Continuous Monitoring to Approximate Total Phosphorus Concentrations Through Implementation of Regression Equations

Ben Hammond
Woolpert Inc. Columbia, SC

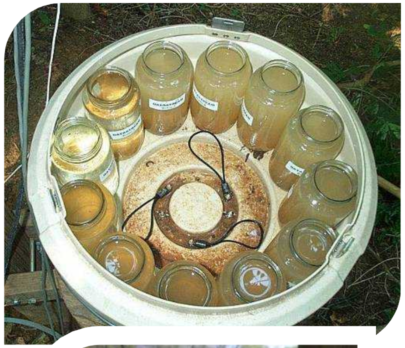
May 2, 2012

WOOLPERT
DESIGN | GEOSPATIAL | INFRASTRUCTURE

What is Our Goal?

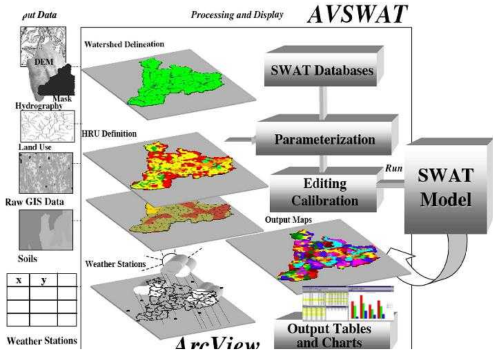
- Comprehensive Water Quality Data
  - Improved / more technology
  - More Data
  - Higher frequency data collection

- Water Quality Awareness
  - Are conditions improving or degrading?

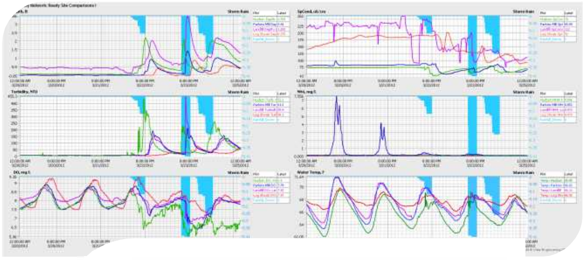


Options for Better Awareness

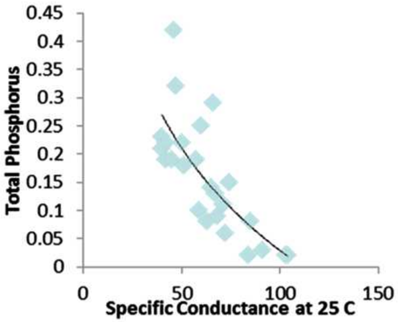
- Watershed Models
  - Scale and detail
  - Minimal disaggregation
  - Assumptions and lumped parameters



- Continuous Monitoring
  - Potential real-time parameter fluctuations
  - Sensors for priority pollutants



- Middle Ground- Regression
  - What we know correlated with what we don't know enough about
  - Fewer assumptions
  - Can guide sample collection in addition to estimating pollutant loads



Don't Be Fooled

- Regression does NOT replace traditional sampling methods ENTIRELY
- Regression Accuracy
  - Is not equal to laboratory accuracy, but…
  - Projected loads estimates are likely better, given the frequency of data collection

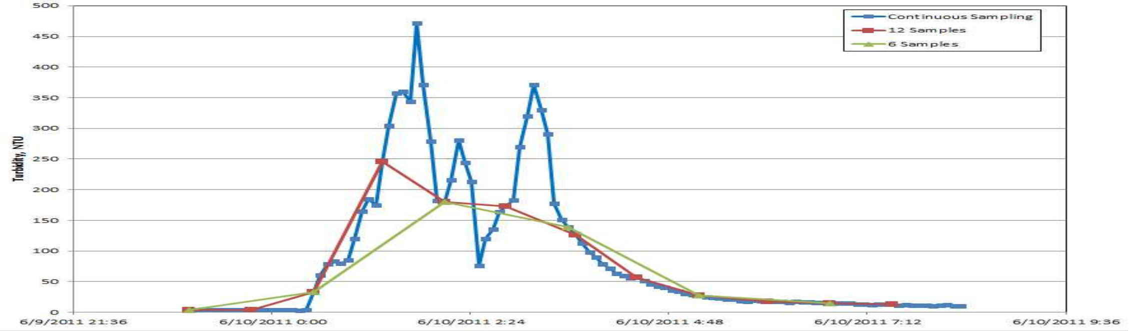
- Regression is not inexpensive
  - Laboratory costs can be cost prohibitive
- You need
  - Time / patience
  - Necessity
  - $$

