Risk Analysis, Vol. 29, No. 1, 2009 DOI: 10.1111/j.1539-6924.2009.01135.x

A Robust Approach to Risk Assessment Based on Species Sensitivity Distributions

Gianna S. Monti,[1] Peter Filzmoser[2] Roland C. Deutsch[3]

The guidelines for setting environmental quality standards are increasingly based on probabilistic risk assessment due to a growing general awareness of the need for probabilistic procedures. One of the commonly used tools in probabilistic risk assessment is the species sensitivity distribution (SSD) which represents the proportion of species affected belonging to a biological assemblage as a function of exposure to a specific toxicant. Our focus is on the inverse use of the SSD curve with the aim of estimating the concentration, HCp, of a toxic compound that is hazardous to p% of the biological community under study. Towards this end, we propose the use of robust statistical methods in order to take into account the presence of outliers or apparent skew in the data, which may occur without any ecological basis. A robust approach exploits the full neighbourhood of a parametric model, enabling the analyst to account for the typical real world deviations from ideal models. We examine two classic HCp estimation approaches and consider robust versions of these estimators. In addition, we also use data transformations in conjunction with robust estimation methods in case of heteroscedasticity. Different scenarios using real data sets as well as simulated data are presented in order to illustrate and compare the proposed approaches. These scenarios illustrate that the use of robust estimation methods enhances HCp estimation.

KEY WORDS: Hazardous concentration, Box-Cox transformation, Model fit, Monte Carlo simulations, Bootstrap, Robust Statistics

SUMMARY

The species sensitivity distribution (SSD) is a commonly used tool in probabilistic risk assessment. It represents the proportion of species affected belonging to a biological assemblage as a function of exposure to a specific toxicant. Different methods exist to estimate the concentration, HCp, of a toxic compound that is hazardous to p% of the biological community under study. Here, statistical estimation methods are proposed which are robust against outliers and more tolerant against strict data requirements, like (log-)normal distribution. A comparison between the traditional and the robust approaches is addressed, especially in presence of outliers or apparent skew in the data, which may occur without any ecological basis. In addition, data transformations are also used in conjunction with robust estimation methods in case of heteroscedasticity. Results show that the use of robust estimation methods enhances HCp estimation.

[1]Department of Economics, Management and Statistics, University of Milano-Bicocca, Milan, Italy, gianna.monti@unimib.it
[2]Institute of Statistics & Mathematical Methods in Economics, Vienna University of Technology, P.Filzmoser@tuwien.ac.at
[3]Institute of Statistics & Mathematical Methods in Economics, Vienna University of Technology, roland.c.deutsch@gmail.com

1. SPECIES SENSITIVITY DISTRIBUTIONS

Environmental risk assessment is a scientific process that identifies and evaluates adversely affected environments, in particular living organisms and their ecosystems. The purpose of risk assessment is to assess the severity and the likelihood of undesirable outcomes to the environment, which might arise due to exposure to a toxic compound. Environmental risk assessment may help risk managers to derive "safe" exposure levels so that the chances of these outcomes occurring are limited to an acceptable minimum e.g. environmental quality standards determined according to the European Water Framework Directive, [10].

The large interspecies difference in sensitivity delivers a high degree of uncertainty in the risk assessment. To reduce this uncertainty it would be opportune to test a wide number of species, however it would be unfeasible from a practical point of view. In this context, species sensitivity distributions (SSDs) are used to quantify the variation in sensitivity across species with probabilistic models under certain data limitations [22].

SSDs are indeed a tool that describe the interspecies variability in sensitivity to a toxic compound through a statistical distribution function such as the log-normal [2], the log-logistic [3] or the log-triangular distributions [13]. Less commonly, the Weibull distribution [31] – the preferred choice for heavy-tailed distributions – as well as nonparametric methods [21,33] are also employed. Others have also used interpolation techniques as the upper end of the distribution is sometimes undefined [6].

The SSD is an extrapolation model [22,2] with the aim of making inference from a few tested species (believed to be representative of the corresponding biological community) to the community level in order to predict the environmental impact of contaminants. In practice, the SSD is estimated by fitting a cumulative distribution function (CDF) to a sample of effect concentrations (such as the median effective concentration, EC50, or the no-observed-effect concentration, NOEC) derived from acute or chronic toxicity tests on the sample species, in which the species are arranged from the most to the least sensitive.

Of the two commonly used effect concentrations, SSDs derived from NOEC are usually considered more conservative than the ones derived from EC50. However, one must be aware that NOEC is not statistically determined, depends upon the experimental design and has no biological meaning. Whenever possible, statistically determined values indicating low or negligible effect (e.g. effective concentration of level 10, EC10, or 1, EC1) should be preferred. The data necessary for reliable estimation of these effect values are often not available or not reported in the literature. For this reason, robust statistical methods and their application as described herein are developed using assemblages of EC50 data, but would readily transfer to any other series of data.

In ecological risk assessment, SSD curves are used in two ways, commonly known as the forward and the inverse approach [30]. The goal of the former is to estimate the proportion of species affected (potentially affected fraction - PAF) at a pre-specified effective concentration. Conversely, the inverse approach is used to estimate the effective concentration associated with a PAF that is of interest to risk assessors, and thus is helpful in deriving environmental quality standards. In this context, the estimator is called *hazardous concentration of level* p (HCp) [30] and is defined as the concentration of a certain toxic compound considered hazardous for p% of species in a community. For our work presented herein, the focus lies on the latter, inverse approach and we will consider HCp-estimation for low values of p such as 0.01, 0.05 and 0.10. We would also like to emphasise that while in its original form, HCp was derived from NOEC data, we use it as a quantity derived from EC50 data due to the issues with NOEC estimation stated above. However, our methods do not depend on this choice and are readily transferrable to NOEC data.

The SSD-approach is not without its problems and the discussion on its applicability is lively and several works in the literature have appeared to contribute to the debate: Newman et al. [21] describe that the use of the log-normal distribution is frequently not supported in actual applications, where lack of data and the presence of outliers or skew can affect the quality of the model fit. An outlier may occur due to variability in the measurement or it may indicate an experimental error (in this case, exclusion from the data set is warranted). If the selected compounds have highly specific mechanisms of action, then the organisms should be homogeneous taxonomically when constructing an SSD curve. In this context, an outlier could be a point that differs from the others although there is no ecological reason that leads to its elimination. Skewness can be caused by different species behaviour for lower

and higher concentration levels. Consequently, the resulting HCp-estimates may fall at the low end of the range of PAF values. Moreover, some data sets may contain distinct taxonomic and ecological groupings[21,34] such that the assumed unimodal distributional model is not completely adequate.

