Monitoring Strategies for Various Stormwater Characterization Objectives

Mostly excerpted from:

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 . 2002. 911 pages
http://unix.eng.ua.edu/~rpitt/Publications/BooksandReports/Stormwater%20Effects%20Handbook%20by%20%20Burton%20and%20Pitt%20book/MainEDFS_Book.html

Plus excerpts from recent research projects

Robert Pitt, Ph.D., P.E., BCEE
Emeritus Cudworth Professor Urban Water Systems
Department of Civil, Construction and Environmental Engineering
University of Alabama, Tuscaloosa, AL 35487 USA

1

Presentation Topics

- Experimental process
- Experimental design (determining number of samples required)
- Basic monitoring strategy
- Issues concerning stormwater that may need to be addressed
- Stormwater sampling options
- Special sampling and handling needs for solids analyses
- Basic data analyses
- Suggestions to minimize non-detectable observations
- Conclusions

2

Logical Experimental Processes:

1) Establish clear study objectives and goals
2) Conduct initial site assessment and preliminary problem identification
3) Review historical site data
4) Formulate a conceptual framework
5) Determine optimal assessment parameters

3

Logical Experimental Processes (cont.):

6) Establish data quality objectives
7) Locate sampling sites
8) Establish field procedures
9) Review QA/QC issues
10) Construct data analysis plan
11) Implement study.

4

Experimental Design – monitoring to consider variability and objectives

- The large variability of stormwater quality requires a major sampling effort to obtain useful data
- Experimental design equations can be used to estimate the number of samples needed to meet the data quality objectives (power analysis)

5

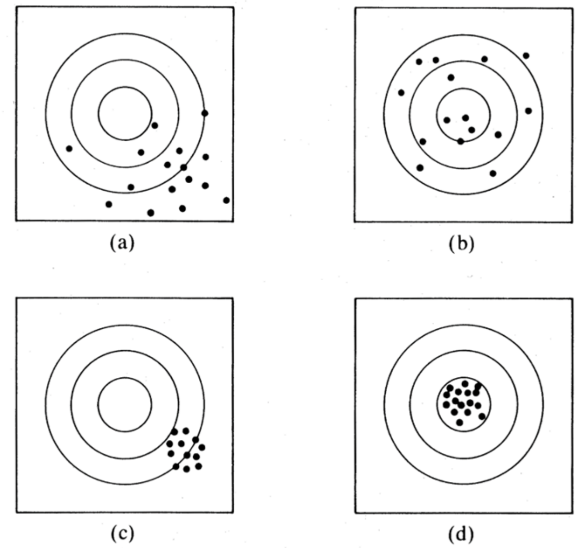
Accuracy Definitions:

(a) low precision, large bias,

(b) low precision, small bias,

(c) high precision, large bias, and

(d) high precision, small bias (the only "accurate" case)



Gilbert 1987

6

Errors in decision making are usually divided into confidence, or type 1 (alpha) and power, or type 2 (beta) errors:

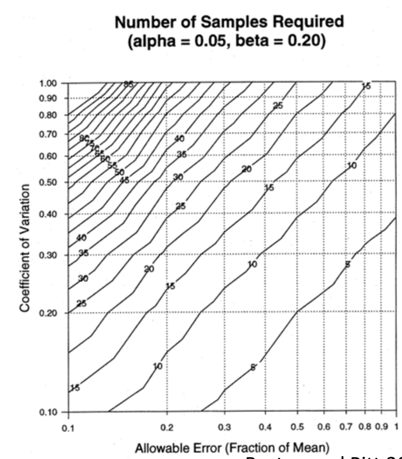
(alpha) (type 1 error) - a false positive, or assuming something is true when it is actually false. An example would be concluding that a tested water was adversely contaminated, when it actually was clean. The most common value of alpha is 0.05 (accepting a 5% risk of having a type 1 error). Confidence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

(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, beta is usually ignored (if ignored, beta 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- beta, or the certainty of not having a false negative.

7

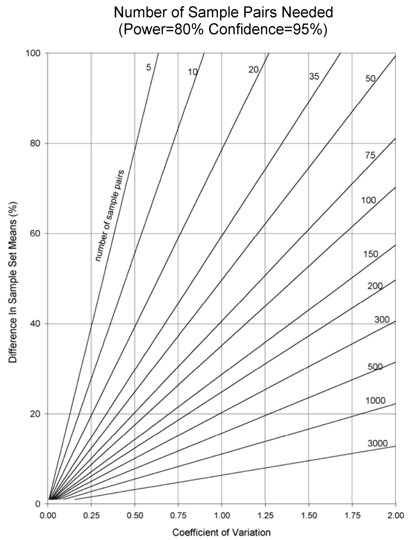
Experimental Design - Number of Samples Needed

The number of samples needed to characterize stormwater conditions for a specific site is dependent on the COV and allowable error. For most constituents and conditions, about 20 to 30 samples may be sufficient for most objectives. Most NPDES Phase 1 sites only have about 10 events, but each stratification category (land use for region of the US) usually has much more.



Burton and Pitt 2002

8



Experimental Design - Number of Samples Needed can be Large

Much information will be needed to confirm performance of stormwater controls for most constituents.

Obviously, easier to confirm removals when the differences between influent and effluent are greatest. Data sets having few samples cannot detect small and moderate differences. A power analysis before the monitoring program needs to be conducted to determine the level of control that can be detected with significance and to ensure that value meets the data quality objectives for the project.

Burton and Pitt 2002

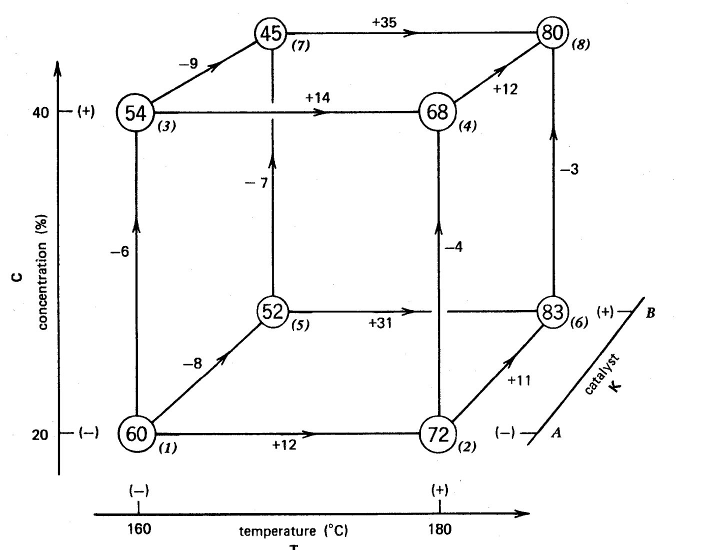
9

Factorial Analysis

a powerful experimental design and analysis tool

- A basic and powerful tool to identify significant factors and significant interacting factors.
- Use as the first step in sensitivity analysis and model building.
- Far superior to "holding all variables constant except for changing one variable at a time" classical approach (which doesn't consider interactions).
- Should be used in almost all experimental evaluations, especially valuable in controlled laboratory tests, and very useful to organize "environmental" test results.

