ANALYSIS OF ESTUARINE SEDIMENT CONTAMINANT AND TOXICITY DATA FOR ELICITING RESPONSES

John F. Paul

paul.john@epa.gov

National Health and Environmental Effects Research Laboratory

U.S. Environmental Protection Agency

Research Triangle Park, North Carolina

Thomas P. O'Connor

tom.oconnor@noaa.gov

NOAA/NOS

National Centers for Coastal Ocean Science

Silver Spring, Maryland

EMAP Symposium

4-6 May 2004

Basic Question

How well we can predict sediment toxicity using sediment contaminant data?

Impetus

When we decide to manage sediment contamination, we should understand how well we can predict the biological change?

Probability is appropriate way to express predictions

Outline of Presentation

- Definitions
- ERLs/ERMs
- Conditional probability analysis (CPA)
- Logistic regression model
- Logistic regression meets CPA & EMAP data
- Summary

Definition and Terminology

Probability **– quantification of the likelihood that something will occur**

Conditional probability**: probability of something occurring when it is known that something else has occurred**

P (Y) = probability of Y (***unconditional probability*****)**

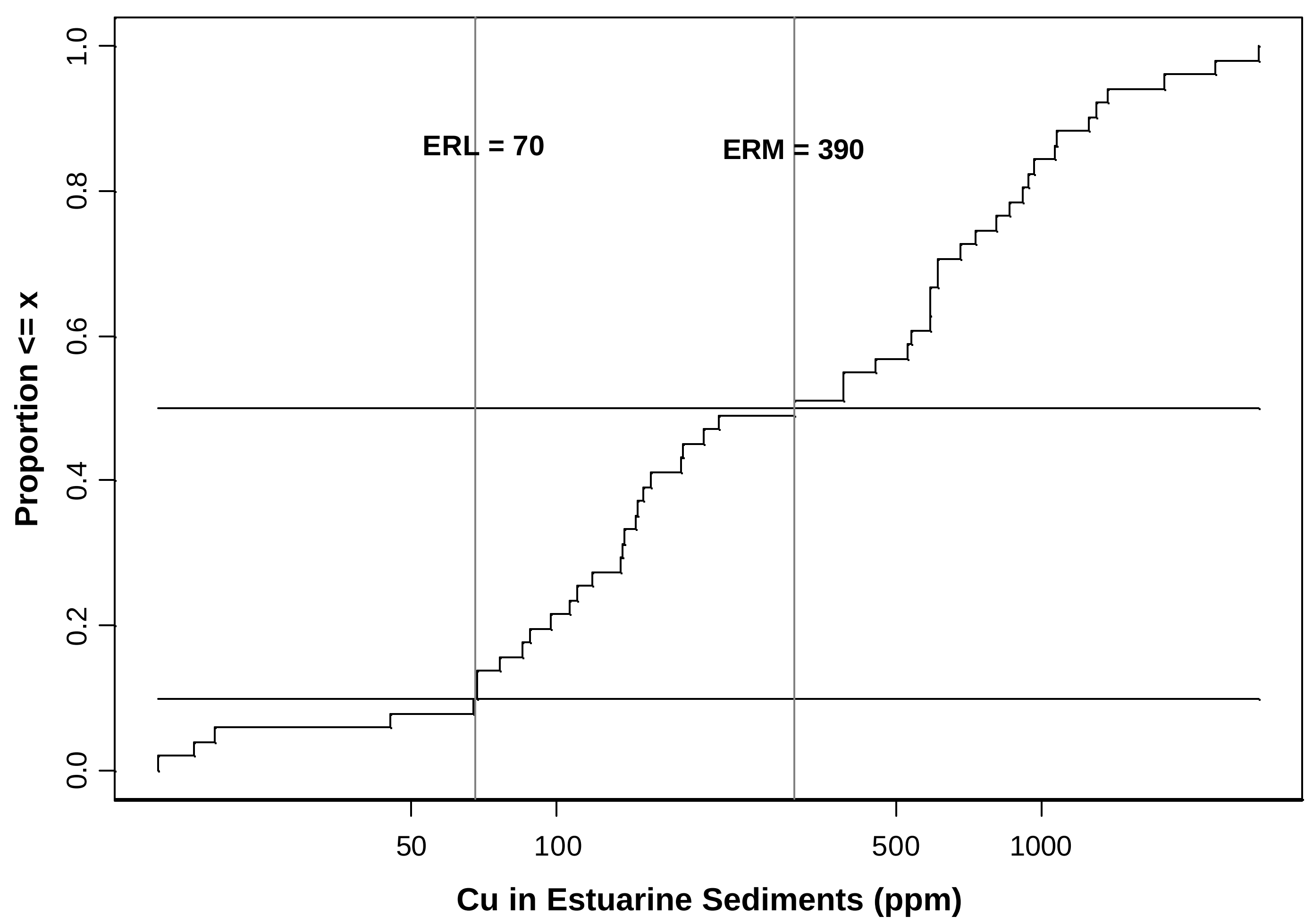
P (Y | X) = probability of Y if X occurs (***conditional probability*****)**

ERLs and ERMs

ERL and ERM are, respectively, the 10th and 50th percentiles among ranked concentrations associated with effects.

They are percentiles, they have no inherent predictive ability

Cumulative Distribution Function for Data Used to Derive Original ERL/ERM for Cu (Long and Morgan 1990)



ERL Concentrations
(Long et al. 1995)

Empirically based as 10th percentile among concentrations found to co-occur with a biological effect

Ag	1.0 ppm	Hg	0.15 ppm
Cd	1.2	Ni	21
Cr	81	Pb	47
Cu	34	Zn	150
As	8.2		

DDE	2.2 ppb	tDDT	1.6ppb
tPCB	23	tPAH	4,000

13 PAH compounds 16 to 700

Conditional Probability Analysis

Formulation as probability of aquatic impact if specific level of sediment contamination is exceeded

