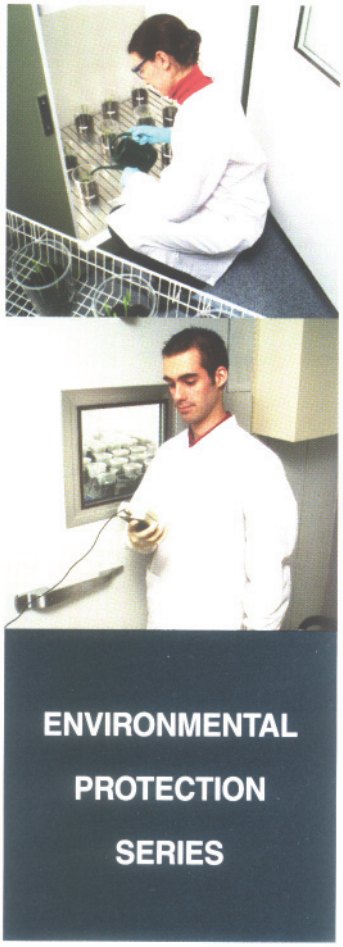
EPS 1/RM/45 – February 2005 (with June 2007 amendments)
Method Development and Applications Section
Environmental Technology Centre
Environment Canada



Biological Test Method: Test for Measuring Emergence and Growth of Terrestrial Plants Exposed to Contaminants in Soil

Environment Canada Environnement Canada

Canada

Biological Test Method:
Test for Measuring Emergence and Growth of Terrestrial Plants Exposed to Contaminants in Soil

Method Development and Applications Section
Environmental Technology Centre
Environment Canada
Ottawa, Ontario

Report EPS 1/RM/45
February 2005 (with June 2007 amendments)

Library and Archives Canada Cataloguing in Publication Data

Main entry under title :

Biological test method. Test for measuring emergence and
growth of terrestrial plants exposed to contaminants in soil

Report ; EPS 1/RM/45
Issued also in French under title: Méthode d'essai biologique.
Essai de mesure de la levée et de la croissance de plantes terrestres
exposés à des contaminants dans le sol.
Includes bibliographical references: p.

ISBN 0-662-40284-7
Cat. no. En49-7/1-45E

1. Growth (Plants) – Regulation – Testing – Methodology.
2. Plants – Effect of pollution on.
3. Soil pollution – Measurement.
4. Toxicity testing – Methodology
I. Canada. Environment Canada
II. Environmental Technology Centre (Canada). Method Development and Applications Section.
III. Series.: Report (Canada. Environment Canada) ; EPS 1/RM/45

QK745.B56 2005 628.1'61 C2005-980121-2

ISBN 0-662-40284-7
Cat No. En49-7/1-45E

Readers' Comments

Comments regarding the content of this report should be addressed to:

Richard Scroggins
Chief, Biological Methods Division
Environmental Technology Centre
Environment Canada
335 River Road
Ottawa, Ontario
K1A 0H3

Cette publication est aussi disponible en français. Pour l'obtenir, s'adresser à:

Publications de la Protection de l'environnement
Environnement Canada
Ottawa (Ontario)
K1A 0H3

Review Notice

This report has been reviewed by the staff of the Environmental Technology Advancement Directorate, Environment Canada, and approved for publication. Mention of trade names or commercial products does not constitute endorsement by Environment Canada for use. Other products of similar value are available.

Abstract

This document provides detailed procedures, conditions, and guidance for preparing for and conducting a biological test for measuring soil toxicity using terrestrial plants. Twelve species options are provided and include: alfalfa (Medicago sativa), *barley* (Hordeum vulgare), *blue grama grass* (Bouteloua gracilis), *carrot* (Daucus carota), *cucumber* (Cucumis sativus), *durum wheat* (Triticum durum), *lettuce* (Lactuca sativa), *northern wheatgrass* (Elymus lanceolatus), *radish* (Raphanus sativus), *red clover* (Trifolium pratense), *red fescue* (Festuca rubra), *or tomato* (Lycopersicon esculentum). *The test is a 14- or 21-day test for effects on seedling emergence and plant growth (measured as shoot and root length and shoot and root dry mass). The method is conducted as a static (i.e., no renewal) test, using one or more samples of contaminated or potentially contaminated soil or one or more concentrations of chemical(s) or chemical product(s) spiked in negative control (or other) soil. Water is added to the test vessels to hydrate soils for the duration of the test.*

The test is conducted at a mean temperature of 24 ± 3 °C in 1-L polypropylene test vessels containing a measured wet weight equivalent to a volume of ~500 mL of test soil. Five or ten seeds (i.e., number of seeds per test vessel is species-specific) are placed into each replicate test vessel. This test uses ≥5 replicated test vessels/treatment for a single-concentration test, and 3–6 replicated test vessels/treatment for a multi-concentration test. The options for test design in a multi-concentration test include an equal number of replicates per treatment (i.e., ≥4) or unequal replicates per treatment (i.e., six per treatment for each negative and other control; four replicates for each of the lowest 4–6 test concentrations; and three replicates for each of the highest five test concentrations). Following a 14- or 21-day exposure (i.e., test duration is species-specific), the number of emerged seedlings in each replicate and each treatment is determined, and the mean percent emergence for each treatment is then compared. Additionally, the shoot and root lengths and the shoot and root dry weights of individual plants surviving in each replicate are determined, and the treatment means compared.

General or universal conditions and procedures are outlined for test preparation and performance. Additional conditions and procedures are stipulated that are specific to the intended use of each test. The biological test method described herein is suitable for measuring and assessing the toxicity of samples of field-collected soil, biosolids, sludge, or similar particulate material; or of natural or artificial soil spiked (mixed) in the laboratory with test chemical(s) or chemical product(s). Instructions and requirements are included on test facilities, sample collection, handling and storing samples, seed source, seed storage and handling, preparing soil or spiked-soil mixtures and initiating tests, specific test conditions, appropriate observations and measurements, endpoints and methods of calculation, and the use of a reference toxicant.

Résumé

Le présent document renferme des indications précises et décrit en détail les procédures et conditions applicables à la préparation et à la conduite d'un essai biologique visant à mesurer la toxicité d'un sol pour des plantes terrestres. Les douze espèces végétales pouvant être utilisées sont les suivantes : luzerne (Medicago sativa)*, orge* (Hordeum vulgare), *boutelou gracieux* (Bouteloua gracilis)*, carotte* (Daucus carota), *concombre* (Cucumis sativus)*, blé dur* (Triticum durum)*, laitue* (Lactuca sativa)*, agropyre du Nord* (Elymus lanceolatus), *radis* (Raphanus sativus)*, trèfle violet* (Trifolium pratense)*, fétuque rouge* (Festuca rubra)*, tomate* (Lycopersicon esculentum). *L'essai, d'une durée de 14 ou de 21 jours, permet de mesurer les effets sur la levée des plantules et la croissance des plantes (d'après la longueur des pousses et des racines et d'après leur masse sèche). Il s'agit d'un essai sans renouvellement faisant appel à un échantillon ou plus de sol contaminé ou susceptible d'être contaminé, ou encore à une concentration ou plus d'au moins une substance ou un produit chimique que l'on mélange avec un sol témoin négatif (ou autre). De l'eau est ajoutée aux récipients d'essai afin d'hydrater le sol pendant la durée de l'essai.*

L'essai est mené à une température moyenne de 24 ± 3 °C; les récipients d'essai en polypropylène, d'une capacité de 1 L, contiennent l'équivalent mesuré (masse humide) de ~500 mL de sol d'essai. Cinq ou dix graines (le nombre de graines par récipient dépend de l'espèce) sont placées dans chaque récipient de répétition. Dans le cas d'un essai à concentration unique, on utilise ≥5 récipients de répétition par traitement; s'il s'agit d'un essai à concentrations multiples, 3–6 récipients de répétition sont employés par traitement. Les options présentées pour le plan d'expérience d'un essai à concentrations multiples incluent un nombre égal de répétitions par traitement (soit ≥4) ou un nombre inégal de répétitions par traitement (soit 6 par traitement pour chaque témoin négatif ou autre; 4 répétitions pour chacune des 4–6 concentrations expérimentales les plus faibles; 3 répétitions pour chacune des 5 concentrations expérimentales les plus élevées). Après une exposition de 14 ou de 21 jours (la durée de l'essai dépend de l'espèce), on détermine le nombre de plantules levées dans chaque répétition et chaque traitement, puis on compare le pourcentage moyen d'émergence pour chaque traitement. On détermine aussi tant la longueur que la masse sèche des pousses et des racines de chaque plante ayant survécu dans chaque répétition, puis on compare les moyennes obtenues.

Le présent document décrit les procédures et conditions générales ou universelles applicables à la préparation et à la conduite de l'essai. Il renferme aussi une description des autres procédures et conditions propres à l'usage prévu des résultats de chaque essai. La méthode d'essai biologique présentée ici convient à la mesure et à l'évaluation de la toxicité d'échantillons de sol, de biosolides, de boue ou de matériau particulaire semblable recueillis sur le terrain, ou encore de sol naturel ou artificiel enrichi, c'est-à-dire mélangé en laboratoire avec une substance ou un produit chimique d'essai ou plus. Des instructions et des exigences sont incluses sur les éléments suivants : installations d'essai; prélèvement, manipulation et entreposage des échantillons; source, entreposage et manipulation des graines; préparation du sol ou des mélanges de sol enrichi; mise en route de l'essai; conditions propres à l'essai; observations et mesures pertinentes; paramètres et méthodes de calcul; utilisation d'un toxique de référence.

Foreword

*This is one of a series of **recommended methods** for measuring and assessing the toxic effect(s) on single species of terrestrial or aquatic organisms, caused by their exposure to samples of toxic or potentially toxic substances or materials under controlled and defined laboratory conditions. Recommended methods are those that have been evaluated by Environment Canada (EC), and are favoured:*

- *for use in EC environmental toxicity laboratories;*
- *for testing which is contracted out by Environment Canada or requested from outside agencies or industry;*
- *in the absence of more specific instructions, such as are contained in regulations; and*
- *as a foundation for the provision of very* explicit *instructions as might be required in a regulatory protocol or standard reference method.*

The different types of tests included in this series were selected because of their acceptability for the needs of programs for environmental protection and management carried out by Environment Canada. These reports are intended to provide guidance and to facilitate the use of consistent, appropriate, and comprehensive procedures for obtaining data on the toxicity to terrestrial or aquatic life of specific test substances or materials destined for or within the environment. Depending on the biological test method(s) chosen and the environmental compartment of concern, substances or materials to be tested for toxicity could include samples of chemical or chemical product, soil or similar particulate material, sediment or similar particulate material, effluent, elutriate, leachate, or receiving water. Appendix A lists the biological test methods and supporting guidance documents published to date by Environment Canada's Method Development and Applications Section in Ottawa, ON.

Words defined in the Terminology section of this document are italicized when first used in the body of the report according to the definition. Italics are also used as emphasis for these and other words, throughout the report.

Table of Contents

Abstract v
Résumé vi
Foreword vii
List of Tables xii
List of Figures xii
List of Abbreviations and Chemical Formulae xiii
Terminology xiv
Acknowledgements xxvi

Section 1
Introduction 1
1.1 Background 1
1.2 Selection of Test Species 2
1.2.1 Alfalfa *(Medicago sativa* L.*)* 4
1.2.2 Barley *(Hordeum vulgare* L.*)* 6
1.2.3 Blue Grama Grass *[Bouteloua gracilis* (HBK) Lag. ex Steud.*]* 7
1.2.4 Carrot *(Daucus carota* L.*)* 8
1.2.5 Cucumber *(Cucumis sativus* L.*)* 8
1.2.6 Durum Wheat *[Triticum durum* (Desf.) or *Triticum turgidum* L. subsp. *durum* (Desf.) Husn. or *Triticum pyramidal* (Percival)*]* 8
1.2.7 Lettuce *(Lactuca sativa* L.*)* 9
1.2.8 Northern Wheatgrass *[Elymus lanceolatus* (Scribn. & J.G. Sm.) Gould; formerly *Agropyron dasystachyum* (Hook.) Scribn.*]* 9
1.2.9 Radish *(Raphanus sativus* L.*)* 10
1.2.10 Red Clover *(Trifolium pratense* L.*)* 10
1.2.11 Red Fescue *(Festuca rubra* L.*)* 11
1.2.12 Tomato *(Lycopersicon esculentum* Mill.*)* 11
1.3 Historical Use of Terrestrial Plants in Toxicity Tests 12

Section 2
Test Organisms 15
2.1 Species 15
2.2 Source 16
2.3 Seed Sorting and Preparation 17
2.4 Seed Storage 17
2.5 Seed Condition 17

Section 3
Test System 19
3.1 Facilities and Apparatus 19
3.2 Initial and Definitive Tests 20
3.2.1 Initial Tests 20
3.2.2 Definitive Test 20
3.3 Lighting 21
3.4 Negative Control Soil 21
3.4.1 Natural Soil 22
3.4.2 Artificial Soil 22

3.5 Positive Control Soil 23
3.6 Reference Soil 24
3.7 Test Soil 25

Section 4
Universal Test Procedures 26
4.1 Preparing Test Soils 29
4.1.1 Range-Finding Test 30
4.2 Beginning the Test 31
4.3 Test Conditions 31
4.4 Criteria for a Valid Test 32
4.5 Hydration of Test Soil During the Test 33
4.6 Observations and Measurements During the Test 33
4.7 Ending the Test 35
4.8 Test Endpoints and Calculations 37
4.8.1 Percent Emergence 38
4.8.2 EC50 38
4.8.3 ICp 41
4.8.3.1 Use of regression analyses 42
4.8.3.2 Linear interpolation using ICPIN 45
4.9 Tests with a Reference Toxicant 46

Section 5
Specific Procedures for Testing Field-Collected Soil or Similar Particulate Material 53
5.1 Sample Collection 53
5.2 Sample Labelling, Transport, Storage, and Analyses 54
5.3 Preparing Sample for Testing 56
5.4 Test Observations and Measurements 60
5.5 Test Endpoints and Calculations 61
5.5.1 Variations in Design and Analysis 62
5.5.2 Power Analysis 63

Section 6
Specific Procedures for Testing Chemical-Spiked Soil 65
6.1 Sample Properties, Labelling, and Storage 65
6.2 Preparing Test Mixtures 66
6.3 Test Observations and Measurements 69
6.4 Test Endpoints and Calculations 70

Section 7
Reporting Requirements 71
7.1 Minimum Requirements for a Test-Specific Report 71
7.1.1 Test Substance or Material 71
7.1.2 Test Organisms 71
7.1.3 Test Facilities 72
7.1.4 Test Method 72
7.1.5 Test Conditions and Procedures 72

7.1.6 Test Results 72
7.2 Additional Reporting Requirements 73
7.2.1 Test Substance or Material 73
7.2.2 Test Organisms 73
7.2.3 Test Facilities and Apparatus 73
7.2.4 Negative Control Soil or Reference Soil 73
7.2.5 Test Method 73
7.2.6 Test Conditions and Procedures 74
7.2.7 Test Results 74

References 75

Appendix A
Biological Test Methods and Supporting Guidance Documents Published by Environment Canada's Method Development and Applications Section 85

Appendix B
Environment Canada Regional and Headquarters Offices 88

Appendix C
Members of the Inter-Governmental Environmental Toxicity Group (as of December, 2004) 89

Appendix D
Members of the Scientific Advisory Group 91

Appendix E
Procedural Variations for Tests of Emergence and Growth in Soil Using Terrestrial Plants, as Described in International Methodology Documents 93

Appendix F
Variations in Recommended Test Species for Tests of Emergence and Growth in Soil Using Terrestrial Plants, as Described in International Methodology Documents 108

Appendix G
Natural and Artificial Negative Control Soils Used for Method Development and the Establishment of Test Validity Criteria 111

Appendix H
Logarithmic Series of Concentrations Suitable for Toxicity Tests 115

Appendix I
Instruction on the Derivation of ICps Using Linear and Nonlinear Regression Analyses 116

List of Tables

1 Characteristics of Plant Species 5

2 Checklist of Recommended Conditions and Procedures for Conducting Definitive Tests of Soil Toxicity Using Terrestrial Plants 27

3 Checklist of Recommended Conditions and Procedures for Conducting Reference Toxicity Tests on Soil Using Terrestrial Plants 48

List of Figures

1 Considerations for Preparing and Performing Soil Toxicity Tests using Terrestrial Plants and Various Types of Test Materials or Substances 3

2 Estimating a Median Effective Concentration by Plotting Emergence on Logarithmic-Probability Paper 40

3 The General Process for the Statistical Analysis and Selection of the Most Appropriate Model for Quantitative Toxicity Data 44

List of Abbreviations and Chemical Formulae

AES	atomic emission spectrophotometry
ANOVA	analysis of variance
$CaCl_2$	calcium chloride
$CaCO_3$	calcium carbonate
$Ca(OH)_2$	calcium hydroxide
CCME	Canadian Council of Ministers of the Environment
cm	centimetre(s)
CV	coefficient of variation
EC50	median effective concentration
ECx	effective concentration for a (specified) percent effect
ERA	ecological risk assessment
ES	effect size
g	gram(s)
h	hour(s)
HCl	hydrochloric acid
HNO_3	nitric acid
HPLC	high pressure liquid chromatography
H_2O	water
H_o	null hypothesis
ICAP	inductively coupled argon plasma
ICp	inhibiting concentration for a (specified) percent effect
KCl	potassium chloride
kg	kilogram(s)
L	litre(s)
lm	lumen(s)
LOEC	lowest-observed-effect concentration
LSD	least significant difference
m	metre(s)
M	mole(s) (concentration)
mg	milligram(s)
mL	millilitre(s)
mm	millimetre(s)
mS	millisiemens
MW	molecular weight
n	sample size
nm	nanometre(s)
NOEC	no-observed-effect concentration
OM	organic matter
P	probability
PAHs	polycyclic aromatic hydrocarbons
SD	standard deviation
s	second
sp.	species (singular)
spp.	species (plural)
TOC	total organic carbon
TM (™)	Trade Mark
v:v	volume-to-volume
WHC	water-holding capacity
wt	weight
wt:wt	weight-to-weight
°C	degree(s) Celsius
α	level of statistical significance (alpha)
µg	microgram(s)
µmhos	micromhos
µmol	micromole(s)
>	greater than
<	less than
≥	greater than or equal to
≤	less than or equal to
%	percentage or percent
=	equals
+	plus
−	minus
±	plus or minus
×	times
÷	divided by
/	per; alternatively, “or” (e.g., shoot/root)
≃	approximately equal to
~	approximately

Terminology

Note: all definitions are given in the context of the procedures in this report, and might not be appropriate in another context.

Grammatical Terms

Must is used to express an absolute requirement.

Should is used to state that the specified condition or procedure is recommended and ought to be met if possible.

May is used to mean "is (are) allowed to".

Can is used to mean "is (are) able to".

Might is used to express the possibility that something could exist or happen.

Technical Terms

Adventitious roots are thin, moderately branching roots that arise from somewhere other than the primary root; for example, roots that arise from the stem or leaves.

Angiosperm is a term used in plant classification referring to plants that flower, and whose ovules (young seeds) are enclosed in an ovary. The ovary matures into a fruit with seeds, following fertilization. The Phylum Magnoliophyta (or Anthophyta) contains all angiosperms and is the largest and most diverse group within the Kingdom Plantae. Two Classes of angiosperms include the Class Magnoliopsida (*Dicotyledons*) and the Class Liliopsida (*Monocotyledon*s).

Annual is a plant that completes its entire life cycle in a single growing season. (See also *biennial* and *perennial*.)

Biennial is a plant that normally requires two seasons to complete its life cycle, growing only roots and leaves in the first season and producing flowers and fruits and then dying in the second season. (See also *annual* and *perennial*.)

Biomass is the total weight (mass) of a group of animals or plants.

Canopy for the purpose of this method is the more or less continuous cover produced by the foliage of plants.

Cespitose means growing in dense clumps or tufts.

Chlorosis is a condition in which the green parts of plants have depressed concentrations of chlorophyll and the leaves are pale green or yellow in colour. This might result from disease, exposure to toxic substances, nutrient deficiencies, or senescence.

Coleoptile is the protective tissue surrounding the growing shoot in a monocotyledonous plant.

Compliance means in accordance with governmental regulations or requirements for issuing a permit.

Conductivity is a numerical expression of the ability of an aqueous solution to carry an electric current. This ability depends on the concentrations of ions in solution, their valence and mobility, and on the solution's temperature. Conductivity is measured at 25 °C, and is reported as micromhos per centimetre (µmhos/cm) or as millisiemens per metre (mS/m); 1 mS/m = 10 µmhos/cm.

Cotyledon is a primary leaf of the developing embryo in seeds; there is only one in monocotyledonous plants, and two in dicotyledonous plants. In many dicotyledonous species, such as the bean, the cotyledons emerge above ground and appear as the first leaves.

Cultivar means a race or variety of plant that has been created or selected intentionally and maintained through cultivation.

Defoliation is the condition in which a plant does not have a normal complement of leaves due to some internal or external cause.

Dessication is when the plant, or portion of plant, is dried.

Dicotyledon is a term used in the classification of plants, that refers to those species having two seed leaves (*cotyledon)*.

Ecological risk assessment (ERA) is the process of identifying and quantifying *risks* to nonhuman organisms and determining the acceptability of those risks.

Emergence occurs following the *germination* of a plant, wherein the early *growth* of a seedling pushes the *epicotyl* through the soil surface. In this test method, emergence refers to the appearance of the seedling shoot 3 mm above the surface of the soil.

Epicotyl is that portion of an embryo or seedling containing the shoot. It is delineated anatomically by the tissue transition zone which separates the epicotyl from the *hypocotyl*.

Epigeal (germination) refers to a type of *germination* where the *hypocotyl* is active and pulls the *cotyledon*s above ground during its growth. Germination begins with the imbibition of water and proceeds with *emergence* of the *radicle* from the seed to form the primary root and secondary roots; elongation of the active hypocotyl follows with the hypocotyl arch penetrating through the soil surface. Epigeal-emerging dicots (e.g., 90% of dicotyledonous plants) have the advantages of being able to commence photosynthesis as soon as the cotyledons emerge, and of being able to expand leaf area rapidly.

Germination refers to the process by which the plant embryo within the *seed* resumes growth after a period of dormancy and the *seedling* emerges from the seed. (See also *epigeal* and *hypogeal*.)

Growth is the increase in size or weight as the result of proliferation of new tissues. In this test method, growth refers to an increase in shoot and root length, as well as an increase in shoot and root dry and wet weights.

Hormesis is an observed stimulation of performance among organisms, compared to the control organisms, at low concentrations in a toxicity test.

Hull is the dry outer covering of a *seed*.

Hypogeal (germination) refers to a type of *germination* where the *hypocotyl* is inactive and the scutellum (*cotyledon*) remains below the ground. The *radicle* emerges first to form the primary root, followed by the *coleoptile*. *Emergence* is largely dependent on elongation of the coleoptile and the first internode. When the

soil surface is reached, light inhibits further growth and true leaves emerge through the hollow sheath. All grasses (e.g., barley) are characterized by hypogeal germination.

Hypocotyl is that portion of an embryo or seedling containing the root or *radicle*. It is delineated anatomically by the tissue transition zone which separates the *epicotyl* from the hypocotyl.

Lux is a unit of illumination based on units per square metre. One lux = 0.0929 foot-candles and one foot-candle = 10.76 lux. For conversion of lux to quantal flux [µmol/(m^2 · s)], the spectral quality of the light source must be known. Light conditions or irradiance are properly described in terms of quantal flux (photon fluence rate) in the photosynthetically effective wavelength range of approximately 400–700 nm. The relationship between quantal flux and lux or foot-candle is highly variable and depends on the light source, the light meter used, the geometrical arrangement, and the possibility of reflections (see ASTM, 1999a). The approximate conversion between quantal flux and lux, however, for full-spectrum fluorescent light (e.g., Vita-Lite® by Duro-Test®), is 1 lux ≃ 0.016 µmol/(m^2 · s) (Deitzer, 1994; Sager and McFarlane, 1997).

Malformation is a structural defect that occurs infrequently and is due to abnormal development.

Monitoring is the routine (e.g., daily, weekly, monthly, quarterly) checking of quality, or collection and reporting of information. In the context of this report, it means either the periodic (routine) checking and measurement of certain biological or soil quality variables, or the collection and testing of soil samples for toxicity.

Monocotyledon is a term used in the classification of plants, that refers to those species having a single seed leaf (*cotyledon*).

Mottling means marked with spots or streaks of different colors (e.g., blotched). This includes the discoloration of leaf margins.

Necrosis refers to dead tissue.

Nodulate is the process of forming nodules, which are small, cylindrical growths, often found on the roots of leguminous plants. These nodules house symbiotic bacteria (*Rhizobia*) that fix atmospheric nitrogen, making it available to the plant.

Perennial is a plant that, under natural conditions, lives for several to many growing seasons. (See also *annual* and *biennial*.)

pH is the negative logarithm of the activity of hydrogen ions in gram equivalents per litre. The pH value expresses the degree or intensity of both acidic and alkaline reactions on a scale from 0–14, with 7 representing neutrality, numbers less than 7 indicating increasingly greater acidic reactions, and numbers greater than 7 indicating increasingly basic or alkaline reactions.

Photoperiod is the duration of illumination and darkness within a 24-h period.

Phytomass is the total weight (mass), either above and/or below ground, of a group of plants.

Pollution is the addition of a substance or material, or a form of energy such as heat, to some component of the environment, in such an amount as to cause a discernible change that is deleterious to some organism(s) or to some human use of the environment. Some national and international agencies have formal definitions of pollution, which should be honoured in the appropriate contexts.

Pretreatment means treatment of a sample of soil, or portion thereof, before exposure of the test organisms.

Protocol is an explicit set of procedures for a test or an experiment, formally agreed upon by the parties involved, and described precisely in a written document.

Quality assurance (QA) is a program within a laboratory, intended to provide precise and accurate results in scientific and technical work. It includes selection of proper procedures, sample collection, selection of limits, evaluation of data, *quality control*, and qualifications and training of personnel.

Quality control (QC) consists of specific actions within the program of *quality assurance*. It includes standardization, calibration, replication, control samples, and statistical estimates of limits for the data.

Rachis is the central stalk or mid-rib of a compound leaf.

Radicle is the end of a plant embryo that gives rise to the first root.

Reference method refers to a specific *protocol* for performing a toxicity test, i.e., a biological test method with an explicit set of test procedures and conditions, formally agreed upon by the parties involved and described precisely in a written document. Unlike other multi-purpose (generic) biological test methods published by Environment Canada, the use of a reference method is frequently restricted to testing requirements associated with specific regulations.

Remediation is the management of a contaminated *site* to prevent, minimize, or mitigate damage to human health or the environment. Remediation can include both direct physical actions (e.g., removal, destruction, and containment of toxic substances) and institutional controls (e.g., zoning designations or orders).

Rhizobia are soil bacteria that fix nitrogen after becoming established inside the root nodules of legumes.

Rhizome is a fleshy, creeping, horizontal, underground stem that often sends out roots and shoots from its nodes. Certain plants reproduce vegetatively by means of their rhizome.

Rhizomatous – see *rhizome*.

Risk is the probability or likelihood that an adverse effect will occur.

Risk assessment – see *Ecological risk assessment*.

Root is usually the below-ground portion of a plant that serves as support, draws minerals from surrounding soil, and sometimes stores food. There are two main types of root systems: the tap root system, in which there is a main primary root larger than the other branching roots, and the fibrous root system, in which there are many slender roots with numerous smaller root branches. (See also *shoot*.)

Seed is a fertilized and ripened plant ovule consisting of the plant embryo, varying amounts of stored food material, and a protective outer seed coat.

Seedling is a young plant that is grown from a seed.

Seed pretreatment is a coating of fungicide applied to seeds before water imbibition.

Seminal root is a seed-born root that develops directly from root growing points present in the seed. Seminal roots consist of the *radicle* and lateral *seminal roots*, and are the first seedling roots to emerge from the seed. These roots serve to anchor and support the young seedling and absorb small amounts of water and nutrients until the permanent root system takes over.

Shoot is the usually above-ground portion of the plant such as the stems and leaves.

Staining is the discoloration of plant parts (roots, vegetative growth) caused by the test substance.

Tetraploid means having four-times the haploid number of chromosomes (i.e., a single set of chromosomes) in the cell nucleus.

Waterlogging is the over-saturation or soaking of agricultural land caused by a rising water table or excessive irrigation. Waterlogging compacts soil and deprives roots of oxygen.

Wilting occurs when plant tissues lose their turgidity and the plant becomes limp.

Withering is the process of drying; plants become limp and desiccated. This frequently is the result of root damage.

Terms for Test Materials or Substances

Artificial soil is a laboratory-formulated soil, prepared to simulate a natural soil using a specific ratio of natural constituents of sand, clay, and peat. Artificial soil may be used as a *negative control soil*, and as a diluent to prepare multiple concentrations of *site soil(s)* or *chemical-spiked soil(s)*.

Batch means the total amount of a particular *test soil* (or specific concentration thereof) prepared for each treatment (concentration) in a test. A batch is any hydrated *test soil* ready for separation into replicates.

Chemical is, in this report, any element, compound, formulation, or mixture of a substance that might be mixed with, deposited in, or found in association with soil or water.

Chemical-spiked soil is natural or *artificial soil* (usually *negative control soil, reference soil*, or other *clean* soil) to which one or more chemicals or chemical products have been added, and mixed thoroughly to evenly distribute the substance(s) throughout the soil at a specific concentration to form a *batch* for use in a soil toxicity test. See also *spiked soil*.

Clean soil is soil that does not contain concentrations of any substance(s) or material(s) causing discernible toxic effects on the test organisms.

Concentration means, for this biological test method, the ratio of the weight of test substance or material to the weight of soil, and is frequently expressed as the weight of test substance or material per kg of dry soil (mg/kg). Concentration might also be expressed as a percentage of the test substance (e.g., *contaminated* site soil) or material per dry weight of soil.

Contaminant is a substance or material that is present in a natural system, or present at increased concentrations, often because of some direct or indirect human activity. The term is frequently applied to substances or materials present at concentrations having the potential to cause adverse biological effects.

Contaminated (soil) means (soil) containing chemical substances or materials at concentrations that pose a known or potential threat to environmental or human health.

Control is a treatment in an investigation or study that duplicates all the conditions and factors that might affect results, except the specific condition being studied. In toxicity tests, the control must duplicate all the conditions of the exposure treatment(s), but must contain no contaminated test material. The control is used as a check for the absence of toxicity due to basic test conditions such as temperature, health of test

organisms, or effects due to their handling. Control is synonymous with *negative control*, unless indicated otherwise.

Control soil – see *negative control soil*.

Definitive (soil toxicity test) means decisive (as opposed to a preliminary, range-finding test). [See also *range-finding (test)*.]

De-ionized water is water that has been purified by passing it through resin columns or a reverse osmosis system, for the purpose of removing ions such as Ca^{++} and Mg^{++}.

Distilled water is water that has been passed through a distillation apparatus of borosilicate glass or other material, to remove impurities.

Fertility (of soil) refers to the potential of a soil to supply nutrient elements in the amounts, forms, and proportions required for optimal plant growth. Soil fertility is measured directly in terms of the ions and compounds important for plant nutrition. The fundamental components of fertility are the essential nutrients (macronutrients including C, H, O, N, P, K, Ca, Mg, S and micronutrients including Fe, Mn, Mo, B, Cu, Zn, and Cl). Indirectly, soil fertility is measured by demonstrating its productivity (i.e., the capacity of the soil to produce plants that supply man with essential food and fibre; Hausenbuiller, 1985).

Hydration water means water used to hydrate test soils, to create a specific *moisture content* suitable for the test organisms. The water used for hydration is normally *test water*, and is frequently de-ionized or distilled water, reverse-osmosis water, de-chlorinated tap water, or nutrient solution, where applicable. Depending on study design and intent, a surface water or groundwater from the *site* might be used instead of de-ionized or distilled water for the hydration of each test soil (including negative control soil). (See also *test water*, *de-ionized water*, and *distilled water*.)

Material is the *substance* or substances from which something is made. A material would have more or less uniform characteristics. Soil, sediment, or surface water are materials. Usually, the material would contain several or many substances.

Moisture content is the percentage of water in a sample of test soil, based on its wet or dry mass. It is determined by measuring both the wet and dry weights of a subsample of the soil. The soil's moisture content is then calculated and expressed on a dry-weight basis, by dividing the mass of water in the subsample (wet mass – dry mass) by the mass of dry soil, and then multiplying by 100. Units for mass (i.e., g or mg) must be the same in each instance.

Negative control (see *control*).

Negative control soil is *clean* soil that does not contain concentrations of one or more contaminants which could affect the emergence, survival, or growth of the test organisms. Negative control soil might be natural soil from an uncontaminated *site*, or artificial (formulated) soil. This soil must contain no added test material or substance, and must enable acceptable emergence, survival, and growth of the test plants during the test. The use of negative control soil provides a basis for interpreting data derived from toxicity tests using test soil(s).

Organic matter (OM) in soil consists primarily of plant and animal residues, at different stages of decomposition, including soil humus. The accumulation of OM within soil is a balance between the return or addition of plant and animal residues and their subsequent loss due to the decay of these residues by soil micro-organisms. For most types of soil, the following equation (AESA, 2001) is suitable for estimating the total OM content of soil from *total organic carbon* (TOC) measurements: % OM = % TOC × 1.78. (See also *total organic carbon.*)

Positive control soil is *contaminated* soil that contains known concentrations of one or more contaminants that adversely affect the emergence, survival, or growth of the test organisms using the biological test method defined herein. Positive control soil might be used as a *reference toxicant* to assess the sensitivity of the test organisms at the time the test material or substance is evaluated, and to determine the precision of results obtained by the laboratory for that reference toxicant.

Product is a commercial formulation of one or more chemicals. (See also *chemical*.)

Range-finding (test) means a preliminary soil toxicity test, performed to provide an initial indication of the toxicity of the test material under defined conditions and to assist in choosing the range of concentrations to be used in a definitive multi-concentration test. [See also *definitive (soil toxicity test)*.]

Reference soil is typically *clean* field-collected soil or formulated (artificial) soil, that is selected for use in a particular toxicity test together with a *negative control soil* and one or more samples of *test soil*. The test soil might be either field-collected *site soil* that is *contaminated* or potentially so, or *chemical-spiked soil*. Reference soil used in a test frequently exhibits physicochemical properties (e.g., *texture*, compactness, *total organic carbon* content, pH) closely matching those of the test soil sample(s), except that it is free from the source of contamination being assessed. In tests involving samples of *site soil*, one or more samples of reference soil are often selected from the general location of test soil sampling, and thus might be subject to other sources of contamination aside from the one(s) being studied. Reference soil is used to describe matrix effects in the test, and may also be used as a diluent to prepare concentrations of the test soil. In tests involving *chemical-spiked soil*, one or more samples of artificial (formulated) soil with differing physicochemical characteristics might be chosen to investigate the influence of certain soil properties (e.g., soil *texture*, or percent *organic matter*) on the toxicity of a chemical mixed in each of these soil types. (See also *negative control soil*, *site soil*, *test soil*, *clean*, *artificial soil*, and *chemical-spiked soil.*)

Reference toxicant is a standard chemical used to measure the sensitivity of the test organisms to establish confidence in the toxicity data obtained for a test material or substance. In most instances, a toxicity test with a reference toxicant is performed to assess the sensitivity of the organisms at the time the test material or substance is evaluated, and the precision and reliability of results obtained by the laboratory for that chemical.

