Day 2: Experimental Design
Number of Samples Needed

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

Plus excerpts from various research projects

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

1

1

Outline of Presentation

- Resources
- Experimental Objectives
- Experimental Design
- Exploratory Data Analyses
- QA/QC
- Handling Non-Detected Results
- Selection of Statistical Tests
- Model Building
- Several Case Study Examples

2

2

Recommended Classic Exploratory Data Analysis Reference Books

Exploratory Data Analysis. John W. Tukey. Addison-Wesley Publishing Co. 1977. This is a classic, basic book with many simple ways to examine data to find patterns and relationships.

The Visual Display of Quantitative Information. Edward R. Tufte. Graphics Press, Box 430, Cheshire, Connecticut 06410. 1983. This is a beautiful book with many examples of how to and how not to present graphical information. He has two other books that are sequels: *Envisioning Information* 1990, and *Visual Explanations: Images and Quantities, Evidence and Narrative*, 1997.

Visualizing Data. William S. Cleveland. Hobart Press, P.O. Box 1473, Summitt, NJ 07902, 1993 and *The Elements of Graphing Data*, 1994 are both continuations of the concept of beautiful and information books on elements of style for elegant graphical presentations of data.

3

3

Recommended Classic Experimental Design Books (with some basic statistical methods)

Statistics for Experimenters. George E. P. Box, William G. Hunter and J. Stuart Hunter. John Wiley and Sons, 1978. This book contains detailed descriptions of basic statistical methods for comparing experimental conditions and model building.

Statistical Methods for Environmental Pollution Monitoring. Richard O. Gilbert. Van Nostrand Company, 1987. This book contains a good summary of sampling designs and methods to identify trends, unusual conditions, etc.

4

4

Recommended Older General Statistics Books

Statistics for Environmental Engineers. Paul Mac Berthouex and Linfield C. Brown. Lewis, 2nd ed. 2001. This excellent book reviews short-comings and benefits of many common statistical procedures, enabling much more thoughtful evaluations of environmental data.

Biostatistical Analysis. Jerrold H. Zar. Prentice Hall. 1996. A highly recommended basic statistics text book for the environmental sciences, especially with its many biological science examples.

Primer on Biostatistics. Stanton A. Glantz. McGraw-Hill. 1992. This is one of the easiest to read and understand introductory texts on basic statistics available.

Recommended Older Books for Specialized Statistical Methods

Nonparametrics: Statistical Methods Based on Ranks. E.L. Lehman and H.J.M. D'Abrera. Holden-Day and McGraw-Hill. 1975. This is a good discussion with many examples of nonparametric methods for the analysis and planning of comparative studies.

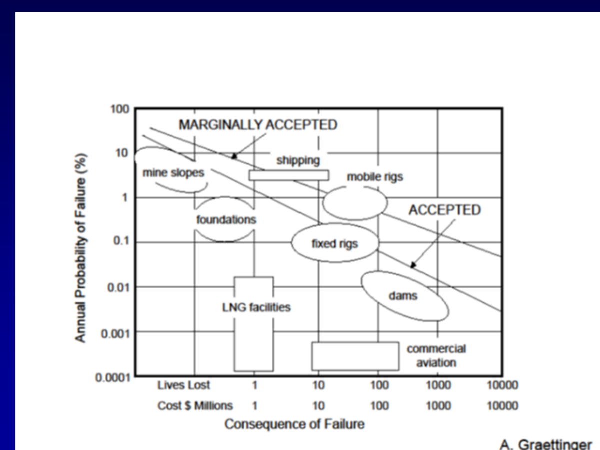
Applied Regression Analysis. Norman Draper and Harry Smith. John Wiley and Sons. 1981. Thorough treatment of one the most commonly used (and misused) statistical tools.

Experimental Objectives

First Question: What Do You Want to Do With Your Data?

- How will the data be used to arrive at conclusions?
- What will the resulting actions be? And,
- What are the allowable errors?

Allowable Errors/Acceptable Risks



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

9

9

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.

10

10

Example Data Analyses for the National Stormwater Quality Database (NSQD)

- Statistical analyses of accepted data conducted at several complementary levels.
- Focusing on robust tests that can handle left-censored data (non-detects) and missing data, plus log-normal distributions and large amounts of sample variability.
- The very large number of data observations result in confident results.

11

11

Data Analyses (cont.)

- Statistical tests using NSQD include:
  - Descriptive tests (probability plots, typical value, variation, range, Lilliefors test for normal distributions, cross tabulations, etc.)
  - Exploratory data analysis (cluster analyses, principal component analyses, Pearson correlation matrix, etc.)
  - Comparison tests (box plots, Kruskal-Wallis ANOVA on ranks test, Wilcoxon rank sum tests, Bonferroni *t*-test, etc.)
  - Trend analyses (plots, Sen's estimator of slope, seasonal Kendall test, etc.)
  - Model building based on significant factors and categories

12

12

Experimental Design

- Numbers of samples to satisfy data quality objectives
- Arrangement of experiments to maximize sensitivity and to identify major factors and interactions

13

13

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; influent vs. effluent).

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

14

14

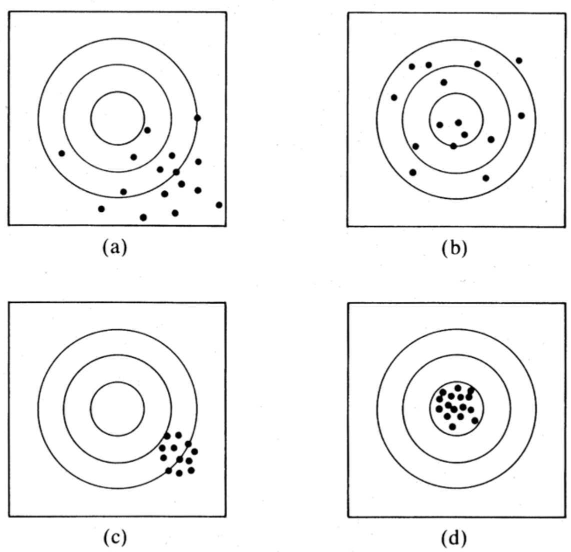
Accuracy Definitions:

(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

15

Error Types

- (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, β is usually ignored (if ignored, β is 0.5). If it is considered, a typical value is 0.2, implying accepting a 20% risk of having a type 2 error. Power is 1-β, or the certainty of not having a false negative.