Incorporates uncertainty and natural variability into analysis

Conditional Probability Analysis

Given –

Y = 1 : impacted aquatic condition

X : sediment contamination

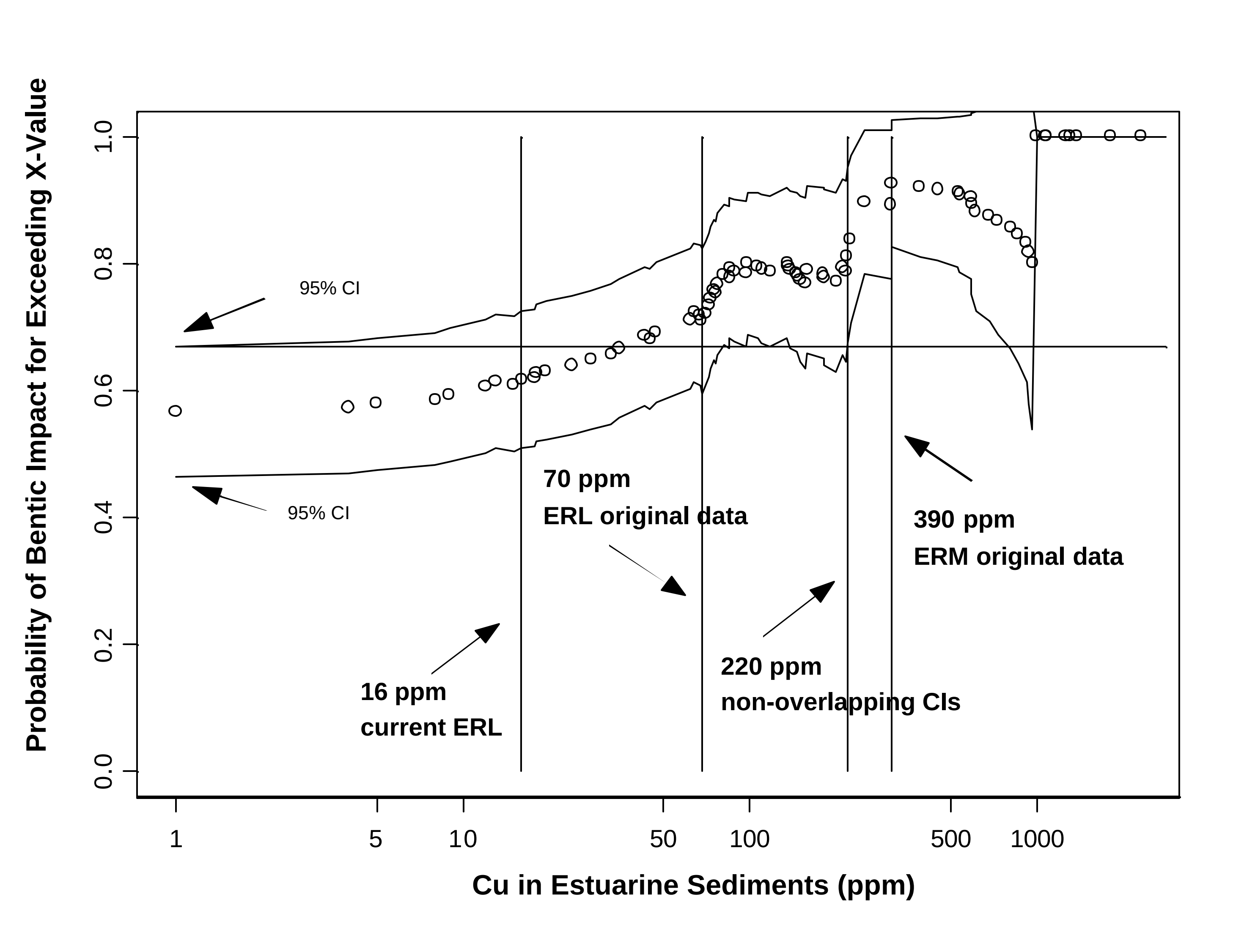
$$\mathbf{P}(\mathbf{Y} = 1 | \mathbf{X} > \mathbf{X}_C) = \frac{\mathbf{P}(\mathbf{Y} = 1, \mathbf{X} > \mathbf{X}_C)}{\mathbf{P}(\mathbf{X} > \mathbf{X}_C)}$$

Two-step approach to calculate P (Y = 1 | X > X_C) –

1. **Identify subset of resource for X > X_C**
2. **Determine fraction of subset with impact**

Calculate P (Y = 1| X > X_C) for all observed values of X_C

Conditional Probability Analysis on Data Used to Derive Original ERL/ERM for Cu



Quantified Predictability

1508 samples (NS&T and EMAP)
<ERL 5% toxic >ERL<ERM 13% >ERM 38%

2475 samples (NS&T and EMAP)
<ERL 5% toxic >ERL<ERM 13% >ERM 41%

2760 samples (NS&T, EMAP, and others)*
<ERL 8% toxic >ERL<ERM 20% >ERM 48%

***(16% of 2160 *Ampelisca abdita* tests showed toxicity,**
51% of 600 *Rhepoxnius abronius* tests showed toxicity)

Redefine ERLs and ERMs

ERL ˜ concentration at which about toxicity is found about 10% of the time

ERM ˜ concentration at which about toxicity is found about 50% of the time

Just coincidence that ERL and ERM are the 10th and 50th percentiles of concentrations co-occurring with effects

