SPARROW MODELING WORKSHOP

Section 4.

Calibration Data Sets

Load Estimation From Stream Monitoring Data

USGS

science for a changing world

Overview of the Development of a SPARROW Calibration Data Set

1. Use of Stream Load Information in the Model

2. Sources and Characteristics of a Calibration Data Set

3. Chesapeake Bay as an Example

4. Load Estimation Methodology

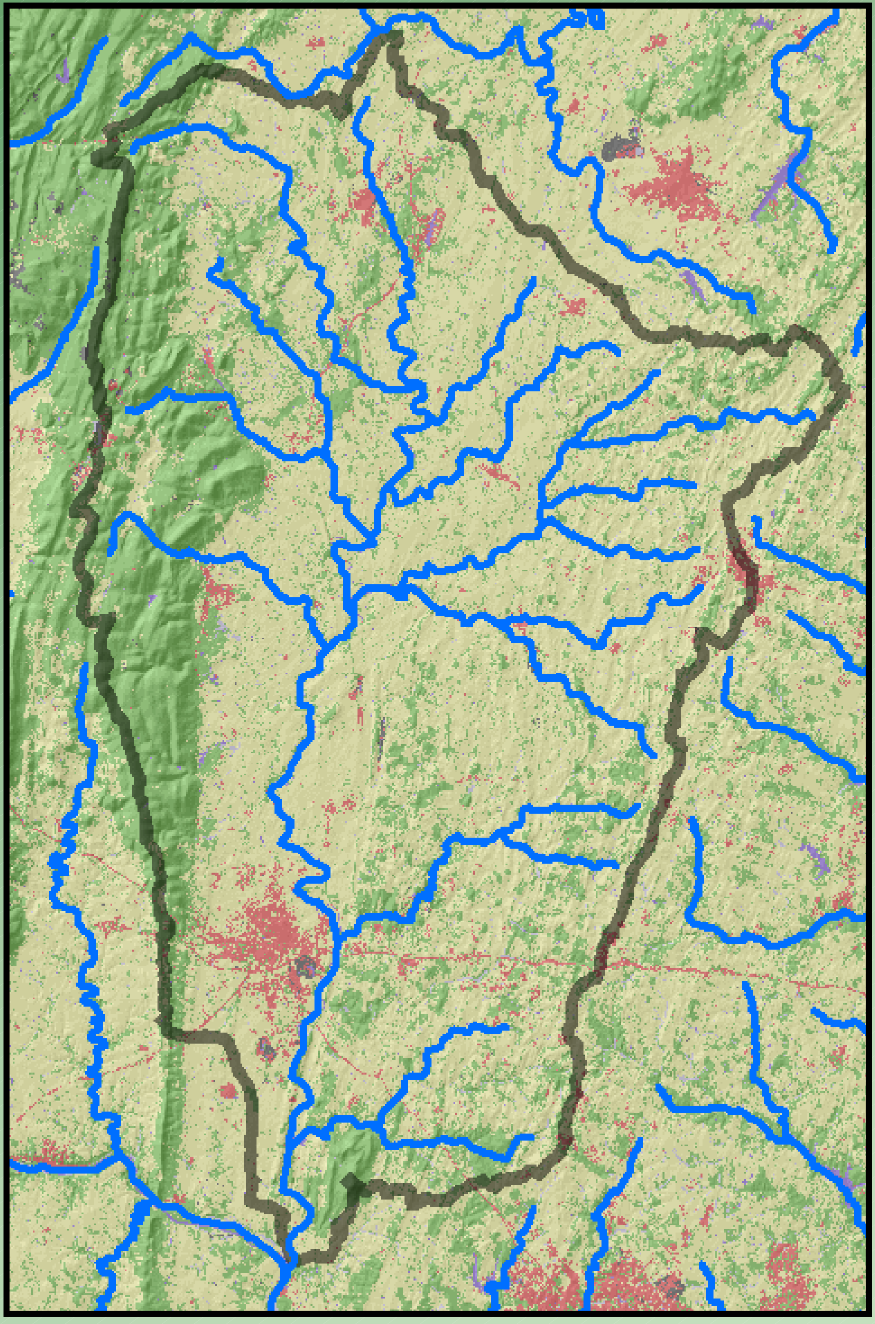
How is

Stream Loading Information

Used in SPARROW?

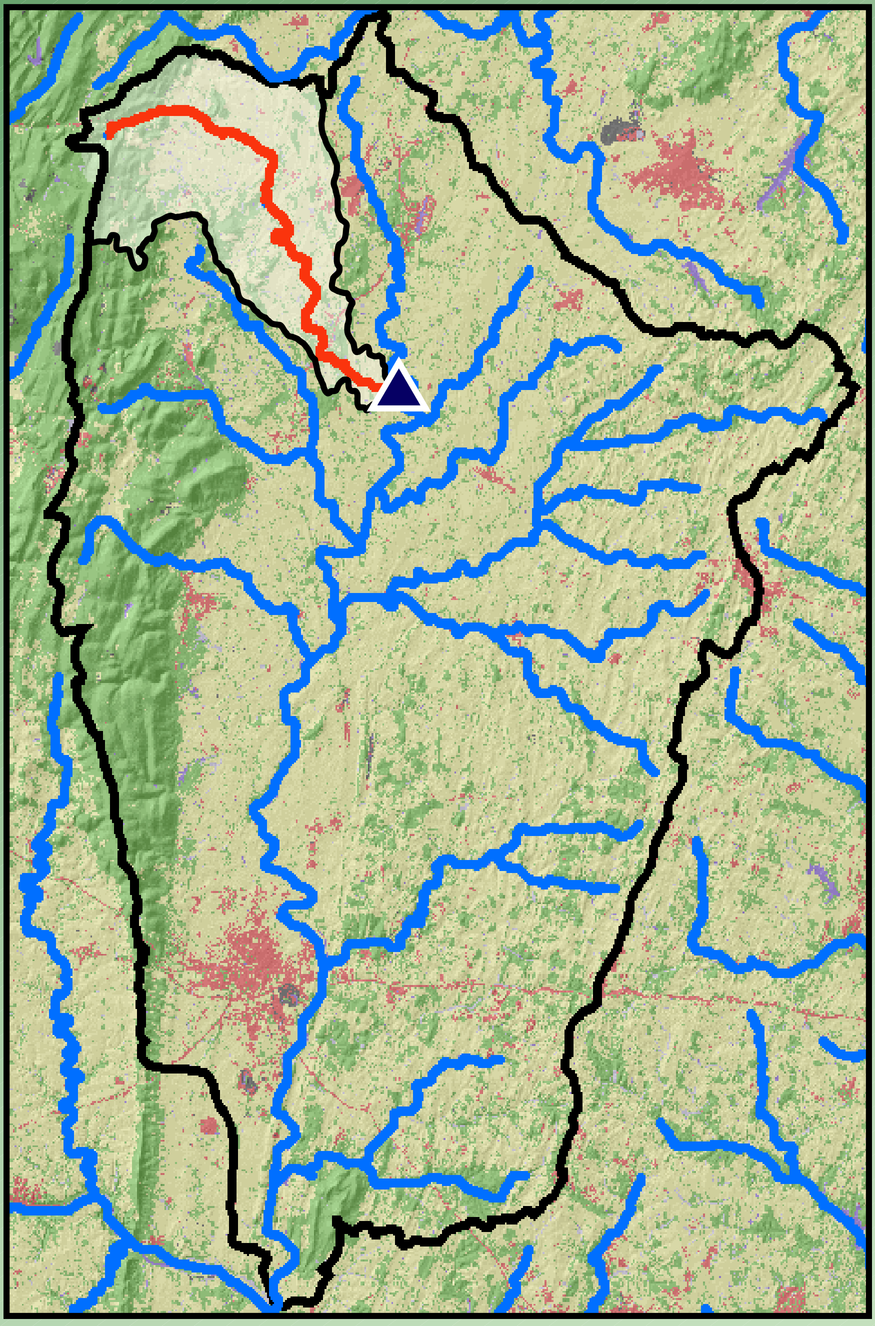
Use of Stream Load Data For Model Calibration

1. Basin of Interest



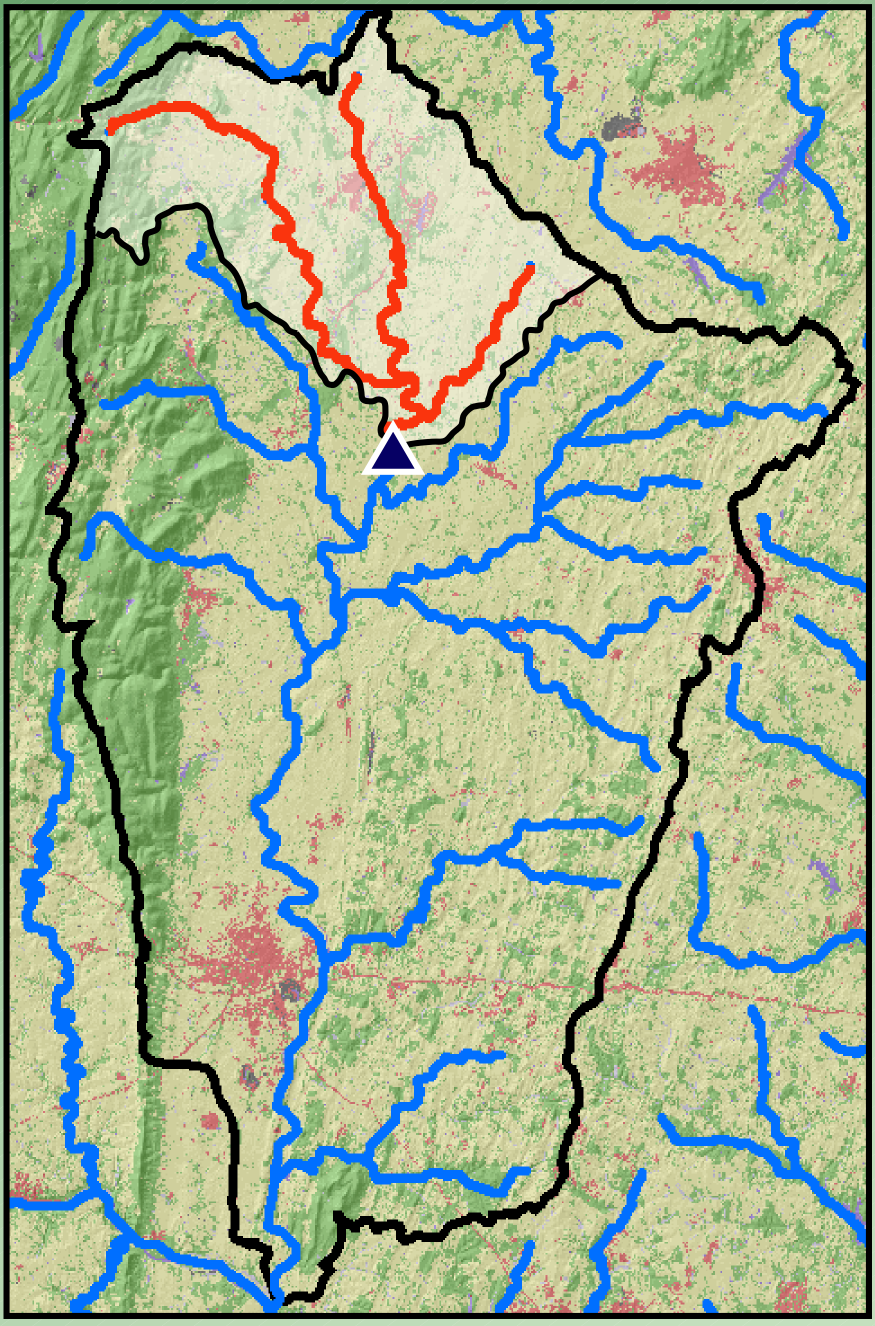
Use of Stream Load Data For Model Calibration