10



Box, Hunter and Hunter 1987

11

Basic Monitoring Strategy

- Characterization of stormwater
- Scale-up of monitoring from pilot to full-scale control devices
- Need flexibility of small units and control to test many variables under large variety of conditions
- Need to verify with full-scale units to check performance under real-world conditions

12

Different Pilot-Scale Treatment Setups

13

Milwaukee, WI, Ruby Garage Public Works Maintenance Yard MCTT Tests (0.1 ha acre site)

14

Minocqua, WI, MCTT Tests (1 ha site)

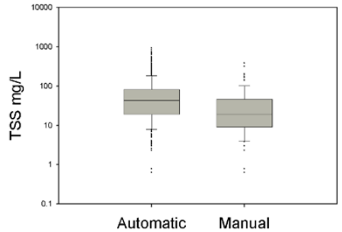
15

Field Samples – Controlled Sediment Tests

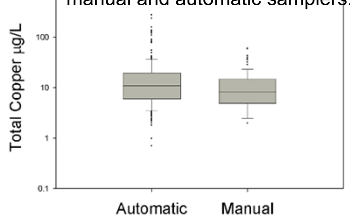
Sample triplicates made using a churn splitter; analytical duplicates made using cone splitter

16

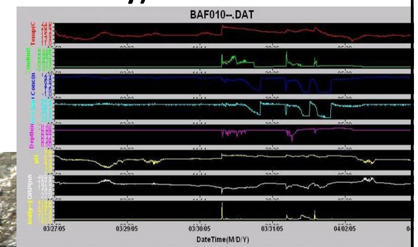
May have small biases with automatic vs. manual sampling, but automatic sampling allows unattended operation under a variety of conditions and captures complete event. Manual sampling can better represent complete range of particulate matter in sample.



NSQD data comparing results obtained in same areas using manual and automatic samplers.



The use of continuous water quality sondes can supplement other sampling programs by providing high resolution data for a variety of constituents (turbidity, temperature, DO, ORP, and conductivity).

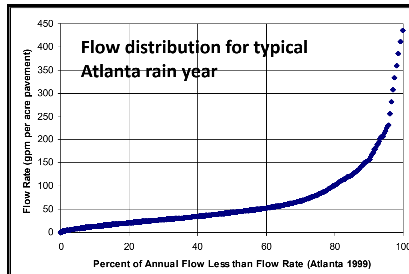


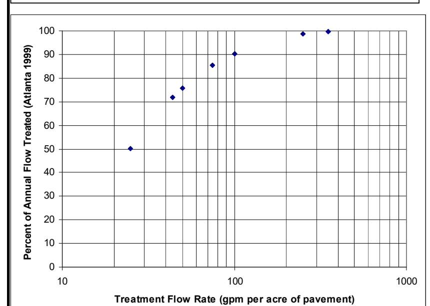
Effluent cascading onto water quality sonde

Sonde data analysis screen showing ten days of high-resolution (every 15 minute) water quality measurements

Issues Concerning Stormwater that May Need to be Addressed

- Rainfall patterns must be considered for area being studied, and accurate flow measurements are necessary as performance is commonly related to hydraulic conditions. Most flow instruments must be calibrated at the site.
- The variability of stormwater quality must be considered when designing a sampling program.
- Incorrectly reported data can have a very large effect on many statistical analyses
- Variability of stormwater quality does not always vary as anticipated ("first-flush" relatively rare, unless mostly paved areas and small drainage areas; little relationship with rain depth of event)
- Sources of flows and pollutants vary with land use and development characteristics



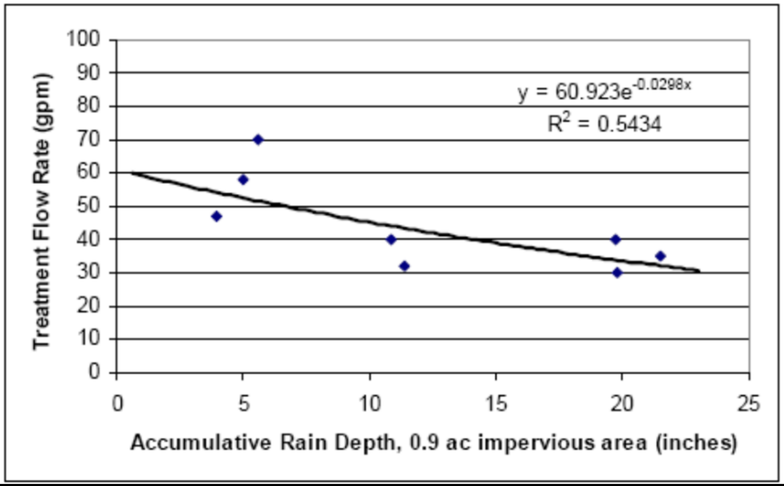


Continuous Simulation can be used to Determine Needed Treatment Flow Rates:

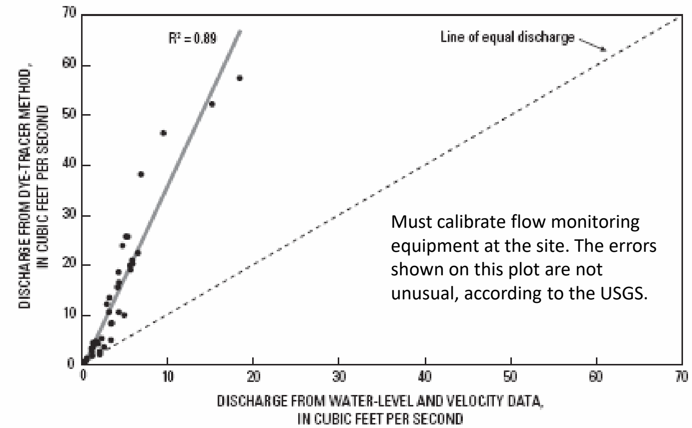
- 90% of the annual flow for SE US conditions is about 170 gpm/acre pavement (max about 450).

- treatment of 90% of annual runoff volume would require treatment rate of about 100 gpm/acre of pavement. More than three times the treatment flow rate needed for NW US.