As a first step in addressing these issues, an evaluation of the quality of the supplied data is highly recommended before any data analysis[18]. Proper outlier detection in SSDs, however, is only marginally discussed in Smith and Cairns[28], Gottschalk and Nowack[14] and sometimes outliers are altogether eliminated[17] reducing a frequently limited sample size even further. However, influential observations are of interest because their inclusion or exclusion in the analysis may lead to substantial changes in the numeric value of the HCp estimate.

In light of these issues, we propose a robust approach to SSD curve estimation. The advantage of this approach is that it exploits the full neighbourhood of a parametric model, thus enabling the analyst to account for typical real world deviations from ideal models such as the classes of normal or log-normal distributions. Robust statistical procedures aim to improve upon the classical parametric statistical results. This is done by taking into account that the assumed models which are employed by the analysts are only approximate. Robust inference produces results that are stable with respect to small changes in the data or to small model deviations[20,16].

Robust methods and their theoretical justification are widely used in statistics: from regression and multivariate analysis to generalised linear models and time series[20]. Procedures for robust estimation are widely available in various software packages, such as the R package `robustbase`[25,29]. To our knowledge, this is the first attempt of incorporating robust procedures in SSD analysis.

Herein, we focus on estimating of HCp by examining robust analogues to the existing classical statistical estimation methods. Furthermore, we also consider Box-Cox transformations of the original data in the presence of heteroscedasticity. We illustrate the usual approaches to SSD and HCp estimation in Section 2.1. Sections 2.2 to 2.4 expand these using a range of robust approaches and data transformations. In Section 3 we present several case studies in order to illustrate and compare the proposed methodology on real data. Section 4 reports the results of an extensive Monte Carlo simulation study with the aim to compare the proposed approaches under controlled circumstances in terms of HCp estimation. We conclude the presentation with a brief discussion of our results. All statistical computations described in the paper were performed using the statistical software environment `R`[23].

2. STATISTICAL METHODS FOR SSD ESTIMATION

2.1 SSD Model Formulation and Fitting Approaches

Let n denote the number of different species tested with respect to a certain compound and let x_i denote the effective concentration data, of the i^{th} species under study. Without loss of generality, we will refer in the following to x_i as an EC50 data point, a median effective/hazardous concentration value.

As the collected EC50s are consequently ordered in increasing fashion, we assume $x_i \leq x_j$ for $1 \leq i < j \leq n$. From these, the relative rank of each species, the so-called plotting positions y_i, are assigned to each of the collected EC50s. Numerous formulæ for these plotting positions have been suggested[27]. As the analysis assumes no a priori distribution of the x_i's, we use Weibull plotting positions, $y_i = i/(n+1)$. Furthermore, as species concentration levels, x, are commonly $\log_{10}$-transformed in toxicological studies, we arrive at the final data points $(\log_{10}(x_i), y_i)$.

A generic model formulation to fit an SSD curve to the observed data points (x, y) can be written as

$$y = F\big(\log_{10}(x), \boldsymbol{\beta}\big) + \varepsilon \,, \tag{1}$$

where $F(., \boldsymbol{\beta})$ describes the relationship between the (transformed) species concentration levels, x, and the corresponding plotting positions, $\boldsymbol{\beta}$ the unknown model parameters and ε the random error structure of the model. To fit SSD models, one of two approaches is generally used. The first, so-called direct approach is based on the standard assumption that EC50 values from the species in the environment under study are distributed according to a log-normal distribution[1,22]. Therefore, a simple SSD-curve model is given by

$$y = F\big(\log_{10}(x), \boldsymbol{\beta}\big) + \varepsilon = \Phi\left(\frac{\log_{10}(x) - \mu}{\sigma}\right) + \varepsilon \,, \tag{2}$$

where $\Phi(.)$ denotes the cdf of a standard normal distribution and $\boldsymbol{\beta} = (\mu, \sigma)$. This type of model (2) is frequently used by the European Commission[9] for fitting SSD-curves. The second approach applies a probit transformation to the plotting positions and

Table I . Coding scheme for the estimation methods based on underlying model (PI–plug-in; PR–probit-regression) and data transformation ($\log_{10}(x)$–log-transformed data; BC(x)–Box-Cox transformation). The modifiers c– and r– refer to the classic and robust approaches, respectively.

	Transformation $g(x)$		
Model	$\log_{10}(x)$	BC(x)	rBC(x)
Direct	c–PI, r–PI	/	/
Backward Regression	c–PR,r–PR	c–BC, r–BC	r–rBC

base the SSD-curve on a simple regression model for the data pairs $(\log_{10}(x_i), \Phi^{-1}(y_i))$, i.e.

$$y = \Phi(\beta_0 + \beta_1 \log_{10}(x)) + \varepsilon. \tag{3}$$

While this formulation is in principle equivalent to model (2), it allows for the use of the rich class of standard linear regression models for statistical inference (e.g. confidence intervals can be calculated based on the assumption of normally distributed errors). However, a major drawback of this approach, as detailed in Hickey and Craig [15, Section 4.2], is that the error terms are generally neither independent nor homoscedastic. Nevertheless, the model is quite popular and has been implemented by the US EPA in their CADDIS software package [12] as fitting a full model is recommended when a full characterisation of the relationship between exposure and PAF is desired as opposed to just deriving HCp for a single pre-specified value of p. In the following, we will refer to this model formulation as the inverse or backward probit regression model (the term backward regression refers to the fact that, unlike common regression models, the plotting positions are actually fixed, while the reported effective median concentrations are the random quantities).

2.2 SSD Curve Estimation

When employing the direct approach, the two parameters are estimated by the sample mean, $\hat{\mu} = \sum_{i=1}^{n} \log_{10}(x_i)/n$, and the sample variance, $\hat{\sigma}^2 = \sum_{i=1}^{n} (\log_{10}(x_i) - \hat{\mu})^2/(n-1)$. Plugging these estimators into (2) gives the so-called "classic" plug-in estimator [11] of the SSD-curve (c–PI in Table I). Conversely, the parameters of the inverse probit regression approach are estimated by the method of least-squares (LS), i.e. $\hat{\boldsymbol{\beta}}_{LS} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} r_i(\boldsymbol{\beta})^2$, where $r_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 g(x_i),\ i = 1, \ldots, n$, denote the residuals of the model fit. Inserted into (3) these estimators yield the "classic" probit regression estimator of the SSD curve (c–PR in Table I).

For a robust SSD curve estimator using the direct modelling approach (2), robust estimators of μ and σ of the normal distribution are called for [20,16], most commonly the median, $\tilde{\mu} = \text{median}_j(\log_{10} x_j)$, and the median absolute deviation (MAD), $\text{MAD} = 1.4826 \cdot \text{median}_i\{\ |\log_{10} x_i - \tilde{\mu}|\ \}$, respectively. By plugging these robust estimators into (2), we obtain the robust plug-in SSD-estimator of the SSD curve (r–PI in Table I).

