Pathogen Source Assessment

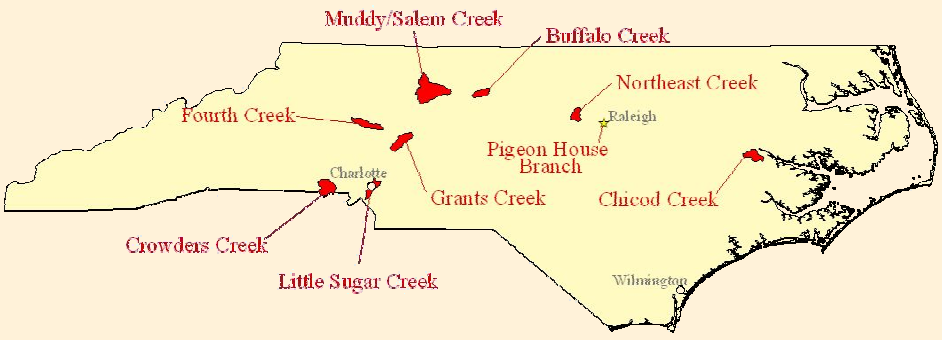
for

TMDL Development and Implementation in

North Carolina

April 28, 2005

North Carolina BST Study



2

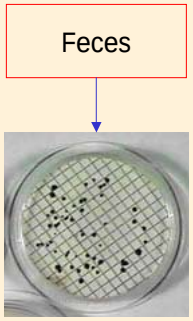
Bacterial Source Tracking

- Laboratory Methods
- Known-Source Library Development
- Statistical Procedures
- Understanding the Reports
- Post-Processing

3

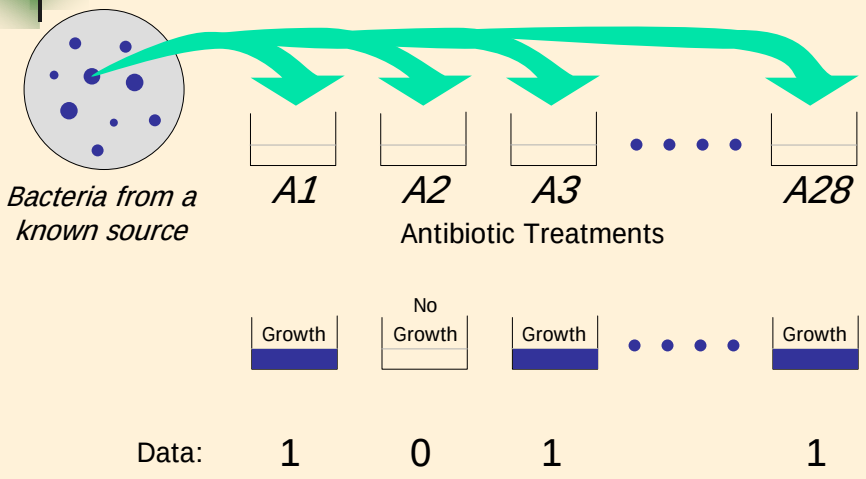
ARA Methodology

- Grow bacteria from known sources of bacteria
- Isolate colonies
- Subject “isolates” to different concentrations and types of antibiotics
- Identify source-related patterns
- Detect similar patterns in water samples



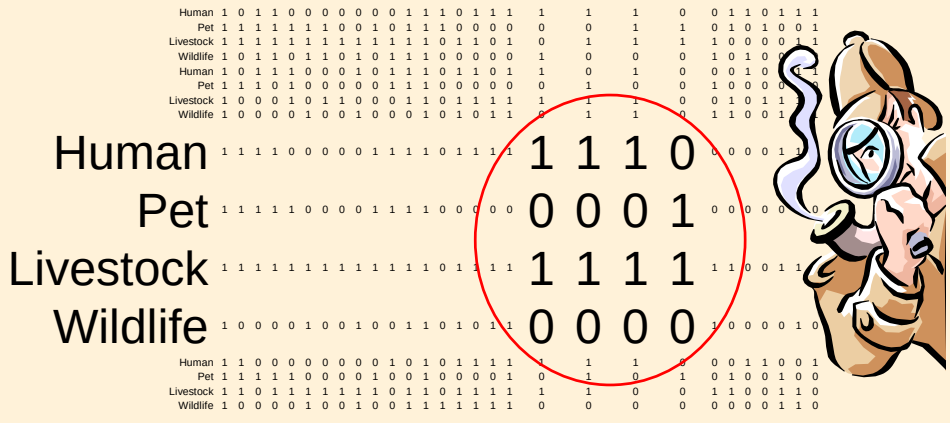
4

Known-Source Library Development



5

Known-Source Library: Data from 1,000+ Isolates



6

Statistical Analysis

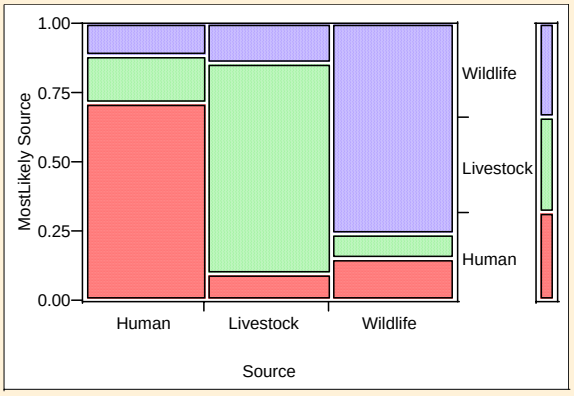
Use known-source library to determine the relationship between Antibiotic Resistance and bacteria source:

Most Likely Source = *f(A1, A2, A3 … A28)*

7

How good is the model?

