Attachment 3-2

Guidance for Developing Ecological Soil Screening Levels (Eco-SSLs)

Eco-SSL Standard Operating Procedure (SOP) #2: Plant and Soil Invertebrate Literature Evaluation, Data Extraction, and Eco-SSL Calculation

OSWER Directive 92857-55

November 2003

This page intentionally left blank

TABLE OF CONTENTS

This Page Intentionally Left Blank

1.0 INTRODUCTION

This standard operating procedure (SOP) describes how nine criteria are used for assessing the applicability of published studies for deriving a terrestrial plant or soil invertebrate Ecological Soil Screening Level (Eco-SSL), provides a set of rules for extracting and reporting the most appropriate study data, and presents the method for calculating an Eco-SSL. Only those studies that meet the study acceptance criteria in SOP #1 (Appendix 3-1) should be evaluated and scored using this SOP. This SOP is intended to ensure that the data most appropriate for deriving an Eco-SSL are selected and used.

2.0 EVALUATION AND SCORING OF STUDY ATTRIBUTES

Nine study evaluation criteria are used to score each reported study (Attachment A). Scoring is based on a three-point scale (i.e., 0, 1, or 2) with a score of "2" being the highest score. The scores for each study evaluation criteria are recorded in a Score Sheet spreadsheet (Attachment B) and summed to generate a total evaluation score for each study. A study must receive a total evaluation score greater than ten to be considered for deriving an Eco-SSL.

Users of this SOP should recognize that toxicity studies reported in published literature were generally not conducted or intended for the purpose of deriving Eco-SSLs. Therefore, the specific information addressed by each criterion may not be reported for each study. Scoring should be objective, however, in some instances professional judgement may be needed to ascertain the appropriate score for a criterion.

Some publications will contain the results of several different experimental designs, or studies, each of which should be scored separately. When a publication reports toxicity data for more than one test species, chemical, or soil type (defined below) these should be considered different studies for the purpose of deriving an Eco-SSL. Studies that vary in other parameters, such as temperature, photoperiod, or species life stage (e.g., immature versus mature), should not be considered different studies for the purpose of deriving an Eco-SSL. Distinct artificial soil types are defined as having different bioavailability scores as determined with study evaluation criteria #1, which is based on pH and percent organic matter. Natural soils from two different sampling locations are considered distinct soils regardless of their bioavailability scores.

When multiple studies are presented in a paper, the reviewer should assign a unique identification code to each study, and document information for each study separately on the Score Sheet. For example, a publication by Jones et al. (Publication No.1022) contains results for three separate experimental designs. In this example, results of each experimental design (i.e., study) should be evaluated and scored separately, and identified on the Score Sheets with an unique experiment identification code such as a, b, and c for the respective publication number (e.g., 1022a, 1022b, 1022c).

A publication may include some studies that do not pass the study acceptance criteria. The reviewer should only score those individual studies that meet the requirements of the study

acceptance criteria (see Appendix 3-1). For example, if a study reports the results of both a topical application and artificial soil study, the topical application study (which does not meet the study acceptance criteria) would not be scored. Reviewers should provide comments on which studies were scored and which were excluded. These comments should be entered in the "comment" field of the Critical Notes spreadsheet (see Section 3.0).

3.0 DATA EXTRACTION

For each study reviewed, a set of Critical Notes (Figure 1) are recorded in a spreadsheet (Attachment C). As with the Study Evaluation Criteria Score Sheet, individual studies are assigned separate experiment identification codes.

Details on the soil parameters including soil pH and percent organic matter (%OM) are recorded. If a study reports the pH at both test initiation and completion only the initial pH should be recorded. If a pH range is reported the arithmetic mean of the minimum and maximum should be calculated and reported. However, if a range is reported and it includes a pH value that is outside of the acceptable soil parameters (i.e., pH < 4 or > 8.5), this study should be rejected and this information should be noted in the comment field of the Critical Notes. Likewise, if %OM is reported as a range for a single soil type and the range extends outside of the acceptable range (i.e., >10%), the study should be rejected and not used for deriving an Eco-SSL. This information should be noted in the comment field of the Critical Notes.

