The Curse of the Rainbow Trout

Graeme L. Hickey*,†; Peter S. Craig†; Frank Coolen†; Andy Hart‡

Introduction

Probabilistic (Ecological) Risk Assessment has recently become a well studied topic. In particular, the principle of Species Sensitivity Distributions (SSDs) and Assessment Factors (AFs) (*a.k.a.* extrapolation and safety factors) has propelled much research. However, in this field there have been a large number of assumptions made which current research have cast doubt upon, including the current proposed methodologies – some of which are implemented in European countries.

My research looks into one of these assumptions; the idea that a community of species is exchangeable. It follows on from EFSA (2005) and Craig (2006) to propose a collection of estimators for setting environmental safety limits for hazardous substances.

Background

An SSD is a probability distribution function $F(\cdot)$ representing the probability that a random species drawn from the relevant assemblage has its toxicological endpoint violated at a given environmental log-concentration (EC) y. It is usually assumed that toxicity data is log-transformed. This is then used to define the Potentially Affected Fraction (PAF) of species in the assemblage at risk from an EC y; $F(y)$. If F represents a Gaussian CDF with mean μ and standard deviation σ, then the PAF = $\Phi(y)$.

SSDs can be used directly in a *forward sense*, *i.e.* to estimate the PAF given an EC y; or used in an *inverse sense*, *i.e.* to estimate an EC y given a required PAF. It is often useful to perform the latter for setting regulatory safety limits and/or pesticide registration. Therefore, given $p \equiv$ PAF, then $y^*=F^{-1}(p)$. However, we introduce uncertainty in the problem by not knowing μ and σ.

Up until now it has been assumed that the small samples of toxicity data for a community could be envisaged as realisations from the F. I.e.

$$y_{1,\ldots,}y_n \sim N(\mu,\sigma^2)$$

where $y_1,\ldots,y_n$ is a conditionally independent sample of (log-) toxicity data from n distinct species present in the community. Using this data and parametric assumptions researchers have tried to estimate the PAF y^*. Statistically, this is analogous to estimating the p^{th} percentile of a distribution.

It is typical that an estimator for y^*, often referred to as the (Log-) Hazardous Concentration to p% of the community (LHC_p), is of the form

$$\bar{y} - \kappa_p \cdot s$$

where $\bar{y}$ and s are the mean and sample standard deviation for the log-toxicity data. κ_p is the assessment (shift-) factor; and $\kappa_p s$ is the Assessment Factor (AF) which is a function of p. If μ and σ are known, then $\kappa_p = K_p$ – the (100-p)$^{\text{th}}$ percentile of a standard normal distribution.

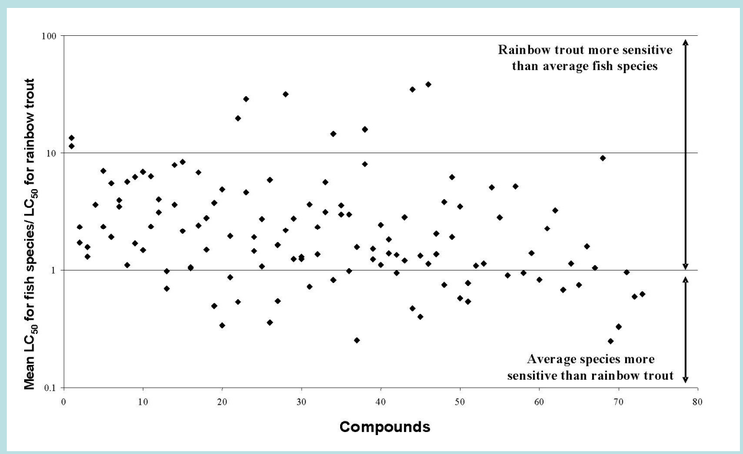
Aldenberg and Jaworska (2000), EFSA (2005) and Craig (2006) have proposed different estimators for κ_p based on different assumptions.

On Non-Exchangeability

There is recognisable evidence that the Rainbow trout is typically more sensitive than other species; *i.e.* the Rainbow trout will tend to feature in the lower half of an SSD. We therefore say that this species is *non-exchangeable* with the other species. A non-parametric hypothesis test applied to a large RIVM aquatic database indicated that this was case; in fact the Rainbow trout featured in the lower sensitive region in 72% of the datasets and was the most biased species as ranked by significance values.

EFSA (2005) noted that if one considers the ratio of the geometric mean of the other n-1 toxicity values to y^* (the toxicity statistic for the Rainbow trout) for each of the N datasets analysed, then one will infer that the Rainbow trout is typically more sensitive than the other species.

In the figure (right) points that lie above 1 imply that the rainbow trout is more sensitive than the other species for that particular substance.



We also explored using the Bayesian Information Criterion as a means to select optimal models by sequentially adding in species effects. It was found that the Rainbow trout was the first to be included under all the modelling assumptions we explored. This included the standard model, but also models where the SSD variance was assumed homogeneous with other "similarly" judged substances, *e.g.* same toxic mode of action.

Re-Modelling

We propose a change to the standard exchangeability assumption. As such, we introduce a predictive distribution for the *special species* (here the Rainbow trout, though it could be another biased species). Furthermore, we maintain the standard assumptions for the remaining n-1 test species (and other species in the community). Our new model for the toxicity data of sample size n is now:

$$y^* \sim N\left(\mu - k, [\varphi\sigma]^2\right)$$
$$y_i \sim N(\mu,\sigma^2) \text{ for } i = 1,\ldots n-1$$

where y^* is the special species' toxicity value. A few important points to mention are:

(1) This is a modified version of a proposal in EFSA (2005). It was assumed there that the predictive mean was μ-$k\sigma$ so that its position was unaffected by the SSD variability.

