Statistical Analyses of Stormwater Characterization and Control Data

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.

Bob Pitt
University of Alabama

Recommended Exploratory Data Analysis Reference Books

Exploratory Data Analysis. John W. Tukey. Addison-Wesley Publishing Co. 1977. This is a 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.

Recommended 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.

Recommended 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 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 Design

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

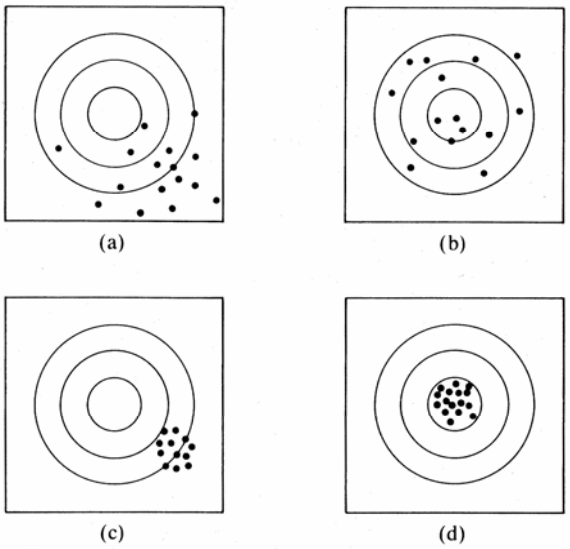
Accuracy Definitions:

(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

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

- 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 notes as CV), the standard deviation divided by the mean (Data set assumed to be normally distributed.)

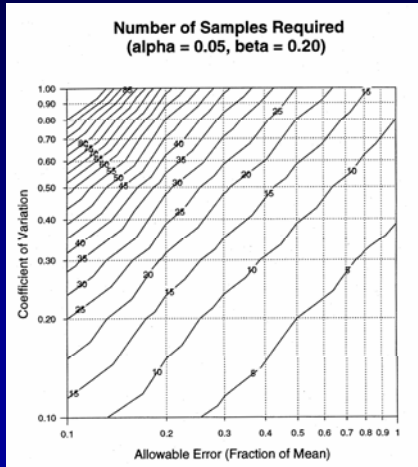
Error Types

- (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 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.

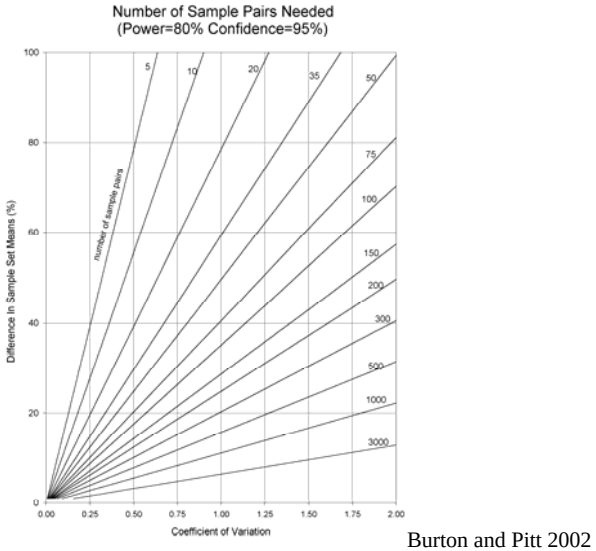
- (beta) (type 2 error) - a false negative, or assuming something is false when it is actually true. An example would be concluding that a tested water was clean when it actually was contaminated. If this was an effluent, it would therefore be an illegal discharge with the possible imposition of severe penalties from the regulatory agency. In most statistical tests, is usually ignored (if ignored, 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-, or the certainty of not having a false negative.

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 Phase 1 sites only have about 10 events, but each stratification category usually has much more.



Burton and Pitt 2002



Burton and Pitt 2002

Experimental Design Example using Preliminary Data

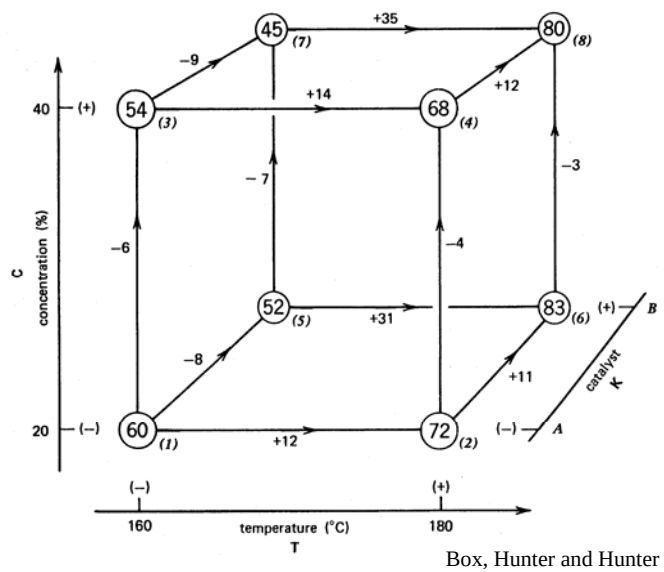
preliminary data set #1	preliminary data set #2
60	26
55	22
65	26
84	22
75	45
38	58
98	25
39	58
55	59
48	45

	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	

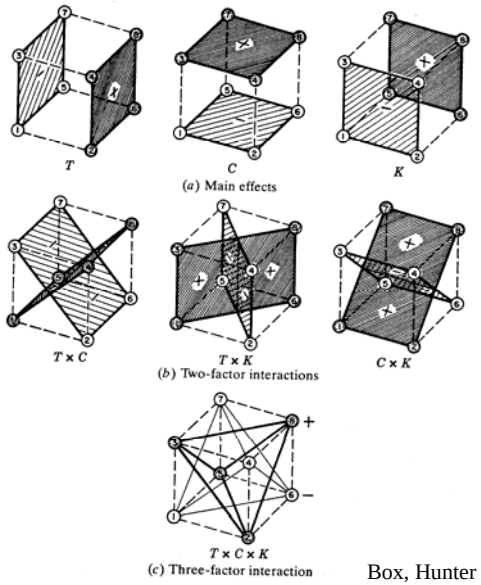
False pos. rate: (Confid.)	α	1- α	$Z_{1-\alpha}$	False neg. rate: (Power)	β	1- β	$Z_{1-\beta}$	# of pairs: n
97.50%	0.025	0.975	1.96	95%	0.05	0.95	1.645	15.2
95%	0.05	0.95	1.645	90%	0.1	0.9	1.28	10.0
95%	0.05	0.95	1.645	80%	0.2	0.8	0.847	7.3
90%	0.1	0.9	1.28	80%	0.2	0.8	0.847	5.3
80%	0.2	0.8	0.847	50%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

Factorial Analysis

- 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

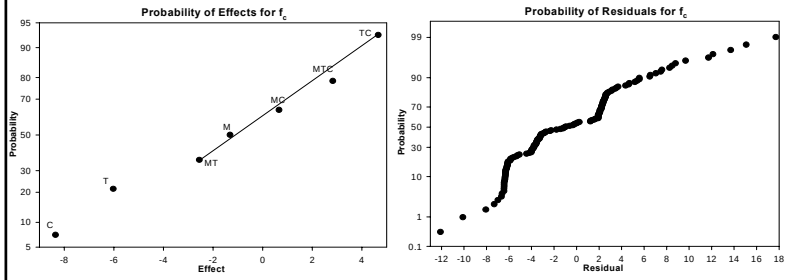
Moisture (Wet=+/Dry=-)	Texture (Clay=+/Sand=-)	Compacted (Yes=+/No=-)	Factorial Group	Average	Standard Error	Number
+	+	+	1	0.23	0.13	18
+	+	-	2	0.43	0.50	27
+	-	+	3	1.31	1.13	18
+	-	-	4	16.49	1.40	12
-	+	+	5	0.59	0.35	15
-	+	-	6	7.78	4.00	17
-	-	+	7	2.25	0.98	21
-	-	-	8	13.08	2.78	24
			overall average	5.27		
			calculated polled S.E	1.90		

Factorial Group	effects	rank	Prob
C	-8.35	1	7.14
T	-6.02	2	21.43
MT	-2.55	3	35.71
M	-1.31	4	50.00
MC	0.66	5	64.29
MTC	2.83	6	78.57
TC	4.66	7	92.86

$f_c = 5.27 \pm (T/2) \pm (C/2)$

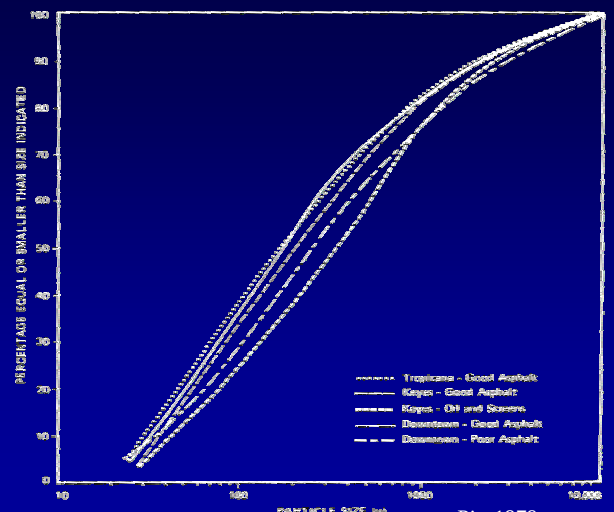
$f_c = 5.27 \pm (-6.02/2) \pm (-8.35/2)$

