A Random Walk Through Sampling Designs:

The Ups and Downs of Probabilistic Monitoring

Jeroen Gerritsen

NEAEB 2008

Top reasons to be statistician

- Deviation is considered normal
- Statisticians feel complete and sufficient
- Statisticians do it both discretely and continuously
- Statisticians can legally comment on someone's posterior distribution
- Statisticians are right 95% of the time
- Statisticians are honestly significantly different
- No one else wants the job

Overview

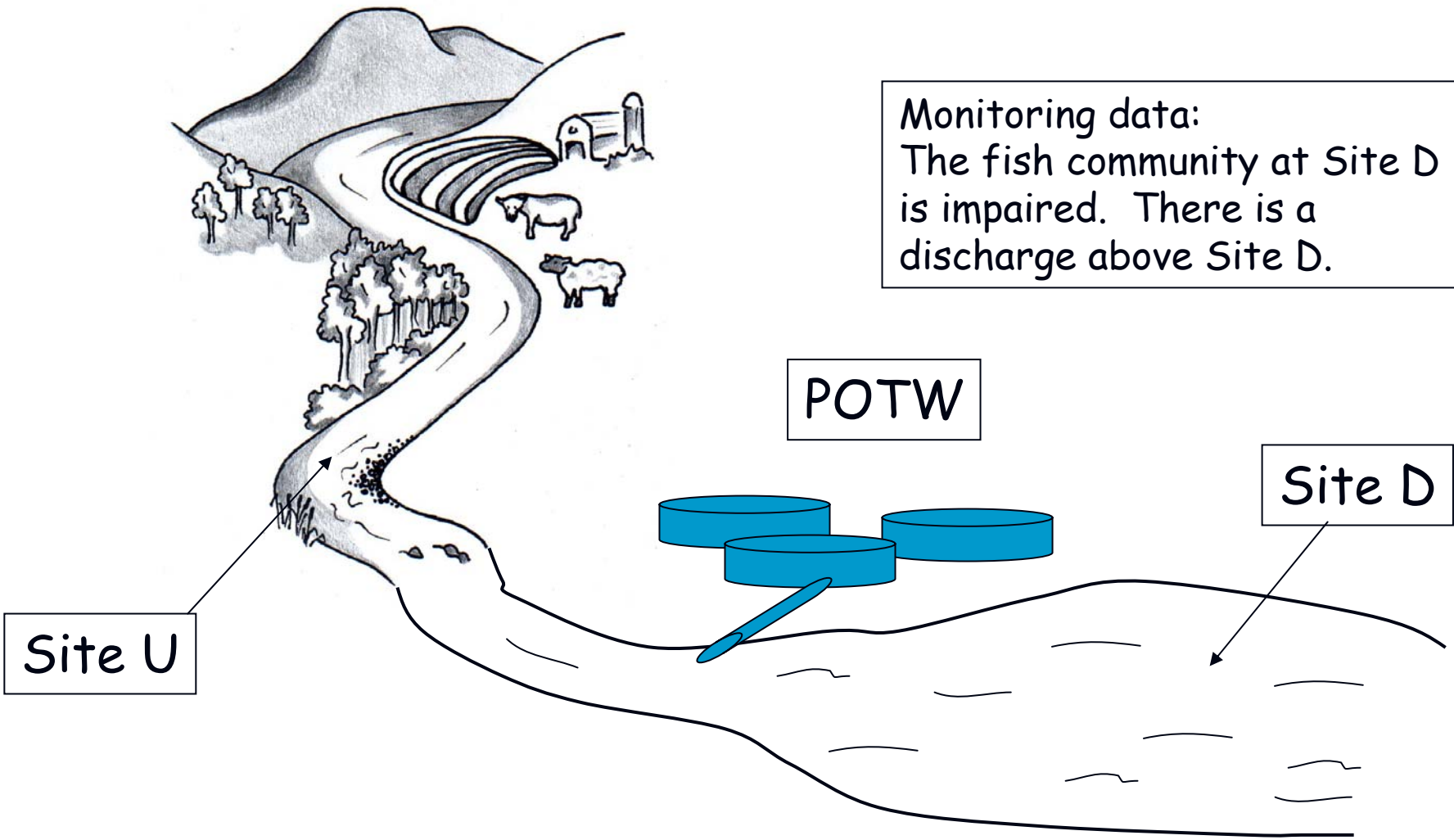
- A cautionary tale
- Inference
- Types of Studies
- Design as a toolbox

Why Do We Need Statistics?

- Describe something
- Evaluate hypothesis
- Assist in management

} Decision-making

A cautionary tale: Scenario (1)



Is wastewater causing impairment?

- Site D has impaired fish community relative to Site U
- Wastewater discharge between U and D
- BOD in discharge
- Hypothesis: excess BOD depresses DO at Site D, causing fish impairment
- Have DO measurements at U and D

Data: DO
(measured early AM)

Month	Site U	Site D	Difference
March	10.2	10.6	-0.4
April	11.3	10.4	0.9
May	9.5	7.2	2.3
June	8.7	6.3	2.4
July	7.5	5.5	2.0
August	7.4	6.9	0.5
September	9.1	8.5	0.6
October	9.7	8.5	1.2
November	10.7	12.0	-1.3
Mean	9.34	8.43	0.91

t-test

- Paired t
- Mean difference = 0.91 mg/L, s = 1.239
- H_0: $d = 0$; H_a: $d <> 0$
- $t = d/s_d$ =0.91/(1.239/sqrt(9)) = 2.99
- $t_{crit\ 8,0.05}$ = 2.306

2.99 > 2.306

TEST IS SIGNIFICANT

Scenario 2
(slightly different data)

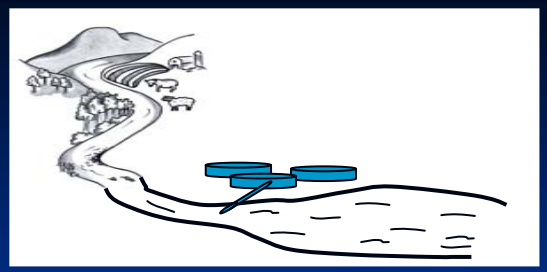
Month	Site U	Site D	Difference
June	8.7	6.9	1.8
July	7.5	2.5	5.0
August	7.4	3.3	4.1
Mean	7.87	4.23	3.63

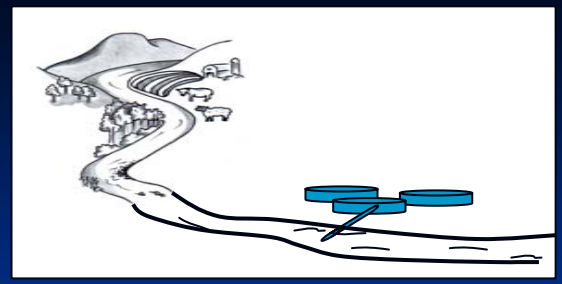
t-test (scenario 2)

- Paired t
- Mean difference = 3.63 mg/L, *s* = 1.65
- H_0: $d = 0$; H_a: $d <> 0$
- $t = d/s_d = 3.63/(1.65/\text{sqrt}(3)) = 3.81$
- $t_{\text{crit } 2,0.05} = 4.303$

3.81 < 4.303

TEST IS NOT SIGNIFICANT





Scenario 1

- DO measured upstream and downstream 9 months
- Upstream mean DO = 9.34; downstream = 8.43
- Difference **IS** significant at $p < 0.05$

Scenario 2

- DO measured upstream and downstream 3 months
- Upstream mean DO = 7.87; downstream = 4.23
- Difference **IS NOT** significant at $p < 0.05$

Which is a stronger case for DO causing impairment?

Scenario 1: What can we infer?

