EPA United States Environmental Protection Agency

Web-based Interspecies Correlation Estimation (Web-ICE) for Acute Toxicity

Sandy Raimondo

USEPA/ORD/NHEERL/GED

12 February 2014

http://www.epa.gov/ceampubl/fchain/webice/

Web-ICE Team (past & present)

USEPA/NHEERL/ORD/NHEERL

Mace Barron
Crystal Jackson
Deborah Vivian
Jill Awkerman
Larisa Lee
Julie Krzykwa
Chris Russom
Sonny Mayer
Marion Marchetto
Anthony DiGirolamo
Brandon Jarvis
Christel Chancy
Nathan Lemoine

USEPA/Regions

Candice Bauer
Edward Hammer
Burt Shepard

USEPA/Program Offices

Kris Garber (OPP)
Elizabeth Riley (OPP)
Thomas Steeger (OPP)
Brian Montague (OPP)
Don Rodier (OPPT)
Vince Nabholz (OPPT)
Wade Lehman (OW)

USGS

Christopher Ingersoll
Ning Wang
John Besser

US Fish & Wildlife Service

Tom Augspurger

Proctor & Gamble

Scott Dyer
Scott Belanger
Jessica Brill
Don Versteeg
Joel Chaney

Environment Canada

Pierre Mineau
Alain Baril
Brian Collins

Computer Sciences Corporation

Wally Schwab
Derek Lane

What is Web-ICE?

http://www.epa.gov/ceampubl/fchain/webice/

- An internet application developed by the US Environmental Protection Agency, Office of Research and Development, Gulf Ecology Division.

- Uses interspecies correlation estimation (ICE) models to estimate acute toxicity to a species, genus, or family from the known toxicity of a surrogate species.

- Contains modules to derive acute hazard levels (e.g., HC5) and endangered species toxicity useful to chemical hazard assessment and Ecological Risk Assessment (ERA).

How can Web-ICE valuable to ERA?

Estimates acute toxicity (LC/EC50) for a species, genus or family from a surrogate species

<u>ICE in ERA</u>

- Populates toxicity database
- Direct toxicity estimation for endangered species
- Allows for species sensitivity comparisons and development of SSDs

Rainbow trout

Atlantic salmon

Discussion Outline

1) **Technical Basis of ICE models**
   - Database & model development
   - Validation and uncertainty analysis
   - User guidelines / Rules of Thumb

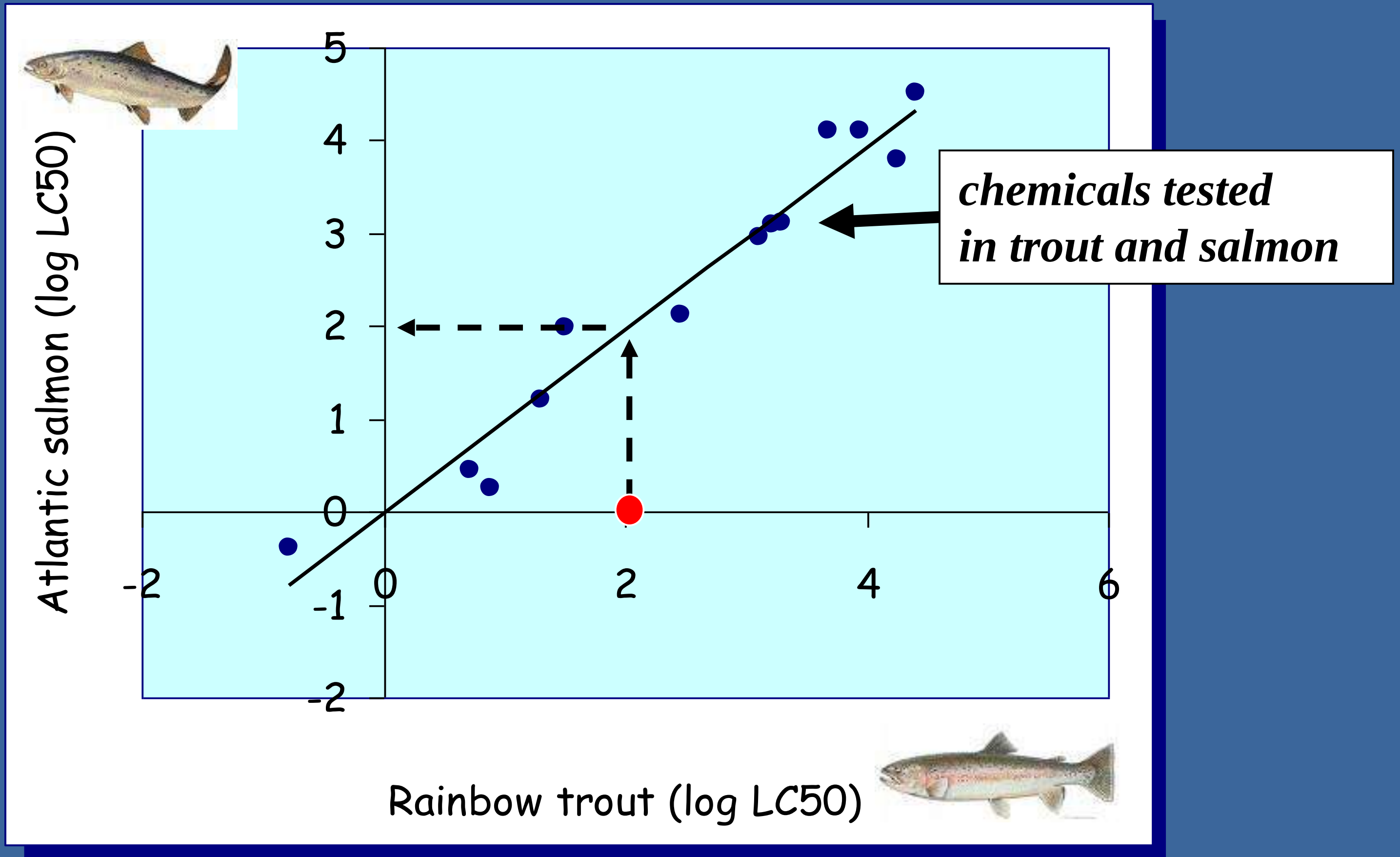
2) **Applications in SSDs**

10:30 – 12:00 **Web-ICE Demonstrations**

Interspecies Correlation Estimation (ICE)

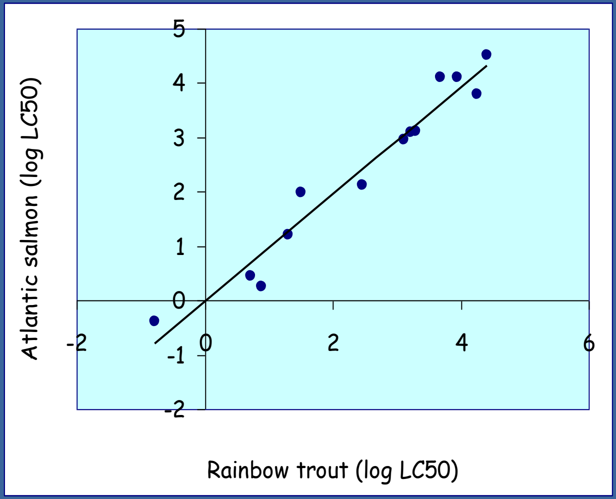
Log-linear least squares regressions of the relationship between the acute toxicity (LC50/LD50) of chemicals tested in two species.

