A Critique of the Use of Distribution-Based Extrapolation Models in Ecotoxicology
Author(s): T. L. Forbes and V. E. Forbes
Source: *Functional Ecology*, Vol. 7, No. 3 (Jun., 1993), pp. 249-254
Published by: British Ecological Society
Stable URL: http://www.jstor.org/stable/2390202
Accessed: 08/10/2013 15:55

Functional Ecology 1993
7, 249–254

ESSAY REVIEW

A critique of the use of distribution-based extrapolation models in ecotoxicology

T. L. FORBES and V. E. FORBES
Department of Marine Ecology and Microbiology, National Environmental Research Institute, Frederiksborgvej 399, PO Box 358, DK-4000 Roskilde, Denmark

Extrapolation in ecotoxicology

A fundamental unanswered question in ecotoxicology concerns the extent to which ecosystem-level effects of pollutants can be understood or predicted from tests at lower levels of organization (Forbes & Forbes 1993). Much attention during the 1970s and early 1980s has been directed towards developing new and better test methods and identifying ideal test species, indicator organisms and biomarkers. Given the impracticality of testing all species for their sensitivity to toxicants, reliance has usually been put on data gathered for a few selected species. In contrast to toxicological studies where data for a few surrogate species are extrapolated to humans, ecotoxicological testing requires extrapolation from a small number of test species to a vast number of species varying in taxonomy, size, life history, physiology and geographic range (Cairns & Mount 1990).

In this paper we examine the features and assumptions of recently developed distribution-based extrapolation models that are currently in use (OECD 1991; Van Leeuwen *et al.* 1992). All these presume an underlying interspecific distribution of sensitivities to a toxicant. We suggest that the rational use of these models as ecotoxicological tools for environmental regulation requires additional basic knowledge in two areas. The first concerns the relationship between structure and function in ecosystems. The second involves the nature of the statistical distribution of toxicity end-points in natural assemblages of species. Because of deficiencies in basic knowledge in these areas, statistical extrapolation does not presently offer an improvement over much simpler arbitrary assessment factors. We conclude that the added complexity inherent in their use is not outweighed by the benefits obtained.

Distribution-based extrapolation models

The legitimacy of extrapolating from acute to chronic toxicity, from toxicity in one species to that in another, and from single-species toxicity to ecosystem effects has been debated at length (Forbes & Forbes 1993). Recently, a series of distribution-based extrapolation methods has been proposed for evaluating the effects of chemicals on all species in a community or ecosystem from single-species data (Stephan & Rogers 1985; Kooijman 1987; Van Straalen & Denneman 1989; Aldenberg & Slob 1991; Wagner & Løkke 1991). These methods differ primarily in the assumed underlying distribution of species sensitivities, often measured as No Observed Effect Concentrations (NOEC). The goal of these procedures is to allow predictions of effects on a larger array of species based on laboratory toxicity data for very few (i.e. 3–8) (Stephan & Rogers 1985; Van Straalen & Denneman 1989; Aldenberg & Slob 1991; Wagner & Løkke 1991). The strongest proponents of these methods have even claimed that they can be '. . . considered to be an acceptable approach for the protection of the structure and function of aquatic ecosystems' (Van Leeuwen *et al.* 1992, p. 271).

Extrapolation models evolved from earlier 'assessment', 'application' or 'safety' factor approaches that attempted to estimate toxicant concentrations above which adverse effects on ecosystems would be expected to occur (OECD 1991). Here we define an assessment factor as simply a number that is applied to single-species toxicity data to adjust the effect concentrations [$L(E)C_{50}$, NOEC, etc.] to estimate a maximum acceptable toxicant concentration for a community or ecosystem (OECD 1991).

Distribution-based extrapolation models are typically based on a number of assumptions related to the biological variability of toxicity. In theory they can be used to quantify a limited type of higher-level response to a toxicant. If the assumptions hold, an exposure level at which, say, only 1% of the community or assemblage would be adversely affected can be defined.