T	C	Calculated Values
+	+	-1.92
+	-	6.43
-	+	4.10
-	-	12.45



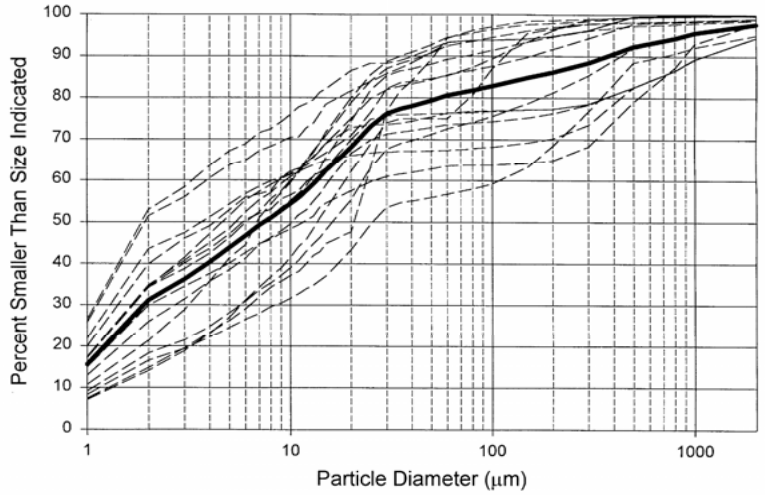
Pitt, *et al.*, 1999

Particle Size Distribution of Street Dirt

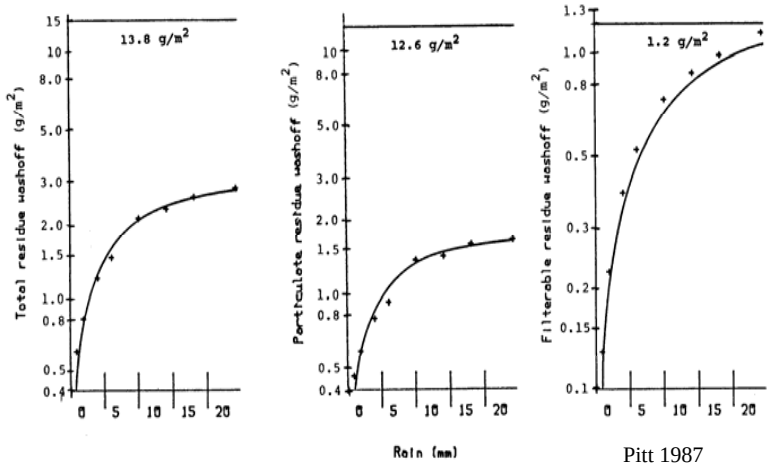


Pitt 1979

Measured Particle Sizes, Including Bed Load Component, at Monroe St. Detention Pond, Madison, WI



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



Pitt 1987

Ratio of Available SS to Total SS Street Dirt Loadings

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

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

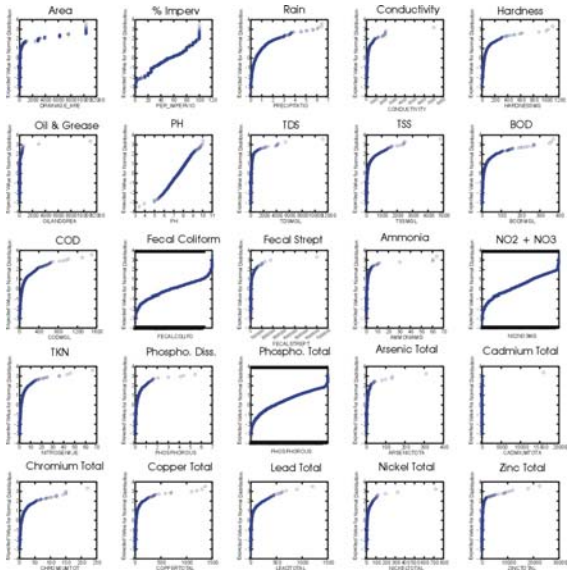
I+T+ (high and rough) : $\hat{Y} = 0.10$
I+T- (high and smooth): $\hat{Y} = 0.18$
I-T+ (low and rough) : $\hat{Y} = 0.02$
I-T- (low and smooth) : $\hat{Y} = 0.10$

Pitt 1987

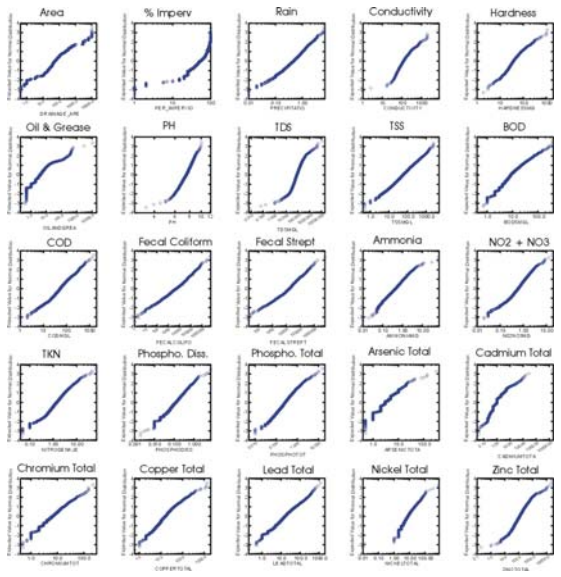
Exploratory Data Analyses

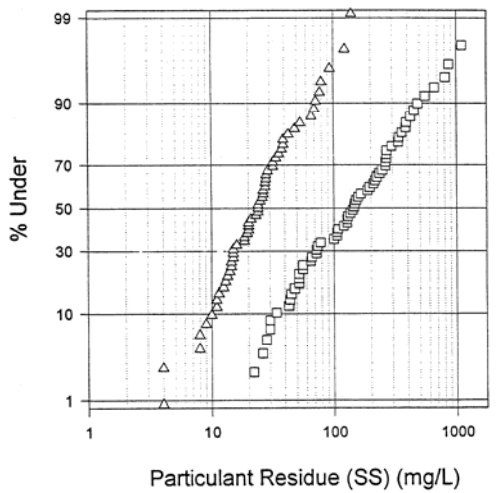
- Basic QA/QC data plots
- Probability plots and histograms
- Scatterplots
- Grouped box and whisker plots
- Simple line plots

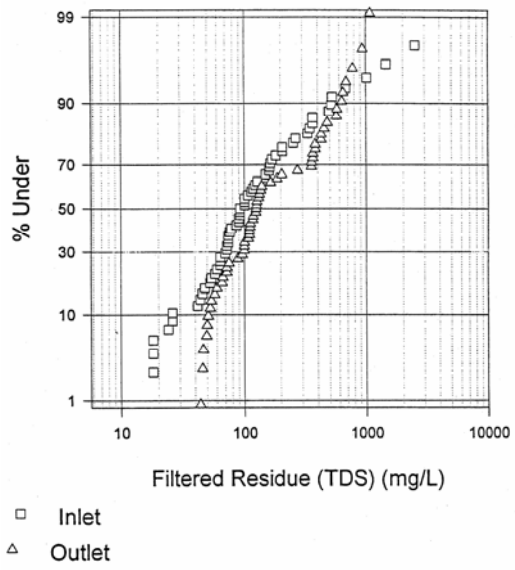
These data plots on regular probability graphs indicate few Normal distributions (pH is most obvious and expected).

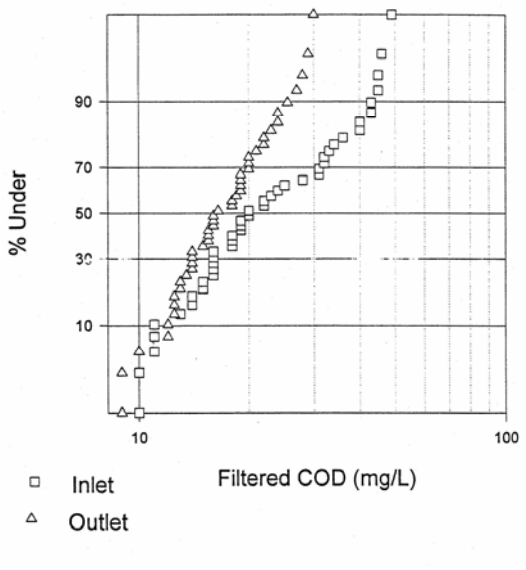


These log-normal probability plots indicate much better straight-line fits, indicating likely log-normal probability distributions of the data.