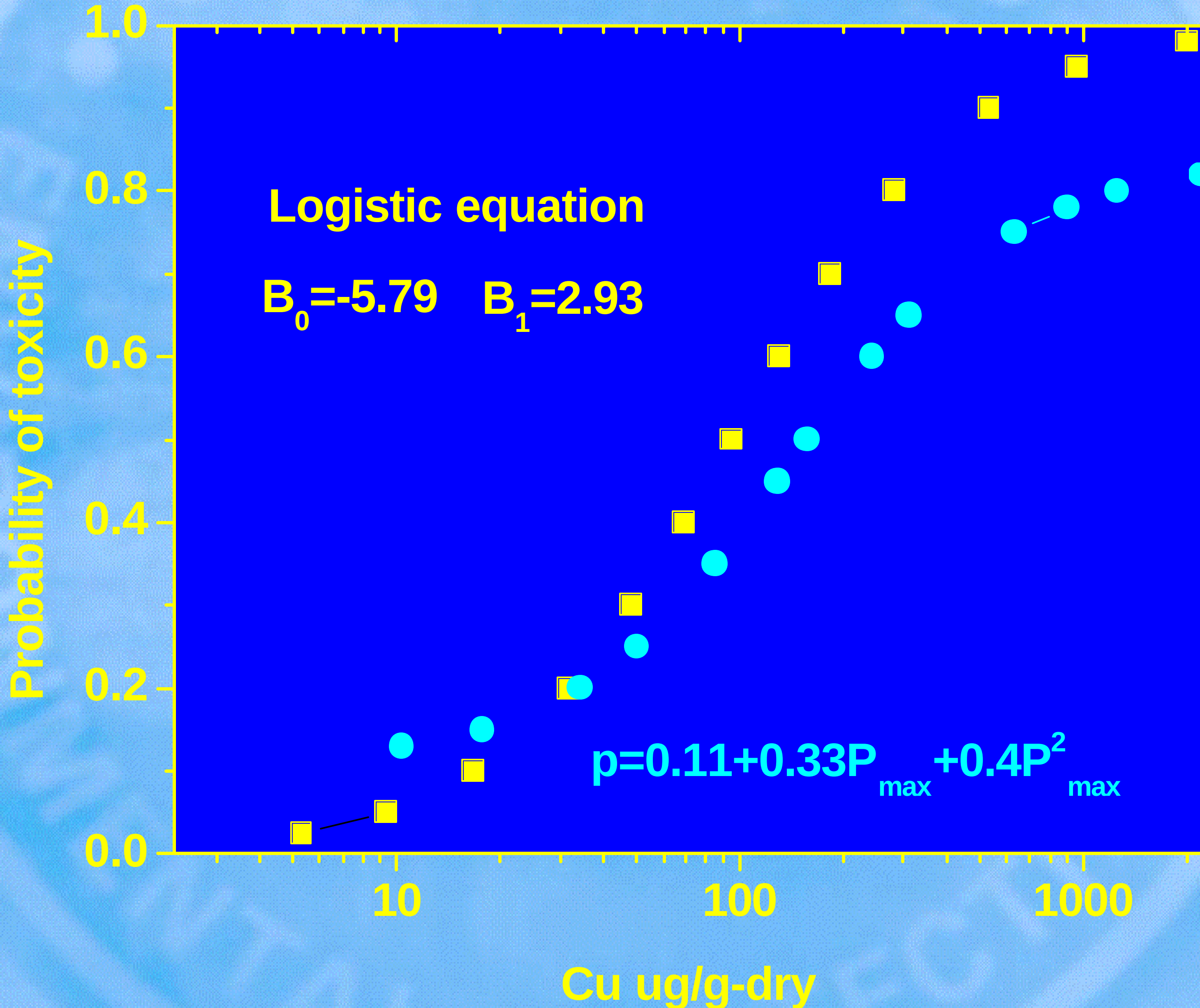
It does not follow that

5 ERL exceedances = 1 ERM exceedance

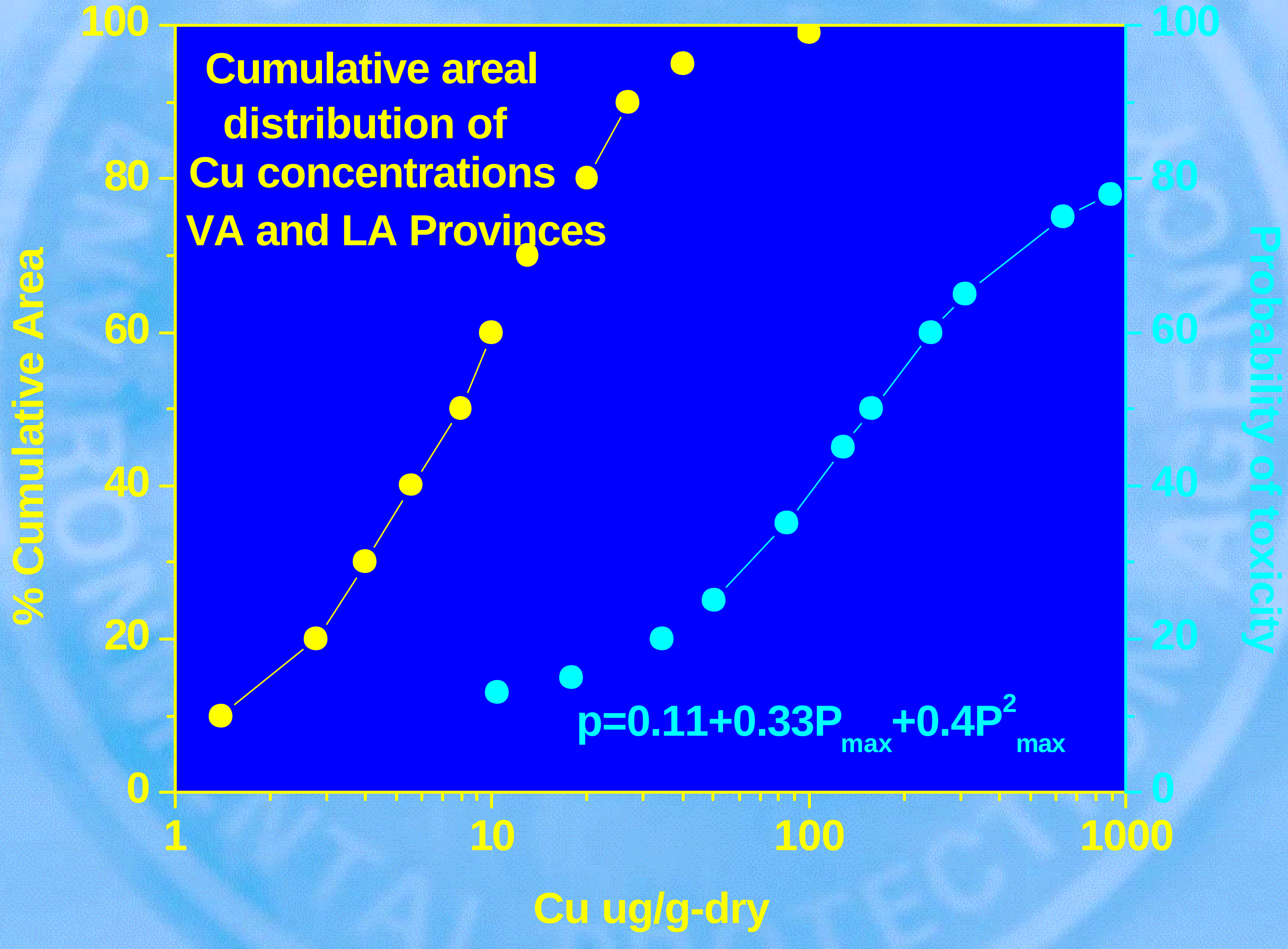
Logistic Regression (Field et al., 2002) yields a continuous relationship between probability (p) of toxicity and concentration of (x)