1. Basin of Interest
2. Start With Headwater Subbasin



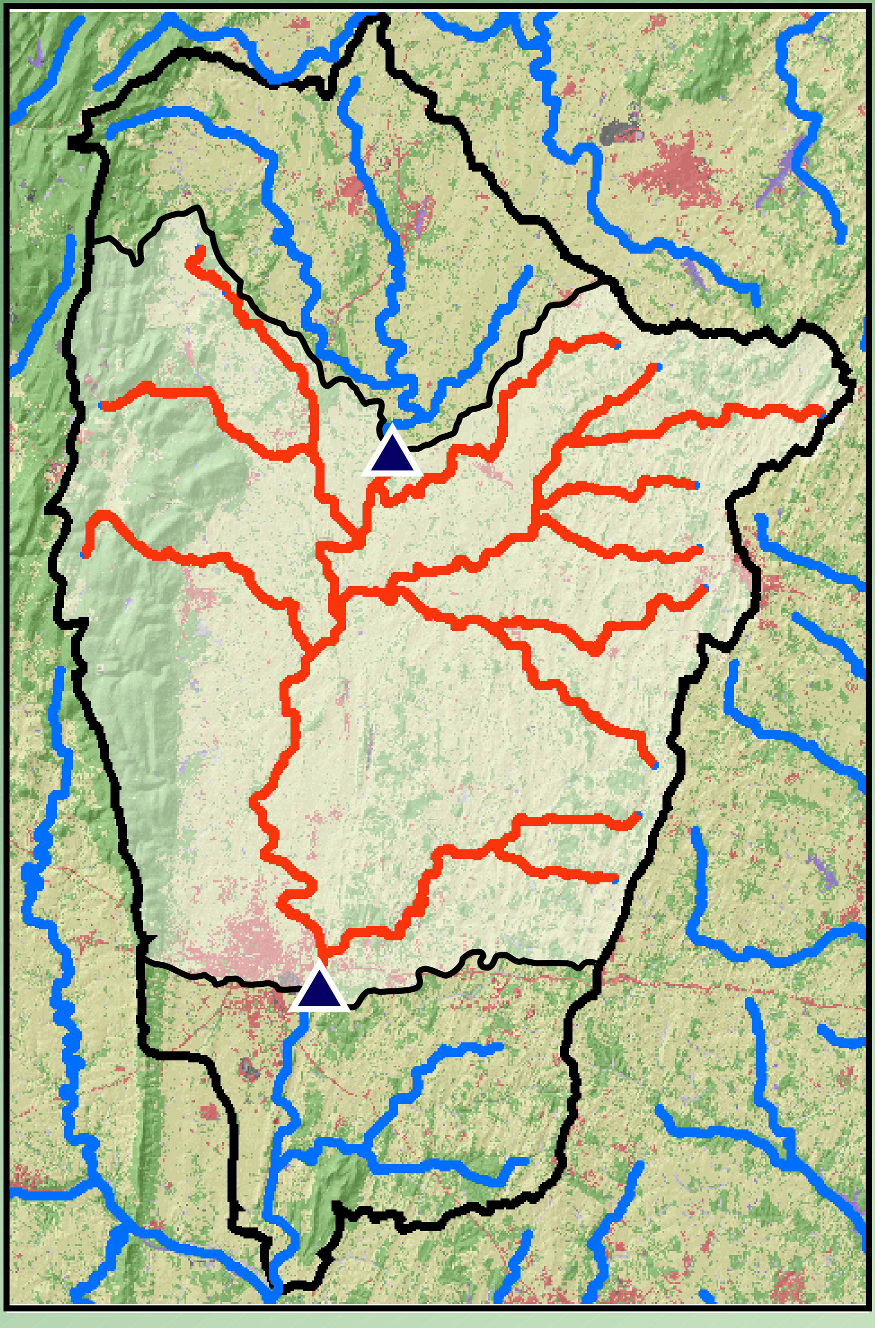
Use of Stream Load Data For Model Calibration

1. Basin of Interest
2. Start With Headwater Subbasin
3. Stations May Relate to Larger Drainages



Use of Stream Load Data For Model Calibration

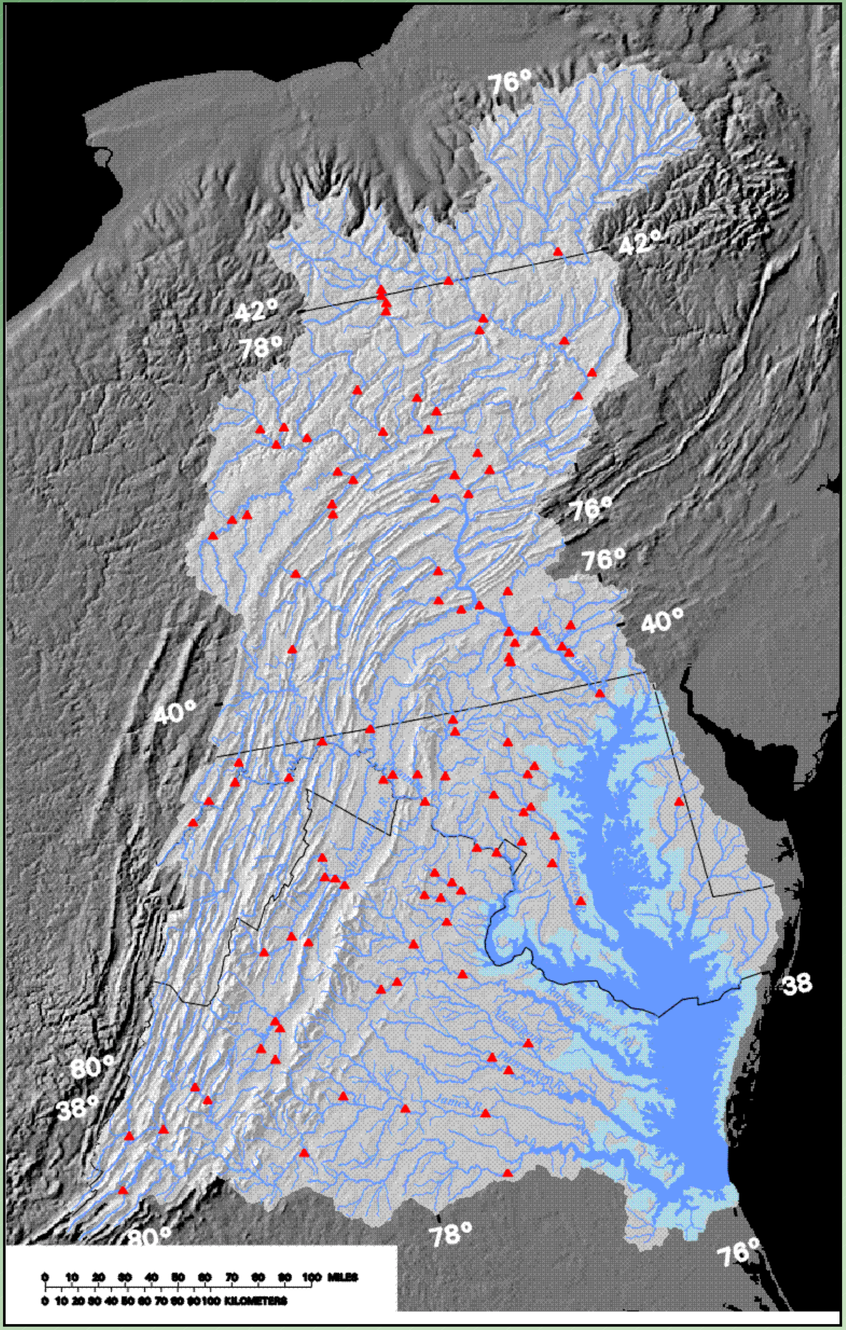
1. Basin of Interest
2. Start With Headwater Subbasin
3. Stations May Relate to Larger Drainages
4. Multiple Stations Define Intervening Drainages



Chesapeake Bay SPARROW

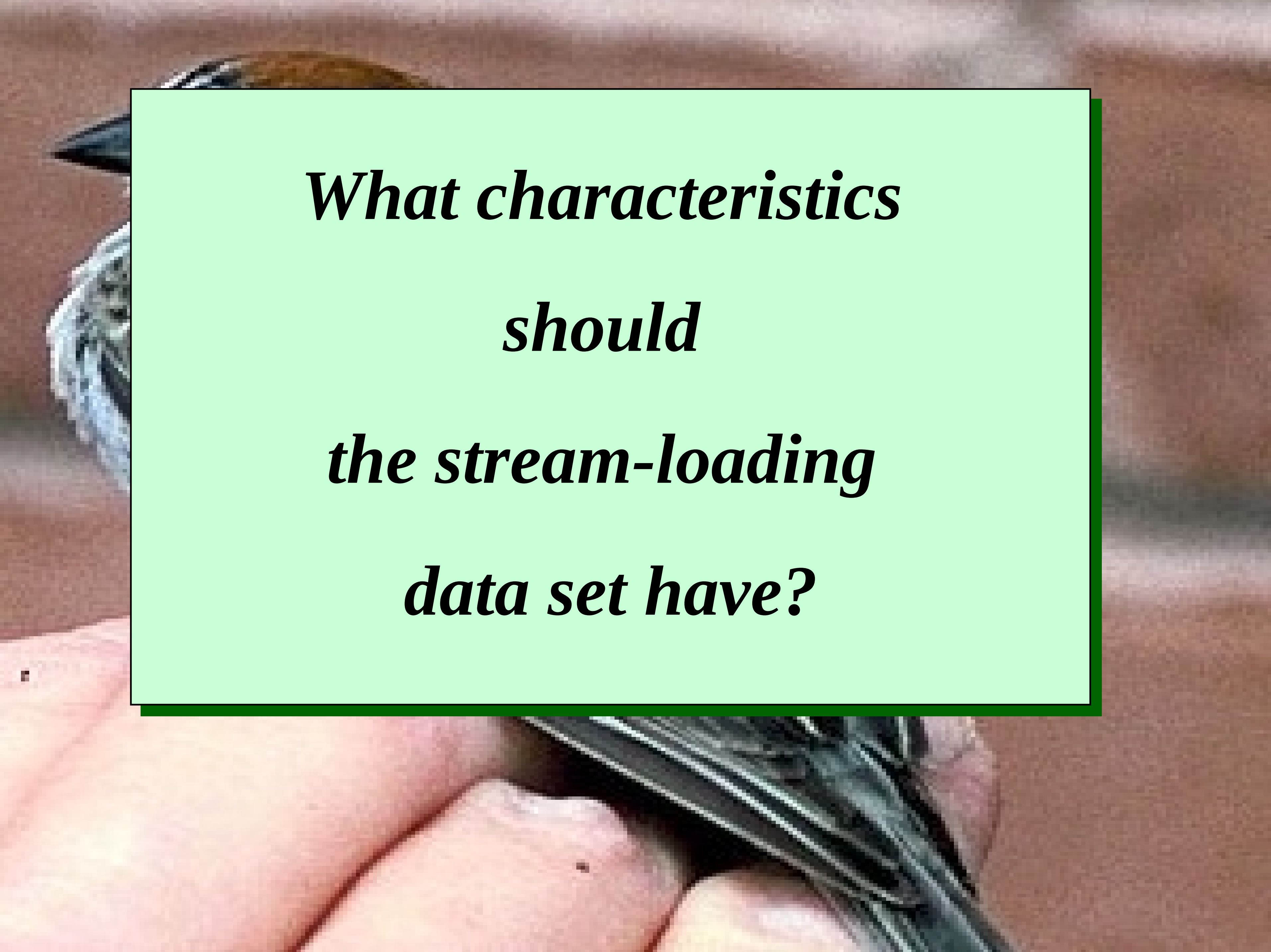
Version I

Network / Load Sites



Sources of Load Data Compiled for National and Regional SPARROW Models

National Model	~ 400 sites	USGS / NASQAN
Chesapeake Bay	~ 130 sites	USGS (~15%) / State (~75%)
New Zealand	37 sites	NZ regional network
North Carolina	44 sites	USGS (~15%) / State (~75%)
New England	82 sites	USGS (~75%) / Other (25%)
Delaware	~150 sites	USGS (~50%) / State (40%)



What characteristics should the stream-loading data set have?

