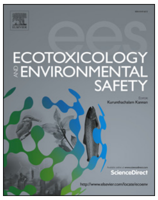
Ecotoxicology and Environmental Safety 125 (2016) 161–169

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

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



Species sensitivity distribution for chlorpyrifos to aquatic organisms: Model choice and sample size

Jinsong Zhao [a,*], Boyu Chen [a,b]

[a] College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China
[b] Suyu District Environmental Monitoring Station, Suqian 223800, China

ARTICLE INFO

Article history:
Received 9 July 2015
Received in revised form
29 November 2015
Accepted 30 November 2015
Available online 14 December 2015

Keywords:
Species sensitivity distribution
Aquatic organism
Organophosphate pesticide
Parametric and nonparametric bootstrap
Sample size
Model choice

ABSTRACT

Species sensitivity distribution (SSD) is a widely used model that extrapolates the ecological risk to ecosystem levels from the ecotoxicity of a chemical to individual organisms. However, model choice and sample size significantly affect the development of the SSD model and the estimation of hazardous concentrations at the 5th centile (HC_5). To interpret their effects, the SSD model for chlorpyrifos, a widely used organophosphate pesticide, to aquatic organisms is presented with emphases on model choice and sample size. Three subsets of median effective concentration (EC_{50}) with different sample sizes were obtained from ECOTOX and used to build SSD models based on parametric distribution (normal, logistic, and triangle distribution) and nonparametric bootstrap. The SSD models based on the triangle distribution are superior to the normal and logistic distributions according to several goodness-of-fit techniques. Among all parametric SSD models, the one with the largest sample size based on the triangle distribution gives the most strict HC_5 with 0.141 $\mu mol\ L^{-1}$. The HC_5 derived from the nonparametric bootstrap is 0.159 $\mu mol\ L^{-1}$. The minimum sample size required to build a stable SSD model is 11 based on parametric distribution and 23 based on nonparametric bootstrap. The study suggests that model choice and sample size are important sources of uncertainty for application of the SSD model.

1. Introduction

Chlorpyrifos (*O*,*O*-diethyl *O*-3,5,6-trichloropyridin-2-pyridinyl phosphorothioate) is one of the most widely used broad spectrum organophosphate pesticides. It acts on the nervous systems of insects by inhibiting acetylcholinesterase (Flaskos, 2012; Giesy et al., 2014). Chlorpyrifos is highly toxic to amphibians and fish and infiltrates aquatic ecosystems via runoff. Because of its long half-life in water, the effects of chlorpyrifos on aquatic ecosystems at different trophic levels are attracting more and more attention (Asselborn et al., 2015; Giddings et al., 2014; Khalil, 2015).

Now, most ecotoxicological data, such as median effective concentration (EC_{50}) or no observed effect concentration (NOEC), are obtained from short-term laboratory experiments on single species at different trophic levels (Vighi et al., 2006). Although they are reproducible, such experiments do not reflect environmental reality. They lack the endpoint and structure that are specific to an ecosystem, and they do not take the dynamic changes of ecosystems into account. Therefore, they lack serious ecological merit (Newman and Unger, 2003). The appearance and application of species sensitivity distribution (SSD) provides an effective way to extrapolate the ecotoxicological endpoint of species to the effects on the ecological community or higher levels (Posthuma et al., 2002).

The SSD model has attracted many researchers and governments to discuss and explore its principles, methods, assumptions, and applications (Newman et al., 2000). However, it still has some inherent issues, such as model choice, data quantity and quality, and sample size (Chapman et al., 1998; Dowse et al., 2013; Forbes and Forbes, 1993; Hopkin, 1993; Smith and Cairns Jr, 1993; van Straalen, 2002; Wheeler et al., 2002). These issues are the root of uncertainty in SSD models.

Currently, there is no theoretical evidence that SSD belongs to a specific distribution (Forbes and Calow, 2002; Xu et al., 2015). In practice, it is often assumed that the available ecotoxicological data of a chemical follows a certain specific probability distribution such as normal, logistic, triangle, or Burr Type III (Aldenberg and Jaworska, 2000; Aldenberg and Slob, 1993; Larras et al., 2013; Stephan et al., 1985; van Straalen, 2002; Xu et al., 2015). However, a simply chosen probability distribution may not be a good fit for a dataset with a large sample size because of its diverse error source and structure or for a dataset with a small sample size because of the difficultly in estimating the parameters. In such cases, it is necessary to use nonparametric approaches without any assumptions, such as basic bootstrap and modified bootstrap procedures that reduce the uncertainty brought by the model choice(Wang et al., 2008; Xing et al., 2014), but the power of these approaches is also affected by the

---

[*] Corresponding author.
E-mail address: jszhao@mail.hzau.edu.cn (J. Zhao).

http://dx.doi.org/10.1016/j.ecoenv.2015.11.039
0147-6513/

sample size (Newman et al., 2000).

Representative and valid ecotoxicological data are the basis for a reliable SSD model (Dowse et al., 2013). The assumption of the test species are sampled randomly from the ecosystem, and their response to a specific stress follows a particular probability distribution (Forbes and Calow, 2002; Posthuma et al., 2002). In fact, they are determined by their availability and are not randomly sampled (Wagner and Løkke, 1991). So far, there is no universal protocol to ensure data quality or to determine the minimum amount of ecotoxicological data for developing an SSD model.