- Low DO caused degraded fish community?
- Discharge caused degraded fish community?
- No, ONLY that the DO at Site D is lower than at Site U

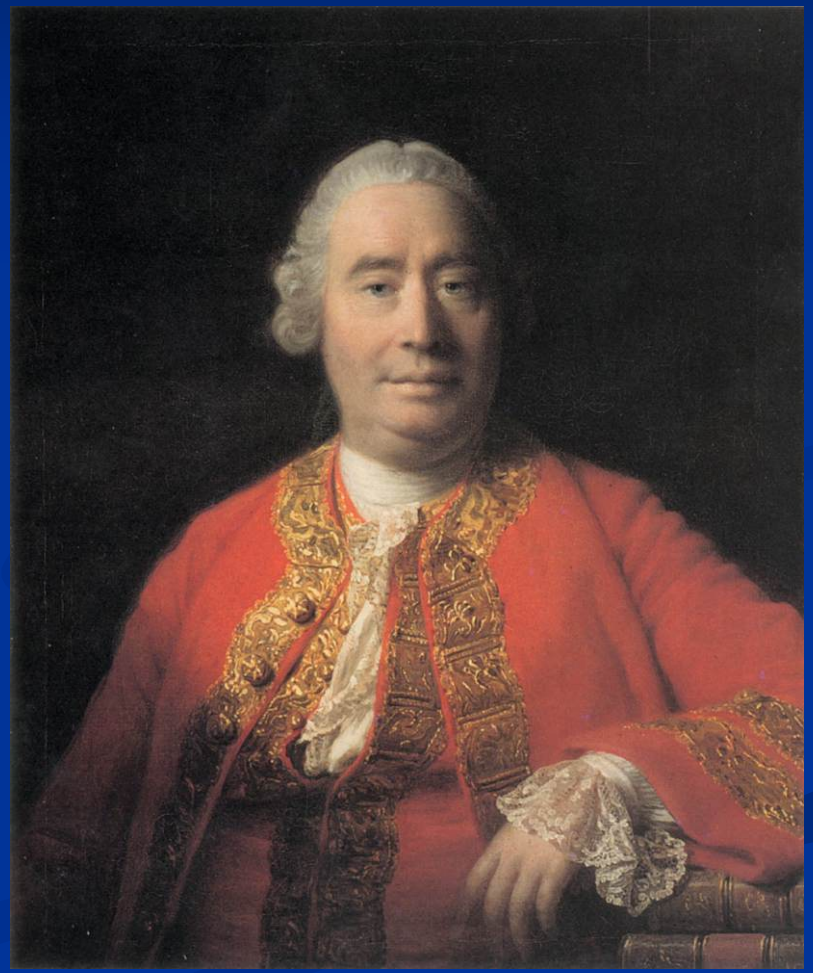
Scenario 2: What can we infer?

- Low DO caused degraded fish community?
- Discharge caused degraded fish community?
- NO, From classical statistics, nothing

BUT, what is the biological significance of DO of 2.5 mg/L?

Inference

- Inductive reasoning (Hume)
  - From repeated observations, we make generalizations about the state of the world
- Statistical inference
  - From repeated observations (a sample) on a class of things, we infer a property of the class (a population)
- We can fool ourselves!



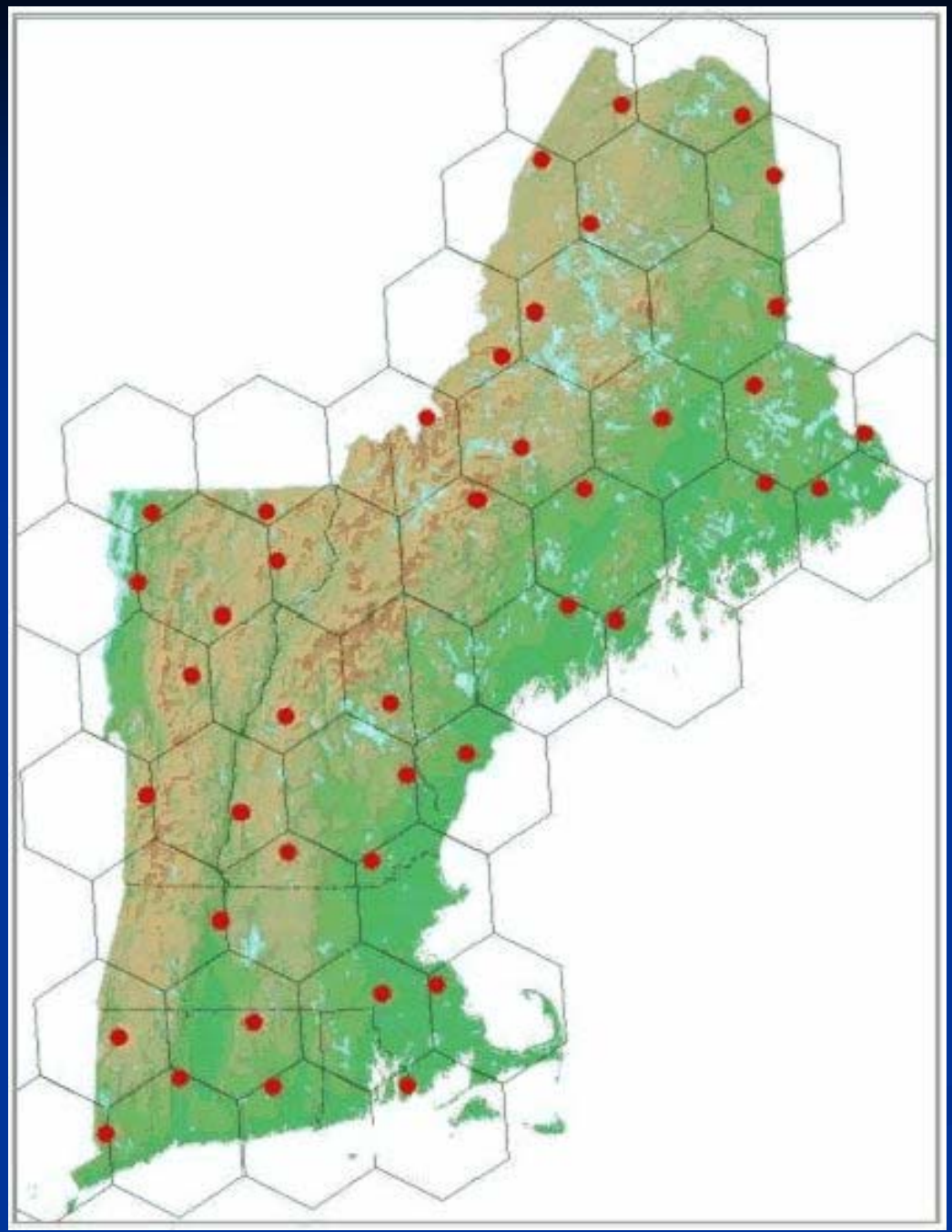
David Hume (1711-1776)

Inference

- Inference is limited to the class of things from which we sampled
- And, how we structured the design around our question
- Our sample must be representative of the class
- How do we get a representative sample?
  - Census (measure every member of class)
  - Random sample
  - Prior knowledge
  - Combination of prior knowledge and random sample

Random Sample

- Question: describe population
- Simple random: every member has equal probability of being sampled
  - Waterbodies: create "list frame"; list of all members of population
  - Can have "bad luck"
- Multi-stage
  - Systematic Random
  - Cluster



Systematic random

NEWS probabalistic sampling design.