Model assumptions

The utility of these models hinges critically on the validity of the assumptions. The two core assumptions are as follows (OECD 1991):

1. The distribution of species sensitivities in natural ecosystems (or statistical population of interest—

whatever it may be) closely approximates the postulated theoretical distribution. The log-triangular, log-logistic and the log-normal distributions have all been suggested as possibilities (OECD 1991). In practice, the selection of very few species as input into these models (i.e. 3 to 11) means that most tests of the shape of the resulting distribution will have very low power.

2. The sensitivities of species used in laboratory tests provide an *unbiased* measure of the variance and mean of the sensitivity distribution of species in natural communities.

Wagner & Løkke (1991) stressed the need to select test species that are representative of the ecosystem because, they argued, if the species represent a small part of the distribution curve, the variance among species will be underestimated and the protection concentration for the toxicant in question will be too high. Wagner & Løkke (1991) also argued that 'A large variation of end-points and taxa in combination with a limited number of test species may cause serious errors in the extrapolation procedure because of wrong assumptions of the distribution model or the lack of representativeness of the selected taxa.' The precise meaning of 'representative' is left unclear. In practice, 'representatives' are selected from different taxa and trophic levels. This may result in the estimation of a toxicity distribution that is biased and thus one in which the variance is either under- or overestimated. For example, selecting a very sensitive species, a moderately sensitive species and a tolerant species, as is often suggested, will overestimate the true community variance because the extreme values are sampled in greater proportion than they actually exist. Conversely, selecting a sensitive species from each of several taxa will underestimate the community variance. Ideally the selection should be random. However, it may be difficult to design sampling schemes in which random rather than 'representative' samples are selected unless the species composition of the community, ecosystem or statistical population of interest has been determined.

If the claims of ecological relevance such as those of Van Leeuwen *et al.* (1992) are to be considered valid, then the following two additional assumptions are required:

3. By protecting species composition, community function is also protected. However, community structure and function are often uncoupled (Schindler 1987; Gray 1989) and, indeed, there are strong theoretical grounds for not expecting a simple mapping between ecosystem structural and functional components (O'Neill *et al.* 1986 and references therein). O'Neill *et al.* (1986) provided convincing evidence that due to differences in spatiotemporal constraints, community or ecosystem structure and function cannot be easily or consistently coupled.

4. Interactions among species in communities/ecosystems can be ignored. This may often be invalid as is illustrated by the common occurrence of 'keystone' species. For example, the structure of the intertidal macrofaunal community on the north-west coast of the USA is largely controlled by a single species of starfish, *Pisaster orchraceus* (Paine 1966, 1974). The extent to which influential species are either more or less susceptible to the effects of pollutants is unknown.

In order to apply these models confidently we need, at least, to know the conditions under which these assumptions definitely do not apply. It would also be helpful to have some idea of how sensitive the extrapolations are to violations of the assumptions. We discuss these problems in the context of the current work done by the Organization for Economic Cooperation and Development (OECD) on extrapolation methods (OECD 1991). We have focused on the OECD report on extrapolation methods in some detail because of its importance in regulatory decision making.

The importance of operational definitions

The lack of operationally defined concepts in the past has been at the root of a number of problems that have seriously impeded hypothesis testing and theory development in ecology (Peters 1991). Ecotoxicologists need to address this problem as well. Until the biological system of interest is defined precisely, it cannot be sampled properly, and the validity of the extrapolation models cannot be determined. We are forced to define the ecosystem, community, or assemblage we wish to protect. Open acknowledgement of this places the problem in proper perspective. For example, Ricklefs (1979) defines ecosystem as 'all the interacting parts of the physical and biological worlds'. For the purpose of testing extrapolation model assumptions, this definition is of little use. We can never test whether the hypothesis that the 'ecosystem sensitivity distribution' follows, for example, the log-logistic distribution unless we know what constitutes the ecosystem. Ultimately, the definition must involve political and cultural as well as scientific considerations because society (or its representatives) must consider the question of what qualifies for protection. Thus the first step in what should be the pre-use evaluation of the current set of extrapolation models must be the development of an operational definition of the ecosystem or community because that is the focus of the modelling effort. We suggest that it is more profitable to think of sensitivity distributions as applying to communities or, even more appropriately, assemblages, rather than ecosystems because of the complete lack of an abiotic component in the theoretical basis of the extrapolation models.