More options are available when employing the inverse probit regression approach (3). At first, one could employ robust parameter estimation which typically uses the iteratively reweighted least squares (IRWLS) algorithm minimizing an objective function based on the residuals $r_i(\boldsymbol{\beta}) = y_i - \hat{\beta}_0 - \hat{\beta}_1 g(x_i)$. Among many choices, we use the class of MM-estimators [35]. This class extends the so-called M-estimator of the regression parameters defined as

$$\hat{\boldsymbol{\beta}}_M = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho\left(\frac{r_i(\boldsymbol{\beta})}{\hat{\sigma}}\right). \tag{4}$$

In this formulation, $\hat{\sigma}$ is a robust scale estimator of the residuals (depending on $\boldsymbol{\beta}$) and $\rho(.)$ is a bounded loss function defining the contribution of each scaled residual to the model fit [16]. However, M-estimators are sensitive to leverage points (outliers in the domain of explanatory variables). A fully robust HCp-estimator can be obtained using the class of MM-estimators, which also achieve high (tuneable) statistical efficiency. Their three-step computation is outlined below [20]:

(1) Determine an initial S-estimator [7] of the regression parameters as $\hat{\boldsymbol{\beta}}_S = \arg\min_{\boldsymbol{\beta}} \hat{\sigma}\left(r_1(\boldsymbol{\beta}), \ldots, r_n(\boldsymbol{\beta})\right)$, where $\hat{\sigma}$ is an M-estimator of the regression scale parameter found as the solution of $\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{r_i(\boldsymbol{\beta})}{\sigma}\right) = \delta$, with $\delta \in (0, \infty)$ a fixed constant.
(2) Based on the residuals from step 1, compute an M-estimator for σ.
(3) Compute an M-estimator of the regression parameters based on $\hat{\boldsymbol{\beta}}_S$ from step 1 and $\hat{\sigma}$ from step 2.

Implementing MM-estimation in the inverse probit regression model results in the robust probit regression estimator (r–PR in Table I).

2.3 HCp Estimation

Once the SSD is derived, one may obtain an estimator of HCp as the solution, x, of $p = F\big(\log_{10}(x), \hat{\boldsymbol{\beta}}\big)$. When considering the direct approach defined in (2), HCp can be derived as [30,22]

$$\log_{10}(\text{HCp}) = \mu - z_{1-p}\sigma, \tag{5}$$

where z_{1-p} is the $(1-p)^{\text{th}}$ lower percentile point of a standard normal distribution. Replacing the unknown model parameters with their respective estimators, classic as well as robust, yields the corresponding "plug-in" estimator, $\widehat{\text{HCp}}$. The standard error may be approximated by the delta method [5] when using the classic estimators. However, application of the delta method for robust estimators is not an easily applicable option and typically the standard error is approximated by a bootstrap estimator [11].

Therefore, to better compare the classic and robust plug-in estimators of HCp, we approximate SE(HCp) via bootstrap and compute confidence intervals via the Wald approximation:

$$\widehat{\text{HCp}} \pm z_{1-\alpha/2}\text{SE}(\text{HCp}). \tag{6}$$

When considering inverse probit regression as defined in (3), HCp is obtained by inverting the predictive regression line, i.e.

$$\log_{10}(\text{HCp}) = \frac{z_p - \beta_0}{\beta_1}. \tag{7}$$

Once more, we can supply the parameter estimators via the classic LS-method or the robust MM-estimators to obtain $\widehat{\text{HCp}}$. As in the direct approach, an estimator of SE(HCp) can also be best obtained via bootstrap and the corresponding confidence interval is computed as in (6).

2.4 Alternative Data Transformations

In cases where the log-normality assumption is violated and the inverse probit regression model (3) is used, the typical $\log_{10}$-transformation of the original data is untenable. In these cases we suggest to use the class of Box-Cox power transformations [4], which extend the $\log_{10}$-transformation described above. This kind of transformation is quite common in bioassay studies [26,24] for heteroscedastic data. One would therefore employ the transformation function $g(x) = (x^\lambda - 1)/\lambda,\ \lambda \neq 0$ on the original EC50s, where λ is typically chosen by a profile likelihood function. With this formulation, we obtain the extended inverse probit regression model

$$y = \Phi(\beta_0 + \beta_1 g(x)) + \varepsilon. \tag{8}$$

We wish to emphasise that for $\lambda = 0$ the classical Box-Cox transformation is $g(x) = \ln(x)$, but in this special case we take it to be the $\log_{10}$-transformation. If implemented successfully, the transformed data does follow approximately a normal distribution with stable variance.

The Box-Cox transformation can be used for both "classic" and robust SSD-curve and HCp-estimation and we denote estimators using these transformations by the acronyms c–BC and r–BC in Table I , respectively.

Marazzi and Yohai [19] proposed also a robust version of the Box-Cox transformation, choosing the parameter λ via a robust residual autocorrelation estimator. Therefore, we also combine the robust SSD-curve estimator based on the inverse probit regression with the robust Box-Cox transformation to yield the final model r–rBC in Table I .

3. AN APPLICATION WITH REAL DATA

To demonstrate the applicability as well as illustrating key differences in the models presented above, we fit these to a data set of distinct species tolerance values determined for an ecotoxicological risk assessment (ERA) case detailed in De Zwart [8].

3.1 Origin and Selection of Toxicity Data

The aquatic ecotoxicity database described in De Zwart [8] and in Hickey and Craig [15] comprises 30,369 acute (EC50 and LC50) and chronic (NOEC) records spanning 3,442 different substances over 1,549 species. This database provides a wide set of toxicity data, covering different endpoints and laboratory test conditions. Eleven fish species and 34 toxic compounds (31 insecticides and 3 herbicides) were selected. The endpoints for the different selected species of fish were 96-h EC50 (96 hours test). We have selected taxonomically comparable species in order to avoid polymodality of data. Furthermore, the selected compounds have a taxon-specific toxic mode of action.

