

STATISTICAL TOOLS TO EVALUATE SPECIES SENSITIVITY DISTRIBUTIONS AND CALCULATE FINAL ACUTE AND CHRONIC VALUES

Michael Bock, PH.D

136 Commercial Street, Suite 401

Portland, ME 04105

207-517-8223

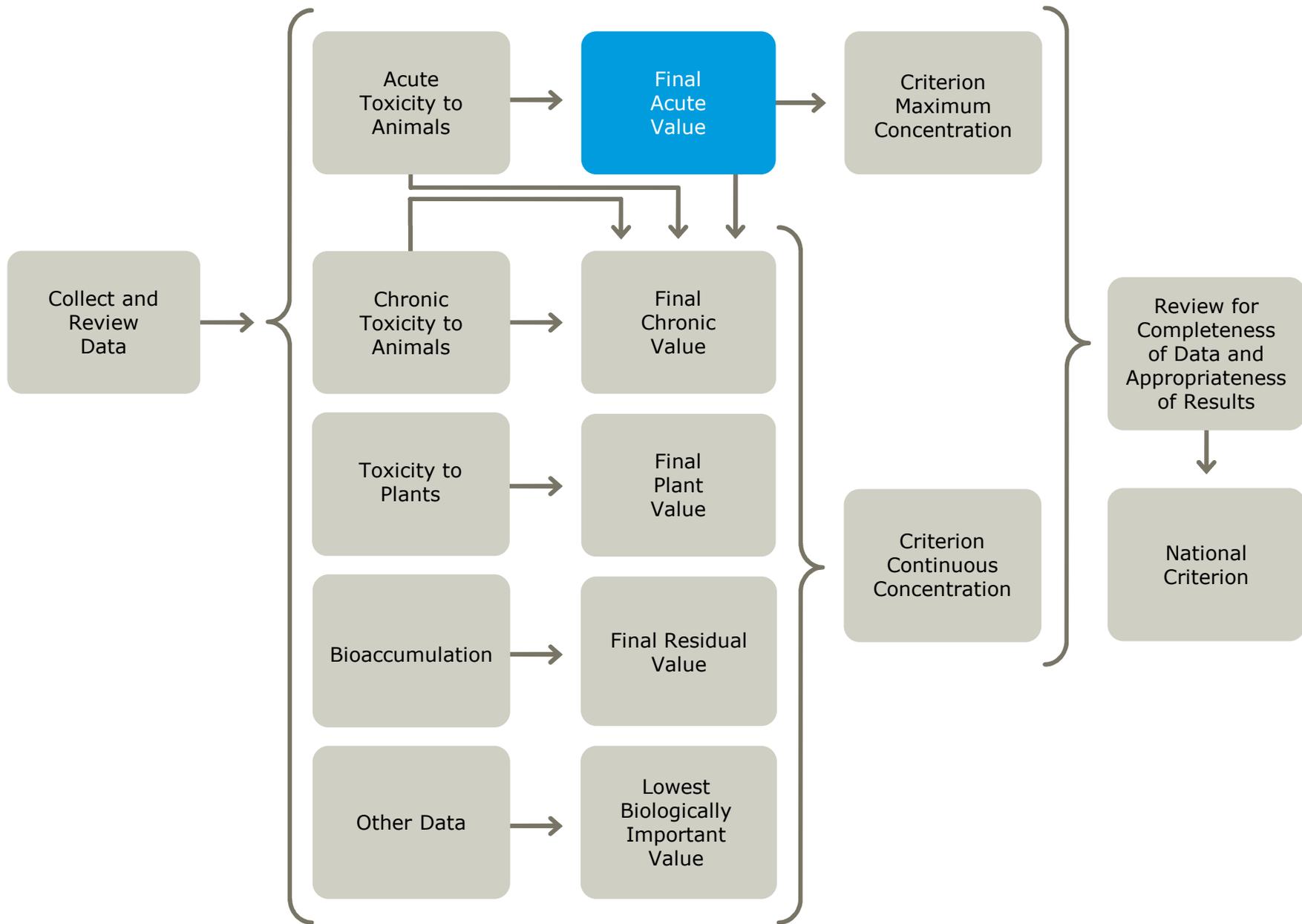
mbock@environcorp.com



Copyright © 2015 Ramboll Environ, Inc. All rights reserved. Copies of this PowerPoint are being distributed for your personal, non-commercial use.