The official OECD approach to the evaluation of extrapolation models seems to take a very casual approach to the objections we have mentioned above. For example, a workshop comparing newly developed extrapolation methods was held in December 1990 in Arlington, Virginia, USA and actually recommended that:

> *The use of the extrapolation methods for refined effects assessment* needs to be carried forward based on species sensitivity distributions, i.e. the log-triangular, the log-logistic, and the log-normal methods (OECD 1991, p. 38). (Emphasis added.)

and

> Because these extrapolation models have reasonable underlying assumptions, and have been *accepted for regulatory use*, there are good bases for their application in assessing the hazards of chemicals (OECD 1991, p. 21). (Emphasis added.)

But how do we know the assumptions *are* reasonable? The answer is that we do not. There are currently no acute or chronic toxicity data based on random samples from a well-defined ecosystem that can be compared to any of the hypothetical species sensitivity distributions. Despite claims to the contrary, the validity of the distributional assumption with regard to natural ecosystems has not yet been adequately demonstrated. For example, the workshop report cites the laboratory toxicity data used by Kooijman (1987) [which was originally compiled and reported by Sloof, Canton & Hermens (1983)], as data which were '. . . a random selection of species'. From what statistical population was it a random sample?

As part of the justification of the approach there has been a statistically misguided tendency to attempt to identify representative species. For the most part, these are those already used in standardized laboratory toxicity tests. For example, the draft report of the Arlington workshop (OECD 1991) stated that:

> . . . variation among species is not a random error since selection procedures require a *diversity* of organisms in the data set. (Emphasis added.)

This is true. However the report continues with:

> This diversity is intended to increase the likelihood of having both tolerant and sensitive species, *in order to improve the estimation* of the desired percentile (p. 33). (Emphasis added.)

This is incorrect. The proper goal of these studies should be to estimate the 'desired percentile' as accurately and reliably as possible. This is best done by randomly sampling the statistical population of interest, which must be clearly defined before it can be sampled. Selection of a 'diversity of organisms' that is somehow 'representative' of the 'ecosystem' will only act to bias the estimated mean and variance of the hypothetical distribution to some unknown degree.

Given that the assumptions of the statistically based extrapolation models have yet to be adequately tested, it is unclear why the OECD report should recommend the use of these much more complicated and cumbersome methods in preference to the previously used and much simpler assessment factors. Previous approaches have recommended that these simple factors be applied to values from chronic or semi-chronic data sets for a given chemical. For example, dividing the lowest of three laboratory-determined NOEC by 10, where one value each has been obtained for a fish, a crustacean and an alga, has often been used in the past for screening new and existing chemicals (OECD 1991).

We think it is fair to suggest that in order to be approved for use by the OECD, the distribution-based extrapolation methods should offer some unique or distinct advantage over the much simpler and cruder methods employed to date, especially if their use is advocated before they have been thoroughly validated.