16

16

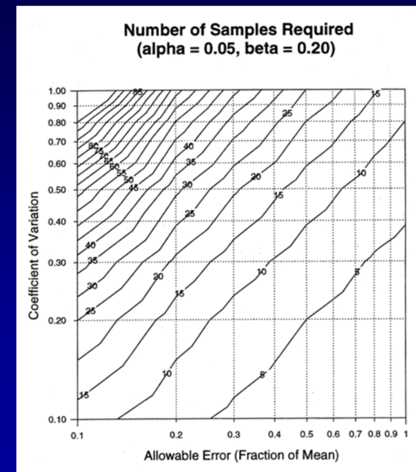
$n = [COV(Z_{1-\alpha} + Z_{1-\beta})/(error)]^2$

- n = number of samples needed
- α= false positive rate (1-α is the degree of confidence. A value of α of 0.05 is usually considered statistically significant, corresponding to a 1-α degree of confidence of 0.95, or 95%.)
- β= false negative rate (1-β is the power. If used, a value of β of 0.2 is common, but it is frequently ignored, corresponding to a β of 0.5.)
- $Z_{1-\alpha}$ = Z score (associated with area under normal curve) corresponding to 1-α. If α is 0.05 (95% degree of confidence), then the corresponding $Z_{1-\alpha}$ score is 1.645 (from standard statistical tables).
- $Z_{1-\beta}$= Z score corresponding to 1-β value. If β is 0.2 (power of 80%), then the corresponding $Z_{1-\beta}$ score is 0.85 (from standard statistical tables). However, if power is ignored and β is 0.5, then the corresponding $Z_{1-\beta}$ score is 0.
- error = allowable error, as a fraction of the true value of the mean
- COV = coefficient of variation (sometimes noted as CV), the standard deviation divided by the mean (Data set assumed to be normally distributed.)
- Z scores are combined and can't be distinguish; statistical analyses assume all alpha. Beta must be addressed in experimental design. 17

17

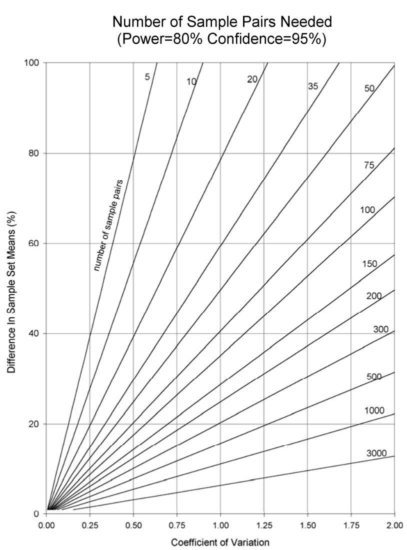
Experimental Design - Number of Samples Needed

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 combined stratification category usually has much more.



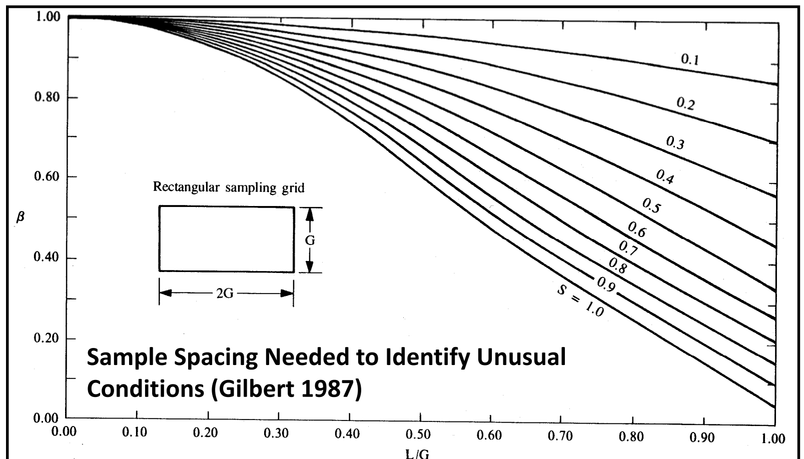
Burton and Pitt 2002

18



Burton and Pitt 2002

19



Sample Spacing Needed to Identify Unusual Conditions (Gilbert 1987)

The figure relates the ratio of the size of a circular hot spot to the rectangular grid dimensions (sampling spacing) to the probability of detection. β is the probability of not finding the spot, while S is shape factor for the hot spot (S = 1 for a circular spot, while S = 0.5 for an elliptical spot). For example, if a semi-elliptical spot was to be targeted (S=0.7), and the acceptable probability of not finding the spot was set at 25% (β = 0.25), the required L/G ratio would be about 0.95, with the rectangular width about equal to the minor radius of the desired target. 20

20

Experimental Design Example using Preliminary Data

preliminary data set #1	preliminary data set #2
60	26
55	22
65	26
84	22
75	45
38	58
98	25
39	58
55	59
48	45

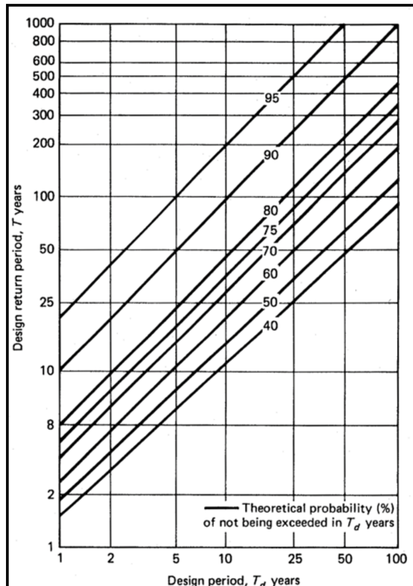
	Set A	Set B
mean:	61.7	38.6
standard deviation:	19.32	16.00
COV:	0.31	0.41

u1 =	61.7
u2=	38.6
u1-u2=	23.1
avg st dev =	17.66
avg COV =	0.36
% difference of means	37.44

Example Sample Needs for Different Levels of Confidence and Power, based on preliminary data:

False pos. rate: (Confid.)	α	1- α	$Z_{1-\alpha}$	False neg. rate: (Power)	β	1- β	$Z_{1-\beta}$	# of required pairs: n
97.50%	0.025	0.975	1.96	95%	0.05	0.95	1.645	15.2
95%	0.05	0.95	1.645	90%	0.1	0.9	1.28	10.0
95%	0.05	0.95	1.645	80%	0.2	0.8	0.847	7.3
90%	0.1	0.9	1.28	80%	0.2	0.8	0.847	5.3
80%	0.2	0.8	0.847	50%	0.5	0.5	0	0.8

Power is determined by the experimental design and by obtaining sufficient numbers of samples. Statistical analyses calculate confidence (but that is actually a measure of confidence plus power)



Design Period and Return Period (McGee 1991)

If a construction project was to last for 2 years, but the erosion control practices need to be certain of survival at least at the 95% level, then a 40-year design storm condition must be used! Similarly, a 1,000-year design flow (one only having a 0.1% chance of occurring in any one year) would be needed if one needed to be 90% certain that it would not be exceeded during a 100-year period.