Treatment Flow Rate Changes during 10 Month Monitoring Period of Prototype UpFlo™ Filter



21

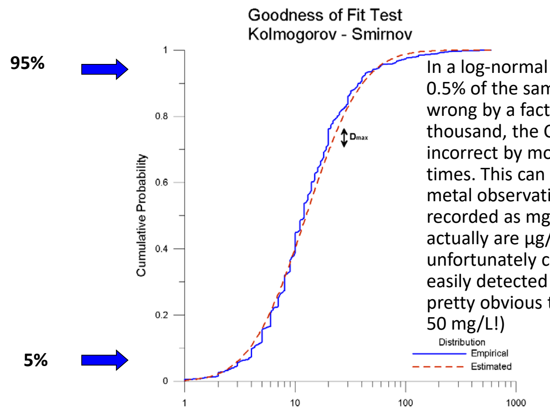


Relation between actual discharges determined using a rhodamine dye tracer and measured discharges, computed by water-level and velocity data, during free-flow conditions and actual runoff events (Selbig and Bannerman 2008)

22

Typical Stormwater Constituent Probability Distribution

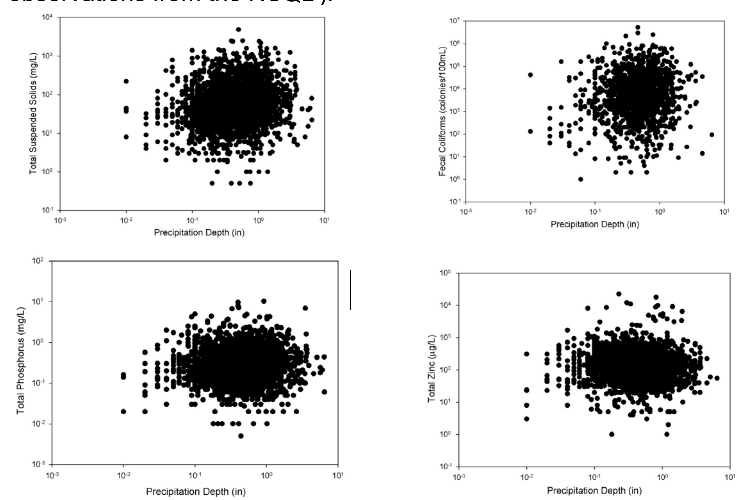
Many stormwater constituent concentrations are log-normally distributed between about the 5th and 95th percentiles of the observations (this example is for several thousand residential stormwater copper observations from the National Stormwater Quality Database, NSQD)



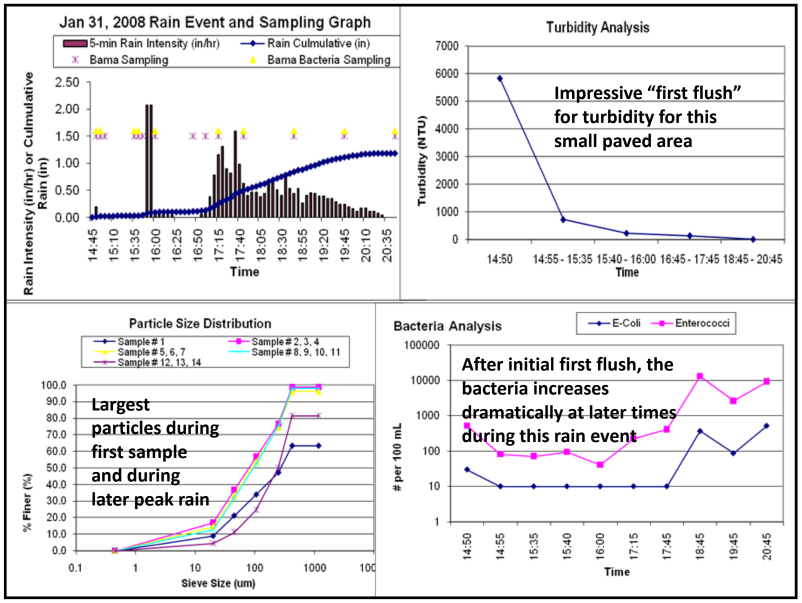
In a log-normal distribution, if 0.5% of the samples are wrong by a factor of a thousand, the COV will be incorrect by more than 10 times. This can occur when metal observations are recorded as mg/L when they actually are µg/L, an unfortunately common, but easily detected error (it's pretty obvious that Cu is not 50 mg/L!)

23

Plots of concentrations vs. rain depth typically show random patterns (several thousand residential stormwater quality data observations from the NSQD).



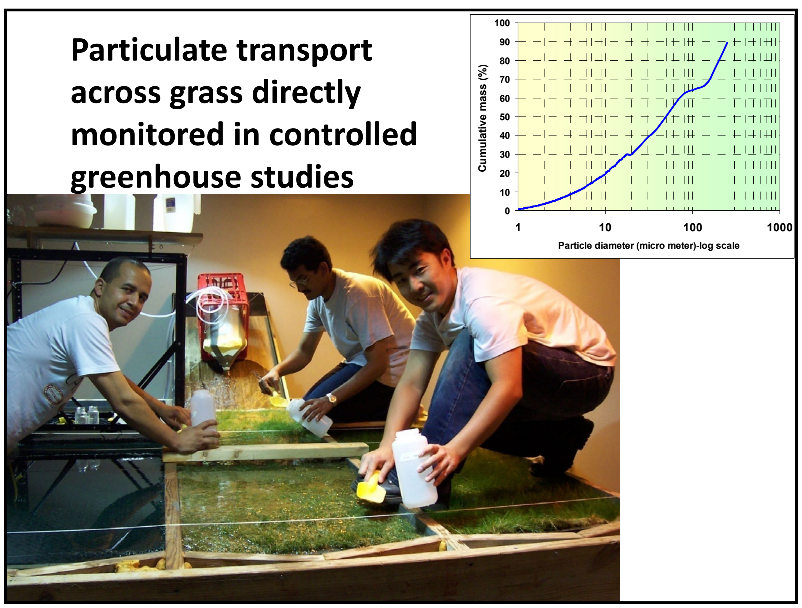
24


Jan 31, 2008 Rain Event and Sampling Graph
5-min Rain Intensity (in/hr)
Rain Culmulative (in)
Bama Sampling
Bama Bacteria Sampling
Rain Intensity (in/hr) or Culmulative Rain (in)
Time
Turbidity Analysis
Turbidity (NTU)
Impressive "first flush" for turbidity for this small paved area
Time
Particle Size Distribution
Largest particles during first sample and during later peak rain
% Finer (%)
Sieve Size (um)
Bacteria Analysis
E-Coli
Enterococci
After initial first flush, the bacteria increases dramatically at later times during this rain event
per 100 mL
Time

