations with a 95% confidence.

station A	station B
23	43
45	87
72	45
4	67
79	24
32	52
	79
	34
	23
	17

t-Test: Two-Sample Assuming Independent Observations

	Variable 1	*Variable 2*
Mean	42.5	47.1
Variance	836.3	582.5444444
Observations	6	10
Hypothesized Mean Difference	0	
df	9	
t Stat	-0.327207199	
P(T<=t) one-tail	0.375496691	
t Critical one-tail	1.833113856	
P(T<=t) two-tail	0.750993382	
t Critical two-tail	2.262158887	

Example of ANOVA using Excel

ANOVA can be used to compare data from different sites, as in the following example. This is a one-way ANOVA that is comparing the variability within each site to the variability between the sites. This is an excellent tool to supplement grouped box and whisker plots that display the data graphically. The following example shows three to six replicate values from each of 5 sites. ANOVA requires that the replicated values are normally distributed, so probability plots should be prepared. One probability plot showing all five sets of data would be especially informative. If the probability lines are parallel, they would also have similar variabilities, another requirement of ANOVA.

site a	site b	site c	site d	site e
78	43	153	14	12
45	79	87	53	9
63	54	245	42	34
54		432	64	14
24		43	23	
		164		

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	5	264	52.8	407.7
Column 2	3	176	58.66667	340.3333
Column 3	6	1124	187.3333	19161.87
Column 4	5	196	39.2	427.7
Column 5	4	69	17.25	128.9167

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	98255.39	4	24563.85	4.411859	0.011642	2.927749
Within Groups	100218.4	18	5567.686			
Total	198473.7	22				

This ANOVA analysis from Excel summarizes the data from each column above the analysis of variance table. The P-value needs to be smaller than the critical values (usually considered to be 0.05). The F critical value shown on the ANOVA table is the F value that would result in a p-value equal to 0.05, so you want a calculated F value to be greater. The F is the ratio of the mean sum of squares (MS) of the "between group" and "within groups" values. The mean sum of squares is the sum of squares (SS) values divided by the degrees of freedom (df).

In this case, the p-value is 0.012, much smaller than 0.05, so at least one site is significantly different from the other sites. Of course, this now begs the question of which one(s) are different from the others? A graphical grouped box and whiskers plot helps evaluate this. In addition, some statistical packages offer a Bonferroni t-test. This is simply a set of t-tests where each site is compared to each other site individually. In many cases, this will help distinguish the important groups, but it also usually results in some ambiguity, especially for many sites, and/or for 2-way ANOVAs. Also, since these are t-tests, the data must meet the t-test requirements (normally distributed, with each group having similar standard deviations), as does ANOVA. Transformations (usually using logs) may be helpful, then the ANOVA (and further tests) are conducted on the log values.

Example Regression Analysis using Excel

order of data collection	x axis	y axis	1st order predicted value	1 st order residuals	2nd order predicted value	2nd order residuals
3	1	45	199.396	154.396	113.0488	68.0488
7	3	73	201.448	128.448	117.6392	44.6392
2	7	105	205.552	100.552	126.7912	21.7912
8	34	180	233.254	53.254	187.5628	7.5628
9	75	253	275.32	22.32	276.5	23.5
1	105	573	306.1	-266.9	339.02	-233.98
5	368	733	575.938	-157.062	794.6412	61.6412
6	533	958	745.228	-212.772	995.7432	37.7432
4	1094	1143	1320.814	177.814	1190.747	47.7468
			sum:	0.05		78.6932

Conduct ANOVA for the regression. Things to consider:
• Want a good R^2 values, but it doesn't end there (0.85 here, pretty good).
• Examine statistical significance of the regression (Significance F), want it to be <0.05 (0.00057 here, excellent)
• Examine statistical significance of intercept and X Variable 1 (P-value), want them to be <0.05 (0.03 and 0.00057 here, fine). If the intercept is not significant, then eliminate it from the equation (force the equation thru zero) and redo the regression, only using the other (slope) terms.
• Examine the 95% range of the coefficients. If zero is in the range, then question the need for the term.

Regression Statistics	
Multiple R	0.913599785
R Square	0.834664568
Adjusted R Square	0.81104522
Standard Error	179.9109753
Observations	9

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1143824.509	1143824.5	35.33817	0.000573169
Residual	7	226575.7131	32367.959		
Total	8	1370400.222			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	198.3675387	73.54501117	2.6972263	0.030762	24.46134607	372.27	24.46	372.27
X Variable 1	1.025987456	0.172591741	5.9445918	0.000573	0.617873131	1.43	0.61	1.434

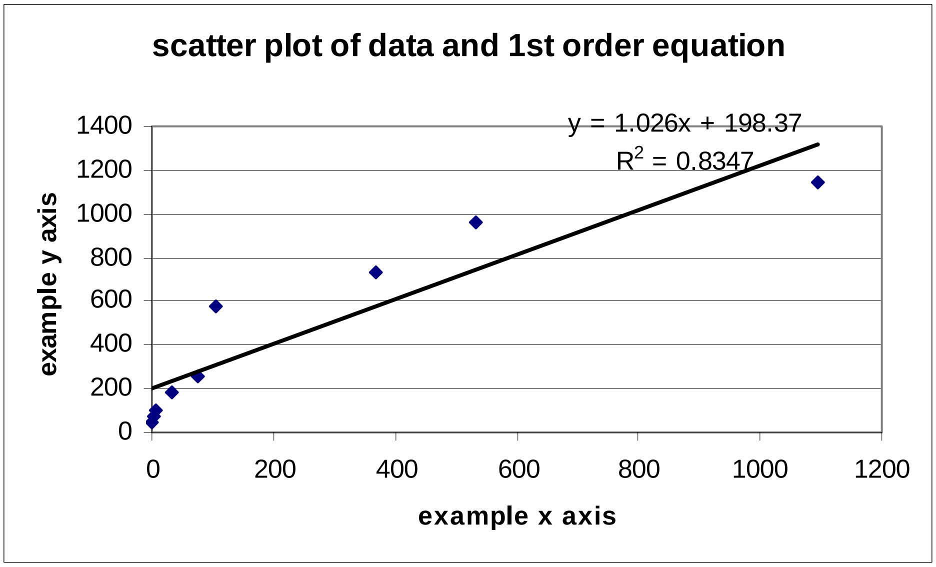
RESIDUAL OUTPUT

Observation	*Predicted Y*	*Residuals*
1	199.3935261	-154.3935261
2	201.445501	-128.445501
3	205.5494509	-100.5494509
4	233.2511122	-53.25111217
5	275.3165979	-22.31659786
6	306.0962215	266.9037785
7	575.9309224	157.0690776
8	745.2188526	212.7811474
9	1320.797815	-177.7978154