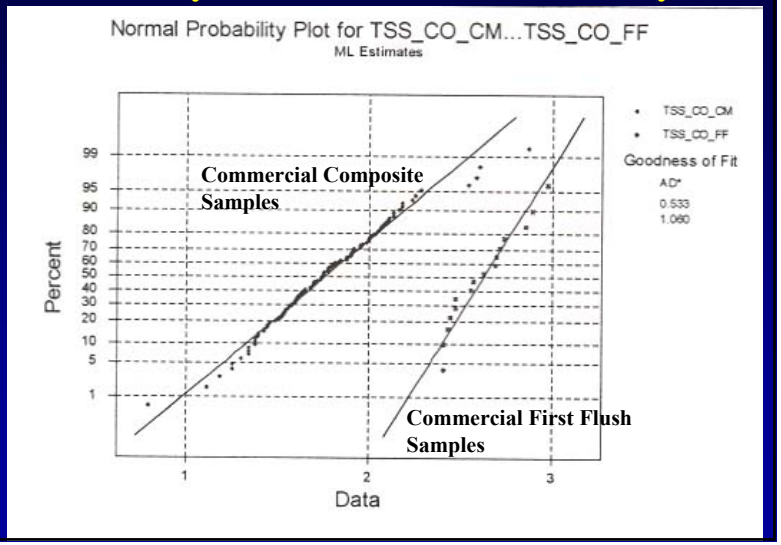




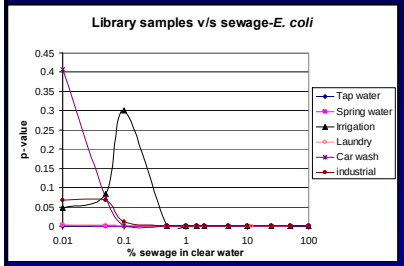




Probability Plots for First-Flush Analyses

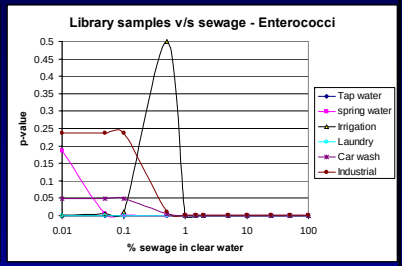


Comparison of Sewage with Dry Weather Source Samples



p=0.05, % Sewage= 0.43

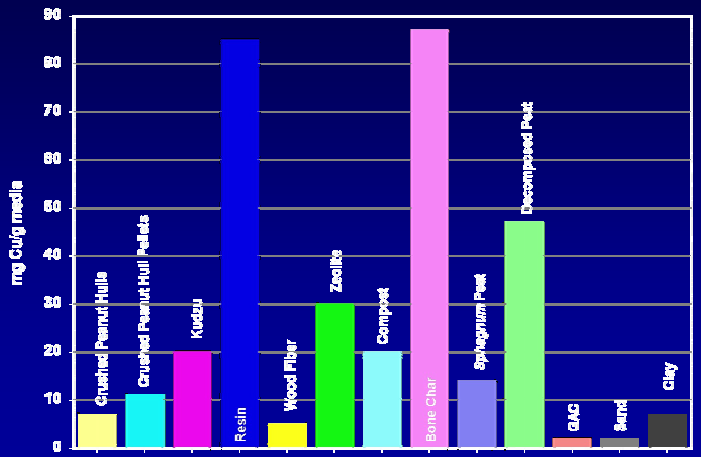
E. coli= 12,000 MPN/100 mL



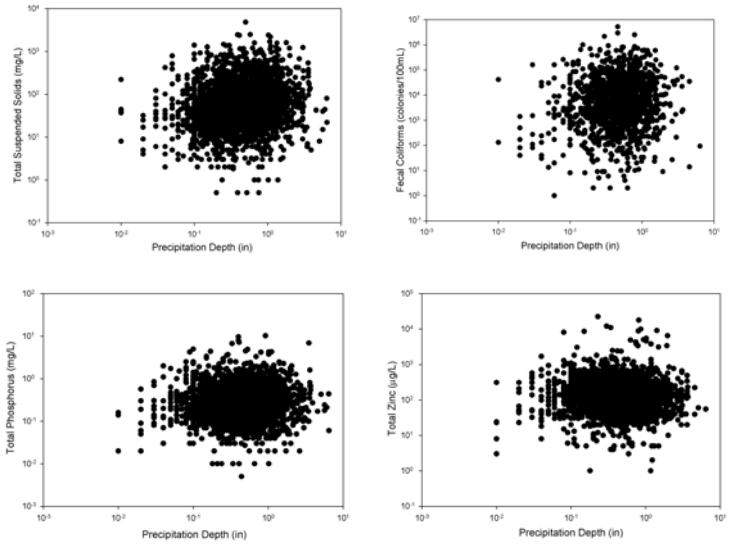
p=0.05, % Sewage= 0.95

Enterococci = 5,000 MPN/100 mL

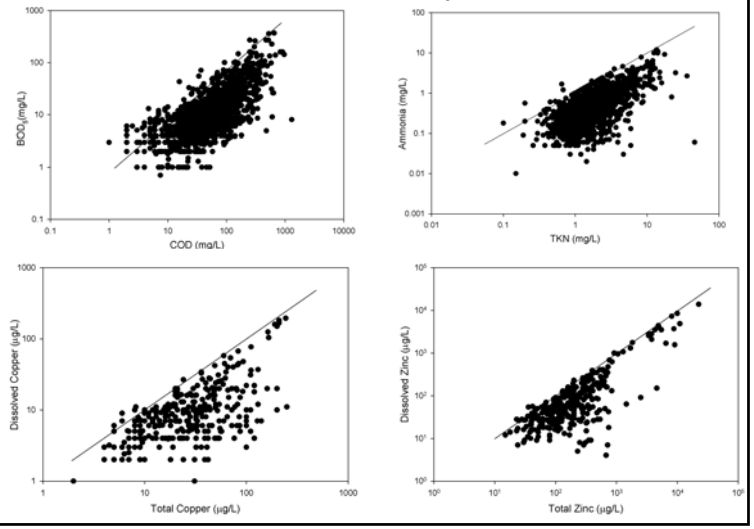
MEDIA CAPACITIES FOR COPPER

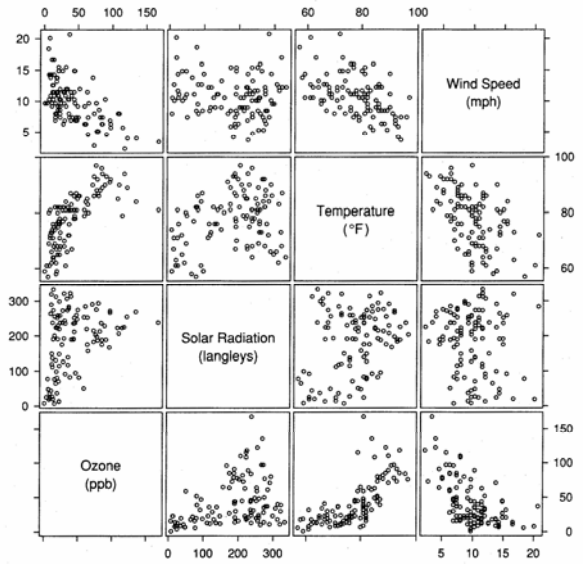


Plots of concentrations vs. rain depth typically show random patterns.

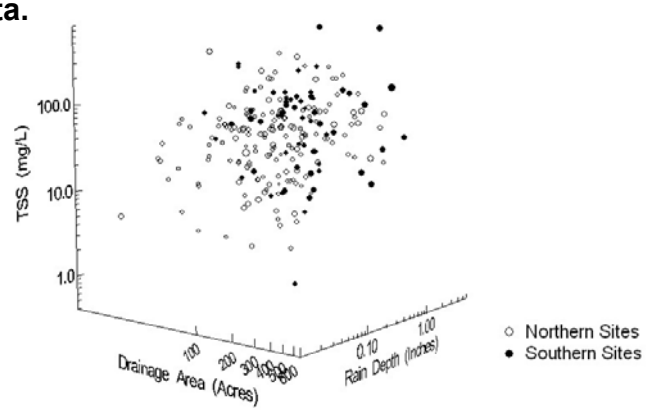


Plots of expected relationships are being used to identify data redundancies that can reduce future analytical costs.





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

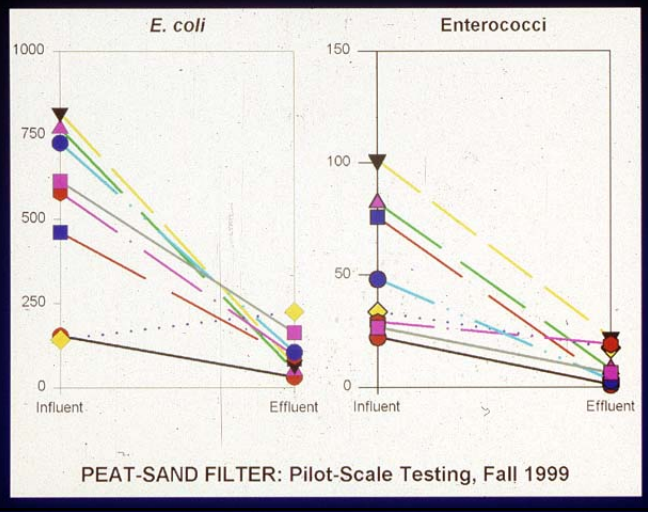


Paired observations of data

Parametric tests (data require normality and equal variance)
- Paired Student's *t*-test (more power than non-parametric tests)

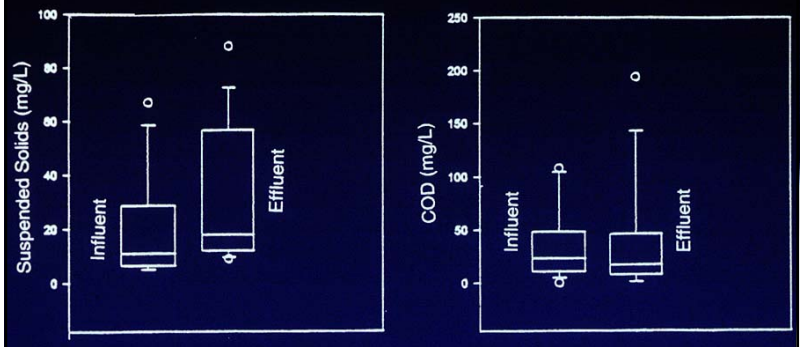
Non-parametric tests
- Sign test (no data distribution requirements, some missing data accommodated)
- Fiedman's test (can accommodate a moderate number of "non-detectable" values, but no missing values are allowed
- Wilcoxon signed rank test (more power than sign test, but requires symmetrical data distributions)

E. coli AND *Enterococci* REMOVAL

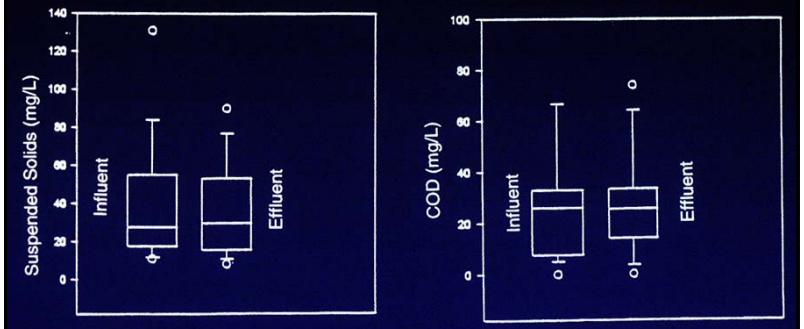


PEAT-SAND FILTER: Pilot-Scale Testing, Fall 1999

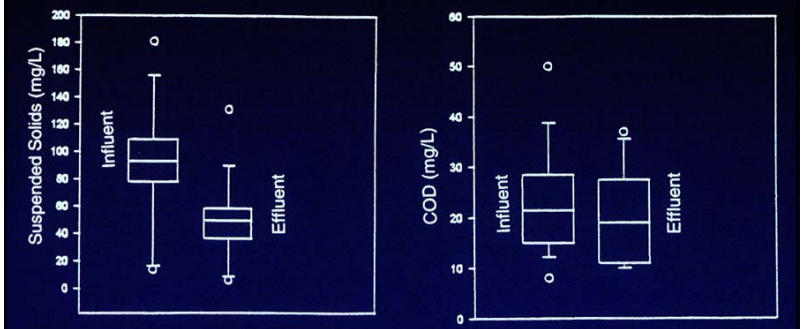
Box Plots - Coarse Screen Unit



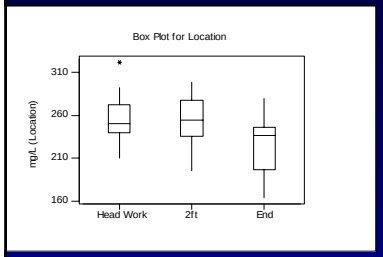
Box Plots - Filter Fabric Unit

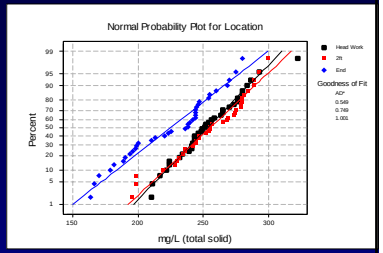


Box Plots - Catchbasin with Sump

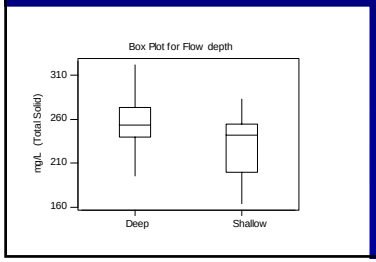


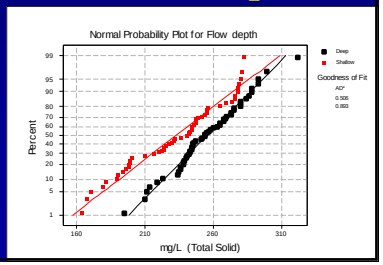
Solids Removal in Swales: Flow Length





Solids Removal in Swales: Flow Depth





Two independent groups of data

Parametric tests (data require normality and equal variance)
- Independent Student's *t*-test (more power than non-parametric tests)

Non-parametric tests
- Mann-Whitney rank sum test (probability distributions of the two data sets must be the same and have the same variances, but do not have to be symmetrical; a moderate number of "non-detectable" values can be accommodated)

Many groups (use multiple comparison tests, such as the Bonferroni *t*-test, to identify which groups are different from the others if the group test results are significant).