$$\text{Log}(\text{Toxicity}_{predicted}) = a + b*\text{Log}(\text{Toxicity}_{surrogate})$$



ICE Models – the basics

1. ICE models start with large database of acute toxicity
   – e.g. ICE v3.1 aquatic invert/fish database = 5487 LC50 values; 1258 chemicals; 180 species
2. All possible pairings of species by common chemical
   – e.g. 180*180 = 32400 potential pairings
3. ICE model = Log-linear least squares regression of common chemicals tested in two species
   – some pairings will not yield any ICE model
   – some models will not be significant ($p>0.05$)
4. Suite of ICE models dependent on toxicity database

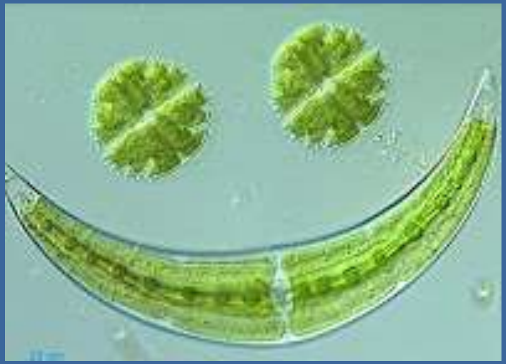


Web-ICE Databases

<u>Aquatic (fish and invertebrates):</u>

- US EPA Office of Pesticide Programs Ecotoxicity Database
- US EPA Office of Pollution Prevention & Toxics PMN, HPV
- US/EPA/ Office of Research and Development
  - Ecotox
  - e.g. Mayer 1987
- Ambient Water Quality Criteria
- US Geologic Survey
- Mayer and Ellersieck 1986
- Open literature sources
  - endangered species
  - mollusks

Web-ICE Databases



NEW Aquatic (algae):

- Procter & Gamble (CRADA)
- EPA Office of Pesticide Program
- Open literature searches
- Ecotox

Wildlife (birds and mammals):

- US EPA Office of Pesticide Programs Ecotoxicity Database
- Environment Canada (Baril et al. 1994)
- Hudson et al. 1984
- Schafer et al. 1983
- Shafer and Bowles 1985
- Safer and Bowles 2004
- Smith 1987

Data Standardization

All Databases

- cas/name consistency
- single compound tested
- a.i. $\geq$ 90%
- chemical & element specific AWQC normalizations (e.g. ammonia, metals)

Aquatic animal

- no sediment, dietary, mixed dose exposures or phototoxicity results
- ASTM/OPPTS standards
- endpoint = mortality/immobilization
- 48h EC/LC50 - daphnids, midges and mosquitoes
- 96h EC/LC50 - fish and other invertebrates
- juvenile only: fish, amphibians, insects, molluscs, decapods
- other groups: all life stages

Aquatic plants (algae)

- 72 or 96 hour EC50, growth rate or biomass

Wildlife

- Single oral dose LD50 (mg/kg), adults only

Web-ICE Databases

Taxa	# Records	# Chemicals	# Species
Aquatic animal	5487	1258	180
Algae	1647	457	57
Wildlife	4329	951	156

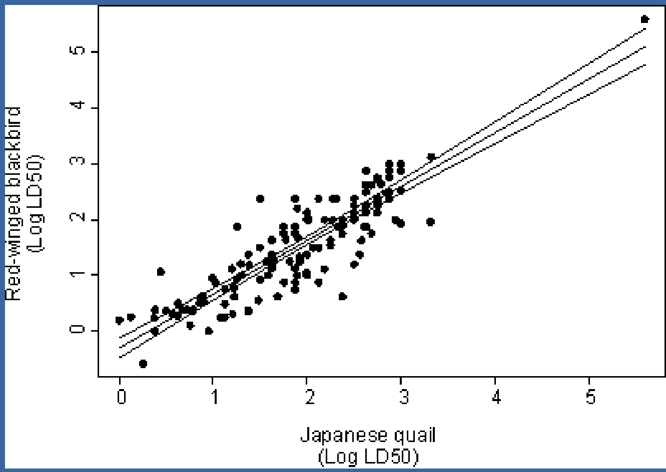
	A	B	C	D	G	H
1	data source	chemical	cas	species	type	toxicity (ug/L)
2	OPP	(S)-Dimethenamid	163515148	Cyprinodon variegatus	F	12000
3	OPP	(S)-Dimethenamid	163515148	Daphnia magna	F	12000
4	OPP	(S)-Dimethenamid	163515148	Lepomis macrochirus	F	10000
5	OPP	(S)-Dimethenamid	163515148	Oncorhynchus mykiss	F	6300
6	ECOTOX	1,1,2,2-tetrachloroethane	79345	Paratanytarsus dissimilis	S	30800
7	ECOTOX	1,1,2,2-tetrachloroethane	79345	Pimephales promelas	F	13400
8	ECOTOX	1,1,2,2-tetrachloroethane	79345	Pimephales promelas	F	20300
9	ECOTOX	1,1-Dichloroethane	75354	Menidia beryllina	S	250000
10	ECOTOX	1,2,4-trichlorobenzene	120821	Oncorhynchus mykiss	F	1530
11	OPP	1,2-Benzenedicarboxaldehyde	643798	Daphnia magna	SR	87
12	OPP	1,2-Benzenedicarboxaldehyde	643798	Daphnia magna	S	80
13	OPP	1,2-Benzenedicarboxaldehyde	643798	Oncorhynchus mykiss	SR	20
14	OPP	1,2-Benzenedicarboxaldehyde	643798	Oncorhynchus mykiss	F	72
15	ECOTOX	1,2-dichlorobenzene	95501	Lepomis macrochirus	S	27000
16	ECOTOX	1,2-dichlorobenzene	95501	Lepomis macrochirus	S	320000
17	ECOTOX	1,2-dichlorobenzene	95501	Menidia beryllina	S	7300
18	ECOTOX	1,2-dichlorobenzene	95501	Menidia beryllina	S	240000
19	ECOTOX	1,2-dichlorobenzene	95501	Oncorhynchus mykiss	F	1580
20	ECOTOX	1,2-dichlorobenzene	95501	Paratanytarsus dissimilis	S	12000
21	OPP	1,3 Dichloropropene	542756	Crassostrea virginica	F	640
22	OPP	1,3 Dichloropropene	542756	Cyprinodon variegatus	F	870
23	OPP	1,3 Dichloropropene	542756	Daphnia magna	S	90
24	OPP	1,3 Dichloropropene	542756	Lepomis macrochirus	F	3700
25	OPP	1,3 Dichloropropene	542756	Lepomis macrochirus	S	6700
26	OPP	1,3 Dichloropropene	542756	Micropterus salmoides	S	3650
27	OPP	1,3 Dichloropropene	542756	Pimephales promelas	S	4100
28	OPP	1,3 Dichloropropene	542756	Sander vitreus	S	1080
29	OPP	1,3,5-Triethylhexahydro-s-triazin	7779273	Daphnia magna	S	15300
30	OPP	1,3,5-Triethylhexahydro-s-triazin	7779273	Daphnia magna	F	25000
31	OPP	1,3,5-Triethylhexahydro-s-triazin	7779273	Lepomis macrochirus	S	30800
32	OPP	1,3,5-Triethylhexahydro-s-triazin	7779273	Oncorhynchus mykiss	S	23300
33	OPP	1,3,5-Triethylhexahydro-s-triazin	7779273	Oncorhynchus mykiss	F	35000
34	OPP	1,4-dichlorobenzene	106467	Lepomis macrochirus	S	6400
35	OPP	1,4-dichlorobenzene	106467	Oncorhynchus mykiss	S	880
36	ECOTOX	1,4-dichlorobenzene	106467	Oncorhynchus mykiss	F	1120
37	ECOTOX	1,4-dichlorobenzene	106467	Paratanytarsus dissimilis	S	13000
38	OPP	1,4-dichlorobenzene	106467	Salvelinus fontinalis	S	1670
39	Boogard	2-(digeranylamino)-ethanol	001	Carassius auratus	S	290
40	Boogard	2-(digeranylamino)-ethanol	001	Cyprinus carpio	S	50.7
41	Boogard	2-(digeranylamino)-ethanol	001	Fundulus diaphanus	S	792
42	Boogard	2-(digeranylamino)-ethanol	001	Lepomis cyanellus	S	640
43	Boogard	2-(digeranylamino)-ethanol	001	Lepomis macrochirus	S	720
44	Boogard	2-(digeranylamino)-ethanol	001	Micropterus dolomieu	S	237
45	Boogard	2-(digeranylamino)-ethanol	001	Micropterus salmoides	S	[illegible]

