FINAL REPORT

Water Quality Data-Mining, Data Analysis, and Trends Assessment

Report prepared by
Pinnacle Consulting Group
Division of
North Wind, Inc.

Authors: Dr. Dave Hargett, Holli Hargett, and Steve Springs

Submitted to
Saluda-Reedy Watershed Consortium

27 July 2005

This research project was funded
by a grant from the
V. Kann Rasmussen Foundation
to
Upstate Forever
on behalf of the
Saluda-Reedy Watershed Consortium

www.saludareedy.org

TABLE OF CONTENTS

LIST OF FIGURES

I. Project Objectives and Overview

A. Objectives

The Pinnacle Consulting Group (Pinnacle), a Division of North Wind, Inc., was tasked by the Saluda-Reedy Watershed (SRW) Consortium (SRWC) to conduct selected data-mining activities on the Consortium's water quality data-warehouse. Ongoing objectives of this data review include evaluation of historical water quality, sediment quality, and stream flows to better understand temporal and spatial trends in the watershed. Better understanding of long-term trends in water quality and flows, and of differences in various portions of the watershed, provide valuable insights with regard to water resource management needs.

This report presents the results of an initial review of key indicator data assessed to determine the status and trends with respect to water quality in the SRW. The key parameters reviewed in this report are metrics related to oxygen levels, and critical primary nutrients, nitrogen and phosphorus. This report focuses on the spatial and temporal trends of these constituent measures. Future work will focus on other water quality parameters and their spatial and temporal trends in the SRW.

B. Data-Management Issues

Our SRW Consortium, led by Pinnacle, has compiled a data-warehouse of environmental data with nearly two million data records. General categories of data include water quality, water quantity, sediment quality, stream flow, precipitation, macroinvertebrate, fish community, fish tissue, and other research and environmental monitoring information.

The bulk of this secondary data was compiled in 2003, with additional data added in 2004 and 2005. All data have been compiled into a Microsoft® Access™ data-warehouse that currently contains over 1.9 million records, assembled from more than 80 sources. Historic data sources span 70+ years for stream flow data and 45 years for the United States Environmental Protection Agency's (USEPA) Storage and Retrieval (STORET) data. This STORET data comes primarily from samples collected and analyzed by the South Carolina Department of Health and Environmental Control (SCDHEC) through its ambient water quality monitoring program. Approximately half the data is related to water quality and biological information, and the other half, to flows and precipitation.

Key challenges in this data-mining review of very long-term trends have included inconsistent sampling locations and frequencies, changes in monitored parameters and detection levels, and shifts in the approach of the ambient water quality monitoring program. Technical issues that have emerged from the statistical review of the data include major swings in sample sizes and data variance, the proper handling of "extreme" values, effective utilization of detectable values below quantifiable concentrations, and interpretive thresholds for screening significance of data trends.

Another concern is comparability of older "Legacy" STORET data (often referred to in this report as "historical" data) with "Recent" STORET data (post-1988). Because of these data quality issues, even the best available data has required extensive processing and transformations to make it readily useable.

C. Water Quality Parameters

Water quality data initially extracted and reviewed for presentation in this report are:

- **Biological Oxygen Demand (BOD),**
- **Dissolved Oxygen (DO),**
- **Ammonia Nitrogen (NH_3),**
- **Total Kjeldahl Nitrogen (TKN),**
- **Nitrate-Nitrite Nitrogen (NO_3-NO_2),** and
- **Total Phosphorus (TP)**.

Biochemical Oxygen Demand (BOD) is a very commonly measured attribute of water quality. BOD reflects the amount of organic matter or organic loading to a water body, manifested as material that requires oxygen to be broken down by biological and chemical processes. BOD is thus inversely proportionate to another very familiar measure of water quality, **Dissolved Oxygen (DO)**. As BOD levels increase, DO levels generally are proportionately decreased. High levels of BOD reflect high loadings of organic matter and potentially other associated contaminants.

Common sources of BOD include natural organic matter and detritus from leaves, woody debris, and the like, as well as high-strength organic materials such as municipal or industrial wastes discharged from wastewater treatment plants (WWTPs). These materials tend to require large amounts of oxygen for degradation, both chemically and by microorganisms. As oxygen is consumed in the breakdown of these substances, less DO is available for fish and other aquatic organisms. Hence, a high concentration of BOD reflects the potential for high stresses on biological organisms in an otherwise healthy stream environment. As a measure of water quality, BOD is a long-recognized measurement technique for which we have substantial BOD data for the SRW going back to the mid-1950s.

Nitrogen is a primary nutrient in both terrestrial and aquatic organisms, a natural part of plant detritus, and also comes from atmospheric sources with precipitation. However, nitrogen in the waters of the SRW comes from numerous sources, some natural, some man-made. Point sources of nitrogen associated with wastewater treatment facilities are significant. Likewise, non-point sources from lawn fertilization, agriculture, animal wastes, and storm water runoff are quite significant (Klaine and Smink). Nitrogen is monitored in its various forms:

- **Ammonia (NH_3)**,
- Organic nitrogen, measured as reduced forms of nitrogen (including NH_3) using the Kjeldahl analytical method, i.e. **Total Kjeldahl Nitrogen (TKN)**, and
- The oxygenated forms of nitrogen, **nitrate (NO_3) and nitrite (NO_2).**
- **Total Nitrogen (TN)** is the sum of both reduced and oxygenated forms of nitrogen (calculated only from observations on same date at same location):
  $TN = TKN + [NO_3+NO_2]$

Total Phosphorus (TP) includes all forms of phosphorus significant to aquatic ecosystems. Phosphorus is also a major nutrient in terrestrial and aquatic environments. Like nitrogen, phosphorus occurs naturally and in man-made forms. Decaying natural vegetation and animal wastes from wildlife are natural sources, as are some mineral forms containing low levels of phosphorus. However, point sources of man-made phosphorus are very significant from human and industrial waste streams. Non-point sources are also very significant from fertilization, agriculture, and domestic and agricultural animal wastes.

Elevated water concentrations of phosphorus are often implicated as a major contributing cause in nuisance algae blooms such as the one that occurred in the Reedy River Arm of Lake Greenwood in 1999 (McKellar and Bulak). Because aquatic ecosystems in this region, such as Lake Greenwood, are especially sensitive to phosphorus concentrations, this constituent is especially critical to monitor as an indicator of water quality.

From a lay-perspective, the most important core indicators of water quality, or conversely, of the impairment of water quality due to pollution, are:

- Measures of oxygen availability and demand, i.e. BOD and DO
- Nitrogen, i.e. TN
- Phosphorus, i.e. TP

These constituents are excellent indicators of the general health of surface waters, and of spatial and temporal trends in water quality. Numerous other water quality constituents are also of interest to the SRWC and will continue to be analyzed going forward.

D. Overview of Watershed Study Area

The study area is divided into logical hierarchical watershed sub-areas for closer examination. The most fundamental break is at the major sub-basin level. In the case of the SRW, those logical breaks are the Rabon Creek, Reedy River, and Saluda River Sub-Basins , as shown in Figure 1. Many of the results presented in this report are presented in the context of these sub-basin areas.

The next logical subdivision is according to hydrologic units, as defined by the United States Geological Survey (USGS) for watershed delineation, of HUC-11 subwatershed units (using an 11-digit nomenclature). As shown in Figure 2, thirteen HUC-11 units are identified in the SRW. The Saluda, Reedy,

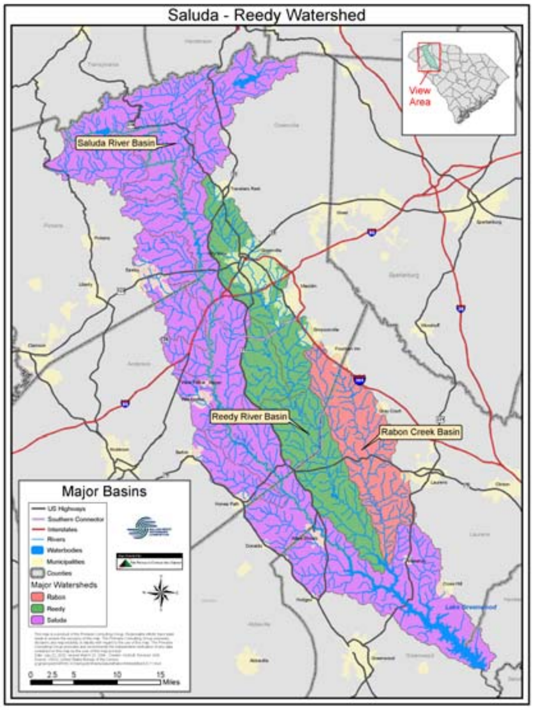


Figure 1. Major Sub-Basins of the SRW

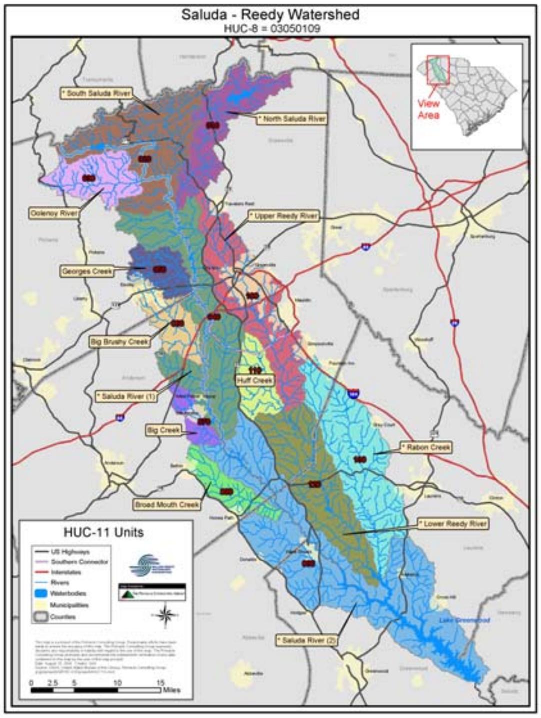


Figure 2. USGS HUC-11 Hydrologic Units of the SRW

and Rabon sub-basins include nine, three, and one HUC-11 units, respectively.

