Ecotoxicological Effects Extrapolation Models[1]

Glenn W. Suter II

Environmental Sciences Division

Oak Ridge National Laboratory[2]

Oak Ridge, Tennessee 37831-6038

Phone: 423/574-7306

FAX: 423/676-8646

E-MAIL: swg@ornl.gov

---

[1]Publication No. _____, Environmental Sciences Division, ORNL.

[2]Managed by Lockheed Martin Energy Research Corp. For the U.S. Department of Energy under contract number DE-AC05-96OR22464.

DISCLAIMER

INTRODUCTION

One of the central problems of ecological risk assessment is modeling the relationship between test endpoints (numerical summaries of the results of toxicity tests) and assessment endpoints (formal expressions of the properties of the environment that are to be protected)(Suter 1993a; U.S. EPA 1992). For example, one may wish to estimate the reduction in species richness of fishes in a stream reach exposed to an effluent and have only a fathead minnow 96 hr LC50 as an effects metric. The problem is to extrapolate from what is known (the fathead minnow LC50) to what matters to the decision maker, the loss of fish species. Models used for this purpose may be termed Effects Extrapolation Models (EEMs) or Activity-Activity Relationships (AARs), by analogy to Structure-Activity Relationships (SARs).

These models have been previously reviewed in Ch. 7 and 9 of (Suter 1993a) and by an OECD workshop (OECD 1992). This paper updates those reviews and attempts to further clarify the issues involved in the development and use of EEMs. Although there is some overlap, this paper does not repeat those reviews and the reader is referred to the previous reviews for a more complete historical perspective, and for treatment of additional extrapolation issues.

This class of models are defined as being limited to empirical models and does not

include mechanistic mathematical simulation models which are treated in other chapters of this volume as well as in (Bartell and others 1992 ; Suter 1993a). Although these empirical models are simple and often imprecise, models of this sort have in general proved to be more predictive of ecological responses than models with a more theoretical basis (Peters 1991). However, there is a continuum from models that are purely descriptive to those that are purely theoretical. Although all of the models discussed below are derived from toxicological data by expert judgement or statistical techniques, their basis in theory ranges from none (e.g., factors), to minimal (e.g., species sensitivity distributions), to well-developed and mechanistic (e.g., allometric scaling models). Although association with theory is reassuring, it does not assure that a statistically fitted model more true or more useful. All models should stand or fall based on their predictive power.

It would be a mistake to portray these empirically based and statistically derived models as being in competition with theoretically based and mathematically derived models. Ecological risk assessments are best conducted using multiple lines of evidence, each with its own strengths and weaknesses. Therefore, one model need not drive out another if both are credible and appropriate. In addition, empirical models may be used to derive parameter values for theoretical models (Barnthouse and others 1990).

This paper begins by describing different techniques for extrapolation modeling and

illustrates them by showing how they have been used to extrapolate between species. Interspecies extrapolations are considered much more often than other extrapolations and a name has even been proposed for them, quantitative species sensitivity relationships (QSSRs) (Notenboom and others 1995). The paper then presents brief discussions of some other extrapolations and concludes with some predictions and recommendations concerning future directions in the field.

TECHNIQUES FOR EXTRAPOLATION

Classification

The simplest way to extrapolate test endpoints is to assume that the observed responses are representative of some class of responses. For example, if the LC50 for fathead minnows in a particular test is x_t then we may assume that half of all fish will die at that concentration, or half of all cyprinids, or simply half of all fathead minnows, depending on how broadly we define the classes. The model is $E_a = E_t + e$, where E_a and E_t are the assessment and test endpoints, respectively.

Classification is the inevitable starting point of extrapolation modeling. That is, we must begin by deciding what differences important enough to be modeled and which are not. If only this method is used, then we must conclude either that the equality applies and

the test endpoint approximately equals the assessment endpoint or that the species, life stage, and other specifics of the test endpoint are sufficiently different from the assessment endpoint that they belong to different classes and therefore there is no useful relationship between them.

Species may be classified on various bases. For example, plants are often classified by growth form and the EPA classifies freshwater fish as warm water and cold water species (Stephan and others 1985). However, most commonly species are classified taxonomically. Usually, vertebrates are effectively classified to the class level. That is, birds are lumped but separated from mammals and bony fishes are lumped but separated from amphibians. However, all invertebrates and all plants are typically lumped in assessments of both aquatic and terrestrial contaminants. In calculation of U.S. water quality criteria, LC50s for all members of a genus are lumped by taking their geometric mean and carrying that genus mean value through the calculations (Stephan and others 1985). Studies based on correlations of LC50s of species at different taxonomic distances indicated that for both freshwater and marine fishes and arthropods, species within genera and genera within families tended to be relatively similar which suggests that they could be lumped (Suter 1993a; Suter and Rosen 1988; Suter and others 1983). The same conclusion was reached by the same method for terrestrial vascular plants (Fletcher and others 1990). Application of analysis of variance to acute toxicity data sets for aquatic organisms classified to the class level (Hoekstra and others 1994) and for birds classified to family (Baril and others 1994),

found taxonomy to be a statistically significant variable.

Multivariate ordination techniques would seem to be a logical technique for identifying clusters of species with similar sensitivities to a set of chemicals. However, a research program at the Dutch National Institute for Public Health and Environmental Protection (RIVM) that applied principal components analysis (PCA) to acute toxicity data for birds and mammals and to acute and chronic data sets for aquatic species did not find clear clusters except for daphnids in the aquatic acute data set (Vaal and others 1994; Van der Wal and others 1995a; Van der Wal and others 1995b). A PCA of avian LD50s for pesticides showed distinct clusters of Icteridae and Phasianidae, but not other taxa (Baril and others 1994). However, only eight species were included.

Factors

If the test endpoint is judged to not belong to the same class as the assessment endpoint, the simplest and most common way to extrapolate between them is to multiply by a factor. The model is $E_a = fE_t$. Factors may be applied to account for uncertainties concerning the nature of the relationship between assessment and test endpoints or for biases in test endpoints as estimators of assessment endpoints. For example, if the assessment endpoint is defined in terms of some sublethal response to chronic exposures and the test endpoint is an acute LC50, then a factor may be applied

to account for the bias associated with the fact that the chronic sublethal responses usually occur lower concentrations. Factors may be derived by analysis of data, but more often they are based on professional judgement. Most often they are factors of ten which makes their imprecision apparent and aids computation. Probably the most commonly used set of factors are those developed by the EPA for use in regulation of industrial chemicals and adapted by an OECD workshop (Table 1).

Distributions

One may assume that the assessment endpoint is a random variate represented by the distribution of test endpoints. The most common form of this extrapolation technique is the species sensitivity distribution. One simply assumes that for a particular chemical the distribution with respect to concentration (or dose) of test endpoints for different species is an estimate of the distribution of concentrations (or doses) at which species exposed in the field will display the endpoint response. For example, if the distribution of 96 hr LC50 values for fish exposed to a chemical is normally distributed (m_t, s_t) then half of fish species in the field would be expected to experience mass mortality after exposure to concentration m_t for 96 hours. This approach was developed by the U.S.EPA for the calculation of water quality criteria (Stephan and others 1985). Distributions used have included the log triangular (Stephan and others 1985), log normal (Wagner and Lokke 1991), and log logistic (Kooijman 1987) (Fig. 1).

The lower fifth percentile of such distributions has been used as the basis for environmental quality standards (Fig. 2). Because of this use, species sensitivity distributions have received much attention and criticism (OECD 1992; Smith and Cairns 1993). Among those who use these models, there is no controversy about the use of log transformation of exposures or the choice of the fifth percentile (no other number fills scientists with as much confidence and comfort as 5%), differences of opinion occur on the following points.

Distribution function - In the absence of evidence that one function is better than another, functions have been chosen largely on the basis of convenience and personal preference. However, the use by the U.S. EPA of the triangular distribution with its lower limit to sensitivity has been criticized by the Dutch and others who prefer distributions with potentially unlimited sensitivity (OECD 1992).

Minimum data set - The EPA requires eight acute values others require as few as three or as many as 20 for establishing distributions (Hoekstra and others 1994; OECD 1992). Also, certain species or representatives of certain taxa or groups (e.g., daphnids or cold water fish) may be required.

Inclusiveness - The EPA includes multicellular aquatic animals (Stephan and others 1985) but others include algae and other plants as well (Aldenberg and Slob 1993; Wagner and Lokke 1991). Although inclusiveness seems desirable, it strains the assumption that the species are drawn from a single unimodal distribution.

Confidence in the fifth percentile - The EPA uses the most likely (median) estimate of the fifth percentile but others have required 95% confidence of not exceeding the fifth

percentile (i.e., the lower 95% tolerance limit) (Aldenberg and Slob 1993).