Reference toxicity test is a test conducted using a *reference toxicant* in conjunction with a soil toxicity test, to appraise the sensitivity of the organisms and/or the precision and reliability of results obtained by the laboratory for that chemical at the time the test material or substance is evaluated. Deviations outside an established normal range indicate that the sensitivity of the test organisms, and the performance and precision of the test, are suspect. A reference toxicity test with plants is performed as a *spiked-soil* test, using a standard chemical.

Sampling station means a specific location, within a *site* or sampling unit (depending on the study design), where the sample(s) of field-collected soil are obtained for toxicity tests and associated physicochemical analyses.

Site means a delineated tract of land that is being used or considered as a study area, usually from the perspective of it being *contaminated* or potentially contaminated by xenobiotics.

Site soil is a field-collected sample of soil, taken from a location thought to be *contaminated* with one or more chemicals, and intended for use in the toxicity test with plants. In some instances, the term includes *reference soil* or *negative control soil* from a site.

Soil is whole, intact material representative of the terrestrial environment, that has had minimal manipulation following collection or formulation. In the natural environment, it is formed by the physical, chemical, and biological weathering of rocks and the decomposition and recycling of nutrients from *organic matter*

originating from plant and animal life. Its physicochemical characteristics are influenced by biological activities (e.g., microbial, invertebrates, and plants) therein, and by anthropogenic activities.

Solvent control soil is a sample of (usually artificial) soil included in a test involving *chemical-spiked soil*, in which an organic solvent is required to solubilize the test chemical before mixing it in a measured quantity of *negative control soil*. The amount of solvent used when preparing the solvent control soil must contain the same concentration of solubilizing agent as that present in the highest concentration of the test chemical(s) in the sample of chemical-spiked soil to be tested. This concentration of solvent should not adversely affect the plants during the test. Any test that uses an organic solvent when preparing one or more concentrations of *chemical-spiked soil* must include a solvent control soil in the test. (See also *artificial soil*, *negative control soil*, and *chemical-spiked soil*.)

Spiked soil is natural or *artificial soil* (usually *negative control soil, reference soil*, or other *clean* soil) to which one or more chemicals, chemical products, or other test substances or materials (e.g., a sample of sludge or drilling mud) have been added in the laboratory, and mixed thoroughly to evenly distribute the substance(s) or material(s) throughout the soil at a specific concentration to form a *batch* for use in a soil toxicity test. (See also *chemical-spiked soil* and *spiking*.)

Spiking refers to the addition of a known amount of chemical(s), chemical product(s), or other test substance(s) or material(s) (e.g., a sample of sludge or drilling mud) to a natural or *artificial soil*. The substance(s) or material(s) is usually added to *negative control soil, reference soil*, or another *clean* soil, but sometimes to a *contaminated* or potentially contaminated soil. After the addition ("spiking"), the soil is mixed thoroughly. If the added test material is a *site soil*, Environment Canada documents typically do not call this spiking, but instead refer to the manipulation as "dilution" or simply "addition". (See also *chemical-spiked soil* and *spiked soil*.)

Stock solution means a concentrated solution of the substance(s) to be tested, following the addition of a measured quantity of this solution to a sample of natural or *artificial soil* and thorough mixing to prepare a *batch* of *chemical-spiked soil*. To prepare the required strength of the stock solution, measured weights or volumes of test chemical(s) or chemical product(s) are added to test water (*de-ionized* or *distilled water*, or equivalent), with or without the inclusion of an organic solvent.

Substance is a particular kind of material having more or less uniform properties. The word substance has a narrower scope than *material*, and might refer to a particular chemical (e.g., an element) or chemical product.

Test battery is a combination of several toxicity tests, normally using different species of test organisms (e.g., a series of soil toxicity tests using earthworms, plants, or springtails, or a series of soil toxicity tests using several species of plants), different biological endpoints (e.g., lethal and various sublethal), and different durations of exposure (e.g., *acute* and *chronic*).

Test soil is a sample of field-collected soil or *chemical-spiked soil* to be evaluated for toxicity to plants. In some instances, the term also applies to any solid-phase sample or mixture thereof (e.g., *negative control soil*, *positive control soil*, *reference soil*, sludge, drilling mud) used in a soil toxicity test.

Test water is water used to prepare *stock solutions*, rinse test organisms, or rinse glassware and other apparatus and for other purposes associated with the biological test method (e.g., to hydrate samples of *test soil*). Test water must be *de-ionized* or *distilled water* or better (e.g., reagent-grade water produced by a system of reverse osmosis, carbon, and ion-exchange cartridges). (See also *hydration water*.)

Texture is defined based on a measurement of the percentage by weight of sand, silt, and clay in the mineral fraction of soils. Classification as to texture confers information on the general character and behaviour of substances in soils, especially when coupled with information on the structural state and *organic matter*

content of the soil. Soil texture is determined in the laboratory by measuring the particle-size distribution using a two-step procedure whereby the sand particles (coarse fragments) are initially separated by sieving from the silt and clay particles; followed by separation of the silt and clay particles by their sedimentation in water. Textural classification systems typically refer to groupings of soil based on specific ranges in relative quantities of sand, silt, and clay. There are three main textural classes:

(i) coarse texture (sands, loamy sands, sandy loams);
(ii) medium texture (loams, silt loams, silts, very fine sandy loams); and
(iii) fine texture (clays, silty clay loams, sandy clay loams, silty clays, and sandy clays).

Further distinction as to texture (e.g., "sandy clay", "silt loam", "loam") can be made based on classification schemes using the relative amounts of percent sand, percent silt, and percent clay in the soil (Hausenbuiller, 1985; ACECSS, 1987).

Total organic carbon (TOC) refers to the organic carbon content of soil exclusive of carbon from undecayed plant and animal residues. The TOC is determined by dry combustion analysis (ISO, 1995). (See also *organic matter*.)

Water-holding capacity (WHC) refers to the maximum quantity of water that a soil can retain, following complete saturation. It is usually determined gravimetrically, and is generally expressed as the percentage of water (by mass; wt water:wt dry soil) retained in a sample of soil that has been saturated with water.

Statistical and Toxicological Terms

Acute means within a short period of exposure (seconds, minutes, hours, or a few days) in relation to the life span of the test organism.

Acute toxicity is a discernible adverse effect (lethal or sublethal) induced in the test organisms within a short period (usually a few days, and for purposes of this document within 5–7 days) of exposure to *test soil(s)*.

Bioassay is a test (= assay) in which the strength or potency of a substance is measured by the response of living organisms. In standard pharmacological usage, a bioassay assesses the unknown potency of a given preparation of a drug, compared to the known potency of a standard preparation. *Toxicity test* is a more specific and preferred term for environmental studies.

Chronic means occurring within a relatively long period of exposure (weeks, months, or years), usually a significant portion of the life span of the organism such as 10% or more.

Coefficient of Variation (CV) is the standard deviation (SD) of a set of data divided by the mean of the data set, expressed as a percentage. It is calculated according to the following formula:
CV (%) = 100 × (SD ÷ mean).

EC50 is the *median effective concentration,* i.e., the concentration (e.g., % or mg/kg) of substance(s) or material(s) in soil estimated to cause some defined toxic effect on 50% of the test organisms. In most instances, the EC50 and its 95% confidence limits are statistically derived by analyzing the percentages of organisms affected (e.g., % seeds emerged) at various test concentrations, after a fixed period of exposure. The duration of exposure must be specified (e.g., 7-day EC50). The EC50 describes *quantal* effects, lethal or sublethal, and is not applicable to continuous (i.e., *quantitative*) effects (see *ICp*). Depending on the study objectives, an ECx other than EC50 (e.g., an EC20) might be calculated instead of or in addition to the EC50.

Endpoint means the measurement(s) or value(s) that characterize the results of a test (e.g., EC50, IC25). It also means the response of the test organisms that is being measured (e.g., seedling *emergence*, or shoot/root length and weight).

Environmental toxicology is a branch of *toxicology* with the same general definition; however, the focus is on ecosystems, natural communities, and wild living species, without excluding humans as part of the ecosystems.

Geometric mean is the mean of repeated measurements, calculated logarithmically. It has the advantage that extreme values do not have as great an influence on the mean as is the case for an arithmetic mean. The *geometric mean* can be calculated as the n^{th} root of the product of the "n" values, and it can also be calculated as the antilogarithm of the mean of the logarithms of the "n" values.

Heteroscedasticity refers herein to data showing heterogeneity of the residuals within a scatter plot (see Figures I.2B and I.2C in Appendix I). This term applies when the variability of the residuals changes significantly with that of the independent variables (i.e., the test concentrations or treatment levels). When performing statistical analyses and assessing residuals (e.g., using Levine's test), for test data demonstrating heteroscedasticity (i.e., non-homogeneity of residuals), there is a significant difference in the variance of residuals across concentrations or treatment levels. (See also *homoscedasticity* and *residual*.)

Homoscedasticity refers herein to data showing homogeneity of the residuals within a scatter plot (see Figure I.2A in Appendix I). This term applies when the variability of the residuals does not change significantly with that of the independent variables (i.e., the test concentrations or treatment levels). When performing statistical analyses and assessing residuals (e.g., using Levine's test), for test data demonstrating homoscedasticity (i.e., homogeneity of residuals), there is no significant difference in the variance of residuals across concentrations or treatment levels. (See also *heteroscedasticity* and *residual*).

ICp is the *inhibiting concentration for a (specified) percent effect*. It represents a point estimate of the concentration of test substance or material that causes a designated percent inhibition (*p*) compared to the control, in a *quantitative* (continuous) biological measurement such as length of shoots attained by individual seedlings at the end of the test.

LOEC is the *lowest-observed-effect concentration*. This is the lowest concentration of a test substance or material for which a statistically significant adverse effect on the test organisms was observed, relative to the control.

NOEC is the *no-observed-effect concentration*. This is the highest concentration of a test substance or material at which no statistically significant adverse effect on the test organisms was observed, relative to the control.

Phytotoxicity means unwanted detrimental deviations from the normal pattern of appearance, growth, and/or function of plants in response to the test material. Phytotoxicity might occur during *germination*, *growth* differentiation, and/or maturation of plants.

Precision refers to the closeness of repeated measurements of the same quantity to each other, i.e., the degree to which data generated from repeated measurements are the same. It describes the degree of certainty around a result, or the tightness of a statistically derived endpoint such as an ICp.

Quantal effects in a toxicity test are those in which each test organism responds or does not respond. For example, a seedling might fail to emerge from *contaminated* test soil. Generally, quantal effects are expressed as numerical counts or percentages thereof. (See also *quantitative*.)

Quantitative effects in a toxicity test are those in which the measured effect is continuously variable on a numerical scale. Examples would be shoot length of emerged seedlings or dry weight of roots at the end of the test. Generally, quantitative effects are determined and expressed as measurements. (See also *quantal*.)

Replicate (*treatment, test vessel,* or *test unit*) refers to a single test vessel containing a prescribed number of organisms in either one concentration of the test material or substance, or in the control or reference treatment(s). A *replicate* of a treatment must be an independent test unit; therefore, any transfer of organisms or test material from one test vessel to another would invalidate a statistical analysis based on the replication (see Sections 5.1 and 5.5.1 herein, and Section 2.5 of EC, 2004a).

Replicate samples are field-replicated samples of soil collected from the same *sampling station*, to provide an estimate of the sampling error or to improve the precision of estimation. A single soil sample from a sampling station is treated as one replicate. Additional samples are considered to be additional replicate samples when they are treated identically but stored in separate sample containers (i.e., not composited).

Residual, in the context of Section 4.8.3.1 and Appendix I, refers to the difference between the predicted estimate (based on the model) and the actual value observed, as determined by subtracting the former from the latter. (See also *heteroscedasticity* and *homoscedasticity*).

Static describes a toxicity test in which the *test soil* (nor any chemical or chemical product therein) is not renewed or replaced during the test.

Sublethal (toxicity) means detrimental to the organism, but below the concentration or level of contamination that directly causes death within the test period.

Sublethal effect is an adverse effect on an organism, below the concentration or level of contamination that directly causes death within the test period.

Toxic means poisonous. A toxic chemical or material can cause adverse effects on living organisms, if present in sufficient amount at the right location. *Toxic* is an adjective or adverb, and should not be used as a noun; whereas *toxicant* is a legitimate noun.

Toxicant is a toxic substance or material.

Toxicity is the inherent potential or capacity of a substance or material to cause adverse effect(s) on living organisms. These effect(s) could be lethal or sublethal.

Toxicity test is a determination of the effect of a substance or material on a group of selected organisms of a particular species, under defined conditions. A toxicity test involving samples of *test soil* usually measures (a) the proportions of organisms affected (*quantal*), and/or (b) the degree of effect shown (*quantitative* or graded), after exposure of the test organisms to the whole sample (e.g., undiluted *site soil*) or specific concentrations thereof.

Toxicology is a branch of science that studies the toxicity of substances, materials, or conditions. There is no limitation on the use of various scientific disciplines, field or laboratory tools, or studies at various levels of organization, whether molecular, single species, populations, or communities. Applied toxicology would normally have a goal of defining the limits of safety of chemical or other agents. (See also *environmental toxicology*.)

Treatment refers to a specific *test soil* (e.g., a *site soil*, *reference soil*, or *negative control soil*) from a particular *sampling station*, or a concentration of *chemical-spiked soil* (or a mixture of test soil diluted with *clean* soil) prepared in the laboratory. Test soils representing a particular *treatment* are typically replicated in a toxicity test. (See also *replicate* and *replicate samples*.)

Visual assessment represents the description of any visual damage to the test species based on observations of phytotoxicity (i.e., *malformation*, *chlorosis*, *necrosis*, *defoliation*, *dessication*, *mottling*, *staining*, *wilting*, or *withering*) observed in test vessels with *contaminated* soil compared to the controls.

Warning chart is a graph used to follow changes over time, in the endpoints for a *reference toxicant*. Date of the test is on the horizontal axis and the effect-concentration is plotted on the vertical logarithmic scale.

Warning limit is plus or minus two standard deviations, calculated logarithmically, from a historic geometric mean of the endpoints from tests with a *reference toxicant*.

Acknowledgements

This document was co-authored by J.A. Miller (Miller Environmental Sciences Inc., Innisfil, ON) and G.L. Stephenson (Aquaterra Environmental, Orton, ON). Dr. D.J. McLeay (McLeay Environmental Ltd., Victora, BC) is sincerely thanked for his contributions on format, structure, and text in most sections and his final review of the method. J.I. Princz (Biological Methods Division, Environmental Technology Centre, EC, Ottawa, ON) and K. Becker-van Slooten (École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland) are also thanked for their contributions to certain sections of the document. Input from J.B. Sprague (Sprague Associates Ltd., Salt Spring Island, BC) with respect to definitions in the Terminology section and statistical guidance herein is gratefully acknowledged. B.A. Zajdlik (Zajdlik & Associates, Rockwood, ON) is thanked for his input to the statistical guidance on regression analyses. N.C. Feisthauer (Stantec Consulting Ltd., Guelph, ON) is thanked for her technical guidance and thorough review of each draft.

R.P. Scroggins (Biological Methods Division, Environmental Technology Centre, EC, Ottawa, ON) acted as Scientific Authority and provided technical input and direction throughout the work. The studies resulting in the biological test method defined herein were developed in part by G.L. Stephenson, as part of her doctoral thesis. Further validation of the test method was directed by Dr. Stephenson, and managed and performed by N.C. Feisthauer (Stantec Consulting Ltd, Guelph, ON) and co-workers with Stantec Consulting Ltd. (formerly ESG International Inc.) (Guelph, ON). Additional method-development studies were carried out by J.I. Princz and K. Becker-van Slooten. R. Pandey of Guelph Chemical Laboratories Ltd. (Guelph, ON) and V. Marsielle-Kerslake of Analytical Services (University of Guelph, Guelph, ON) are thanked for providing assistance on methods for physicochemical analyses.

The interlaboratory studies undertaken to validate the biological test method were coordinated by J.I. Princz and performed by the following participating laboratories: Environment Canada's Atlantic Environment Science Centre (Moncton, NB), Environment Canada's Soil Toxicology Laboratory at the Environmental Technology Centre (Ottawa, ON), Ontario Ministry of the Environment (Etobicoke, ON), Saskatchewan Research Council (Saskatoon, SK), EVS Consultants Ltd. (North Vancouver, BC), and Hydroqual Laboratories Ltd. (Calgary, AB). Names of laboratory personnel participating in these tests are provided in the technical report on these studies (EC, 2005a).

We gratefully acknowledge the many useful comments provided by each member of Environment Canada's Scientific Advisory Group (SAG) responsible for the document's initial and final reviews, as follows: C. Boutin (National Wildlife Research Centre, CWS, Environment Canada, Ottawa, ON), M. Dixon (Ontario Ministry of Environment, Brampton, ON), J. Gorsuch (Health, Safety, and Environment, Eastman Kodak Company, Rochester, NY), L. Kaputska (Ecological Planning and Toxicology Inc., Corvallis, OR), C. Kjaer (National Environment Research Institute, Silkeborg, Denmark), G. Linder (Heron Works Farm, Brooks, OR), R. Petrie (Office of Pesticides, Environmental Fate, and Effects Division, USEPA, Washington, DC), T. Pfleeger (Western Ecology Division, USEPA, Corvallis, OR), A. Renoux (SANEXEN Services Environnementaux Inc., Longueuil, PQ), R. Roshon (Stantec Consulting Ltd., Guelph, ON), J. Staveley (ARCADIS Geraghty and Miller, Inc., NC), and B. Zajdlik (Zajdlik Associates, Rockwood, ON). Contact information for each SAG member is provided in Appendix D. Support and input from various members of the Inter-Governmental Environmental Toxicity Group (see Appendix C) is also acknowledged.

In addition to the members of the SAG members who reviewed this document, the following people also reviewed the initial and/or final drafts and provided useful comments: K. Becker-van Slooten (École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland), C. Boutin (National Wildlife Research Centre, CWS, Environment Canada, Ottawa, ON), E. Canaria (EVS Environment Consultants, North Vancouver, BC), S. Campich (École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland), N. Feisthauer (Stantec Consulting Ltd., Guelph, ON), J. Gorsuch (Health, Safety, and Environment, Eastman Kodak Company, Rochester, NY),

S. Goudey (Hydroqual Laboratories Ltd., Calgary, AB), F. Heimbach (Bayer CropScience Development - Ecotoxicology, Monheim, Germany), G. Lazarovitz (Agriculture and Agri-Food Canada, London, ON), D. McLeay (McLeay Environmental Ltd., Victora, BC), M. Moody (Environment/Minerals Branch, Saskatchewan Research Council, Saskatoon, SK), K. Potter (National Guidelines and Standards Office, Environment Canada, Gatineau, QC), J. Princz (Biological Methods Division, Environmental Technology Centre, EC, Ottawa, ON), L. Regier (Hydroqual Laboratories Ltd., Calgary, AB), A. Renoux (SANEXEN Services Environnementaux Inc., Longueuil, QC), C.A.M. van Gestel (Institute of Ecological Science, Vrije Universiteit, Amsterdam, The Netherlands), and R. Wagemakers (Hydroqual Laboratories Ltd, Calgary, AB).

Introduction

1.1 Background

The *Method Development and Applications Section (MDAS)* of Environment Canada is responsible for the development, standardization, and publication (see Appendix A) of a series of biological test methods for measuring and assessing the *toxic* effect(s) on single species of terrestrial or aquatic organisms, caused by their exposure to samples of test *materials* or *substances* under controlled and defined laboratory conditions. In 1994, MDAS, the Canadian Association of Petroleum Producers (CAPP), and the federal Panel for Energy Research and Development (PERD) initiated a multi-year program to research, develop, validate, and publish a number of standardized biological test methods for measuring the *toxicity* of samples of *contaminated* or potentially contaminated *soil*, using appropriate species of terrestrial test organisms. The goal was to develop new biological test methods that were applicable to diverse types of Canadian soil using terrestrial species that were representative of Canadian soil ecosystems. The initial phase of this multi-year program involved a comprehensive review of existing biological test methods, used globally to evaluate the toxicity of *contaminated* soils to plants and soil invertebrates. The resulting report recommended that Environment Canada support the development, standardization, and publication of a number of single-species biological test methods for measuring soil toxicity, including those using terrestrial plants (Bonnell Environmental Consulting, 1994). This recommendation was endorsed by both the headquarters and regional offices of Environment Canada (Appendix B) and the Inter-Governmental Environmental Toxicity Group (IGETG) (Appendix C).

Since 1994, several years of research have been completed under the direction of the MDAS on the selection of suitable and sensitive test organisms for measuring soil toxicity to meet Canadian regulatory and *monitoring* requirements, and on the development of appropriate biological test methods. A technical report was produced describing species selection criteria and processes, as well as the results of testing associated with the development of a terrestrial plant *toxicity test* for the assessment of *contaminated* soils (Aquaterra Environmental, 1998a). Other technical reports written concurrently describe tests for assessing the toxicity of soils; specifically a test for mortality and reproductive inhibition of a small soil-dwelling arthropod (Collembola: *Onychiurus folsomi*; Aquaterra Environmental, 1998b) and tests for mortality, avoidance behaviour, and reproductive inhibition of earthworms (Aquaterra Environmental, 1998c).

Numerous soil toxicity tests have been coordinated or supported by Environment Canada, using various terrestrial plant species exposed to samples of soil contaminated with pesticides, metals, petrochemical wastes, volatile hydrocarbons, or prospective *reference toxicants*. These studies (Aquaterra Environmental, 1998a; Stephenson *et al.*, 1999a, b, 2000a; Aquaterra Environmental and ESG, 2000; ESG, 2000, 2001, 2002; ESG and Aquaterra Environmental, 2002) focussed on the development and standardization of biological test methods for determining the *sublethal* toxicity of samples of *contaminated* soil to plants. Based on the results of these studies, together with the findings of a series of interlaboratory method validation studies (EC, 2005a); Environment Canada proceeded with the preparation and finalization of a biological test method for conducting soil toxicity tests that measure *emergence* and *growth* inhibition of terrestrial plant species, as described in this report.

A Scientific Advisory Group (see Appendix D) of international experts experienced with the design and implementation of soil toxicity tests using terrestrial plants provided key references which were reviewed and considered as part of this undertaking. These individuals also served actively in providing a critical peer review of the initial draft of this methodology document. The experience of the international scientific community when performing similar soil toxicity tests using terrestrial plants (see Appendices E and F) was relied on heavily when preparing this biological test method.

Detailed procedures and conditions for preparing and performing this biological test method are defined herein. Universal procedures for preparing and conducting soil toxicity tests using selected species of agricultural crop, market-garden, or grassland plants are described. Options for test species include: alfalfa *(Medicago sativa)*, barley *(Hordeum vulgare)*, blue grama grass *(Bouteloua gracilis)*, carrot *(Daucus carota)*, cucumber *(Cucumis sativus)*, durum wheat *(Triticum durum),* lettuce *(Lactuca sativa),* northern wheatgrass *(Elymus lanceolatus*; formerly named *Agropyron dasystachyum)*, radish *(Raphanus sativus)*, red clover *(Trifolium pratense)*, red fescue *(Festuca rubra),* and tomato *(Lycopersicon esculentum)*. Guidance is also provided for specific sets of conditions and procedures which are required or recommended when using this biological test method for evaluating different types of *substances* or *materials* (e.g., samples of field-collected soil or similar particulate waste, or samples of one or more *chemicals* or chemical *products* experimentally mixed into or placed in contact with natural or formulated soil). The biological *endpoints* for this method are: (a) *seedling emergence*, and (b) plant *growth* (measured as live *shoot* and *root* length and shoot and root dry mass) measured at the end of the test.

The flowchart in Figure 1 illustrates the universal topics covered herein, and lists topics specific to testing samples of field-collected soil, similar particulate waste (e.g., sludge, drilling mud, or dredged material), or soil spiked experimentally with chemical(s) or chemical product(s).

This biological test method is intended for use in evaluating the *sublethal* toxicity of samples of material such as:

(1) field-collected soil that is *contaminated* or potentially contaminated;

(2) soils under consideration for removal and disposal or *remediation* treatment;

(3) dredged material destined or under consideration for land disposal after dewatering;

(4) industrial or municipal sludge and similar particulate wastes that might be deposited on land; and

(5) clean or *contaminated* soil (natural or *artificial*), spiked with one or more chemicals or chemical products (e.g., for *risk assessment* of new or current-use chemicals).

In formulating this biological test method, an attempt has been made to balance scientific, practical, and cost considerations, and to ensure that the results will be sufficiently precise for most situations in which they will be applied. It is assumed that the user has a certain degree of familiarity with soil toxicity tests. Explicit instructions that might be required in a regulatory *protocol* are not provided in this report, although it is intended as a guidance document useful for that and other applications.

For guidance on the implementation of this and other biological test methods, and on the interpretation and application of *endpoint* data for soil toxicity, the reader should consult Sections 4.12, 5.5, and 5.6.4 in Environment Canada's "*Guidance Document on Application and Interpretation of Single-Species Tests in Environmental Toxicology*" (EC, 1999a).

1.2 Selection of Test Species

Phase I (Bonnell Environmental Consulting, 1994) of the soil toxicity test method development program (see Section 1.1) produced a list of potential species to be investigated for inclusion in a future Canadian terrestrial plant soil toxicity test method. This list includes both the traditional agricultural crop species and some more "ecologically relevant" species.

During Phase II (1995–1998) of the project, 30 plant species were screened to assess their suitability for use in toxicity tests.[1] Preliminary

[1] The 30 plant species screened were: alfalfa, american sloughgrass, barley, bluejoint, bluestem (little), bromegrass (fringed), bromegrass (mountain), cabbage, canola, carrot, clover, corn, cucumber, flax, grama grass, junegrass, lettuce, oat, onion, radish red fescue, ryegrass soybean, timothy, tomato, turnip, wheat, wheatgrass (northern), wheatgrass (streambank), and wheatgrass (western) (Stephenson, 2003a).

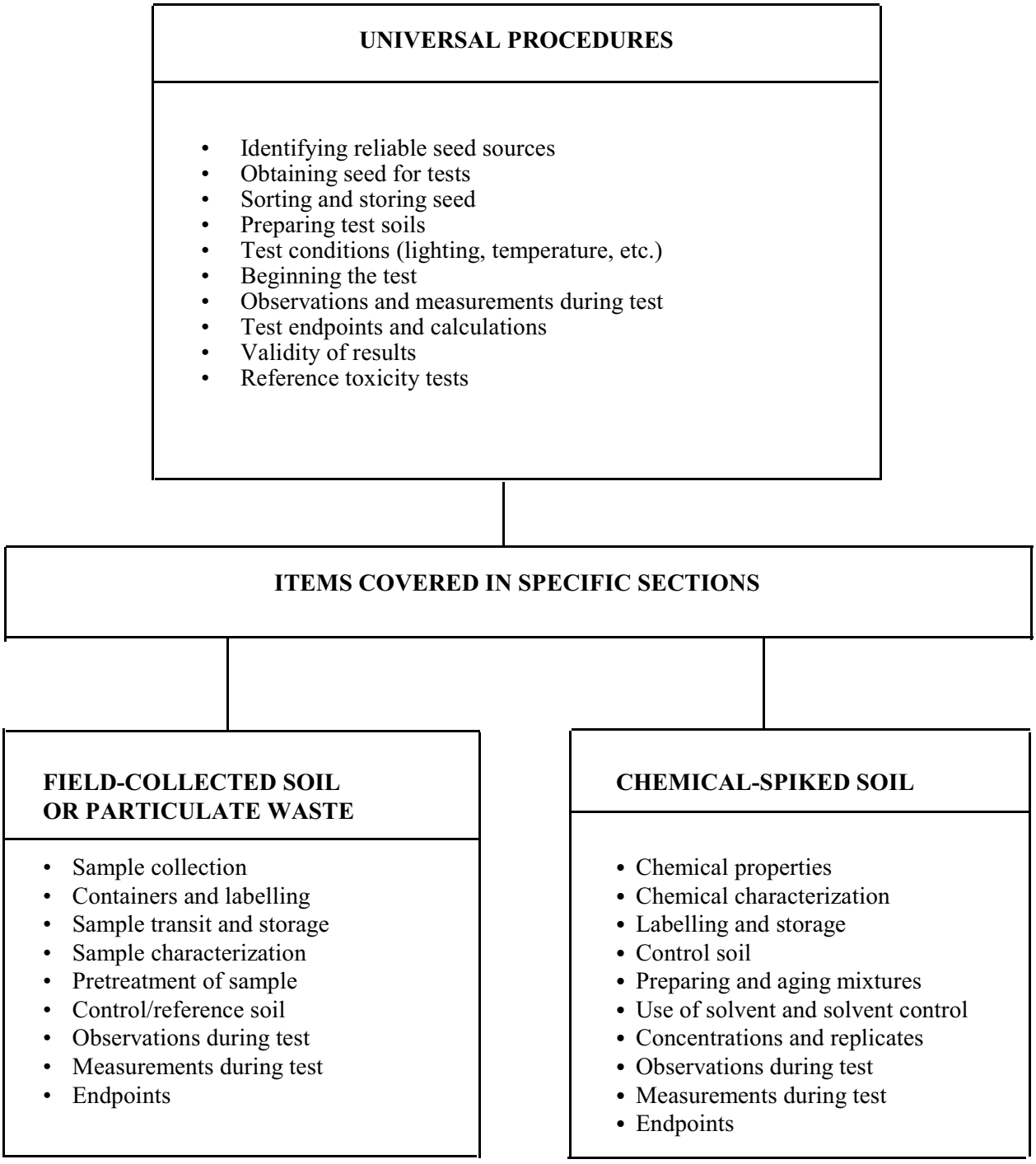


Figure 1 Considerations for Preparing and Performing Soil Toxicity Tests Using Terrestrial Plants and Various Types of Test Materials or Substances

screening was performed with two *negative control soils* (a formulated *artificial soil* and a field-collected natural soil) which were spiked with boric acid (Aquaterra Environmental, 1998a; Stephenson *et al*., 1997; Stephenson, 2003a). The results of these reference tests were used in conjunction with predetermined selection criteria to focus or reduce the list of 30 candidate test species, while ensuring that some of the biological and ecological diversity of plants was incorporated into test species selection (Stephenson, 2003a). *A priori* selection criteria included plant characteristics that were considered:

(1) extremely important [e.g., type of *germination* (*epigeal* or *hypogeal*), class of *angiosperm* (*monocotyledon* vs. *dicotyledon*), and phenology and life history traits (*biennial*, *perennial*, *annual)*];

(2) moderately important (e.g., time to germination, crop vs. non-crop species, and nature of the photosynthetic system); and

(3) less important (e.g., above vs. below-ground crop species).

Each species was also assessed according to six criteria used to evaluate its amenability to the prospective new biological test method and its associated procedures (e.g., test duration, ease of root separation, sufficient *biomass* at the end of a test, seed size, time to emergence, and effect of soil on early seedling *emergence* and *growth*), in addition to its relative sensitivity to boric acid in soil. Species recommended for *definitive* plant tests were those with a known range of sensitivities to boric acid; including those considered to be sensitive (alfalfa, northern wheatgrass, carrot, cucumber, and radish), moderately sensitive (lettuce, timothy, red fescue, and grama grass), and insensitive or tolerant (canola and corn) to the *toxicants* tested (Aquaterra Environmental, 1998a). Timothy was dropped from further investigation because its fragile roots made it difficult to work with, and corn was chosen over canola to represent the "tolerant" species (Stephenson, 2003a).

In February 2003, Environment Canada hosted a workshop in Vancouver, BC, on the toxicological assessment of Canadian soils and development of standardized test methods. One of the recommendations of the workshop participants was to expand the current battery of test organisms in Environment Canada's draft plant method to include more species (EC, 2004b). As a result, numerous soil toxicity tests have been undertaken by Environment Canada to further expand the list of potential test species described herein (EC, 2005b). In addition, corn was dropped from earlier drafts of this test method because it is one of the least sensitive test species, and workshop participants agreed that this "tolerant species" should not be included as an option in the test method document (EC, 2004b).

The twelve terrestrial plant species (and varieties, where applicable) selected for use in this test method are described in detail in the following subsections and summarized in Table 1.

1.2.1 Alfalfa (Medicago sativa L.)

Alfalfa *(Medicago sativa* L.*)*, also called lucerne, is one of the oldest cultivated forage crops in North America and is one of the most palatable and nutritious (i.e., rich in protein, vitamins, and minerals). It has a very high yield compared with that of other crops and is an integral component of many crop rotations because of its ability to fix nitrogen, improve soil structure, and control weeds in subsequent crops (Sullivan, 1992). The distribution of alfalfa is worldwide. In Canada, alfalfa is grown mainly as forage and fodder for livestock; however, alfalfa seeds are also sprouted for human consumption as a vegetable (Munro and Small, 1997). Alfalfa is the most important forage legume in Canada, and is grown in almost all provinces. Most of the Canadian crop is used as bailed hay (2×10^6 hectares), and some is used as silage and pasture ($2–3 \times 10^6$ hectares) (Munro and Small, 1997). Alfalfa has also naturalized in many areas throughout Canada.