Seven of these HUC-11 units were selected as representative units for detailed analysis of spatial trends in the SRW. The seven selected HUC-11 units capture the range of watershed characteristics affecting water quality, to include urban, industrial, and rural mixes of land cover and land use. These HUC-11s are also generally more rigorously monitored and, therefore, have more year-round data.

E. Analytical Data Review and Statistical Processes

1. Statistical Tools

Based on character, quality, and technical utility of various information sources, records were screened and processed to identify records that met data quality objectives, then imported into JMP™ (SAS Institute Inc., Cary, NC) statistical analysis package. Data were analyzed using established statistical methodology for temporal, spatial, parametric, and cross-parametric trends.

2. Context for Statistical Comparisons

Statistical analysis was designed to answer these simple questions:

- Have concentrations changed over time?
- If so, do these changes exhibit a pattern, and where have these changes occurred?

To examine the first question, each parameter was analyzed from two perspectives:

- An historical look based on all data available in the STORET system, generally going back to ca 1955, with reference to The Clean Water Act of 1972.
- A closer look at the last 15 years of data (a generally-accepted review period for trend analysis) available in the STORET system, from 1988 to 2002 (some 2003 and 2004 data has since become available, but was not included in this work due to timing).

To examine the second question with respect to spatial patterns of change, each of the two major datasets was analyzed from three perspectives:

- A broad overview of the entire SRW;
- A look at the three major sub-basins of the SRW; and
- A close look at the seven major HUC-11 subwatersheds of the SRW.

For each parameter, analyses of both the "historical" and "recent" datasets followed the same statistical sequence:

a) To provide a simple visual graphic of possible changes over time, actual mean concentrations watershed-wide (calculated for each year) were plotted against year.

b) To determine if concentration was related to time (measured in years), a simple linear regression or bivariate analysis tested a basic relationship between these two continuous variables: **Does concentration watershed-wide vary across the specified time frame as expressed through a simple linear relationship?** Results indicate overall magnitude, direction, and statistical significance of any change.

c) To determine if concentration was influenced by sub-basin or HUC-11 characteristics in the SRW itself, regardl0ess of time, a one-way analysis of variance looked at how concentration differed across discrete hydrologic sub-groups: **Within the specified time frame, do significant differences in concentrations exist among the three sub-basins? the seven primary HUC-11 subwatersheds?** Results indicate overall magnitude, direction, and statistical significance of any differences within each group.

d) To determine if the relationship between concentration and time was influenced these sub-basin or HUC-11 characteristics within the SRW, a multiple linear regression simultaneously looked at how concentration changes over time differed across these hydrologic sub-groups: **Within the specified time frame, do significant differences in concentrations over time exist among the three sub-basins? the seven primary HUC-11 subwatersheds?** Results provide the strongest basis and most useful information for understanding spatial and temporal distinctions in water quality in the SRW.

3. Explanation of Statistical Graphical Output

In JMP™, the "Fit Y by X" platform was used for statistical analysis. A one-way analysis of variance (ANOVA) table was used to determine if at least one sample mean was statistically significant from the others. Pairs of means were compared for pairwise significant differences using Tukey-Kramer HSD (Honestly Significant Difference) test, and graphically illustrated with comparison circles. A 5% level of significance was used throughout.

The reader is directed to the Appendix for an interpretative guide to understanding graphical presentation of the statistical analyses.

4. Handling of Extreme Records

Data analysis of any type requires a check for extreme values, or outliers, in the dataset. Outliers distort several calculations that are critical to the statistical inference process, thus possibly resulting in flawed interpretations and conclusions. These water quality data sets, often with many thousands of observations for each parameter over decades of collections, contained a number of extreme values, always on the high end. Consultation with a few key state-level experts, and review of the USEPA STORET specifications, convinced us that all data entries have passed local and/or state scrutiny, are assumed to be correct, and can be used "as is" (personal communications with David Chestnut, SCDHEC Bureau of Water). To make decisions on any of these extreme values would have put us on a slippery slope, as we had no basis for determining legitimacy or not.

Therefore, with hundreds and thousands of data points to help minimize such effects, all observations were used. Invariably, the more extreme data values do create some distortion, but the large data sets help dampen that variance.

II. Analysis Results and Trends Assessment

(Note: Portions of these results were presented to the NALMS (North American Lake Management Society) Southeastern Region Conference in April 2005, entitled "Long-Term Water Quality Trends in the Saluda-Reedy Watershed: Early Insights from Data-Mining.")

For benefit of perspective, **almost 100 separate statistical analyses** went into the preparation of the following results and discussion. Because we were using secondary data, i.e. data that had been previously collected by primary researchers for a purpose other than the one at hand, we did not have the luxury of experimental design, with methodology predetermined. So data were examined from many perspectives, with the goal of pragmatic and straightforward statistical analyses that produced actionable results.

In **every single analysis** we executed, the effects of the dependent variable "concentration" is statistically significantly affected by the independent variable, be it time or watershed feature, at a probability level of $p<0.0001$ or less. Interestingly, adjusted-R^2 values for these same analyses of variance are relatively low, ranging from 0.001 to 0.24, indicating that the simple models we used only explain 0.1 to 24 percent of the variability in the data. Yet, with thousands of observations, the impacts of time and watershed features are so overwhelmingly powerful that their effects are obvious, despite extreme records.

A. Long-Term Water Quality, 1955-2002

Looking across the entire SRW at major trends in water quality reveals several highly useful insights. Note that this initial discussion addresses the entirety of the period of record, and is referred to as the "Long-Term" or historical overview. These long-term trends are very useful in understanding the overall history of this watershed, and appreciating the real and dramatic impacts of Clean Water Act (CWA) of 1972, and related shifts in regulatory programs affecting water quality. However, some of the oldest data may not have the same integrity as newer data, for many reasons. In addition, the last 15 years of information is typically used for trend assessments. So, the most recent years of available data, 1988-2002, are broken out as a subset for subsequent discussion in this report (see part B in this section, beginning Page 12).

1. Watershed-Wide Assessments

a) Oxygen: Historical Trends in BOD and DO

Figure 3 presents BOD conditions across the watershed (nearly 12,000 samples) from the beginning of systematic water quality monitoring in 1956, through 2002. The highly elevated condition of waters across the watershed with respect to BOD loading prior to ca 1970 is evident. With the advent of the 1972 CWA,

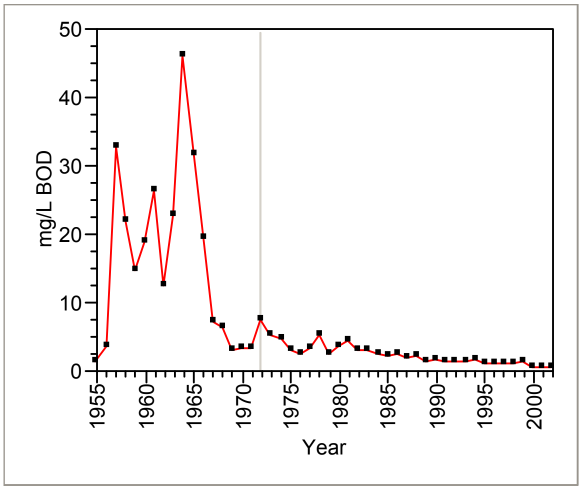


Figure 3. Long-Term BOD Trends Across the Entire SRW (n=11,965 records)

]

municipalities and industries were required to clean up their discharges, and clearly the resultant changes in practices have had the desired effect in improving water quality.

Figure 4 shows the general trends in DO across the watershed for the same period (nearly 25,000 measurements). DO essentially mirrors the BOD trends, with generally low and erratic levels of DO prior to the late 1960s. For purposes of reference, many fish cannot survive and reproduce if DO levels are below ~ 5.0 mg/L. After 1972, concentrations of DO stabilized on a watershed basis, to an overall concentration of ~ 8 mg/L.

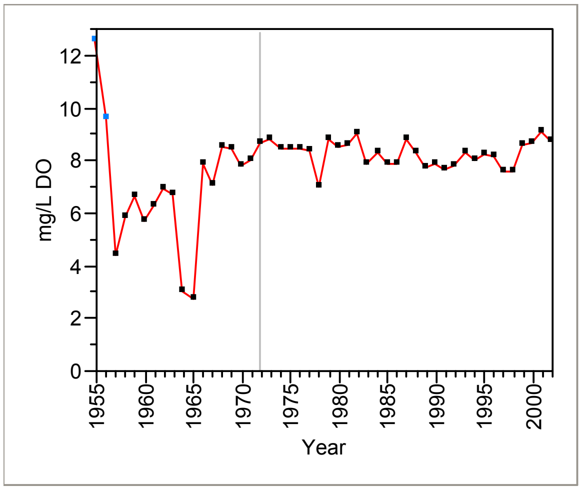


Figure 4. Long-Term DO Trends Across the Entire SRW (n=24,755)

b) Nitrogen: Historical Trends in TN, NH_3, and NO_3-NO_2

Data for nitrogen, available since 1972, reveals a similar trend (Figure 5). Note that TN is not calculated prior to this date because TN is a summation of several other nitrogen analytes, and analyses of all the various components of TN were not routinely conducted prior to the late 1960s. Still, this graphic depicts a trend of progressively more stable and improving TN values over the last 20 years.