Parametric tests (data require normality and equal variance)
- One-way ANOVA for single factor, but for >2 "locations" (if 2 "locations, use Student's *t*-test)
- Two-way ANOVA for two factors simultaneously at multiple "locations"
- Three-way ANOVA for three factors simultaneously at multiple "locations"
- One factor repeated measures ANOVA (same as paired *t* test, except that there can be multiple treatments on the same group)
- Two factor repeated measures ANOVA (can be multiple treatments on two groups)

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).

Many Groups (cont.)

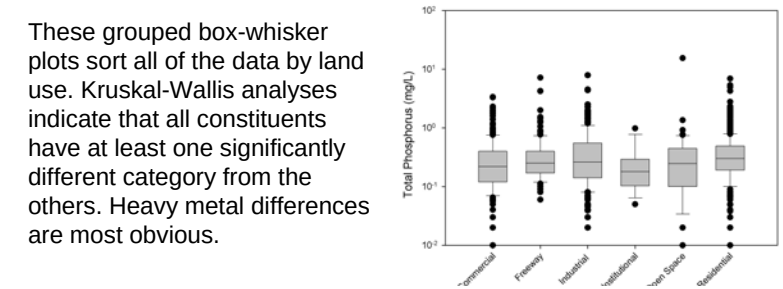
Nominal observations of frequencies (used when counts are recorded in contingency tables)

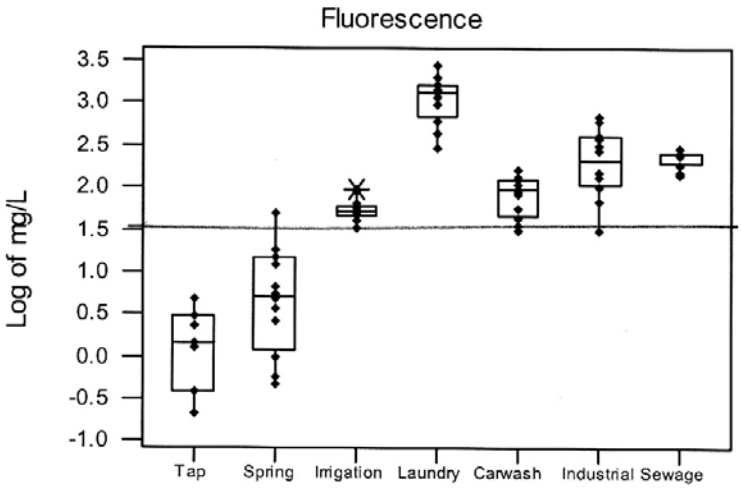
- Chi-square (X^2) test (use if more than two groups or categories, or if the number of observations per cell in a 2X2 table are > 5).

- 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)

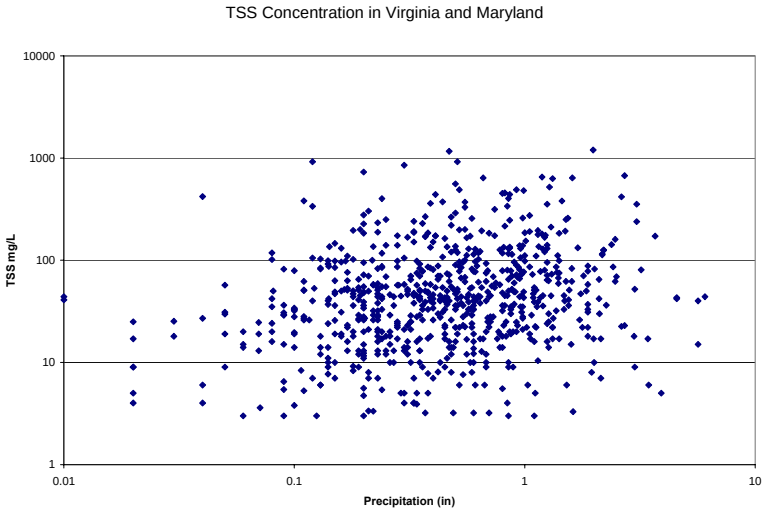
These grouped box-whisker plots sort all of the data by land use. Kruskal-Wallis analyses indicate that all constituents have at least one significantly different category from the others. Heavy metal differences are most obvious.

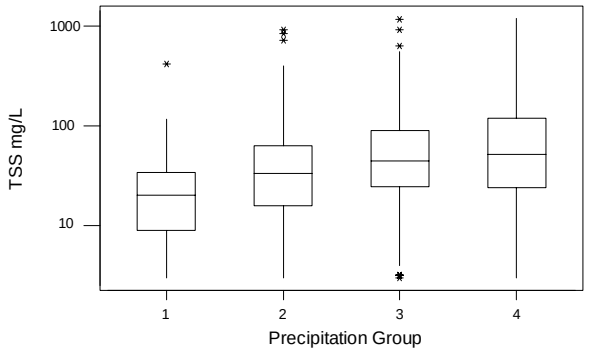




Example 2-way ANOVA

- Want to investigate the differences between different strata.
- Are the variations between groups more important than the variations within the groups?
- What about interactions between different variables?
- ANOVA requires normally distributed data. In most stormwater cases, log-transformed values need to be used.





Group 1: <0.1 inch of rain
Group 2: between 0.1 and 0.35 inches of rain
Group 3: between 0.35 and 1 inch of rain
Group 4: greater than 1 inch of rain

262 total cases

ANOVA

Analysis of Variance For **LTSS**
No Selector

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	631.103	631.103	3993.4	≤ 0.0001
Pgp	3	6.29608	2.09869	13.28	≤ 0.0001
Ssn	3	4.63302	1.54434	9.772	≤ 0.0001
Error	255	40.2994	0.158037		
Total	261	50.5397			

The rain group factor and the season factor are both highly significant. The prior 2-way ANOVA found that the interaction term was not significant; the ANOVA was therefore re-run without that term.

Coefficients

Coefficients of: LTSS on Pgp

Level of Pgp	Coefficient	std. err.	t Ratio	prob
1	-0.3302	0.06357	-5.195	≤ 0.0001
2	-0.003782	0.04182	-0.09044	0.9280
3	0.1966	0.04005	4.91	≤ 0.0001
4	0.1374	0.05064	2.712	0.0071

Expected Cell Means

Expected Cell Means of: LTSS on Pgp

Level of Pgp	Expected Cell Mean	Cell Count
1	1.165	25
2	1.491	86
3	1.692	104
4	1.632	47

Scheffe Post Hoc Tests

The first and third rain categories are significant.

Coefficients

Coefficients of: LTSS on Ssn

Level of Ssn	Coefficient	std. err.	t Ratio	prob
FA	-0.1701	0.04282	-3.974	< 0.0001
SP	0.07531	0.04581	1.644	0.1015
SU	0.178	0.04421	4.026	< 0.0001
WI	-0.08311	0.0405	-2.052	0.0412

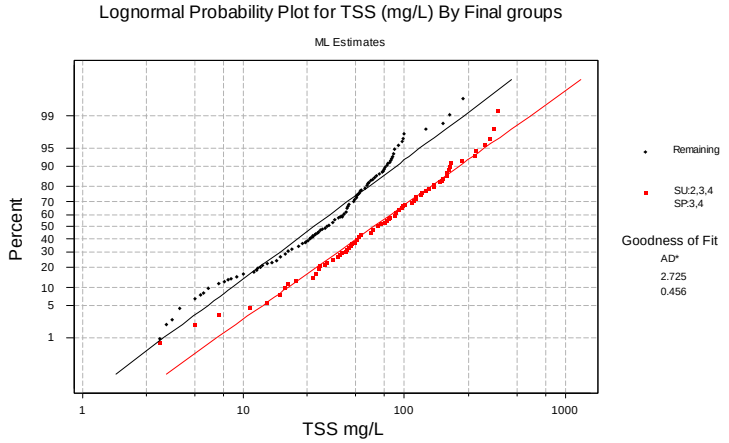
Expected Cell Means

Expected Cell Means of: LTSS on Ssn

Level of Ssn	Expected Cell Mean	Cell Count
FA	1.325	67
SP	1.57	54
SU	1.673	60
WI	1.412	81

Scheffe Post Hoc Tests

Only Fall and Summer are significant.



Further analyses resulted in two main groups of data.

Example 1-way ANOVA

- Is at least one member of a group significantly different from the other members?
- Complement analysis with group box-whisker plot
- This doesn't identify which one(s) is(are) different.
- If a significant member, should be able to recognize from box-whisker plot and with Bonferroni T-test (multiple pair-wise comparisons).

1-way ANOVA

Site A	Site B	Site C	Site D	Site E
78	43	153	14	12
45	79	87	53	9
63	54	245	42	34
54		432	64	14
24		43	23	
		164		

Are any of these sites different from the others?

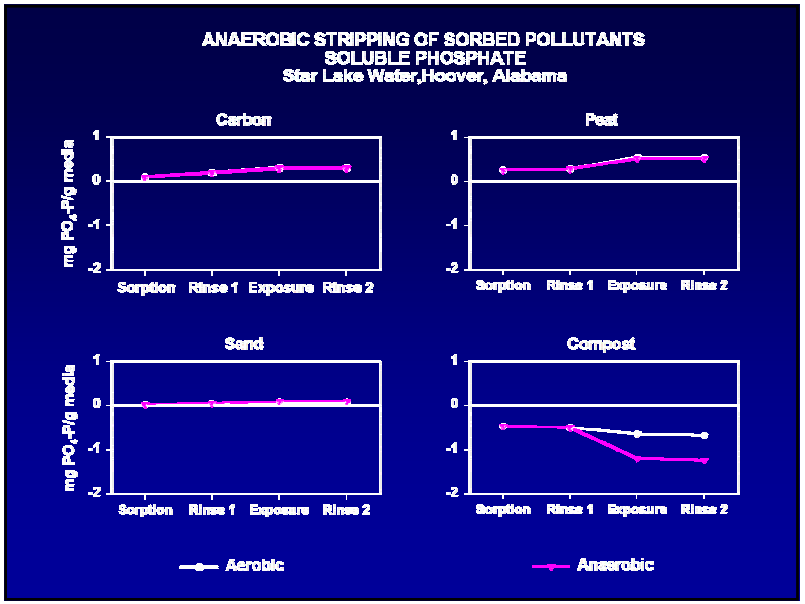
ANOVA Single Factor (using Excel)