Desirable Characteristics of Stream Load Data Set

1. As large a number of sites (degrees of freedom) as possible

2. Spatially distributed set of sites

3. Set of sites that are representative of spatial variability in geographic characteristics

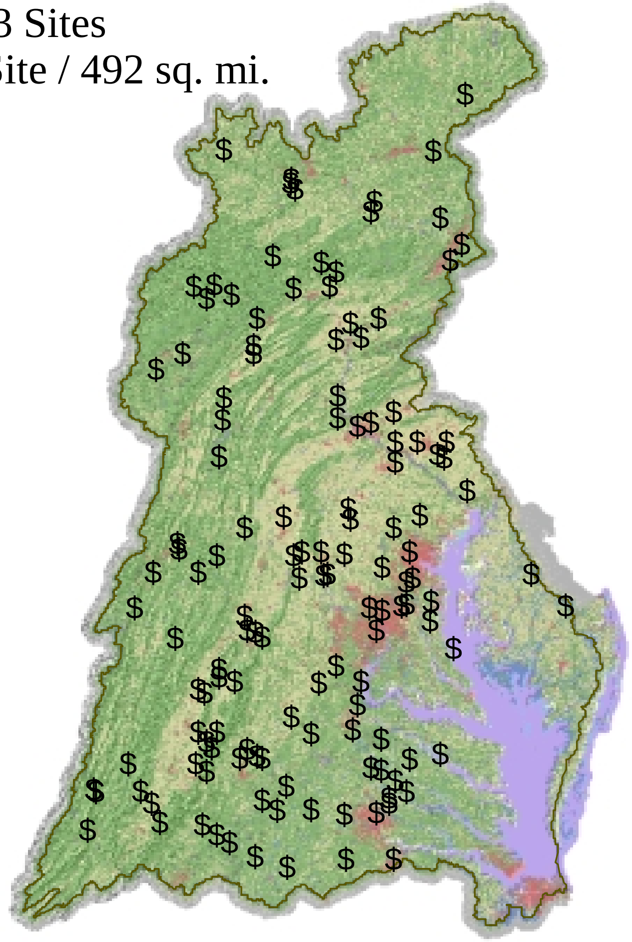
Effect of Sample Size

Number of Load Sites

Current Sampling Density

133 Sites

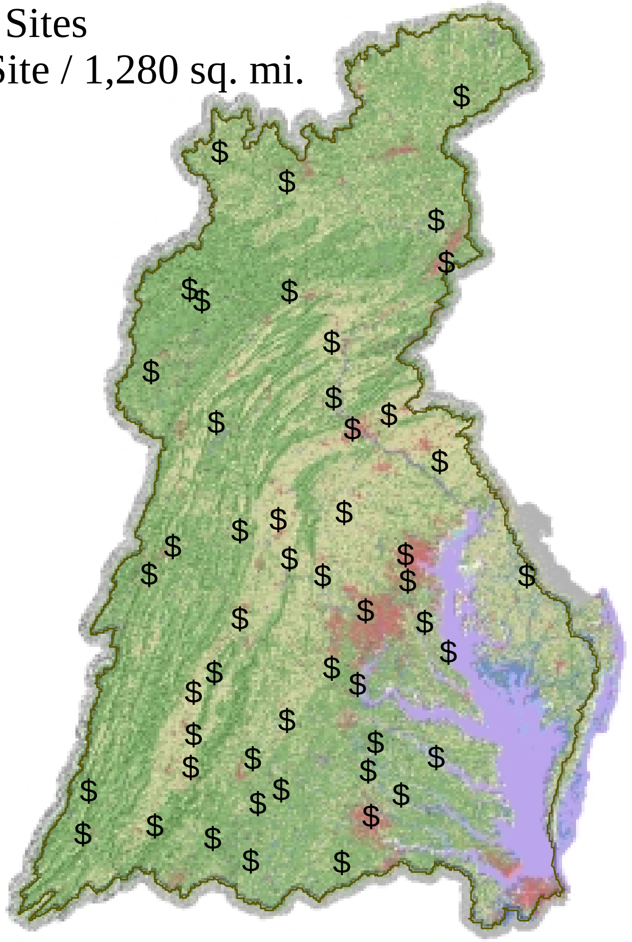
1 Site / 492 sq. mi.



Potential Sampling Density

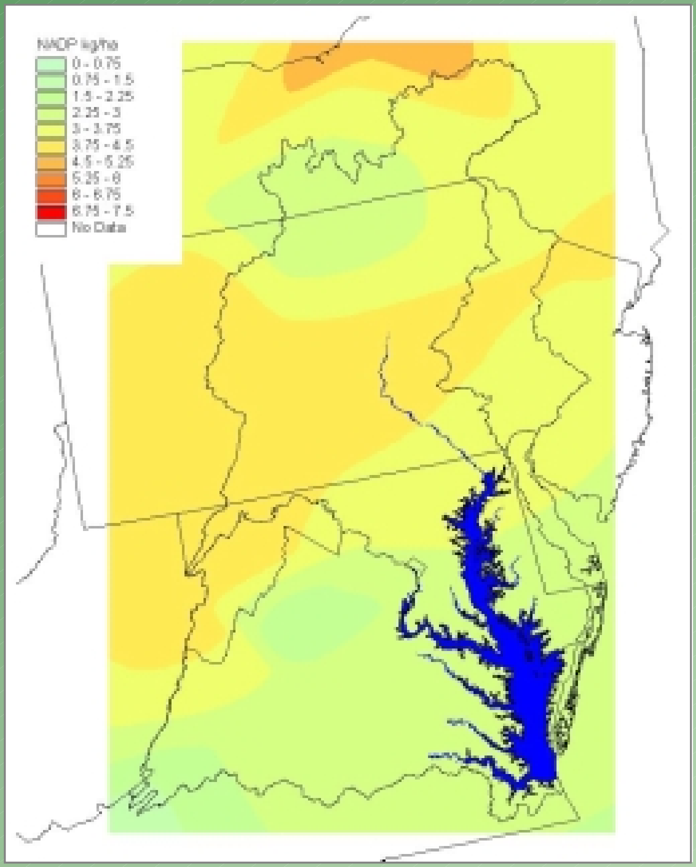
50 Sites

1 Site / 1,280 sq. mi.

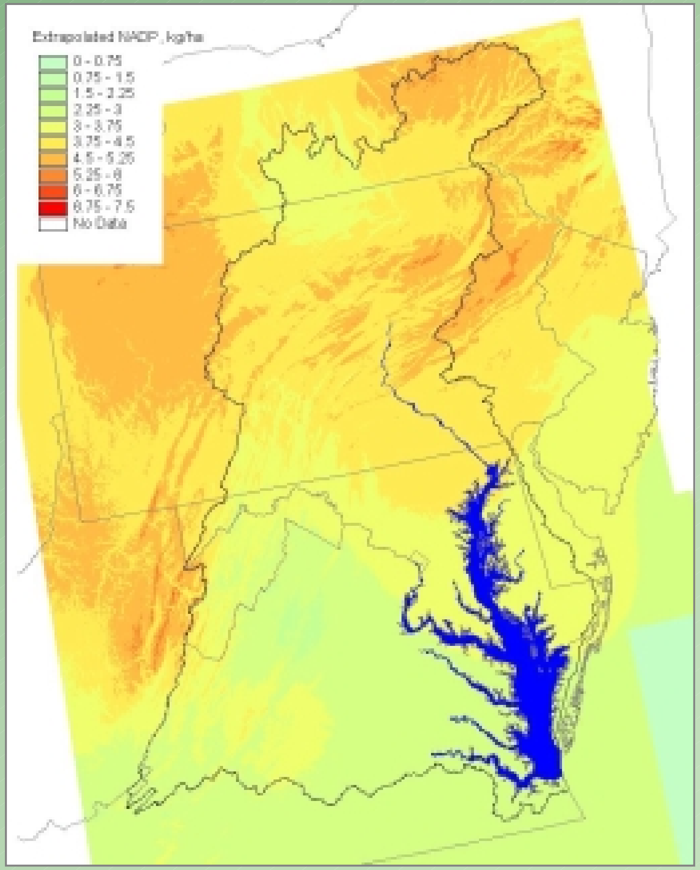


Nitrate Wet-Deposition Data Sets
Chesapeake Bay Region

Interpolated NADP Data



Enhanced NADP Data



Options for Maximizing the Number of Stream Load Sites

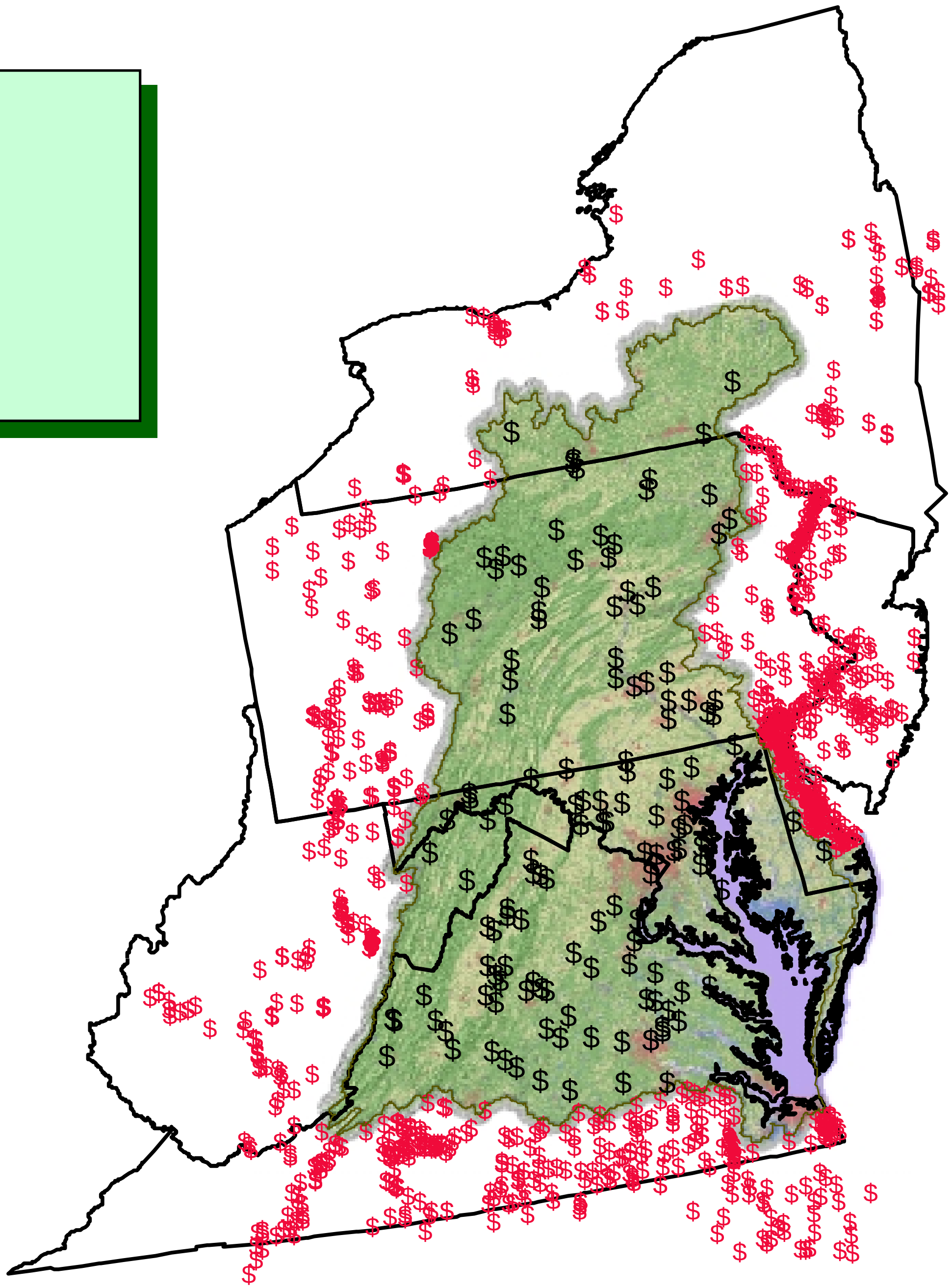
1. Consider extending the size of the study area.

2. Consider using a more detailed stream-reach data set.

Extending the Size of the Study Area

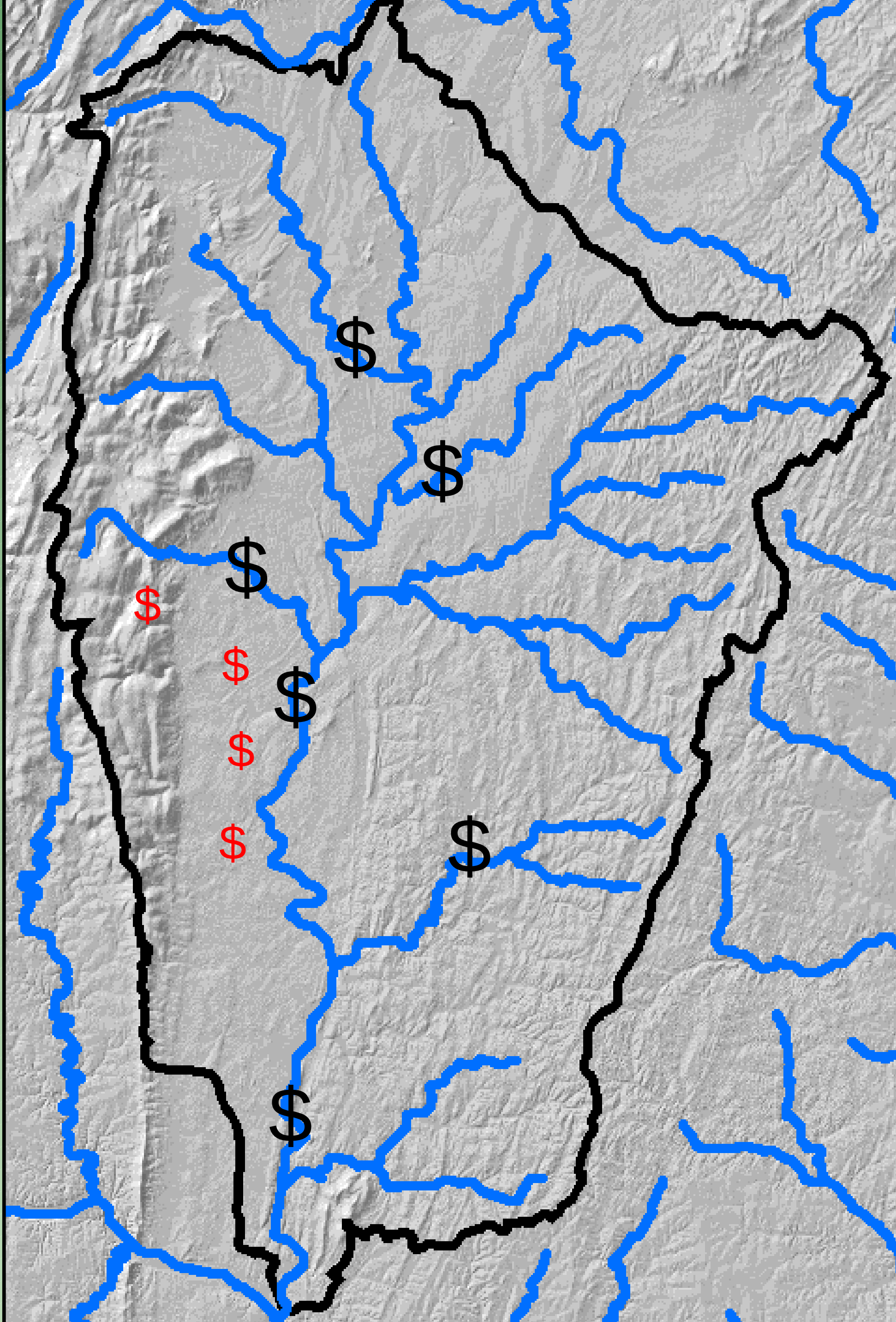
1. Benefit – potentially provides a larger number of sites