Observations of Purple Cows

from Albert and Horwitz (1988)

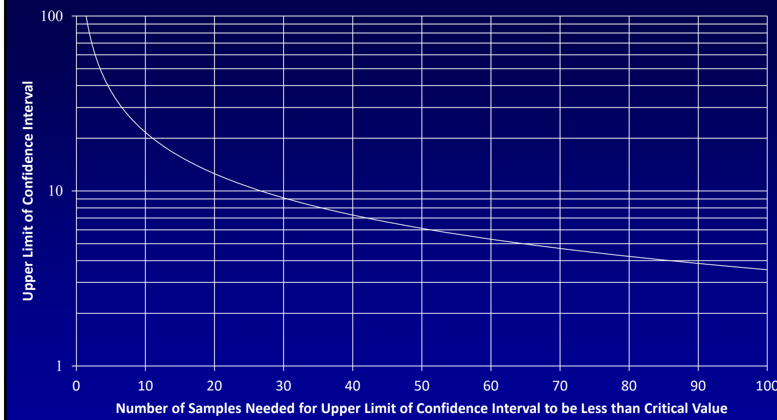
(how to be certain that something does not exist; how many samples are needed if all are non-detected?)

What is the actual percentage of cows that are purple (at a 95% confidence level), based on observations? The following formula can be used to calculate the upper limit of the 95% confidence interval:

$$(1-0)^n - (1-x)^n = 0.95, \text{ or } 1- (1-x)^n = 0.95$$

where n is the number of negative observations and x is the upper limit of the 95% confidence interval. Therefore, for a sampling of 20 cows (n = 20), the actual percentage of cows that are purple is between 0.0% and 13.9% (x = 0.139). If the sample was extended to 40 cows (n = 40), the actual percentage of cows that are purple would be between 0.0% and 7.2% (x = 0.072). The upper limit of both of these cases is well above zero and, for most people, these results generally conflict with common sense. Obviously, the main problem with the purple cow example is the violation of the need for random sampling throughout the whole population. Also, the confidence interval includes the zero value (the likely correct answer). In discussions of regression, the confidence intervals of the equation coefficients need to be examined. If doing a trend analysis, for example, if the confidence interval of the "slope" term includes the zero value, the trend is not considered significant.

Confidence Interval Upper Limits for Non-Detected Observations

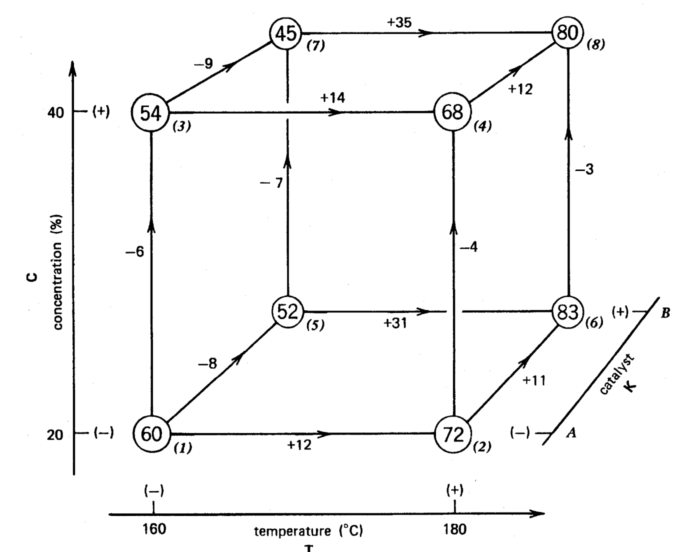


Therefore, to be certain at the 95% level that all observations would not be detected (5% upper limit of CI), about 60 samples are needed. If only have 10 samples, only 75% confident that all observations would not be detected.

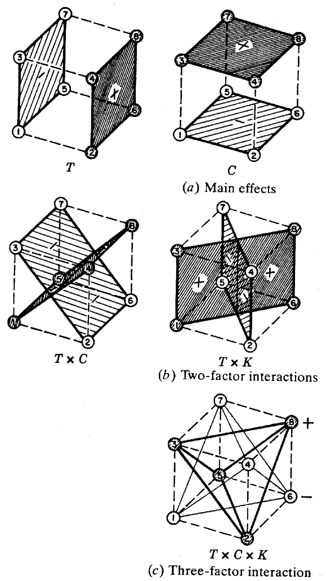
Factorial Analysis

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.



Box, Hunter and Hunter 1987

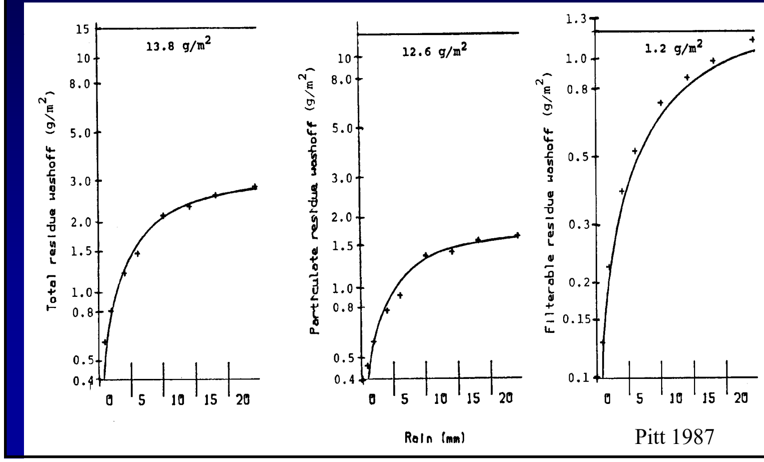


Box, Hunter and Hunter 1987

29

Washoff Plots for Heavy Rain Intensities, Dirty Streets, and Rough Pavement Textures

(conducted 8 controlled washoff tests to cover all combinations of rough vs. smooth texture, high vs. low rain intensity, and clean vs. dirty street dirt load)



Pitt 1987

30

Ratio of Available Street Dirt Loadings to Total SS Washoff Quantity

the rain intensity and pavement texture were the only significant factors affecting availability of street dirt for washoff

$$I = 0.08 \pm 0.04$$
$$T = -0.08 \pm 0.05$$

$$\hat{Y} = 0.097 + 0.04(I) - 0.04(T)$$

I+T+ (high and rough) : $\hat{Y} = 0.10$
I+T- (high and smooth): $\hat{Y} = 0.18$
I-T+ (low and rough) : $\hat{Y} = 0.02$
I-T- (low and smooth) : $\hat{Y} = 0.10$

31

Factorial Tests of Infiltration Rates for Different Biofilter Media Sand Mixtures

Media texture (median particle size), media uniformity, organic content (peat additions), and compaction were tested in sixteen columns for complete 2^4 full factorial set of tests covering all combinations, plus replicates.