Alfalfa is a long-lived, *perennial*, dicotyledenous legume that exhibits *epigeal germination*. It belongs to the family Fabaceae (pea family, also known as Leguminosae) and is classified as an above-ground agricultural crop with a C_3 photosynthetic system. Seeds are a bright olive-green to yellow, and are medium-sized (~2.6 × 1.5 mm) (see Table 1). Alfalfa seed must be placed in contact with moist

Table 1 Characteristics of Plant Species

Plant	Seed Size (mm)	Germination	Monocot vs. Dicot[1]	Seedling Emergence[2] (days)	Life cycle	Soil Type Preference	Tolerance
Alfalfa	2.6 × 1.5	epigeal	dicot	3–4	perennial	loamy, well drained	tolerates drought; winter hardiness; intolerant of flooding, waterlogging, poor soil drainage
Barley	9.0 × 3.4	hypogeal	monocot	2–3	annual	well drained, fertile loams and lighter clay soils; loamy to heavy soils tolerated	tolerates saline soil, heat and drought; does not grow well at pH < 6.0; intolerant of waterlogging
Blue Grama Grass	4.9 × 1.0	hypogeal	monocot	3–5	perennial	fine- to coarse-textured including clay, site, fine loams, sandy loams, sand and gravelly soils	tolerant of cold, drought, and shade; intolerant of salt; seed viability greater at higher temperatures
Carrot	3.6 × 1.5	epigeal	dicot	4–5	biennial	all soil types; grows best in medium- to -light loose, sandy loam soils with good WHC	tolerates a wide pH range (4.2–8.7) but grows best at pH 6.5–7.8; intolerant of drought
Cucumber	7.7 × 3.6	epigeal	dicot	3–4	annual	most well-drained soils; grows best in heavier clay loam or salty loam soils high in organic matter	requires pH at or near neutral with high amount of nitrogen
Durum Wheat	8.0 × 4.0	hypogeal	monocot	2–3	annual	tolerates sandy, loamy and clay soils, but requires well-drained conditions	prefers dry conditions, hot days, cool nights; tolerates wide pH range intolerant of cold and long winters
Lettuce	3.8 × 1.3	epigeal	dicot	3–4	annual	will grow in fine sandy loams, clay soils and muck soils, but prefers soil high in organic matter	requires cool temperatures for germination; optimal growing temp. 15–18 °C; prefers pH 6.0–8.0
Northern Wheatgrass	7.5 × 1.3	hypogeal	monocot	4–5	annual/ perennial	tolerates a range of soil types, but prefers medium- to coarse-textured	tolerant of moderate flooding, but is known for its drought tolerance; prefers basic soils (pH 6.0–9.5)
Radish	2.9 diameter	epigeal	dicot	2–3	biennial	grows well in a variety of soil types	tolerant of low fertile soil; prefers cooler temperatures; prefers neutral soil (pH 6.0–7.0), but can tolerate slightly acidic soils (pH 5.5–6.8); poor salt tolerance

Plant	Seed Size (mm)	Germination	Monocot vs. Dicot[1]	Seedling Emergence[2] (days)	Life cycle	Soil Type Preference	Tolerance
Red Clover	2.0 × 1.5	epigeal	dicot	3–4	perennial/ biennial	well drained highly fertile loam soil; loams, silt loams and heavy soils are better than light sandy or gravelly soils	tolerant of wide pH range (4.5–8.2), but prefers near neutral pH of 6.6–7.6; better than alfalfa at tolerating soils of low pH, low fertility and/or poor drainage; moderately drought tolerant
Red Fescue	6.6 × 0.9	hypogeal	monocot	4–5	perennial	can grow on clay loam, and sandy soils provided moisture is adequate	tolerant of soils that are saline, acidic (pH 4.5), and low in fertility; tolerates moist soils, some waterlogging, cold winters, and some drought
Tomato	3.0 × 2.4	epigeal	dicot	4–5	perennial	light, warm, sandy soils and heavier soils	tolerant of of pH 5.5–7.5; prefers warm days (21–28 °C) and cool nights (15–20 °C); sensitive to low light and adverse temperatures; intolerant of waterlogging or high humidity (over 80%)

[1] Monocot = Monocotyledon.
Dicot = Dicotyledon.

[2] Number of days until seedlings start to appear.

soil to germinate, and for best seedling survival, seeds should be planted approximately 0.6 cm deep. Seedlings are unable to emerge from the soil if planted too deep and emergence is greatly reduced when seeds are planted deeper than 1.3 cm. Seedlings emerge 3–4 days after being planted. Alfalfa has demonstrated a range of emergence in control soils of approximately 70–90% (ESG and Aquaterra Environmental, 2002; Stephenson, 2003a; EC, 2005b).

Alfalfa typically has a deep taproot, although some varieties have different root systems. The roots form nodules in association with *Rhizobium* spp. bacteria, which fix atmospheric nitrogen. Alfalfa is tolerant of drought and exhibits winter hardiness, and although it grows best in loamy, well-drained soils, it is tolerant of soils having a variety of *textures*. The response of alfalfa in toxicity tests with boric acid appears to be unaffected by soil type (Stephenson, 2003a). Alfalfa is intolerant of flooding, *waterlogging*, or poor soil drainage (Sullivan, 1992). In toxicity tests, well-defined concentration-response relationships were observed for exposures to boric acid, copper sulphate, diuron, and petroleum hydrocarbons such as crude oil, condensates and amines in soils (Aquaterra Environmental, 1998a; Aquaterra Environmental and ESG, 2000; ESG, 2001; ESG and Aquaterra Environmental, 2002; EC, 2005b).

1.2.2 Barley (Hordeum vulgare *L.*)

Barley *(Hordeum vulgare* L.*)* is a cereal crop harvested for its grain (beer, food, and fodder) and for straw. It is one of the most ancient of the cultivated grains, with evidence of its cultivation dating back more than 5000 years (Magness *et al*., 1971). Barley is cultivated extensively throughout Canada and is considered to be a highly significant agricultural crop species in both Canada and the United States (US) (Duke, 1983; Bonnell Environmental Consulting, 1994). *Hordeum vulgare* L. is a six-rowed barley with a tough *rachis* or spiked stem (Magness *et al.*, 1971). Barley is a plant species commonly used in Canadian toxicity testing laboratories.

Barley, a C_3, monocotyledonous *annual* with *hypogeal germination*, is a member of the Poaceae (formerly named Graminae) family, also known as the family of "true grasses". Barley is a fast growing, above-ground crop with a seed size of 9.0 × 3.4 mm (see Table 1). Seeds are sown at a depth of 1.3–4.5 cm (no greater than 5 cm). Seedlings are vigorous and emerge within 2–3 days of planting. Barley is reported to be tolerant of saline soils, heat, and drought (Duke, 1983). It is also tolerant of a soil pH range of 4.5–8.3; however, it does not grow well in very acid soils (i.e., below pH 6.0) (McLeod, 1982; Duke, 1983; Stoskopf, 1985). Barley can be grown on many soil types including well-drained, fertile loams and lighter, clay soils. Loamy to heavy soils are tolerated, but *waterlogging* is not (Valenzuela and Smith, 2002).

Although the varieties might differ with respect to their efficacy of germination, of the three varieties tested (var. CDC Buck-huskless, Bedford, and Chapais) var. Chapais consistently germinated and emerged at >96% in both the *artificial soil* and a field-collected *negative control soil* (ESG and Aquaterra Environmental, 2002; Stephenson, 2003a; EC, 2005b). Barley roots are strong and fibrous and can be easily separated from the soil with minimal breakage. Growth is rapid and plants quickly produce large amounts of *phytomass*, which makes barley a good choice for use in soil toxicity tests. Plant metrics (e.g., shoot/root lengths, shoot/root wet/dry masses) generally exhibit a classic concentration-response relationship in soil toxicity tests with petroleum hydrocarbons (e.g., motor gasoline), metals, and pesticides (Aquaterra Environmental and ESG, 2000; ESG, 2000, 2001; ESG and Aquaterra Environmental, 2002).

1.2.3 Blue Grama Grass [Bouteloua gracilis (HBK) Lag. ex Steud.]

Blue grama grass is a densely tufted prairie grass, native to much of North America. It is common in Alberta, east to Manitoba and south through the Rocky Mountains, Great Plains, and Midwest States to Mexico. It is uncommon in the Northwest Territories, British Columbia, and the northeastern United States. Blue grama grass is a valuable and highly palatable forage for domestic livestock as well as deer and elk. It can form dense cover, and as such, is an important soil-building grass (Anderson, 2003).

The use of blue grama grass in laboratory soil toxicity tests is virtually unknown; however, its value as an ecologically relevant species to Canadian ecosystems is evident by its vast distribution.

Blue grama grass, a *perennial* monocot with C_4 metabolism and *hypogeal germination*, is another "true grass", belonging to the Poaceae (formerly named Graminae) family. It has a medium-sized seed (4.9 × 1.0 mm), from which plants are readily established (see Table 1). Seed viability appears to be temperature dependent and has been shown to be greater at higher temperatures (Aquaterra Environmental, 1998a; Anderson, 2003). Emergence occurs within 3–5 days, with seedlings developing rapidly. This grass emerges at rates of 70–89% in various control soils (EC, 2005b). Seedlings develop a single *seminal root* that is short-lived, and therefore, survival depends on the development of *adventitious roots*, which occur approximately 14-days after the seedling emerges. The fibrous root system is dense and shallow, and there are conflicting reports of *rhizome* formation (Anderson, 2003). Grama grass reproduces primarily by tiller and tufts (e.g., *cespitose*) formation. It has a high water use efficiency, which increases under warm climatic conditions and might decrease with increasing water availability (Anderson, 2003).

Grama grass occupies a range of soil types from fine- to coarse-textured, including: clay, silt, fine loams, sandy loams, sand, and gravelly soils. Good growth occurs in the well-drained soils found in open plains, foothills, and mesas (Aquaterra Environmental, 1998a; Anderson, 2003). Grama grass is cold, drought, and shade tolerant; however, it is fairly intolerant of salt. Because of its wide adaptation, ease of establishment, and economic value, grama grass is used extensively for conservation purposes, rangeland seeding, and landscaping (Anderson, 2003). Grama grass exhibited a strong concentration-response relationship with boric acid and copper sulphate when studied for use as a potential test species for inclusion in this test method (Aquaterra Environmental, 1998a; Aquaterra Environmental and ESG, 2000; and EC, 2005b). However, when using this species as a test organism, maximum test temperature and duration are recommended in order to yield adequate *phytomass* for endpoint measurements.

1.2.4 Carrot (Daucus carota *L.*)

Daucus carota contains both wild and domesticated forms of carrot and has numerous variants. There is little agreement on the most appropriate nomenclature for the many forms that have been described. All of the domesticated forms, however, are in the subspecies *sativus* (Hoffm.) Arcangeli (Munro and Small, 1997). The domesticated carrot is a below-ground, market-garden, crop species that is harvested annually. It is one of the most important of cool-climate root crops worldwide, and is one of the most valuable crops in Canadian vegetable production (Munro and Small, 1997). The carrot is cultivated primarily for its enlarged fleshy taproot that is widely consumed raw, cooked, or as juice. The carrot is sometimes used as fodder, as oil, as a sweetening agent, as a coffee substitute and/or in liqueurs (Munro and Small, 1997).

The carrot, a C_3 *biennial dicotyledon*, is a member of the Apiaceae family (previously named Umbelliferae). It has a medium-sized seed (3.6 × 1.5 mm) that should be planted at depths of 1–2 cm (see Table 1). It does well in all soil types; however, it grows best in medium-to-light, loose, sandy loam soils with good *water-holding capacity*. The carrot is extremely sensitive to soil conditions, and good single, thick tap roots can only be produced in soils that permit their easy penetration. The carrot tolerates a wide pH range (4.2–8.7), but grows best in soils with a pH ranging from 6.5–7.8 (Duke, 1983) and mean temperatures of 16–21 °C (Huxley *et al*., 1992; Munro and Small, 1997). The carrot is intolerant of drought. This dicotyledon reproduces biennially by seed, with *epigeal germination*. In the laboratory, carrot seedlings emerged in approximately 4–5 days and percent emergence ranged from 64–87% and 76–86% in reference and artificial control soils, respectively (Aquaterra Environmental, 1998a; EC, 2005b). The toxicity of boric acid to the carrot appears to be unaffected by soil type (Stephenson, 2003a). Concentration-response relationships for both root and shoot growth were classic (i.e., the severity of effect increased with increasing exposure *concentration*), in tests with boric acid and tests with copper sulphate (Aquaterra Environmental, 1998a; Aquaterra Environmental and ESG, 2000; EC, 2005b).

1.2.5 Cucumber (Cucumis sativus L.)

The cucumber *(Cucumis sativus)* is an ancient Old World vegetable that likely originated in India. It is an important crop worldwide and is one of the most important vegetables in Canada, representing about 5% of the value of the fresh vegetable industry (Bonnell Environmental Consulting, 1994; Munro and Small, 1997). The cucumber has a wide distribution in Canada, and is grown in Nova Scotia, central New Brunswick, eastern and western Ontario, central Manitoba, southern British Columbia, and northern Alberta.

The cucumber is a rapid-growing, above-ground, C_3 crop species that is harvested annually. This *dicotyledon* belongs to the family Cucurbitaceae or the "gourd family". It grows well in most well-drained soils; however, it does best in heavier clay loam or salty loam soils that are high in *organic matter* (Munro and Small, 1997). The cucumber prefers higher temperatures, and develops deep root systems. It requires soils with a pH at or near neutral and with high amounts of nitrogen (Munro and Small, 1997). The seed of a cucumber is relatively large (mean seed size of approximately 7.7 × 3.6 mm; see Table 1), enabling the plant to produce more *phytomass* in a short period of time, by relying on internal energy reserves. Seedlings exhibit *epigeal germination* and emerge in 3–4 days. In toxicity tests with boric acid, cucumbers (var. Marketer) exhibited a concentration-response relationship for both root and shoot growth in terms of length and wet mass measurements (Aquaterra Environmental, 1998a; EC, 2005b). Root growth was significantly reduced in response to exposure to boric acid, but the above-ground *biomass* was affected to a lesser degree. Percent emergence in toxicity tests in both reference and artificial control soils ranged from 90–98% and the nature of the soil had no effect on the observed toxicity of boric acid (Aquaterra Environmental, 1998a; EC, 2005b). The cucumber has been shown to be relatively sensitive to both organic and inorganic *contaminants* in soil (Aquaterra Environmental and ESG, 2000; ESG and Aquaterra Environmental, 2002).

1.2.6 Durum Wheat [Triticum durum *(Desf.) or* Triticum turgidum *L. subsp.* durum *(Desf.) Husn. or* Triticum pyramidal *(Percival)*]

Durum wheat, also known as "hard wheat", is an *annual* grass that is planted in the spring and

harvested in late summer. It is the only *tetraploid* species of wheat cultivated today, and is the hardest of all wheats. Compared to *Triticum aestivum* (i.e., bread wheat), far less durum wheat is grown in North America. It accounts for roughly 8% of global wheat production, and leading producers include the European Union, Canada, and the US (Small, 1999). Canada produces some of the highest-quality amber durum wheat in the world, and has an Annual Production Average (APA) of about 4.09×10^6 tonnes. Approximately 80% of the durum wheat grown in Canada between 1992 and 1995 was exported (AFBMI, 1998). On average, durum wheat has a higher protein content than bread wheat. It is grown primarily for the production of pasta products, such as spaghetti and macaroni, and for couscous and bulgar (Small, 1999; Vaughan and Geissler, 1997).

Durum wheat is a *monocotyledon* belonging to the family Poaceae (formerly named Graminae) (i.e., grass family). It is an above-ground cereal crop with a C_3 photosynthetic system, *hypogeal germination* like most cereals, and a fibrous root system. Durum wheat is suited to a dry climate with hot days and cool nights, does well under dry conditions, and has a low resistance to cold and to long winters (Vaughan and Geissler, 1997). Durum wheat is tolerant of many soil types including light (sandy), medium (loamy), and heavy (clay) soil, but requires well-drained conditions. It tolerates a wide range of soil pH. Durum wheat is characterized by its large (8.0 × 4.0 mm), ovate-shaped, amber-coloured seed (see Table 1) which should be planted at a depth of about 2.5 cm (OMAF, 2002). Seedlings emerge in 2–3 days, and germination ranged from 83–92% in toxicity tests involving both artificial and field-collected soils. In addition, durum wheat has demonstrated acceptable concentration-response relationships in toxicity tests with boric acid (EC, 2005b).

1.2.7 Lettuce (Lactuca sativa *L.*)

Lettuce is an above-ground, market-garden, crop plant that is harvested annually. It is the Western World's most popular salad plant with year-round demand. Canada grows over 50 000 tons of lettuce yearly, mostly in Quebec; however, this represents less than 1/5 of the lettuce actually consumed in Canada (Munro and Small, 1997).

Lettuce, a C_3 dicotyledonous *annual*, with *epigeal germination*, is a member of the sunflower family (i.e., Asteraceae, formerly named Compositae). It has a medium-sized seed of 3.8 × 1.3 mm, and seedlings begin to emerge in 3–4 days (see Table 1). Lettuce is a cool-season crop, requiring cooler temperatures for germination. Optimal growing temperatures for lettuce range from 15–18 °C (minimum 7 °C/maximum 24 °C) (Munro and Small, 1997). Lettuce germinates and grows best when water is not limited and when the appropriate light is provided. Care must be taken in choosing varieties for use in toxicity testing, because various *cultivars* of lettuce have substantially different light intensity requirements for optimal *emergence* and *growth*. For example, the seed germination of some cultivars of lettuce is negatively photoblastic (i.e., seeds germinate only in the dark, and white light inhibits germination) (Stephenson, 2003a). Lettuce prefers soil high in *organic matter*, but will grow in various soils including fine sandy loams, loams, clay soils, and muck soils. The ideal soil pH for lettuce ranges between 6.0 and 8.0 (Munro and Small, 1997). The initial tap root can become quite fibrous when the plant is mature. In whole soil toxicity tests, emergence ranged from 75–88% for artificial soil and from 79–94% for field-collected soils (Aquaterra Environmental, 1998a; EC, 2005b). For the Grand Rapid and Butter Crunch varieties used in these studies, toxicity tests with boric acid exhibited acceptable concentration-response relationships for both shoot and root metrics. Lettuce is sensitive, however, to differences in certain physicochemical characteristics of artificial or natural *clean* soil used in a test. A "soil effect" was observed in seedling emergence tests using samples of both artificial soil and clean field-collected soil (Stephenson, 2003a).

1.2.8 Northern Wheatgrass [Elymus lanceolatus *(Scribn. & J.G. Sm.) Gould; formerly* Agropyron dasystachyum *(Hook.) Scribn.*]

Northern wheatgrass *(Elymus lanceolatus)*, also known as thickspike wheatgrass, is widely distributed throughout North America from Alaska, south through Canada, into Northern California. It is common in the northern Rocky Mountains and in the prairies from British Columbia to Ontario (Scher, 2002). Northern wheatgrass is a long-lived, cool-season native grass that is highly beneficial to soil systems. The deep root system provides excellent soil stabilization and strong sod formation (Bonnell

Environmental Consulting, 1994; Scher, 2002). This species of grass is capable of forming "tall-grass prairies" at well-drained sites. It is valued as forage for livestock and wildlife and is commonly used in re-vegetation of oil and gas well-sites, pipeline construction areas, roadsides, and other construction sites. Northern wheatgrass is an important re-vegetation species because it forms tight sod under dry conditions, has good seedling strength, and does well in low-*fertility* soils and at eroded sites (Scher, 2002).

Northern wheatgrass is a non-crop, C_3, *annual* and *perennial monocotyledon* that belongs to the family Poaceae. It is another of the optional species of "true grasses" that may be used in this biological test method. Northern wheatgrass is very winter hardy because of its three-way root system (i.e., creeping roots that reproduce asexually or by vegetative propagation, dense shallow roots to 25 cm, and deep roots to at least 60 cm). This grass can also reproduce sexually by seeds. It is strongly *rhizomatous*, but develops via *hypogeal germination* from a slender seed (7.5 × 1.3 mm) (see Table 1). The seeds generally have good viability (e.g., 95% emergence; Stephenson, 2003a); however, emergence can be low (67–77%) with some batches of seed (EC, 2005b). Seedlings have good vigour, emerging in 4–5 days, and under good conditions, they can experience rapid development. Northern wheatgrass grows on a wide range of soil types, but prefers medium- to coarse-textured soils. It will tolerate moderate flooding, but is known for its drought tolerance (Scher, 2002). It also prefers basic soils (pH of 6.0–9.5). Although little data exist regarding the sensitivity of this species to contaminants, it has been shown to exhibit a strong concentration-response relationship when exposed to boric acid, copper sulphate, diuron, or petroleum hydrocarbons (e.g., condensates, crude oil, motor gasoline) in soil (Aquaterra Environmental, 1998a; Aquaterra Environmental and ESG, 2000; ESG, 2001; ESG and Aquaterra Environmental, 2002; EC, 2005b).

1.2.9 Radish (Raphanus sativus *L.*)

The radish *(Raphanus sativus)* is a minor agricultural crop, but is important and popular in gardens and markets (Bonnell Environmental Consulting, 1994). In Canada, 6000 tons are produced annually, mostly in Ontario, Quebec, and British Columbia (Munro and Small, 1997). Although it is used primarily as a salad vegetable in North America, radishes are used in other parts of the world for the production of soap and a drying oil, and as livestock feed.

The radish is a below-ground, cool-season crop species that is harvested annually for its bulbous edible tap root. It is a C_3 *biennial dicotyledon* belonging to the mustard family (Brassicaceae or Cruciferae). The medium-sized seed (2.9 mm in diameter) germinates rapidly and seedlings begin to emerge within 2–3 days under moist soil conditions (see Table 1). The seed undergoes *epigeal germination* and produces a strong root that is easily separated from the soil (Stephenson, 2003a). Radishes are not demanding as to the soil type, and grow well in a variety of soils (Munro and Small, 1997). Stephenson (2003a), however, found that the nature of the control soil had a significant effect on the performance of the species. In screening tests, 98% emergence of radishes (var. Cherry Belle, Champion) was observed in the *artificial soil*, but only 65% emerged in the field-collected negative control soil (Stephenson, 2003a). In other tests, however, emergence was shown to be consistently high (i.e., >92%) in all soil types (i.e., artificial as well as sandy, silt, and clay loam soils) (ESG and Aquaterra Environmental, 2002; EC, 2005b). Radish is tolerant of soils with low fertility. It prefers cooler temperatures and neutral soil (pH 6.0–7.5), but can tolerate slightly acidic soils (pH 5.5–6.8). Radish has a low tolerance of salty soils. Stephenson (2003a) found that a concentration-response relationship for either shoot or root length was not observed in screening tests with boric acid, and that both of these metrics were influenced significantly by the nature of the control soil. The radish proved to be relatively sensitive to metals and pesticides in other studies (Aquaterra Environmental and ESG, 2000; ESG and Aquaterra Environmental 2002). This is one of the two test species (lettuce is the other species) currently recommended in most regulatory test *protocols* for measuring soil toxicity.

1.2.10 Red Clover (Trifolium pratense *L.*)

Red clover *(Trifolium pratense)* is grown widely across North America and occurs from coast-to-coast in Canada, both as a cultivated crop and naturally (Bonnell Environmental Consulting, 1994). It is extensively grown for pasturage, hay, and green manure, and is the most commonly planted forage

legume, after alfalfa. Compared to alfalfa, however, red clover has less digestible protein, slightly more total digestible nutrients, and a slightly higher net energy value (Duke, 1983; USDA-NRCS, 2000). Red clover is also valued as a highly significant species for maintaining soil structure, and is used frequently in reclamation studies (Bonnell Environmental Consulting, 1994).

Red clover is a short-lived, C_3, *perennial dicotyledonous* legume that can be grown under conditions which are either too wet or too acidic for alfalfa (OMAF, 2002). Under some conditions (i.e., warmer climates) red clover is grown as a *biennial*. Red clover, like alfalfa, is a member of the pea family (Fabaceae). It has a relatively small seed (2.0 × 1.5 mm; see Table 1) has good seedling vigour, and is relatively easy to establish (USDA-NRCS, 2000). Red clover has *epigeal germination* and has demonstrated good emergence under laboratory conditions in tests using either artificial soil (68–90%) or samples of *clean* field-collected soils (88–92%) (EC, 2005b). Emergence typically begins 3–4 days after planting. It grows best on well-drained, highly fertile loam soil, but has also adapted to wetter soils. For red clover, loams, silt loams, and even heavy soils are better than light sandy or gravelly soils (Duke, 1983). It is tolerant of a wide pH range (4.5–8.2); however, this plant species prefers a near-neutral pH for nodulation and is most productive on soil that is within a pH range of 6.6–7.6. Red clover is better than alfalfa at tolerating and growing on soils of low pH or those with low fertility and/or poor drainage. Red clover has a deep tap root and is moderately drought tolerant (USDA-NRCS, 2000). Roots *nodulate* naturally from free-living *rhizobia*. Red clover is not commonly used in laboratory toxicity tests as yet, but it has demonstrated good concentration-response relationships in tests with boric acid (EC, 2005b).

1.2.11 Red Fescue (Festuca rubra *L.*)

Red fescue has a wide range across the Northern Temperate Zone, occurring throughout Canada from British Columbia to Newfoundland (Bonnell Environmental Consulting, 1994). It is a valuable species of forage grass in Alberta, where it grows better in poor soils than bluegrass or timothy; it can also out-compete alfalfa. In particular, Creeping (var.) red fescue is a dense, sod-forming grass that establishes and spreads vigorously on most soil types (OMAF, 2002). Its solid root system and thick top-growth make it an excellent grass for stream-bank or grass waterway protection. It is noted for its extended growth period and its retained nutritional value in the fall (OMAF, 2002). It is considered a valuable stabilizer, and an excellent soil-and-sod builder.

Red fescue is a long-lived, C_3, monocotyledonous species, that is a member of the grass family (Poaceae, formerly named Graminae). Red fescue is a cool-season, *perennial*, ground cover that is drought resistant, saline tolerant, tolerant of acidic soils, and hardy of cold winters (Walsh, 1995). It reproduces by a relatively large seed (6.6 × 0.9 mm), and exhibits *hypogeal germination* (see Table 1). Red fescue can also grow vegetatively by *rhizome* formation and can develop deep, extensive systems of fibrous roots. Seed viability is good, ranging from 78–95% (Aquaterra Environmental 1998a; EC, 2005b); however, vigour is moderate. Seedlings typically emerge 4–5 days after planting. Red fescue tolerates moist soils and some *waterlogging*. It can grow on clay, loam, and sandy soils, provided that the moisture is adequate; however, this species can tolerate some drought. It is also tolerant of low fertility and low pH (4.5) in soils (Walsh, 1995). A concentration-response relationship for shoot and root metrics was demonstrated for red fescue in screening test with boric acid (Aquaterra Environmental, 1998a; EC, 2005b).

1.2.12 Tomato (Lycopersicon esculentum *Mill.*)

The tomato *(Lycopersicon esculentum)* is probably Canada's most popular home-garden vegetable. It is the second-most consumed vegetable worldwide per capita, next to the potato (Munro and Small, 1997). About 2×10^6 hectares of tomatoes are planted annually worldwide. The Canadian domestic supply represents about 60% of all tomatoes used in the country and more than 10% of the commercial value of fresh vegetables consumed in Canada (Munro and Small, 1997).

The tomato is a C_3, tropical *perennial* that is grown as an *annual* in temperate parts of the world. It is a dicotyledon belonging to the family Solanaceae (also known as the nightshade or potato family). Tomatoes are grown in light, warm, sandy soils for early crops; however, heavier soils are best for maximum production. The tomato is intolerant of *waterlogging* or high humidity (over 80%), both of which promote disease, but it will tolerate a pH

range of 5.5–7.5 (Munro and Small, 1997). Its medium-sized seed (3.0 × 2.4 mm) demonstrates *epigeal germination* and has a fairly good viability rate (74–95%) (Aquaterra Environmental, 1998a; EC, 2005b). Seedlings begin to appear 4–5 days after planting (see Table 1). When grown from seed, the tomato forms a strong taproot. However, injury to the tap root during transplanting or potting of seedlings tends to modify the natural taproot into a more fibrous one. The tomato needs moderately high daytime temperatures (21–28 °C) and moderately cool nighttime temperatures (15–20 °C) for optimal growth (Munro and Small, 1997). The tomato is sensitive to low light and adverse temperatures, and has demonstrated a good concentration-response relationship response in toxicity tests with boric acid (EC, 2005b).

1.3 Historical Use of Terrestrial Plants in Toxicity Tests

The development of biological test methods for soil toxicity testing lags behind that for other media (e.g., water and sediment) (Bonnell Environmental Consulting, 1994). This delay is partially due to the fact that research and regulators have been focussed on the aquatic environment, and partially due to the fact that soil is a complex medium with many problems inherent in its lack of homogeneity. The variety of exposure routes available to investigators (e.g., via pore water, soil vapours, or direct contact with soil particles), coupled with the high cost of running soil toxicity tests, have often led practitioners to rely on extrapolations from aquatic test methods to soil-based exposures (Bonnell Environmental Consulting, 1994).

The use of pesticides in agriculture began in the late 1940s, and by the late 1960s and early 1970s, became routine. This led to the need to assess the effects of organic chemical pesticides on commercial agricultural crop species (Kaputska *et al*., 1995; Boutin and Rogers, 2000). Assessment of soil quality before the 1980s primarily involved evaluating the physicochemical properties of soil, and not until the 1980s did the initial use of standardized biological test methods for measuring soil toxicity emerge from agencies responsible for pesticide registration and application [e.g., the United States Environmental Protection Agency (USEPA), and the Office of Pesticides Programs (Holst and Ellanger, 1982)].

The first standardized whole-soil toxicity test with terrestrial plants, applicable to both pesticide and non-pesticide exposures in artificial soil, was a seedling emergence test guideline (#208) published by the Organization for Economic Co-operation and Development (OECD, 1984a). This method, however, was developed to assess *chemical-spiked soils* only. In 1989, the USEPA recommended test methods for the toxicity assessment of contaminated *site soils*, whereby the *contaminated* soil was amended with a *clean* control soil in a dilution series (USEPA, 1989). Since the establishment of the joint European Economic Community (EEC)/OECD guidelines, several other agencies such as the International Standards Organisation (ISO) and, in the US, the American Society for Testing and Materials (ASTM) have also developed whole-soil toxicity test methods for selected species of terrestrial plants exposed to samples of chemical-spiked soil and/or contaminated site soil (ISO, 1993a, 1995; ASTM, 1999b).

The toxicity of *site soils* became a "new" concern in the mid 1980s, and regulatory programs such as SUPERFUND in the United States, and the National Contaminated Sites Remediation Program (NCSRP) in Canada, were established to address the urgent need for guidance on the assessment and *remediation* of high-priority contaminated sites. Under the NCSRP, a review of existing whole-organism *bioassays* for soil, freshwater sediment, and fresh water (Keddy *et al*., 1995) was conducted to lead to the establishment of a suite of tests that could be used immediately for contaminated-site assessment in Canada (Bonnell Environmental Consulting, 1994). Keddy *et al.* (1995) concluded that most of the existing methods or procedures for measuring the toxicity of samples of soil from contaminated sites were inadequate for proper ecotoxicological assessment, and recommended that attempts be made to develop a suite of standardized biological test methods for soil that used test species and conditions applicable to Canadian soil ecosystems. The Canadian Council of Ministers of the Environment (CCME) published a framework for *ecological risk assessment* (ERA) in 1994 (CCME, 1994) which had a subsequent impact on the management of

contaminated sites (CCME, 1996, 1997). The ERA approach, which relied on the results of single-species toxicity tests, led to the need to develop reliable, reproducible, and realistic soil toxicity tests with ecologically relevant terrestrial test species for the assessment of contaminated site soils (Bonnell Environmental Consulting, 1994). In the late 1990s, biological assessments in the form of toxicity testing were becoming a useful complement to chemical analyses, especially when applied to site-specific *risk assessments*.

Today, plants are widely used as test organisms in single-species toxicity tests intended to measure the toxicity of pure chemicals, chemical products, or samples of soil contaminated or potentially contaminated with chemicals in the field or (for experimental purposes) in the laboratory. In Canada, results of soil toxicity tests are used to:

(i) derive national soil quality criteria,

(ii) establish site-specific, risk-based, cleanup objectives (e.g., remediation targets), and

(iii) assess the efficacy of remediation technologies (Stephenson *et al*., 2002).

Extensive reviews on the use of plant toxicity tests as "ecological assessment tools" for appraising the toxicity of contaminated or potentially contaminated soils have been carried out (Wang, 1991, 1992; Wang and Freemark, 1995; Kaputska, 1997; Meier *et al*., 1997; Saterbak *et al*., 1999). In some cases, standard methods have been modified or unique methods have been developed in order to obtain relevant data (Pfleeger *et al*., 1991; Sheppard, 1994; Chaineau *et al*., 1997). Data-base reviews have been summarized in reports discussing trends of plant toxicity to various contaminants (Kenaga, 1981; Miller *et al*., 1985; Boutin and Rogers, 2000). *Toxic* effects of plant exposure to contaminated soils have been documented in laboratory studies involving samples of soil spiked or contaminated with:

- pesticides (Fletcher *et al*., 1995, 1996; Boutin *et al*., 2000, 2004),
- metals (Godbold and Hüttermann, 1985; Kaputska *et al*., 1995; Kjaer and Elmegaard, 1996; Rader *et al*., 1997; Kjaer *et al*., 1998; Redente *et al*., 2002; Lock and Janssen, 2003),
- petroleum hydrocarbons (Chaineau *et al*., 1997; Wong *et al*., 1999;), and
- other chemicals (Siciliano *et al*., 1997; Kalsch and Römbke, 1999).

Various plant species have been recommended for *phytotoxicity* testing by different agencies (See Appendices E and F; and ASTM, 1999b). The test species most commonly recommended among international agencies include: lettuce, cabbage, cucumber, soybean, oat, perennial ryegrass, corn, tomato, rice, and carrot. Fletcher *et al*. (1985; 1988) reviewed the PHYTOTOX data base and provided a summary of the most commonly used terrestrial plants. These plants included wheat, pea, tomato, oats, beans, apple, soybean, corn, and barley.

The number of species recommended for use in a *test battery* depends primarily on the purpose of the study and the regulatory requirements. The ISO (1995) recommends a minimum of two species, OECD (1984a) recommends a minimum of three species, and ASTM (1999b) recommends a minimum of five species. The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) and the Toxic Substances Control Act (TSCA), as well as the latest USEPA Office of Prevention, Pesticides and Toxic Substances (OPPTS) draft (1996) all recommend 10 species of terrestrial plants for inclusion in a test battery. The OECD (2000) draft test method recommends three species for testing general chemicals, and 6–10 species for testing crop protection products. The proposed Canadian guidelines for non-target plant testing with chemical pesticides (Boutin *et al*., 1995) suggest 10 species for non-herbicide testing and 30 species for herbicide testing. The recommended ratio of *monocotyledons* to *dicotyledons* to be used in a *test battery* is generally 1:2, and is fairly consistent among international agencies. Boutin and Rogers (2000) conducted an extensive review of Canadian and American data bases and found that monocotyledons show similar sensitivities, whereas dicotyledons vary in their sensitivities. In a test battery, therefore, it might be appropriate to test fewer monocotyledons than the commonly

recommended monocotyledon to dicotyledon ratio of 1:2 (Boutin and Rogers, 2000). Cole *et al*. (1993) and Brown and Farmer (1991) also provide rationale for selecting a variety of test species for testing the effects of pesticides on non-target plant species.

Many plant species and numerous phytotoxic assessment endpoints have been used to characterize the effects of toxicants on vegetation (Markwiese *et al*., 2000). To date, the seedling germination test and the root elongation test are the *acute phytotoxicity* tests most widely used (Kaputska, 1997). Unfortunately, the germination test is relatively insensitive to many substances, primarily because the embryonic plant survives using the nutritional reserves stored in the seed and is therefore effectively isolated from the environment (Kaputska, 1997). In typical root elongation tests, roots are exposed to water extracts and soluble test soil constituents, which do not involve any exposure to whole soil.

The seedling emergence test differs from seedling germination tests, in that different endpoints are measured. Most seedling emergence tests have been modelled after the OECD Terrestrial Plant Growth Test (OECD, 1984a), in which seeds of recommended test species are exposed to potentially contaminated site soils, or to a dilution series (i.e., *site soils* amended with control soils), followed by the measurement of the number of seedlings that emerge from the soil to a minimum height of 3 mm. Generally, seedling emergence is not as sensitive an endpoint as growth metrics (e.g., shoot and root lengths and weights) that can be obtained from early seedling growth tests. These early seedling emergence-and-growth tests overcome some of the deficiencies of the seed germination and root elongation tests discussed earlier (Kaputska, 1997; Stephenson *et al*., 2002). The ASTM (1999b) has developed an early seedling growth test with a test duration that is relatively longer than the seedling emergence test (i.e., >14 days). Its measurement endpoints include shoot and root length, shoot and root wet and dry mass, and seedling emergence and seedling survival at the end of the test (Stephenson *et al*., 2002). The OECD is currently revising their biological test method to include both a test for seedling emergence and growth, as well as a test for vegetative vigour whereby the test substance is applied to the leaves and above-ground portions of the test organisms (OECD, 2000a). The ASTM has also included a life-cycle test with *Brassica rapa* (a variety of turnip, that has been genetically modified for rapid assessment) in an annex of their standard guide (ASTM, 1999b). It is a test that goes from seed-to-seed, thereby covering the complete life cycle of the test organism.