$$p = \frac{\exp(B_0 + B_1(\log_{10}(x))}{1+ \exp(B_0 + B_1(\log_{10}(x))}$$

$$\ln(p/(1-p)) = B_0 + B_1(\log_{10}(x))$$

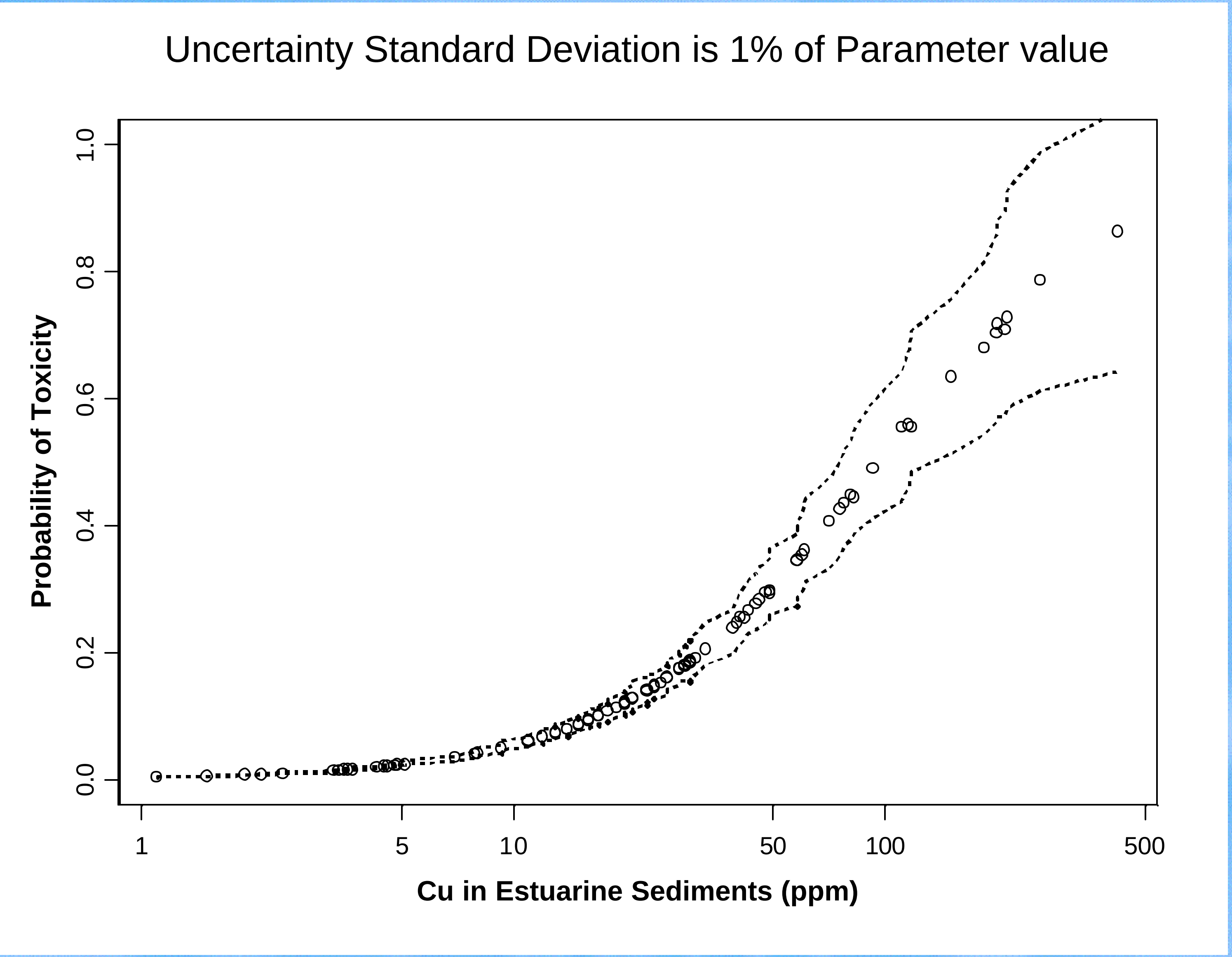


Logistic equation
B_0=-5.79 B_1=2.93
$p=0.11+0.33P_{max}+0.4P^2_{max}$
Probability of toxicity
1.0
0.8
0.6
0.4
0.2
0.0
10
100
1000
Cu ug/g-dry

