

Chapter 2

Evaluation and Selection of Methods

2-1.0 Introduction

National water quality criteria methodology was established in the United States in 1985 (USEPA 1985). Since then several methods have been developed around the world, incorporating recent advances in the field of aquatic toxicology in a variety of different approaches. A recent review (The Phase I report, TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) compared these methodologies and summarized the differences between them. The findings of that report are briefly reviewed in Chapter 1, Tables 1.1 and 1.2. While EPA methods provided a good basis for criteria calculation, many of the newer approaches had valuable procedures that could improve criteria generation. Of particular concern were pesticides with toxicity datasets that do not meet the eight taxa requirements of EPA's guidelines (USEPA 1985). There is little guidance on how to derive criteria for such compounds in the EPA 1985 guidelines. Therefore, a new methodology was devised to handle a variety of data sets, consisting of a combination of features from existing methodologies with refinements based on recent research in aquatic ecotoxicology and environmental risk assessment.

This chapter describes the development of the new methodology. Criteria derivation is a process that can be broken down into a number of steps, starting with data collection and ending with numeric criteria that are protective of aquatic life with as much certainty as possible given the amount of data available. In this chapter, methods that may be used to achieve each step are presented, evaluated and selected for use in the new methodology. The focus of this chapter is choosing best elements for the new methodology, although some desirable elements were indicated in the Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) and for those aspects the reader will be referred to the analysis in that report. For most aspects of the new methodology further analysis was done before choosing a procedure and it is described here. This chapter also documents any new procedures derived for the new methodology.

A full set of chlorpyrifos data was collected, processed according to suggested methods, and utilized to compare models. Eleven other pesticide data sets from USEPA (1980a; 1980b; 1980c; 1980d; 1980e; 1980f; 1980g; 1986b; 2003c; 2005a) were also utilized.

2-1.1 Level of organization to protect and goal

The narrative objective of the Central Valley Regional Water Quality Control Board is to maintain waters free of "toxic substances in concentrations that produce detrimental physiological responses in plant, animal, or aquatic life" (CVRWQCB 2004). The Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) discusses how the goal of most water quality criteria methodologies is to protect ecosystems, and to do so, many aim for protection at the species level. Also discussed in the Phase I report is how the disappearance of a single species could lead to the unraveling of community structure due to complex

interactions among species, suggesting that ecosystems might not be fully protected if water quality criteria are derived by a method that does not have the goal of protecting all species. The new method will have the goal of protection at the species level in order to fully protect natural ecosystems and meet the policy mandate.

The goal of this method is to extrapolate from available pesticide toxicity data for a limited number of species to a concentration that should not produce detrimental physiological effects in aquatic life. These criteria aim to protect all species in the ecosystem. The development of this methodology focused on the Sacramento and San Joaquin River watersheds of the California Central Valley and this ecosystem is specifically discussed in several instances. The resulting method, however, is generally appropriate for any freshwater ecosystem in the United States. Additionally, simple modifications could be made to adapt this method for saltwater criteria or other geographic areas.

2-2.0 Data

Criteria derivation requires ecotoxicological effects data. Good criteria must be based on good quality data of adequate taxonomic diversity. Physical-chemical data are also important for proper interpretation of toxicity test data, for estimation of freely dissolved fractions of chemicals, for understanding possible effects of water quality characteristics on toxicity, and for estimation of toxicity for some classes of chemicals. The most reliable, most certain criteria are derived from the largest and best quality data sets. The new methodology includes guidance on what kinds of data should be collected, where to collect it, and how to evaluate its quality.

2-2.1 Kinds of data to collect

For thorough evaluation of pesticide effects it is necessary to collect physical-chemical and ecotoxicity data. Although not used directly in derivation of aquatic life criteria, dietary exposure effects data for wildlife and humans should also be collected for assessment of potential hazards due to pesticide bioaccumulation. The new methodology includes a table summarizing the kinds of physical-chemical, ecotoxicity and other data required for criteria derivation. Because of the large variety of ecotoxicity data that may be found, the remainder of this section provides specific details regarding the kinds of ecotoxicity data that should be collected for criteria derivation. The kinds of physical data to collect are reviewed in section 2-2.5.1 because it is tied into the data evaluation process.

The goal is to collect high quality data from as many taxonomic groups as possible. This methodology is for derivation of criteria for use in the United States, only data for freshwater species from families with reproducing populations residing in North America will be used in criteria derivation. However, all available data should be collected as they may be used as supporting information or for derivation of acute-to-chronic ratios if they are otherwise of high quality. The choice to consider geographic distribution at the family level, rather than the species level - as is done in current

USEPA methodologies (USEPA 1985; 2003d) - is based on work by the USEPA showing that interspecies toxicity correlations work well at the family level (USEPA 2003e, see discussion in section 2-2.2.2). A summary table that lists all of these data requirements is included in the new methodology (Table 3.3; Chapter 3).

2-2.1.1 Acute vs. chronic toxicity data

Both acute and chronic toxicity data are needed for criteria derivation. As discussed in the Phase I report for this project (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009), the difference between acute and chronic toxicity is not always clear and thus it is important that this new methodology define what kinds of tests are considered acute vs. chronic. Good general definitions of acute and chronic are given in USEPA (2003d):

Acute toxicity: Concurrent and delayed adverse effect(s) that results from an acute exposure and occurs within any short observation period which begins when the exposure begins, may extend beyond the exposure period, and usually does not constitute a substantial portion of the life span of the organism.

Chronic toxicity: Concurrent and delayed adverse effect(s) that occurs only as a result of a chronic exposure. Chronic exposure is exposure of an organism for any long period or for a substantial portion of its life span.

These general definitions are helpful, but for clarity more specific guidance is needed. Following are definitions of acute and chronic data from existing methodologies. Most of these are incorporated into the new methodology:

Acute

- 1) Crustacean or insect tests lasting 24-96 h; (RIVM 2001; Siepmann & Finlayson 2000; USEPA 1985; 2003d);
- 2) Fish, mollusk or amphibian tests lasting 96 h (RIVM 2001);
- 3) Shellfish embryo, larval, or older life-stage tests lasting 96 h (USEPA 1985; 2003d).

Chronic

- 1) Algae, bacteria, or protozoa tests lasting 3-4 d (RIVM 2001);
- 2) Fish, mollusk or amphibian early life-stage tests and 28-d growth tests (RIVM 2001);
- 3) Single-celled organism tests of any duration (USEPA 1985; 2003d);
- 4) Any test that takes into account the number of young produced, regardless of duration (USEPA 2003d);
- 5) Full life-cycle (ranging from 7 d for mysids to 15 months for salmonids), partial life-cycle (all major life stages exposed in less than 15 months; specifically for fish that require more than a year to reach sexual maturity), and early life-stage tests (ranging from 28-60 d; also specifically for fish (USEPA 1985; 2003d).

Endpoints in acute tests may be survival or immobility and endpoints in chronic tests may be survival, growth, reproduction, or measures of population growth rate. Other endpoints that have been linked to survival, growth or reproduction may also be included. See section 2-2.1.3 for further discussion of endpoints.

Definitions 1 and 3 in the chronic list both apply to single-celled organisms. Likewise, definitions 2 and 4 in the chronic list apply to early life-stage, or short-term chronic tests. The USEPA definitions (3 and 4 above in the chronic list) are included in the new methodology as they will result in the inclusion of a broader range of tests than those from the RIVM (Netherlands' National Institute for Public Health and the Environment) document. For example, algae, bacteria and protozoa tests shorter than 3 d are included by the USEPA definition, but excluded by the RIVM definition. Life cycles of plants vary widely and procedures for conducting toxicity tests with plants are not well developed. Currently plant toxicity test usually measure endpoints generally associated with chronic toxicity, such as growth and reproduction. Therefore, explicit definitions for acute plant tests are not included, and all plant and algal toxicity data will be considered chronic toxicity data.

Typically, very few chronic data are available for a given chemical. Methods are available for deriving chronic toxicity values and chronic water quality criteria from acute toxicity data. These methods are discussed in 2-2.2.3 and 2-3.2.5.

2-2.1.2 Hypothesis tests vs. regression analysis

Toxicity values are usually expressed as values such as: LC_{50} , EC_{50} , NOEC or LOEC. These names come from analyzing ecotoxicity data by two main methods: regression analysis and hypothesis tests. The Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) reviews both these approaches in detail. Major points of the discussion are summarized here because this information is important for interpreting toxicity data and deciding which endpoints are appropriate. The different analyses both have advantages and disadvantages for use in criteria derivation.

Regression analysis is most commonly used for acute toxicity tests, but can also be applied to chronic tests. In regression analysis an equation is derived that describes the relationship between concentrations and effects (Stephan & Rogers 1985). Thus it is possible to make point estimates of toxicant concentrations that will cause an effect (EC_x) to x percent of organisms. It is also possible to predict effects for a given level of toxicant. The effect may be lethality, expressed as (LC_x). Acute toxicity data are most often expressed as LC_{50} , which is the concentration that is lethal to 50% of organisms. Also for invertebrates, the effect of immobility may be expressed with an EC_{50} , where EC_{50} is the concentration that affects the mobility of 50% of organisms.

The other widely used method for analysis of ecotoxicity data is hypothesis tests, which are typically used for life-cycle, partial life-cycle, and early life-stage tests. Hypothesis tests compare treatment groups to a control group to determine which of the treatment groups is significantly different from the control (Stephan & Rogers 1985). A

no observed effect concentration (NOEC) or no observed effect level (NOEL) and a lowest observed effect concentration or level (LOEC or LOEL, respectively) may be derived from this type of analysis. Some methodologies use the geometric mean of the NOEC and LOEC to calculate a maximum acceptable toxicant concentration (MATC).

Many problems with hypothesis testing are described in the literature. They are summed up succinctly by Stephan & Rogers (1985) who point out that the effect value obtained from a hypothesis test is dependent on what toxicant concentrations were actually tested and the selection of α (type I error rate), which is usually arbitrarily chosen at 0.05. Hence, Hoekstra & Van Ewijk (1993) have shown examples where NOEC values from tests can be misinterpreted as actual no effect levels. In contrast, regression analysis determines a relationship between concentration and effect, and so provides a means to interpolate for estimation of effects at untested concentrations. Also, with regression analysis, the confidence limits will change according to α or with an increase in variability, but the point estimate will not change. Bruce *et al.* (1992) observe another shortcoming of hypothesis testing, namely, that when results are reported solely as a NOEC value, information on the concentration-response curve and variability in the data is lost.

There is apparent agreement among toxicologists that regression analysis provides better effect level estimates than hypothesis tests (Stephan & Rogers 1985; Bruce *et al.* 1992; Grothe *et al.* 1996; Moore & Caux 1997). Regression methods are commonly used and widely accepted for analysis of acute toxicity data, but for analysis of chronic data, hypothesis tests have been more widely used. Most chronic toxicity studies reported in the literature use hypothesis tests to determine the NOEC and the LOEC, making this form most useful. In addition, although regression methods are preferred, there is little agreement among scientists as to what level of statistical effect may be considered a no biological effect level (e.g., EC₅, EC₁₀, or some other value). The challenge, then, is to find a way to ensure that toxicity values derived from hypothesis tests are reliable values for use in criteria derivation.

A good approach is that proposed by participants in a 1994 workshop in The Netherlands (Van Der Hoeven *et al.* 1997). They concluded that NOEC data may be used as a summary statistic in ecotoxicity testing if the following are reported: a) the minimum significant difference; b) the actual observed difference from control; c) the statistical test used; and d) the test concentrations. These factors can be used to make case-by-case judgments as to whether a reported NOEC is reliable. For example, an extremely high minimum significant difference can be a sign of a poorly conducted test, in which high variability within treatment groups has obscured differences between groups. Thus, the minimum significant difference should be reported as a measure of within-test variability for NOEC data (USEPA 2000; Denton *et al.* 2003). Likewise, the response at the NOEC should be reasonable compared to control (e.g., a 50% reduction compared to the control should certainly not be considered “no effect”), the statistical methods should be appropriate, and the test should be designed with an appropriate dilution factor (test-specific, but ranges from 1.5-3.2 in various methods from the

Organization for Economic Co-operation and Development, OECD; the American Public Health Association, APHA; and the American Society for Testing and Materials, ASTM).

For the new methodology, acute toxicity data should be in the form of LC₅₀ or EC₅₀ values derived from regression analysis. The use of chronic data expressed as EC_x values (from regression analysis) is only acceptable if species-specific studies are available to show what level of *x* represents a biological no-effect level (see discussion above). Species-specific studies are also required to determine what levels of effects can be detected in toxicity tests (Denton & Norberg-King 1996).

Chronic data expressed as results of hypothesis tests are acceptable, but must be evaluated to ensure that reported toxicity values are reasonable estimates of no-effect levels. The new methodology includes the following factors in test reliability rating schemes for results of hypothesis tests: minimum significant difference, observed difference from control at the NOEC, LOEC or MATC, statistical method used and test concentrations. Absence of these parameters, or unacceptable results for them, will not necessarily eliminate a test from the data set, but will reduce its reliability score.

The question still remains as to which hypothesis test value (NOEC, LOEC, MATC), should be used for criteria derivation. In a well-designed, well-conducted toxicity test, the true no-effect level lies between the NOEC and LOEC (each of which must be one of the toxicant concentrations actually tested). Although interpolation between these toxicity values is not strictly allowed (Stephan & Rogers 1985), the MATC, which is the geometric mean of the LOEC and NOEC, represents an accepted way of estimating the true no effect level. Therefore, the MATC is the value used in the new methodology to calculate the chronic criterion.

2-2.1.3 Endpoints

As discussed in the Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) most derivation methodologies use ecotoxicity data only from studies in which the evaluated endpoints were survival and/or immobility, and growth and/or reproduction, including effects such as histopathological effects on reproductive organs, spermatogenesis, fertility, pregnancy rate, number of eggs produced, egg fertility, and hatchability (RIVM 2001). These so-called traditional endpoints are favored because they can be readily linked to population-level effects. Linkages between effects at successive levels of organization are what define whether or not an observed effect is biologically significant (Suter & Barnthouse 1993). Non-traditional endpoints, such as endocrine disruption, enzyme induction, enzyme inhibition, behavioral effects (other than immobility), histological effects, stress protein induction, changes in RNA or DNA levels, mutagenicity, and carcinogenicity, have had very few links established between these effects seen at the individual level and effects at the population, community or ecosystem level. Many of these non-traditional endpoints are merely markers of exposure, with no link between that exposure and adverse effects on survival, growth or reproduction. Generally, non-traditional endpoints can be used as supporting information, but not directly in criteria derivation.

Two exceptions to this generalization are the endocrine disrupting effects of tributyltin and the inhibition of acetylcholinesterase by pesticides. Segner (2005) discusses three cases in which population-level effects in wildlife could be linked to environmental substances with endocrine activity: reductions in dogwhelk (*Nucella lapillus*) populations due to imposex caused by exposure to tributyltin; reduction in predatory bird populations due to egg-shell thinning caused by exposure to DDE; and decline in Atlantic salmon populations due to effects of 4-nonylphenol on the ability of smolts to osmoregulate. However, only in the case of tributyltin is there a strong case for endocrine disruption as the mechanism of the observed toxic effects. As a result, the recent “Ambient Aquatic Life Water Quality Criteria for Tributyltin (TBT) –Final” (USEPA 2003b) utilizes data from several studies of imposex in gastropods to set the final chronic criterion.

In the case of effects on acetylcholinesterase (AChE) activity, mortality or immobility has been linked to enzyme inhibition of > 90% in larval walleye and > 71% in juvenile walleye in acute exposures to chlorpyrifos (Phillips *et al.* 2002), and 50-70% in *Daphnia magna* exposed to a variety of chemicals (Barata *et al.* 2004; Printes & Callaghan 2004). On the other hand, Ferrari *et al.* (2004) found that goldfish exposed to azinphos-methyl could withstand cholinesterase reductions > 90% without mortality. Printes & Callaghan (2004) found that different cholinesterase-inhibiting pesticides had different inhibition levels associated with mortality. These studies indicate that although AChE activity has been associated with mortality the association is species- and chemical-specific.

Whether AChE inhibition (or other biochemical endpoint) data can be used in criteria derivation has to be decided on a case-by-case basis. For example, in a study of Chinook salmon, Wheelock *et al.* (2005) determined a 96-h acetylcholinesterase inhibition NOEC for chlorpyrifos of 1.2 µg/L, LOEC of 7.3 µg/L and MATC of 3.0 µg/L. At the LOEC, inhibition was 85% in brain tissue and 92% in muscle. Raw data from this study were analyzed by linear regression (Excel v. 11.2.5) to roughly estimate that 7.5% mortality would be expected at the MATC concentrations of 3.0 µg/L. Thus, in this case, a level of chlorpyrifos that causes less than 85% enzyme inhibition (i.e., at the MATC) is expected to result in just 7.5% mortality. At the LOEC of 7.3 µg/L chlorpyrifos, where 85-92% enzyme inhibition occurs, 18.2% mortality is expected. These levels of mortality would have to be shown to be both statistically and biologically significant for the AChE inhibition results to be used in criteria derivation. The statistical significance may be determined by analysis of toxicity test variability. For example, Denton & Norberg-King (1996) determined that a reduction of 11.2-14.9% in survival compared to controls could be detected in fish toxicity tests for some species. Biological significance is determined by a link to survival, growth or reproduction because these are the endpoints recognized in the new methods (see above). In this example, AChE inhibition is linked to one of the recognized endpoints: survival. To be able to use AChE, or other biochemical, data for criteria derivation in this methodology, it would be necessary for a toxicity study to derive an inhibition concentration (IC_x) value, where x is

equal to the enzyme inhibition level that is linked to statistically significant change in mortality, growth, or reproduction for a given chemical and species.

Another class of non-traditional endpoints is those that are directly linked to population-level effects, but are rarely determined in single-species toxicity studies. Population-level endpoints are suggested by Whitehouse *et al.* (2004) as a way to more directly predict toxic effects of chemicals on ecosystems. Using the population parameters r (intrinsic rate of population growth) and λ (factor by which a population increases in a given time), Whitehouse *et al.* (2004) found r to be a more sensitive endpoint than reproduction in tests with *Daphnia magna* exposed to zinc, and found EC_{20} values based on r to be in good agreement with a NOEC value determined for effects of 17α -ethinyloestradiol in fathead minnows (Lange *et al.* 2001). On the other hand Forbes & Calow (1999) found that, in most cases, r and λ were equally sensitive, or less sensitive to toxicant exposures compared to individual traits (e.g., reproduction). Whitehouse *et al.* (2004) note that while their intra-species examples of using population-level endpoints worked out well, there is little evidence to support their use across species for a given chemical. They attempted to collect data to expand the model across species, but out of 385 potentially usable studies, the needed mortality and fecundity data could only be obtained for six. Due to lack of data, and lack of evidence that they are any more protective of ecosystems than traditional endpoints, population-level endpoints are not generally used in the new methodology, but can be used if available.

For the new methodology, results of tests using individual level endpoints other than survival/immobility, growth and reproduction may be used to derive criteria if those endpoints have been adequately linked to effects on survival, growth, reproduction, or population-level parameters, such as r or λ . Population level endpoints, such as r and λ can be used if they come from studies rated relevant and reliable, but only if a more sensitive endpoint is not available for that species.

2-2.1.4 Multispecies (field/semi-field, laboratory) data

While there is much debate in the literature about whether or not single-species toxicity tests are good predictors of ecosystem effects, multispecies data is problematic for use in criteria derivation due its paucity and variability. Several studies have shown that the repeatability, reproducibility and ecological realism of these mesocosms were poor enough to preclude the use of such data in predictive risk assessment, or for extrapolation to natural ecosystems (Crane 1997, Kraufvelin 1999, Sanderson 2002, Hansen *et al.* 2003). As discussed previously in the Phase 1 report, water quality criteria derived from single-species tests are protective of ecosystems in many cases (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009). Also, several studies showed that laboratory-derived NOECs were predictive of field effects (Borthwick *et al.* 1985, Crane *et al.* 1999, Persoone & Janssen 1994)

Due to these problems, and given the relative cost-effectiveness, reproducibility and reliability of single-species toxicity tests, most methodologies do not utilize multispecies and field data for criteria derivation. Some methodologies do not use field or

semi-field data directly, but do use them as a comparison to criteria derived from single-species data (RIVM 2001, OECD 1995). In some cases a final criterion may be adjusted if strong multispecies evidence indicates that the single-species criterion is over- or underprotective (USEPA 1985, USEPA 2003c, Zabel & Cole 1999, RIVM 2001). In light of the above, multispecies are not used for criteria derivation in the new methodology. Although not useful for direct derivation of criteria, field or semi-field data are very useful for comparison to criteria derived from single-species data (OECD 1995; RIVM 2001), and may provide justification for adjustment of a final criterion (RIVM 2001; USEPA 1985; 2003d; Zabel & Cole 1999). If toxicity values obtained for appropriate endpoints (i.e., those related to survival, growth, or reproduction) in high quality multispecies studies are lower than the derived criteria, then criteria may need to be adjusted downward. Adjustment of criteria upward is not recommended, as single species data have indicated this concentration to be protective and raising the criterion may cause toxicity to sensitive species.

2-2.1.5 Data from multipathway exposures

As discussed in Phase I (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009), until food web or other models are further developed to incorporate multipathway exposures into criteria derivation, the best approach is to do water-only assessments. If studies show these criteria to be underprotective, and if the substance has a log K_{OW} between 5 and 7, then dietary uptake studies specific to the compound and species affected should be done to determine if exposure has been significantly underestimated. Water-only exposures are used for criteria derivation by the new methodology, but if derived criteria for chemicals with log K_{OW} between 5 and 7 appear to be underprotective, then targeted studies are recommended to determine more precisely if dietary uptake is significant for particular species/chemical combinations. For pesticides with log K_{OW} between 5 and 7, laboratory toxicity test data should be carefully reviewed to ensure that feeding regimes eliminated, or minimized any effects from interaction of the pesticide with food particles (e.g., reduction of test solution concentration due to partitioning into the food particles, or introduction of a dietary exposure route if animals ingest food that has had a chance to sorb pesticide).

2-2.1.6 Toxicity data that incorporate time

To address the effect of time on toxicity of chemicals, it would be ideal to be able to derive criteria that could be defined for any given exposure duration, thus eliminating the need to define separate acute and chronic criteria. In other words, different criteria would be derived for different exposure scenarios. Thus, environmental managers would be able to determine compliance with water quality standards for anything from brief pulse exposures to extended exposures of essentially infinite time. This could be done by, for example, segregating acute toxicity data according to exposure duration and deriving criteria for 24-, 48-, 72-h, etc. durations. Alternatively, toxicity test data could be analyzed using time-to-event methods to describe dose-time-response surfaces (Sun *et al.* 1995), or by kinetic models that describe the time course of toxicity (USEPA 2005b). As discussed in the Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009),

these models are data and resource intensive. Toxicity data, as they are currently (typically) reported, do not include information regarding multiple time points or kinetics. Thus, these time-to-event models are not feasible for general use in criteria derivation at this time and derivation of separate acute and chronic criteria is the approach adopted for the new methodology. However, as discussed in section 2-2.2.3, if data are available, concentration-time-effect models may be used to estimate chronic toxicity. Also if site- or situation-specific (i.e., for pulse exposures) criteria are desired, studies could be designed and conducted specifically to gather data for those models and the resulting toxicity data could be used to derive time-specific criteria.

2-2.2 Filling data gaps with estimation techniques

A common challenge in derivation of water quality criteria is that very few usable data are available. This is of particular concern in the case of chronic toxicity data. This section presents approaches for estimation of acute and chronic toxicity. Some, such as acute-to-chronic ratios (ACRs) are widely used and accepted. Some, such as quantitative structure activity relationships (QSARs) are widely accepted for some kinds of toxicants, but are still under development for most toxicants. Others, such as time-concentration-effect models are newer, have been validated for a large number of fish species, but are very data intensive procedures that are not feasible with most currently available data.

2-2.2.1 Quantitative Structure Activity Relationships (QSARs)

QSARs can be useful for estimation of acute and chronic toxicity for non-specific modes of action (narcotic) chemicals, but for reactive and specifically-acting chemicals, QSAR models are not reliable for estimation of toxicity (Phase I report; TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009). The ANZECC & ARMCANZ (2000) guidelines caution that QSARs should not be used as black boxes and considerable chemical expertise is required in their use. Using them only to predict toxicity of narcotic chemicals minimizes the risk of inappropriate use. Many industrial chemicals (e.g., solvents or other hydrocarbons) have a narcotic mode of action, but most pesticides have specific modes of action. Therefore, QSARs are of limited use for pesticide criteria derivation until such time as they are more fully developed for specifically-acting chemicals.

For chemicals with non-specific modes of action, and for which no other ecotoxicity data exist, QSARs could be used to estimate toxicity, and those estimated toxicity values can be utilized in the species sensitivity distribution and/or the assessment factor (AF) criteria derivation methods. However, if at least one measured toxicity datum is available, the AF method should be used to derive criteria (section 2-3.2). Since pesticides used in California must have been tested at least minimally for registration, there should not be cases where no measured toxicity data are available. Therefore, the new methodology does not include a QSAR component as a way to supplement data sets. However, QSARs are included as an option for estimating toxicity to threatened and endangered species for narcotic acting chemicals. The RIVM (2001) guidelines provide

19 QSARs for estimating chronic effects of narcotic chemicals on aquatic species representing nine different taxa. Other QSARs are available in the literature.

2-2.2.2 Interspecies correlations for estimation of toxicity

The USEPA has developed interspecies correlation estimation software (ICE v. 1.0), which can be used to estimate acute toxicity of all kinds of compounds to aquatic species, genera and families having little to no measured toxicity data, based on species with adequate data sets (Asfaw *et al.* 2003). Toxicity estimates made by interspecies correlations work well within taxonomic families, but less well as taxonomic distance increases. The ICE models generate estimated acute toxicity values with confidence limits to quantify uncertainty. This technique is promising, but needs further development and validation before being used to generate data for criteria derivation. However, this model provides an excellent tool for assessment of potential effects of derived criteria on threatened and endangered species (discussed further in section 2-5.3) and it is included for this use in the new methodology for that purpose.

2-2.2.3 Estimating chronic toxicity from acute data

In many cases, acute toxicity data are abundant, but there is very little acceptable chronic data available, making it difficult to derive chronic criteria. Two approaches are available for estimation of chronic toxicity from acute data. The first, the acute-to-chronic ratio (ACR) approach, is readily usable, even with very small data sets. The ACR approach is discussed in section 2-3.2.5 because the technique is applied during the criteria derivation process. The second approach is the use of time-concentration-effect (TCE) models, and those are described in this section because they generate individual toxicity data that may be used in SSD derivation procedures.

Analysis of ecotoxicity data by time-concentration-effect (TCE) methods provides a means to extrapolate from acute to chronic exposures. The USEPA has developed an acute-to-chronic estimation program that uses TCE models for estimating chronic toxicity from acute data (Ellersieck *et al.* 2003). Three slightly different models are included in the ACE v. 2.0 software package. These are the accelerated life testing model (Mayer *et al.* 2002; Sun *et al.* 1995), the multifactor probit analysis model (Lee *et al.* 1995; Mayer *et al.* 2002), and the linear regression analysis model (Mayer *et al.* 2002; Mayer *et al.* 1994). Each model considers three important factors in toxicity: exposure concentration, degree of response, and time course of effect. Analysis of these three factors from short-term (acute) toxicity tests allows for prediction of effects over long-term (chronic) exposures. The models only work for mortality data, thus will not provide the most sensitive estimate of chronic effects in cases where sub-lethal effects are more sensitive toxicity indicators. The Australia/New Zealand guidelines (ANZECC & ARMCANZ 2000) allow for the use of these same models for estimation of chronic toxicity from acute data.

Chronic values estimated by the models in the ACE program have been validated within species for 7 species of fish exposed to 17 different chemicals (Mayer *et al.* 2002),

and across species for a variety of invertebrates and fish exposed to chlorpyrifos and ammonia (Whitehouse *et al.* 2004). Although TCE models hold promise for estimation of chronic toxicity, lack of appropriate data presents a barrier to their use. Typically, acute toxicity data are reported as LC₅₀ or EC₅₀ values at one or more time points. However, the TCE models must be populated with response data for each concentration at multiple time points. This kind of information is normally collected in acute toxicity tests, but is rarely reported. Whitehouse *et al.* (2004) contacted authors of 385 toxicity studies to try to request the raw data needed for TCE modeling. Only 85 replies were received, and only 34 of those actually contained the needed data. When data are available, TCE models should be used to supplement chronic toxicity data sets. For the new methodology the USEPA ACE program may be used to estimate chronic toxicity from acute data if appropriate data are available. NOEC values derived from ACE may be used along with other chronic data in AF or SSD methods.

2-2.3 Data sources and literature searches

To ensure that toxicity and physical-chemical data are collected without bias, a thorough search of numerous sources is necessary. While review articles and criteria documents are good starting points for data collection, original sources should be found and evaluated whenever possible. The Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) identified data sources from several existing criteria derivation methodologies. All of the sources mentioned are useful, and, as such, are simply compiled and presented, along with other useful sources, in a table (or tables) for the new methodology. The Dutch and Danish methodologies specify that literature searches should go back to 1970 and 1985, respectively (RIVM 2001; Samsøe-Petersen & Pedersen 1995). However, such limits are not necessary. Since pesticides are synthetic compounds (for the most part), any data available from the time a pesticide was first developed should be collected and evaluated.

The new methodology includes tables of physical-chemical and ecotoxicity data sources as well as instructions to conduct literature searches that encompass the lifetime of the pesticide in question.

2-2.4 Data summaries of ecotoxicity data

In reviewing ecotoxicity data it is important to have a systematic way of rating the quality of a given study, but to rate diverse studies fairly requires that similar information be obtained from each. This particularly applies to single-species toxicity data; multi-species laboratory and field or semi-field tests are usually too complex to lend themselves to simplified summarization. To facilitate the data evaluation process in The Netherlands (RIVM 2001) data are put into data tables with the following headings: species (including scientific name), species properties (e.g., age, weight, life stage), analysis of test compound (measured or not, Y of N), test type (flow-through, static-renewal, static), substance purity, test water, pH, water properties (e.g., hardness, salinity), exposure time, test criterion (e.g., LC₅₀ or NOEC), ecotoxicological endpoint (growth, reproduction,

mortality, immobilization, morphological effects, histopathological effects), LC₅₀ values, NOEC values, notes, and reference information.

This sort of summary is helpful, but does not include enough information for thorough evaluation of the quality of the study. Critical information missing from the Dutch data summary tables includes: where test organisms reside, control response, source of test organisms with confirmation that they were collected from non-polluted sites, detailed test design information, detailed water quality information, concentrations of any carrier solvents used, statistical methods used, and whether responses recorded at NOEC and LOEC concentrations are reasonable. All of this information is needed to adequately rate the quality of test data according to the methods outlined in section 2-2.5. Therefore, a new, more detailed data summary table has been developed to use with the new methodology. The table includes all of the elements mentioned above, as well as space to record any additional information that may be important to the study.

2-2.5 Data evaluation

In this section, the issue of data quality is explored. High quality data are both relevant and reliable. The EU Technical Guidance Document on Risk Assessment (TGD) describes reliability as the inherent quality of a test relating to test methodology and the way the that the performance and results of the test are described, while relevance refers to the extent to which a test is appropriate for a particular hazard or risk assessment (ECB 2003). Reliable data are from studies for which test reports describe the test in detail and indicate that tests were conducted according to generally accepted standards. Relevance is judged by whether a study included appropriate endpoints, was conducted under relevant conditions, and if the substance tested was representative of the substance being assessed. Existing approaches for evaluation of physical-chemical and ecotoxicity data are discussed below, and the strongest elements of many are included in the new methodology.

2-2.5.1 Physical-chemical data

In the context of deriving water quality criteria, physical-chemical data are relevant to the extent that they enhance the interpretation of toxicity data. For example, the organic carbon-water partition coefficient (K_{oc}) can be used to predict concentrations of freely dissolved hydrophobic contaminants in water and whether dietary uptake might be important, while the acid dissociation constant (pK_a) can be used to predict the predominant form of an ionizable compound, and the half-life ($t_{1/2}$) can be helpful in determining if separate criteria are needed for degradation products. Toxicity tests can be evaluated based on whether test solution renewals were adequate considering the half-life ($t_{1/2}$) of the compound, or whether reported toxicity values (i.e., LC₅₀, NOEC, etc.) are reasonable given the water solubility of the compound. Octanol-water partition coefficients (K_{ow}) may be used, in limited cases, to predict toxicity of chemicals using Quantitative Structure Activity Relationships (QSARs) if toxicity data are lacking. To the extent that physical-chemical parameters are affected by temperature, attention should be paid to whether a reported value was measured at a relevant temperature; if not, then

physical-chemical values should be adjusted as necessary (RIVM 2001; Schwarzenbach *et al.* 1993).

The Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) concluded that reliable physical-chemical data are those determined by current, standard methods (e.g., ASTM, OECD, APHA) applied and performed correctly for the chemical of interest. Non-standard methods may also be appropriate, but only if valid reasons are given for deviation from standard methods, or if studies were done prior to the existence of standard methods, but generally followed currently acceptable methods. In regard to pesticides, which vary widely in characteristics such as hydrophobicity, water solubility, and ionizability, it is particularly important to verify that reported partition coefficients were determined correctly. Thus it is not acceptable to simply use a value reported in a handbook without verifying the value via the original reference. An exception to this is that recommended K_{OW} values may be taken without further review from the LOGKOW database (Sangster Research Laboratories 2004) because all values in this database have been reviewed by partition coefficient experts.