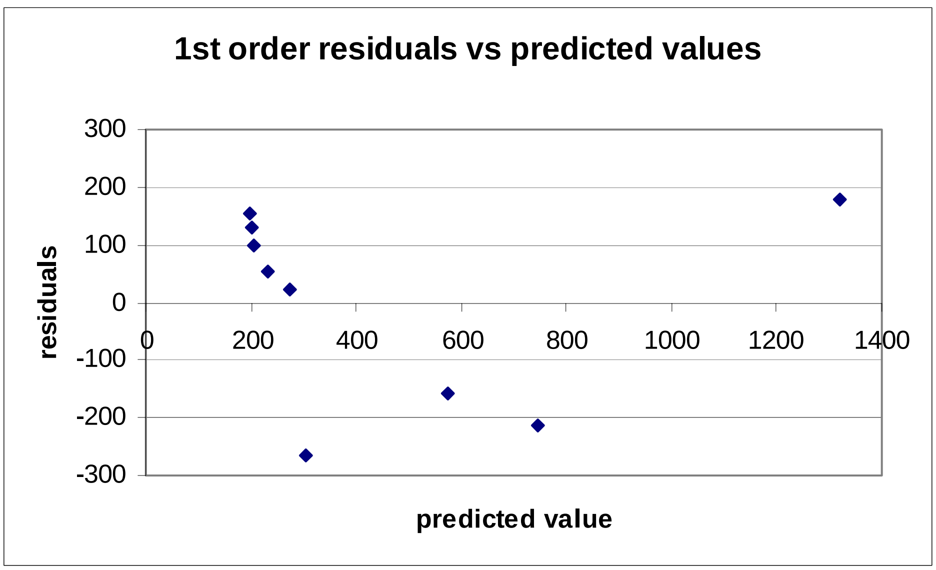
PROBABILITY OUTPUT

Percentile	*Y*
5.555556	45
16.66667	73
27.77778	105
38.88889	180
50	253
61.11111	573
72.22222	733
83.33333	958
94.44444	1143

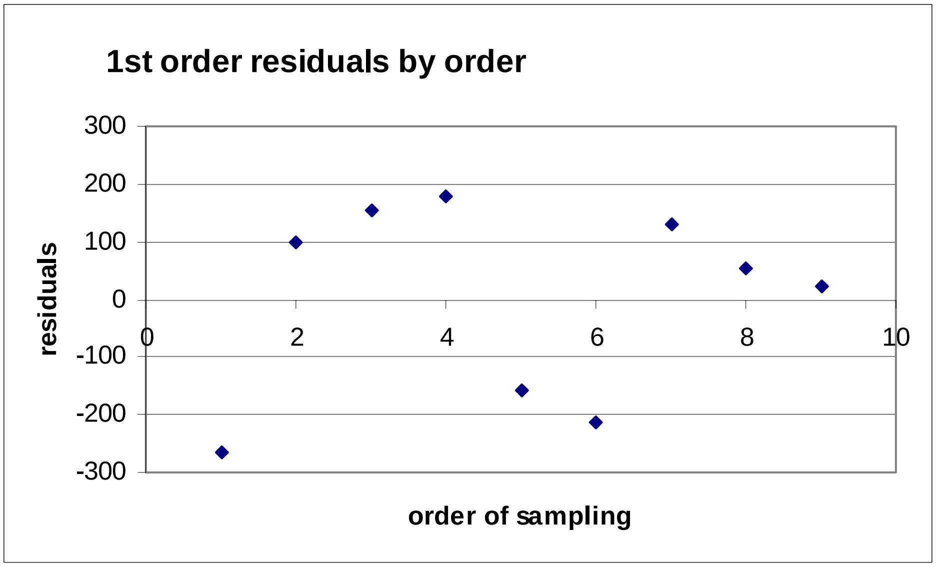
Then plot the data and the regression equation. Does it look "good"? In this case, there appears to be a bowing of the data compared to the first order polynomial regression:



Plot the residuals to see if they form an undesired pattern. Want to be a random band centered about the zero residual value. In this case, they seem to have a distinct bow pattern (except for one data point). Therefore, consider higher order equation.