Then Why Regression

- Higher Frequency = More Detail

- More Detail = More comprehensive analysis
  - Spatial: Where is the problem?
  - Temporal: When does the problem occur?
  - Hydrologic: Is stormwater runoff a problem?

- Traditional Methods ALONE can generate more questions than answers
  - What happened when you were not collecting samples?



Regression Analysis

- Six continuous monitoring stations
  - In series along the same river
  - YSI Multi-parameter sondes
  - Trend and hot spot analysis
  - Total phosphorus regression
- Two years of data collection
  - Sample n varied at all 6 sites
- 1 year of validation…so far
- Work in progress

Regression Analysis

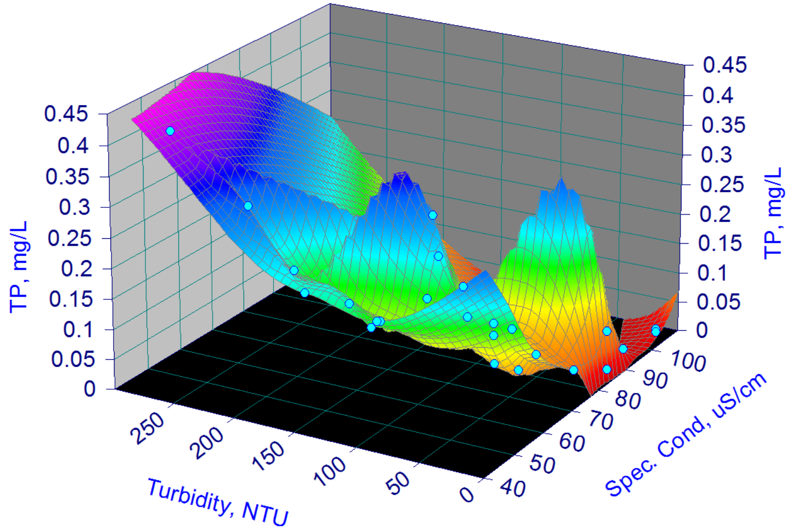
- Sonde data as surrogates
  - Turbidity
  - Specific conductivity
  - Temperature
  - pH
  - Dissolved oxygen

Collection Time	Sonde Dissolved Oxygen (field)	Sonde pH	Sonde Specific Conductance at 25 C	Sonde Temperature	Sonde Turbidity	Total Phosphorus
12/8/2008 10:30	11.09	7.13	264	48	11.09	0.07
4/23/2009 10:45	9.76	6.62	193	58.88	6.1	0.05
5/21/2009 10:30	8.49	6.92	240	65.7	4.4	0.1
9/18/2009 8:30	7.12	6.88	63	70.87	198.7	0.24
9/18/2009 11:15	7.28	6.98	92	71.33	184.3	0.33
10/12/2009 9:15	8.34	7.37	195	67.35	4.4	0.09
10/12/2009 10:15	8.43	7.34	183	66.7	15	0.14
10/12/2009 11:15	8.64	7.26	126	65.14	71.4	0.27
10/12/2009 13:00	8.74	7.1	79	64.01	117.7	0.22
10/12/2009 14:15	8.84	7.13	73	63.91	92	0.27
12/2/2009 9:45	10.53	7.06	124	49.18	134.4	0.14
12/2/2009 11:15	10.69	7.02	105	48.88	109.9	0.24
12/2/2009 13:30	10.99	7.01	78	48.17	168.5	0.3
12/3/2009 8:15	10.42	6.81	70	50.87	94.3	0.18
12/3/2009 13:45	10.07	6.79	76	52.23	69.4	0.22
5/3/2010 13:30	7.8	6.98	69	69.35	94	0.2
5/3/2010 15:45	7.88	7.01	69	69.02	81.7	0.23
5/4/2010 9:00	8.36	7.15	158	67.38	24.9	0.12
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