A number of web-based programs are available that can be used to estimate various physical-chemical parameters. These include ClogP3 (www.biobyte.com), the USEPA's Estimation Program Interface Suite (www.epa.gov/oppt/exposure/docs/episuite.htm), and KowWin (www.syrres.com/esc/est_soft.htm). Several databases containing physical-chemical data are also available on the web, including the Environmental Fate Database (www.syrres.com/esc/efdb.htm), EXTOXNET (extoxnet.orts.edu), PHYSPROP (www.syrres.com/esc/physprop.htm), TOXNET (toxnet.nlm.nih.gov), and the US Department of Agriculture, Agricultural Research Service (USDA ARS) Pesticide Properties Database (<http://www.ars.usda.gov/services/docs.htm?docid=14199>). These are helpful resources, but should be used with caution. Estimated parameters should only be used in the absence of measured data. As with physical-chemical values taken from handbooks, any values taken from online databases should be fully referenced as to their original sources and those sources should be checked to ensure that the values were appropriately obtained.

While it is preferable to use only physical-chemical data from studies that can be reviewed and verified, in reality, the only information available may be in a handbook citing data that are not available for review. Such values should be used with caution. In some cases, values can be considered reliable without further review. For example, recommended values from the LOGKOW database (Sangster Research Laboratories 2004) due to the careful review process for values in the database, and physical-chemical data from unpublished manufacturer studies are acceptable due to their wide acceptance. If several values are available for the same physical-chemical parameter, and all were obtained by acceptable methods, the geometric mean of the values measured at the same temperature should be used.

The new methodology provides general guidance on acceptability of physical-chemical data based on information in several existing methodologies (OECD 1995;

RIVM 2001; USEPA 1985; 2003d). More specifically, tables are provided specifying acceptable test methods for determination of physical-chemical parameters such as log K_{OW} , water solubility, dissociation constants (pK_a), and hydrolysis or other degradation rates. Also, a table of on-line databases and calculators that may be used to obtain or estimate physical-chemical parameters is provided.

2-2.5.2 Ecotoxicity data evaluation

By studying criteria derivation methodologies for Phase I and working through the process of collecting and evaluating chlorpyrifos effects data it has become clear that ecotoxicity data need to be evaluated on three points: 1) relevance to criteria derivation; 2) documentation; and 3) acceptability. Documentation and acceptability together define the reliability of a study. The ECOTOX (2006) system for rating documentation of aquatic and terrestrial toxicity data from laboratory and field is widely accepted and is included in the new methodology, with some modifications by way of added detail. A similar system is used for evaluation of acceptability and relevance. The elements upon which a test is judged are similar for rating documentation and acceptability, but in the former case, scores are based solely upon whether or not an item was documented, while in the latter case, scores are based upon whether or not a given parameter was within accepted testing guidelines and organism tolerances. Weighting of scores for acceptability is based upon test acceptability criteria as stated in standard methods (e.g., USEPA, OECD, APHA or ASTM toxicity test methods). For example, control response and temperature control, which are common measures of tests acceptability, are weighted more heavily as measures of reliability than water hardness and alkalinity, which are not as critical. As there are no standard methods for multi-species lab, field or mesocosm/microcosm, or wildlife studies, these studies should be rated primarily on documentation, but also on a few key acceptability criteria as described by OECD (1995) and RIVM (2001). The new methodology includes detailed data scoring schemes, with point values assigned to each toxicity test element (see Tables 3.6-3.10, Chapter 3).

The elements for judging relevance of a study are somewhat different, but can be weighted and rated in a similar fashion. The relevance scores should be weighted such that failure of just one of a number of critical factors immediately renders the study irrelevant for criteria derivation. Tests that involve in vitro exposures of organs or tissues (i.e., were not whole-body exposures) and tests which report toxicity values greater than 2x the water solubility of the pesticide are not useful even as supporting information and can be eliminated without further consideration (see discussion below). For compounds with log K_{OW} between 5 and 7, laboratory test feeding regimes should eliminate or minimize interaction of pesticide with food particles (see discussion in section 2-2.1.5). All other effects studies should be evaluated for relevance based on the following critical factors: controls must be documented and must meet minimum test acceptability requirements; tests must be with species belonging to families that reside in North America; endpoints must be linked to survival, growth or reproductive effects; tests must produce toxicity values (i.e., no $>$ or $<$ values); tests must be with freshwater species, and; tests must be conducted with pesticide that is $\geq 80\%$ pure (i.e., no formulations and no mixtures). Only single-species toxicity tests need to be evaluated for relevance

because they are the only ones that may be used directly for criteria derivation. Tests that fail to meet any one of these criteria may not be used to derive criteria or for derivation of acute-to-chronic ratios, but may be used as supporting information as long as they are rated highly enough (details of rating scheme follow).

The decision to include ecotoxicity data in cases where the reported toxicity values are up to 2 x greater than the water solubility of the pesticide is based on a recent review by Shen & Wania (2005) as well as data from the PAN and PHYSPROP databases (PAN 2006; PHYSPROP 2006). These sources show that water solubility values reported for a given pesticide can vary by greater than 100-fold, but it is common for reported water solubility values to vary by a factor of 2. Thus it seems reasonable to accept toxicity data that are within a factor of two of the geometric mean of acceptable solubility values.

For single-species toxicity data in this evaluation scheme, documentation and acceptability scores are averaged, resulting in one score for reliability. The reliability and relevance scores are then used to give the final rating. The scores for the two categories (relevance and reliability) are set (see Table 3.11) so that studies with the highest scores are considered relevant/reliable (R), studies receiving middle scores are considered less relevant/less reliable (L), and lowest scoring studies are considered not relevant/not reliable (N). When the two categories are combined, the final ratings fall into one of three categories: 1) those that may be used for criteria derivation (rating = RR); 2) those that may be used as supporting information (rating = RL, LR, LL); and 3) those that are not usable at all (any with an N in the rating).

To establish a rating scale, the chlorpyrifos data set collected for this report was evaluated using the scoring systems detailed in Tables 3.6-3.8 of Chapter 3. The chlorpyrifos set was broken down as follows: scores in the 75th or higher percentile of all scores were rated reliable; scores between the median and the 75th percentile were rated less reliable; and scores below the median were rated unreliable. Similarly, relevance scores in the 90th or higher percentile were rated relevant; scores between the median and the 90th percentile were rated less relevant; scores below the median were rated not relevant. The 75th percentile of scores is suggested for the reliability rating because, in the case of the chlorpyrifos data set, higher percentiles were too restrictive, resulting in rejection of too much data that others have accepted for criteria derivation (Siepmann & Finlayson 2000; USEPA 1986a). On the other hand, the selection of the 75th percentile resulted in rejection of a few tests accepted by others, indicating that this rating system is a bit more rigorous than those used previously. The relevance scoring system was designed to include six major requirements for a study to be used in criteria derivation. Lack of one of these requirements would lower the score below 90 so only studies scoring above 90 should be used for criteria derivation. Based on this translation of numeric scores into ratings for the chlorpyrifos data set, a numeric scale in Table 3.11 (Chapter 3) was established for rating other pesticide data sets in the new methodology.

All other chlorpyrifos effects data (i.e., not single-species) were evaluated on documentation and reliability using the scoring schemes in Tables 3.9 and 3.10. Aquatic

and terrestrial studies receiving R or L ratings may be used as supporting data in the criteria review process; studies rated N may not be used.

The data evaluation system presented here has been incorporated into the new methodology. The system includes procedures for scoring relevance and reliability of single-species, multispecies, laboratory, field, semi-field, microcosm and mesocosm studies, as well as a table describing ranges of scores that define the categories RR, RL, RN, etc.

2-2.6 Data quantity—ecotoxicity

The new methodology includes specific instructions regarding number and taxonomic diversity of data required for criteria derivation by SSD and AF methods. First discussed is the number of data required.

Several current methodologies use toxicity data from five species to derive criteria by statistical extrapolation. The Australian/New Zealand methodology (ANZECC & ARMCANZ 2000) uses five single-species chronic NOEC values. OECD (1995) guidelines use statistical extrapolations by Aldenberg & Slob (1993) or Wagner and Løkke (1991) that require at least five chronic NOECs. The California Department of Fish and Game has derived criteria using the USEPA (1985) SSD method with fewer than the eight required families, using professional judgment to determine that species in the missing categories were relatively insensitive and their addition would not lower the criteria (Siepmann & Jones 1998; Menconi & Beckman 1996).

Although more data would improve fit, the use of five data for statistical extrapolations by parametric techniques has been supported (Aldenberg & Luttik 2002; Okkerman *et al.* 1991). According to Aldenberg & Slob (1993) the risk of under-protection of a 50% confidence limit estimate of the hazardous concentration for 5% of organisms (HC₅; based on a log-logistic distribution) decreases considerably as sample size is increased from 2 to 5, but less so as it is increased from 5-10 and from 10-20. Looking ahead at the chosen distribution (as described in section 2-3.1), the Burr Type III distribution and software, BurrliOZ v. 1.0.13, is used with minimum of five data (ANZECC & ARMCANZ 2000). Therefore five data is the minimum data required for criteria derivation by the new method.

When five ecotoxicity data are not available (data sets as small as one datum), only assessment factor derivation methods are appropriate (see sections 2-3.0 and 2-3.2 for description of assessment factors). Minimal data sets available for derivation of criteria in California will be those required for registration under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA, U. S. Code Title 7 1947) and those required by the California Department of Pesticide Regulation (California DPR 2005). According to 40 CFR Part 158.490 (USEPA 1993) the minimum data required for registration under FIFRA is an LC₅₀ for a fish and an LC₅₀ for a freshwater invertebrate. All other kinds of aquatic toxicity data are only conditionally required depending on planned pesticide usage, potential for transport to water, whether any acute LC/EC₅₀ values were < 1 mg/L,

whether estimated environmental concentrations are > 0.01 times any LC/EC₅₀, or if data indicate reproductive toxicity, persistence, or bioaccumulative potential. It is possible that for many new chemicals, only the two acute toxicity data will be available. The DPR has tiered data requirements (California DPR 2005). The minimum data set includes LC₅₀ values for one warm water fish, one cold water fish, and for a freshwater invertebrate. Further testing is required for the same reasons discussed for FIFRA. Again, it is possible that no more than the minimum data will be available for criteria derivation for new pesticides. An assessment factor criteria derivation method is needed for these very small data sets.

The overall quantity of data is one consideration, but it is also important to specify the range of taxonomic diversity that should be represented in the data set. The USEPA methodologies (USEPA 1985; 2003d) have the most comprehensive, specific taxonomic requirements among the reviewed methodologies and are used for the new methodology, with some modification to reflect species of importance to the Central Valley of California and to reduce the required number of data from 8 to 5.

First, consider the following taxonomic requirements in the USEPA (1985; 2003d) methods:

For derivation of acute or chronic criteria by the SSD method (minimum of 8 acute or chronic data):

- a. The family Salmonidae in the class Osteichthyes;
- b. One other family (preferably a commercially or recreationally important, warm water species) in the class Osteichthyes (e.g., bluegill, channel catfish);
- c. A third family in the phylum Chordata (e.g., fish, amphibian);
- d. A planktonic crustacean (e.g., cladoceran, copepod);
- e. A benthic crustacean (e.g., ostracod, isopod, amphipod, crayfish);
- f. An insect (e.g., mayfly, dragonfly, damselfly, stonefly, caddisfly, mosquito, midge);
- g. A family in a phylum other than Arthropoda or Chordata (e.g., Rotifera, Annelida, Mollusca);
- h. A family in any order of insect or any phylum not already represented.

With regard to plants, the USEPA (1985) guidelines require results from at least one test with a freshwater alga or vascular plant, while the Great Lakes methodology (USEPA 2003d) indicates that plant data are desirable, but not required. In both cases, if plants are among the most sensitive (as is likely with herbicides), then tests with plants representing at least two phyla are required.

For derivation of acute criteria by the assessment factor method:

- a. At least one datum from the family Daphniidae; species must be from the genus *Daphnia*, *Ceriodaphnia*, or *Simocephalus* (USEPA 2003d).

For derivation of an acute-to-chronic ratio (minimum of 3 chronic data):

- a. At least one fish;
- b. At least one invertebrate;
- c. At least one acutely sensitive freshwater species (the other two may be saltwater species).

With these requirements in mind, and with the minimal data sets available from pesticide registration, the following taxonomic requirements are in the new methodology:

For derivation of acute or chronic criteria by the SSD method (minimum of 5 data):

- a. The family Salmonidae;
- b. A warm water fish;
- c. A planktonic crustacean, of which one must be in the family Daphniidae in the genus *Ceriodaphnia*, *Daphnia*, or *Simocephalus*;
- d. A benthic crustacean;
- e. An insect (aquatic exposure);

The rationale for exclusion of items c, g, and h in USEPA list is as follows:

- c. Two fish species (one warm water and one cold water) are sufficient to represent the phylum Chordata;
- g. Rotifers, annelids and mollusks are typically insensitive to pesticides (e.g., Giesy *et al.* 1999);
- h. This category is very general and simply fills out the eight minimum data required by the USEPA SSD method (USEPA 1985; 2003d).

For determination of acute criteria by the assessment factor method (minimum of 1 datum):

- a. The family Daphniidae in the genus *Ceriodaphnia*, *Daphnia*, or *Simocephalus*.
- b. Additional data must be from different families as per the list of those required for the SSD method. For example, to derive insecticide criteria, if data are available from two acceptable studies, then one must be from the family Daphniidae and the other must be either a member of the family Salmonidae, a warm water fish, a benthic crustacean or an insect. If three data are available, then one must be in the family Daphniidae and the others must be from two other, different families, and so on such that each additional datum contributes toward completing the minimum data set required for the SSD method.

Item b. is added to ensure that the magnitude of the assessment factor is only reduced in cases where data are available for multiple families, and to encourage generation of data that would complete the minimum SSD set. For determination of acute-to-chronic ratios, the requirements in USEPA guidance (USEPA 1985; 2003d) are acceptable; including the use of saltwater species if not enough freshwater data are available.

Additionally, alga or vascular aquatic plant data must be included for herbicides. The plant requirement for herbicides is added because herbicides are expected to be more toxic to plants than to animals. Since life cycles of plants vary widely, procedures for conducting toxicity tests with plants are not well developed, and explicit definitions for acute plant tests are not included, the methodology for herbicides will differ slightly.

Plant data are not required for the acute distribution, but an acute criterion should still be derived with animal data according to the requirements above. If the chemical is an herbicide and plants are the most sensitive group, options for chronic data are:

- 1) Fit a distribution with only alga or vascular aquatic plant data, if there are data from at least 5 different species that were rated RR.
- 2) If there are not data from 5 different species following the requirements or a distribution cannot be fit, use the lowest NOEC value from an important alga or vascular aquatic plant species that has measured concentrations and the endpoint is biologically relevant.

Option 2 is similar to the EPA Final Plant Value (FPV) from their 1985 methodology. An alternative method is needed for herbicides because without acute alga or plant data an ACR cannot be calculated for plants. Additionally the AF procedure included in this methodology would not be appropriate as these factors were derived for acute data based on animal requirements, however, AF may still be used to derive an acute value for herbicides based on animal data.

Many pesticides are also fungicides. The mode of action for fungicides, however is not specific to fungi, so fish or invertebrates may be very sensitive. Similar to herbicides, ecological risk assessment for fungicides is a topic that has not received much attention. Maltby *et al.* (2009) found that using plant and animal data in the SSD appeared to derive protective criteria; however, there are very limited data on fungi to assess effects on non-target fungi. The new method will require the same 5 taxa for animal taxa described above for fungicides.

2-2.7 Data reduction

Data that have been rated as relevant and reliable require further reduction prior to criteria derivation so that no species receive undue weight in the derivation process. For example, if there are multiple data for a particular species, then some method has to be specified for reducing those data into a single point within the SSD for that species. As discussed in the Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) the geometric mean is a reasonable approach for reducing data from multiple tests to a single number for criteria derivation. No species should be represented more than once in the final SSD.

SSD procedures assume that toxicity data in the distribution represent an unbiased sample of the system to be protected (Forbes & Calow 2002). Current USEPA procedures

(USEPA 1985; 2003d) utilize genus mean toxicity values on the grounds that this reduces potential bias due to overabundance of data for species from a few genera, and it minimizes statistical problems that arise due to non-random sampling (i.e., the close relationship between organisms within a genus prevents organisms from responding independently). This approach does not seem entirely justified because there is considerable variability of sensitivity between species within a genus in some cases. For example, Harmon *et al.* (2003) report an EC₅₀ (immobility) for *Daphnia ambigua* exposed to chlorpyrifos of 0.035 µg/L, while Van Der Hoeven & Gerritsen (1997) report an EC₅₀ (immobility) for *Daphnia pulex* of 0.42 µg/L, a 10-fold difference. In addition, sets of toxicity data acceptable for criteria derivation are usually quite small (e.g., in Erickson & Stephan 1988, data sets ranged from 8 to 45 values); thus, general statements about the intra-genus variability are not well-supported. For these reasons, and since no other existing methodology uses the genus-level approach, the new methodology utilizes data at the species level for both acute and chronic criteria derivation.

Specific data reduction methods used in existing methodologies were described in Phase I (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009). Following is a compilation of those methods that are used for reducing data to species mean acute values (SMAVs) and species mean chronic values (SMCVs) in the new methodology.

- 1) Calculate SMAVs/SMCVs as the geometric mean of toxicity values from one or more acceptable tests with the same endpoints (ANZECC & ARMCANZ 2000; ECB 2003; OECD 1995; RIVM 2001; USEPA 1985; 2003d);
- 2) If data are available for life stages that are at least a factor of two more resistant than another life stage for the same species, then do not use the data for the more resistant life stage to calculate the SMAV because the goal is to protect all life stages (RIVM 2001; USEPA 1985; 2003d);
- 3) If data are available for one species, but for multiple endpoints, then use the data for the most sensitive endpoint (ANZECC & ARMCANZ 2000; ECB 2003; OECD 1995; RIVM 2001); if multiple endpoints are equally sensitive, then note both endpoints, but use only one value for criteria calculation;
- 4) If a NOEC is not explicitly reported in chronic toxicity studies, but statistical analysis was done, the NOEC may be determined as the highest reported concentration not statistically different from the control ($p < 0.05$, RIVM 2001); the NOEC is not used in criteria derivation, but is needed for calculation of the MATC;
- 5) Similarly, if a LOEC is not explicitly reported in chronic toxicity studies, it may be determined as the lowest reported concentration that is statistically different from the control ($p < 0.05$); the LOEC is not used in criteria derivation, but is needed for calculation of the MATC;
- 6) If a MATC is not reported, it may be calculated as the geometric mean of the NOEC and LOEC;

7) If no toxicity values were reported, but raw data are available, calculate toxicity values using appropriate statistical methods (ECB 2003);

8) If a MATC is expressed as a range of values, recalculate the MATC as the geometric mean of the high and low values (RIVM 2001);

9) If reasons for differences between tests for the same species/endpoints are found, then data may be grouped according to appropriate factors (e.g., pH or temperature; ECB 2003). Selection of the appropriate value to use in criteria derivation should be based on standard test parameters. Tests conducted under non-standard conditions (vs. standard conditions as defined in standard test methods) may be used to derive quantitative relationships between those conditions and toxicity (as in USEPA 1985; 2003d). If such a relationship is established then toxicity values derived under non-standard conditions may be translated to standard conditions and added to the criteria derivation data set. If no quantitative relationship can be derived then tests conducted under non-standard conditions should not be used for criteria derivation, but may be used as supporting information.

10) If data are available for multiple time points from crustacean or insect acute toxicity studies use the latest time point (i.e., 96-h tests are preferred over tests of < 96 h);

11) For a given species, use data from flow-through tests in which concentrations were measured, if it is available. If such data are not available, then data from static or static-renewal tests and/or tests in which concentrations were not measured may be used as long as they are rated otherwise relevant and reliable.

12) Further reduction may be needed in the course of SSD analysis. If data cannot be described by or fit to a distribution, then the set should be examined for outliers and/or bimodality. The new methodology includes detailed guidance for detection of outliers and exclusion of outliers. If data are bimodally distributed (as determined visually), use only the lower of the two groups for criteria derivation (ANZECC & ARMCANZ 2000); the effects of data exclusions on the criteria must be explored and explained (ECB 2003).

Data reduction procedures excluded from this list include those that equate NOECs with some percentage reduction from control responses and those that use factors to estimate NOECs from LOECs, with factor size dependent on level of response (ECB 2003; RIVM 2001). As discussed in section 2-2.1.2, there is little agreement as to what levels of effect constitute “no effect” (in terms of statistical or biological significance) so these extrapolations are unreliable. Also excluded was a procedure to exclude data for species in cases where toxicity values differ by more than a factor of 10 (USEPA 1985; 2003d). The approach for excluding data is not explained (USEPA 1985; 2003d) and step 12 in the data reduction procedure should adequately manage outliers. Also excluded was a procedure whereby multiple data for a given species were reduced by consideration of which studies best reflected regional environmental parameters (ECB 2003). This procedure makes little sense for laboratory toxicity data obtained in standardized tests,

but could be useful in choosing field or semi-field data to support the criteria derivation process.

Based on this discussion, guidance is provided in the new methodology for selecting or calculating values for use in criteria derivation. Instructions are given for how to reduce multiple data for a given chemical/species combination to a single species mean acute or chronic value, and for how to manage bimodal distributions and outliers.

2-3.0 Criteria Calculation

Criteria need to have components of magnitude, duration, and frequency to be most useful to environmental managers (conclusion of Phase I report; TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009). Exposure duration is partially addressed through the derivation of separate acute and chronic criteria (USEPA 1985; 2003d). Further consideration of allowable duration and frequency of exceedances are discussed in sections 2-3.3 and 2-3.4. The use of species sensitivity distribution (SSD) and assessment factor (AF) extrapolation methods for determination of criteria magnitudes are addressed in this section.

The aim of both SSD and AF methods is to extrapolate from available toxicity data for a limited number of species to toxicity values that will be protective of all species in an ecosystem. The assessment factor (AF) method involves multiplying the lowest value of a set of toxicity data by a factor to arrive at a criterion. The SSD method involves the use of one of several similar statistical extrapolation techniques (described in the next section) to determine the criterion. The Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) concluded that if at least five data are available from five taxonomic groups (defined in section 2-2.6), then the SSD method should be used. If fewer than five data are available, then the AF method should be used. There are a number of approaches to each of these methods, and the purpose of this section is to compare approaches using several example data sets, and to select the best one for inclusion in the new methodology.

2-3.1 SSD methodology

The SSD method uses a statistical probability distribution to determine the criterion. A probability distribution that is used to model the variability of species sensitivities to a toxicant (i.e., toxicity data) is also known as a species sensitivity distribution (SSD). Such distributions can be used to estimate concentrations that are likely to fall below the sensitivity of a major portion of species, often 95% of species, thus aiming to protect most species in the ecosystem. (The significance of the percentile cutoff is discussed in section 2-3.1.2, and for more general description of an SSD see TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009.)

There are five points on which SSD methodologies may differ: 1) the shape of the distribution; 2) the kinds and quantity of data; 3) the level of confidence associated with derived criteria; 4) how data are aggregated; and 5) what percentile cutoff is the best

predictor of no-effect concentrations. Item 2 has been addressed in sections 2-2.1 and 2-2.6. Items 1, 3 and 4 will be evaluated in more detail in the following sub-sections. Item 5 was discussed in the Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) and the 5th percentile was determined to be the best predictor of no-effect levels. However, since this is a critical issue in criteria derivation, the discussion is briefly revisited here.

2-3.1.1 Distribution shape and fit

Of the five points above, the shape of the distribution is one of the most important considerations in choosing a distribution. The shape of the distribution determines how well the distribution fits the data. The SSD approach assumes that the toxicity data are a random sample of all species and that if all species were sampled, they could be described by some defined distribution. When a specific distribution is chosen the assumption is that toxicity data from all species in an ecosystem (if it were obtainable) would fit the shape of that specific distribution. The goal is to try to select a distribution that minimizes violations of that distributional assumption and fits the data well.

In the Phase I report, distributions used by major agencies around the world were reviewed for estimation of community or ecosystem effects based on single-species toxicity tests (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009). In this section the fit of three most commonly used distributions are tested: the log-triangular, log-normal and Burr Type III distribution. The goodness of fit and distributional assumptions of these three techniques were evaluated in this section by fitting the distribution to existing data sets and testing the fit of the distribution. The data sets used include the chlorpyrifos acute toxicity data collected for this project (a more detailed table of this data set is included in Chapter 4), as well as acute pesticide data sets taken from Erickson & Stephan (1988) and diazinon and atrazine data sets from USEPA criteria documents (USEPA 2003c; 2005a). These data are shown in Table 2.1. Toxicity values in the USEPA diazinon and atrazine sets that were reported as > or < values were excluded from this analysis, as they are not usable in any but the log-triangular distribution method. The highest DDT value and the two highest aldrin toxicity values were excluded because they were greater than two times the water solubility of the compound (as discussed in section 2-2.5.2.)

Figure 2.1 shows the log-normal distributional fits along with box plots and histograms of each data set (output from JMP IN v. 5.1.2, JMP 2004). Figure 2.2 shows log-triangular, log-normal and Burr Type III fits for these data sets (constructed using Excel v. 11.2.3). The log-triangular distributions were constructed according to the following parameters (Evans *et al.* 2000):

$$CDF = \frac{(x - a)^2}{(b - a)(c - a)} \text{ for } a \leq x \leq c; \quad (2.1)$$

$$\text{and } CDF = \frac{(b - x)^2}{((b - a)(b - c))} \text{ for } c \leq x \leq b \quad (2.2)$$

$$Mean = \frac{(a + b + c)}{3} \quad (2.3)$$

where:

CDF = cumulative distribution function

x = value in data set

a = minimum value in data set

b = maximum value in data set

c = mode

Table 2.1 Acute toxicity data sets. Cpf = chlorpyrifos; Txp = toxaphene; End = endrin; Lin = lindane; Ald = aldrin; Diel = dieldren; Hept = heptachlor; Chl = chlordane; Endos = Endosulfan; Dia = diazinon; Atr = atrazine; all toxicity values in µg/L.

Rank	Cpf ¹	DDT ²	Txp ²	End ²	Lin ²	Ald ²	Diel ²	Hept ²	Chl ²	Endos ²	Dia ²	Atr ²
1	0.035	0.36	0.8	0.15	2.0	4.0	2.5	0.9	3.0	0.34	0.3773	3000
2	0.0427	1.1	1.3	0.32	10	4.5	4.5	1.1	6.3	0.83	0.7764	5300
3	0.06	1.4	1.962	0.33	10.5	6.1	5.0	1.8	15	2.3	1.048	6300
4	0.0654	1.6	2	0.41	32	7.4	6.1	2.8	26	3.2	1.587	6700
5	0.100	1.7	2.3	0.44	32	8.0	8.0	7.8	26	3.7	2.04	14700
6	0.150	1.7	3	0.46	40	9.0	8.1	13.1	37	3.8	6.51	20000
7	0.220	1.9	3.1	0.47	44	10	10.8	23.6	40	5.8	10.7	27000
8	0.25	1.9	3.446	0.54	44	16	15	24	45	6.0	16.82	49000
9	1.0	2.4	3.7	0.69	45	21	20	26	56	88	25	60000
10	4.7	2.6	3.822	0.75	48	27	22	29	57	261	425.8	
11	6	3.0	4.874	0.76	55.6	27	24	42	58		459.6	
12	8	3.0	5.782	0.78	64	28	39	47.3	59		602	
13	10	3.2	6	0.85	67.1	32	41	61.3	82		723	
14	15.96	3.5	6.7	1.0	68	34	130	78	190		800	
15	178	3.9	10	1.1	83	42	213	81.9			1643	
16	806	4.0	10.12	1.2	90	45.9	250	101			2166	
17	2410	4.3	10.8	1.3	138	50	567	148			3198	
18		4.9	11.85	1.5	141.1	143	620	320			7804	
19		5.0	12	1.8	207	180	740				7841	
20		7.3	13	2.1	460						8000	
21		7.8	13	3.1	485						9000	
22		7.8	13.78	4.7	676						11000	
23		8.0	14.59	5.9							11640	
24		8.5	14.6	32								
25		9.3	15.68	34								
26		10	16.71	60								
27		12	17.61	64								
28		14	18	352								
29		17	20									
30		18	24									
31		25	26									
32		33	31.75									
33		40	40									
34		48	73.48									
35		48	140									
36		54	210									
37		67	500									
38		68										
39		175										
40		192										
41		362										

¹ Data collected for this project

² USEPA criteria documents (USEPA 1980a; b; c; d; e; f; g; 1986a; b; 2003c; 2005a)

The mode of each log-triangular distribution was determined by dividing the data into bins and taking the average of the maximum and minimum values in largest bin.

The log-normal and Burr Type III distributions in Fig. 2.2 were constructed using Excel (v. 11.2.3). The Burr III distribution has the following cumulative distribution function:

$$F(x) = \frac{1}{\left[1 + \left(\frac{b}{x}\right)^c\right]^k} \quad (2.4)$$

where:

$F(x)$ = probability of x ;
 b , c and k are fit parameters;
 x = data point in the distribution

As $k \rightarrow \infty$, the Burr III distribution approaches the reciprocal Weibull distribution, and as $c \rightarrow \infty$, the Burr III distribution approaches the reciprocal Pareto distribution. Thus, these so called limiting distributions are used in cases where $k > 100$ and $c > 80$, respectively (Campbell *et al.* 2000; CSIRO 2001).

The cumulative distribution function for the reciprocal Weibull distribution is:

$$F(x) = \exp^{-\alpha x^{-\beta}} \quad (2.5)$$

where:

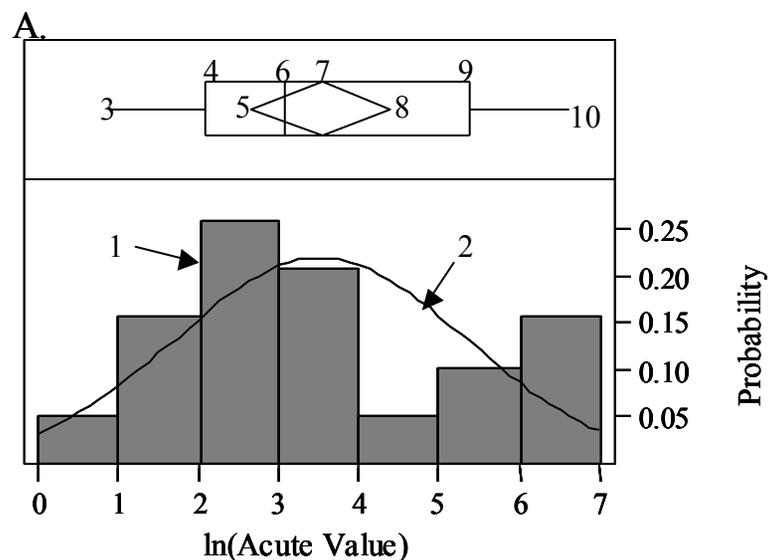
$F(x)$ = probability of x ;
 α and β are fit parameters;
 x = data point in the distribution

For the reciprocal Pareto distribution the cumulative distribution function is:

$$F(x) = \left(\frac{x}{x_0}\right)^\theta \quad (2.6)$$

where:

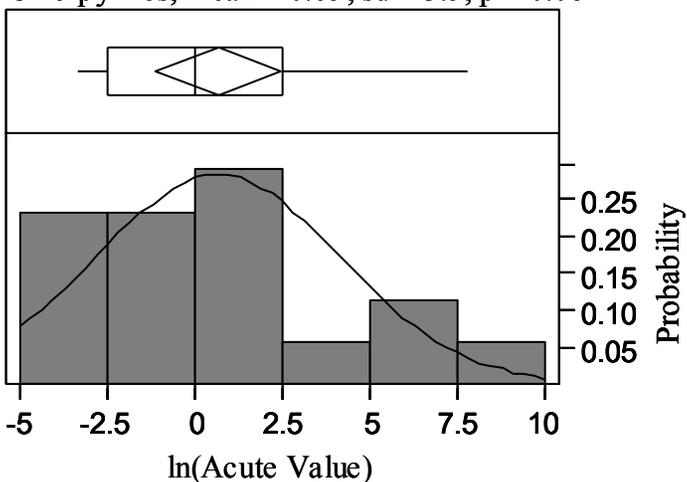
$F(x)$ = probability of x ;
 X_0 and θ are fit parameters;
 X = data point in the distribution



1. Histogram of data set
2. Normal distribution fit
3. Minimum
4. Lower quartile
5. Lower 95% confidence limit of mean
6. Median
7. Mean
8. Upper 95% confidence limit of mean
9. Upper quartile
10. Maximum

B.

Chlorpyrifos; mean = 0.65; sd = 3.5; p = 0.08



DDT; mean = 2.2; sd = 1.6; p = 0.26

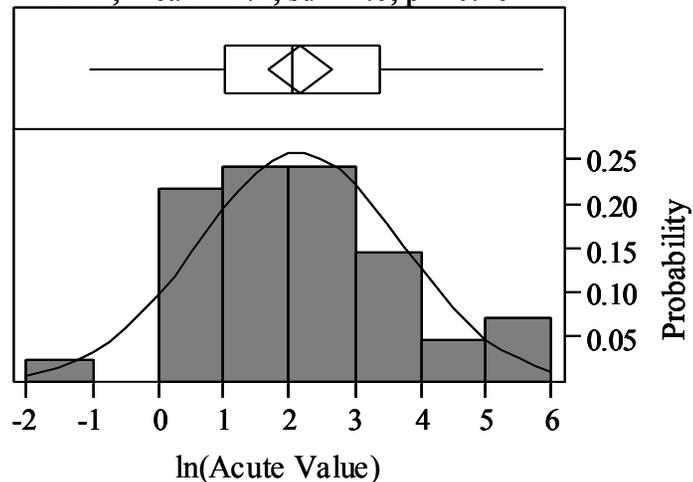


Figure 2.1. Tests for log-normal distribution of data sets. A. Key to distribution diagrams. B. Distributions for each of 12 data sets; $p < 0.05$ indicates lack of fit. ■ indicates outliers (outside 1.5 times the interquartile range).