Table 1 is a condensed version of one found in the draft report from the Arlington workshop on extrapolation (OECD 1991, also their Table 1). It was used to compare levels of concern calculated using chronic NOEC data determined for a limited set of organisms (an alga, a crustacean and a fish) as well as values for 11 (eight additional) species using both new and old extrapolation methods. The first column

Table 1. Comparison of 'extrapolation methods' based on laboratory NOEC values for eight selected compounds. The data are from an OECD draft report of the workshop on extrapolation methods (1991). See text for detailed discussion

Compound	Lowest NOEC divided by 10	95CV	Difference	% Difference
K_2CrO_4	0·01	0·010	0	0
NaBr	1	0·21	−0·79	−79
TPBS	0·1	0·042	−0·058	−58
2,4 DCA	0·0032	0·0080	0·0048	150
p-NT	0·1	0·10	0	0
DNOC	0·01	0·0027	−0·0073	−73
Dimethoate	0·0032	0·00036	−0·00284	−89
Pentachlorophenol	0·0032	0·00030	−0·0029	−91

of extrapolated values in Table 1 corresponds to the lowest NOEC from the set of three species divided by 10, and the second is what was termed the 'log-logistic distribution 95% confidence value' (here abbreviated 95CV). By definition, the 95CV is a value determined such that it will protect 95% of the species comprising the hypothetical ecosystem sensitivity distribution 95% of the time, given the validity of the underlying assumptions (OECD 1991; Aldenberg & Slob 1991; Wagner & Løkke 1991).

Are the newer and much more complex extrapolation methods an improvement over the previous simple and arbitrary methods? We note that although the OECD draft report recommended carrying forward the use (not just the testing of assumptions) of these refined extrapolation methods, we could find no specific statements in the report justifying this conclusion. Because these methods have been recommended for use with three or more data values and because there is a general consensus among ecotoxicologists that chronic or semi-chronic NOEC are more relevant as a basis for extrapolation, we compared the predictions generated using the lowest NOEC from the set of three species divided by 10 and the 95CV method using the same data set evaluated in the OECD report (Table 1).

Comparing the two columns we note that two predictions are the same regardless of the method used. Of the remaining six, the 95CV method predicts five values which range from approximately 60 to 90% lower. In the case of 2,4 DCA, the 95CV method predicts a value that is higher by 150% (Table 1).

At first it looks as if the 95CV method may tend to be a little more conservative and possibly predict lower values in general. In order to test the hypothesis that the two methods produce values that are indeed different, we compared them using the non-parametric Wilcoxon signed ranks test for paired observations (i.e. two observations on each chemical) (Hollander & Wolfe 1973). We used a non-parametric test because the variances of the two groups were extremely heterogeneous (s^2_{NOEC} 0·345, s^2_{95CV} 0·074; Bartlett's test, df=1, P=0·0006). We also know from Monte Carlo simulations that estimated 95CV values, especially for small numbers of test species (≤5), deviate markedly from normality [see Aldenberg & Slob (1991) their Fig. 3].

Comparison of the paired results for each of the eight chemicals in columns 2 and 3 in Table 1 using the Wilcoxon signed ranks test provides very little evidence for a difference in the predictions of the two extrapolation methods (P=0·116). We conclude that advocating the newer methods simply on the basis of different or more conservative predictions cannot be justified.

If the newer extrapolation methods generate values indistinguishable from the arbitrary factor methods, perhaps their greater theoretical underpinning can justify the increased complication. While it may be possible to verify the distributional assumptions for some operationally defined system of interest, an additional and more serious theoretical problem remains. This problem involves the well-recognized lack of incorporation of higher-level ecological processes (Van Leeuwen 1990; OECD 1991).

Structure, function and pollutants

The degree to which structure and function can be uncoupled during pollutant stress is currently under debate by ecotoxicologists (reviewed in Forbes & Forbes 1993). One school of thought maintains that detectable structural changes will occur before functional ones are manifest. The proponents of this line of reasoning suggest that feedback mechanisms at the population, community and ecosystem levels may often tend to preserve function even while significant structural changes are occurring (e.g. Schindler 1987; Gray 1989). At the other end of the spectrum are those who maintain that ecosystem functional changes in response to pollution are either likely to occur first or are more easily detected than structural changes (e.g. Crow & Taub 1979). Yet ecosystem theory suggests that what we are able to observe as the structural and functional components of ecosystems cannot always be simply related to each other (O'Neill *et al.* 1986).