Distributions may be fit and percentiles calculated by any statistical software. However, convenient software is available for this purpose, including calculation of both HC_5 and its lower, one-tailed 95% confidence limit for log normal, log logistic, and log triangular distributions (Aldenberg 1993).

Regression

If the assessment endpoint can be approximated by some test endpoint and if values of that test endpoint and the test endpoint from which we wish to extrapolate are available for several chemicals, then we may create an extrapolation model by regressing the one endpoint against the other. The linear version of the model is $E_a = a + bE_t + e$. For example, if the assessment endpoint is acute lethality in rainbow trout and a fathead minnow LC50 is available, then a regression of rainbow trout LC50s against fathead minnow LC50s for chemicals that were tested on both species under similar conditions could be used to perform the extrapolation (Kenaga 1978).

In that simple form, regression models can be used only to extrapolate among commonly tested species. However, clustering species in the taxonomic hierarchy permits one to predict the responses of untested as well as tested species (Suter

1993a; Suter and Rosen 1988; Suter and others 1983). If the test and endpoint species are in the same genus then a species to species model is used; if the species belong to different genera within a family, then a regression of those two genera is used, etc. (Fig. 3). For example, if one wishes to predict the response of largemouth bass from rainbow trout, the lowest taxonomic level that they have in common is the class *Osteichthyes* so the extrapolation is between the orders *Perciformes* and *Salmoniformes* (Fig. 4). In performing an interfamily extrapolation, one assumes that the relative sensitivity of the two species are represented by the relative sensitivities of the families to which they belong and that the uncertainty in that assumption is represented by the variance in the model. For North American aquatic species, most of the needed equations are available although many possible combinations of taxa are missing (Suter 1993a). This is because most of the missing equations are for low level extrapolations (species within genera) while most extrapolations required for assessments are relatively high level (orders within classes). For those that are missing, it is not possible to estimate relative sensitivity or extrapolation-specific variance. However, one may assume that the weighted mean prediction intervals on extrapolations at the same taxonomic level are estimates of the variance for the specific extrapolations (Calabrese and Baldwin 1994; Suter and others 1987). The prediction interval is the appropriate expression of variance for these uncertainty factors because we are interested in intervals that contain a pair of species at that taxonomic level with 95% confidence (or some other level) (Calabrese and Baldwin 1994; Suter and others 1987).

A third regression-based approach to taxonomic extrapolation is regression of all members of a taxon against a standard test species (Barnthouse and Suter 1986; Holcombe and others 1988; Suter and Rosen 1988; Suter and others 1987). This approach is less precise than the taxonomic hierarchy approach and does not work when the E_t is for nonstandard test species, but it does estimate the sensitivity of standard species relative to larger sets of species and the uncertainty concerning predicted effects on a randomly chosen species (Suter 1993a).

Scaling

Extrapolations can be made by assuming that the differences in sensitivity among organisms and species are differences of physical scale. The simplest and most common example of this is the expression of doses to wildlife as dose per unit mass (mg/kg) which amounts to an assumption that toxicity is a function of the dilution of the toxicant in the mass of the organism. The model in extrapolation by scaling is, $E_a = E_t + e$ when both the endpoint species and test species are appropriately scaled. The formal analysis of the consequences of organism size in physiology, ecology, pharmacology, and other branches of biology is termed allometry.

The most commonly used allometric model is a power function of weight, $E_x = a\,W^b$. This form has been adopted by toxicologists because various physiological processes

including metabolism and excretion of drugs and other chemicals are approximated by that form (Davidson and others 1986; Peters 1983). Exponents for various processes range from 0.6 to 0.8. Some ecotoxicologists followed the EPA's (EPA 1986) practice of using 2/3 in human health risk assessments for wildlife (Opresko and others 1993). This practice is conservative for humans and mammalian wildlife in that it makes large species such as deer more sensitive than the small rodents that are typically used in mammalian toxicity testing while making small wild species approximately equal in sensitivity. More recently, the EPA has used the less conservative 3/4 power for piscivorous wildlife (EPA 1993b), and others have followed their lead (Sample and others 1996). Acute mammalian toxicity data sets yield exponents that are closer to 3/4 than 2/3 on average, but are consistent with either value (Goddard and Krewski 1992; Travis and Morris 1992; Watanabe and others 1992).

Little attention has been paid to allometric models for avian toxicology. However, use of the same models for birds as mammals with the same exponents was supported by allometric models of avian physiology (Peters 1983) and pharmacology (Pokras and others 1993). In fact, Pokras et al. (1993) present models for the extrapolation of effective doses of drugs from mammals to birds based on a common exponent of 3/4 but with a higher *a* value for birds. However, Mineau et al. (in press) performed allometric regression analyses on 37 pesticides with between six and 33 species of birds (Mineau and others in press). They found that for 78% of chemicals the exponent was greater than 1 with a range of 0.63 to 1.55 and a mean of 1.1 (Fig. 5). If a

chemical has an exponent greater than one for birds, that implies that most avian species would be more sensitive than standard test species such as mallard ducks and bobwhite quail. It is not clear why these toxicological results would differ from the pharmacological experience. Use of an exponent of one is consistent with the mean exponent from the only available allometric study of avian toxicology (Mineau and others in press), and, because it reduces to scaling to weight, it is parsimonious, making neither small nor large birds more sensitive to a given dose (Sample and others 1996).

Fractional exponents of weight would be expected when effects are a function of some time integral of internal dose. This is because the mechanism controlling integral internal exposure is drug or toxicant clearance which is more rapid in small species due to their more rapid metabolism. However, if effects are a function of peak internal dose, as may be the case for lethal effects of cholinesterase inhibitors, dilution in the mass of the organism (exponent of one) would be the expected exposure model. These sorts of toxicokinetic and toxicodynamic considerations could provide the basis for scaling models that are specific to chemical classes.

The scaling approach has not been used for aquatic ecological risk assessments although Patin (1982) argued that sensitivity of aquatic species is a function of size (Patin 1982). The reasonableness of this proposition is supported by observations that chemical uptake, elimination, and body burden are power functions of weight (Newman

and Mitz 1988; Newman and Heagler 1991) and that intraspecies variation in sensitivity (time to death) of mosquitofish to NaCl was a function of weight as well as concentration (Newman and others 1994). However, interspecific extrapolation based on allometry has apparently not been investigated for aquatic species beyond Patin's (1982) initial effort.

An alternative proposed for fish is scaling toxicity of lipophilic chemicals to lipid content (Geyer and others 1994). This approach is based on the theory of "survival of the fattest" which proposes that tissues that are the site of toxic action are protected by partitioning chemicals to lipids (Lassiter and Hallam 1990). Greyer et al. (1994) found that acute toxicity of lindane in 16 fish species was a linear function of lipid content (%) above 5% lipid and curvilinear below that level (Fig. 6). This relationship is likely to be more complex for chronic exposures in which lipid stores fluctuate with season and reproductive cycle and toxic effects of the chemicals may include effects on lipid accumulation or metabolism.

ACUTE-CHRONIC EXTRAPOLATIONS

After extrapolation between species, the extrapolation that is most often addressed is that between acute and chronic toxicity. Conventionally, acute toxicity test endpoints are median lethal concentrations or doses (LC50 or LD50) from tests of a few days

duration and chronic test endpoints are concentrations or doses that constitute thresholds for statistically significant differences from controls in any of a variety of lethal and sublethal responses (NOELs or LOELs) in tests extending over more than 10% of an organism's life span. Therefore, extrapolation between acute and chronic toxicity involves extrapolation between short and long durations, between lethality and various responses, and between a biological effect level and a statistically significant difference. The acute-chronic extrapolation is typically performed using a generic factor (e.g., 10) or a chemical-specific factor. In the latter case, acute and chronic tests are performed for a particular species-chemical combination and the ratio of the resulting test endpoints is then applied to acute test endpoints for other species to estimate the chronic test endpoint (Mount and Stephan 1967).

Because the acute-chronic distinction is such a hodge-podge, it has been recommended that it be abandoned, that exposure duration be treated as a continuous variable, and that the individual response parameters be distinguished (Suter and others 1987). Extrapolation between response parameters (e.g., between mortality and fecundity) has been addressed using factors (Mayer and others 1986) and regression models (Suter and others 1987).

Similarly, extrapolation between different exposure durations may be performed using ratios of response levels at different durations (chronicity factors) or regression-based techniques. One may simply assume a concentration-duration function and fit it to test

endpoints for different durations (e.g., 24, 48, 72, and 96 hour LC50s). One may then use the model to extrapolate to other durations including indefinitely long exposures (Green 1965; Mayer and others 1994).

