Estimating and Presenting Uncertainties in Empirical Data and Modeling Results

Dwayne R.J. Moore
Cantox Environmental, Inc.

CANTOX ENVIRONMENTAL

Overview

- Sources of uncertainty
- Methods for describing and presenting uncertainty
- Wildlife case study
- Recommendations for estimating and presenting uncertainty

CANTOX ENVIRONMENTAL

Introduction to Uncertainty Analysis

- Majority of ERAs deterministic and conservative (*i.e.*, screening level)
- However ...
  - degree of conservatism is opaque, undefined and uncontrolled
  - difficult to characterize risk, except in extreme situations
- 95th %ile x 95th %ile x 95th %ile ≠ 95th %ile
- EPA guidance now available
  - Guiding Principles for Monte Carlo Analysis
  - Superfund, ECOFRAM, Office of Pesticides

CANTOX ENVIRONMENTAL

Types of Uncertainty

- Variability
- Incertitude
- Model uncertainty

CANTOX ENVIRONMENTAL

Variability

- Arises from natural stochasticity
- Parameter variability due to
  - spatial distribution of chemical concentration
  - temporal fluctuations in rainfall
  - genetic differences among individuals
- Not reducible by empirical effort, but can be better understood

CANTOX ENVIRONMENTAL

Incertitude

- Arises from incomplete knowledge
- Parameter uncertainty due to
  - limited sample size
  - possible biases in sampling design
  - use of surrogate data
- Reducible with empirical effort

CANTOX ENVIRONMENTAL

Methods for Describing Model Uncertainty

- Simple statistics
- Interval analysis
- Monte Carlo
- Bayesian methods
- Second-order Monte Carlo
- Probability bounds analysis
- Many others

CANTOX ENVIRONMENTAL

Interval Analysis

How?

- circumscribe estimates with bounds, $a = [a_1, a_2]$, where $a_1 \leq a_2$
- addition: $[a_1, a_2] + [b_1, b_2] = [a_1+b_1, a_2+b_2]$
- subtraction: $[a_1, a_2] - [b_1, b_2] = [a_1-b_2, a_2-b_1]$
- multiplication: $[a_1, a_2] \times [b_1, b_2] = [a_1b_1, a_2b_2]$, $0 \leq a,b$
- division: $[a_1, a_2] \div [b_1, b_2] = [a_1/b_2, a_2/b_1]$, $0 \leq a,b$

Why?

- natural for scientists and easy to explain to others
- works no matter where uncertainty comes from

Why not?

- paradoxical: can't give exact value but can give exact bounds
- ranges can grow quickly, giving wide results
- repeated variables cumbersome to handle optimally

CANTOX ENVIRONMENTAL

Monte Carlo Simulation

How?

- replace each point estimate with a probability distribution
- repeatedly sample from each, tally answers in a histogram

Why?

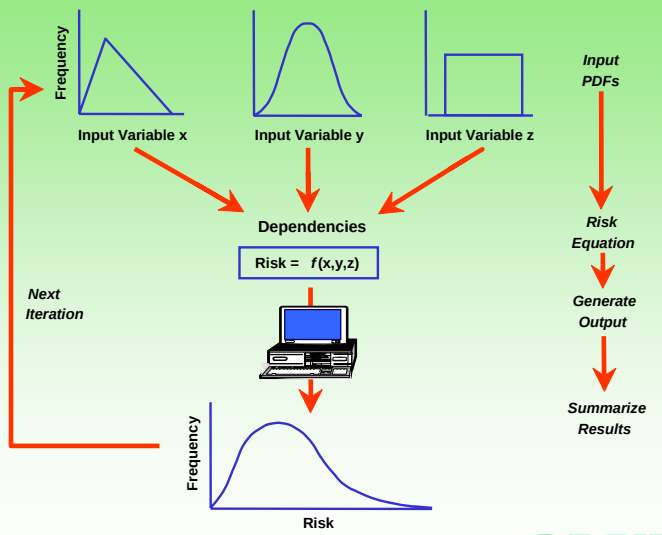
- simple to implement
- fairly simple to explain
- summarizes entire distribution of risk

Why not?

- requires a great deal of empirical information
- usually need to guess some things

CANTOX ENVIRONMENTAL

Steps Involved



CANTOX ENVIRONMENTAL

Second-order Monte Carlo

How?

- let parameters of input distributions be distributions as well
- nest Monte Carlo analyses

Why?