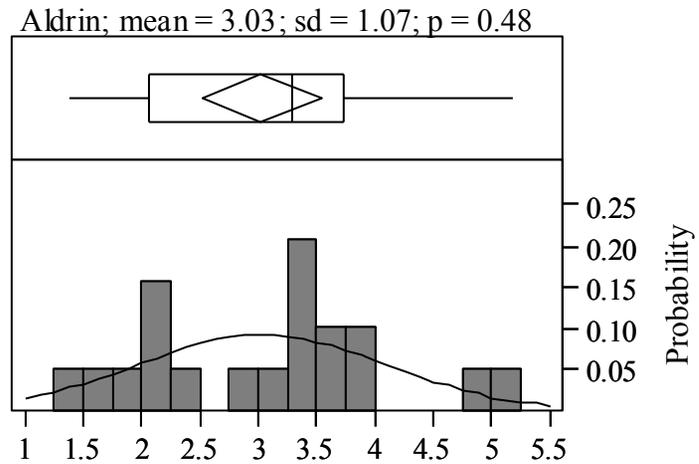
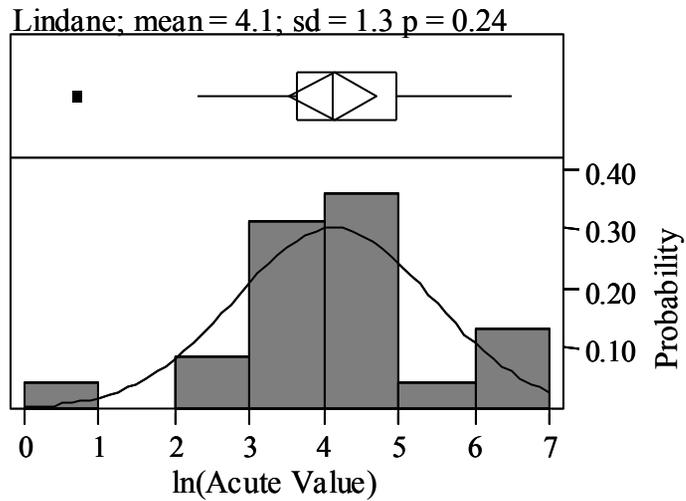
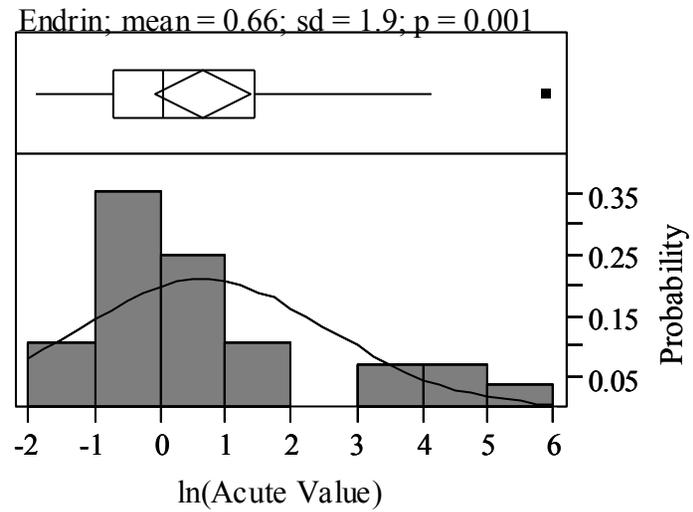
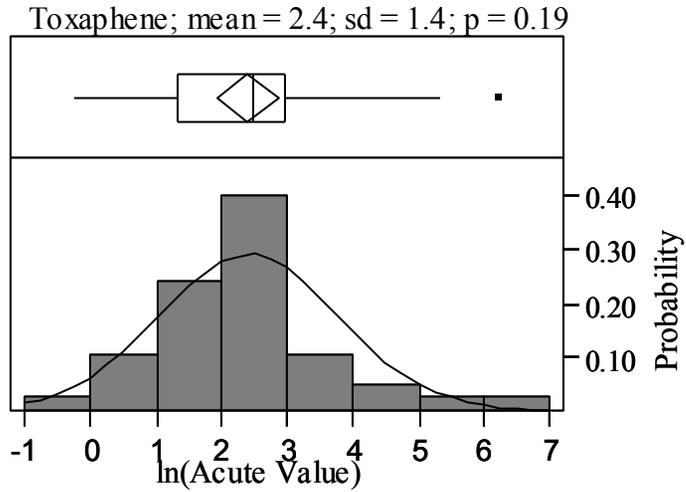
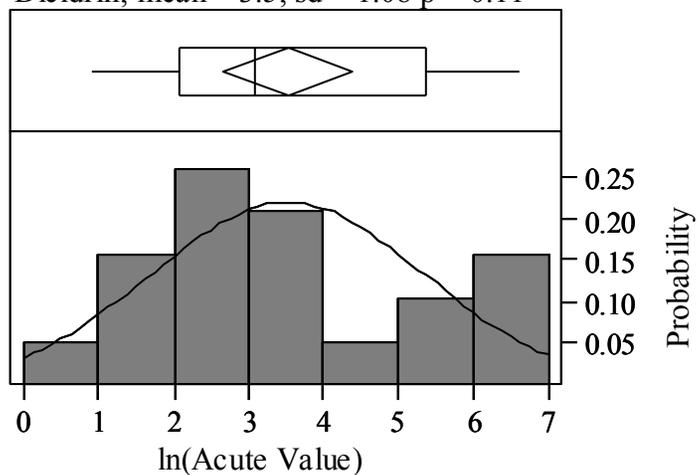
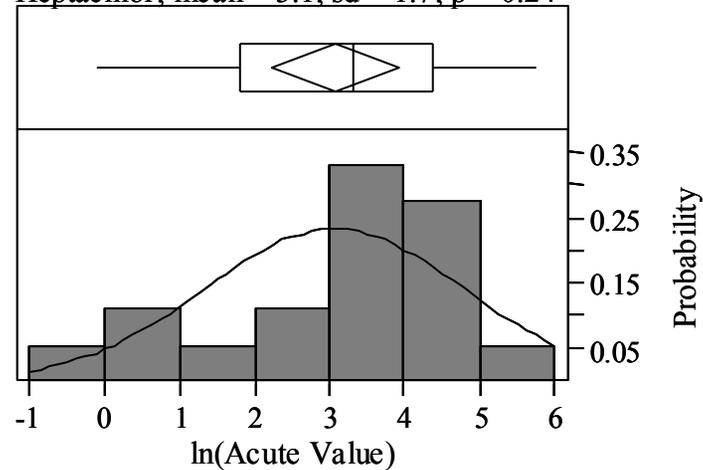


Figure 2.1 (cont.). Tests for log-normal distribution of data sets. A. Key to distribution diagrams. B. Distributions for each of 12 data sets; $p < 0.05$ indicates lack of fit. ■ indicates outliers (outside 1.5 times the interquartile range).

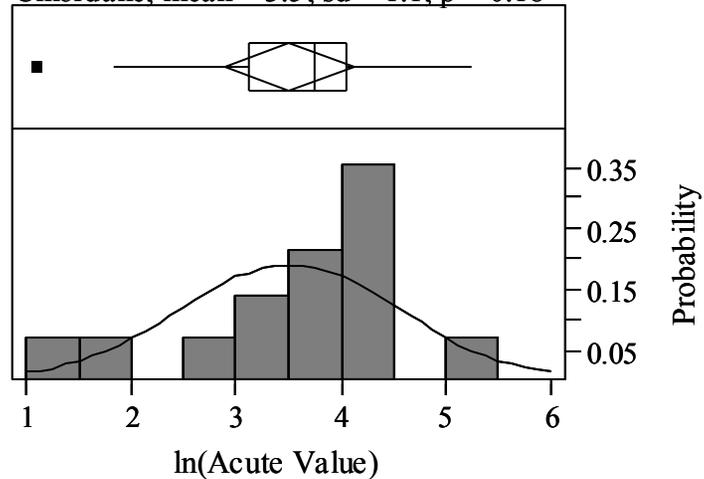
Dieldrin; mean = 3.5; sd = 1.08 p = 0.11



Heptachlor; mean = 3.1; sd = 1.7; p = 0.24



Chlordane; mean = 3.5; sd = 1.1; p = 0.18



Endosulfan; mean = 1.7; sd = 2.0; p = 0.20

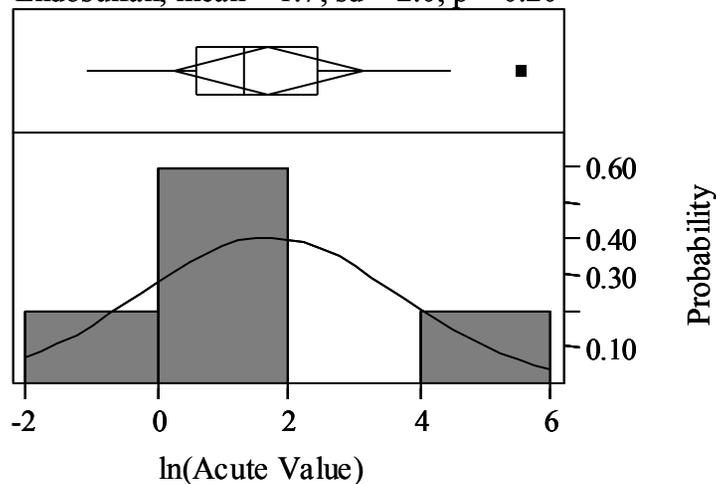


Figure 2.1 (cont.). Tests for log-normal distribution of data sets. A. Key to distribution diagrams. B. Distributions for each of 12 data sets; $p < 0.05$ indicates lack of fit. ■ indicates outliers (outside 1.5 times the interquartile range).

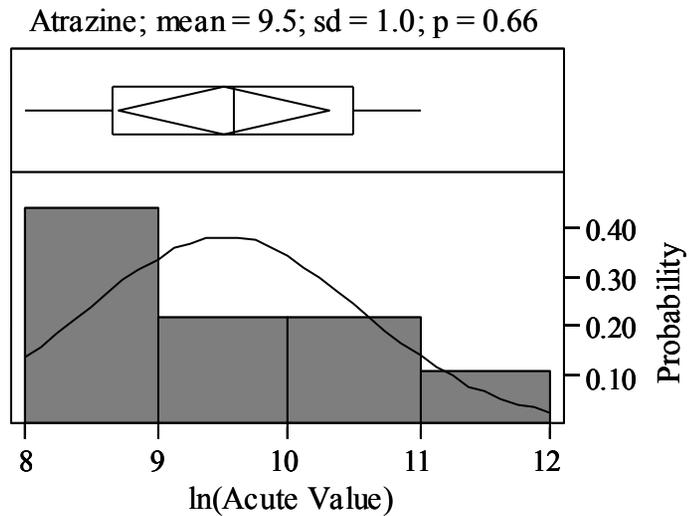
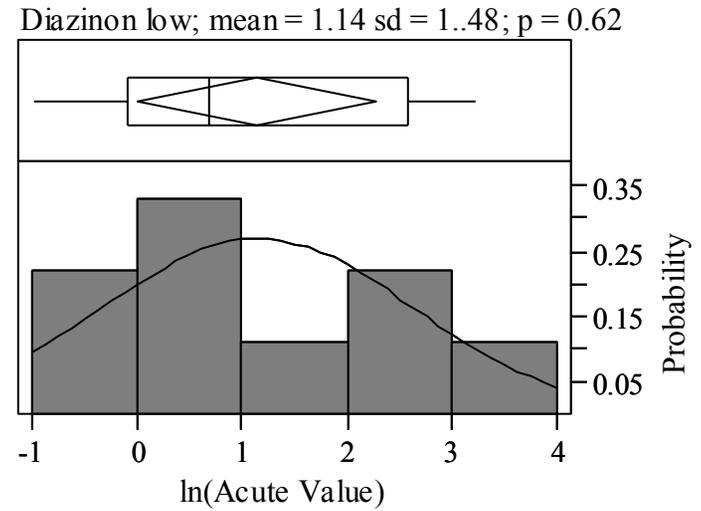
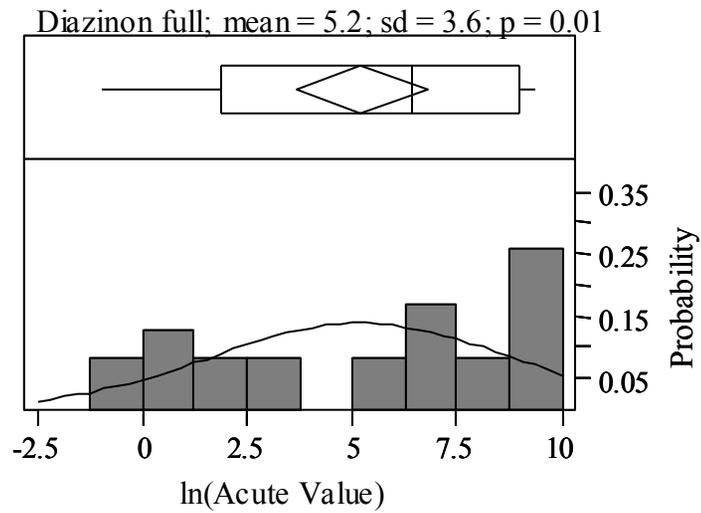


Figure 2.1 (cont.). Tests for log-normal distribution of data sets. A. Key to distribution diagrams. B. Distributions for each of 12 data sets; $p < 0.05$ indicates lack of fit. ■ indicates outliers (outside 1.5 times the interquartile range).

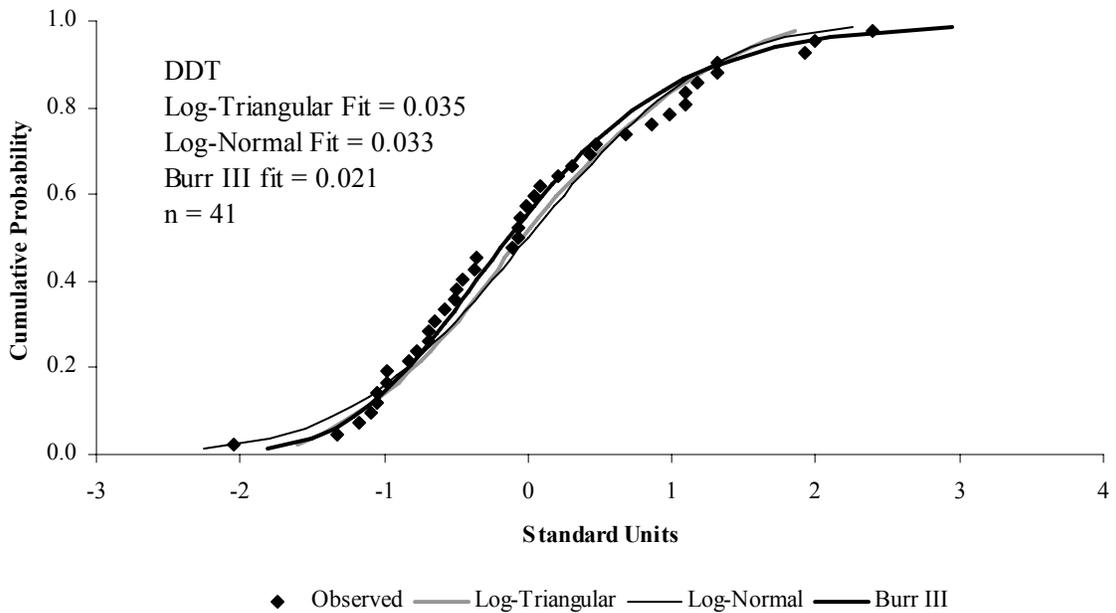
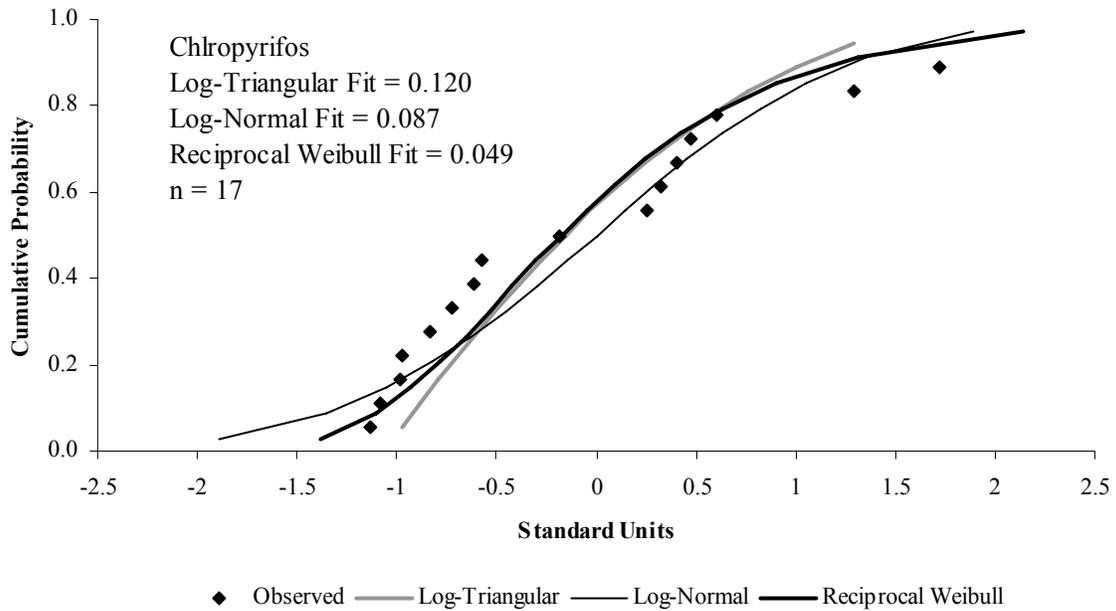


Figure 2.2. Comparison of fits of pesticide toxicity data to log-triangular, log-normal and Burr Type III distributions (including Reciprocal Weibull).

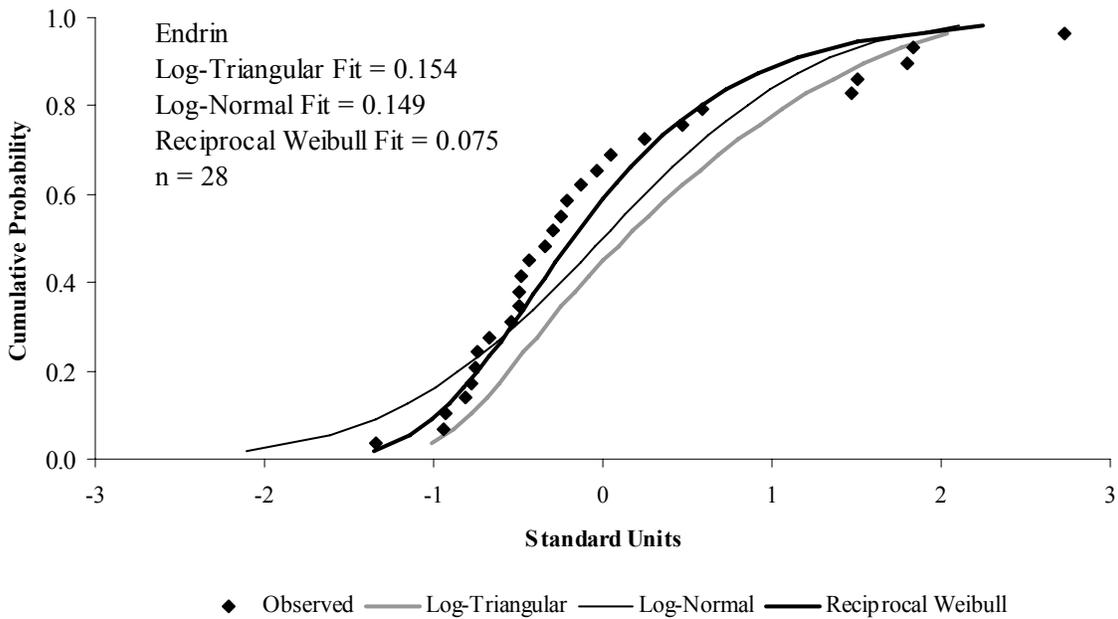
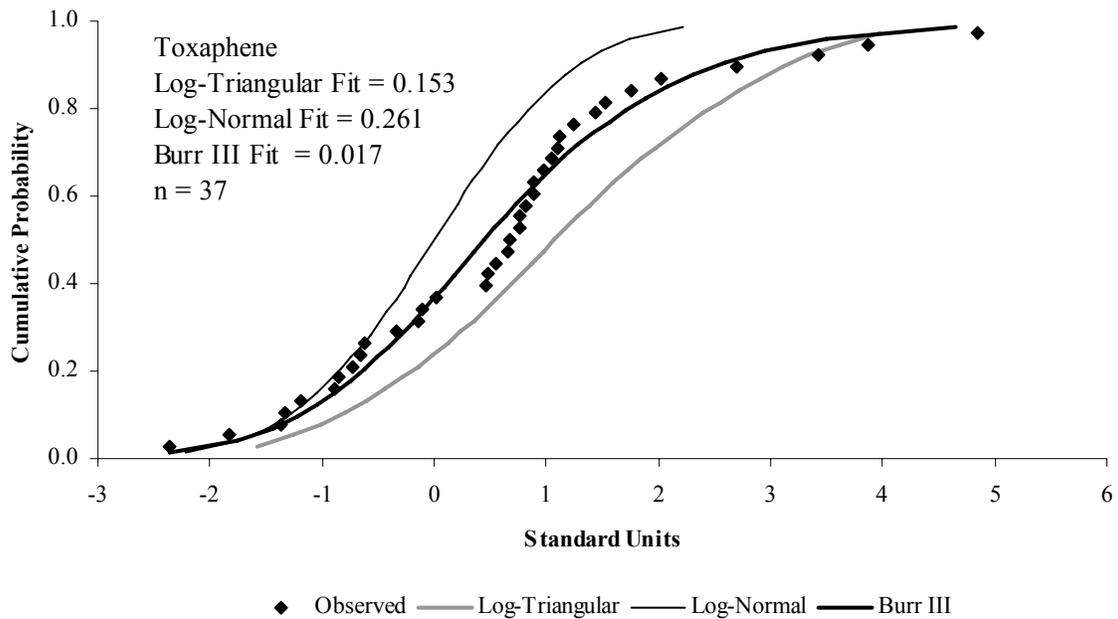


Figure 2.2 (cont.). Comparison of fits of pesticide toxicity data to log-triangular, log-normal and Burr Type III distributions (including Reciprocal Weibull).

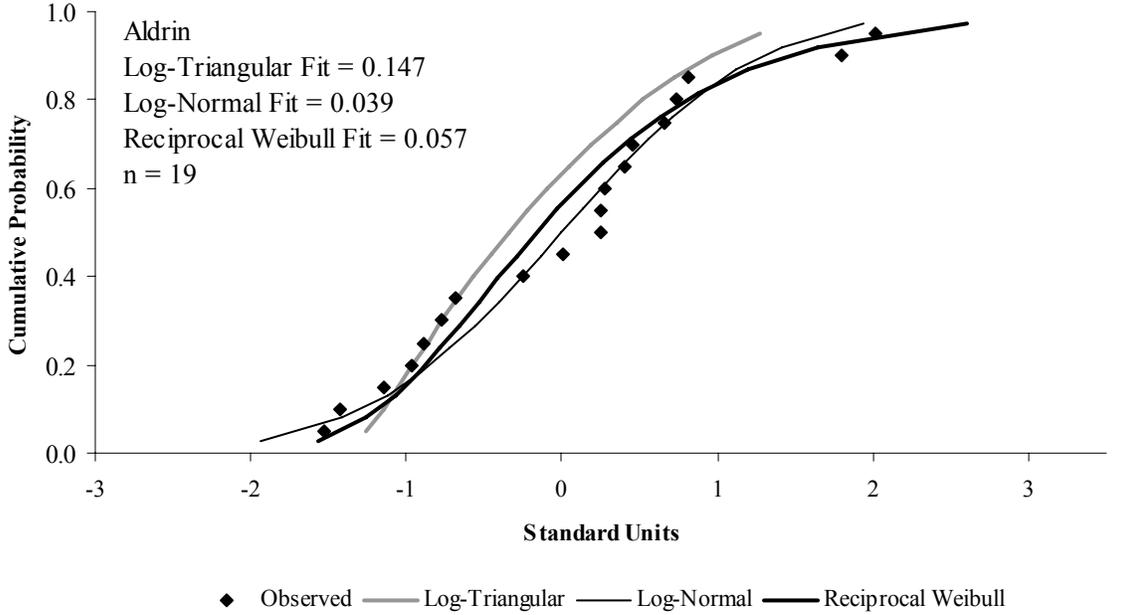
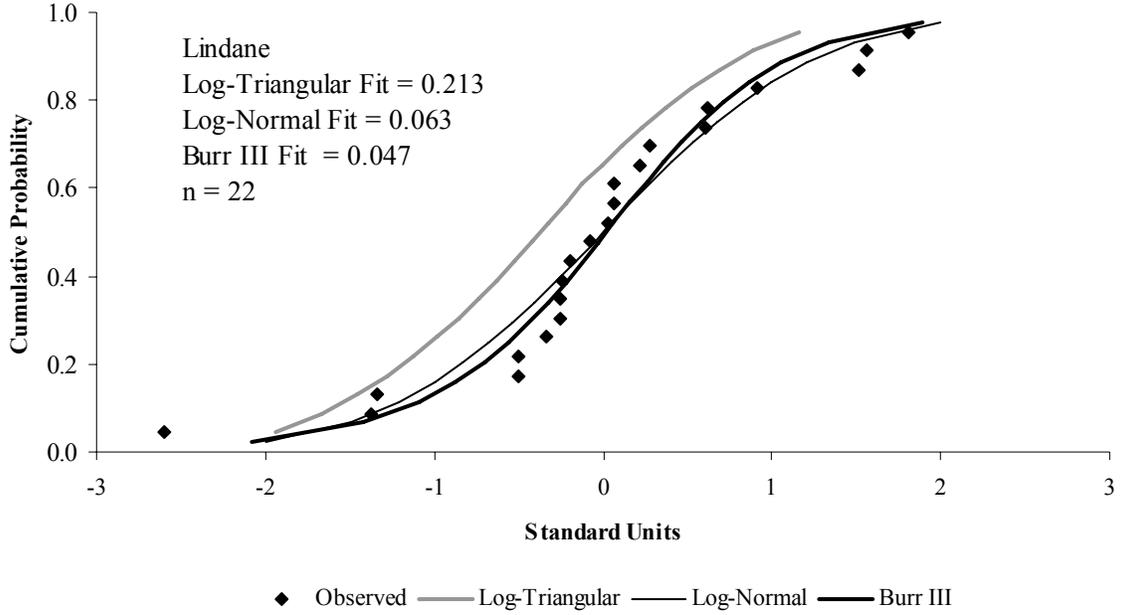


Figure 2.2 (cont.). Comparison of fits of pesticide toxicity data to log-triangular, log-normal and Burr Type III distributions (including Reciprocal Weibull).

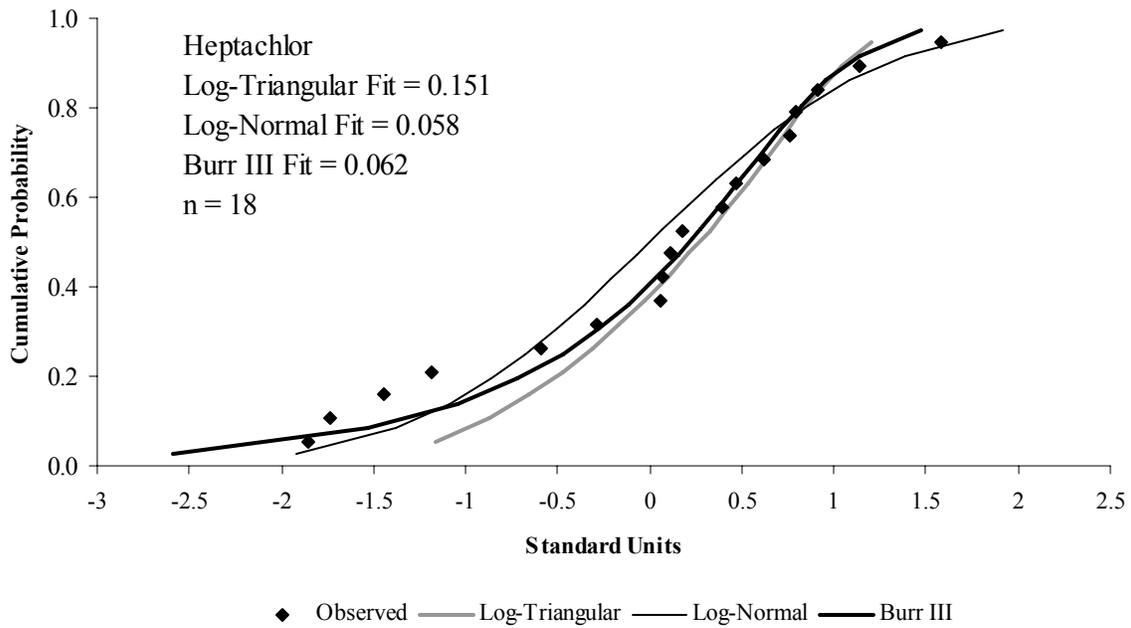
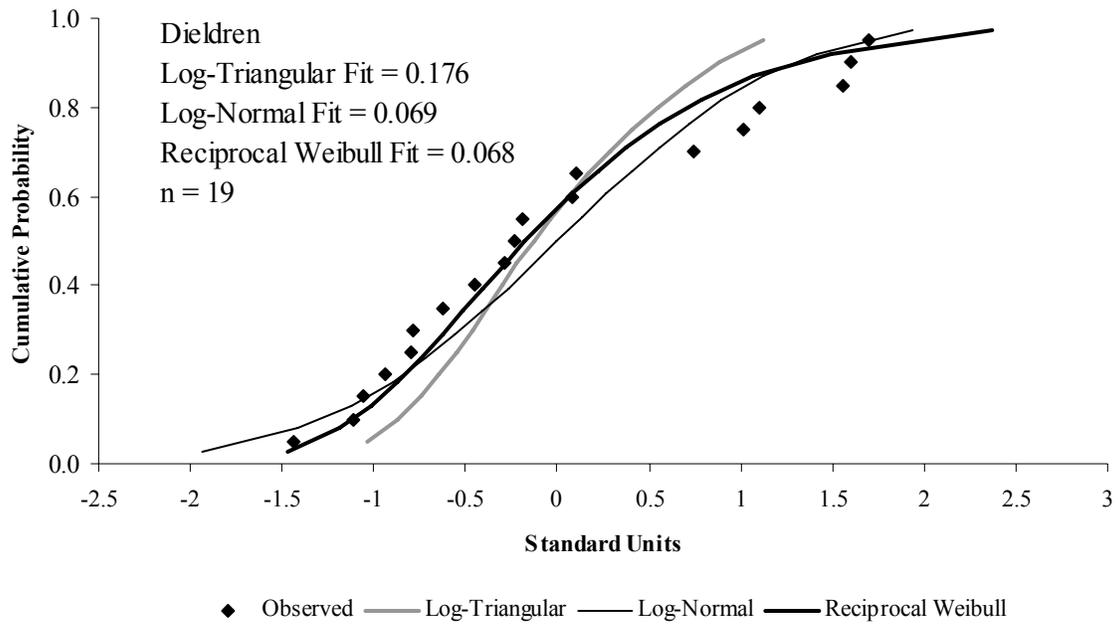


Figure 2.2 (cont.). Comparison of fits of pesticide toxicity data to log-triangular, log-normal and Burr Type III distributions (including Reciprocal Weibull).