- Paired Samples used to develop regression equations

Methods

- Best Fit Plots, ANOVA Statistics, Residuals
  - Is there a linear relationship
- Simple and Multiple Linear Regression
  - Evaluated using Predicted Residual Sum of Squares (PRESS) statistic and RMSE.
- This is not data fitting



Methods

- Site Specific Equations
  - We considered watershed scale as well as upstream and downstream of a major continuous point source
  - Site specific statistically more significant as a result of many inputs between sites
- Storm vs. Baseflow Equations
  - Equations specific to the conditions that were monitored during the regression development process
  - Not necessarily intended to detect all sources, e.g. SSOs

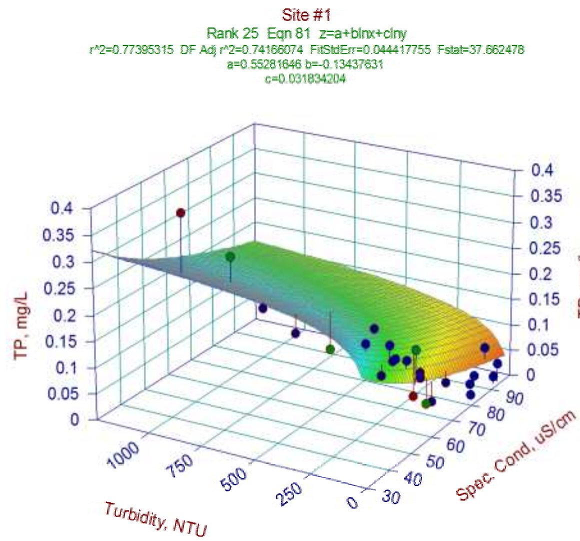
Methods

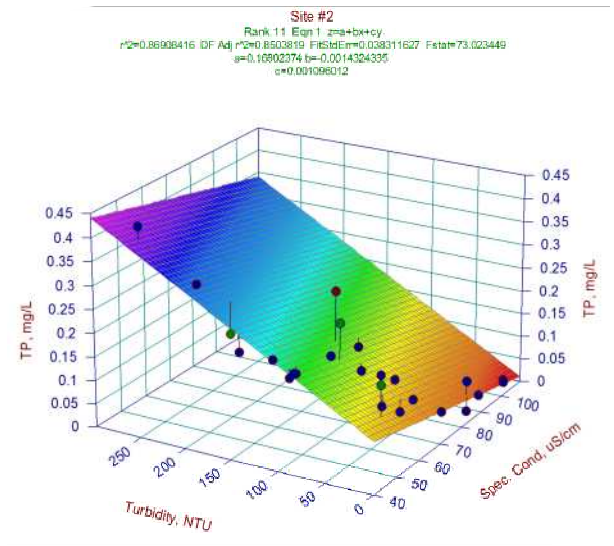
- Choosing surrogates
  - We let data decide
  - Turbidity and SC were best
  - Why not flowrate as a surrogate? Will include in updates
  - Continued work might include characteristic variables, essentially establishing two relationships using one equation