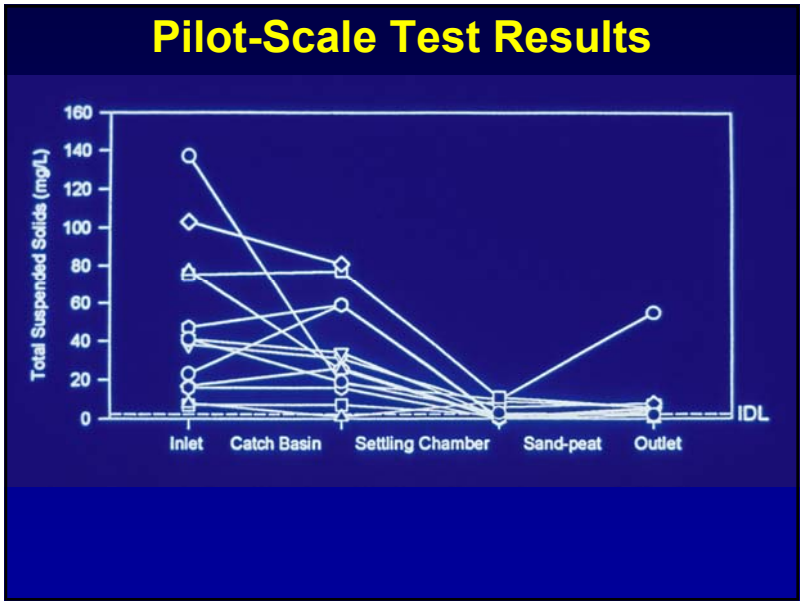
SUMMARY

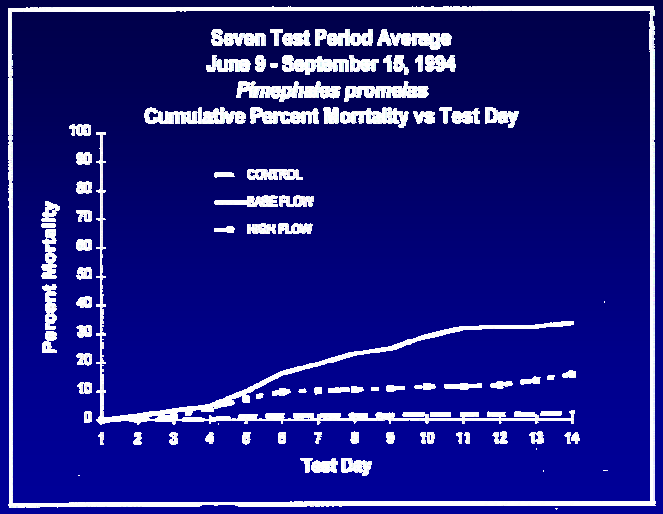
Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	5	264	52.8	407.7
Column 2	3	176	58.66667	340.3333
Column 3	6	1124	187.3333	19161.87
Column 4	5	196	39.2	427.7
Column 5	4	69	17.25	128.9167

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	98255	4	24564	4.41	0.0116	2.9277
Within Groups	100218	18	5567			
Total	198473	22				

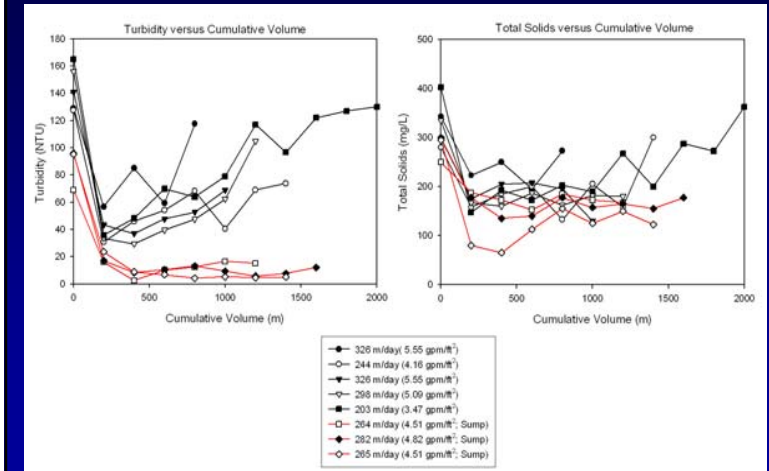






Crunkilton, *et al.* (1996)

Design Configuration Optimization using Pool Sand Filter Media



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

Linear equation fitting (statistically-based models)

- Simple linear regression (y=b0+b1x, with a single independent variable, the slope term, and an intercept. It is possible to simplify even further if the intercept term is not significant).

- Multiple linear regression (y=b0+b1x1+b2x2+b3x3+…+bkxk, having k independent variables. The equation is a multi-dimensional plane describing the data).

- Stepwise regression (a method generally used with multiple linear regression to assist in identifying the significant terms to use in the model.)

- Polynomial regression (y=b0+b1x1+b2x2+b3x3+…+bkxk, having one independent variable describing a curve through the data).

Non-linear equation fitting (generally developed from theoretical considerations)

- Nonlinear regression (a nonlinear equation in the form: y=bx, where x is the independent variable. Solved by iteration to minimize the residual sum of squares).

Model Building Steps

1) Re-examine the hypothesis of cause and effect (an original component of the experimental design previously conducted and was the basis for the selected sampling activities).

2) Prepare preliminary examinations of the data, as described previously (most significantly, prepare scatter plots and grouped box/whisker plots).

3) Conduct comparison tests to identify significant groupings of data. As an example, if seasonal factors are significant, then cause and effect may vary for different times of the year.

4) Conduct correlation matrix analyses to identify simple relationships between parameters. Again, if significant groupings were identified, the data should be separated into these groupings for separate analyses, in addition to an overall analysis.

Modeling Building (cont.)

5) Further examine complex inter-relationships between parameters by possibly using combinations of hierarchical cluster analyses, principal component analyses (PCA), and factor analyses.

6) Compare the apparent relationships observed with the hypothesized relationships and with information from the literature. Potential theoretical relationships should be emphasized.

7) Develop initial models containing the significant factors affecting the parameter outcomes. Simple apparent relationships between dependent and independent parameters should lead to reasonably simple models, while complex relationships will likely require further work and more complex models.

Plots to Assist in Model Building

- Simple Correlation Matrices
- Hierarchical Cluster Analyses
- Principal Component Analyses (PCA) and Factor Analyses

Simple Data Associations

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

- Spearman Rank Order Correlation (a non-parametric equivalent to the Pearson test).

Emery (Industrial)

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

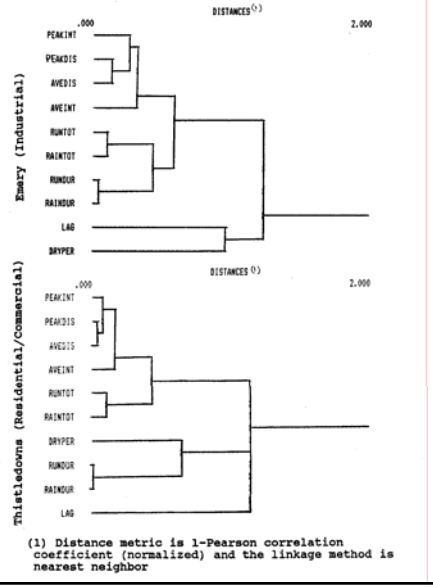
Thistledowns (Residential/Commercial)

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

Complex Data Associations (typically only available in advanced software packages)

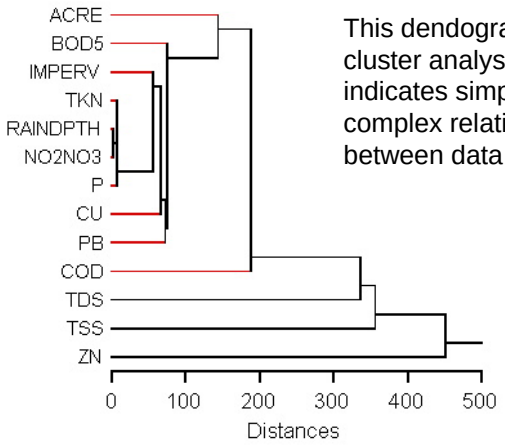
- Hierarchical Cluster Analyses (graphical presentation of simple and complex inter-relationships. Data should be standardized to reduce scaling influence. Supplements simple correlation analyses).

- Principal Component Analyses (identifies groupings of parameters by factors so that variables within each factor are more highly correlated with variables in that factor than with variables in other factors. Useful to identify similar sites or parameters).



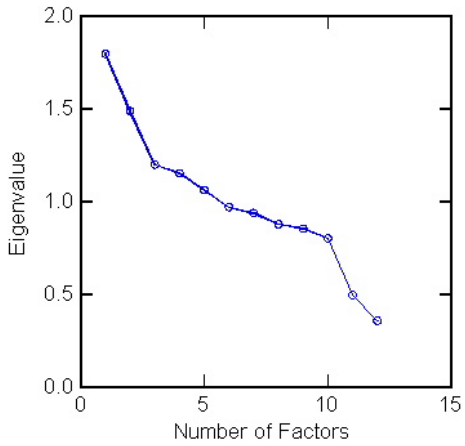
(1) Distance metric is 1-Pearson correlation coefficient (normalized) and the linkage method is nearest neighbor

Cluster Tree



This dendogram from a cluster analysis indicates simple and complex relationships between data subsets.

Scree Plot



Principal Component Analysis (PCA) is a sophisticated procedure where information is sorted to determine the constituents needed to explain the variance of the data.

Regression Analyses

1) Formulate the objectives of the curve-fitting exercise (a subset of the experimental design previously conducted).

2) Prepare preliminary examinations of the data, as described previously (most significantly, prepare scatter plots and probability plots of the data, plus correlation evaluations to examine independence between multiple parameters that may be included in the models)

3) Identify alternative models from the literature that have been successfully applied for similar problems (part of the previously conducted experimental design activities in order to identify which parameters to measure, or to modify or control).

4) Evaluate the data to ensure that regression is applicable and make suitable data transformations.

Regression (cont.)