If you wish to cite to the paper, please include full attribution.

For any other use, please contact Ramboll Environ, Inc. by telephone at 207-517-8223 or e-mail mbock@environcorp.com for permission.



FAV DEFINED

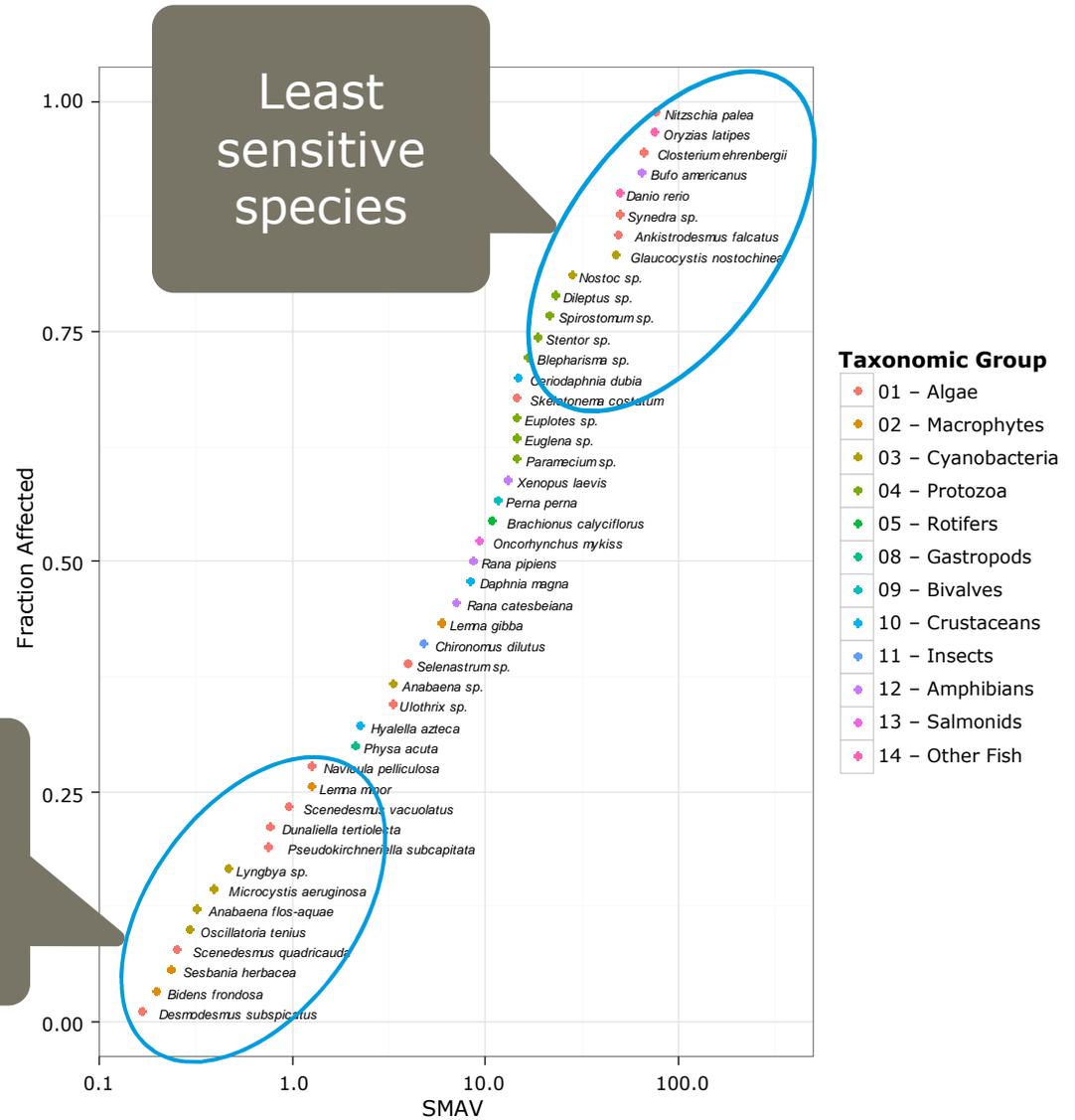
IV. Final Acute Value

Appropriate measures of acute (short-term) toxicity of the material to a variety of species of aquatic animals are used to calculate the Final Acute Value. The Final Acute Value is an estimate of the concentration of the material corresponding to a cumulative probability of 0.05 in the acute toxicity values for the genera with which acceptable acute tests have been conducted on the material. However, in some cases, if the Species Mean Acute Value of a commercially or recreationally important species is lower than the calculated Final Acute Value, then that Species Mean Acute Value replaces the calculated Final Acute Value in order to provide protection for that important species.

FAV is equivalent to an HC₅ from a Species Sensitivity Distribution (SSD)

SPECIES SENSITIVITY DISTRIBUTIONS

Species are ranked from lowest to highest SMAV and plotted against fraction affected



CURRENT APPROACH TO THE FAV (ERICKSON AND STEPHAN 1988)

- Assumed the GMAV follow a log-triangular distribution
- Compile GMAVs and order from low to high
- Assign ranks to GMAVs and calculate cumulative probability $P=R/(N+1)$
