









ssdtools v2: An R package to fit Species Sensitivity Distributions


Joe Thorley ¹, Rebecca Fisher ^{2,3}, David Fox ^{4,5}, and Carl Schwarz ⁶

1 Poisson Consulting, Canada 2 Australian Institute of Marine Science, Australia 3 Oceans Institute, The University of Western Australia, Australia 4 The University of Melbourne, Australia 5 Environmetrics Australia, Australia 6 Simon Fraser University, Canada  Corresponding author

DOI: [10.21105/joss.07492](https://doi.org/10.21105/joss.07492)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: Fabian Scheipl 

Reviewers:

- [@flor14](#)
- [@nanhung](#)

Submitted: 02 October 2024

Published: 15 January 2025

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary

Species sensitivity distributions (SSDs) are cumulative probability distributions that are used to estimate Hazard Concentrations (HC_x) - the concentration of a chemical that is expected to affect a given $x\%$ of species. HC_5 values, which are intended to protect 95% of species, are often used for the derivation of environmental quality criteria and ecological risk assessment for contaminated ecosystems (Posthuma et al., 2001). The Hazard Proportion (HP_u) is the proportion of species affected by a given concentration x .

ssdtools is an R package (R Core Team, 2024) to fit SSDs using Maximum Likelihood (Millar, 2011) and estimate HC_x and HP_u values by model averaging (Schwarz & Tillmanns, 2019) across multiple distributions (Thorley & Schwarz, 2018). The shinyssdtools R package (Dalgarno, 2021) provides a Graphical User Interface to ssdtools.

Since the publication of Thorley & Schwarz (2018), the ssdtools R package has undergone two major updates. The first update (v1) included the addition of four new distributions (inverse Pareto, Burr Type III and the log-normal log-normal and log-logistic log-logistic mixtures) and a switch to the R package TMB (Kristensen et al., 2016) allowing full control over model specification. The second major release (v2) includes critical updates to ensure that the HC_x and HP_u estimates satisfy the *inversion principle* as well as bootstrap methods to obtain confidence intervals (CIs) with more appropriate coverage (Fox et al., 2024).

Statement of need

SSDs are a practical tool for the determination of safe threshold concentrations for toxicants in fresh and marine waters, and are implemented in some form for risk assessment and water quality criteria derivation throughout multiple jurisdictions globally (BC Ministry of Environment and Climate Change Strategy, 2019; Lepper, 2005; US EPA, 2020; Warne et al., 2018).

The selection of a suitable distribution has been identified as one of the most important and difficult choices in the use of SSDs (Newman et al., 2000). Since the original implementation (v0), ssdtools (Thorley & Schwarz, 2018) has used model averaging to allow estimation of HC_x and HP_u values using multiple distributions (Schwarz & Tillmanns, 2019). The method, which is described in detail by Fox et al. (2021) in the SSD context, provides a level of flexibility and parsimony that is difficult to achieve with a single distribution.

Technical details

Distributions

Ten distributions are currently available in `ssdtools`. The original version (v0) of `ssdtools` provided the two parameter log-normal (`lnorm`), log-logistic (`llogis`), log-Gumbel (`lgumbel`, also known as the inverse Weibull), gamma (`gamma`), Weibull (`weibull`) and Gompertz (`gompertz`) distributions. In the first major update (v1), the two parameter inverse Pareto (`invpareto`), three parameter Burr Type III (`burrIII3`) and five parameter log-normal log-normal (`lnorm_lnorm`) and log-logistic log-logistic (`llogis_llogis`) mixture distributions were added. Together with the Burr Type III, the inverse Pareto and inverse Weibull provide the underlying distributions of the SSD fitting software `BurrIIOz` (Barry & Henderson, 2012) while the mixture distributions were added to accommodate bimodality (Fox et al., 2021). Since v1, `ssdtools` has by default fitted the `lnorm`, `llogis`, `lgumbel`, `gamma`, `weibull` and `lnorm_lnorm` distributions.

