



# Monitoring Network Optimization

USGS Hydrologic Workshop II  
November 1, 2005

# Presentation overview

- Part 1. Data requirements
- Part 2. Statistical methods
- Part 3. Optimization method

**Case study: Great Smoky Mountains Water  
Quality Monitoring Network (GRSM)**

# Why optimize?

- **Must meet budget constraints**
- **Reallocation of funds to other monitoring efforts**
- **Determine if additional monitoring efforts are needed**
- **Reduce duplicated efforts**
- **Assessment of historical data**

# What data are available?

- Land cover
- Soils
- Vegetation
- Geology
- Watershed characteristics
- Stream information
- Historical water quality data (DLF)
- Biological monitoring data (DLF)
- Streamflow

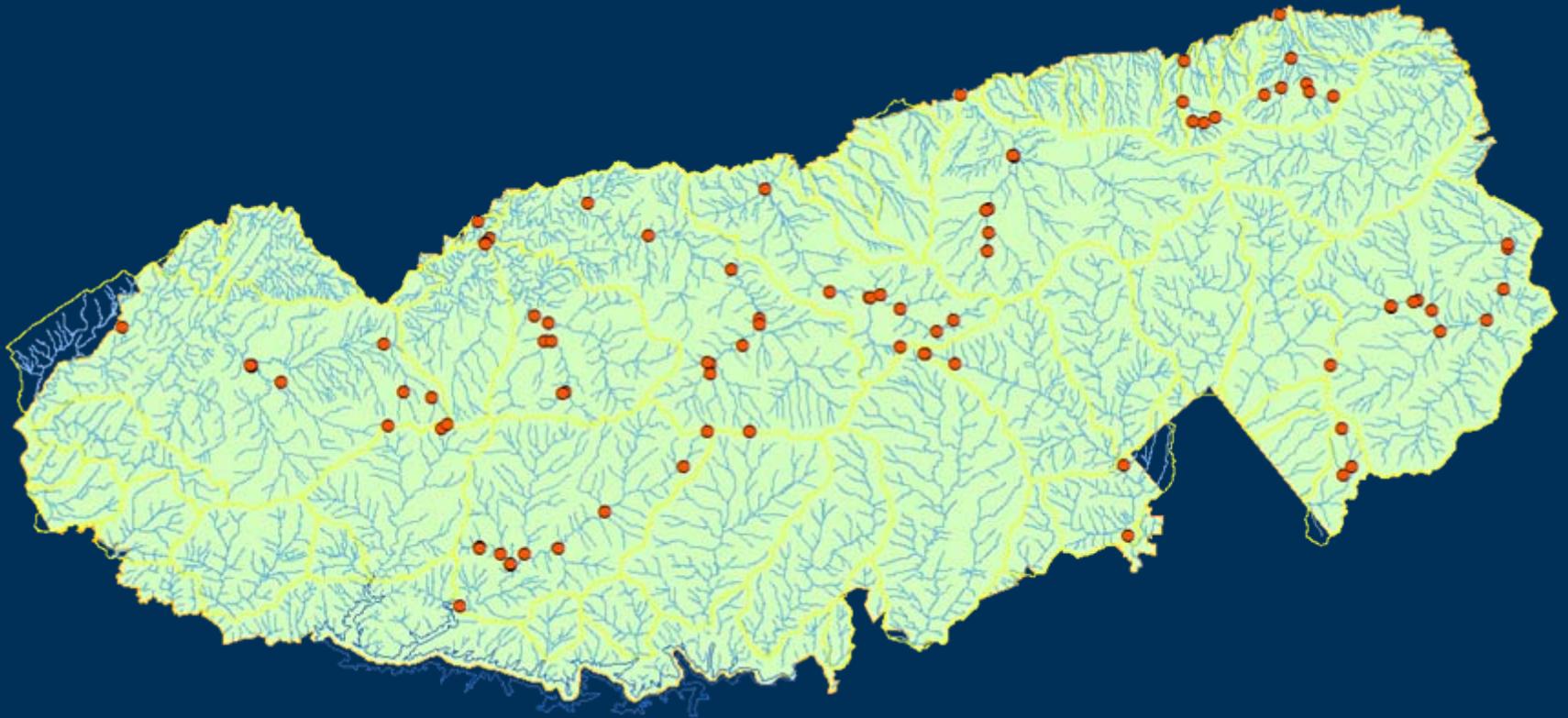


# GRSM Data

- **Water Quality – pH, ANC, conductivity, nitrate, sulfate, chloride, sodium, and potassium**
  - Quarterly grab samples
  - Period from 1996-2001
- **Watershed characteristics**
  - Geology
  - Stream morphology
  - Vegetation
- **Collocation information**
  - Benthic study
  - Brook trout study
- **Costs**
  - Laboratory
  - Site access



# Great Smoky Mountains Network



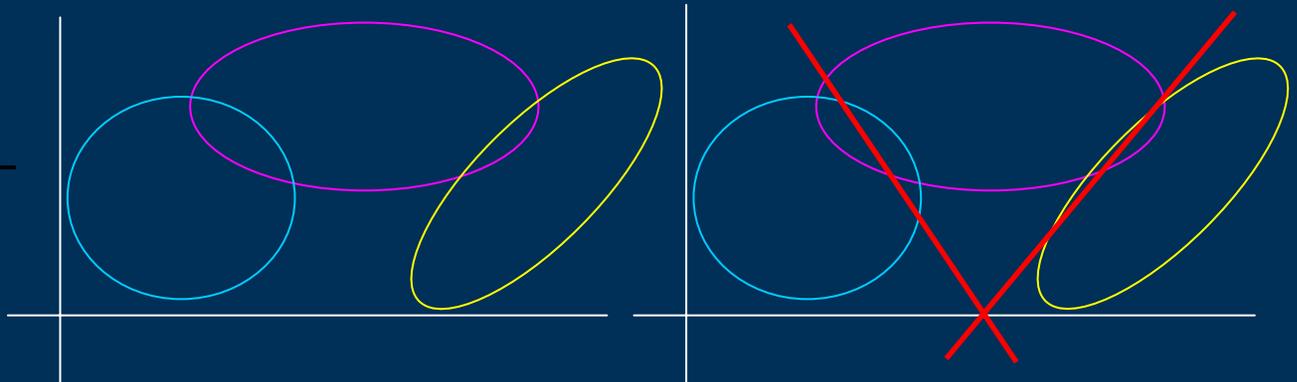