The methodology documents summarized in Appendix E have been used as guidance in developing Environment Canada's standardized biological test method for performing a test that measures the *toxic* effects of prolonged exposure to *chemical-spiked soil* or *site soil* on the emergence and growth of terrestrial plants. This (Environment Canada's) new biological test method, is defined herein.

Test Organisms

2.1 Species

Test organisms to be used in this biological test method must be selected from the specified group of 12 terrestrial plant species listed here. Acceptable choices for plant species include five *monocotyledons* and seven *dicotyledons*, as follows:

monocotyledons
- barley *(Hordeum vulgare)*
- blue grama grass *(Bouteloua gracilis)*
- durum wheat *(Triticum durum)*
- northern wheatgrass *(Elymus lanceolatus*; formerly named *Agropyron dasystachyum)*
- red fescue *(Festuca rubra)*

dicotyledons
- alfalfa *(Medicago sativa)*
- carrot *(Daucus carota)*
- cucumber *(Cucumis sativus)*
- lettuce *(Lactuca sativa)*
- radish *(Raphanus sativus)*
- red clover *(Trifolium pratense)*
- tomato *(Lycopersicon esculentum)*

For some plant species, specific varieties are recommended for use in this test method. The varieties, where named, are based on those that have been used successfully in the development of the test method described herein, and include: barley (var. Chapais), durum wheat (var. Durum), red fescue (var. Creeping), carrot (var. Royal Chantenay), cucumber (var. Marketmore 76 or var. Marketer), lettuce (var. Buttercrunch or var. Grand Rapid), radish (var. Champion or Cherry Belle), and tomato (var. Heinz 1439).

The selection of multiple test species should depend on the species' sensitivity to the substance or material being tested, if known. The number and type of species selected for comprehensive studies (i.e., multi-species effects) will depend on the purpose of the study and the regulatory requirements under which the tests are being conducted. The ratio of monocotyledons to dicotyledons selected will vary depending on the study objectives.[2] The ecological, agricultural, and physiological significance of the 12 terrestrial plant species to be used in this biological test method are summarized in Section 1.2.

The test must be started using certified seed (i.e., seed is certified for purity and percent germination).[3] The plant seed used for testing must be from the same lot number for each of the individual plant species, and should be free of fungicides, insecticides, repellents, or other chemical pretreatment agents. It is preferable to use untreated seed; however, if untreated seed is unavailable and/or the objectives of a specific test would be better met with the use of treated seed, then treated seed may be used.[4] Details on seed purchasing, sorting, storing, and condition are provided in Sections 2.2, 2.3, 2.4, and 2.5.

---

[2] A summary of test species and ratio of *monocotyledons* to *dicotyledons* recommended by various international agencies is provided in Section 1.3 and listed in Appendices E and F. Criteria to consider when selecting species for a *test battery* include: evaluation of response to reference toxicant (emergence, shoot/root length, and shoot/root mass); performance and sensitivity in *range-finding* tests (i.e., % emergence in *control*, test duration, ease of root separation, sufficient *biomass* at end of test, time to emergence, effect of soil on growth); type of germination *(epigeal/hypogeal)*; monocotyledonous or dicotyledonous; crop or non-crop species; type of *seed pretreatment*; nature of the photosynthetic system; source availability and quality of seed; type of root formation; phenology and life history traits; and critical variable requirements (e.g., pH, nutrients) (ESG, 2002).

[3] Genetically modified *cultivars* should not be used for testing unless they are of specific concern (e.g., they are at *risk* of exposure at a specific site).

[4] If treated seeds are used for testing, it is recommended that side-by-side tests be carried out with non-treated seed versus treated seed, to determine the effect of using treated seed on the test results.

Each plant species has unique characteristics that affect its performance in a toxicity test; therefore, certain test procedures and conditions (i.e., number of seeds used to initiate a test, test duration, and test validity criteria) are modified on a species-specific basis to accommodate these requirements (see Sections 4.2, 4.3, and 4.4).

2.2 Source

Seeds used to initiate a soil toxicity test should be obtained from commercial seed companies or government seed banks. When purchasing certified seed, the supplier should provide proof that the seed has not been pretreated with any substances. The following information should be obtained when purchasing seed:

- species (Latin and common names),
- variety,
- grade,
- year of collection,
- packet size (g or kg),
- lot #,
- *cultivar*,
- rating for % germination,
- date of germination rating,
- date of purchase,
- shelf life, and
- name of supplier.

The date the seed package is opened should also be recorded. Seed should generally be purchased at least annually [5], preferably within one year of the date of the seed germination rating; however, a given lot of seed may be used as long as the seed can meet the *control* performance criteria described herein (see Section 4.4), and provided that the sensitivity of the seed does not change significantly over time as determined by the results of *reference toxicity tests* (see Section 4.9).

[5] This recommendation is only a general guideline since seed viability will vary from year to year. It might be advantageous in certain instances, therefore, to use older seed with a relatively high (known) emergence rate vs. newly purchased seed with a lower (and/or unknown) emergence rate.

Sources that have been used to secure quality seed for toxicity testing include:

William Dam Seeds Ltd.
Box 8400
Dundas, ON
Canada L9H 6M1
Phone: (905) 628-6641
Fax: (905) 627-1729
Web site: http://www.damseeds.com
Species: alfalfa, carrot, cucumber, lettuce, radish, red clover, red fescue, tomato

Rosebank Seed Farms Ltd.
7340 Perth Line 24
RR #2, Staffa, ON
Canada N0K 1Y0
Phone: 1-888-289-9934
Fax: (519) 345-9930
Species: barley

Pickseed Canada Inc.
Box 3230
Sherwood Park, AB
Canada T8A 2A6
Phone: (780) 464-0350
Fax: (780) 464-0305
Web site: http://www.pickseed.com
Species: northern wheatgrass

Alberta Nursery & Seeds Ltd.
Box 20
Bowden, AB
Canada T0M 0K0
Phone: (403) 224-3544
Fax: (403) 224-2455
Web site: http://www.gardenersweb.ca
Species: radish, red fescue

Prairie Moon Nursery
Route 3, Box 1633
Winona, NM
USA 55987-9515
Phone: 1-866-417-8156
Fax: (507) 454-5238
Web site: http://prairiemoonnursery.com
Species: blue grama grass

Early's Farm & Garden Centre Inc.
2615 Lorne Ave.
Saskatoon, SK
Canada S7J 0S5
Phone: 1-800-667-1159
Fax: (306) 931-7110
Web site: www.earlysgarden.com
Species: carrot, lettuce, radish, northern wheatgrass

David T. Gehl, Seed Increase Unit
Sustainable Production Systems
Indian Head Research Farm
Agriculture and Agri-Food Canada
P.O. Box 760
Indian Head, SK
Canada, S0G 2K0
Phone: (306) 695-2274
Fax: (306) 695-3445
E-mail: gehl@agr.gc.ca
Species: durum wheat

Prairie Habitats
Box 10, Argyle
Manitoba, Canada, R0C 0B0
Web site: http://www.prairiehabitats.com
Species: blue grama grass

A current list of seed suppliers can be obtained by contacting:

Soil Toxicology Laboratory
Environmental Technology Centre
Environment Canada
335 River Road
Ottawa, ON, K1A 0H3
Phone: (613) 990-9544

2.3 Seed Sorting and Preparation

The seeds of a given species vary in size, shape, and, in some cases, colour. These differences in external features of the seed are often associated with different rates of germination or even different germination requirements (ASTM, 1999b). To minimize variance in test results, the seed should be "hand-sorted" or screened[6] to ensure uniformity in size, colour, and "quality". Quality of seed refers to those seeds without a blemished seed coat or irregular shape. Separation of broken or damaged seeds, empty *hulls*, and other vegetative debris from the seed lot is important. It is especially necessary to sort the seed of northern wheatgrass and barley, since for these species there might be a number of hulls without seed. Some of the test species vary in size, and it is important to use seed that is as uniform as possible to reduce the variation within and among the test vessels (Aquaterra Environmental, 1998a). Uniformity in colour of seed might be a consideration for some species such as alfalfa.[7] Seeds that have evidence of fungal contamination on the seed coat, or seeds that appear to be damaged, must be discarded.

2.4 Seed Storage

The seeds should be kept in their original paper packages and stored in the dark, in labelled, sealed containers (e.g., zip-lock bags) at 4 ± 2 °C. Test seed must remain refrigerated until the day of test initiation (Day 0), at which time the seed must be removed from the refrigerator and brought to room temperature. Seed must not be stored in a freezer since this could cause damage to the seeds (e.g., damage due to dehydration, or that due to expansion of moisture during freezing and in some instances the resultant splitting of the seed).

2.5 Seed Condition

The sensitivity of each new lot of seed used in a *definitive* soil toxicity test must be measured using a

---

[6] Several wire-mesh sizing screens can be nested, with the one containing the largest holes on top and those with successively smaller holes in sequence below. Pour seeds into the top screen and shake the stack of screens until all the seeds are separated. The size class containing the most seeds should be selected for testing. The other size classes might be stored separately for future testing (USEPA, 1989).

[7] Dark-coloured alfalfa seeds typically have low rate of germination (~10%), while the light-coloured seeds have a high percent germination (~90%) (ASTM, 1999b).

7- or 10-day (i.e., depending on the species) reference toxicity test (see Section 4.9). Ideally, a reference toxicity test should be performed together with each definitive soil toxicity test. However, laboratories routinely undertaking soil toxicity tests may choose instead to conduct routine reference toxicity tests (i.e., at least once every two months) using a portion of each lot of seed used for definitive soil toxicity tests. All tests with a *reference toxicant*(s) should be performed using the conditions and procedures outlined in Section 4.9. Species-specific, test-related criteria used to judge the validity of a particular definitive soil toxicity test (and, indirectly, the condition of the lot of seed used in the test), based on the performance of the test organisms in the negative control soil, are given in Section 4.4.

Test System

3.1 Facilities and Apparatus

Tests must be performed in an environmental chamber or equivalent facility having acceptable temperature and lighting control (see Section 4.3).[8] The test facility should be well ventilated to prevent personnel from being exposed to harmful fumes, and it should be isolated from physical disturbances or any contaminants that might affect the test organisms. The area used to prepare *test soils* should also be properly ventilated.

The test facility should be isolated from the area where samples are stored or prepared, to prevent the possibility of contamination of test vessels and their contents from these sources. The ventilation system should be designed, inspected, and operated to prevent air within the sample handling and storage facilities, or those where chemicals are processed or tested, from contaminating the area of the laboratory where tests are conducted.

The mean air temperature of the test facility must be maintained at 24 ± 3 °C . However, for those test facilities with growth chambers that can accommodate a cyclical temperature change, it is recommended that the temperature be maintained at a mean air temperature of 24 ± 3 °C during the day, and 15 ± 3 °C at night. The relative humidity of the test facility should be maintained at a minimum of 50%.[9]

Any construction materials that might contact the organisms, water, or test vessels within this facility must be nontoxic and should minimize sorption of chemicals. Borosilicate glass, nylon, high-density polyethylene, high-density polystyrene, polycarbonate, fluorocarbon plastics, Teflon™, Nalgene™, porcelain, fibreglass, and type 316 stainless steel should be used whenever possible to minimize chemical sorption and leaching. The use of *toxic* materials including copper, zinc, brass, galvanized metal, lead, and natural rubber must be avoided.

The test facility must have the basic instruments required to monitor the quality (e.g., temperature and pH) of the *test soil* and associated *test (hydration) water*. Additionally, the laboratory should be equipped to facilitate prompt and accurate analysis of the *moisture content* of test soils. Equipment requirements include a drying oven which can be set at 90 °C for drying test organisms and 105 °C for drying soils, a weighing balance accurate to the nearest 0.1 mg, a light meter, and a pH meter. Safety apparatus, including a respirator with dust protection, gloves, laboratory clothing, and glasses for eye protection, are required when preparing mixtures and aliquots of test soil.

All test vessels, equipment, and supplies that might contact *site soils*, *test soils*, *test (hydration) water*, *stock solutions*, or test solutions must be clean and rinsed with *de-ionized* or *distilled* water (i.e., *test water*), before being used. All nondisposable materials should be washed after use. The following cleaning procedure is recommended (EC, 1997a, b, 2001, 2004c):

1. soak in tap water (with or without detergent added) for 15 minutes, then scrub with detergent or clean in an automatic dishwater;

2. rinse twice with tap water;

3. rinse carefully with fresh, dilute (10%, v:v[10]) nitric (HNO_3) or hydrochloric acid (HCl)

[8] Greenhouses are not considered acceptable alternatives for test facilities since they are typically too variable in terms of controlling light, temperature, and humidity conditions.

[9] The relative humidity of the test facility will affect the frequency of soil hydration required throughout the test. If the humidity is low (i.e., <50%), soils in the test vessels will dry out faster and more frequent watering will be necessary.

[10] To prepare a 10% solution of acid, carefully add 10 mL of concentrated acid to 90 mL of de-ionized water.

(metal-free grade) to remove scale, metals, and bases;

4. rinse twice with de-ionized water (or other *test water*);

5. rinse once with full-strength, pesticide-grade acetone to remove organic compounds and with reagent-grade (e.g., HPLC grade, ≥98.5% purity) hexane for oily residues (use a fume hood);[11]

6. allow organic solvent to volatilize from dishware in fume hood and rewash with detergent (scrub if necessary); and

7. rinse three times with de-ionized water (or other *test water*).

Test vessels and apparatus that might contact soil or *test (hydration) water* should be thoroughly rinsed with test (hydration) water, immediately before being used in the test.

3.2 Initial and Definitive Tests

3.2.1 Initial Tests

Before *definitive* plant toxicity tests are performed for the first time by a testing laboratory, it is recommended that a minimum of five *control* performance tests with one or more samples of uncontaminated natural or *artificial soil* intended (or under consideration) for use in one or more definitive soil toxicity tests as a *negative control soil* (see Section 3.4) be undertaken by laboratory personnel. Additionally, a minimum of five reference toxicity tests should be performed using one or more samples of a candidate artificial or natural negative control soil intended for routine use in conjunction with definitive soil toxicity tests (see Section 4.9). These initial tests are recommended to confirm that acceptable performance of each test species can be achieved in a candidate natural or artificial negative control soil using that laboratory and the procedures specified in this report.

The conditions and procedures used to perform these initial tests with *negative control soil* should be identical and according to Section 4. The conditions and procedures used to perform these initial tests with one or more *reference toxicant*s should be identical and according to Section 4.9. Each set of initial tests with *negative control soil* or *reference toxicant(s)* should be performed on each terrestrial plant species intended for use in future definitive toxicity tests.

Data from the control performance tests (≥5) must show that the criteria for test validity (see Section 4.4) can be met for the intended test species using a natural or *artificial soil* intended for use as negative control soil in a definitive soil toxicity test. Data from the initial reference toxicity tests (≥5) should be compared by calculating and appraising the magnitude of the *coefficient of variation* (CV) for the respective series of tests and endpoint values (see Section 4.9).

3.2.2 Definitive Test

Test vessels to be used in *definitive* tests must be inert to test and reference substances or *contaminant* mixtures (i.e., the test or reference substances, or mixtures thereof, should not adhere to or react in any way with the test vessel). The volume of the vessel should be sufficiently large to accommodate seedling growth for the duration of the test. It is important that the size, shape, colour, and composition of the vessel be appropriate for the plant species chosen. The vessels should have a sealable lid and should not interfere with light quality within. The test vessel recommended for emergence-and-growth tests with terrestrial plants described herein is a 1-L clear polypropylene container, with a clear polypropylene lid.[12]

---

[11] Rinsing Plexiglas™ with acetone or hexane is not recommended, since the Plexiglas™ can become pitted and etched by these solvents and can turn from transparent to opaque.

[12] Advantages of using the 1-L polypropylene vessels include: 1) they do not need to be washed and are disposable; 2) they are clear so that water pooling on the bottom of the containers (i.e., soil saturation has been exceeded) can be observed when watering from above; 3) they are relatively inexpensive and readily available; 4) they can accommodate from 300 mL to close to 1 L of soil without changing the ratio of surface area to volume substantially, because of their tapered shape; 5) they come in a variety of sizes (e.g., 500 mL and 1000 mL) and have snap-on lids that are easily removed and replaced for

Alternatively, 1-L glass jars sealed with transparent lids (e.g., hinged glass lids) may be used.[13] The vessels must be covered for the first seven days of the test or until the plants reach the top of the container, whichever comes first.

3.3 Lighting

Tests should be illuminated using full-spectrum fluorescent or equivalent lighting. The light fluence rate, measured adjacent to the level of the soil surface, must be 300 ± 100 µmol/($m^2 \cdot s$) (i.e., equivalent to 18 750 ± 6250 lux). Since light intensity tends to vary in a given space, it should be measured at several points within the testing area. The light fluence rate within the testing area should not vary by more than ± 15% of the selected light fluence rate.[14]

3.4 Negative Control Soil

Each soil toxicity test must include *negative control soil* as one of the experimental *treatments*. Negative control soil is essentially free of any contaminants that could adversely affect the performance of plants during the test. The use of negative control soil provides a measure of test acceptability, evidence of the health and performance of the test organisms, assurance as to the suitability of the test conditions and procedures, and a basis for interpreting data derived from the *test soils*.

A soil toxicity test may use *clean* (uncontaminated) natural soil and/or *artificial soil* as the negative control soil. The selection of an appropriate negative control soil depends on considerations such as the study design, physicochemical characteristics of the *test soil*(s), and the availability of suitable *clean* natural soil with acceptable properties.[15] There should also be prior experimental evidence that the soil chosen for use as negative control soil will sustain seedling emergence and growth that consistently and reliably meet the criteria for test validity defined herein (see Section 4.4).

The biological test method described herein has been developed and tested using five negative control soils with diverse physicochemical characteristics (Aquaterra Environmental, 1998a; Stephenson *et al*., 1999a, b, 2000a; Aquaterra Environmental and ESG, 2000; ESG, 2000, 2001, 2002; ESG and Aquaterra Environmental, 2002; EC, 2005b). These *clean* soils included one *artificial soil* and four natural soils (i.e., samples of sandy loam and silt loam agricultural soils from southern Ontario, a clay loam prairie soil from Alberta, and a forest loam soil from the Canadian Shield in northern Ontario). These soils differed in composition with respect to the physicochemical characteristics that could potentially influence the fate and effects of contaminants. All of the field-collected soils originated from uncontaminated areas that had not been subjected to any direct application

---

watering; 6) the lids don't change substantially or interfere with light fluence; and 7) phytotoxic observations can be made without removing the lids.

[13] The 1-L glass jars proved to be a successful alternative in a study where a highly volatile test substance reacted with the 1-L polypropylene vessels, thereby compromising their integrity (Stephenson *et al*., 2001a).

[14] Light intensity, and its control thereof, can be as important, if not more so, than the pH and temperature during plant toxicity tests. The light fluence rate throughout the entire test area should be checked before initiating the test. The distance between the plant *canopy* and the light source can be increased or reduced in order to achieve the appropriate lighting conditions. Alternatively, the portion of the test area that is within 15% of the selected light fluence rate can be "mapped out" to designate the boundaries of adequate versus inadequate light fluence rate (EC, 1999b).

[15] The Canadian Council of Ministers of the Environment (CCME) provides a comprehensive Web site on Canadian Environmental Quality Guidelines including those for soil (www.ccme.ca). This information is useful when reviewing analytical data (e.g., values for metals or PAHs) for samples of field-collected soil from a location under consideration as a source of natural soil suitable for use as negative control soil in toxicity tests. The summary table of CCME's Environmental Quality Guidelines can be accessed directly at www.ccme.ca/assets/pdf/e1_06.pdf. These Web sites and associated links will assist the investigator(s) reviewing the physicochemical characteristics of presumably *clean* natural soils under consideration for use as negative control soil in soil toxicity tests. The CCME can also be contacted by toll-free phone (1-800-805-3025) or e-mail (info@ccme.ca).

of pesticides in recent previous years and were therefore considered to be "*clean*". The origin and physicochemical characteristics of these natural soils are further described in Appendix G. The test validity criteria for the various plant species described in Section 4.4 are based on the performance data for these plants in negative control soil, that were generated for each of these five diverse soils (Aquaterra Environmental, 1998a; Stephenson *et al.*, 1999a, b, 2000a; Aquaterra Environmental and ESG, 2000; ESG, 2000, 2001, 2002; ESG and Aquaterra Environmental, 2002; EC, 2005b).

3.4.1 Natural Soil

Negative control soil may be natural soil collected from a *clean* (uncontaminated) site which is known to have been free of pesticide or fertilizer applications for at least five years. Before using a sample of *clean* field-collected soil as *negative control soil* in a definitive toxicity test, the test laboratory must have previous experimental evidence showing that natural soil from this source can meet the criteria that must be achieved for the results of a toxicity test to be considered valid (see Section 4.4).

Accordingly, initial tests involving a sample of this soil must be performed using the intended plant test species, to confirm that the test organisms are able to meet the criteria for test validity (see Section 3.2.1). Thereafter, and assuming that the preceding results for these preliminary tests are satisfactory, it is recommended that samples of natural soil selected for possible use as *negative control soil* in soil toxicity tests (as well as samples of candidate *reference soil*) be analyzed for the following physicochemical characteristics:

- pH,
- particle size distribution,
- *conductivity*,
- *texture*,
- *fertility*,
- *total organic carbon* content (%),
- *organic matter* content (%),
- cation exchange capacity,
- major cations,
- total nitrogen,
- total phosphorus,
- bulk density,
- WHC,
- metals,
- petroleum hydrocarbons (including PAHs),
- organophosphorus insecticides,
- organochlorine insecticides, and
- a suite of herbicides (e.g., atrazine and other triazine herbicides, picloram, acrolein, and sulfonylureas).

Pesticide and metal concentrations should not exceed the Canadian Soil Quality Guidelines, if available (see footnote 15). If seeds from a natural seedbank germinate in the sample(s) of natural soil at any time (i.e., during storage or testing), then these seedlings must be removed. If the results of both the preliminary biological tests and the physicochemical analyses are satisfactory, a larger sample of this natural soil can be collected, air dried to a *moisture content* of between 10 and 20%, coarse-screened (4–6 mm), transferred to clean, thoroughly rinsed plastic pails, and stored in darkness at 4 ± 2 °C until required. Plastic pails should not be used for the collection and storage of soils if there are concerns about chemical constituents of the plastic leaching into the soil.

3.4.2 Artificial Soil

Negative control soil may be *artificial soil* formulated in the laboratory. The use of artificial soil offers a consistent, standardized approach and is advantageous when testing the toxicity of chemicals or chemical products spiked in negative control soil (Section 6).

In keeping with the formulation of artificial soil recommended by OECD (1984b, 2000b), USEPA (1989), ISO (1991, 1993b, 1998) and ASTM (1999c) for earthworm testing, and that recommended in Environment Canada's soil toxicity tests using earthworms and springtails (EC, 2004c, 2005c), the following ingredients should be used to prepare artificial soil to be used in the biological test method described herein:

- 10% *Sphagnum* sp. peat, air dried and sieved through a 2-mm mesh screen (ASTM, 1999c)

- 20% kaolin clay with particles <40 µm (ASTM, 1999c)

- 70% "grade 70" silica sand (USEPA, 1989; ASTM, 1999c)

The ingredients should be mixed thoroughly in their dry form using a mechanical stirrer and/or gloved hands.[16] Reagent-grade calcium carbonate should be added to the dry mixture in a quantity sufficient to attain a pH for the *artificial soil* ranging within 6.5–7.5 once it is hydrated.[17] Thereafter, the mixture should be hydrated gradually using *test water* (i.e., de-ionized or distilled water) until its *moisture content* is ~20% (which is ~28% of the soil's *water-holding capacity*), while mixing further until the soil is visibly uniform in colour and texture. As necessary, reagent-grade calcium carbonate should be added to the hydrated mixture in a quantity sufficient to maintain a pH ranging within 6.5–7.5. Samples of pH-adjusted *artificial soil* should be stored in darkness at 20 ± 2 °C for a minimum of three days before being used in a toxicity test, to enable adequate time for pH equilibration.[17] Thereafter, *artificial soil* can be stored at 4 ± 2 °C. As and when required for a soil toxicity test, a suitable quantity of stored artificial soil should be hydrated further using test water until its moisture content is ~70% of the water-holding capacity.

3.5 Positive Control Soil

The use of one or more samples of *positive control soil* is recommended for inclusion in each series of soil toxicity tests with terrestrial plants, to assist in interpreting the test results. In choosing a positive control soil, the intent is to select a *toxic* soil that will elicit a response in the test organisms which is predictable based on earlier toxicity tests with this material. The positive control soil might be a sample of negative control soil that is spiked with a *reference toxicant* for which historic data are available on its toxicity to plants using specified test conditions and procedures. For the test method described herein, one or more reference toxicants must be used as a positive control soil when appraising the sensitivity of the test organisms and the *precision* and reliability of results obtained by the laboratory for that material (see Section 4.9). A test might also include a sample of negative control soil (natural or artificial; see Section 3.4) that has been spiked experimentally (Section 6) with one or more *toxic* chemicals or chemical products of particular concern when evaluating the sample(s) of *test soil*, at a concentration toxic to the plant species used according to the biological test method

---

[16] It is recommended that the dry ingredients initially be mixed (to incorporate the calcium carbonate) using a mechanical stirrer. Mixing should be completed using a gloved hand, to ensure that all of the soil from the corners of the container have been well mixed. Personnel must take the appropriate precautions for protection to prevent the inhalation of and contact with these ingredients.

[17] The amount of calcium carbonate ($CaCO_3$) required to adjust the pH of artificial soil to within this range depends on the nature (i.e., acidity) of the ingredients (and, in particular, that of the *Sphagnum* sp. peat). A quantity of 10–30 g of $CaCO_3$ for each kg of peat might prove adequate. A pH as low as 4.5 can occur when the soil is first formulated without the addition of $CaCO_3$. The initial pH adjustment should attempt to raise pH to range within 7.0–7.5, since the pH of artificial soil typically drops slightly (to 6.5–7.0) during the three-day equilibration period, before it stabilizes. The pH of stored samples of artificial soil should be checked regularly (e.g., once every two weeks) to ensure that it has not changed dramatically; adjustments should be made as necessary by adding additional quantities of $CaCO_3$ (Aquaterra Environmental, 1998a; G.L. Stephenson, personal communication, Aquaterra Environmental, Orton, ON, 2001).

A mixture of formulated artificial soil can also be stored dry, followed by partial hydration to ~20% moisture content, storage at 20 ± 2 °C for a minimum 3-day period, and subsequent hydration to ~70% WHC when required for use in a toxicity test. If storing formulated artificial soil dry, it is necessary to partially hydrate (to ~20% moisture) and equilibrate thereafter (for ≥3 days) to provide conditions for pH equilibrium similar to that recommended herein using artificial soil stored partially hydrated. Using this optional approach, the interim storage as partially hydrated artificial soil is necessary to enable the addition of more water (and, in certain instances, the addition of a chemical solution) as required when finalizing the pH and moisture content (i.e., adjusted to ~70% WHC) of artificial test soil. Storage of artificial soil that is partially hydrated, rather than dry, is considered a preferred approach since it enables laboratory personnel to more quickly hydrate to the desired moisture content (i.e., ~70% WHC) while ensuring pH equilibrium, and reduces any further delay in time associated with the dry storage of artificial soil.

described herein. In some instances, a test might include a positive control soil that is comprised of a highly contaminated sample of field-collected soil or sludge shown previously to be consistently toxic to terrestrial plants according to the biological test method described herein. [18]

3.6 Reference Soil

One or more samples of *reference soil* might be included in a soil toxicity test using terrestrial plants.[19] The type and nature of the sample(s) of soil used as reference soil in a particular study depend on the experimental design and the study's objectives. If the toxicity of samples of field-collected soil from a contaminated or potentially contaminated *site* is under investigation, the reference soil included in the study might be one or more samples of field-collected soil taken from a *clean* (uncontaminated) *site* where the physicochemical properties (e.g., organic carbon content, *organic matter* content, particle size distribution, *texture*, pH) represent the sample(s) of test (*contaminated*) soil as much as possible. Ideally, the reference soil is collected near the *site*(s) where samples of *test soil* are collected, but it is removed from the source(s) of contamination. One or more samples of field-collected *clean* reference soil from sites removed from the test site(s) might also be chosen due to their known lack of toxicity in previous tests with plants, and their possession of physicochemical characteristics similar to the samples of test soil. The sample(s) of field-collected reference soil used in a study could be tested for *toxic* effects at full strength only, or this soil could be mixed with the sample(s) of test soil to prepare a range of concentrations to be included in a multi-concentration test[20] (see Sections 3.7 and 5.5, as well as the introductory comments in Section 4). Samples of reference soil should not be collected from sites known to have received applications of pesticides or fertilizers within the past five years or more.

An investigator might choose to include one or more samples of *artificial soil* as reference soil in a particular test. For instance, these samples could be used in multi-concentration tests with *site soils* or *chemical-spiked soils* to investigate the influence of certain physicochemical characteristics (e.g., a number of artificial reference soils prepared to provide a range of differing values for *texture* and/or percent organic matter content; Sheppard and Evenden, 1998; Stephenson *et al.*, 2002) on the toxicity of a contaminated site soil or a chemical-spiked soil. Multiple samples of clean field-collected soil collected from various sites, which differ markedly with respect to one or more physicochemical characteristics, might also be used for this purpose. For such a study, a portion of each reference soil used to prepare a series of concentrations of the the test soil should be included in the test without dilution (i.e., 100% reference soil).

Each test involving one or more samples of reference soil must include a sample of negative control soil (see Section 3.4). Conversely, certain tests (e.g., one involving a series of concentrations of chemical-spiked soil prepared using artificial or natural negative control soil) need not involve a sample of reference soil. For tests with field-collected site soil, the inclusion of one or more samples of reference soil from a neighbouring site is a preferred approach for comparative purposes (see Section 5.5); a decision to dilute site soil with reference soil (rather than negative control soil)

---

[18] If the *positive control soil* is comprised of a highly contaminated sample of field-collected soil, it is important that its phytotoxic potential is stable over time (i.e., the sample is old enough that the bioavailability has stabilized).

[19] The use of field-collected *reference soil* might not be appropriate for certain toxicity tests such as those using samples of sludge (Section 5) or *chemical-spiked soil* (Section 6). Table 6 of Appendix E summarizes the applications of reference soil described in certain other test-method documents.

[20] Alternatively, the series of test concentrations used in a multi-concentration test could be prepared using negative control soil. The choice might be influenced by whether or not the candidate *reference soil*s are known to likely be non-toxic in the test to which they are to be applied, or a desire to prepare a range of concentrations of test soil using a *clean* soil with characteristics (e.g., *texture*, organic matter content) that closely match those of the test soil.

when preparing multiple concentrations for testing depends on the study objectives.

3.7 Test Soil

This biological test method is intended to measure the toxicity of one or more samples or mixtures of contaminated or potentially contaminated soil (*test soil*), using terrestrial plants as test organisms. The sample(s) of test soil might be either field-collected soil from an industrial or other site of concern, or industrial or municipal biosolids (e.g., dredged material, municipal sludge from a sewage treatment plant, composted material, or manure) under consideration for possible land disposal. A sample of field-collected test soil might be tested at a single concentration (typically, 100%) or evaluated for toxicity in a multi-concentration test whereby a series of concentrations are prepared by mixing measured quantities with either *negative control soil* or *reference soil* (see Section 5). The test soil might also be one or more concentrations of a *chemical-spiked soil*, prepared in the laboratory by mixing one or more chemicals or chemical products with *negative control soil*, *reference soil,* or *site soil* (see Section 6).

Universal Test Procedures

General procedures and conditions described in this section for toxicity tests with terrestrial plants apply when testing the toxicity of samples of soil, particulate waste (e.g., sludge), or chemical, and also apply to their associated reference toxicity tests. More specific procedures for conducting tests with field-collected samples of soil or other similar particulate material (e.g., sludge, de-watered mine tailings, drilling mud residue, compost, biosolids) are provided in Section 5. Guidance and specific procedures for conducting tests with *negative control soil* or other soil spiked (amended) experimentally with chemical(s) or chemical product(s) are given in Section 6.

All aspects of the test system described in Section 3 must be incorporated into these universal test procedures. Those conditions and procedures described in Section 2 for seed storage, handling, and sorting in preparation for soil toxicity tests, also apply. A summary checklist in Table 2 describes recommended conditions and procedures to be universally applied to each test with samples of contaminated or potentially contaminated soil, as well as those for testing specific types of test materials or substances. These could include samples of *site soil*, biosolids (e.g., dredged material, sludge from a sewage treatment plant, composted material, or manure), or *negative control soil* (or other soil, *contaminated* or *clean*) spiked in the laboratory with one or more test chemicals or chemical products.

This biological test method uses terrestrial plant seed as test organisms, and measures seedling *emergence* and *growth* (shoot and root length and dry mass) inhibition as the biological endpoints. Test organisms are chosen from a list of 12 species approved for use in this test method (see Section 1.2). Test duration is 14 or 21 days [21], depending on the species chosen and the *biomass* needed for determination of the endpoint measurement(s) (see Section 4.3). The test soils are hydrated during the test, but not renewed. This definitive test method was applied and validated by six participating laboratories in a series of concurrent 14-day multi-concentration tests using red clover (*Trifolium pratense*) in artificial soil spiked with boric acid (EC, 2005a).[22]

[21] The incipient lethal level (or incipient EC50, the concentration at which an effect can be expected in 50% of the test organisms after an indefinitely long exposure) might not be reached for some test substances (e.g., some herbicides) in the time period outlined in this method. For further guidance on determining and applying the incipient lethal concentration, see EC (1999a; 2004a).

[22] In this series of tests, each of the participating laboratories was able to achieve valid test results (i.e., all test validity criteria specified herein for red clover were met) with the exception of one laboratory, which failed to meet the validity criterion for root length (i.e., the laboratory met only two of the three validity criteria specified for red clover). The data produced by the laboratory that failed to achieve all three test validity criteria were removed from the final data analysis, as were results from a second laboratory. Even though the second laboratory produced valid test results, the data were questionable due to low initial and final soil pH (i.e., initial and final pHs were all <6.5 and <6.0, respectively; whereas other laboratoriess had initial and final pHs >7.0). Seedling emergence was unaffected by boric acid, and therefore EC50s for emergence were not calculated. Data for shoot length generated during these tests yielded a mean IC50 of 981 mg boric acid/kg artificial soil (dry wt) with individual laboratories ranging from 837–1179 mg/kg. The IC50s for root length ranged from 643–864 mg/kg artificial soil (dry wt) with a mean IC50 of 725 mg/kg. Mean IC50s for shoot and root dry mass were 754 and 588 mg boric acid/kg artificial soil (dry wt), respectively; with data for shoot and root dry mass ranging between 705 and 803 mg/kg and 430 and 692 mg/kg, respectively for the various laboratories involved. The interlaboratory *coefficient of variation* for each of the endpoints produced in the series of tests, including IC50s for shoot and root lengths, and shoot and root dry mass were 15%, 14%, 7%, and 22%, respectively. These are considered to be acceptable levels of *precision* between laboratories (EC, 2005a).