To reinforce the TN trend, Figures 6 and 7 illustrate long-term trends for ammonia, a

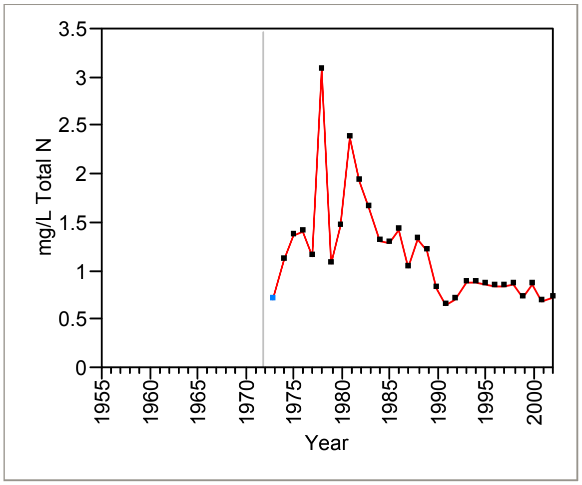


Figure 5. Long-Term TN Trends Across the Entire SRW (n=5,015)

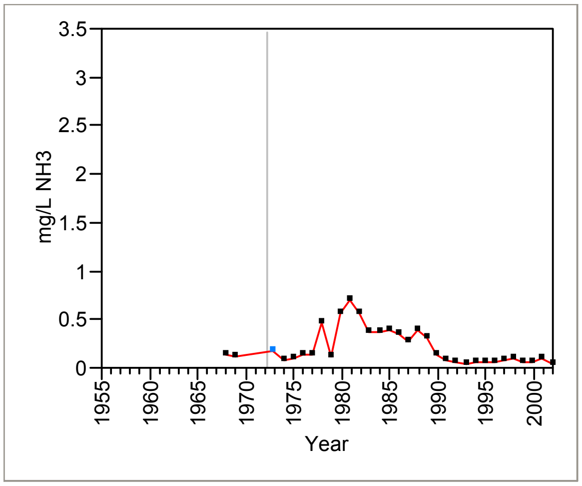


Figure 6. Long-Term NH_3 Trends Across the Entire SRW (n=5,374)

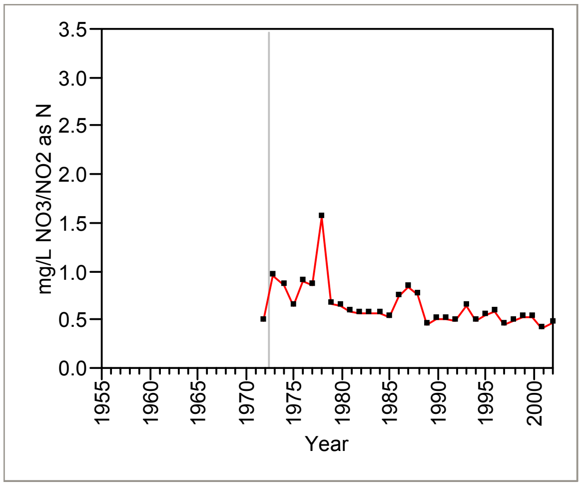


Figure 7. Long-Term NO_3-NO_2 Trends Across the Entire SRW (n=9,568)

major component of nitrogen-loading to aquatic systems, and for nitrate-nitrite nitrogen, the oxygenated forms of nitrogen, respectively. Both graphics demonstrate steady improvement since the late 1970s.

c) Phosphorus: Historical Trends in TP

TP in the watershed, as presented in Figure 8, exhibits a similar pattern of erratic, high levels in the 1970s, followed by steady improvements beginning in the late 1980s.

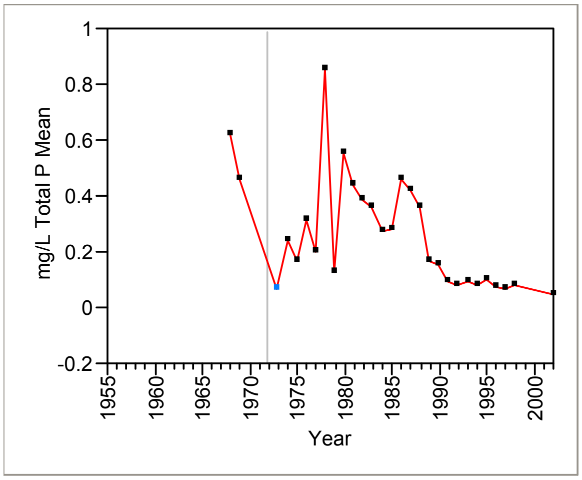


Figure 8. Long-Term TP Trends Across the Entire SRW (n=6,482)

d) Summary of Long-Term Water Quality Watershed-Wide, 1955-2002

Subsequent data analysis, as described in the protocol on pages 4-5, addressed this question: **Does concentration watershed-wide vary across the specified time frame as expressed through a simple linear relationship?** Results substantiated what is clearly apparent in the plots: that every one of these relationships is statistically significant. With exception of DO, the slope of all predicted lines is negative, meaning the relationships are negative, that is, as year increases, the predicted concentrations decrease. The relationship for DO is just the opposite, with a positively-sloped prediction line, i.e. as year increases, predicted DO concentrations also increase.

The general improvement in water quality for each of these major parameters over this 50-year period is very encouraging, particularly important because this broad, overall trend results from examining the many thousands of measurements on a watershed-wide basis.

Another important consideration is that, although conditions are very substantially improved, some indicators, such as TP, still have room for improvement. For example, a huge mass of phosphorus (P) was discharged routinely up until the 1980s. This macronutrient has a strong affinity for sediments and, as sediments have been deposited in SRW reservoirs, large masses of phosphorus have been deposited, as well. These observations have helped the SRWC team of scientists focus now on the ongoing questions and concerns revolving around the potential bioavailability of phosphorus-enriched sediment as a source of nutrients for possible algae blooms (McKellar and Bulak). These nutrient-enriched sediments, along with shallow conditions in the Reedy Arm of Lake Greenwood, may have contributed to the dramatic algae bloom that plagued this portion of Lake Greenwood during the summer of 1999. Ongoing research by the SRWC will provide additional insights on that scenario.

The following sections drill deeper into the database to evidence conditions on a more localized basis, and determine if concentration changes over time may reflect spatial trends within portions of the watershed.

2. Assessments by Major Sub-Basins

a) Oxygen: Historical Trends in BOD and DO

BOD data for Saluda and Reedy Rivers, and Rabon Creek sub-basins are presented in Figure 9. *(Note: The reader is referred to* *the Appendix for an interpretive guide to these statistical graphical tools.)* Data reflected in this figure, and others like it, represent all the BOD data for the period of record, so time is not a factor. Mean concentrations of BOD in the Reedy are significantly higher ($p<0.05$) than in the other two sub-basins. Subsequent multiple regression indicates that these differences among the sub-basins remain significant over time. Such results reflect the very high waste loading to the Reedy during these 50 years.

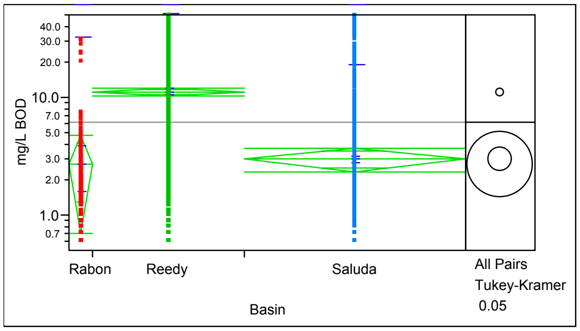


Figure 9. Comparison of Long-Term BOD Levels Across Major SRW Sub-Basins (n=11,965)

Figure 10 presents the same analysis for DO in the three sub-basins. Here, all three data populations are different from each other, with the Reedy having the lowest mean (~ 6.3 mg/l) and the Saluda, the highest mean DO level (~ 8.3 mg/l). Subsequent multiple simple linear regression indicates these differences among the three sub-basins remain significant over time.

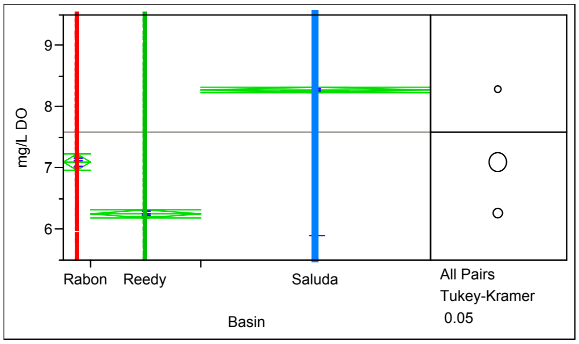


Figure 10. Comparison of Long-Term DO Levels Across Major SRW Sub-Basins (n=24,755)

Another useful perspective on this data is provided by trend lines calculated by decade for each sub-basin. Figures 11 and 12 illustrate the general improvement in both BOD and DO over the past 50 years, and generally reflects that history of wastewater discharges.