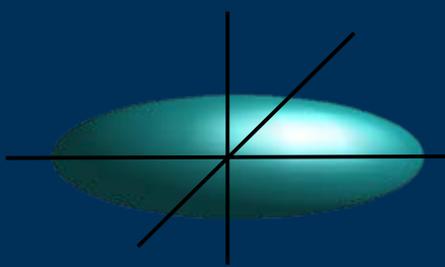
# The statistics toolbox

- Data screening (descriptive statistics)
- Principal components analysis (PCA)
- Cluster analysis (CA)
- Discriminant analysis (DA)
- Robust PCA



# Multivariate statistical methods

- Principal components analysis – reduce the dimensionality of the data
- Cluster analysis – group similar sampling sites together the use cluster centroid distance as a measure of variability explained within each cluster
- Discriminant analysis – validation test for the clusters that were formed using cross-validation method

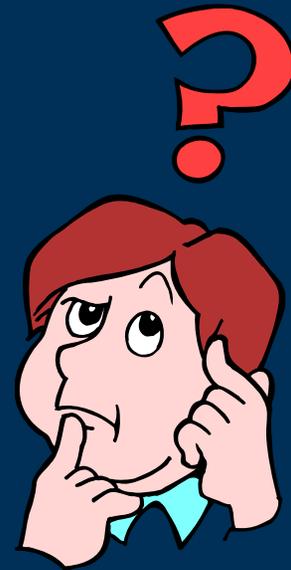


# Optimization needs

- Mechanism for assigning benefits to sampling sites
- Objective function to score and compare different network designs
- Knowledge of any special circumstances that may need to be addressed in the benefit assignment or programming phase

# Special considerations (GRSM)

- Small clusters should remain intact – only clusters with large memberships should be targeted
- Ensure that all water quality, geology, morphology, and vegetation clusters are represented in the final network