Question: What is condition of streams in New England

Stage 1: Select hex

Stage 2: Select site from list frame of NHD stream miles in hex

Inference: streams of New England

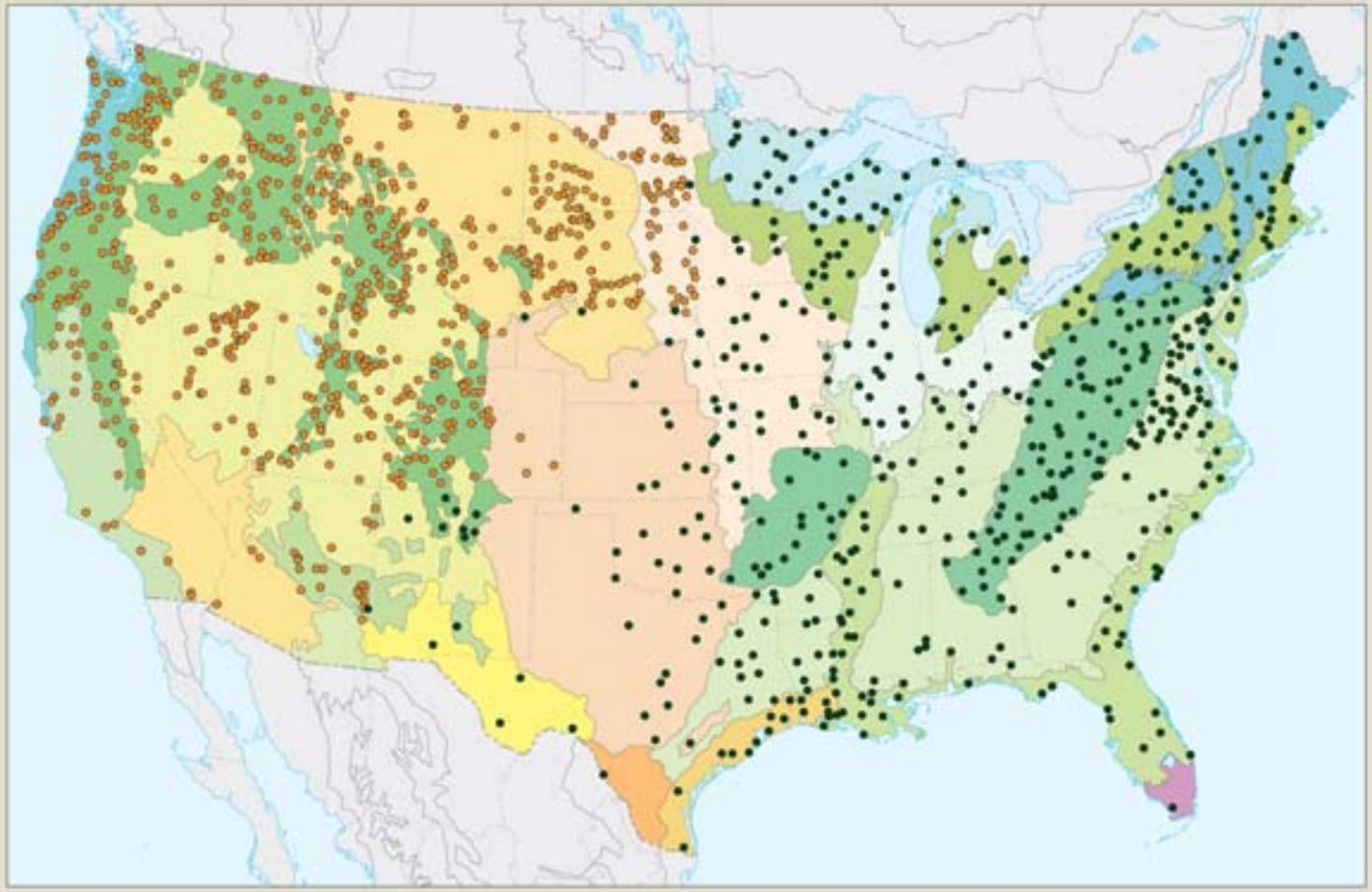
Why the emphasis on probability sampling?

- Late 1980's: What is status of nation's waters? Getting better?
- EPA could not answer
  - 305(b) reports worthless
    - States sampled where they felt like it
    - Criteria meant different things in different states
- EMAP, REMAP were result
- Then: WSA, NLA, Large Rivers, etc.

Stratification

- WSA: stratified on stream order
  - EPA Region (10)
  - WSA Aggregated Ecoregion (9)
  - Within each EPA region – Ecoregion combination, construct list frame of NHD streams for each order 1 -5. Selection probability adjusted among stream orders
- Ensured representative sample, known uncertainty for EPA region, ecoregion, stream order
- What are the questions, what are the inferences?

WSA Sites


5.2 Mixed Wood Shield
5.3 Atlantic Highlands
6.2 Western Cordillera
7.1 Marine West Coast Forest
8.1 Mixed Wood Plains
8.2 Central Plains
8.3 Southeastern Plains
8.4 Ozark, Ouachita-Appalachian Forests
8.5 Mississippi Alluvial / Southeast Coastal Plains
9.2 Temperate Prairies
9.3 West-Central Semi-Arid Prairies
9.4 South-Central Semi-Arid Prairies
9.5 Texas-Louisiana Coastal Plain
9.6 Tamaulipas-Texas Semi-Arid Plain
10.1 Western Interior Basins and Ranges
10.2 Sonoran and Mohave Deserts
10.4 Chihuahuan Desert
11.1 Mediterranean California
12.1 Western Sierra Madre Piedmont
13.1 Upper Gila Mountains
15.4 Everglades

Questions

- What is the condition of the Nation's streams?
  - State?
  - Ecoregion?
  - EPA region?
- What is the response of aquatic biota to stressors?
- Has this river improved since permit limits were tightened?
- Why is this river impaired (what is the cause)?
- What will be the effect of climate change on our rivers?
- More?

Types of Studies

- Manipulated experiments
- Prospective and retospective studies
- Sample surveys
- Pure observational studies

Experiments

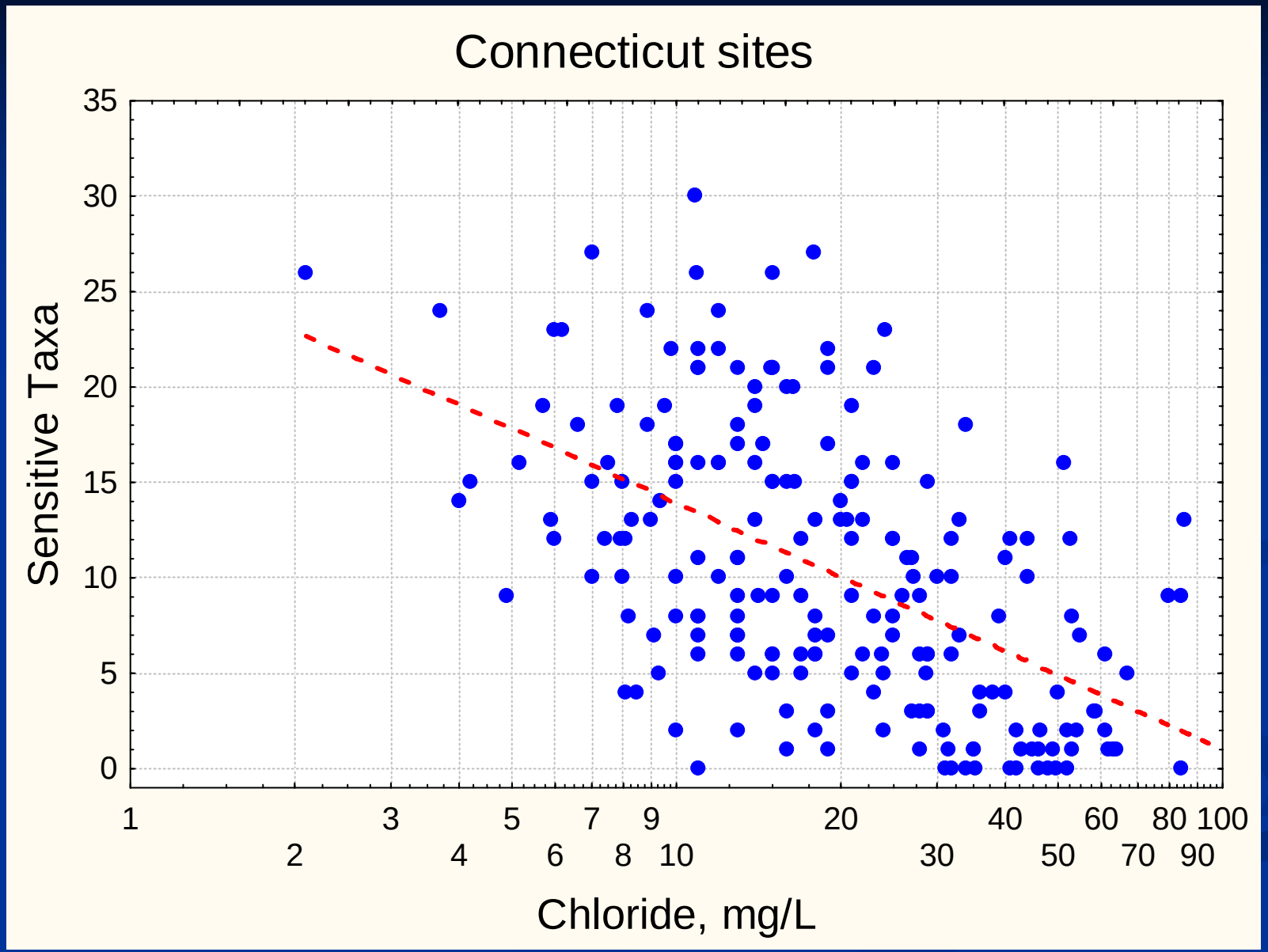
- Questions: Usually examine cause, e.g.
  - Does P cause algal blooms
  - Is Al toxic
- The "Gold Standard", but not often available
- Set up and control of system
- Scientific conclusions come from the logic and design of the experiment
- Inference may be limited, but if randomized and repeated, can generalize to cause and effect

