Comparing statistical techniques for uncertainty assessment of species sensitivity distributions: effect of sample size.

F. Verdonck[1], J. Jaworska[2], C. Janssen[1] and P.A. Vanrolleghem[1].
[1] Ghent University, Belgium
[2] Procter & Gamble, ETC, Belgium

A species sensitivity distribution (SSD) reflects variability in sensitivities of various species to a toxicant. In EU environmental risk assessment regulations, the 5th-percentile of the SSD (HC5: hazardous concentration) is used to set the quality criteria. Several techniques are available to characterize uncertainty of a SSD and HC5, that give different results. Once a technique has been selected, the question arises how many species are needed to calculate HC5 with a desired accuracy.
A comparison is made between maximum likelihood estimation (assuming lognormal and loglogistic distribution), Bayesian approaches (assuming lognormal distribution) and nonparametric bootstrapping (using Hazen plotting and the interpolated empirical distribution function) techniques to characterise uncertainty and variability using small toxicity data sets. Using the parametric techniques and assuming a lognormal or loglogistic distribution results in accurate HC5 confidence intervals, provided the underlying distribution is the proper one. However, these techniques are very sensitive to outliers. Nonparametric techniques overestimate the HC5 confidence intervals but do not depend on an assumption of an underlying distribution and are less sensitive to outliers.
The choice of an appropriate sample size (for determining a SSD) is essential. Sample sizes can be very small (five data points already appeared to be sufficient) when a lognormal or loglogistic distribution was the real underlying one and the parametric techniques were used. Ten data points were sufficient for the simulated data sets when using the nonparametric bootstrap with the interpolated empirical distribution function and the Hazen plotting system.