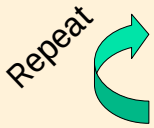
- Rate of Correct Classification (RCC)
  - Human 71%
  - Livestock 76%
  - Wildlife 76%
  - Average (ARCC) 74%



8

Is the library large enough?

- Randomization Test

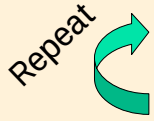


  - Randomly assign sources to samples
  - Create new model
  - Calculate ARCC
  - With 4 source categories, the randomized ARCC should be close to 25%
  - ARCC >> 25% indicates false clustering
    - library is too small

9

Is the library representative?

- Jackknife Analysis



  - Hold back one known-source sample
  - Create model
  - Assess model's ability to categorize held-back sample

10

Jackknifing

Sample	Isolate	A1	A2	A3	A4	•	•	•	A28	Category	Score
1	1	1	1	0	1				1	Human	1
1	2	1	1	1	0				1	Human	1
1	3	1	1	1	1				1	Human	1
1	4	1	1	0	1				0	Human	1
1	5	1	1	0	1				1	Human	1
1	6	1	1	0	1				1	Human	1
1	7	1	0	0	1				1	Human	0
1	8	1	1	0	1				1	Human	1
2	1	1	1	0	0				1	Domestic Animal	1
2	2	0	1	0	0				1	Domestic Animal	1
2	3	1	1	0	0				1	Domestic Animal	0
2	4	1	0	0	0				0	Domestic Animal	0
2	5	1	1	0	0				1	Domestic Animal	1
2	6	1	1	0	0				1	Domestic Animal	1
2	7	1	1	0	0				1	Domestic Animal	1
2	8	1	1	0	1				1	Domestic Animal	0
•											
•											
•											
1000	1	0	1	1	1				0	Wildlife	1
1000	2	0	1	0	1				1	Wildlife	1
1000	3	0	0	1	1				0	Wildlife	1
1000	4	0	1	0	1				0	Wildlife	1
1000	5	0	1	1	1				1	Wildlife	1
1000	6	0	0	1	0				0	Wildlife	0
1000	7	1	1	1	1				1	Wildlife	1
1000	8	0	1	1	0				0	Wildlife	1

Score

79%

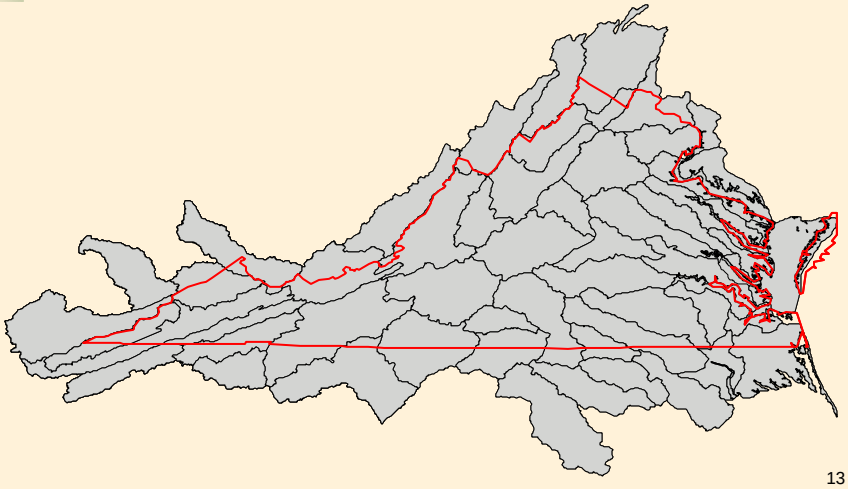
11

Why not throw all of the data into one big library?

- Geographical Differences
  - Combine smaller libraries as appropriate
  - Too big: ARCC goes down
  - Too small: False Clustering goes up
- Temporal Differences
  - Need to update libraries

12

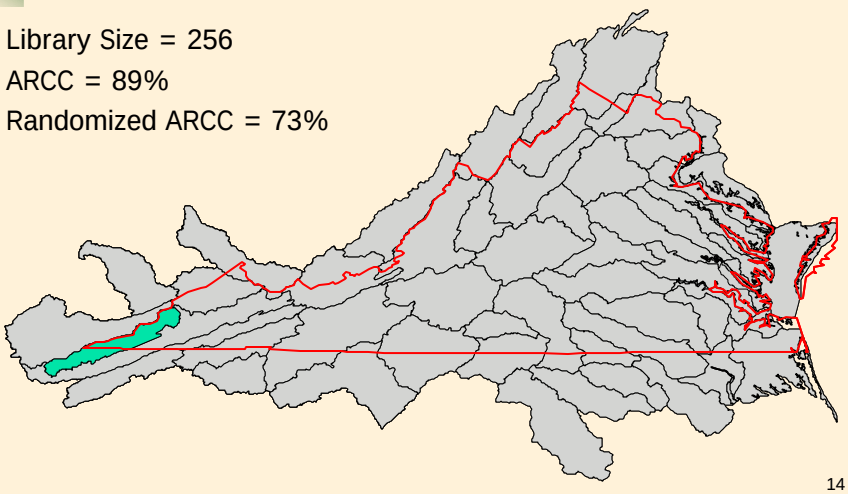
Case Study - Virginia:
Libraries Built within HUC Boundaries