Prospective and retrospective studies ("natural experiments")

- Questions: most often on relationships between variables we measure, e.g.,
  - Relation of organic loading to community composition
- Random assignment is beyond our control; we assume nature has randomized for us.
- Pseudoreplication may be a problem
- Optimize range of explanatory variables.
- Inference: associations

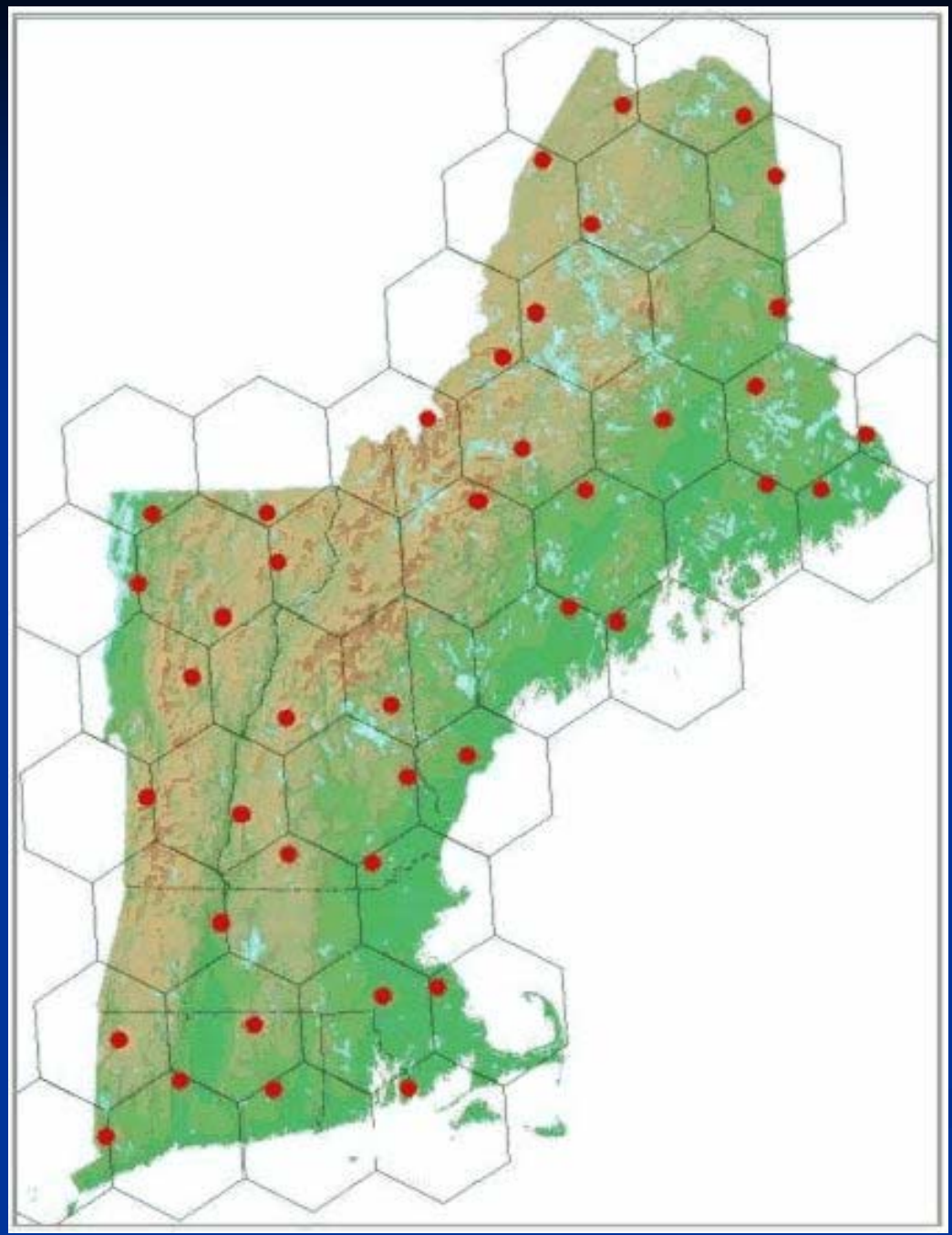
Questions

- What is the condition of the Nation's streams?
  - State?
  - Ecoregion?
  - EPA region?
- What is the response of aquatic biota to stressors?
- Has this river improved since permit limits were tightened?
- Why is this river impaired (what is the cause)?
- What will be the effect of climate change on our rivers?
- More?


Connecticut sites
Sensitive Taxa
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
20
30
40
50
60
70
80
90
100
Chloride, mg/L