- Select the four GMAVs with P closest to 0.05 (often 4 lowest GMAVs)

$$S^2 = \frac{\sum ((\ln GMAV)^2) - ((\sum \ln GMAV))^2 / 4}{\sum (F) - ((\sum (\sqrt{P}))^2 / 4)}$$

$$L = (\sum (\ln GMAV) - S(\sum (\sqrt{P}))) / 4$$

$$A = S(\sqrt{0.05}) + L$$

$$FAV = e^A$$

LIMITATIONS OF THE CURRENT APPROACH

- The assumption of the log-triangular distribution is not verified for each dataset
- The assumed log-triangular distribution is not supported by recent research
- The current approach heavily weights 4 data points
- Advances in distribution fitting methods justify a re-evaluation

ESTIMATING THE SSD AND THE HC₅

- Distribution fitting using regression models (bivariate calculation)
- Distribution fitting using MVUE and MLE methods (univariate calculation)
- Distribution fitting using Bayesian methods (univariate calculation)
- Non-parametric methods (univariate calculation)

REGRESSION MODELS

-  Fit GMAV versus fraction affected (percentiles)
-  Check goodness of fit versus the regression models
-  Choose the regression model based on GOF and calculate HC₅
-  **How to calculate fraction affected / percentiles?**

REGRESSION MODELS



There are many methods for calculating the percentiles

Six common methods shown

m = fraction affected

k = rank

n = sample size

These methods converge at large sample sizes (**n** > 100)