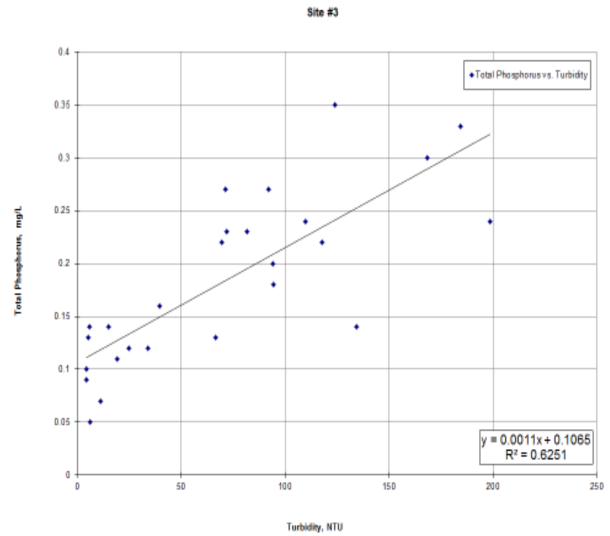
Results of Regression Analysis

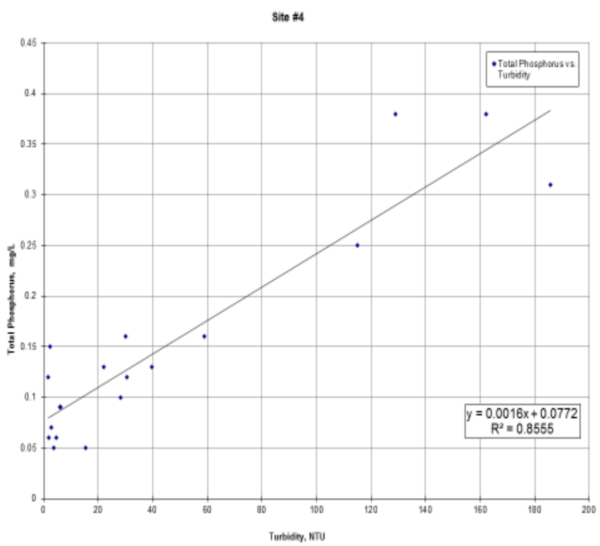
- 6 site-specific equations
- PRESS statistic used to select site-specific equation
- Single and Multiple Linear equations

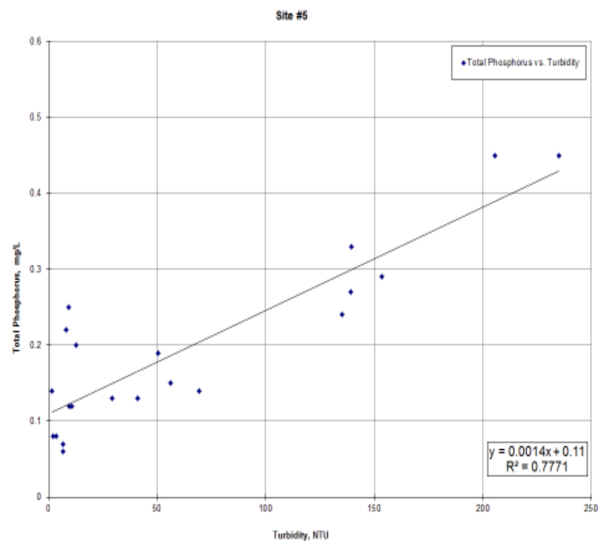
Site	Regression Equation	n	PRESS	RMSE	R2	Adjusted R2	F
Site #1	0.553 + 0.073*log(turb) - 0.309*log(SC)	25	0.060	0.043	0.77	0.75	38
Site #2	0.168 - 0.001*SC + 0.001*Turb	25	0.041	0.043	0.87	0.86	73
Site #3	0.107 + 0.001*turb	26	0.074	0.049	0.63	0.61	40.0
Site #4	0.077 + 0.002*turb	20	0.044	0.039	0.86	0.85	107
Site #5	0.110 + 0.001*turb	21	0.045	0.053	0.78	0.77	66
Site #6	-0.207 + 0.110*logturb + 0.095*logSC	19	0.040	0.041	0.39	0.31	5.1