Lab column construction for flow test using bioretention media: a) bottom of the columns secured with a fiberglass window screen, b) bioretention media, and c) compaction

Lab column setup for determining the infiltration rates of the soil media

32

32

Full 2^4 Factorial Design

Case	Texture	Uniformity	Organic content	Compaction	Average Fc for test conditions (cm/hr)
1	+	+	+	+	9.1
2	+	+	+	-	20.9
3	+	+	-	+	5.2
4	+	+	-	-	5.8
5	+	-	+	+	110
6	+	-	+	-	282
7	+	-	-	+	1,000
8	+	-	-	-	1,030
9	-	+	+	+	6.7
10	-	+	+	-	46.4
11	-	+	-	+	2.8
12	-	+	-	-	15.8
13	-	-	+	+	7.1
14	-	-	+	-	41.9
15	-	-	-	+	5.5
16	-	-	-	-	8.1

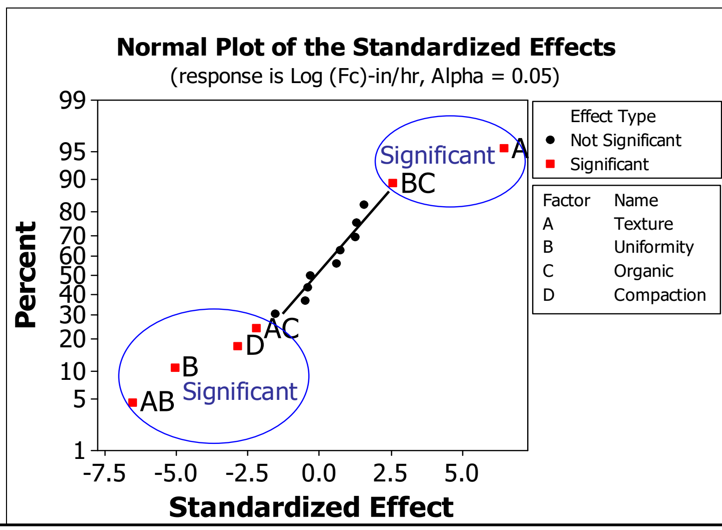
33

Estimated Effects for Fc

Estimated Effects for Fc (cm/hr)	
Texture	290.9
Uniformity	-296.4
Organic	-193.5
Compaction	-37.7
Texture*Uniformity	-298.7
Texture*Organic	-211.0
Texture*Compaction	-15.2
Uniformity*Organic	206.9
Uniformity*Compaction	21.4
Organic*Compaction	-26.9
Texture*Uniformity*Organic	207.1
Texture*Uniformity*Compaction	25.2
Texture*Organic*Compaction	-12.2
Uniformity*Organic*Compaction	17.4

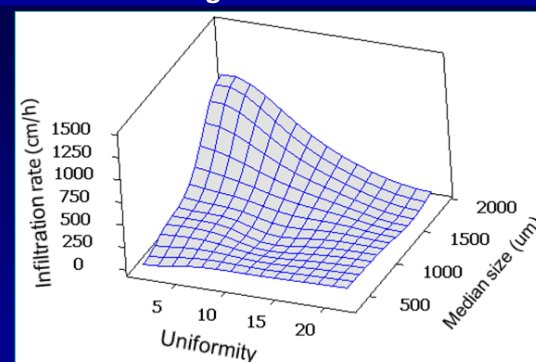
34

Probability Plot of the Calculate Effect Levels to Identify Outliers from Random Effects



35

Resulting Response Surface Model of Significant Effects Affecting Infiltration Rates



Surface plot for uniformity and texture vs. final infiltration rate for low organic content conditions. Higher infiltration rate values were observed for a mixture having low uniformity and higher median size values, as expected.

36

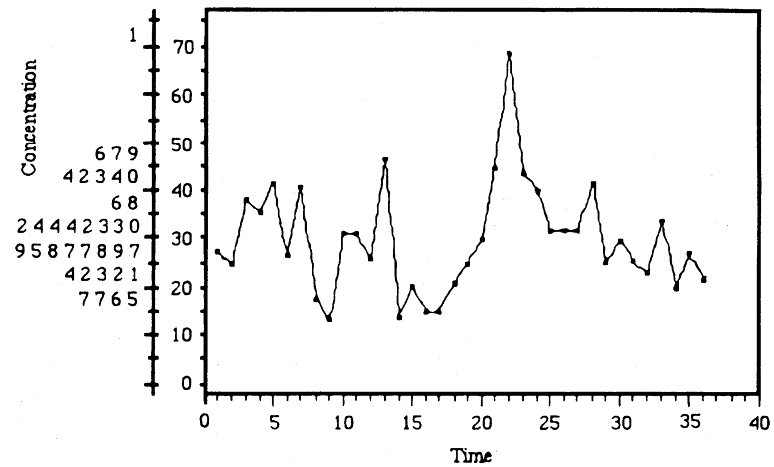
Exploratory Data Analysis Plots

- Prepare simple plots reflecting hypothesized data relationships:
  - Line plots
  - Scatter plots
  - 3D plots
  - Specialized plots (survival)
- If you can't "see" the relationship visually, statistical analyses aren't likely to be effective
- Illustrate the pattern and support with statistically measured significance and power
- Distinguish statistical significant vs. engineering importance

