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) (i.e., extrapolation and safety factors) has propelled much research. However, in this field there have been a large number of assumptions made which cannot research have cast doubt upon, including the current propos-
ed 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 from on 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 $F(y)$ representing the probability that a random species drawn from the relevant assemblage has its toxicological endpoint violates at a given environmental 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 $y$. If $F$ represents a Gaussian CDF with mean $\mu$ and standard deviation $\sigma$, then the PAF is $(\Phi(y))$. 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 given a required PAF. It is often useful to perform the latter for setting regulatory safety limits and/or pesticide registrations. Therefore, given $p \neq$ PAF, then $\Phi^{-1}(p)$.

However, we introduce uncertainty in the problem by not knowing $\mu$ and $\sigma$.

Up until now it has been assumed that the small samples of toxicity data for a community could be envisaged as realisations from the $F$. $Y =_Y \sim N(\mu, \sigma^2)$

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 assume 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 the 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.

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 additional methodologies to re-estimate an SSD. We 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 or otherwise "similarly" judged substances, e.g. same toxic mode of action.

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 assume 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 the 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.

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 additional methodologies to re-estimate an SSD. We 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 or otherwise "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:

$$y_i = \mathbf{N}(\mu, \sigma^2) \quad \text{for } i = 1, \ldots, n-1$$

where $y_j$ 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 $\mu$ so that its position was unaffected by the SSD variability.

(2) $\mu$ and $\sigma$ 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) $\mu$ and $\sigma$ are difficult to estimate in this simple model. However, this is in trade-off for mathematically tractable and transparent risk calculations.

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 $\hat{\mu}$ and $\hat{\sigma}$ based on Model 1 and Model 2.

Estimators

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

$$\hat{\mu} = \hat{Y}_{\hat{K}_p}$$

The Aldenberg and Jaworska (2000) class of estimators for $\mu$, with certain level $\gamma$ are typically of the form

$$\hat{\mu} = \hat{Y}_{\hat{K}_p} \quad \text{for } i \geq 1,2$$

where $F_y$ is the CDF of the non-centre $\gamma$ distribution with $\alpha$ degrees of freedom and non-centrality parameter $\kappa_\gamma$. The EFSA (2005) estimator of $\kappa_\gamma$ is

$$\hat{\kappa}_\gamma = F_y^{-1}/\{c\gamma-1, K_{\gamma}\}$$

where $F_y$ is the CDF of a Student $\gamma$ distribution with $\gamma$ degrees of freedom and non-centrality parameter $\kappa_\gamma$. For Model 2 the estimators are typically of the form of

$$\hat{\gamma} = \hat{\gamma}_{\gamma_{\text{adj}}}$$

where

$$\gamma_{\text{adj}} = \{\gamma-1\}^2 + 2\gamma \{\gamma-1\} + 2\gamma$$

The estimator of $\gamma$ remains the same as before in this instance.

When one has determined the optimal estimate of the LHC, 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

$$\text{PEC} < 1$$

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 Hazard Concentration and Fraction Affected for Normal Species Sensitivity Distributions (SSDs) and Assessment Factors (AFs) Ed methodologies – some of which are implemented in European countries. Another important model assumption was also proposed in EFSA (2005). If there is access to a large relevant database of substances/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 $\mu$, $\sigma^2$ are distributedInverse-Gamma($\alpha$, $\beta$) a priori. We can then simultaneously estimate $\mu$ and $\sigma$ from the database alongside $\alpha$ and $\beta$. 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 as Model 1.

On the LHC scale, we estimated the species non-exchangeability parameters for the two modelling assumptions, as well as $\mu$ and $\sigma$ for Model 2. Estimations were based on standard non-informative prior distributions.