(2) k and φ are assumed to be known. It is therefore required that we estimate them from a sufficiently large (and relevant) database so that uncertainty has little impact on inferences.

(3) k and φ are difficult to estimate in this simpler model. However, this is in trade-off for mathematically tractable and transparent risk calculations.

Another important model assumption was also proposed in EFSA (2005). If there is access to a large relevant database of substance/species toxicity values such that the substance under current assessment (SUCA) is 'similar' to those in the database, then we augment the model as follows. We assume, from a Bayesian perspective (though applicable in a frequentist paradigm), that $\sigma_1^2,\ldots,\ \sigma_N^2$ are distributed Inverse-Gamma(α,β), *a priori*. We can then simultaneously estimate α and β from the database alongside k and φ. We will refer to this extended model as Model 2. The former model, *i.e.* where we only estimate the non-exchangeability parameters from additional substance-toxicity data, we refer to this as Model 1.

On the $\log_{10}$ scale, we estimated the species non-exchangeability parameters for the two modelling assumptions, as well as α and β for Model 2. Estimation was based on standard non-informative prior distributions.

	k	φ	α	β
Model 1	0.195	0.701	NA	NA
Model 2	0.205	0.656	1.523	0.314

We concluded that the uncertainty in estimating the parameters above was small enough to consider them known. They are applicable to the Rainbow trout only. Moreover, there is little difference between k and φ based on Model 1 and Model 2.

Estimators

Under the assumption of exchangeability, the estimators for the LHC_p under Model 1 are typically of the form

$$\bar{y} - \kappa_p \cdot s$$

The Aldenberg and Jaworska (2000) class of estimators for κ_p with certainty level γ is

$$\kappa_p = \frac{1}{\sqrt{n}} F_T^{-1}\left(\gamma; n-1, K_p\sqrt{n}\right)$$

where F_T is the CDF of a non-central t-distribution with n-1 degrees of freedom and non-centrality parameter $K_p\sqrt{n}$.

The EFSA (2005) estimator of κ_p is

$$\kappa_p = F_t^{-1}(p; n-1)\sqrt{1+\frac{1}{n}}$$

where F_t is the CDF of a Student t-distribution with n-1 degrees of freedom. For Model 2 the estimators are typically of the form

$$\bar{y} - \kappa_p \cdot s_{adjusted}$$

where

$$s_{adjusted} = \sqrt{\frac{(n-1)s^2 + 2\beta}{(n-1)+2\alpha}}$$

The estimator of κ_p remains the same as before in this instance.

If one assumes that a special species was tested and knows the non-exchangeability parameters, then it can be shown that the *optimal* estimators for Model i (i=1,2,) are of the form:

$$\hat{\mu} - \kappa_p^* s_i \text{ for } i = 1,2$$

Where

$$\hat{\mu} = \frac{\varphi^{-2}(y^*+k) + \sum_{i=1}^{n-1} y_i}{\varphi^{-2}+n-1},$$

$$s_1^2 = \frac{1}{n-1}\left[\varphi^{-2}(y^*+k-\hat{\mu})^2 + \sum_{i=1}^{n-1}(y_i-\hat{\mu})^2\right],$$

and

$$s_2^2 = \frac{(n-1)s_1^2 + 2\beta}{(n-1)+2\alpha}.$$

The optimal estimator of κ_p^* for Model 1 based on Aldenberg and Jaworska's work is:

$$\kappa_p^* = \frac{1}{\sqrt{\varphi^{-2}+n-1}} F_T^{-1}\left(\gamma; n-1, K_p\sqrt{\varphi^{-2}+n-1}\right)$$

The optimal estimator of κ_p^* for Model 1 and 2 based on the EFSA (2005) risk measure is:

$$\kappa_p^* = F_t^{-1}(p; n-1)\sqrt{1+\frac{1}{(\varphi^{-2}+n-1)}}$$

When one has determined the optimal estimate of the LHC_p, one can then compare it to the Predicted Environmental Concentration (PEC). One may also apply other assessment factors to account for other uncertainties *e.g.* Acute to Chronic ratios *etc.* If

$$\frac{PEC}{HC_p} < 1$$

then one *may* not authorise the registration of the toxicant. If the ratio is >1 then, assuming coherence of higher tier risk assessments, one would grant permission for registration of the toxicant.

References

Aldenberg, T. and Jaworska, J. S. (2000). Uncertainty of the Hazardous Concentration and Fraction Affected for Normal Species Sensitivity Distributions. Ecotoxicol. Environ. Safety. **46** 1-18.

Craig, P. S. (2006). Uncertainty Factors in Ecotoxicological Risk Management. *Unpublished.*

EFSA (2005). Opinion of the Scientific Panel on Plant Health, Plant Protection Products and their Residues on a Request from EFSA Related to the Assessment of the Acute and Chronic Risk to Aquatic Organisms with Regard to the Possibility of Lowering the Uncertainty Factor if Additional Species were Tested. The EFSA Journal. **301** 1-45.

This research was kindly funded by the EPSRC and a Defra-CSL Seedcorn Grant.

*g.l.hickey@durham.ac.uk