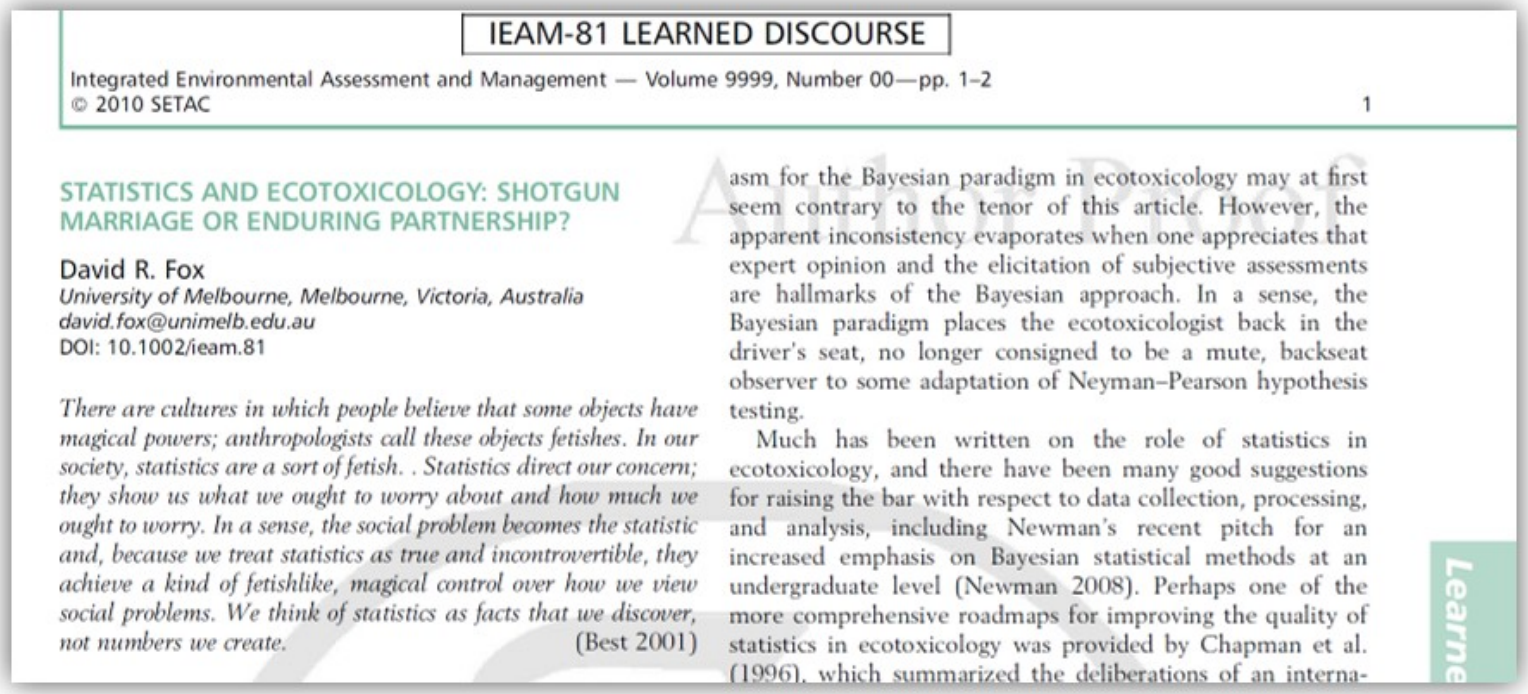
Statistical Ecotoxicology

Failure to Launch?

Prof. David Fox

Revitalising the Marriage



IEAM-81 LEARNED DISCOURSE

Integrated Environmental Assessment and Management — Volume 9999, Number 00—pp. 1–2

STATISTICS AND ECOTOXICOLOGY: SHOTGUN MARRIAGE OR ENDURING PARTNERSHIP?

David R. Fox
University of Melbourne, Melbourne, Victoria, Australia
david.fox@unimelb.edu.au
DOI: 10.1002/ieam.81

There are cultures in which people believe that some objects have magical powers; anthropologists call these objects fetishes. In our society, statistics are a sort of fetish. . Statistics direct our concern; they show us what we ought to worry about and how much we ought to worry. In a sense, the social problem becomes the statistic and, because we treat statistics as true and incontrovertible, they achieve a kind of fetishlike, magical control over how we view social problems. We think of statistics as facts that we discover, not numbers we create. (Best 2001)

asm for the Bayesian paradigm in ecotoxicology may at first seem contrary to the tenor of this article. However, the apparent inconsistency evaporates when one appreciates that expert opinion and the elicitation of subjective assessments are hallmarks of the Bayesian approach. In a sense, the Bayesian paradigm places the ecotoxicologist back in the driver's seat, no longer consigned to be a mute, backseat observer to some adaptation of Neyman–Pearson hypothesis testing.

Much has been written on the role of statistics in ecotoxicology, and there have been many good suggestions for raising the bar with respect to data collection, processing, and analysis, including Newman's recent pitch for an increased emphasis on Bayesian statistical methods at an undergraduate level (Newman 2008). Perhaps one of the more comprehensive roadmaps for improving the quality of statistics in ecotoxicology was provided by Chapman et al. (1996), which summarized the deliberations of an interna-

Statistics - Struggling for a 'place in the sun'?

The field of ecotoxicology includes concepts arising from:

disciplines such as toxicology

biology

analytical chemistry

environmental chemistry

organic chemistry

physiology

ecology, genetics

microbiology

biochemistry

immunology

molecular biology

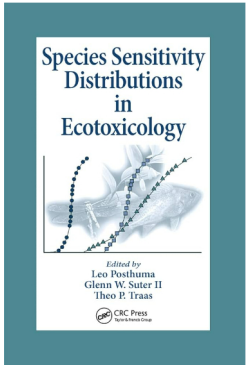
soil sciences

water sciences

air sciences

economics

Statistics - The cornerstone of SSD Modelling



Suppose RVs $X1$ and $X2$ both take values x. We are interested in the probability of $X1$ exceeding $X2$, that is, $\Pr(X1 > X2)$, or equivalently $\Pr(X1 - X2 > 0)$. Thus, we consider a new RV: $Z = X1 - X2$ for the difference, taking values z, and require: $\mathrm{EXF}_{X1-X2}(0)$.

The analytical derivation of the probability of failure integrals closely follows Papoulis (1965: p. 189, see his figure 7-2), Mood et al. (1974: pp. 185–186), or Hsu (1997: p. 137). The difference of a pair of values x_1 and x_2 exceeds value z, if $x_2 < x_1 - z$. Consequently, we have to sum (integrate) the *joint probability* of x_1 and x_2 over all values satisfying this inequality:

$$\mathrm{EXF}_{X1-X2}(z) = \iint_{x1-x2>z} \mathrm{PDF}_{X1,X2}(x_1,x_2)dx_1dx_2 = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{x1-z} \mathrm{PDF}_{X1,X2}(x_1,x_2)dx_2\right]dx_1$$

We now assume that the RVs of $X1$ and $X2$ are *independent*:

$$\mathrm{PDF}_{X1,X2}(x_1,x_2) = \mathrm{PDF}_{X1}(x_1)\cdot\mathrm{PDF}_{X2}(x_2)$$

that is, the joint probability density factorizes into the univariate PDFs.

It follows that

$$\mathrm{EXF}_{X1-X2}(z) = \int_{-\infty}^{\infty}\mathrm{PDF}_{X1}(x_1)\cdot\left[\int_{-\infty}^{x1-z}\mathrm{PDF}_{X2}(x_2)dx_2\right]dx_1$$

$$= \int_{-\infty}^{\infty}\mathrm{PDF}_{X1}(x)\cdot\mathrm{CDF}_{X2}(x-z)dx$$

The required exceedence of $Z = X1 - X2$ at $z = 0$ equals

$$\Pr(X1 > X2) = \mathrm{EXF}_{X1-X2}(0) = \int_{-\infty}^{\infty}\mathrm{PDF}_{X1}(x)\cdot\mathrm{CDF}_{X2}(x)dx \qquad (5.5)$$

An alternative expression for this exceedence can be derived as follows. Instead of integrating $\mathrm{PDF}_{X1,X2}(x_1,x_2)$ over the region $x_2 < x_1 - z$ for pairs of values x_1 and x_2 given z, one could have done the double integration over the region $x_1 > x_2 + z$:

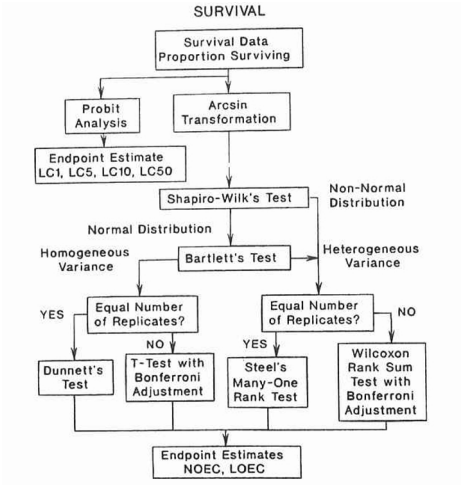
$$\mathrm{EXF}_{X1-X2}(z) = \int_{-\infty}^{\infty}\left[\int_{x2+z}^{\infty}\mathrm{PDF}_{X1,X2}(x_1,x_2)dx_1\right]dx_2 = \int_{-\infty}^{\infty}\mathrm{EXF}_{X1}(x+z)\cdot\mathrm{PDF}_{X2}(x)dx$$

The exceedence of $Z = X1 - X2$ at 0 equals:

$$\Pr(X1 > X2) = \mathrm{EXF}_{X1-X2}(0) = \int_{-\infty}^{\infty}\left(1-\mathrm{CDF}_{X1}(x)\right)\cdot\mathrm{PDF}_{X2}(x)dx \qquad (5.6)$$

A Protocol Statistical Analaysis of Fathead Minow Larval Survival and Growth TestA Protocol Statistical Analaysis of Fathead Minow Larval Survival and Growth Test

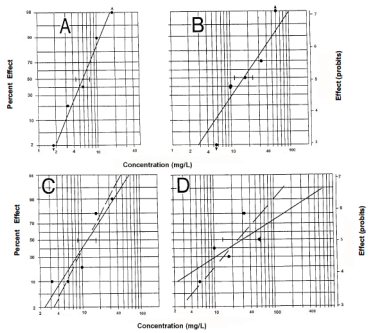
Source: Weber et al. (1989)



A Protocol Statistical Analaysis of Fathead Minow Larval Survival and Growth Test

Source: Weber et al. (1989)

Outdated advice: Example #1 - Canadian guidance document shows how to use graph paper to fit a probit model!



Outdated advice: Example #2 - OECD guidance document 54 recommends transforming data instead of using a more appropriate statistical modelling framweork.

4.3.3. Transformation of data

65. Many standard parametric methods (e.g. ANOVA, t-tests, linear regression analysis) assume normally distributed data and homogenous variances. In practice, the data often deviate from these assumptions, and if so, the inferences resulting from these standard methods may be disturbed. A variance-stabilising transformation is often applied to the data, and then the statistical analysis is performed on the transformed data. Examination of residual plots (plot of the residuals vs. the predicted values) and tests of heterogeneity of variance (e.g., Levene, Bartlett, Hartley's F-max, or Cochran's Q) can help to identify instances when the variances among the concentration groups are unequal. References on this topic include Box and Cox (1964), Box and Hill (1974), Box and Tidwell (1962), Draper and Cox (1969).