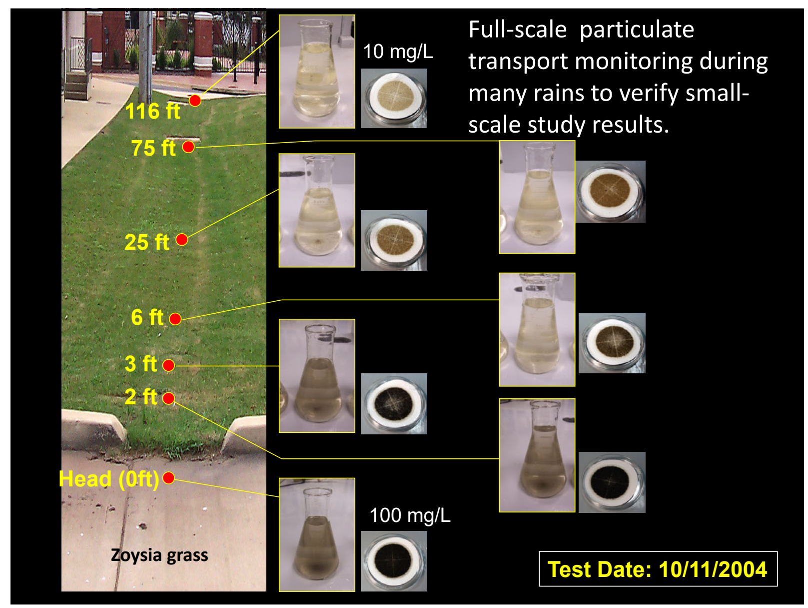
25


Stormwater Quality Variability Can be High, Even for One Location During a Single Storm Event – Can Reflect Different Sources Contributing Flows at Different Times of Runoff

26


Particulate transport across grass directly monitored in controlled greenhouse studies
Cumulative mass (%)
Particle diameter (micro meter)-log scale

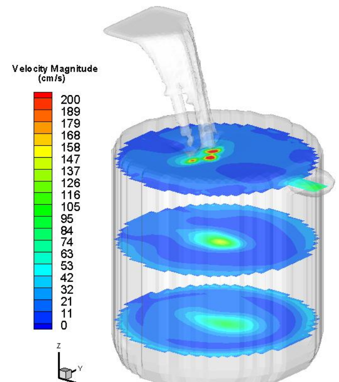
27


Full-scale particulate transport monitoring during many rains to verify small-scale study results.
10 mg/L
116 ft
75 ft
25 ft
6 ft
3 ft
2 ft
Head (0ft)
Zoysia grass
100 mg/L
Test Date: 10/11/2004

28

Physical and 3D-CFD Modeling

Scour tests of previously deposited sediment in sumps



CFD modeling being verified by full-scale 3D flow field measurements

29

Velocity and shear stress for different slopes and depths (2 ft pipe)

Depth/ Diameter ratio	Velocity (ft/sec) 0.1% slope	Shear stress (lb/ft^2) 0.1% slope	Velocity (ft/sec) 2% slope	Shear stress (lb/ft^2) 2% slope
0.1	0.91	0.0081	4.1	0.16
0.5	2.3	0.031	10	0.62
1.0	2.3	0.031	10	0.62

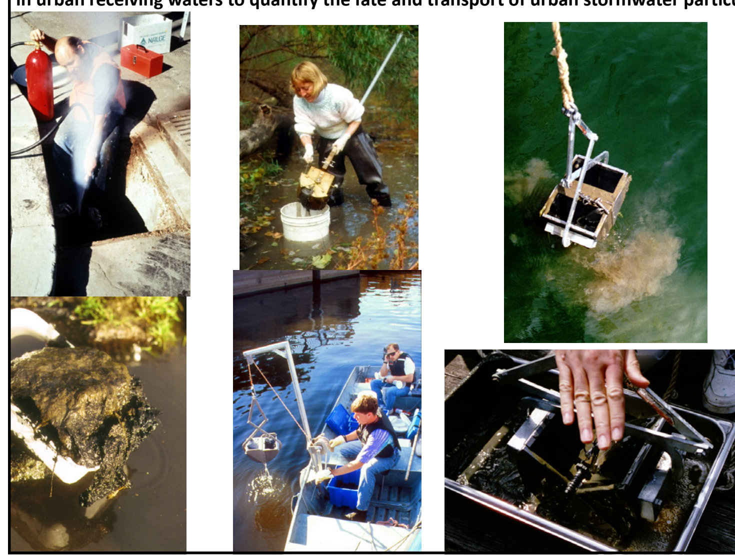
Pipes having small slopes allow particles >100 µm to settle and form permanent deposits, while pipes with large slopes will likely have moving beds of larger material.

30

Coarse Floatable Control and Monitoring Also Important in Many Areas

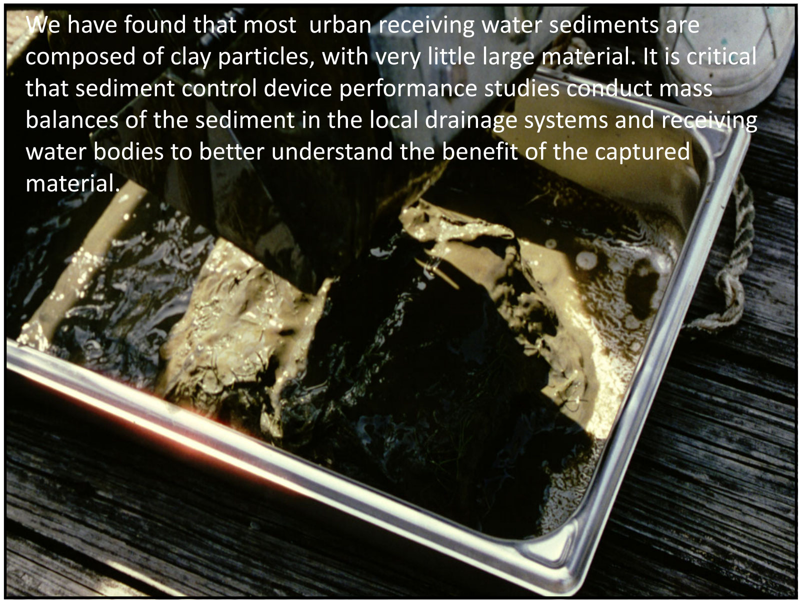
31

We have monitored sediment transport in storm drainage systems and accumulated sediment in urban receiving waters to quantify the fate and transport of urban stormwater particulates.



32

We have found that most urban receiving water sediments are composed of clay particles, with very little large material. It is critical that sediment control device performance studies conduct mass balances of the sediment in the local drainage systems and receiving water bodies to better understand the benefit of the captured material.