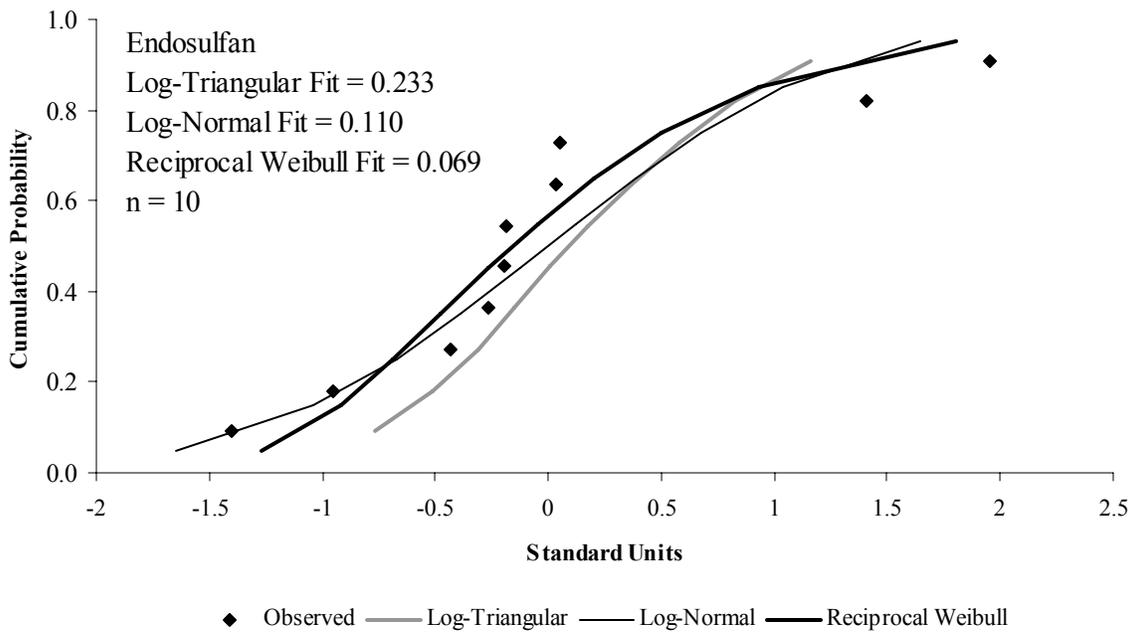
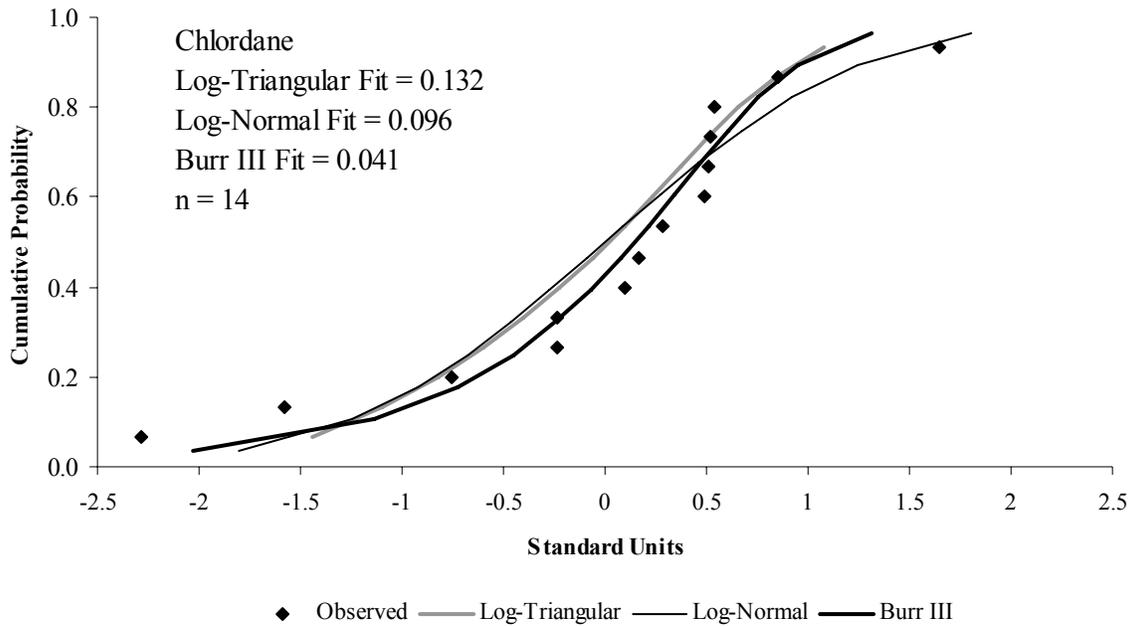


Figure 2.2 (cont.). Comparison of fits of pesticide toxicity data to log-triangular, log-normal and Burr Type III distributions (including Reciprocal Weibull).

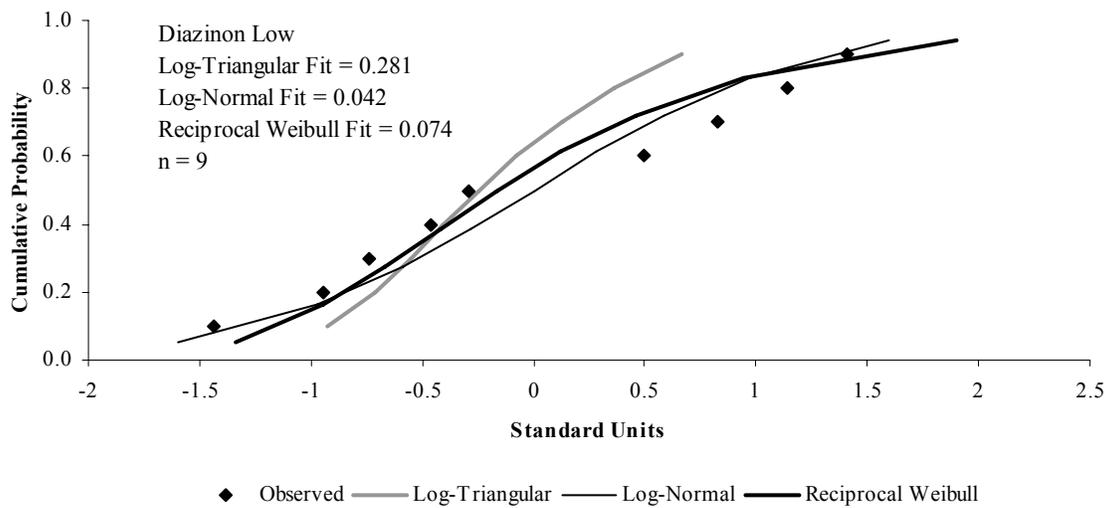
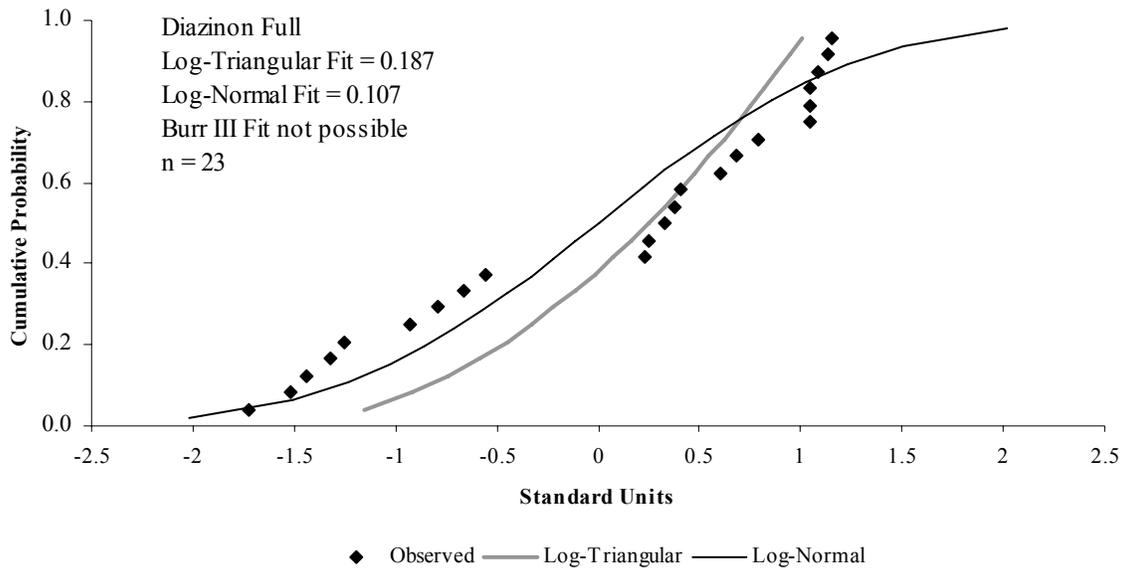


Figure 2.2 (cont.). Comparison of fits of pesticide toxicity data to log-triangular, log-normal and Burr Type III distributions (including Reciprocal Weibull).

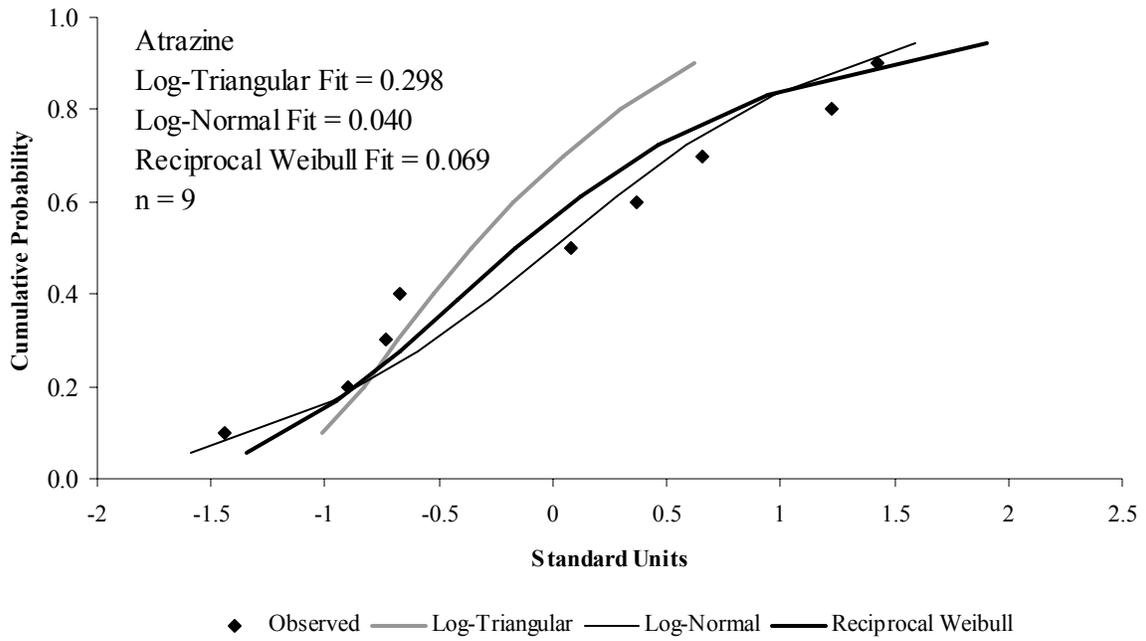


Figure 2.2 (cont.). Comparison of fits of pesticide toxicity data to log-triangular, log-normal and Burr Type III distributions (including Reciprocal Weibull).

Table 2.2 shows the results of curve-fitting for the 12 data sets in Table 2.1. The BurrliOZ v. 1.0.13 program (CSIRO 2001) was used to fit Burr Type III distributions to each data set, except the full diazinon set, which is bimodal (histogram, Fig. 2.1; observed data, Fig. 2.2). Using only the lowest 9 diazinon toxicity values, a reciprocal Weibull distribution (one of the Burr III family) was fit.

Table 2.2. Burr III family distribution fit parameters for data sets in Table 2.1. Parameters c, b and k apply to the Burr III distributions; α and β apply to the reciprocal Weibull distribution.

Pesticide	Distribution	c or α	b or β	k
Chlorpyrifos	Reciprocal Weibull	0.6979	0.3855	-----
DDT	Burr III	0.8576	0.7528	5.077
Toxaphene	Burr III	1.0852	3.6524	2.366
Endrin	Reciprocal Weibull	0.8914	0.7889	-----
Lindane	Burr III	1.552	76.521	0.8204
Aldrin	Reciprocal Weibull	15.908	1.095	-----
Dieldrin	Reciprocal Weibull	6.653	0.7061	-----
Heptachlor	Burr III	2.147	92.256	0.2830
Chlordane	Burr III	3.275	67.499	0.3572
Endosulfan	Reciprocal Weibull	1.743	0.667	-----
Diazinon Full	Could not fit	-----	-----	-----
Diazinon Low	Reciprocal Weibull	1.454	0.8207	-----
Atrazine	Reciprocal Weibull	34375	1.157	-----

Erickson & Stephan (1988) utilized the following formula to measure goodness of fit to a number of different distributions:

$$\frac{\sum (X_R - E(X_R))^2}{\sum_n (X_R - \bar{X}_R)^2} \quad (2.7)$$

where:

- n = number of data in set;
- X_R = observed value of X at rank R ;
- $E(X_R)$ = expected value of X at rank R ;
- \bar{X}_R = mean value of X for all ranks.

and:

$$E(X_R) = \hat{L} + \hat{S} \bullet E(Z_R) \quad (2.8)$$

where:

\hat{L} = location parameter estimate (mean)

\hat{S} = scale parameter estimate (standard deviation)

Using equations 2.7 and 2.8, the data in Table 2.1 were tested for goodness of fit to the log-triangular, log-normal and Burr Type III distributions used in USEPA (1985; 2003d), RIVM (2001), and ANZECC & ARMCANZ (2000), respectively. Results are shown in Table 2.3 and are included in Fig. 2.2.

Table 2.3 Comparison of fit of log-triangular, log-normal, and Burr Type III distributions for data sets from Table 2.1. Lower number (shaded) indicates better fit. Where both are shaded, fits are equally good. Fit is measured as described in Erickson & Stephan (1988); data were log-transformed prior to analysis.

	Log-Triangular	Log-Normal	Burr Type III
Chlorpyrifos	0.120	0.087	0.049
DDT	0.035	0.033	0.021
Toxaphene	0.153	0.261	0.017
Endrin	0.154	0.149 ¹	0.075
Lindane	0.213	0.063	0.047
Aldrin	0.147	0.039	0.057
Dieldrin	0.176	0.069	0.068
Heptachlor	0.151	0.058	0.062
Chlordane	0.132	0.096	0.041
Endosulfan	0.233	0.110	0.069
Diazinon full	0.187	0.107 ¹	No fit
Diazinon low	0.281	0.042	0.074
Atrazine	0.298	0.040	0.069

¹ Normal fit rejected by ETX (Van Vlaardingen *et al.* 2004) and JMP (2004) analyses.

The log-triangular distribution is not the best fit in any case, although it does well with the large DDT data set. Eleven of the data sets may be described by log-normal distributions ($p > 0.05$, Fig. 2.1), while that of endrin and the full, bimodal diazinon set may not. Burr III distributions were fit to 12 of the data sets, but could not be fit to the full diazinon set. For toxaphene, lindane, dieldrin and heptachlor, the log-normal and Burr III fits were equally good (ratio of fit numbers < 1.5). For six of the sets, the Burr Type III distribution fits better than either the log-triangular or log-normal (Table 2.3), while the log-normal distribution is the best fit in two cases. When the log-normal fit is better, it is only slightly better (ratio of fit numbers = 1.5-1.8). On the other hand, there are cases where the Burr III fit is much better (ratio of fit numbers = 1.6-15). Although the goodness-of-fit test used in Table 2.3 indicates that the log-normal distribution is better than the log-triangular for the full diazinon set, Fig. 2.1 indicates that this set cannot be described by a normal distribution ($p < 0.05$). Thus, the Burr Type III result of “no fit” (Table 2.3) is more accurate. For the lower portion of the diazinon set, the Burr Type III distribution is the best fit. For endrin, the only distribution that fits is the Burr Type III.

The Burr family of distributions provides a better fit than the log-triangular distribution in all cases tested, and provides an equivalent or far better fit than the log-normal distribution in most cases. This is expected because the Burr III family of distributions approximates the log-normal and log-triangular distributions (CSIRO 2001). Based on the fit of the SSD alone, the Burr III family of distributions is the best candidate for use in the new methodology for derivation of criteria by the SSD technique. However, a few other factors important in choosing an SSD technique. Species sensitivity distribution methodologies of USEPA (1985; 2003d), RIVM (2001) and ANZECC & ARMCANZ (2000) are discussed and compared further in section 2-3.1.5. This discussion and others in the following sections will be considered and an SSD will be chosen in section 2-3.1.6.

2-3.1.2 Percentile cutoff

To use an SSD method for criteria derivation requires selection of a percentile of the distribution as a cutoff point. This is often interpreted to mean that species lying above this point in the distribution will be protected as long as the concentration of chemical is below the concentration at the selected percentile, but species lying below the percentile would be harmed. Van Straalen & Van Leeuwen (2002) note that it is not correct to interpret the 5th percentile to mean that 5% of species will be harmed (as was argued, for example, by Lillebo *et al.* 1988, regarding the USEPA 1985 methodology). Rather, this approach is one method for derivation of a predicted no-effect concentration, and although the choice of the 5th percentile is purely a pragmatic one, it has been validated by field studies. Solomon *et al.* (2001) note that any percentile may be chosen as long as it can be validated against knowledge and understanding of ecosystem structure and function.

The USEPA rationale for choosing the 5th percentile is simply that criteria values derived using the 10th or 1st percentiles seemed too high and too low, respectively, and since the 5th falls between those, it was selected (Stephan 1985). By the USEPA methodology (USEPA 1985; 2003d) chronic criteria are derived directly from the 5th percentile of maximum acceptable toxicant concentrations (MATC) values. Acute criteria are derived from EC₅₀ or LC₅₀ data. Since 50% effect is not acceptable, the 5th percentile values are divided by a safety factor of 2 to arrive at the final acute criterion value. This figure was based on 219 acute toxicity tests with various chemicals, which showed that the mean concentration that did not cause mortality greater than control was 0.44 times the LC₅₀ (34 FR 97, p 21508-21218). The inverse of 0.44 (2.27) was rounded to 2 for use in EPA methods. Subsequent studies have shown good agreement between USEPA criteria and no-effect concentrations determined in experimental stream studies (USEPA 1991). The Dutch guidelines (RIVM 2001) use the 5th percentile for derivation of environmental limits. Specific reasons for this choice are not given, but the 5th percentile has been validated against field NOECs in studies by Emans *et al.* (1993) and Okkerman *et al.* (1993). The Australia/New Zealand guidelines (ANZECC & ARMCANZ 2000) consider the question more rigorously, but still arrive at the 5th percentile level for the simple reasons that it works well in the Dutch guidelines (RIVM 2001) and it gives criteria that agree with NOEC values from multi-species tests. The reason for not

regularly using a lower percentile is that the uncertainty is very high in the extreme tail of the distribution and the uncertainty can contribute more to derived criteria than the data. However, the Australia/New Zealand guidelines do use the 1st percentile as a default value for high conservation ecosystems, for bioaccumulative substances, and for cases where an important species is not protected at the 5th percentile level. To provide further information to water quality managers in Australia/New Zealand, other percentile levels are also calculated so that criteria are given based on the 1st, 5th, 10th and 20th percentiles.

Other researchers have also found good correlation between criteria derived from the 5th percentile of single-species SSDs and NOECs determined in multi-species tests (Hose & Van Den Brink 2004; Maltby *et al.* 2005; Versteeg *et al.* 1999). On the other hand, Zischke *et al.* (1985) found that a laboratory-derived criterion concentration of pentachlorophenol was not protective of invertebrates and fish in outdoor experimental channels. Maltby *et al.* (2005) determined that concentrations of pesticides derived from the 5th percentile of species sensitivity distributions with 95% confidence was protective of species in freshwater ecosystems, but concentrations derived with 50% confidence were not protective and required application of a safety factor.

The 5th percentile SSD cutoff, which has been validated against multi-species NOECs in many cases, is a level that balances the desire to select a percentile near zero with the need to avoid utilizing the highly uncertain tails of the distributions. The new methodology will use the 5th percentile, but, since a few studies have shown this level to be underprotective, criteria derived from this value will be evaluated against available data from tests with multi-species, ecosystem, sensitive species and threatened or endangered species. If evidence suggests that the 5th percentile will not be protective, criteria may be adjusted downward. The recommended means of making such an adjustment is to use either a lower 95% confidence limit estimate of the 5th percentile (see discussion in section 2-3.1.3), or a median or 95% confidence limit estimate of the 1st percentile.

2-3.1.3 Level of confidence

With SSD methods it is necessary to decide what level of certainty is desired in the resulting concentration. The USEPA approach (USEPA 1985; 2003d) does not provide a means to determine levels of confidence for derived criteria; all are derived as the median estimate of the 5th percentile, meaning that the true value may be greater or less than the estimated value with equal probability. All other SSD methodologies result in a criterion derived from a specified percentile level and a specified level of confidence. Uncertainty in an extrapolated value is due to the risk that the extrapolated value is wrong (Aldenberg & Slob 1993). The distribution around the extrapolated value can be used to determine lower boundaries for the extrapolated value (Aldenberg & Slob 1993; Kooijman 1987; Van Straalen & Denneman 1989; Wagner & Lokke 1991). By evaluating this uncertainty, it is possible to state that the true 5th percentile value falls above (or below) the extrapolated value with, say, a 50%, 90%, 95% or other level of certainty. These distributional confidence limits assume that the uncertainty in the fitted

distribution is the greatest source of uncertainty in the criteria calculation and it will outweigh other sources of uncertainty, such as:

- a) The uncertainty of an LC₅₀ or MATC calculation, which could be expressed as the confidence limits of the reported LC₅₀;
- b) The uncertainty of the reported toxicity test concentrations from the method in which they were determined;
- c) The effect of the conditions of a particular lab set up or batch of organisms, which is indicated by different studies for the same species that report different LC₅₀s.

The differences in species sensitivities usually vary by several orders of magnitude (as seen in Table 2.1) and are likely to overshadow the other calculable sources of uncertainty listed above. Since the distribution models the large species-to-species variation, the uncertainty in the fit of the distribution is likely the best available quantitative measure of uncertainty in the criteria.

While different confidence levels may be calculated, the most statistically robust is the 50%, or median, estimate (ANZECC & ARMCANZ 2000; EVS 1999; Fox 1999). In cases where there is evidence that a median 5th percentile estimate will not be adequately protective, a lower 95% confidence limit estimate of the 5th percentile may be used instead. Variability in the tails of the distributions tends to compound, rather than clarify, uncertainties. Ultimately, the selection of certainty levels is a policy decision, but one that can be informed by understanding the limitations of values derived from distributional tails. The new methodology presents a method for derivation of criteria with multiple levels of certainty so that environmental managers can choose values that best suit their needs.

2-3.1.4 Aggregation of taxa and outliers

As discussed in section 2-3.1.1, one challenge in the use of SSDs is to fit the data to an appropriate distribution prior to extrapolation. One way to achieve a better fit is to break data into groups rather than to pool it all together in one SSD. Data may be grouped according to toxicant mode of action, habitat (e.g., lentic vs. lotic), reproductive strategy or life cycle (Solomon & Takacs 2002). The Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) concluded that it is important to include all species in the criteria derivation procedure. However, it is reasonable, especially in construction of SSDs, to separate species into groups if multimodal distributions are evident. On the other hand, if there is no justifiable difference between apparent groups (e.g., vertebrates and invertebrates, or plants and animals) then the data should be pooled for criteria derivation. Since these data are not expected to be normally distributed, no statistical test for outliers will be included. Most of the distributions discussed in this report are well adept for handling outliers, so critical examination of the data will be the main course of action suggested. Apparent ‘outliers’ should be scrutinized for possible sources of error (typographical errors, units reported incorrectly, inappropriate methods). Inability to fit

the distribution is more likely to be caused by a bimodal group of data in which there is not enough data to actually characterize the two subsets. In this case, non-distributional methods (assessment factors) should be defaulted to, which would be similar to the method for data sets as small as the subset is likely to be. If the distribution is unimodal and cannot be fit with a larger data set it may be reasonable to exclude outliers, provided that there is some rationale for the difference in sensitivity and that the criteria be adjusted if not protective of sensitive species.

2-3.1.5 Comparison of methods

In the previous sections, important specific aspects of SSD methods have been evaluated including the fit of different SSDs, the percentile cutoff, confidence level, and aggregation of data. In this section the methods of three agencies are more broadly reviewed. Also in this section these three techniques are compared by deriving example criteria. Two of them are the widely used SSD techniques of RIVM (2001) and the USEPA (1985; 2003d). Also included is the ANZECC & ARMCANZ (2000) technique, which represents an improvement over RIVM (2001) in that it provides a way to fit distributions to data sets (a technique supported by OECD 1995). Results of running all three methods with all data sets in Table 2.1 are given in Table 2.4. Before that, each method is described, assumptions are stated, and advantages and disadvantages are listed.

2-3.1.5.1 Assumptions common to the three methods

All three methods use a SSD to extrapolate to the 5th percentile. Several assumptions apply to such SSD methods and are given here:

- 1) Surrogate species are good representatives of species of concern;
- 2) Protecting species from direct adverse effects will also protect them from indirect adverse effects;
- 3) Effects that occur on a species in laboratory tests will generally occur on the same species in comparable field situations;
- 4) Extrapolation of the 5th percentile of single-species toxicity values will produce a value that is protective of the all species in an ecosystem;
- 5) Protecting the most sensitive species will protect all species in an ecosystem;
- 6) Surrogate species represent a random sampling of all species in an ecosystem.

2-3.1.5.2 USEPA (1985; 2003d)

EPA methods use a log triangular distribution to calculate a final acute value (FAV), which is also the 5th percentile estimate. To calculate the FAV, the total number of SMAVs and usually the lowest 4 SMAVs are used to calculate the FAV. This calculation can be done by hand or using a spreadsheet and the equations are included in the Phase I report (see Erickson & Stephan 1988 for derivation). Note that the USEPA (1985; 2003d) SSD method is defined in terms of genus mean acute values (GMAVs), but, as discussed in section 2-2.7, species mean acute values (SMAVs) are used in the

new methodology. Thus, for the criteria derived in Table 2.4 for comparison, SMAVs are substituted for GMAVs.

The final chronic value (FCV) may be derived in the same manner if enough chronic data are available, however, the FCV is typically derived by application of an acute-to-chronic ratio (ACR; section 2-3.2.5) to the FAV.

Assumptions specific to this method:

- 1) No species succumbs to infinitesimal, or tolerates infinite, concentrations of toxicant (Erickson & Stephan 1988);
- 2) Data sets represent independent random samples from symmetrical log-triangular distributions;
- 3) Aquatic ecosystems can tolerate some stress and occasional adverse effects, therefore protection of all species at all times and places is not necessary;
- 4) Censored data (data expressed as < or > a value) can be used if not in the lowest 4 values.

Minimum data required: 8

Advantages:

- 1) Fitting to only the four toxicity values nearest the 5th percentile eliminates problems that arise when toxicity data sets do not meet the log-triangular distribution assumption;
- 2) Focuses on sensitive end.

Disadvantages:

- 1) There is no biological basis for selecting a triangular distribution (ANZECC & ARMCANZ 2000);
- 2) Not all of the data are used to fit distribution;
- 3) No associated confidence levels can be calculated;
- 4) Requires the most data of the three methods.

2-3.1.5.3 RIVM (2001) formerly MHSPE (1994)

Environmental risk limits (ERLs; 5th percentile estimates) are derived using the SSD method of Aldenberg & Jaworska (2000). That is, HC_p values (hazardous concentrations affecting p% of species) are calculated based on a log-normal SSD. Equations are included in the Phase I report and a computer program called ETX 2.0 is available for making these calculations (Van Vlaardingen *et al.* 2004; available for free at <http://www.rivm.nl/rvs/overig/risico/methoden/ETX.jsp>). Calculations are usually done with data in the form of NOEC to calculate chronic criteria, but can also be done with acute data.

Assumptions specific to this model:

- 1) No toxicity thresholds exist;

2) Data sets represent independent random samples from symmetrical log-normal distributions.

Minimum data required: 4

Advantages:

- 1) All of the data are used to fit the distribution;
- 2) Associated confidence levels can be calculated;
- 3) Requires the least data of the three methods.

Disadvantages:

- 1) Data sets that do not fit the symmetrical log-normal distributions would be problematic.

2-3.1.5.4 ANZECC & ARMCANZ (2000)

The Australian/New Zealand guidelines use the same method as the Dutch, but with a curve-fitting procedure that overcomes the problem of data that do not fit an assumed distribution. Using the program BurrliOZ v. 1.0.13 (Campbell *et al.* 2000; CSIRO 2001), which is available for free at <http://www.cmis.csiro.au/Envir/burrlioz/>, data are fit to either the Burr III or one of the limiting distributions and the median 5th percentile value is calculated. The equations are shown in Chapter 3 section 3-3.2.1 (Burr 1942, also discussed in section 2-3.1.1). As in the Dutch procedure, NOEC values are usually used to calculate a chronic criterion, but this model could be used with acute data.

Assumptions specific to this model:

- 1) No toxicity thresholds exist;
- 2) Data sets represent independent random samples from symmetrical Burr Type III distributions.

Minimum data required: 5

Advantages:

- 1) All of the data are used to fit the distribution;
- 2) Associated confidence levels can be calculated.

Disadvantages:

- 1) Data sets that do not fit symmetrical Burr Type III distributions would be problematic.

One other advantage of this method over the USEPA (1985; 2003d) and RIVM (2001) methods is that if a Burr III distribution, or one of the limiting distributions, cannot be fit to a data set, then no 5th (or other) percentile value can be calculated. It may still be possible to derive a 5th percentile value if modification of the data set is warranted due to bimodality or the presence of outliers. Using either the USEPA and RIVM

methods, an unwitting user can determine a 5th percentile value whether or not the distributional assumptions are met. Such values would be unreliable.

2-3.1.5.5 Results and discussion of SSD model comparison

Table 2.4 shows the 5th percentile values derived for each of 12 pesticides using each of the three SSD methods. The USEPA method results in one value, the median estimate of the 5th percentile, while each of the others (ANZECC & ARMCANZ 2000; RIVM 2001) result in a median estimate, as well as a lower 95th percentile estimate. The true 5th percentile value has an equal certainty of falling above or below the median estimates, but has a 95% certainty of falling above the lower 95th percentile estimate. Other levels of certainty may be calculated with the ANZECC & ARMCANZ (2000) method as well, but these calculations become less and less reliable in the extreme tails of the distribution.

Many of the 5th percentile values derived by the various methods are similar. For example, the median 5th percentile values derived for DDT, toxaphene, aldrin, dieldrin, heptachlor, diazinon (low), and atrazine are within a factor of 2 by all three methods. The endrin median 5th percentile value is 2.5-2.75 times lower by the RIVM (2001) method compared to the other two, but that data set violated the assumption of log-normality, thus the RIVM (2001) method is not a good choice for the endrin data. The chlorpyrifos median 5th percentile value determined by the RIVM (2001) method is a factor of 4.6-6.6 lower than those obtained by the other methods. Although the chlorpyrifos data fit a normal distribution, the box plot in Fig. 2.1A reveals that the distribution is right-skewed, which would cause the 5th percentile value to fall at a lower value than if the distribution were not skewed. The 5th percentile values obtained by USEPA (1985; 2003d) and ANZECC & ARMCANZ (2000) for endrin are very similar, in spite of the skew of the data. For lindane, the median 5th percentile value obtained from the USEPA method is less than half that obtained by the other two methods. This is likely due to the presence of a low outlier in the lindane set (see Fig. 2.2; outliers were left in for this analysis since they are part of the final USEPA criteria sets). As discussed in the Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) and GLEC (2003), low outliers tend to lower the 5th percentile values derived by the USEPA method. The same phenomenon occurs for chlordane, where the USEPA value is less than half that of the other two. For endosulfan, the ANZECC & ARMCANZ (2000) method produces a median value that is more than twice the values obtained by the other two methods. In this case, Figure 2.1 indicates a high outlier, which is handled better by the Burr III family of distributions than either the log-triangular or log-normal (Table 2.3).

Table 2.4 Results of analyzing pesticide data sets (Table 2.1) with SSD methods of USEPA (1985), the Netherlands (RIVM 2001), and Australia/New Zealand (ANZECC & ARMCANZ 2000); all values represent 5th percentile values (not criteria) expressed in µg/L; A/NZ = Australia/New Zealand; NC = not calculable.

Pesticide	USEPA (median)	RIVM (median)	RIVM (95th)	A/NZ (median)	A/NZ (95th)	Lowest Value in Data Set¹
Chlorpyrifos	0.033	0.005	0.0003	0.023	0.018	0.035
DDT	0.845	0.659	0.322	0.97	0.61	0.36
Toxaphene	1.21	1.15	0.590	1.54	1.04	0.8
Endrin	0.20	0.081	0.027	0.22	0.15	0.15
Lindane	2.65	6.82	2.77	7.4	2.18	10
Aldrin	3.76	3.45	1.54	4.59	3.52	4
Dieldrin	2.67	1.62	0.415	3.10	2.22	2.5
Heptachlor	0.768	1.22	0.322	0.67	0.07	0.9
Chlordane	2.10	5.65	2.11	5.21	1.64	3
Endosulfan	0.183	0.188	0.017	0.44	0.24	0.34
Diazinon All	0.449	0.460	0.043	NC	NC	0.3773
Diazinon Low	0.260	0.251	0.036	0.41	0.25	0.3773
Atrazine	2,514	2,276	572	3,233	2,477	3,000

¹From Table 2.1

An important consideration in assessing the various methods is whether or not they will produce protective criteria. The acute criterion will be derived by dividing the 5th percentile value by 2, so it is of interest to compare the lowest value in each data set with the resulting criterion. The last column of Table 2.4 shows the lowest value from each data set. For most of the pesticides, criteria derived from any of the median 5th percentile values would be lower than the lowest value. However, for DDT the median values derived by the USEPA (1985; 2003d) and ANZECC & ARMCANZ (2000) methods would result in criteria of 0.42 and 0.48 µg/L, respectively, which are higher than the lowest value of 0.36 µg/L. In such a case, the Australia/New Zealand approach provides the option of using the 95th percentile value to produce a number below the lowest value. By the USEPA method, an additional safety factor would have to be applied to derive a protective criterion. The new methodology includes a step of checking derived criteria against available data and adjusting the criteria if they do not appear to be protective (section 2-5.0).