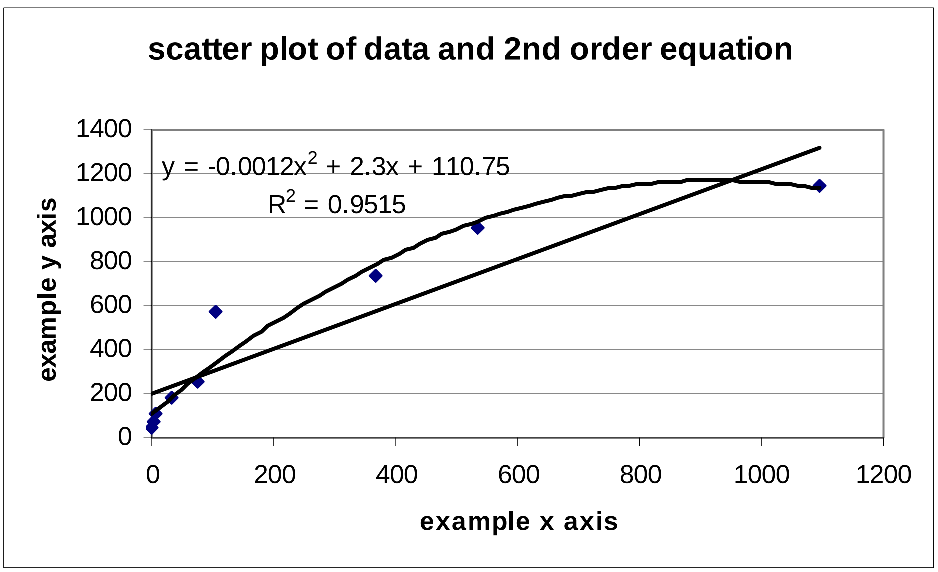


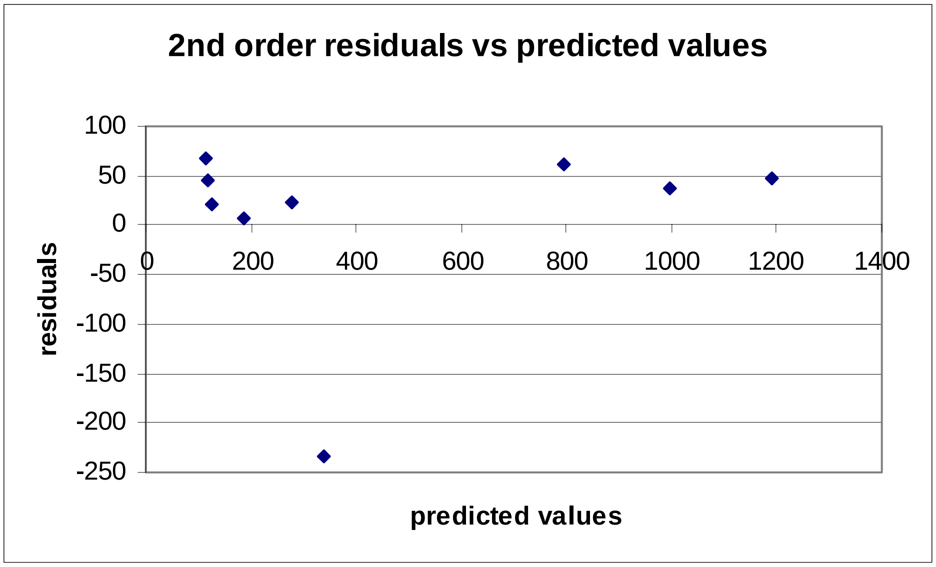
Residuals by order. Want random pattern, with no obvious carryover or serial correlation between observations. This pattern looks OK.

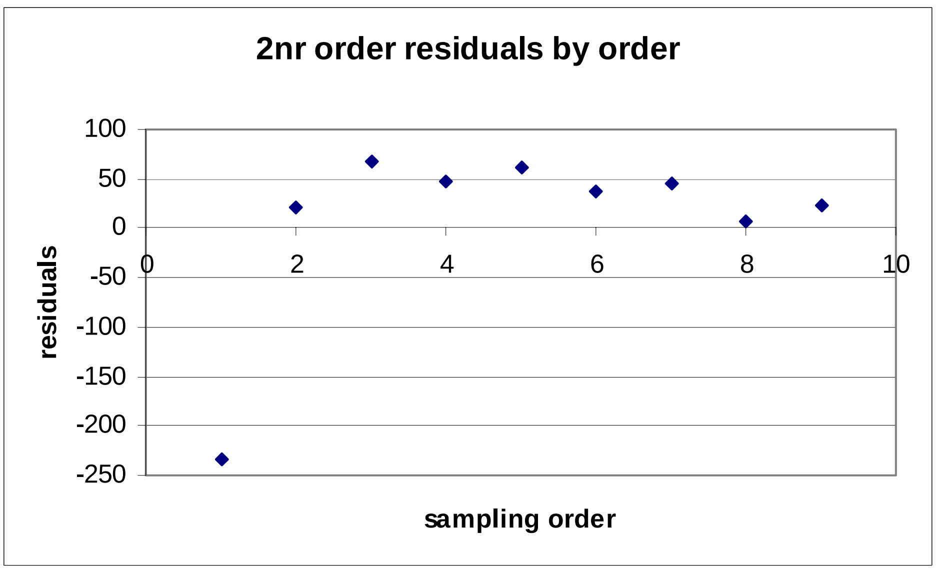


Also prepare probability plot of residuals to show random nature of the values (approximate straight line) (plot not shown).

The second order equation was then evaluated (ANOVA not shown). The data plot and regression line, plus residuals is shown below and look much better.







However, this regression example is still flawed. The X values are not evenly distributed; they indicate a bunching of low values (6 values are <200, while only 3 are between 200 and 1100). This is a common problem with many measurements where negative values are not possible and there is no physical limit to high values. In this case, log-transforming the values may be a suitable solution, and the regression repeated using the log values in the regression instead of the actual values.

Other Statistical Tests Available in Excel

The above examples only show a few of the many statistical tests. With the "Data Analysis" selection (after adding the "Analysis ToolPak" under Add-Ins) under "Tools", a long list of statistical tests is available, including:

ANOVA (single factor, two-factor with replication, and two-factor without replication)
Correlation

Covariance
Descriptive Statistics
Exponential Smoothing
F-Test two-sample for variances
Fourier analyses
Histogram
Moving average
Random number generator
Rank and percentile
Regression
Sampling
t-test (paired two sample for means, two-sample assuming equal variances, and two-sampling assuming unequal variances)
z-test (two sample for means)

In addition to these, there are a number of low-cost statistical packages that can also be added to Excel for selected specialized analyses. However, it may be worthwhile to invest in a complete statistical package. While these can be expensive, their capabilities, especially when dealing with large datasets, can be important when a larger variety of tests needs to be considered. As an example nonparametric analyses, exploratory data analyses, extended graphical capabilities, and multivariate analyses, are all important tools when conducting environmental research.

Wilcoxon Rank-Sum Test

Attached tables from Lehmann, E.L. *Nonparametrics; Statistical Methods based on Ranks*. McGraw-Hill 1975 and from Conover, W.J. *Practical Nonparametric Statistics*, 2^{nd} Edition. John Wiley and Sons, 1980.