EcoToxdata

Model Development

Models are only developed for species within the same database (e.g., no algae to fish models)

Log10(predicted toxicity) = a + b*Log10(surrogate toxicity)



Taxa	Species	Genus	Family
Aquatic animal	780	289	374
Algae	58	44	0
Wildlife	560	0	292

- All species/taxa are paired with each other by common chemical within single database
- Geometric means for more than one toxicity value per species/chemical pair
- Excluded species/chemical pairs where the minimum and maximum tox values were greater than 10-fold
- Three or more common chemicals per pair are required to develop a model
- Only models with a significant relationship (p-value < 0.05) are included in Web-ICE

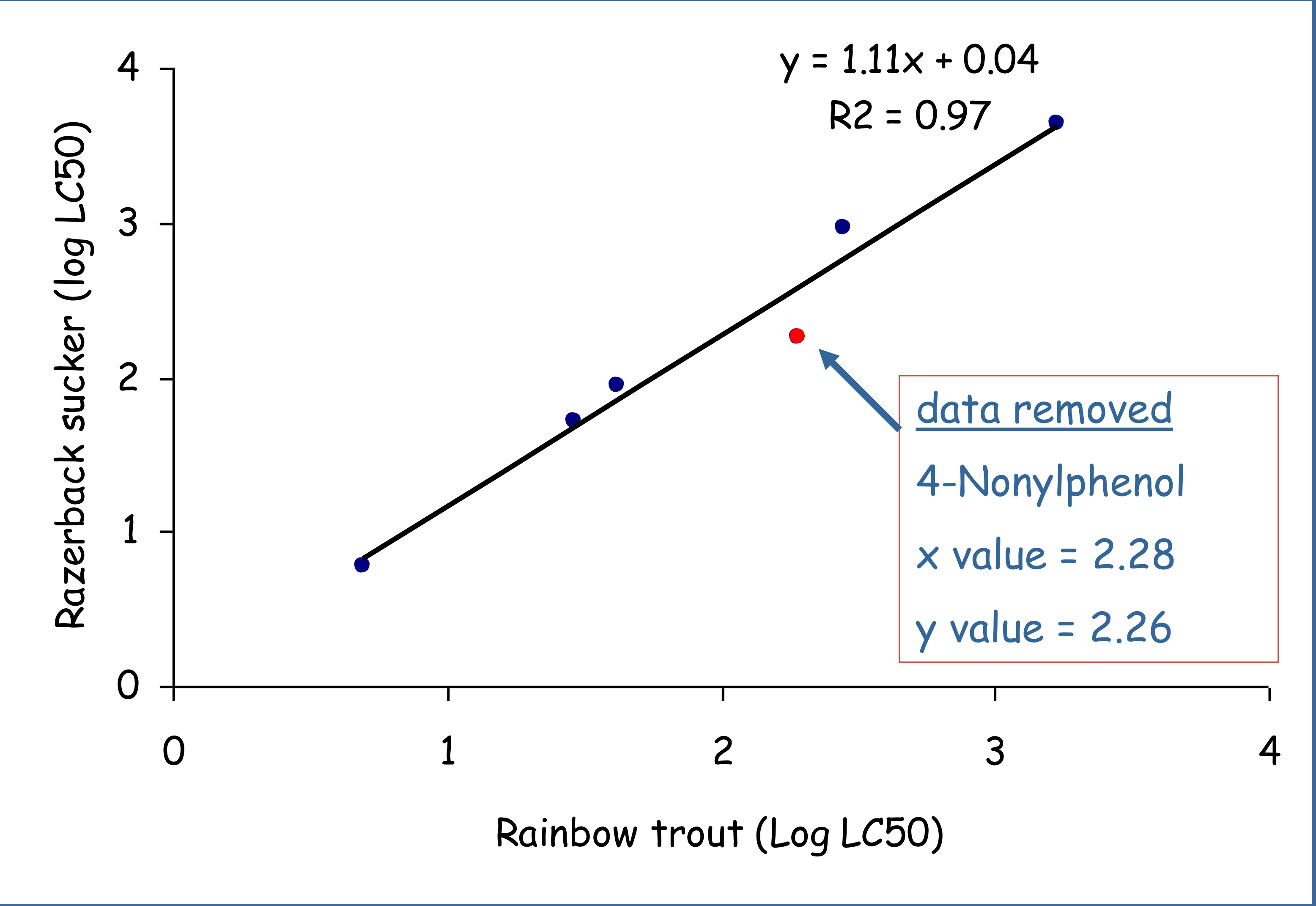
Model Validation & Uncertainty Analysis

- Only included models significant at the $p<0.05$ level
- Leave-1-out cross-validation performed on models $N \geq 4$