This dual organization arises from structural constraints that operate on organisms and functional constraints that operate on processes (see O'Neill *et al.* 1986 and references therein). An exception to this general uncoupling would be expected to occur under conditions where a single species dominates a function, i.e. in communities that possess a 'functional keystone species' (after Paine 1966). While data are lacking, it seems reasonable to suggest that there are species whose removal would affect ecosystem processes (O'Neill *et al.* 1986). Finally, we would expect ecosystems with impoverished biota to show a much greater structure–function coupling when perturbed by pollutants. If there is little or no functional redundancy, then removal of a species may severely affect ecosystem processes. We suggest that much of the apparent inconsistency of results to date may be resolved by studies of functional redundancy and the degree of structural complexity of pollutant-stressed communities.

Parsimony and the choice of tools

But what about the method of dividing the lowest of a set of NOEC data by some assessment factor? The most often cited objection concerns the arbitrary and non-theoretical nature of the factor. An additional important objection is as follows. As data for more species are accumulated for a particular chemical, the lowest value of the lot can only get lower. Thus as more toxicity information becomes available for a compound, the 'acceptable' concentrations must decrease. This makes many scientists and regulators uneasy. A better approach would be to divide the median or geometric mean value of the available data set by some arbitrary value. This would be relatively robust to wildly different input values based on selections of 'representative' species. Choice of the median can have the additional advantage of making no distributional assumptions. The method could be made as conservative as desired simply by adjusting the extrapolation or assessment factor. Although application of a constant factor to the median value is arbitrary it would have the beneficial effect of reducing the variance in the extrapolated values without the need for employing untested assumptions.

There is another argument for using an arbitrary assessment factor. The method is less complex. If the predictions of two or more methodologies are indistinguishable, parsimony demands that we do not use increasingly complex methods unless they have clearly demonstrated an increased utility. Mathematical treatment of the problem should be trimmed to the minimum necessary to address the problem without obscuring the main point. In our view the more complex extrapolation methods have not provided predictive or heuristic improvements over the cruder methods.

It has been suggested that the recently developed extrapolation models (e.g. that of Van Straalen & Denneman 1989) are a 'useful tool in ecotoxicological effects assessment but should only be used if more than three (semi) chronic test data are available' (Van Leeuwen 1990). Specifically, the current test data should be supplemented by 'internationally accepted test guidelines on at least a mollusc and an insect species' (Van Leeuwen 1990). This amounts to putting the 'cart before the horse'. First, as mentioned above these methods have not been shown to be either more accurate or more conservative than the simpler assessment factor approaches. Second, and most important, focusing on the collection of more 'representative' species for use in a model of questionable validity is misplaced effort and a waste of increasingly limited resources.

There is also a more insidious problem with the recommendation to use the extrapolation methods described above. By failing to test rigorously the assumptions of these new models before incorporating them into the ecotoxicological toolbox we run a serious risk of deluding ourselves and the public that we have made substantial progress when in fact we have not. Just because these models have a more sophisticated outward appearance and seem to incorporate reasonable assumptions a priori does not justify their uncritical incorporation into regulatory protocols. Most ecotoxicologists justifiably abhor the use of arbitrary assessment factors, but we must not let the wish for improved methods obscure critical scientific evaluation of their use. Thus we recommend that the older methods remain in use. Immediate improvements in assessment factor approaches such as use of a median or geometric mean rather than lowest observed NOEC value could be employed to improve the stability of the estimates in the face of increasing information. While unarguably crude, the less refined 'assessment factor' approaches should be used as the most parsimonious current approach, a monument to our ignorance and a spur to rapid, rigorous and critical evaluation of the next generation of methods.