Model Fitting

In the first major update (v1), the dependency `fitdistrplus` (Delignette-Muller & Dutang, 2015) was replaced by `TMB` (Kristensen et al., 2016) for fitting the available distributions via Maximum Likelihood (Millar, 2011). The move to `TMB` means the likelihood function is hand coded in C++, which allows full control over model specification and improved handling of censored data. The change is internal and does not directly affect the user interface.

Model Averaging

In both the original (Thorley & Schwarz, 2018) and updated versions, the Akaike Information Criterion (AIC), AIC corrected for small sample size (AICc) and Bayesian Information Criterion (BIC) can be calculated for each distribution (Burnham & Anderson, 2002). Information criterion based model weights have the properties $0 \leq w_i \leq 1$ and $\sum_{i=1}^m w_i = 1$ where w_i is the weight of the i^{th} of the m models (Burnham & Anderson, 2002). Except in the case of censored data, `ssdtools` uses AICc based weights for model averaging.

The first two implementations of `ssdtools` used the weighted arithmetic mean to obtain a model-averaged estimate of HC_x :

$$\widetilde{HC}_x = \sum_{i=1}^m w_i HC_x^{(i)}$$

where $HC_x^{(i)}$ is the HC_x estimate for the i^{th} model.

The weighted arithmetic mean is conventionally used for averaging model parameters or estimates (Burnham & Anderson, 2002). However, in the case of HC_x and HP_u values, the estimator \widetilde{HC}_x fails to satisfy the *inversion principle* (Fox et al., 2024) which requires

$$[HP_u]_{u=HC_\theta} = \theta$$

This inconsistency has been rectified in `ssdtools` v2 by estimating the model-averaged HC_x (denoted \widehat{HC}_x) directly from the model-averaged cumulative distribution function (*cdf*)

$$G(u) = \sum_{i=1}^m w_i F_i(u)$$

where $F_i(\cdot)$ is the *cdf* for the i^{th} model and w_i is the model weight as before. \widehat{HC}_x is then obtained as the solution to

$$u : G(u) = x$$

or, equivalently

$$u : G(u) - x = 0$$

for the proportion affected x . Finding the solution to this last equation is referred to as *finding the root(s)* of the function $G(u) - x$. As of `ssdtools` v2, methods such as `ssd_hc()` and `ssd_hp()` now use the *inversion principle* by default when `multi_est = TRUE`. To estimate the values using the original weighted arithmetic mean set `multi_est = FALSE`.

Confidence Intervals

`ssdtools` generates confidence intervals for HC_x and HP_u values via bootstrapping. By default all versions of `ssdtools` use parametric bootstrapping for non-censored data as it has better coverage than the equivalent non-parametric approach used in some other SSD modelling software such as `BurrIioz` (see [Fox et al., 2022](#)). The first two versions of `ssdtools` both calculated the model averaged CI from the weighted arithmetic mean of the CIs of the individual distributions. Unfortunately, this approach has recently been shown to have poor coverage ([Fox et al., 2024](#)) and is inconsistent with the *inversion principle*.

Consequently, v2 also offers a parametric bootstrap method for non-censored data that uses the joint cdf to generate data before refitting the original distribution set and solving for the newly estimated joint cdf (see details in [Fox et al., 2024](#)). This so-called 'multi' method can be implemented with (`ci_method = "multi_free"`) and without (`ci_method = "multi_fixed"`) re-estimation of the model weights. In order to implement the 'multi' method of bootstrapping described above, v2 also provides the probability density (`ssd_pmulti()`), cumulative distribution (`ssd_qmulti()`) and random generation (`ssd_rmulti()`) functions for multiple distributions.

However, although the 'multi' method has good coverage it is computationally slow. To overcome this limitation, the default method (`ci_method = "weighted_samples"`) provided by the current update is a faster heuristic based on taking bootstrap samples from the individual distributions proportional to their weights ([Fox et al., 2024](#)).