2-3.1.6 SSDs in the new methodology

The ANZECC & ARMCANZ (2000) methodology, utilizing the Burr Type III distributions with the BurrliOZ program offers the best combination of best fit, data requirements, appropriate distributional assumptions, and flexibility in choosing protection and confidence levels as discussed in these last several sections. In addition, all of the distributions historically used for SSD analysis are considered by using the Burr III family: the Burr Type III distribution approximates the log-triangular and log-normal

distributions, and the log-logistic distribution is a special case of the Burr III distribution (CSIRO 2001).

While the ANZECC & ARMCANZ (2000) methodology does have advantages, it is also noteworthy that all of the methods currently in use appear to derive protective criteria. In the Netherlands, the log-normal distribution was selected over a log-log distribution (Aldenberg & Slob 1993) because, although the distributions are not all that different, and results obtained are not different, the normal distribution provides powerful statistical tools (RIVM 2001). Likewise, the OECD (1995) concludes that the log-normal, log-logistic and triangular distribution methods give very similar results. Here too it was found that, with some exceptions, median 5th percentile values obtained by the ANZECC & ARMCANZ (2000) method are comparable to those obtained by the USEPA (1985; 2003d) methodology.

The new methodology uses the ANZECC & ARMCANZ (2000) SSD methodology utilizing the BurriOZ program to calculate 5th percentile values with 50% and 95% confidence that the true 5th percentile value lies above the derived value. The acute criterion derived by the SSD methodology is equal to the median 5th percentile value divided by 2. The safety factor of 2 is applied because the SSD is constructed with toxicity values that indicate a 50% effect level (section 2-3.1.2). A 5th percentile value derived from chronic data is the chronic criterion without further adjustment.

2-3.2 AF methodology

When fewer than five data from an appropriate assortment of taxa are available, the SSD method cannot be used for criteria derivation. In such cases, an assessment factor (AF) method must be used. As discussed in Phase I (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009), assessment factors are recognized as a conservative approach for dealing with uncertainty in assessing risks posed by chemicals (Chapman *et al.* 1998). Assessment factors (also called safety factors, application factors, extrapolation factors) are usually applied to account for a wide range of possible effects and situations for which no data exist, including: lack of tests with relevant species; persistence or bioaccumulative potential of substances; genotoxic potential; laboratory to field extrapolation; acute-to-chronic extrapolation; variations in mesocosm types for multispecies tests; absence of most sensitive species in multispecies tests; mixture effects; experimental variability, and; lack of data (Phase I, TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009). Further factors may be applied in some cases based on the professional judgment of the risk assessor. In all cases, the more toxicity data that are available for species of different trophic levels, different taxonomic groups, and different lifestyles, the smaller the applied factor.

Table 2.5 summarizes assessment factors used by the methodologies reviewed by TenBrook & Tjeerdema (2006, TenBrook *et al.* 2009). They range from 1-1000 or more, and, with the exception of ACRs, are primarily based on the premise that 10 is a widely used factor in toxicology and should be applied for each step of an extrapolation from data at hand to real-world application.

Table 2.5 Assessment factors used in existing methodologies.

Methodology	Range of factors	Applied to	Reference
Australia/ New Zealand	Acute: NA ¹ Chronic: 10-1000 Default ACR: 10+	Acute: NA Chronic: NOEC (single- or multispecies)	ANZECC & ARMCANZ (2000)
California (draft)	Acute: NA Chronic: 10	Acute: NA Chronic: LOEC	Lillebo <i>et al.</i> (1988)
Canada	Acute: NA Chronic: 10-100+ Default ACR: 10	Acute: NA Chronic: LC/EC ₅₀ , LOEC	CCME (1999)
European Union/ Denmark	Acute: NA Chronic: 1-1000 Default ACR: 10	Acute: NA Chronic: LC/EC ₅₀ , NOEC, QSAR estimates	OECD (1995); Samsoe-Petersen & Pedersen (1995); ECB (2003)
France	Acute: NA Chronic: 1-1000	Acute: NA Chronic: LC/EC ₅₀ , NOEC	Lepper (2000)
Germany	Acute: NA Chronic: 10-1000 Default ACR: 10	Acute: NA Chronic: LC/EC ₅₀ , NOEC	Irmer <i>et al.</i> (1995)
Great Lakes	Acute: 4.3-21.9 Chronic: NA Default ACR: 2 for FACR ² 18 for SACR ³	Acute: LC/EC ₅₀ Chronic: NA	USEPA (2003d)
North Carolina	Acute: 3 Chronic: 1 Default ACR: 100 for t _{1/2} > 96 h 20 for t _{1/2} < 96 h	Acute: LC ₅₀ Chronic: MATC	North Carolina DENR (2003)
South Africa	Acute: 1-100 Chronic: 1-1000	Acute: LC ₅₀ Chronic: Final Acute Value	Roux <i>et al.</i> (1996)
Spain	Acute: NA Chronic: 1-100+	Acute: NA Chronic: LC/EC ₅₀ , NOEC	Lepper (2000)
The Netherlands	Acute: NA Chronic: 1-1000	Acute: NA Chronic: LC/EC ₅₀ , NOEC	RIVM (2001)
United Kingdom	Acute: 2-10 Chronic: 1-100	Acute: LC/EC ₅₀ Chronic: NOEC	Zabel & Cole (1999)
USEPA	Acute: 2 Chronic: NA Default ACR: 2	Acute: Final Acute Value Chronic: NA	USEPA (1985)
USEPA	Acute: 100-1000 Chronic: 10 Default ACR: 10	Acute: LC/EC ₅₀ Chronic: MATC	Nabholz (1991)

¹ Not applicable

² Final acute-to-chronic ratio

³ Secondary acute-to-chronic ratio

2-3.2.1 Appropriate use of assessment factors

An important point to keep in mind in using assessment factors is that application of empirically based factors to toxicity data does not quantify uncertainty, but does reduce the probability of underestimating risk. At the same time, the use of AFs also greatly increases the possibility of overestimating risk (Chapman *et al.* 1998). It is worth restating and evaluating some of the specific points by Chapman *et al.* (1998) regarding the use of assessment factors, keeping in mind that each of the points needs to be evaluated in the context of water quality criteria derivation, as opposed to ecological risk assessment. Ecological risk assessment seeks to estimate risk based on a specific set of exposure and effects data, usually for a specific site. Numeric water quality criteria are derived considering only effects data along with a few exposure factors that directly affect toxicity. Criteria must be protective of aquatic life, and therefore must err on the side of conservatism when data are lacking. Criteria may be site-specific, but more often must be valid for a range of sites. When data are lacking, criteria will likely represent an over-estimation of risk. More data are required for extrapolated protective values to approach true values. With this in mind, consider each of the points raised by Chapman *et al.* (1998).

1) Data supercede extrapolation; that is, if data are available, they should be used.

This point simply reinforces the idea that more data will result in better estimates of risk, and therefore, better estimates of appropriately protective criteria.

2) Extrapolation requires context; use of assessment factors should be based on existing scientific knowledge.

This statement is true, but somewhat contradictory. In fact, assessment factors are used to fill gaps in scientific knowledge. With the exception of measured acute-to-chronic ratios (ACRs; discussed in section 2-3.2.5), many existing criteria derivation methodologies use standardized factors of 10, 20 and 100, despite lack of supporting data (Chapman *et al.* 1998). Further, assessment factors are often based on policy rather than empirical science. One methodology that has made use of existing scientific knowledge is that of the Great Lakes (USEPA 2003d), which uses empirically derived factors and default ACRs. Details of the Great Lakes factors are discussed further in section 2-3.2.4. This approach is adapted for the new methodology.

3) Extrapolation is not fact; estimates of effect levels obtained using assessment factors should only be used as screening values, not as threshold values (criteria).

All criteria are extrapolated values, and while those obtained by application of large factors to small data sets have a high level of conservatism and uncertainty, it is a policy decision whether or not to use them as threshold values. The new methodology includes an AF method to be used for data sets that are inadequate for SSD analysis. Less conservative, more certain numbers may be derived if more data become available.

4) Extrapolation is uncertain; assessment factors should encompass a range rather than being a single value.

With the exception of the German methodology (Irmer *et al.* 1995), all existing methodologies utilize ranges of factors, and in all cases factors get smaller as data sets get larger. The AF procedure in the new methodology includes a range of factors.

5) All substances are not the same; assessment factors should be scaled relative to different substances, potential exposures and nature of effects.

As noted in the response to point 4, ranges of factors are used by nearly all existing methodologies. Among the reasons for using larger factors are lack of data, persistence, bioaccumulative potential, mixture toxicity, and potential for genotoxic effects. In effect, assessing factors for each of these variables achieves the scaling suggested by Chapman *et al.* (1998). Another example of how this is done is that the state of North Carolina utilizes different default ACRs depending on the half-life of the chemical of concern (North Carolina DENR 2003, shown in Table 2.5). The new methodology includes other means of addressing bioaccumulative potential, bioavailability, and mixture toxicity and therefore does not incorporate these elements into assessment factors. Genotoxic effects are of concern in human health risk assessment, but not for protection of aquatic life (see section 2-2.1.3 regarding non-traditional endpoints), and thus are not incorporated into assessment factors either.

6) Unnecessary overprotection is not useful; assessment factors for individual extrapolation steps should not exceed 10, and may be much lower.

This statement regarding overprotection is true, but should not be addressed by limiting the size of assessment factors. Chapman *et al.* (1998) found ACRs ranging from 1-20,000. Thus, for acute-to-chronic extrapolation, factors should definitely not be limited to 10. Also, empirically derived factors, such as those in the Great Lakes methodology (USEPA 2003d), should not be limited to 10. The best way to minimize overprotection is to expand available acute and chronic data sets.

2-3.2.2 Toxicity values

One of the choices to be made in using the AF method is what kind of toxicity data to use. For the new methodology, separate acute and chronic criteria are derived. Therefore, acute criteria are derived from LC₅₀ or EC₅₀ data. The situation is not so simple for chronic criteria. As shown in Table 2.5 most methodologies from the United States (Nabholz 1991; North Carolina DENR 2003; USEPA 1985; 2003d) use MATC values, while most other methodologies from around the world use NOEC values for derivation of chronic criteria; two utilize LOEC values (CCME 1999; Lillebo *et al.* 1988). The Netherlands (RIVM 2001) and the EU risk assessment technical guidance document (ECB 2003) accept EC₁₀ values to represent NOECs. As discussed in section 2-2.1.2 hypothesis test results can be used in criteria derivation, as long as results are evaluated with respect to test design, minimum significant difference (type II error), and

effect levels at the NOEC and LOEC. As discussed in section 2-2.1.2, the MATC is the hypothesis test value used in the new methodology to derive the chronic criteria.

2-3.2.3 Magnitude of factors

As shown in Table 2.5, the magnitude of assessment factors ranges widely among existing methodologies. Aside from measured ACRs, little to no justification is given for the magnitude of most assessment factors (Phase I, TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009). The exception is the Great Lakes methodology (USEPA 2003d), which utilizes final acute value factors and default acute-to-chronic ratios that are empirically based and theoretically supported (Host *et al.* 1995). These factors are discussed further in the following sections and the methods of Host *et al.* (1995) are used to derive factors for the new methodology.

2-3.2.4 Acute factors

Most AF methods utilize standard factors that get larger and larger as more extrapolation steps are required. Such factors are widely used despite having little scientific basis (Chapman *et al.* 1998). The acute toxicity factors used in the Great Lakes methodology are based on work by Host *et al.* (1995) in which they described both empirical and theoretical methods for derivation of factors using data sets for all kinds of chemicals and based on log-triangular data distributions. Their methods were adapted to derive acute factors for the new methodology. The major differences in the approach used here are the use of only pesticide data (vs. all kinds of contaminants), and the use of Burr III distributions (vs. log-triangular distributions). This method requires that full data sets be available for several compounds, thus at this point it is only possible to derive acute (vs. chronic) factors by this method.

The goal of this process is to estimate a 5th percentile value when fewer than 5 data are available. The magnitude of the assessment factors needs to be set to achieve this. The factors are used as divisors for the lowest value in data sets that contain only 1-4 toxicity values.

As per Host *et al.* (1995), the following procedure was applied to each individual pesticide data set (from Table 2.1):

- 1) Ninety-nine (99) subsets of 5 toxicity values were randomly selected with the restriction that the first value had to be for an invertebrate from the genus *Daphnia*, *Ceriodaphnia*, or *Simocephalus*. These organisms were required because they are known to usually be among the most sensitive and because data for one these species should be available based on EPA pesticide registration requirements. Each successive sample had to fulfill a different requirement in the SSD minimum data set (section 2-2.6). The selection of which family to use for the second and subsequent toxicity values in each subset was made randomly;

- 2) For each subset of 5 toxicity values, subsets of 1-4 toxicity values were also created, resulting in 99 subsets of 1-5 toxicity values;
- 3) The lowest acute value in each subset of size 1-5 toxicity values was used as the numerator for calculating the assessment factor;
- 4) Each of the 99 5-sample subsets was used to generate 5th percentile values using BurrliOZ v. 1.0.13 (Campbell *et al.* 2000; CSIRO 2001);
- 5) The geometric mean of the 99 5th percentile values was used as the denominator for calculating the assessment factor;
- 6) This procedure yielded 99 factors for each subset size;
- 7) The 95th percentile of the 99 factors was determined for each subset size.

This procedure was followed for chlorpyrifos, DDT, toxaphene, endrin, lindane, aldrin, dieldrin, heptachlor, chlordane, and endosulfan. Atrazine was not included since the data set is from a draft document. Diazinon was not included because the above procedure could not be applied to the data set. The bimodality of the full diazinon set resulted in many subsets that could not be fit to a Burr Type III distribution. The lower portion of the diazinon data did not include representatives from all five families required for the minimum SSD data set.

In accordance with USEPA (2003d), 95th percentile factors (from step 7, above) for all pesticides were compiled and the median of those factors for each subset size was selected to be the summary assessment factor (i.e., a factor to apply to all pesticides; Table 2.6). The summary factors for each sample size are shown in Table 2.7 along with the estimated 5th percentile toxicity values obtained for each pesticide using the summary factors together with the geometric mean lowest value for each subsample size. Diazinon was not used to derive factors, but the diazinon data set was used to estimate 5th percentile toxicity values using the factors. In all cases, the 5-sample factor produced an estimated 5th percentile value that is comparable to, or below, the median 5th percentile value determined from applying the SSD procedure to the full data set. The 4-, 3-, and 2-sample factors all produced estimated 5th percentile values below the value generated by the SSD method. Most of the 1-sample factors produced similar results. The exceptions are endrin and endosulfan for which the 1-sample factor overestimated the 5th percentile value by 8-10-fold. This occurred because, for those two pesticides, the family Daphniidae was the most tolerant and the 1-sample subsets had very high toxicity values. To ensure that criteria derived by the assessment factor method are protective, even when based on a single toxicity value, an additional factor of 10 should be assessed. As shown in Table 2.6, this brings the endrin and endosulfan 1-sample values very near the SSD median value. The additional factor of 10 will lead to very conservative criteria in cases where Daphnids are among the most sensitive species, but such conservatism is reasonable when relying on a single datum to make predictions for an ecosystem.

Table 2.6 Compilation of 95th percentile of factors for subsets of 1-5 samples; the median values in the last row are the summary factors for each sample size.

Subset size	5	4	3	2	1
Chlorpyrifos	5.56	11.81	52.95	52.95	52.95
DDT	4.08	4.08	4.08	4.08	4.08
Toxaphene	3.86	5.93	8.01	8.01	8.01
Endrin	3.65	5.1	8.49	148.49	1633.43
Lindane	4.03	4.12	4.37	18.83	61.48
Aldrin	5.13	5.13	5.32	5.32	5.32
Dieldrin	3.24	3.24	7.64	63.86	74.95
Heptachlor	3.48	16.3	32.71	97.02	97.02
Chlordane	2.23	3.86	5.94	8.62	8.62
Endosulfan	2.02	13.66	22.04	522.68	1550.24
Median	3.8	5.1	7.8	36	57

Table 2.7 Median 5th percentile toxicity value estimates for sample sizes of 1-5 acute toxicity values using summary pesticide assessment factors. All values are in µg/L.

Sample size	5	4	3	2	1	1	SSD 5 th percentile (median)	Lowest value in data set ¹
Factor	3.8	5.1	7.8	36	57	57 x 10		
Chlorpyrifos	0.01	0.012	0.009	0.003	0.002	0.0002	0.022	0.035
DDT	0.25	0.20	0.13	0.03	0.02	0.002	0.97	0.36
Toxaphene	0.68	0.61	0.56	0.18	0.22	0.022	1.54	0.8
Endrin	0.10	0.09	0.08	0.03	1.80	0.18	0.22	0.15
Lindane	4.0	3.4	2.9	1.1	9.2	0.92	7.4	10
Aldrin	1.8	1.5	1.3	0.40	0.48	0.048	4.59	4
Dieldrin	1.2	1.0	0.82	0.40	3.5	0.35	3.10	2.5
Heptachlor	0.36	0.36	0.44	0.24	1.1	0.11	0.67	0.9
Chlordane	2.2	2.1	2.0	0.63	1.0	0.10	5.21	3
Endosulfan	0.09	0.09	0.09	0.06	4.6	0.46	0.44	0.34
Diazinon	0.22	0.22	0.22	0.22	0.22	0.022	0.20	0.3373

¹From Table 2.1

To derive an acute criterion by this method, the lowest value from an acceptable data set is divided by the appropriate factor from Table 2.6 (depending on sample size). The resulting value represents an estimate of the median 5th percentile value and is divided by 2 to determine the acute criterion. In all cases, for all pesticides shown in Table 2.6, criteria derived using the proposed factors are below the lowest values in each data set and would be expected to be protective.

2-3.2.5 Acute-to-chronic ratios (ACRs)

If at least five chronic data are available from five different families, the SSD method should be used to derive chronic criteria. However, when chronic data are

lacking, the use of ACRs is necessary to extrapolate from acute to chronic toxicity. The ACR is calculated by dividing an acute LC/EC₅₀ value by a chronic value (e.g., MATC) derived from the same test, or from tests conducted by the same laboratory under identical conditions (USEPA 1985; 2003d). There are three basic approaches to deriving ACRs: 1) derive chemical-specific, multispecies ACRs using acute and chronic values derived from the same tests (ANZECC & ARMCANZ 2000; USEPA 1985; 2003d); 2) derive chemical-specific, multispecies ACRs using whatever chronic data are available, combined with one or more default ACR values (USEPA 2003d), and; 3) use default ACR values for all chemicals. As these approaches represent a stepwise procedure depending on available data, all are appropriate for inclusion in the new methodology.

2-3.2.5.1 Single-chemical, multispecies ACR based on measured data

The first approach is used in both the USEPA methodologies (USEPA 1985; 2003d) and in the Australia/New Zealand methodology (ANZECC & ARMCANZ 2000). However, only the USEPA methodologies give clear guidance for the procedure. The Great Lakes guidance, which is updated from the 1985 version, is presented here. The procedure requires acute and chronic data from organisms in at least three different families including a fish, an invertebrate, and at least one other acutely sensitive species. If there are not enough freshwater data to fulfill the ACR data requirements, then saltwater species may be used because freshwater and saltwater ACRs have been shown to be comparable (USEPA 1985) and this approach has been accepted in numerous criteria derivations (Siepmann & Finlayson 2000; USEPA 1980a; b; c; d; 2003a; 2005a). For each chronic value (MATC) having at least one corresponding appropriate acute value, an ACR is calculated by dividing the geometric mean of all acceptable flow-through acute tests by the chronic value. Static tests are acceptable for midges, daphnids and other zooplankton. For fish, the acute test(s) should be conducted with juvenile or younger fish. For all species, the acute test(s) should be part of the same study and use the same dilution water as the chronic test. If acute tests were not conducted as part of the same study, but were conducted as part of a different study in the same laboratory and dilution water, then they may be used. If no such acute tests are available, results of acute tests conducted in the same dilution water in a different laboratory may be used. If no such acute tests are available, an ACR is not calculated by this method.

The species mean acute-to-chronic ratio (SMACR) is calculated for each species as the geometric mean of all ACRs available for that species. For some materials, the ACR seems to be the same for all species, but for other materials the ratio seems to increase or decrease as the SMAV increases. Thus the multispecies ACR can be obtained in one of three ways, depending on the data available:

1. If the SMACR seems to increase or decrease as the SMAVs increase, the ACR is calculated as the geometric mean of the ACRs for species whose SMAVs are close to the acute 5th percentile value.

The USEPA methodologies (USEPA 1985; 2003d) do not define what is meant by “SMAVs close to” the 5th percentile value. A definition for use in the new

methodology can be developed based on the second approach to derivation of interspecies ACRs (item 2 below), which uses the geometric mean of SMACRs as long as they are within a factor of 10 of each other. Thus, it is reasonable to define species with “SMAVs close to” the 5th percentile as those whose SMACRs are within a factor of 10 of the SMACR of the species whose SMAV is nearest the 5th percentile value.

2. If no major trend is apparent and the ACRs for all species are within a factor of 10, the ACR is calculated as the geometric mean of all of the SMACRs.

3. If the most appropriate SMACRs are less than 2.0, and especially if they are less than 1.0, acclimation has probably occurred during the chronic test. In this situation, the final ACR should be assumed to be 2.0, so that the chronic criterion is equal to the acute criterion.

If the available SMACRs do not fit one of these cases, use the procedure described in section 2-3.2.5.2 to derive an ACR based partially on measured values and partially on assumed values.

2-3.2.5.2 Single-chemical, multispecies ACR based on measured and assumed values

If not enough data are available for calculation of an ACR by the preceding procedure, then an ACR is derived by using any available measured ACRs plus enough assumed ACRs to give a total of 3 ACRs (USEPA 2003d). For example, if no measured ACRs are available, then three assumed ACRs are used. If two measured values are available, then just one assumed value is used. The magnitude of the assumed, or default, ACR is discussed in the following section.

2-3.2.5.3 Default ACRs

In the Phase I report TenBrook & Tjeerdema (2006; TenBrook *et al.* 2009) concluded that there is no evidence that default ACR values are appropriate for pesticides in general. Nonetheless, adequate chronic data are often not available and some means of estimating an ACR is needed. The Great Lakes guidance uses a default ACR of 18 (USEPA 2003d), which represents the 80th percentile value of all available ACRs from USEPA criteria documents (Host *et al.* 1995). This seems a reasonable approach because it is based on ACRs that have been derived from carefully reviewed studies. Some of the very high ACRs reported in the literature have been rejected by USEPA upon such review (e.g., diazinon ACRs determined by Kenaga 1982). For the new methodology the default ACR used in the Great Lakes guidance was recalculated to include only pesticide data from Host *et al.* (1995), as well as the ACR in the California Department of Fish and Game diazinon criteria document (Siepmann & Finlayson 2000) and a new chlorpyrifos value calculated by the new methodology. Table 2.8 shows this calculation. Based on this data set, a default ACR of 12.4 was calculated for the new methodology, with the following caveats: 1) if data sets collected according to this methodology lead to different ACR values, those values may be substituted into this table and the default ACR recalculated; 2) if previously calculated ACRs are shown to be invalid based on data sets

collected according to this methodology, then those values should be removed and the default ACR recalculated; and 3) if additional pesticide ACR values become available, the default ACR should be recalculated.

The chronic criterion is calculated by dividing the acute 5th percentile value (derived by SSD method or estimated by AF method) by the ACR (derived by one of the three methods in sections 2-3.2.5.1 through 2-3.2.5.3):

$$\text{Chronic Criterion} = 5^{\text{th}} \text{ Percentile Value} \div \text{ACR} \quad (2.9)$$

Table 2.8 Calculation of default acute-to-chronic ratio (ACR).

Chemical	ACR
Chlordane ¹	14
Chlorpyrifos ²	2.2
Diazinon ³	3.0
Dieldrin ¹	8.5
Endosulfan ¹	3.9
Endrin ¹	4.0
Lindane ¹	25
Parathion ¹	10
80 th percentile	12.4

¹ Taken from Host *et al.* (1995); originally from USEPA criteria documents.

² Derived in Chapter 4 of this report .

³ Siepmann & Finlayson (2000).

2-3.3 Averaging periods

Criteria derived according to either the SSD or AF methods are stated in terms of how much of a chemical may be in the water without causing harm (i.e. in terms of magnitude), but without consideration of for how long (duration) or how often (frequency) that level may be exceeded without harm. Section 2-3.4 addresses the frequency component. This section explores the question of duration.

Derivation of separate acute and chronic criteria, as is done in the new methodology, provides a duration component, but criteria derived from studies conducted under constant exposure scenarios do not account for the possibility of pulsed, or otherwise uneven, exposures. Such exposures are common in the Sacramento and San Joaquin River basins (Phase I, TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) and criteria need to reflect that. Time-to-event models could potentially provide a way to express criteria for any given exposure duration, but as discussed in section 2-2.1.6 such models are not currently feasible for use in criteria derivation. The best readily usable approach for determination of an appropriate duration component is to consult the literature. This is the approach used by the USEPA (1985; 2003d) to set both acute and chronic averaging periods. The averaging period is the period of time over which the

receiving water concentration is averaged for comparison with criteria concentrations (USEPA 1994). There are two aspects to consider in setting an averaging period. First is to set the period long enough such that toxicity might occur due to an exceedance, and second is to set the period short enough that the effects of concentration fluctuations on the average concentration are minimized. For example, the USEPA (1985; 2003d) sets the acute averaging period at 1-h based primarily on the fact that ammonia exerts its effects in that time frame, but also because if the period were longer, peak concentrations would be masked in the averaging process. Similarly, 4-7 days has been shown to be long enough to observe the equivalent of chronic toxicity (USEPA 2002), but short enough to minimize the effect of concentrations fluctuations (USEPA 1991). As discussed in Phase I (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009), the 4-day averaging period used in the USEPA methodologies is reasonable and is used in the new methodology. Since the 1-h average is based on ammonia toxicity and may not be appropriate for pesticide criteria. This is explored in the following section.

2-3.3.1 Acute averaging period

To establish an appropriate acute averaging period for the new methodology, a literature review was conducted to try to determine the time-course of acute pesticide toxicity. While many studies report pesticide effects after very short toxicant exposures, only a handful consider environmentally relevant concentrations. Often researchers expose organisms at concentrations many-fold higher than 96-h LC₅₀ values determined in continuous exposure tests, or at concentrations higher than would ever be expected in the environment (e.g., Barry *et al.* 1995a; Barry *et al.* 1995b; Jarvinen *et al.* 1988; Naddy *et al.* 2000; Peterson *et al.* 2001).

Among studies that consider environmentally relevant concentrations, Cold & Forbes (2004) report adverse effects on survival and reproduction in *Gammarus pulex* after 1-h pulses of esfenvalerate as low as 0.05 µg/L, with effects observed for as long as 2 weeks after the exposure. Forbes & Cold (2005) likewise report effects on larval survival and development rates of the midge *Chironomus riparius* with esfenvalerate exposures as brief as 1-h. Heckman and Friberg (2005) found similar results with pulsed exposures of streams to λ-cyhalothrin. A 1-h pulse application of lambda-cyhalothrin at ≥ 0.1 µg/L caused a nearly instantaneous increase in drift of stream macroinvertebrates, with increases occurring within 2 h at concentrations ≥ 0.01 µg/L (Lauridsen & Friberg 2005). In that same study, *Gammarus pulex* drift increased within 3 h after application at 0.001 µg/L. Holdway *et al.* (1994) reported significant 96-h mortality to larval rainbowfish (*Melanotaenia fluviatilis*) with 1-h pulse exposures to esfenvalerate at 0.32 µg/L. Significant mortality occurred in 2 d to *Daphnia magna* exposed to a 24-h pulse of fenvalerate at 3.2 µg/L (Reynaldi & Liess 2005). For the organophosphates pirimiphos-methyl and temephos Brown *et al.* (2002) reported 24-h LC₅₀ values from 1-h pulse exposures of rainbowfish (*Melanotaenia duboulayi*) that were lower than estimated environmental concentrations. Schulz & Liess (2001) report reductions in emerged individuals of the insect *Limnephilus lunatus* at 154 d, as well as reduction in dry weight and reduction in biomass at 240 d after 1-h pulse exposures to fenvalerate. Sublethal responses in Schulz & Liess (2001) were observed at 0.001 µg/L and lethal effects at 0.1

µg/L. In tests of 1-h vs. 10-h equivalent doses of fenvalerate (measured in µg h), Schulz & Liess (2000) reported long-term effects on emergence success and dry weight of caddisfly larvae, with significantly stronger effects from the stronger 1-h pulses vs. the lower level 10-h pulses. Chronic lethal and sublethal effects were observed in *Daphnia magna* after a 24-h pulse exposure of fenvalerate, with complete mortality occurring after 10 d at 3.2 µg/L (Reynaldi & Liess 2005).

Another type of data that is useful in this analysis is effects data from intermediate time points in standard toxicity tests. This type of data is regularly collected, but rarely reported. Several studies of acute toxicity of chlorpyrifos and diazinon with *Neomysis mercedis* (mysid), *Ceriodaphnia dubia*, and *Physa* sp. (pond snail) by the California Department of Fish and Game (CDFG 1992a; b; c; d; e; f; g; h; i; j; k; 1998a; b), and one of diazinon with *Lepomis macrochirus* (bluegill) by CIBA-GEIGY (1987) were obtained for review. All of these studies indicate that chlorpyrifos and diazinon are not particularly fast-acting toxicants, with mortality occurring at each 24-h observation period throughout the test (as opposed to only at the earliest observation period).

Based on the available pesticide literature, the 1-h acute averaging period utilized by the USEPA (1985; USEPA 2003d) is reasonable for the new methodology. This value is conservative for some pesticides that do not exert effects quickly (e.g., chlorpyrifos and diazinon), but should be protective for pyrethroids that exhibit latent effects after very short acute exposures, and for organophosphates (e.g., pirimiphos-methyl and temephos) that have been shown to exert their effects with 1-h pulse exposures of environmentally relevant magnitude.

To summarize, criteria derived by the new methodology include an expression of allowable exposure duration. For acute criteria, a 1-h averaging period is established, while for chronic criteria a 4-d period is established.

2-3.4 Allowable frequency of exceedance

In addition to magnitude and duration, it is necessary to consider the frequency with which a pesticide concentration may exceed a criterion without causing harm to aquatic organisms. For the new methodology, the allowable frequency of exceedance is based on a review of studies of the ability of organisms to recover from brief exposures to pesticides. Generally, studies of post-exposure recovery are addressed to one of several levels of organization: ecosystem, community, population, species and individual. All levels should be considered in order to meet the mandate of the Central Valley Regional Water Quality Control Board that “waters shall be maintained free of toxic substances in concentrations that produce detrimental physiological responses in...aquatic life” (CVRWQCB 2004).

2-3.4.1 Review of the literature

Yount & Niemi (1990) provide a good review of studies of recovery of lotic communities and ecosystems after physical or chemical disturbances or stresses. They

make an important distinction between press disturbances that occur over long periods of time causing sustained alterations in ecosystems, versus pulse disturbances that occur over shorter periods causing brief ecosystem alteration. The allowable duration of criteria exceedances included in this methodology (1 h for acute criteria; 1 h to 4 d for chronic criteria) are pulse disturbances by this definition. Thus, the same allowable frequency of exceedance applies to both acute and chronic criteria. This review focuses on studies of pulse, rather than press, disturbances. To examine recovery times, the review includes studies of recovery from brief, mild excursions of pesticide concentrations to toxic levels, as well as studies of recovery from catastrophic events (i.e., large spills). The latter are not really relevant for determination of time-to-recovery in cases of non-catastrophic events, but do provide good guidance on general aspects of ecosystem recovery.

The ability of an ecosystem to recover from a disturbance is dependent on several factors, which have been described by various authors in slightly different terms at different times. Cairns & Dickson (1977) give seven factors that determine how rapidly the recovery process will be after an ecosystem has suffered damage from one or more stressors (chemical or physical):

- 1) severity and duration of the stress
- 2) number and kinds of stressors
- 3) residual effects on the physical environment (e.g., from dredging or building dams)
- 4) presence of epicenters or refugia for recolonization
- 5) innate vulnerability of the system (e.g., lack of structural/functional redundancy)
- 6) inertia of the system (resistance to change)
- 7) resilience of the system (ability to readjust after exposure to a stressor)

Later, Cairns (1990) discussed, and quantified, six similar factors that affect an ecosystem's ability to recover from major ecological disasters: a) existence of nearby sources of re-colonizing organisms; b) voluntary or involuntary transportability of eggs, spores, larvae, flying adults, or other life stage; c) condition of habitat following stress; d) presence and persistence of residual toxicants following stress; e) chemical-physical environmental quality following stress; and f) potential for management/other agencies to assist in remediation. By rating each of these factors with a score of 1 = poor, 2 = moderate, or 3 = good, the following simple recovery index was defined:

$$\text{Recovery Index} = a \times b \times c \times d \times e \times f \quad (2.10)$$

A score of 400+ indicates that the ecosystem has an excellent chance of rapid recovery; 55-399 indicates a fair to good chance of rapid recovery; and < 55 indicates a poor chance. Rapid recovery is defined as having 40-60% of species reestablished within the first year after a major exposure event, and as many as 95% reestablished within three years. Using this system (and recognizing that excursions above water quality criteria are not in the realm of ecological disasters, but are rare, brief, and mild events of limited

scope), the following scores may be assigned to ecosystems in the Sacramento and San Joaquin River basins that have had excursions above water quality criteria:

a = 1-3; unaffected nearby tributaries are expected to be present, except in highly urbanized or heavily agricultural areas.

b = 1-3; depending on degree of transportability of species in damaged community

c = 3; exceedance is not expected to result in habitat alteration

d = 1-3; depending on chemical

e = 3; no chemical-physical alterations expected

f = 3; regulatory intervention expected

These values give a recovery index ranging from 27 to 729 indicating that the chance of rapid recovery within three years following a major ecological disaster is poor to excellent. It is reasonable to assume that rapid recovery (as defined by Cairns 1990) is likely in much of the Sacramento and San Joaquin River basins following brief, mild, limited-scope excursions above criteria levels. However, in cases where all nearby tributaries are affected by the toxicant, where species in the affected area are not readily transportable, and/or where the chemical either does not readily dissipate or has an ongoing source, recovery may be hindered. In such cases, site-specific frequency components of criteria may need to be derived.