68. If a transformation is used, it is also necessary to back-transform the means and confidence intervals to the original scale, when reporting results. It is not correct to back-transform the standard errors. It is important to understand that the back-transformed means differ from the arithmetic means of the original data. These back-transformed means should be interpreted as estimates of the median of the underlying data distribution, if the transformed data are normally (or at least symmetrically) distributed. In the special case of a log-transformation, the back-transformed mean is the geometric mean of the original data, and this value estimates the median of the underlying lognormal distribution.

The statistical grenade

Sometimes the ‘best’ advice may be: ill-conceived, wrong, intuitively appealing but statistically reckless

"Ban Statistical Significance"

2 Don't Say "Statistically Significant"

The *ASA Statement on P-Values and Statistical Significance* stopped just short of recommending that declarations of "statistical significance" be abandoned. We take that step here. We conclude, based on our review of the articles in this special issue and the broader literature, that it is time to stop using the term "statistically significant" entirely. Nor should variants such as "significantly different," "$p < 0.05$," and "nonsignificant" survive, whether expressed in words, by asterisks in a table, or in some other way.

SPECIAL ISSUE ON P-VALUE

Statistical Inference in the 21st Century: A World Beyond 'p<0.05' is a special issue of *The American Statistician*.

› Altern Lab Anim. 2013 Mar;41(1):19-31. doi: 10.1177/026119291304100105.

Species Sensitivity Distribution estimation from uncertain (QSAR-based) effects data

Tom Aldenberg [1], Emiel Rorije

Affiliations + expand

PMID: 23614542 DOI: 10.1177/026119291304100105

Free article

Abstract

In environmental risk assessment, Species Sensitivity Distributions (SSDs) can be applied to estimate a PNEC (Predicted No-Effect Concentration) for a chemical substance, when sufficient data on species toxicities are available. The European Chemicals Agency (ECHA) recommendation is 10 biological species. The question addressed in this paper, is whether QSAR-predicted toxicities can be included in SSD based PNEC estimates, and whether any modifications need to be made to account for the uncertainty in the QSAR-model estimates. This problem is addressed from a probabilistic modelling point of view. From classical analysis of variation (ANOVA), we review how the error-in-data SSD problem is similar to separation into between-group and within-group variance. ECHA guidance suggests averaging similar endpoint data for a species, which is consistent with group means, as in ANOVA. This exercise reveals that error-in data reduces the estimation of the between species variation, i.e. the SSD variance, rather than enlarging it. A Bayesian analysis permits the assessment of the uncertainty of the SSD mean and variance parameters for given values of mean species toxicity error. This requires a hierarchical model. Prototyping this model for an artificial five-species data set seems to suggest that the influence of data error is relatively minor. Moreover, when neglecting this data error, a slightly conservative estimate of the SSD results. Hence, we suggest including (model-predicted) data as model point estimates and handling the SSD as usual. The Bayesian simulation of the error-in-data SSD leads to

More Noise Does Not Mean More Precision: A Review of Aldenberg and Rorije (2013)

David R. Fox[1,2]

[1]*Environmetrics Australia, Melbourne, Australia;* [2]*University of Melbourne, Melbourne, Australia*

Summary — This paper provides a critical review of recently published work that suggests that the precision of hazardous concentration estimates from Species Sensitivity Distributions (SSDs) is improved when the uncertainty in the input data is taken into account. Our review confirms that this counter-intuitive result is indeed incorrect.

Key words: *ANOVA, Bayesian predictive distribution, beta-content tolerance intervals, components of variation, species sensitivity distributions, uncertainty estimation.*

Address for correspondence: *David R. Fox, PO Box 7117, Beaumaris, Victoria, Australia 3193.*
E-mail: *david.fox@environmetrics.net.au*

Assessment factors in species sensitivity distributions for the derivation of guideline values for aquatic contaminants

David R. Fox [A B *] and Graeme E. Batley [C]

+ Author Affiliations

* Correspondence to: david.fox@environmetrics.net.au

Handling Editor: Kevin Wilkinson

Environmental Chemistry 19(4) 201-209 https://doi.org/10.1071/EN22061

Submitted: 12 June 2022 Accepted: 30 July 2022 Published: 7 September 2022

Environmental context. The use of assessment factors applied to guideline values derived using species sensitivity distributions adds an unnecessary level of conservatism. Using an adequate toxicity dataset and applying the latest model-averaging software will yield values of greatest reliability.

Abstract. The development of the Species Sensitivity Distribution (SSD) more than 30 years ago was in direct response to the many criticisms concerning the use of subjective Assessment (or Application) Factors (AFs) in widespread use at the time. While not perfect, SSD modelling is statistically defensible whereas AFs are not. While intuitively appealing, we believe recent guidance recommending the use of AFs in conjunction with SSD modelling is concerning and

- or an opinion on a personal website

7:34 am Sat 16 Nov

Not Secure — debtox.info

VPN 71%

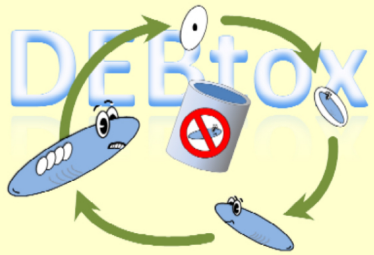
DEBtox information
Making sense of ecotoxicity test results

Main menu

- Home
- E-books
- Publications
- Software
- Courses
- Contact, links, news

Submenu

- About DEBtox
- About GUTS
- About DEBkiss
- About ECx and NOEC



Home / About ECx and NOEC

Download a preprint of my opinion paper on statistical approaches

Jager T (preprint). It's about time: moving away from statistical analysis of ecotoxicity data. Preprint version.

What is the traditional approach to ecotox data analysis?

The classical approaches to analyse ecotoxicity data are statistical (data-driven) in nature:

- Hypothesis testing to find the highest concentration without significant effects, relative to the control, at the end of the test (e.g., yielding an NOEC).
- Curve fitting and interpolation to find the concentration associated with *x*% effect, relative to the control, at the end of the test (e.g., an EC*x* or LC*x*).

Note the wording "at the end of the test": these approaches only deal with the dose and not with time. These methods are extremely common in both regulatory and academic investigations, and feature in every ecotoxicological textbook. They are so common that most ecotoxicologists are probably not even aware of their severe limitations:

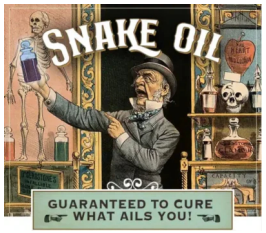
- These methods make poor use of the data as only the results at the end of the test are used.
- If multiple endpoints are observed in a test (e.g., growth and reproduction), they are treated as independent and unrelated traits (which is biological nonsense).
- The summary statistics depend on the duration of exposure. This is easily missed as exposure time is standardised in routine test protocols. How toxicity changes over time is depends on the chemica, on the species, and on the test conditions (e.g., temperature). Furthermore, it depends on how the trait is expressed (e.g., using body size as length or as weight).
- Because of the previous point, summary statistics cannot be compared between species and between chemicals. Every subsequent analysis that is done with these numbers (such as QSARs and SSDs) is therefore also questionable.
- Since these methods are descriptive, they offer no insight into the mechanisms underlying the toxic response: we don't learn anything from them.
- Since these methods are descriptive, they cannot be used to provide meaningful extrapolations to other conditions, e.g., different exposure patterns, different exposure durations, different environmental conditions.

Unrealistic for GVs

The NOEC has a number of additional limitations that make it even more unatractive than the ECx:

- The NOEC has to be one of the test concentrations, so its value depends on test design.
- We cannot generate a confidence interval on the NOEC.
- The NOEC is a fallacy against statistical principles: drawing conclusions from non-significant results. A lack of statistically significant effects does not mean no effect.
- Because of the previous point, the NOEC becomes higher (less protective) with increasing variability in the observations. The actual level of effect at the NOEC can be quite high in standard toxicity tests.

Beware snake-oil merchants



This (non-statistical) expert testified in a NZ environment court. His statistical analysis was rubbish:

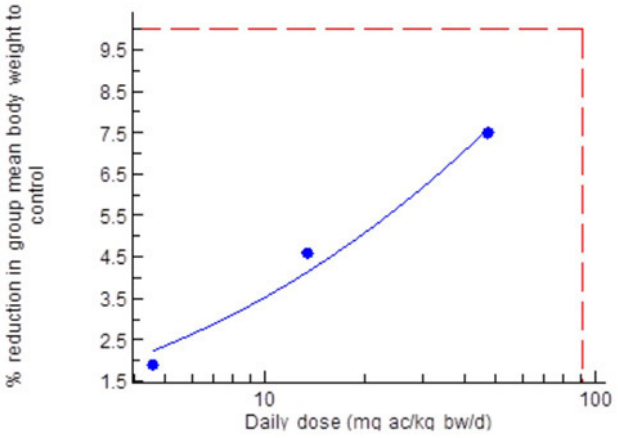
58. EFSA (2023) recommends to always consider the 10% as the effect level which is biologically relevant for all different parameters assessed for birds.

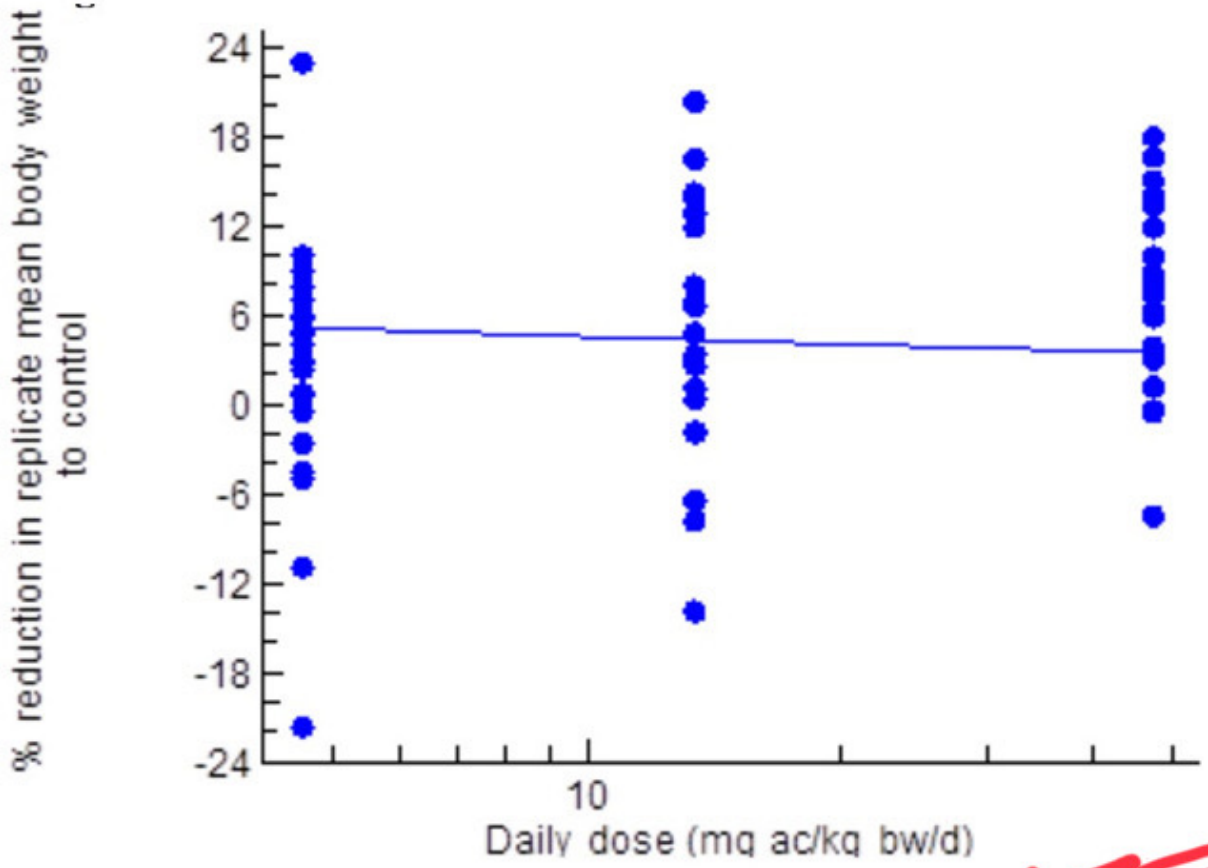
Level of effect observed in the bobwhite quail long term study

The day 14 chick bodyweight does demonstrate a dose/response relationship over the three tested exposure concentrations (Figure 2).