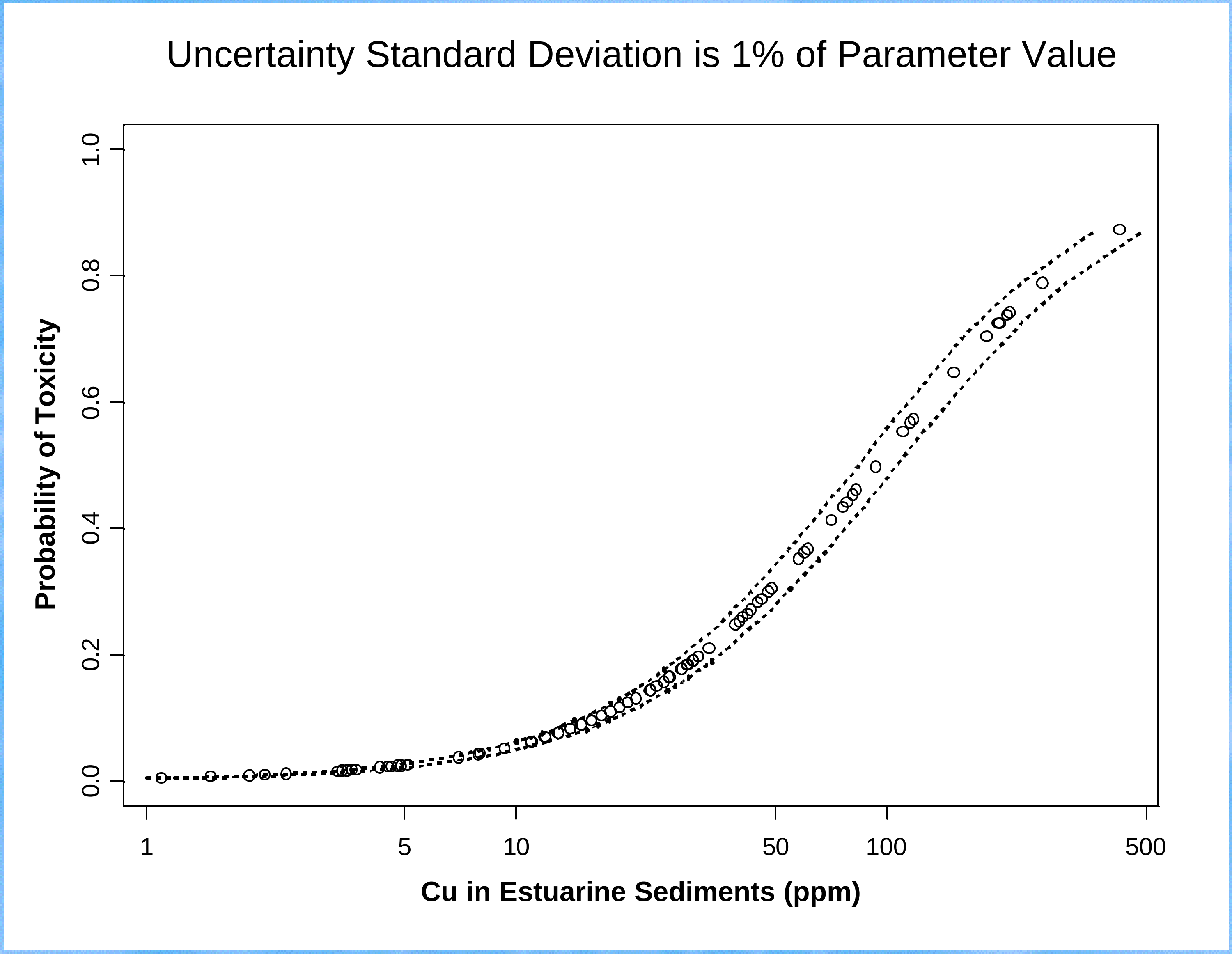


Cumulative areal distribution of Cu concentrations VA and LA Provinces
p=0.11+0.33P$_{max}$+0.4P$^2_{max}$
% Cumulative Area
Probability of toxicity
Cu ug/g-dry
0
20
40
60
80
100
1
10
100
1000

Simulations with Cu Levels (EMAP-VP 1990-93 Data) Using Logistic Regression with Uncertainty in Regression Parameters

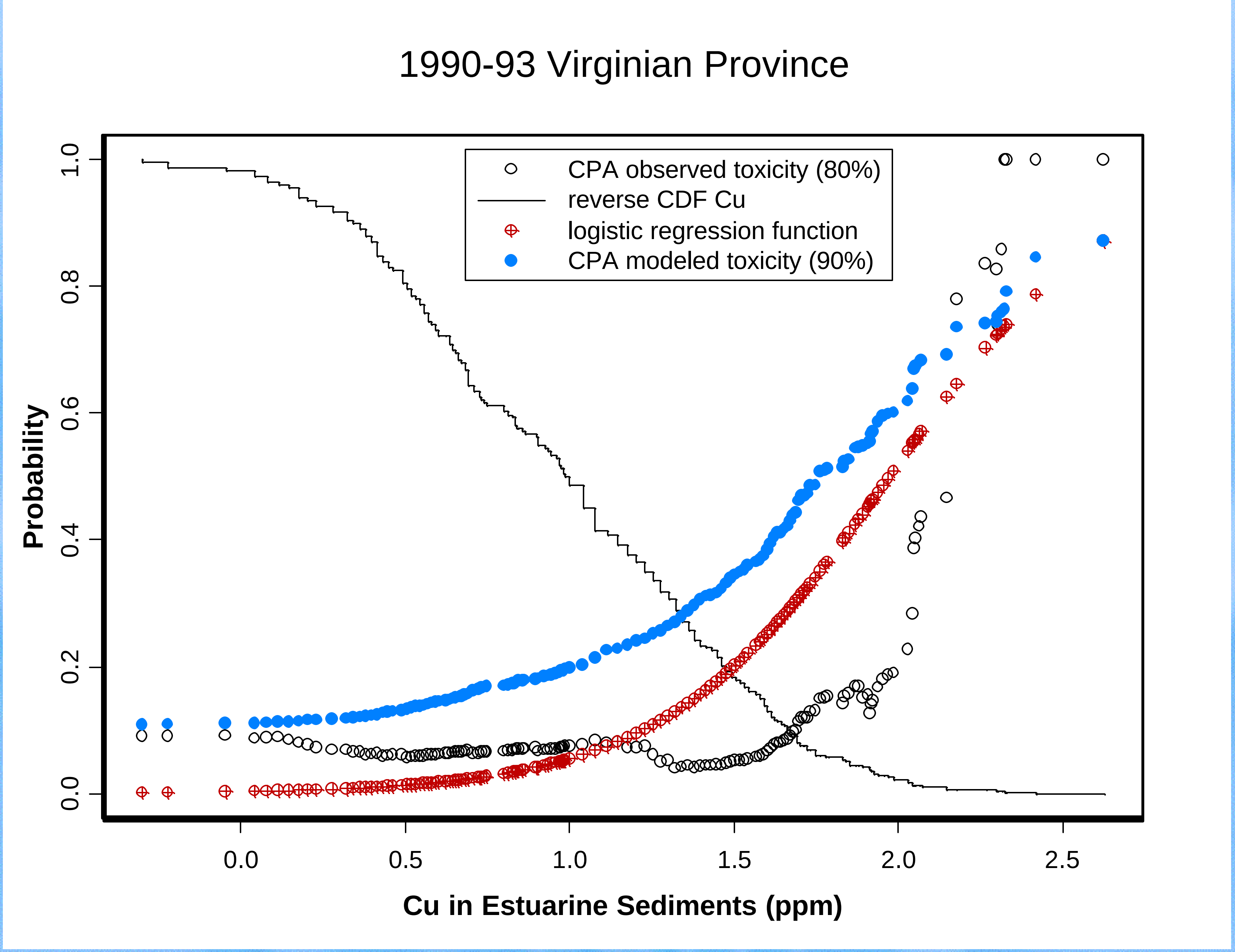


Simulations with Probability of Toxicity Using Logistic Regression with Uncertainty in Regression Parameters