3.2 Statistical Analysis

For the purpose of illustrating the characteristics of the methods from Table I , we selected four different chemicals yielding four distinct scenarios:

Aldrin (no outliers): for this compound, the recorded effective concentration values feature no apparent outliers and can be well described as log-

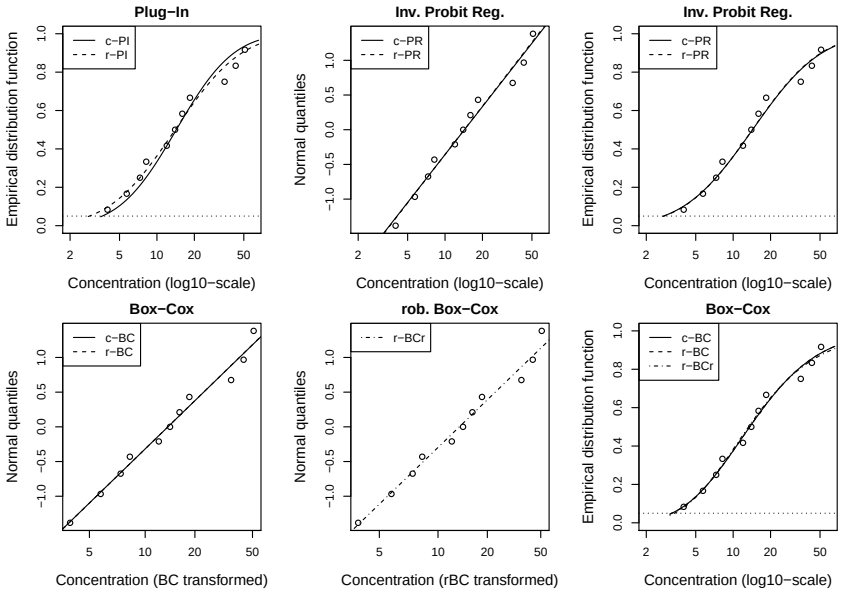

Fig. 1. Species sensitivity distribution estimates for the compound Aldrin (no outliers). The panel on the top left displays the empirical distribution function using the Weibull plotting positions from Section 2.1 and concentrations on a $\log_{10}$-scale. In addition, the SSD estimates of the classical (solid line) and robust (dashed line) plug-in estimators are superimposed. The middle upper panel displays the classical (solid line) and robust (dashed line) inverse probit regression estimates on a transformed y-axis, i.e. normal percentiles. These estimates are back-transformed to the risk scale in the top right panel. The estimates for classic (solid line) and robust (dashed line) regression are displayed on the bottom left panel with transformed y-axis. The bottom center panel displays the result for robust regression with robust Box-Cox transformation again on a transformed y-axis. All three estimates are then presented as back-transformed estimates in the bottom right panel. Furthermore, all the SSD-curves are plotted only for risk values of 5% (represented by the dotted horizontal line) and above.

Table II . Comparison of HC5 values (μg/L) calculated from the estimated SSD curves for different chemical compounds for the seven estimation methods: classic (c-) and robust (r-) plug-in (PI), probit-regression for log-transformed (PR) and (robust rBC) Box-Cox (BC) transformed data, along with 95% confidence intervals (CI) based on 1000 bootstrap samples.

	Compound			
Method	Aldrin	Endrin	Endosulfan	Carbaryl
c–PI	3.63	0.27	0.42	1656
CI	[2.89,4.30]	[0.13,0.42]	[0.16,0.69]	[1245,2039]
r–PI	2.86	0.43	0.68	2485
CI	[1.60,4.80]	[0.11,0.94]	[0.17,1.58]	[1165,3980]
c–PR	2.75	0.14	0.18	1235
CI	[2.32,3.16]	[0.05,0.25]	[0.02,0.41]	[929,1504]
r–PR	2.73	0.10	0.85	1994
CI	[2.27,3.15]	[0.03,0.19]	[0.35,1.59]	[1514,2448]
c–BC	3.21	0.45	0.99	772
CI	[2.84,3.52]	[0.39,0.51]	[0.94,1.03]	[526,1005]
r–BC	3.29	0.45	1.08	718
CI	[2.89,3.67]	[0.38,0.51]	[1.02,1.14]	[453,964]
r–rBC	3.50	0.51	0.99	2409
CI	[3.12,3.86]	[0.43,0.58]	[0.91,1.07]	[1735,3004]

normal (see Figure 1). In this case, all methods perform similarly and yield HC5 risk values between 2.7 μg/L and 3.6 μg/L (see Table II). Further, note that $\lambda \approx 0$ for both, the standard as well as the robust, Box-Cox transformations of x_i, thus making all regression based SSD-estimates indistinguishable from each other. Table II also shows 95% confidence intervals of $HC5^{(1)}$ based on 1000 bootstrap samples. These intervals are quite comparable, with the exception of r-PI that are a bit wider than the others.

Endrin (skewness): this is an example of a skewed distribution of effective concentrations (see Figure 2). In this case a Box-Cox or robust Box-Cox transformation proves useful as these transformations yield a more linear data pattern on the transformed concentration scale (see the upper center as well as the bottom left and bottom center panels in Figure 2). There is no practical difference in the use of classical or robust regression. The estimated HC5 risk values for the (robust) Box-Cox are approximately 0.45 μg/L, while omitting this transformation leads to a more conservative value, i.e. 0.14 μg/L for the inverse probit regression

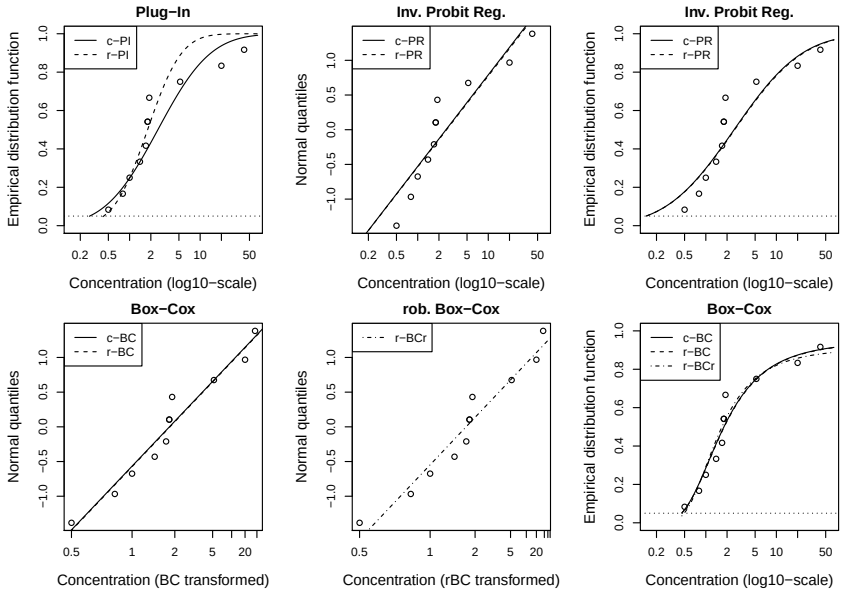

Fig. 2. Species sensitivity distribution estimates for the compound Endrin (skewed data). For a detailed description of the six panels refer to Figure 1.

(see Table II along with the corresponding 95% confidence interval). Further, note the striking differences between robust and classical plug-in estimates (upper left panel), with the former also yielding a HC5 value close to 0.45 μg/L.