Figure 1. Critical Notes

- Paper identification code
- Study identification code
- First author and year of publication
- Chemical
- Receptor
- Common name
- Species name
- Soil pH
- Percent organic matter (OM)
- Bioavailability score
- Total evaluation score
- Ecologically relevant endpoint (ERE)
- Preferred toxicity parameter
- Preferred toxicity value
- Secondary toxicity parameter
- Secondary toxicity value
- Other available toxicity parameters and concentrations
- Preference level
- Comments

Toxicity values are reported in the Critical Notes spreadsheet. Toxicity values are chemical concentrations related to measurements of a defined ecologically relevant endpoint (ERE). The EREs are presented in Table 1. Toxicity values should be reported as milligram per kilogram (mg/kg) dry weight of the chemical in soil. If the concentrations are reported in units other than mg/kg, or are reported as the concentration of a salt, the reviewer should convert these values to mg/kg of the chemical. If the concentrations are reported as wet weight of the chemical in soil, they should be converted to dry weight concentration based on the percent soil moisture reported in the publication. Any calculations or assumptions by the reviewer must be noted in the comment field of the Critical Notes. If the toxicity value is reported as a range of concentrations rather than a point estimate, no value should be recorded on the Critical Notes and a notation should be made in the comment field.

Table 1. Ecologically Relevant Endpoints (ERE) for Deriving Eco-SSLs	
Ecologically Relevant Endpoint	**Definition**
Reproduction	Measures of the effect of toxicants on offspring production. Examples of EREs associated with reproduction include changes in fecundity, number of progeny produced (eggs, cocoons, etc.), rate of reproduction (hatching rates, etc.), rate of maturation, sexual development, change in sex expression, and sterility number or proportion of abnormal progeny.
Population	Measurements and endpoints regarding a group of soil invertebrates of the same species occupying the same area at a given time. Measurement includes population dynamics. Examples of EREs associated with population include changes in size and age class structures, changes in sex ratio, intrinsic population growth rate, survivability of subsequent generations, diversity, evenness, index to population size (count, number, abundance), life table data, population density (number/area).
Growth	Broad category which encompasses measures of weight/mass and length. EREs associated with growth and development responses such as a change in body weight.
Biomass (plants only)	Measurement of plant products including seed production, standing corp biomass, seedling emergence, shoot length/growth, root elongation/growth, fresh or dry mass, or yield.
Physiology (plants only)	For the purposes of developing Eco-SSLs, only plant studies reporting EREs associated with physiological responses will be used. Physiological endpoints for plants include net photosynthesis (CO_2 uptake, oxygen release), decrease in chlorophyll content or chlorophyll fluorescence, increased deformation, membrane damage, desiccation/decrease in water content, detrimental changes in dormancy measures, decreased flowering, and increased senescence.

Toxicity values are categorized in the Critical Notes by standard measurements of dose-response relationships, or toxicity parameters. Acceptable toxicity parameters include No Observed Adverse Effect Concentrations (NOAEC), Lowest Observed Adverse Effect Concentrations (LOAEC), or Effect Concentrations for 10 - 20% of the population (EC_{10} and EC_{20}). Acute toxicity data (e.g., LC_x) and, $EC_{<10}$ or $EC_{>20}$ are not considered acceptable toxicity parameters for deriving Eco-SSLs. If these are the only parameters reported for a study this information should be recorded in the comment field.

If a publication does not identify acceptable toxicity parameters, but sufficient data are provided, the reviewer should determine and record the toxicity value under the appropriate toxicity parameter. For example, if a study does not identify LOAECs and NOAECs but treatments with and without significant adverse effects are reported, the reviewer should record these toxicity values as LOAECs and NOAECs and note in the comment field that these toxicity parameters were assigned by the reviewer, instead of reported in the publication.