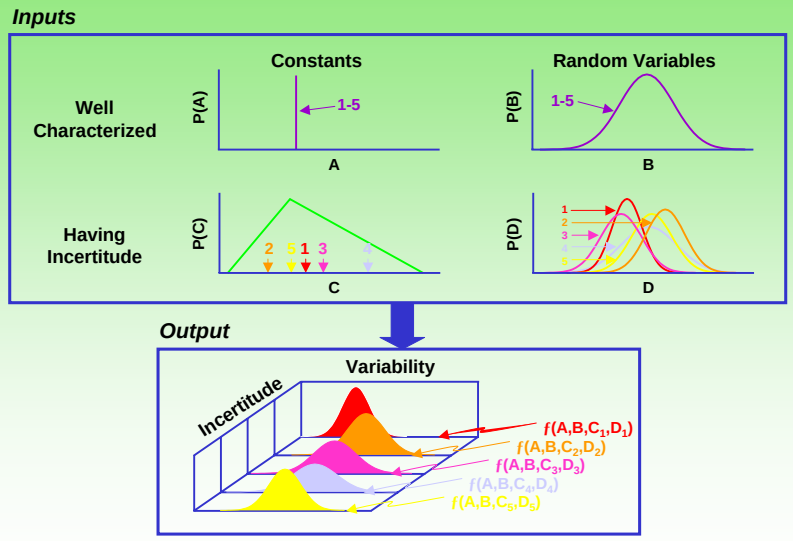
- partitions variability and incertitude
- plots get narrower with better empirical information

Why not?

- difficult to include some sources of uncertainty
- results are cumbersome to interpret and explain
- often difficult to separate incertitude and variability

CANTOX ENVIRONMENTAL

Steps Involved



CANTOX ENVIRONMENTAL

Probability Bounds

How?

- specify what you are sure about
- establish bounds on probability distributions
- select dependencies (no assumption, independence, perfect, etc.)

Why?

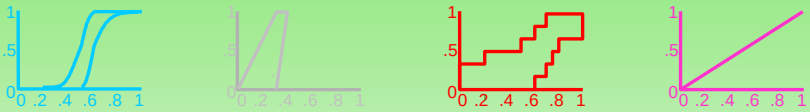
- partitions variability and incertitude
- puts bounds on Monte Carlo results
- bounds get narrower with better empirical information
- outputs easy to understand

Why not?

- cannot handle second-order probabilities
- may not be able to use subtle information to tighten bounds
- optimum bounds difficult to compute when variables repeated

CANTOX ENVIRONMENTAL

Steps Involved

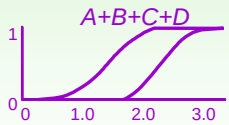


A = {lognormal, mean=[.5,.6], variance=[.001,.01]}

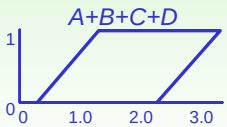
B = {min=0, max=.5, mode=.3}

C = {data = (.2, .5, .6, .7, .75, .8)}

D = {shape = uniform, min=0, max=1}



Under independence



Without independence

CANTOX ENVIRONMENTAL

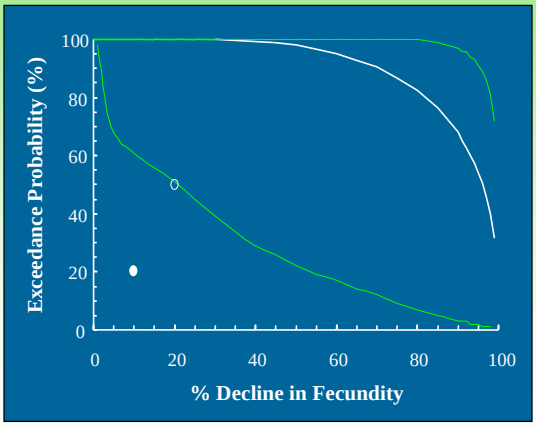
Advantages of 2-D Techniques

- With 1-D Monte Carlo, difficult to express confidence in the predictions
- Two output distributions may appear similar, but one may have low incertitude while the other has high incertitude
- Additional research effort only beneficial for high incertitude situations
- 2-D Monte Carlo and probability bounds analysis can be used to address this issue

CANTOX ENVIRONMENTAL

Presenting Modeled Risks

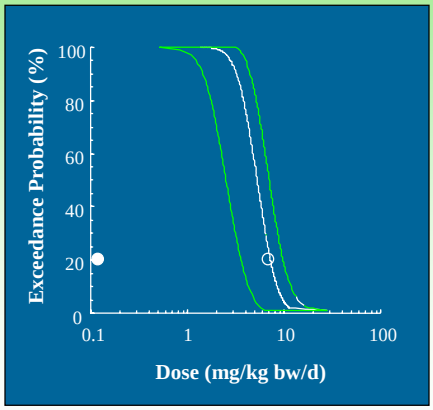
- If risk curve available:
  - probability of 10% or greater effect is less than 20% = LOW RISK
  - probability of 20% or greater effect is greater than 50% = HIGH RISK
  - all other outcomes = INTERMEDIATE RISK