# Determining costs and benefits (GRSM)

- Total network cost of \$69,200
  - \$19,200 per year for access and sampling time (640 man-hours X \$30/man-hour)
  - \$50,000 per year for laboratory, technical, administration, and overhead (approx. \$602 per site/year)
- Total Benefit = 1.2 X \$69,200 = \$83,040
  - Basis: Benefit should outweigh cost
  - Basis: 20 percent return is a modest expectation
  - $BENEFIT_{TOTAL} = \$83,040$
- Cost of p-sites:

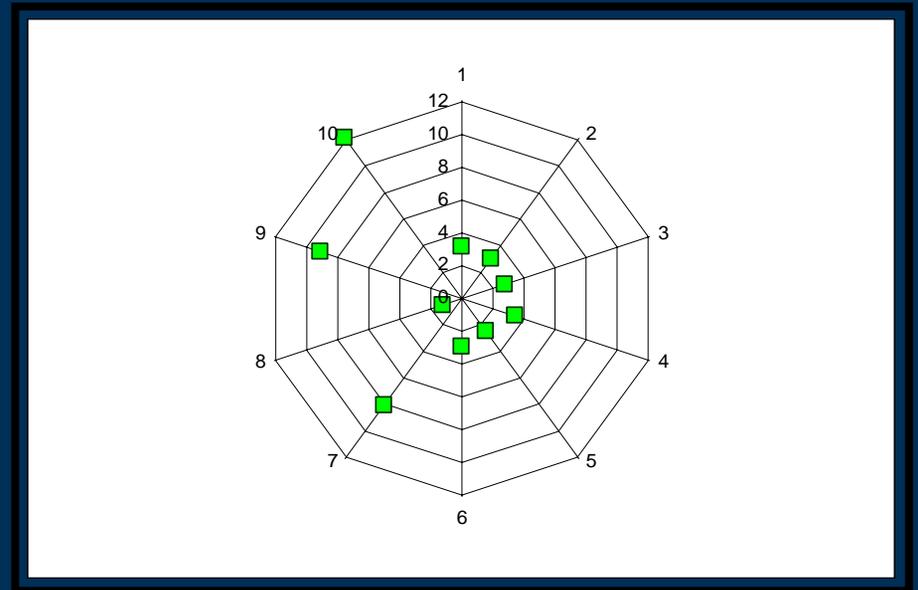
$$COST_p = \sum_p LABCOST + \sum_p ACCESS$$



# Apportioning for site benefits (GRSM)

- Sites ranked using distance from centroid
- Ranks are then summed across categories - one score for each  $\Psi_i = \omega_1 W_i + \omega_2 G_i + \omega_3 M_i + \omega_4 V_i + \omega_5 C_i$
- All scores are then summed for apportionment total,  $\Psi_{TOTAL}$

$$BENEFIT_i = \frac{\Psi_i}{\Psi_{TOTAL}} \times BENEFIT_{TOTAL}$$



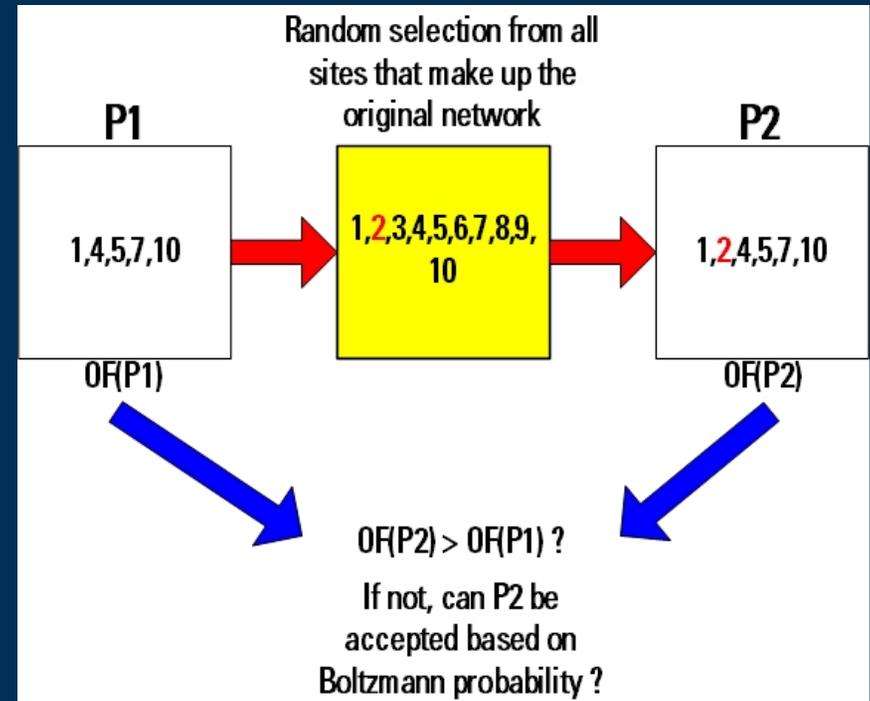
# Optimization using simulated annealing