ORGANISM TO COMMUNITY/ECOSYSTEM

If one assumes that responses of communities or ecosystems are simply the aggregate responses of the organism-level responses of individual species, the interspecies extrapolations described above can be reinterpreted as extrapolations from organismal test endpoints to community or ecosystem endpoints. That is, if fathead minnows are assumed to represent all other individual fish species they may also represent the aquatic community as a whole. This interpretation is commonly applied to species sensitivity distributions. That is, rather than assuming that one is protecting species with 95% confidence, it is assumed that one is protecting the community by protecting 95% of species. The same assumptions can be applied to regressions of all species against standard test species.

Alternatively one may extrapolate from organism level tests to microcosms or mesocosms which are assumed to represent ecosystems. Microcosm and mesocosm NOECs have been regressed against the lowest reported laboratory LC50 and NOEC values (Sloof and others 1986). These models have similar precision to regressions

between orders of fish.

A direct approach to extrapolation from laboratory species to the field is to regress field measurements of the assessment endpoint against the test endpoint. Hartwell et al. regressed fish diversity (Margalef's index) and IBI in a number of streams against a combined score for tests of stream water and sediment with multiple test species (Hartwell and others 1995a; Hartwell and others 1995b) (Fig. 7). They found that toxicity of water and sediment was correlated with diversity but not IBI. These models provide the most complete and realistic extrapolation.

BETWEEN MEDIA/COMMUNITIES

Far more ecotoxicological data exist for freshwater aquatic species than for communities inhabiting other media, and it would be desirable to extrapolate from freshwater toxicity to toxicity in other media. In some cases the media are not qualitatively different and one may simply use factors or regression models to extrapolate between media. For example, regressions of standard chronic test endpoints for the most commonly tested salt water fish and crustacean against the most commonly tested freshwater species resulted in equations with slopes approximately equal to one and intercepts of approximately zero (Fig. 8)(Suter and Rosen 1988).

The lack of data is particularly problematical for organisms in soil and sediment. One solution is to extrapolate from the aquatic organisms to soil and sediment organisms. This is done by assuming that the organisms are exposed to the aqueous phase of soil or sediment, that the aqueous concentration can be estimated, and that the sensitivity of organisms to aqueous phase exposures is the same in all media. The U.S.EPA calculates sediment quality criteria for neutral organic chemicals, by assuming that the chemicals are at equilibrium and partition between the aqueous phase and the organic fraction of the solid phase (EPA 1993a). They then use aquatic toxicity data to estimate the effects of that exposure on sediment organisms. Lokke has tentatively proposed that an extrapolation from aquatic to soil organisms can be made by assuming that soil exposures are to only the soil pore water, that concentration in pore water can be estimated from soil concentrations using the distribution coefficient for bulk soil K_d, and the sensitivity of soil and sediment organisms to the aqueous phase chemical is the same as the sensitivity of aquatic organisms (Lokke 1994).

EXTRAPOLATION OF MODE OF ACTION

Although the extrapolation of mode of action between species is a critical issue in human health risk assessment (e.g., is a rat carcinogen also a human carcinogen?), the issue is seldom considered in ecological risk assessments. However, because ecological risk assessments are concerned with estimating reproductive effects which

are relatively costly and difficult to measure. Therefore, there is considerable need for models to predict whether a chemical has a specific mode of action on the reproductive system or for some other reason causes reproductive effects at low exposures relative to adult lethality. This issue has gained some urgency with the recent emphasis on chemicals that affect the reproductive system by acting as agonists or antagonists to endocrine hormones. Further, since multigenerational tests of pesticides are nearly always performed on rats, it would be useful to extrapolate from rats to birds. However, only 44% of chemicals that were found to be avian reproductive toxicants were also reproductive toxicants in rats (Mineau and others 1994).

MULTIPLE EXTRAPOLATIONS

Although extrapolation models are usually used singly in ecotoxicological assessments, it is often apparent that one extrapolation does not incorporate all of the differences between the measurement and assessment endpoints. The simplest approach is factor chains: $E_a = (f_1 f_2 f_3 \ldots f_n) E_t + e$, where each factor accounts for a particular difference between the measurement and assessment endpoint. This approach is very conservative in practice because it amounts to an assumption that all of the differences between the endpoints are simultaneously extreme (Suter 1993a) and (National Research Council 1994). For example, the assessment endpoint species is extremely sensitive, and the endpoint life stage is extremely sensitive, and the effects of

differences in duration are large, simultaneously.

Regression models lend themselves to multiple extrapolations. The output of one regression model (e.g., cyprinid LC50 to salmonid LC50) may be used as input to another (e.g., LC50 to EC25 for weight of juveniles per egg) to generate an extrapolation from a fathead minnow LC50 to brook trout early life-stage effects (Suter 1993a; Suter and others 1983). The principle difficulty in this approach is correct propagation of uncertainty through the chain of models.

Multiple extrapolations made be performed using multiple techniques. For example, in calculation chronic water quality criteria, the EPA uses species sensitivity distributions to estimate a fifth percentile acute value and then applies an acute/chronic factor to estimate a fifth percentile chronic value (Stephan and others 1985).

All of these multiple extrapolations depend on the assumption that the individual extrapolations are concordant. For example, it is assumed in calculation chronic water quality criteria and in the example of multiple regression models that relative sensitivities of species are the same in acute and chronic exposures. That assumption is unlikely to be perfectly true. For example, there are more modes of action involved in the various chronic effects than in acute lethality which is likely to result in increased variance among species. On the other hand, if the mode of action is the same for acute and chronic toxicity, then interspecies variance may be lower in the chronic tests

because differences in response rate would be less important.

UNCERTAINTIES

These empirical models deal adequately with uncertainty estimable from the data used, but not the uncertainty inherent in the selection of data.

Representative species - Extrapolation models such as species sensitivity distributions that use sets of test species to represent endpoint species depend on the assumption that test species are representative. Although test species have not been randomly or systematically selected, there is no reason in general to believe that there has been a bias in their selection. However, some taxa such as fishes are clearly over represented in data sets relative to their abundance in nature.

Representative life stages - Most of the data used to develop extrapolation models are for a single life stage. The absence of life stages that may be more sensitive than tested stages is particularly problematical when extrapolation models are limited to interspecies extrapolations and acute or subchronic data are used. Some sensitive life stages such as reproducing adults are seldom represented (Suter and others 1987).

Representative chemicals - Extrapolation models such as factors and regression

models that use multiple chemicals depend on the assumption that tested chemicals represent all chemicals of concern. However, chemicals are not randomly chosen for testing. For example, because ecotoxicological data sets are dominated by pesticides, they can have an inordinate influence. Partitioning data sets into chemical classes reduces but does not eliminate this uncertainty.

Small data sets - Many extrapolation models are derived from small data sets which has obvious implications for their reliability.

Data quality and consistency - Poor quality data obviously can cause bad predictions and inconsistent data obscures relationships and increases variance. However, these issues must be carefully considered and balance when choosing data sets. For example, the very large data set generated at the former Columbia National Fisheries Research Laboratory would be considered low quality because of the use of static rather than flow-through exposures (Mayer and Ellersieck 1986). However, it is unlikely that the bias associated with static testing would be significantly different in different species of fish. Therefore, the large size and consistency of this data set makes it useful for extrapolation modeling.

Model extrapolation - Often one type of data is used to develop a model which is then used for extrapolation of another type. For example, interspecies extrapolation models developed with acute lethality data are used to perform interspecies extrapolations of

chronic data. This extrapolation of the model to other uses is usually unacknowledged but may be important.

Incomplete extrapolation - In most assessments, one or a few extrapolations are modeled and the others are assumed to negligibly contribute to uncertainty. For example, it is common to explicitly extrapolate between taxa but not from the laboratory to the field (OECD 1992; Smith and Cairns 1993).

The importance of these considerations depends on their relative contribution to total uncertainty, but the analysis of relative uncertainty must be made in the proper context. For example, many studies have pointed out that variance among species is small relative to variance among chemicals. Such comparisons tend to reassure the reader that uncertainties in the extrapolation models are relatively insignificant. However, these comparisons are based on analyses of sets of heterogeneous chemicals. Real assessments are likely to be concerned with estimating the risks of a set of alternative cholinesterase inhibiting pesticides rather than comparing a pesticide to ethane. In that context, differences in species and life stage are relatively large contributors to uncertainty in the results of the assessment.

CONCLUSIONS AND RECOMMENDATIONS

Although extrapolation models have been an important tool for ecotoxicological assessments for decades, they have been neglected as a research topic. Far more effort has been devoted to developing new toxicity tests than to developing models to relate the test results to relevant endpoints. In part, this is because of inertia. For example, the EPA Office of Pollution Prevention and Toxic (OPPT) developed a set of factors for their assessments of new industrial chemicals based on a short, unpublished, and unreviewed literature review (Branch 1984)(Table 1). OPPT continues to use and defend these simple factors because more complex models have not been demonstrated to be more appropriate (Zeeman 1995). However, the choice of extrapolation method can clearly make a large difference in the results of assessments. A comparison of HC_5 values derived by the van Straalen and Denneman model and concern levels derived using the OPPT factors applied to sets of three LC50s for eight chemicals found that the results differed by factors of 2.6 to 1020 (Okkerman and others 1991). Similarly, a comparison of the results of applying seven published extrapolation models to a chemical with a fathead minnow LC50 or 100 mg/L found a factor of 19,200 difference in an estimated effects threshold (Suter 1993b).