2. Cost – requires additional work to develop expanded spatial data sets



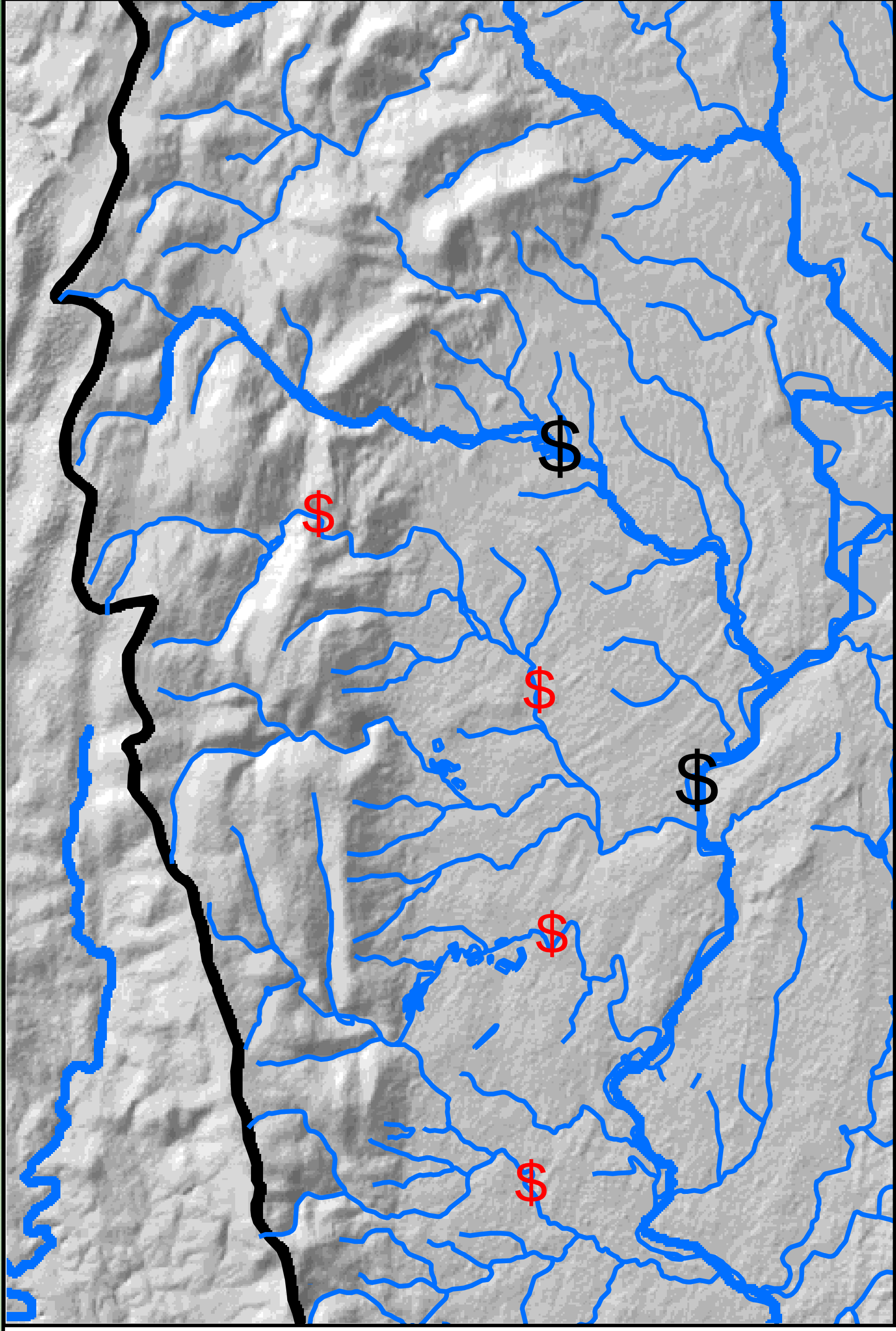
Maximize the Number Of Load Sites Sites Off of Network

Scale of network
can limit the number
of load sites included



Maximize the Number Of Load Sites Sites Off of Network

Shifting to more spatially detailed network is one way to increase the number of sites included.



Desirable Characteristics of Stream Load Data Set

1. As large a number of sites (degrees of freedom) as possible

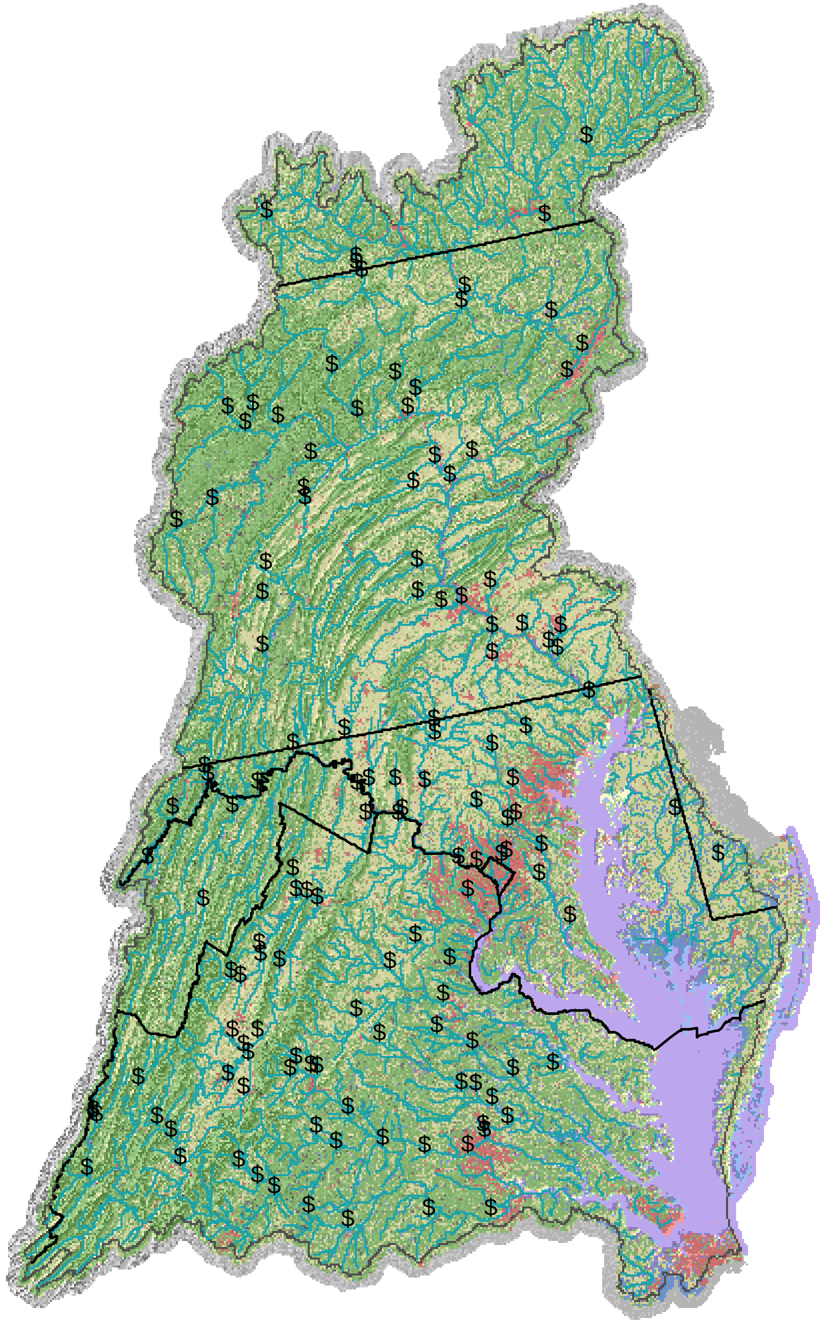
2. Spatially distributed set of sites

3. Set of sites that are representative of spatial variability in geographic characteristics

Jurisdictional Variations In Monitoring Programs

Need spatially distributed set of sites.

Often need to deal with spatial variations in network due to jurisdictional differences or changes in monitoring programs.



Desirable Characteristics of Stream Load Data Set

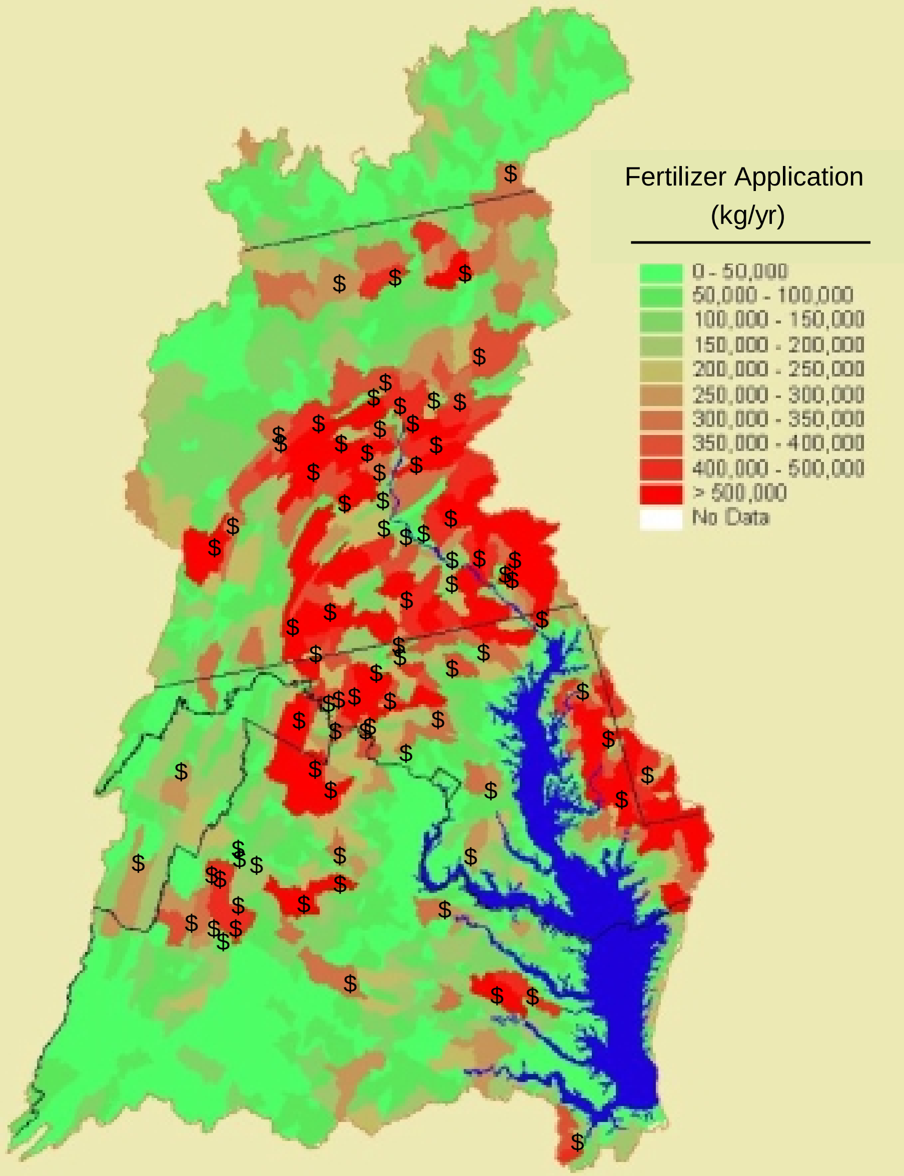
1. As large a number of sites (degrees of freedom) as possible