$$m = k/n$$

$$m = (k - 0.5)/n$$

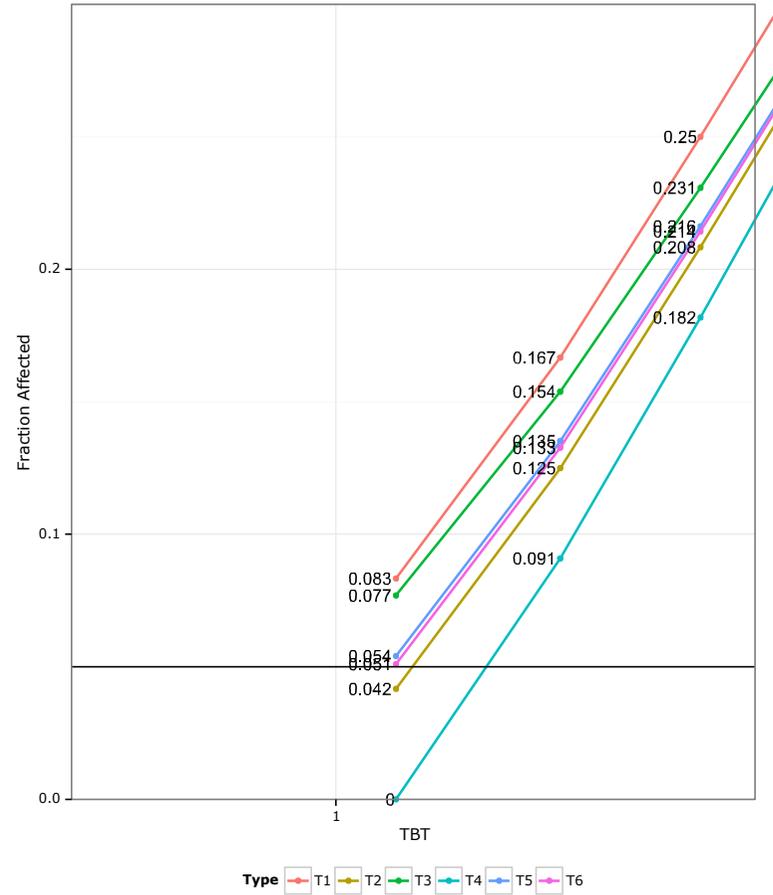
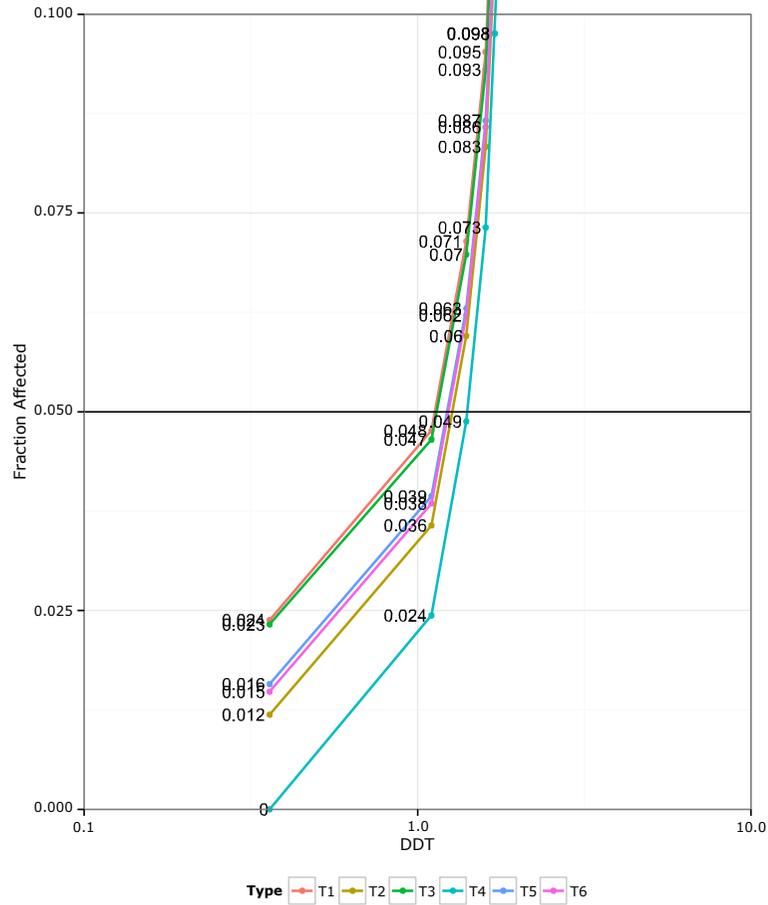
$$m = k/(n+1)$$

$$m = (k-1)/(n-1)$$

$$m = (k - 1/3)/(n + 1/3)$$

$$m = (k - 3/8)/(n+1/4)$$

THE FRACTION AFFECTED PROBLEM (DDT/TBT)