Table 5: Alternative bird reproduction endpoints for risk assessment

Species level	Endpoint	Rationale
Non-threatened species	ED_{10} 91.7 mg ac/kg bw/d LOC = 1	Biological relevance. <10% effects on chick body weight, most sensitive parameter.
Threatened species	ED_{10} 91.7 mg ac/kg bw/d LOC = 0.1	

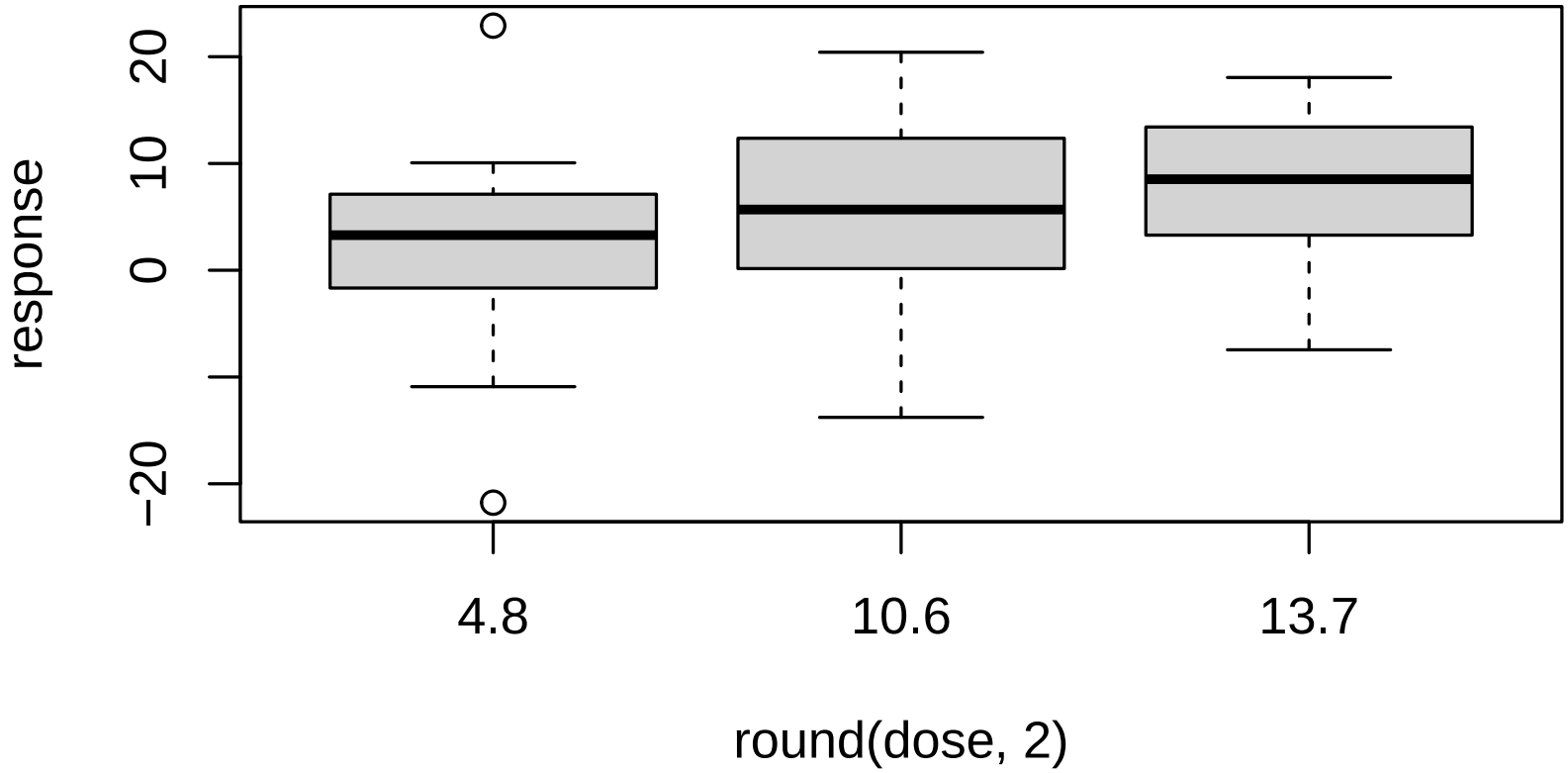




WTF!

Figure 3: Example of 4 parameter log logistic dose/response model – response of day 14 chick body weights

term	df	sumsq	meansq	statistic	p.value
dose	1	279.899	279.89897	4.008913	0.0504883
Residuals	52	3630.597	69.81917	NA	NA

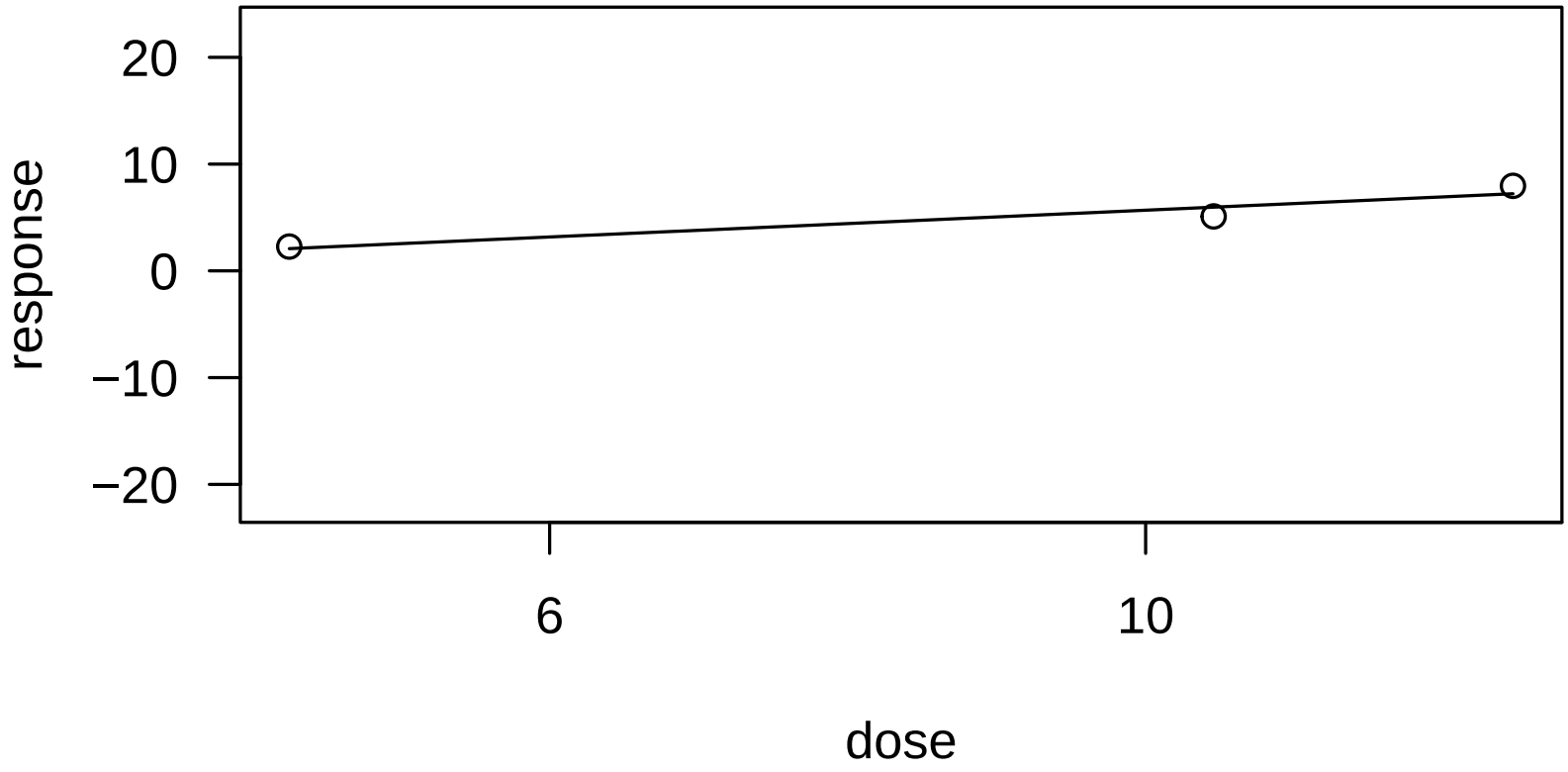


No significant `dose` effect.

Fitted model plot.

```
Loading required package: MASS

'drc' has been loaded.
Please cite R and 'drc' if used for a publication,
for references type 'citation()' and 'citation('drc')'.

Attaching package: 'drc'
The following objects are masked from 'package:stats':

    gaussian, getInitial
```

term	curve	estimate	std.error	statistic	p.value
b	(Intercept)	-0.2070017	2.998112	-0.0690440	0.9452300
c	(Intercept)	-42.5452686	633.037977	-0.0672081	0.9466841
d	(Intercept)	52.2791855	732.931765	0.0713289	0.9434206
e	(Intercept)	8.4725297	NaN	NaN	NaN




```
Warning in sqrt(diag(varMat)): NaNs produced
```

Cannot reliably estimate 4 parameters from 3 replicated doses.