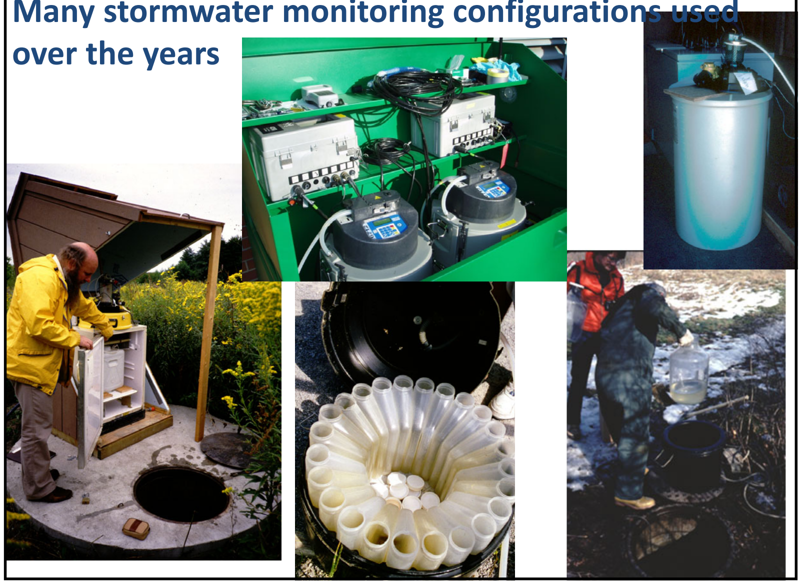
33

Stormwater Sampling

- Important to monitor sources, transport, and fate of stormwater pollutants.
- Need to program automatic samplers to collect samples under a wide-range of flow conditons.

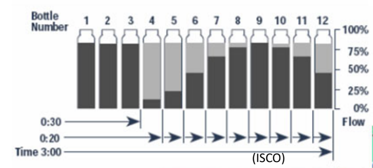
34

Many stormwater monitoring configurations used over the years



35

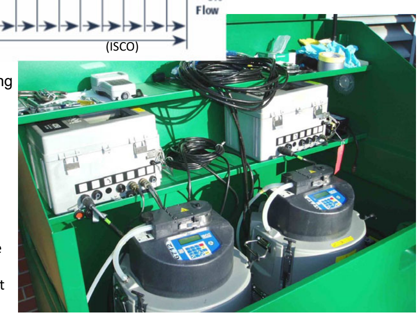
It is difficult to program an automatic sampler to collect flow-weighed samples over a wide range of flow conditions.



use a large sample base in order to accommodate a wide range of runoff events

use time-compositing instead of flow-weighted sampling and then manually composite the sample using the available flow data

Most commonly, set single sampler before each rain based on expected rain amount (when visiting site before rain to verify operation readiness)



use two samplers located at the same location, one optimized for small events, the other optimized for larger events

36

Many types of runoff monitoring have been used to understand their transport and fate, from small source areas to outfalls.

37

Fate of stormwater pollutants and actual receiving water effects need to be directly monitored

38

Stirred then settled sample, showing settleable solids (collected with automatic sampler during Madison, WI, high-efficiency street cleaner tests)

39

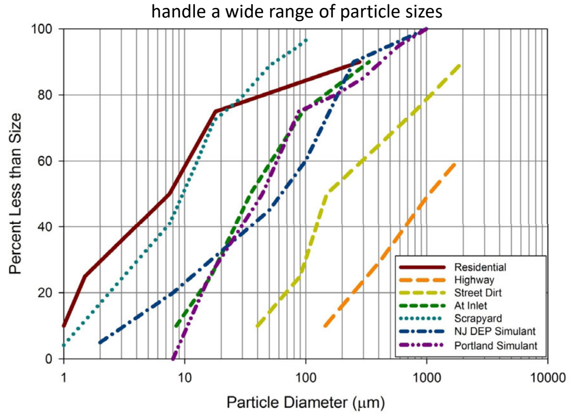
Special Sampling and Handling Needs – solids processing

- A wide range of sample characteristics need to be considered in a sampling program
- Automatic samplers are not effective in collecting large particles; recovery of particles >250 µm is usually <50%, while they can be close to 100% effective for particles <100 µm.
- In most cases, the actual errors in annual mass discharges are <10%. However, complete mass balances need to be done as part of control practice monitoring to quantify the errors and to identify the large particle fraction.
- Particle size information is one of the most important stormwater characteristics affecting treatability, transport, and fate.
- Cone splitters need to be used to divide samples for analyses and SSC (suspended sediment concentration) should be used instead of TSS for the most repeatable results.
- Discrete particle size pollutant analyses on different particle sizes can also be important for treatability and fate analyses.

40

Particle Size Distributions for Stormwater

Suitable sampling and measurement methods must be able to handle a wide range of particle sizes



Generally, larger median particle sizes at source areas and inlets and smaller median particle sizes at outfalls to receiving waters. Stream sediments will accumulate the largest particulates discharged from outfalls.

41

USGS and WI DNR Monitoring Facility for Hydrodynamic Separator Tests, Madison, WI

42

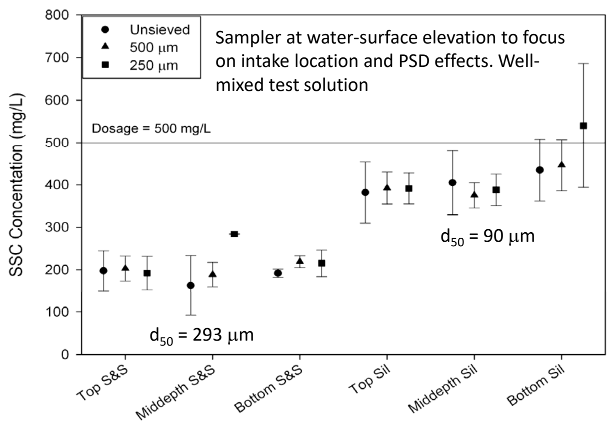
Results of Verification Monitoring of Hydrodynamic Separator (Madison, WI)

Sampled solids load in	1623 kg
Sampled solids load out	1218 kg
Trapped (by difference)	405 kg (25% removal)
Actual trapped total sediment by measuring captured material	536 kg (33% actual removal)
Total solids not captured by automatic samplers	131 kg out of 1623+131 kg missed (8%)

USGS data

43

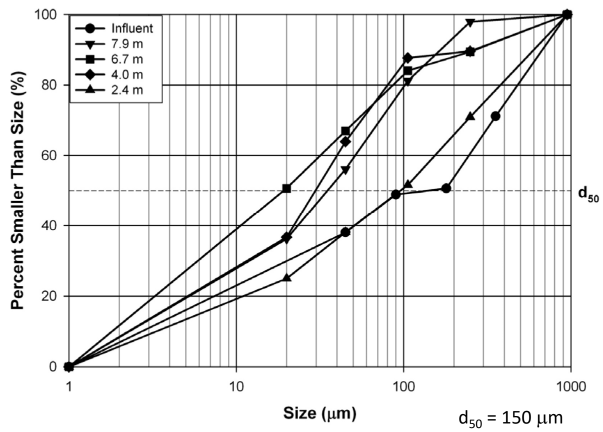
Effect of Intake Location and Solids PSD



- Well-mixed water column required to not see biases in intake location. Smaller particle sizes less subjected to bias because they stay suspended.
- Sand-sized particles much more problematic; sands > 250 µm not highly recovered.