REGRESSION BASED TOOLS

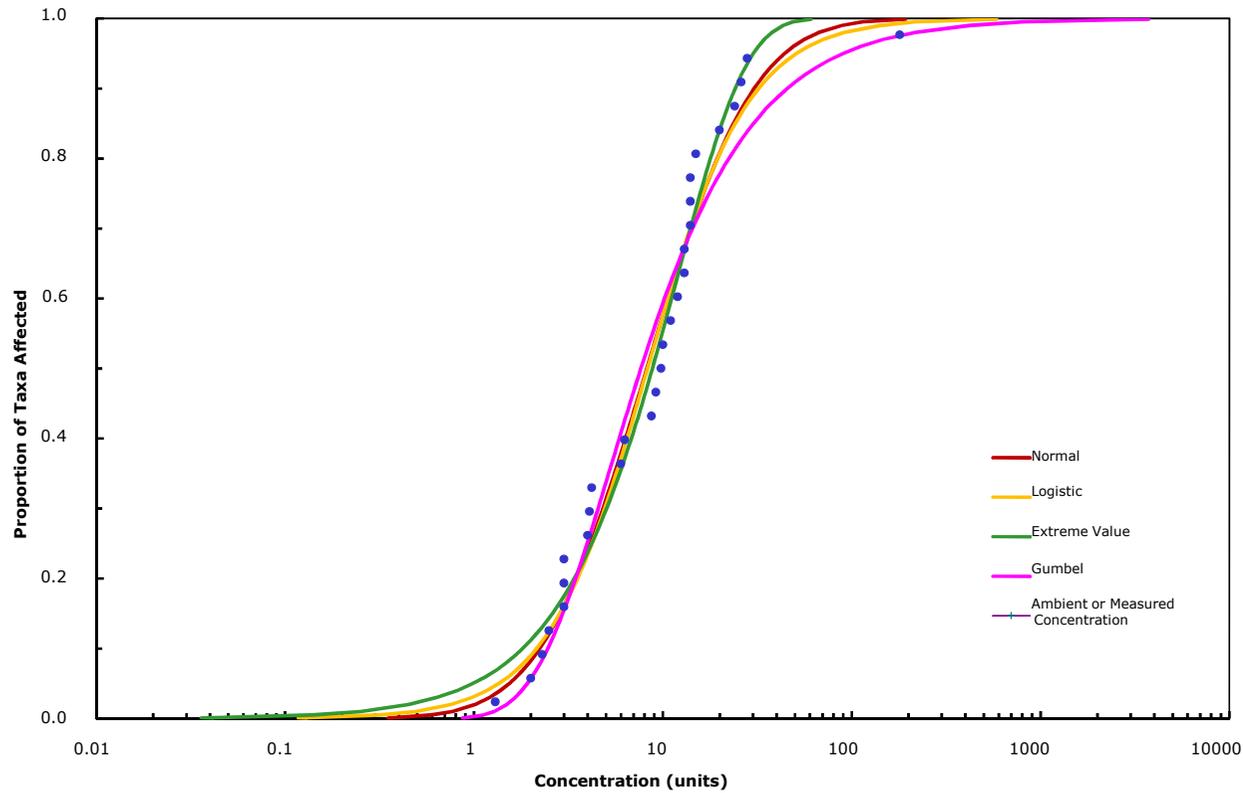
USEPA SSD Generator

- Provides one method for fraction affected
- Provides a single distribution (log-probit)