OECD Guidance document 54 - Revision process has commenced

34th SETAC Europe Annual Meeting | 05 - 09 May 2024 | Sevilla

Activities to revise the OECD Document No. 54 on statistical analysis of ecotoxicity data

Benjamin Daniels, Thomas Gräff, Pia Kotschik & Susanne Walter-Rohde
German Environment Agency (UBA), Dessau-Roßlau, Germany

Background

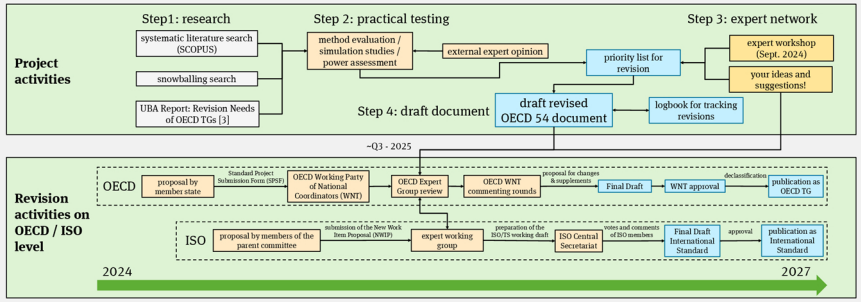
- OECD No. 54 (2006) [1] summarizes the **most important statistical methods** for data analysis from ecotoxicological studies
- More suitable statistical methods and test approaches are nowadays available
- Some approaches described in OECD No. 54 are no longer considered as state of the art
- Choice and application of the statistical method has direct impact on all OECD Test Guidelines

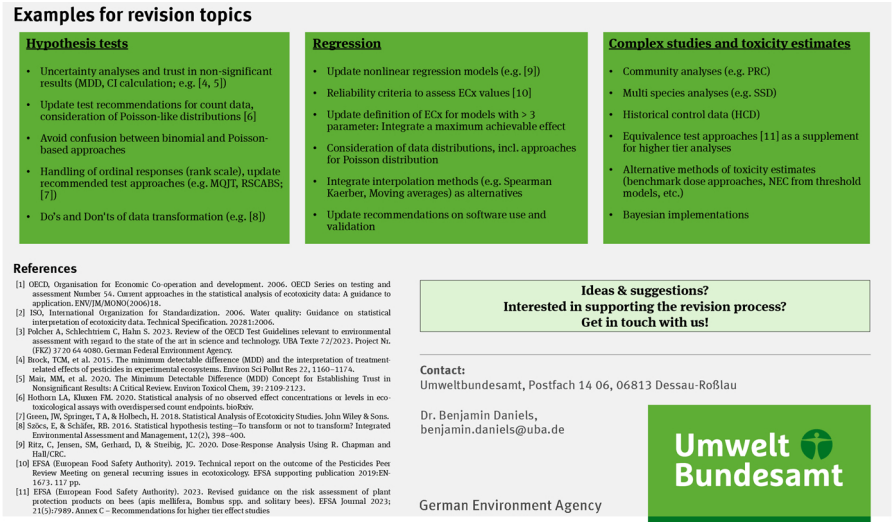
  → evaluation of the effects of regulated chemicals is directly affected!

Objectives

- Update of methods and statistical procedures
- Provide more practical guidance on the selection & comparison of hypothesis tests and model fitting approaches
- Create a close link to the update of the ISO/TS 20281 [2] to ensure harmonized guidance
- Draft an updated OECD Document No. 54
- Support the revision process at OECD level

Intended revision process




Examples for revision topics
Hypothesis tests
• Uncertainty analyses and trust in non-significant results (MDD, CI calculation; e.g. [4, 5])
• Update test recommendations for count data, consideration of Poisson-like distributions [6]
• Avoid confusion between binomial and Poisson-based approaches
• Handling of ordinal responses (rank scale), update recommended test approaches (e.g. MQJT, RSCABS; [7])
• Do's and Don'ts of data transformation (e.g. [8])
Regression
• Update nonlinear regression models (e.g. [9])
• Reliability criteria to assess ECx values [10]
• Update definition of ECx for models with > 3 parameter: Integrate a maximum achievable effect
• Consideration of data distributions, incl. approaches for Poisson distribution
• Integrate interpolation methods (e.g. Spearman Kaerber, Moving averages) as alternatives
• Update recommendations on software use and validation
Complex studies and toxicity estimates
• Community analyses (e.g. PRC)
• Multi species analyses (e.g. SSD)
• Historical control data (HCD)
• Equivalence test approaches [11] as a supplement for higher tier analyses
• Alternative methods of toxicity estimates (benchmark dose approaches, NEC from threshold models, etc.)
• Bayesian implementations
References
[1] OECD, Organisation for Economic Co-operation and development. 2006. OECD Series on testing and assessment Number 54. Current approaches in the statistical analysis of ecotoxicity data: A guidance to application. ENV/JM/MONO(2006)18.
[2] ISO, International Organization for Standardization. 2006. Water quality: Guidance on statistical interpretation of ecotoxicity data. Technical Specification. 20281:2006.
[3] Polcher A, Schlechtriem C, Hahn S. 2023. Review of the OECD Test Guidelines relevant to environmental assessment with regard to the state of the art in science and technology. UBA Texte 72/2023. Project No. (FKZ) 3720 64 4080. German Federal Environment Agency.
[4] Brock, TCM, et al. 2015. The minimum detectable difference (MDD) and the interpretation of treatment-related effects of pesticides in experimental ecosystems. Environ Sci Pollut Res 22, 1160–1174.
[5] Mair, MM, et al. 2020. The Minimum Detectable Difference (MDD) Concept for Establishing Trust in Nonsignificant Results: A Critical Review. Environ Toxicol Chem, 39: 2109-2123.
[6] Hothorn LA, Knuren FM. 2020. Statistical analysis of no observed effect concentrations or levels in eco-toxicological assays with overdispersed count endpoints. bioRxiv.
[7] Green, JW, Springer, T A, & Holbech, H. 2018. Statistical Analysis of Ecotoxicity Studies. John Wiley & Sons.
[8] Szöcs, E, & Schäfer, RB. 2016. Statistical hypothesis testing–To transform or not to transform? Integrated Environmental Assessment and Management, 12(2), 398–400.
[9] Ritz, C, Jensen, SM, Gerhard, D, & Streibig, JC. 2020. Dose-Response Analysis Using R. Chapman and Hall/CRC.
[10] EFSA (European Food Safety Authority). 2019. Technical report on the outcome of the Pesticides Peer Review Meeting on general recurring issues in ecotoxicology. EFSA supposting publication 2019:EN-1673. 117 pp.
[11] EFSA (European Food Safety Authority). 2023. Revised guidance on the risk assessment of plant protection products on bees (apis mellifera, Bombus spp. and solitary bees). EFSA Journal 2023; 21(5):7989. Annex C – Recommendations for higher tier effect studies
Ideas & suggestions?
Interested in supporting the revision process?
Get in touch with us!
Contact:
Umweltbundesamt, Postfach 14 06, 06813 Dessau-Roßlau
Dr. Benjamin Daniels,
benjamin.daniels@uba.de
German Environment Agency
Umwelt Bundesamt

Current approaches to C-R and SSD Modelling

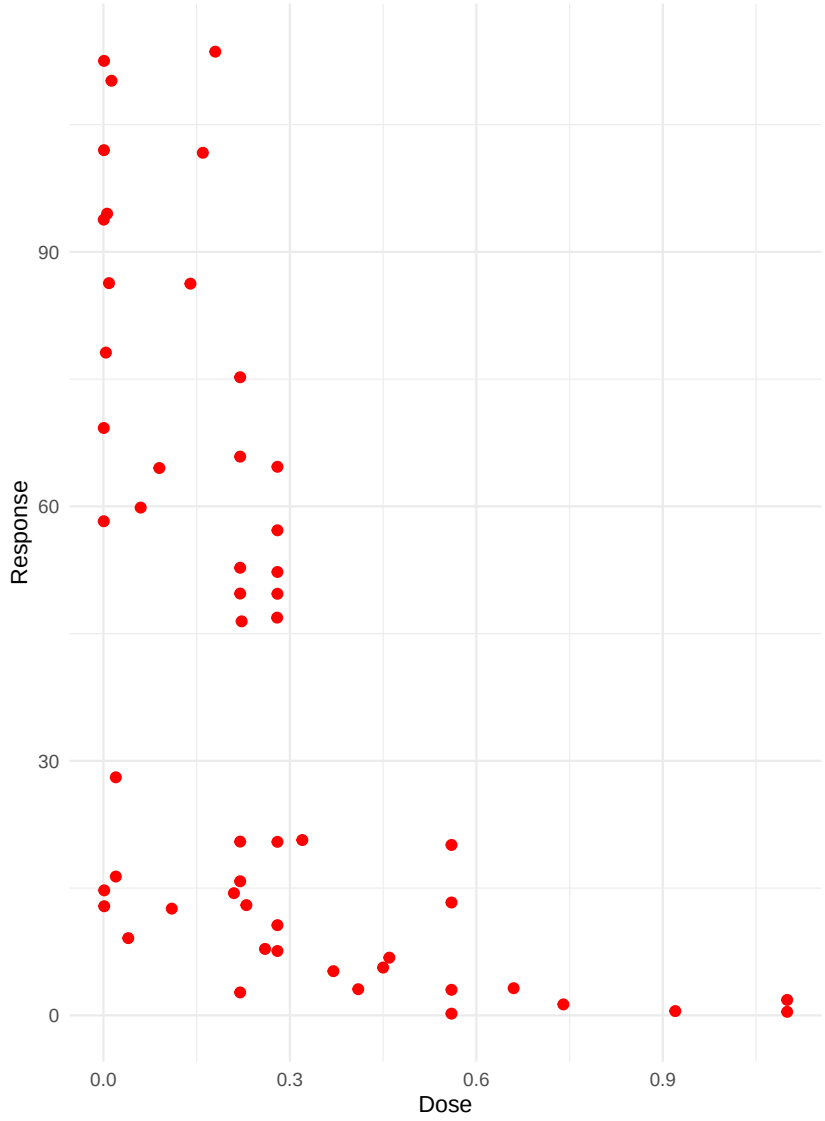
	Analysis type			
	C-R Modelling →	**Toxicity metrics →**	**SSD Modelling →**	**Environmental Protection**
Mode				
Old school	• ANOVA + Dunnett's Test • Doesn't model the functional relationship between concentration and response	• NOEC / LOEC (no statement of precision/uncertainty)	• Simple (single) distribution-fitting • begs the question "which distribution?" • Application of Assessment Factor(s) AFs	Completely unknown when based on HCx/AF
'Contemporary'	• R packages: - drc - drda - Bayesnec - BMD and others • USEPA BMDS • RIVM PROAST	• NEC (Frequentist and Bayes) • NSEC • ECx • BMD (Standard errors;confidence intervals; posterior distributions)	Comprehensive SSD modelling tools: • ssdtools - R package • shiny(ssdtools) - interactive, online • SSDToolbox - USEPA software (MATLAB) • MOSAIC - online (University Lyon) Note: ssdtools and SSDToolbox utilise model-averaging	• Quantifiable although dependent on a number of critical assumptions • Backed up by solid statistical analysis • Results are reproducable (for same data, same set of distributions)

Example: Survival time of daphnids versus dinoseb concentration

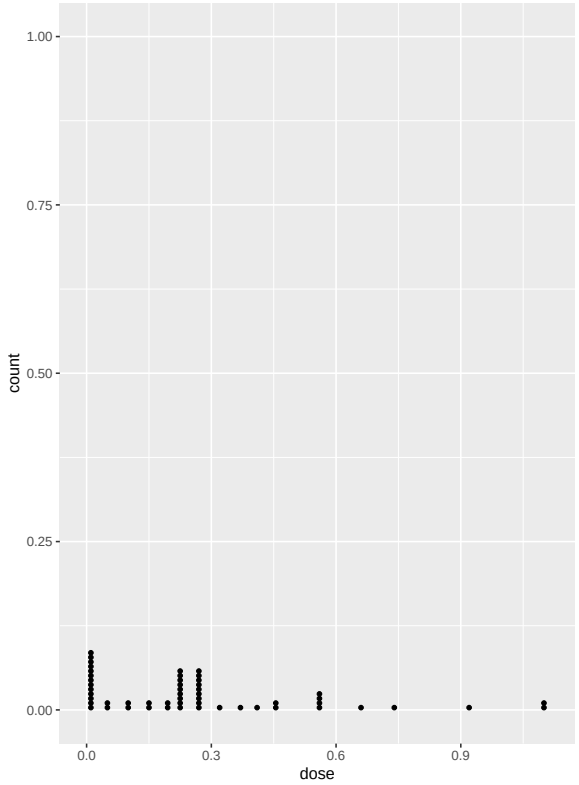
Source: Chevre et al. (2009)