In the same vein, Yount & Niemi (1990) report that in cases where rapid recovery has been observed in lotic systems it is because: 1) the life history characteristics of organisms allowed rapid recolonization and repopulation; 2) unaffected up- and downstream areas were available to supply new organisms; 3) lotic systems have high rates of flushing; and 4) organisms that live in lotic systems have evolved a lot of flexibility and adaptability because lotic systems are variable environments.

Yount & Niemi (1990) also summarized studies of ecosystem recovery following application of piscicides for eradication of undesired fish. In studies of rotenone treatments, fish recovery took from 12-16 months and benthic macroinvertebrate recovery took 12 months (Charles 1958; Little 1966). Similarly, in California macroinvertebrates recovered from rotenone applications within 6 months (Cook & Moore 1969), while in Scotland they recovered within one year (Morrison 1977). Macroinvertebrates had not recovered one year after application of toxaphene in Alaska (Meehan & Sheridan 1966). Jacobi & Degan (1977) report full recovery of macroinvertebrates one year after antimycin application, and Minckley & Mihalick (1981) observed complete recovery of benthic invertebrates three years after antimycin application (in this case, recovery may have occurred sooner, but the investigators did not check).

Whiles & Wallace (1995) studied macroinvertebrates in headwater streams exposed to methoxychlor in four seasonal treatments over four years. They concluded that abundance measures were not necessarily the best measure of recovery, but that ecosystem structure is important. They found that ecosystem recovery was dependent on the life cycles of the taxa making up the system. For example, organisms with shorter life

cycles or extended flight capability are able to recolonize more rapidly leading to more rapid ecosystem recovery. The ecosystem in this study took two years to recover nearly to its pre-treatment structure, but even then there were slight taxonomic and developmental stage differences.

In a review of studies of ecosystem recovery after damage from runoff of pesticides after forest application, Yount & Niemi (1990) report recovery times ranging from 2-3 months for some species of insects exposed to DDT (Hoffman & Drooz 1953) to more than four years for other insects (Hastings *et al.* 1961). Ephemeroptera exposed to aldrin required more than 19 months to recover, while Trichoptera and Chironomidae recovered in less than 19 months (Moye & Luckmann 1964). Arthropod biomass was reduced by exposure to fenitrothion, but recovered in 50 d (Eidt 1981). Considering fish species in situations of exposure due to runoff, Warner & Fenderson (1962) reported recovery of trout populations three years after DDT exposure. Keenleyside (1959) concluded that Atlantic salmon populations were able to recover if DDT application was not repeated more than once every three years. Contrarily, when forests were sprayed with DDT during hatching and smolt migration periods, Atlantic salmon recovery required 4-6 years due to near elimination of an entire age class (Elson 1967). In areas where DDT had been sprayed repeatedly for several years, recovery took longer than 9 years. The longer recovery periods in this study were related to the relatively long (5-6-year) life cycle of Atlantic salmon, and to the fact that the DDT was applied on a watershed scale resulting in isolation of the affected areas from refugia that might have provided migrants to repopulate the affected area. Such a drastic exposure from watershed-scale pesticide applications are not likely to occur in the Sacramento and San Joaquin River basins, thus the probability of eliminating, or nearly eliminating, an age-class from an entire basin is remote.

Studies of recovery times are also summarized by Yount & Niemi (1990) for cases of direct application of pesticides to water bodies. Corbet (1958) reported recovery of insect larvae in 40 d after DDT application to streams in Uganda. After treatment with methoxychlor Chironomidae recovered in 1-2 weeks, Plecoptera in 5 weeks, and blackflies in 2-4 weeks (Fredeen 1975; 1983). Snail populations recovered after treatment with the molluscicide, Bayluscid, after 10 months in hard water, and after 22 months in soft water (Harrison & Rattray 1966). After treatment with 3-trifluoromethyl-4-nitrophenol (TFM), benthic organisms in Lake Superior and Lake Michigan recovered in one year (Torblaa 1968), while Dermott & Spence (1984) observed recolonization of invertebrates in streams within three weeks of TFM treatment. In a follow-up experiment, Jeffrey *et al.* (1986) found that benthic invertebrates had not recovered more than 35 days after TFM treatment of a stream. The difference between the two studies was that the Jeffrey *et al.* (1986) study was conducted in colder weather when convective currents caused mixing of TFM 55 cm into the hyporheic zone. The Dermott & Spence (1984) study was conducted in warmer weather when the convective forces were not present and the uncontaminated hyporheic zone served as a refuge from TFM exposure.

A few other studies of recovery after pesticide exposure were reviewed by Yount & Niemi (1990). Ghetti & Gorbi (1985) simulated an accidental spill of parathion that

would produce measurable, but not catastrophic, effects on macroinvertebrates. Their test stream recovered in three months with macroinvertebrate density equaling or exceeding that of a reference stream after 117 days. However, macroinvertebrate trophic structure required two years for recovery (Wallace *et al.* 1986). In Nigeria, an isolated pool was treated with the insecticide Gammalin-20 in order to kill fish. After one month, no fish were found, but after three months the total number of taxa had returned to the pre-treatment state (Victor & Ogbeibu 1986).

One approach to measuring the ability of individuals or species to recover from pulse exposures to pesticides is to compare LC₅₀ values obtained in continuous exposure experiments with those obtained in pulsed exposure experiments. For example, Parsons & Surgeoner (1991a) found no difference in LC₅₀ values for mosquito larvae exposed to four 1-h pulses of carbaryl compared to those exposed continuously for four hours. They concluded that this indicated that there was no recovery during the 12-h intervals between pulses.

Turning to more recent studies, Liess & Schulz (1999) observed that in streams exposed to ethyl parathion, 4 of 11 species of macroinvertebrates that had disappeared after treatment recovered in 6 months; 9 species had recovered after 11 months; 2 species remained at low density for the full year of the study. Cold & Forbes (2004) found that *Gammarus pulex* recovered in 2 weeks from a 1-h pulse exposure to esfenvalerate. Heckmann & Friberg (2005) showed that macroinvertebrate community structure had recovered within two weeks from two 30-min pulses of lambda-cyhalothrin. *Daphnia magna* recovered to control levels of total neonates per female and population growth rate within 21 d after a 24-h pulse exposure to fenvalerate (Reynaldi & Liess 2005). Schulz & Liess (2001) observed chronic effects on populations of insect larvae more than 240 d after 1-h exposures to fenvalerate. Parsons & Surgeoner (1991b) reported that mosquito larvae exposed to 0.5-4 h pulses of permethrin, carbaryl and carbofuran were able to recover from immobility, but not after 8-24 h exposures.

2-3.4.2 Allowable frequency of exceedance-conclusion

In setting an allowable frequency of exceedance of a water quality criterion, the question is how much time it would take for organisms at various organizational levels to recover from brief pulse exposures to contaminants. Yount & Niemi (1990) conclude that ecosystem recovery from pulse exposures generally occurs in less than three years, and often in less than one year. Species that are slowest to recover are those with the longest life cycles. Likewise, Niemi *et al.* (1990) concluded that most ecosystems are able to recover from disturbances in less than three years except in cases where physical habitat was altered, the system was isolated, or residual pollutant remained. The majority of reviewed studies that consider community, population, or species-level effect indicate that recovery occurs in three years or less. The only exception is the study by Elson (1967), but, as discussed, the exposure conditions of that study were extreme and not really relevant to cases of brief, mild excursions above a water quality criterion.

Based on this review, three years between exposure events should allow full recovery from effects of an excursion above either acute or chronic water quality criteria in the Sacramento and San Joaquin River basins. This is in agreement with USEPA (1985; 2003d) methodologies, although the 3-year frequency component was supported by minimal data when it was first proposed. Acute and chronic criteria derived by the new methodology include a statement that exceedances should not occur more than once every three years.

2-4.0 Water quality effects

Because water quality criteria are derived from laboratory studies conducted in carefully controlled systems, it is necessary to consider the effects that water quality characteristics may have on the toxicity of a chemical in the environment. For pesticides, the major concerns are the effects of suspended and dissolved particulate matter on bioavailability, the effects of pesticide mixtures, and the effects of temperature, pH, or other water quality parameters on toxicity.

2-4.1 Bioavailability

The issue of bioavailability was discussed in the Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009), but this important and complex topic merits further exploration because there is a level of uncertainty in deriving criteria with data from laboratory studies conducted in clean water (i.e., solid-free) and then using those criteria to protect aquatic organisms in an environment with varying levels of solids. Two questions need to be addressed. First, which phase or phases of a chemical in water are bioavailable? And, second, when bioavailability is well understood for a particular pesticide, what is the best way to determine compliance with water quality criteria? The first question is addressed through a literature review. Options for addressing the second question include equilibrium partitioning models, direct analysis of pesticides in different phases, and the use of passive sampling devices to estimate concentrations of bioavailable pesticides. Each of these is discussed below.

Staples *et al.* (1985) consider bioavailable chemicals to be those available to exert toxicity or to bioaccumulate. They hypothesized that most neutral organic chemicals must be in solution in order to be bioavailable, and they cite studies with polynuclear aromatic hydrocarbons (PAHs), crude oil components and polychlorinated biphenyls (PCBs) that support this (Anderson *et al.* 1977; Halter & Johnson 1977; Neff 1979; Roesijadi *et al.* 1978a; Roesijadi *et al.* 1978b; Rossi 1977). Likewise, DiToro *et al.* (1991) conclude that compounds bound to either sediments or dissolved organic carbon (DOC) are not bioavailable. Reductions in accumulation and toxicity of synthetic pyrethroids in *Daphnia magna* have been related to the binding of these compounds to DOC (Day 1991), and reduction in bioconcentration of chlorobenzuron in *Daphnia magna* was attributed to binding of the compound to dissolved humic material (Steinberg *et al.* 1993). Kukkonen & Oikari (1991) showed that the total concentration of dissolved organic matter (DOM) in water is one of the main factors controlling bioavailability of organic contaminants. Pyrethroid uptake and toxicity in *Daphnia magna* and *Chironomus*

tentans in water-sediment systems was reduced with increasing sediment organic carbon concentration (Maund *et al.* 2002). There is a species-specific aspect to the effects of solids on bioavailability. Sediment-biota accumulation factors (BSAF) for PAHs were reduced for the marine polychaete, *Nereis diversicolor*, in sediments amended with activated carbon, but there was no change in the BSAF for the gastropod *Hinia reticulata* (Cornelissen *et al.* 2006).

The studies discussed so far provide evidence that solids in water affect bioavailability of organic contaminants, but they do not necessarily support the hypothesis that only freely dissolved contaminants are bioavailable. Several studies refute the “dissolved = bioavailable” concept. Dissolved humic material decreased the toxicity of diazinon, 4-chloroaniline, and 4-nitrophenol, had no effect on the toxicity of tetrabromobisphenol-A, o-toluidine, 3,4-dichloroaniline, and pentachlorophenol, but increased the toxicity of 2,4-dichlorophenol and 2,4,5-trichlorophenol to *Daphnia magna* (Steinberg *et al.* 1992). Fewer walleye survived exposure to chlorpyrifos-humic acid (HA) complexes than to either HA alone or chlorpyrifos alone, and no differences were seen in cholinesterase inhibition between chlorpyrifos-HA and aqueous chlorpyrifos exposures (Phillips *et al.* 2003). Schnürer *et al.* (2006) showed that glyphosate was microbially degraded even when sorbed to soil.

Without pesticide-specific, species-specific, site-specific information regarding which phases are bioavailable, compliance must be based on measurement of total pesticide concentration in water. If bioavailability information is available for a specific case, then there are several approaches that regulators may use to determine compliance. Each is discussed below.

Case 1: Pesticide is bioavailable in all three phases or no information is available. If studies show that a pesticide is bioavailable in solid, dissolved-solid and freely-dissolved phases, or if nothing is known about bioavailability for a particular pesticide on a site-specific basis, then compliance must be determined on the basis of the total concentration of pesticide in water.

Case 2: Pesticide is bioavailable in fewer than three phases. In this case, regulators still have the conservative option to determine compliance based on total pesticide concentration. However, if site-specific information is available, then compliance determination may be refined by consideration of just the bioavailable fraction or pesticide in water. The most direct approach is to measure pesticides in each phase individually, and then determine the total bioavailable concentration by adding together the results from each bioavailable phase. Exploring analytical options is beyond the scope of this project, but several studies have measured pesticide concentrations in three phases by various methods (e.g., Eadie *et al.* 1990; Liu *et al.* 2004; Rogers 1993). Two other options for determining compliance based on bioavailable pesticide are discussed below. The first is a modeling approach applicable in cases where only the dissolved fraction is bioavailable. The second is the use of passive sampling devices, which are applicable to any combination of phase bioavailability.

To address the case of bioavailability in the freely-dissolved phase, Staples *et al.* (1985) developed a simple model to describe the relationship between total and dissolved concentrations of a contaminant in water based on the concentration of suspended sediment in the water and the solid-water partition coefficient. This model is the one used in RIVM (2001), for converting total concentrations to dissolved concentrations:

$$C_{dissolved} = \frac{C_{total}}{1 + (K \cdot S)} \quad (2.11)$$

where: $C_{dissolved}$ = concentration of chemical in dissolved phase
 C_{total} = total concentration of chemical in water
 K = solid-water partition coefficient (L/kg); expressed as K_{oc}/f_{oc}
 S = concentration of sediment in water (kg/L)

One problem with this approach is that it makes no distinction between suspended solids and dissolved solids, which both affect bioavailability (DiToro *et al.* 1991; McCarthy *et al.* 1985), but which can have very different partition coefficients (Delle Site 2001). Measured DOC-water partition coefficients (K_{DOC}) for fluoanthene were incorrectly estimated in 11 different sediment pore waters using models that assume $K_{DOC} = K_{OC}$ (Brannon *et al.* 1995). In a study of phenanthrene binding and sorption to humic acids (HA), Laor *et al.* (1998) found that partition coefficients for dissolved HA were at least an order of magnitude higher than coefficients for mineral-associated HA. Thus, trying to describe the partitioning process based only on a K_{oc} value is an oversimplification.

To improve this model, researchers have expanded it to include three phases: freely-dissolved, adsorbed to dissolved organic matter, and adsorbed to solids (Eadie *et al.* 1990; Liu *et al.* 2004). The Great Lakes criteria derivation methodology uses a three-phase model for derivation of human health and wildlife criteria (Eadie *et al.* 1990; USEPA 2003d). The three-phase model is an improvement over the two-phase model, but it does not acknowledge that partition coefficients vary considerably depending on the nature of the solids. Normalizing the partition coefficient to organic carbon is a common approach to reduce that variability, but even K_{oc} values vary depending on the nature of the organic carbon. Delle Site (2001) notes that the sorptivity of organic matter depends on the relative proportion of humic and fulvic acids, lipids and humins (based on studies by Chiou *et al.* 1987; Chiou *et al.* 1986; Garbarini & Lion 1986; Gauthier *et al.* 1987). Kukkonen & Oikari (1991) reported that the degree of aromaticity and portion of hydrophobic acids in DOM are important controlling factors in the sorption of organic compounds. The log K_{OC} values for nonylphenol were 1.71, 3.08, 4.15, 4.50, and 4.71 for cellulose, chitin, lignin, humic acid, and natural sediment, respectively (Burgess *et al.* 2005). In a study of chlorpyrifos binding to colloidal materials, Wu & Laird (2004) found that chlorpyrifos sorbed strongly to a calcium-humate and did not desorb, but moderately sorbed to and desorbed from a river sediment. They concluded that both the organic and inorganic materials in suspended sediment affect the adsorption and desorption of chlorpyrifos. In light of these studies, simple partitioning models are not useful for making general predictions about phase distributions for organic contaminants. It may be

possible to use them in site-specific situations in which partition coefficients are available for the specific types of solids present in the site water.

Simple partitioning models also assume that the mechanism for reduced bioavailability and toxicity of organic contaminants in the presence of solids is adsorption or binding of the contaminant to the solids. However, Steinberg *et al.* (1992) found that dissolved humic material and sunlight enhanced diazinon degradation, which is another mechanism reducing toxicity. This is not surprising in that photolysis of DOM is a known source of hydroxyl radicals in waters (Takahashi *et al.* 1988).

Semi-permeable membrane devices (SPMDs) are passive sampling devices that are intended to mimic uptake of bioavailable contaminants (Huckins *et al.* 1990). Using laboratory-determined uptake rate constants, SPMDs can be used to determine time-integrated ambient water concentrations of bioavailable contaminants. These devices work well both qualitatively and quantitatively for hydrophobic organic chemicals (Huckins *et al.* 1990; Lu & Wang 2003), but do not give reliable quantitative results for polar organics (Alvarez *et al.* 2004). Using SPMDs for quantitative analysis of ambient waters is problematic because each device has a specific membrane permeability, and uptake rates are affected by flow rates, temperatures, and fouling (Huckins *et al.* 1990). For measuring hydrophobic compounds, performance reference compounds (PRCs) have been developed to counter these variables (Huckins *et al.* 2002), but attempts to develop PRCs for polar organics have not been successful (Alvarez *et al.* 2004).

SPMDs are promoted for their ability to take time-integrated water samples such that contaminant pulses can be detected. However, this type of sampling is not useful for determination of short-term variability in ambient water contaminant concentrations (Gustafson & Dickhut 1997; Prest *et al.* 1998), and thus would be of little value in measuring compliance with acute criteria that must be met on the basis of 1-h average concentrations. Other problems with SPMDs include the fact that compounds are subject to photolysis if devices are deployed in shallow water or near the surface (USGS 2000), and that they are not accurate measures of bioaccumulation because devices do not mimic biological processes (i.e., metabolism, excretion chemical, movement, feeding) that affect equilibrium concentrations in organisms (Huckins *et al.* 2004).

There is much evidence that solids in natural waters affect the uptake and toxicity of organic contaminants. As discussed in Phase I (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) pesticide loadings to surface waters are typically due to storm or agricultural runoff, and since suspended solids are also higher than normal during runoff events, it would be ideal if criteria could be expressed, or compliance could be determined, in terms that reflect this covariance of pesticides and suspended solids. This is particularly important in the Central Valley of California where suspended solids levels are quite variable. The US Geological Survey reports levels ranging from 1-330 mg/L in samples from various streams in the Sacramento River Basin and from 1-5280 mg/L in the San Joaquin River Basin (USGS 2005a; b).

As the discussion in this section has shown, there is no simple way to incorporate the effects of solids on bioavailability into either criteria derivation or compliance determinations. It is not correct to assume that freely dissolved compounds are equivalent to bioavailable compounds. Further, it is too simplistic to assume that general partition coefficients are valid for all kinds of solids. Thus, there is no general way to predict bioavailability from physical-chemical parameters and water quality data. However, if studies are available that show which fraction, or fractions, of a particular pesticide is bioavailable, then it may be of interest to directly measure the pesticide concentrations in those fractions to determine compliance. For pesticides which are only bioavailable in the freely dissolved phase, and for which K_{OC} and K_{DOC} values are available, use of a three-phase partitioning model is an option for translating measured total pesticide concentrations into dissolved concentrations. If technical limitations of passive sampling devices can be overcome, they offer another option for estimating bioavailable pesticide in water samples, but only for determination of compliance with chronic criteria. The new criteria derivation methodology includes brief guidance regarding how to address bioavailability.

2-4.2 Mixtures

Various approaches to addressing toxicity of mixtures were discussed in Phase I (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009). Lydy *et al.* (2004) also provides a good review of pesticide mixtures. They both concluded that there really is no way to derive criteria for all of the potential mixtures of pesticides that could occur in a water body. Nonetheless, there are many models available for determination of mixture toxicity. The question is whether such models, which are designed to predict toxicity of a mixture to a single species, can be adapted for use in determination of compliance to a water quality criterion that applies to multiple species. The CVRWQCB (2004) has done this by substituting water quality criteria for LC/EC₅₀ values in a simple concentrations addition model and using the results to assess compliance. Felsot (2005) has shown that this is also possible with a toxic equivalence model.

How to best model chemical mixtures will depend on the nature of the mixture. All models must be applicable to mixtures of two or more components. Where models will differ is in whether they apply to chemicals of similar or different modes of action, and whether the mixtures show additive toxicity, or if there are known interactions leading to antagonistic (less than additive) or synergistic (greater than additive) effects. Pesticide modes of action are often known; pesticide interactions are not as well-studied. The following discussion is broken down according to the mixture models that are available. Each is described and discussed in terms of its applicability to particular types of mixtures or to determination of compliance with water quality criteria. The SSD approach to mixtures (Traas *et al.* 2002) was explored in Phase I (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009), but is not discussed further here as it is fairly complex, it is not readily adaptable for use in compliance determination, and it requires that SSDs be available for each component of the mixture, which would limit its applicability.

2-4.2.1 Additivity

There are two basic models to describe additive mixture toxicity. The concentration addition model is used for chemicals with similar modes of action, while the independent action model is used for chemicals with different modes of action (Plackett & Hewlett 1952).

2-4.2.1.1 Concentration addition—for similar modes of action

The basic concentration addition model is expressed as (acc. to Olmstead & LeBlanc 2005; PapeLindstrom & Lydy 1997):

$$\sum_{i=1}^n \frac{C_i}{ECx_i} = 1 \text{ TU} \quad (2.12)$$

where:

C_i = concentration of the i th chemical in the mixture

ECx_i = concentration of the i th chemical that elicits the same response (x) as the full mixture

TU = toxic unit

If the sum is 1TU, then the mixture exhibits additivity. If the sum is greater than 1 TU, the mixture is less than additive (antagonistic); if less than 1 TU, the mixture is greater than additive (synergistic). For example, if the EC_{50} is the chosen level of effect, then it is expected that the sum of the concentration: EC_{50} ratios for each component of the mixture will equal 1 TU. If less than 1 TU is required to elicit 50% effect in the mixture, then the mixture is showing more than additivity (synergism); if more than 1 TU is required, then the mixture is showing less than additivity (antagonism).

The CVRWQCB (2004) adapts this equation for determination of compliance with water quality criteria:

$$\sum_{i=1}^n \frac{C_i}{O_i} < 1.0 \quad (2.13)$$

where:

C_i = concentration of toxicant i in water

O_i = water quality objective/criterion for toxicant i

As long as the sum is < 1.0 , the water body is considered to be in compliance with respect to the mixture.

Felsot (2005) argued that actual toxicity values (e.g., LC_{50} s) should be used in these equations instead of water quality objectives for criteria to more accurately reflect

toxicity of the mixture. This argument was made in regards to acute criteria that include a safety factor of 2 after determination of an acute value. However, for determination of compliance, the question is simply whether concentrations of chemicals in a water body are below criteria levels. To answer this question, using water quality criteria in these equations is appropriate.

As an alternative to equation 2.13, Felsot (2005) suggested using the relative potency factor (RPF) approach to determine compliance in cases of additive toxicity for compounds with similar modes of action. The RPF approach is analogous to the toxic equivalency factor (TEF) approach used in assessing toxicity of dioxin and dioxin-like compounds (Van den Berg *et al.* 1998). By the RPF approach, one chemical (usually the most toxic) is chosen to be the reference chemical and the potency of all other similarly-acting chemicals is expressed relative to the reference. The potency of each chemical is divided by the potency of the reference chemical and this ratio, the RPF, is multiplied by measured concentrations of each non-reference chemical to produce concentrations in terms of equivalents of the reference chemical. Compliance with the objective for the reference chemical is based on the sum of the measured concentration of the reference chemical plus the concentrations of the equivalents. Mathematically, this is expressed as:

$$TE_{total} = C_R + \sum_n^i TE_i \quad (2.14)$$

where:

TE_{total} = total toxic equivalents of mixture ($\mu\text{g/L}$)

C_R = Concentration of reference chemical ($\mu\text{g/L}$)

TE_i = toxic equivalents of i th component of the mixture ($\mu\text{g/L}$)

and

$$TE_i = RPF_i * C_i \quad (2.15)$$

where:

RPF_i = relative potency factor of the i th component of the mixture

C_i = concentration of the i th component of the mixture ($\mu\text{g/L}$)

and

$$RPF_i = \frac{EC_{xR}}{EC_{xi}} \quad (2.16)$$

where:

EC_{xR} = concentration of reference chemical causing $x\%$ effect when tested alone ($\mu\text{g/L}$)

EC_{xi} = concentration of mixture component i causing $x\%$ effect when tested alone ($\mu\text{g/L}$)

For compliance determination, a multispecies RPF is needed, so equation 2.16 can be written:

$$RPF_i = \frac{Criterion_{xR}}{Criterion_{xi}} \quad (2.17)$$

where:

$Criterion_{xR}$ = water quality criterion of reference chemical ($\mu\text{g/L}$)

$Criterion_{xi}$ = water quality criterion of the i th chemical ($\mu\text{g/L}$)

Using these equations, if $TE_{total} \leq$ the criterion for the reference compound, then the water body is in compliance.

Both the concentration addition approach and the TEF approach are valid for determination of compliance in cases of additive toxicity when chemicals in the mixture have similar modes of action. Both are included in the new methodology allowing regulators to select which one works best for them.

2-4.2.1.2 Response addition—for independent modes of action

For chemicals that do not have similar modes of action, the response addition, or independent action, model is used. It is expressed mathematically as (acc. to Belden & Lydy 2006):

$$R_{mix} = 1 - \prod_n^{l-1} (1 - R_l) \quad (2.18)$$

where:

R_{mix} = response of the mixture (i.e., percent response)

R_l = response expected from the l th component the mixture

This model is not applicable for determination of compliance. To illustrate, consider an example in which two pesticides are each present at concentrations equal to its water quality criterion. Since the criteria are based on a concentration that might affect 5% of species, the R_l value for each pesticide is 0.05 and the equation is:

$$R_{mix} = 1 - (1 - 0.05) * (1 - 0.05) = 0.10, \text{ or } 10\% \quad (2.19)$$

Thus, the mixture could harm as many as 10% of species, which is unacceptable. To ensure compliance, each of the two chemicals would have to be present at a level that might be expected to harm 2.5% of species, so that R_{mix} would be at or below 5%. Estimates of percentile values below the 5th percentile are highly variable (Phase I, TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009), and yet all values of interest for

application of the independent action model to compliance determination for mixtures lie below this level (values above the 5th percentile would indicate non-compliance with individual criteria). While the independent action model works well for predicting toxicity of mixtures when individual toxicities are known, it is not adaptable for compliance determination.

2-4.2.2 Non-additivity; synergism and antagonism

Chemical mixtures may display non-additive toxicity in the form of either antagonistic or synergistic effects. This indicates an interaction between chemicals such that the response observed for a mixture is either less than (antagonism) or greater than (synergism) that predicted by additivity models. The concept of synergy is often used in reference to cases where one chemical present at non-toxic concentrations increases the toxicity of a second chemical, but it can be applied to mixtures in which both chemicals are at toxic levels. Mu & LeBlanc (2004) utilized the coefficient of interaction (K) to define this relationship. First described by Finney (1942), the basic equation is:

$$K_x = \frac{EC50_0}{EC50_x} \quad (2.20)$$

where:

K_x = coefficient of interaction at synergist/antagonist concentration x

$EC50_0$ = EC_{50} of chemical in absence of synergist/antagonist

$EC50_x$ = EC_{50} of chemical in presence of synergist/antagonist at concentration x

When a measured concentration of a chemical is multiplied by the K_x value for a given concentration of a synergist/antagonist, the value that results is an adjusted, or effective, concentration of the chemical. Mathematically, this is expressed as:

$$C_a = C_m(K) \quad (2.21)$$

where:

C_a = adjusted, or effective, concentration of chemical

C_m = concentration measured

K = coefficient of interaction

For application to compliance determination equation 2.21 could be used and the effective concentration compared to the water quality criterion. Additionally, the effective concentration could be used in additivity models described in section 2-4.2.1.1. The difficulty is in determination of an appropriate K value. Rider & LeBlanc (2005) fit logistic functions to describe the relationship between K values and piperonyl butoxide (PBO) concentration, such that K values could be estimated for a wide range of PBO concentrations. Unfortunately, K values derived in that manner are not generally applicable. In the case of Rider & LeBlanc (2005) they are specific to the interaction of

PBO with either malathion or parathion and the toxicity of binary or ternary mixtures of those chemicals to *Daphnia magna*.

Equation 2.21 can be modified, in theory, for mixtures containing both synergists and antagonists, or multiple synergists/antagonists, (LeBlanc, pers. comm. 2006):

$$C_a = C_m(K_1K_2...K_n) \quad (2.22)$$

where:

C_a and C_m are as defined in equation 2.21
 $K_1, K_2, K_n = K$ values for synergist/antagonist 1, 2...n

Dr. LeBlanc cautions that as more K values are strung together, the error of each term will lead to large error in the adjusted concentration. Thus, this approach should not be used for compliance determination, but may be used to assess research needs.

To use the interaction coefficient concept in determination of water quality criteria compliance would require the establishment of relationships between K values and synergist/antagonist concentrations. That is, for pesticides that commonly occur together, it might be worth the research effort to establish relationships (i.e. predictive equations) between K and concentrations of known synergists/antagonists for a range of species. This issue is discussed further at the end of section 2-4.2.3.

2-4.2.3 Combined models

Each of the models discussed so far apply to only one type of mixture effect. In the environment, it would not be unusual to find complex mixtures that include chemicals that show all three of the basic mixture effects: additivity with similar modes of action, additivity with different modes of action, and interaction leading to synergism or antagonism. Olmstead & LeBlanc (2005) developed a model that combines concentration addition and response addition models in to one. Rider & LeBlanc (2005) expanded on that model to include an interaction component. The basic model equation is:

$$R = 1 - \prod_{i=1}^N \left\{ 1 - \frac{1}{1 + \frac{1}{\left(\sum_{i=1}^n \frac{k_{a,i}(C_a) \times C_i}{EC50_i} \right)^{\rho'}}} \right\} \quad (2.23)$$

where:

R = response of the mixture (percent of individuals responding)
 N = number of cassettes (cassette = group of chemicals of similar mode of action)

I = I th cassette

n = number of chemicals

i = i th chemical

$k_{a,i}$ = interaction coefficient for chemical a (synergist/antagonist) interacting with chemical i

C_a = concentration of chemical a in the mixture

C_i = concentration of chemical i in the mixture

$EC50_i$ = EC_{50} for chemical i alone

ρ' = average power (slope) of dose-response curves of chemicals in cassette I

This model integrates all aspects of mixture toxicity, but can only be applied to one species at a time. Gerald LeBlanc was contacted to discuss the possibility of adapting this model for determination of criteria compliance (pers. comm. 2006). The proposal was put forth that some of the variables in equation 2.23 be redefined as follows:

R = response to the mixture (percent of species responding)

$k_{a,i}$ = multi-species interaction coefficient

$EC50_i$ = replace with water quality criterion

ρ' = average slope of species sensitivity distributions in cassette I

Using the model in this way, when $R > 0.05$, the criterion would not be met, because the criterion is based on the 5th percentile of the SSD. Likewise, for $R \leq 0.05$, the mixture would be in compliance.

LeBlanc's response was that this approach seemed reasonable in all respects except that it may not be possible to derive a reliable multi-species K value. His concern was that since K is a mechanism-dependent value, assuming a common value across species would be equivalent to assuming similar toxicity mechanisms across species. He gave the example of a chemical mixture containing an androgen receptor antagonist and an androgen synthesis inhibitor. In vertebrates, these chemicals interact synergistically, but in invertebrates they do not because invertebrates lack androgen receptors. Thus it may not be possible to derive a valid ecosystem K value for this mixture. Before it can be used for compliance assessment, the Rider & LeBlanc (2005) model would have to be validated for use across species. However, K-values for individual species could be used to assess the potential harm from non-additive toxicity on a species by species basis.

2-4.2.4 Conclusions on mixtures

Among the mixture approaches presented, only the concentration addition models are readily applicable for determination of compliance with water quality criteria. The interaction coefficient concept could be used if further research provided K values applicable to multiple species over a range of synergist/antagonist concentrations. The combination model of Rider & LeBlanc (2005) that incorporates concentration addition, response addition and interaction components holds promise as a way to assess toxicity

of complex mixtures, but further research and testing is needed to find a way to utilize it in compliance determination.

The new methodology includes both of the concentration addition approaches discussed in section 2-4.2.1.1. The non-additive model (section 2-4.2.2) is also included, with the caveat that it can only be used if reliable K values are available (either a multispecies value, or individual species values). The response addition and combined models are not incorporated into the methodology. Regulators can choose among the models and apply them to determine compliance with water quality criteria or to assess the potential for harm due to non-additive toxicity.

2-4.3 Other water quality effects

As described in USEPA (1985; 2003d), if data are available to establish quantitative relationships between water quality characteristics and toxicity, then criteria should be expressed as equations reflecting that relationship. Both USEPA methodologies (1985; 2003d) provide detailed instructions for an acceptable method for determination of acute and chronic criteria in cases where toxicity to two or more species is related to a water quality characteristic (hardness, pH, temperature, etc.). The USEPA approach is included in the new methodology. Details are not reproduced here, but may be found in Chapter 3, section 3-5.3.