Plotting

As well as fitting SSDs and providing methods for estimating HC_x and HP_u values, from v1 `ssdtools` has extended the `ggplot2` R package ([Wickham, 2016](#)) by defining `geom_ssdpoint()`, `geom_ssdsegment()`, `geom_hcintersect()` and `geom_xribbon()` geoms and a discrete color-blind scale `scale_color_sdd()` for SSD plots. The current version (v2) adds `scale_fill_ssd()` for a discrete color-blind fill scale and `ssd_label_comma()` and `ssd_label_comma_hc()` for formatting of x-axis labels.

Example of use

The following code fits the six default distributions to the boron example data set from `ssddata` ([Fisher & Thorley, 2021](#)) and prints the goodness of fit table complete with information criteria:

```
library(ssdtools)
fits <- ssd_fit_dists(ssddata::ccme_boron)
ssd_gof(fits)

# A tibble: 6 × 9
  dist      ad      ks      cvm      aic      aicc      bic delta weight
  <chr>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 gamma    0.440 0.117 0.0554 238. 238. 240. 0.005 0.357
2 lgumbel  0.829 0.158 0.134 244. 245. 247. 6.56 0.013
3 llogis   0.487 0.0994 0.0595 241. 241. 244. 3.39 0.066
4 lnorm    0.507 0.107 0.0703 239. 240. 242. 1.40 0.177
```

```
5 lnorm_lnorm 0.320 0.116 0.0414 240. 243. 247. 4.98 0.03
6 weibull      0.434 0.117 0.0542 238. 238. 240. 0 0.357
```

The model averaged HC_5 estimate with 95% CIs can then be obtained using:

```
ssd_hc(fits, ci = TRUE)
```

```
# A tibble: 1 × 11
```

dist	proportion	est	se	lcl	ucl	wt	method	nboot	pboot	samples
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>	<dbl>	<I<list>>
1 average	0.05	1.26	0.819	0.397	3.41	1	parametric	1000	1	<dbl [0]>

And all of the distributions plotted via:

```
autoplot(fits)
```

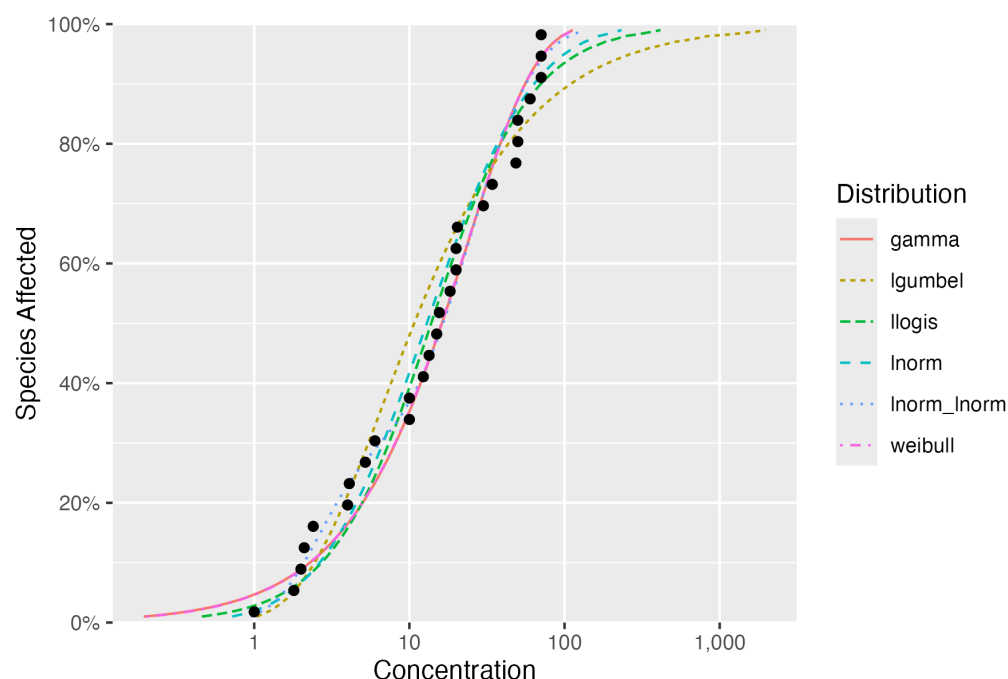


Figure 1: Species sensitivity distributions for the six default distributions with the Boron species concentration data.