Endosulfan (one upper outlier): the presence of an upper outlier (see Figure 3) yields large differences between the classical and robust SSD-estimates except when a Box-Cox transformation was used. Obviously, the classical methods are very much influenced by the outlier and tend to be overtly conservative for lower risk values. This is evidenced by a HC5 estimate of approximately 1 μg/L for the (reliable) robust methods, while the (non-reliable) classic methods are much lower (see Table II). However, both classical as well as robust Box-Cox transformations provide a transformation where the outlier follows more or less a linear trend, and thus no longer affects the classic regression estimate. Therefore, the resulting HC5 values are very similar.

Carbaryl (two lower outliers): In this case (shown in Figure 4), two lower outliers are evident. As expected this leads to quite different fits between the classical and the robust methods. The robust HC5 estimates (both plug-in and inverse regression) are both approximately 2000 μg/L or above, while the classical methods yield more conservative (lower) estimates (see Table II). An interesting difference can be found between the regression estimates based on classical Box-Cox and robust Box-Cox transformations: the non-robust version results in a transformation where all data points are forced into a linear trend, while for the robust version the outliers are allowed to deviate from the trend pattern of the remaining data points. The resulting deviation is visible in terms of large residuals, and the corresponding observations should be carefully investigated since the two methods result in very different HC5 estimates.

These four common scenarios provide a good overview on the differences in SSD-curve estimation between the methods introduced in Sections 2.1-2.4. However, these scenarios do not allow for a full characterisation and performance evaluation of the HCp-estimators. For this reason, we perform an extensive Monte Carlo simulation study outlined in the next section.

4. SIMULATION STUDY

In the following simulation study we focus on the estimation of HCp values, and compare the performance of the models from Table I . Specifically, we are interested in comparing low levels of risk, i.e. HC1, HC5 and HC10, which represent the concentrations of the toxicant hazardous to 1%, 5% and 10% of the biological community, respectively.

In order to investigate the robustness of the HCp estimators, we generated $M = 2,000$ Monte Carlo

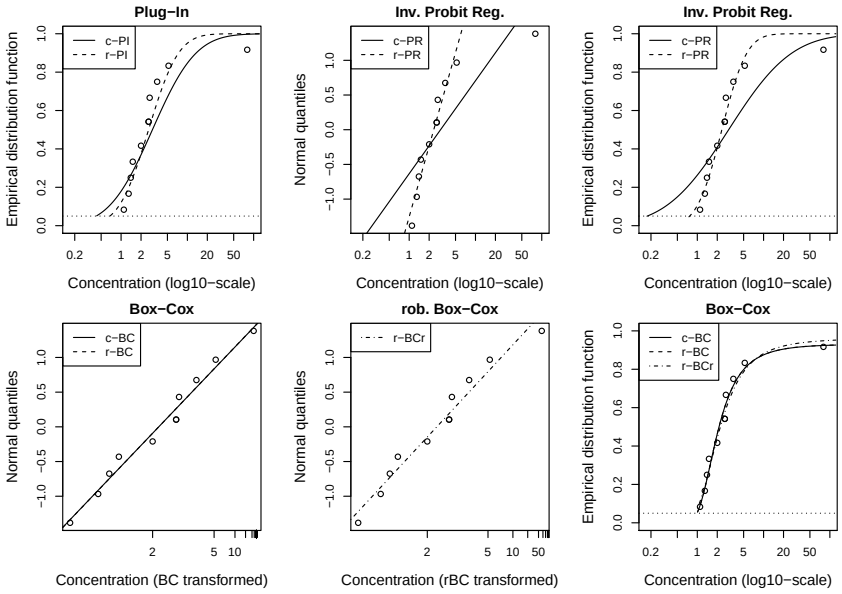

Fig. 3. Species sensitivity distribution for the compound Endosulfan (one upper outlier). For a detailed description of the six panels refer to Figure 1.

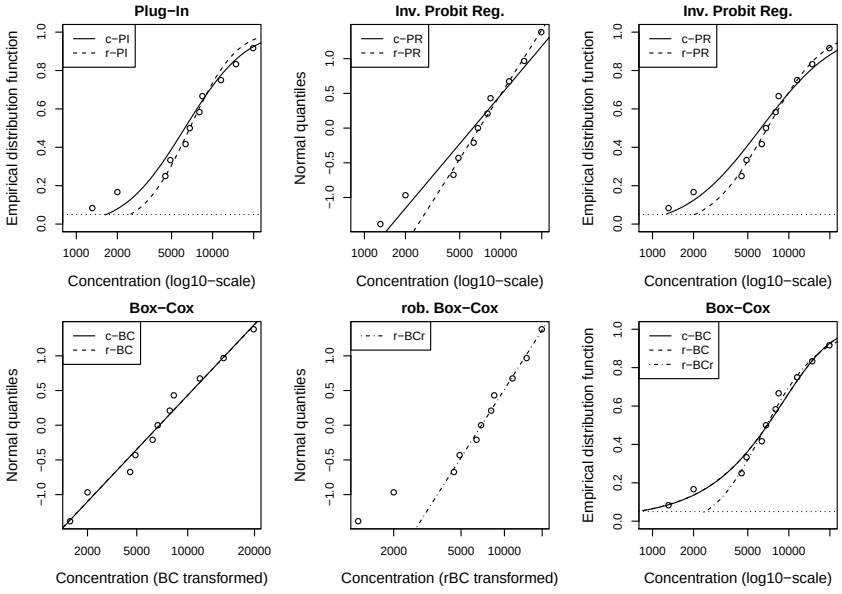

Fig. 4. Species sensitivity distribution for the compound Carbaryl (two lower outliers). For a detailed description of the six panels refer to Figure 1.

replications of the concentrations of $n = 15$ different species from a pre-specified distribution (see below). We evaluated the performance of the HCp estimators in two ways:

(1) **Distribution:** we display all replications within a pre-specified data distribution in a boxplot for each of the methods. These provide an overview of variability and bias across methods of $\widehat{\text{HCp}}$.

(2) **Coverage:** for all simulation settings we computed bootstrap-based confidence intervals for HCp based on 100 bootstrap replications. In each setting, empirical coverage rates were taken as the proportion of times

that the 95% bootstrap-based confidence interval covered the true HCp-value from the corresponding distribution used in the simulation run. Notice then that with the target coverage rate of 0.95 the approximate standard error of each estimated coverage rate is $\sqrt{(0.05)(0.95)/2000} = 0.005$ and does not exceed 0.012.

To study the sensitivity of the estimation methods to the underlying data distribution, we simulated data from six different distributions: log-normal and normal distribution, both commonly assumed to be underlying the data, two skewed distributions, exponential and a chi-squared distribution, and two more heavy-tailed distributions, t-distributions with 3 and 10 degrees of freedom (df). For each assumed data distribution, the HCp value was determined as the solution of $p = F(x)$ and is hereafter referred to as the "true" HCp value. The results for this first run are displayed in Figure 5 (Distribution) and Table III (Coverage). As the results are very similar for all risk values in all our simulation runs, we only report the ones concerning HC5 (complete results available from the authors).