One example reflected in these trends is the discharges to Rabon Creek from a notorious wastewater treatment plant in service until the 1960s, when the discharge was redirected to the Reedy basin. Dramatic water quality improvements in BOD and DO concentrations are obvious in Rabon Creek. Improvements in DO have been sustained in the Reedy, but appear to be declining slightly in the Saluda and the Rabon basins, possibly due to increases

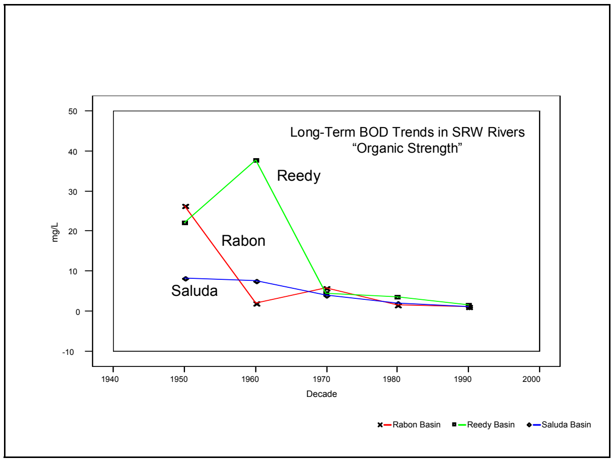


Figure 11. Comparison of Long-Term BOD Levels Across Major SRW Sub-Basins

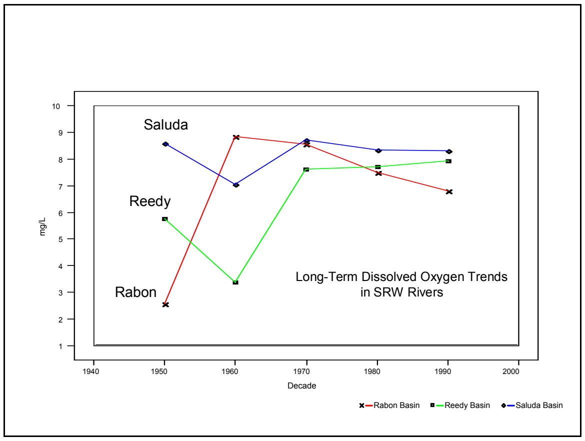


Figure 12. Comparison of Long-Term DO Levels Across Major SRW Sub-Basins

in waste loading and non-point source loadings increasing with suburban development.

b) Nitrogen: Historical Trends in TN

Sub-basin comparisons in total nitrogen concentrations are presented in Figure 13. This inter-basin analysis confirms the TN concentrations in the Reedy to be dramatically and statistically different from the Rabon and Saluda waters. The mean concentration of TN for the Reedy across all samples is 2.0 mg/l, compared to 0.7 and 0.5 mg/l for the Saluda and Rabon waters, respectively. These conditions, likewise, demonstrate the historically high loading of wastewaters to the Reedy.

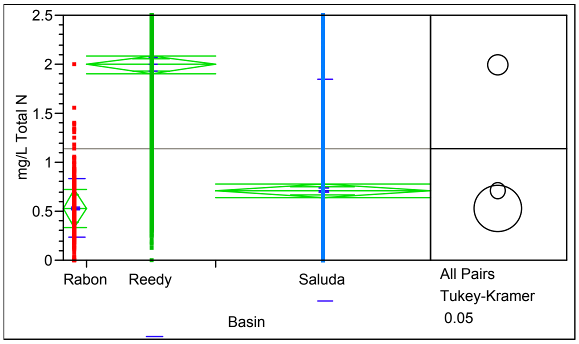


Figure 13. Comparison of Long-Term TN Levels Across Major SRW Sub-Basins (n=5,015)

Looking at TN across sub-basins, the waters in each have enjoyed steady improvement in water quality. Figure 14 shows those trends by decade, comparing means calculated for the 1970s, 1980s, and 1990s.

c) Phosphorus: Historical Trends in TP

Trends in phosphorus over the period of record are presented in Figure 15. As with the other key parameters, the mean concentrations of TP in the Reedy (0.39 mg/l) are significantly higher than those for the Saluda (0.18 mg/l) and Rabon samples (0.10 mg/l).

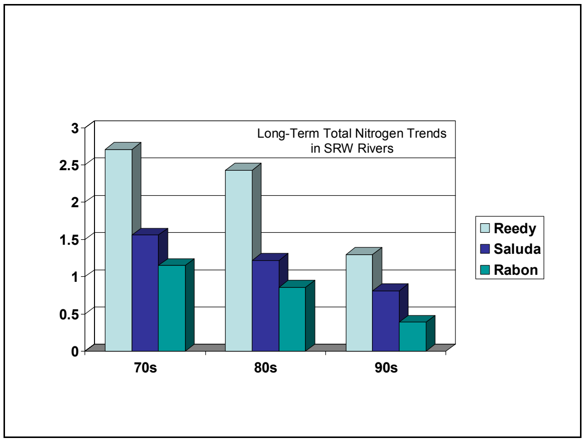


Figure 14. Comparison of Long-Term TN Trends Across Major SRW Sub-Basins

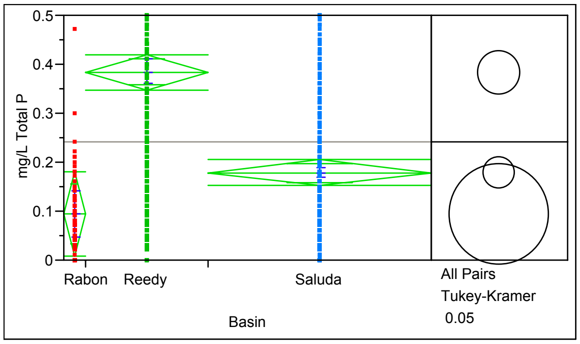


Figure 15. Comparison of Long-Term TP Levels Across Major SRW Sub-Basins (n=6,482)

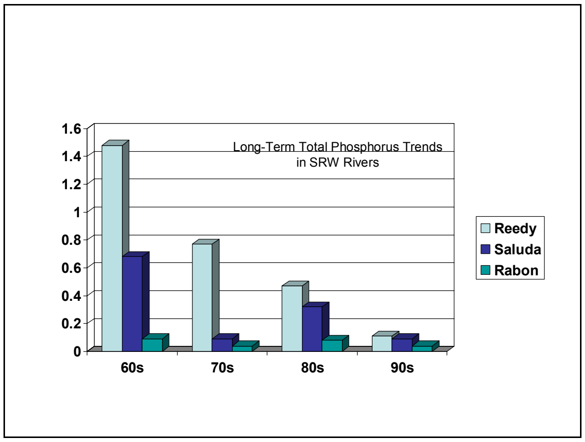


Figure 16. Comparison of Long-Term TP Trends Across Major SRW Sub-Basins

Figure 16 presents the trends in TP by decade across the sub-basins. Each basin shows significant improvement in TP concentrations since the 1960s.

The remarkable improvements in phosphorus concentrations in the Reedy reflect the measures undertaken to improve wastewater treatment technologies at the major treatment works (Western Carolina Sewer Authority's Mauldin Road and Lower Reedy WWTPs). One caution in interpreting the dramatic, roughly 10-fold improvement in phosphorus concentrations in the Reedy, is to raise the question of the fate of the huge mass of TP discharged in prior decades. In looking at TP data on this scale, it becomes readily apparent that the historic loading to the Reedy may be manifested in significantly enriched sediments accumulating in the reservoirs of the watershed.

3. Assessments by Selected HUC-11 Subwatershed Units

The next appropriate level of scrutiny of the key parameters indicative of water quality is at the HUC-11 level. Because there are 13 HUC-11 units in the entirety of the SRW watershed, we simplified this analysis to assess seven HUC-11 subwatershed units that represent the range of land-use/land-cover conditions and water quality conditions in the overall watershed. The watersheds are:

- North Saluda (above confluence, rural)
- South Saluda (above confluence, rural)
- Urban Saluda (middle reaches, urban)
- South Saluda (lower rural reaches)
- Upper Reedy (highly urbanized)
- Lower Reedy (lower rural reaches)
- Rabon (generally rural)

a) <u>Oxygen: Historical Trends in BOD and DO</u>

The trends in BOD for seven HUC-11 basins are presented in Figure 17. The reader is cautioned to recognize that the data reflected in this figure cover data collected for the entire period of record. Breaking the data down to the HUC-11 level further differentiates the historic condition of the Upper Reedy relative to all other HUC-11 units. The Upper Reedy unit has a mean of 13.6 mg/l BOD, which is statistically significantly from all other units. The Urban Saluda reach has a mean of 4.2 mg/l, and all other units are less than 2.7 mg/l.

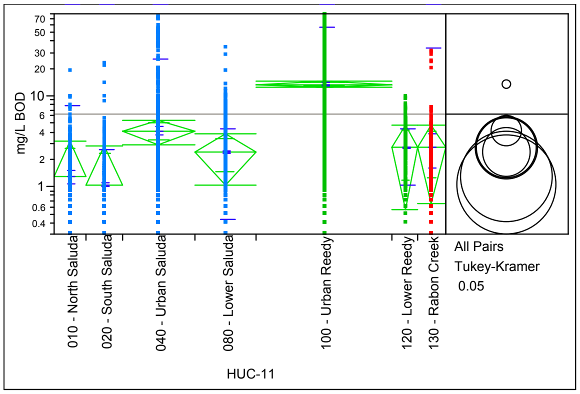


Figure 17. Comparison of Long-Term BOD Levels Across Selected SRW HUC-11 Subwatersheds (n=10,327)

The related data for DO show a similar pattern for HUC-11 units. Figure 18 presents the DO summary for the seven selected HUC-11s.

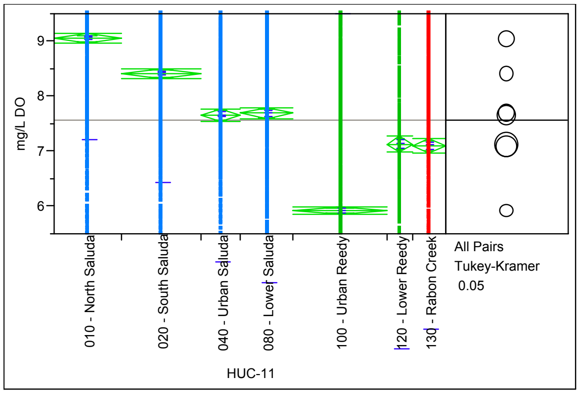


Figure 18. Comparison of Long-Term DO Levels Across Selected SRW HUC-11 Subwatersheds (n=22,920)

Once again, the Urban Reedy is distinctively different from the other HUC-11 units. The mean value for DO for the Upper Reedy was 5.9 mg/l, compared to 7.1 for the Rabon and Lower Reedy units. The depressed levels of DO in the Upper Reedy is in marked contrast to the North Saluda (mean=9.1 mg/l) and the South Saluda (mean=8.4 mg/l), which still have remarkably favorable DO conditions, reflecting their dominantly forested and urban watershed conditions.

b) Nitrogen: Historical Trends in TN

Figure 19 presents the statistical comparison for TN in the seven HUC-11 sub-basin units. Both Upper Reedy and Lower Reedy units are significantly different from the other units, with mean concentrations of 2.1 and 1.8 mg/l, respectively. The Urban Saluda had a mean TN value of 1.0 mg/l. Again, the North Saluda and the South Saluda demonstrate comparably pristine conditions, with low TN values of only 0.30 and 0.21 mg/l, respectively.