When a study provides NOAEC and LOAEC values, these data are used to calculate a Maximum Acceptable Threshold Concentration (MATC). The MATC is the geometric mean of the NOAEC and LOAEC values:

$$\mathbf{GM = exp(average(LnY_1, Y_2, Y_3 \ldots Y_n))}$$

If data for more than one toxicity parameter are reported for a study (e.g., EC_{20} and EC_{10}), all acceptable toxicity data are reported on the Critical Notes Form and a single preferred toxicity parameter is selected for Eco-SSL derivation based on the following hierarchy:

$$\mathbf{EC_{20} > MATC > EC_{10}}$$

If a study reports more than one toxicity value for the same type of toxicity parameter, a preferred toxicity value for invertebrates is selected according to the following hierarchy of EREs:

Reproduction > Population > Growth

For plants the most sensitive measurement of biomass production is selected. If no measurement of biomass production is reported some physiological endpoints are acceptable EREs for plants. If a publication reports multiple "preferred" toxicity values for the same study (e.g., two reproductive EC_{20} values), the lowest value is recorded on the Critical Notes.

4.0 ECO-SSL DERIVATION

The first step in deriving an Eco-SSL is to sort the studies by their literature evaluation score. Studies with a total evaluation score $\leq$10 (out of 18 possible points) are removed from further consideration for deriving an Eco-SSL. Studies that receive an evaluation score >10 are then ranked by bioavailability score. The Eco-SSL is calculated as the geometric mean of the toxicity values at the highest bioavailability score for which sufficient data exists ($\geq$3 data points). For example, if there are two high bioavailability data points (score = 2) and five medium bioavailability data points (score = 1) the Eco-SSL would be calculated as at the medium bioavailability score using all seven data points. If there are less than three acceptable data points an Eco-SSL can not be calculated.

An Eco-SSL is calculated by taking the geometric mean (GM) of the preferred toxicity values (where $N \geq 3$) with the highest bioavailability scores:

$$\mathbf{GM = exp(average(LnY_1\ Y_2))}$$

Where Y_1 = NOAEC and Y_2 = LOAEC

5.0 QUALITY ASSURANCE REVIEW

A quality assurance review of the data in studies used to derive an Eco-SSL must be preformed by a panel of experts. The reviewers must verify that all of the studies that will be used to derive an Eco-SSL were correctly evaluated and scored. All publications are checked by at least two individuals, and then reported to the panel of experts for final quality assurance evaluation. The Quality Assurance review consists of the following multi-step process:

- The Literature Acceptance Criteria Checklist (SOP# 1) are reviewed to verify that all of the Acceptance Criteria were met.
- The study evaluation scores are checked to verify that only studies that scored greater than 10 are considered for deriving Eco-SSLs.
- Each publication is reviewed to ensure that toxicity data for all of the available studies are reported on the Critical Notes.
- The soil pH, OM and bioavailability score for each study is verified.
- Selection of the appropriate ecological endpoints, toxicity parameter(s) and toxicity values are verified.
- The preferred toxicity parameter and toxicity value for each study are verified.
- The units of the toxicity value (wet weight/dry weight, chemical form) are verified.
- Calculations (e.g., converting organic carbon to organic matter) and any assumptions made by the reviewer (e.g., assigning toxicity values where values are not reported) documented in the comment field are verified.

ATTACHMENT A
STUDY EVALUATION CRITERIA

Criteria #1 Testing was Done Under Conditions of High Bioavailability

Bioavailability of metals and polar organic compounds is influenced by pH and soil organic matter. The scoring is intended to favor relatively high bioavailability. If the authors do not present the organic matter content, but present another measure of organic content (total organic carbon, particulate organic carbon, or organic carbon) these measurements are converted to organic matter content by multiplying them by a factor of 1.72. If neither the percent organic matter nor the organic carbon content is provided in the publication but this information is available from the United States Department of Agriculture (USDA) National Cooperative Soil Survey (NRCS) database, the data reported in the soil database may be used.

Natural Soils

Natural soils are scored using one of the three Bioavailability Tables provided below. These tables are the same as those reported in Chapter 2 where very high or high = 2, medium = 1, and low or very low = 0.