The claim that no other extrapolation method is better than simple factors of 10, 100, and 1000 can be made because there is no good basis for the comparison. That is, there is no agreement about what the assessment endpoints should be (i.e., what the models should predict) and no set of data from the field that is agreed to represent the responses of those assessment endpoints. Therefore, no validation of extrapolation

models has been possible. One possible standard endpoint is mass mortalities (Suter 1993b). Nearly everyone agrees that streams littered with dead fish or a park littered with dead geese is undesirable, and the ability to estimate concentrations causing mass mortalities is a minimal goal for assessment models. An appropriate and more challenging standard assessment endpoint for validation purposes would be a 5% reduction in species richness. Its advantages include the following:

1. The fifth percentile of species sensitivity distributions is used for regulation of chemicals and effluents in the U.S., the Netherlands, and elsewhere.
2. Species richness is sensitive to toxic effects relative to other ecosystem or community level endpoints (Dickson and others 1992; Hartwell and others 1995b).
3. The concern with preservation of biodiversity implies that species richness is a societally valued endpoint.
4. Species richness data are likely to be available for most field studies of toxic effects.
5. A 5% reduction in species richness is likely to be detectable in many field studies of toxic effects.

A validation study for extrapolation models is long overdue, and one based on these endpoints is quite feasible.

REFERENCES

Aldenberg T. 1993. E_TX 1.3a, a program to calculate confidence limits for hazardous concentrations based on small samples of toxicity data. Bilthoven, The Netherlands: RIVM.

Aldenberg T, Slob W. 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. Ecotox. Environ. Safety 25:48-63.

Baril A, Jobin B, Mineau P, Collins BT. 1994. A consideration of inter-species variability in the use of the median lethal dose (LD_{50}) in avian risk assessment. Hull, PQ: Canadian Wildlife Service. Report nr No. 216.

Barnthouse LW, Suter GW, II. 1986. User's manual for ecological risk assessment. : Oak Ridge National Laboratory. Report nr ORNL-6251.

Barnthouse LW, Suter GW, II, Rosen AE. 1990. Risks of toxic contaminants to exploited fish populations: Influence of life history, data uncertainty, and exploitation intensity. Environ. Toxicol. Chem. 9:297-311.

Bartell SM, Gardner RH, O'Neill RV. 1992 . Ecological Risk Estimation. Ann Arbor, Michigan: Lewis Publishers.

Branch EE. 1984. Estimating "concern levels" for concentrations of chemical substances in the environment. Washington, D.C.: U. S. Environmental Protection Agency.

Calabrese EJ, Baldwin LA. 1994. A toxicological basis to derive a generic interspecies uncertainty factor. Environ. Health Perspect 102:14-17.

Davidson IWF, Parker JC, Beliles RP. 1986. Biological bassis for extrapolation across mammalian species. Regul. Toxicol. Pharmacol. 6:211-237.

Dickson KL, Waller WT, Kennedy JH, Ammann LP. 1992. Assessing the relationship between ambient toxicity and instream biological response. Environmental Toxicology and Chemistry 11:1307-1322.

EPA. 1986. Guidelines for carcinogenic risk assessment. Federal Register 51:33992-34003.

EPA. 1993a. Technical basis for deriving sediment quality criteria for nonionic organic contaminants for the protection of benthic organisms by using equilibrium partitioning. Washington, D.C.: Office of Water, U.S. Environmental Protection Agency. Report nr EPA-822-R-93-001.

EPA. 1993b. Wildlife criteria portions of the proposed water quality criteria for the Great Lakes system. Washington, D.C.: Office of Science and Technology, U.S. Environmental Protection Agency. Report nr EPA/822/R-93/006.

Fletcher JS, Johnson FL, McFarlane JC. 1990. Influence of greenhouse versus field testing and taxonomic differences on plant sensitivity to chemical treatment. Environ. Toxicol. Chem. 9:769©776.

Geyer HJ, Scheunert I, Bruggemann R, Matthies M, Steinberg CEW, Zitko V, Kettrup A, Garrison W. 1994. The relevance of aquatic organisms' lipid content to the toxicity of lipophilic chemicals: toxicity of lindane to different fish species. Ecotoxicol. Environ. Safety 28:53-70.

Goddard MJ, Krewski D. 1992. Interspecies extrapolation of toxicity data. Risk Analysis

12:315-317.

Green RH. 1965. Estimation of tolerance over an indefinite time period. Ecology 46:887.

Hartwell SI, Dawson CE, Durell EQ, Jordahl DH. 1995a. Integrated measures of ambient toxicity and fish community diversity. Annapolis, Maryland: Maryland Department of Natural Resources, Chesapeake Bay Research and Monitoring Division. Report nr CBRM-TX-95-2.

Hartwell SI, Dawson CE, Jordahl DH, Durell EQ. 1995b. Demonstrating a method to correlate measures of ambient toxicity and fish community diversity. Annapolis, Maryland: Maryland Department of Natural Resources, Chesapeake Bay Research and Monitoring Division. Report nr CBRM-TX-95-1.

Hoekstra JA, Vaal MA, Notenboom J, Sloof W. 1994. Variation in the sensitivity of aquatic species to toxicants. Bull. Environ. Contam. Toxicol. 53:98-105.

Holcombe GW, Phipps GL, Veith GD. 1988. Use of aquatic lethality tests to estimate safe chronic concentrations of chemicals in initial ecological risk assessments. In: Lewis GWSIhîEM, editor. Aquatic Toxicity and Hazard Assessment: Eleventh Symposium. Philadelphia, PA: American Society for Testing and Materials. p 442-467.

Kenaga EE. 1978. Test organisms and methods useful for early assessment of acute toxicity of chemicals. Environ. Sci. Technol. 12:1322-1329.

Kooijman SALM. 1987. A safety factor for LC_{50} values allowing for differences in sensitivity among species. Water Res. 21:269-276.

Lassiter RR, Hallam TG. 1990. Survival of the fattest: implications for acute effects of lippophilic chemicals on aquatic populations. Environ. Toxicol. Chem. 9:585-595.

Lokke H. 1994. Ecotoxicological extrapolation: tool or toy. In: Donker M, Eijsackers H, Heimbach F, editors. Ecotoxicology of Soil Organisms. Boca Raton, FL: Lewis Publishers. p 411-425.

Mayer FL, Krause GF, Buckler DR, Ellersieck MR, Lee G. 1994. Predicting chronic lethality of chemicals to fishes from acute toxicity test data: concepts and linear regression analysis. Environ. Toxicol. Chem. 13(671-678).

Mayer FL, Mayer KS, Ellersieck MR. 1986. Relationship of survival to other endpoints in chronic toxicity tests with fish. Environ. Toxicol. Chem. 5:737-748.

Mayer FLJ, Ellersieck MR. 1986. Manual of acute toxicity: Interpretation and data base for 410 chemicals and 66 species of freshwater animals. Washington, D.C.: U.S. Fish and Wildlife Service. Report nr Resource Pub. 160.

Mineau P, Boersma DC, Collins B. 1994. An analysis of avian reproduction studies submitted for pesticide registration. Ecotoxicol. Environ. Safety 29:304-329.

Mineau P, Collins BT, Baril A. in press. On the use of scaling factors to improve interspecies extrapolation of acute toxicity in birds. Regul. Toxicol. Pharmacol.

Mount DI, Stephan CE. 1967. A method for establishing acceptable toxicant limits for fish--malathion and butoxyethanol ester of 2,4-D. Trans. Am. Fish. Soc. 96:185-193.

National Research Council. 1994. Science and Judgement in Risk Assessment.

Washington. D.C.: National Academy Press.

Newman M, Mitz SV. 1988. Size dependence of zinc elimination and uptake from water by mosquitofish *Gambusia affinis* (Baird and Girard). Aquatic Toxicol. 12:17-32.

Newman MC, Heagler MG. 1991. Allometry of metal bioaccumulation and toxicity. In: Newman MC, McIntosh AW, editors. Metal Ecotoxicology, Concepts and Applications. Chelsea, MI: Lewis Pub. p 91-130.

Newman MC, Keklak MM, Doggett MS. 1994. Quantifying animal size effects in toxicity: a general approach. Aquatic Toxicol. 28:1-12.