The model averaged cdf with 95% CIs (with the model averaged HC_{10} indicated by a dotted line) can be plotted using:

```
predictions <- ssdtools::predict(fits, ci = TRUE)
ssd_plot(ssddata::ccme_boron, predictions,
         hc = 0.1, xlims = c(NA, 3000),
         shape = "Group", color = "Group", label = "Species",
         xlab = "Concentration (mg/L)"
) +
  scale_color_ssd()
```

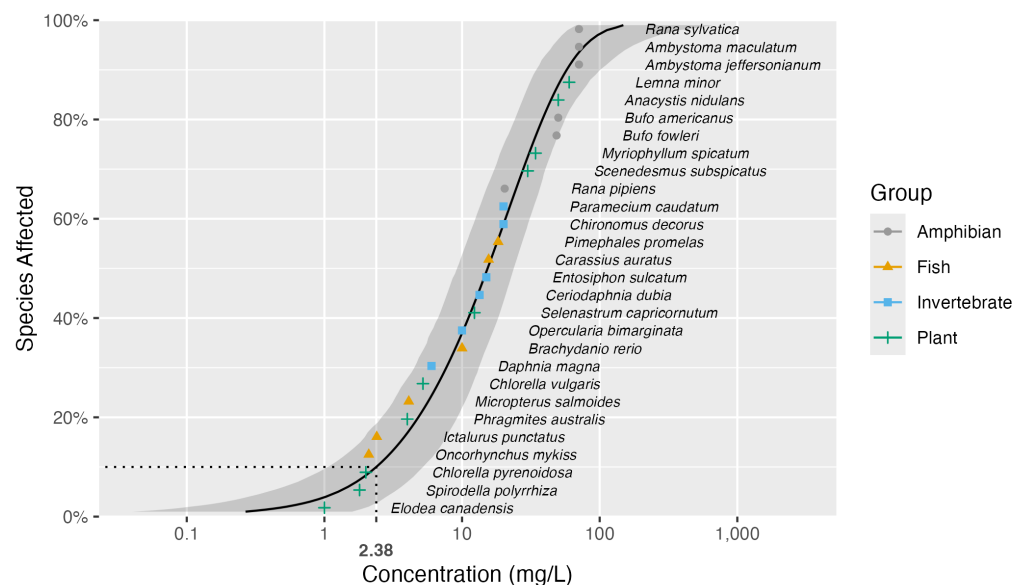


Figure 2: Model averaged species sensitivity distribution with 95% CI based on the six default distributions with Boron species concentration data. The HC_{10} value is indicated by the dotted line.

Acknowledgements

We acknowledge contributions from Angeline Tillmanns, Seb Dalgarno, Kathleen McTavish, Heather Thompson, Doug Spry, Rick van Dam, Graham Batley, and Ali Azizisharzi. Development of ssdtools was funded by the Ministry of Environment and Climate Change Strategy, British Columbia and the Department of Climate Change, Energy, the Environment and Water, Australia.

References

- Barry, S., & Henderson, B. (2012). Burrlioz 2.0 manual. *Commonwealth Science and Industrial Research Organisation*. <https://research.csiro.au/software/burrlioz>
- BC Ministry of Environment and Climate Change Strategy. (2019). *Derivation of Water Quality Guidelines for the Protection of Aquatic Life in British Columbia*. Water quality guideline series, WQG-06. Province of British Columbia, Victoria, BC, Canada. https://www2.gov.bc.ca/assets/gov/environment/air-land-water/water/waterquality/water-quality-guidelines/derivation-protocol/bc_wqg_aquatic_life_derivation_protocol.pdf
- Burnham, K., & Anderson, D. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer New York. <https://doi.org/10.1007/b97636>
- Dalgarno, S. (2021). shinyssdtools: A web application for fitting Species Sensitivity Distributions (SSDs). *Journal of Open Source Software*, 6(57), 2848. <https://doi.org/10.21105/joss.02848>
- Delignette-Muller, M., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, 64(4), 1–34. <https://doi.org/10.18637/jss.v064.i04>
- Fisher, R., & Thorley, J. (2021). ssddata: Species Sensitivity Distribution Data. <https://doi.org/10.32614/cran.package.ssddata>
- Fox, D., Fisher, R., & Thorley, J. (2024). *Final report of the joint investigation into SSD*