QUANTITATIVE BIOAVAILABILITY FOR CATIONIC METALS IN NATURAL SOILS

	Low OM (< 2%)	Medium OM (2 - 6%)	High OM (> 6 -10%)
4 ≤ Soil pH ≤ 5.5	V. High (score = 2)	High (score = 2)	Medium (score = 1)
5.5 < Soil pH < 7	High (score = 2)	Medium (score = 1)	Low (score = 0)
7 ≤ Soil pH ≤ 8.5	Medium (score = 1)	Low (score = 0)	V. Low (score = 0)

QUANTITATIVE BIOAVAILABILITY FOR ANIONIC METALS IN NATURAL SOILS

	Low OM (< 2%)	Medium OM (2 - 6%)	High OM (> 6 - 10%)
4 ≤ Soil pH ≤ 5.5	Medium (score = 1)	Low (score = 0)	V. Low (score = 0)
5.5 < Soil pH < 7	High (score = 2)	Medium (score = 1)	Low (score = 0)
7 ≤ Soil pH ≤ 8.5	V. High (score = 2)	High (score = 2)	Medium (score = 1)

QUANTITATIVE BIOAVAILABILITY FOR ORGANIC CHEMICALS IN NATURAL SOILS

Soil Type	Chemical Type	Organic Matter (%)		
		< 2	2 - 6	> 6 - 10
4 ≤ Soil pH ≤ 5.5	Pesticides/PCBS (Log Koc > 3.5)	High (score = 2)	Medium (score = 1)	Low (score = 0)
	Other Organics (Log Koc < 3.5)	V. High (score = 2)	High (score = 2)	Medium (score = 1)
5.5 < Soil pH < 7	Pesticides/PCBS (Log Koc > 3.5)	Medium (score = 1)	Low (score = 0)	Low (score = 0)
	Other Organics (Log Koc < 3.5)	High (score = 2)	Medium (score = 1)	Low (score = 0)
7 ≤ Soil pH ≤ 8.5	Pesticides/PCBS (Log Koc > 3.5)	Low (score = 0)	Low (score = 0)	Low (score = 0)
	Other Organics (Log Koc < 3.5)	Medium (score = 1)	Low (score = 0)	Low (score = 0)

Standard Artificial Soils

Standard artificial soils (i.e., ASTM, ISO, OECD, i.e., 10% OM, 20% Kaolinite, 69% sand, 1% $CaCO_3$) with pH of 4.0 to 8.5 are assigned a medium bioavailability score of 1. All other artificial soils are scored according to the Bioavailability Tables for natural soils.

Criteria #2 Experimental Designs for Laboratory and Field Studies are Documented and Appropriate

There are two sections that may apply for this criterion. Apply the criteria in 2A when the paper describes laboratory studies. Criteria 2A is used for laboratory studies and Criteria 2B are used when a field study is described. Experimental design can significantly influence the quality of a study. Higher quality studies use experimental design(s) sufficiently robust to allow analysis of the test variables and discriminate non-treatment effects.

Criteria #2A: Experimental Designs for Laboratory Studies

The following scores are applied to laboratory studies. A **Score of 2** is assigned if a standard method or protocol is cited (e.g., US EPA, OECD, ASTM, ISO), or if a standard method is not cited but the study includes a description of the experimental design, the test conditions, and the nature of the test units, as indicated in the following subsections.

Experimental Design. The number of exposure concentrations must be ≥6 including a control, the exposure concentrations (nominal or measured), the number of test organisms per test unit (i.e., loading rate), and the time of observations must be reported in the publication. In addition, if an ANOVA or factorial design was used, there must be at least 3 replicate test units per exposure concentration; or if the study used a regression design, there must be at least two replicates and the toxicity estimate must encompass the range of responses needed to describe the dose-response (e.g., interpolation).

Test Conditions. Test conditions reported in the publication should include, at a minimum: exposure temperature. If it is a plant study, it must also report photoperiod (or conditions, e.g., natural light June-August), and type (e.g. sunlight) or intensity of light.

Test Units. Volume or dimensions, and material comprising the test unit, amount/type of soil in each test unit.

A **Score of 1** is assigned if an analysis of variance (ANOVA) or factorial design is used and the number of exposure concentrations is 4 or 5 including a control, or if number of replicate test units are 2 (duplicates). A Score of 1 is also assigned in the following instances:

- the study uses a regression design and the number of exposure concentrations is 4 or 5 including a control, or ≥6 without replication (i.e., only one test unit per exposure concentration); or
- the reported toxicity estimate (e.g., effect concentration or ECx) encompasses the range of responses needed to describe the dose-response,
- the extrapolation does not exceed 10% of the highest test concentration.
- the conditions described in the preceeding experimental design are met but those in either Test Conditions and/or Test Units are not met.