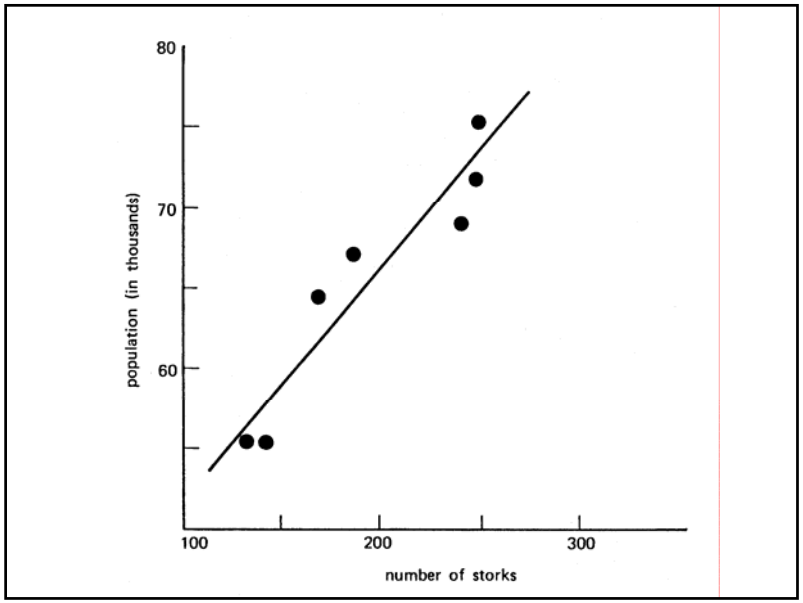
5) Apply regression procedures to the selected alternative models.

6) Evaluate the regression results by examining the coefficient of determination (R2) and the results of the analysis of variance of the model (standard error analyses and p values for individual equation parameters and overall model).

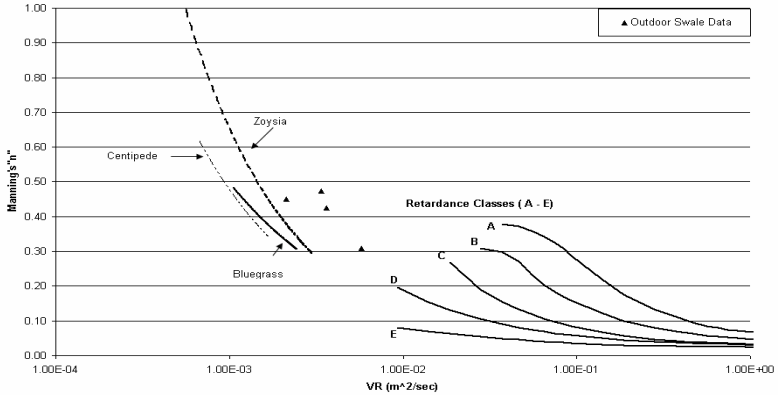
7) Conduct an analysis of the residuals (as described below).

8) Evaluate the results and select the most appropriate model(s).

9) If not satisfied, it may be necessary to examine alternative models, especially based on data patterns (through cluster analyses and principal component analyses) and re-examinations and modification of the theoretical basis of existing models. Statistical based models can be developed using step-wise regression routines.

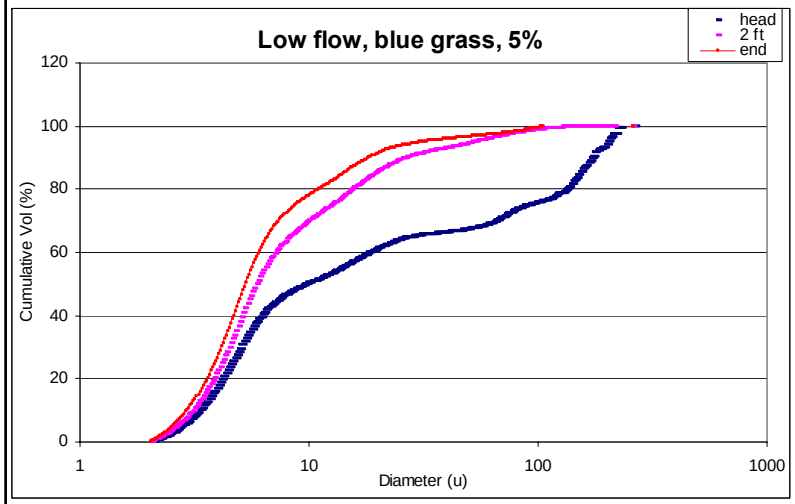


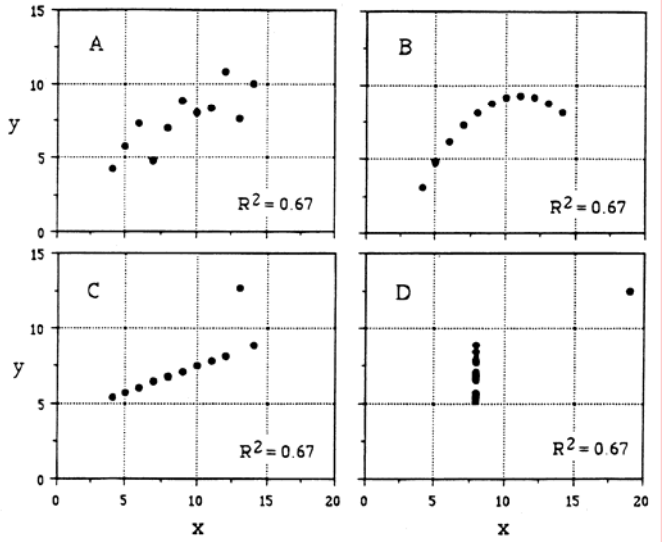
Indoor vs. Historical Stillwater, Oklahoma, Retardance Curves



From such graphs swale hydraulic characteristics can be predicted on the basis of flow rate, cross sectional geometry, slope, and vegetation type.

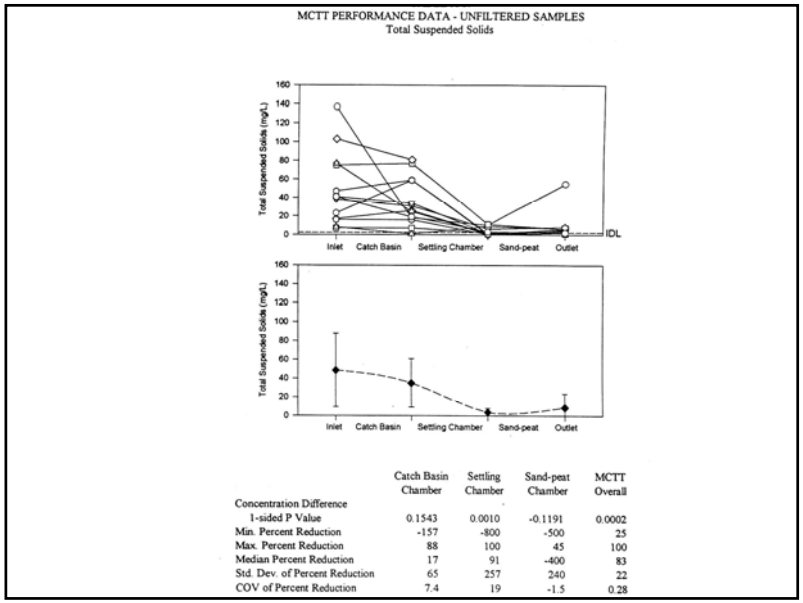
Preferential Capture of Large Particles in Grass Swales





Regression Example with ANOVA

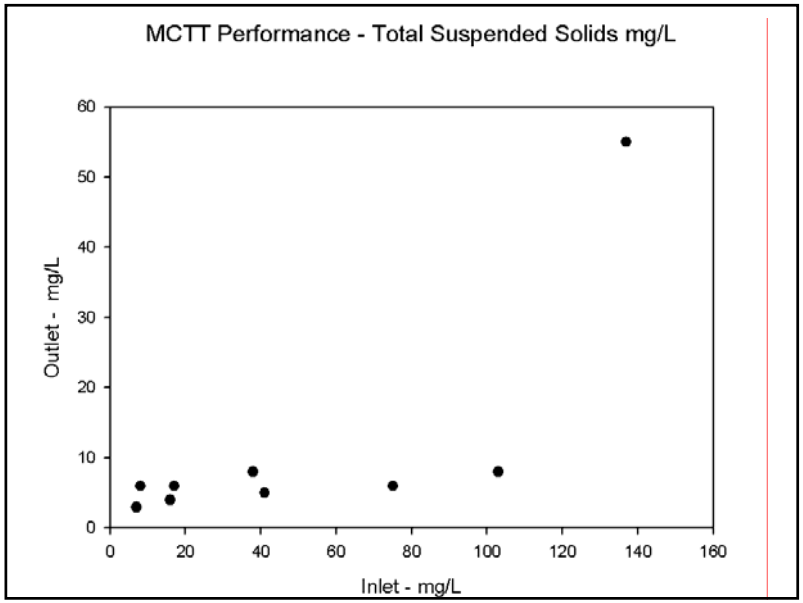
- Examining treatment data with regression and associated plots and ANOVA

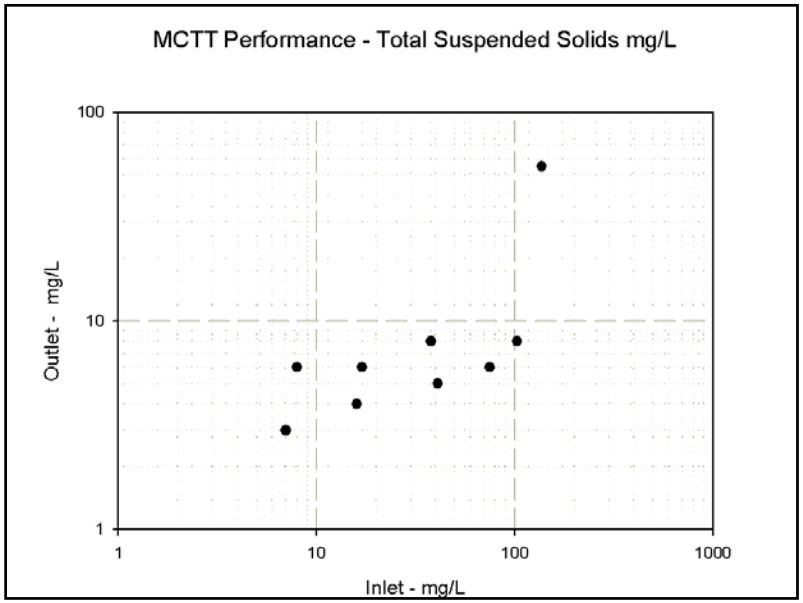


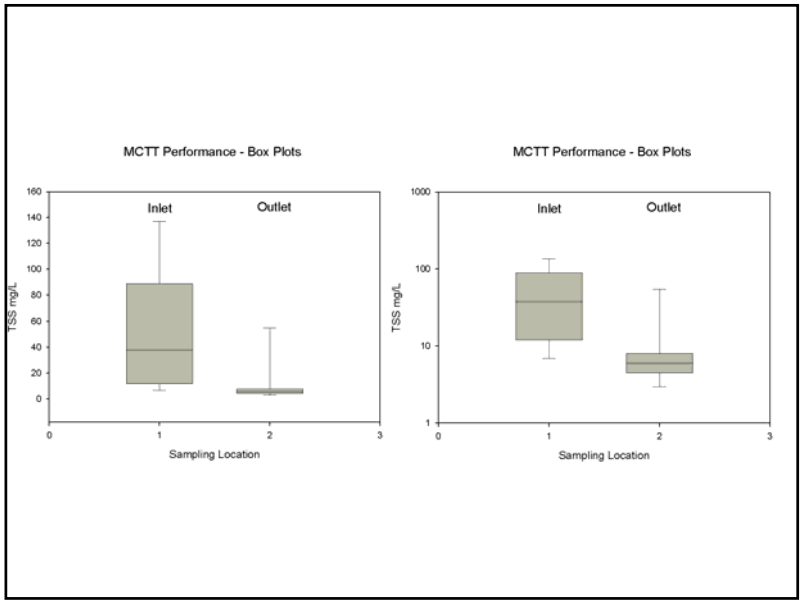
	Catch Basin Chamber	Settling Chamber	Sand-peat Chamber	MCTT Overall
Concentration Difference				
1-sided P Value	0.1543	0.0010	-0.1191	0.0002
Min. Percent Reduction	-157	-800	-500	25
Max. Percent Reduction	88	100	45	100
Median Percent Reduction	17	91	-400	83
Std. Dev. of Percent Reduction	65	257	240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