2. Spatially distributed set of sites

3. Set of sites that are representative of spatial variability in geographic characteristics

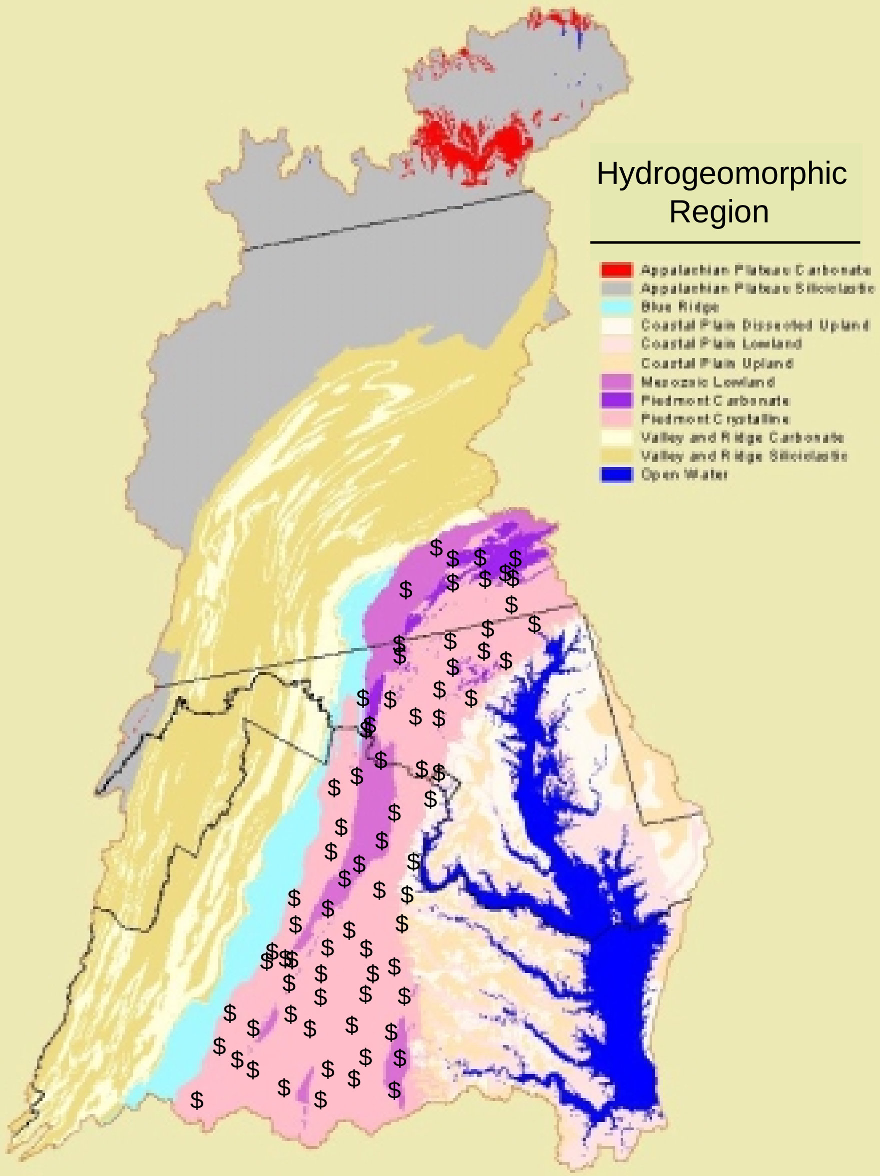
Network that is Representative of Spatial Variability in Sources

Bias or
"mis-identification" of
source / load relationship
in
targeted monitoring
network.



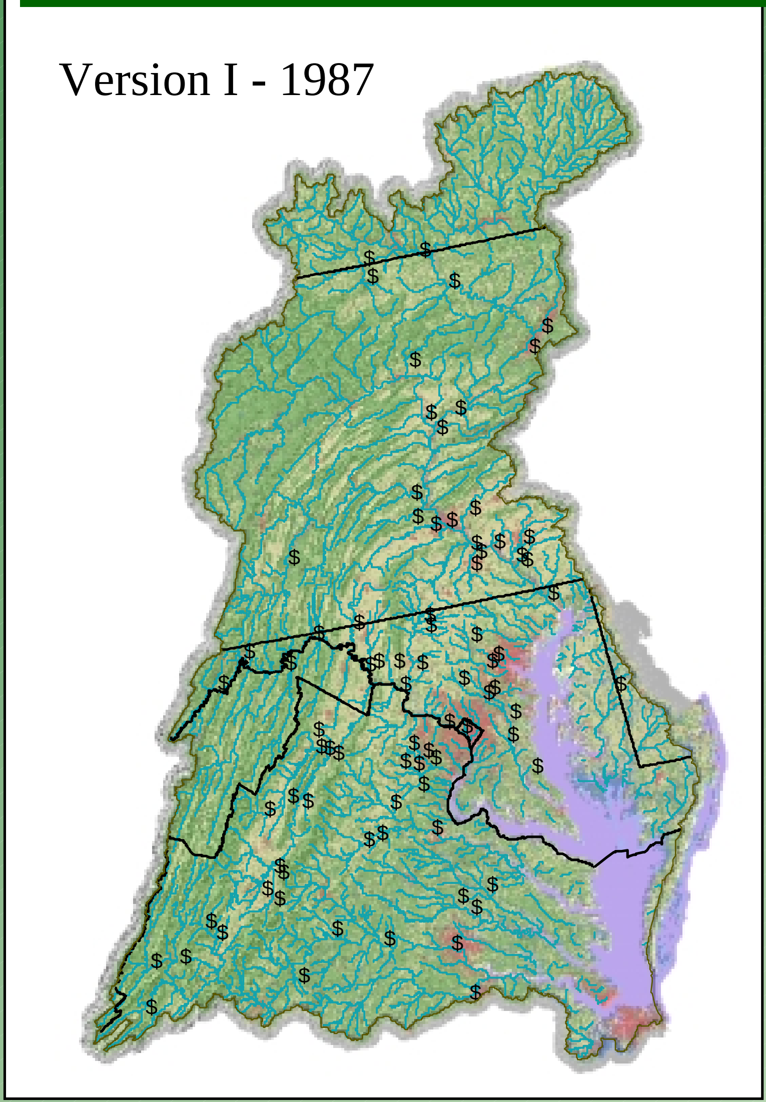
Network that is Representative of Spatial Variability in Basin Characteristics

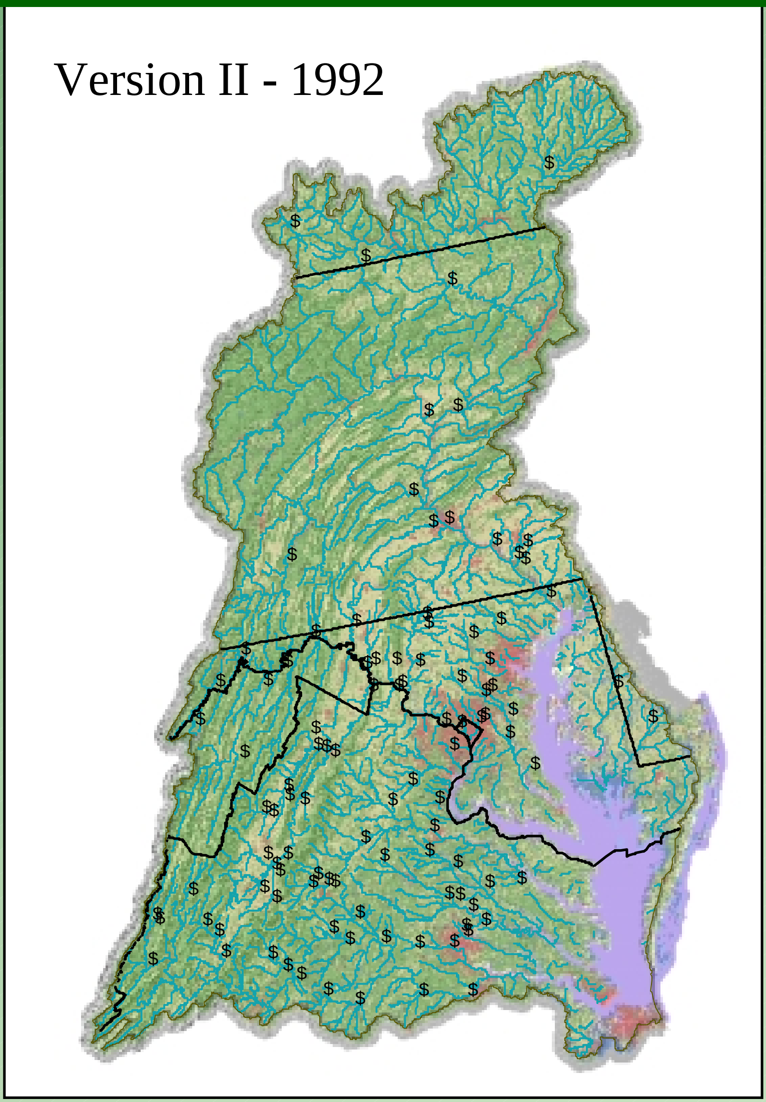
Bias or
“mis-identification” of
land-to-water delivery
in
targeted monitoring
network.



Chesapeake Bay SPARROW

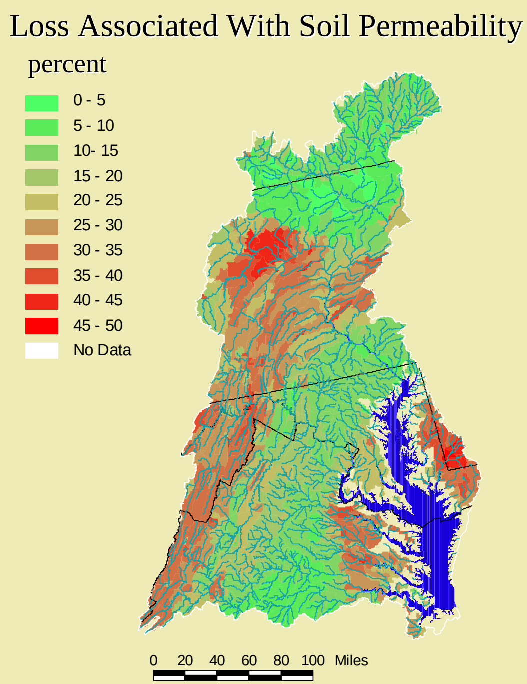
Version I / Version II Total Nitrogen Load Sites



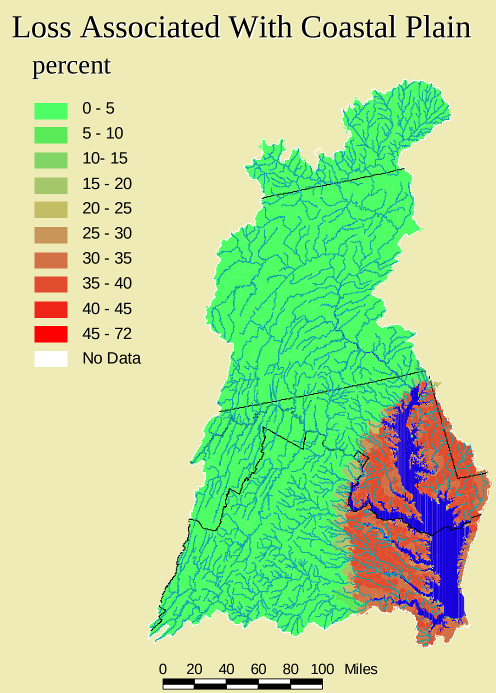


Chesapeake Bay SPARROW

Version I / Version II Land-to-Water Loss Terms



Version I - 1987

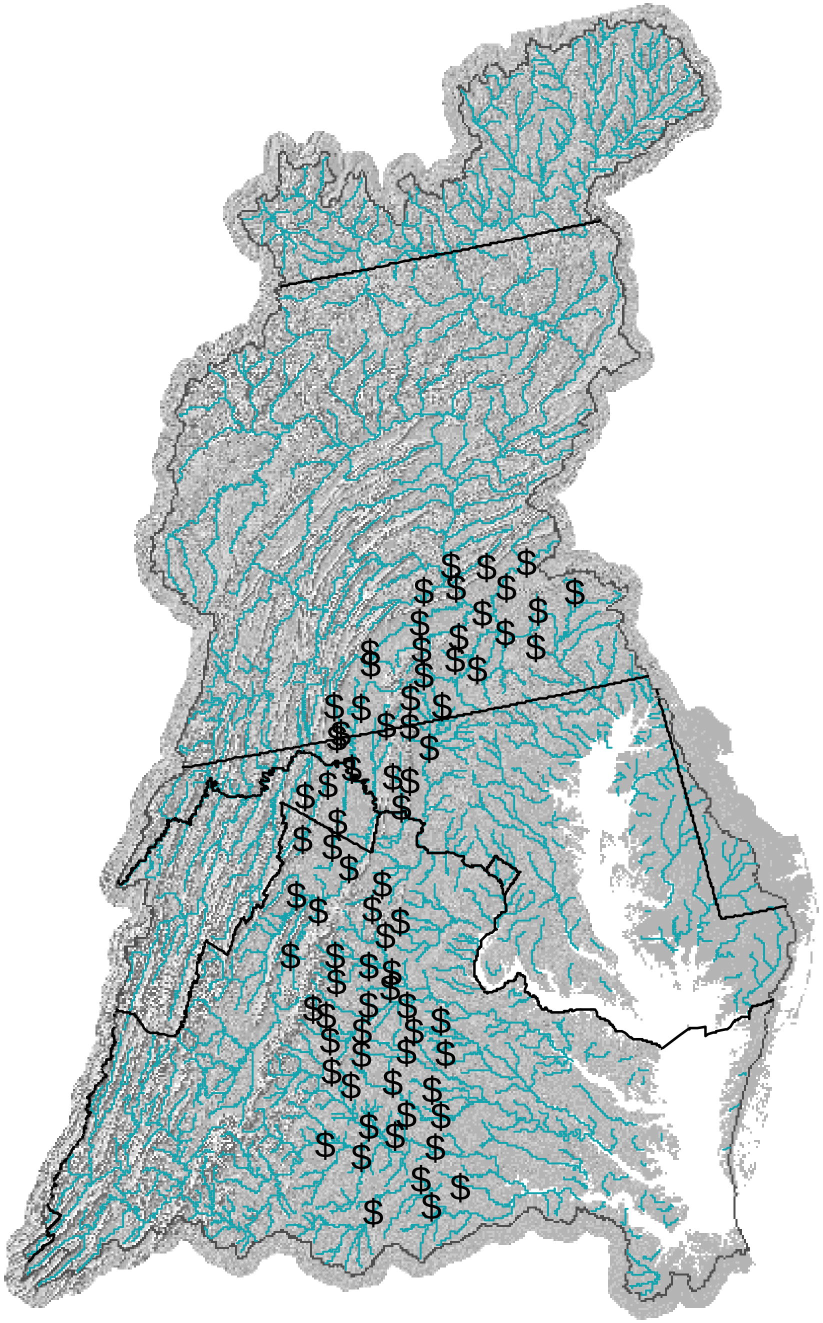


Version II - 1992

Network that is Representative of Spatial Variability in Stream Characteristics

Bias or
“mis-identification” of
instream loss.