37

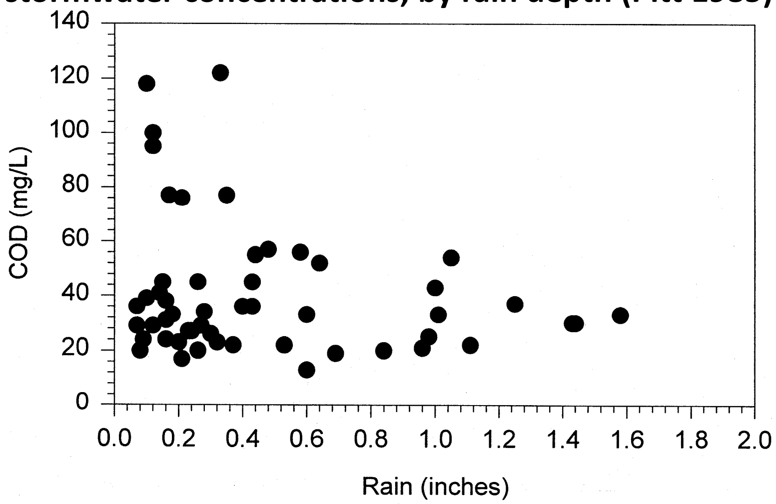
37

Digidot Plot, real-time time series and histogram (Berthouex and Brown 1994)



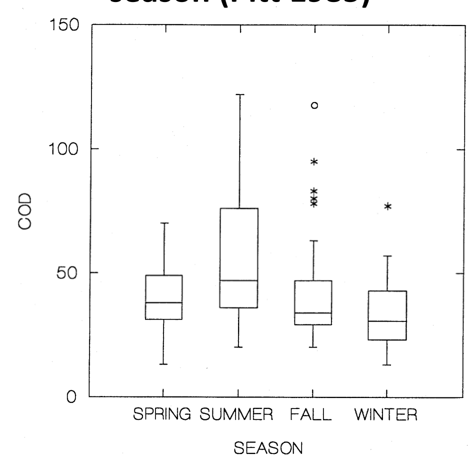
38

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



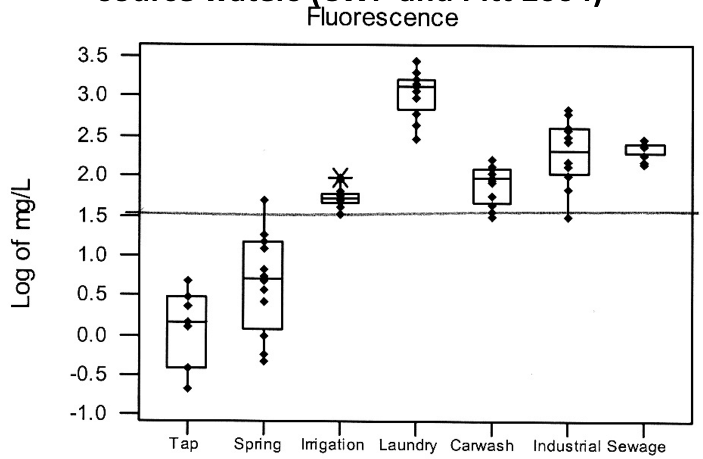
39

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



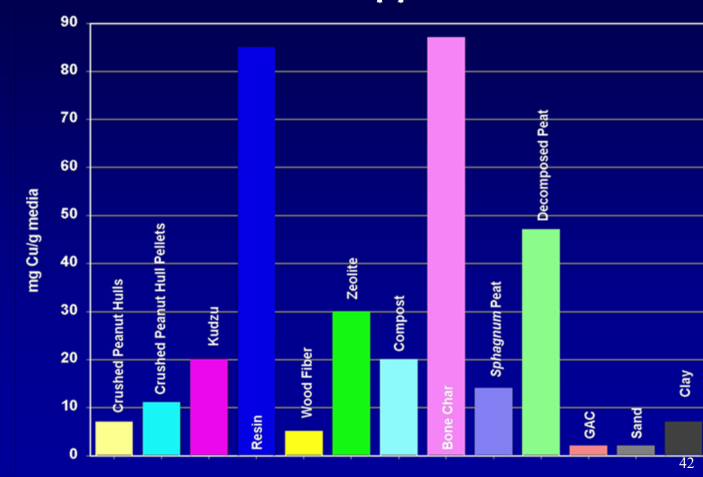
40

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



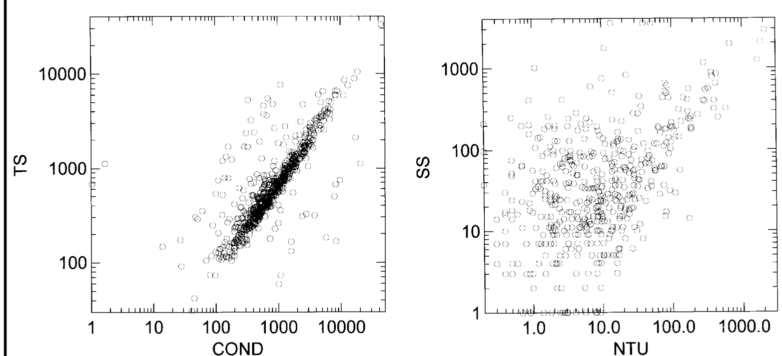
41

Bar chart contrasting media capacities for copper



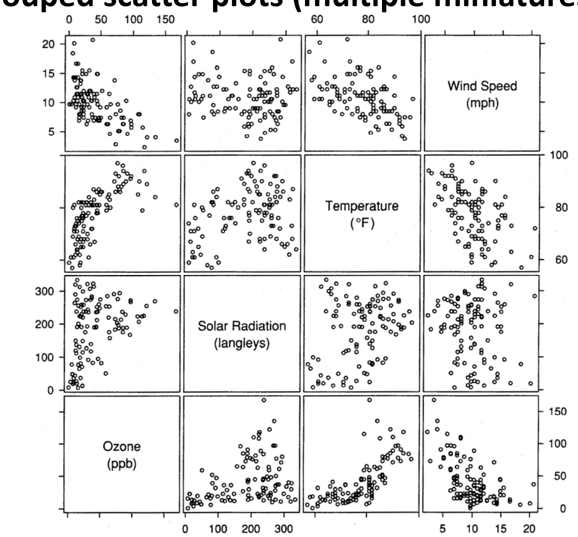
42

Scatterplots showing strong correlation (0.84) between total solids and conductivity, but surprisingly weak correlation (0.53) between suspended solids and NTU in water collected from underground telecommunications facilities (Pitt, *et al*. 1999)