---

dose	response
0.0013	12.865
0.0013	14.736
0.0006	58.245
0.0006	69.240
0.0040	78.127
0.0090	86.312
0.0060	94.501
0.0006	93.801
0.0009	101.988
0.0010	112.514
0.0130	110.169
0.0200	16.365
0.0200	28.060
0.0400	9.104
0.0600	59.855
0.1100	12.581
0.0900	64.520
0.1400	86.251
0.1600	101.679
0.1800	113.599
0.2200	2.704
0.2300	12.995
0.2100	14.405
0.2200	15.803
0.2200	20.481
0.2224	46.446
0.2200	49.721
0.2200	52.762
0.2200	65.862
0.2200	75.219
0.2600	7.832


Response
90
60
30
0
0.0
0.3
0.6
0.9
Dose

Next - look at the frequency distribution of the **dose** variable

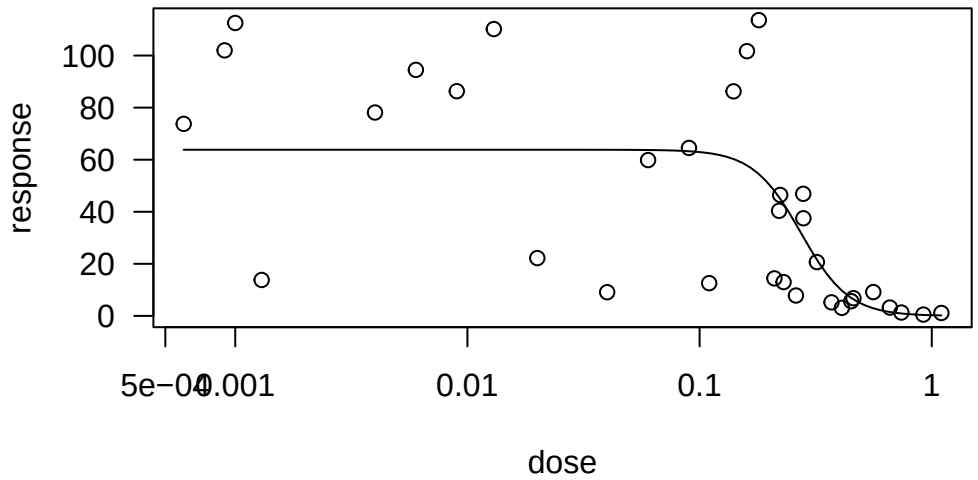


Note there are a number of concentrations having no replication. This presents no difficulties for fitting a C-R model but is a problem for ANOVA and hence NOEC computation.

Playing with the sliders we can visually fit a reasonable model. For example, `b = 6.5` and `e = 0.34`. These would be reasonable starting values for formal model-fitting using `drc` for example.Let's try `drc`.

`drc` Modelling of dinoseb data

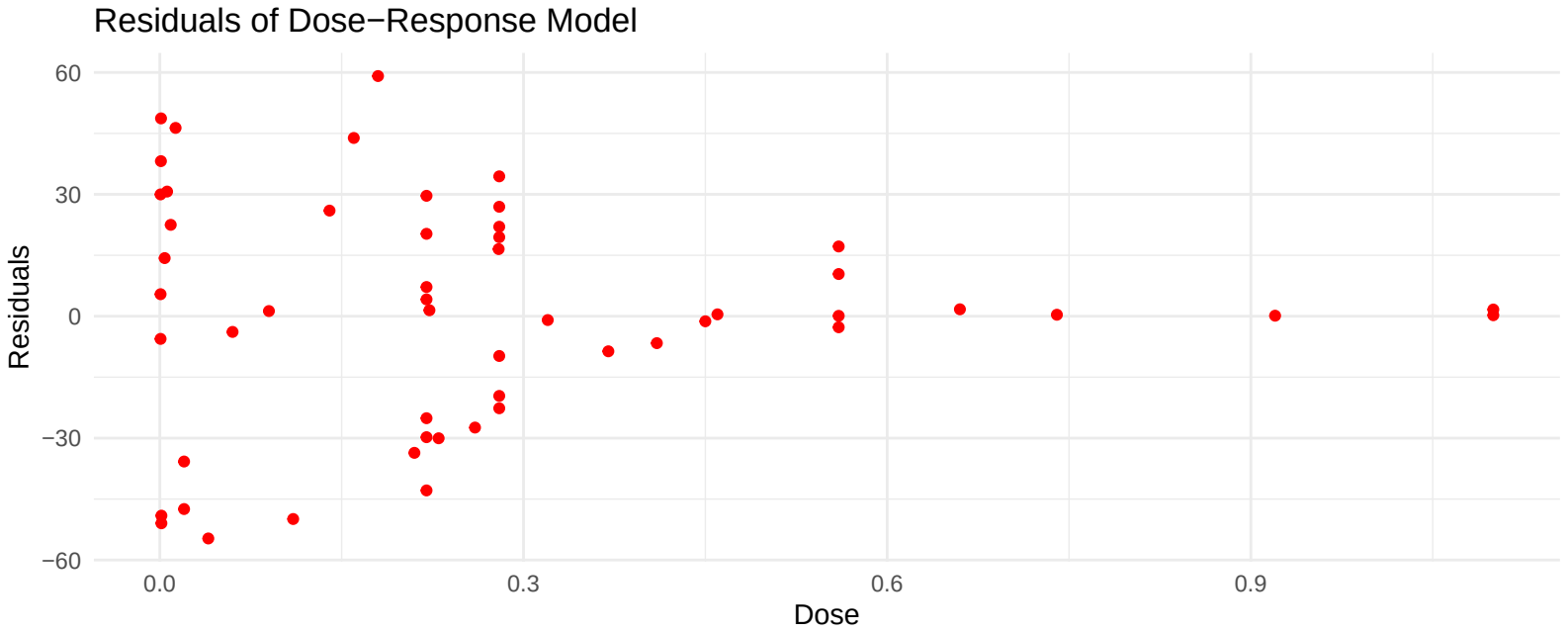
```
Model fitted: Log-logistic (ED50 as parameter) with lower limit at 0 (3 parms)

Parameter estimates:

               Estimate Std. Error t-value   p-value
b:(Intercept)  4.230796   2.134889  1.9817   0.05302 .
d:(Intercept) 63.820667   6.690872  9.5385 7.692e-13 ***
e:(Intercept)  0.273154   0.033564  8.1382 1.018e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error:

 28.56357 (50 degrees of freedom)
```



Note, `drc` always plots on a logarithmic scale. Let's look at a plot on a *linear* `dose` scale.

	ModelDf	RSS	Df	F value	p value
ANOVA	20	8910.78	NA	NA	NA
DRC model	50	40793.87	30	2.385357	0.023065

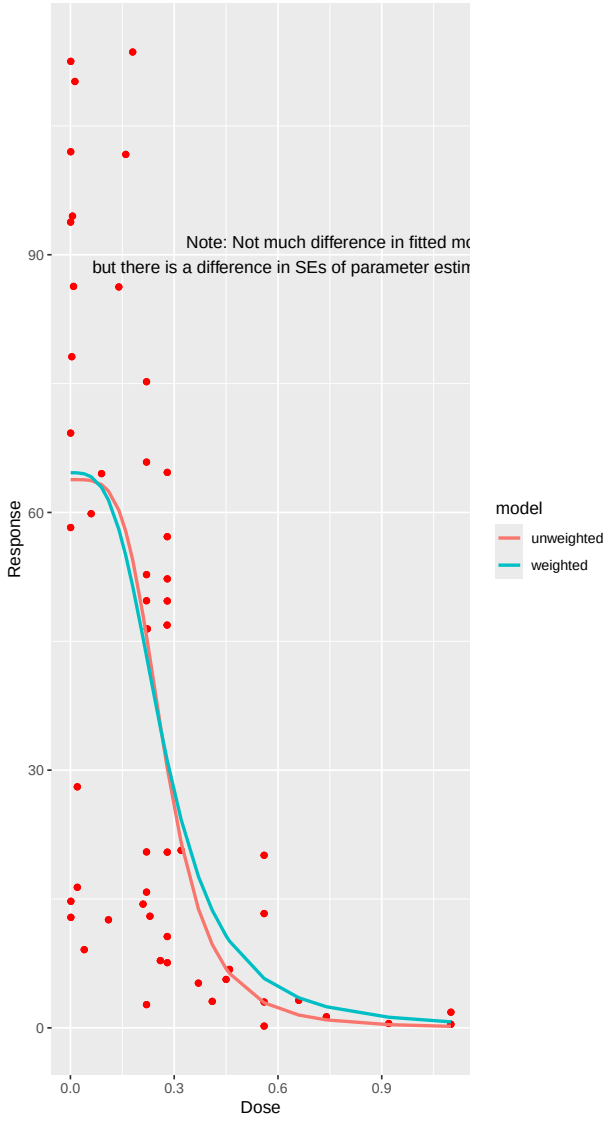

```
Assess Model fit
```



The residual plot reveals a serious issue - that of heteroscedacisity - in otherwords, non-constant variance. This is a severe violation of a key assumption.Dealing with it is unfortunately not straightforward due to the non-linear link function relating `response` and `dose`.We can use other packages for fitting non-linear, weighted regression models.

```
library(nlstools)
library(broom)
library(ggtext)
# Starting values for the 3-parameter log-logistic model
start_values <- list(d = 64, b = 6.5, e = 0.34)
```

```r
# Fit the model using nls() with a 3-parameter log-logistic form
ll.1 <- nls(response ~ d / (1 + exp(b * (log(dose) - log(e)))),
                         data = df, start = start_values)

# Use fitted values to determine weights
fv1 <- predict(ll.1)

ll.2 <- nls(response ~ d / (1 + exp(b * (log(dose) - log(e)))),
                              data = df, start = start_values, weights = 1 / fv1)

# Add fitted values to the data frame for plotting
tmp<-data.frame(dose=rep(df$dose,2),response=rep(df$response,2),pred=c(predict(ll.1),predict(ll.2)),
    model=rep(c("unweighted","weighted"),each=53))


# Plot observed data and fitted model
ggplot(tmp, aes(x = dose, y = response)) +
  geom_point(color = "red") +  # Original data points
  geom_line(aes(y = pred, color = model),linewidth=1) +  # Fitted model line
  labs(x = "Dose", y = "Response") +
  annotate("text",label="Note: Not much difference in fitted model, \nbut there is a difference in SEs of parameter e
           x=0.8,y=90)
```


Note: Not much difference in fitted mo
but there is a difference in SEs of parameter estim
Response
90
60
30
0
0.0
0.3
0.6
0.9
Dose
model
unweighted
weighted