- modelling and ssdtools implementation for the derivation of toxicant guideline values in Australia and New Zealand. Report prepared for the department of climate change, energy, the environment and water.* Environmetrics Australia; Australian Institute of Marine Science. <https://doi.org/10.25845/xtvt-yc51>
- Fox, D., Fisher, R., Thorley, J., & Schwarz, C. (2022). *Joint investigation into statistical methodologies underpinning the derivation of toxicant guideline values in Australia and New Zealand.* Environmetrics Australia; Australian Institute of Marine Science. <https://doi.org/10.25845/fm9b-7n28>
- Fox, D., van Dam, R., Fisher, R., Batley, G., Tillmanns, A., Thorley, J., Schwarz, C., Spry, D., & McTavish, K. (2021). Recent Developments in Species Sensitivity Distribution Modeling. *Environmental Toxicology and Chemistry*, 40(2), 293–308. <https://doi.org/10.1002/etc.4925>
- Kristensen, K., Nielsen, A., Berg, C., Skaug, H., & Bell, B. (2016). TMB: Automatic differentiation and laplace approximation. *Journal of Statistical Software*, 70(5), 1–21. <https://doi.org/10.18637/jss.v070.i05>
- Lepper, P. (2005). Manual on the methodological framework to derive environmental quality standards for priority substances in accordance with article 16 of the water framework directive (2000/60/EC). *Schmallenberg, Germany: Fraunhofer-Institute Molecular Biology and Applied Ecology*, 15, 51–52. <https://www.wrrl-info.de/docs/manual-derivation-qs.pdf>
- Millar, R. B. (2011). *Maximum likelihood estimation and inference: With examples in R, SAS, and ADMB.* Wiley. <https://doi.org/10.1002/9780470094846>
- Newman, M. C., Ownby, D. R., Mézin, L. C. A., Powell, D. C., Christensen, T. R. L., Lerberg, S. B., & Anderson, B.-A. (2000). Applying species-sensitivity distributions in ecological risk assessment: Assumptions of distribution type and sufficient numbers of species. *Environmental Toxicology and Chemistry*, 19(2), 508–515. <https://doi.org/10.1002/etc.5620190233>
- Posthuma, L., Suter II, G., & Traas, T. (2001). *Species sensitivity distributions in ecotoxicology.* CRC press. <https://doi.org/10.1002/etc.4373>
- R Core Team. (2024). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. <https://www.R-project.org/>
- Schwarz, C., & Tillmanns, A. (2019). *Improving Statistical Methods for Modeling Species Sensitivity Distributions* (No. WSS2019-07). Province of British Columbia. https://a100.gov.bc.ca/pub/acat/documents/r57400/2_1568399094009_8398900200.pdf
- Thorley, J., & Schwarz, C. (2018). ssdtools: An R package to fit species sensitivity distributions. *Journal of Open Source Software*, 3(31), 1082. <https://doi.org/10.21105/joss.01082>
- US EPA. (2020). *Species sensitivity distribution (SSD) toolbox.* Center for Computational Toxicology; Exposure, Durham, NC. <https://doi.org/10.23645/epacomptox.11971392.v2>
- Warne, M., Batley, GE, van Dam, R., Chapman, J., Fox, D., Hickey, C., & Stauber, J. (2018). *Revised Method for Deriving Australian and New Zealand Water Quality Guideline Values for Toxicants – update of 2015 version.* Prepared for the Revision of the Australian and New Zealand Guidelines for Fresh and Marine Water Quality. Australian and New Zealand Governments and Australian state and territory governments, Canberra, 48 pp. <https://www.waterquality.gov.au/sites/default/files/documents/warne-wqg-derivation2018.pdf>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. <https://doi.org/10.1007/978-0-387-98141-3>