The Wilcoxon Rank-Sum test is suitable to compare two sets of independent data, with few restrictions. The two data sets, however, each have to be symmetrical. You should use the Sign Test when comparing paired data. These tests should only be used if more powerful tests cannot be used. Specifically, these tests are useful when non-detects, or over-range, data are present. These non-parametric tests are based solely based on the ranks of the observed data, not their values. Multiple non detects (and over-range) values in any data set can be dealt with by calculating the average ranks:

Example of handing ties using average ranks:

Ranked Value	Rank	Rank (with averages for ties)
<1	1	2 =(1+2+3)/3
<1	2	2
<1	3	2
5.2	4	4
30.5	5	5
29.2	6	6
161.6	7	7
344.8	8	8
488.2	9	9
>2419.2	10	10.5 = (10+11)/2
>2419.2	11	10.5
Sum of ranks:	66	66

The Wilcoxon Rank-Sum test is illustrated using the following *E. coli* data:

Roof - Birds	Roof – no birds
145.5	<1
461.1	30.5
18.7	2
1413.6	5.2
410.6	344.8
>2419.2	161.6
>2419.2	29.2
2	<1
<1	>2419.2
517.2	6.3
	2

The first step is to sort all of the observations together (but remember which data belong to which data set):

Rank	Tied Ranks	Roof - Birds	Roof – no birds
1	2	<1	
2	2		<1
3	2		<1
4	5		2
5	5		2
6	5	2	
7	7		5.2
8	8		6.3
9	9	18.7	
10	10		29.2
11	11		30.5
12	12	145.5	
13	13		161.6
14	14		344.8
15	15	410.6	
16	16	461.1	
17	17	517.2	
18	18	1413.6	
19	20	>2419.2	
20	20	>2419.2	
21	20		>2419.2
Number of observations		10	11
Sum of ranks		134	97

The test statistic "a" is then calculated:

$$a = W_{xy} = W_r - \frac{1}{2}n(n+1)$$

Where W_r is the largest sum of ranks (134 in this example) and n is the number of observations in the set having the largest sum of ranks (10 here). This tests that W_r, the largest sum of ranks, is < than the other sum.

$$W_{xy} = W_r - \frac{1}{2}n(n+1) = 134 - \frac{1}{2}(10)(11) = 79$$

In the tables, k_1 is the number of observations in the smallest set (10 here) while k_2 is the number of observations in the larger set (11 here).

In an example of using the table, assume that $k_1 = 3$ and $k_2 = 7$ and that W_{xy} was calculated to be 16. In this example, P is seen to be 0.9083. The one-tail test is therefore P = 91% (marginally significant that the largest median value is greater than the smaller median value of the other data set, with an α= 1-0.983 = 0.092). For a two-tailed test (testing for a difference, with no prior knowledge of one expected to be larger than the other), alpha = 2(0.092) = 0.18 and P = 1-0.18 = 0.82, indicating that there was not a significant difference in the two medians.

One should also check the sum of rank calculations:

$$\sum ranks = \frac{N(N+1)}{2} = \frac{21(22)}{2} = 231$$

where N = n + m = 21, the number of observations in both sets combined.

The Conover tables (attached) can be used for data sets having as many as 20 elements each. However, they only show the critical test statistic values associated with the 0.001, 0.005, 0.01, 0.05, and 0.10 p values. These are also for 2-tailed tests (where one is testing for a difference, but it is not known in advance which is larger). For 1-tailed tests (when the larger one is a-priori known, then the p values should be halved). In the above example, the number of observations are 10 and 11, and the W_{xy} test statistic is 79. This corresponds closely to a p of 0.01 for a 2-tailed test, indicating a very high probability that they are different (significant results are usually indicated if the p is $\leq$ 0.05).

Conover (1980) Appendix A7 Table for Mann-Whitney Test statistic (same as Wilcoxon Rank-Sum test statistic):

2-tailed (½ p for 1-tail when want to test > or < specifically) use 2-tail tests if just for differences

{Same as Wilcoxon Rank-Sum test}

Table A7 Quantiles of the Mann–Whitney Test Statistic[a]

n	p	m = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	.001	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	.005	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4
	.01	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	5	5
	.025	3	3	3	3	3	3	4	4	4	5	5	5	5	5	5	6	6	6	6
	.05	3	3	3	4	4	4	5	5	5	5	6	6	7	7	7	7	8	8	8
	.10	3	4	4	5	5	5	6	6	7	7	8	8	8	9	9	10	10	11	11
3	.001	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7
	.005	6	6	6	6	6	6	6	7	7	7	8	8	8	9	9	9	9	10	10
	.01	6	6	6	6	6	7	7	8	8	8	9	9	9	10	10	11	11	11	12
	.025	6	6	6	7	8	8	9	9	10	10	11	11	12	12	13	13	14	14	15
	.05	6	7	7	8	9	9	10	11	11	12	12	13	14	14	15	16	16	17	18
	.10	7	8	8	9	10	11	12	12	13	14	15	16	17	17	18	19	20	21	22
4	.001	10	10	10	10	10	10	10	10	11	11	11	12	12	12	13	13	14	14	14
	.005	10	10	10	10	11	11	12	12	13	13	14	14	15	16	16	17	17	18	19
	.01	10	10	10	11	12	12	13	14	14	15	16	16	17	18	18	19	20	20	21
	.025	10	10	11	12	13	14	15	15	16	17	18	19	20	21	22	22	23	24	25
	.05	10	11	12	13	14	15	16	17	18	19	20	21	22	23	25	26	27	28	29
	.10	11	12	14	15	16	17	18	20	21	22	23	24	26	27	28	29	31	32	33
5	.001	15	15	15	15	15	15	16	17	17	18	18	19	19	20	21	21	22	23	23
	.005	15	15	15	16	17	17	18	19	20	21	22	23	23	24	25	26	27	28	29
	.01	15	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
	.025	15	16	17	18	19	21	22	23	24	25	27	28	29	30	31	33	34	35	36
	.05	16	17	18	20	21	22	24	25	27	28	29	31	32	34	35	36	38	39	41
	.10	17	18	20	21	23	24	26	28	29	31	33	34	36	38	39	41	43	44	46