13

Case Study - Virginia:
Start with a HUC

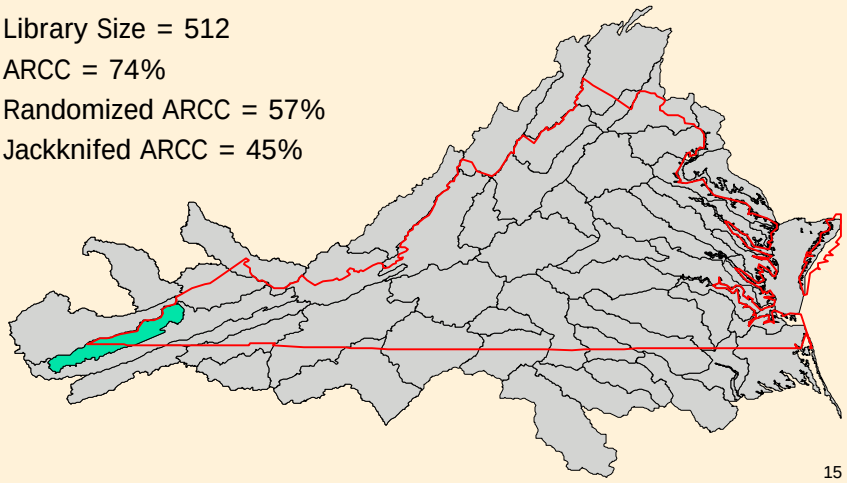
Library Size = 256
ARCC = 89%
Randomized ARCC = 73%



14

Case Study - Virginia:
Start with a HUC

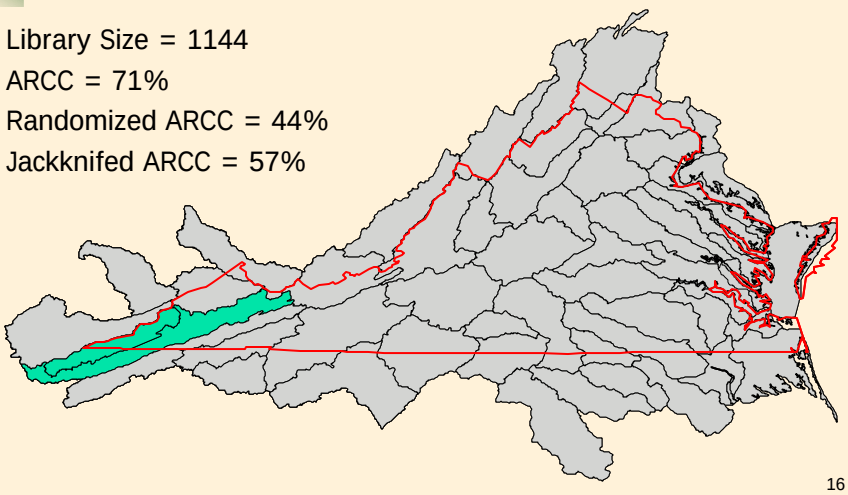
Library Size = 512
ARCC = 74%
Randomized ARCC = 57%
Jackknifed ARCC = 45%



15

Case Study - Virginia:
Add data from nearby HUCs

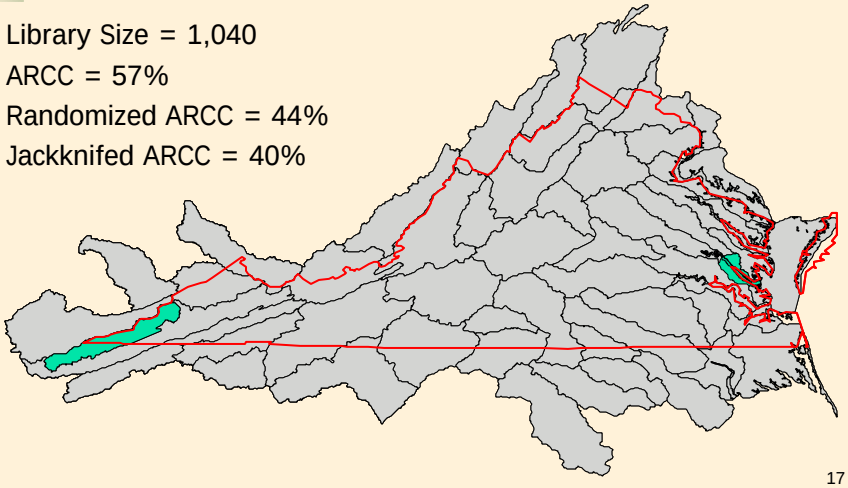
Library Size = 1144
ARCC = 71%
Randomized ARCC = 44%
Jackknifed ARCC = 57%



16

Case Study - Virginia:
Why not distant HUCs?

Library Size = 1,040
ARCC = 57%
Randomized ARCC = 44%
Jackknifed ARCC = 40%



17

Categorization

- Appropriate categories for analysis need to be selected, based on:
  - Known Sources
  - Management Concerns

Species	Categorization	Categorization	Categorization
Human	Human	Human	Human
Horse	Non-Human	Domestic Animals	Livestock
Cattle			
Swine			
Poultry			
Sheep			
Goats			
Cats			Pets
Dogs			
Raccoons		Wildlife	Wildlife
Deer			
Beaver			
Waterfowl			
Muskrat			
etc. …			

18

What source categories are appropriate?