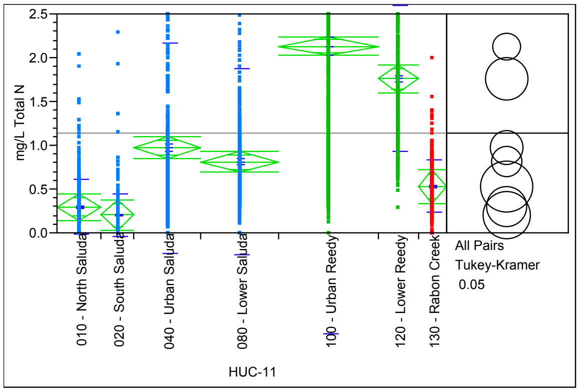


Figure 19. Comparison of Long-Term TN Levels Across Selected SRW HUC-11 Subwatersheds (n=4,684)

c) Phosphorus: Historical Trends in TP

Levels of phosphorus over the period of record for the seven HUC-11 units are presented in Figure 20. Both Upper and Lower Reedy units have much higher means for TP (0.40 mg/l for each) than all other HUC-11 units. The Urban Saluda has an intermediate mean value of 0.28 mg/l, significantly greater than the other four HUC-11 units, which range from 0.07 to 0.10 mg/l. These findings are generally as expected, given the known history of TP discharges and loadings to the SRW waters.

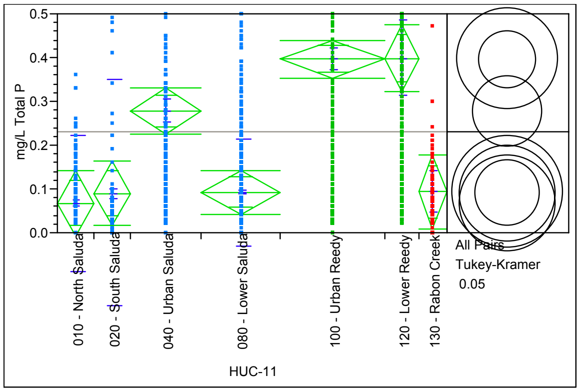


Figure 20. Comparison of Long-Term TP Levels Across Selected SRW HUC-11 Subwatersheds (n=5,686)

B. Recent Water Quality, 1988-2002

Over the past two decades, since 1988, significant improvements in the SCDHEC ambient water quality monitoring program, advances in analytical protocols and sensitivities, enhanced regulatory functions, improved performance in waste treatment by industry and municipalities, and the stewardship of resources have all been advanced. As a result, the STORET data reflects a substantially improved measure of water quality trends. This more recent water quality data is markedly distinctive from the older data. Based on a review of data trends, and in consideration of feedback from SCDHEC water quality specialists (Chestnut), we made a break in the data population at 1988. Data from this period forward provides the highest quality information of greatest value to our decision-making with regard to management of the watershed's water resources. The following analysis of trends reflects this more recent data.

1. Watershed-Wide Assessments

a) Oxygen: Recent Trends in BOD

Figure 21 presents the trends in BOD for the 1988-2002 period. BOD concentrations for this period reflect a strong downward trend. This is interpreted to reflect reductions in overall waste-loading to the collective waters of the entire SRW. Figure 22 reflects the same data modeled via a simple linear regression, which shows that the drop in BOD levels across the entire SRW over these 15 years is, indeed, highly significant ($p<0.001$).

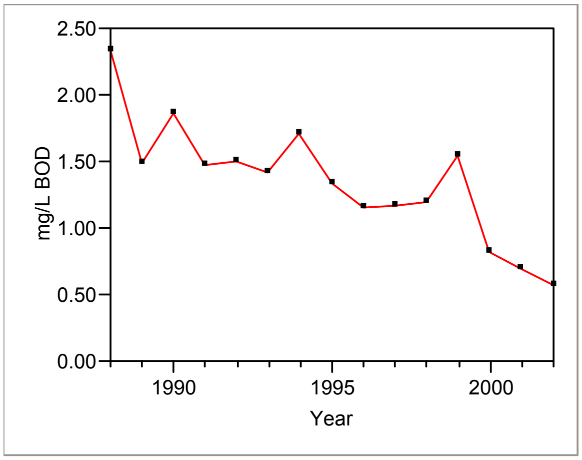


Figure 21. Recent BOD Levels Across the Entire SRW, 1988-2002 (n=4,638)

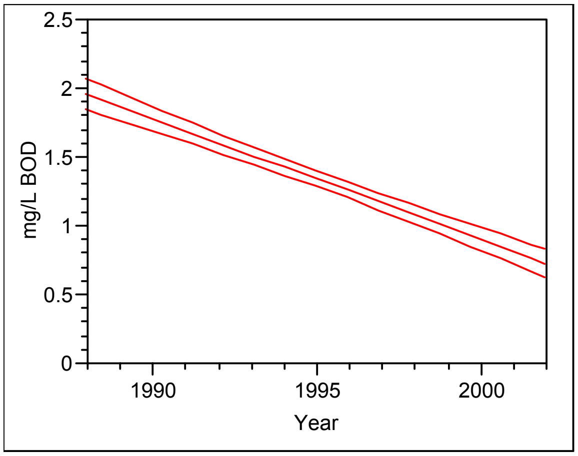


Figure 22. Simple Linear Regression of BOD Across the Entire SRW, 1988-2002

b) Nitrogen: Recent Trends in TN

Figure 23 presents the trends in Total Nitrogen for the 1988-2002 period. TN levels for this period reflect an initially strong downward trend for the 1988 to 1991 period, and a steady declining trend from 1993 to 2002.

This is interpreted to reflect reductions in overall waste loading to SRW waters. Figure 24 shows the same trend modeled via simple linear regression.

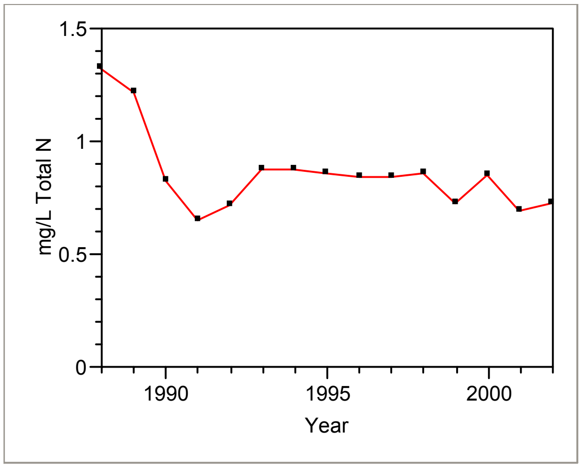


Figure 23. Recent TN Levels Across the Entire SRW, 1988-2002 (n=3,146)

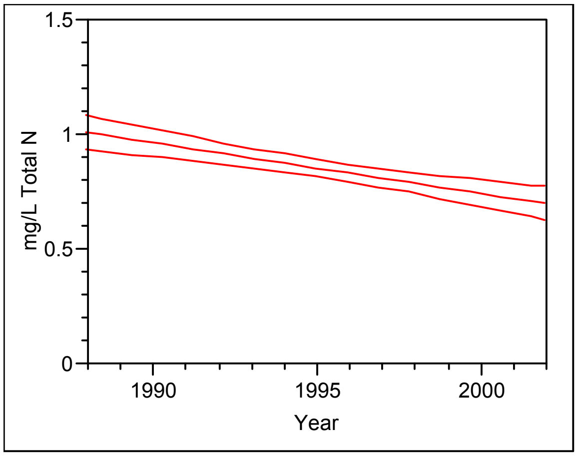


Figure 24. Simple Linear Regression of TN Across the Entire SRW, 1988-2002

c) Phosphorus: Recent Trends in TP

Figure 25 presents the trends in Total Phosphorus for the 1988-2002 period. TP concentrations for this period reflect a strong initial downward trend for the 1988-1991 period, and a steady declining trend during 1993-2002.

This is interpreted to reflect reductions in overall waste loading to the waters of the SRW, primarily as a function of improved treatment technologies implemented at WWTPs. Figure 26 reflects the same trend modeled via a simple linear regression, with the cautious reminder that such a simple model demonstrates a strong linear relationship and does not reflect the true asymptotic nature of this data.