Figure 5, shows that the performance of the estimators in terms of bias and mean squared error depends strongly on the underlying data distribution. While bias in estimation, i.e. the difference between the pre-determined "true" value of the parameter and its estimate, is evident across all cases, it is most pronounced for the exponential distribution, maybe due to the fact that this distribution approaches zero exponentially fast. The plug-in approaches perform well for log-normal and symmetric distributions, but bias is evident for skewed distributions. On the other hand, the inverse probit regression methods performs better for skewed distributions and are the methods of choice for heavy-tailed distributions (t3, t10). The Box-Cox transformations perform particularly well for the normal and log-normal distributions, but may introduce bias and larger variability in the other cases. It should also be noted that the variability in estimation is generally higher when using robust methods.

Table III shows the empirical coverage rates as derived from our simulations. The majority of methods as applied to different distribution shapes result in coverage rates frequently falling below nominal coverage, that is they cover the actual HC5 value less often than the desired nominal coverage rate of 0.95. This effect is most pronounced for cases where the underlying data distribution is skewed and is probably due to the fact that such distributions are less suited to describe the gradual increase of affected species at low concentrations [30]. On the other hand, coverage rates of 1 result from an upward bias of the estimated confidence intervals, which are based on the underlying skewed distributions.

Table III . Empirical coverage rates of the bootstrap c onfidence intervals for HC5 based on the seven estimation methods from Table I : classic (c-) and robust (r-) plug-in (PI), probit-regression for log-transformed (PR) and (robust - rBC) Box-Cox (BC) transformed data. The data was generated from six different distributions and the target coverage rate is 0.95.

	Method						
Model	c-PI	r-PI	c-PR	r-PR	c-BC	r-BC	r-rBC
lnorm	0.92	0.99	0.96	1.00	0.91	0.97	0.90
norm	0.83	0.96	0.98	1.00	0.93	0.98	0.98
exp	0.46	0.79	0.72	0.90	0.74	0.87	0.72
chi5	0.62	0.89	0.90	0.97	0.81	0.92	0.81
t3	0.86	0.93	0.98	0.99	0.97	1.00	0.97
t10	0.93	0.98	0.99	1.00	0.98	1.00	1.00

We also investigated the same performance criteria in the presence of outliers. We did this by generating data from a mixed distribution, where 90% (80%) of the data points were generated from a log-normal distribution with mean equal to 5 and standard deviation equal to 1 on the log scale ("lnorm(5,1)"). The remaining 10% (20%) of the simulated data points were generated as three types of shift-outliers from a log-normal distribution with standard deviation of 0.1 and mean values of 4, 6, and 8, respectively. Thus, in the first case (mean 4) the outliers are well within the range of the main data distribution, while the other situations lead to much heavier tails.

Comparing the results pictured in Figures 6 and 7, we come to similar conclusions for both outlier scenarios: the approaches based on (robust) Box-Cox perform quite well, independent from the position of the outliers. Inverse probit regression leads to small variance, but higher bias (underestimation), depending on the position of the outliers. In case of heavier tailed distributions, the robust plug-in approach performs better than the classical one in terms of bias and mean squared error.

Finally, the coverage of the 5 percent risk value shown in Tables IV and V also show a similar behaviour for the 10% and 20% outlier cases. Generally, the robust methods are closer to

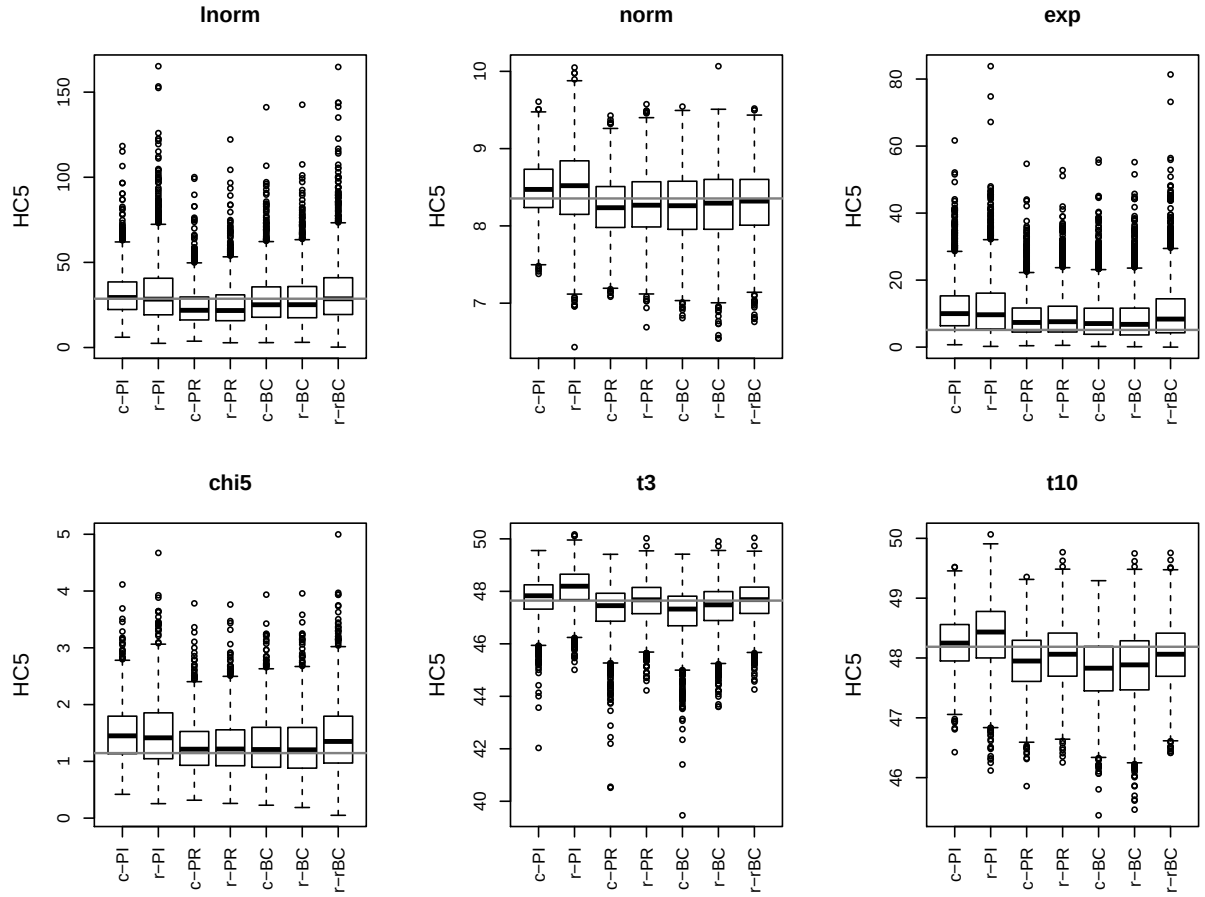

Fig. 5. Empirical distribution results for $\widehat{\text{HC5}}$. The panels represent the empirical distributions for the methods from Table I for each of six distributions (clockwise from top left): Lognormal(5,1), Normal(10,1), Exponential(5), t(10), t(3) and $\chi^2(5)$. The horizontal line in each panel indicates the true value of HCp, which corresponds to the theoretical p^{th} percentile point, for the respective underlying data distribution. For more details, please refer to the text.