Network needs to be
representative of
range of stream sizes,
channel shapes and “nesting”.



How was the stream-load data set built for the Chesapeake Bay SPARROW model?

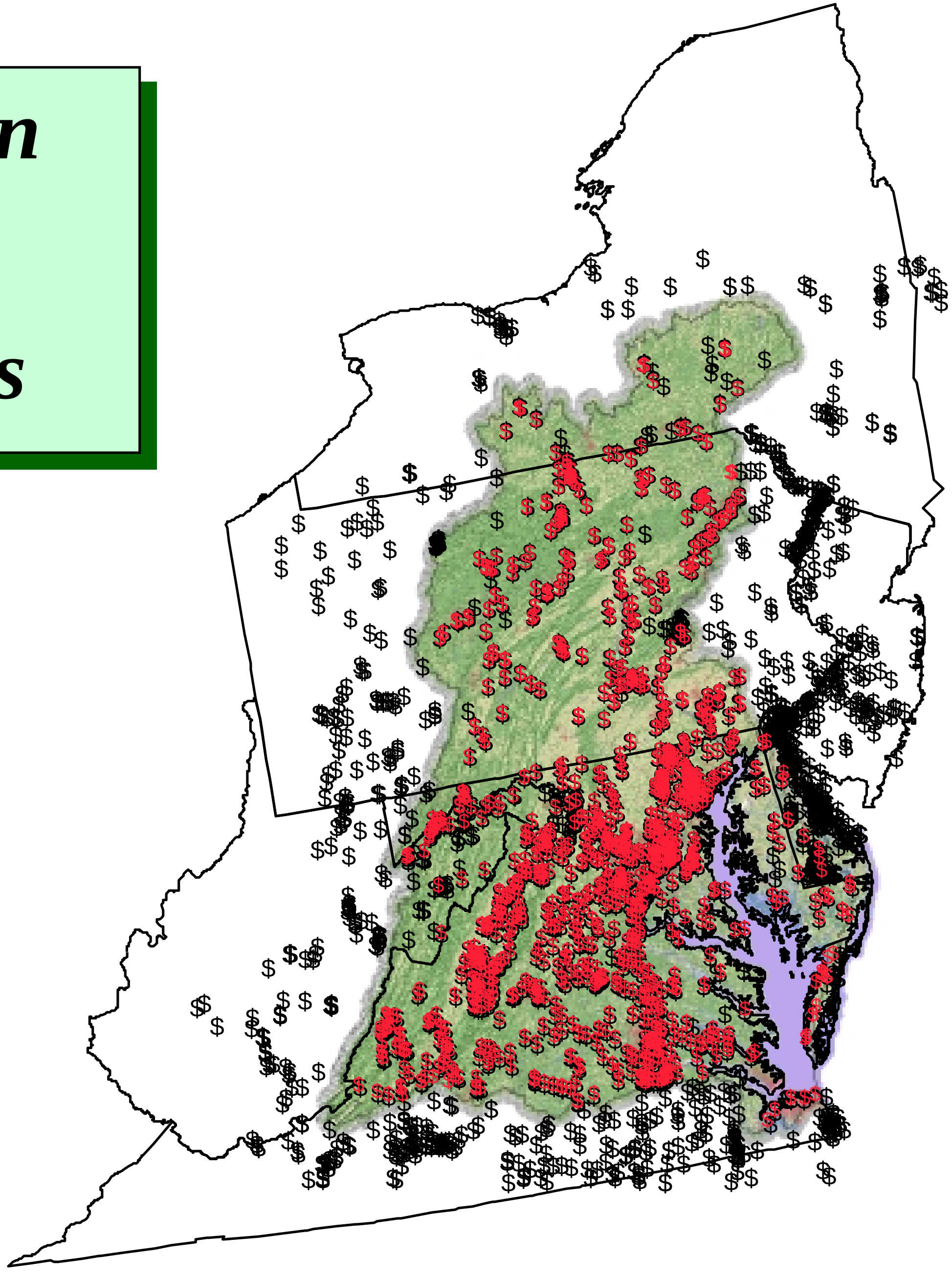
Initial Identification Of Appropriate Water-Quality Sites

Initial Retrieval

- 2,451 potential sites identified in and around Chesapeake Bay watershed.

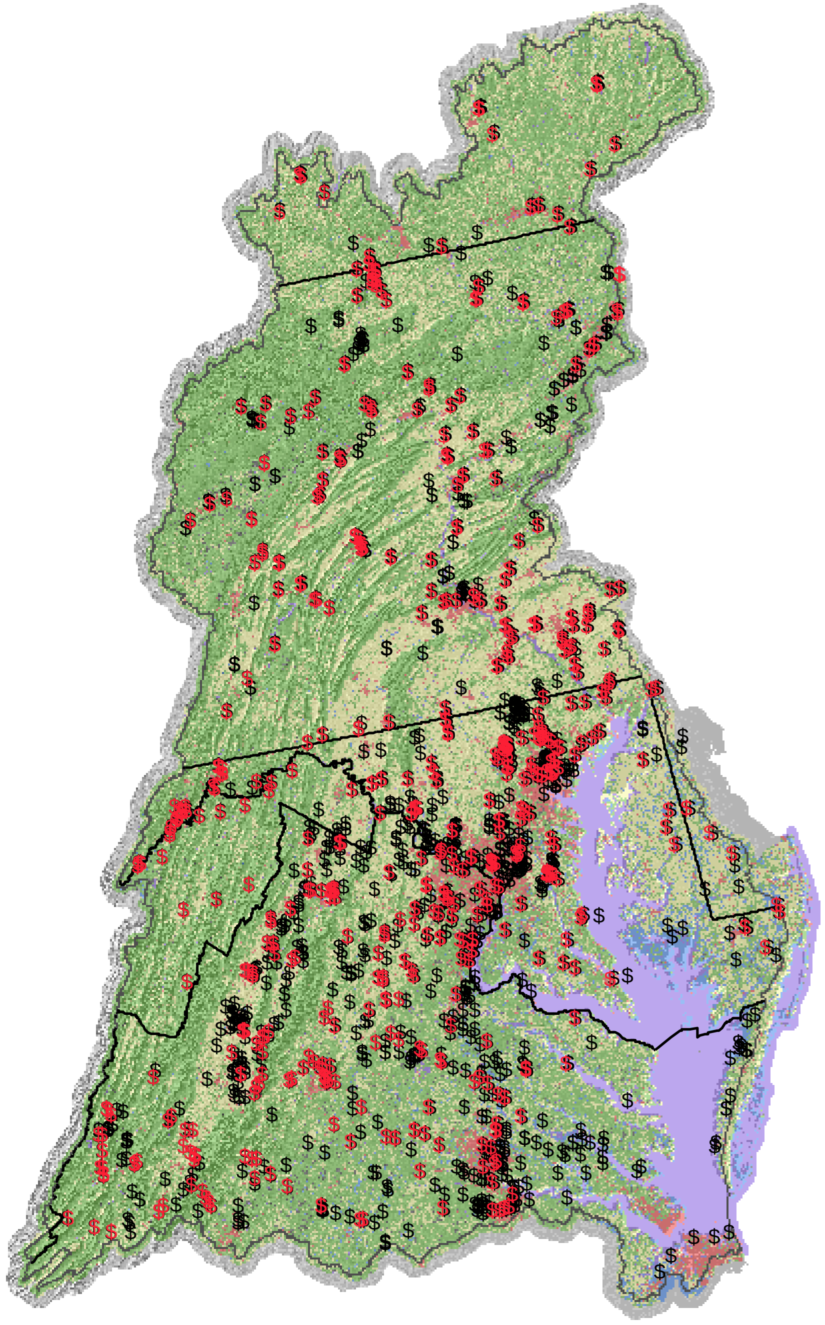
Initial Screening

- Select sites within Chesapeake Bay Drainage.
- Yields approximately 1,300 sites.