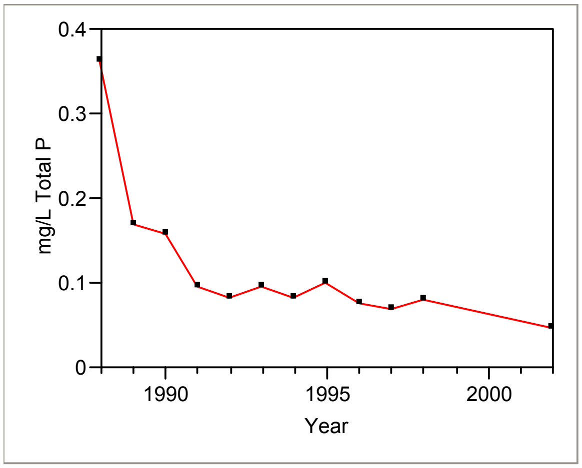


Figure 25. Recent TP Levels Across the Entire SRW, 1988-2002 (n=3,600)

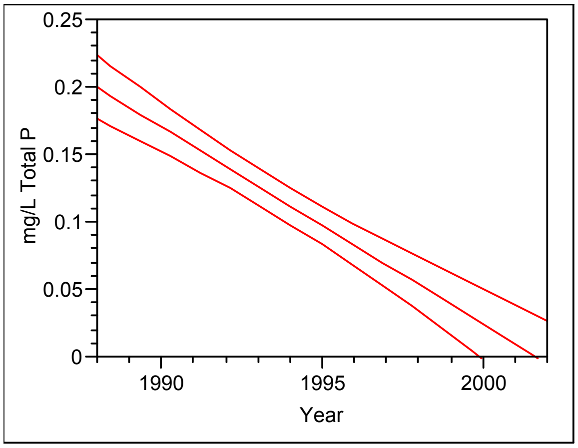


Figure 26. Simple Linear Regression of TP Levels Across the Entire SRW, 1988-2002

2. Assessments by Major Sub-Basins

a) <u>Oxygen: Recent Trends in BOD</u>

It is also important to examine the recent data for trends with respect to the major SRW sub-basins. Figure 27 presents the statistical analysis of data for the Rabon, Reedy and Saluda basins for this period. Clearly, the Reedy, with 1.7 mg/l, has a much higher and statistically significantly different mean from the other sub-basins, each with a mean of 1.2 mg/l.

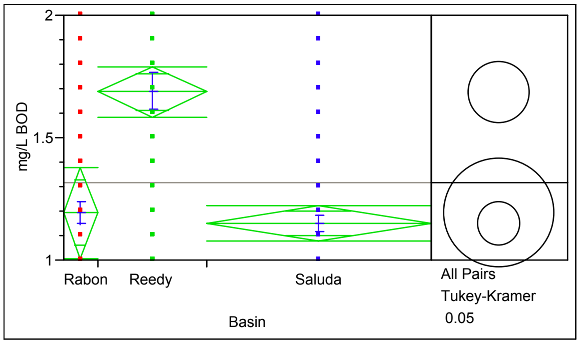


Figure 27. Comparison of Recent BOD Levels Across Major SRW Sub-Basins, 1988-2002 (n=4,638)

b) <u>Nitrogen: Recent Trends in TN</u>

Review of overall TN concentrations for the "Recent" period confirms the Reedy sub-basin of the SRW reflects significantly higher nitrogen concentrations than the Saluda and Rabon sub-basins (Figure 28). The mean for the Reedy data group is 1.4 mg/l, more than twice the 0.6 and 0.5 mg/l means for the Saluda and Rabon data groups, respectively.

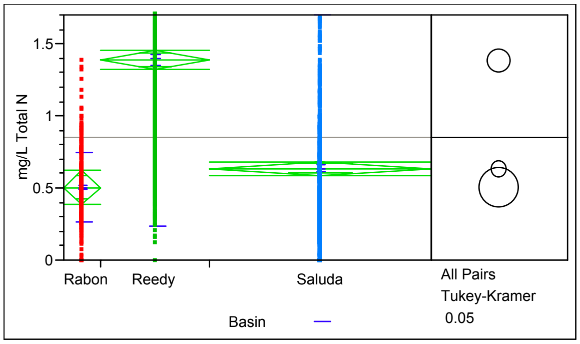


Figure 28. Comparison of Recent TN Levels Across Major SRW Sub-Basins, 1988-2002 (n=3,146)

c) <u>Phosphorus: Recent Trends in TP</u>

As illustrated in Figure 29, review of TP concentrations for the "Recent" period confirms that all three sub-basins are significantly different from each other, with the Reedy sub-basin of the SRW having the highest phosphorus concentrations over these 15 years, with a mean of 0.15 mg/l. The mean for the Saluda data group is significantly less with 0.11 mg/l, as is the mean for the Rabon data group, with 0.04 mg/l.

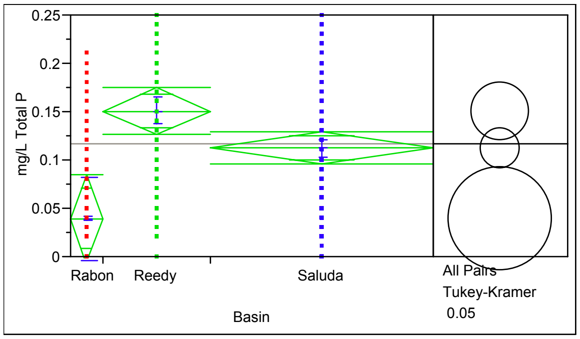


Figure 29. Comparison of Recent TP Levels Across Major SRW Sub-Basins, 1988-2002 (n=3,600)

3. Assessments by Selected HUC-11 Subwatershed Units

a) <u>Oxygen: Recent Trends in BOD</u>

Dissecting this data further to the representative HUC-11s illustrates recent distinctions in waste loading to these subwatersheds (Figure 30). Interestingly, for this more recent period, the highest mean BOD concentrations are manifested in the Lower Reedy and Lower Saluda HUC-11s, with mean concentrations of 2.1 and 1.8 mg/l, respectively. The Urban Reedy has the next highest HUC-11 at 1.6 mg/l. All other HUC-11s are below 1.2 mg/l. The North Saluda and South Saluda HUC-11s have means of 0.7 and 0.6 mg/l, respectively. The tests of statistical significance show that the Lower Reedy is higher than all other HUC-11s except the Lower Saluda, and that the North Saluda and South Saluda are the least of all HUC-11s. As explained below, this finding masks additional trends that are evident from subsequent data analysis of specific subwatersheds, and shifts in impacts attributable to storm water and other non-point sources.

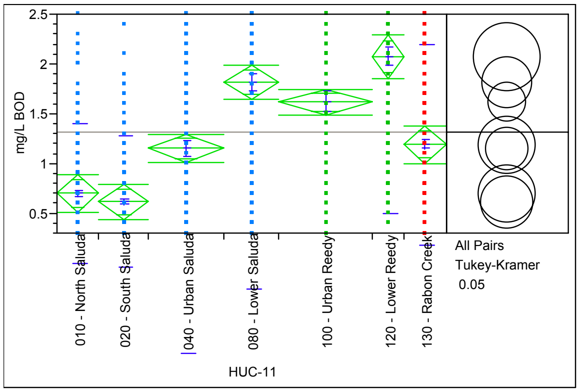


Figure 30. Comparison of Recent BOD Levels Across Selected SRW HUC-11 Subwatersheds, 1988-2002 (n=3,979)

These assessments underscore the importance of evaluating both the “Long-Term” data as well as the “Recent” data. The long-term data is distorted to some extent by the older data which includes significantly elevated concentrations of BOD during the 1950s-1970s. By comparison, BOD trend analysis for all HUC-11’s were higher for the entire “long-term” period. Means for the Lower Reedy and Lower Saluda HUC-11’s were 1.7 and 2.4 mg/l respectively, higher than the values presented here for the “Recent” period. By contrast, the value for the Urban Reedy was 13.6 mg/l for the “Long-Term” period, compared to 1.6 for the “Recent” period. Obviously, waste loadings and concomitant BOD concentrations for all HUC-11s are improved, but those that were most severely impacted during the 1950s-1970s are most dramatically improved. This comports with known reductions in waste load discharges to the Urban Reedy, Urban Saluda, and Rabon subwatersheds. Hence, both periods of analysis are quite useful, as are the comparisons of these periods.

Examining the temporal trend analysis within each of the seven representative HUC-11s also underscores this distinction. Figure 31 presents a multiple simple linear regression of BOD Concentration against Year for the seven HUC-11s in the Recent period. The remarkable trends exposed through this analysis is the increasing trend in BOD levels in the Lower Reedy over this period compared to decreases in all other HUC-11s. This upward trend is assumed to reflect continuing increases in both point source discharges to the Lower Reedy, as well as impacts attributable to non-point source loading associated with continuing urbanization over the period. These trends should be monitored closely in the future to examine potential degradation in water quality and aquatic habitat.

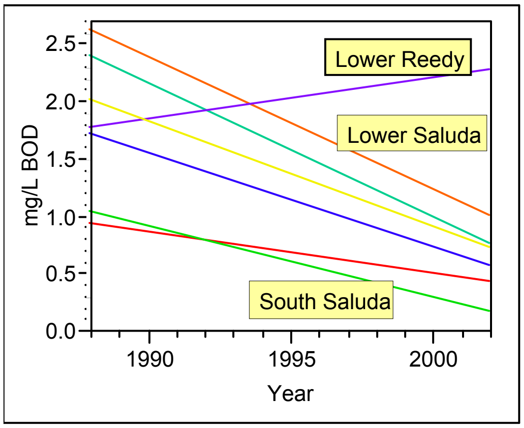


Figure 31. Multiple Simple Linear Regression of Recent BOD Levels Across Selected SRW HUC-11 Subwatersheds, 1988-2002 (n=3,979)

b) Nitrogen: Recent Trends in TN

The closer look at TN data for the Recent period by HUC-11 sub-watershed reveals additional important insights. Figure 32 presents the statistical summary graphics for the seven representative HUC-11s. For this period the Urban Reedy and the Lower Reedy HUC-11s had means of 1.45 and 1.36 mg/l, respectively, values which are statistically significantly higher than the means for the other five HUC-11s. The Urban Saluda was next highest, with a mean value of 1.03, also significantly different from the lower four HUC-11s. The North and South Saluda HUC-11s had values of 0.24 and 0.20 mg/l, respectively, significantly the lowest of all examined subwatersheds.