Comparison of parameter estimates

Parameter estimates for Model 1

```
Attaching package: 'dplyr'
The following object is masked from 'package:MASS':

    select
The following objects are masked from 'package:stats':

    filter, lag
The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

Attaching package: 'kableExtra'
The following object is masked from 'package:dplyr':

    group_rows
```

Parameter estimates for Model 2

term	estimate	std.error	statistic	p.value
d	63.8205	6.8651	9.2964	0.0000
b	4.2309	2.3618	1.7914	0.0793
e	0.2732	0.0324	8.4195	0.0000

term	estimate	std.error	statistic	p.value
d	64.6201	8.1635	7.9157	0.0000
b	3.2470	0.6866	4.7288	0.0000
e	0.2736	0.0368	7.4380	0.0000

Selecting a Model

`drc` has many model-fitting options:

Toxicity metrics estimated from a C-R Model

- NOEC
- ECx
- LCx

- NEC
- NSEC (New!)
- BMD

Toxicity metrics

TABLE 1 Toxicity estimates currently used for estimating no and low-effects for using in SSD modeling

Toxicity estimate	Definition	Statistical method
NSEC	No significant effect concentration—the concentration at which the modeled mean response is statistically indistinguishable from the mean control response	Interpolation from a CR model (Fisher & Fox, 2023)
NEC	No effect concentration—the minimum concentration above which an effect is predicted to occur	Parameter estimate of a CR threshold model (Fox, 2010; Pires et al., 2002; Van Der Hoeven, 1997)
NOEC	No observable effect concentration—the highest tested concentration at which the mean response is statistically indistinguishable from the mean control response	Dunnett's test (based on ANOVA)
EC*x*/IC*x*/LC*x*	*x*% effect/inhibition/lethal concentration—the concentration that is expected to cause a specified effect in *x*% of a group of organisms or *x*% effect (EC*x*); an *x*% reduction in a nonquantal measurement such as fecundity or growth (IC*x*); or be lethal to *x*% of a group of organisms (LC*x*)	Interpolation from a CR model

Note: Definitions are adapted from Warne et al. (2015), with the exception of NSEC, which is described in Fisher and Fox (2023). Abbreviations: CR, concentration-response; SSD, species sensitivity distribution.

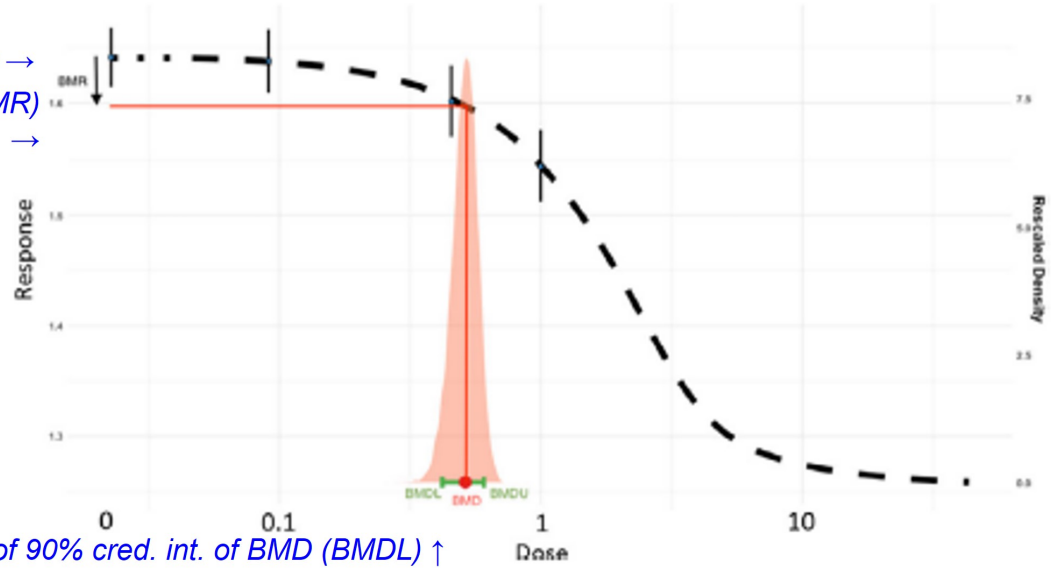
Benchmark Dose (BMD)

$BMDL_x$ = dose below which the change in response is likely to be smaller than x%.

where the term 'likely' is defined by the statistical credible level, usually 95%-level.

1) Fitted reponse →

2) Benchmark reponse (BMR) →



4) Lower 95% of 90% cred. int. of BMD (BMDL) ↑

3) Benchmark dose (BMD) ↑

Figure 1: Key concepts for the BMD approach. The observed mean responses plus or minus the observed standard deviation are plotted as vertical lines. The dashed curve is a fitted dose–response model, either one of the 16 individual dose–response models (see Section 2.5.1) or the averaged model. This curve determines the point estimate of the BMD, which is generally defined as a dose that corresponds to a low but biologically relevant[8] change in response, denoted the benchmark response (BMR). The density shows the posterior distribution of the BMD and the green line at the bottom of the density indicates the boundaries of the two-sided 90% credible interval of the BMD (defined by the 5% left and right tail probabilities of that posterior distribution). The BMDL is the 95% one-sided lower bound of the 90% credible interval for the BMD. Likewise, the BMDU is the 95% one-sided upper bound of the 90% credible interval for the BMD. It should be noted that the estimated background response (the median response of the control group) does not necessarily coincide with the observed background response. The BMR is defined as a change with regard to the background response predicted by the fitted model

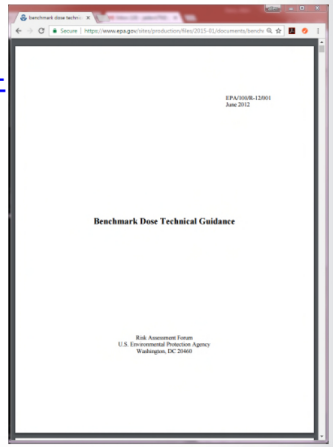
---

EFSA on-line BMD app

https://efsa.openanalytics.eu/

EPA's BMD Technical Guidance

- *Final* version of the EPA's Benchmark Dose Technical Guidance document was published in 2012: https://www.epa.gov/risk/benchmark-dose-technical-guidance
- Other guidance documents relevant to BMD modeling available at: http://epa.gov/iris/backgrd.html
- EPA's Statistical Working Group periodically updates recommended model practices



Traditional Dichotomous Models

Model name	Functional form	# of Parameters[a]	Low Dose Linearity	Model fits
Multistage	$\gamma + (1-\gamma)\left[1 - \exp\left\{-\sum_{j=1}^{k} \beta_j X^j\right\}\right]$	1+k	Yes, if $\beta_1 > 0$ No, if $\beta_1 = 0$	All purpose
Logistic	$\frac{1}{1+\exp\{-(\alpha+\beta X)\}}$	2	Yes	Simple; no background
Probit	$\Phi(\alpha+\beta X)$	2	Yes	Simple; no background
Log-logistic	$\frac{\gamma+(1-\gamma)}{1+\exp\{-[\alpha+\beta\ln(X)]\}}$	3	No	All purpose; S-shape with plateau at 100%
Log-probit	$\gamma+(1-\gamma)\,\Phi\{\alpha+\beta\ln(X)\}$	3	No	All purpose; plateau S-shape with plateau at 100%
Gamma	$\gamma+(1-\gamma)\left[\int_0^{\beta x} t^{\alpha-1}e^{t}dt\right]/\Gamma(\alpha)$	3	No	All purpose
Weibull	$\gamma+(1-\gamma)[1-\exp\{-\beta X^{\alpha}\}]$	3	No	"Hockey stick" shape
Dichotomous Hill	$v\times g+\frac{(v-v\times g)}{1+\exp\{-a-b\times\ln(X)\}}$	4	Yes	Symmetrical, S-shape with plateau

[a] Background parameter = γ. Background for hill model = v × g

Continuous Model Forms

Model Name	Functional Form	# of Parameters	Model Fits
Polynomial[a]	$\beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_n X^n$	1 + n	All purpose, can fit non-symmetrical S-shaped datasets with plateaus
Power	$\gamma + \beta X^{\Phi}$	3	L-shaped
Hill	$\gamma + \frac{(v \times X^n)}{(k^n + X^n)}$	4	Symmetrical, sigmoidal, S-shape with plateau
Exponential[b] Model 2 Model 3 Model 4 Model 5	$a \times \exp\{\pm 1 \times b \times X\}$ $a \times \exp\{\pm 1 \times (b \times X)^d\}$ $a \times [c - (c-1) \times \exp\{\pm 1 \times b \times X\}]$ $a \times [c - (c-1) \times \exp\{\pm 1 \times (b \times X)^d\}]$	 2 3 3 4	All purpose (Models 2 & 3) Symmetrical and asymmetrical S-shape with plateau (Models 4 & 5)

[a] The stand-alone Linear model in BMDS is equal to a first-order polynomial model
[b] Nested family of 4 related models described by Slob (2002) and included in the PROAST software of RIVM

The No Significant Effect Concentration (NSEC)

Environmental Toxicology and Chemistry—Volume 42, Number 9—pp. 2019–2028, 2023
Received: 24 January 2023 | Revised: 23 February 2023 | Accepted: 13 March 2023

Hazard/Risk Assessment

Introducing the No-Significant-Effect Concentration

Rebecca Fisher[a,b,*] and David R. Fox[c,d]

[a]Australian Institute of Marine Science, Crawley, Western Australia, Australia
[b]Oceans Institute, University of Western Australia, Crawley, Western Australia, Australia
[c]Environmetrics Australia, Beaumaris, Victoria, Australia
[d]Department of Infrastructure Engineering, University of Melbourne, Parkville, Victoria, Australia

Abstract: The no-effect concentration (NEC) is the preferred threshold metric for single-species toxicity tests applied to derive safe concentration thresholds for contaminants in the environment for use in species sensitivity distributions. However, the NEC is only suitable when concentration–response (C-R) data exhibit a threshold response. We describe an alternative toxicity estimate, the no-significant-effect concentration (NSEC), which is better suited to C-R data for which the response is a monotonically decreasing function of concentration and no threshold effects are evident. We use a flexible, three-parameter sigmoidal function to describe the C-R relationship and detail both Bayesian and frequentist approaches to estimation and inference for the NSEC. While the NSEC is conceptually linked to the traditional no-observed-effect concentration (NOEC), it is a substantial improvement over the NOEC because it decouples the estimate from being directly dependent on the placement of treatment concentrations as well as admitting statements of precision of the resulting toxicity estimate. *Environ Toxicol Chem* 2023;42:2019–2028. © 2023 Commonwealth of Australia and The Authors. *Environmental Toxicology and Chemistry* published by Wiley Periodicals LLC on behalf of SETAC.