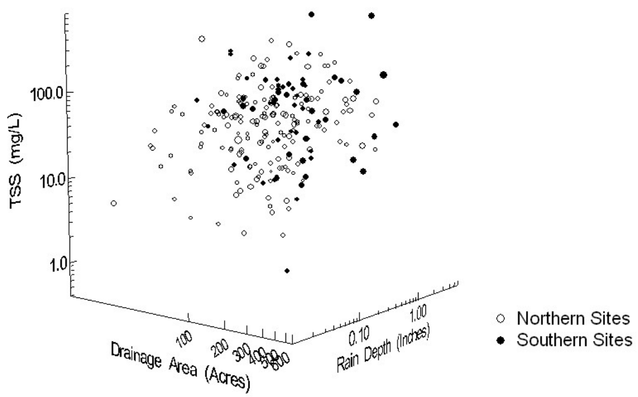
43

Grouped scatter plots (multiple miniatures)



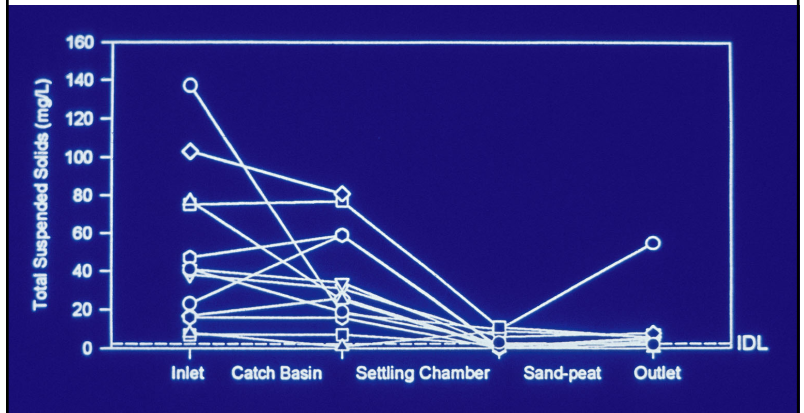
44

3-D scatterplot showing lack of obvious relationship between rain depth, geographical area, and drainage area for residential suspended solids data.



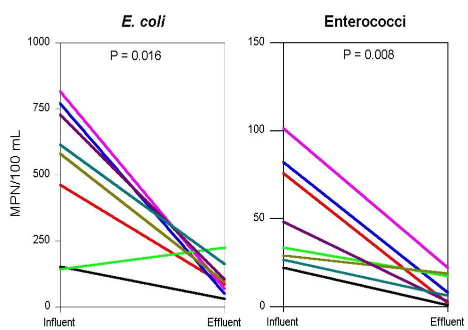
45

Line plot showing pilot-scale test results for MCTT



46

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

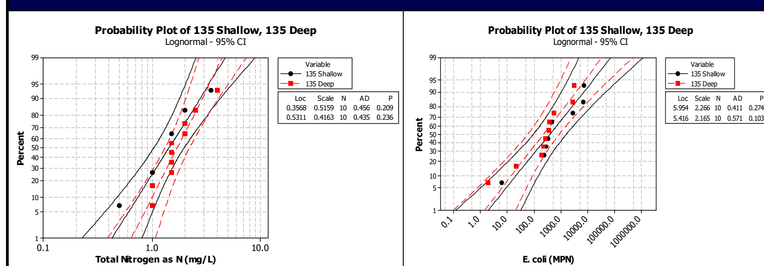


PEAT-SAND FILTER: Pilot-Scale Testing, Fall 1999

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

47

Log-normal probability plots and Anderson-Darling test statistics contrasting bacteria in shallow and deep water below dry wells



- Most of the data are seen to overlap within the limits of the 95% confidence limits, indicating that the data are likely from the same population (likely no significant differences).
- The data seem to generally fit a straight line on log-normal plots, indicating likely log-normal data distributions, and as supported by the Anderson-Darling (AD) test statistic.

48

48

QA/QC Data Reviews

- Quality assurance/quality control reviews of the submitted data for the NSQD were a major project effort and were based on:
  - data trends,
  - relationships between constituents,
  - analytical methods,
  - reasonableness of data (comparisons with historical benchmarks),
  - detection limits,
  - sampling methods,
  - sampling locations,
  - extreme values,
  - completeness of descriptions,
  - etc.

49

49

Specific Challenges of QA/QC and Statistical Methods

- Effects of left-censored data (non-detects) on data summaries and statistical tests and what are the best ways of handling these data?
- Effects of a few incorrect data in large databases (for example, if 1% of the data are off by 1000x (possible for metal reports when concentrations are actually µg/L but are reported as mg/L), can increase the COV by 10x! Fewer bad data actually make this effect on COV worse, while more have less impact.
- How much can the data be subdivided into interesting groups before we lose the ability to distinguish them (slicing and dicing)?

50

50

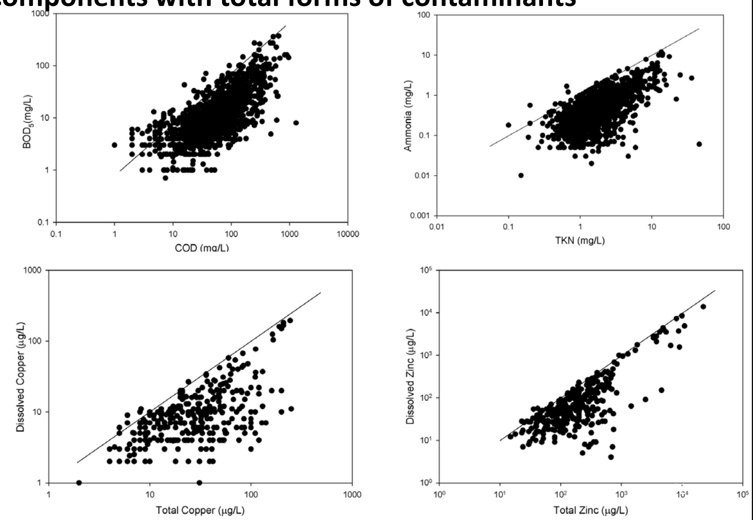
Effects of Transcription Errors

- "Common" to have metal results reported as mg/L when they really are µg/L. These should be easy to identify knowing reasonable concentration expectations (similar problem with phosphorus).
- Location errors are also relatively common.
- Many other errors are more difficult to detect/correct.
- Spent maybe 75% of our time on the NSQD doing QA/QC reviews, including comparing electronic reports with original paper lab results.
- Can't rely on submitted data without review.
- How much error can be tolerated?