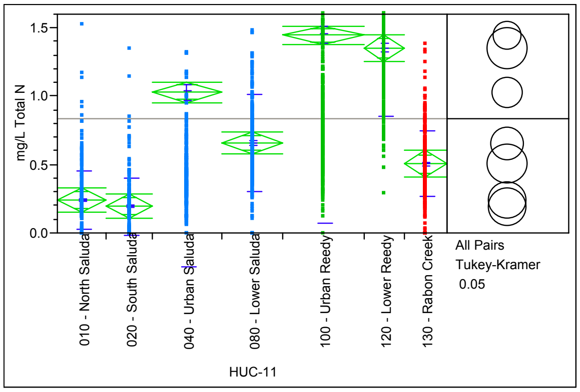


Figure 32. Comparison of Recent TN Levels Across Selected SRW HUC-11 Subwatersheds, 1988-2002 (n=2,880)

These rankings are identical to those for the "Long-Term" period (Figure 19). However, comparison of mean TN concentrations reflected in the most highly impacted subwatersheds, the Urban and the Lower Reedy, for the "Long-Term" versus the "Recent" periods, indicates a roughly 50 percent improvement in mean concentration, dropping from 50-year means of 2.14 and 1.76 mg/l, respectively, to recent 15-year means of 1.45 and 1.36 mg/l. Changes in the other HUC-11s for TN are small.

Figure 33 presents the multiple simple linear regression for the seven HUC-11 data groups. The distinction in the relative magnitudes and the downward trends of the datasets are readily evident. The other remarkable distinction evident here is the modeled increase in TN over the 15-year period in the Urban Saluda. This trend is assumed to reflect increases in point source and non-point source impacts in this sub-water-shed. These findings should be closely monitored in the future to discern any significant degradation in water quality and aquatic habitat.

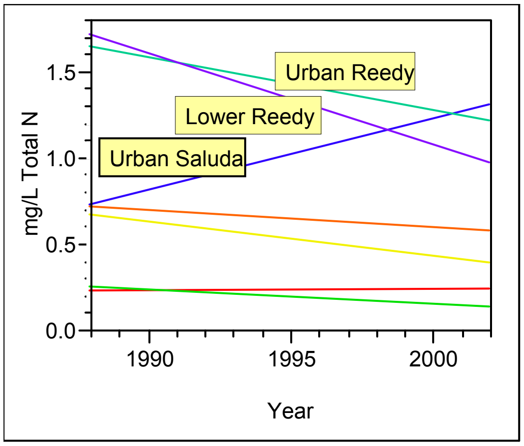


Figure 33. Multiple Simple Linear Regression of Recent TN Trends Across Selected SRW HUC-11 Subwatersheds, 1988-2002 (n=2,880)

c) Phosphorus: Recent Trends in TP

The closer examination of TP data for the 15-year period across HUC-11 sub-watersheds reveals additional important insights. Figure 34 presents the statistical summary graphics for the seven representative HUC-11s. For this period the Urban Saluda HUC-11 has the highest TP concentrations, with a mean of 0.21 mg/l. The Urban Reedy and the Lower Reedy HUC-11s have means of 0.16 and 0.14 mg/l, respectively. These three HUC-11s are not different from each other, but are significantly different from the other four HUC-11s, which have means of 0.03 to 0.07mg/l.

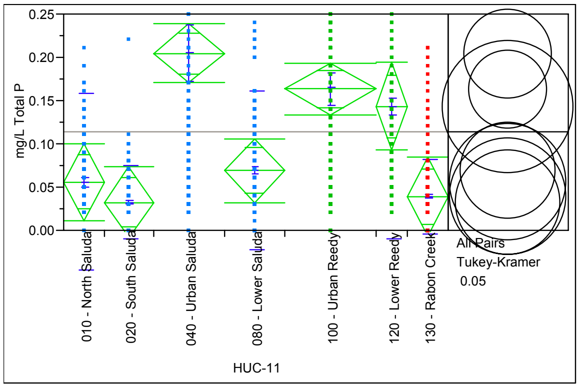


Figure 34. Comparison of Recent TP Levels in Selected SRW HUC-11 Subwatersheds, 1988-2002 (n=3,083)

As with TN concentrations, TP concentrations for the "Recent" data set are lower for all seven HUC-11s as compared to the "Long-Term" data set. Several HUC-11s have overall improvements of as much as 60 percent comparing these periods, but the Urban Saluda has improved the least, with only about a 30 percent reduction. This suggests that although waste treatment technologies and discharge loadings have been reduced substantially, those reductions are being offset by impacts attributable to development.

Figure 35 presents the multiple simple linear regressions for TP across the HUC-11 data groups. The distinction in the relative magnitudes of the datasets are readily evident here with the Urban Saluda, Urban Reedy and Lower Reedy HUC-11's having distinctively different data trends. Obviously, the higher mean for Urban Saluda seen in Figure 34 is a function of its higher levels in the late 1980s. The graphic representation in Figure 35 also illustrates a benefit of looking at data from different perspectives and the limitations of using a simple linear regression to analyze what are actually curvilinear, asymptotic data functions.

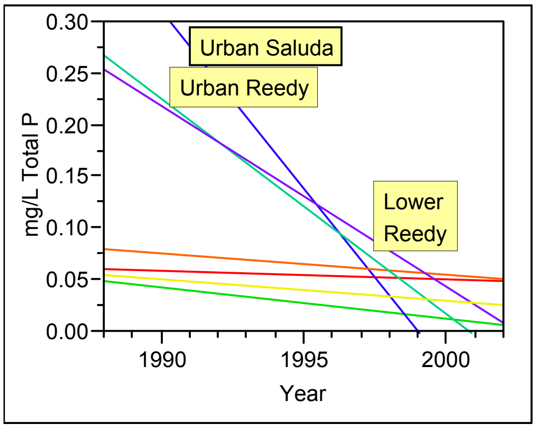


Figure 35. Multiple Simple Linear Regression of Recent TP Levels in Selected SRW HUC-11 Subwatersheds, 1988-2002 (n=3,083)

III. SUMMARY AND NEXT STEPS

A. Synthesis and Effect

The work undertaken thus far through the data-mining task has been enormously useful in developing a better understanding of water quality trends in the watershed. The methods established are now well-defined, many of the critical nuances of the data are now understood, and threshold detection concentrations can be transformed into quantifiable data. The number of issues and opportunities for inquiry with regard to study of water quality trends is virtually endless. The work performed to date represents the proverbial "tip of the iceberg" of data assessment possibilities.

This study has provided a wide range of insights, as summarized below.

1. Long-term recovery from historical pollutant loading: Most of the areas within the SRW demonstrate marked improvement for most water quality parameters over the last 30 years. These improvements are manifested in each of the major parameters analyzed thus far.

2. Impact of the Clean Water Act (CWA) of 1972: Many of the improvements in water quality can be traced back to the effective implementation of the CWA of 1972.

3. Continued & chronic impact to urban / industrial stream reaches: The urban portions of the watershed are, without question, the most severely impacted historically, and they continue to suffer the influences of urban stormwater runoff. Although these stream reaches have recovered substantially, their recovery appears to be offset by nonpoint impacts associated with ongoing development and increased magnitudes of existing and/or new point discharges.

4. Recent impacts to reaches downstream of development: Areas that continue to be subject to changes in land use quickly demonstrate degradation of water quality. The analysis of waste loading and nutrient loading indicated increasing stresses on the areas that have experienced the most significant changes in land use over the past 15 years.

5. Upward trends in BOD & TN loading – suburban reaches: Some urbanizing areas demonstrate clear increasing trends in concentrations of oxygen-demanding substances and total nitrogen. Increased point and nonpoint sources are potential contributing factors in these trends.

6. Distinctions in "long-term" vs. "recent" impacts: The older data is distinctive from the newer data, reflecting shifts in land use and environmental practices, as well as changes and improvements in monitoring methodology and technology. Analysis of both datasets are warranted as each illustrates changes in various factors affecting water quality.

7. Utility of statistical assessment of large historic datasets: The data-mining processes established thus far demonstrate the power and utility of using data-mining, data processing, and statistical analysis techniques on large water quality datasets in a watershed-wide study of this scope.

8. Potential utility for management and policy decisions: The analyses performed thus far provide significant insights to support public policy and environmental regulations within the watershed.

9. Opportunities for ongoing inquiry & trends analysis: Progressive inquiry into the data

reveals additional opportunities for deeper investigation. Additional data-mining activities that warrant priority analysis include partitioning the data into “wet”, “normal”, and “dry” periods which may then be analyzed to examine the effects of climate and streamflow influences. Additional parameters of special interest with regard to water quality and aquatic health deserve additional data-mining attention. Specifically, mercury, copper and other parameters associated with water quality impairment deserve additional review.

10. <u>Need to correlate trends with land cover and/or land use changes</u>: The data suggest strong links between water quality trends and land use. It is critical that the SRWC follow up and better define these relationships.

B. Future Analysis

1. Stream Flow - Monthly and Yearly Statistics

In the course of data assessment numerous questions relevant to SRW water resources management were identified. One fundamental issue that we have identified, and needs to be addressed, is the temporal character of stream hydrology, and long-term trends in stream flows. Priorities for analysis of this data include review of stream flows for monthly mean, minimum and peak flows, and annual mean and peak flows.

An important related task that we have identified as a priority future task is identification of statistically defined “wet”, “normal” and “dry” years. This framework would then be used to partition water quality data from the data-warehouse to examine relationships between climate/hydrology and water quality, and whether or not the trends in water quality are significantly different under these scenarios.

2. Identification of Temporal and Spatial Patterns of Chemical Pollutant Trends

Future analysis and inquiries will examine additional chemical parameters, and include pollutant loading patterns, analysis of sediment and other quality trends, and integration of land cover factors, precipitation data, and event-specific analysis into the data-mining process. Also, ongoing analysis will examine water quality trends according to climatologically-defined “wet, normal, and dry” periods.

IV. REFERENCES

Chestnut, David, Senior Scientist, Bureau of Water, Water Quality Monitoring Section, South Carolina Department of Health & Environmental Control. Personal correspondence via e-mail and telephone conversation, Mar 2005.