CANTOX ENVIRONMENTAL

Presenting Modeled Risks, continued

- If risk curve not available:
  - probability of exceeding lower toxicity threshold is less than 20% = LOW RISK
  - probability of exceeding the upper toxicity threshold is greater than 20% = HIGH RISK
  - all other outcomes = INTERMEDIATE RISK



CANTOX ENVIRONMENTAL

Case Study: Mink Exposed to tPCBs in Housatonic River Area

CANTOX ENVIRONMENTAL

Exposure Modeling

- Used standard total daily intake (*TDI*, mg/kg bw/d) model

$$TDI = \left[FMR \bullet \sum_{i=1}^{n} \frac{C_i \cdot P_i}{GE_i \cdot AE_i} \right] \cdot Pt$$

- where,
  - *FMR* = free metabolic rate (kcal/kg/d)
  - C_i = concentration in prey item (mg/kg ww)
  - P_i = proportion of prey item in the diet
  - AE_i = assimilation efficiency for mink consuming prey item
  - GE_i = gross energy of prey item (kcal/kg)
  - *Pt* = proportion of time in contaminated area
- Probabilistic modeling conducted using both Monte Carlo and probability bounds analysis

CANTOX ENVIRONMENTAL

Incorporating Uncertainty in Empirical Data

- Uncertainty due to limited sample size
  - calculate confidence limits on measures of centrality and dispersion
  - for lognormal data (e.g., concentration), EPA recommends Land *H*-statistic
  - for normal data, use Student's *t*
  - use Chebyshev, bootstrap resampling or jackknife when sample size is small
- In mink ERA, UCL from Land *H*-statistic used to represent prey concentration in Monte Carlo analyses
- LCL and UCL used in p-bounds analyses

CANTOX ENVIRONMENTAL

Prey Concentration

Chemical:	tPCBs
Medium:	Whole Body Fish
Location:	Reach 5
Units:	mg/kg

Concentration	LN Concentration
44.91539	3.804780498
51.8434	3.948227636
61.90942	4.125672349
32.76553	3.489377048
54.73981	4.002591233
44.633739	3.798490053
57.42488	4.050477659
70.090022	4.249780444
54.879926	4.005147635
106.315842	4.666414305
42.008099	3.737862433
40.102283	3.691433265

Sample Size (n)	27		
Arithmetic Mean	3.812209	of LN Concentration	
Standard Deviation (s_y)	0.435637	of LN Concentration	
Land $H_{0.95}$	1.902969		
Land $H_{0.05}$	-1.63421		
$UCL_{0.95}$ on the Mean	58.53824		
$UCL_{0.05}$ on the Mean	43.27087		
Interpolation Calculations			
Lower s_y	0.4	Closest value below ac	
Upper s_y	0.5	Closest value above a	
Actual s_y	0.435637		
	Lower n	*Upper n*	
Sample Size	21	31	Closest val
$H_{0.95}$ for Lower s_y	1.905	1.856	Lookup val
$H_{0.95}$ for Upper s_y	1.989	1.928	Lookup val
$H_{0.95}$	1.934935	1.881659	

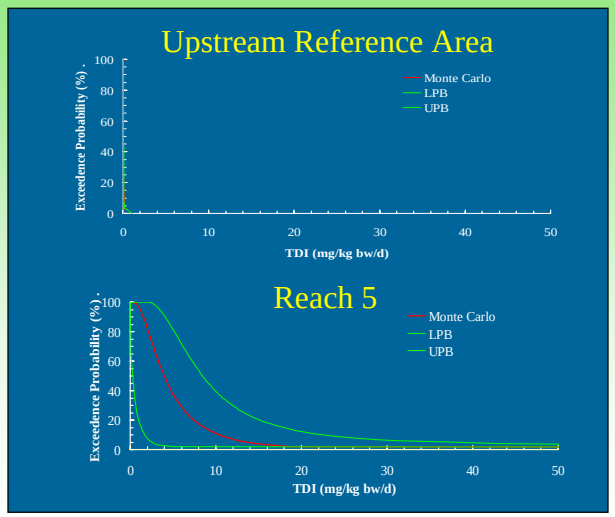
CANTOX ENVIRONMENTAL

Other Sources of Empirical Uncertainty

- Uncertainty arising from non-detects
  - Conducted bounding analysis for each prey concentration variable
  - If ratio of maximum possible mean (ND=DL) to minimum possible mean (ND=0) <1.3, point estimate assuming ND=0.5 DL used
  - Otherwise, uniform distribution (min to max) and distribution-free range used in Monte Carlo and p-bounds analyses, respectively
- Analogous approach may be used to deal with analytical uncertainty