6	.001	21	21	21	21	21	21	23	24	25	26	26	27	28	29	30	31	32	33	34
	.005	21	21	22	23	24	25	26	27	28	29	31	32	33	34	35	37	38	39	40
	.01	21	21	23	24	25	26	28	29	30	31	33	34	35	37	38	40	41	42	44
	.025	21	23	24	25	27	28	30	32	33	35	36	38	39	41	43	44	46	47	49
	.05	22	24	25	27	29	30	32	34	36	38	39	41	43	45	47	48	50	52	54
	.10	23	25	27	29	31	33	35	37	39	41	43	45	47	49	51	53	56	58	60
7	.001	28	28	28	28	29	30	31	32	34	35	36	37	38	39	40	42	43	44	45
	.005	28	28	29	30	32	33	35	36	38	39	41	42	44	45	47	48	50	51	53
	.01	28	29	30	32	33	35	36	38	40	41	43	45	46	48	50	52	53	55	57
	.025	28	30	32	34	35	37	39	41	43	45	47	49	51	53	55	57	59	61	63
	.05	29	31	33	35	37	40	42	44	46	48	50	53	55	57	59	62	64	66	68
	.10	30	33	35	37	40	42	45	47	50	52	55	57	60	62	65	67	70	72	75
8	.001	36	36	36	37	38	39	41	42	43	45	46	48	49	51	52	54	55	57	58
	.005	36	36	38	39	41	43	44	46	48	50	52	54	55	57	59	61	63	65	67
	.01	36	37	39	41	43	44	46	48	50	52	54	56	59	61	63	65	67	69	71
	.025	37	39	41	43	45	47	50	52	54	56	59	61	63	66	68	71	73	75	78
	.05	38	40	42	45	47	50	52	55	57	60	63	65	68	70	73	76	78	81	84
	.10	39	42	44	47	50	53	56	59	61	64	67	70	73	76	79	82	85	88	91
9	.001	45	45	45	47	48	49	51	53	54	56	58	60	61	63	65	67	69	71	72
	.005	45	46	47	49	51	53	55	57	59	62	64	66	68	70	73	75	77	79	82
	.01	45	47	49	51	53	55	57	60	62	64	67	69	72	74	77	79	82	84	86
	.025	46	48	50	53	56	58	61	63	66	69	72	74	77	80	83	85	88	91	94
	.05	47	50	52	55	58	61	64	67	70	73	76	79	82	85	88	91	94	97	100
	.10	48	51	55	58	61	64	68	71	74	77	81	84	87	91	94	98	101	104	108
10	.001	55	55	56	57	59	61	62	64	66	68	70	73	75	77	79	81	83	85	88
	.005	55	56	58	60	62	65	67	69	72	74	77	80	82	85	87	90	93	95	98
	.01	55	57	59	62	64	67	69	72	75	78	80	83	86	89	92	94	97	100	103
	.025	56	59	61	64	67	70	73	76	79	82	85	89	92	95	98	101	104	108	111
	.05	57	60	63	67	70	73	76	80	83	87	90	93	97	100	104	107	111	114	118
	.10	59	62	66	69	73	77	80	84	88	92	95	99	103	107	110	114	118	122	126

Table A7 (Continued)

n	p	m = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	.001	66	66	67	69	71	73	75	77	79	82	84	87	89	91	94	96	99	101	104
	.005	66	67	69	72	74	77	80	83	85	88	91	94	97	100	103	106	109	112	115
11	.01	66	68	71	74	76	79	82	85	89	92	95	98	101	104	108	111	114	117	120
	.025	67	70	73	76	80	83	86	90	93	97	100	104	107	111	114	118	122	125	129
	.05	68	72	75	79	83	86	90	94	98	101	105	109	113	117	121	124	128	132	136
	.10	70	74	78	82	86	90	94	98	103	107	111	115	119	124	128	132	136	140	145
	.001	78	78	79	81	83	86	88	91	93	96	98	102	104	106	110	113	116	118	121
	.005	78	80	82	85	88	91	94	97	100	103	106	110	113	116	120	123	126	130	133
12	.01	78	81	84	87	90	93	96	100	103	107	110	114	117	121	125	128	132	135	139
	.025	80	83	86	90	93	97	101	105	108	112	116	120	124	128	132	136	140	144	148
	.05	81	84	88	92	96	100	105	109	111	117	121	126	130	134	139	143	147	151	156
	.10	83	87	91	96	100	105	109	114	118	123	128	132	137	142	146	151	156	160	165
	.001	91	91	93	95	97	100	103	106	109	112	115	118	121	124	127	130	134	137	140
	.005	91	93	95	99	102	105	109	112	116	119	123	126	130	134	137	141	145	149	152
13	.01	92	94	97	101	104	108	112	115	119	123	127	131	135	139	143	147	151	155	159
	.025	93	96	100	104	108	112	116	120	125	129	133	137	142	146	151	155	159	164	168
	.05	94	98	102	107	111	116	120	125	129	134	139	143	148	153	157	162	167	172	176
	.10	96	101	105	110	115	120	125	130	135	140	145	150	155	160	166	171	176	181	186
	.001	105	105	107	109	112	115	118	121	125	128	131	135	138	142	145	149	152	156	160
	.005	105	107	110	113	117	121	124	128	132	136	140	144	148	152	156	160	164	169	173
14	.01	106	108	112	116	119	123	128	132	136	140	144	149	153	157	162	166	171	175	179
	.025	107	111	115	119	123	128	132	137	142	146	151	156	161	165	170	175	180	184	189
	.05	109	113	117	122	127	132	137	142	147	152	157	162	167	172	177	183	188	193	198
	.10	110	116	121	126	131	137	142	147	153	158	164	169	175	180	186	191	197	203	208

	.001	120	120	122	125	128	133	135	138	142	145	149	153	157	161	164	168	172	176	180
	.005	120	123	126	129	133	137	141	145	150	154	158	163	167	172	176	181	185	190	194
15	.01	121	124	128	132	136	140	145	149	154	158	163	168	172	177	182	187	191	196	201
	.025	122	126	131	135	140	145	150	155	160	165	170	175	180	185	191	196	201	206	211
	.05	124	128	133	139	144	149	154	160	165	171	176	182	187	193	198	204	209	215	221
	.10	126	131	137	143	148	154	160	166	172	178	184	189	195	201	207	213	219	225	231
	.001	136	136	139	142	145	148	152	156	160	164	168	172	176	180	185	189	193	197	202
	.005	136	139	142	146	150	155	159	164	168	173	178	182	187	192	197	202	207	211	216
16	.01	137	140	144	149	153	158	163	168	173	178	183	188	193	198	203	208	213	219	224
	.025	138	143	148	152	158	163	168	174	179	184	190	196	201	207	212	218	223	229	235
	.05	140	145	151	156	162	167	173	179	185	191	197	202	208	214	220	226	232	238	244
	.10	142	148	154	160	166	173	179	185	191	198	204	211	217	223	230	236	243	249	256
	.001	153	154	156	159	163	167	171	175	179	183	188	192	197	201	206	211	215	220	224
	.005	153	156	160	164	169	173	178	183	188	193	198	203	208	214	219	224	229	235	240
17	.01	154	158	162	167	172	177	182	187	192	198	203	209	214	220	225	231	236	242	247
	.025	156	160	165	171	176	182	188	193	199	205	211	217	223	229	235	241	247	253	259
	.05	157	163	169	174	180	187	193	199	205	211	218	224	231	237	243	250	256	263	269
	.10	160	166	172	179	185	192	199	206	212	219	226	233	239	246	253	260	267	274	281
	.001	171	172	175	178	182	186	190	195	199	204	209	214	218	223	228	233	238	243	248
	.005	171	174	178	183	188	193	198	203	209	214	219	225	230	236	242	247	253	259	264
18	.01	172	176	181	186	191	196	202	208	213	219	225	231	237	242	248	254	260	266	272
	.025	174	179	184	190	196	202	208	214	220	227	233	239	246	252	258	265	271	278	284
	.05	176	181	188	194	200	207	213	220	227	233	240	247	254	260	267	274	281	288	295
	.10	178	185	192	199	206	213	220	227	234	241	249	256	263	270	278	285	292	300	307
	.001	190	191	194	198	202	206	211	216	220	225	231	236	241	246	251	257	262	268	273
	.005	191	194	198	203	208	213	219	224	230	236	242	248	254	260	265	272	278	284	290
19	.01	192	195	200	206	211	217	223	229	235	241	247	254	260	266	273	279	285	292	298
	.025	193	198	204	210	216	223	229	236	243	249	256	263	269	276	283	290	297	304	310
	.05	195	201	208	214	221	228	235	242	249	256	263	271	278	285	292	300	307	314	321
	.10	198	205	212	219	227	234	242	249	257	264	272	280	288	295	303	311	319	326	334