- Heuristic method based on the thermodynamics of heating a body to a temperature such that all bonds have been broken between molecules
- Controlled cooling is then applied such that the molecules can arrange themselves to a minimal energy state
- Simulated annealing escapes local minima/maxima
- Maximize the objective function

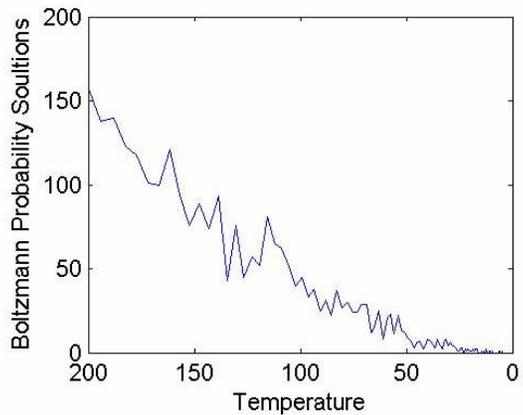
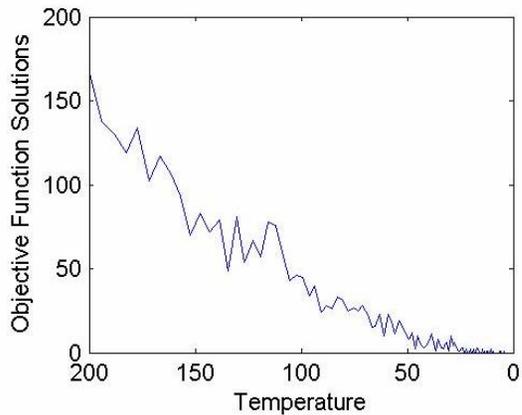
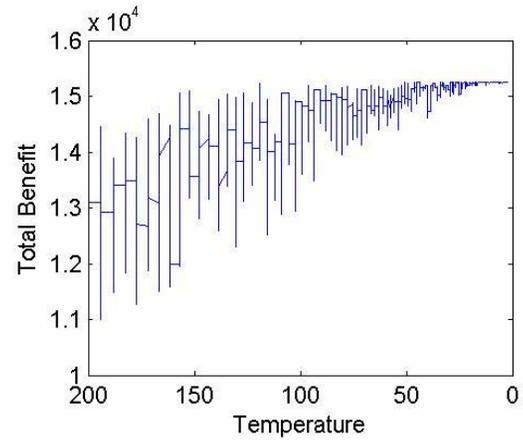
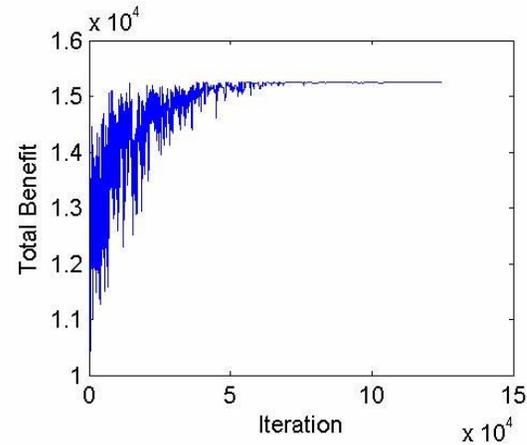
$$NETBENEFIT_p = \sum_p BENEFIT - \sum_p COST$$