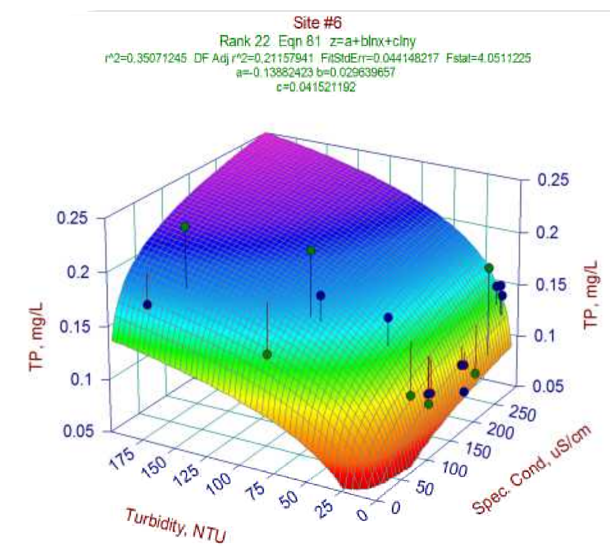

Site #1
Rank 25 Eqn 81 z=a+blnx+clny
r^2=0.77395315 DF Adj r^2=0.74166074 FitStdErr=0.044417755 Fstat=37.662478
a=0.55281646 b=-0.13437631
c=0.031834204
TP, mg/L
TP, mg/L
Turbidity, NTU
Spec. Cond, uS/cm


Site #2
Rank 11 Eqn 1 z=a+bx+cy
r^2=0.86906416 DF Adj r^2=0.8503819 FitStdErr=0.038311627 Fstat=73.023449
a=0.16802374 b=-0.0014324335
c=0.001096012
TP, mg/L
TP, mg/L
Turbidity, NTU
Spec. Cond, uS/cm

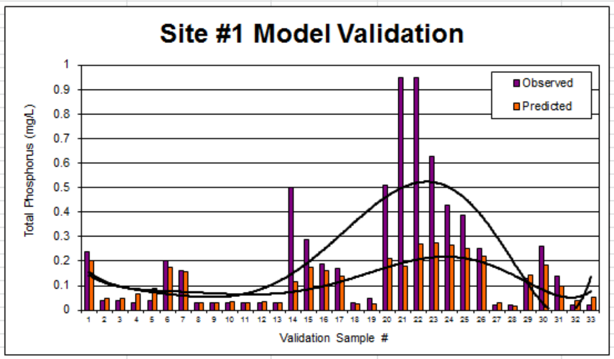

Site #3
Total Phosphorus vs. Turbidity
y = 0.0011x + 0.1065
R² = 0.6251
Total Phosphorus, mg/L
Turbidity, NTU

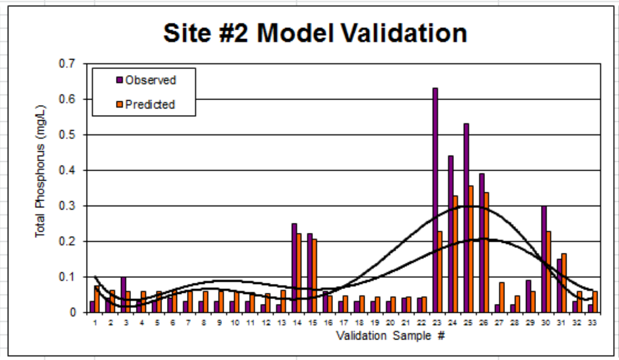

Site #4
Total Phosphorus vs. Turbidity
y = 0.0016x + 0.0772
R² = 0.8555
Total Phosphorus, mg/L
Turbidity, NTU

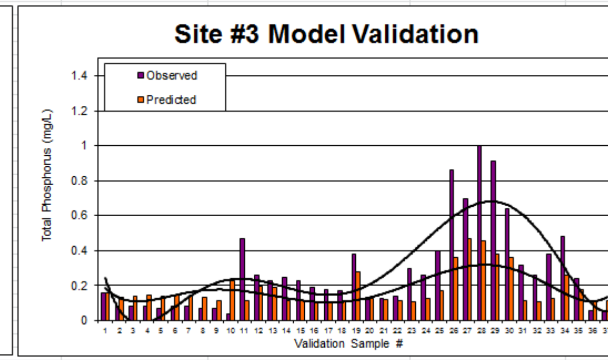

Site #5
Total Phosphorus vs. Turbidity
y = 0.0014x + 0.11
R² = 0.7771
Total Phosphorus, mg/L
Turbidity, NTU