Table A7 (CONTINUED)

n	p	$m=2$	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	.001	210	211	214	218	223	227	232	237	243	248	253	259	265	270	276	281	287	293	299
	.005	211	214	219	224	229	235	241	247	253	259	265	271	278	284	290	297	303	310	316
	.01	212	216	221	227	233	239	245	251	258	264	271	278	284	291	298	304	311	318	325
20	.025	213	219	225	231	238	245	251	259	266	273	280	287	294	301	309	316	323	330	338
	.05	215	222	229	236	243	250	258	265	273	280	288	295	303	311	318	326	334	341	349
	.10	218	226	233	241	249	257	265	273	281	289	297	305	313	321	330	338	346	354	362

For n or m greater than 20, the pth quantile w_p of the Mann–Whitney test statistic may be approximated by

$$w_p = n(N+1)/2 + x_p\sqrt{nm(N+1)/12}$$

where x_p is the pth quantile of a standard normal random variable, obtained from Table A1, and where $N = m + n$.

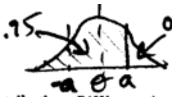
[a] The entries in this table are quantiles w_p of the Mann–Whitney test statistic T, given by Equation 5.1.1, for selected values of p. Note that $P(T < w_p) \leq p$. Upper quantiles may be found from the equation

$$w_p = n(n+m+1) - w_{1-p}$$

Critical regions correspond to values of T less than (or greater than) but not equal to the appropriate quantile.

The following Table B is from Lehmann (1975) for Wilcoxon Rank-Sum test statistics:

one-tail test reads

0.95  0.05 if $P(W_{XY} \le a) = 0.05$, $\alpha = 0.05$ $P=95$

if two-sided $\alpha = 0.10$ $P=90$

TABLE B. Wilcoxon rank-sum distribution: $P(W_{XY} \leqslant a)$

See p 99

k_1	a	$k_2 = 3$	$k_2 = 4$	$k_2 = 5$	$k_2 = 6$	$k_2 = 7$	$k_2 = 8$	$k_2 = 9$	$k_2 = 10$	$k_2 = 11$	$k_2 = 12$
3	0	.0500	.0286	.0179	.0119	.0083	.0061	.0045	.0035	.0027	.0022
	1	.1000	.0571	.0357	.0238	.0167	.0121	.0091	.0070	.0055	.0044
	2	.2000	.1143	.0714	.0476	.0333	.0242	.0182	.0140	.0110	.0088
	3	.3500	.2000	.1250	.0833	.0583	.0424	.0318	.0245	.0192	.0154
	4	.5000	.3143	.1964	.1310	.0917	.0667	.0500	.0385	.0302	.0242
	5	.6500	.4286	.2857	.1905	.1333	.0970	.0727	.0559	.0440	.0352
	6	.8000	.5714	.3929	.2738	.1917	.1394	.1045	.0804	.0632	.0505
	7	.9000	.6857	.5000	.3571	.2583	.1879	.1409	.1084	.0852	.0681
	8	.9500	.8000	.6071	.4524	.3333	.2485	.1864	.1434	.1126	.0901
	9	1.0000	.8857	.7143	.5476	.4167	.3152	.2409	.1853	.1456	.1165
	10		.9429	.8036	.6429	.5000	.3879	.3000	.2343	.1841	.1473
	11		.9714	.8750	.7262	.5833	.4606	.3636	.2867	.2280	.1824
	12		1.0000	.9286	.8095	.6667	.5394	.4318	.3462	.2775	.2242
	13			.9643	.8690	.7417	.6121	.5000	.4056	.3297	.2681
	14			.9821	.9167	.8083	.6848	.5682	.4685	.3846	.3165
	15			1.0000	.9524	.8667	.7515	.6364	.5315	.4423	.3670
	16				.9762	.9083	.8121	.7000	.5944	.5000	.4198
	17				.9881	.9417	.8606	.7591	.6538	.5577	.4725
	18				1.0000	.9667	.9030	.8136	.7133	.6154	.5275
4	0		.0143	.0079	.0048	.0030	.0020	.0014	.0010	.0007	.0005
	1		.0286	.0159	.0095	.0061	.0040	.0028	.0020	.0015	.0011
	2		.0571	.0317	.0190	.0121	.0081	.0056	.0040	.0029	.0022
	3		.1000	.0556	.0333	.0212	.0141	.0098	.0070	.0051	.0038
	4		.1714	.0952	.0571	.0364	.0242	.0168	.0120	.0088	.0066
	5		.2429	.1429	.0857	.0545	.0364	.0252	.0180	.0132	.0099
	6		.3429	.2063	.1286	.0818	.0545	.0378	.0270	.0198	.0148
	7		.4429	.2778	.1762	.1152	.0768	.0531	.0380	.0278	.0209
	8		.5571	.3651	.2381	.1576	.1071	.0741	.0529	.0388	.0291
	9		.6571	.4524	.3048	.2061	.1414	.0993	.0709	.0520	.0390
	10		.7571	.5476	.3810	.2636	.1838	.1301	.0939	.0689	.0516
	11		.8286	.6349	.4571	.3242	.2303	.1650	.1199	.0886	.0665
	12		.9000	.7222	.5429	.3939	.2848	.2070	.1518	.1128	.0852
	13		.9429	.7937	.6190	.4636	.3414	.2517	.1868	.1399	.1060
	14		.9714	.8571	.6952	.5364	.4040	.3021	.2268	.1714	.1308
	15		.9857	.9048	.7619	.6061	.4667	.3552	.2697	.2059	.1582
	16		1.0000	.9444	.8238	.6758	.5333	.4126	.3177	.2447	.1896
	17			.9683	.8714	.7364	.5960	.4699	.3666	.2857	.2231
	18			.9841	.9143	.7939	.6586	.5301	.4196	.3304	.2604
	19			.9921	.9429	.8424	.7152	.5874	.4725	.3766	.2995
	20			1.0000	.9667	.8848	.7697	.6448	.5275	.4256	.3418
	21				.9810	.9182	.8162	.6979	.5804	.4747	.3852
	22				.9905	.9455	.8586	.7483	.6334	.5253	.4308
	23				.9952	.9636	.8929	.7930	.6823	.5744	.4764
	24				1.0000	.9788	.9232	.8350	.7303	.6234	.5236