- Human/Non-human
  - Where human waste is the only pollutant of concern
- Human/Domestic Animals/Wildlife
  - A good catch-all grouping
- Human/Pets/Wildlife
  - Watersheds with NO livestock
- Human/Livestock/Pets/Wildlife
  - Watersheds with all sources present
- Others…

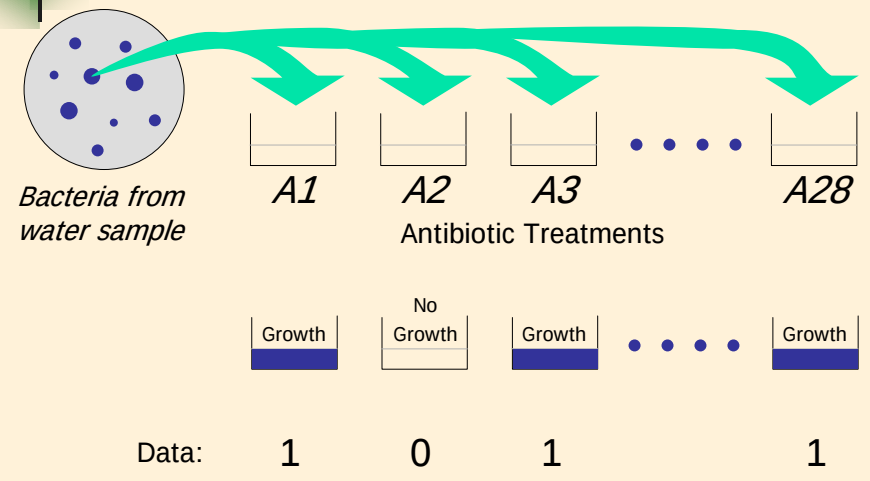
19

What species should be included?

- Major contributors
  - Waterfowl
  - Aquatic mammals
  - Warm blooded animals that frequent the stream corridor, or have a large, wide-spread population
- Opportunity/Availability
- Does it matter if one or two species are not included?
  - No, the statistical procedures used account for that

20

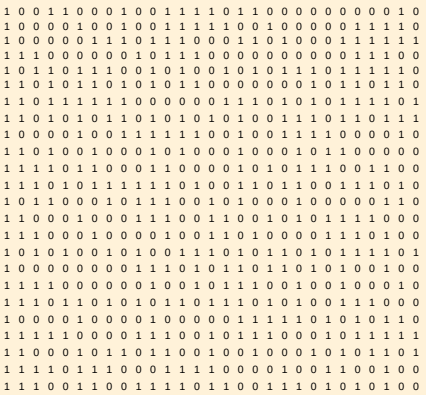
Water Sample Analysis



21

Water Sample Analysis

Water Sample



- 10% Human
- 40% Wildlife
- 30% Livestock
- 20% Pets

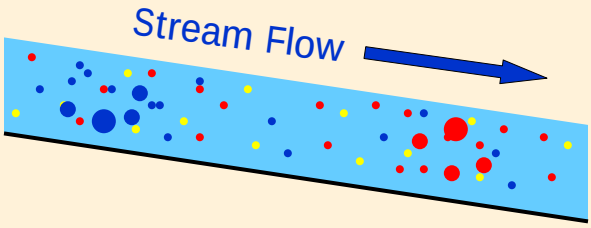
22

Statistical Analyses: Where is there uncertainty?

- **Natural Variability:** How is the bacteria distributed in the stream?
  - "Clumping"
  - Temporal variation
- **Field Sampling:** Was the water sample representative of stream conditions at the time of sampling?
  - Good standard water sampling procedures
- **Statistical Significance:** Have enough isolates been analyzed to draw conclusions about the larger population?
  - Sample size requirements similar to assessing survey results
- **Lab Methodology:** How accurate is the BST methodology?

23

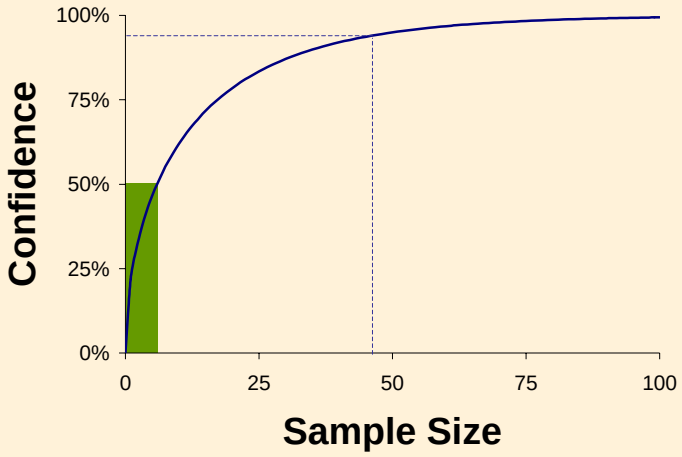
Natural Variability



- Analyses often assume homogeneity in the stream
- As with other pollutants (and more than some), there are non-homogeneities

24

Statistical Significance



25

Uncertainty in the Methodology

- Some resistance patterns will be duplicated in different sources
- The statistical analysis gives the "most likely" source
- The various analyses of the library indicate the degree of uncertainty