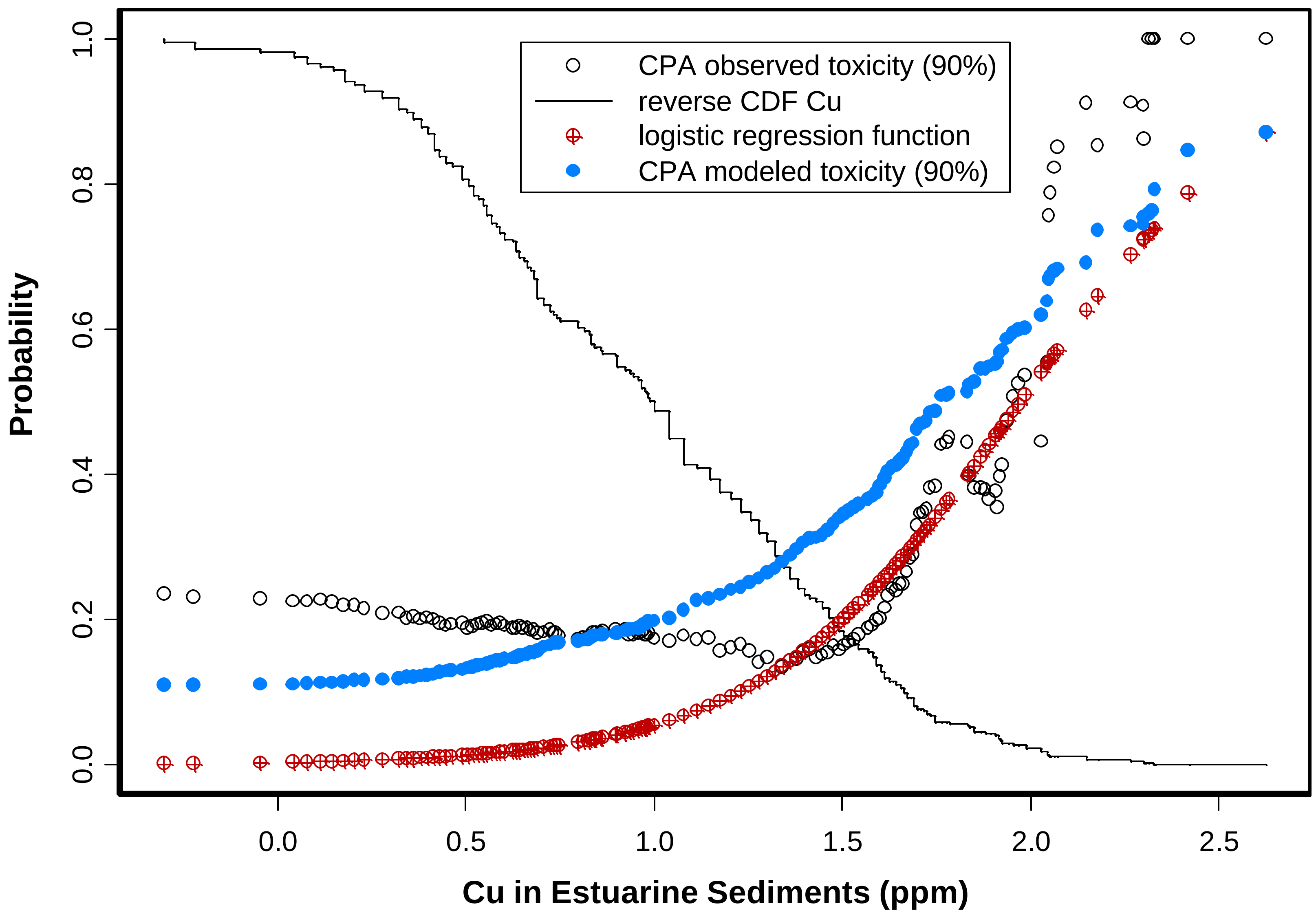
$$\mathbf{p} = \frac{\mathbf{e}^{(\mathbf{B}_0+\mathbf{B}_1(\mathbf{log}_{10}(\mathbf{X})))}}{1+\mathbf{e}^{(\mathbf{B}_0+\mathbf{B}_1(\mathbf{log}_{10}(\mathbf{X})))}}$$