Total Suspended Solids mg/L

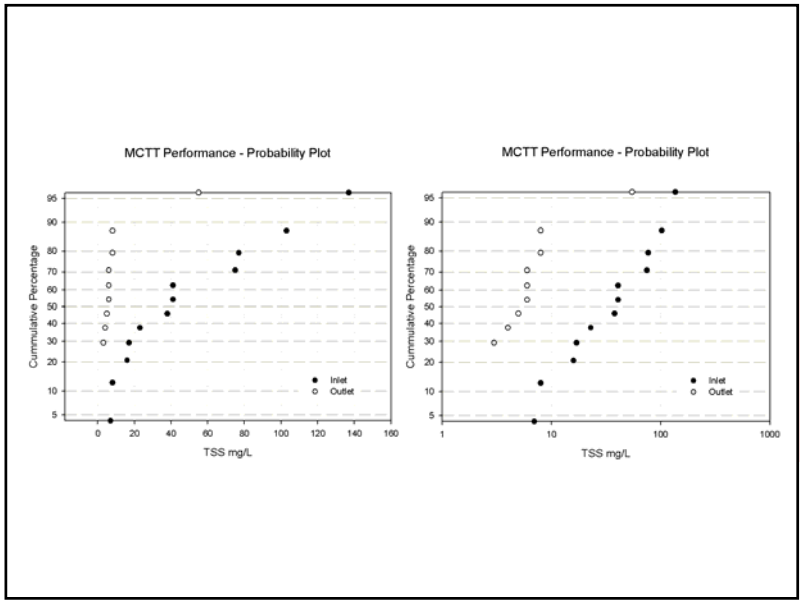
STORM	INLET	OUTLET
1	137	55
2	7	3
3	8	6
4	38	8
5	17	6
6	16	4
7	23	<2.5
8	75	6
9	77	<2.5
10	41	5
11	103	8
12	41	<2.5

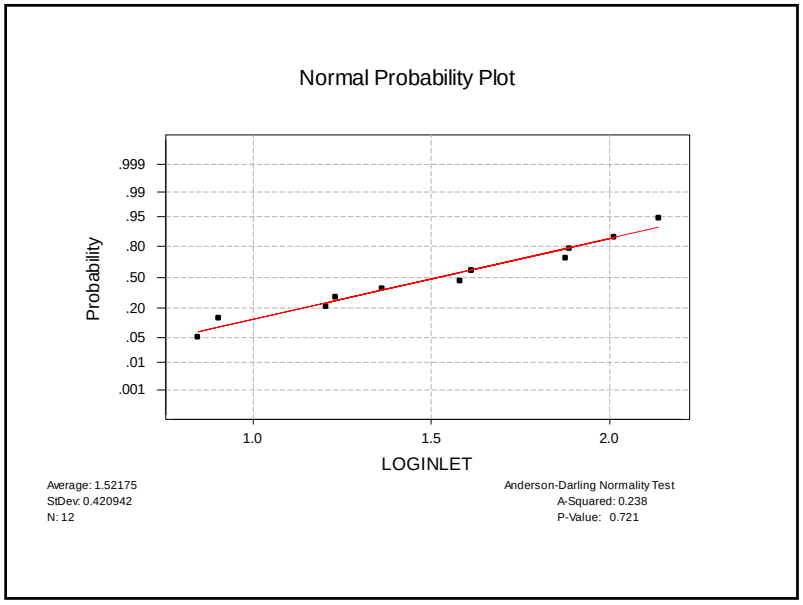


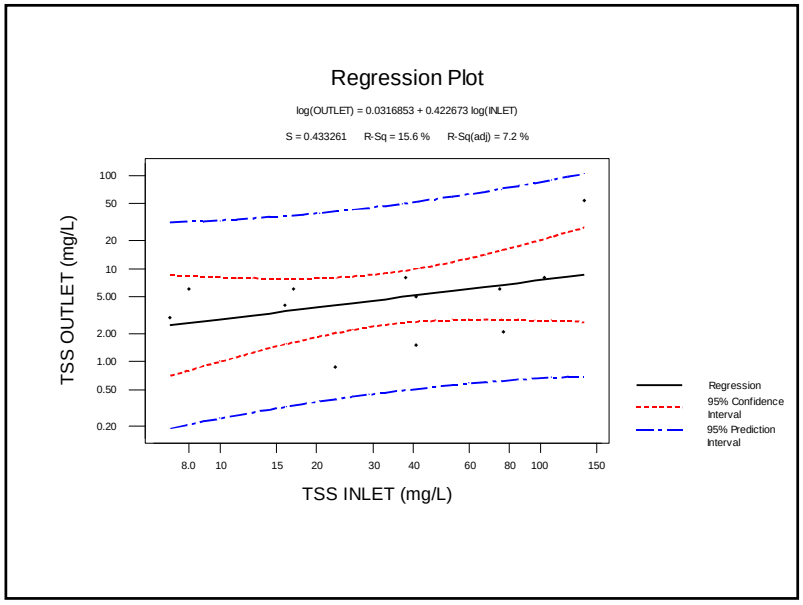




	Influent	Effluent
N	12	12
Detected Observations	12	9
Mean	48.6	11.22
Median	39.5	5.5
StDev	41.1	16.5
SE Mean	11.9	5.5
Minimum	7	3
Maximum	137	55
Q1	16.3	2.7
Q3	76.5	7



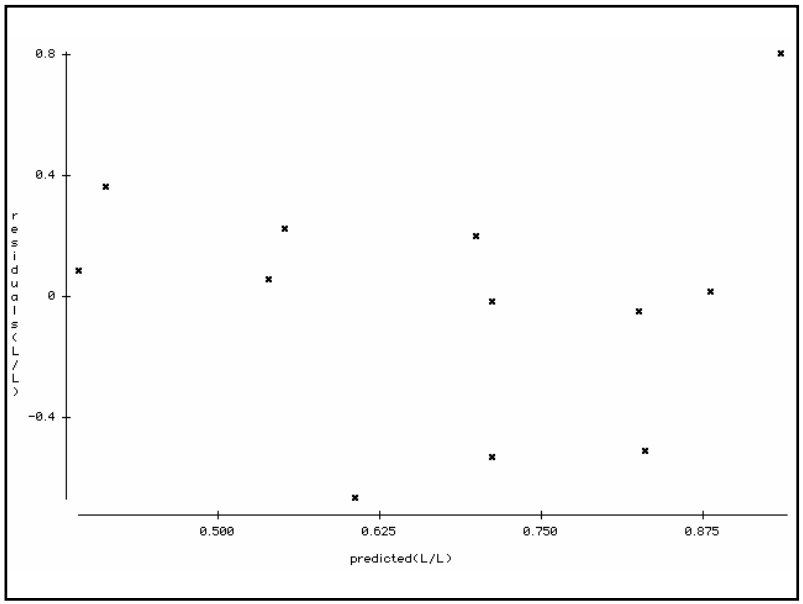




Dependent variable is: **LOGOUTLET**
No Selector
R squared = 15.6% R squared (adjusted) = 7.2%
s = 0.4332 with 12 - 2 = 10 degrees of freedom

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

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

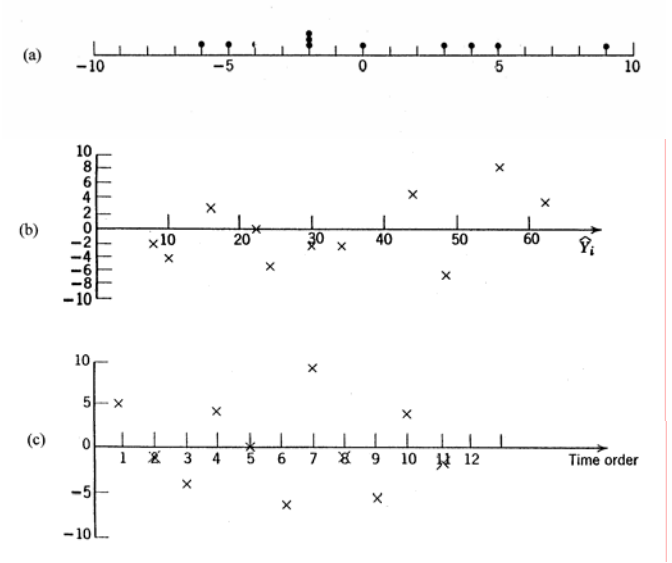


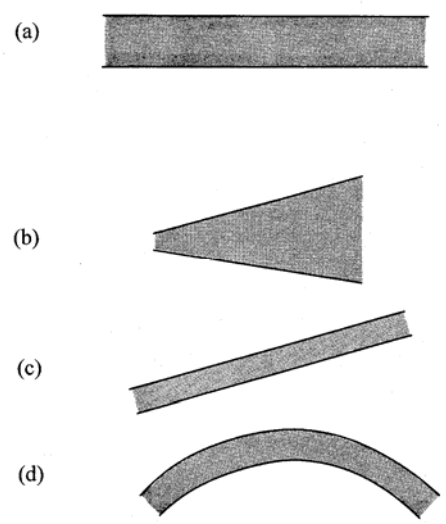
Residual Analyses of Regression Model

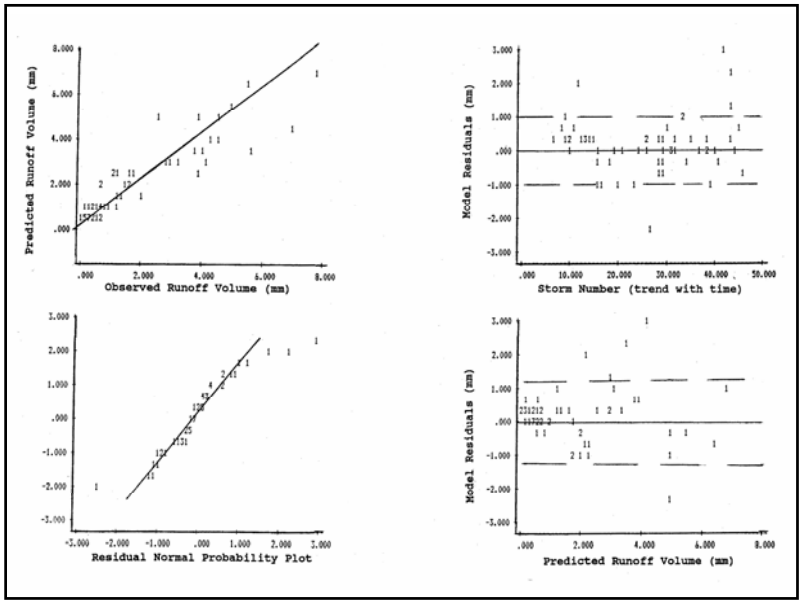
- the residuals are independent
- the residuals have zero mean
- the residuals have a constant variance (S^2)
- the residuals have a normal distribution (required for making F-tests)