Table 2 Checklist of Recommended Conditions and Procedures for Conducting Definitive Tests of Soil Toxicity Using Terrestrial Plants

Universal	
Test type	– whole soil toxicity test; no renewal (static test)
Test duration	– 14 days for barley, cucumber, durum wheat, lettuce, radish, red clover, or tomato; and – 21 days for alfalfa, blue grama grass, carrot, northern wheatgrass, or red fescue
Approved test species	– monocotyledons: barley (*Hordeum vulgare*), blue grama grass *(Bouteloua gracilis)*, durum wheat *(Triticum durum)*, northern wheatgrass *(Elymus lanceolatus*; formerly named *Agropyron dasystachyum)*, and red fescue *(Festuca rubra)*; dicotyledons: alfalfa *(Medicago sativa)*, carrot *(Daucus carota)*, cucumber *(Cucumis sativus)*, lettuce *(Lactuca sativa)*, radish *(Raphanus sativus)*, red clover *(Trifolium pratense),* and tomato *(Lycopersicon esculentum)*
Number of concentrations	– minimum of 9, plus negative control; recommend 11, plus negative control
Number of replicates	For single-concentration test (e.g., site soil tested at 100% concentration only): – ≥5 replicates/treatment For multi-concentration test: – ≥4 replicates/treatment for equal replicate test design; or – regression design; unequal replicates among test treatments: - 6 replicates for negative control soil - 4 replicates for lowest 4–6 test concentrations - 3 replicates for highest 5 test concentrations
Number of seeds per replicate	– 5 seeds/vessel for barley, cucumber, durum wheat, lettuce, northern wheatgrass, radish, red clover, red fescue, or tomato; and – 10 seeds/vessel for alfalfa, blue grama grass, or carrot
Negative control soil	– depends on study design and objectives; *clean* field-collected soil or artificial soil if testing site soils; recommend artificial soil for tests with chemical(s) or chemical product(s) spiked in soil
Test vessel	– polypropylene cups (1 L), covered for 7 days or until plants reach top of container
Amount of soil/ test vessel	– identical wet wt, equivalent to a volume of ~500 mL; ~350 g dry wt if artificial soil
Moisture content, test soils	– for soil preparation, hydrate to the optimal % of its water-holding capacity (WHC) if field-collected soil (see Section 5.3), or to ~70% of WHC if artificial soil; during test, hydrate to saturation
Air temperature	– daily range, constant 24 ± 3 °C; alternatively, day: 24 ± 3 °C, night: 15 ± 3 °C
Humidity	– ≥50%
Lighting	– full spectrum fluorescent: mimic natural light spectrum (e.g., Vita Lite® by Duro-Test®); 300 ± 100 μmol/(m^2 · s) adjacent to the level of the soil surface; 16 h light:8 h dark
Watering	– hydration water sprayed over soil surface until saturation, about every two days when covered and once per day after covers are removed, or whenever soil appears dry; weak nutrient solution might be necessary depending on fertility of soil and length of test

Measurements during test	– soil moisture content in each treatment/concentration at start; pH in each treatment/concentration at start and end; temperature in test facility, daily or continuously; humidity in test facility; light intensity once during test
Observations during test	– number of emerged seedlings at end of test in each test vessel; shoot/root length and shoot/root dry mass at test end; number of surviving plants at test end showing an atypical appearance (e.g., chlorosis, lesions); optionally, Day-7 seedling emergence (%) and shoot/root wet mass at test end
Biological endpoints	– emergence of seedlings during test; length of longest shoot and longest root at test end; dry weight of entire shoot and root structures (oven-dried at 90 °C until constant mass) at test end; appearance of surviving plants at test end; optionally, wet weight of shoot and root at test end
Statistical endpoints	– mean (± SD) percent emergence in each treatment/concentration at test end (Day 14 or Day 21); mean (± SD) length of longest shoots and roots in each treatment at test end (Day 14 or Day 21); mean (± SD) dry wt of shoots and roots in each treatment at test end (Day 14 or Day 21); if multi-concentration test: 14- or 21-day EC50 for inhibition of % emergence, data permitting; 14- or 21-day ICp for each of mean shoot length, root length, shoot dry wt, and root dry wt of individual plants surviving in each concentration at test end
Test validity	– invalid if **any** of the following occurs in negative control soil at test end: • mean % emergence is <60% for carrot, cucumber, or tomato; <70% for alfalfa, barley, blue grama grass, lettuce, northern wheatgrass, red clover, or red fescue; <80% for durum wheat; or <90% for radish • mean % survival of emerged seedlings in negative control soil at test end is <90% • mean percentage of control seedlings exhibiting phytotoxicity or developmental anomalies is >10% • mean root length is <40 mm for tomato; <70 mm for blue grama grass, red clover, or red fescue; <80 mm for carrot; <100 mm for lettuce; <110 mm for northern wheatgrass or radish; <120 mm for alfalfa or cucumber; or <170 mm for barley; or <200 mm for durum wheat • mean shoot length is <20 mm for lettuce; <30 mm for red clover; <40 mm for alfalfa; <45 mm for carrot; < 50 mm for blue grama grass, radish, or tomato; < 60 mm for cucumber; <80 mm for red fescue; <100 mm for northern wheatgrass; <150 mm for barley; or <160 mm for durum wheat
Test with reference toxicant	– must perform at least once every two months, or in conjunction with definitive test(s) with soil samples; use boric acid; prepare and test ≥5 concentrations plus a negative control, using artificial soil as a substrate; ≥3 replicates/concentration; 5 or 10 seeds per replicate (i.e., species-specific); follow procedures and conditions for a reference toxicity test described in Section 4.9; determine % emergence and 7-day or 10-day (species dependent) ICp for shoot length (including 95% confidence limits); express as mg boric acid/kg, dry wt

Field-Collected Soil

Transport and storage	– seal in plastic and minimize air space; transport in darkness (e.g., using an opaque cooler, plastic pail or other light-tight container); do not freeze or overheat during transportation; store in dark at 4 ± 2 °C; test should start within two weeks, and must start within six weeks unless soil contaminants are known to be stable
Negative control soil	– either natural, uncontaminated field-collected soil or artificial soil, for which previous plant tests have shown that all criteria for test validity could be regularly met
Reference soil	– one or more samples for tests with field-collected soil; ideally taken from site(s) presumed to be clean but near sites of test soil collection; characteristics including percent organic matter, particle size distribution, and pH similar to test soil(s)
Characterization of test soils	– at least percent moisture, WHC, pH, conductivity, percent total organic carbon (TOC), percent organic matter, and particle sizes (% sand, % silt, % clay); optionally, contaminants of concern [e.g., metals, polycyclic aromatic hydrocarbons (PAHs), pesticides]
Preparation of test soils	– if necessary, remove debris and indigenous macro-organisms using forceps; if necessary, press through a sieve of suitable mesh size (e.g., 4–6 mm); mix; determine soil moisture content; hydrate with de-ionized or distilled water (or, if and as necessary, dehydrate) to optimal percentage of its

	WHC (see Section 5.3); mix; dilute with control or reference soil if multi-concentration test; ensure homogeneity

Soil Spiked with Chemical(s) or Chemical Substance(s)

Negative control soil	– recommend artificial soil, or a *clean* field-collected soil
Characterization of chemical(s) or chemical substance(s)	– information on stability, water solubility, vapour pressure, purity, and biodegradability of chemical(s) or chemical substance(s) spiked into negative control soil should be known beforehand
Solvent	– de-ionized water is the preferred solvent; if an organic solvent is used, the test must include a solvent control
Preparation of mixtures	– procedure depends on the nature of the test substance(s) and the test design and objectives; chemical/soil mixtures may be prepared manually or by mechanical agitation; test substance(s) may be added as measured quantities in solution (i.e., in water or an organic solvent), directly as a liquid substance, or as a solid material comprised partly or completely of the test substance(s); ensure homogeneity
Concentration of chemical(s) or chemical substance(s) added	– normally measure at beginning and end of test, in high, medium, and low concentrations as a minimum

4.1 Preparing Test Soils

Each test vessel (see Section 3.2.2) placed within the test facility must be clearly coded or labelled to enable identification of the sample and (if diluted) its concentration. The date and time when the test is started must be recorded, either directly on the labels or on separate data sheets dedicated to the test. The test vessels should be positioned such that observations and measurements can be made easily. *Treatments* should be positioned randomly within the test facility (EC, 1997a, b, 2001, 2004c) and rotated regularly (e.g., while watering).

On the day of the start of the test, which is the day the seeds are initially exposed to samples of test material or substance (i.e., Day 0), each sample or subsample of *test soil* or similar particulate material, including *negative control soil* and, if used, *reference soil*, should be mixed thoroughly[23] (see Sections 5.3 and 6.2) to provide a homogeneous mixture consistent in colour, texture, and moisture.

If field-collected samples of *site soil* are being prepared for testing, large particles (stones, thatch, sticks, debris) should be removed before mixing, along with any vegetation or macroinvertebrates observed (see Section 5.3).

Test soils for terrestrial plant testing are prepared on the day of test initiation (i.e., Day 0). The quantity of each test soil mixed as a *batch* should be enough to set up the *replicates* of that *treatment* (see Table 2) plus an additional amount for the physicochemical analyses to be performed (Section 4.6) and a surplus to account for the unused soil that adheres to the sides of the mixing container. The *moisture content* (%) of each test soil should be known or determined, and adjustments made as necessary by mixing in *test water* (or, if and as necessary, by dehydrating the sample) until the desired moisture level is achieved (see Sections 5.3 and 6.2). Quantitative measures of the homogeneity of a *batch* might be made by taking aliquots of the mixture for measurements such as particle size analysis, *total organic carbon* (%), *organic matter* content (%), *moisture content* (%), and concentration of one or more specific chemicals.

[23] Any liquid that has separated from a sample or subsample of test soil during transport and/or storage must be remixed into the sample.

Immediately following the mixing of a *batch*, an identical wet weight of test soil equivalent to a volume of ~500 mL should be transferred to each *replicate* test vessel.[24] The soil added to each test vessel should be smoothed (but not compressed) using a spoon, by gently shaking the vessel back and forth horizontally, or by gently tapping the vessel ≥3 times on the benchtop or with a hand.

For a single-concentration test [e.g., *site soil* tested at 100% concentration only; a particular concentration of test soil; or a chemical tested at one concentration (e.g., Maximum Label Rate)], a minimum of five replicate test vessels and five replicate *negative control* vessels must be set up by adding an identical wet weight (equivalent to a volume of ~500 mL) of the same *batch* to each replicate vessel. For a multi-concentration test, either equal or unequal replication across *treatments* can be used. If replication is equal across treatments, at least four replicate test vessels must be set up for each treatment. If replication is unequal across treatments (see Section 4.8), six replicate vessels should be prepared for the negative control soil, four replicate vessels should be prepared for the lowest 4–6 test concentrations, and three replicate vessels should be prepared for the highest five test concentrations. [25] For any test that is intended to estimate the ICp in a definitive soil test (see Section 4.8), at least nine concentrations plus a negative control soil must be prepared, and more (≥11) are recommended to improve the likelihood of bracketing each endpoint sought.[26]

Concentrations should be chosen to span a wide range, including a low concentration that obtains effects like the negative control, and a high concentration that results in "complete" or severe effects. It is a common mistake to anticipate the endpoint and bracket it with a closely spaced series of concentrations, all of which might turn out to be either too low or too high. To keep the wide range of concentrations, and also obtain the important mid-range effects, it might be necessary to use additional *treatments* in order to split the selected range more finely. In any case, a consistent geometric series should be used. Additional guidance on selecting test concentrations that applies here is found in EC (2004a).

4.1.1 Range-Finding Test

In the case of appreciable uncertainty about sample toxicity, it is often beneficial to run a *range-finding* test for the sole purpose of establishing more closely the concentrations to be used for the *definitive* test, in which instance the number of replicates per concentration could be reduced (see Section 6.2). Conditions and procedures for the range-finding test are similar to the definitive test (see Table 2); however, the experimental design differs.

The *range-finding* test is a short-term test (7 days for barley, cucumber, durum wheat, lettuce, red clover, radish, and tomato; and 10 days for alfalfa, blue grama grass, carrot, northern wheatgrass, and red fescue), with ≥6 concentrations of test chemical or test soil[27], and only duplicate vessels (i.e., two replicates) per treatment. The test species must be

---

[24] The wet weight of soil required to achieve a volume of ~500 mL depends on the moisture content, bulk density, and other characteristics of the soil, and will vary from sample to sample. Accordingly, the wet weight of each sample required to achieve this volume should be determined by transferring the amount of sample required to fill a preweighed (or tared) 1-L test vessel to a 500-mL mark scribed on its side, followed by smoothing the surface and gently tapping the container on the bench top, three times. Thereafter, the wet weight of that quantity should be determined and recorded, and an identical wet weight added to each *replicate* test vessel.

[25] A greater number of replicates can be used and the distribution of replicates across treatments can be balanced (i.e., ≥4 replicates per treatment); however, the unbalanced nature of the regression design (i.e., unequal replicates among treatments) was developed to keep the level of effort comparable to that of an ANOVA design in terms of the total number of test vessels per test (Stephenson, 2003a).

[26] The large number of test treatments are needed to show the shape of the concentration-response relationship and to choose the appropriate linear or nonlinear regression model. Also, they contribute to the success of the computer calculations of the ICp and increase the probability of deriving a value (EC, 2004a).

[27] For *range-finding* tests, a suitable range of test concentrations for a test chemical spiked in soil might be 1000, 100, 10, 1, 0.1, and 0.01 mg/kg, or some common multiple thereof. For a multi-concentration test with a sample of contaminated (or potentially contaminated) field-collected soil, a suitable range of concentrations to use in a range-finding test might be 100, 50, 25, 12.5, 6.25, and 3% (Stephenson *et al*., 2001a).

the same as that to be used in the definitive test (see Section 2.1), and the number of seeds per replicate should be the same as those used in the definitive test (see Table 2 and Section 4.2). Negative control soil, air temperature and lighting conditions, percent moisture of soils, watering, and measurements during the test, are the same as those described for the definitive test (see Table 2). Shoot length and root length can be used to predict where the *sublethal* endpoints for growth will be in the definitive test.[28] In most cases, the endpoints for growth in the definitive test will be at lower concentrations than those observed for the *range-finding* test, due to the longer test duration in the definitive test. The number of emerged seedlings at the end of the range-finding test should also be observed and recorded to determine whether the test validity criteria for seedling emergence in the definitive test are likely to be met (see Section 4.4).

4.2 Beginning the Test

Following the addition of *test soil* to each test vessel, 5 or 10 sorted seeds, depending on the species (see Section 2.1) are planted in the soil within each test vessel, in order of increasing test concentration. For species requiring only five seeds (i.e., barley, cucumber, durum wheat, lettuce, northern wheatgrass, radish, red clover, red fescue, and tomato), four seeds are distributed equally around one seed within the centre of the soil in each test vessel. For alfalfa, blue grama grass, and carrot, which require 10 seeds per test vessel, nine seeds are distributed equidistant around one centre seed. Using fine forceps, each seed should be planted to a depth that is twice the diameter of the seed itself. The seeds are covered with the surrounding test substrate by tapping the test substrate with a stainless steel spatula or glass rod.[29] After the seeds have been added to each test vessel, the vessels are hydrated by spraying the soil surface with *hydration water* using a fine-mist spray bottle. Enough water is added to bring the moisture content of the soils close to saturation (see Section 4.5). Following hydration, lids (see Section 3.2.2) should be placed on the test vessels, to minimize loss of moisture.

4.3 Test Conditions

- This is a 14- or 21-day soil toxicity test, during which the soil in each test vessel is not renewed. The test duration for barley, cucumber, durum wheat, lettuce, radish, red clover, and tomato is 14 days, whereas that for alfalfa, blue grama grass, carrot, northern wheatgrass, and red fescue (i.e., species that produce less *phytomass* and/or take longer to germinate) is 21 days.

- The test vessel is a 1-L clear polypropylene container. Its contents (i.e., a 500-mL volume of *test soil*) are covered with a clear polypropylene lid (see Section 3.2.2).

- For a single-concentration test, at least five replicate test vessels must be set up for each *test soil* (i.e., each *treatment*). For a multi-concentration test, the use of an unequal number of replicate test vessels per test concentration and control, depending on concentration and treatment, is recommended. A minimum of six replicates for controls, four replicates in the lowest 4–6 test concentrations, and three replicates in the highest five test concentrations, should be prepared (see Section 4.1).[25]

- The test must be conducted at a constant mean air temperature of 24 ± 3 °C; or a daily mean air

[28] The effect on seedling growth can be visibly reflected in the above-ground *biomass* of the seedling; however, root endpoints are often more sensitive than shoot endpoints. Both shoot and root lengths, therefore, should be measured in a range-finding test in order to provide a strong likelihood that the concentration range chosen for the definitive test will bracket both shoot and root endpoints.

[29] To avoid variability in the planting depth of seed, which can lead to variability in percent emergence, the following procedure may be used. When preparing each replicate, add all but a portion (~10%) of the soil to the test vessel. The surface of the soil in the test vessel is gently flattened, and the seed is placed on the soil surface in the species-specific pattern described in this section. A plastic template, with pre-marked holes that are just slightly larger than the diameter of the seed, is useful to achieve a uniform distribution of very small seeds within the test vessel. After the seed has been equally distributed on the surface of the soil according to the pattern described herein, the remainder of the soil (i.e., the ~10% that was not originally added to the replicate vessel) is used to cover the seed to a uniform depth. This is repeated for each of the test replicates (G. Lazarovits, personal communication, Agriculture and Agri-Food Canada, London, ON, 2004).

temperature of 24 ± 3 °C and a nightly mean air temperature of 15 ± 3 °C for those facilities that can accommodate daily changes in test temperatures (see Section 3.1).

- Test vessels must be illuminated with a 16-h light and 8-h dark daily *photoperiod*. Full-spectrum fluorescent lights or equivalent which mimic a natural light spectrum (e.g., Vita Lite® by Duro-Test®) should be used. Light intensity adjacent to the surface of the soil in each test vessel must be 300 ± 100 µmol/($m^2 \cdot s$) (i.e., equivalent to 18 750 ± 6250 lux) (see Section 3.3).

4.4 Criteria for a Valid Test

For a valid test, each of the following five test criteria must be achieved[30]:

1. The mean percent *emergence* for individual plant species grown in negative control soil for the duration of the test must be:

   - ≥60% for carrot, cucumber, or tomato;
   - ≥70% for alfalfa, barley, blue grama grass, lettuce, northern wheatgrass, red clover, or red fescue;
   - ≥80% for durum wheat; or
   - ≥90% for radish.

2. The mean percent survival for emerged seedlings grown in negative control soil for the duration of the test must be ≥90%.[31]

3. The mean percentage of seedlings grown in negative control soil for the duration of the test, that exhibit *phytotoxicity* and/or developmental anomalies, must be ≤10%.[32]

4. The mean root length for individual plant species grown in negative control soil for the duration of the test must be:

   - ≥40 mm for tomato;
   - ≥70 mm for blue grama grass, red clover, or red fescue;
   - ≥80 mm for carrot;
   - ≥100 mm for lettuce;
   - ≥110 mm for northern wheatgrass or radish;
   - ≥120 mm for alfalfa or cucumber;
   - ≥170 mm for barley; or
   - ≥200 mm for durum wheat.

5. the mean shoot length for individual plant species grown in negative control soil for the duration of the test must be:

   - ≥20 mm for lettuce;
   - ≥30 mm for red clover;
   - ≥40 mm for alfalfa;
   - ≥45 mm for carrot;
   - ≥50 mm for blue grama grass, radish, or tomato;
   - ≥60 mm for cucumber;
   - ≥80 mm for red fescue;
   - ≥100 mm for northern wheatgrass;
   - ≥150 mm for barley; or
   - ≥160 mm for durum wheat.

[30] The test validity criteria presented here are based on control data generated in many studies carried out during the development of the method. These studies involved various toxicants such as petroleum hydrocarbons, pesticides, metals, and boric acid as well as the control data of five different types of negative control soils (see Appendix G) (Aquaterra Environmental, 1998a; Stephenson *et al.*, 1999a, 1999b, 2000a; EC, 2000; Aquaterra Environmental and ESG, 2000; ESG and Aquaterra Environmental, 2002; and EC, 2005b). The *coefficient of variation* (CV) based on all of the data used to develop the test validity criteria for percent emergence ranged from 3.2% for radish to 16.3% for barley. For shoot length in control soils, the CV ranged from 9.0% for lettuce to 32.2% for tomato, and for root length, the CV ranged from 10.7% for durum wheat to 27.3% for blue grama grass. These levels of variability are considered to be acceptable in terms of intra- and interlaboratory *precision*.

[31] The mean percent survival is calculated from the percentage of emerged plants, in each test vessel containing negative control soil, that survive to the end of the test. For instance, if only four of the five emerged seedlings in a given vessel survive to the end of the test, the percent survival for that vessel would be 80%. However, the mean percent survival is the average percent survival for emerged plants in all of the test vessels containing negative control soil.

[32] Phytotoxicity and/ or developmental anomalies could include phytotoxic symptoms such as *chlorosis*, *defoliation*, *desiccation*, *malformation, mottling*, *staining*, *necrosis*, *withering*, or *wilting*; and/or overt evidence of atypical formative effects such as leaf and/or stem deformation or apparent signs of impaired rates of growth and development (USEPA, 1996).

4.5 Hydration of Test Soils During the Test

Test soils are hydrated to "near-saturation" as needed, throughout the test. Hydrating to near-saturation means, in this instance, that water is added to the surface of the soil until ~0.5 cm of water is temporarily (≤ 1 h) visible pooling at the bottom of the test vessel following its addition. *Hydration water*, at 24 ± 3 °C, should be sprayed onto the surface of the soil using a fine-mist spray bottle on Day 0, just after the seed has been added to the test vessels, and again every 48 hours, or as needed, until the lids of the test vessels are removed (see Section 4.6). Thereafter, and if and as required, water should be added at least once every 24 hours to achieve near-saturation daily throughout the test.[33] A weak nutrient solution [e.g., a half strength of Hoagland's nutrient solution (Hoagland and Aaron, 1950)] should be used to water each test vessel if it is determined that the test soil is too deficient in nutrients to sustain healthy plant growth in the negative control soil for the duration of the experiment.[34]

The location of the test vessels in the environmental chamber or the testing area should be randomly varied each time that water is added to test vessels, so that the test organisms within these vessels are randomly exposed to any slight variations in test conditions (i.e., lighting, temperature, humidity, or ventilation) that might exist in the testing area.

4.6 Observations and Measurements During the Test

The biological endpoints for the test are seedling *emergence*, root and shoot length, and root and shoot dry mass at the end of the test (i.e., on Day 14 or Day 21, depending on the test species). Determining the number of seedlings emerged in each test vessel on Day 7 is also useful and frequently done (i.e., to determine the 7-day *emergence* rate for a single-concentration test, or the 7-day EC50 for a multi-concentration test), although such observations are

---

[33] The rate of watering depends on the rate of water loss from the soil, and might vary between test vessels. The rate of water loss is influenced by soil type, organic matter content, root mass, the development of shoot *canopy* during the test, and the humidity of the test facility. The watering regime changes over the duration of the test. Initially, since the test vessels have lids, watering might not be necessary. If watering is necessary, however, it can be accomplished at this point by gently spraying the soil surface using a spray bottle, and subsequently observing (over one hour) to see if a condition of "near-saturation" has been achieved (as per guidance herein). After the lids are removed (Section 4.6), but before a substantial root mass has formed, water loss might be relatively fast and the soils might dehydrate to a substantial depth. At this time, water might still be added to the soil surface by spraying. The judicious use of a spray bottle ensures that the surface soils do not get unduly disturbed (i.e., the integrity of the surface-soil structure is maintained) and that the water is distributed uniformly over the surface of the soil column within each test vessel. As the root mass becomes appreciable, there is less water lost directly from the soil via evaporation but the roots can take up more water. At this point, water may be added by gently pouring it over the soil surface while being careful not to exceed the saturation of the soil column.

[34] Potential nutrient deficiency should be considered when preparing for and conducting this soil toxicity test. However, whether or not any nutrient solution is used during a toxicity test depends on the study objectives (e.g., site-specific objectives, potential use of land from the site of soil collection, etc). At the Environment Canada soil toxicity workshop, participants proposed two strategies related to nutrient deficiency in soil. The first was to use a more appropriate species (i.e., one that might not require the use of a nutrient solution), and the second was to run side-by-side tests with and without the addition of nutrient solution to two sets of replicate test vessels (EC, 2004b). Stephenson *et al.* (2001b) used three different types of water and a nutrient solution to determine their influence on barley emergence and growth in an artificial control soil and in artificial soil spiked with copper sulphate. These investigators found that there was no effect of water type (de-ionized water, municipal tap water, or well water) on the emergence and growth of barley; however, shoot and root growth were enhanced in artificial control soil watered with nutrient solution. The toxicity of copper sulphate to barley was also affected by the use of nutrient solution. Copper sulphate was more toxic to barley when seedlings were watered with a nutrient solution (Stephenson *et al.*, 2001b). Wilke *et al.* (2003) showed that the effects of pollutants (i.e., zinc or a pesticide containing Metamitron™) are greatly influenced by the nutritional status of test soils. Without the addition of a nutrient solution to two different types of soil (i.e., a sand and a silty loam sand), the growth of turnip was not affected by either toxicant compared to the controls, however, when nutrient solution was added, significant growth reductions between treatments and controls were observed (Wilke *et al.*, 2003). Therefore, the influence of a nutrient solution on the growth of seedlings in various types of soils, as well as its potential influence on soil contaminants should be considered when generating toxicity data.

optional. Depending on the study objectives, root and shoot wet mass might also be determined at the end of the test; however, these endpoints are also optional. Throughout the test, observations should be made and recorded of the number of emerged and the state or condition of the emerged plants, each time the test soils are hydrated (see Section 4.5).

Seedling *emergence* is measured visually by counting the number of seedlings that have emerged 3 mm above the soil surface in each test vessel. The lids must be removed from all of the test vessels for the remainder of the test on Day 7 or just before the seedlings reach a height that would contact the lid of the test vessel, whichever occurs first.

A *visual assessment* of the health and condition of the plants (e.g., *phytotoxicity*) in each test vessel should also be made and recorded when the plants first appear, as well as each time a test vessel is watered thereafter.[35] Observations might include:

- *chlorosis* (loss of pigment),
- *necrosis* (localized dead tissue),
- *defoliation* (loss of leaves),
- *dessication* (dried leaves or stems),
- *malformation* (structural defects),
- *mottling* (marked or spotted),
- *staining* (discolouration),
- *wilting* (limp),
- *withering* (in the process of drying),
- discoloured or deformed leaves or stem,
- overt signs of delayed emergence, or
- impaired development and/or growth.

The number of seedlings in the control test vessels that are alive at the end of the test should be counted, to determine whether the test validity criterion for percent survival of emerged plants in negative control soil has been met (see Section 4.4).

Air temperature in the test facility (Section 3.1) must be measured daily (e.g., using a maximum/minimum thermometer) or continuously (e.g., using a continuous chart recorder). The humidity should be measured periodically (Section 3.1).

The light fluence rate must be measured at least once during the test period at points approximately the same distance from the light source as the soil surface and at several locations in the test area (see Section 3.3).

In at least one replicate of each *treatment* (including the *negative control soil* and, if used, *reference soil*), the pH must be measured and recorded at the beginning and end of the test, and the *moisture content* must be measured and recorded at the beginning of the test only.[36] The initial (Day 0) measurements should be made using subsamples of each *batch* of *test soil* used to set up replicates of a particular *treatment* (see Section 4.1).[37] The final (i.e., Day 14 or Day 21) measurements should be made using subsamples of the replicates of each treatment to which plants were exposed, following the end-of-test observations of plant emergence, condition, and growth.

Soil pH should be measured using a calcium chloride ($CaCl_2$) slurry method (modified from Hendershot *et al*., 1993).[38] For these analyses, 4 g of hydrated

---

[35] Each time a test vessel is watered, the general condition of the plants therein should be assessed and observations recorded. These obeservations can be useful for interpretation of the results. For example, it is important to note when symptoms of stress (e.g., wilting, discoloration) first occur, and whether they get progressively worse or not. It is also important to record any signs of phytotoxicity for the plants in the negative control soil treatment, again for interpretation of the test results at the end of the test. Phytotoxic signs can be indicative of effects from physical factors such as lights too close to the foliage *canopy*, soils too dry between watering, nutrients in soil are either limited, or in excess (G.L.Stephenson, personal communication, Aquaterra Environmental, Orton, ON, 2002).

[36] The moisture content of test soils can be quite variable at the end of the test, depending on how much time has lapsed between hydrating the test soils and processing the vessels at the end of the test. This, in addition to the fact that test vessels are watered to near-saturation throughout the test, means that the measurement of moisture content at the end of the test is not required.

[37] Additional soil for each batch should be prepared for physicochemical analyses of the test-initiation (Day-0) conditions. One or more additional replicates of each test soil should be prepared and placed into a test vessel within the test facility. These replicates (with seed added) should be reserved for physicochemical analyses of test-end (i.e., Day-14 or Day-21) conditions.

[38] The method by Hendershot *et al.* (1993) includes a step that involves air drying the sample for 48 h before analyzing the pH. The experience of Environment Canada investigators is that this step is needlessly time-

soil[39] is placed into a 30-mL glass beaker (~3 cm in diameter and ~7 cm high) with 20 mL of 0.01 *M* $CaCl_2$.[40] The suspension should be stirred intermittently for 30 min (e.g., once every 6 min). The slurry should then be left undisturbed for ~1 h. Thereafter, a pH probe is immersed into the supernatant and the pH recorded once the meter reading is constant.

---

consuming (K. Doe, personal communication, Atlantic Environmental Science Centre, Environment Canada, Moncton, NB, 2004; J. Princz, personal communication, Biological Methods Division, Environment Canada, Ottawa, ON, 2004), and does not appreciably modify the pH relative to that for hydrated (i.e., as per the toxicity test) soil (Courchesne *et al.*, 1995; J. Princz, personal communication, Biological Methods Division, Environment Canada, Ottawa, ON, 2004).

Becker-van Slooten *et al*. (2004) assessed three different soil slurry methods for measuring pH. The need for this testing was identified during Environment Canada's soil toxicity workshop in Vancouver, BC (February, 2003) where certain participants recommended that a commonly used and "universally standardized" method for measuring soil pH be incorporated into each of Environment Canada's soil toxicity test methods (EC, 2004b). The following three methods for measuring soil pH were compared: 1) 1 *M* KCl in water; 2) 0.01 *M* $CaCl_2$ in water; and 3) water only. Results of this investigation showed that there were advantages and disadvantages with each of these methods for measuring pH. However, based on practical considerations and the recommendations of the workshop participants (i.e., that a widely used method for characterizing soil pH be applied), the 0.01 *M* $CaCl_2$ method was recommended as the most appropriate for Environment Canada's soil toxicity test methods (Becker-van Slooten *et al*., 2004).

[39] It might be necessary to use a lower soil:$CaCl_2$ solution ratio (e.g., 2 g of soil to 20 mL of $CaCl_2$) for soils with a high organic matter content (i.e., for soils where the slurry does not yield a supernatant).

[40] To prepare 0.01 *M* $CaCl_2$, dissolve 2.940 g of calcium chloride dihydrate ($CaCl_2 \cdot 2H_2O$) with distilled water, in a 2000-mL volumetric flask. The conductivity of the $CaCl_2$ solution should be between 224 and 240 mS/m at 25 °C, and the pH should range within 5.5–6.5 at 25 °C (Hendershot *et al.*, 1993). If the pH is outside this range, it should be adjusted to the range using a hydrogen chloride (HCl) or calcium hydroxide [$Ca(OH)_2$] solution. If the conductivity is not within the acceptable range, a new solution must be prepared.

The *moisture content* of each *test soil* is measured by placing a 3–5 g subsample of each test soil into a pre-weighed aluminum weighing pan, and measuring and recording the wet weight of the subsample. Each subsample should then be placed into a drying oven at 105 °C until a constant weight is achieved; this usually requires a minimum of 24 hours. The dry weight of each subsample should then be measured and recorded. Soil moisture content must be calculated (on a dry-weight basis) by expressing the moisture content as a percentage of the soil dry weight:

$$\text{Moisture content (\%)} = \frac{\text{wet weight (g)} - \text{dry weight (g)}}{\text{dry weight (g)}} \times 100$$

It is important that the moisture content (%) calculation be based on dry weight (not wet weight), since the results of these calculations are used with calculations of water-holding capacity (also calculated based on dry weight) to express the optimal moisture content in test soils (see Section 5.3).

Depending on the nature of the test and the study design, concentrations of chemical(s) or chemical product(s) of concern might be measured for test soils or selected concentrations thereof, at the beginning and end of the test. For a test using a sample of field-collected *site soil*, the chemical(s) or chemical product(s) measured will depend on the contaminant(s) of concern (see Section 5.4). For a multi-concentration test with *chemical-spiked soil*, such measurements should be made for the high, medium, and low strengths tested, as a minimum (see Section 6.3). Aliquots for these analyses should be taken as described previously for pH and moisture content; analyses should be according to proven and recognized (e.g., SAH, 1992; Carter, 1993) analytical techniques.

4.7 Ending the Test

The test is terminated after 14 days of exposure for barley, cucumber, durum wheat, lettuce, radish, red clover, and tomato; and after 21 days of exposure for alfalfa, blue grama grass, carrot, northern wheatgrass, and red fescue. At that time, the number of live and apparently dead plants in each test vessel should be determined and recorded, and any abnormal patterns in morphology, growth, and development (i.e., relative to the plants in the negative control soil) also recorded. Photographs might be taken to visually record the concentration-

response relationship in the above-ground *phytomass*. Even if no shoots are visible above the soil surface, the soil should be checked for root material in case roots developed from the seed but no shoot material was produced. These observations are for qualitative purposes only (i.e., for this test method, a seedling must emerge 3 mm above the soil surface to be considered "emerged") and, if roots develop, where no shoot material was produced, it should be noted. Thereafter, each test vessel must be processed separately to keep the seedlings within each replicate isolated from those in each of the other replicate test vessels.

The plants must be carefully separated from the test soil and from the roots of the other plants. This can be achieved by gently loosening the soil and root matrix from the test vessel and removing all soil that can be easily removed without disturbing the root matrix. In some cases, roots can be more easily separated from the soil after the soil is first saturated with water and allowed to soak for several minutes. The remaining soil and plant mass are placed into a pan of water. The roots can then be held under a gentle stream of tap water, or they can be sprayed with water from a spray bottle, to gently dislodge as many of the remaining soil particles as possible. This also aids in separating the roots of the plants from each other. The plants are then placed onto a moistened, labelled sheet of paper towel, one for each test vessel, and covered with plastic to minimize water loss until measurements can be made and recorded. Measurements of shoot and root lengths are made from the transition point between the *hypocotyl* and the root to the longest leaf tip when the leaves are gently straightened, and to the tip of the longest root when the roots are gently straightened. Shoot and root length for each plant in each replicate are measured with a ruler, and recorded in millimetres.