For two independent samples, test H0 (no diff) against a shift of the medians of the Ys to the right of the Xs (Ys are better than Xs)

tests that W_Y < the other sum

largest sum of ranks; n = # of observ. in "largest sum of ranks" set

$$W_{XY} = W_Y - \tfrac{1}{2}n(n+1)$$

example: $W_Y = 133$ $n = 10$ [P > 52.9%] that Ys > Xs

	X	Y
Σ ranks =	57	133
# of obs	9	10

$a = W_{XY} = 133 - \frac{1}{2}(10)(11) = 78$ ∴ off chart, use table C

$k_1 = 9$ $k_2 = 10$

check: 408

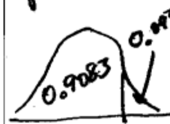
$\Sigma\Sigma$ ranks $= 57 + 133 = 190 \equiv \frac{N(N+1)}{2} = \frac{19(20)}{2} = 190$ where $N = m+n$

TABLE B. Wilcoxon rank-sum distribution: $P(W_{XY} \leqslant a)$ (Continued)

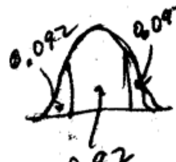
k_1	a	$k_2 = 5$	$k_2 = 6$	$k_2 = 7$	$k_2 = 8$	$k_2 = 9$	$k_2 = 10$
5	0	.0040	.0022	.0013	.0008	.0005	.0003
	1	.0079	.0043	.0025	.0016	.0010	.0007
	2	.0159	.0087	.0051	.0031	.0020	.0013
	3	.0278	.0152	.0088	.0054	.0035	.0023
	4	.0476	.0260	.0152	.0093	.0060	.0040
	5	.0754	.0411	.0240	.0148	.0095	.0063
	6	.1111	.0628	.0366	.0225	.0145	.0097
	7	.1548	.0887	.0530	.0326	.0210	.0140
	8	.2103	.1234	.0745	.0466	.0300	.0200
	9	.2738	.1645	.1010	.0637	.0415	.0276
	10	.3452	.2143	.1338	.0855	.0559	.0376
	11	.4206	.2684	.1717	.1111	.0734	.0496
	12	.5000	.3312	.2159	.1422	.0949	.0646
	13	.5794	.3961	.2652	.1772	.1199	.0823
	14	.6548	.4654	.3194	.2176	.1489	.1032
	15	.7262	.5346	.3775	.2618	.1818	.1272
	16	.7897	.6039	.4381	.3108	.2188	.1548
	17	.8452	.6688	.5000	.3621	.2592	.1855
	18	.8889	.7316	.5619	.4165	.3032	.2198
	19	.9246	.7857	.6225	.4716	.3497	.2567
	20	.9524	.8355	.6806	.5284	.3986	.2970
	21	.9722	.8766	.7348	.5835	.4491	.3393
	22	.9841	.9113	.7841	.6379	.5000	.3839
	23	.9921	.9372	.8283	.6892	.5509	.4296
	24	.9960	.9589	.8662	.7382	.6014	.4765
	25	1.0000	.9740	.8990	.7824	.6503	.5235
6	0		.0011	.0006	.0003	.0002	.0001
	1		.0022	.0012	.0007	.0004	.0002
	2		.0043	.0023	.0013	.0008	.0005
	3		.0076	.0041	.0023	.0014	.0009
	4		.0130	.0070	.0040	.0024	.0015
	5		.0206	.0111	.0063	.0038	.0024
	6		.0325	.0175	.0100	.0060	.0037
	7		.0465	.0256	.0147	.0088	.0055
	8		.0660	.0367	.0213	.0128	.0080
	9		.0898	.0507	.0296	.0180	.0112
	10		.1201	.0688	.0406	.0248	.0156
	11		.1548	.0903	.0539	.0332	.0210
	12		.1970	.1171	.0709	.0440	.0280
	13		.2424	.1474	.0906	.0567	.0363
	14		.2944	.1830	.1142	.0723	.0467
	15		.3496	.2226	.1412	.0905	.0589
	16		.4091	.2669	.1725	.1119	.0736
	17		.4686	.3141	.2068	.1361	.0903
	18		.5314	.3654	.2454	.1638	.1099
	19		.5909	.4178	.2864	.1942	.1317
	20		.6504	.4726	.3310	.2280	.1566
	21		.7056	.5274	.3773	.2643	.1838
	22		.7576	.5822	.4259	.3035	.2139
	23		.8030	.6346	.4749	.3445	.2461
	24		.8452	.6859	.5251	.3878	.2811
	25		.8799	.7331	.5741	.4320	.3177
	26		.9102	.7774	.6227	.4773	.3564
	27		.9340	.8170	.6690	.5227	.3962
	28		.9535	.8526	.7136	.5680	.4374
	29		.9675	.8829	.7546	.6122	.4789
	30		.9794	.9097	.7932	.6555	.5211