Sample	Isolate	A1	A2	A3	A4	• • •	A28	Category
1	1	1	1	0	1		1	Human
1	2	1	1	1	0		1	Human
1	3	1	1	1	1		1	Human
1	4	1	1	0	1		0	Human
1	5	1	1	0	1		1	Human
1	6	1	1	0	1		1	Human
1	7	1	0	0	1		1	Human
1	8	1	1	0	1		1	Human
2	1	1	1	0	0		1	Domestic Animal
2	2	0	1	0	0		1	Domestic Animal
2	3	1	1	0	0		1	Domestic Animal
2	4	1	0	0	0		0	Domestic Animal
2	5	1	1	0	0		1	Domestic Animal
2	6	1	1	0	0		1	Domestic Animal
2	7	1	1	0	0		1	Domestic Animal
2	8	1	1	0	1		1	Domestic Animal
•								
•								
•								
1000	1	0	1	1	1		0	Wildlife
1000	2	0	1	0	1		1	Wildlife
1000	3	0	0	1	1		0	Wildlife
1000	4	0	1	0	1		0	Wildlife
1000	5	0	1	1	1		1	Wildlife
1000	6	0	0	1	0		0	Wildlife
1000	7	1	1	1	1		1	Wildlife
1000	8	0	1	1	0		0	Wildlife

How can we address the uncertainty in the data?

- Natural Variability
  - Sample more, if greater certainty is needed
  - Be careful not to give too much weight to 1 sample
- Field Sampling
  - Use good stream sampling techniques
- Statistical Significance
  - Analyze more isolates per sample, if greater certainty is needed
- Lab Methodology/Analysis
  - Use false-positive rates to assess results

27

Using the False-Positive Rate

- How often does the analysis indicate the category of interest when the source category is something else (False-Positive)?
  - Analyze known-source library
  - FP Rate = (Sum FPs)/(# of “other” isolates analyzed)
- Use FP Rate to determine if percentages are significantly different from zero

28

BST Report

- Chapter 1: Introduction
- Chapter 2: Objectives
- Chapter 3: Methods
  - Source Sampling Strategy
  - Stream Sampling Locations
- Chapter 4: Known-Source Library Development
  - Details follow…
- Chapter 5: Results
  - Details follow…

29

Chapter 4: Known-Source Library Development

- Initial Libraries
  - High RCCs, but high random RCCs
  - Potential for overfitting
- Regional Libraries
  - Acceptable RCCs
  - Lower random RCCs
  - Acceptable Jackknife RCCs
- Urban Library
  - Urban library did not perform as well as regional libraries

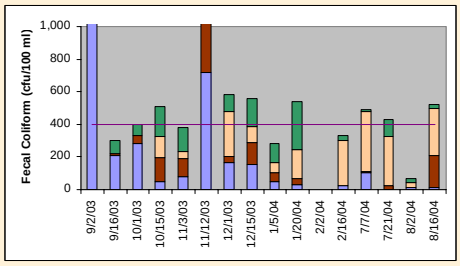
30

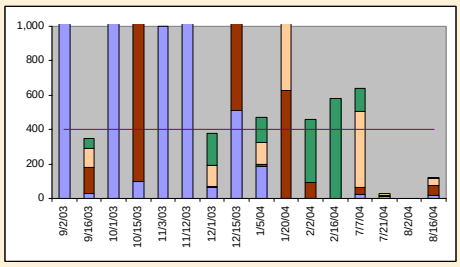
Chapter 5: Results

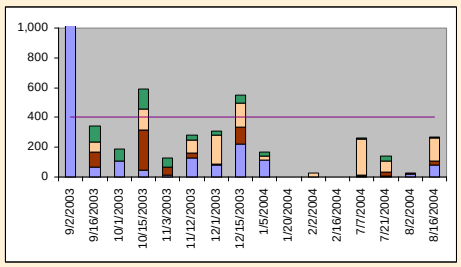
Station ID	Sample Date	Lab ID	*E. coli* Value	*E. coli* Qual	Fecal Coliform Value	Fecal Coliform Qual	Isolates	Bacteria Source Human	Livestock	Pets	Wildlife
CROWDH2O2	6/23/04	2NC98	1200		250		48	0%	**48%**	**19%**	**33%**
CROWDH2O2	7/6/04	2NC103	570		210		48	0%	8%	10%	**82%**
CROWDH2O2	7/20/04	2NC104	600		340		48	0%	15%	4%	**81%**
CROWDH2O2	8/11/04	2NC119	260		140		37	**22%**	0%	**73%**	5%
CROWDH2O2	8/24/04	2NC132	390		260		48	0%	0%	4%	**96%**
CROWDH2O2	9/8/04	2NC136	6000		3100		48	4%	**40%**	12%	**44%**
CROWDH2O2	9/20/04	2NC144	640		420		48	2%	**54%**	6%	**38%**
CROWDH2O2	10/4/04	2NC150	1300		480		48	2%	**25%**	**56%**	17%