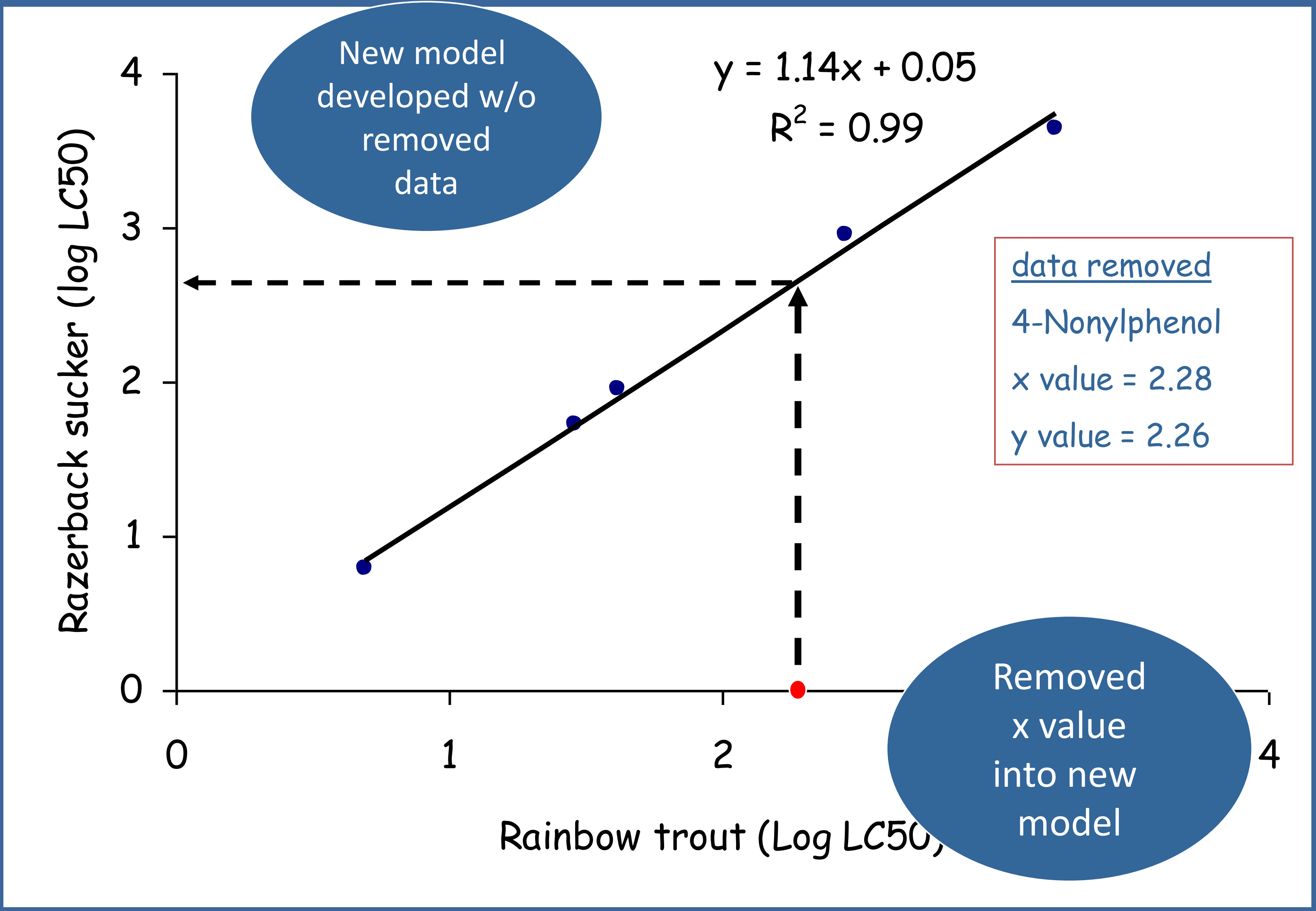
<u>leave-1-out cross-validation</u>

Each data point is removed, one at a time, and the model is rebuilt with remaining data. Removed surrogate data are used to estimate removed predicted data from rebuilt model.

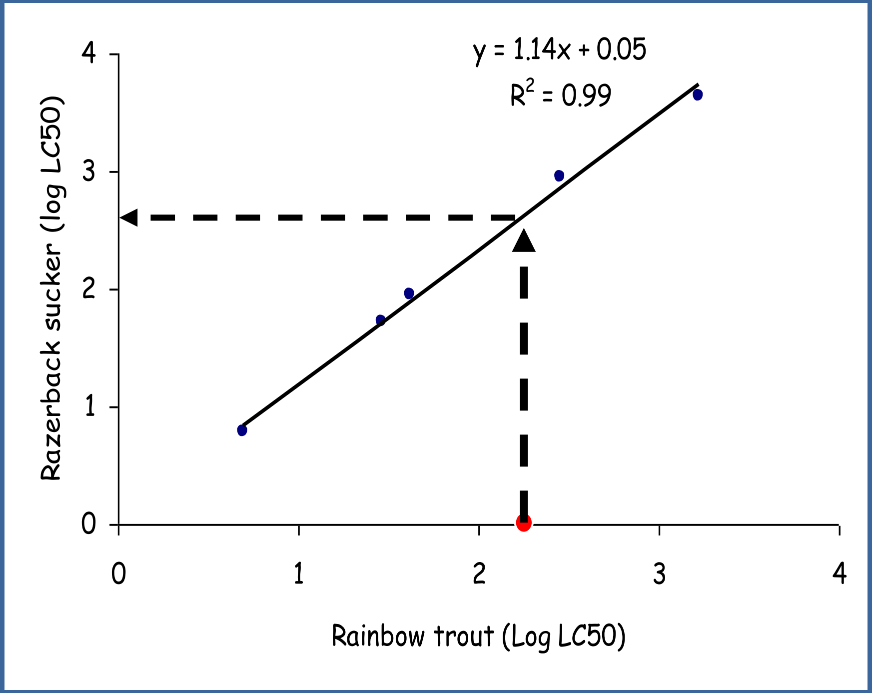
Model Validation: Leave-1-Out



Model Validation: Leave-1-Out



Model Validation: Leave-1-Out



<u>data removed</u>

4-Nonylphenol

x value = 2.28

y value = 2.26

New model predicts 2.64

compare to actual value (2.26)

"fold difference" of untransformed data=

10^(2.64)/10^(2.26)= 2.46

*** "fold diff" is maximum of predicted/actual or actual/predicted

Model Uncertainty Analyses

- Used “fold difference” of cross-validation to identify areas of model uncertainty
- Species level models
  - <u>Aquatic:</u> 10914 data points removed and validated from 696 models
  - <u>Wildlife:</u> 11846 data points from 538 models
- Model “Cross-validation success rate” = % removed data within 5-fold
- Taxonomic Distance
- Model parameters
- Mode of Action

Taxonomic Distance

Measures the taxonomic relatedness of the surrogate and predicted taxa

<u>Surrogate:</u>
Rainbow trout

Taxonomic distance	– shared taxonomic level	
1 – genus	Cutthroat trout (*Oncorynchus clarki*)	
2 - family	Atlantic salmon (*Salmo salar*)	
3 - order		
4 - class	Sheepshead minnow (*Cyprinodon variegatus*)	
5 - phylum	S. leopard frog (*Rana sphenocephala*)	
6 - kingdom	Daphnid (Daphnia magna)	

Model uncertainty & taxonomic distance

Aquatic animals in same family ~ 90% within 5-fold, 95% within 10-fold

Shared taxonomic level	Significant models (N)	Percentage within predicted range			
		5-fold	10-fold	50-fold	> 50-fold
Genus (1)	254	94	2	3	1
Family (2)	700	91	5	3	1
Order (3)	208	86	10	3	0
Class (4)	4432	80	9	8	3
Phylum (5)	764	61	14	17	8
Kingdom (6)	4556	56	15	19	10

Raimondo et al. 2010 Environmental Science and Technology

Model uncertainty & taxonomic distance

Wildlife in same order ~ 90% within 5-fold, 97% within 10-fold

Percentage of all datapoints in cross-validation category

Common level	Number datapoints	5-fold	10-fold	50-fold	> 50 fold
Genus (1)	48	100	0	0	0
Family (2)	1452	92	6	2	0
Order (3)	2238	90	7	3	0.3
Class (4)	5706	85	10	5	0.2
Phylum (5)	2402	76	13	9	1.5

Raimondo et al. 2007 Environmental Science and Technology

Uncertainty and Model Parameters
Aquatic

Regression Tree:

Is “fold difference” predicted by:

intercept

slope

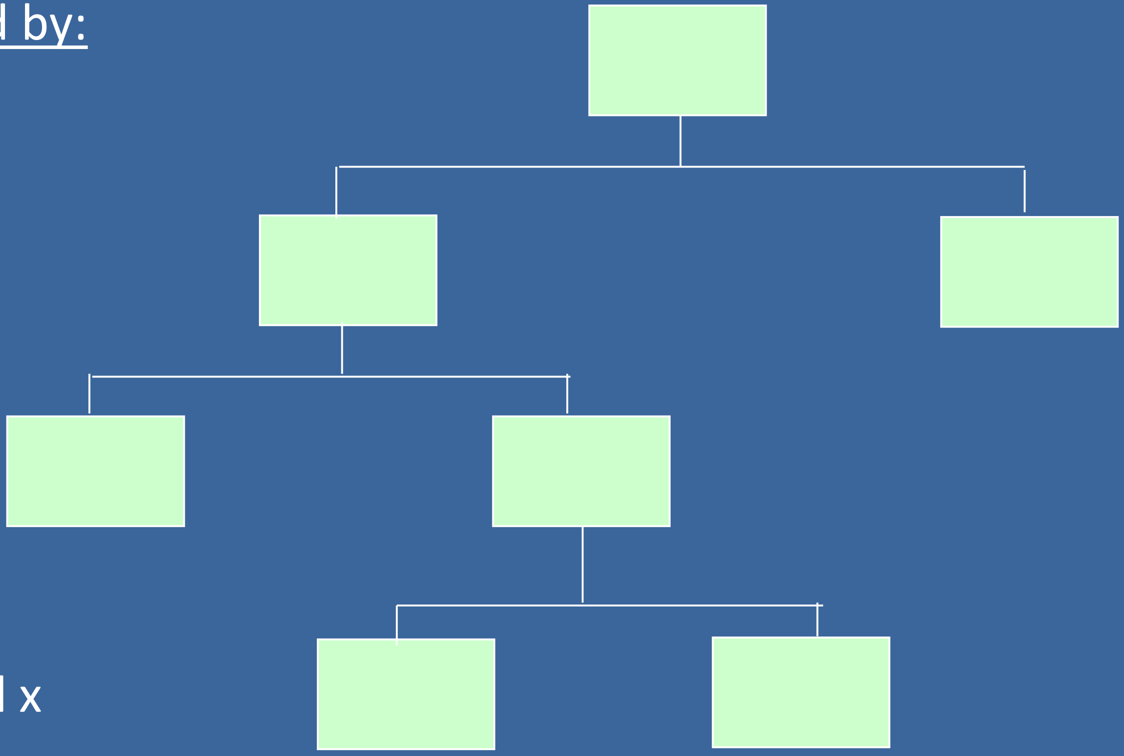
R^2

df

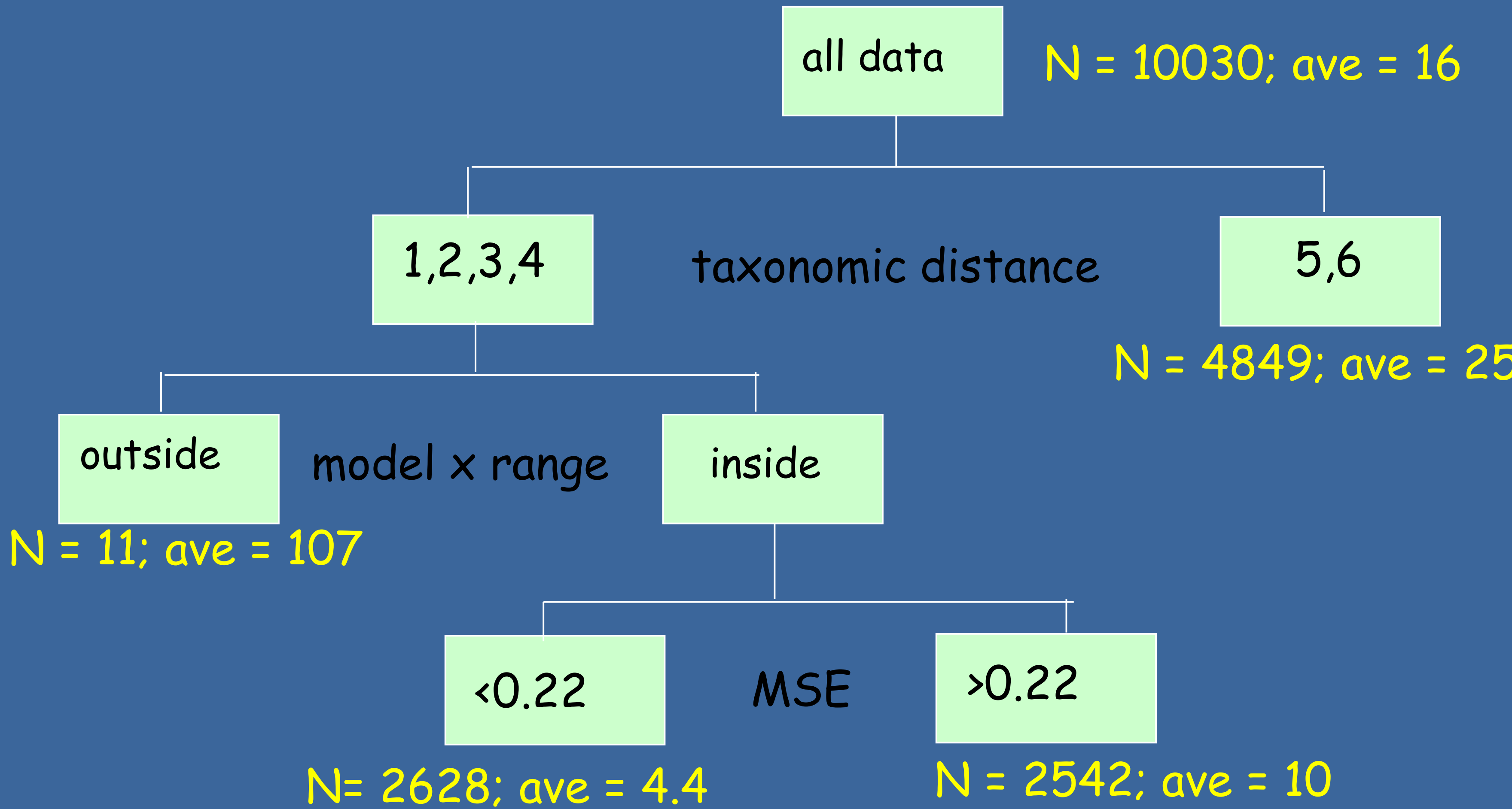
MSE

taxonomic distance

input value relative to model x range



Regression Tree - *Aquatic* (Decision Tree, Pattern Recognition)



Raimondo et al. 2010 Environmental Science and Technology

User Guidance:
Selecting a Model with Low Uncertainty

No one attribute defines model robustness!