Keywords: Species sensitivity distribution; no-observed-effect concentrations (NOECs); Concentration–response modeling; Ecosystem protection; Ecotoxicology; Toxicity estimate

NSEC Explained

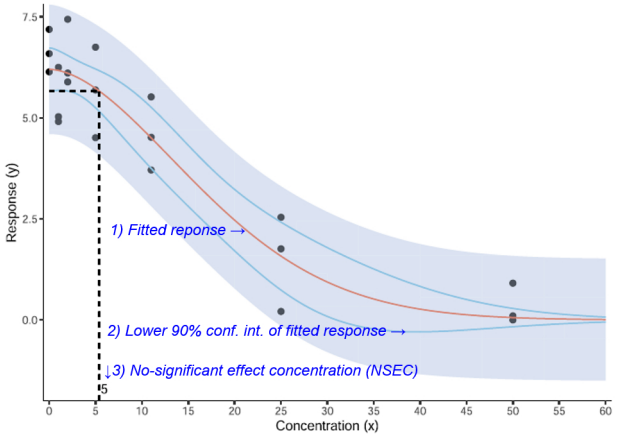
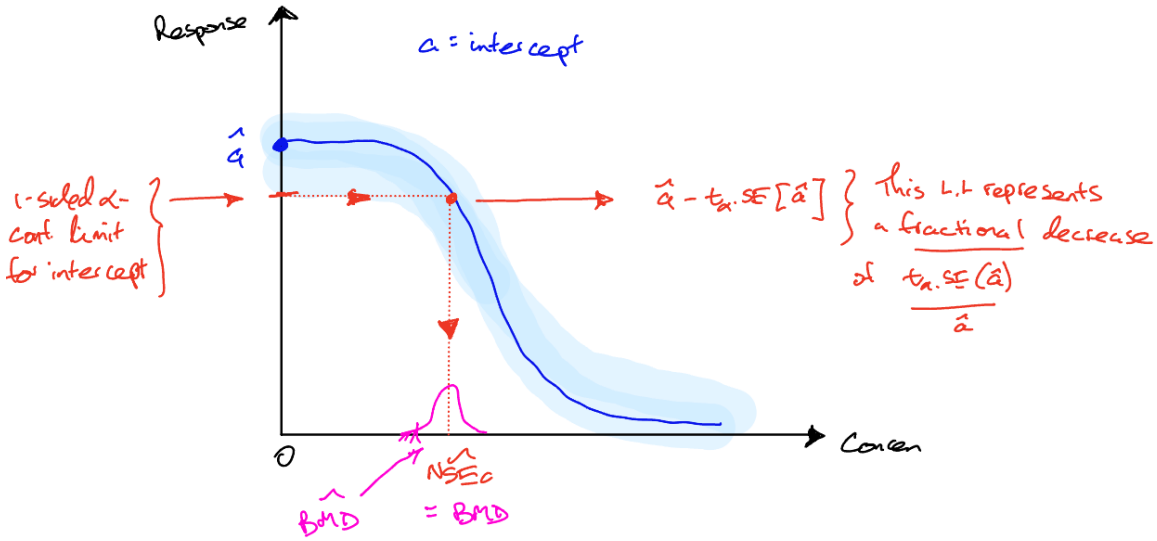


FIGURE 3: Fish growth data (solid circles). Red curve is estimated mean response function given by Equation 4. Blue lines denote limits of 90% confidence interval for mean response; 90% prediction band is represented by the shaded region. Horizontal dashed line is at the lower limit of the confidence interval for concentration = 0, and vertical dashed line is at the corresponding concentration.

Relationship between the NSEC and BMD



if BMR = $\frac{t_{\alpha} SE(\hat{a})}{\hat{a}}$ then NSEC = BMD.

But $\widehat{BMD}$ is taken to be lower conf./credibility limit of the sampling distribution of $\widehat{BMD}$

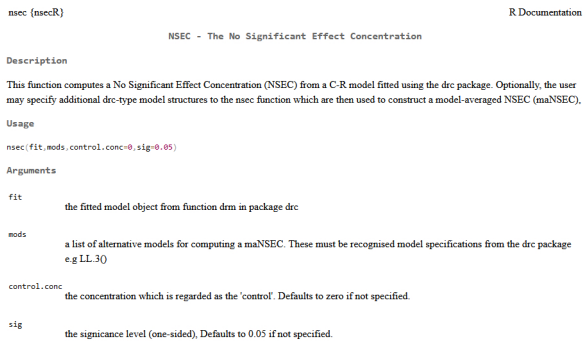
$\therefore$ $\widehat{BMD}$ will be < $\widehat{NSEC}$

Computing the NSEC

R Package `nsecR` can be installed from github by typing:

```
remotes::install_github("environmetrics/nsecR")`
```

in your `R` / `Rstudio` console.



nsec {nsecR} R Documentation

NSEC - The No Significant Effect Concentration

Description

This function computes a No Significant Effect Concentration (NSEC) from a C-R model fitted using the drc package. Optionally, the user may specify additional drc-type model structures to the nsec function which are then used to construct a model-averaged NSEC (maNSEC),

Usage

```
nsec(fit,mods,control.conc=0,sig=0.05)
```

Arguments

fit	the fitted model object from function drm in package drc
mods	a list of alternative models for computing a maNSEC. These must be recognised model specifications from the drc package e.g LL.3()
control.conc	the concentration which is regarded as the 'control'. Defaults to zero if not specified.
sig	the signicance level (one-sided), Defaults to 0.05 if not specified.

nsecR Example

```
library(nsecR)
library(drc)
library(dplyr)
library(kableExtra)
library(broom)
```

```r
#  First, fit a model(s) to your data

df<-read.csv(file="CHEVRE - EFFECTS OF DINOSEB ON DAPHNIA.DAT.csv")

ll.mod<-drm(response ~ dose, fct = LL.3(),data=df)

summary(ll.mod)
```

```
Model fitted: Log-logistic (ED50 as parameter) with lower limit at 0 (3 parms)

Parameter estimates:

               Estimate Std. Error t-value   p-value
b:(Intercept)  4.230796   2.134889  1.9817   0.05302 .
d:(Intercept) 63.820667   6.690872  9.5385 7.692e-13 ***
e:(Intercept)  0.273154   0.033564  8.1382 1.018e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error:

 28.56357 (50 degrees of freedom)
```

```r
modelFit(ll.mod)
```

```
Lack-of-fit test

          ModelDf   RSS Df F value p value
```

```
ANOVA          20  8911
DRC model      50 40794 30  2.3854  0.0231
```

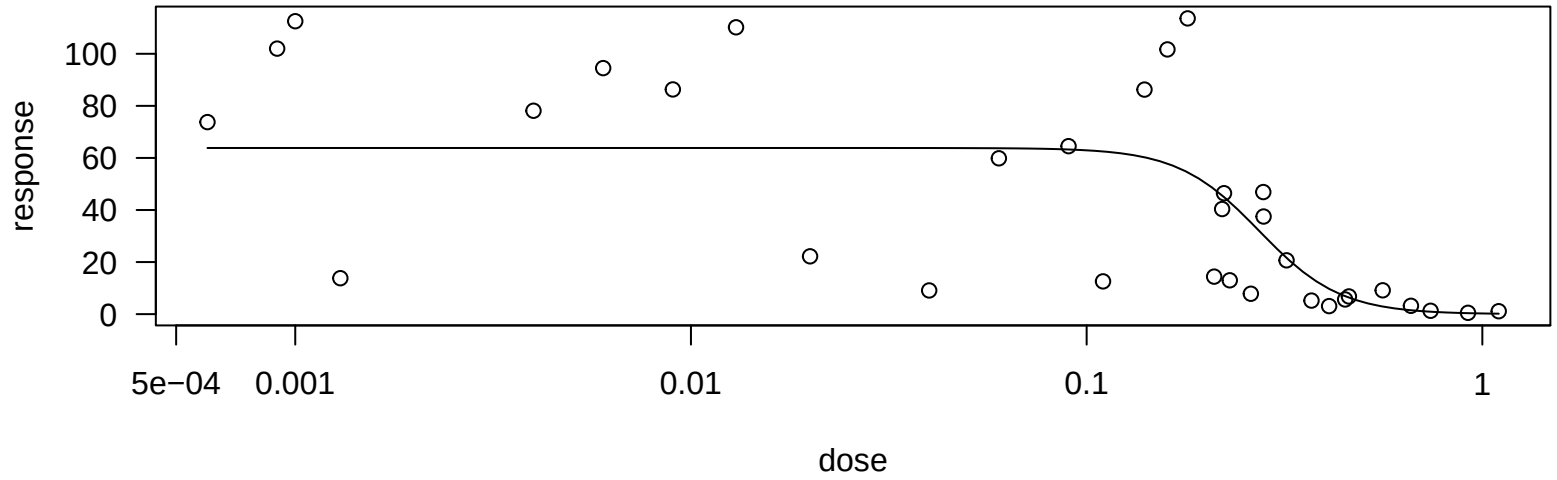
---

nsecR Example

Next compute the NSEC:

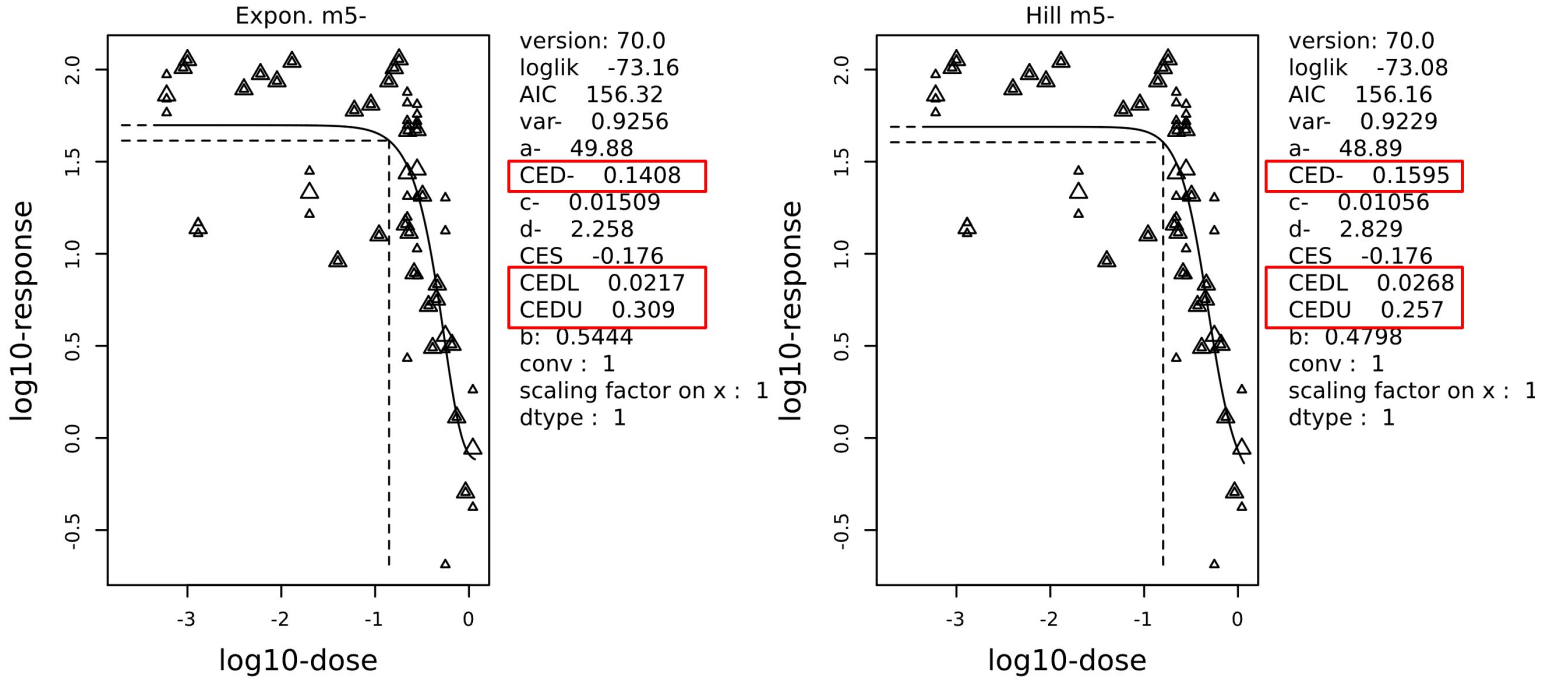
```
     ED52.6074008067397 Weight
LL.3          0.1895549    0.5
LL.3          0.1895549    0.5
```

```
[1] 0.1895549
```