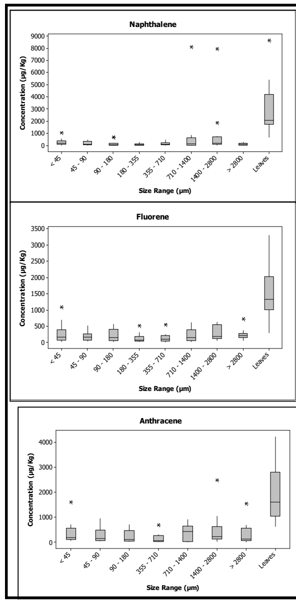
44

Sampler Height Effects

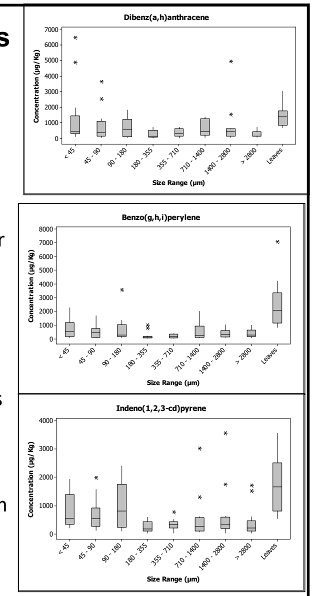


In this example, sampler heights >2.5 m resulted in fewer larger particles in sampler.

45

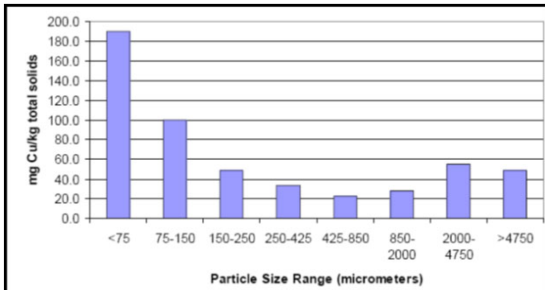


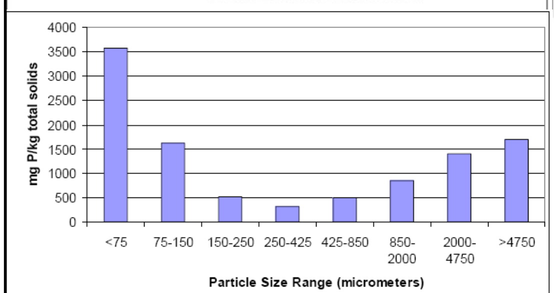
Concentrations of PAHs by Particle Sizes (urban stream sediments)



The smaller (<180μm) and larger (>710μm) particles have higher concentrations of PAHs than the intermediate sizes, but the largest mass is associated with the more abundant intermediate sizes for the urban stream sediments

46



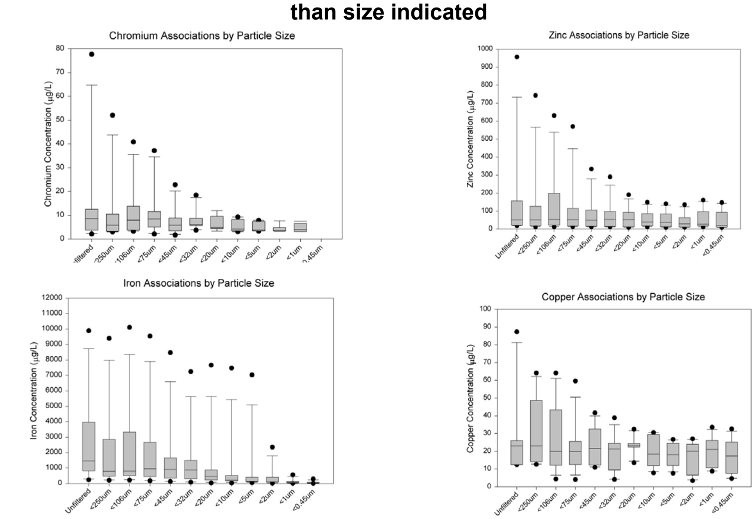


Example copper and phosphorus concentrations as a function of particle size, showing typically higher concentrations with smaller particles

47

Tuscaloosa, AL, Stormwater Outfall Samples

Residual stormwater concentrations after removal of particles larger than size indicated



48

The main objectives of most monitoring studies may be divided into two general categories:

- Characterization (quantifying a few simple attributes of the parameter of interest), and/or
- comparisons (to standards or reference conditions).

Other common objectives include identifying hot spots, examining trends, etc.

Basic Data Analyses

- The most common goals for stormwater monitoring programs are comparisons (influent vs. effluent), characterization (for different conditions), and model building (relating effluent to influent conditions).
- Simple exploratory data analysis plots are very helpful (scatter plots, line graphs, histograms).
- Probability analyses are very important to compare the data sets directly and to help select the best and correct statistical tests
- ANOVA and residual analyses must be conducted with regression analyses to verify that the test assumptions have been met.

Selection of Statistical Tests Based on Probability Distribution and Other Characteristics

Comparing Paired Observations of Data

Parametric tests (data require normality and equal variance)
- Paired Student's *t*-test (more power than non-parametric tests but only if data requirements are met)

Non-parametric tests
- Sign test (no data distribution requirements, some missing data accommodated)
- Friedman'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)

Comparing Two Independent Groups of Data

Parametric tests (data require normality and equal variance)
- Independent Student's *t*-test (more power than non-parametric tests, but only if data distribution requirements are met)

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)

Comparing 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)

53

Many Groups (cont.)

Non-parametric tests:

- 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).

54

Many Groups (cont.)