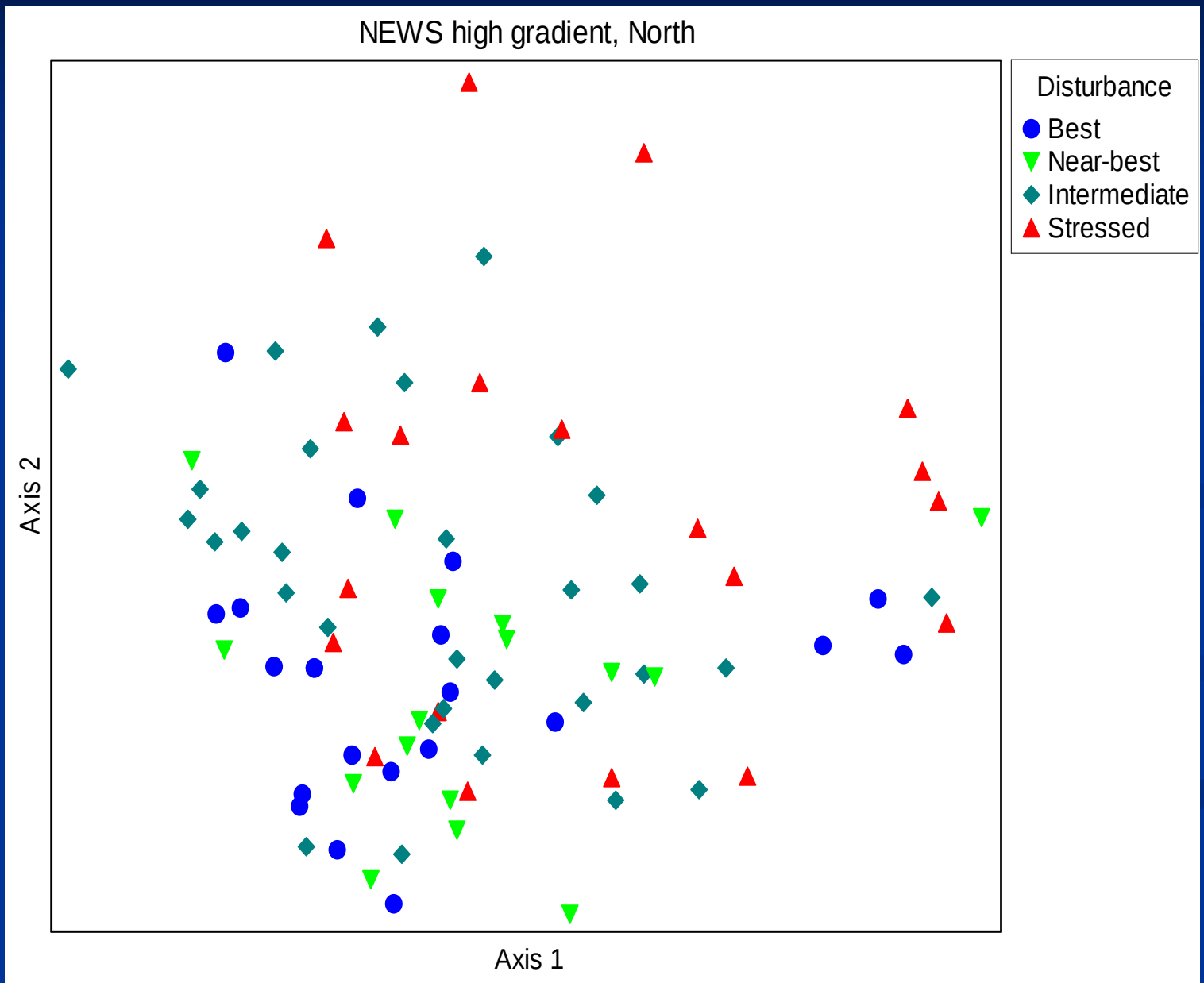
Sample Surveys

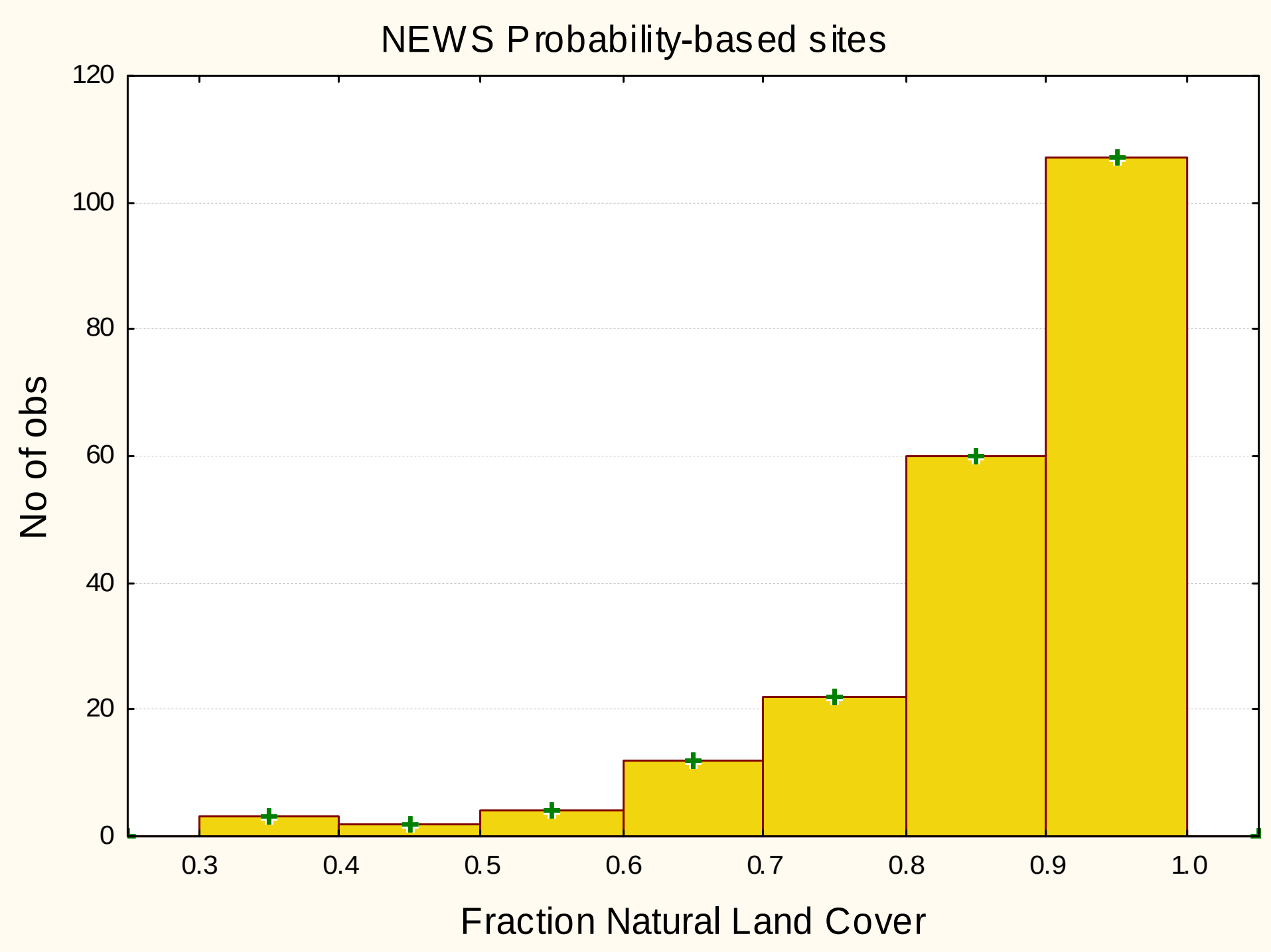
- Question: Descriptions of populations and differences among populations
  - Status
  - Trends
- Probability-based sample from defined statistical population(s)
- Inference: generalizable to the population, depend on representative
- Predictive associations may be problematic
  - Regression, other models