Notenboom J, Vaal MA, Hoekstra JA. 1995. Using comparaative ecotoxicology to develop quantitative species sensitivity relationships (QSSR). Environ. Sci. & Pollut. Res. 2:242-243.

OECD. 1992. Report of the OECD workshop on the extrapolation of laboratory aquatic toxicity data to the real environment. Paris: Organization for Economic Cooperation and Development. Report nr OCDE/GD(92)169.

Okkerman PC, Plassche VD, Sloof W, Van Leeuwen CJ, Canton JH. 1991. Ecotoxicological effects assessment,: A comparison of several extrapolation procedures. Ecotoxicology and Environmental Safety 21:182-193.

Opresko DM, Sample BE, Suter GW, II. 1993. Toxicological benchmarks for wildlife. Oak Ridge, TN: Oak Ridge National Laboratory. Report nr ES/ER/TM-86.

Patin SA. 1982. Pollution and the Biological Resources of the Oceans. London: Butterworth Scientific.

Peters RH. 1983. The Ecological Implications of Body Size. Cambridge: Cambridge

University Press.

Peters RH. 1991. A Critique for Ecology. Cambridge: Cambridge University Press. 366 p.

Pokras MA, Karas AM, Kirkwood JK, Sedgewick CJ. 1993. An introduction to allometric scaling and its uses in raptor medecine. In: Redig PT, Cooper JE, Remple JD, Hunter DB, editors. Raptor Biomedecine. Minneapolis: U. Minnesota Press. p 211-224.

Sample BE, Opresko DM, Suter GW, II. 1996. Toxicological benchmarks for wildlife. Oak Ridge, TN: Oak Ridge National Laboratory. Report nr ES/ER/TM-86/R3.

Sloof W, van Oers JAM, de Zwart D. 1986. Margins of uncertainty in ecotoxicological hazard assessment. Environ. Toxicol. Chem. 5:841-852.

Smith EP, Cairns J, Jr. 1993. Extrapolation methods for setting ecological standards for water quality: statistical and ecological concerns. Ecotoxicology 2:203-219.

Stephan CE, Mount DI, Hanson DJ, Gentile JH, Chapman GA, Brungs WA. 1985. Guidelines for deriving numeric National Water Quality Criteria for the protection of aquatic organisms and their uses. : U.S. Environmental Protection Agency, Washington, D.C. Report nr PB85-227049.

Suter GW, II. 1993a. Ecological Risk Assessment. Boca Raton, FL: Lewis Publishers. 538 p.

Suter GW, II. 1993b. New concepts in the ecological aspects of stress: the problem of extrapolation. Science of the Total Environment Supplement 1993:63-76.

Suter GW, II, Rosen AE. 1988. Comparative toxicology for risk assessment of marine

fishes and crustaceans. J. Environ. Sci. Technol. 22:548-556.

Suter GW, II, Rosen AE, Linder E, Parkhurst DF. 1987. Endpoints for responses of fish to chronic toxic exposures. Environ. Toxicol. Chem. 6:793-809.

Suter GW, II, Vaughan DS, Gardner RH. 1983. Risk assessment by analysis of extrapolation error, a demonstration for effects of pollutants on fish. Environmental Toxicology and Chemistry 2:369-378.

Travis CC, Morris JM. 1992. On the use of 0.75 as an interspecies scaling factor. Risk Analysis 12:311-313.

U.S. EPA. 1992. Framework for ecological risk assessment. : Risk Assessment Forum, Washington, D.C. Report nr EPA/630/R-92/001.

Vaal MA, Van der Wal JT, Hoekstra JA. 1994. Ordering aquatic species by their sensitivity to chemical compounds: a principal component analysis of acute toxicity data. Bilthoven, The Netherlands: National Institute for Public Health and Environmental Protection. Report nr no. 719102 028.

Van der Wal JT, Luttik R, Vaal MA, Hoekstra JA. 1995a. Ordering birds and mammals by their senstivity to chemical compounds: a principal component analysis. Bilthoven, The Netherlands: National Institute for Public Health and Environmental Protection. Report nr no. 719102 036.

Van der Wal JT, Vaal MA, Hoekstra JA, Hermens JLM. 1995b. Ordering chemical compounds by their chronic toxicity to aquatic species: a principal component analysis. Bilthoven, The Netherlands: National Institute for Public Health and Environmental Protection. Report nr no. 719102 041.

Wagner C, Lokke H. 1991. Estimation of ecotoxicological protection levels from NOEC toxicity data. Water Research 25:1237-1242.

Watanabe K, Bois FY, Zeise L. 1992. Interspecies extrapolation: a reexamination of acute toxicity data. Risk Analysis 12:301-310.

Zeeman MG. 1995. Ecotoxicity testing and estimation methods developed under section 5 of the Toxic Substances Control Act (TSCA). In: Rand G, editor. Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment. Washington, D.C.: Taylor and Francis. p 703-715.

Table 1. Factors for estimating environmental concern levels from different ecotoxicological data sets for the U.S. EPA and OECD (OECD 1992; Zeeman 1995).

U.S. EPA Data Sets	OECD Data Sets	Assessment/Extrapolation Factor
Limited (e.g., only one acute LC50 via SAR/QSAR)	Lowest acute LC50, EC50, or QSAR estimate for 1-2 aquatic species	1000
Base set acute toxicity (e.g., fish and daphnid LC50s and algal EC50)	Lowest acute LC50, EC50, or QSAR estimate in a set of at minimum algae, crustaceans, and fish	100
Chronic toxicity MATCs	Lowest chronic NOEC value or QSAR estimate in a set comprising at minimum alga, crustaceans, and fish	10
Field test data		1

Figure Captions

Fig. 1 Comparison alternative functions used to estimate species sensitivity distributions. Redrawn from (Aldenberg 1993).

Fig. 2 The estimated species sensitivity distribution (solid line) and the proportion of species not protected by a given HCp (shaded area for p = 5%). The probability distribution of HCp (dashed line) and the likelihood of protecting less than 9%% of species at HCp (line-hatched area) when the number of species tested is five. Redrawn from (Aldenberg and Slob 1993).

Fig. 3 A diagrammatic representation of the use of phylogenetic relatedness to extrapolate between species. If species (Sx) are within the same genus (Gx), the model extrapolates directly between the two species. If the species are in different genera in the same family (Fx), data for all members of the two genera are aggregated and the model developed to extrapolate between genera. If they are in different families in the same order (Ox), then toxicity data are aggregated within families, etc.

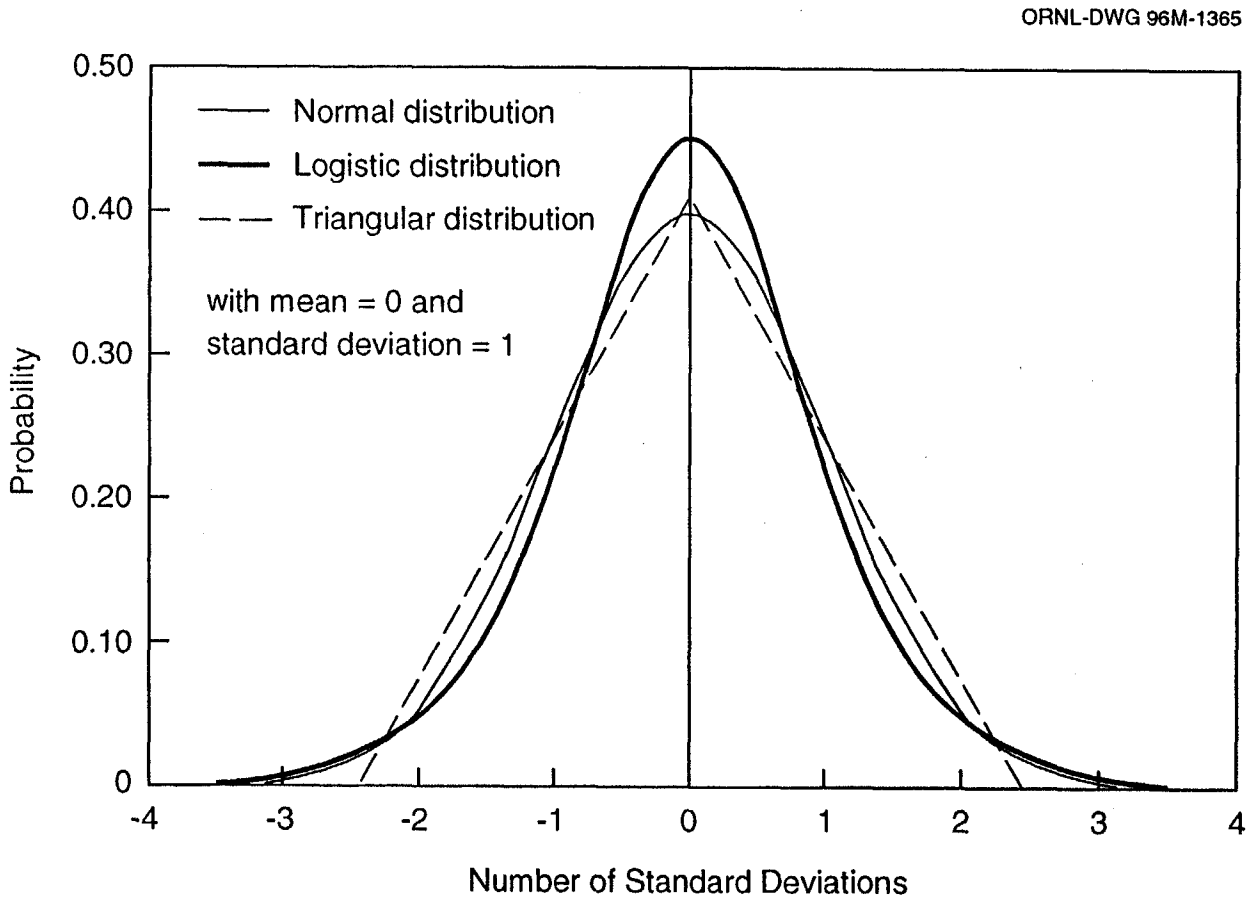
Fig. 4 Regression of log LC50 values for members of the Order *Perciformes* against LC50 values for the same chemicals for members of the Order *Salmoniformes*. From (Suter and others 1983).