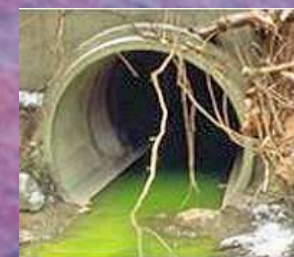
Rules of Thumb:

- low MSE (< 0.94) *(based on re-analysis)*
- high R^2 (> ~0.6)
- Close confidence intervals *(input value should be relatively close to the range of surrogate data used to create the model)*

Weight of Evidence

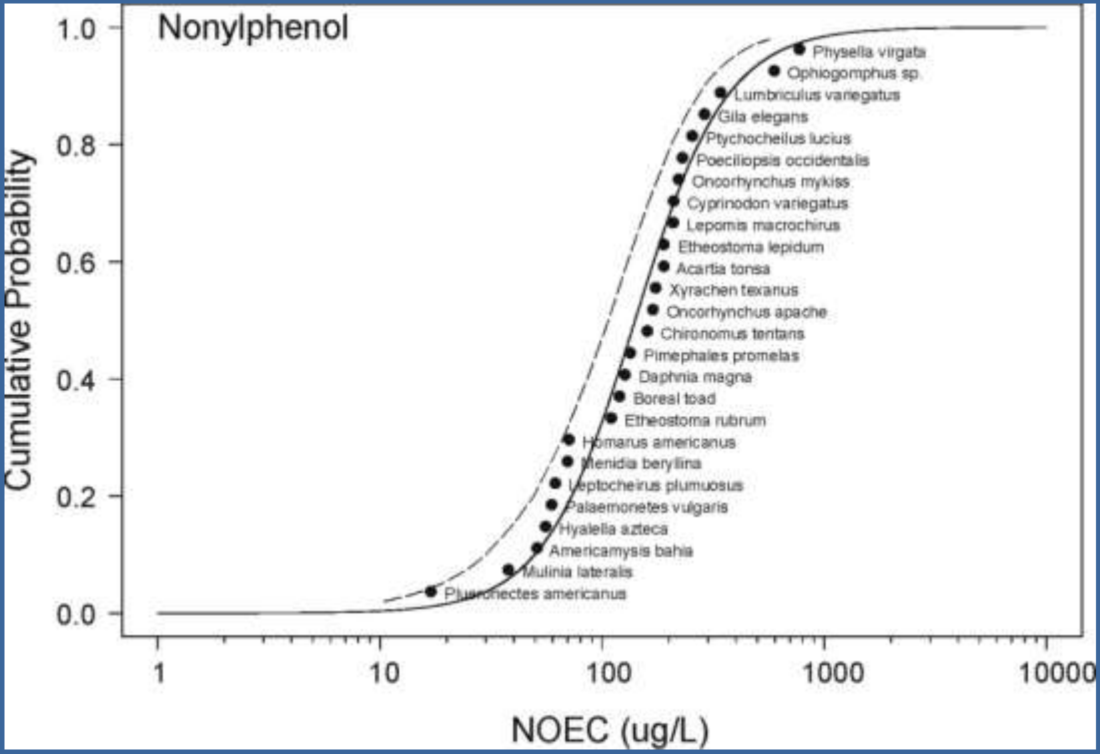
- high cross-validation success rate
- close taxonomic distance (within class)

Applications:
Ecological Risk Assessment

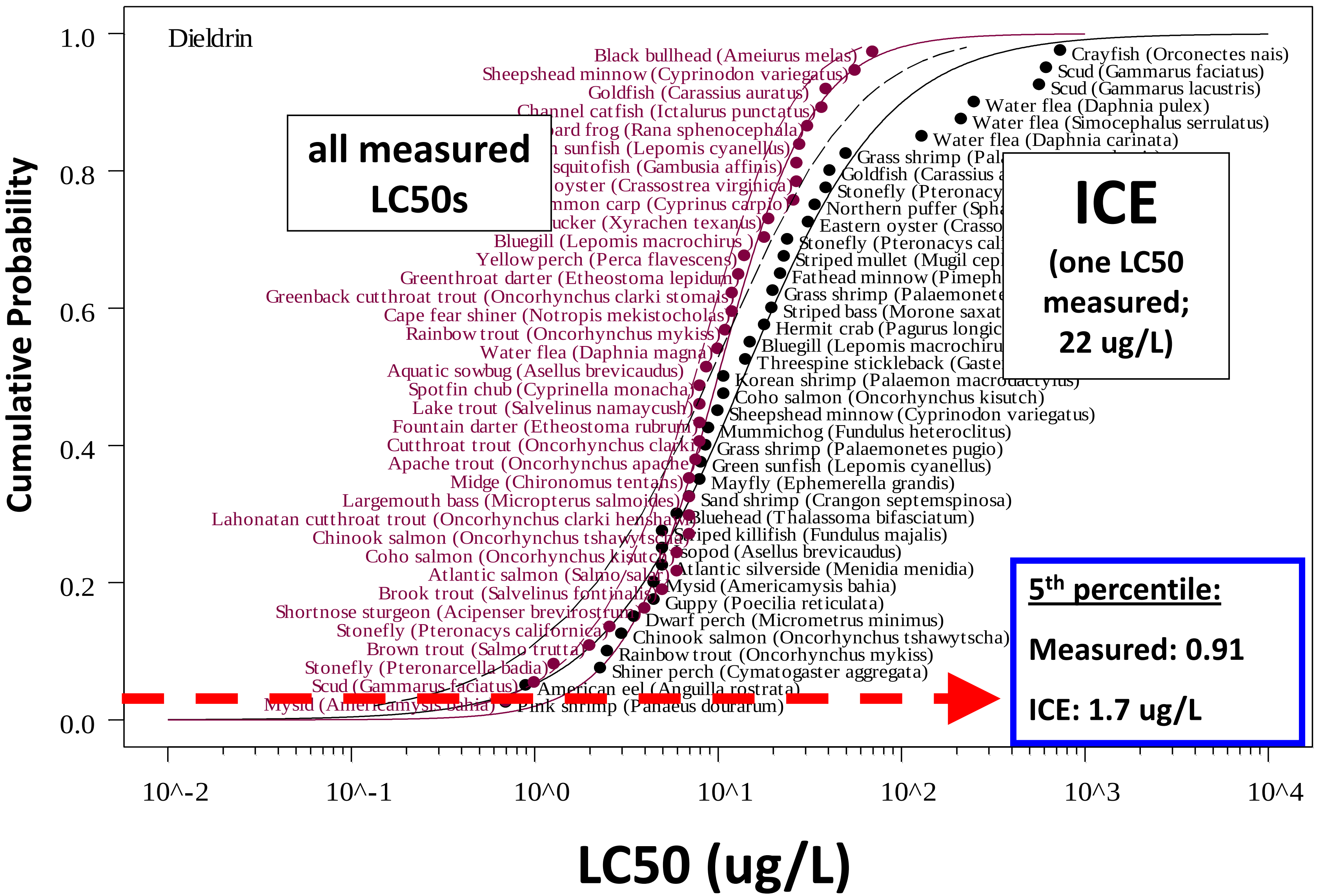


Applications:
Species Sensitivity Distributions

- Supplement datasets to populate SSDs
- Develop SSDs from single toxicity value
- link to QSAR
- ICE-generated SSDs are comparable to SSDs built from measured data



ICE and Measured SSDs



How Do ICE and Measured SSDs Compare for Aquatic Organisms?