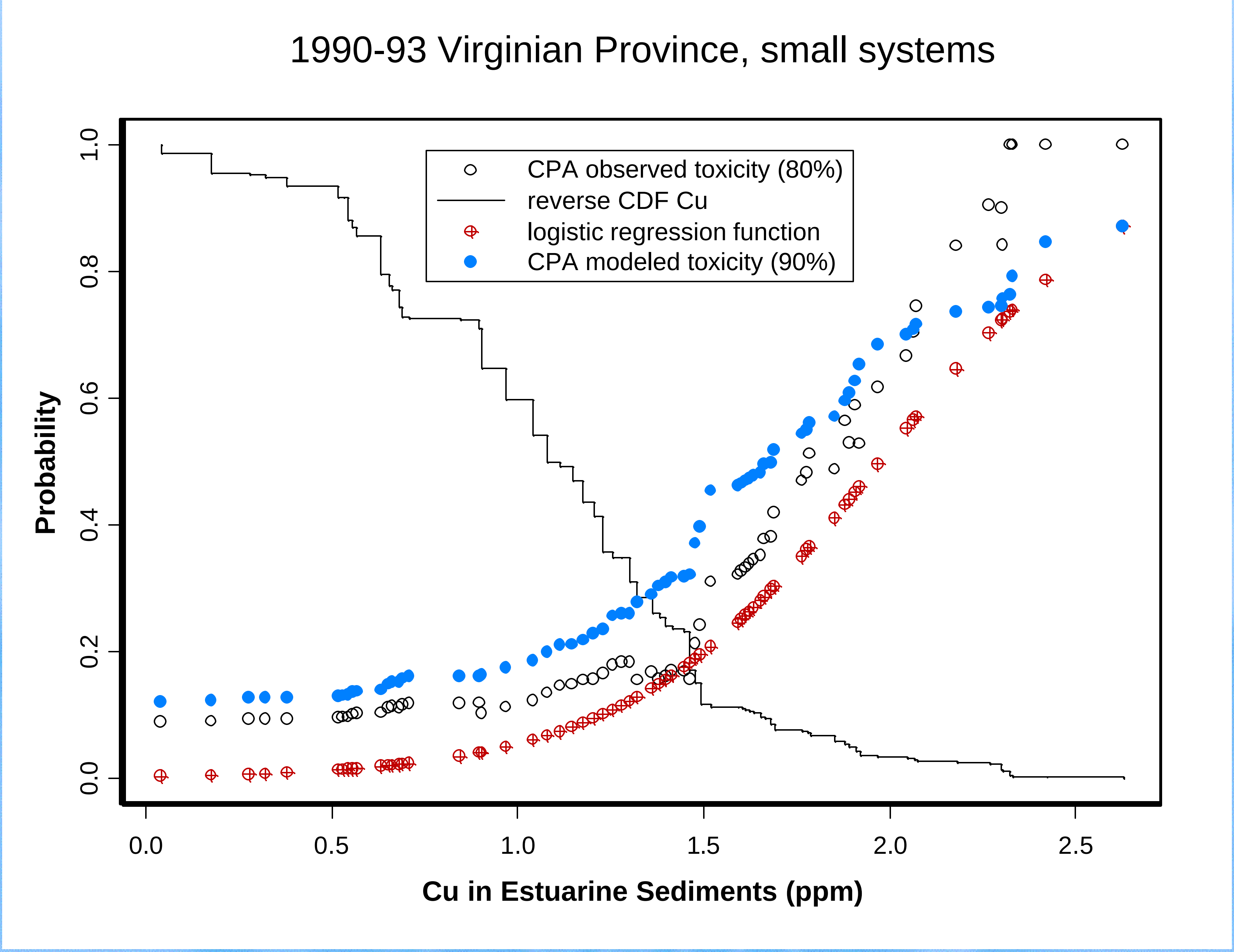
$$\mathbf{P}(\mathbf{Y} = 1 \mid \mathbf{X} > \mathbf{X}_C)$$



1990-93 Virginian Province
CPA observed toxicity (80%)
reverse CDF Cu
logistic regression function
CPA modeled toxicity (90%)
Probability
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
2.5
Cu in Estuarine Sediments (ppm)