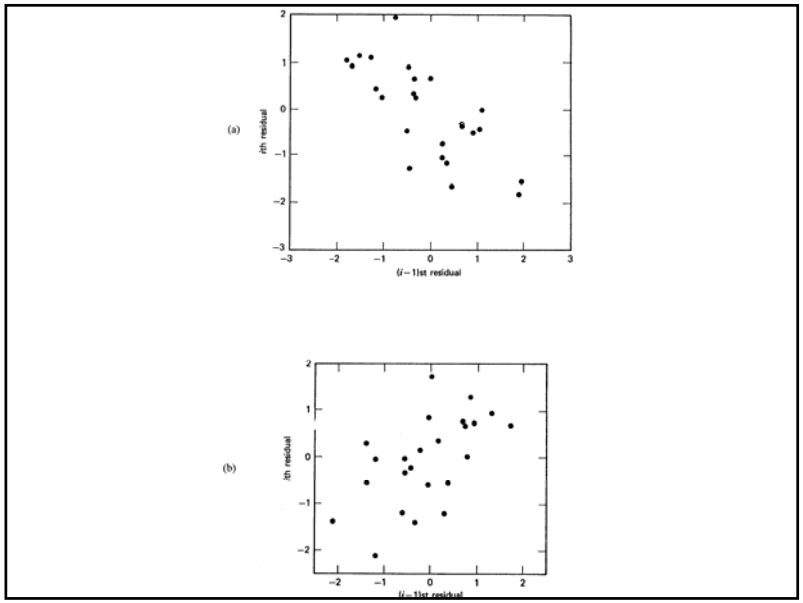
Plots to Check Residuals

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









Data Trends

- Graphical methods (simple plots of concentrations versus time of data collection).

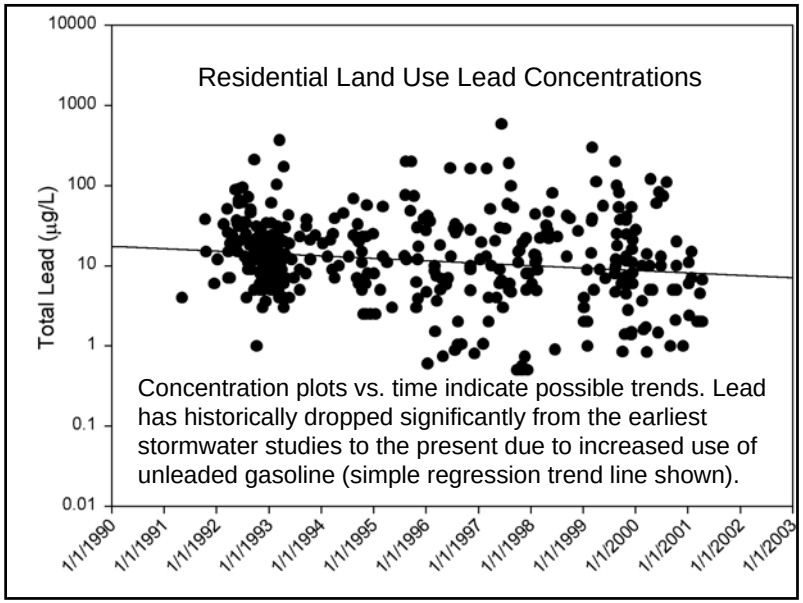
- Regression methods (perform a least-squares linear regression on the above data plot and examine ANOVA for the regression to determine if the slope term is significant. Can be misleading due to cyclic data, correlated data, and data that are not normally distributed).

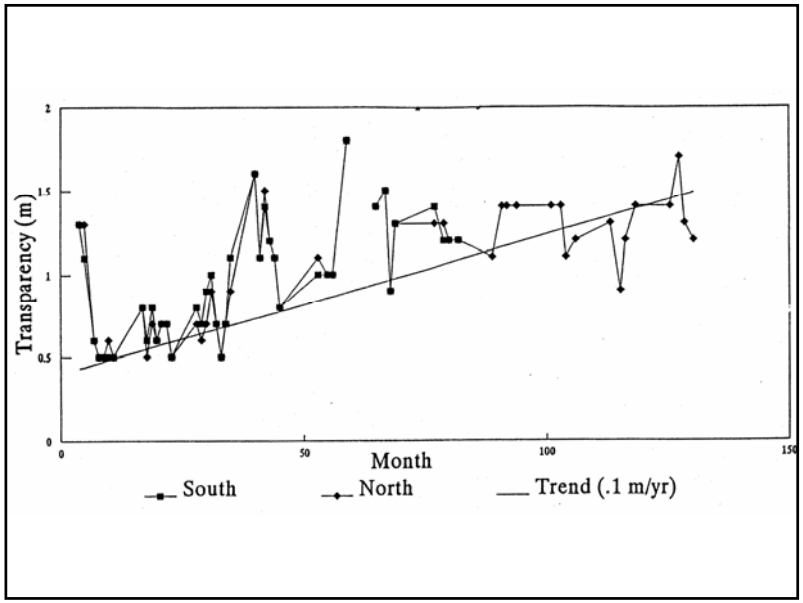
- Mann-Kendall test (a nonparametric test that can handle missing data and trends at multiple stations. Short-term cycles and other data relationships affect this test and must be corrected).

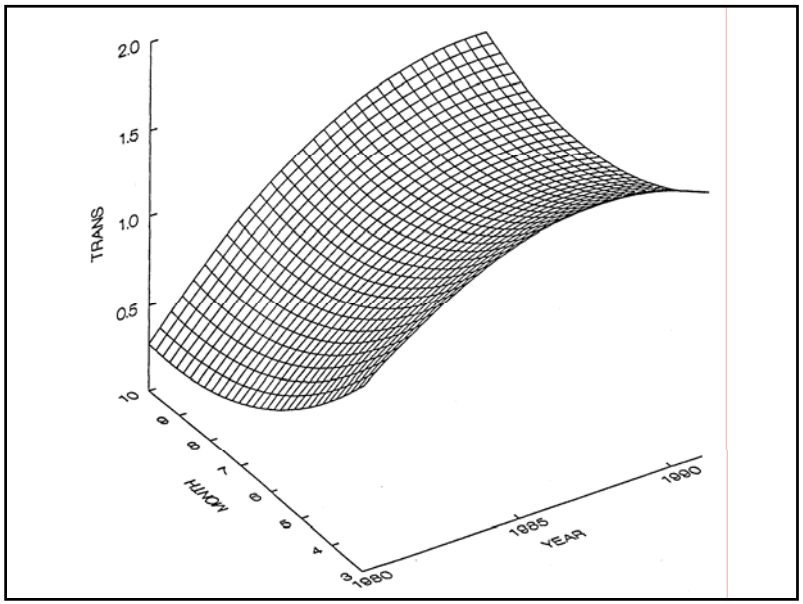
Data Trends (cont.)

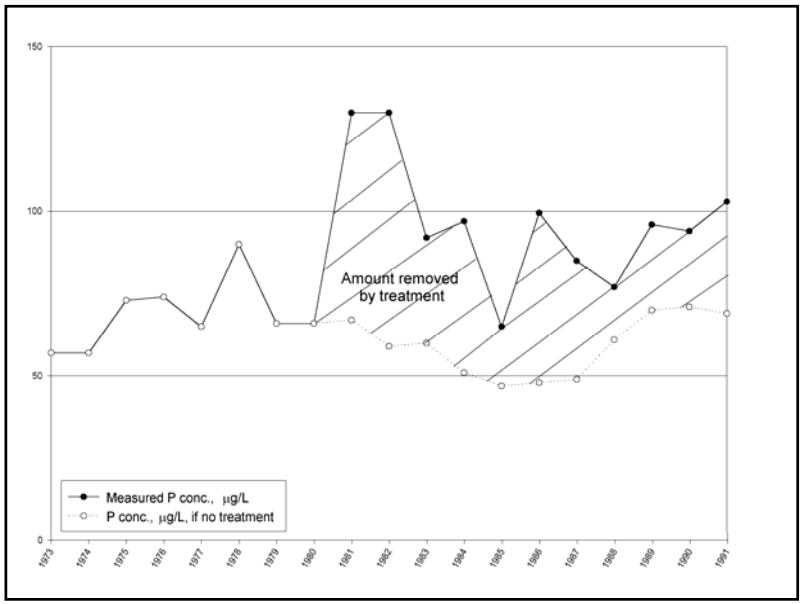
- Sen's estimator of slope (a nonparametric test based on ranks closely related to the Mann- Kendall test. It is not sensitive to extreme values and can tolerate missing data).

- Seasonal Kendall test (preferred over regression methods if the data are skewed, serially correlated, or cyclic. Can be used for data sets having missing values, tied values, censored values, or single or multiple data observations in each time period. Data correlations and dependence also affect this test and must be considered in the analysis).


Residential Land Use Lead Concentrations
Total Lead (µg/L)
10000
1000
100
10
1
0.1
0.01
1/1/1990
1/1/1991
1/1/1992
1/1/1993
1/1/1994
1/1/1995
1/1/1996
1/1/1997
1/1/1998
1/1/1999
1/1/2000
1/1/2001
1/1/2002
1/1/2003
Concentration plots vs. time indicate possible trends. Lead has historically dropped significantly from the earliest stormwater studies to the present due to increased use of unleaded gasoline (simple regression trend line shown).


Transparency (m)
Month
South
North
Trend (.1 m/yr)


TRANS
HUNOW
YEAR
1980
1985
1990


Amount removed by treatment
Measured P conc., µg/L
P conc., µg/L, if no treatment