- Aquatic Species Proof of Concept (Dyer et al. 2006)
  - ICE-based SSDs had HC5s within an order of magnitude of measured HC5s
- Aquatic Species, Expanded Study (Dyer et al. 2008)
  - 55 AWQC chemicals, 4 surrogate species (*P. promelas, O. mykiss, C. varigatus, D. magna*)
  - Using fish surrogate to predict fish and invertebrate surrogate to predict invertebrates yielded HC5s similar to measured (average factor of 3.0 ± 6.7 over 7 orders of toxic magnitude)

Environ. Sci. Technol. **2008**, *42*, 3076–3083

Comparison of Species Sensitivity Distributions Derived from Interspecies Correlation Models to Distributions used to Derive Water Quality Criteria

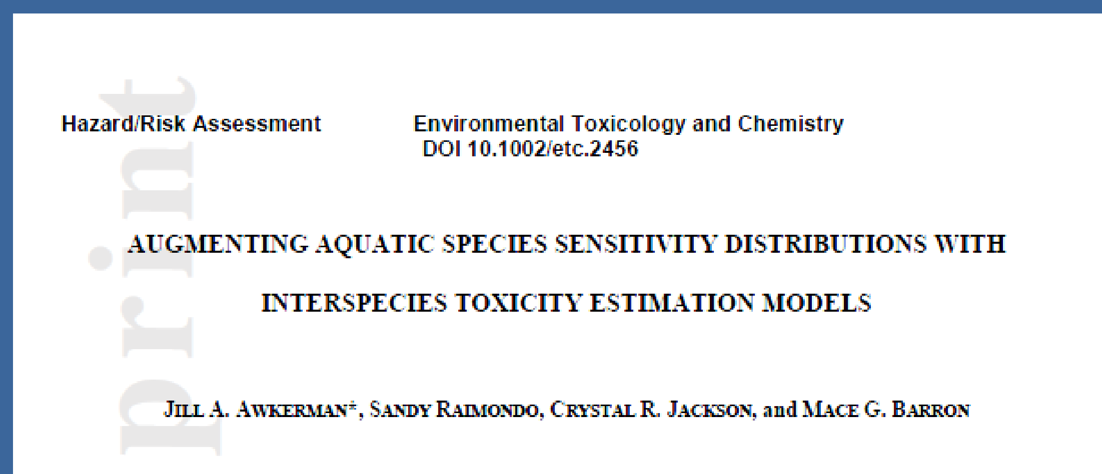
SCOTT D. DYER,*,† DONALD J. VERSTEEG,† SCOTT E. BELANGER,† JOEL G. CHANEY,† SANDY RAIMONDO,‡ AND MACE G. BARRON‡
The Procter and Gamble Company, 11810 East Miami River

Introduction

Ecological risk assessments typically require characterizing the effects of multiple chemicals on a diversity of ecological receptors using toxicity data for only a limited number of species. Regulatory activities such as REACH (Registration, Evaluation and Authorization of Chemicals ((*1, 2*); http://europa.eu.int/comm/environment/chemicals/reach.htm), ICCA (International Council of Chemical Associations) High Production Volume (HPV) Chemicals Challenge (http://www.icca-chem.org/), and Canada's Domestic Substance List (*3*) will also create new demands for toxicity data. Note that the ICCA HPV Chemicals challenge is in cooperation with OECD and its member countries which then also include other HPV challenge programs such as those of the United States Environmental Protection Agency (USEPA) and Japan.

Augmenting Datasets with ICE-Generated Data

- Limited measured data augmented with ICE toxicity values (49 chemicals with ≥ 10 species)
- SSD sensitivity analysis
- ICE prediction uncertainty contributed less to HC5 variability than species composition
- ICE prediction variability influence on HC5 dependent on location in SSD