2-5.0 Check criteria against ecotoxicity data

Once derived according to methods discussed in the procedures in section 2-3.0, criteria must be evaluated to ensure that they are set at levels that will protect against adverse effects to: 1) particularly sensitive species, 2) species within ecosystems, and 3) threatened and endangered species (TES).

2-5.1 Sensitive species

Derived criteria should be compared to studies of the most sensitive species in the data set to ensure that these species will be protected. If a calculated criterion is higher than toxicity values reported for a particularly sensitive species, then the criterion may require downward adjustment, for example, by using the lower 95% or 99% confidence interval estimate of the 5th percentile, rather than the median.

2-5.2 Ecosystem and other studies

As recommended in section 2-2.1.4, criteria should be evaluated against field or semi-field data to judge whether they will be protective of all species within ecosystems. If not, then criteria may need to be adjusted downward. This is consistent with several criteria derivation methodologies (OECD 1995; RIVM 2001; USEPA 1985; 2003d; Zabel & Cole 1999) and is included in the new methodology.

2-5.3 Threatened and endangered species

A number of threatened and endangered species (TES) live in the waters of the Central Valley. Due to their protected status, it is likely that very little toxicity test data will be available for these species (TES). However, it is important to ensure that they are protected by water quality criteria. Certainly, if data of acceptable quality are available for TES, then those data should be included in sets used for criteria derivation. Derived criteria should be checked against toxicity values for TES to ensure that criteria will be protective of those species. Since criteria were calculated to be estimates of ecosystem no effect levels, they should be protective of TES. However, it is worthwhile to use available data and tools to confirm this.

When effects data are lacking for TES, both the USEPA (2003d) and the Australia/New Zealand (ANZECC & ARMCANZ 2000) guidelines suggest that studies with appropriate surrogate species may be used to set criteria. Deciding on appropriate surrogate species is left to professional judgment. More rigorous methods for estimating toxicity to TES based on surrogates are provided by both QSARs and interspecies correlation estimates (ICE). The ICE model (Asfaw *et al.* 2003) is available for free download at <http://www.epa.gov/ceampubl/fchain/index.htm> (documentation included in Chapter 3, Appendix 3A). It can be used to estimate toxicity for any chemical for which a toxicity value is available for a surrogate, but, so far, has only been developed for estimation of acute toxicity. Also, the ICE model works best when making correlations within families. It can be used for larger taxonomic distances, but estimates will not be as good. QSARs are available for both acute and chronic toxicity (e.g., RIVM 2001), but, as discussed in section 2-2.2.1, application of QSARs is not well developed for chemicals with non-narcotic modes of action, which severely limits their usefulness in estimating pesticide toxicity.

To assess whether criteria derived by the new methodology will be protective of TES (based on the most recent list available), the following procedure is included:

For comparison to acute criteria:

- 1) Compare criteria to toxicity values from acceptable studies of effects on TES.
- 2) If no toxicity values are available for a TES, but an acceptable acute toxicity value is available for a surrogate species in the same family or genus as the TES, then use the ICE (v. 1.0) program to estimate a toxicity value for the TES; compare this estimated value to the acute criterion.
- 3) If no surrogate value is available, and if the chemical of interest has a narcotic mode of action, select a QSAR (e.g., from OECD 1995; RIVM 2001) that can be used to estimate toxicity based on a log K_{OW} value.

For comparison to chronic criteria:

- 1) Compare criteria to toxicity values from acceptable studies of effects on TES.

2) If the chemical of interest has a narcotic mode of action, select QSARs (e.g. from OECD 1995; RIVM 2001) that can be used to estimate toxicity based on a log K_{OW} value.

If no data for the TES or acceptable surrogates are available, or if QSARs are not applicable, then it will not be possible to assess whether changes in the criteria are required to be protective of these species. If any of the above comparisons reveal that a criterion is higher than any of the TES toxicity values (or estimated values), then the criterion may need to be adjusted downward.

2-6.0 Partitioning to other environmental compartments

While partitioning of a pesticide to other environmental compartments is a concern for environmental managers is it not the goal of water quality criteria. This section is intended to check for agreement between the water quality criteria and existing guidelines for wildlife, humans, air and sediment. If these criteria are found to be in conflict with any existing guidelines, the concern should be flagged for review by environmental managers, but the criteria should not be adjusted.

2-6.1 Bioaccumulation/secondary poisoning

This methodology is concerned with setting water quality criteria for the protection of aquatic life, thus it is not directly concerned with the protection of terrestrial wildlife or human health. However, for potentially bioaccumulative chemicals it is important to be sure that water quality criteria are set at levels that do not lead to unacceptable levels of chemicals in food items. The new methodology includes a procedure for checking calculated chronic criteria for the possibility of secondary poisoning of wildlife, or possible human health effects, due to bioaccumulation in fish or other food items. Acute criteria do not require this check because they are intended to protect against short periods of elevated pesticide concentrations, making the equilibrium model inappropriate. For wildlife, this requires the availability of studies that demonstrate adverse effects from dietary intake of toxicants; for human health, this requires the availability of FDA action limits for the chemical of concern. The procedure described here is based on the discussion in section 7.3.2 of the Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009). The discussion is framed in terms of fish tissue, but the procedure may be applied to any potential food items.

The first step in the process is to determine if the chemical of interest has the potential to bioaccumulate. The OECD (1995) provides useful guidance, which is incorporated into the new methodology. According to this guidance, chemicals are likely to bioaccumulate if they have a log $K_{OW} > 3$, molecular weight < 1000 , molecular diameter $< 5.5 \text{ \AA}$, and molecular length $< 5.5 \text{ nm}$. The latter two parameters are not readily available for many chemicals, but may be use as guidelines if available. Chemicals are not expected to bioaccumulate if they are reactive and/or readily metabolized.

The next steps only apply if a chemical is determined to have bioaccumulative potential, and if dietary toxicity data or FDA action levels are available. Based on an equation in the EU risk assessment technical guidance document (ECB 2003), the Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) proposed the following equation for translating dietary NOEC or LC₅₀ values, or FDA action, levels to water NOEC values:

$$NOEC_{water} = \frac{NOEC_{oral-predator}}{BCF_{fish} \cdot BMF_{fish}} \quad (2.24)$$

or:

$$NOEC_{water} = \frac{LC_{50,oral-predator}}{BCF_{fish} \cdot BMF_{fish}} \quad (2.25)$$

where:

$NOEC_{water}$ = NOEC in water; concentration in water below this level is not expected to lead to bioaccumulation to harmful levels in food items;

$NOEC_{oral-predator}$ = dietary NOEC value for wildlife or FDA action level;

$LC_{50,oral-predator}$ = dietary LC₅₀ value for wildlife (mg pesticide/kg food);

BCF_{fish} = bioconcentration factor; ratio of concentration of chemical in tissue due to water-only exposure to concentration in water; whole-body, wet weight value (ECB 2003; OECD 1995);

BMF_{fish} = biomagnification factor in food item; ratio of concentration of chemical in predator to concentration in prey items; based on lipid-normalized values, if available (ECB 2003)

If no measured BMF is available, then an appropriate default value (Table 2.9; based on log K_{ow} or BCF) should be used (ECB 2003).

Table 2.9 Default BMF values (ECB 2003).

log K_{ow}	BCF	BMF
< 4.5	< 2,000	1
4.5 - < 5	2,000-5,000	2
5 - 8	5,000	10
> 8 - 9	2,000-5,000	3
> 9	< 2,000	1

Alternatively, if a bioaccumulation factor (BAF) is available for fish, then equation 2.24 may be modified to:

$$NOEC_{water} = \frac{NOEC_{oral-predator}}{BAF_{fish}} \quad (2.26)$$

where:

$NOEC_{water}$ = NOEC in water;

$NOEC_{oral-predator}$ = dietary NOEC for wildlife or FDA action level (mg pesticide/kg food);

BAF_{fish} = bioaccumulation factor in fish; ratio of concentration of chemical in tissue due to water plus dietary exposure to concentration in water.

In this form, the equation is similar to the USEPA (USEPA 1985; 2003d) method for determination of a final residue value, but for the new methodology, the BAF may not simply be replaced with a BCF value. If no BAF value is available, then equation 2.24 or 2.25 must be used, and if no measured BMF value is available, then the appropriate default value should be used (Table 2.9). If multiple BCF, BAF or BMF values are available for a chemical, the geometric mean of all acceptable values should be used.

To determine compliance the $NOEC_{water}$ derived from either equation 2.24, 2.25, or 2.26 would be compared to the chronic water quality criterion. If it is above the criterion, then the no adjustment of the criterion is necessary. If the $NOEC_{water}$ is below the criterion, then this should be indicated in the final criteria statement that these criteria may not be protective of all beneficial uses based on the bioaccumulation/secondary poisoning section and that additional review is needed.

2-6.2 Harmonization with sediment and air criteria

The final element that regulators may wish to consider is whether a pesticide that is present in the water at criteria levels might have the potential to move from the water compartment into another environmental compartment (i.e., sediment, biota, air) where it may exceed levels of concern established for those compartments. This is the concept of harmonization as discussed in the Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009). Consideration of bioaccumulation and secondary poisoning is a specific case of harmonization of criteria between two compartments. As this is an assessment of equilibrium conditions, it is only necessary to consider chronic criteria concentrations, as discussed for bioaccumulation (section 2-6.1). This procedure is not necessary if levels of concern have not been established, and toxicity data are not available, for air, sediment or biota compartments.

Steady-state environmental models are tools that may be used to assess harmony, or coherence, across environmental media. Acceptable models are available for free over the internet. The Exposure Analysis Modeling System (EXAMS, Burns 2004) is available from the USEPA Center for Exposure Assessment Modeling (CEAM; <http://www.epa.gov/ceampubl/swater/index.htm>). Mackay's (2001) Fugacity-Based Environmental Equilibrium Partitioning Models Levels I, II and III, are available for free download from the Canadian Environmental Monitoring Center (CEMC; <http://www.trentu.ca/cemc/welcome.html>). To estimate equilibrium concentrations of

pesticides in various compartments at equilibrium, the easiest-to-use and most appropriate model is the Level I fugacity model. The EXAMS and Level II and III fugacity models can be used, but are more complex and require more site-specific input. The Level I fugacity model requires several input parameters including water solubility, vapor pressure, melting point and log K_{OW} for prediction of steady-state concentrations of chemicals in air, water, suspended sediment, bed sediment and biota. Note that steady state environmental models determine equilibrium concentrations in various compartments. The equilibrium that exists between any two compartments may be described by the following equation (based on a simple first-order kinetic model):

$$C_1k_{12} = C_2k_{21} \quad (2.27)$$

where:

C_1 = concentration of pesticide in compartment 1

C_2 = concentration of pesticide in compartment 2

k_{12} = rate constant for transfer of pesticide from compartment 1 to compartment 2

k_{21} = rate constant for transfer of pesticide from compartment 2 to compartment 1

Equation 2.27 can be rearranged:

$$C_1/C_2 = k_{21}/k_{12} = K \quad (2.28)$$

where:

K = the equilibrium constant between the two compartments

In the simulations that will be run to assess harmonization in the new methodology, the concentration of pesticide in water will be set at the chronic criterion level by adjusting the total mass of pesticide in the system. According to equation 2.28, as long as C_1 is constant, C_2 will also be constant for a given equilibrium constant. Thus, the only kinds of model input changes that will affect final concentrations in non-water compartments are those that affect the equilibrium constant. For example, changing lipid levels in fish, or organic carbon content in suspended sediments will cause changes in equilibria, but changing concentrations of solids, or volumes of air or water will not. Model simulations can be run over a range of values to provide information applicable to a variety of site-specific situations.

Model outputs, based on having a chemical of concern at its chronic criterion level in water, should be compared to appropriate levels of concern (if any) established for the non-water compartments (e.g., sediment or air quality criteria or FDA action levels). If the steady-state concentrations in all compartments are below their respective levels of concern, then the water quality criterion is acceptable. If not, then this should be indicated in the final criteria statement that these criteria may not be protective of all beneficial uses based on the harmonization procedure and that additional review is needed. Before such an adjustment is made, though, further site-specific modeling is

recommended. The new methodology includes a harmonization procedure and a list of available, acceptable models with documentation.

2-7.0 Assumptions and limitations to methodology

The assumptions, limitations and uncertainties involved in criteria generation should be included in the criteria methodology and reports to inform environmental managers of the accuracy and confidence in criteria. This chapter was written to discuss many of the assumptions and limitations as individual procedures were chosen. Major assumptions, limitations and uncertainties will be reviewed here, followed by recommendations for data to improve the criteria derivation process.

One benefit of calculating criteria using a statistical distribution is that it provides a quantitative measure of the uncertainty in the estimate. Several calculations of distributional estimates will be included in criteria reports (as discussed in section 2-3.1.3) and the uncertainty in the resulting criteria can be seen by comparing the median estimate to the estimate with 95% confidence. In this method the confidence level of these estimates is calculated assuming that the uncertainty in the fitted distribution is the greatest source of uncertainty in the criteria calculation. The variation in species sensitivities is likely to outweigh other calculable sources of uncertainty (see section 2-3.1.3), such as the uncertainty of an LC_{50} , which could be expressed as the confidence limits of the reported LC_{50} .

However, these confidence intervals do not include uncertainty from the assumptions listed in section 2-3.1.5.1, which are universal to any method using SSDs for protection of all species. These may be the most important assumptions influencing the effectiveness of criteria, unfortunately the uncertainty associated with these assumptions is nearly impossible to quantify at this time. Also, while this method was designed to incorporate many considerations not yet in many methodologies, it is impossible to account for all. Some, such as sublethal effects and additive effects with other compounds (i.e., additive effects of compounds with different modes of action) are too complex to incorporate at this time. These particular issues and others could lead to underprotection, while others could lead to overprotection. Further, the models included will likely often be limited by available data.

Few chronic data are often available, which will be limiting for any method. This method employs acute to chronic ratios, which are a fairly common means to provide protective estimates when chronic data are lacking. These values are based in the assumption that it is possible to extrapolate from toxicity data at one life stage to an entire life cycle. When sufficient pesticide specific data are not available, this method also provides a default acute to chronic ratio and assessment factors. These factors were derived empirically using pesticide data to be relevant for the goal of this method; however, they are limited by the relatively small amount and diversity of pesticide data available.

Other limitations are likely to be encountered when deriving a particular criterion and these should be discussed in individual criteria reports. Final criteria statements should briefly review any data limitations that affect the procedure used to determine the final criteria, making it obvious how the final criterion was derived. An example of an important limitation affecting the derivation process would be missing taxa requirement that required use of assessment factors. Also any other considerations (as described in section 2-4.0 through 2-6.0) should be included that may be important for policy makers to consider.

2-7.1 Data generation to improve criteria derivation

In the process of deriving this methodology specific areas were identified where more data would help to derive better criteria. This section highlights some of the procedures most in need of additional data and research.

For many pesticides one of the largest limiting factors in the certainty of calculated criteria is the amount of data available, especially for newer compounds. While a large amount of toxicity data may be generated for pesticide registration, a pesticide may be registered with as little as 3 acute toxicity studies on freshwater organisms: one invertebrate and one or two fish. While in practice more data are generated, it is uncommon to have a dataset sufficiently large to use established methods such as EPA 1985. Chronic data is especially needed, as pesticide regulation requires only one fish early-life stage test and one invertebrate life cycle test for freshwater chronic studies (40 CFR 158.630). It would also be helpful if chronic studies were reported with both hypothesis testing and regression analysis (EC_x) as regression methods are now often preferred, but most existing data is from hypothesis tests, as discussed in section 2-2.1.2.

It will be impossible to have large data sets for all compounds, making it necessary to have some means to set limits for pesticides with limited data. The ACR and AF methods are derived from richer pesticide data sets to use as estimates for compounds without much data. However most of the richer pesticide data sets available now do not reflect many of compounds currently in use. The default ACR would benefit from the generation and incorporation of more multispecies pesticide ACRs, making the default ACR a better representative of currently used pesticides. Similarly, the assessment factors would benefit from the addition of more acute pesticide data sets. To accommodate additional data, this chapter includes the procedure used to derive ACRs and AFs, for their recalculation as more data become available.

Given the frequent data limitations, criteria reports would benefit from periodic review and incorporation of the most recent literature.

2-8.0 Guideline format

As discussed in Phase I (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009), the methodology must be understandable, navigable and usable by environmental managers.

The new methodology, presented in Chapter 3 of this Phase II report, includes a flow chart outlining the derivation procedure, explicit guidance and instructions for each step of the process, details of calculations, and numerous tables of information. For clarity, background and supporting information have been separated from the methodology itself. The rationale for selecting various components and approaches is discussed in the Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) and in this chapter (Chapter 2) of the Phase II report. A complete criteria derivation is also included in Chapter 4 of the Phase II report to demonstrate application of the methodology.

2-9.0 Conclusion

Each aspect of criteria derivation identified in the Phase I report (TenBrook & Tjeerdema 2006, TenBrook *et al.* 2009) as potential components of the new methodology has been evaluated in more detail. Methods have been selected for data collection, evaluation, and reduction. The SSD approach utilized in the Australia/New Zealand criteria derivation guidelines (ANZECC & ARMCANZ 2000) was selected for use in the new methodology, but with the modification that it may be used to derive acute or chronic criteria when at least 5 toxicity data representing 5 different families are available.

To arrive at the acute criterion the 5th percentile values derived from acute SSDs is divided by two, while 5th percentile values derived from chronic data would become chronic criteria. For smaller acute data sets, an AF method similar to that in the Great Lakes methodology (USEPA 2003d) is included, but new factors, specific to pesticides, have been developed. In cases when fewer than 5 chronic data are available, chronic criteria are to be derived from acute criteria divided by an acute-to-chronic ratio.

Criteria are stated in terms of magnitude, duration, and frequency. For acute criteria a 1-h averaging period is established, while for chronic criteria a 4-d averaging period is established. For both acute and chronic criteria, an allowable frequency of criteria exceedance of once-in-three-years is established.

To determine compliance with derived criteria, guidance is given for how to account for bioavailability, pesticide mixture effects, and water quality effects. Methods are also presented for evaluating criteria in terms of their ability to protect species (including threatened and endangered species) within ecosystems, and their potential for partitioning into other environmental media.

2-10.0 References

- Aldenberg T, Jaworska JS. 2000. Uncertainty of the hazardous concentration and fraction affected for normal species sensitivity distributions. *Ecotoxicol Environ Saf* 46:1-18.
- Aldenberg T, Slob W. 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. *Ecotoxicol Environ Saf* 25:48-63.

- Alvarez DA, Petty JD, Huckins JN, Jones-Lepp TL, Getting DT, Goddard JP, Manahan SE. 2004. Development of a passive, in situ, integrative sampler for hydrophilic organic contaminants in aquatic environments. *Environ Toxicol Chem* 23:1640-1648.
- Anderson JW, Moore LJ, Blaylock JW, Woodruff DL, Keissa SL. 1977. Bioavailability of sediment-sorbed naphthalenes to the sipunculid worm, *Phascolosoma agassizii*. In: Wolfe DA, editor. *Fate and Effects of Petroleum Hydrocarbons in Marine Ecosystems and Organisms*. Elmsford, NY: Pergamon Press. p 275-285.
- ANZECC, ARMCANZ. 2000. Australian and New Zealand guidelines for fresh and marine water quality. Report Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australia.
- Asfaw, A., M.R. Ellersieck, and F.L. Mayer. 2003. Interspecies correlation estimations (ICE) for acute toxicity to aquatic organisms and wildlife. II. User manual and software. U.S. Environmental Protection Agency Report No. EPA/600/R-03/106, Washington, DC. 20 p. + software.
- ASTM. 2002. Standard guide for conducting acute toxicity tests on test materials with fishes, macroinvertebrates, and amphibians. American Society for Testing and Materials. E 729-96.
- Barata C, Solayan A, Porte C. 2004. Role of B-esterases in assessing toxicity of organophosphorus (chlorpyrifos, malathion) and carbamate (carbofuran) pesticides to *Daphnia magna*. *Aquat Toxicol* 66:125-139.
- Barry MJ, Logan DC, Vandam RA, Ahokas JT, Holdway DA. 1995a. Effect of Age and Weight-Specific Respiration Rate on Toxicity of Esfenvalerate Pulse-Exposure to the Australian Crimson-Spotted Rainbow Fish (*Melanotaenia-Fluviatilis*). *Aquat Toxicol* 32:115-126.
- Barry MJ, Ohalloran K, Logan DC, Ahokas JT, Holdway DA. 1995b. Sublethal Effects of Esfenvalerate Pulse-Exposure on Spawning and Non-Spawning Australian Crimson-Spotted Rainbowfish (*Melanotaenia-Fluviatilis*). *Arch Environ Contam Toxicol* 28:459-463.
- Belden JB, Lydy MJ. 2006. Joint toxicity of chlorpyrifos and esfenvalerate to fathead minnows and midge larvae. *Environ Toxicol Chem* 25:623-629.
- Borthwick PW, Clark JR, Montgomery RM, Patrick JM Jr, Loes EM. 1985. Field confirmation of a laboratory-derived hazard assessment of the acute toxicity of fenthion to pink shrimp, *Penaeus duorarum*. In: *Aquatic Toxicology and Hazard Assessment: Eighth Symposium. ASTM STP 891*, Bahner RC, Hansen DJ, eds. American Society of Testing and Materials, Philadelphia, PA, pp. 177-189.
- Brannon JM, Pennington JC, Davis WM, Hayes C. 1995. Fluoranthene K-DOC in Sediment Pore Waters. *Chemosphere* 30:419-428.
- Brown MD, Carter J, Thomas D, Purdie DM, Kay BH. 2002. Pulse-exposure effects of selected insecticides to juvenile Australian crimson-spotted rainbowfish (*Melanotaenia duboulayi*). *J Econ Entomol* 95:294-298.
- Bruce RD, Versteeg DJ. 1992. A statistical procedure for modeling continuous toxicity data. *Environ Toxicol Chem* 11: 1485-1494.

- Burgess RM, Pelletier MC, Gundersen JL, Perron MM, Ryba SA. 2005. Effects of different forms of organic carbon on the partitioning and bioavailability of nonylphenol. *Environ Toxicol Chem* 24:1609-1617.
- Burns LA. 2004. Exposure analysis modeling system (EXAMS): User manual and system documentation, Revision G. Report US Environmental Protection Agency, Washington, DC.
- Burr IW. 1942. Cumulative frequency functions. *Annals of Mathematical Statistics* 13:215-232.
- Cairns J. 1990. Lack of Theoretical Basis for Predicting Rate and Pathways of Recovery. *Environ Manage* 14:517-526.
- Cairns JJ, Dickson KL. 1977. Recovery of streams from spills of hazardous materials. In: Cairns JJ, Dickson KL, Herricks EE, editors. *Recovery and restoration of damaged ecosystems*. Charlottesville, VA: University Press of Virginia. p 24-42.
- California DPR. 2005. Registration Desk Manual. In: Regulation CDoP, editor. Report California Department of Pesticide Regulation, Sacramento, CA.
- Campbell E, Palmer MJ, Shao Q, Warne M, Wilson D. 2000. BurliOZ: A computer program for calculating toxicant trigger values for the ANZECC and ARMCANZ water quality guidelines, Perth, WA. Report.
- CCME. 1999. A protocol for the derivation of water quality guidelines for the protection of aquatic life. Canadian Environmental Quality Guidelines. Report Canadian Council of Ministers of the Environment, Ottawa.
- CDFG. 1992a. Test No. 61, chronic, chlorpyrifos, *Ceriodaphnia dubia*. Elk Grove, CA: California Department of Fish and Game. Report nr 61.
- CDFG. 1992b. Test No. 133, acute, chlorpyrifos, *Neomysis mercedis*. Elk Grove, CA: California Department of Fish and Game. Report nr 133.
- CDFG. 1992c. Test No. 137, acute, chlorpyrifos, *Ceriodaphnia dubia*. Elk Grove, CA: California Department of Fish and Game. Report nr 137.
- CDFG. 1992d. Test No. 139, acute, chlorpyrifos, *Ceriodaphnia dubia*. Elk Grove, CA: California Department of Fish and Game. Report nr 139.
- CDFG. 1992e. Test No. 142, acute, chlorpyrifos, *Neomysis mercedis*. Elk Grove, CA: California Department of Fish and Game. Report nr 142.
- CDFG. 1992f. Test No. 143, acute, chlorpyrifos, *Neomysis mercedis*. Elk Grove, CA: California Department of Fish and Game. Report nr 143.
- CDFG. 1992g. Test No. 150, acute, chlorpyrifos, *Ceriodaphnia dubia*. Elk Grove, CA: California Department of Fish and Game. Report nr 150.
- CDFG. 1992h. Test No. 157, acute, diazinon, *Ceriodaphnia dubia*. Elk Grove, CA: California Department of Fish and Game. Report nr 157.
- CDFG. 1992i. Test No. 162, acute, diazinon, *Neomysis mercedis*. Elk Grove, CA: California Department of Fish and Game. Report nr 162.
- CDFG. 1992j. Test No. 163, acute, diazinon, *Ceriodaphnia dubia*. Elk Grove, CA: California Department of Fish and Game. Report nr 163.
- CDFG. 1992k. Test No. 168, acute, diazinon, *Neomysis mercedis*. Elk Grove, CA: California Department of Fish and Game. Report nr 168.
- CDFG. 1998a. Test No. 122, acute, diazinon, *Ceriodaphnia dubia*. Elk Grove, CA: California Department of Fish and Game. Report nr 122.

- CDFG. 1998b. Test No. 132, acute, diazinon, *Physa* spp. Elk Grove, CA: California Department of Fish and Game. Report nr 132.
- Chapman PM, Fairbrother A, Brown D. 1998. A critical evaluation of safety (uncertainty) factors for ecological risk assessment. *Environ Toxicol Chem* 17:99-108.
- Charles JR. 1958. Final report on population manipulation studies in three Kentucky streams. *Proceedings of the Southeastern Association of Game and Fish Commissioners* 11:155-185.
- Chiou CT, Kile DE, Brinton TI, Malcolm RL, Leenheer JA, Maccarthy P. 1987. A Comparison of Water Solubility Enhancements of Organic Solutes by Aquatic Humic Materials and Commercial Humic Acids. *Environ Sci Technol* 21:1231-1234.
- Chiou CT, Malcolm RL, Brinton TI, Kile DE. 1986. Water Solubility Enhancement of Some Organic Pollutants and Pesticides by Dissolved Humic and Fulvic-Acids. *Environ Sci Technol* 20:502-508.
- CIBA-GEIGY. 1987. Static acute toxicity of diazinon AG500 to bluegill (*Lepomis macrochirus*), EPA guidelines no. 72-1. Report for study conducted by Springborn Life Sciences, Inc., Wareham, MA for CIBA-GEIGY Corporation. Greensboro, NC.
- Cold A, Forbes VE. 2004. Consequences of a short pulse of pesticide exposure for survival and reproduction of *Gammarus pulex*. *Aquat Toxicol* 67:287-299.
- Cook SF, Moore RL. 1969. The effects of rotenone treatment on the insect fauna of a California stream. *Trans Am Fish Soc* 98:539-544.
- Corbet PS. 1958. Some effects of DDT on the fauna of the Victoria Nile. *Rev Zool Bot Afr* 57:73-95.
- Cornelissen G, Breedveld GD, Naes K, Oen AMP, Ruus A. 2006. Bioaccumulation of native polycyclic aromatic hydrocarbons from sediment by a polychaete and a gastropod: Freely dissolved concentrations and activated carbon amendment. *Environ Toxicol Chem* 25:2349-2355.
- CSIRO. 2001. BurrliOZ v. 1.0.13: Commonwealth Scientific and Industrial Research Organization, Australia.
- Crane M. 1997. Research needs for predictive multispecies tests in aquatic toxicology. *Hydrobiologia* 346: 149-155. Crane M, Attwood C, Sheahan D, Morris S. 1999. Toxicity and bioavailability of the organophosphorus insecticide pirimiphos methyl to the freshwater amphipod *Gammarus pulex* L. in laboratory and mesocosm systems. *Environ Toxicol Chem* 18: 1456-1461.
- CVRWQCB. 2004. The water quality control plan (basin plan) for the California Regional Water Quality Control Board Central Valley Region, fourth edition, the Sacramento and San Joaquin River basins. Report Central Valley Regional Water Quality Control Board, Rancho Cordova, CA.
- Day KE. 1991. Effects of Dissolved Organic-Carbon on Accumulation and Acute Toxicity of Fenvalerate, Deltamethrin and Cyhalothrin to *Daphnia-Magna* (Straus). *Environ Toxicol Chem* 10:91-101.
- Delle Site A. 2001. Factors affecting sorption of organic compounds in natural sorbent/water systems and sorption coefficients for selected pollutants. A review. *Journal of Physical and Chemical Reference Data* 30:187-439.

- Denton DL, Fox JF, Fulk FA. 2003. Enhancing toxicity performance by using a statistical criterion. *Environ Toxicol Chem* 22:2323-2328.
- Denton D, Norberg-King TJ. 1996. Whole effluent toxicity statistics: a regulatory perspective. In: Grothe DR, Dickson KL, Reed-Judkins DK, editors. *Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts*: SETAC Press. p 83-102.
- Dermott RM, Spence HJ. 1984. Changes in Populations and Drift of Stream Invertebrates Following Lampricide Treatment. *Can J Fish Aquat Sci* 41:1695-1701.
- DiToro DM, Zarba CS, Hansen DJ, Berry WJ, Swartz RC, Cowan CE, Pavlou SP, Allen HE, Thomas NA, Paquin PR. 1991. Technical Basis for Establishing Sediment Quality Criteria for Nonionic Organic-Chemicals Using Equilibrium Partitioning. *Environ Toxicol Chem* 10:1541-1583.
- Eadie BJ, Morehead NR, Landrum PF. 1990. 3-Phase Partitioning of Hydrophobic Organic-Compounds in Great-Lakes Waters. *Chemosphere* 20:161-178.
- ECB. 2003. Technical guidance document on risk assessment in support of commission directive 93/67/EEC on risk assessment of new notified substances, commission regulation (EC) no. 1488/94 on risk assessment for existing substances, directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market. Part II. Environmental risk assessment. In: European Chemicals Bureau ECJRC, editor. Report European Communities.
- ECOTOX. 2006. ECOTOX code list. Report United States Environmental Protection Agency, Washington D. C.
- Eidt DC. 1981. Recovery of Aquatic Arthropod Populations in a Woodland Stream after Depletion by Fenitrothion Treatment. *Can Entomol* 113:303-313.
- Ellersieck, M.R., A. Asfaw, F.L. Mayer, G.F. Krause, K. Sun, and G. Lee. 2003. Acute-to-chronic estimation (ACE v. 2.0) with time-concentration-effect models: User manual and software. U.S. Environmental Protection Agency Report No. EPA/600/R-03/107, Washington, DC. 26 p. + software.
- Elson PF. 1967. Effects on Wild Young Salmon of Spraying DDT over New Brunswick Forests. *Journal of the Fisheries Research Board of Canada* 24:731-&.
- Emans HJB, Vanderplassche EJ, Canton JH, Okkerman PC, Sparenburg PM. 1993. Validation of Some Extrapolation Methods Used for Effect Assessment. *Environ Toxicol Chem* 12:2139-2154.
- Erickson RJ, Stephan CE. 1988. Calculation of the final acute value for water quality criteria for aquatic organisms. Report Environmental Research Laboratory-Duluth, United States Environmental Protection Agency, Duluth, MN.
- Evans M, Hastings N, Peacock B. 2000. *Statistical Distributions, Third Edition*. New York, NY: John Wiley & Sons, Inc.
- EVS. 1999. A critique of the ANZECC and ARMCANZ (1999) water quality guidelines. Prepared for Minerals Council of Australia and Kwinana Industries Council. Final report. Vancouver, B. C.
- Felsot AS. 2005. A critical analysis of the draft report, "Amendments to the water quality control plan for the Sacramento River and San Joaquin River basins for the control of diazinon and chlorpyrifos runoff into the lower San Joaquin River," (Karkoski et al. 2004) and supporting documents. Prepared for the Central Valley Regional Water Quality Control Board, Sacramento, CA.