31

Muddy & Salem Creeks

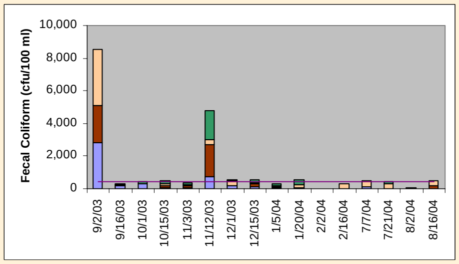


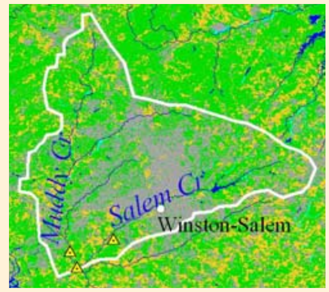


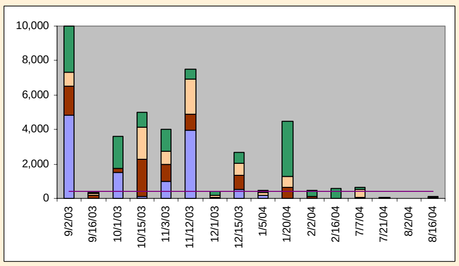


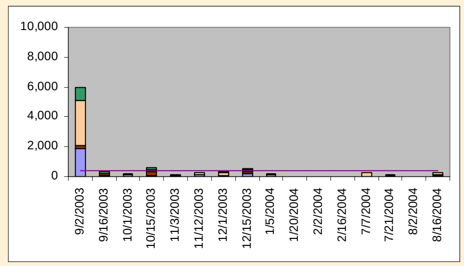
Human Livestock Pets Wildlife

Muddy & Salem Creeks



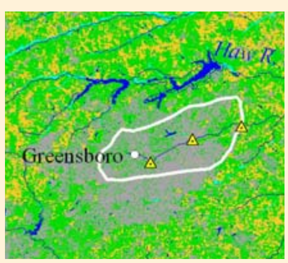


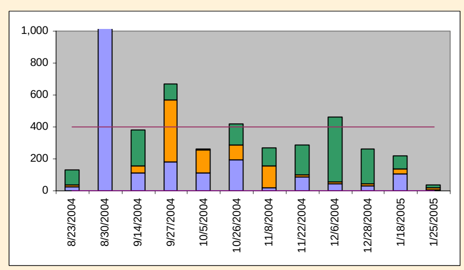


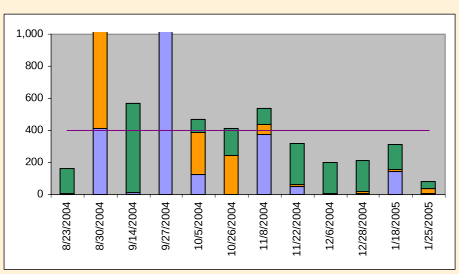


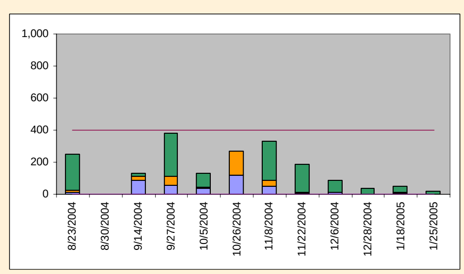
Human Livestock Pets Wildlife

North Buffalo Creek



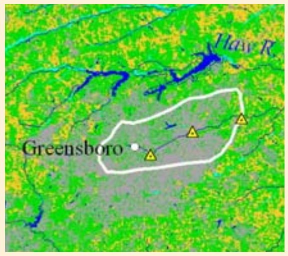


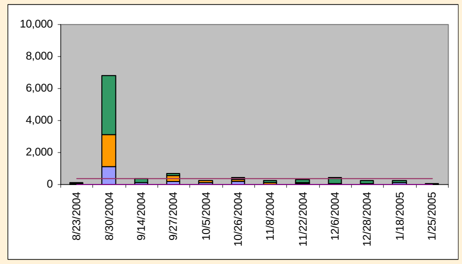


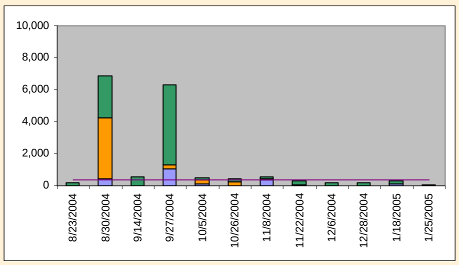


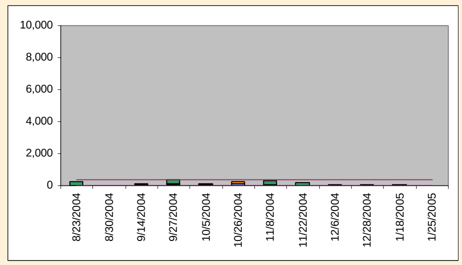
Human Domestic Wildlife

North Buffalo Creek



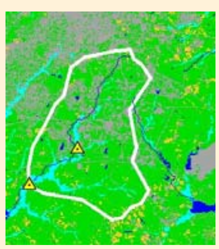