Now, $\widehat{NSEC} = 0.1896$ and from this part of the output `ED52.6074...` we see that at `dose=0.1896` the response is `52.6074`. We also see that the `response` intercept $\left(\hat{d}\right)$ is at `63.8207`. Thus, `52.6074` represents a `17.6%` $\left\{= 1 - \frac{52.6074}{63.8207} = 0.1757\right\}$ 'effect'.Comapre this to the 'equivalent' BMD output:

term	estimate
b	4.2307958
d	63.8206673
e	0.2731545



Model-averaged nsecR (maNSEC) Example

Having fitted a single model and computing the NSEC, it's trivial to compute a *model-averaged* NSEC (maNSEC) by including additional models.Possible models are those available in the `drc` package, namely:

Let's try a maNSEC by adding the following models:

LL.4

W1.3

W1.4

LN.3

LN.4

```
nsec(ll.mod, mods = list(LL.4(), W1.3(), W1.4(),
                         LN.3(), LN.4()))
```

```
 Model-averaged NSEC =  0.1899
```

From individual to population - The Species Sensitivity Distribution (SSD)

- What is an SSD? It's simply a theoretical probability model (a *cumulative distribution function* or *cdf*) fitted to a (usually) small collection of toxicity metrics for a particular chemical in a particular environment - for example `copper` in a `freshwater` environment.
- Putting aside the current debate about TKTD models *versus* SSD models (and *statistical science* more generally), the SSD is remains one of the most credible and scientifically defensible means of establishing **default guideline**

fct	description
LL.2	Log-logistic (ED50 as parameter) with lower limit at 0 and upper limit at 1
LL.3	Log-logistic (ED50 as parameter) with lower limit at 0
LL.3u	Log-logistic (ED50 as parameter) with upper limit at 1
LL.4	Log-logistic (ED50 as parameter)
LL.5	Generalized log-logistic (ED50 as parameter)
W1.2	Weibull (type 1) with lower limit at 0 and upper limit at 1
W1.3	Weibull (type 1) with lower limit at 0
W1.4	Weibull (type 1)
W2.2	Weibull (type 2) with lower limit at 0 and upper limit at 1
W2.3	Weibull (type 2) with lower limit at 0
W2.4	Weibull (type 2)
BC.4	Brain-Cousens (hormesis) with lower limit fixed at 0
BC.5	Brain-Cousens (hormesis)
LL2.2	Log-logistic (log(ED50) as parameter) with lower limit at 0 and upper limit at 1
LL2.3	Log-logistic (log(ED50) as parameter) with lower limit at 0
LL2.3u	Log-logistic (log(ED50) as parameter) with upper limit at 1
LL2.4	Log-logistic (log(ED50) as parameter)
LL2.5	Generalised log-logistic (log(ED50) as parameter)
AR.2	Asymptotic regression with lower limit at 0
AR.3	Shifted asymptotic regression
MM.2	Michaelis-Menten
MM.3	Shifted Michaelis-Menten

	NSEC	Weight
LL.4	0.1904	0.0925
W1.3	0.1846	0.2056
W1.4	0.1885	0.0879
LN.3	0.1925	0.2596
LN.4	0.1957	0.1053

values or **DGVs** for toxicants in the environment.

- Although SSD modelling is very mature (>30 years old) there has been a renewed research push with some important advances having being recently made.
- The motivation for SSD modelling was to put DGV determination on a more rational/objective footing. It displaced the previous method of using subjective **Assessment Factors**.
- Yet we still have jurisdictions around the world recommending AFs be applied to DGVs derived fom an SSD!!

- The minimum sample size (number of data). This issue is subject to an ongoing debate. While OECD, 2007 proposes a minimum of eight NOECs on species from different taxonomic groups, EC, 2003a recommends 10 NOECs (and preferably more than 15) on species from eight taxonomic groups. Similar proposals have been made by Gibbons and Coleman, 2001 and de Bruijn *et al.*, 1999.
- How multiple data for one species are dealt with, e.g. averaging comparable data, or selecting the most sensitive endpoint when various data are available.
- Statistical fitting procedures. That is, the method must be mentioned and explained, where the log-normal distribution is the preferred one for pragmatic reasons. In addition, a statistical method is to be used to test the goodness of fit. In addition to the Kolmogorov-Smirnov test, the Anderson–Darling goodness of fit test can be used as a criterion for the choice of a parametric distribution for data-rich data sets, because it gives more weight to the tails of the distribution. Results should be discussed in regard to the graphical representation of the species distribution. If the data do not fit any distribution, the left tail of the distribution (the lowest effect concentrations) should be analysed more carefully. Any choice of a specific distribution function should be clearly explained.
- Estimated parameter. That is, the concentration corresponding with the point in the species sensitivity distribution (SSD) profile below which 5% of the species occur may be derived with a 50% confidence interval associated with this concentration, as an intermediate value in the determination of the PNEC.
- Estimation of the PNEC. That is, the intermediate value may be divided by an appropriate assessment factor, if needed, to reflect the further uncertainties identified. If mesocosm studies are available, they should also be evaluated to decide on the assessment factor.

Deviations from these recommendations can be made on a case-by-case basis, through consideration of sensitive endpoints, sensitive species, mode of toxic action and/or knowledge from structure activity considerations.

The PNEC should also be derived by applying the assessment factor approach on the same database.

Figure 1: DEWHA, Commonwealth of Australia (2009)

Environmental Toxicology and Chemistry—Volume 40, Number 2—pp. 293–308, 2021
Received: 22 June 2020 | Revised: 13 July 2020 | Accepted: 30 October 2020

Critical Review

Recent Developments in Species Sensitivity Distribution Modeling

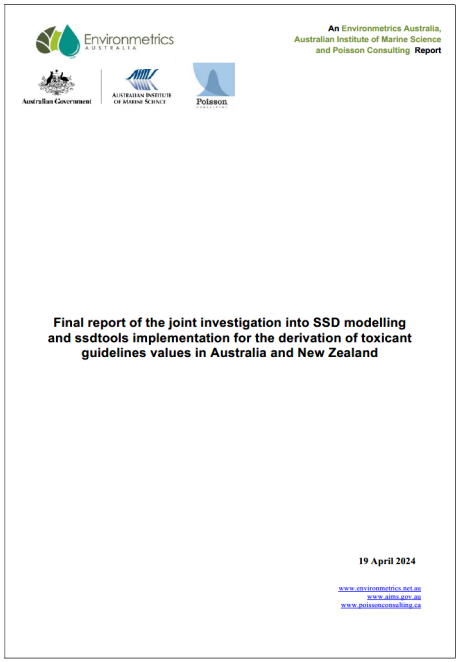
D.R. Fox,[a,b,*] R.A. van Dam,[c] R. Fisher,[d] G.E. Batley,[e] A.R. Tillmanns,[f] J. Thorley,[g] C.J. Schwarz,[h] D.J. Spry,[i] and K. McTavish[i]

[a]Environmetrics Australia, Beaumaris, Victoria, Australia
[b]University of Melbourne, Parkville, Victoria, Australia
[c]WQadvice, Adelaide, South Australia, Australia
[d]Australian Institute of Marine Science and the University of Western Australia Oceans Institute and School of Plant Biology, Crawley, Western Australia, Australia
[e]CSIRO Land and Water, Lucas Heights, New South Wales, Australia
[f]British Columbia Ministry of Environment and Climate Change Strategy, Victoria, British Columbia, Canada
[g]Poisson Consulting, Nelson, British Columbia, Canada
[h]StatMathComp Consulting, Vancouver, British Columbia, Canada
[i]Environment and Climate Change Canada, Gatineau, Quebec, Canada

Abstract: The species sensitivity distribution (SSD) is a statistical approach that is used to estimate either the concentration of a chemical that is hazardous to no more than x% of all species (the HCx) or the proportion of species potentially affected by a given concentration of a chemical. Despite a significant body of published research and critical reviews over the past 20 yr aimed at improving the methodology, the fundamentals remain unchanged. Although there have been some recent suggestions for improvements to SSD methods in the literature, in general, few of these suggestions have been formally adopted. Furthermore, critics of the approach can rightly point to the fact that differences in technical implementation can lead to marked differences in results, thereby undermining confidence in SSD approaches. Despite the limitations, SSDs remain a practical tool and, until a demonstrably better inferential framework is available, developments and enhancements to conventional SSD practice will and should continue. We therefore believe the time has come for the scientific community to decide how it wants SSD methods to evolve. The present study summarizes the current status of, and elaborates on several recent developments for, SSD methods, specifically, model averaging, multimodality, and software development. We also consider future directions with respect to the use of SSDs, with the ultimate aim of helping to facilitate greater international collaboration and, potentially, greater harmonization of SSD methods. *Environ Toxicol Chem* 2021;40:293–308. © 2020 SETAC

Keywords: Species sensitivity distrbution; Statistical inference; Hazardous concentration; Computer software

Australia-New Zealand-Canada Collaboration



`ssdtools V2.1`

`shinyssdtools` is (IMHO) the most advanced SSD modelling tool currently available. There are others - for example `MOSAIC` (Sandrine Charles' group at the University of Lyon) and `ssdtoolbox` (a MALAB executable from Matt Emerson at USEPA), but neither can match the full suite of modelling tools available in `ssdtools`. `ssdtools` is an `R` package available from CRAN and there is a companion, interacive on-line version called `shiny(ssdtools)`.

`shiny(ssdtools)`

Launch Report Card

- increased awareness of the role and relevancy of statistical methods
- (gradually) increasing uptake of `R`
- recent multi- agency, cross-jurisdictional R&D efforts => increased harmonisation of approaches
- development of interactive on-line tools
- increased participation of statisticians/quantitative biologists/R-programmers
- jurisdictions that are committed to re-writing statistical guidance

- NOECs - the ecotoxicological 'cockroach'
- AFs - Always Fraught
- End the 'Quixotic Quest'
- 'Friendly Fire' and debates in inappropriate fora
- Jurisdictional 'silos'

- Inertia

Challenges and Opportunities

- Addressing and compensating for small, biased samples in SSD and HCx estimation;
- Development of statistical methods to validate predictions from TKTD models and SSD-based approaches;
- How to design a C-R experiment that maximizes information content for minimum cost / maximum precision;
- How, or establish if it's possible to set an HCx for a mixture of chemicals;
- How to *seamlessly* integrate the *temporal* dimension into SSD modelling and HCx estimation rather than marginalising it;
- Strategies for error propagation to incorporate uncertainties arising from data collection process, imprecise model specification, and statistical treatment of data;
- Refinement of modelling capabilities to undertake external and internal exposure assessments;
- Elicitation of 'expert' opinion in the setting of Bayesian priors and protocols for reaching consensus when used these are used in a regulatory context.

Thank you for listening!

A pdf copy of this presentation is available at:

`https://environmetrics.net/resources/documents-and-reports/`