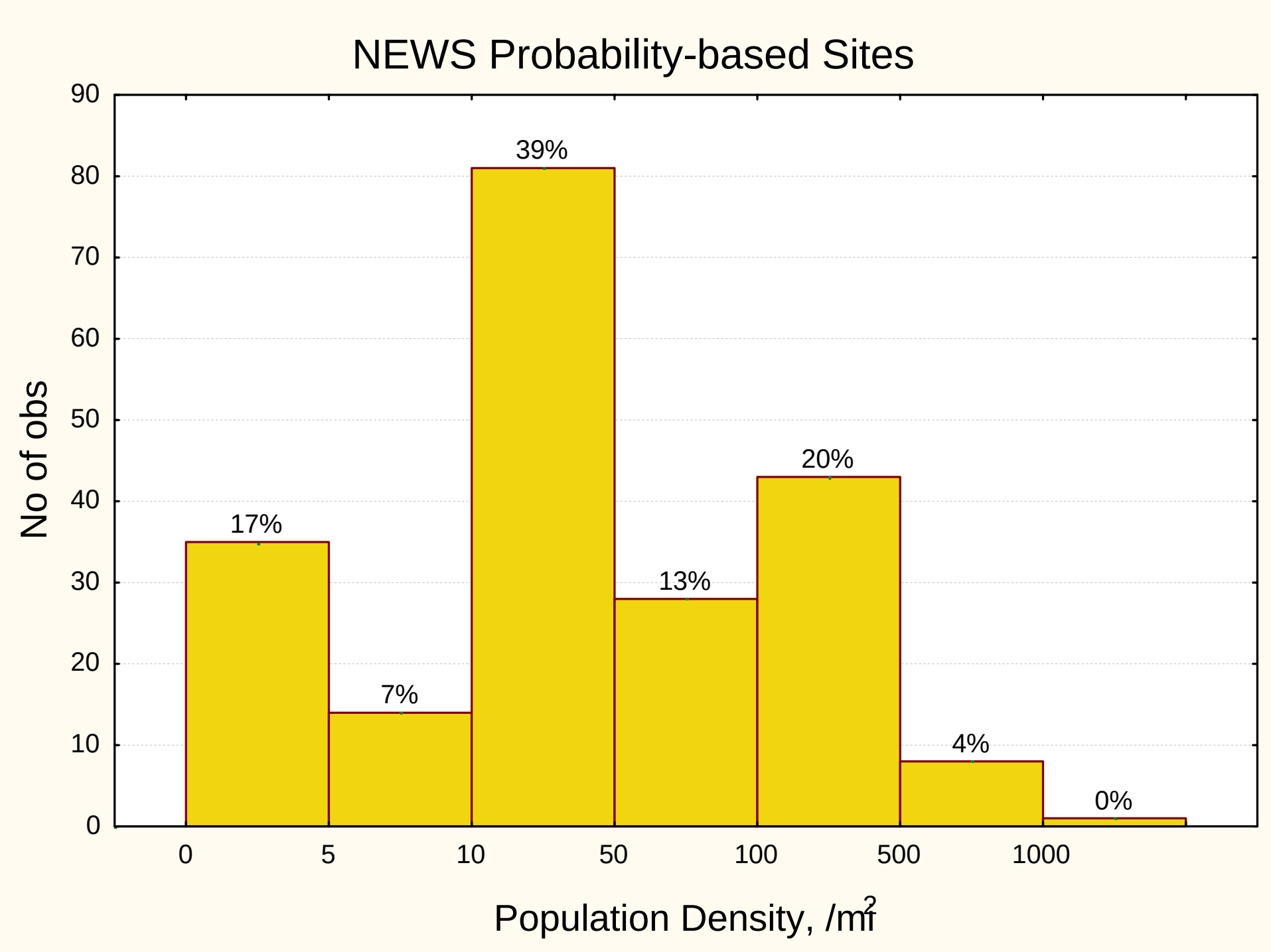


NEWS probabalistic sampling design (Systematic – random). Additional hexagonal overlays were used to select sites within states.

Problem?


NEWS high gradient, North
Disturbance
Best
Near-best
Intermediate
Stressed
Axis 2
Axis 1


NEWS Probability-based sites
120
100
80
60
40
20
0
No of obs
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Fraction Natural Land Cover


NEWS Probability-based Sites
No of obs
90
80
70
60
50
40
30
20
10
0
17%
7%
39%
13%
20%
4%
0%
0
5
10
50
100
500
1000
Population Density, /m2

Model development

- Bivariate normal distribution with linear relationship between X and Y:

  $Y = 4 - 0.67X + \varepsilon$

  $X = N(3,1)$ Normal, with mean 3 and s.d. = 1

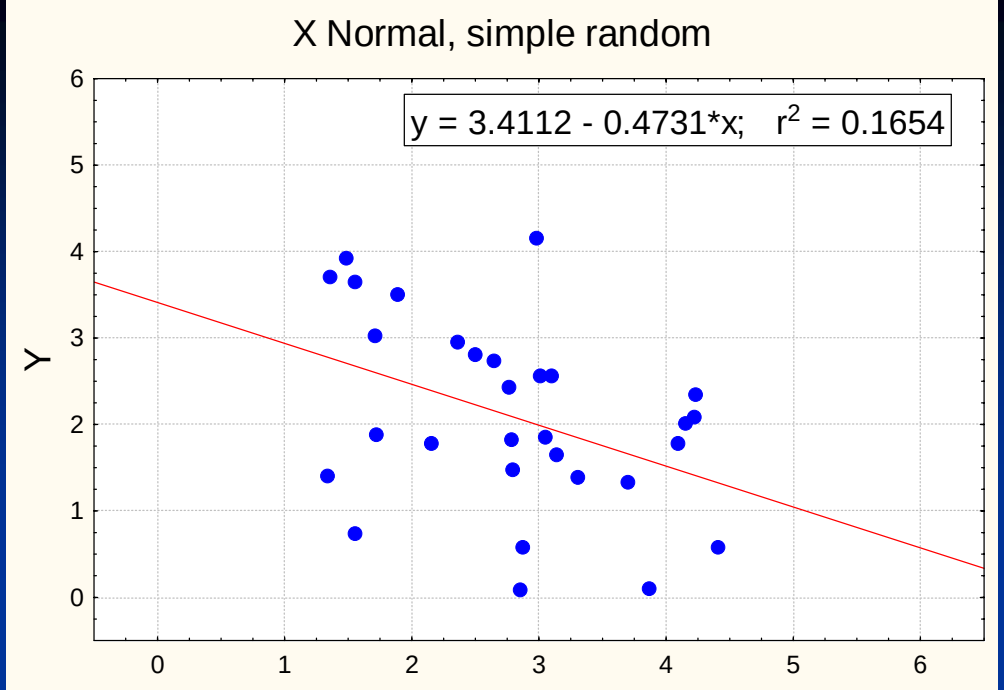
  $\varepsilon = N(0,1)$ Normal, with mean 0 and s.d. = 1

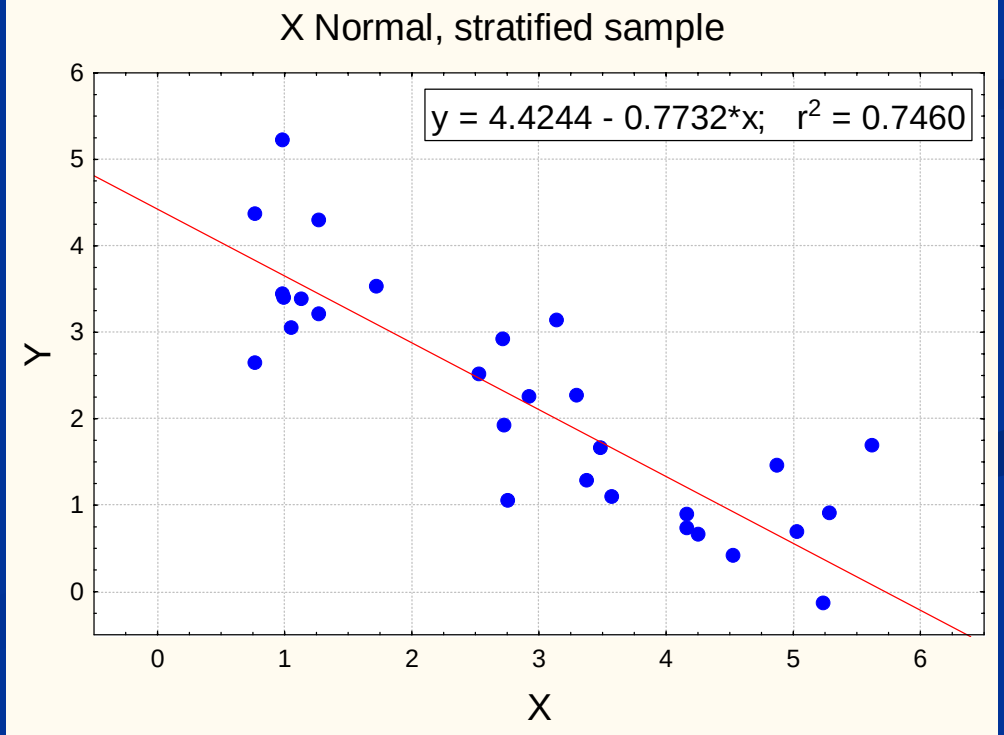
- If we randomly sample from this distribution, how accurately can we estimate the linear relationship (regression)

X is normally distributed

Simple random

Sample extreme X values





Effect of distributions

- Unstratified sample increases risk of poor model
  - Unstratified
    - $r^2 = 0.07 – 0.68$
    - 20% of regression models had $r^2 < 0.2$
  - Stratified
    - $r^2 = 0.32 – 0.80$
    - 0% of regression models had $r^2 < 0.2$
    - 7% of regression models had $r^2 < 0.4$
- What was question and inference space of NEWS?