Environment Canada SSD Master

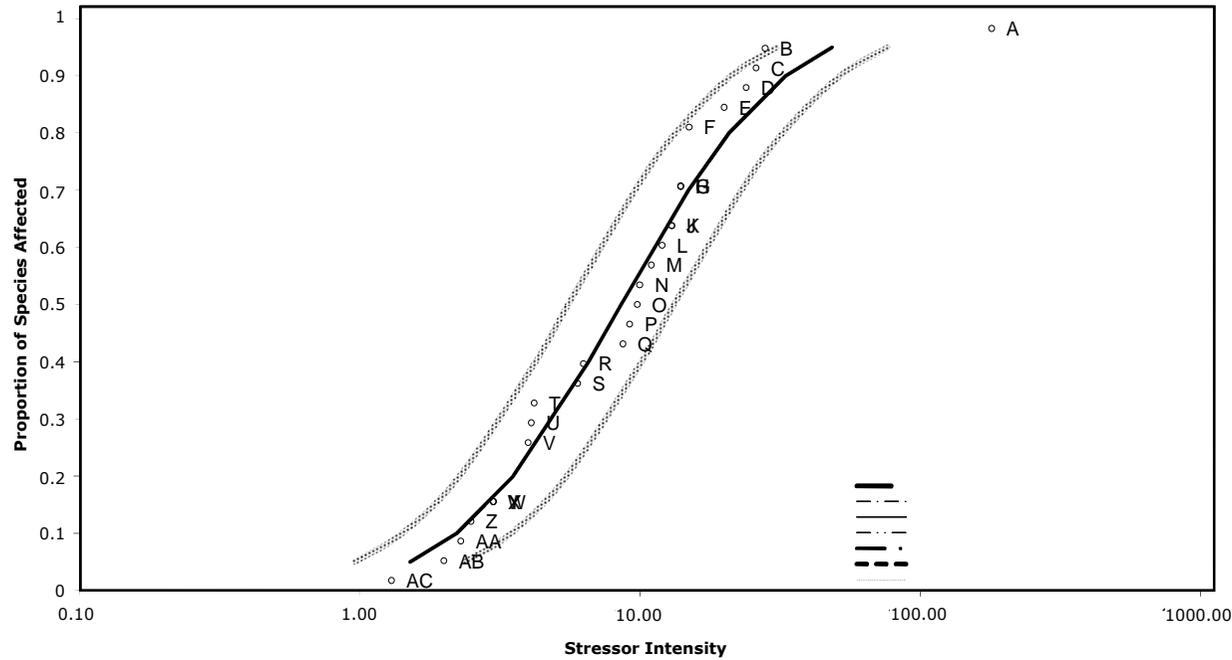
- Two methods for fraction affected
- Five distributions (log-normal, log-logistic, extreme value, Weibull, Gumbell)

SSD MASTER EXAMPLE (TOXAPHENE)



	Filben HC ₅	Hazen HC ₅
Log-normal	1.53	1.58
Log-logistic	1.35	1.40
Extreme Value	0.98	1.03
Gumbel	1.91	1.96
Weibull	NC	NC

SSD GENERATOR EXAMPLE (TOXAPHENE)



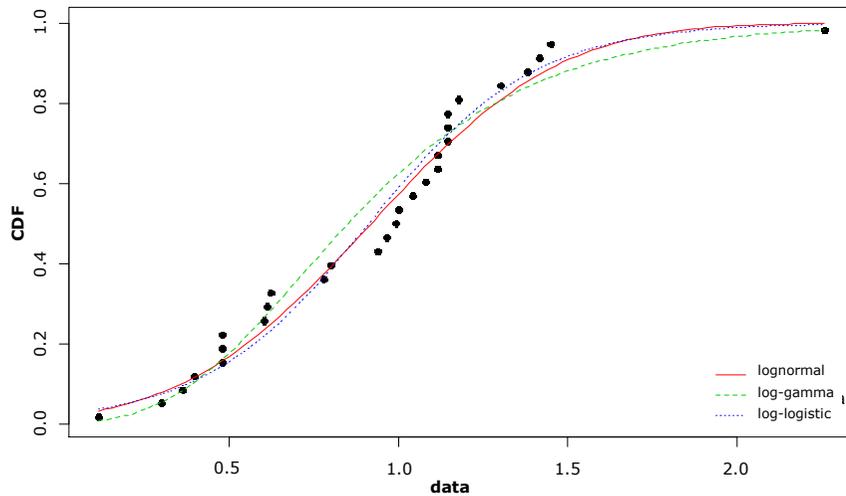
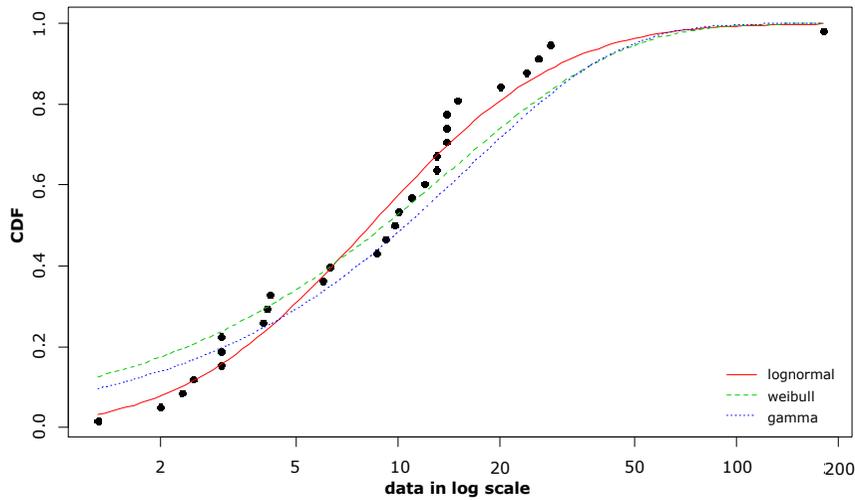
	HC ₅
Log-probit	1.514