Hargett, David (presenter), Holli Hargett, Todd Scott, Steve Springs, and Bob Golus. 2005. ***Long-Term Watershed-Wide Water Quality Trends: Insights from Data-Mining.*** Presentation at 14th Annual Southeastern Lake Management Conference, 13-15 Apr, Asheville, NC.

Klaine, Stephen J., and John A Smink. 2005 ***Final Report on Characterization of Hydrologic Input to Lake Greenwood.*** Clemson University Report to Saluda-Reedy Watershed Consortium.

McKellar, Hank, and Jim Bulak. 2005. ***Model Calibration Data for Dynamic Water Quality Simulations of a Eutrophic Reservoir: Nutrient-Algae-Oxygen Interactions in Lake Greenwood, South Carolina.*** South Carolina Department of Natural Resources, Freshwater Fisheries Research Lab, Final Report to Saluda-Reedy Watershed Consortium.

SAS Institute. 2002. ***JMP™ Statistics and Graphics Guide, Version 5.*** SAS Institute, Inc., Cary, NC.

South Carolina Department of Health and Environmental Control. 2001. ***Water Classifications and Standards: 61-68, Water Classifications and Standards.*** Bureau of Water State Regulations, Office of Environmental Quality Control, Columbia, SC.

South Carolina Department of Health and Environmental Control. 2004. ***State of South Carolina Integrated Report for 2004, Part II: Assessment and Reporting.*** SCDHEC, Columbia, SC.

United States Environmental Protection Agency. 1997. ***Biological Criteria: Technical Guidance for Survey Design and Statistical Evaluation of Biosurvey Data.*** EPA/822/B-97/002. Health and Ecological Criteria Division, Office of Water, Washington, DC.

United States Environmental Protection Agency. 2000. ***Guidance for Data Quality Assessment: Practical Methods for Data Analysis, EPA QA/G-9, QA00 Update***. EPA/600/R-96/084. Office of Environmental Information, Washington, DC.

United States Environmental Protection Agency. 2003. ***Guidance for 2004 Assessment, Listing, and Reporting Requirements Pursuant to Sections 303(d) and 305(b) of the Clean Water Act.*** Assessment and Watershed Protection Division, Office of Wetlands, Oceans, and Watersheds, Office of Water, Washington, DC

United States Environmental Protection Agency. 2005. Website: Welcome to STORET, EPA's largest computerized environmental data system. **http://www.epa.gov/storet/index.html**

V. APPENDIX: UNDERSTANDING JMP™ GRAPHICAL RESULTS

JMP™, a product of SAS Institute, Inc., promotes itself as "The Statistical Discovery Software." It provides convenient and powerful tools to help facilitate interpretation and understanding of statistical analyses and output. As explained in JMP™'s ***Statistics and Graphics Guide***, graphical presentations produced in the analytical process help us to understand the results through visual representations. From the chapter entitled ***One-Way ANOVA***, here are some simple explanations and guidelines for interpreting many of the plots presented in this report:

1. Each point plotted on the Y-axis for concentration is actually the ***mean*** calculated for each unit of time, i.e. year or decade, plotted on the X-axis.

2. In these simple plots of mean concentration by year, means are connected to facilitate trend assessment. This connection of points, however, does NOT imply any statistical model for continuity from one year to the next.

3. The Y-axis in most plots is a typical linear scale. The range for BOD, however, is too wide to capture as such, so its Y-axis is a log scale.

4. The standard ANOVA can also perform multiple comparison tests and visually represent these comparisons. The test used here is Tukey or Tukey-Kramer HSD (honestly significant difference), an exact alpha-level ($\alpha = 0.05$) test if sample sizes are the same and conservative ($\alpha < 0.05$ for each pair, to protect the overall 0.05) if sample sizes are different.

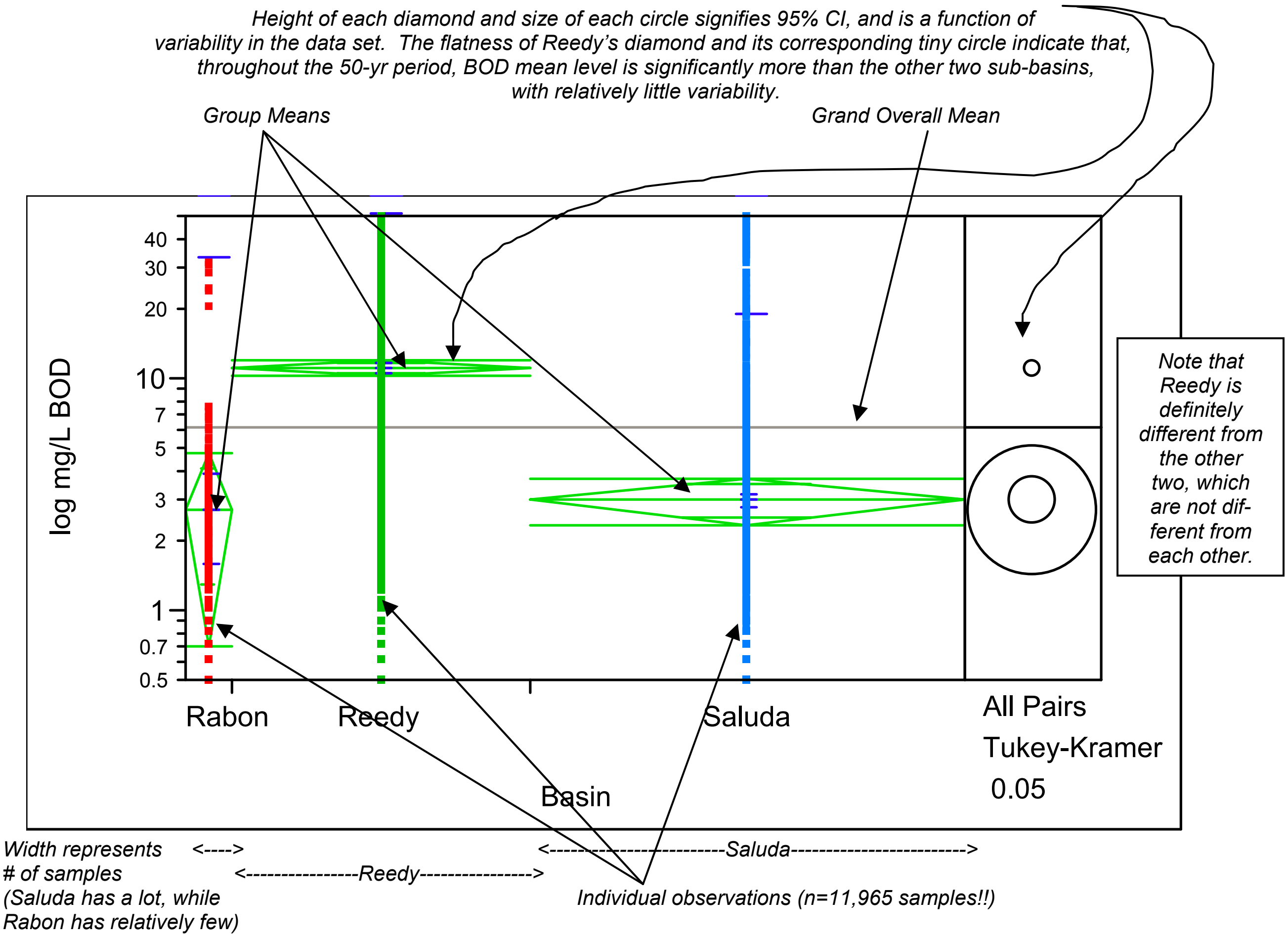
   Another test commonly used for multiple pairwise comparison tests is Student's *t*-test, but it was not used here because it is not quite as rigorous or conservative as Tukey's, especially when so many variables are involved, meaning one runs the risk of concluding differences when, in fact, they do not actually exist. In this study, however, so many of the differences are so overwhelmingly significant, with $p<0.001$ in many of the analyses, that *t*-tests would likely produce similar results.

5. Since only the *y* variable is continuous, JMP™'s one-way ANOVAs produce plots with means diamonds and comparison circles instead of continuous lines or scatterplots.
   - Each plot of means diamonds shows:
     - the overall grand mean across the middle; and
     - data points above each group along the X-axis.
   - Each means diamond illustrates:
     - its group mean across its center;
     - the 95% confidence interval (C.I.) associated with this mean, as shown by the diamond's height; and
     - the sample size of each *x* variable, because the width of each diamond along the *x*-axis is proportional (narrower diamonds are usually taller because fewer data points yield a less precise estimate of the group mean).

You can compare each pair of group means visually by examining how the accompanying comparison circles intersect. The outside angle of intersection tells you whether group means are significantly different at the 95% C.I.:

- Circles for means that are significantly different either do not intersect or intersect slightly so that the outside angle of intersection is less than 90 degrees.
- If the circles intersect by an angle of more than 90 degrees, or if they are nested, the means are not significantly different.
- The 95% C.I. determines circle size; smaller circles represent less data variability and more precise estimates of means, whereas larger circles indicate more variability.

Figure 9 (from Page 9) provides a good example of means diamonds and comparison circles:



Height of each diamond and size of each circle signifies 95% CI, and is a function of variability in the data set. The flatness of Reedy's diamond and its corresponding tiny circle indicate that, throughout the 50-yr period, BOD mean level is significantly more than the other two sub-basins, with relatively little variability.

Group Means

Grand Overall Mean

Note that Reedy is definitely different from the other two, which are not different from each other.

Width represents # of samples (Saluda has a lot, while Rabon has relatively few)

<---->

<----------------Reedy---------------->

<-------------------------Saluda------------------------->

Individual observations (n=11,965 samples!!)

6. Since both *x* and *y* variables are continuous, multiple linear regressions produce plots with simultaneous predicted lines of fit for the sub-groups. Slope of each indicates direction and rate of change in concentration over time. Multiple comparisons are made using an output table of predicted means and Tukey-Kramer HSD designations of significant differences.