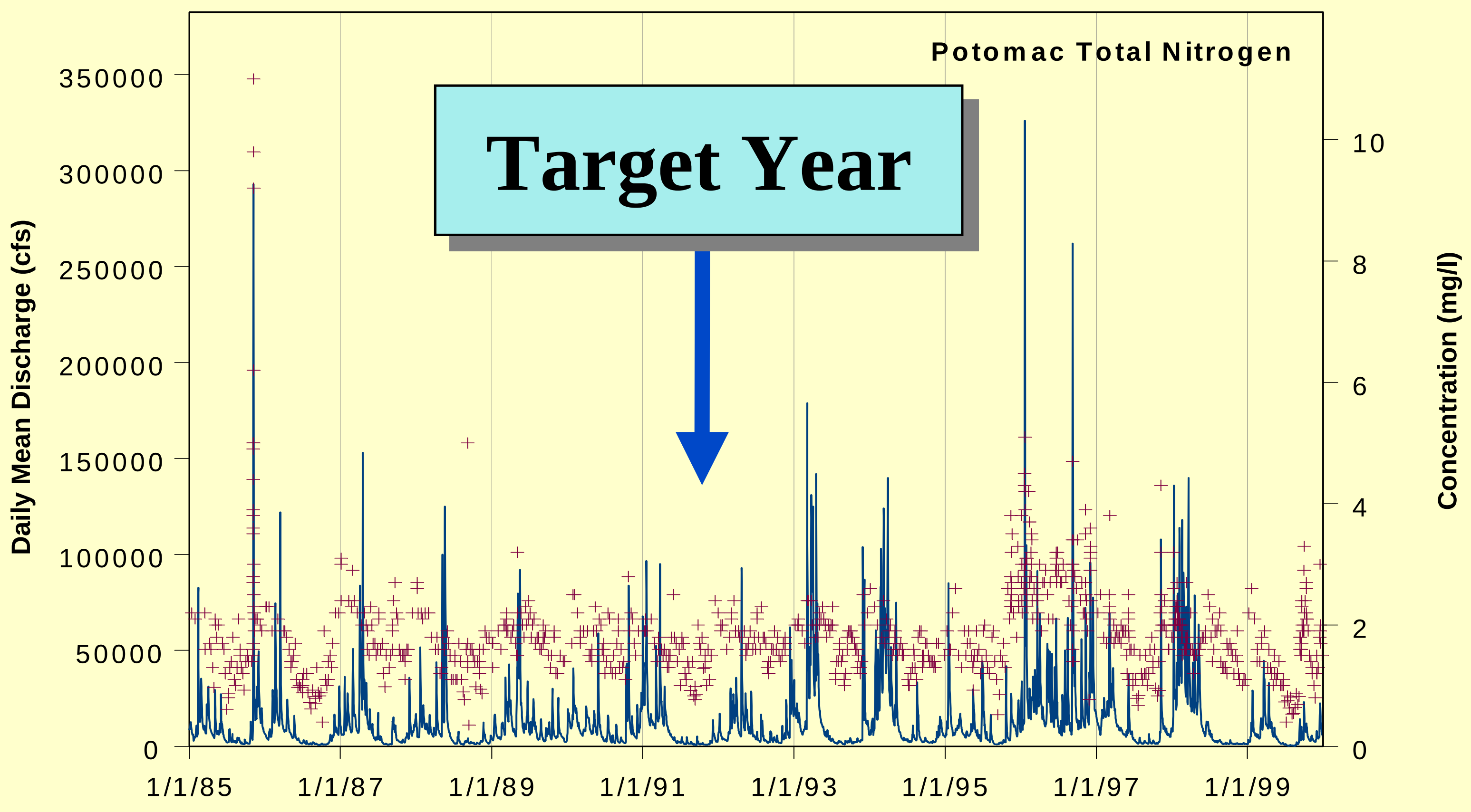


Step 2 - Identify Sites With A Minimum Number of Samples

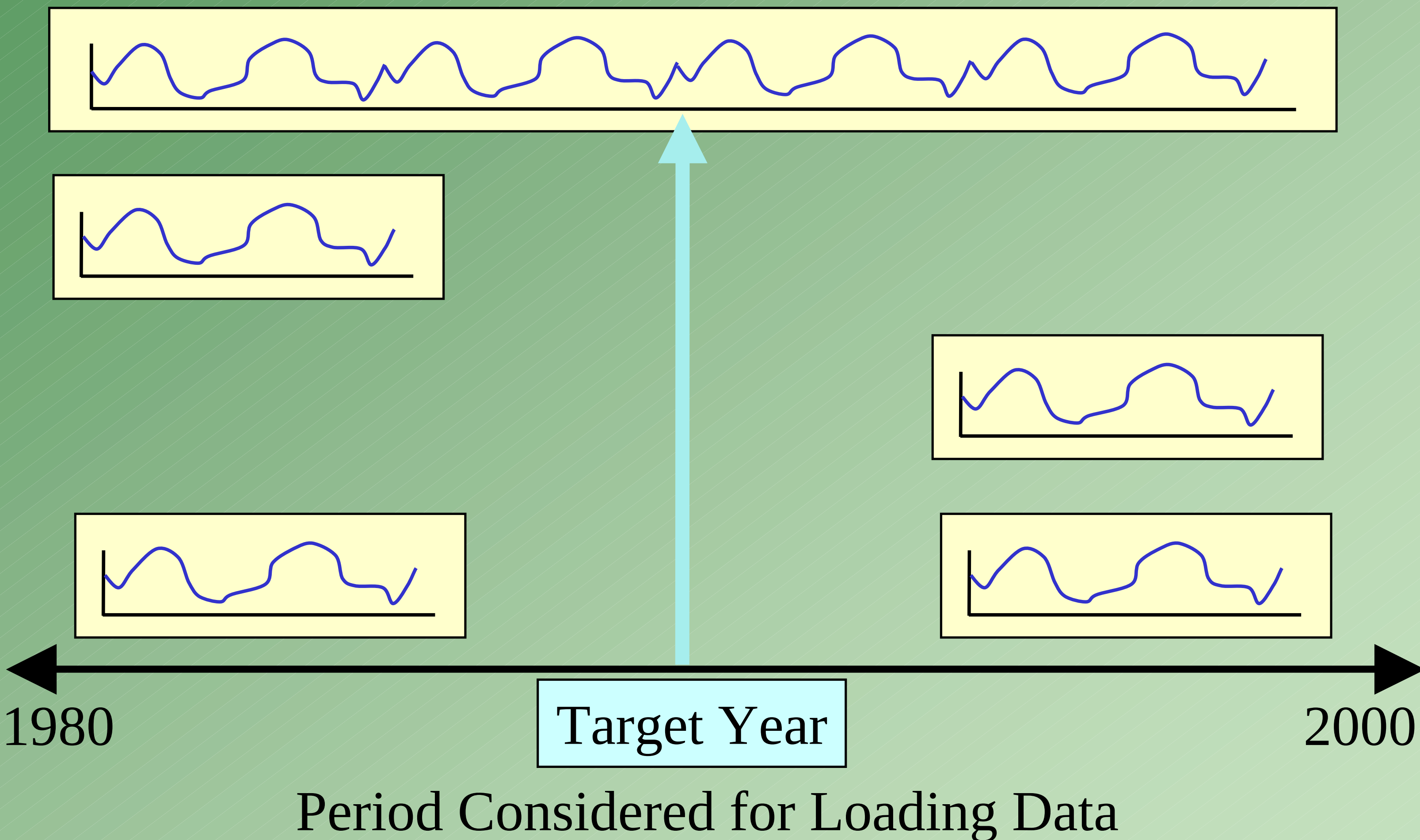
- Select only those sites that have at least 10 water-quality measurements.
- Yields 582 sites.



Step 3 - Identify Sites With An Appropriate Period of Record

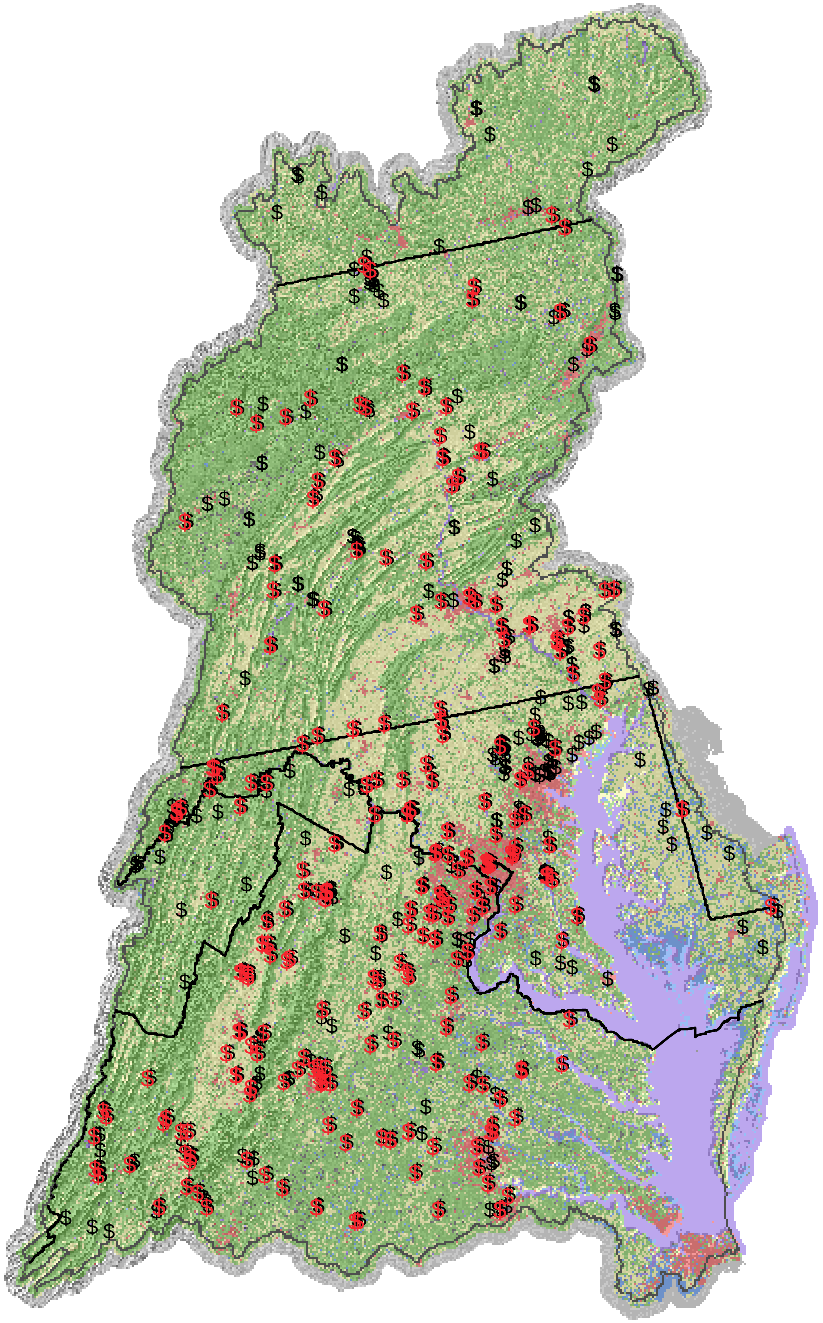


Step 3 - Identify Sites With An Appropriate Period of Record



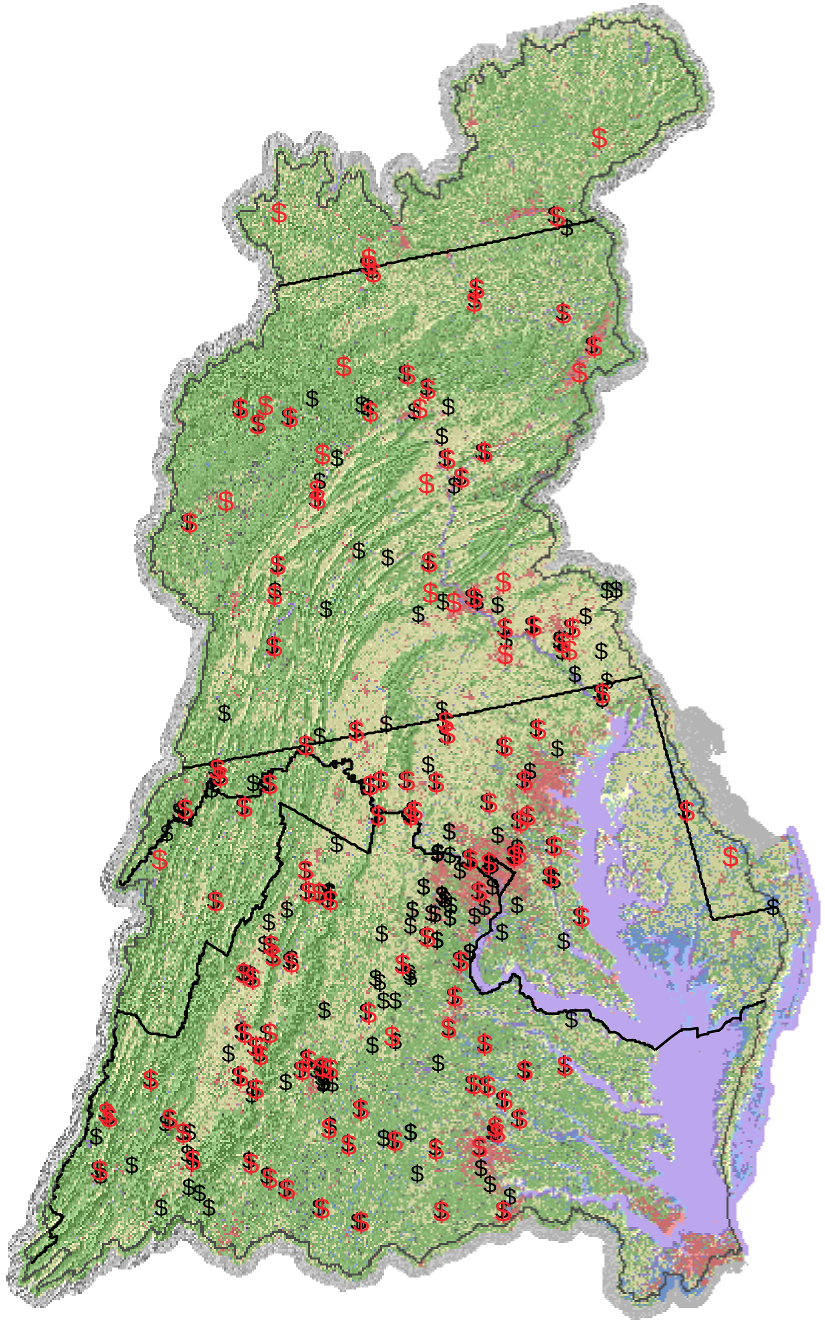
Step 3 - Identify Sites With An Appropriate Period of Record

- Select only those sites that have records within two years of target year.
- Yields 274 sites.