The shoots and roots are then separated from each other at the point at which there is a discernible transition between root and shoot tissue, and from the seed itself, using a scalpel. The remaining seed is discarded. The shoot and root structures from each replicate test vessel are weighed separately, as two groups (i.e., shoots and roots). The entire rinsed shoot *biomass* from each test vessel must be transferred as a group to a damp paper towel or blotting paper. Thereafter, they should be placed into a clean aluminum weighing pan (1–2.5 g) that has been previously numbered, weighed, and held in a desiccator.[41] This process is repeated with the entire rinsed root *biomass* from each test vessel. If wet mass is being determined, the aluminium pans containing shoots and roots are weighed immediately with an analytical balance that measures consistently to 0.1 mg. The dry mass must be determined and is done so in a similar way once the plants are dried in an oven at 90 °C until a constant weight is achieved (this usually takes a minimum of 24 h) (Aquaterra Environmental and ESG, 2000). Upon removal from the oven, the weighing pans are moved immediately to a desiccator. Once cooled, each weighing pan should be individually and randomly removed from the desiccator and weighed immediately [42] to the nearest 0.1 mg on a balance that measures accurately to this limit. Mean dry weight per surviving plant is calculated for each replicate (see Section 4.8.3).

Although it is the intention of Environment Canada to use mean shoot dry weight and mean root dry weight as additional test validity criteria for definitive tests, there is insufficient data at this time on which to base minimum weight requirements for control plants. It is recommended, however, that for definitive tests:

- The mean shoot dry weight per surviving plant, for individual plant species grown in negative control soil for the duration of the test be:
  - ≥1.0 mg for red fescue;
  - ≥1.5 mg for blue grama grass;
  - ≥2.0 mg for carrot;
  - ≥2.5 mg for lettuce;
  - ≥4.0 mg for red clover;
  - ≥5.0 mg for tomato;
  - ≥7.0 mg for northern wheatgrass;
  - ≥8.0 mg for alfalfa;
  - ≥20 mg for radish;
  - ≥25 mg for durum wheat;
  - ≥35 mg for barley; or
  - ≥40 mg for cucumber, and

[41] If any deposits (e.g., wax) associated with the weighing pans are cause for concern with respect to providing weighing errors, the weighing pans should be oven-dried for at least 48 h to achieve a constant weight (EC, 1997a, b; 2001; 2004c).

[42] The dried plants can take up water vapour readily, so weighing should be rapid and the time standardized among weighing pans.

- The mean root dry weight per surviving plant, for individual plant species grown in negative control soil for the duration of the test be:
  - ≥0.2 mg for tomato;
  - ≥0.5 mg for blue grama grass, carrot, or red fescue;
  - ≥1.0 mg for lettuce or red clover;
  - ≥3.0 mg for northern wheatgrass or radish;
  - ≥4.0 mg for alfalfa;
  - ≥7.0 mg for cucumber; or
  - ≥25.0 mg for barley or durum wheat.

During the series of dry-weight determinations for the groups of plants from a test, the first weighing pan should be returned to the desiccator, and weighed again at the end of all weighings. This serves as a check on any sequential gain of water by the weighing pans in the desiccator over time, which can occur when each weighing pan is removed for its weight determination. The change in weight of the first weight pan over time should not be >5%; if it is, all weighing pans should be re-dried for ≥2 h and then re-weighed.

Following the removal of plants from each test vessel, subsamples of each *test soil* (including the *negative control soil* and, if included in the test, *reference soil*) should be taken for pH determination (Section 4.6). Analyses for other chemical constituents (i.e., concentrations of contaminants) should also be made at this time using representative subsamples of each test soil (Section 4.6).[43]

4.8 Test Endpoints and Calculations

The percent *emergence* in each test vessel at the end of the test (Day 14 for barley, cucumber, durum wheat, lettuce, radish, red clover, and tomato; and Day 21 for alfalfa, blue grama grass, carrot, northern wheatgrass, and red fescue) must be calculated and reported for each test. The mean (± SD) percent *emergence* for all replicate groups of plants exposed to each *treatment* for 14 or 21 days must also be calculated and reported. Any optional observations of *emergence* taken on Day 7 (see Section 4.6) should also be calculated and reported as percent *emergence* in each test vessel, as well as mean (± SD) percent *emergence* for each treatment.

For a single-concentration test (see Section 4.1), the mean (± SD) value for the percent *emergence* of plants at test end, as determined for each treatment, is compared with that for the sample(s) of *reference soil* or, as necessary and appropriate, compared with that for the *negative control soil* (see Section 5.5). For a multi-concentration test (see Sections 4.1, 5.3, and 6.2), the 14-day or 21-day EC50 for emergence must be calculated and reported (data permitting).[44] If 7-day observations of percent emergence in each concentration were made during a multi-concentration test, it is recommended that the 7-day EC50 for emergence also be calculated and reported (data permitting). Environment Canada's guidance document on statistical methods for estimating endpoints of toxicity tests (EC, 2004a) provides definitive direction and advice for calculating EC50s, which should be followed (see Section 4.8.2, herein).

The *growth* endpoints for this test are based on shoot and root length, as well as shoot and root dry weight, of surviving plants in each *replicate* and each *treatment* as measured at the end of the 14- or 21-day test period. Shoot and root wet weight are additional (optional, but recommended) endpoints. A significant reduction in the length or weight of the plants is considered indicative of an adverse *toxic* effect of the treatment on the growth of test plants. For a single-concentration test (see Sections 5.3 and 6.2), the mean (± SD) values for shoot and root length, and shoot and root dry weight, of plants surviving in the *test soil* at test-end is determined and compared to those values for the sample(s) of *reference soil* or, as necessary and appropriate, compared to those values for the *negative control soil*. A *Student's t-test* or other appropriate statistic (EC, 2004a) should be used for this comparison. For a multi-concentration test (see Sections 5.3 and 6.2), the 14- or 21-day *ICp* for growth inhibition

[43] If soaking of soils is necessary to ease removal of plants (see 2nd paragraph of Section 4.7), aliquots for analyses of chemical constituents (i.e., concentration of contaminants) should be collected before the soil is soaked and the plants are removed. Alternatively, independent replicates might be set up for this purpose alone (see footnote 37).

[44] If there is no concentration-response for the emergence data in a multi-concentration test (see Sections 4.1, 5.3, and 6.2), the emergence data in each treatment, expressed as a percentage of the control, must be plotted to determine whether there is an adverse effect on seedling emergence (see Section 4.8.1). Enhanced emergence or inhibition of emergence in each treatment relative to the control, and the presence of a concentration-response relationship, can then be determined from the histogram.

represented by each endpoint measurement (i.e., decreased mean length of individual plant shoots and roots, and decreased mean dry weights of individual plant shoots and roots) must be calculated and reported (data permitting).[45]

Environment Canada (2004a) provides direction and advice for calculating ICps, which should be followed; Section 4.8.3 (including Appendix I) gives further guidance in this regard. Initially, regression techniques (see Section 4.8.3.1) must be applied to multi-concentration data intended for calculation of an ICp.[46] In the event that the data do not lend themselves to calculating the 14- or 21-day ICps for the growth inhibition using the appropriate regression analysis (see Appendix I), linear interpolation of these data using the program ICPIN should be applied in an attempt to derive an ICp (see Section 4.8.3.2).

4.8.1 Percent Emergence

The mean and standard deviation of seedling *emergence* are calculated for each test concentration. The percent effect is then calculated for each treatment using the following formula:

$$\text{Percent effect} = \frac{(\text{mean treatment emergence} - \text{mean control emergence})}{\text{mean control emergence}} \times 100$$

The percent effect is then plotted against test concentration in a histogram, with the median line representing the control response, or the 0% effect. All histogram bars above the median line (+ve percent effect) indicate that there is an enhanced *emergence* relative to that in the control, and histogram bars below the median line (-ve percent effect) indicate that there is an inhibition of *emergence* relative to that in the control. The magnitude and consistency of the percent effect among treatments indicates whether or not there is a concentration-response relationship. If there is an obvious, visible adverse effect in an exposure-dependent manner (i.e., there is a visual concentration-response relationship), the 14-day or 21-day EC50 must be calculated and reported (data permitting).

4.8.2 EC50

When a multi-concentration test with soil mixtures is conducted (Section 6.2), the *quantal* seedling emergence data for a specific period of exposure must be used to calculate (data permitting) the appropriate *median effect concentration (EC50)* for inhibition of percent emergence, together with its 95% confidence limits. For barley, cucumber, durum wheat, lettuce, radish, red clover, and tomato, a multi-concentration test must determine the 14-day EC50 for inhibition of percent emergence (at test end); and for alfalfa, blue grama grass, carrot, northern wheatgrass, and red fescue, the 21-day EC50 must be determined (at test end). The seven-day EC50 (i.e., that based on emergence data collected on Day 7 of the test) for inhibition of percent emergence, might also be determined and reported, data permitting (see Section 4.6).

To estimate an EC50, emergence data at the specified period of exposure are combined for all replicates at each concentration (including the replicate control groups). If emergence is not ≥50% in at least one concentration, the EC50 cannot be estimated. If there is complete emergence at a specific concentration, that information is used as a

---

[45] Historically, investigators have frequently analyzed *quantitative* sublethal data from multi-concentration tests by calculating the no-observed-effect concentration (NOEC) and the lowest-observed-effect concentration (LOEC). Disadvantages of these statistical endpoints include their dependence on the test concentrations chosen and the inability to provide any indication of *precision* (i.e., no 95% or other confidence limits can be derived) (NERI, 1993; EC, 2004a). Given these disadvantages, ICp is the required statistical endpoint for growth data derived in a multi-concentration test using terrestrial plants.

[46] Regression is the method of choice for estimating ICp. It involves fitting the data mathematically to a selected model and then calculating the statistical endpoint using the model that best describes the exposure-concentration response relationship. Nonlinear regression techniques were originally recommended by Stephenson *et al.* (2000b) for several reasons including: the relationship that exists between exposure concentration and plant response is typically nonlinear; the *heteroscedasticity* of the data is rarely reduced by transformation; the more standard bootstrap simulation technique has several limitations for these types of data; and nonlinear regression can fit effect distributions showing *hormesis*. By using standard mathematical techniques, a regression can be well described in terms that convey useful information to others, effects at high and low concentrations can be predicted, and confidence intervals can be estimated. Deficiencies of the smoothing and interpolation method can be largely remedied (EC, 2004a).

0% effect of emergence. However, if successive concentrations yield a series of 100% emergence, only the highest concentration of the series should be used in estimating the EC50 (i.e., the zero-effect that is "closest to the middle" of the distribution of data). Similarly, if there was a series of successive complete inhibition of emergence at the high concentrations in the test, only one value of 100% effect would be used, i.e., the one at the lowest concentration. Use of only one 0% and one 100% effect applies to any form of statistical analysis and to plotting on a graph.

The guidance provided by Environment Canada (2004a) on choosing statistical test methods to be applied to *quantal* (e.g., EC50) data should be consulted when choosing the statistical test to be applied to such data for toxicity tests using plants. Probit and/or logit regressions are the "preferred" methods (EC, 2004a), provided that two or more concentrations showing partial effects are included in the data. The probit analysis also gives the slope of the line, which should be reported. If probit or logit do not work because of only one partial effect, use the Spearman-Kärber method with no trim. If no partial effect is evident, use the binomial method. The binomial estimate might differ somewhat from the others, and this estimate should only be used as a last resort. Formal confidence limits are not estimated using the binomial method; instead, outer limits of a range are provided, within which the EC50 and its true confidence limits would lie.

Various computer programs may be used to calculate the EC50. Stephan (1977) developed a program to estimate EC50s using probit, moving average, and binomial methods, and adapted it for the IBM-compatible personal computer. Use of this program, which was modified in 1989 to include estimates using the Spearman Kärber method with no "trimming" (i.e., with no deletion of data from the calculations), is available on diskette[47] from Environment Canada (address in Appendix B), and is recommended. Other satisfactory computer and manual methods may be used (e.g., SAS, 1988 or version 3.5 of TOXSTAT, 1996; see EC, 2004a for additional information). Programs using the trimmed Spearman-Kärber method are available for personal computers; however, this method (with trimming) should be applied cautiously to EC50 estimates according to EC (2004a), because divergent results might be obtained by operators who are unfamiliar with the implications of trimming ends of the concentration-response data. However, there are situations where application of the trimmed Spearman Kärber method is warranted (see EC, 2004a for guidance).

Any computer-derived EC50 should be checked by examining a plot, on logarithmic-probability scales, of percent emergence at a defined period of exposure for the various test concentrations (EC, 2004a). Any major disparity between the estimated EC50 derived from this plot and the computer-derived EC50 must be resolved. A hand-plotted graph is recommended for this check (EC, 2004a). A computer-generated plot (e.g., SigmaPlot™; Version 8.0.2 or later)[48] could be used if it were based on logarithmic-probability scales. If there has been an error in entering the data, however, a computer-generated plot would contain the same error as the mathematical analysis, and so the investigator should carefully check for correct placement of points (EC, 2004a).

A manual plot of emergence (mortality)/concentration data to derive an estimated EC50 is illustrated in Figure 2. This (hypothetical) figure is based on test concentrations of 1.8, 3.2, 5.6, 10, and 18 mg chemical/kg soil (dry-weight basis) causing emergence inhibition of 0, 20, 40, 90, and 100% of seedlings exposed to the respective concentrations for a specified period of time. The concentration expected to inhibit the emergence of 50% of the seedlings can be read by following across from 50% (broken line) to the intersection with the best-fit line, then down to the horizontal axis for an estimated EC50 (5.6 mg/kg, dry wt).

In fitting a line such as that in Figure 2, more emphasis should be assigned to points that are near 50% inhibition of emergence. Logarithmic-probability paper (*log-probit*, as in Figure 2) can be purchased in good technical bookstores, ordered through them, or photocopied (see blank graph in EC, 2004a).

---

[47] Through the courtesy of Dr. Charles E. Stephan (USEPA, Duluth, MN).

[48] Available for purchase from SYSTAT Software, Inc., 501 Canal Boulevard, Suite C, Point Richmond, CA 94804-2028, USA, phone no. 1-800-797-7401; see Web site www.systat.com/products/SigmaPlot/.

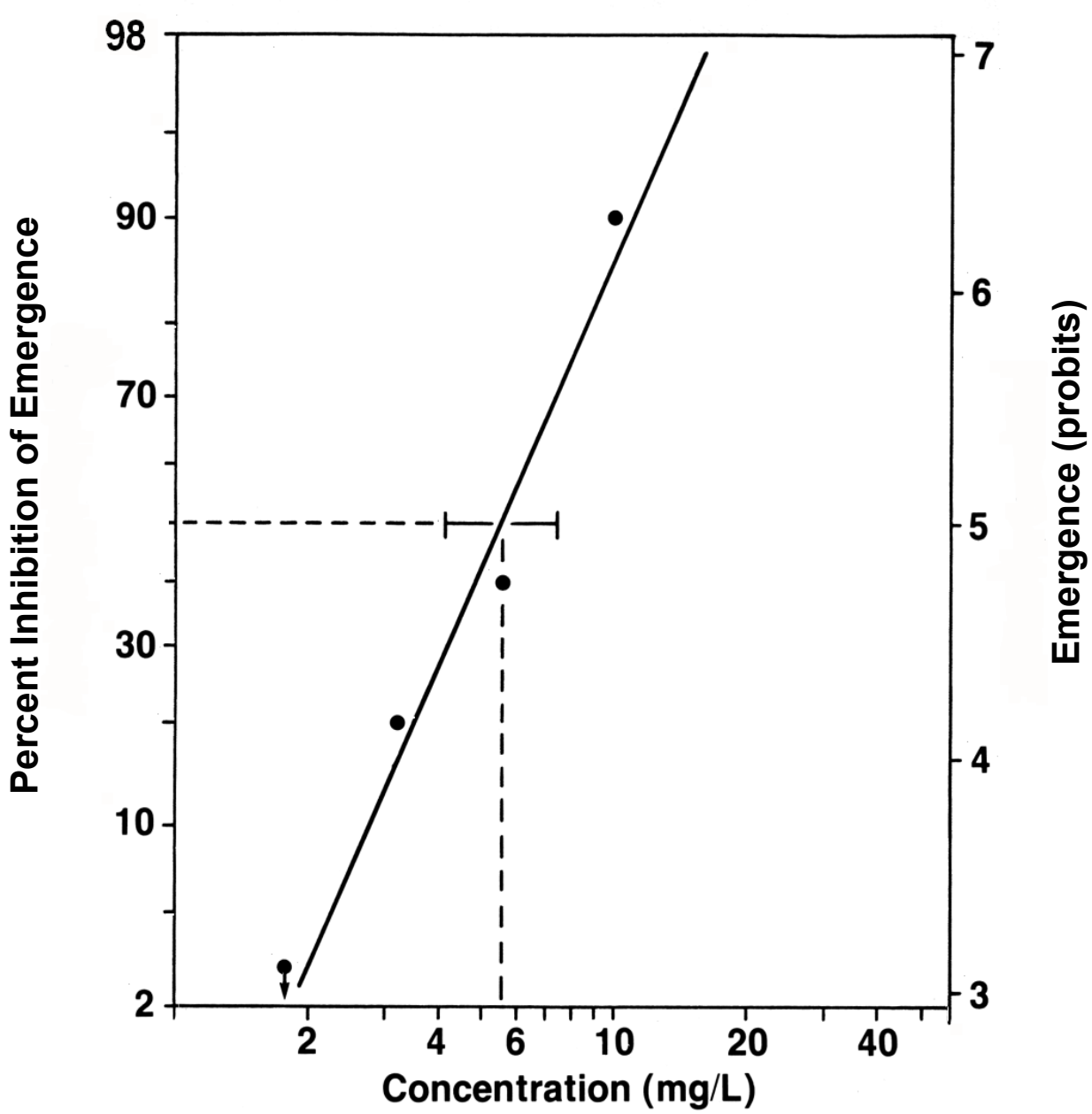


Figure 2 Estimating a Median Effective Concentration by Plotting Emergence on Logarithmic-Probability Paper

For the regular set of data in Figure 2, computer programs gave very similar estimates to the graphic one. Some of the computed EC50s (and 95% confidence limits) were:

Stephan (1977) method:
- probit: 5.58 (4.24 and 7.37)
- moving average: 5.58 (4.24 and 7.33)
- binomial: 6.22 (between 1.8 and 10)

SAS (1988) probit analysis: 5.58 (4.26 and 7.40)

TOXSTAT (1996) method (version 3.5)
- probit: 5.58 (4.38 and 7.12)
- Spearman Kärber, zero trim: 4.64 (4.40 and 7.23)
- logit: 5.63 (4.39 and 7.22)

Table 4.2 in EC (2004a) provides additional examples of computed data for *acute* quantal tests using various computer programs.

4.8.3 ICp

When a multi-concentration test for effects of exposure of terrestrial plants to *spiked soil* mixtures is conducted, the *quantitative* continuous data representing growth inhibition (i.e., shoot and root length, and shoot and root dry mass) must be used to calculate an *ICp* (*inhibiting concentration for a specified percent effect*) for each of these four endpoints, data permitting (see introductory paragraphs of Sections 4.8 and 6.2). The ICp is a quantitative estimate of either:

(1) the concentration causing a fixed percent reduction in the mean length of individual plant shoots at test end;

(2) the concentration causing a fixed percent reduction in the mean length of individual plant roots at test end;

(3) the concentration causing a fixed percent reduction in the mean dry weight of individual plant shoots at test end; or

(4) the concentration causing a fixed percent reduction in the mean dry weight of individual plant roots at test end.

The ICp is calculated as a specified percent reduction for each endpoint (e.g., the IC25 and/or IC20, for a 25% and/or 20% reduction, respectively). The desired value of *p* is selected by the investigator, and 25% or 20% is currently favoured. Any ICp that is calculated and reported must include the 95% confidence limits.

In the analyses of *growth*, the length and weight measurements of individual shoots or roots in each replicate (test vessel) are pooled for each of these measurements, and the mean of these lengths and weights are used in the analyses. For length measurements, the mean length of individual shoots (or roots) in each replicate is calculated. For dry weight measurements, the mean weight of individual shoots (or roots) in each replicate is calculated as the total dry weight of all of the plant shoots (or roots) that survived in the test vessel divided by the number of plants that survived in that vessel to the end of the test.[49, 50]

---

[49] In a dual-effect test (i.e., one that measures growth and survival), it is often preferable to analyze the *sublethal effect* (i.e., growth) separately from any mortality, to estimate the endpoint (i.e., ICp) (EC, 2004a). This is accomplished by dividing the mean weight of shoots or roots in each replicate by the number of seedlings that survived to the end of the test. Alternatively, productivity can be analyzed by measuring the weight of all shoots or all roots as a group in each replicate (i.e., no division by surviving seedlings). Productivity is a population indicator which combines both effects (i.e., survival and growth) and therefore provides less opportunity for interpretation of the data. Futhermore, if there is an interest in calculating the "productivity", the investigator can do so once the individual effects have been determined (i.e., growth and survival have been assessed separately) (A. Renoux, personal communication, SANEXEN Environmental Ltd., Varennes, QC, 2004). Further guidance on assessing data from dual-effect tests is provided by EC (2004a; Section 8).

[50] To measure the dry weight of individual shoots (or roots) within each replicate, the total dry weight of plant shoots (or roots) in each replicate is divided by the number of plants that survived in that replicate to the end of the test. For example, if ten alfalfa seeds were planted in a given replicate, but only seven plants survived to the end of the test, then the dry weight measurement for that replicate would be the dry weight of all seven shoots, divided by seven. The same holds true for whole roots or shoots lost (e.g., washed down the drain) during the processing of a test vessel.

For plants that break apart during the processing of test vessels (i.e., a portion of the shoot or root breaks off),

The mean lengths and weights from all the replicates of a given treatment (concentration) are used to calculate the average for the treatment; this is the average individual shoot and root length and shoot and root dry weight of surviving plants per concentration. These data are compared to the average individual lengths of shoots and roots and the average individual weights of shoots and roots in the *negative control*, obtained by the same procedure. If there are no emerged plants in a replicate (test vessel), that replicate does not contribute to the average for the treatment. If there are no emerged plants in all replicates at a given concentration, that concentration does not have an average length or weight of emerged plants and cannot be used in the analysis for comparison with the average length or weight in the negative control.

As indicated in the introductory paragraphs of Section 4.8, separate ICps for individual shoot and root length and shoot and root dry mass must be calculated and reported (data permitting) upon completion of the test. These calculations must be made using the appropriate linear or nonlinear regression analyses (see Section 4.8.3.1). If, however, regression analyses fail to provide meaningful ICps for shoot/root length or shoot/root dry weight, the ICPIN analyses described in Section 4.8.3.2 should be applied to the corresponding data.

professional judgement should be used to determine what to include in the final length and weight measurements. If the portion of the root or shoot that broke off is recovered (i.e., not lost), then it can be included in both the length and weight measurements. If, however, the portion of the shoot or root that broke off is lost, then the length measurement for that shoot or root should be excluded from the final length determination for the replicate. For dry weight measurements, the decision as to what to include (or not to include) depends on the estimated proportion of the root (or shoot) that is missing. For example, if the portion of a missing root is less than half of the root (i.e., based on the length of the other roots in the replicate), then the missing portion would not have a large overall effect on the final dry weight per root, and therefore, the remaining portion of the root may be included in the dry weight measurement for that replicate. If, however, the portion of the missing root is estimated to be more than half of the root, then that root should be excluded from the weight analysis (i.e., the root is not dried and weighed and the final "per plant" dry weight for the replicate is based only on the number of roots that were dried and weighed).

4.8.3.1 Use of regression analysis. Upon completion of a *definitive* 14- or 21-day multi-concentration test, separate ICps (including their respective 95% confidence limits) for the individual mean lengths and dry weights of shoots and roots, must be calculated using linear and/or nonlinear regression procedures. These values may be calculated using a series of linear and nonlinear regression models (data permitting) proposed by Stephenson *et al*. (2000b) that have been re-parameterized, based on techniques applied by van Ewijk and Hoekstra (1993), to automatically generate the ICp and its 95% confidence limits for any value of 'p' (e.g., IC25 or IC50). The proposed models for application consist of one linear model, and the following four nonlinear regression models: exponential, Gompertz, logistic, and logistic adjusted to accommodate *hormesis*[51]. Further guidance on the use of these linear and nonlinear regression models for calculating ICps is provided by Stephenson (2003a) and Stephenson *et al*. (2000b). The reader is also strongly advised to consult EC (2004a) for additional guidance on the general application of linear and non-linear regression for the analysis of *quantitative* toxicity data. Instruction for the appropriate application of linear and non-linear regression, using Version 11.0 of the statistical program SYSTAT[52], is provided in Appendix I. However, any statistical software capable of linear and nonlinear regression may be used when

[51] A hormetic response (i.e., *hormesis*) might be found in sublethal observations at the lowest concentration(s), i.e., performance at such concentration(s) is enhanced relative to that in the negative control. For instance, the shoot and root lengths might be longer for seedlings grown in soil with low concentrations than for those grown in the control treatment, or the dry weights of shoots or roots might be substantially greater relative to those for seedlings grown in the control. This is not a flaw in the testing. Rather, it is a real biological phenomenon. To calculate the ICp when this phenomenon occurs, the data should be analyzed using the hormesis model. The hormetic effects are included in the regression, but do not bias the estimate of the ICp. An estimated IC25 would still represent a 25% reduction in performance from that of the control.

[52] The latest (e.g., Version 11.0) version of SYSTAT™ is available for purchase by contacting SYSTAT Software, Inc., 501 Canal Boulevard, Suite C, Point Richmond, CA 94804-2028, USA, phone no. 1-800-797-7401; see Web site www.systat.com/products/Systat/.

calculating the respective ICps and their associated 95% confidence limits. Appendix I provides instruction on the use of regression models to derive the most appropriate ICps for reduced plant growth, assessed using shoot and root length and dry weight metrics.

The five models recommended for application follow. Further information on these specific models is presented in Appendix I.

Exponential model:

$$Y = a \times (1 - p)^{(C \div ICp)}$$

where:

- Y = root or shoot length or dry mass
- a = the y-intercept (i.e., the control response)
- p = desired value for 'p' (e.g., 0.25 for a 25% inhibition)
- C = the test concentration as a logarithm
- ICp = the ICp for the data set

Gompertz model:

$$Y = t \times \exp[\log(1 - p) \times (C \div ICp)^b]$$

where:

- Y = root or shoot length or dry mass
- t = the y-intercept (i.e., the control response)
- exp = the exponent of the base of the natural logarithm
- p = desired value for 'p' (e.g., 0.25 for a 25% inhibition)
- C = the test concentration as a logarithm
- ICp = the ICp for the data set
- b = a scale parameter (estimated to be between 1 and 4) that defines the shape of the equation

Hormesis model:

$$Y = t \times [1 + (h \times C)] \div \{1 + [(p + (h \times C)) \div (1 - p)] \times (C \div ICp)^b\}$$

where:

- Y = root or shoot length or dry mass
- t = the y-intercept (i.e, the control response)
- h = describes the hormetic effect (estimated to be small, usually between 0.1 and 1)
- C = the test concentration as a logarithm
- p = desired value for 'p' (e.g., 0.25 for a 25% inhibition)
- ICp = the ICp for the data set
- b = a scale parameter (estimated to be between 1 and 4) that defines the shape of the equation

Linear model:

$$Y = [(-b \times p) \div ICp] \times C + b$$

where:

- Y = root or shoot length or dry mass
- b = the y-intercept (i.e., the control response)
- p = desired value for 'p' (e.g., 0.25 for a 25% inhibition)
- ICp = the ICp for the data set
- C = the test concentration as a logarithm

Logistic model:

$$Y = t \div \{1 + [p \div (1 - p)] \times (C \div ICp)^b\}$$

where:

- Y = root or shoot length or dry mass
- t = the y-intercept (i.e, the control response)
- p = desired value for 'p' (e.g., 0.25 for a 25% inhibition)
- C = the test concentration as a logarithm
- ICp = the ICp for the data set
- b = a scale parameter (estimated to be between 1 and 4) that defines the shape of the equation

The general process for the statistical analysis and selection of the most appropriate regression model (linear or non-linear) for *quantitative* toxicity data is outlined in Figure 3. The selection process begins with an examination of a scatter plot or line graph of the test data to determine the shape of the concentration-response curve. The shape of the curve is then compared to available models so that one or more appropriate model(s) that best suits the data is (are) selected for further examination (see Figure I.1, Appendix I, for an example of five potential models).

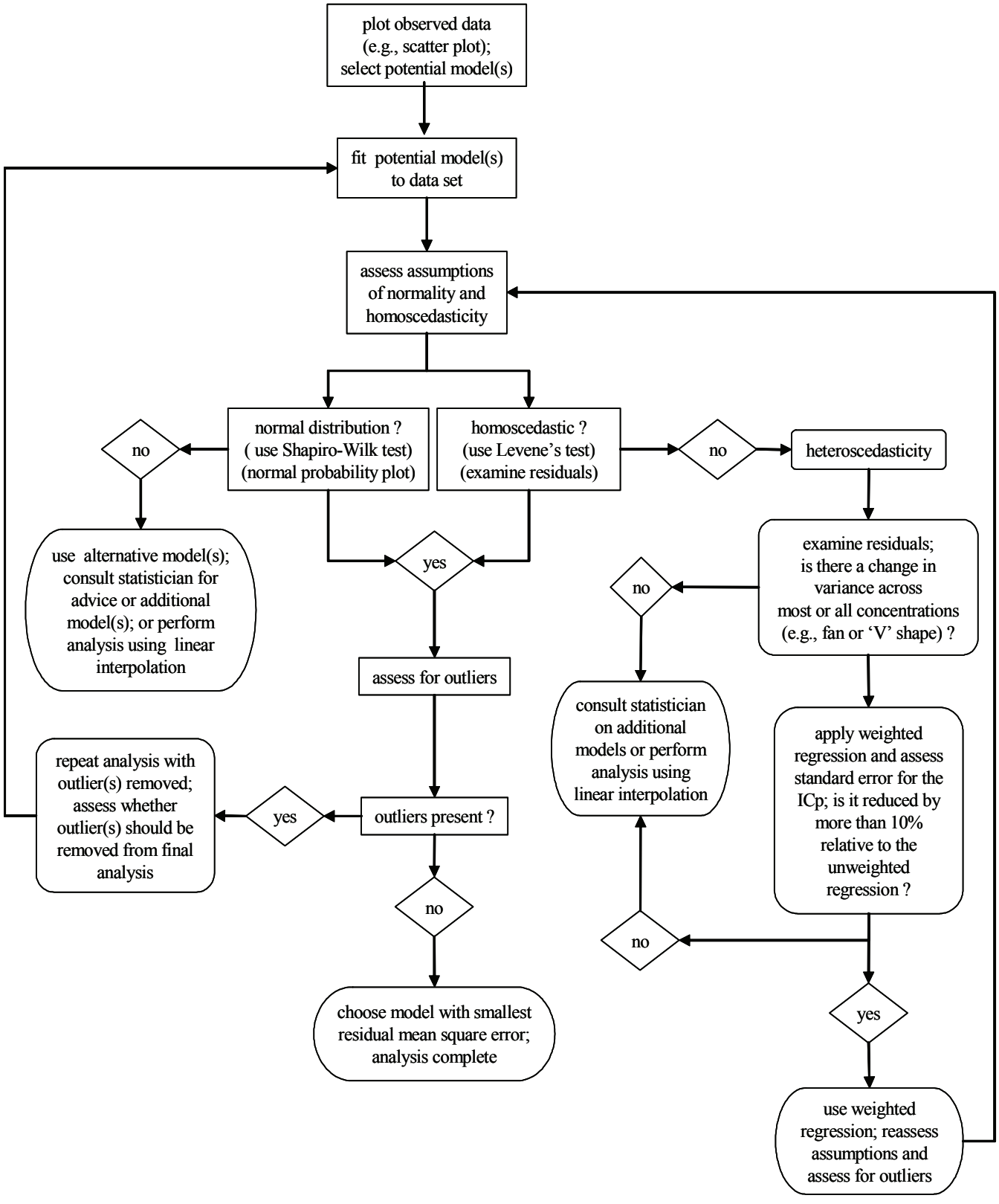


Figure 3 **The General Process for the Statistical Analysis and Selection of the Most Appropriate Model for Quantitative Toxicity Data** (adapted and modified from Stephenson *et al*., 2000b)

Once the appropriate model(s) is (are) selected for further consideration, assumptions of normality and *homoscedasticity* of the *residuals* are assessed. If the regression procedure for one or more of the examined models meets the assumptions, the data (and regression) are examined for the presence of outliers. If an outlier has been observed, the test records and experimental conditions should be scrutinized for human error. If there are one or more outliers present, the analysis should be performed with and without the outlier(s), and the results of the analyses compared to examine the effect of the outlier(s) on the regression. Thereafter, a decision must be made as to whether the outlier(s) should be removed from the final analysis. The decision should take into consideration natural biological variation, and biological reasons that might have caused the apparent anomaly. Additional guidance on the presence of outliers and unusual observations is provided in Appendix I (Section I.2.4), as well as in EC (2004a). If there are no outliers present or none are removed from the final analysis, the model that demonstrates the smallest residual mean square error is selected as the model of best choice. Additional guidance from a statistician familiar with dealing with outlier data is also advised.

Normality should be assessed using the *Shapiro-Wilk's test* as described in EC (2004a). A normal probability plot of the *residuals* may also be used during the regression procedure, but is not recommended as a stand-alone test for normality as the detection of a 'normal' or 'non-normal' distribution depends on the subjective assessment of the user. If the data are not normally distributed, then the user is advised to try another model, consult a statistician for further guidance on model selection, or to perform the less-desirable linear interpolation (using ICPIN, see Section 4.8.3.2) method of analysis.

Homoscedasticity of the *residuals* should be assessed using *Levene's test* as described in EC (2004a), and by examining the graphs of the residuals against the actual and predicted (estimated) values. Levene's test provides a definite indication of whether the data are homogeneous (e.g., as in Figure I.2A of Appendix I) or not. If the data (as indicated by Levene's test) are *heteroscedastic* (i.e., not homogeneous), then the graphs of the residuals should be examined. If there is a significant change in the variance and the graphs of the residuals produce a distinct fan or 'V' pattern (refer to Figure I.2B, Appendix I for an example), then the data analysis should be repeated using weighted regression. Before choosing the weighted regression, the standard error of the ICp is compared to that derived from the unweighted regression. If there is a difference of greater than 10% between the two standard errors[1], then the weighted regression is selected as the regression of best choice. However, if there is less than a 10% difference in the standard error between the weighted and unweighted regressions, then the user should consult a statistician for the application of additional models, given the test data, or the data could be re-analyzed using the less-desirable linear interpolation (using ICPIN, see Section 4.8.3.2) method of analysis. This comparison between weighted and unweighted regression is completed for each of the selected models while proceeding through the process of final model selection (i.e., model and regression of best choice). Some non-divergent patterns might be indicative of an inappropriate or incorrect model (refer to Figure I.2C, Appendix I, for an example), and the user is again urged to consult a statistician for further guidance on the application of additional models.