DISTRIBUTION FITTING TOOLS

- Distribution fitting is included in most statistical packages
- Two fitting methods
 - Maximum Likelihood Estimator (MLE)
 - More robust to outliers and generally preferred
 - Minimum Variance Unbiased Estimator (MVUE)
 - Many distributions are available (e.g., normal, lognormal, logistic, etc.)

MLE DISTRIBUTION FITTING (TOXAPHENE)

EMPIRICAL AND THEORETICAL CDFS



	HC ₅
Log-normal	1.59
Weibull	0.42
Gamma	0.535
Log-gamma	1.957
Log-logistic	1.567

OTHER METHODS

Aldenberg et al. 2002

- Use Bayesian methods to derive the coefficients used to calculate the HC₅ assuming a lognormal distribution

Traditional Bayesian distribution fitting

- Many distributions are available (e.g., normal, lognormal, logistic, etc.)

Newman et al 2000, modified as suggested by Erickson and Stephan 2001

- Uses a bootstrap analysis to calculate a robust HC₅ by resampling the GMAV data
- Can be used when the data does not follow a discernible distribution

RESULTS

	Family	Toxaphene	DDT	DDT (n=15)
Stephan et al 1988 (log-triangular)	Modified Regression	1.88	0.86	2.28
SSD Master	Regression			
Log-normal		1.58	0.53	0.82
Log-logistic		1.40	0.43	0.64
Extreme Value		1.03	0.18	0.32
Log-Gumbell		1.96	0.94	1.35
SSD Generator	Regression	1.514	0.588	1.186
MLE	Distribution			
Log-normal		1.68	0.63	1.44
Weibull		0.503	0.09	0.64
Gamma		0.727	0.04	0.66
Log-gamma		2.015	1.02	1.87
Log-logistic		1.631	0.53	1.22
Aldenberg	Bayesian	1.52	0.57	1.20
Modified Newman	Non-parametric	2.2	0.88	1.84

CONCLUSIONS

Erickson and Stephan (1988) is not the optimal tool to calculate FAVs

- Newer tools provide robust and comparable results
- This tool can not be used to calculate other percentiles leading to the use of different methods for different purposes (e.g. risk assessment, NRDA, etc.)

Regression based tools have significant issues

- Convert a univariate problem to a multivariate problem
- Sensitive to how percentiles are calculated, especially at small sample sizes
- The tools are often unstable

Distribution based tools perform well

- Many different distributions can be tested
- When combined with graphical analysis, the best fit can be easily selected

Modified Newman's method may be best for some data sets

- No distribution is assumed
- Robust to outliers

PROPOSED APPROACH



Comprehensive data evaluation to test for:

1. Representativeness
2. Outliers



Run univariate distribution fitting

1. Evaluate goodness of fit statistics
2. Evaluated distribution plots for:
 - Skewness
 - Fit, especially at the lower tail



If goodness of fit is poor, use Newman's non-parametric method

THANK YOU