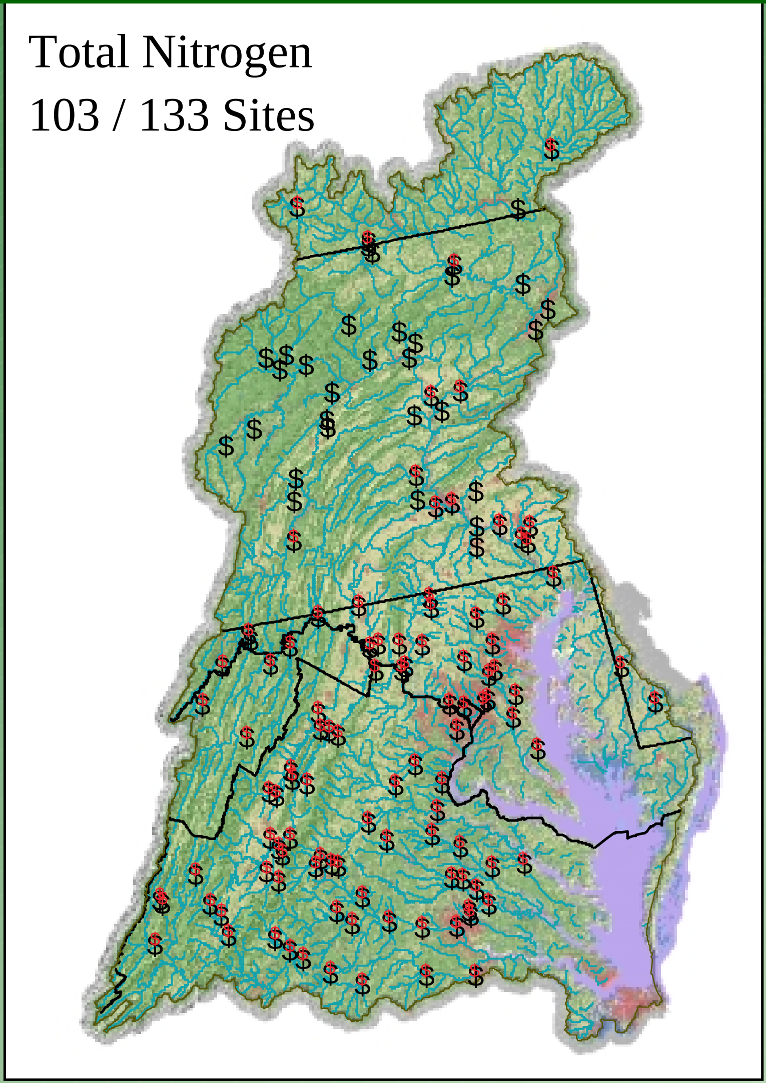
Step 4 – Associate Water-Quality Sites With Stream Gage And Network

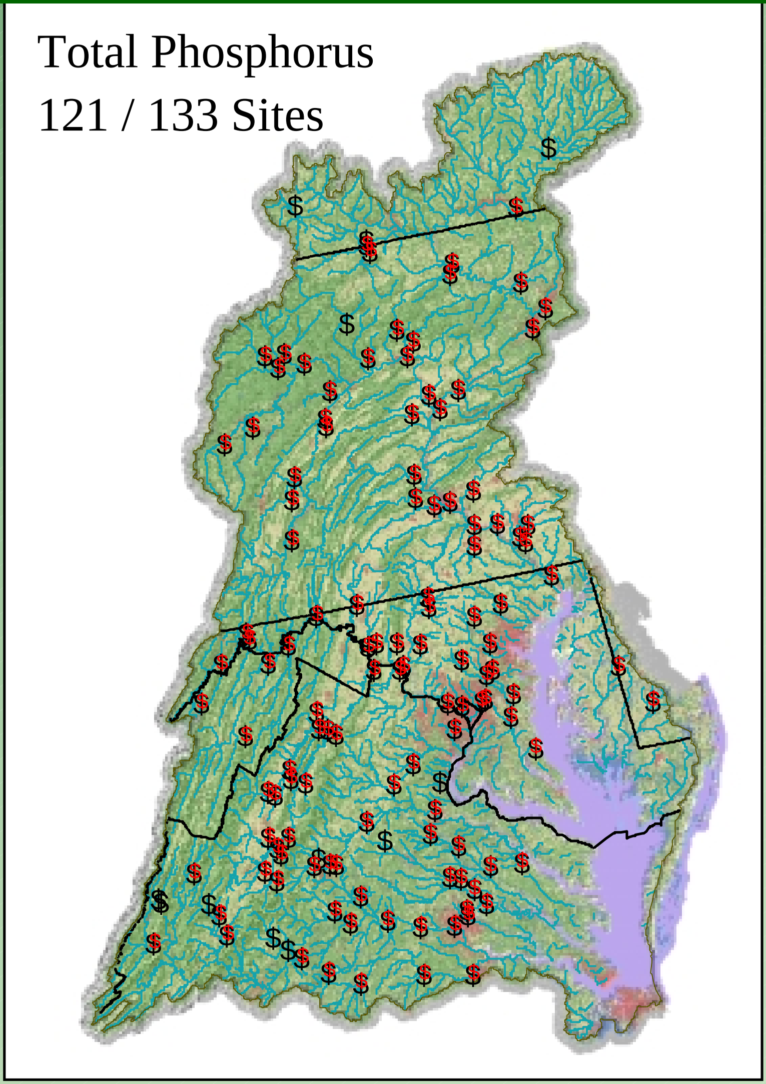
- Eliminate sites that are not co-located with stream-gages.
- Eliminate sites that are not on the stream-reach network.
- Yields 133 sites.



Chesapeake Bay SPARROW

Total Nitrogen / Total Phosphorus Load Sites





How are

the load values

calculated?

Steps for Load Calculation

1. QA/QC checks.
2. Compile water-quality and discharge data sets.
3. Set up load estimation software.

QA/QC Checks

1. Obvious outliers
2. Naming conventions among networks
3. Mis-matches in water-quality / discharge periods of record

Load Estimation Numerical Methodology

1. ESTIMATOR
2. LOADEST
3. Greg Schwarz – enhanced ESTIMATOR (FLUXMASTER)
4. Others

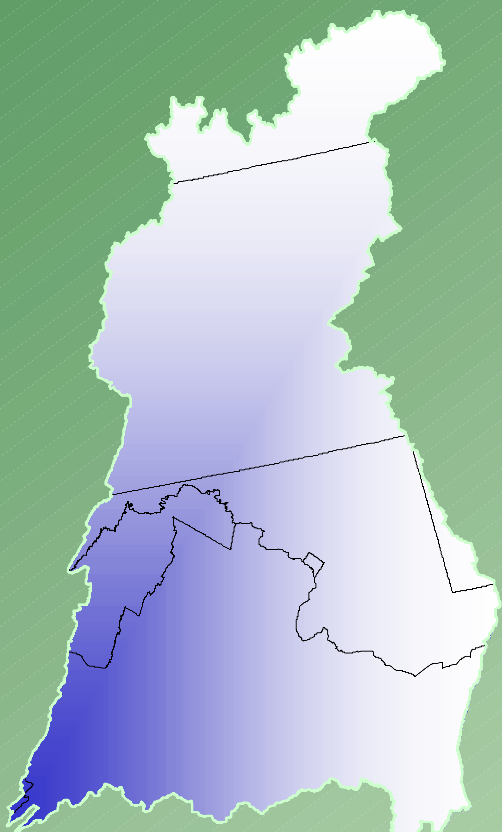
Unique to Load Estimation for SPARROW

Must correct for temporally defined spatial variations in hydrology.

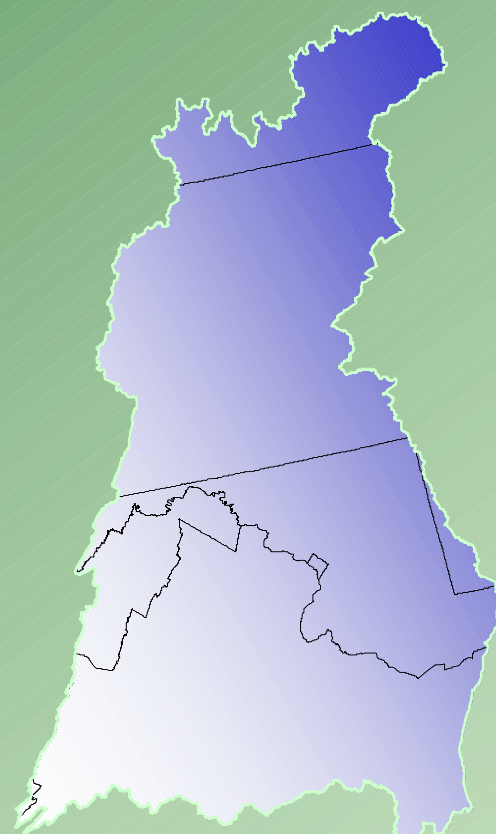
Load Estimation for SPARROW

Year to Year Spatial Variation in Rainfall

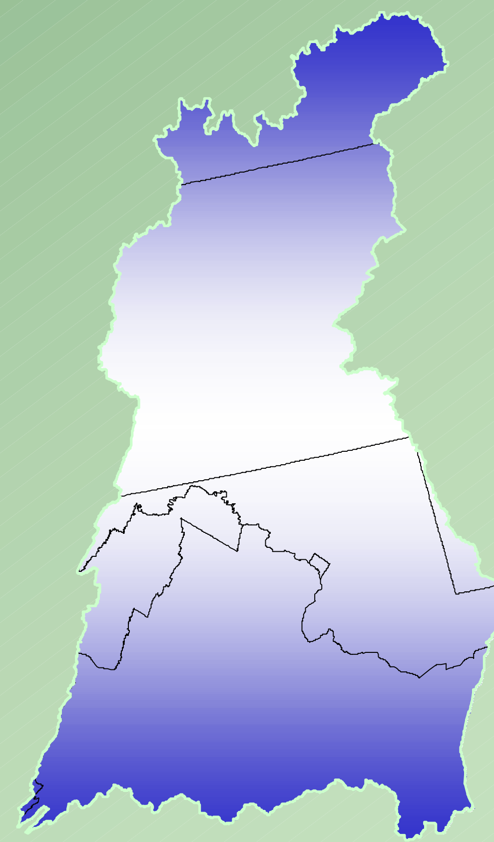
Rainfall Year 1



Rainfall Year 2



Rainfall Year 3



Approach for Compensating for Year to Year Spatial Variation in Hydrology

1. Base load estimates used for calibration on long-term average hydrology.
2. Compile discharge records for as long as possible, typically 1950 to present.

Alternative Methods for Calculating Loads Based on Long-Term Hydrology

1. Calibrate using actual discharge and concentration data – predict using synthetic annual time series of daily mean discharge calculated from long-term record.

2. Calibrate and predict using actual discharge and concentration data for as long of a period as possible – use mean annual load for SPARROW calibration.

Summary and Conclusions

1. A stream loading data base is a key component of any SPARROW model.

2. There may be limitations in the data available for developing stream-load data base.

3. Developing a stream-load data base for a SPARROW model is labor-intensive and one of the major tasks involved.

Fluxmaster: A New Model for Load Estimation

Greg Schwarz
USGS, NAWQA, Nutrient Synthesis

Capabilities

- Batch processing for all stations/constituents
- Computes long-term trend-adjusted mean load and flow at monthly, quarterly and annual frequencies
- Computes daily trend-adjusted load and flow (non-batch mode)
- All estimates include an estimate of prediction error variance
- Estimates of load and flow variability at daily, monthly, quarterly and annual frequencies

Capabilities (continued)

- Accommodates censored data and discontinuous flow or water-quality record
- Decision rules built in to modify model if record is too short
- Flexible model specification
  - Standard ESTIMATOR variables built in
  - Straightforward inclusion of other variables
  - Error term in flow model can be specified as any arbitrary ARMA process
  - Error term in load model can be first-order autoregressive

Capabilities (continued)

- Diagnostics and graphics provided
- Accommodates common flow data input formats
- Uses ESTIMATOR methods (with some modification)
- New Simulated Maximum Likelihood method for serially correlated load model
- Conditional estimates available for serially correlated load model

Program Description

- SAS program divided into 8 sections:
  - Fluxmaster_header.sas – Model specification (modified by user)
  - Main_macros.sas – Controls model execution
  - Makelst_macros.sas – Creates macro variables used to control execution
  - Setdata_macros.sas – Macros used to input flow and WQ data
  - Load_calibration_macros.sas – Macros used to calibrate the WQ and flow models
  - Load_prediction_macros.sas – Macros used to predict/de-trend/aggregate WQ and flow
  - Graph_macros.sas – Macros used to graph predictions and residuals
  - Compile_macros.sas – Macros used to compile results

Calibration

- Flow
  - Model used to trend-adjust flow data
  - Can specify any ARMA model
  - Estimation fully accommodates missing data
- Load
  - Estimation uses a close approximation of the Adjusted Maximum Likelihood Estimation (Cohn)
  - Model accommodates censored, serially correlated data.

Load Prediction

- Can impose adjustment for trend which applies to both flow and load prediction
- All predictions include estimates of standard error of prediction
- Methods for non-serially correlated loads
  - Prediction method approximates ESTIMATOR approach to correct for retransformation bias
- New methods for prediction with serially correlated loads
  - Conditional estimates (estimates conditioned on observed loads) are available using a generalization of the Simulated Maximum Likelihood method
  - Retransformation bias corrected using parametric bootstrap methods

Program Output - Calibration

- Adjusted and unadjusted maximum likelihood coefficient estimates
- Standard errors of coefficients
- Covariance matrix
- Fit statistics
- Graphs of predicted vs observed

Program Output - Prediction

- Station attributes
  - ID, lat/lon, area, county, HUC, etc.,.
- Flow Summarization
  - Trend-adjusted, long-term flow at annual, quarterly, monthly, etc., frequencies; trend-adjusted flows for individual years, months and days
  - Flow variability at daily, quarterly, monthly and annual frequencies
  - 7Q10

Program Output – Prediction (continued)

- Load Prediction (by contaminant)
  - Trend-adjusted or trend-unadjusted mean annual, quarterly or monthly load, and loads by year, quarter, month and day.
  - Standard error of predictions
  - Variance of annual, quarterly, monthly, and daily loads.
  - Graph of predicted and observed daily predictions over time
  - Concentration/flow covariance for generating time-weighted (as opposed to flow-weighted) concentrations