To obtain the hazardous concentrations at the 5th centile (HC_5) of chlorpyrifos and evaluate the effects of model choice and sample size, a series of SSD models is built based on the acute ecotoxicological data (EC_{50}) of chlorpyrifos to aquatic organisms; detailed comparisons are performed among several parametric and nonparametric models with different sample sizes; and, finally, a minimum sample size that is vital to develop a usable SSD model is appraised.

2. Materials and methods

2.1. Data source

From the perspective of reducing uncertainty, EC_{50} of chlorpyrifos is chosen as the ecotoxicological data upon which to build SSD models. The EC_{50} of chlorpyrifos (see Table A) are obtained from ECOTOX, a ecotoxicological database provided by the U.S. Environmental Protection Agency (U.S. Environmental Protection Agency, 2015). To ensure reliability, the EC_{50} obtained following standardized protocols in the laboratory are chosen. In the case of multiple data points for one species, the geometric mean is used as the estimation of EC_{50} for the species (European Commission, 1996; Raimondo et al., 2008; RIVM, 2001).

To quantitatively compare the influence of data sources on an SSD model, three datasets are grouped based on different experimental media and exposure types, namely EC_{50} from (1) both marine and freshwater media (FW&SW) with the ratio 1:4 for data from marine media to freshwater media, (2) only freshwater media (FW), and (3) only freshwater under static exposure (FW|S).

2.2. Choice and building of SSD model

To evaluate the uncertainty from choosing SSD models, the widely used normal (Aldenberg and Jaworska, 2000), logistic (Aldenberg and Slob, 1993; Larras et al., 2013), and triangle (Stephan et al., 1985) distributions are selected to build the SSD model.

The cumulative distribution function of the logistic distribution is:

$$F(x;\ \mu,\ s) = \frac{1}{1 + e^{-\frac{x-\mu}{s}}}$$

where x is the EC_{50} of chlorpyrifos in logarithmic scale, and μ and s are 2 parameters of the probability distribution function.

The cumulative distribution function of normal distribution is:

$$F(x;\ \mu,\ \sigma) = \Phi(\frac{x-\mu}{\sigma}) = \frac{1}{2}[1 + \mathrm{erf}(\frac{x-\mu}{\sigma\sqrt{2}})]$$

where μ and σ are 2 parameters of the probability distribution function, and erf is the error function.

The cumulative distribution function of the triangle distribution is:

$$F(x;\ a,\ b,\ c) = \begin{cases} 0,\ x < a \\ \frac{(x-a)^2}{(b-a)(c-a)},\ a \le x \le c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)},\ c < x \le b \\ 1,\ b < x \end{cases}$$

where a and b are the lower and upper threshold, and c is mode of the probability distribution function.

All parameters are estimated by the maximum likelihood approach (Venables and Ripley, 2002). The confidence interval of each parameter is obtained based on the profile likelihood method. The Kolmogorov–Smirnov and Anderson–Darling tests are used to evaluate the goodness-of-fit (Stephens, 1986). Akaike information criterion (AIC) and Bayesian information criterion (BIC) are used as measures for selecting the optimum SSD model.

2.3. Uncertainty in SSD model and HC_5

To quantitatively address the uncertainty of the SSD model affected by the EC_{50} of chlorpyrifos, parametric and nonparametric bootstrap are used to generate a simulated dataset (Davison and Hinkley, 1997) on which a corresponding SSD model is fitted with maximum likelihood estimation. The median and confidence interval of each parameter are used to evaluate the uncertainty in the parameters of the SSD model.

To quantitatively measure the uncertainty in HC_5 of chlorpyrifos to aquatic organisms, one nonparametric and three parametric methods are used. For parametric methods, HC_5 is obtained from the SSD model fitting to the normal, logistic, and/or triangle distribution based on (1) the original datasets of EC_{50}, simulated data with the same sample size generated by (2) parametric bootstrap and (3) by nonparametric bootstrap; and for the nonparametric method, HC_5 is estimated (4) directly from simulated data generated by nonparametric bootstrap. The confidence interval for HC_5 is estimated with the delta method (Oehlert, 1992) for case (1) and the percentile method for the other 3 cases.

2.4. Determination of minimum sample size

To determine the minimum sample size, the HC_5s based on a series of simulated samples with different size are estimated. The procedure is described as follows: For a given sample size, 5001 samples are generated from the parent sample using random sampling with replacement (Davison and Hinkley, 1997). Based on those samples, the HC_5 and its confidence interval can be estimated according to Section 2.3. This process is replicated for each sample size from 2–N, the sample size of the parent sample.

The minimum sample size is detected by change point analysis on HC_5 obtained with above procedure. A change point analysis is used to determine a tipping point that separates the response variable (HC_5 in this study) into two groups, each with distinct characteristics, such as the mean and the variance (Qian et al., 2003). The tipping point, which corresponds to the minimum sample size that is the lowest requirement of the ecotoxicological data points for building a stable SSD model, is detected based on the variance changing under a significance level at 0.05.

All simulations and statistical analyses are performed with *R* (Version 3.11) (R Core Team, 2014), a language and environment for statistical computing, and add-on packages for *R* such as *fitdistrplus* (Delignette-Muller and Dutang, 2015), *triangle* (Carnell, 2013), and *ADGofTest* (Bellosta, 2011) for building SSD models and evaluating HC_5 as well as *changepoint* (Killick and Eckley, 2014) for detecting minimum sample size.