Table IV . Average (trimmed) coverage of the 5 percent risk value. The data are simulated from log-normal distribution with one shift-outlier added. The estimation methods are classic (c-) and robust (r-) plug-in (PI), and probit-regression for log-transformed (PR) and (robust rBC) Box-Cox (BC) transformed data.

	Method						
Model	c-PI	r-PI	c-PR	r-PR	c-BC	r-BC	r-rBC
1*lnorm(4,0.1)	0.97	1.00	0.91	0.99	0.93	0.98	0.96
1*lnorm(6,0.1)	0.89	0.98	0.97	1.00	0.91	0.98	0.91
1*lnorm(8,0.1)	1.00	0.99	0.91	1.00	0.88	0.94	0.90

Table V . Average (trimmed) coverage of the 5 percent risk value. The data are simulated from log-normal distribution with two shift-outliers added.The estimation methods are classic (c-) and robust (r-) plug-in (PI), and probit-regression for log-transformed (PR) and (robust rBC) Box-Cox (BC) transformed data.

	Method						
Model	c-PI	r-PI	c-PR	r-PR	c-BC	r-BC	r-rBC
2*lnorm(4,0.1)	0.99	1.00	0.87	0.99	0.92	0.99	0.96
2*lnorm(6,0.1)	0.83	0.97	0.98	1.00	0.91	0.97	0.91
2*lnorm(8,0.1)	1.00	1.00	0.81	0.99	0.89	0.95	0.83

the nominal level than their classical counterparts. Overall, method r-BC shows the best behaviour.

5. DISCUSSION AND CONCLUSIONS

We presented a range of possibilities that extend the two common approaches (plug-in and inverse probit regression) in SSD-curve estimation

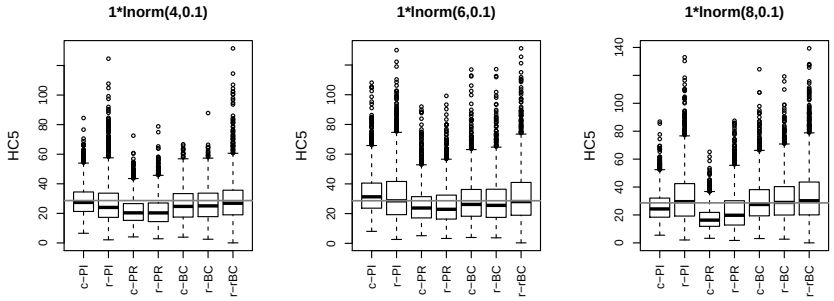

Fig. 6. Simulation results for the 5 percent risk value, compared to true value as determined by the simulation scenario. The data are simulated from log-normal distribution, and one shift-outlier is added.

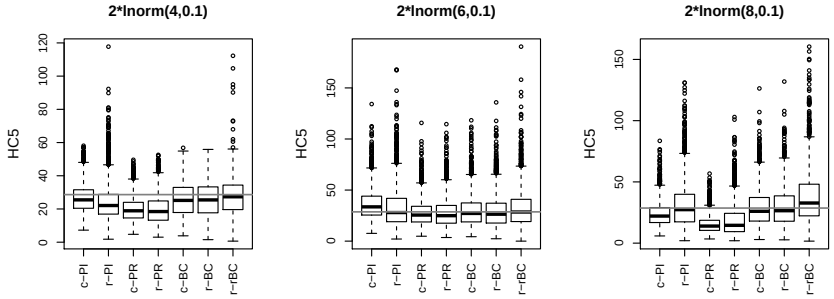

Fig. 7. Simulation results for the 5 percent risk value, compared to the the theoretical 5^{th} percentile point of the underlying simulation distribution. The data are simulated from log-normal distribution, and two shift-outliers are added.

to robust methods. These include robust approaches to parameter estimation as well as the often applied Box-Cox data transformation.

Based on sample data, we illustrated that robust methods manage to fit data with heavier tails well and can therefore be viewed as a flexible modelling approach able to cope with deviations from (log-)normality. In the case of high outliers, our example suggests that the robust SSD-curve estimators lead to estimates that more closely approximate the empirical data than the corresponding classic methods – except for cases where a Box-Cox transformation is employed. In these situations, the classical methods are strongly influenced by the presence of outliers, and tend to greatly underestimate HCp for small p.

If lower outliers are present, robust methods tend to emphasise on the remaining data values leading to generally higher estimates of HCp in the lower tail. Specifically, regression estimators based on Box-Cox transformations result in a linear trend while the robust Box-Cox transformation yields a linear trend only for the majority of the data which are not considered as outliers.

In case of skewed distributions, Box-Cox transformations on the data result in a more linear data pattern and bias is further reduced by the use of robust methods. Their combination is therefore the preferred choice for skewed data.

In addition, we performed a detailed simulation study in which we investigated the performance of the different HCp estimators. Of these, the robust probit regression for Box-Cox transformed data had the best overall performance in terms of bias, variability, reliability and coverage. In case of 'regular' data patterns, a plug-in approach or a robust inverse probit regression performs also well. In conclusion, we have determined that the use of robust methods can greatly enhance SSD-curve and HCp estimation as these are apparently more flexible in handling a wide range of data situations.

A few caveats need to be addressed: we observed that the coverage rate of the confidence intervals for HCp deviated in many cases from the nominal coverage of 95%, even in situations where the distributional assumptions were met. Additionally, the empirical coverage rates are quite sensitive to lowering the number of species under study as indicated by a preliminary simulation study not

discussed herein (results available from the authors). We attribute these results to the fact that due to the small number of data points generated (which is frequently present in practice) the Wald-type confidence intervals become inappropriate. As observed from the boxplots, empirical sampling distributions are rarely symmetric adding further evidence to the inadequacy of this popular confidence interval construction. Alternative confidence interval methods are called for. However, universal confidence interval construction methods for small sample sizes are still an open research problem and would exceed the scope of this paper.

One must also be aware that any point on the SSD curve represents one species in the community and that any species fills an ecological niche. The SSD and HCp approach represents a substantial improvement in comparison with the traditional deterministic approaches for the definition of a PNEC (Predicted No Effect Concentration) based on a few toxicity data and the application of uncertainty factors (see i.e. [9]). Nevertheless, the SSD approach should not be used blindly as the presence of a keystone or endangered species at the lower end of an SSD curve may represent a critical situation [32].