4.8.3.2 Linear interpolation using ICPIN. If regression analyses of the endpoint data (see Section 4.8.3.1) fail to provide an acceptable ICp for growth, linear interpolation using the computer program called *ICPIN* should be applied. This program (Norberg-King, 1993; USEPA, 1994b, 1995) is not proprietary, is available from the USEPA, and is included in most computer software for environmental toxicology, including TOXSTAT. The original instructions for ICPIN from USEPA are clearly written and make the program easy to use

[53] The value of 10% is only a "rule-of-thumb" based upon experience. Objective tests for the improvement due to weighting are available, but beyond the scope of this document. Weighting should be used only when necessary, as the procedure can introduce additional complications to the modeling procedure. A statistician should be consulted when weighting is necessary.

(Norberg-King, 1993).[54] An earlier version was called BOOTSTRP.

Analysis by ICPIN does not require equal numbers of replicates in different concentrations. The ICp is estimated by smoothing of the data as necessary, then using the two data-points adjacent to the selected ICp (USEPA, 1994b, Appendix L; USEPA, 1995, Appendix L). The ICp cannot be calculated unless there are test concentrations both lower and higher than the ICp; both those concentrations should have an effect reasonably close to the selected value of p, preferably within 20% of it. At present, the computer program does not use a logarithmic scale of concentration, and so Canadian users of the program must enter the concentrations as logarithms. Some commercial computer packages have the logarithmic transformation as a general option, but investigators should make sure that it is actually retained when proceeding to ICPIN. ICPIN estimates confidence limits by a special "bootstrap" technique since usual methods would not be valid. Bootstrapping performs many resamplings from the original measurements. The investigator must specify the number of resamplings, which can range from 80 to 1000. At least 400 is recommended here, and 1000 would be beneficial.[55]

If there are several adjacent high concentrations with no emerged plants, only the lowest of that string of concentrations should be used in the analysis (i.e., the concentration closest to the middle of the series of concentrations used in the test). Normally, there is no particular benefit to including the additional concentrations, since they offer nothing to the analysis (i.e., the data consist only of zero mean weights and zero mean lengths).

Besides determining and reporting the computer-derived ICps for length and weight of individual plants at test end, separate graphs of percent reduction for each of shoot and root lengths and shoot and root dry weights should be plotted against the logarithm of concentration, to check the mathematical estimations and to provide visual assessments of the nature of the data (EC, 2004a).

If the ICPIN program is used when there is a hormetic effect, an inherent smoothing procedure could change the control value and bias the estimate of ICp. Accordingly, before statistical analysis, hormetic values at low concentration(s) should be arbitrarily replaced by the control value. This is considered a temporary expedient until a superior approach is established (EC, 2004a). The correction is applied for any test concentration in which the average effect (i.e., the geometric average of the replicate means) is higher ("better") than the average for the control. To apply this correction, replace the observed mean weights (or mean lengths) of the replicates in the hormetic concentration(s), with the means of replicates in the control. The geometric average for that/those concentration(s) will then be the same as that for the control.

4.9 Tests with a Reference Toxicant

Table 13 of Appendix E summarizes the guidance for performing reference toxicity tests given in other documents describing procedures and conditions for conducting emergence-and-growth tests in soil using plants. Described herein are the procedures and conditions to be followed when performing reference toxicity tests in conjunction with a 14-day or 21-day test of *emergence* and *growth* using plants. The procedures herein also apply to tests for assessing the acceptability and suitability of batches of seed for use in soil toxicity tests; and should be applied to assess intralaboratory *precision* when a laboratory is inexperienced with the biological test method defined in this document and is initially setting up to perform it (see Sections 2.5, and 3.2.1).

The routine use of a *reference toxicant* is necessary to assess, under standardized test conditions, the

---

[54] The instructions in Norberg-King (1993) are sometimes misleading on the identity of "replicates". The term is used in such a way that it would apply to numbers or weights of individual organisms within the same vessel. This slip of wording does not affect the functioning of the program. Some commercial programs have been less user-friendly for entry of data and analysis.

[55] ICPIN has some deficiencies which is why it is recommended only in cases where the use of regression fails to provide an acceptable ICp. Its interpolation method is an inefficient use of data, sensitive to peculiarities of the two concentrations used. The program fails to adopt logarithm of concentration, which would introduce a slight bias towards a higher value of ICp. A modification of the bootstrap method has now remedied a problem of overly narrow confidence limits; however, regression analyses provide better methods of estimating the ICp and its 95% confidence limits (EC, 2004a) (see Section 4.8.3.1).

relative sensitivity of a lot of terrestrial plant seed being used. Tests with a *reference toxicant* also serve to demonstrate the *precision* and reliability of data produced by the laboratory personnel for that *reference toxicant*, under standardized test conditions. A reference toxicity test, conducted according to the procedures and conditions described herein, must be performed according to one of the following regimes:

(1) at least once every two months using the same lot of seed being used to provide test organisms for soil toxicity tests over an extended period (i.e., ≥2 months); or

(2) at the same time as the definitive soil toxicity test(s), using seed taken from the same lot number as those used for the definitive test(s) (see Section 2.5).

A laboratory that frequently performs soil toxicity tests using terrestrial plants might choose to routinely (e.g., every two months) monitor the sensitivity of their seed to one or more *reference toxicant*s, while including a reference toxicity test using a portion of the seeds used to start a definitive soil toxicity test. Alternatively, a laboratory might choose to monitor the sensitivity of their seed to a reference toxicant less frequently, and to perform a reference toxicity test at the time that each definitive soil toxicity test is performed.

Each reference toxicity test performed in conjunction with the *definitive* test for soil toxicity must be conducted as a *static* multi-concentration growth test. The duration of the reference toxicity test is seven days if the species of organisms is alfalfa, barley, cucumber, durum wheat, lettuce, radish, red clover, or tomato. For reference toxicity tests with blue grama grass, carrot, northern wheatgrass, and red fescue, the test duration is 10 days. In each instance, the ICp for shoot length is determined at the end of the test. A summary checklist in Table 3 describes the conditions and procedures that must be applied to each reference toxicity test. Additional conditions and procedures described in Section 4 for performing a multi-concentration test with samples of *test soil* apply equally to each reference toxicity test. Procedures given in Section 6 for the preparation and testing of chemicals spiked in *negative control soil* also apply here, and should be referred to for further information. Environment Canada's guidance document on using negative control sediment spiked with a reference toxicant (EC, 1995) provides useful information that is also applicable when performing reference toxicity tests with *negative control soil* spiked with a reference toxicant.

The reference toxicity test should be performed using 1-L polypropylene containers as test vessels (Section 3.2.2) and a 500-mL aliquot of *test soil* representing each treatment (concentration) in each test vessel. The number of replicate test vessels per concentration must be ≥3. The number of seeds per vessel is species-specific, and differ slightly from those specified for definitive tests. Reference toxicity tests with barley, cucumber, durum wheat, lettuce, radish, red clover, red fescue, and tomato must include five seeds per vessel, whereas for alfalfa, blue grama grass, carrot, and northern wheatgrass, 10 seeds per vessel are required (see Table 3).

Procedures for starting and ending a reference toxicity test should be consistent with those described in Sections 4.2 and 4.7. Test conditions for temperature and light, described in Section 4.3, apply. Test observations and measurements given in Section 4.6 should be followed. Observations and measurements should be as described in Section 4.6; however, only percent *emergence* and individual shoot length should be determined at the end of the test.

To be valid, the mean percent *emergence* at the end of the test for plants held in the control treatment used in a particular reference toxicity test must be:

- ≥60 % for tomato;
- ≥70 % for blue grama grass, carrot, lettuce, northern wheatgrass, red clover, or red fescue;
- ≥80 % for alfalfa, barley, cucumber, or durum wheat; or
- ≥90 % for radish.

Additionally, the mean shoot length for individual plant species grown in negative control soil for the duration of the test must be:

- ≥10 mm for lettuce or red clover;
- ≥20 mm for alfalfa, blue grama grass, or tomato;
- ≥40 mm for carrot, cucumber, radish, or red fescue;
- ≥50 mm for northern wheatgrass;
- ≥100 mm for barley; or
- ≥120 mm for durum wheat.

Table 3 Checklist of Recommended Conditions and Procedures for Conducting Reference Toxicity Tests on Soil Using Terrestrial Plants

Test type	–	whole soil reference toxicity test; no renewal (static test)
Test duration	–	7 days for alfalfa, barley, cucumber, durum wheat, lettuce, radish, red clover, or tomato; and
	–	10 days for blue grama grass, carrot, northern wheatgrass, or red fescue
Approved test species	–	monocotyledons: barley *(Hordeum vulgare)*, blue grama grass *(Bouteloua gracilis*), durum wheat *(Triticum durum)*, northern wheatgrass *(Elymus lanceolatus*; formerly named *Agropyron dasystachyum),* and red fescue *(Festuca rubra)*;
		dicotyledons: alfalfa *(Medicago sativa)*, carrot *(Daucus carota)*, cucumber *(Cucumis sativus)*, lettuce *(Lactuca sativa)*, radish *(Raphanus sativus)*, red clover *(Trifolium pratense)*, and tomato *(Lycopersicon esculentum)*
Number of concentrations	–	minimum of five test concentrations, plus a negative control
Number of replicates	–	≥ 3 replicates/concentration
Number of seeds per replicate	–	5 seeds/vessel for barley, cucumber, durum wheat, lettuce, radish red clover, red fescue, or tomato; and
	–	10 seeds/vessel for alfalfa, blue grama grass, carrot, or northern wheatgrass
Negative control soil	–	artificial soil
Test vessel	–	polypropylene cups (1 L), covered for seven days or until plants reach top of container
Amount of soil/ test vessel	–	identical wet wt, equivalent to a volume of ~500 mL; ~350 g dry wt for artificial soil
Moisture content	–	for soil preparation, hydrate to ~70% of water-holding capacity (WHC) for artificial soil; during test, hydrate to saturation, as needed
Air temperature	–	daily range, constant 24 ± 3 °C; alternatively, day: 24 ± 3 °C, night: 15 ± 3 °C
Humidity	–	≥50%
Lighting	–	full spectrum fluorescent: mimic natural light spectrum (e.g., Vita Lite® by Duro-Test®); 300 ± 100 μmol/(m^2 · s) adjacent to the level of the soil surface; 16 h light:8 h dark
Watering	–	hydration water sprayed over soil surface until saturation, about every two days when covered and once per day after covers are removed, or whenever soil appears dry
Measurements during test	–	soil moisture content in each treatment/concentration at start; pH in each treatment/concentration at start and end; temperature in test facility, daily or continuously; humidity in test facility; light intensity once during test

Observations during test	–	number of emerged seedlings at the end of the test in each test vessel and shoot length at test end; number of surviving plants showing an atypical appearance (e.g., chlorosis, lesions)
Biological endpoints	–	emergence of seedlings during test and length of longest shoot at test end; appearance of surviving plants at test end
Statistical endpoints	–	mean (± SD) percent emergence in each treatment/concentration at test end (i.e., on Day 7 or Day 10); mean (± SD) length of longest shoots in each treatment/test concentration at test end (Day 7 or Day 10); 7-day or 10-day ICp for shoot length
Test validity	–	invalid if **any** of the following occurs in negative control soil at test end: • mean percent emergence is <60% for tomato; <70% for blue grama grass, carrot, lettuce, northern wheatgrass, red clover, or red fescue; <80% for alfalfa, barley, cucumber, or durum wheat; or <90% for radish • mean percent survival of emerged seedlings in negative control soil at test end is <90% • mean percentage of control seedlings exhibiting phytotoxicity or developmental anomalies is >10% • mean shoot length is <10 mm for lettuce or clover; <20 mm for alfalfa, blue grama grass, or tomato; <40 mm for carrot, cucumber, radish, or red fescue; <50 mm for northern wheatgrass; <100 mm for barley; or <120 mm for durum wheat

In addition, the mean percent survival of emerged seedlings in the negative control soil must be ≥90 % at the end of the test; and the mean percentage of seedlings grown in negative control soil that exhibit phytotoxicity and/or developmental anomalies must be ≤10% in order for the results of a reference toxicity test to be declared valid.

Test endpoints to be calculated and reported include the mean percent *emergence* in each treatment on Day 7 or Day 10, depending on the species. The 7-day or 10-day ICp (including its 95% confidence limits) for shoot length must also be calculated. Results for a reference toxicity test should be expressed as mg reference chemical/kg soil, on a dry-wt basis.

Appropriate criteria for selecting the reference toxicant to be used in conjunction with a definitive test for soil toxicity include the following (EC, 1995):

- chemical readily available in pure form;
- stable (long) shelf life of chemical;
- can be interspersed evenly throughout *clean* substrate;
- good concentration-response curve for test organism;
- stable in aqueous solution and in soil;
- minimal hazard posed to user; and
- concentration easily analyzed with precision.

The reference toxicity test requires a minimum of six treatments (i.e., *negative control soil* and five concentrations of *reference toxicant*). Reagent-grade boric acid (H_3BO_3)[56] is recommended for use as the reference toxicant when performing soil toxicity tests with plants, although other chemicals

[56] Boric acid has been used historically as a soil chemo-sterilant and has been found to be an effective biocide. Boric acid dissociates readily in water with neutral pH and is readily absorbed, accumulated, and translocated by the roots of plants. It is relatively persistent in soils in laboratory conditions, and does not readily photodecompose or volatilize (Stephenson *et al.*, 1997).

may be used if they prove suitable.[57] Each test concentration should be made up according to the guidance in Sections 4.1 and 6.2, using *artificial soil* (Section 3.4.2) as substrate.

Routine reference toxicity tests (e.g., those performed once every two months or in conjunction with each definitive test for soil toxicity) using boric acid [or another suitable reference chemical, such as copper sulphate ($CuSO_4$)] spiked in *negative control soil* should consistently apply the same test conditions and procedures described herein. A series of test concentrations should be chosen[58], based on preliminary tests, to bracket the ICp and enable calculation of the 7-day or 10-day ICp for shoot length.

Once sufficient data are available (EC, 1995), all comparable ICps for a particular *reference toxicant* derived from these toxicity tests must be plotted successively on a *warning chart*. A separate warning chart must be prepared for each plant species used in definitive toxicity tests. Each new ICp for the same reference toxicant should be examined to determine whether it falls within ± 2 SD of values obtained in previous comparable tests using the same reference toxicant and test procedure (EC, 1997a, 1997b, 2001). A separate warning chart must be prepared and updated for each dissimilar procedure (e.g., differing plant species or differing reference toxicants). The warning chart should plot logarithm of concentration on the vertical axis against date of the test or test number on the horizontal axis. Each new ICp for the reference toxicant should be compared with established limits of the chart; the ICp is acceptable if it falls within the *warning limits*.

The logarithm of concentration (including ICp as a logarithm) should be used in all calculations of mean and standard deviation, and in all plotting procedures. This simply represents continued adherence to the assumption by which each endpoint value was estimated based on logarithms of concentrations. The warning chart may be constructed by plotting the logarithmic values of the mean and ± 2 SD on arithmetic paper, or by converting them to arithmetic values and plotting those on the logarithmic scale of semi-log paper. If it were demonstrated that the ICps failed to fit a log-normal distribution, an arithmetic mean and SD might prove more suitable.

---

[57] Aquaterra Environmental (1998a) initially evaluated the performance of boric acid as a candidate *reference toxicant* for use in conjunction with *acute toxicity* tests for measuring soil toxicity to terrestrial plant species. Subsequent studies by EC (2005a, b) using boric acid spiked in artificial soil confirmed the usefulness of boric acid as a suitable reference toxicant when performing bi-monthly 7-day or 10-day reference tests.

[58] ASTM (1999b) recommends a 0.5 dilution series for reference toxicity tests using boric acid. Specifically, ASTM (1999b) recommends that a 640, 320, 160, 80, 40, 20, and 10 mg/kg soil dry weight dilution series will bracket the sensitivity of most plant species to boric acid. Environment Canada (2005b) found that some species were not sensitive to concentrations as high as 640 mg boric acid/kg soil dry wt, and therefore recommended the inclusion of higher concentrations for reference toxicity tests using this chemical. A dilution series consisting of 2000, 1125, 630, 360, and 200 mg/kg soil dry wt should bracket the sensitivity of each of the test species recommended herein; however, the dose response should be modified to suit each species being tested.

Environment Canada (2005b) demonstrated endpoint values ranging from 452 mg boric acid/kg soil (dry wt) for carrot to 1603 mg boric acid/kg soil (dry wt) for alfalfa in their results for 7- and 10-day IC50s for shoot length, with boric acid mixed in artificial soil, using all 12 plant species, and the test method for a reference toxicity test, described herein. See Appendix H for guidance in selecting an appropriate series of test concentrations (assuming a log-concentration response) for use in toxicity tests with this or other chemicals to be used in a reference toxicity test.

As part of a series of interlaboratory studies performed to validate Environment Canada's reference toxicity test described in Section 4.9, six laboratories undertook concurrent seven-day reference toxicity tests with cucumber (*Cucumis sativus* var. Marketmore 76) and multiple concentrations of boric acid spiked in artificial soil. Each of the six participating labs achieved the validity criteria for percent emergence of control seedlings, but only five of the six participating labs achieved the validity criteria for mean shoot length of control seedlings. The data produced by the laboratory that failed to achieve the validity criteria was removed from the final data analysis. Based on data from the remaining five laboratories, the mean seven-day IC50 for seedling shoot length for boric acid in artificial soil was 693 mg H_3BO_3/kg dry wt, with values for individual laboratories ranging from 379–961 mg/kg. The CV of 30% for these IC50s showed acceptable interlaboratory *precision* (EC, 2005a).

The mean of the available values of log (ICp), together with the upper and lower warning limits (± 2 SD), should be recalculated with each successive ICp for the *reference toxicant* until the statistics stabilize (EC, 1995, 1997a, b, 2001). If a particular ICp fell outside the warning limits, the sensitivity of the test organisms and the performance and precision of the test would be suspect. Since this might occur 5% of the time due to chance alone, an outlying ICp would not necessarily indicate abnormal sensitivity of the seed, nor unsatisfactory precision of toxicity data. Rather, it would provide a warning that there might be a problem. A thorough check of all test conditions and procedures should be carried out. Depending on the findings, it might be necessary to repeat the reference toxicity test or purchase new seed before undertaking further soil toxicity tests.

Results that remained within the warning limits might not necessarily indicate that a laboratory was generating consistent results. Extremely variable historic data for a *reference toxicant* would produce wide warning limits; a new data point could be within the warning limits but still represent undesirable variation in test results. A CV of no more than 30%, and preferably 20% or less, has been suggested as a reasonable limit by Environment Canada (EC, 1995, 2004a) for the mean of the available values of log (ICp) (see preceding paragraph). For this biological test method, the CV for mean historic data derived for reference toxicity tests performed using boric acid should not exceed 30%.

Concentrations of *reference toxicant* in all *stock solutions* should be measured chemically using appropriate methods (e.g., analytical methods involving AES with ICAP scan, for concentration of boron). Test concentrations of reference toxicant in soil are prepared by adding a measured quantity of the *stock solution* to negative control soil[59], and mixing thoroughly.[60] Upon preparation of the mixtures of the reference toxicant in soil, aliquots should be taken from at least the negative control soil as well as the low, middle, and high concentrations.[61] Each aliquot should either be analyzed directly, or stored for future analysis (i.e., at the end of the test) if the ICp for shoot length, based on nominal concentrations, was found to be outside the warning limits. If stored, sample aliquots must be held in the dark at 4 ± 2 °C. Stored aliquots requiring chemical measurement should be analyzed promptly upon completion of the reference toxicity test. The 7-day or 10-day ICp for shoot length should be calculated based on the measured concentrations if they are appreciably (i.e., ≥20%) different from nominal ones and if the accuracy of the chemical analyses is satisfactory.

If boric acid is used as a *reference toxicant*, the following analytical method applies (OMEE, 1996). A 1–5 g subsample of soil spiked with boric acid is dried at 105 °C to constant weight. A 1-g aliquot is then extracted using an 0.01 *M* solution of $CaCl_2$, by boiling a slurry of soil in 50 mL of this extraction solution and then re-adjusting the final volume to 50 mL using more extraction solution. The 50-mL extract is then filtered through a #4 Whatman™ filter, and diluted to a final volume of 100 mL. A

---

[59] Section 6.2 "*Preparing Test Mixtures*" includes an example showing the amounts of de-ionized or distilled water and boric acid to be added to artificial soil, to prepare a given treatment for a reference toxicity test with a specific concentration of boric acid in artificial soil. The calculations in this example show the amount of water necessary to adjust the moisture content of the artificial soil to a fixed percentage (i.e., 70%) of the soil's water-holding capacity, while taking into account the volume of the *stock solution* of boric acid as part of the overall adjustment for soil moisture content.

[60] An accepted procedure is to add a precalculated volume of *stock solution* (using volumetric and/or graduated pipets) to a glass Erlenmeyer™ flask, diluting to a graduated mark using de-ionized water, and then adding a measured volume to the soil. The flask is then rinsed three times with de-ionized water, and the rinsate is added to the soil. The mixture of soil and stock solution is then mixed thoroughly (for approximately three minutes) with a mechanical mixer (e.g., a hand-held mixer with revolving stainless steel beaters) until the soil appears homogeneous in colour, texture, and moisture content. During the mixing process, the soil in the mixing bowl should also be stirred intermittently using a large stainless steel spoon to facilitate homogenization.

[61] If the ICp for each reference toxicity test is to be based on measured concentrations, it is recommended that one or more aliquots of the chemical-in-soil mixture representing each test concentration be collected and analyzed. If, however, the endpoints for each test are based on nominal concentrations, sampling and analysis of aliquots from at least the low, middle, and high test concentrations is recommended.

blank sample is prepared in a similar manner. The filtrate is analyzed for elemental boron using ICAP/AES. The boric acid concentration in the soil is then calculated using the following equation:

$$\text{boric acid (mg/kg, dry wt)} = \frac{\mu\text{g B/mL (measured)} \times \text{final volume (mL)} \times MW_{\text{boric acid}}/MW_{\text{boron}}}{1000\ (\mu\text{g}) \times \text{weight of sample (mg dry wt)}} \times 10^6$$

The analytical limit of detection for boric acid in soil is reportedly 1 mg boric acid/kg soil dry wt in most instances (Stephenson, 2003b).

Specific Procedures for Testing Field-Collected Soil or Similar Particulate Material

This section provides specific instructions for preparing and testing samples of field-collected (site) soil or similar particulate material, in addition to the procedures discussed in Section 4.

Detailed guidance for the collection, handling, transport, storage, and analyses of field-collected soil is given in a number of reports specific to these subjects (e.g., van Ee *et al.*, 1990; Webster and Oliver, 1990; USEPA, 1991; Keith, 1992; Klute, 1986; Carter, 1993; OMAFRA, 1999). In the absence of guidance specific to these subjects from Environment Canada, such reports should be consulted and followed (in addition to the guidance provided here), when collecting samples of field-collected soil and preparing them for toxicity tests with terrestrial plants using the biological test method described herein.

5.1 Sample Collection

Crépin and Johnson (1993) provide a useful summary of field-sampling design and appropriate techniques for sample collection. Field surveys of soil toxicity using biological tests with terrestrial plants and/or other suitable, soil associated test organisms (e.g., EC, 2004c, 2005c) are frequently part of more comprehensive surveys (e.g., Callahan *et al*., 1991; Menzie *et al*., 1992; and Saterbak *et al*., 2000). Such surveys could include a *test battery* to evaluate the toxicity of soil together with tests for bioaccumulation of contaminants, chemical analyses, biological surveys of epifaunal and/or infaunal organisms, and perhaps the compilation of geological and hydrographic data. Statistical correlation can be improved and costs reduced if the samples are taken concurrently for these tests, analyses, and data acquisitions.

Samples of soil to be used in the biological test method described herein (Section 4), might be taken quarterly, semiannually, or annually from a number of contaminated or potentially contaminated sites for *monitoring* and *compliance* purposes. Samples of soil might also be collected on one or more occasions during field surveys of sites for spatial (i.e., horizontal or vertical) or temporal definition of soil quality. One or more sites should be sampled for *reference* (presumably *clean*) *soil* during each field collection[62].

The number of stations to be sampled at a study site and the number of *replicate samples* per station will be specific to each study. This will involve, in most cases, a compromise between logistical and practical constraints (e.g., time and cost) and statistical considerations. Webster and Oliver (1990), Crépin and Johnson (1993) and OMAFRA (1999) provide guidance on the sampling design; van Ee *et al.* (1990) and USEPA (1991) address issues related to *quality assurance* and *quality control*.

For certain *monitoring* and regulatory purposes, multiple replicates (i.e., separate samples from different grabs or cores taken at the same site) should be taken at each *sampling station*, including one or more reference stations. Each of these field replicates should be tested for its toxicity to terrestrial plants using five or more test vessels per replicate sample (Section 4.1). The use of power analysis (see Section 5.5.2) with endpoint data obtained in previous tests of the same type, performed with previous samples from the same or similar sites, will assist in determining if additional laboratory replicates need to be tested with each field replicate. Also, some of the statistical tests have requirements for a minimum number of replicates. For certain other purposes (e.g., preliminary or extensive surveys of the spatial

[62] Ideally, a *reference soil* is collected near the site(s) of concern. It possesses geochemical characteristics (e.g., *texture*, organic carbon content, *organic matter* content, pH) similar to those of the field-collected *test soil*(s) but without anthropogenic contaminants. It is not unusual for nearby reference sites to have some degree of contamination from anthropogenic chemicals, and in some instances, reference soil might be toxic or otherwise unacceptable for use in a soil toxicity test, because of naturally occurring physical, chemical, or biological properties.

distribution of toxicity), the survey design might include only one sample from each station, in which case the sample would normally be homogenized and split between five replicate test vessels. The latter approach precludes any determination of mean toxicity at a given sampling location (station), and completely prevents any conclusion on whether a station is different from the control or reference, or from another location. It does, however, allow a statistical comparison of the toxicity of that particular sample with the reference or control or with one or more samples from other locations. It is important to realize that any conclusion(s) about differences, which arise from testing single field samples lacking replication, cannot be extended to make any conclusion(s) about the sampling locations.

Sites for collecting *reference soil* should be sought where the geochemical properties of the soil are similar to soil characteristics encountered at the test sites. Matching of *total organic carbon* content (%) or organic matter content (%) might not be warranted in cases where *pollution* (e.g., from or within sewage or industrial sludge) is responsible for the high organic carbon content of test soils. Preliminary surveys to assess the toxicity and geochemical properties of soil within the region(s) of concern and at neighbouring sites are useful for selecting appropriate sites at which to collect reference soil.

Samples of municipal or industrial sludge (e.g., sewage sludge, dewatered mine tailings, or biosolids from an industrial clarifier or settling pond) might be collected for the assessment of their *toxic* effect(s) on plants, and for geochemical and contaminant analyses. Other particulate wastes being considered for land disposal might also be collected for toxicity and physicochemical evaluation.

Guidance for various soil sampling plans and procedures is available in the technical literature (e.g., Petersen and Calvin, 1986; Keith, 1992; Crépin and Johnson, 1993). Procedures used for sample collection (i.e., core, grab, or composite) will depend on the study objectives and the nature of the soil or other particulate material being collected. A shovel, auger, or soil corer (preferably stainless steel) is frequently used for collecting soil samples. The surface of the location where each sample is to be collected should be cleared of debris such as twigs, leaves, stones, thatch, and litter. If the location is an area of grass or other herbaceous plant material, the plants should be cut to ground level and removed before the sample is collected. Removal of the vegetation should be done such that removal of soil particles with the roots is minimal. Dense root masses (e.g., grasses) should be removed and then shaken vigorously to remove soil particles adhering to the roots. The soil sample to be collected for toxicity and evaluation of chemistry should be taken from one or more depths that represent the layer(s) of concern (e.g., a surficial layer of soil, or one or more deeper layers of soil or subsoil if there is concern about historical deposition of contaminants).

The required volume of soil per sample should be calculated, before a sampling program is initiated. This calculation should take into account the quantity of soil required to prepare laboratory replicates for soil toxicity tests, as well as that required for particle size characterization, *total organic carbon* content(%), *organic matter* content (%), *moisture content* (%), and specific chemical analyses. A volume of at least 5–7 L of soil per sample is normally required, although this will depend on the study objectives/design (e.g., single-concentration or multi-concentration test) and the nature of the chemical analyses to be performed, and possibly also on the nature of the soil (e.g., need for removal of excess water and/or debris in the laboratory, which can reduce the sample volume). To obtain the required sample volume, it is frequently necessary to combine subsamples retrieved using the sampling device. The same collection procedure should be used at all field sites sampled.

5.2 Sample Labelling, Transport, Storage, and Analyses

Containers for transport and storage of samples of field-collected soil or similar particulate material must be made of nontoxic material. The choice of container for transporting and storing samples depends on both sample volume and the potential end uses of the sample. The containers must either be new, thoroughly cleaned, or lined with high-quality plastic. Thick (e.g., 4 mil) plastic bags are

routinely used for sample transport and storage. If plastic bags are used, it is recommended that each be placed into a second clean, opaque sample container (e.g., a cooler or a plastic pail with a lid) to prevent tearing and support the weight of the sample and to maintain darkened conditions during sample transport (ASTM, 1999b). Plastic containers or liners should not be used if there are concerns about the plastic affecting the characteristics of the soil (e.g., compounds from plastic leaching into the soil).

Following sample addition, the air space in each container used for sample transport and storage should be minimized (e.g., by collapsing and taping a filled or partially filled plastic bag). Immediately after filling, each sample container must be sealed, and labelled or coded. Labelling and accompanying records made at this time must include at least a code or description that identifies sample type (e.g., grab, core, composite), source, precise location, land use information, replicate number, and date of collection; and should include the name and signature of sampler(s). Persons collecting samples of soil should also keep records that describe details of:

- the nature, appearance, and volume of each sample;
- the sampling procedure and apparatus;
- any procedure used to composite or subsample grabs or cores in the field;
- the number of replicate samples taken at each *sampling station*;
- the sampling schedule;
- the types and numbers of containers used for transporting samples;
- any field measurements (e.g., temperature, pH, soil moisture content) of the soil at the collection site;
- procedures and conditions for cooling and transporting the samples;
- observations of environmental conditions at the time of sampling (e.g., raining); and
- observations of soil fauna and vegetation at the collection site.

Soil samples should not freeze or become overheated during transport or storage. It is recommended that samples be kept in darkness (i.e., held in light-tight, opaque transfer containers such as coolers or plastic pails with lids) during transport, especially if they might contain PAHs or other chemicals or chemical products that could be photoactivated or otherwise altered due to exposure to sunlight. As necessary, gel packs, regular ice, or other means of refrigeration should be used to assure that the temperature of the sample(s) remains cool (e.g., 7 ± 3 °C) during transit.

The date the sample(s) is received at the laboratory must be recorded. Sample temperature upon receipt at the laboratory should also be measured and recorded. Samples to be stored for future use must be held in airtight containers. If volatile contaminants are in the soil or of particular concern, any air "headspace" in the storage container should be purged with nitrogen gas, before capping tightly. Samples must not freeze or partially freeze during transport or storage (unless they are frozen when collected), and must not be allowed to dehydrate. If, however, one or more samples are saturated with excess water upon arrival at the laboratory (e.g., sampling occurred during a significant rainfall event), the sample(s) may be transferred to plastic sheeting for a brief period (e.g., one or more hours) to enable the excess water to run off or evaporate. Thereafter, the sample(s) should be returned to the transport container(s) or transferred to one or more airtight containers for storage.

It is recommended that samples be stored in darkness at 4 ± 2 °C. These storage conditions must be applied in instances where PAHs or other light-sensitive contaminants are present, or if the samples are known to contain unstable volatiles of concern. It is also recommended that samples of soil or similar particulate material be tested as soon as possible after collection. The soil toxicity test(s) should begin within two weeks of sampling, and preferably within one week. The test must begin within six weeks, unless it is known that the soil contaminants are aged and/or weathered and therefore considered stable.

Dry sieving (i.e., press sieving; not wet sieving) of samples through a coarse-mesh sieve is desirable to remove large particles (see Section 5.3). This procedure may be performed in the field. Undesirable coarse material (e.g., large gravel or stones, large debris, large indigenous macroinvertebrates, or large plant material) may also be removed in the field before sample transport.

In the laboratory, each sample of field-collected soil should be thoroughly mixed (Section 5.3), and representative subsamples taken for physicochemical characterization. Each sample (including all samples of *negative control soil* and *reference soil*) must be characterized by analyzing subsamples for at least the following:

- particle size distribution (% sand, % silt, and % clay),
- *total organic carbon* content (%),
- *organic matter* content (%),
- *moisture content* (%),
- water-holding capacity (% based on dry wt of soil),
- total nitrogen,
- total phosphorus,
- pH, and
- conductivity.

Additionally, the following analyses should be performed:

- *texture*,
- *fertility*,
- C:N ratio,
- cation exchange capacity,
- major cations,
- organophosphorous insecticides,
- organochlorine insecticides, and
- a suite of herbicides.

Other analyses could include:

- bulk density,
- total inorganic carbon,
- total volatile solids,
- biochemical oxygen demand,
- chemical oxygen demand,
- oxidation-reduction potential,
- metals, and
- petroleum hydrocarbons (including PAHs).

Unless indicated otherwise, identical chemical, physical, and toxicological analyses should be performed with subsamples representative of each replicate sample of field-collected soil (including *reference soil*) taken for a particular survey of soil quality, together with one or more subsamples of *negative control soil*.

5.3 Preparing Sample for Testing

Field-collected soil or similar particulate waste material must not be sieved with water, as this would remove contaminants present in the interstitial water or loosely sorbed to particulate material. Large gravel or stones, debris, indigenous macroinvertebrates, or plant material should normally be removed using forceps or a gloved hand. If a sample contains a large quantity of debris (e.g., plant material, wood chips, glass, plastic, large gravel) or large macroinvertebrates, these may be removed by pressing the soil through a coarse sieve (e.g., mesh size of 4–9 mm; EC, 2000).

Qualitative descriptions of each sample of field-collected *test soil* should be made and recorded at the testing laboratory, including information on sample colour, texture, and the presence and description of roots, leaves, and macroscopic soil organisms. Unless research or special study objectives dictate otherwise, each sample of field-collected test material should be homogenized in the laboratory before use (USEPA, 1989).[63] Mixing can affect the concentration and bioavailability of contaminants in the soil, and sample homogenization might not be desirable for all purposes.

As indicated in Section 3.7, one or more samples of field-collected *test soil* might either be tested at a single concentration only (typically, 100%), or evaluated for toxicity in a multi-concentration test whereby a series of concentrations are prepared by mixing measured quantities with either negative control soil or reference soil. When performing a multi-concentration test, the following series of concentrations of test soil (mixed in negative control soil or reference soil), which spans the range of 100–1% test soil using nine concentrations, might prove suitable: 100%, 80%, 65%, 50%, 30%, 15%, 7.5%, 3%, 1%, and 0%. Guidance on other concentration series that might prove as or more

[63] One of the reasons for routinely homogenizing samples is to mix into the soil, any pore water which rises to the surface during sample shipment and storage. Homogenization is also necessary to redistribute the sample constituents that have compacted and layered according to particle size during transport and storage.

suitable is found in Section 6.2, along with that for preparing test mixtures which might apply equally when performing a multi-concentration test with one or more samples of field-collected soil. Refer to Section 4.1, for additional guidance when selecting test concentrations. In each instance, the test must include a treatment comprised solely of negative control soil (see Section 3.4).