k_1	a	$k_2 = 7$	$k_2 = 8$	$k_2 = 9$	$k_2 = 10$
7	0	.0003	.0002	.0001	.0001
	1	.0006	.0003	.0002	.0001
	2	.0012	.0006	.0003	.0002
	3	.0020	.0011	.0006	.0004
	4	.0035	.0019	.0010	.0006
	5	.0055	.0030	.0017	.0010
	6	.0087	.0047	.0026	.0015
	7	.0131	.0070	.0039	.0023
	8	.0189	.0103	.0058	.0034
	9	.0265	.0145	.0082	.0048
	10	.0364	.0200	.0115	.0068
	11	.0487	.0270	.0156	.0093
	12	.0641	.0361	.0209	.0125
	13	.0825	.0469	.0274	.0165
	14	.1043	.0603	.0356	.0215
	15	.1297	.0760	.0454	.0277
	16	.1588	.0946	.0571	.0351
	17	.1914	.1159	.0708	.0439
	18	.2279	.1405	.0869	.0544
	19	.2675	.1678	.1052	.0665
	20	.3100	.1984	.1261	.0806
	21	.3552	.2317	.1496	.0966
	22	.4024	.2679	.1755	.1148
	23	.4508	.3063	.2039	.1349
	24	.5000	.3472	.2349	.1574
	25	.5492	.3894	.2680	.1819
	26	.5976	.4333	.3032	.2087
	27	.6448	.4775	.3403	.2374
	28	.6900	.5225	.3788	.2681
	29	.7325	.5667	.4185	.3004
	30	.7721	.6106	.4591	.3345
	31	.8086	.6528	.5000	.3698
	32	.8412	.6937	.5409	.4063
	33	.8703	.7321	.5815	.4434
	34	.8957	.7683	.6212	.4811
	35	.9175	.8016	.6597	.5189

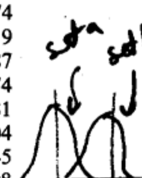
if $W_{YX} = 16$
$k_1 = 3$
$k_2 = 7$
$p = 0.9083$



one-tail test, $p \approx 91\%$ (marginally sign that largest unequal smaller)

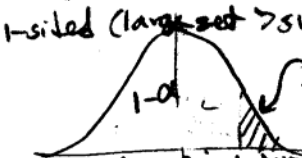


two-tail test
$p \approx 82\%$
not sign. difference
∴ same



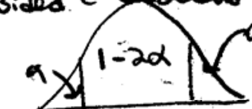
if test for $b > a$, then 1-sided

1-sided (largest set > smallest set)



∴ reject $Y > X$

2-sided ("different" sets)



if test for $b \neq a$, then 2-sided

k_1	a	$k_2 = 8$	$k_2 = 9$	$k_2 = 10$	k_1	a	$k_2 = 9$	$k_2 = 10$	k_1	a	$k_2 = 10$
8	0	.0001	.0000	.0000	9	0	.0000	.0000	10	0	.0000
	1	.0002	.0001	.0000		1	.0000	.0000		1	.0000
	2	.0003	.0002	.0001		2	.0001	.0000		2	.0000
	3	.0005	.0003	.0002		3	.0001	.0001		3	.0000
	4	.0009	.0005	.0003		4	.0002	.0001		4	.0001
	5	.0015	.0008	.0004		5	.0004	.0002		5	.0001
	6	.0023	.0012	.0007		6	.0006	.0003		6	.0002
	7	.0035	.0019	.0010		7	.0009	.0005		7	.0002
	8	.0052	.0028	.0015		8	.0014	.0007		8	.0004
	9	.0074	.0039	.0022		9	.0020	.0011		9	.0005
	10	.0103	.0056	.0031		10	.0028	.0015		10	.0008
	11	.0141	.0076	.0043		11	.0039	.0021		11	.0010
	12	.0190	.0103	.0058		12	.0053	.0028		12	.0014
	13	.0249	.0137	.0078		13	.0071	.0038		13	.0019
	14	.0325	.0180	.0103		14	.0094	.0051		14	.0026
	15	.0415	.0232	.0133		15	.0122	.0066		15	.0034
	16	.0524	.0296	.0171		16	.0157	.0086		16	.0045
	17	.0652	.0372	.0217		17	.0200	.0110		17	.0057
	18	.0803	.0464	.0273		18	.0252	.0140		18	.0073
	19	.0974	.0570	.0338		19	.0313	.0175		19	.0093
	20	.1172	.0694	.0416		20	.0385	.0217		20	.0116
	21	.1393	.0836	.0506		21	.0470	.0267		21	.0144
	22	.1641	.0998	.0610		22	.0567	.0326		22	.0177
	23	.1911	.1179	.0729		23	.0680	.0394		23	.0216
	24	.2209	.1383	.0864		24	.0807	.0474		24	.0262
	25	.2527	.1606	.1015		25	.0951	.0564		25	.0315
	26	.2869	.1852	.1185		26	.1112	.0667		26	.0376
	27	.3227	.2117	.1371		27	.1290	.0782		27	.0446
	28	.3605	.2404	.1577		28	.1487	.0912		28	.0526
	29	.3992	.2707	.1800		29	.1701	.1055		29	.0615
	30	.4392	.3029	.2041		30	.1933	.1214		30	.0716
	31	.4796	.3365	.2299		31	.2181	.1388		31	.0827
	32	.5204	.3715	.2574		32	.2447	.1577		32	.0952
	33	.5608	.4074	.2863		33	.2729	.1781		33	.1088
	34	.6008	.4442	.3167		34	.3024	.2001		34	.1237
	35	.6395	.4813	.3482		35	.3332	.2235		35	.1399
	36	.6773	.5187	.3809		36	.3652	.2483		36	.1575
	37	.7131	.5558	.4143		37	.3981	.2745		37	.1763
	38	.7473	.5926	.4484		38	.4317	.3019		38	.1965
	39	.7791	.6285	.4827		39	.4657	.3304		39	.2179
	40	.8089	.6635	.5173		40	.5000	.3598		40	.2406
						41	.5343	.3901		41	.2644
						42	.5683	.4211		42	.2894
						43	.6019	.4524		43	.3153
						44	.6348	.4841		44	.3421
						45	.6668	.5159		45	.3697
										46	.3980
										47	.4267
										48	.4559
										49	.4853
										50	.5147