Therefore, the presence of outliers may warrant further investigation on whether these unusually high or low values are due to methodological reasons or the actual sensitivity of the species. Outliers are often incorrectly interpreted as "wrong" measurements, but in reality they can occur due to varying responses across species. In many cases, however, we are not dealing with obvious outliers being different from the remaining observations, but with a distribution that deviates from (log-)normality. Unlike strict parametric approaches that rely on idealised model distributions, robust methods are more flexible and lead to reliable results even in case of such deviations.

ACKNOWLEDGMENT

Research partially financially supported by the Italian Ministry of University and Research, FAR (Fondi di Ateneo per la Ricerca) 2014. We would like to thank Professor Marco Vighi for his constructive comments on our article and also for having made available the data for the application.

REFERENCES

1.Aldenberg, T. and Jaworska, J. (2000). Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. *Ecotoxicology and Environmental Safety*, 46(1):1 – 18.

2.Aldenberg, T., Jaworska, J., and Traas, T. (2002). Normal species sensitivity distributions and probabilistic ecological risk assessment. In: Posthuma, L, Suter, GW and Traas, TP Eds. *Species Sensitivity Distributions in Ecotoxicology*. Boca Raton, Lewis Publishers, pages 49–102.

3.Aldenberg, T. and Slob, W. (1993). Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. *Ecotoxicology and Environmental Safety*, 25(1):48 – 63.

4.Box, G. and Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society (B)*, 26:211–252.

5.Casella, G. and Berger, R. (2002). *Statistical Inference*. Duxbury Press, Belmont, CA, 2nd edition.

6.Cormier, S. and Suter II, G. (2013). A method for deriving water-quality benchmarks using field data. *Environmental Toxicology and Chemistry*, 32(2):255–262.

7.Davies, P. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, 15(3):pp. 1269–1292.

8.De Zwart, D. (2002). Observed regularities in SSDs for aquatic species. In: Posthuma, L, Suter, GW and Traas, TP Eds. *Species Sensitivity Distributions in Ecotoxicology*. Boca Raton, Lewis Publishers, pages 133–154.

9.EC (2003). *Technical Guidance Document (TGD) on Risk Assessment of Chemical Substances (2nd edition)*. European Commission, European Chemical Bureau, Joint Research Centre, EUR 20418 EN/2, Ispra, Italy.

10.EC (European Commission) (2000). Directive 2000/60/EC of the european parliament and of the council of 23 october 2000 establishing a framework for community action in the field of water policy. *Off. J. Eur. Communities*, L327:1–72.

11.Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York, NY, USA.

12.EPA, U. (2010). Causal analysis/diagnosis decision information system (caddis). *US EPA (Environmental Protection Agency)*, Office of Research and Development,

Washington, DC. Available online at https://www.epa.gov/caddis. Last updated September 23, 2010.

13. Erickson, R. and Stephan, C. (1985). *Calculation of the final acute value for water quality criteria for aquatic organisms.* National Technical Information Service, Springfield, VA, USA.
14. Gottschalk, F. and Nowack, B. (2013). A probabilistic method for species sensitivity distributions taking into account the inherent uncertainty and variability of effects to estimate environmental risk. *Integrated Environmental Assessment and Management*, 9(1):79–86.
15. Hickey, G. and Craig, P. (2012). Competing statistical methods for the fitting of normal species sensitivity distributions: Recommendations for practitioners. *Risk Analysis*, 32(7):1232–1243.
16. Huber, P. and Ronchetti, E. (2009). *Robust Statistics.* Wiley, New York, 2nd edition edition.
17. Jesenska, S., Nemethova, S., and Blaha, L. (2013). Validation of the species sensitivity distribution in retrospective risk assessment of herbicides at the river basin scalethe scheldt river basin case study. *Environmental Science and Pollution Research*, 20:6070–6084.
18. Klimisch, H., Andreae, M., and Tillmann, U. (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regulatory Toxicology and Pharmacology*, 25(1):1 – 5.
19. Marazzi, A. and Yohai, V. (2006). Robust Box–Cox transformations based on minimum residual autocorrelation. *Computational Statistics & Data Analysis*, 50(10):2752 – 2768.
20. Maronna, R., Martin, R., and Yohai, V. (2006). *Robust Statistics.* John Wiley & Sons, Ltd.
21. Newman, M., Ownby, D., Mzin, L., Powell, D., Christensen, T., Lerberg, S., and Anderson, B. (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.
22. Posthuma, L., Suter, G., and Traas, T. (2002). *Species Sensitivity Distribution in Ecotoxicology.* Lewis Publishers, Boca Raton, FL.
23. R Core Team (2015). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
24. Ritz, C. and Streibig, J. (2005). Bioassay analysis using R. *Journal of Statistical Software*, 12(5).
25. Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., and Maechler, M. (2015). *robustbase: Basic Robust Statistics.* R package version 0.92-4.
26. Sakia, R. (1992). The Box–Cox transformation technique: A review. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 41(2):pp. 169–178.
27. Shabri, A. (2002). A comparison of plotting formulas for the pearson type III distribution. *Journal Teknologi*, 36(C):61–74.
28. Smith, E. and Cairns, J. (1993). Extrapolation methods for setting ecological standards for water quality: statistical and ecological concerns. *Ecotoxicology*, 2(3):203–219.
29. Todorov, V. and Filzmoser, P. (2009). An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, 32(3):1–47.
30. van Straalen, N. (2002a). Theory of ecological risk assessment based on species sensitivity distributions. In: Posthuma, L, Suter, GW and Traas, TP Eds. *Species Sensitivity Distributions in Ecotoxicology.* Boca Raton, Lewis Publishers, pages 37–48.
31. van Straalen, N. (2002b). Threshold models for species sensitivity distributions applied to aquatic risk assessment for zinc. *Environmental Toxicology and Pharmacology*, 11(34):167 – 172.
32. Vighi, M., Finizio, A., and Villa, S. (2006). The evolution of the environmental quality concept: from the us epa red book to the european water framework directive. *Environmental Sciences and Pollution Research*, 13:9 – 14.
33. Wang, Y., Wu, F., Giesy, J., Feng, C., Liu, Y., Qin, N., and Zhao, Y. (2015). Non-parametric kernel density estimation of species sensitivity distributions in developing water quality criteria of metals. *Environmental Science and Pollution Research*, 22(18):13980–13989.
34. Wheeler, J., Grist, E., Leung, K., Morritt, D., and Crane, M. (2002). Species sensitivity distributions: data and model choice. *Marine Pollution Bulletin*, 45(112):192 – 202.
35. Yohai, V. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15(2):pp. 642–656.