A **Score of 0** is assigned in all other cases.

Criteria #2B: Experimental Designs for Field Studies

A **Score of 2** is assigned if the study includes a description of the experimental design, the test conditions, and the nature of the test plots, as described in the following subsections.

Experimental Design.. If experimental plots are used, the study should report the number of exposure concentrations, the number of replicate plots per exposure concentration, the location or method of selecting the sampling locations, and the time of

sampling or number of sampling times. If transects are used, the method for selecting the location of the transects, the number of transects, the location or method of selecting the sampling locations along the transects, and the time of sampling, or number of sampling times, should be reported.

Test Conditions. Information on the physico-chemical characteristics of the soil should be reported and, at a minimum, include: soil texture or particle size description (sand, silt, or clay, or some combination thereof), pH, and organic matter content.

Test Plots. The size of the test plots, or length of transects should be reported or they can be cited elsewhere.

A Score of 1 is assigned in the following instances:

- the experimental design is an ANOVA design and has 5 exposure concentrations including controls
- less than 3 replicate test units per exposure concentration are reported,
- a regression design is used with <6 treatments, including a control and no replication,
- the test conditions, the test units, or the test plots, are only partially described or are not reported and are not cited elsewhere.

A Score of 0 is assigned in all other cases.

Criteria #3: Concentration of Test Substance in Soil is Reported

The concentration of the chemical tested must be reported unambiguously. It is unacceptable, for instance to report application rates (e.g., lbs./acre, to 500 ppm in sludge applied at 10 tons per acre). Studies that only report application rates are not accepted and should not be used to derive an Eco-SSL. In some cases, greenhouse studies may report soil mass of pots that would make it possible to convert an application rate to a concentration, however, this is rare. Pot volume alone is not be an adequate parameter to calculate concentrations as one would have to approximate the mass. If the concentrations are reported on a wet weight or fresh weight basis, this is recorded in the Comments field, along with any information that allows conversion to dry weight.

A **Score of 2** is assigned if the measured concentrations are reported. A Score of 1is assigned if the toxicity values are based on nominal concentrations and are used in calculating toxicity values. A **Score of 0** is assigned in all other cases.

Criteria #4: Control Responses are Acceptable

Negative controls are a crucial part of toxicity tests in order to distinguish treatment effects from non-treatments effects. A **Score of 2** is assigned if a standardized procedure is followed and negative control values are within procedural guidelines of the standard procedure cited. A Score of 2 is also assigned if non-standardized procedure are used and the control values are within an acceptable range (e.g., earthworms mortality <10%, plants germination < 20%). A **Score = 1** is assigned f results of control are not reported or are ambiguous. A **Score of 0** is assigned if control results are not within an acceptable range.

Criteria #5: Chronic or Life Cycle Test

Chronic toxicity tests, or those assessing long-term adverse sub-lethal impacts on the life-cycle phases of an organism, are considered superior to acute toxicity tests. **A Score of 2** is applied if chronic exposures, or life-cycle phase studies are used. A **Score of 1** is applied if acute tests are used and a **Score of 0** if very short term exposures are used (i.e., for physiological measurements).

Criteria #6: **Chemical Dosing Procedure is Reported and Appropriate for Chemical and Test**

Chemical dosing procedure may affect the outcome of a test. The type of chemical dosing procedure depends on the chemical and the type of test. Typically dosing procedure should include:

(A) The form or species of the chemical used in the test,
(B) The carrier or vehicle used to deliver the chemical (e.g., solvent, water, etc.)
(C) How the carrier is handled following dosing (i.e., allowed to volatilize, controls, etc.),
(D) How the soil and chemical are mixed to ensure homogeneity.

A Score of 2 is assigned if a study includes information for the dosing procedure including items items A to D (above). A Score of 1 is assigned if information for items A and B are included but not information for items C or D. A **Score of 0** is assigned if the study does not specify details of the dosing procedure, the dosing procedures cannot be inferred, or it does not meet the other scoring criteria.