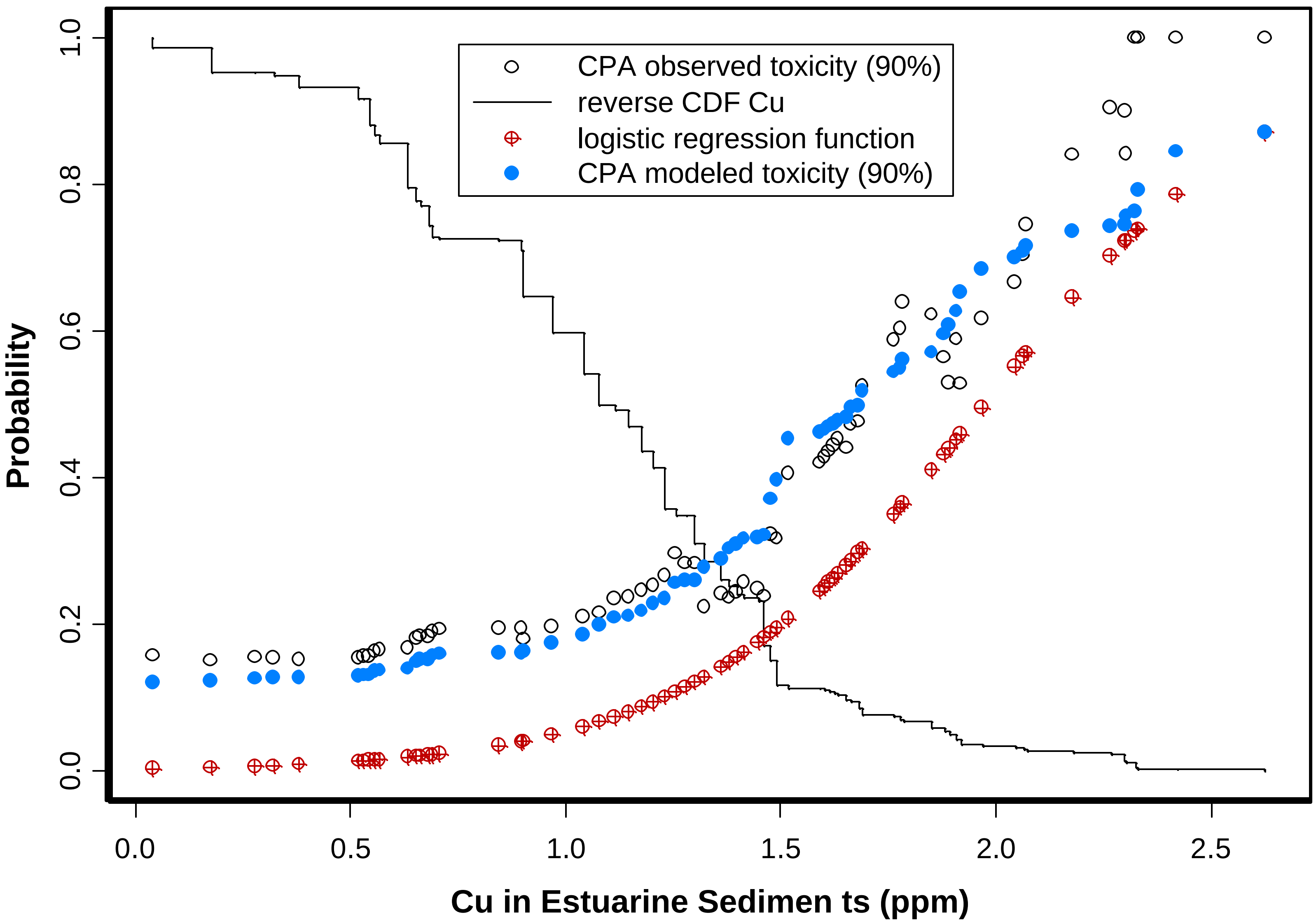
1990-93 Virginian Province





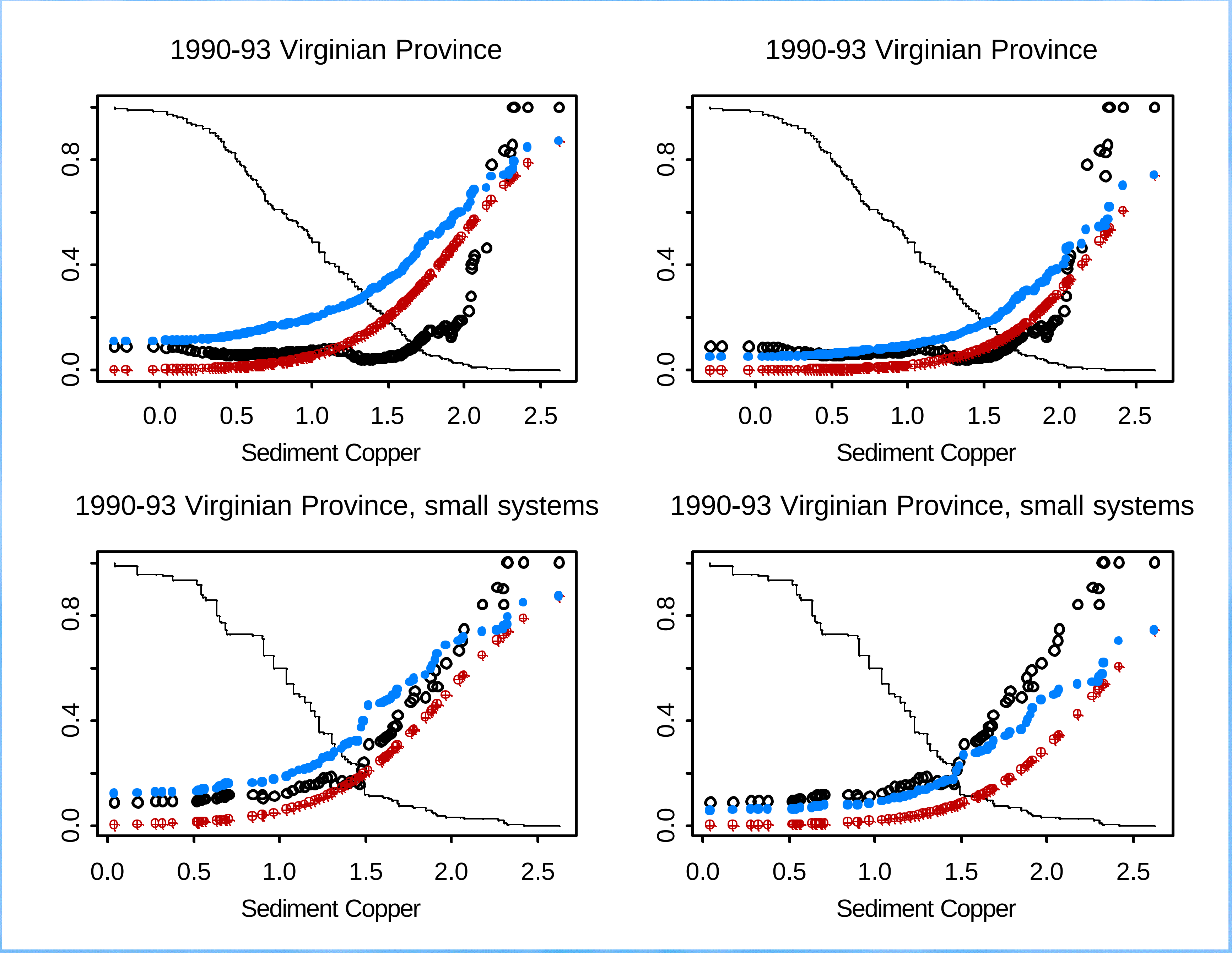
1990-93 Virginian Province, small systems
CPA observed toxicity (80%)
reverse CDF Cu
logistic regression function
CPA modeled toxicity (90%)
Probability
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
2.5
Cu in Estuarine Sediments (ppm)

1990-93 Virginian Province, small systems



Modeled toxicity (90% cut)

Modeled toxicity (80% cut)



Conclusions

Only approximate comparison of logistic regression predictions with observations because of differing definitions of toxic sediments

Small estuarine systems in Virginian Province observed to have most overall extensive sediment toxicity

Overall best comparison of predictions with observations for small systems

Predictions over or underestimate observations depending upon Cu levels in sediments and estuarine geomorphology

Conditional Probabiliy Analysis appears to be a robust approach for comparing predictions with observations

Messages We Want to Leave With You

ERLs/ERMs are not criteria and have no intrinsic predictive value

Incorporation of additional analyses with data can provide some predictive capability

May need to incorporate, at a minimum, estuarine geomorphology into predictions

Advantages of probability-based designs with consistently collected suite of indicators

Conditional Probability Analysis provides different analysis perspective

Approaches for Identifying Changepoint (Thresholds)

- **non-overlapping confidence intervals**
- **nonparametric deviance reduction**
- **change in curvature of fitted curve**

$$P(y = 1|x > x_C) = \begin{cases} \mathbf{1 + (D_0 - 1) / \left(1 + \exp(B_0 (x_C - x_0))\right), \text{ for } x_C > x_0} \\ \mathbf{1 + (D_0 - 1) / \left(1 + \exp(B_1 (x_C - x_0))\right), \text{ for } x_C \le x_0} \end{cases}$$