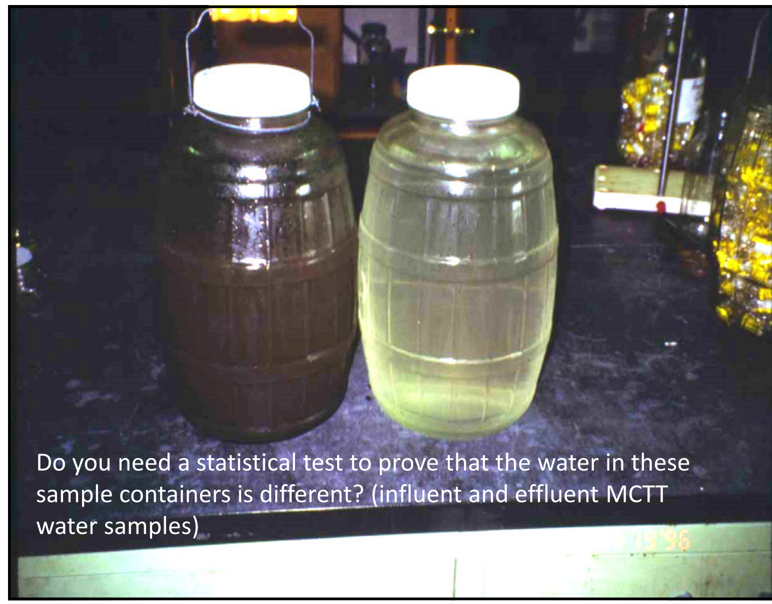
Nominal observations of frequencies (used when counts are recorded in contingency tables)

- 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).

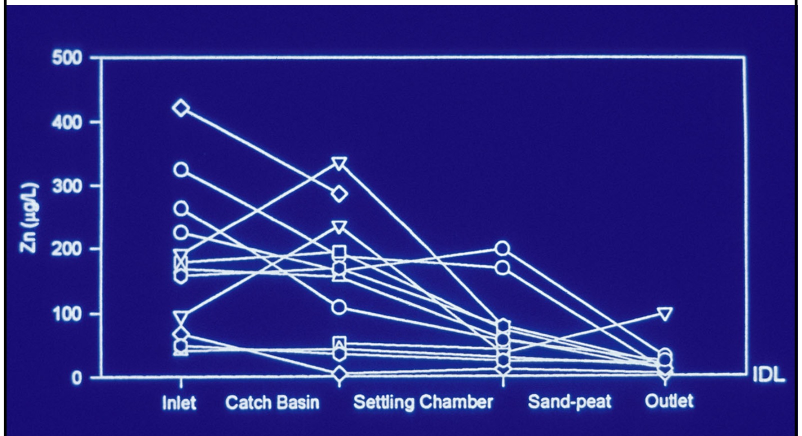
- McNamar's test (use for a "paired" contingency table, such as when the same individual or site is examined both before and after treatment)

55



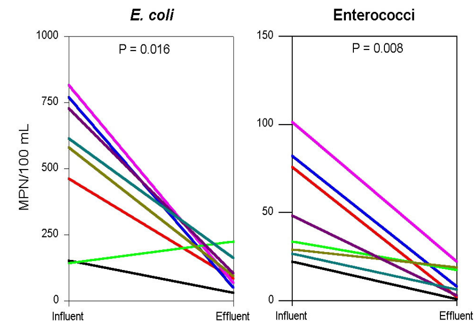
56

MCTT Pilot-Scale Test Results



Simple line plots like this can help understand the role of different treatment processes under different conditions

Line plot showing pilot-scale removal of bacteria in filter media.

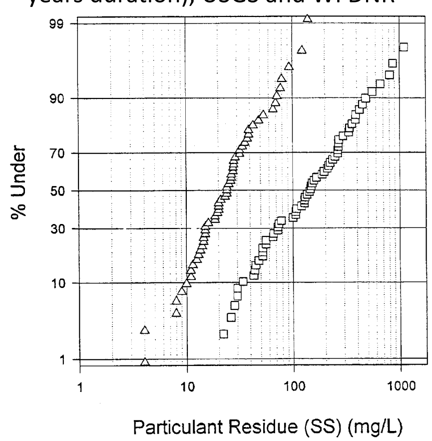


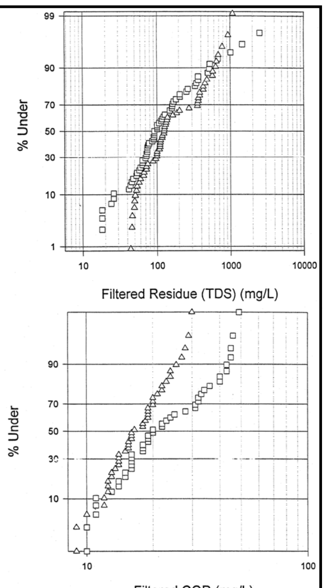
PEAT-SAND FILTER: Pilot-Scale Testing, Fall 1999

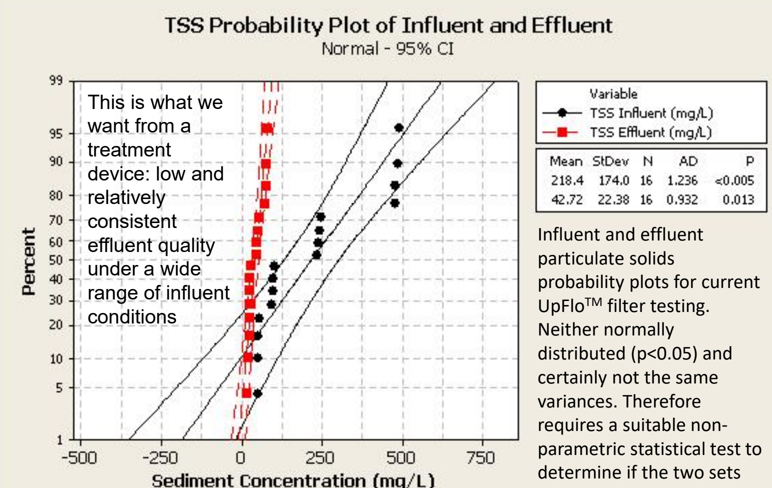
Significant reductions observed with relatively few pairs of observations due to consistent and high levels of reductions observed

Clark 1996 and 2000

Probability Plots of data from the Monroe St., Madison, WI, wet detention pond monitoring project (7 years duration), USGS and WI DNR







Influent and effluent particulate solids probability plots for current UpFlo™ filter testing. Neither normally distributed ($p<0.05$) and certainly not the same variances. Therefore requires a suitable non-parametric statistical test to determine if the two sets are from the same population. Certainly doesn't look it, but need the numbers!