51

51

Obvious plots, such as these scatterplots comparing components with total forms of contaminants

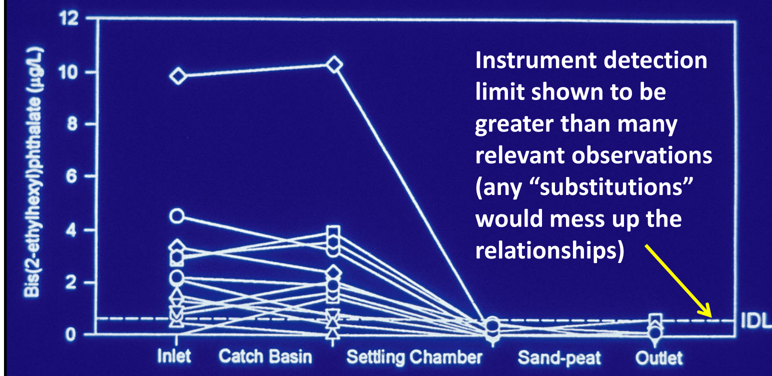


52

Data Substitutions for Non-Detectable Concentrations

- Non-detected ("left-censored") values present special problems in analyzing data (right-censored data can commonly occur for bacteria data and present special problems also).
- If only a few (<5 to <15%) of the observations are below the detection limit, these problems are not very serious.
- However, if the detection limit available results in many left-censored data (say between 25 and 75% of the observations), statistical analyses are severely limited.
- From a statistical (and engineering) viewpoint, it would be better if all concentrations determined by the analytical procedure be reported, even if they are below the designated "formal" detection limit, set using an extreme 99% confidence limit for regulatory reporting purposes. Values reported by the instruments are much better than random or constant substitutions that have no relevance to the process.

Pilot-Scale test results showing removal of a phthalate in different unit processes of the MCTT



Replacing left-censored data by a value half of the detection limit

- Estimating or replacing by half of the detection limit for levels of censoring smaller than 5% does not have a significant effect on the mean, standard deviation and coefficient of variation values.
- Substituting the censored observations by half of the detection limit produces smaller values than when using Cohen's maximum likelihood method (extrapolation of probability relationship thru "boot strapping"}. Replacing the censored observations by half of the detection limit is not recommended for levels of censoring larger than 15%.

Replacing left-censored data by probability plot extrapolations (Cohen's maximum likelihood, bootstrapping, etc.)

- The censored observations in the NSQD database were replaced using estimated values using Cohen's maximum likelihood method for each site before statistical tests.
- Because this method uses the detected observations to estimate the non-detected values, it was found to be not very accurate, and therefore not recommended, when the percentage of censored observations was larger than 40%.

Sign Test Useful for Missing Paired Data

Ranked untreated effluent	Untreated effluent concentration (sorted) (µg/L)	Paired treated effluent concentration (µg/L)	Untreated > Treated?
1	4.2	0.45	yes
2	1.5	0.52	yes
3	1.1	0.36	yes
4	0.90	1.16	no
5	0.86	nd (<0.15 µg/L)	yes
6	0.85	0.30	yes
7	0.85	nd (<0.15 µg/L)	yes
8	0.75	0.30	yes
9	0.74	0.83	no
10	0.72	nd (<0.15 µg/L)	yes
11	0.53	nd (<0.15 µg/L)	yes
12	0.51	nd (<0.15 µg/L)	yes
13	0.40	nd (<0.15 µg/L)	yes
14	0.30	nd (<0.15 µg/L)	yes
15	nd (<0.15 µg/L)	nd (<0.15 µg/L)	? (not usable)
16	nd (<0.15 µg/L)	0.37	no
17	nd (<0.15 µg/L)	nd (<0.15 µg/L)	? (not usable)
18	nd (<0.15 µg/L)	nd (<0.15 µg/L)	? (not usable)

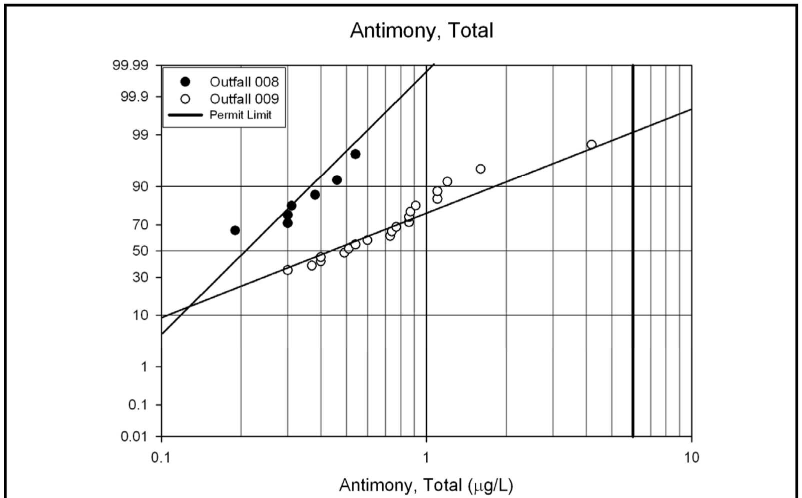
18 total events, but only 15 events with at least one non-detected value that was clearly larger or smaller than paired value

3 of the usable events had smaller concentration at the untreated site

Sign Test results (3 out of 15): p = 0.018 (a statistically significant difference was observed and not likely associated with random variation with >98% confidence)

57

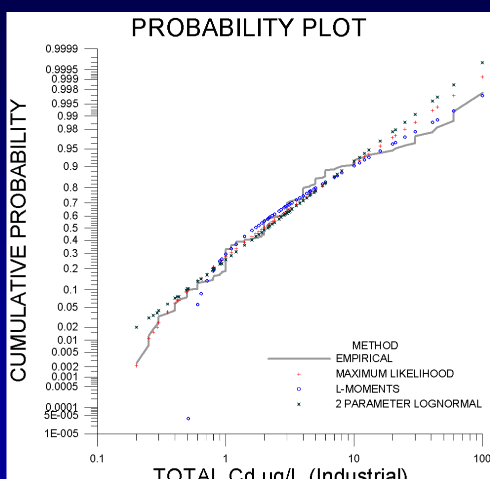
57



Truncated probability distribution plots reflecting missing data due to non-detectable observations. 40 to 60% of the data below the detection limits, but all are below the permit limit. No need to substitute values for missing observations.