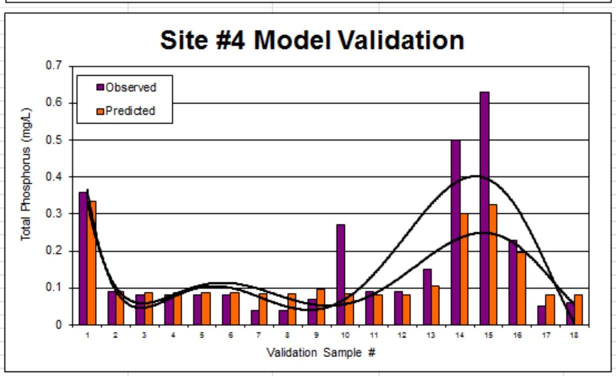

Site #6
Rank 22 Eqn 81 z=a+blnx+clny
r^2=0.35071245 DF Adj r^2=0.21157941 FitStdErr=0.044148217 Fstat=4.0511225
a=-0.13882423 b=0.029639657
c=0.041521192
TP, mg/L
TP, mg/L
Turbidity, NTU
Spec. Cond, uS/cm

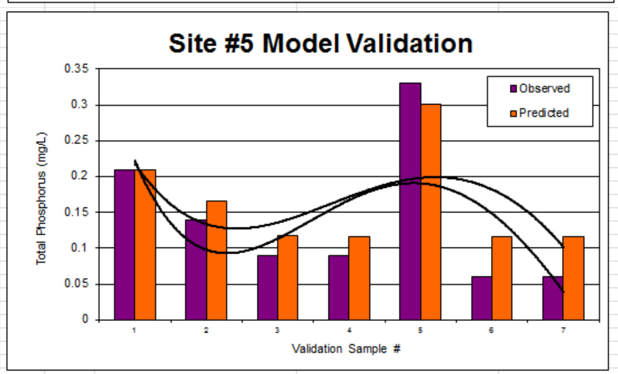
Regression Validation Results

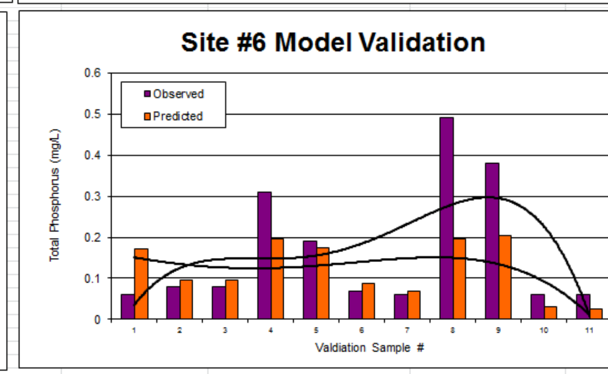






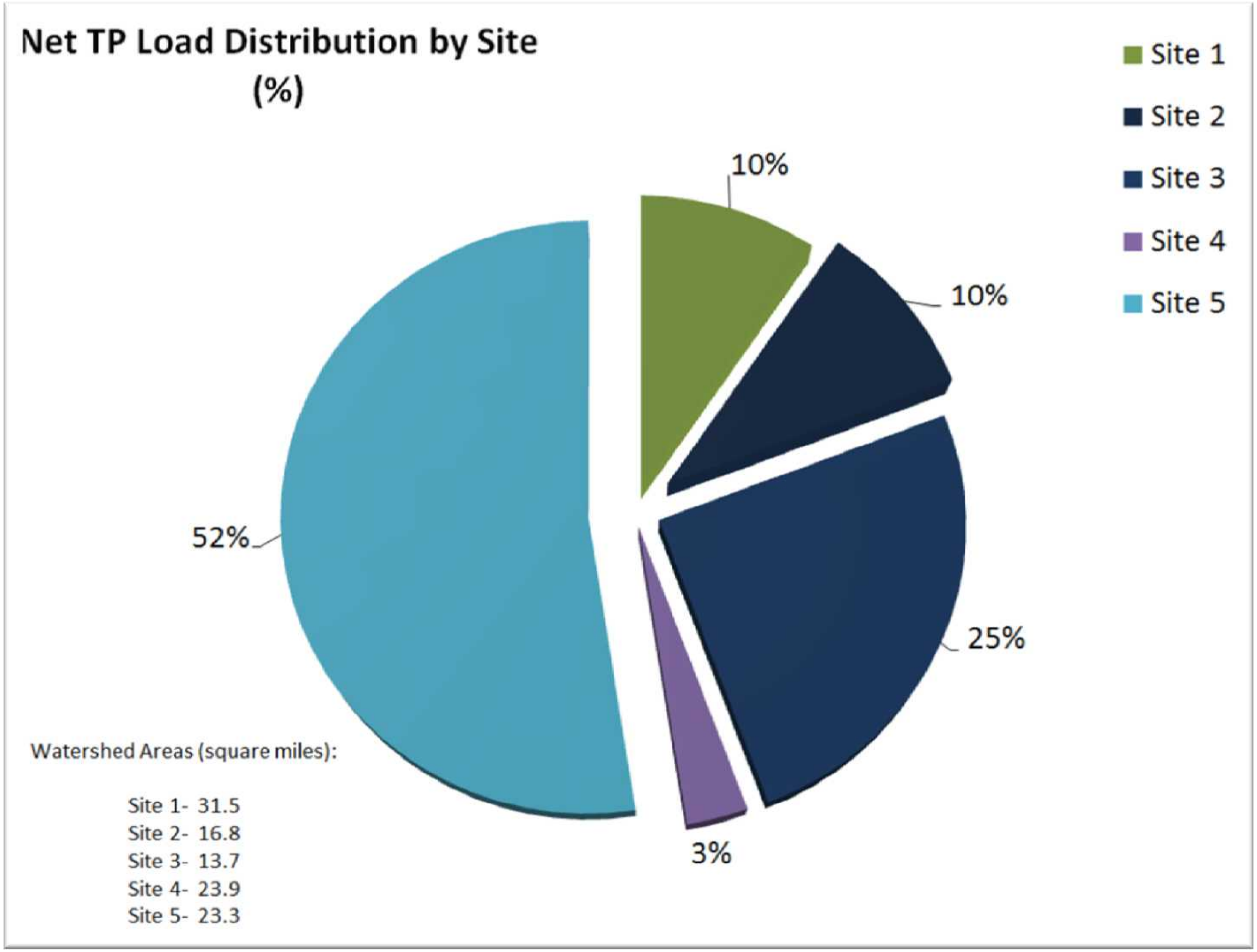






Note: Validation Samples are not serially correlated

Application of Regression Equations



Conclusions

- Identify the problem…then address the problem
- Study results are encouraging
- Enhancements to existing regression equations
  - Flow as surrogate
  - Characteristic variable
  - Antecedent moisture condition
- The goal is to get “it” right
  - Appropriate management of our financial resources can lead to saving or improving our natural resources.

References

+ Rasmussen, T.J., Ziegler, A.C., and Rasmussen, P.P., 2005, Estimation of constituent concentrations, densities, loads, and yields in lower Kansas River, northeast Kansas, using regression models and continuous water-quality monitoring, January 2000 through December 2003: US Geological Survey Scientific Investigations Report 2005-5165, 117p.

+ Rasmussen, T.J., Lee, C.J., and Ziegler, A.C., 2008, Estimation of constituent concentrations, loads, and yields in streams of Johnson County, northeast Kansas, using continuous water-quality monitoring and regression models, October 2002 through December 2006: US Geological Survey Scientific Investigations Report 2008-5014, 103 p.

+ Jones, Amber Spackman, Stevens, David K., Horsburgh, Jeffery S., and Mesner, Nancy, 2011, Surrogate Measures for Providing High Frequency Estimates of Total Suspended Solids and Total Phosphorus Concentrations: Journal of the American Water Resources Association, April 2011.

+ Helsel, Dennis R., and Hirsch, Robert M., USGS TWRI Book 4 Chapter A3, Statistical Methods in Water Resources