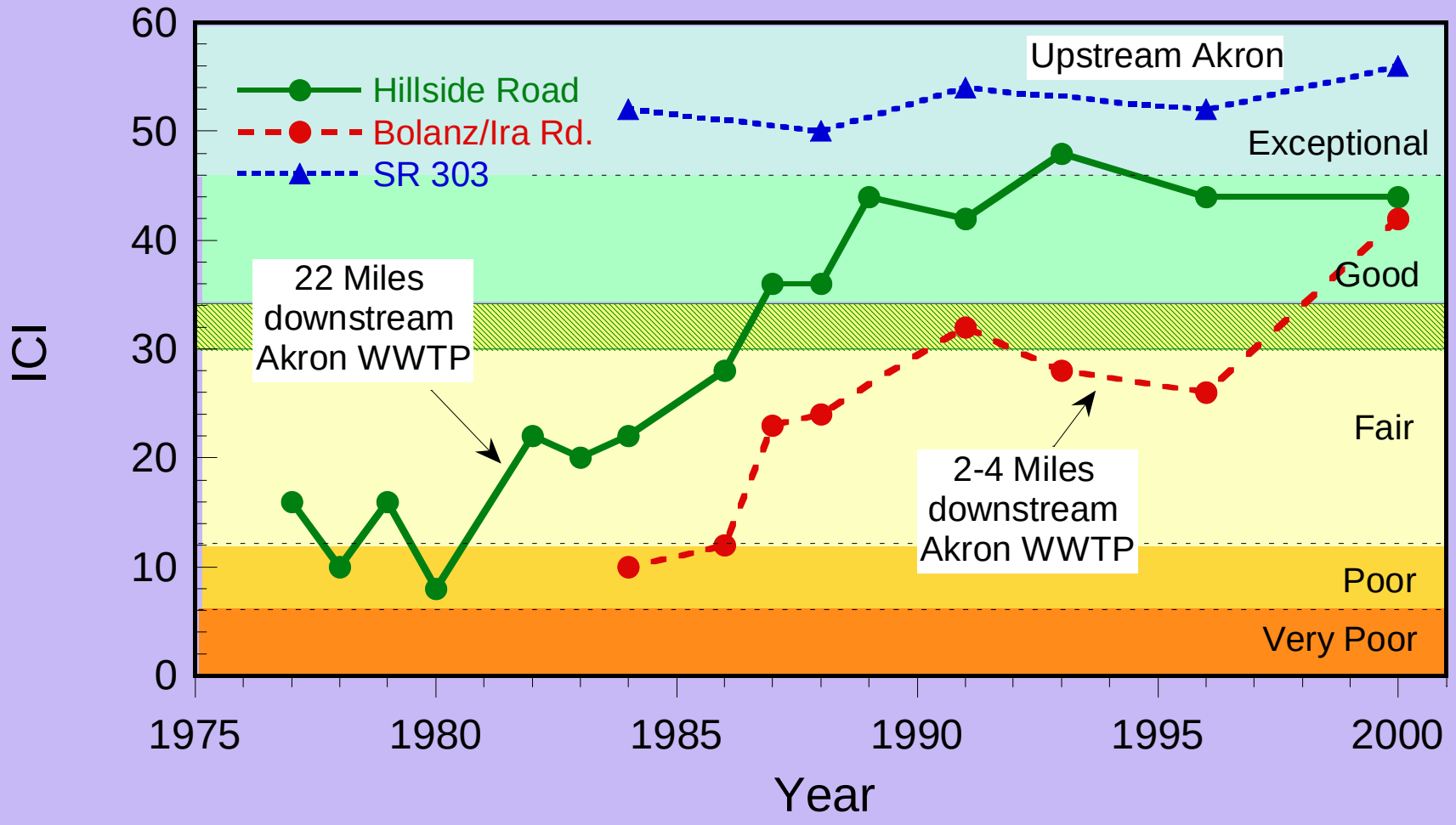
Questions

- What is the condition of the Nation's streams?
  - State?
  - Ecoregion?
  - EPA region?
- What is the response of aquatic biota to stressors?
- Has this river improved since permit limits were tightened?
- Why is this river impaired (what is the cause)?
- What will be the effect of climate change on our rivers?
- More?

Longitudinal studies

- Sites followed through time
- Why?
  - Effectiveness of management: NPDES, BMP, watershed activities
  - Sites faced with future development pressure
  - Climate change – these could be probability selected initially