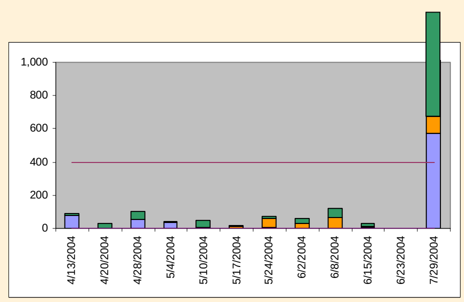


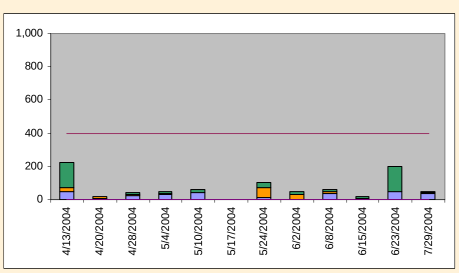


Human Domestic Wildlife

Northeast Creek

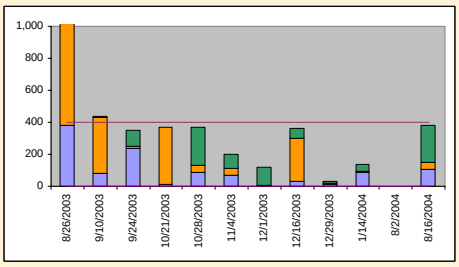


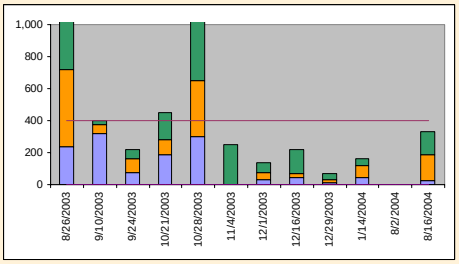


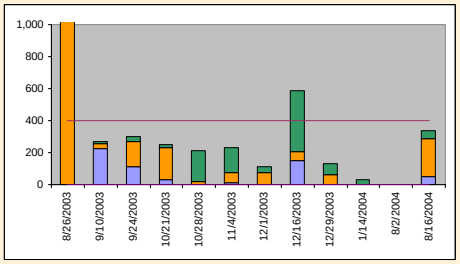


Human Domestic Wildlife

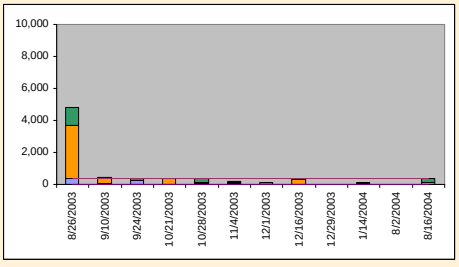
Little Sugar Creek

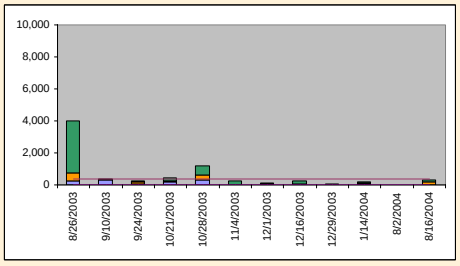


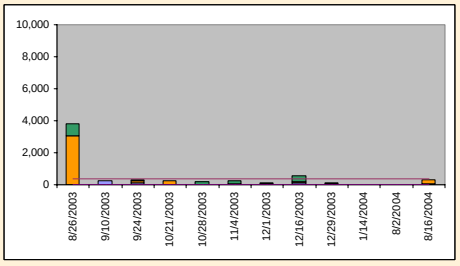




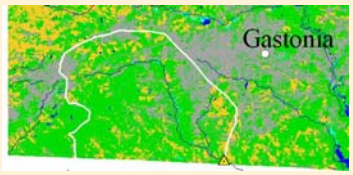
Little Sugar Creek

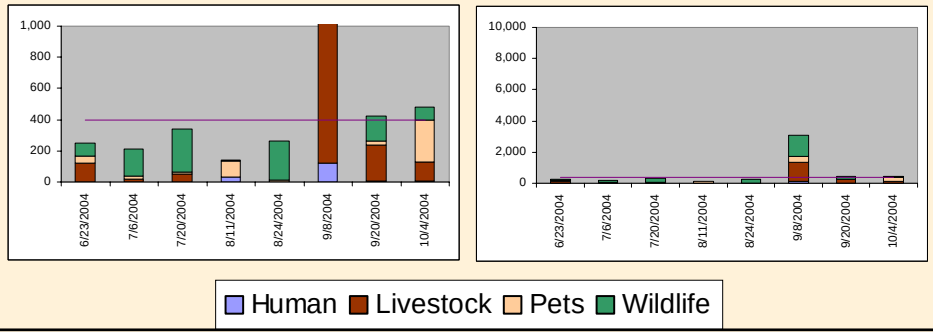




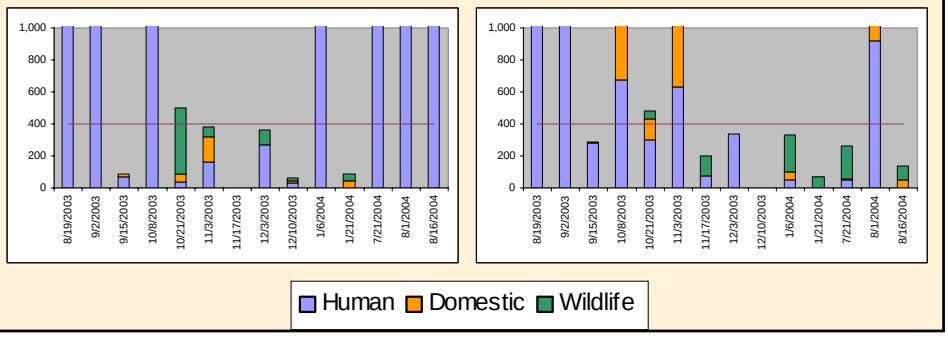


Crowders Creek

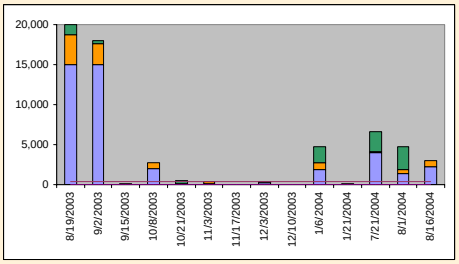


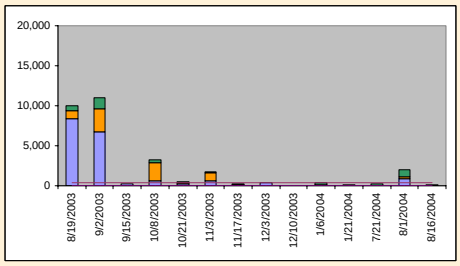


Pigeon House Branch

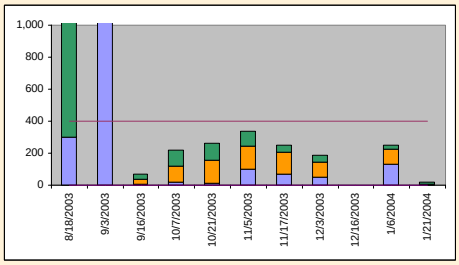


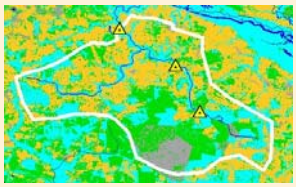