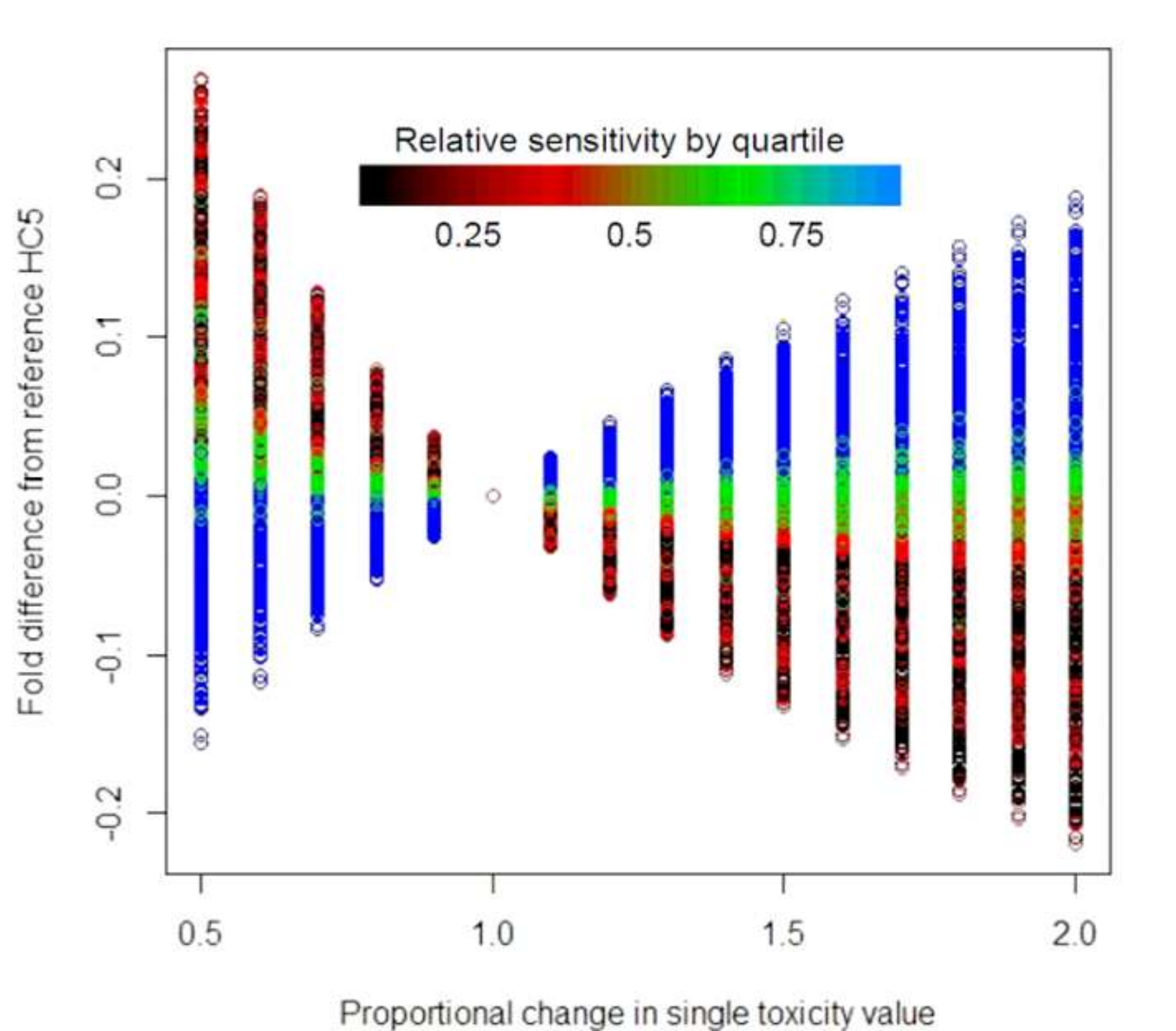

Hazard/Risk Assessment

Environmental Toxicology and Chemistry
DOI 10.1002/etc.2456

AUGMENTING AQUATIC SPECIES SENSITIVITY DISTRIBUTIONS WITH INTERSPECIES TOXICITY ESTIMATION MODELS

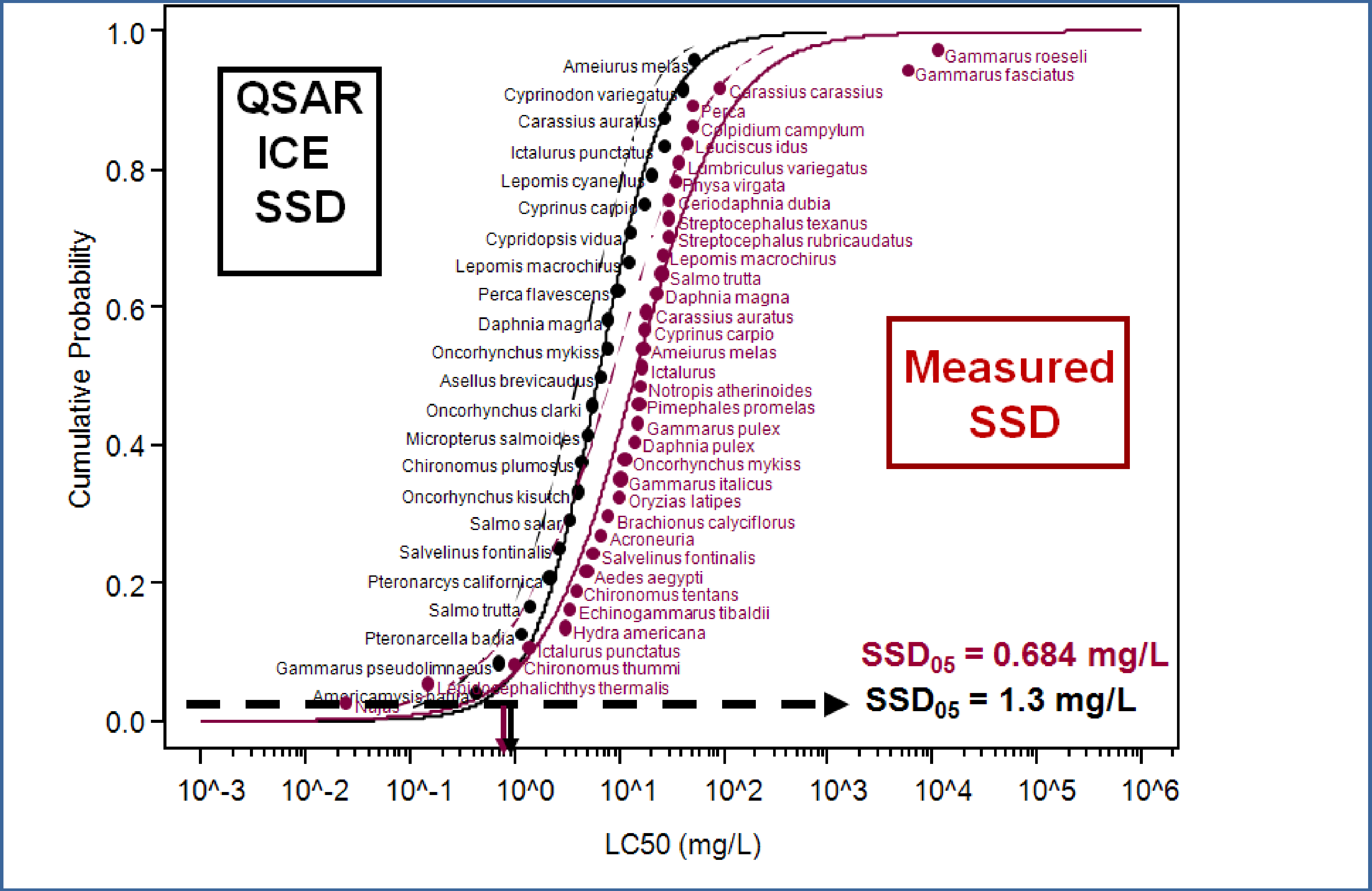
JILL A. AWKERMAN*, SANDY RAIMONDO, CRYSTAL R. JACKSON, and MACE G. BARRON

Influence of SSD data variability on HC5



Web-ICE and QSAR

QSAR tox estimate used as input into ICE SSD generator



QSAR → Web-ICE → SSD

- *in silico* approach
- 10 chems/4 MOAs; daphnid, fathead

- Compared measured SSD HC5 to in silico HC5
  - Results variability with MOA & QSAR uncertainty
  - Variability reduced with identical species composition
    - 80% HC5s within 5-fold of measured using ECOSAR QSAR value

Aquatic Toxicology 116–117 (2012) 1–7

Contents lists available at SciVerse ScienceDirect

Aquatic Toxicology

journal homepage: www.elsevier.com/locate/aquatox

Evaluation of in silico development of aquatic toxicity species sensitivity distributions

Mace G. Barron*, Crystal R. Jackson, Jill A. Awkerman

U.S. EPA, GED, 1 Sabine Island Drive, Gulf Breeze, FL 32561, USA

How Do ICE and Measured SSDs Compare for Wildlife?

- Used ICE database of acute toxicity (23 chemicals with > 15 species)
- For each chemical, removed records from database to generate independent ICE models
- Created SSDs
    1. measured data
    2. ICE data – 5 surrogates (Japanese Quail, Mallard, Northern bobwhite, red-winged blackbird, Norway rat)
- 100% HC5s within a factor of 5 using bird & mammal surrogates

Environ. Sci. Technol. **2008**, *42*, 3447–3452

Development of Species Sensitivity Distributions for Wildlife using Interspecies Toxicity Correlation Models

JILL A. AWKERMAN,* SANDY RAIMONDO, AND MACE G. BARRON

U.S. Environmental Protection Agency, Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, Florida 32561

species with known uncertainty (2). In practice, a least-squares model II regression is developed from the known LD50 values of two species tested for the same group of chemicals. The model is then used to approximate toxicity of one species to a chemical from the known toxicity of the other species (the surrogate). ICE models for wildlife have been determined to be most accurate for two species within the same order, but may be used to estimate toxicity from birds to mammals (2). Confidence intervals calculated with ICE estimates are based on the linear relationship and mean model error and may be used to assess the robustness of model predictions. In aquatic species, acute toxicity values generated from ICE models have been used to populate SSDs and have been recommended for generating reasonable

Can we use only rodent data and ICE to develop reliable wildlife SSDs?

- Database of rodent oral dose LD50 data (Hazardous Substances Data Bank)
- Mouse: 109 chemicals; Rat: 175 chemicals
  - Rodents most often in least sensitive quartile; not great surrogates
- Created measured and surrogate ICE SSDs:
  - rat (53 chems) and mouse (32 chems)
- Compared HD5s
  - 78% of HD5 values within 5x of measured


Journal of Toxicology and Environmental Health, Part A, 72: 1604–1609, 2009
ISSN: 1528-7394 print / 1087-2620 online
DOI: 10.1080/15287390903232491

Taylor & Francis

Estimation of Wildlife Hazard Levels Using Interspecies Correlation Models and Standard Laboratory Rodent Toxicity Data

Jill A. Awkerman, Sandy Raimondo, and Mace G. Barron
U. S. Environmental Protection Agency, Gulf Ecology Division, Gulf Breeze, Florida, USA

Application of Web-ICE

Defensibility

- cross-validation of models provides user with estimate of model performance
- all models significant

User guidance

- clearly outlined
- uncertainty analyses identify “Rules of Thumb” (Aquatic complete in 2008)
- Web-ICE alerts users to potentially ill-fitting data

Demonstration of use in SSDs

- Aquatic and wildlife
- Species composition
- Taxonomic distance of surrogate and predicted

EPA
United States
Environmental Protection
Agency

Web ICE

http://www.epa.gov/ceampubl/fchain/webice/