The evaluator should exercise judgement regarding technical details of all four components (A thru D above), and if questionable or unacceptable methods are used, the scores should be lowered by 1 (i.e., the score becomes either 1 or 0) and the rationale for scoring should be stated in the comment section**.**

Criteria #7: Dose-Response Relationship is Reported or can be Established from Reported Data

A benchmark concentration is intended to represent the location on the dose-response curve that is the threshold between absence and presence of the effects of concern for a relevant ecological endpoint. Two methodologies can be used to identify this benckmark concentration. The first is a method that generates a no observed effect concentration (NOEC) and a lowest observed effect concentration (LOEC). The NOEC is the concentration that did not cause statistically significant effects when compared to controls. The LOEC is the lowest concentration that resulted in statistically significant effects when compared to controls. The threshold lies somewhere between these two values. The second method involves a statistical model to calculate a dose response curve and estimate an effect concentration for some percentage of the population (EC_{xx}), usually between an EC_5 and an EC_{20}. Lethal concentration (LC_{xx}) and $EC_{<10}$ or $EC_{>20}$ are not used for calculating an Eco-SSL and are not scored but the information is recorded on the Critical Notes form. Tests with relatively small upper and lower confidence limits around the NOEC or LOEC and ECx values are preferred. Studies where at least two test concentrations produced adverse effects < 100% are also preferred.

A **Score of 2** is applied if an EC_{10}- EC_{20}; are reported or a NOEC and LOEC within 3x of each other. A **Score of 1** is applied if the difference between the NOEC or LOEC is > 3x but < 10x. A **Score of 0** is applied if a study does not report an ECx; or the difference between the NOEC and LOEC > 10, or only a NOEC *or* LOEC.

Criteria # 8: The Statistical Tests used to Calculate the Benchmark and the Level of Significance were Described

When no observed effect concentrations (NOECs) and lowest observed effect concentrations (LOECs) are reported, an ANOVA or other statistical test should be conducted to determine that the NOEC is the highest test concentration that did not produce a statistically significant effect and the LOEC is the lowest concentration tested that did produce a significant effect when compared to the control. When EC values are reported, the confidence levels around these values should be reported and should be based on a 95% probability level.

A **Score of 2** is applied if the results of the ANOVA or statistical method based on a P = 0.05; or the 95% CI of the ECx are presented. A **Score of 1** is applied if the ANOVA was completed but a P level is not provided, or the P level is > 0.05. If EC data are presented, a **Score of 1** is applied if the 95% CI is not reported or a CI of 90% is used. A **Score of 0** is applied if no

NOEC, LOEC, or EC/LCx data are reported, or if they are reported but no description of the methods used for their calculation are provided.

Criteria #9: The Origin of the Test Organisms is Described

The results of a toxicity test can be influenced by the condition of the test organisms. Test organisms should be healthy and have had no exposure above background to contamination prior to testing.

A Score = 2 is applied if the source and condition of the test organisms are known and described (for seeds unambiguous information should be provided on species identity), and the organisms come from a non-contaminated or commercial source. A **Score of 1** is assigned if the organisms are obtained from a non-commercial source that is not adequately described, or sufficient information is not provided about either the seed stock or the commercial source. A **Score of 0** is applied if organisms are from a known contaminated site, or adequate information is not provided about the seed stock nor the commercial source.

ATTACHMENT B
STUDY EVALUATION CRITERIA SCORE SHEET

Criterion	Title	IP#	IP#	IP#	IP#	IP#	IP#
1	Testing was Done Under Conditions of High Bioavailability (See Soil Evaluation Matrix).						
2	Experimental Designs were Documented and Appropriate.						
3	Concentration of Substance of Interest in Soil was Reported.						
4	Control Measures were Applied.						
5	Chronic or Life Cycle Test was Used.						
6	Chemical Dosing Procedure was Reported and Appropriate for Chemical and Test.						
7	A Dose-Response Relationship is Reported or can be Estimated from Reported Data.						
8	The Statistical Tests used to Calculate the Benchmark and the Levels of Significance were Described.						
9	The Origin of the Test Organisms were Described.						
Total Score							

ATTACHMENT C
CRITICAL NOTES SPREADSHEET

Chemical	Receptor	IP Number	Study ID	First Author, Pub. Year	Species Common Name	Total Eval. Score	Bio-availability Score Soil	pH	OM	ERE	EC20	NOAEC	LOAEC	MATC	EC10	EC50	Level	COMMENTS	