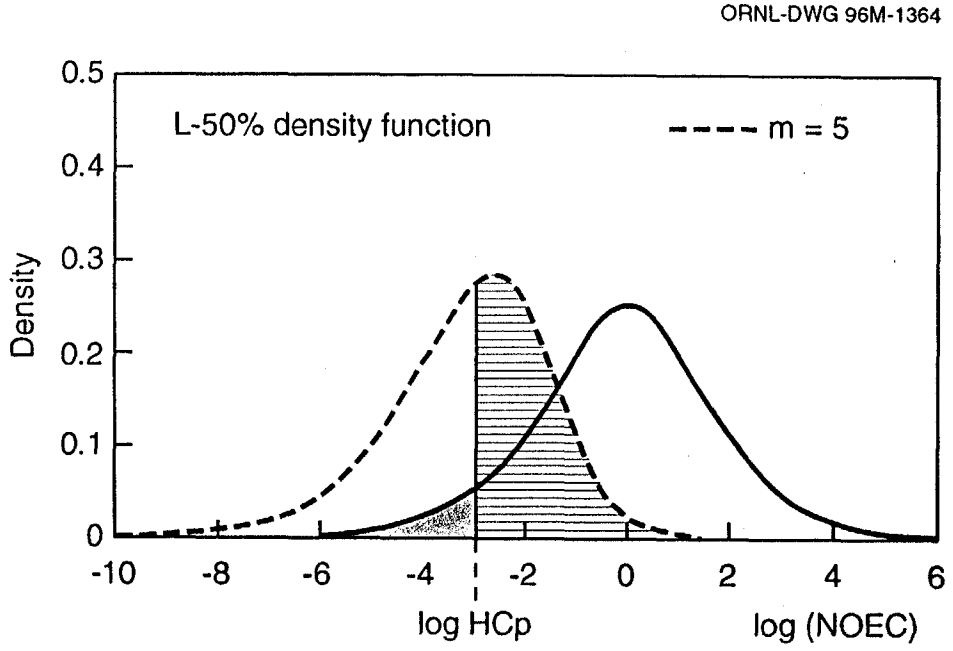
Fig. 5 Regression of log LD50 against weight of the test species for the carbamate pesticide methiocarbe (n = 32). Redrawn from (Mineau and others in press).

Fig. 6 Regression of 48-h LC50 values against percent lipid for various fish species exposed to lindane. Dashed lines delimit 95% confidence intervals. Redrawn from (Geyer and others 1994).

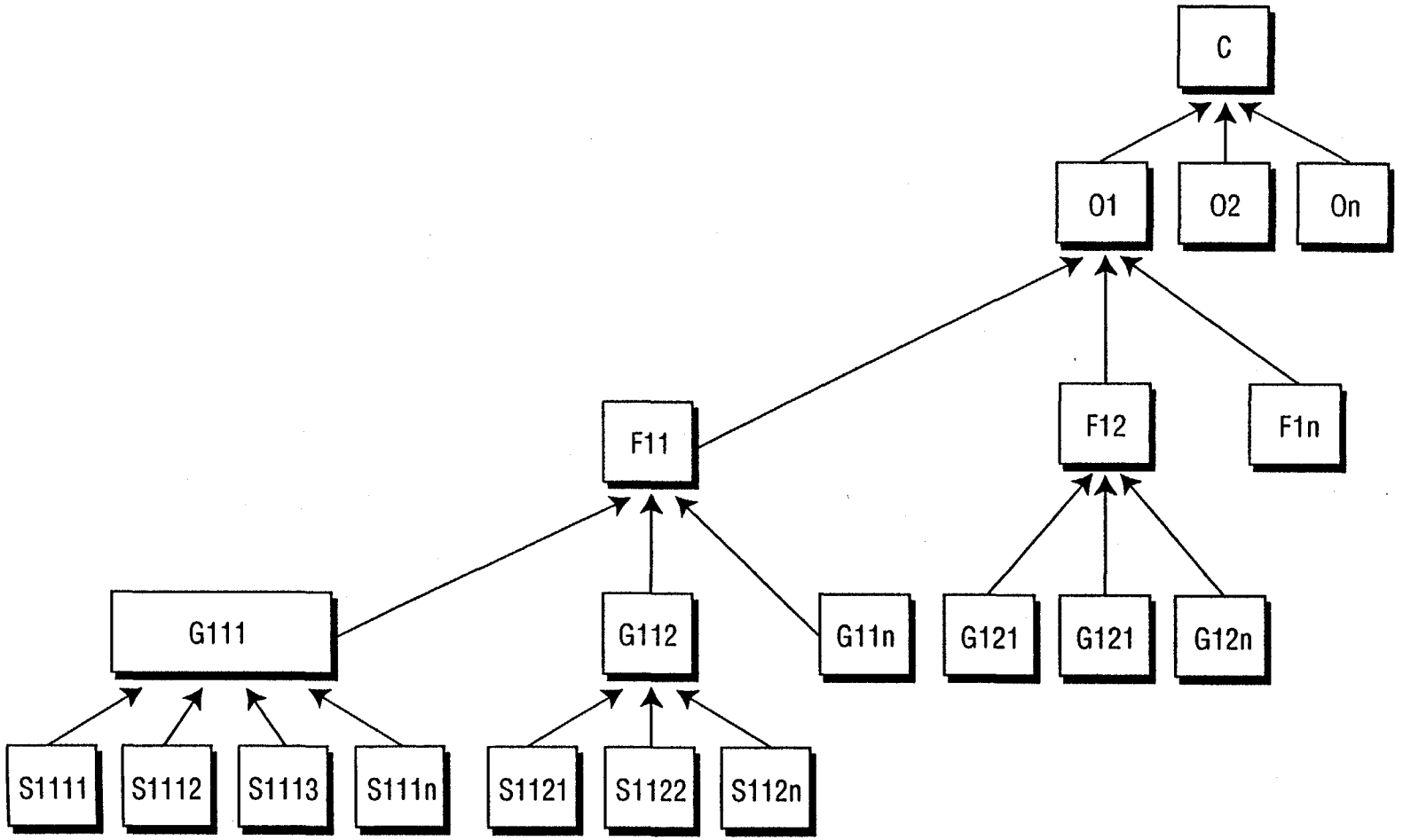
Fig. 7 Regression of Margalef's diversity for bottom-dwelling fish and all resident fish species against the combined risk score, a value that integrates the results of multiple tests of the toxicity of ambient waters. Redrawn from (Hartwell and others 1995a).

Fig. 8 Regression of Maximum Acceptable Toxicant Concentrations (MATCs) for a standard salt water test fish and crustacean species against MATCs for standard freshwater test species. Dark points are metals, all others are organic chemicals. From (Suter and Rosen 1988).