Longitudinal Monitoring
3 Sites on Cuyahoga River, Ohio

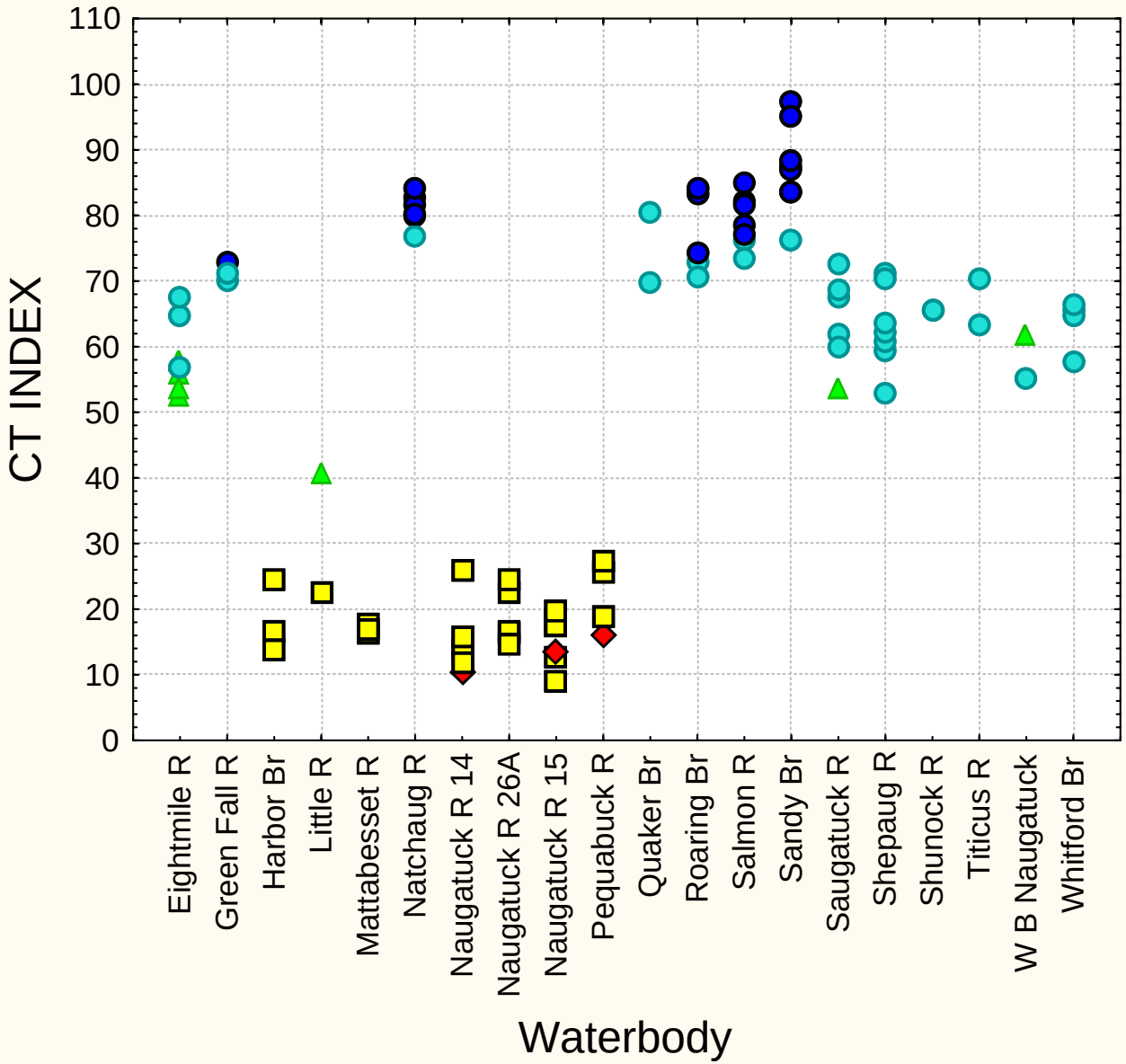


Source: J. DeShon, Ohio EPA

Questions

- What is the condition of the Nation's streams?
  - State?
  - Ecoregion?
  - EPA region?
- What is the response of aquatic biota to stressors?
- Has this river improved since permit limits were tightened?
- Why is this river impaired (what is the cause)?
- What will be the effect of climate change on our rivers?
- More?

Multi-year variability

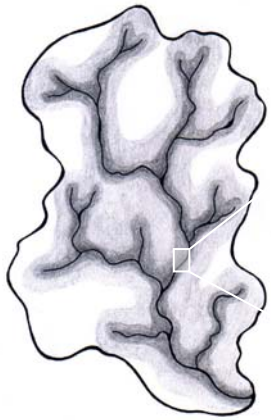


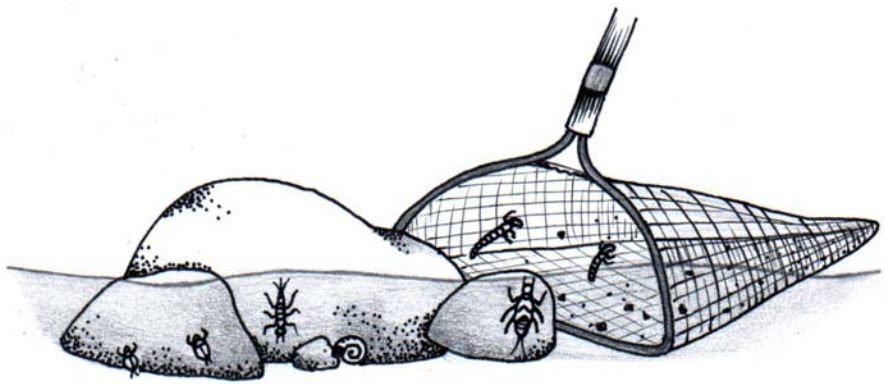
BCG Level

Solutions

- Probability-based sampling for statewide condition assessment
  - Stratify on stressors
- CWA is about more than just statewide condition!
  - NPDES – does it result in better condition?
    - Longitudinal and case-control studies
  - TMDL – how to deal with stressors?
    - Stressor-Response model development (causal assessment)
  - Nonpoint source management, watershed management
  - Biological monitoring is necessary to inform **all** management activities, not just 305(b)

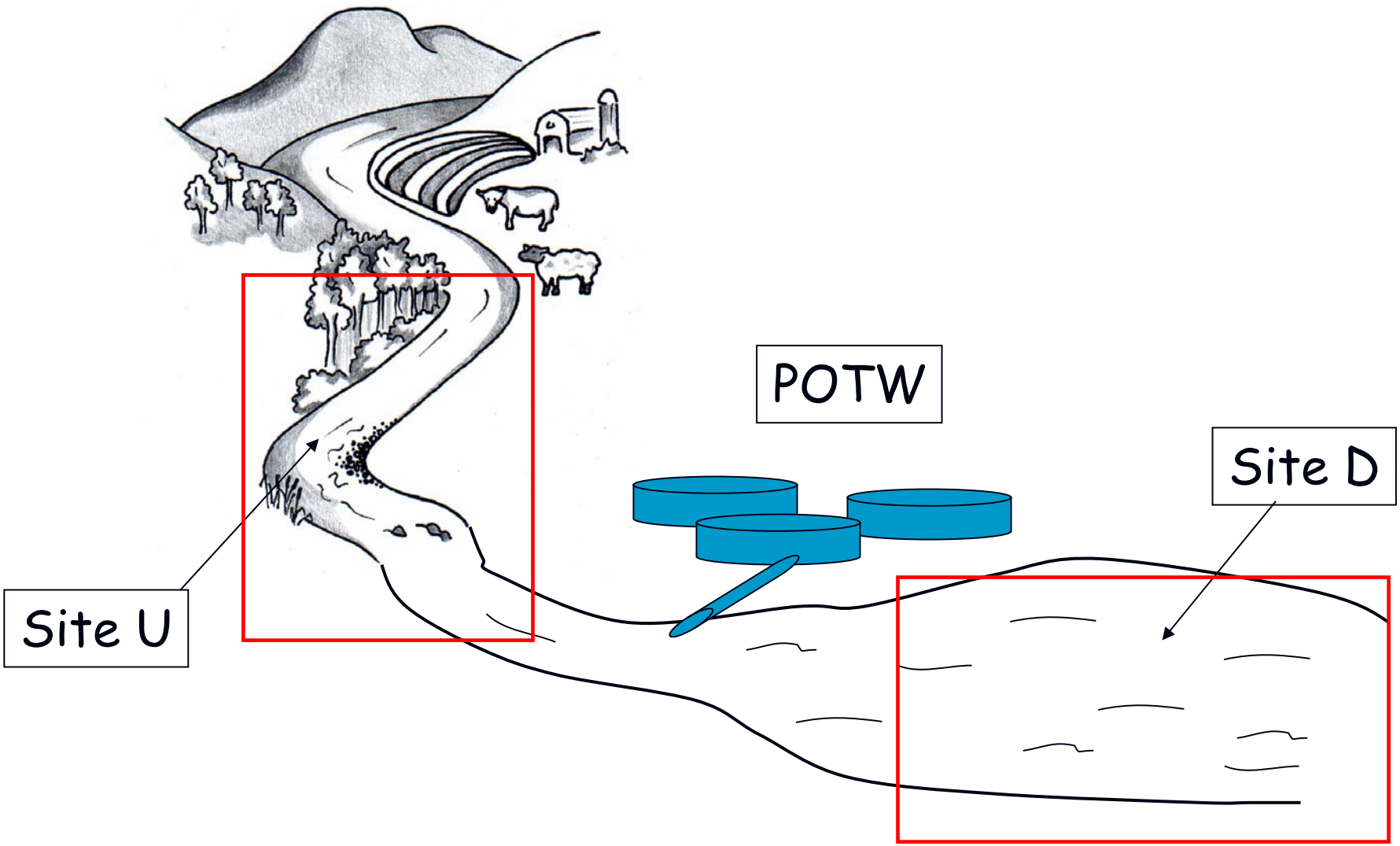
Stratification and sampling methods to control, account for, natural factors





Variability exists at multiple spatial scales!

Stratification to enhance response models and management



What and how far to stratify?

- Depends on principal questions and objectives
- Natural covariates
  - Ecoregion
  - Order
  - Gradient (slope)
- Sources, stressors , confounding factors
  - Land use
  - Discharges
  - Future changes

Conclusions

- Remember question, remember inference space!
- Probability-based surveys address national and some statewide needs
  - Judicious stratification
- Don't throw the baby out with the selected bathwater
  - Longitudinal studies will remain necessary to inform whether management is working
  - Effects of climate change (longitudinal and probability)
  - Stress-response to help identify causal relationship
  - Historic longitudinal sites should not be dropped