# Basics of Simulated Annealing

- Start with a network (P1)
- Randomly choose one site
  - If IN the P1 network, test OF for removal (P2)
  - If OUT of the P1 network, test OF for addition (P2)
- IF  $OF(P2) < OF(P1)$ , Can P2 still be accepted using the Boltzmann probability?
  - As temp gets lower it becomes harder for a network to be accepted using the Boltzmann probability
- Continues until the termination loop is satisfied



# Objective function tracking

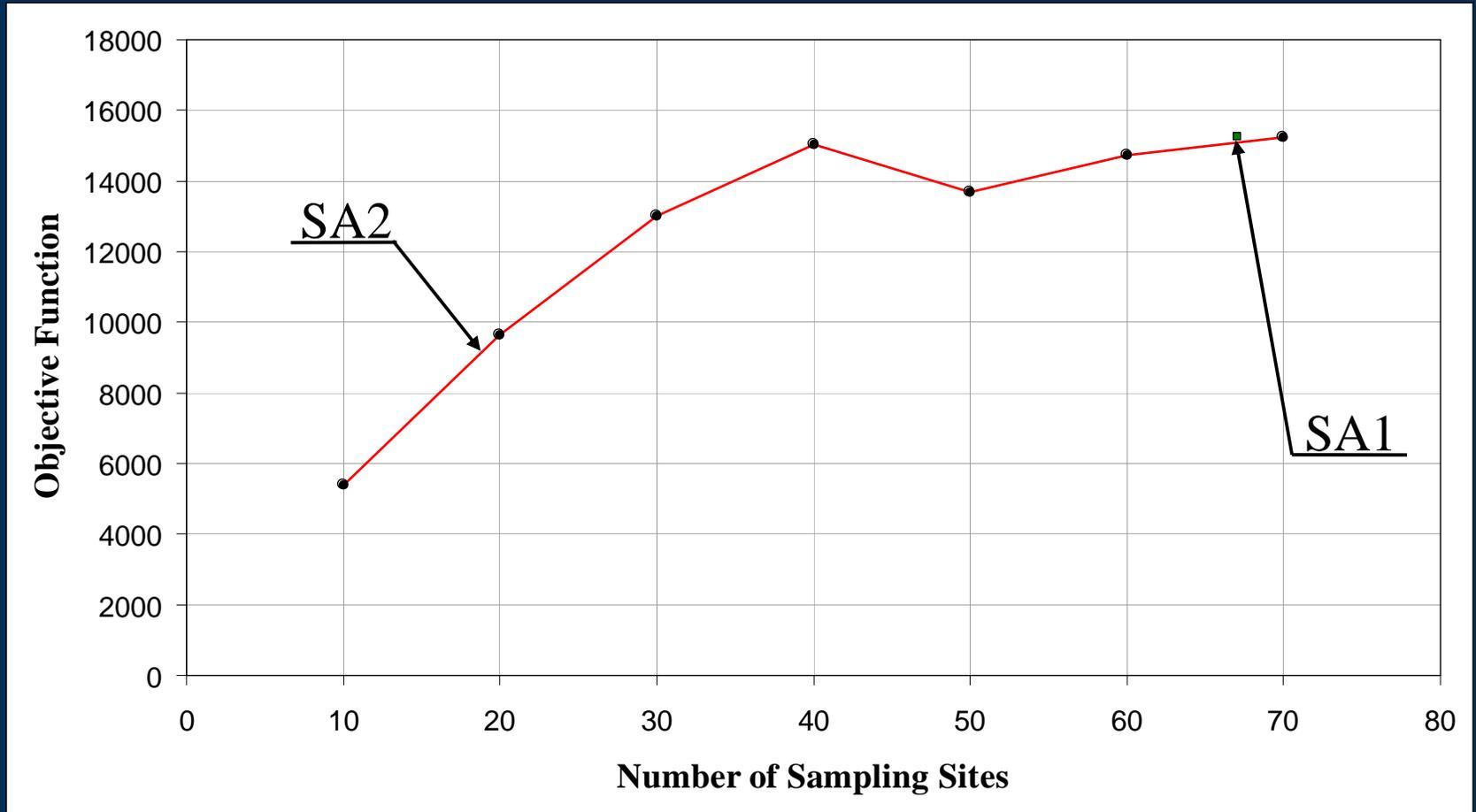


# Network optimization

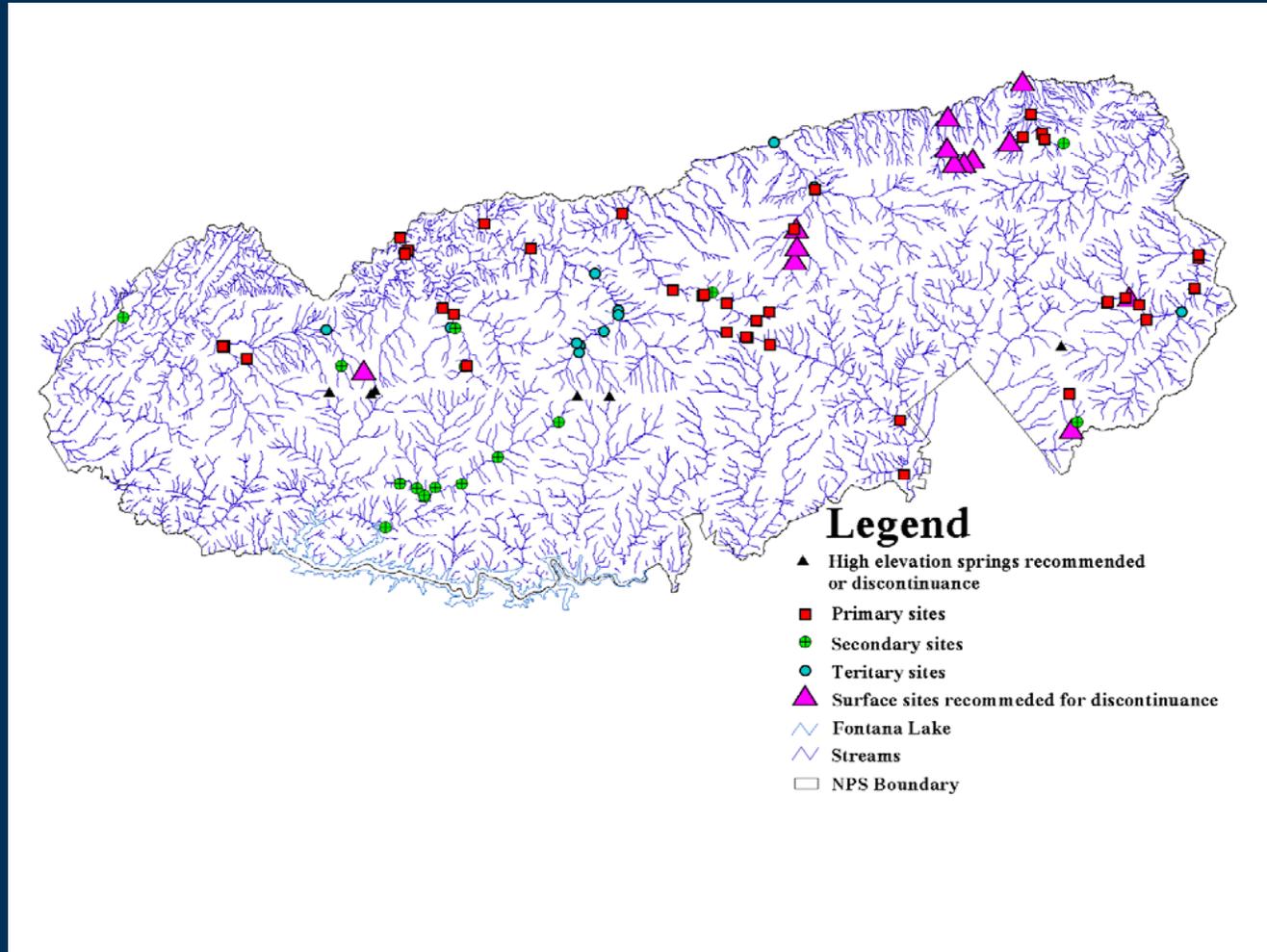
Simulated annealing program written for two cases