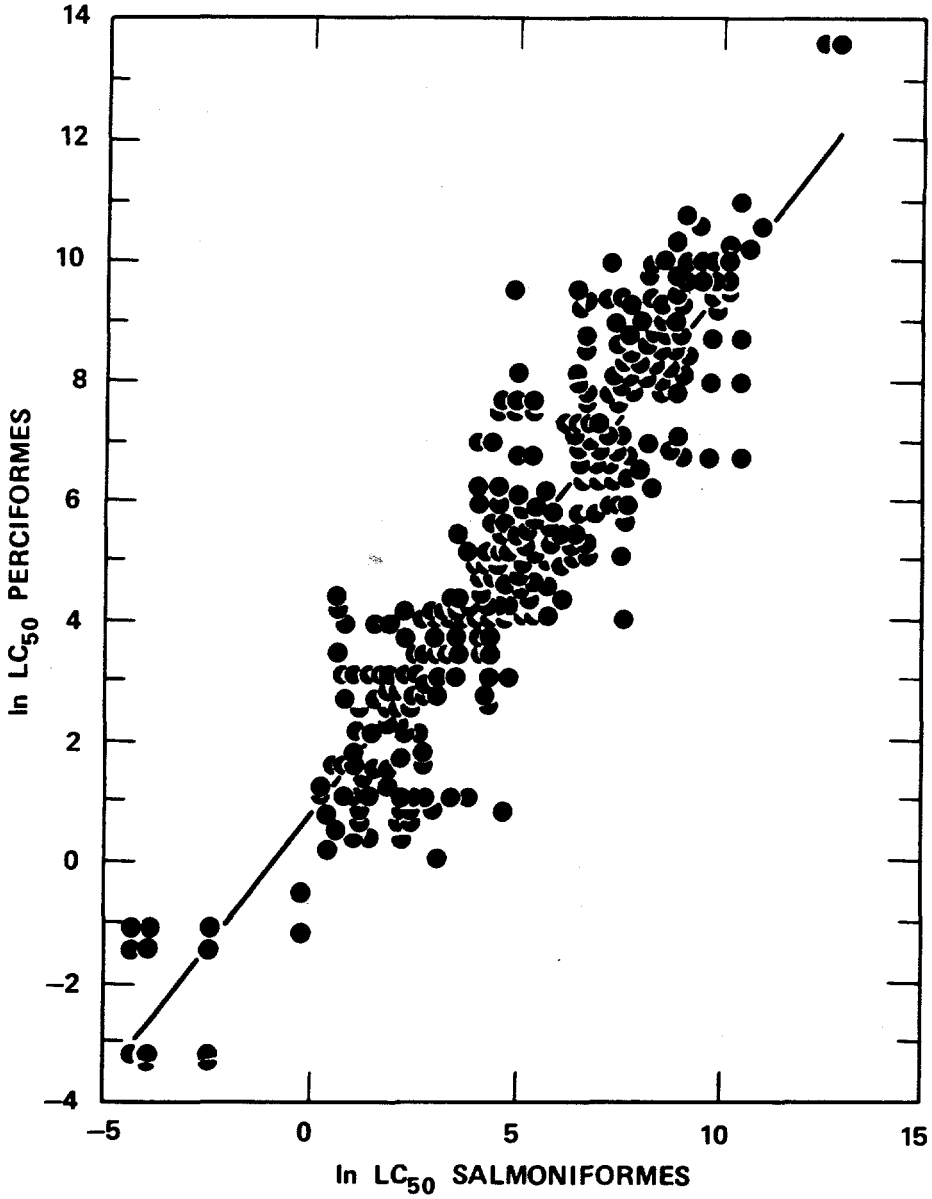

ORNL-DWG 96M-1365
0.50
0.40
0.30
0.20
0.10
0
Probability
Normal distribution
Logistic distribution
Triangular distribution
with mean = 0 and
standard deviation = 1
-4
-3
-2
-1
0
1
2
3
4
Number of Standard Deviations

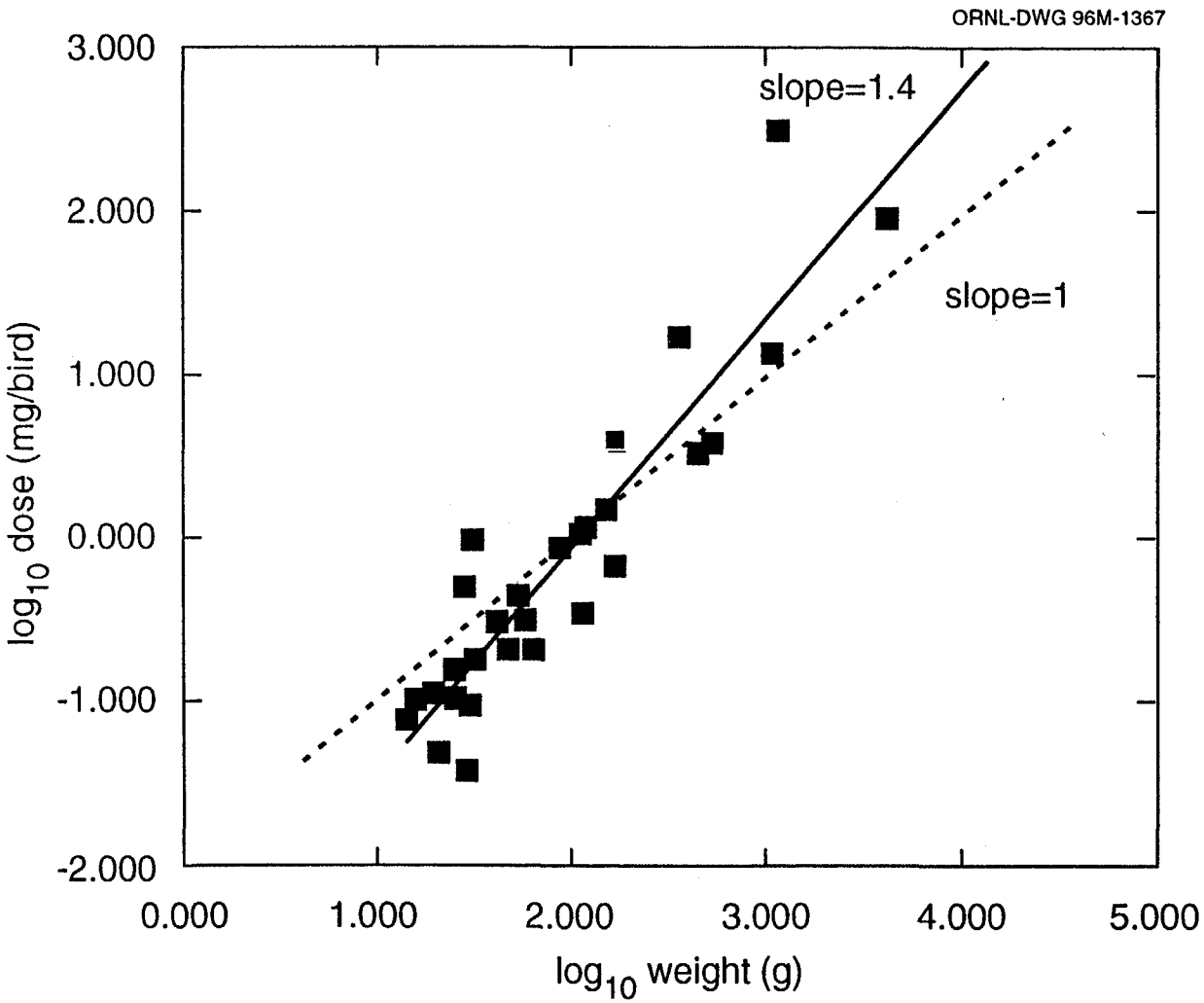

ORNL-DWG 96M-1364
L-50% density function
m = 5
Density
0.5
0.4
0.3
0.2
0.1
0
-10
-8
-6
-4
-2
0
2
4
6
log HCp
log (NOEC)