- Ferrari A, Venturino A, de D'Angelo AMP. 2004. Time course of brain cholinesterase inhibition and recovery following acute and subacute azinphosmethyl, parathion and carbaryl exposure in the goldfish (*Carassius auratus*). *Ecotoxicol Environ Saf* 57:420-425.
- Finney DJ. 1942. The analysis of toxicity tests on mixtures of poisons. *Ann Appl Biol* 29:82-94.
- Forbes VE, Calow P. 1999. Is the per capita rate of increase a good measure of population-level effects in ecotoxicology? (vol 18, pg 1544, 1999). *Environ Toxicol Chem* 18:2664-2664.
- Forbes VE, Calow P. 2002. Species sensitivity distributions revisited: A critical appraisal. *Hum Ecol Risk Assess* 8:473-492.
- Forbes VE, Cold A. 2005. Effects of the pyrethroid esfenvalerate on life-cycle traits and population dynamics of *Chironomus riparius* - Importance of exposure scenario. *Environ Toxicol Chem* 24:78-86.
- Fox DR. 1999. Setting water quality guidelines--A statistician's perspective. *SETAC News* 19:17-18.
- Fredeen FJH. 1975. Effects of a Single Injection of Methoxychlor Black-Fly Larvicide on Insect Larvae in a 161-Km (100-Mile) Section of North Saskatchewan River. *Can Entomol* 107:807-817.
- Fredeen FJH. 1983. Trends in numbers of aquatic invertebrates in a large Canadian river during four years of black fly larviciding with methoxychlor (Diptera: Simuliidae). *Quaestiones Entomologicae* 19:53-92.
- Garbarini DR, Lion LW. 1986. Influence of the Nature of Soil Organics on the Sorption of Toluene and Trichloroethylene. *Environ Sci Technol* 20:1263-1269.
- Gauthier TD, Seitz WR, Grant CL. 1987. Effects of Structural and Compositional Variations of Dissolved Humic Materials on Pyrene Koc Values. *Environ Sci Technol* 21:243-248.
- Ghetti PF, Gorbi G. 1985. Effects of acute pollution on macroinvertebrates in a stream. *Verhandlungen Internationale Vereinigung für Theoretische und Angewandte Limnologie* 22:2426-2431.
- Giesy JP, Solomon KR, Coats JR, Dixon KR, Giddings JM, Kenaga EE. 1999. Chlorpyrifos: ecological risk assessment in North American aquatic environments. *Reviews of Environmental Contamination and Toxicology* 160:1-129.
- GLEC. 2003. Draft compilation of existing guidance for the development of site-specific water quality objectives in the state of California. Report Great Lakes Environmental Center, Columbus, OH.
- Grist EPM, Leung KMY, Sheeler JR, Crane M. 2002. Better bootstrap estimation of hazardous concentration thresholds for aquatic assemblages. *Environ Toxicol Chem* 21: 1515-1524.
- Grothe DR, Kickson KL, Reed-Judkins DK, eds. 1996. Whole effluent toxicity testing: An evaluation of methods and prediction of receiving system impacts. SETAC Press, Pensacola, FL.
- Halter MT, Johnson HE. 1977. A model system to study the desorption and biological availability of PCB in hydrosols. In: Mayer FL, Hamelink JL, editors. *Aquatic*

- Toxicity and Hazard Evaluation ASTM STP 634*. Philadelphia, PA: American Society for Testing and Materials. p 178-195.
- Hanson ML, Sanderson H, Solomon KR. 2003. Variation, replication, and power analysis of *Myriophyllum* spp. microcosm toxicity data. *Environ Toxicol Chem* 22: 1318-1329.
- Harmon SM, Specht WL, Chandler GT. 2003. A comparison of the daphnids *Ceriodaphnia dubia* and *Daphnia ambigua* for their utilization in routine toxicity testing in the Southeastern United States. *Arch Environ Contam Toxicol* 45:79-85.
- Harrison AD, Rattray EA. 1966. Biological effects of mollusciciding natural waters. *S Afr J Sci* 62:236-241.
- Hastings E, Kittams WH, Pepper JH. 1961. Repopulation by aquatic insects in streams sprayed with DDT. *Ann Entomol Soc Am* 54:436-437.
- Heckmann LH, Friberg N. 2005. Macroinvertebrate community response to pulse exposure with the insecticide lambda-cyhalothrin using in-stream mesocosms. *Environ Toxicol Chem* 24:582-590.
- Hoekstra JA, Van Ewijk. 1993. Alternatives for the no-observed-effect level. *Environ Toxicol Chem* 12: 187-194.
- Hoffman CH, Drooz AT. 1953. Effects of a C-47 airplane application of DDT on fish-food organisms in two Pennsylvania watersheds. *Am Midl Nat* 50:172-188.
- Holdway DA, Barry MJ, Logan DC, Robertson D, Young V, Ahokas JT. 1994. Toxicity of Pulse-Exposed Fenvalerate and Esfenvalerate to Larval Australian Crimson-Spotted Rainbow Fish (*Melanotaenia fluviatilis*). *Aquat Toxicol* 28:169-187.
- Hose GC, Van Den Brink PJ. 2004. Confirming the species-sensitivity distribution concept for endosulfan using laboratory, mesocosm, and field data. *Arch Environ Contam Toxicol* 47:511-520.
- Host GE, Regal RR, Stephan CE. 1995. Analyses of acute and chronic data for aquatic life. Report United States Environmental Protection Agency, Washington, DC.
- Huckins JN, Petty JD, Lebo JA, Almeida FV, Booij K, Alvarez DA, Clark RC, Mogensen BB. 2002. Development of the permeability/performance reference compound approach for in situ calibration of semipermeable membrane devices. *Environ Sci Technol* 36:85-91.
- Huckins JN, Prest HF, Petty JD, Lebo JA, Hodgins MM, Clark RC, Alvarez DA, Gala WR, Steen A, Gale Ret al. 2004. Overview and comparison of lipid-containing semipermeable membrane devices and oysters (*Crassostrea gigas*) for assessing organic chemical exposure. *Environ Toxicol Chem* 23:1617-1628.
- Huckins JN, Tubergen MW, Manuweera GK. 1990. Semipermeable membrane devices containing model lipid: A new approach to monitoring the bioavailability of lipophilic contaminants and estimating their bioconcentration potential. *Chemosphere* 20:533-552.
- Irmer U, Markard C, Blondzik K, Gottschalk C, Kussatz C, Rechenberg B, Schudoma D. 1995. Quality targets for concentrations of hazardous substances in surface waters in Germany. *Ecotoxicol Environ Saf* 32:233-243.
- Jacobi GZ, Degan DJ. 1977. Aquatic macroinvertebrates in a small Wisconsin trout stream before, during and two years after treatment with the fish toxicant antimycin. *US Bureau of Sport Fisheries and Wildlife, Investigations in Fish Control* 81:1-24.

- Jagoë RH, Newman MC. 1997. Bootstrap estimation of community NOEC values. *Ecotoxicol* 6: 293-306.
- Jarvinen AW, Tanner DK, Kline ER. 1988. Toxicity of Chlorpyrifos, Endrin, or Fenvalerate to Fathead Minnows Following Episodic or Continuous Exposure. *Ecotoxicol Environ Saf* 15:78-95.
- Jeffrey KA, Beamish FWH, Ferguson SC, Kolton RJ, Macmahon PD. 1986. Effects of the Lampricide, 3-Trifluoromethyl-4-Nitrophenol (Tfm) on the Macroinvertebrates within the Hyporheic Region of a Small Stream. *Hydrobiologia* 134:43-51.
- JMP. 2004. Statistical discovery software. Version 5.1.2: SAS Institute, Inc., Cary NC.
- Keenleyside MHA. 1959. Effects of spruce budworm control on salmon and other fishes in New Brunswick. *Canadian Fish Culturist* 24:17-22.
- Kooijman S. 1987. A Safety Factor for Lc50 Values Allowing for Differences in Sensitivity among Species. *Water Res* 21:269-276.
- Kukkonen J, Oikari A. 1991. Bioavailability of Organic Pollutants in Boreal Waters with Varying Levels of Dissolved Organic Material. *Water Res* 25:455-463.
- Lange R, Hutchinson TH, Croudace CP, Siegmund F. 2001. Effects of the synthetic estrogen 17 alpha-ethinylestradiol on the life-cycle of the fathead minnow (*Pimephales promelas*). *Environ Toxicol Chem* 20:1216-1227.
- Laor Y, Farmer WJ, Aochi Y, Strom PF. 1998. Phenanthrene binding and sorption to dissolved and to mineral-associated humic acid. *Water Res* 32:1923-1931.
- Lauridsen RB, Friberg N. 2005. Stream macroinvertebrate drift response to pulsed exposure of the synthetic pyrethroid lambda-cyhalothrin. *Environ Toxicol* 20:513-521.
- Lee GH, Eilersieck MR, Mayer FL, Krause GF. 1995. Predicting Chronic Lethality of Chemicals to Fishes from Acute Toxicity Test Data - Multifactor Probit Analysis. *Environ Toxicol Chem* 14:345-349.
- Lepper P. 2000. Towards the derivation of quality standards for priority substances in the context of the Water Framework Directive. Final report of the study contract no. B4-3040/2000/30673/MAR/E1. Fraunhofer-Institute Molecular Biology and Applied Ecology, Munich.
- Liess M, Schulz R. 1999. Linking insecticide contamination and population response in an agricultural stream. *Environ Toxicol Chem* 18:1948-1955.
- Lillebo HP, Shaner S, Carlson D, Richard N. 1988. Water quality criteria for selenium and other trace elements for protection of aquatic life and its uses in the San Joaquin Valley. Technical Committee Report: Regulation of agricultural drainage to the San Joaquin River. Appendix D. Report California State Water Resources Control Board, Sacramento, CA.
- Little JD. 1966. Reclamation of Pine Creek, Tennessee. *Proceedings of the Southeastern Association of Game and Fish Commissioners* 19:302-315.
- Liu WP, Gan JJ, Lee S, Kabashima JN. 2004. Phase distribution of synthetic pyrethroids in runoff and stream water. *Environ Toxicol Chem* 23:7-11.
- Lu Y, Wang Z. 2003. Accumulation of organochlorinated pesticides by triolein-containing semipermeable membrane device (triolein-SPMD) and rainbow trout. *Water Res* 37:2419-2425.

- Lydy MJ, Belden JB, Wheelock CE, Hammock BD, Denton DL. 2004. Challenges in regulating pesticide mixtures. *Ecology and Society* 9(6):1.
- Mackay D. 2001. *Multimedia Environmental Fate Models: The Fugacity Approach, Second Edition*. Boca Raton, FL: Lewis Publishers.
- Maltby L, Blake N, Brock TCM, Van Den Brink PJ. 2005. Insecticide species sensitivity distributions: Importance of test species selection and relevance to aquatic ecosystems. *Environ Toxicol Chem* 24:379-388.
- Maltby L, Brock TCM, Van Den Brink PJ. 2009. Fungicide risk assessment for aquatic ecosystems: importance of interspecific variation, toxic mode of action, and exposure regime. *Environ Sci Technol*. doi 10.1021/es901461c.
- Maund SJ, Hamer MJ, Lane MCG, Farrelly E, Rapley JH, Goggin UM, Gentle WE. 2002. Partitioning, bioavailability, and toxicity of the pyrethroid insecticide cypermethrin in sediments. *Environ Toxicol Chem* 21:9-15.
- Mayer FL, Eilersieck MR, Krause GF, Sun K, Lee G, Buckler DR. 2002. Time-concentration-effect models in predicting chronic toxicity from acute toxicity data. In: Crane M, Newman MC, Chapman PF, Fenlon J, editors. *Risk Assessment with Time to Event Models*. Boca Raton, FL: Lewis Publishers.
- Mayer FL, Krause GF, Buckler DR, Eilersieck MR, Lee GH. 1994. Predicting Chronic Lethality of Chemicals to Fishes from Acute Toxicity Test Data - Concepts and Linear-Regression Analysis. *Environ Toxicol Chem* 13:671-678.
- McCarthy JF, Jimenez BD, Barbee T. 1985. Effect of Dissolved Humic Material on Accumulation of Polycyclic Aromatic-Hydrocarbons - Structure Activity Relationships. *Aquat Toxicol* 7:15-24.
- Meehan WR, Sheridan WL. 1966. Effects of toxaphene on fishes and bottom fauna of Big Kitoi Creek, Afognak Island, Alaska. Fish and Wildlife Service Resource Publication No. 12. Report United States Department of the Interior, Washington, D. C.
- MHSPE. 1994. Intervention values and target values--soil quality standards. In: Protection DoS, editor. Report Directorate-General for Environmental Protection, The Hague, The Netherlands.
- Minckley WL, Mihalick P. 1981. Effects of chemical treatment for fish eradication on stream-dwelling invertebrates. *Journal of the Arizona-Nevada Academy of Science* 16:79-82.
- Menconi M, Beckman J. 1996. Hazard assessment of the insecticide methomyl to aquatic organisms in the San Joaquin river system. Admin. Rep. 96-6. California Department of Fish and Game, Environ. Serv. Div., Rancho Cordova, CA.
- Moore DRJ, Caux PY. 1997. Estimating low toxic effects. *Environ Toxicol Chem* 16: 794-801.
- Morrison BRS. 1977. The effects of rotenone on the invertebrate fauna of three hill streams in Scotland. *Fisheries Management* 8:128-139.
- Mount DR, Ankley GT, Brix KV, Clements WH, Dixon DG, Fairbrother A, Hickey CW, Lanno RP, Lee CM, Munns WR, Ringer RK, Staveley JP, Wood CM, Erickson RJ, Hodson PV. 2003. Effects assessment: Introduction. In: *Reevaluation of the State of the Science for Water-Quality Criteria Development*, Reiley MC, Stubblefield WA, Adams WJ, Di Toro DM, Hodson PV, Erickson RJ, Keating FJ Jr, eds., SETAC Press, Pensacola, FL.

- Moye WC, Luckmann WH. 1964. Fluctuations in Populations of Certain Aquatic Insects Following Application of Aldrin Granules to Sugar Creek Iroquois County Illinois. *J Econ Entomol* 57:318-&.
- Mu XY, LeBlanc GA. 2004. Synergistic interaction of endocrine-disrupting chemicals: Model development using an ecdysone receptor antagonist and a hormone synthesis inhibitor. *Environ Toxicol Chem* 23:1085-1091.
- Nabholz JV. 1991. Environmental-Hazard and Risk Assessment under the United-States Toxic-Substances Control Act. *Sci Total Environ* 109:649-665.
- Naddy RB, Johnson KA, Klaine SJ. 2000. Response of *Daphnia magna* to pulsed exposures of chlorpyrifos. *Environ Toxicol Chem* 19:423-431.
- Neff JM. 1979. *Polycyclic Aromatic Hydrocarbons in the Aquatic Environment*. London: Applied Science. 262 p.
- Newman MC, Ownby DR, Mézin LCA, Powell DC, Christensen TRL, Lerberg SB, Anderson B-A. 2000. Applying species-sensitivity distributions in ecological risk assessment: assumptions of distribution type and sufficient numbers of species. *Ecotoxicol Environ Chem* 19: 508-515.
- Niemi GJ, Devore P, Detenbeck N, Taylor D, Lima A, Pastor J, Yount JD, Naiman RJ. 1990. Overview of Case-Studies on Recovery of Aquatic Systems from Disturbance. *Environ Manage* 14:571-587.
- North Carolina DENR. 2003. "Redbook" surface water and wetland standards. In: Quality DoW, editor. Report North Carolina Department of Environment and Natural Resources.
- OECD. 1995. OECD environment monographs No. 92, OECD environmental health and safety publications, series on testing and assessment, No. 3, guidance document for aquatic effects assessment. Report Organization for Economic Co-operation and Development, Paris.
- Okkerman PC, Van Den Plassche EJ, Slooff W, Van Leeuwen CJ, Canton JH. 1991. Ecotoxicological effects assessment: a comparison of several extrapolation procedures. *Ecotox Environ Safe* 21: 182-193.
- Okkerman PC, Vanderplassche EJ, Emans HJB, Canton JH. 1993. Validation of Some Extrapolation Methods with Toxicity Data Derived from Multiple Species Experiments. *Ecotoxicol Environ Saf* 25:341-359.
- Olmstead AW, LeBlanc GA. 2005. Toxicity assessment of environmentally relevant pollutant mixtures using a heuristic model. *Integr Environ Assess Manage* 1:114-122.
- PAN. 2006. Pesticide Action Network Pesticide Database; <http://www.pesticideinfo.org/Index.html>. Report.
- PapeLindstrom PA, Lydy MJ. 1997. Synergistic toxicity of atrazine and organophosphate insecticides contravenes the response addition mixture model. *Environ Toxicol Chem* 16:2415-2420.
- Parsons JT, Surgeoner GA. 1991a. Acute Toxicities of Permethrin, Fenitrothion, Carbaryl and Carbofuran to Mosquito Larvae During Single-Pulse or Multiple-Pulse Exposures. *Environ Toxicol Chem* 10:1229-1233.
- Parsons JT, Surgeoner GA. 1991b. Effect of Exposure Time on the Acute Toxicities of Permethrin, Fenitrothion, Carbaryl and Carbofuran to Mosquito Larvae. *Environ Toxicol Chem* 10:1219-1227.

- Persoone G, Janssen CR. 1994. Field validation of predictions based on laboratory toxicity tests. In: *Freshwater Field Tests for Hazard Assessment of Chemicals*. Hill IR, Heimbach F, Leeuwangh P, Matthiessen P, eds. CRC Press, Boca Raton, FL, pp. 379-397.
- Peterson JL, Jepson PC, Jenkins JJ. 2001. Effect of varying pesticide exposure duration and concentration on the toxicity of carbaryl to two field-collected stream invertebrates, *Calineuria californica* (Plecoptera : Perlidae) and *Cinygma* sp (Ephemeroptera : Heptageniidae). *Environ Toxicol Chem* 20:2215-2223.
- Phillips TA, Summerfelt RC, Wu J, Laird DA. 2003. Toxicity of chlorpyrifos adsorbed on humic colloids to larval walleye (*Stizostedion vitreum*). *Arch Environ Contam Toxicol* 45:258-263.
- Phillips TA, Wu JG, Summerfelt RC, Atchison GJ. 2002. Acute toxicity and cholinesterase inhibition in larval and early juvenile walleye exposed to chlorpyrifos. *Environ Toxicol Chem* 21:1469-1474.
- PHYSPROP. 2006. Physical Properties Database; www.syrres.com/esc/physprop.htm. Report.
- Plackett RL, Hewlett PS. 1952. Quantal Responses to Mixtures of Poisons. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 14:141-163.
- Printes LB, Callaghan A. 2004. A comparative study on the relationship between acetyl cholinesterase activity and acute toxicity in *Daphnia magna* exposed to anticholinesterase insecticides. *Environ Toxicol Chem* 23:1241-1247.
- Reynaldi S, Liess M. 2005. Influence of duration of exposure to the pyrethroid fenvalerate on sublethal responses and recovery of *Daphnia magna* Straus. *Environ Toxicol Chem* 24:1160-1164.
- Rider CV, LeBlanc GA. 2005. An integrated addition and interaction model for assessing toxicity of chemical mixtures. *Toxicol Sci* 87:520-528.
- RIVM. 2001. Guidance document on deriving environmental risk limits in The Netherlands. Report National Institute of Public Health and the Environment.
- Roesijadi G, Anderson JW, Blaylock JW. 1978a. Uptake of Hydrocarbons from Marine-Sediments Contaminated with Prudhoe Bay Crude-Oil - Influence of Feeding Type of Test Species and Availability of Polycyclic Aromatic-Hydrocarbons. *Journal of the Fisheries Research Board of Canada* 35:608-614.
- Roesijadi G, Woodruff DL, Anderson JW. 1978b. Bioavailability of Naphthalenes from Marine-Sediments Artificially Contaminated with Prudhoe Bay Crude-Oil. *Environ Pollut* 15:223-229.
- Rogers HR. 1993. Speciation and Partitioning of Priority Organic Contaminants in Estuarine Waters. *Colloids and Surfaces a-Physicochemical and Engineering Aspects* 73:229-235.
- Rossi SS. Bioavailability of petroleum hydrocarbons from water, sediments and detritus to the marine annelid, *Neanthes arenaceodentata*; 1977; Washington, D. C. American Petroleum Institute. p 621-626.
- Roux DJ, Jooste SHJ, MacKay HM. 1996. Substance-specific water quality criteria for the protection of South African freshwater ecosystems: Methods for derivation and initial results for some inorganic toxic substances. *S Afr J Sci* 92:198-206.

- Samsøe-Petersen L, Pedersen F. 1995. Water quality criteria for selected priority substances. In: Institute WQ, editor. Report Danish Environmental Protection Agency, Copenhagen, Denmark.
- Sanderson H. 2002. Pesticide studies—replication of micro/mesocosm studies. *Environ Sci Pollut R* 6: 429-435.
- Sangster Research Laboratories. 2004. LOGKOW. A databank of evaluated octanol-water partition coefficients (Log P); <http://logkow.cisti.nrc.ca/logkow/index.jsp>. Report Canadian National Committee for CODATA.
- Schnurer Y, Persson P, Nilsson M, Nordgren A, Giesler R. 2006. Effects of surface sorption on microbial degradation of glyphosate. *Environ Sci Technol* 40:4145-4150.
- Schulz R, Liess M. 2000. Toxicity of fenvalerate to caddisfly larvae: chronic effects of 1- vs 10-h pulse-exposure with constant doses. *Chemosphere* 41:1511-1517.
- Schulz R, Liess M. 2001. Toxicity of aqueous-phase and suspended particle-associated fenvalerate: Chronic effects after pulse-dosed exposure of *Limnephilus lunatus* (Trichoptera). *Environ Toxicol Chem* 20:185-190.
- Schwarzenbach RP, Gschwend PM, Imboden DM. 1993. *Environmental Organic Chemistry*. New York, NY: John Wiley & Sons, Inc.
- Segner H. 2005. Developmental, reproductive, and demographic alterations in aquatic wildlife: Establishing causality between exposure to endocrine-active compounds (EACs) and effects. *Acta Hydrochim Hydrobiol* 33:17-26.
- Shen L, Wania F. 2005. Compilation, evaluation, and selection of physical-chemical property data for organochlorine pesticides. *Journal of Chemical and Engineering Data* 50:742-768.
- Siepmann S, Finlayson B. 2000. Water quality criteria for diazinon and chlorpyrifos. Report California Department of Fish and Game.
- Siepmann S, Jones MR. 1998. Hazard assessment of the insecticide carbaryl to aquatic organisms in the Sacramento-San Joaquin river system. Admin. Rep. 98-1. California Department of Fish and Game, Office of Spill Prevention and Response, Rancho Cordova, CA.
- Solomon KR, Giddings JM, Maund SJ. 2001. Probabilistic risk assessment of cotton pyrethroids: I. Distributional analyses of laboratory aquatic toxicity data. *Environ Toxicol Chem* 20:652-659.
- Solomon KR, Takacs P. 2002. Probabilistic risk assessment using species sensitivity distributions. In: Posthuma L, Suter GW, Traas TP, editors. *Species Sensitivity Distributions in Ecotoxicology*. New York, NY: Lewis Publishers. p 285-314.
- Speijers GJA, Franken MAM, Van Leeuwen FXR, Van Egmond HP, Boot R, Loeber JG. 1986. Subchronic oral toxicity study of patulin in the rat. Report 618314001. RIVM Bilthoven, The Netherlands.
- Staples CA, Dickson KL, Rodgers JHJ, Saleh FY. 1985. A model for predicting the influence of suspended sediments on the bioavailability of neutral organic chemicals in the water compartment. In: Cardwell RD, Purdy R, Bahner RC, editors. *Aquatic Toxicology and Hazard Assessment: Seventh Symposium ASTM STP 854*. Philadelphia, PA: American Society for Testing and Materials. p 417-428.

- Steinberg CEW, Sturm A, Kelbel J, Lee SK, Hertkorn N, Freitag D, Kettrup AA. 1992. Changes of Acute Toxicity of Organic-Chemicals to Daphnia-Magna in the Presence of Dissolved Humic Material (Dhm). *Zeitschrift Fur Wasser- Und Abwasser-Forschung-Journal for Water and Wastewater Research-Acta Hydrochimica Et Hydrobiologica*:326-332.
- Steinberg CEW, Xu Y, Lee SK, Freitag D, Kettrup A. 1993. Effect of Dissolved Humic Material (Dhm) on Bioavailability of Some Organic Xenobiotics to Daphnia-Magna. *Chem Spec Bioavail* 5:1-9.
- Stephan CE. 1985. Are the "Guidelines for deriving numerical national water quality criteria for the protection of aquatic life and its uses" based on sound judgments? In: Cardwell RD, Purdy R, Bahner RC, editors. *Aquatic Toxicology and Hazard Assessment: Seventh Symposium, ASTM STP 854*. Philadelphia, PA: American Society for Testing and Materials. p 515-526.
- Stephan CE, Rogers JW. 1985. Advantages of using regression analysis to calculate results of chronic toxicity tests. *Aquatic Toxicology and Hazard Assessment: Eighth Symposium*: American Society for Testing and Materials. p 328-338.
- Sun K, Krause GF, Mayer FL, Eilersieck MR, Basu AP. 1995. Predicting Chronic Lethality of Chemicals to Fishes from Acute Toxicity Test Data - Theory of Accelerated Life Testing. *Environ Toxicol Chem* 14:1745-1752.
- Suter GWI, Barnhouse LW. 1993. Assessment concepts. In: Suter GWI, editor. *Ecological Risk Assessment*. Boca Raton, FL: Lewis Publishers. p 21-47.
- 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.
- Takahashi N, Ito M, Mikami N, Matsuda T, Miyamoto J. 1988. Identification of Reactive Oxygen Species Generated by Irradiation of Aqueous Humic-Acid Solution. *J Pestic Sci* 13:429-435.
- TenBrook PL, Tjeerdema RS. 2006. Methodology for derivation of pesticide water quality criteria for the protection of aquatic life in the Sacramento and San Joaquin River Basins. Phase I: Review of existing methodologies. Final Report. Report Central Valley Regional Water Quality Control Board, Rancho Cordova, CA.
- TenBrook PL, Tjeerdema RS, Hann P, Karkoski J. 2009. Methods for deriving pesticide aquatic life criteria. *Reviews of Environmental Contamination and Toxicology*.199:19-109.
- Torblaa RL. 1968. Effects of lamprey larvicides on invertebrates in streams. Report United States Department of the Interior, Washington, D. C.
- Traas TP, VandeMeent D, Posthuma L, Hamers T, Kater BJ, De Zwart D, Aldenberg T. 2002. The potentially affected fraction as a measure of ecological risk. In: Posthuma L, Suter GWI, Traas TP, editors. *Species Sensitivity Distributions in Ecotoxicology*. New York, NY: Lewis Publishers. p 315-344.
- U. S. Code Title 7. 1947. Federal Insecticide, Fungicide and Rodenticide Act. As amended through 2004 ed. Report.
- USEPA. 1980a. Ambient water quality criteria for aldrin/dieldrin. Report United States Environmental Protection Agency, Washington D. C.
- USEPA. 1980b. Ambient water quality criteria for chlordane, EPA 440/5-80-027. Report United States Environmental Protection Agency, Washington D. C.

- USEPA. 1980c. Ambient water quality criteria for endosulfan, EPA 440/5-80-046. Report United States Environmental Protection Agency, Washington D. C.
- USEPA. 1980d. Ambient water quality criteria for endrin, EPA 440/5-80-047. Report United States Environmental Protection Agency, Washington D. C.
- USEPA. 1980e. Ambient Water Quality Criteria for Heptachlor, EPA 440/5-80-062. Report United States Environmental Protection Agency, Washington D. C.
- USEPA. 1980f. Ambient Water Quality Criteria for Hexachlorocyclohexane, EPA 440/5-80-054. Report United States Environmental Protection Agency, Washington D. C.
- USEPA. 1980g. Ambient Water Quality Criteria for DDT, EPA 440/5-80-038. United States Environmental Protection Agency, Washington D. C.
- USEPA. 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses, PB-85-227049. Report United States Environmental Protection Agency, National Technical Information Service, Springfield, VA.
- USEPA. 1986a. Ambient water quality criteria for chlorpyrifos, EPA 440/5-86-005. Report United States Environmental Protection Agency, Washington, D. C.
- USEPA. 1986b. Ambient Water Quality Criteria for Toxaphene, EPA 440/5-86-006. Report United States Environmental Protection Agency, Washington D. C.
- USEPA. 1991. Technical Support Document for Water Quality-based Toxics Control, EPA/505/2-90-001. Report United States Environmental Protection Agency, Washington D. C.
- USEPA. 1993. 40 CFR 158.490 Wildlife and aquatic organisms data requirements. Report United States Environmental Protection Agency, Washington D. C.
- USEPA. 2002. Understanding and accounting for method variability in whole effluent toxicity application under the National Pollutant Discharge Elimination System Program. EDs: Denton DL, Fox J, Fulk FA, Greenwald K, Narvaez M, Norberg-King TJ, Phillips L. EPA/833R-00-003. Office of Water. Washington DC.
- USEPA. 2002. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms, 4th edition, EPA-821-R-02-013. Report United States Environmental Protection Agency, Washington, D. C.
- USEPA. 2003a. 2003 Draft update of ambient water quality criteria for copper, EPA 822-R-03-026. Report United States Environmental Protection Agency, Washington D. C.
- USEPA. 2003b. Ambient aquatic life water quality criteria for tributyltin (TBT)--Final, EPA 822-R-03-031. Report United States Environmental Protection Agency.
- USEPA. 2003c. Draft ambient life water quality criteria for atrazine, EPA-822-R-03-023. In: Office of Water OoSaT, Health and Ecological Criteria Division, editor. Report United States Environmental Protection Agency, Washington D. C.
- USEPA. 2003d. Water quality guidance for the Great Lakes system. *Federal Register* 40.
- USEPA. 2005a. Aquatic life ambient water quality criteria, diazinon, final, EPA-822-R-05-006. Report United States Environmental Protection Agency, Washington D. C.
- USEPA. 2005b. Science Advisory Board consultation document, proposed revisions to aquatic life guidelines, water-based criteria. In: Subcommittee W-bC, editor. Report United States Environmental Protection Agency, Washington D. C.

- USGS. 2000. SPMD Basics. Report USGS Columbia Environmental Research Center.
- USGS. 2005a. Water resources data, California water year 2004, Volume 3, Southern Central Valley basins and the Great Basin from Walker River to Truckee River. Report United States Geological Survey, Sacramento, CA.
- USGS. 2005b. Water resources data, California water year 2004, Volume 4, Northern Central Valley basins and the Great Basin from Honey Lake basin to Oregon state line. Report United States Geological Survey, Sacramento, CA.
- Van den Berg M, Birnbaum L, Bosveld ATC, Brunstrom B, Cook P, Feeley M, Giesy JP, Hanberg A, Hasegawa R, Kennedy SW *et al.* 1998. Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife. *Environ Health Perspect* 106:775-792.
- Van Der Hoeven N, Gerritsen AAM. 1997. Effects of chlorpyrifos on individuals and populations of *Daphnia pulex* in the laboratory and field. *Environ Toxicol Chem* 16:2438-2447.
- Van Der Hoeven N, Noppert F, Leopold A. 1997. How to measure no effect .1. Towards a new measure of chronic toxicity in ecotoxicology, introduction and workshop results. *Environmetrics* 8:241-248.
- Van Der Hoeven N. 2001. Estimating the 5-percentile of the species sensitivity distributions without any assumptions about the distribution. *Ecotoxicol* 10: 25-34.
- Van Straalen NM, Denneman CAJ. 1989. Ecotoxicological Evaluation of Soil Quality Criteria. *Ecotoxicol Environ Saf* 18:241-251.
- Van Straalen NM, Van Leeuwen CJ. 2002. European history of species sensitivity distributions. In: Posthuma L, Suter GWI, Traas TP, editors. *Species Sensitivity Distributions in Ecotoxicology*. Boca Raton, FL: Lewis Publishers, CRC Press. p 19-34.
- Van Vlaardingen PLA, Traas TP, Wintersen AM, Aldenberg T. 2004. ETX 2.0 A program to calculate hazardous concentrations and fraction affected, based on normally distributed toxicity data. Report National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands.
- Versteeg DJ, Belanger SE, Carr GJ. 1999. Understanding single-species and model ecosystem sensitivity: Data-based comparison. *Environ Toxicol Chem* 18:1329-1346.
- Victor R, Ogbeibu AE. 1986. Recolonization of Macrobenthic Invertebrates in a Nigerian Stream after Pesticide Treatment and Associated Disruption. *Environmental Pollution Series a-Ecological and Biological* 41:125-137.
- Wagner C, Lokke H. 1991. Estimation of Ecotoxicological Protection Levels from Noec Toxicity Data. *Water Res* 25:1237-1242.
- Wallace JB, Vogel DS, Cuffney TF. 1986. Recovery of a headwater stream from an insecticide-induced community disturbance. *J N Am Benthol Soc* 5:115-126.
- Warner K, Fenderson OC. 1962. Effects of DDT spraying for forest insects on Maine trout streams. *J Wildl Manage* 26:86-93.
- Wheelock CE, Eder KJ, Werner I, Huang HZ, Jones PD, Brammell BF, Elskus AA, Hammock BD. 2005. Individual variability in esterase activity and CYP1A levels in Chinook salmon (*Oncorhynchus tshawyacha*) exposed to esfenvalerate and chlorpyrifos. *Aquat Toxicol* 74:172-192.

- Whiles MR, Wallace JB. 1995. Macroinvertebrate production in a headwater stream during recovery from anthropogenic disturbance and hydrologic extremes. *Can J Fish Aquat Sci* 52:2402-2422.
- Whitehouse P, Crane M, Grist E, O'Hagan A, Sorokin N. 2004. Derivation and expression of water quality standards; opportunities and constraints in adopting risk-based approaches in EQS setting. Report UK.
- Wu JG, Laird DA. 2004. Interactions of chlorpyrifos with colloidal materials in aqueous systems. *Journal of Environmental Quality* 33:1765-1770.
- Yount JD, Niemi GJ. 1990. Recovery of Lotic Communities and Ecosystems from Disturbance - a Narrative Review of Case-Studies. *Environ Manage* 14:547-569.
- Zabel TF, Cole S. 1999. The derivation of environmental quality standards for the protection of aquatic life in the UK. *Journal of the Chartered Institution of Water and Environmental Management* 13:436-440.
- Zischke JA, Arthur JW, Hermanutz RO, Hedtke SF, Helgen JC. 1985. Effects of Pentachlorophenol on Invertebrates and Fish in Outdoor Experimental Channels. *Aquat Toxicol* 7:37-58.