Pigeon House Branch

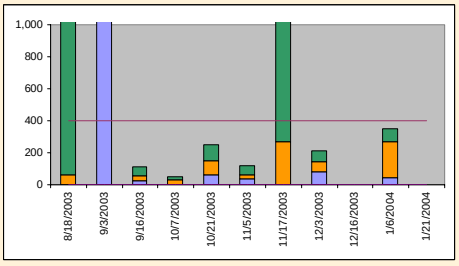


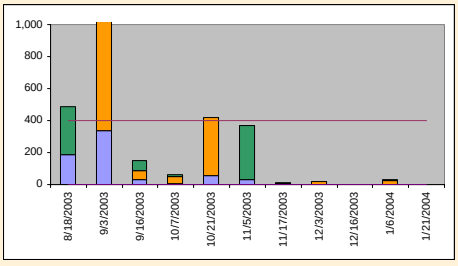


Chicod Creek

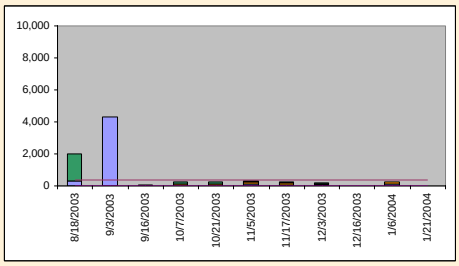


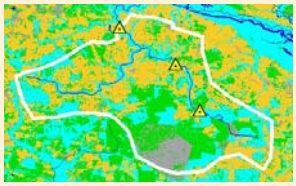


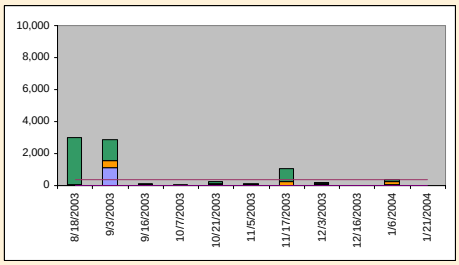


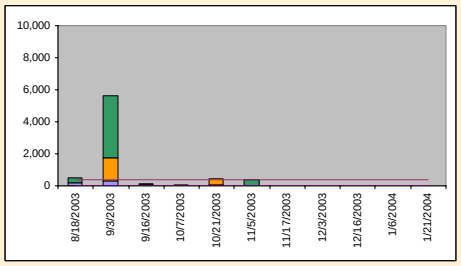


Chicod Creek



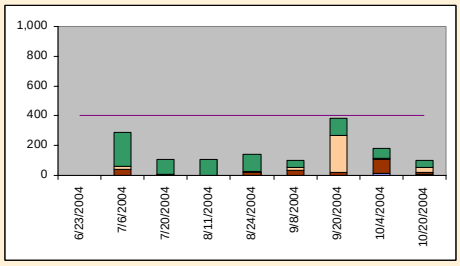






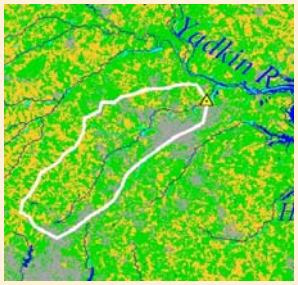
Human Domestic Wildlife

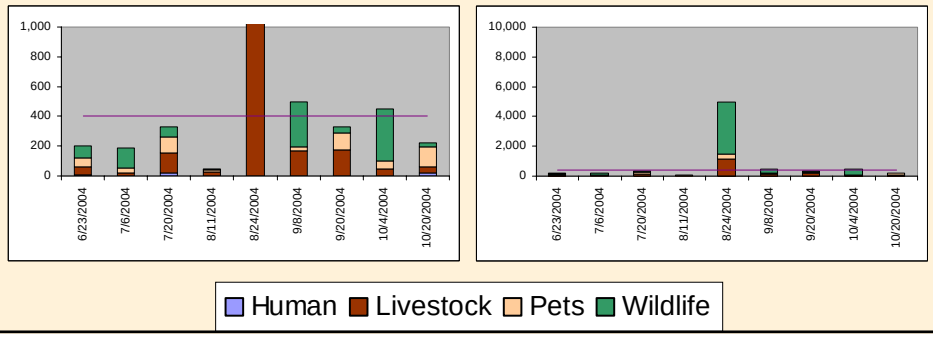
Fourth Creek



Human Livestock Pets Wildlife

Grants Creek





Summarizing Your BST Results

- **Average** of each source category
- **Isolates** weighted average
- **Concentration** weighted average
- **Flow** weighted average
- **Combined** weighted average

46

Questions?

47