- **First case (SA1) – Simulated annealing is performed on the network to determine the overall optimum network configuration**
- **Second case (SA2) – user-specified (n) number of sites desired in the final network. The optimized network will contain exactly n-sites**
  - **Provides a validation for SA1 results**
  - **Provides a logical format for considering other sampling sites to be retained or discontinued**

# SA2 results – $n$ best sites



# Redesigned Network (GRSM)



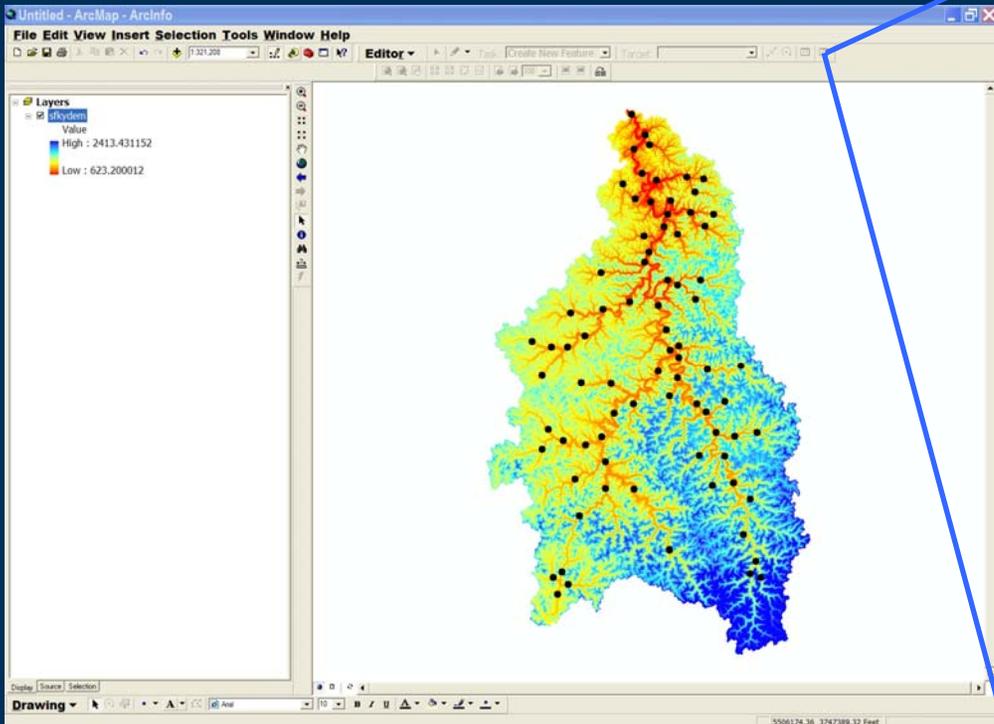
# Sensitivity analysis

- Vary weighting factors
- Test individual categories
- Vary the cost multiplier for benefits

# Temporal assessment

- Resampling of data at different sampling frequencies
- Compare trend test results at different sampling frequencies to the trend from the original high-frequency data (MIN)
  - Boxplot analysis
  - Mann-Kendall test for trend
  - Time series regression
- Identify frequency where dependency becomes an issue using the autocorrelation function (MAX)
- Confidence level to reliably detect a trend within a certain number of years

# ArcMap Tool Application



### Network Optimization Prototype

Input Files    Output Files

**Individual Weightings**

Biological Index	<input type="text" value="1.00"/>
Ecological Index	<input type="text" value="1.00"/>
Water Quality	<input type="text" value="1.00"/>
Morphology	<input type="text" value="1.00"/>
Geology	<input type="text" value="1.00"/>
Vegetation	<input type="text" value="1.00"/>

**Cost Multiplier**

**Type of Optimization**

Full Network

User-specified Network Size

Number of sites desired in final network

**Questions?**