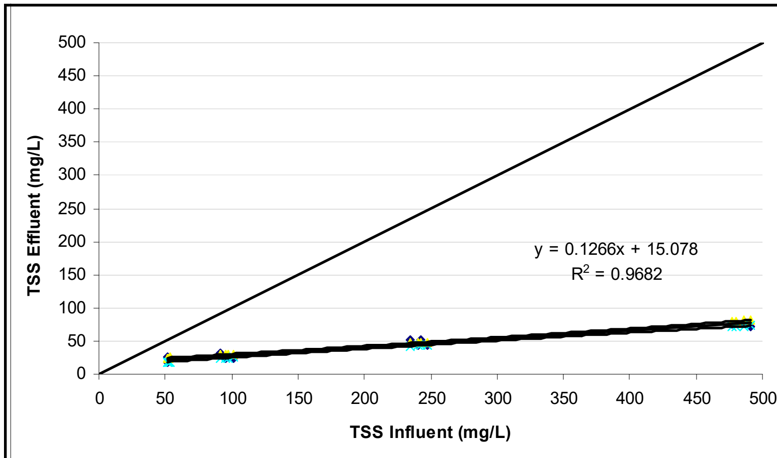
Mann-Whitney-Wilcoxon Test

n1	16
n2	16
Observed T	376
Expected T	264
Std. Dev. T	26.533
Test Statistic	4.221
p-Value (lower tail)	0.99999
p-Value (upper tail)	1.22E-05
p-Value (two tail) (Reject H0, if p-Value < 0.05)	2.43E-05
Significant Diff?	Yes

H0: Influent and Effluent Concentration is Same

Ha: Influent and Effluent Concentration is Differ

61



Scatter plot with calculated regression line and 95% confidence intervals (very narrow CIs because of good fit). Equation needs to be verified with ANOVA and residual analyses.

62

ANOVA to verify that the equation coefficients are significant (if not, remove the offending coefficient and re-analyze) and if the total equation is significant. In this case, both coefficients and the equation are highly significant, with each p<0.001)

SUMMARY OUTPUT

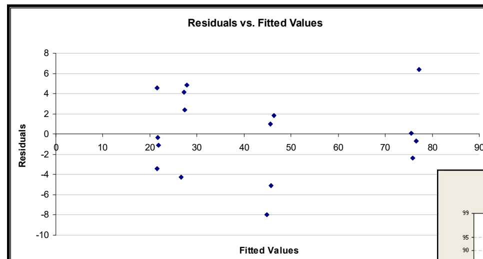
Regression Statistics	
Multiple R	0.984
R Square	0.968
Adjusted R Square	0.966
Standard Error	4.132
Observations	16

ANOVA

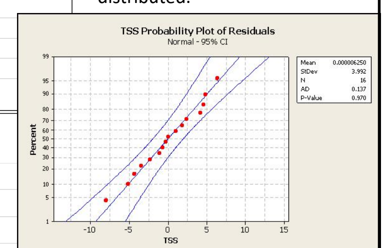
	df	SS	MS	F	Significance F
Regression	1	7275.225	7275.225	426.021	0.000
Residual	14	239.080	17.077		
Total	15	7514.305			

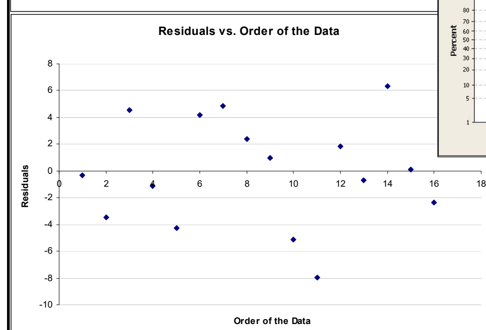
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 90.0%	Upper 90.0%
Intercept	15.078	1.691	8.914	0.000	11.450	18.706	12.099	18.057
X Variable 1	0.127	0.006	20.640	0.000	0.113	0.140	0.116	0.137

63



Must check residual behavior to verify that they are random against the fitted values and the order of the data collection, and that they are normally distributed.





The scatterplots of the residuals appear to be random, and the Anderson-Darling test shows that the residuals are normally distributed (p>0.05).

64

Detection Limits to Minimize Occurrence of Non-Detectable Observations

- Simple solution: Statistical analysis problems would not occur (substitution strategies and biases) if appropriate analytical methods were used to analyze the samples.
- It is very important to select analytical methods capable of detecting the desired range of concentrations in the samples in order to reduce the numbers of censored observations to acceptable levels.
- Use minimum detection limits to obtain manageable non-detection frequencies (<5%)

65

Suggested Analytical Detection Limits for Stormwater Monitoring Programs to Obtain <5% Non-detects (must be verified with local data)

	Residential, commercial, industrial, freeway land uses	Open space land use
Conductivity	20 μS/cm	20 μS/cm
Hardness	10 mg/L	10 mg/L
Oil and grease	0.5 mg/L	0.5 mg/L
TDS	10 mg/L	10 mg/L
TSS	5 mg/L	1 mg/L
BOD_5	2 mg/L	1 mg/L
COD	10 mg/L	5 mg/L
Ammonia	0.05 mg/L	0.01 mg/L
NO_2+NO_3	0.1 mg/L	0.05 mg/L
TKN	0.2 mg/L	0.2 mg/L
Dissolved P	0.02 mg/L	0.01 mg/L
Total P	0.05 mg/L	0.02 mg/L
Total Cu	2 μg/L	2 μg/L
Total Pb	3 μg/L (residential 1 μg/L)	1 μg/L
Total Ni	2 μg/L	1 μg/L
Total Zn	20 μg/L (residential 10 μg/L)	5 μg/L

66

Statistical Methods for use with Non-Detectable Concentrations

- Sign test (when at least one observation of the pair being compared is observed in order to determine which is larger) (honest).
- Truncated probability plots of the data sets (only show the plots for the occurrence range of the observed data) (honest).
- Substitute half of the detection limits for the non-detectable values if <5% are not detected (but can't do paired comparisons using those data) (greater uncertainty).
- Extrapolate using probability plot methods to non-detected region if <40% are not detected (again, can't do paired comparisons of data, but useful to estimate frequency of exceedance, etc.) (may be misleading).

67

Conclusions

- There have been many stormwater monitoring strategies used over the years and we have learned a great deal about stormwater characteristics. It is possible to select a suitable approach based on the monitoring objectives, and to understand the limitations of the method.
- It is important to examine as many elements of the urban area stormwater pollutant mass balance as possible during monitoring activities to appreciate the component being investigated.
- Special sampling and handling is needed to obtain the best particulate solids information.
- The study objectives may require a large sampling effort to obtain statistically valid results.
- Basic data analyses are easy to perform, but care must be taken to ensure that the methods used are appropriate.

68

Conclusions, cont.

- Many tools are freely available to assist in statistically evaluating water quality data.
- Simple data plots need to be supplemented with statistical tests.
- More care needs to be spent in experimental design and planning for specific evaluations.
- Factorial tests combine good experimental design with data evaluations.
- Analytical methods must be selected to minimize non-detected values for critical constituents.
- Be very cautious with data substitutions of non-detected values.
- QA/QC is a necessary component to ensure accurate data for analyses.