Acknowledgements

We thank P. Calow and an anonymous referee whose critical comments greatly improved an earlier version of this manuscript. Partial funding was provided by the Carlsberg Foundation (VF) and a Danish National Science Council postdoctoral fellowship (TF).

References

Aldenberg, T. & Slob, W. (1991) *Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data*. National Institute of Public Health and Environmental Protection (RIVM), report no. 719192992, The Netherlands.

Cairns, J., Jr & Mount, D.I. (1990) Aquatic toxicology: part 2. *Environmental Science and Technology* **24,** 154–161.

Crow, M.E. & Taub, F.B. (1979) Designing a microcosm bioassay to detect ecosystem level effects. *International Journal of Environmental Studies* **13,** 141–147.

Forbes, V.E. & Forbes, T.L. (1993) *Ecotoxicology in Theory and Practice: A Critique of Current Approaches*. Chapman & Hall, London.

Gray, J.S. (1989) Effects of environmental stress on species rich assemblages. *Biological Journal of the Linnean Society* **37,** 19–32.

Hollander, M. & Wolfe, D.A. (1973) *Nonparametric Statistical Methods*. John Wiley & Sons, New York.

Kooijman, S.A.L.M. (1987) A safety factor for LC_{50} values allowing for differences in sensitivity among species. *Water Research* **21,** 269–276.

O'Neill, R.V., DeAngelis, D.L., Waide, J.B. & Allen, T.F.H. (1986) *A Hierarchical Concept of Ecosystems. Monographs in Population Biology*, vol. 23. Princeton University Press, Princeton, New Jersey.

Organization for Economic Cooperation and Development (1991) *Draft report of the OECD workshop on the extrapolation of laboratory aquatic toxicity data to the real environment*. OECD, Paris.

Paine, R.T. (1966) Food web complexity and species diversity. *American Naturalist* **100,** 65–75.

Paine, R.T. (1974) Intertidal community structure. Experimental studies on the relationship between a dominant competitor and its principal predator. *Oecologia* **15,** 93–120.

Peters, R.H. (1991) *A Critique for Ecology*. Cambridge University Press, Cambridge.

Ricklefs, R.E. (1979) *Ecology*, 2nd edn. Chiron Press, New York.

Schindler, D.W. (1987) Detecting ecosystem responses to anthropogenic stress. *Canadian Journal of Fisheries and Aquatic Sciences* **44,** 6–25.

Sloof, W., Canton, J.H. & Hermens, J.L.M. (1983) Comparison of the susceptibility of 22 freshwater species to 15 chemical compounds. I. (sub)acute toxicity tests. *Aquatic Toxicology* **4,** 113–128.

Stephan, C.E. & Rogers, J.W. (1985) Advantages of using regression analysis to calculate results of chronic toxicity tests. *Aquatic Toxicology and Hazard Assessment: Eighth Symposium*, ASTM STP 891 (eds. R.C. Bahner & D.J. Hansen), pp. 328–338. American Society for Testing and Materials, Philadelphia.

Van Leeuwen, K. (1990) Ecotoxicological effects assessment in the Netherlands: recent developments. *Environmental Management* **14,** 779–792.

Van Leeuwen, C., Van der Zandt, P.T.J., Aldenberg, T., Verhaar, H.J.M. & Hermens, J.L.M. (1992) Application to QSARS, extrapolation and equilibrium partitioning in aquatic effects assessment. I. Narcotic industrial pollutants. *Environmental Toxicology and Chemistry* **11,** 267–282.

Van Straalen, N.M. & Denneman, G.A.J. (1989) Ecotoxicological evaluation of soil quality criteria. *Ecotoxicology and Environmental Safety* **18,** 241–251.

Wagner, C. & Løkke, H. (1991) Estimation of ecotoxicological protection levels from NOEC toxicity data. *Water Research* **25,** 1237–1242.

Received 7 September 1992; revised 4 November 1992; accepted 9 November 1992