CANTOX ENVIRONMENTAL

Mink Exposure to Total PCBs

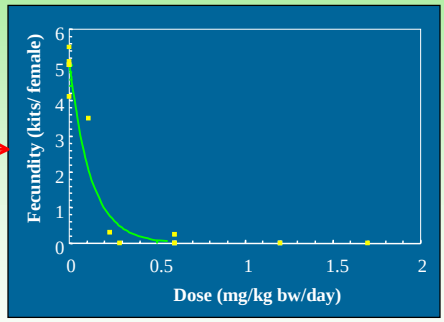


CANTOX ENVIRONMENTAL

Effects Assessment

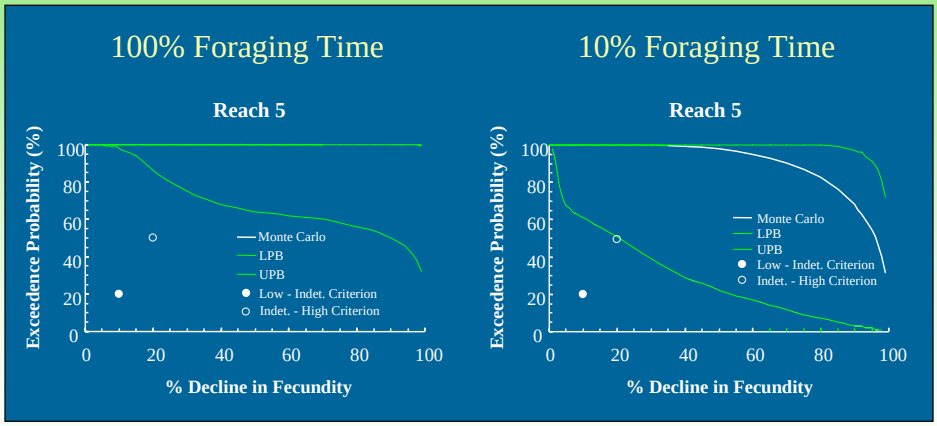
Hierarchy for Choosing Effects Metrics

1. Dose-response relationship from single study
2. Dose-response relationship from combined similar studies
3. NOAEL and LOAEL
4. Field-based threshold range
5. Threshold range spanning sensitive and tolerant species



CANTOX ENVIRONMENTAL

Risk Curves – Mink Exposed to Total PCBs



CANTOX ENVIRONMENTAL

Other Lines of Evidence

- Field surveys indicated reduced abundance and few, if any, residents in contaminated areas
- Feeding study with contaminated fish indicated:
  - 20% reduction in survival of kits from 0 to 6 weeks with approx. 1% contaminated fish in diet
  - Increased incidence of jaw lesions as % contaminated fish in the diet increased
- In this scenario, other lines of evidence reduced uncertainty about modeled risk estimates

CANTOX ENVIRONMENTAL

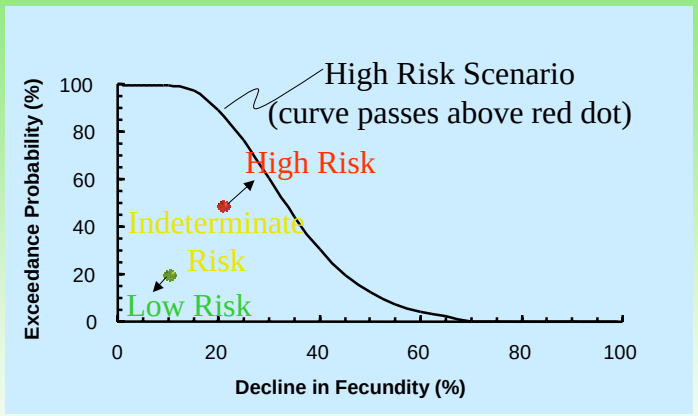
Recommendations

- Use simple techniques (e.g., interval analysis) to identify risk scenarios
- Conduct sensitivity analyses to identify critical variables – guides data collection
- Use Monte Carlo analysis in higher tier assessments – but combine with a 2D method to determine degree of incertitude
- Collect other lines of evidence, conduct weight-of-evidence evaluation
- List sources of uncertainty not included in analyses – discuss possible influence on risk estimates
- Engage risk manager, trustees, industry, public, etc throughout process

CANTOX ENVIRONMENTAL

Importance to Decision Making

- Small Piscivorous Mammals Exposed to tPCBs in Bayou D'Inde -



- Ideally want to push risk down to green dot
- At a minimum, push risk down to red dot
- Need interested parties to agree on risk criteria

CANTOX ENVIRONMENTAL