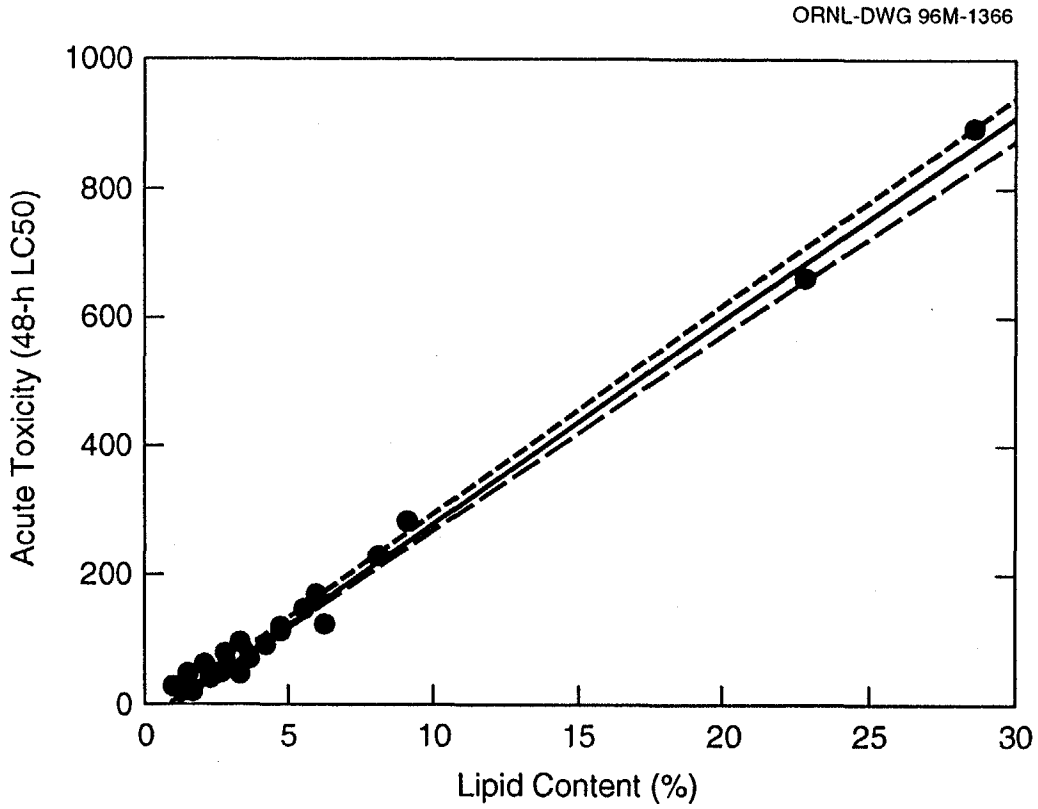
ORNL-DWG 96M-1360

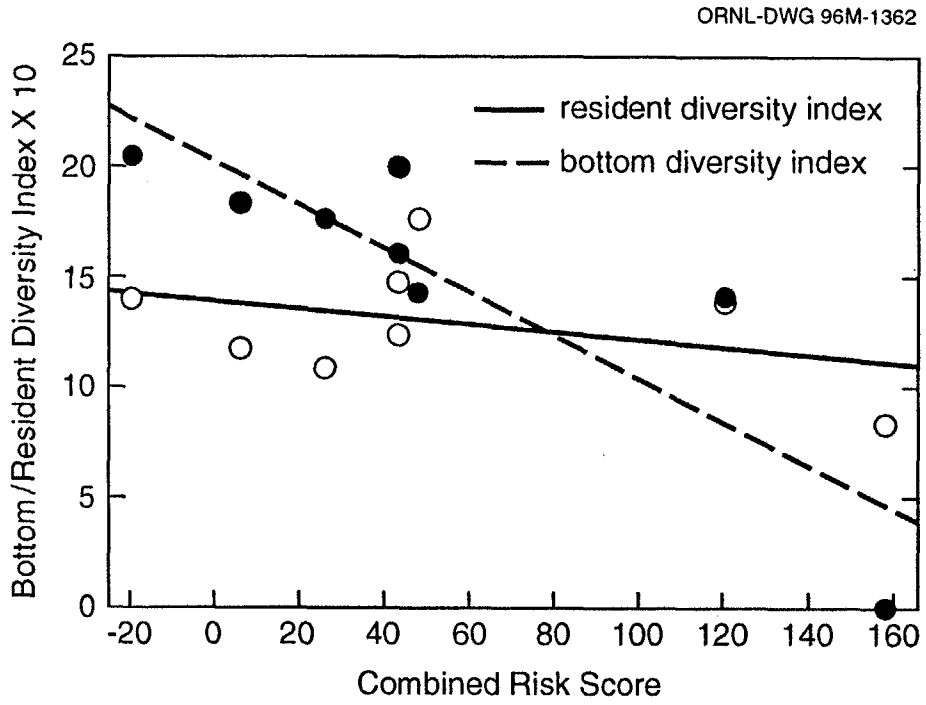


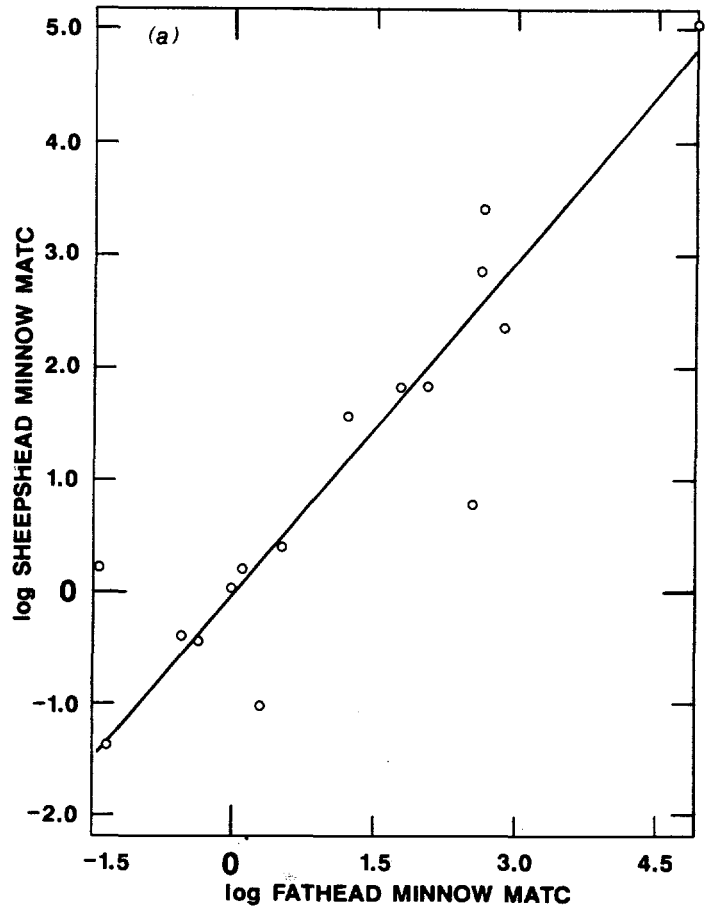
ORNL–DWG 82–15274

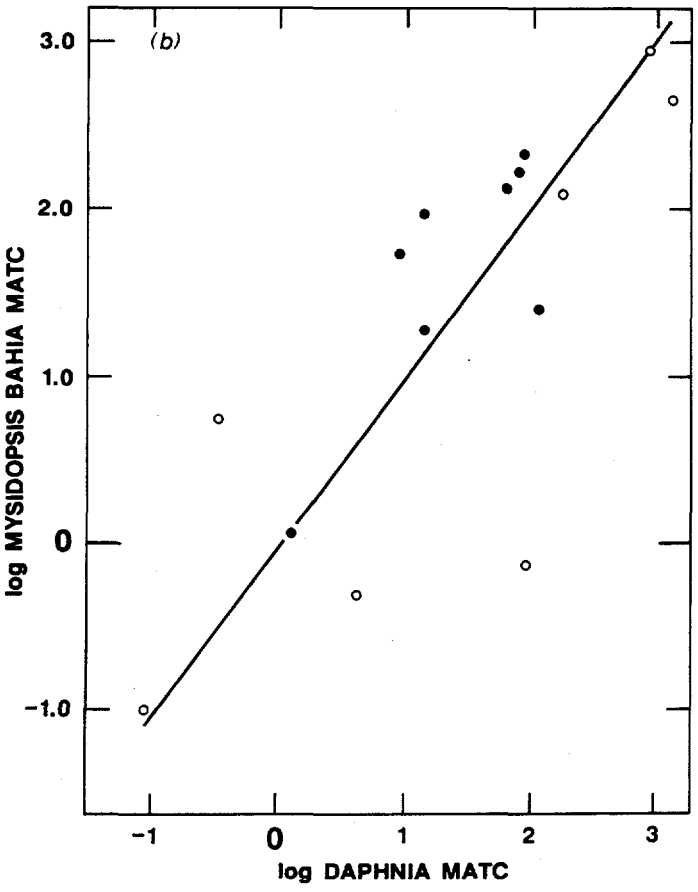



ORNL-DWG 96M-1367
slope=1.4
slope=1
3.000
2.000
1.000
0.000
-1.000
-2.000
0.000
1.000
2.000
3.000
4.000
5.000
log10 dose (mg/bird)
log10 weight (g)


ORNL-DWG 96M-1366
Acute Toxicity (48-h LC50)
1000
800
600
400
200
0
0
5
10
15
20
25
30
Lipid Content (%)


ORNL-DWG 96M-1362
Bottom/Resident Diversity Index X 10
25
20
15
10
5
0
-20
0
20
40
60
80
100
120
140
160
Combined Risk Score
resident diversity index
bottom diversity index


(a)
5.0
4.0
3.0
2.0
1.0
0
-1.0
-2.0
log SHEEPSHEAD MINNOW MATC
-1.5
0
1.5
3.0
4.5
log FATHEAD MINNOW MATC


(b)
3.0
2.0
1.0
0
-1.0
log MYSIDOPSIS BAHIA MATC
-1
0
1
2
3
log DAPHNIA MATC