58

2 parameter (log-normal) vs. 3 parameter (adding skew) to extrapolate probability plot



2 Parameter

$\mu_y = 3.96$

$\sigma_y = 6.32$

3 Parameter

$\mu_y = 4.06$

$\sigma_y = 8.22$

59

59

Descriptive statistics for lead (mg/L) comparing analyses only on observed data, replacing nd data with prob. plot extrapolations, and by replacing with half of the nd

	Deleted	Estimated	Half Detection
Observations	1728	1591	1728
% Detected	76.62 %		
Minimum (dl = 0.20)	0.20	0.03	0.10
Maximum	1200	1200	1200
Average	41.11	34.94	34.77
Median	17	13	14
Std. Deviation	80.8	75	72.5
Coef. Variation	1.96	2.15	2.08

60

60

Detection Limits to Minimize Occurrence of Non-Detectable Observations

- Simple solution: Problems would not exist 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%)

61

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

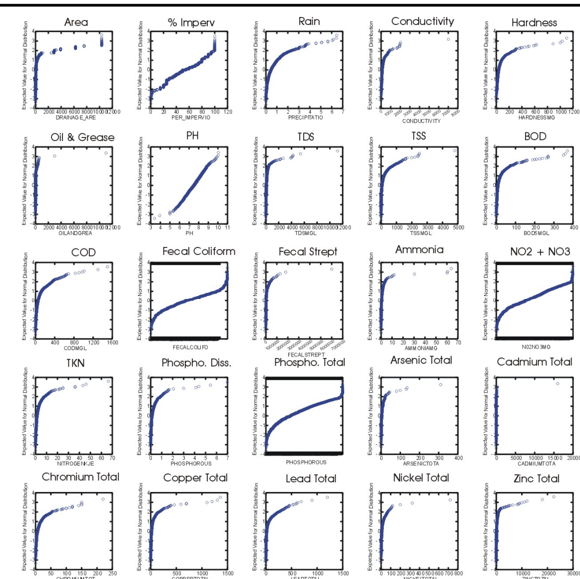
62

Statistical Methods for use with Non-Detectable Concentrations

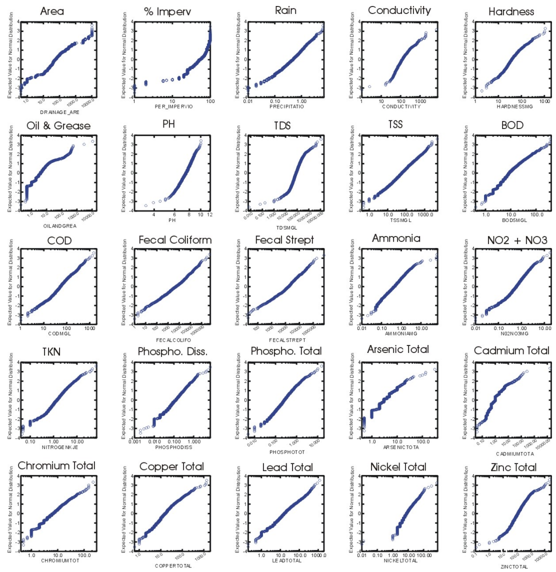
- Sign test (when at least one observation of the pair being compared is observed in order to determine which is larger) (honest).
- Truncated probability plots of the data sets (only show the plots for the occurrence range of the observed data) (honest).
- Substitute half of the detection limits for 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).

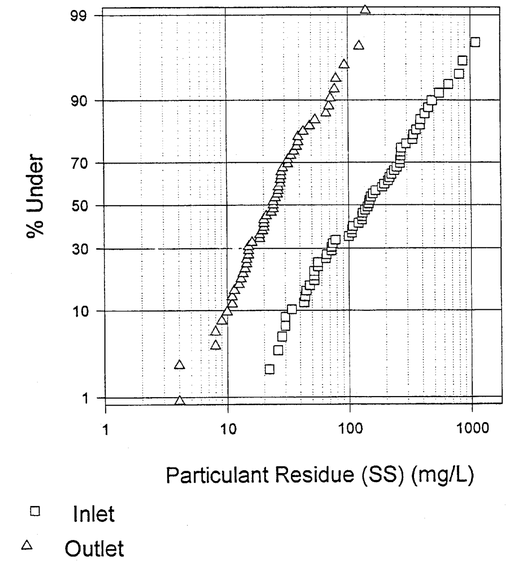
63

These data plots on regular probability scales indicate few Normal distributions (pH is most obvious and expected).

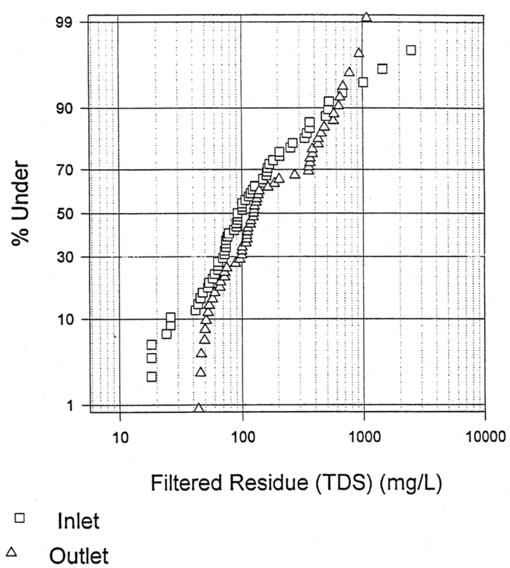


These log-normal probability plots indicate much better straight-line fits, indicating likely log-normal probability distributions of the data.

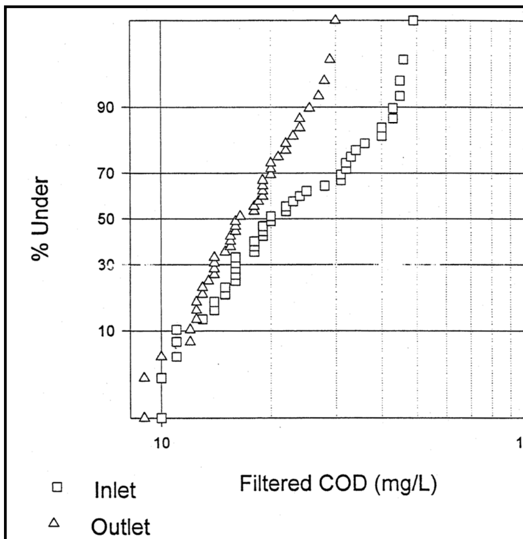




TSS influent vs. effluent concentrations Monroe St. wet pond, Madison, WI (USGS data) showing large differences with treatment.



little treatment benefit for TDS, and



odd relationship for filtered COD (large removal at high concentrations; low to no treatment at low concentrations)

Conclusions

- 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.

69