To achieve a homogeneous sample, transfer it to a clean, rigid mixing container (e.g., a large stainless steel or plastic bowl) or for larger volumes of soil, to clean plastic sheets, spread out on the floor. The sample should be mixed manually (using a gloved hand or a nontoxic device such as a stainless steel spoon) or mechanically (e.g., using a domestic hand-held mixer with beaters at low speed, or a hand-held wire egg beater) until its texture and colour are homogeneous. While mixing, care should be taken to ensure that the impact of mixing on soil structure is minimal and that the structure is not destroyed entirely. As soon as the texture and colour of the sample appears to be homogeneous, mixing should be discontinued.

For each sample included in a test, mixing conditions including duration and temperature must be as similar as possible. If there is concern about the effectiveness of sample mixing, subsamples of the soil should be taken after mixing, and analyzed separately to determine the homogeneity of particle sizes, chemical(s) of interest, etc. Any moisture that separates from a sample during its transport and/or storage must be remixed into it, if possible.

The *moisture content* of a given sample of field-collected *test soil* should be standardized during its preparation by determining its *water-holding capacity* (WHC) and then hydrating the soil to an optimal moisture content based on a percentage of this value. The optimal percentage of the WHC for each sample of field-collected soil must be determined before sample preparation and test initiation. In order to do so, the moisture content of each homogenized sample (i.e., each sample of test soil, including the negative control soil) must be determined (Sections 4.1 and 4.6). Thereafter, the WHC of each sample must be determined using a recognized standard procedure (see following three paragraphs). A subsample of each soil sample is then hydrated to a homogeneous, crumbly consistency with clumps approximately 3–5 mm in diameter.[64] Based on the initial moisture content of the sample, the WHC of the sample, and the amount of water added to achieve the desired soil consistency, the sample's optimal moisture content can be calculated and expressed as a percentage of the WHC for each soil. Once this target (or optimal) percentage of the WHC has been determined, the moisture content of each sample of test soil (including the negative control soil) can be standardized to the selected (sample-specific) moisture content. *Test water* (i.e., de-ionized or distilled water) should be added to each sample with a moisture content that is less than the pre-determined optimal percentage of its WHC, until this moisture content is achieved [65] (Aquaterra Environmental, 1998a). If a sample is too wet, it should be spread as a thin layer on a clean sheet of

---

[64] An unpublished study, carried out by Environment Canada (J. Princz, personal communication, Biological Methods Division, Environmental Technology Centre, Ottawa, ON, 2004), determined the optimal moisture content for each of the diverse types of soil used while developing the biological test method described herein (see Section 3.4 and Appendix G), based on a percentage of each sample's WHC. The optimal percentage of the WHC of these soils ranged from approximately 45–50% for the silt and sandy loam soils to 60% for the clay loam soil. These values were considered optimal since, at these levels of saturation, the soil mixed well, and formed an acceptable structure (i.e., the resulting macro-aggregation of soil particles). Experience indicates that the actual moisture content of the test soils hydrated to optimal conditions can vary greatly (e.g., 20% for sandy loam soil to 50% for clay loam soil), depending on the bulk density and the WHC of the sample(s) of field-collected soil being tested (ESG and Aquaterra Environmental, 2002; Becker-van Slooten *et al*., 2003).

[65] An alternate approach sometimes used by certain investigators is to standardize (and adjust) the moisture content of each sample of field-collected soil to a fixed concentration, such as 35–45% of its dry weight (ASTM, 1999b; EC, 2000). However, a disadvantage of this approach is that certain samples of field-collected soil can appear to be very wet and have standing water on the surface after hydration to only 35–45% of their dry weight; whereas other *site soils* can appear considerably dryer after the same level of hydration (ASTM, 1999b; EC, 2000). Accordingly, the use of this alternate approach is not recommended here.

plastic (e.g., a new plastic garbage bag) or a clean, non-reactive (e.g., stainless steel or plastic) tray, and allowed to dry by evaporation at ambient (~20 °C) room temperature. Rehydration to the pre-determined optimal percentage of its WHC might be necessary. Upon adjustment of a sample's moisture content to the desired percentage of its WHC, the moisture content (%) of the hydrated soil must be determined and the percent WHC and percent moisture content recorded and reported.

The WHC (and the percent WHC that is optimal for biological testing) of a particular soil is generally unique to each soil type, and is ultimately the result of the interaction of many variables associated with soil structure (e.g., micro/macro-aggregation, pore space, bulk density, *texture*, *organic matter* content). There are a number of methods that can be used to determine WHC; however, most of these methods require measurements to be made on an intact soil sample (e.g., soil core) where characteristics (structural aggregations, pore space, bulk density, *texture*, and *organic matter* content) are preserved during collection. The USEPA (1989) has described an appropriate method for toxicity testing using unconsolidated materials (such as samples of field-collected soils that have been dried, sieved, and homogenized; or samples of soil formulated in the laboratory from constituents).[66] This method is outlined here.

For this method, ~130 g (wet wt) of sample is placed into an aluminum pan or petri dish (15 × 1 cm), and dried at 105 °C until a constant weight is achieved (this usually takes a minimum of 24 h). Thereafter, 100 g of the oven-dried soil is placed into a 250-mL glass beaker with 100 mL of distilled or de-ionized water. The resulting slurry is mixed thoroughly with a glass stir rod. A folded filter paper (185-mm diameter Fisherbrand P8 coarse porosity, qualitative creped filter paper; catalogue no. 09-790-12G) is placed into a glass funnel (with a top inside diameter of 100 cm and a stem length of 95 cm). The folded filter paper should be level with the top of the glass funnel. Using a pipette, up to 9 mL of distilled or de-ionized water is slowly added to the filter paper to wet the entire surface. The funnel and hydrated filter paper are then weighed. To obtain the initial weight for the mass of the funnel plus hydrated filter paper plus dried soil (see "I" in Equation 1), the weight of the dried soil (100 g) is added to the weight of the funnel and the wet filter paper.

The funnel is then placed into a 500-mL Erlenmeyer flask and the soil slurry is slowly poured onto the hydrated filter paper held in the funnel. Any soil remaining on the beaker and stir rod is rinsed into the funnel with the least amount of water necessary to ensure that all of the solid material has been washed onto the filter. The funnel is then tightly covered with aluminum foil and allowed to drain for three hours at room temperature. After three hours, the funnel containing the hydrated filter paper and wet soil is weighed. This weighing represents the final weight for the mass of the funnel plus hydrated filter paper plus (wet) soil (see "F" in Equation 1).

The water-holding capacity for the subsample of soil in the funnel, expressed as percentage of soil dry mass, is then calculated using the following equation:

$$WHC = \frac{F - I}{D} \times 100 \quad \text{(Equation 1)}$$

where:

WHC = water-holding capacity (%)
F = mass of funnel + hydrated filter paper + wet mass of soil
I = mass of funnel + hydrated filter paper + dry mass of soil
D = 100 g (i.e., dry mass of soil)

[66] Certain participants at a soil toxicity testing workshop sponsored by EC in Vancouver, BC (February 2003) considered the determination of WHC and a percentage of that capacity to be the most appropriate way of expressing soil moisture content (EC, 2004b). This led to a testing program to compare two different methods for estimating the WHC of soil (i.e., as per Annex C in ISO, 1999 or according to USEPA, 1989) as well as a somewhat different method for expressing soil moisture content, as a percentage of the soil's water-filled pore space (WFPS). The results of this investigation showed that each method had distinct advantages and disadvantages; however, the USEPA (1989) method for measuring WHC was recommended for use in EC's soil toxicity test methods when adjusting (if and as necessary) the moisture content of soil samples (Becker-van Slooten *et al*., 2004).

The WHC of each sample of *test soil* should be determined in triplicate, using three subsamples.

The percentage of water (i.e., P_w) that is added to a sample of field-collected soil to achieve the desired hydration (i.e., the optimal percentage of the WHC) can be calculated as follows:[67]

$MC = [(W - D) / D] \times 100$ [Equation 1]
$P_W = [WHC \times (P_{WHC} / 100)] - MC$ [Equation 2]
$V_W = (P_W \times M) / 100$ [Equation 3]
$M_W = (M_D \times W) / D$

W = wet mass of substrate (g)
D = dry mass of substrate (g)
WHC = water-holding capacity (% of dry mass)
P_{WHC} = percentage of WHC desired (%)
MC = initial moisture content of substrate (%)
P_W = percentage of water to add to soil (%)
M_D = total mass of soil required for experiment (expressed as dry wt)
V_W = volume of water to add to soil (mL)
M_W = total mass of soil required for experiment (expressed as wet wt based on initial MC)

Calculations for a 25% concentration of a contaminated soil in negative control soil:

For a definitive plant test using this example, it is assumed that a total mass of 1225.00 g dry weight (wt) of soil is sufficient to satisfy the requirement for each treatment (i.e., 400.00 g dry wt per replicate × 3 replicates + 25.00 g dry wt extra soil for pH, etc.). To simplify the calculations, this example assumes that 400 g (dry wt) of either type of soil is sufficient to provide the 500-mL aliquot of soil to be added to each of three replicate test vessels per treatment (see Section 4.1).

For a 25% concentration of contaminated soil in negative control soil, 25% of the total mass of soil, on a dry-wt basis, must consist of the contaminated soil:

= 1225.00 g dry wt × (25/100)
= 306.25 g dry wt of contaminated soil

The remainder of the test soil required to prepare this treatment (i.e., 75 %) will consist of the negative control soil:
= 1225.00 g dry wt × (75/100) [or 1225.00 g dry wt - 306.25 g dry wt]
= 918.75 g dry wt of negative control soil

Therefore, the final total mass of soil required, based on wet weight, is 1361.37 g [1150.79 g wet wt at the soil's initial moisture content (i.e., M_{Wnc}) + 210.58 mL of water] for the negative control soil, and 388.44 g [384.76 g wet wt at the soil's initial moisture content (i.e., M_{Wc}) + 3.68 mL of water] for the contaminated soil.

The final moisture content for each soil would be 48.18 % {[(1361.37 – 918.75)/918.75] × 100} for the negative control soil, and 26.84 % {[(388.44 – 306.25)/306.25] × 100} for the contaminated soil.

---

[67] The following example provides calculations that pertain to the hydration of samples of a contaminated field-collected soil and a negative control soil, when preparing a test concentration of 25% for use in a definitive test with plants involving three replicates per treatment.

Assumptions:

Soil #1: Negative Control (nc) Soil

W_{nc} = 2.3934 g
D_{nc} = 1.9108 g
WHC_{nc} = 80.30 %
P_{WHCnc} = 60.00 %
MC_{nc} = 25.26%
P_{Wnc} = 22.92 %
M_{Dnc} = 918.75 g dry wt
V_{Wnc} = 210.58 mL
M_{Wnc} = 1150.79 g wet wt

Soil #2: Contaminated (c) Soil

W_c = 7.0575 g
D_c = 5.6174 g
WHC_c = 67.10 %
P_{WHCc} = 40.00 %
MC_c = 25.64 %
P_{Wc} = 1.20 %
M_{Dc} = 306.25 g dry wt
V_{Wc} = 3.68 mL
M_{Wc} = 384.76 g wet wt

$P_W = [WHC \times (P_{WHC}/100)]- MC$ (Equation 2)

where:

P_W = percentage of water to add to the soil (%)
WHC = water-holding capacity (%)
MC_i = initial moisture content of the soil

The volume of water (i.e., V_w) that should be added to a sample of field-collected soil to achieve the desired hydration (i.e., the optimal percentage of the sample's water-holding capacity) can be calculated as follows:[67]

$V_W = (P_W \times M)/100$ (Equation 3)

where:

V_W = volume of water to add to the soil (mL)
P_W = percentage of water to add to the soil (%)
M = total mass of soil required for test (expressed as dry wt)[68]

The final moisture content of the negative control soil (i.e., 48.18%) represents 60% of that soil's water-holding capacity (48.18 ÷ 80.30 = 0.60). The final moisture content of the contaminated soil (i.e., 26.84%) represents 40% of that soil's water-holding capacity (26.84 ÷ 67.10 = 0.40).

[68] For tests with samples of field-collected soil, the amount of soil added to each test vessel is based on the wet weight of that soil that is equivalent to a volume of ~500 mL (see Section 4.1). However, "M" (i.e., the total mass of soil required for the test) is expressed as dry weight in the formula used to calculate the volume of water to be added to a sample of field-collected soil to achieve the desired hydration (see Equation 3). To calculate the amount of soil required per test vessel (by dry wt), a subsample of "wet" soil is placed into a test vessel (e.g., 1-L polypropylene cup), to determine the correct volume of soil required on a wet-weight basis. For example, assume that (for a given sample) this volume is equivalent to 500 g wet wt and that the wet and dry weights of a subsample of this soil, previously determined for the purpose of calculating the sample's water-holding capacity, are 4.1507 g and 2.7813 g, respectively. The dry weight equivalent to a 500-mL volume of this sample (which has a wet weight of 500 g) can be calculated as follows:

(500 g × 2.7813 g) ÷ 4.1507 g = 335 g

Except for research-oriented toxicity tests intended to determine the influence of pH on sample toxicity, the pH of samples of field-collected soil must not be adjusted. Studies intending to investigate the effect of pH on sample toxicity should conduct two side-by-side tests, whereby one or more sets of treatments is adjusted to a fixed pH value using calcium carbonate or a suitable acid or base, and the pH of one or more duplicate sets of treatments is not adjusted.

Immediately following sample hydration (or dehydration) and mixing, subsamples of test material required for the toxicity test and for physicochemical analyses must be removed and placed into labelled test vessels (see Section 4.1), and into the labelled containers required for the storage of subsamples for subsequent physicochemical analyses. Any remaining portions of the homogenized sample that might be required for additional toxicity tests using plants or other test organisms (e.g., according to EC, 2004c and EC, 2005c) should also be transferred to labelled containers at this time. All subsamples to be stored should be held in sealed containers with minimal air space, and must be stored in darkness at 4 ± 2 °C (Section 5.2) until used or analyzed. Just before it is analyzed or used in the toxicity test, each subsample must be thoroughly remixed to ensure that it is homogeneous.

5.4 Test Observations and Measurements

A qualitative description of each field-collected test material should be made at the time the test is set up. This might include observations of sample colour, texture, and homogeneity, and the presence of plants or macroinvertebrates. Any changes in the appearance of the test material observed during the test or upon its termination, should be noted and reported.

This mass of soil can be rounded up to 350 g dry weight, thereby providing a little extra soil, if necessary. Therefore, for the example provided here, the mass of this sample of soil required for each replicate (expressed as dry wt) is 350 g. The total mass ("M") can then be calculated simply by multiplying the dry mass required for each replicate (in this instance, 350 g dry wt) by the number of replicates to be used in the test (i.e., for this example, three replicates).

Section 4.6 provides guidance and requirements for the observations and measurements to be made during or at the end of each test. These observations and measurements apply and must be made when performing the soil toxicity test described herein using one or more samples of field-collected (site) soil.

Depending on the test objectives and experimental design, additional test vessels might be set up at the beginning of the test (Section 4.1) to monitor soil chemistry. These could be destructively sampled during and at the end of the test. Test organisms might or might not be added to these extra test vessels, depending on the study's objectives. Measurements of chemical concentrations in the soil within these vessels can be made by removing aliquots of the soil for the appropriate analyses (see Section 5.2).

5.5 Test Endpoints and Calculations

The common theme for interpreting the results of tests with one or more samples of field-collected *test soil*, is a comparison of the biological effects for the test (site) soil(s) with the effects found in a *reference soil*. The reference sample should be used for comparative purposes whenever possible or appropriate, because this provides a site-specific evaluation of toxicity (EC, 1997a, b, 2001, 2004c). Sometimes the reference soil might be unsuitable for comparison because of toxicity or atypical physicochemical characteristics. In such cases, it would be necessary to compare the *test soils* with the *negative control soil*. Results for the *negative control soil* will assist in distinguishing contaminant effects from noncontaminant effects caused by soil physicochemical properties such as particle size, *total organic carbon* content (%), and *organic matter* content (%). Regardless of whether the *reference soil* or *negative control soil* is used for the statistical comparisons, the results from *negative control soil* must be used to judge the validity and acceptability of the test (see Section 4.4).

Analyses of the results will differ according to the purposes and particular designs of the test. This section covers the analytical procedures, starting with the simplest design and proceeding to the more complex designs. Standard statistical procedures are generally all that is needed for analyzing the results. Investigators should consult EC (2004a) for guidance on the appropriate statistical endpoints and their calculation. As always, the advice of a statistician familiar with *toxicology* should be sought for the design and analyses of tests.

Analysis of variance (ANOVA) involving multiple comparisons of endpoint data derived for single-concentration tests involving field replicates of field-collected soil from more than one sampling location is commonly used for statistical interpretation of the significance of findings from soil toxicity tests. This hypothesis-testing approach is subject to appreciable weaknesses. Notably, any increased variability within the test will weaken its power to distinguish *toxic* effects (i.e., less toxicity is concluded). Similarly, use of only a few replicates instead of many replicates will weaken the discrimination of a test and will lead to a conclusion of less apparent toxicity, other things being equal (see Section 5.5.2). There is no alternative to hypothesis testing, when comparing toxicity data for multiple samples of field-collected soil (i.e., field replicates of soil from more than one sampling location) that use only one concentration (usually full strength, i.e., 100% sample). There are alternatives for comparing point estimates of toxicity if various concentrations of each sample of field-collected soil are tested and multiple endpoint values for ICp or EC50 are determined (see Section 6.4). Section 9 in EC (2004a) should be consulted for guidance when comparing multiple ICps or multiple EC50s.

The parametric analyses involving ANOVA for comparative data from single-concentration tests with multiple samples of field-collected soil (i.e., field replicates of soil from more than one sampling location) assume that the data are normally distributed, that the treatments are independent, and that the variance is homogeneous among the different treatments. As the first step in analysis, these assumptions should be tested using the *Shapiro-Wilk's Test* for normality and *Bartlett's Test for Homogeneity of Variance* (Eisenhart *et al.*, 1947; Sokal and Rohlf, 1969). If the data satisfy these assumptions, analysis may proceed. If not, data could be transformed (e.g., as square roots, logarithms, or as arcsine square root for *quantal* data which are to be used in statistical analysis; Mearns *et al.*, 1986). The tests for normality and

homogeneity might then show conformance with normality and homogeneity, and in fact that is a likely outcome of a transformation. Assumptions should be re-tested following any transformation of data. Parametric tests are reasonably robust in the face of moderate deviations from normality and equality of variance; therefore, parametric analysis (e.g., ANOVA and multiple comparison) should proceed, even if moderate nonconformity continues after transformation. Excluding a data set for minor irregularities might lose a satisfactory and sensitive analysis and forgo the detection of real effects of toxicity.[69] Analysis by nonparametric statistical procedures should also proceed in parallel, with the more sensitive (lower endpoint) of the two analyses providing the final estimates of toxicity. Section 3 in EC (2004a) should be consulted for guidance when comparing the findings for single-concentration tests involving field replicates of samples from multiple locations, using parametric or non-parametric tests.

Guidance in Section 6 (including that in Section 6.2 for performing range-finding tests, and that in Section 4.8 for calculating test endpoints) should be followed if a multi-concentration test is performed using one or more samples of field-collected soil diluted with *negative control soil* or *clean* reference soil. Section 9 in EC (2004a) should be consulted when comparing such point estimates of toxicity for multiple samples of field-collected soil.

5.5.1 Variations in Design and Analysis

A very preliminary survey might have only one sample of *test soil* (i.e., contaminated or potentially contaminated *site soil*) and one sample of *reference soil*, without replication. Simple inspection of the results might provide guidance for designing more extensive studies.

If there is a single test sample and a single reference sample, with equal replication for each, a standard *Student's t-test* would be suitable for analysis (Paine and McPherson, 1991; EC, 1997a, b, 2001). The t-test is fairly robust and handles unequal numbers of replicates in the test and reference samples, as well as moderately unequal variances in the two groups (Newman, 1995; USEPA, 1995).

A preliminary evaluation might conceivably be conducted with samples from many stations, but without either field replicates or laboratory (within-sample) replicates. The objective might be to identify a reduced number of *sampling stations* deserving of more detailed and further study. Opportunities for statistical analysis would be limited. The nonreplicated test data could be compared with the reference data using outlier detection methods (USEPA, 1994a; Newman, 1995; EC, 1997a, b, 2001, 2004a, c). A sample would be considered *toxic* if its result was rejected as an extreme value when considered as part of the data for the *reference soil* and/or the *negative control soil*.

A more usual survey of soils would involve the collection of replicate samples from several places by the same procedures, and their comparison with replicate samples of a single *reference soil* and/or *negative control soil*. There are several pathways for analysis, depending on the type and quality of data, but often there would be an analysis of variance (ANOVA) followed by one of the multiple-comparison tests. In the ANOVA, the *reference soil* would also be treated as that from a "location".

In these multi-location surveys, the type of replication would influence the interpretation of results. If field replicates were collected at each of the sampling locations, and no laboratory replicates were used, a one-way ANOVA would evaluate the overall difference in test results with respect to sampling location, over and above the combined variability of sampling the location and running the test. It would be unusual but much more powerful, to have field replicates for all sampling locations and also laboratory replicates of each field replicate. If that were done, the laboratory replicates would

---

[69] Tests for normality and homogeneity become less meaningful with the small number of samples from individual *sampling stations* typically collected in studies of *environmental toxicology*. Plotting and examining the general nature of the distribution of toxicity and its apparent deviations can be more revealing and is recommended (EC, 2004a). Equality in sample sizes and the magnitude of variation are probably more important factors for the outcome of parametric analysis, but they have received scant attention in toxicology. The robustness of ANOVA is shown by its ability to produce realistic probabilities if the distribution of data is reasonably symmetrical, and if treatment variances are within threefold of each other (Newman, 1995).

become the replicates in a nested one-way ANOVA, and would be the base of variability for comparing differences in the samples. The ANOVA could be used to determine (a) if there was an overall difference in test results for samples with respect to their sampling location, and (b) whether there was an overall difference in replicates taken at the various locations. After an ANOVA, the analysis would proceed to one or more types of multiple-comparison test, as described in the following text.

If only laboratory replicates and no field replicates were tested, there could be no conclusions about differences due to sampling location (see also Section 5.1). The laboratory replicates would only show any differences in the samples that were greater than the baseline variability in the within-laboratory procedures for setting up and running the test. Sample variability due to location would not really be assessed in the statistical analysis, except that it would contribute to any difference in test results associated with sampling location.

If it were desired to compare the test results for the replicate samples from each sampling location with those for the reference soil, to see if the toxicity of the two sources of soil (locations) differed, *Dunnett's test* should be used. It assumes normality and equal variance, and is based on an experiment-wise value of α (the probability of declaring a significant difference when none actually exists). If replication was unequal, investigators could use the *Dunn-Sidak modification of the t-test*, or alternatively the *Bonferroni adjustment of the t-test* (p. 189 in Newman, 1995; Appendix D in USEPA, 1995; Section 7.5.1 in EC, 2004a).

In a multi-location survey, an investigator might wish to know which of the samples from various sampling locations showed results that differed statistically from others as well as knowing which ones were different from the reference and/or negative control sample(s). Such a situation might involve sampling from a number of locations at progressively greater distances from a point source of contamination, in which instance the investigator might want to know which sampling locations provided samples that had significantly higher toxicity than others, and thus which locations were particularly deserving of cleanup. *Tukey's test* is designed for such an analysis; this test is commonly found in statistical packages and can deal with unequal sample sizes.[70]

If it were desired to compare the toxicity of the samples from each sampling location with that for the reference sample(s), but the data do not conform to requirements of normality and equal variance, the ANOVA and subsequent tests would be replaced by nonparametric tests. *Steel's Many-One Rank test* would be used if replication were equal, while unequal replication would require use of the *Wilcoxon Rank Sum test* with Bonferroni's adjustment.

5.5.2 Power Analysis

An important factor to consider in the analysis of the results for toxicity tests with soil is the potential for declaring false positives (i.e., calling a *clean* site *contaminated*; Type I error) or false negatives (i.e., calling a *contaminated* site *clean*; Type II error). Scientists are usually cautious in choosing the level of significance (α) for tolerating false positive results (Type I error), and usually set it at P = 0.05 or 0.01. Recently, toxicologists have been urged to report both α and statistical power (1 - β), i.e., the probability of correctly rejecting the null hypothesis (H_0) and not making a Type II error. There are several factors that influence statistical power, including:

- variability of replicate samples representing the same treatment;
- α (i.e., the probability of making a Type I error);
- *effect size (ES)*, (i.e., the magnitude of the true effect for which you are testing); and

---

[70] An alternative approach is available (EC, 1997a, b, 2001, 2004c). For equal replicates, *Fisher's Least Significant Difference (LSD)* is recommended. It is based on a smaller "pairwise error rate" for α in comparing data for samples from any given location with those for samples from another location, but holds the overall value of α to the pre-selected value (usually 0.05). *LSD* is seldom included in software packages for toxicity, but it is described in some textbooks (e.g., Steel and Torrie, 1980). Instead, *Tukey's test* is recommended here, partly because *LSD* might declare significant differences too readily. *LSD* is also intended for only a few of all the possible comparisons in a set of data, and those comparisons would have to be specified in advance.

- *n* (i.e., the number of samples or replicates used in a test).

Environment Canada's guidance document on statistical methods for environmental toxicity tests (EC, 2004a) provides further information and guidance on errors of Types I and II.

Power analysis can be used *a priori* to determine the magnitude of the Type II error and the probability of false positive results. It can also be used to ascertain the appropriate number of field and laboratory replicates for subsequent surveys involving this test, or to assist in the selection of future sampling sites. It is always prudent to include as many replicates in the test design as is economically and logistically warranted (see Section 5.1); power analysis will assist in this determination. A good explanation of the power of a test, and how to assess it, can be found in USEPA (1994a). Guidance on power analysis is provided in EC (2004a).

Many investigators have difficulty with power analysis, and do not apply it due to its perceived complexity and the differing formulae specific to various statistical tests. In view of this complexity, the *Minimum Significant Difference* may be applied as an alternative approach (i.e., as an "index of power"; see EC, 2004a for guidance).

Specific Procedures for Testing Chemical-Spiked Soil

This section gives guidance and instructions for preparing and testing *negative control soil* spiked experimentally with chemical(s) or chemical product(s). These recommendations and instructions apply to the biological test method described in Section 4. Guidance in EC (1995) for *spiking* negative control sediment with chemical(s) and conducting toxicity tests with chemical/sediment mixtures is also relevant here, for *chemical-spiked soil*. Further evaluation and standardization of procedures for preparing chemical-spiked soil provided herein (Section 6.2) might be required before soil toxicity tests with plants or other appropriate soil organisms are applied to evaluate specific chemical/soil mixtures for regulatory purposes.

The cause(s) of soil toxicity and the interactive *toxic* effects of chemical(s) or chemical product(s) in association with otherwise *clean* soil can be examined experimentally by spiking negative control soil (Section 3.4) with these substances. The *spiking* might be done with one or more chemicals or chemical products. Other options for toxicity tests with plants, performed using the procedures described herein, include the *spiking* of chemical(s) or chemical product(s) in *reference soil* (Section 3.6) or *test soil* (Section 3.7). Toxicity tests using soil spiked with a range of concentrations of test chemical(s) or chemical product(s) can be used to generate data that estimate statistical endpoints based on threshold concentrations causing specific *sublethal effects* (see Section 4.8).

In Section 6.2, procedures are described for preparing test mixtures of chemical-spiked soil. Section 6.3 describes making observations and measurements during and at the end of the toxicity test, and Sections 4.8 and 6.4 provide procedures for estimating test endpoints for multi-concentration tests. These procedures also apply to the mixing of multiple concentrations of field-collected test soil (including particulate waste material such as sludge or other dredged material intended for land disposal) in negative control soil or reference soil, and to performing multi-concentration tests and determining statistical endpoints for these mixtures (see Section 5, and especially 5.5). Multi-concentration tests with *positive control soil* (Section 3.5) or one or more *reference toxicant*s spiked in negative control soil (Section 4.9) are also performed using the procedures and statistical guidance described in this section. Additionally, the influence of the physicochemical characteristics of natural or artificial negative control soil on chemical toxicity can be determined with spiked-soil toxicity tests according to the procedures and statistical guidance described in this section.

6.1 Sample Properties, Labelling, and Storage

Information should be obtained on the properties of the chemical(s) or chemical product(s) to be spiked experimentally in the negative control soil.[71] Information should also be obtained for individual chemicals or chemical products (e.g., pesticides or other commercial formulations) on their concentration of major or "active" ingredients and impurities, water solubility, vapour pressure, chemical stability, dissociation constants, adsorption coefficients, toxicity to humans and terrestrial organisms, and biodegradability. Where aqueous solubility is in doubt or problematic, acceptable procedures previously used for preparing aqueous

[71] Some studies might require the *spiking* (mixing) of one or more concentrations of chemical(s), chemical product(s), or *test soil* (e.g., contaminated or potentially contaminated field-collected soil or waste sludge) in either *negative control soil* or *reference soil*. Other applications could include the spiking of chemical(s) or chemical product(s) in one or more samples of test soil. For such studies involving samples of contaminated soil or similar particulate material (e.g., domestic or industrial sludge), instructions on sample characterization given in Section 5.2 should be followed. Sample(s) of field-collected negative control soil, reference soil, contaminated soil, or particulate waste to be evaluated in spiked-soil toxicity tests should be collected, labelled, transported, stored, and analyzed according to instructions provided in Sections 5.1 and 5.2.

solutions of the chemical(s) should be obtained and reported. If an acceptable procedure for solubilizing the test chemical(s) in water is not available, preliminary testing for its solubility in test water of a non-aqueous solvent should be conducted and confirmed analytically. Other available information such as the structural formulae, nature and percentage of significant impurities, presence and amounts of additives, and n-octanol:water partition coefficient, should be obtained and recorded. Any pertinent Material Safety Data Sheets (MSDSs) should be obtained and reviewed.

Chemical(s) to be tested should be at least reagent grade, unless a test on a formulated commercial product or technical grade chemical(s) is required. Chemical containers must be sealed and coded or labelled upon receipt. Required information (chemical name, supplier, date received, person responsible for testing, etc.) should be indicated on the label and/or recorded on a separate datasheet dedicated to the sample, as appropriate. Storage conditions (e.g., temperature, protection from light) are frequently dictated by the nature of the chemical.

6.2 Preparing Test Mixtures

On the day of the start of the toxicity test (i.e., Day 0), the mixture(s) of chemical(s) or chemical product(s) spiked in *negative control soil* should be prepared and transferred to test vessels (see Sections 4.1 and 4.2). Each *batch* of *test soil* representing a particular treatment (concentration) should be prepared in a quantity sufficient to enable all test replicates of that treatment (concentration) to be set up along with any additional replicates or quantities required for physicochemical analyses (Section 6.3) or the performance of other soil toxicity tests using earthworms or other soil organisms (e.g., those performed according to EC, 2004c or EC, 2005c).

The use of *artificial soil* (Section 3.4.2) to prepare each test mixture is recommended because it offers a consistent, standardized soil for comparing results for other chemicals or chemical products tested similarly in the same laboratory or by others (e.g., according to USEPA, 1989, ISO, 1993a, 1995; ASTM, 1999b; and OECD, 2000a). If used, the formulation for artificial soil provided in Section 3.4.2 should be followed. The quantity of artificial soil required for the test(s) should be prepared, hydrated to ~20% moisture content, adjusted if and as necessary to a pH within the range of 6.5–7.5[72], aged for a minimum three-day period, and stored at 4 ± 2 °C until required (see Section 3.4.2). The final *moisture content* (including that due to the addition of a measured aliquot of a test chemical or chemical product dissolved in *test water*, with or without an organic solvent) of any *chemical-spiked soil* prepared using artificial soil should be ~70% of the water-holding capacity of the final mixture (Section 3.4.2), for each treatment (concentration).[73] The

---

[72] If, however, the test chemical(s) or chemical product(s) are anticipated to modify soil pH and the intent of the study is to nullify this influence, the (aqueous) pH of each *batch* (concentration) should be adjusted to a standard value (e.g., pH 6.5). Studies for determining the extent to which an acidic or basic test substance modifies the toxicity of soil spiked with a range of concentrations of this substance, due to the influence of pH *per se*, should involve two side-by-side tests. One test adjusts the pH of each test concentration to a standard value (e.g., pH 6.5) using the required (differing, depending on concentration) quantity of calcium carbonate, and the other uses an identical quantity of calcium carbonate for each treatment sufficient to attain the "standard" pH (e.g., 6.5) in the *negative control* treatment.

[73] The following example provides calculations that show the volume of both water (de-ionized or distilled) and a stock solution of a reference toxicant (boric acid) to be added to a sample of artificial soil with an existing moisture content, to create a treatment with a moisture content that is 70% of the WHC for the artificial soil. The calculations take into account the volume of a stock solution of boric acid added when preparing the treatment, as part of the overall adjustment for soil moisture content. To simplify the calculations, this example assumes that 400 g (dry wt) of artificial soil (AS) is sufficient to provide the 500-mL aliquot of soil to be added to each test vessel when performing a reference toxicity test involving three replicate test vessels per treatment.

The equations shown in Section 5.3 for calculating WHC and adjusting soil moisture content to a certain percentage of this value apply equally here. For this example, assume that the following assumptions apply (see Section 5.3 for equations and associated definitions of these terms).

Assumptions:

Wet mass of artificial soil (AS)	= 3.2486 g
Dry mass of AS	= 2.6924 g

Moisture content (MC)
of AS = [(3.2486 – 2.6924)/2.6924] × 100
= 20.66% (initial moisture content)

Water-holding capacity (WHC) of AS = 72.10%
Percentage of WHC desired (P_{WHC}) = 70.00%

Dry mass of AS required for test (M_D) =
[400.00 g per rep × 3 reps] + 25.00 g extra
= 1225.00 g dry wt

Wet mass of AS required for test (M_W)=
(1225.00 × 3.2486)/2.6924
= 1478.06 g wet wt

Calculations to prepare a treatment comprised of 2000 mg boric acid per kg artificial soil (dry wt):

The stock solution consists of 25 g of H_3BO_3 in 1 L of de-ionized water.

The amount of boric acid required, on a dry-mass basis is:
H_3BO_3 = (2 g H_3BO_3/1000 g soil dry wt) × 1225.00 g dry wt
= 2.45 g H_3BO_3

The amount of stock solution required, on a volume basis, is:

H_3BO_3 = 2.45 g H_3BO_3/(25 g H_3BO_3/1000 mL of water)
= 98.00 mL stock solution

The percentage of water (P_W) required for addition to this treatment to achieve the desired percentage of WHC (70%) is:

P_W = [WHC × (P_{WHC}/100)] – MC
= [72.10 × (70.00/100)] – 20.66
= 29.81 %

The volume of water (V_W) required for addition to this treatment to achieve the desired percentage of WHC (70%) is):

V_W = (P_W × M_D)/100
= (29.81 × 1225.00 g dry wt)/100
= 365.17 mL of water required

However, as part of this required volume, 98.00 mL of the stock solution is to be added for dosing; therefore, an additional volume of water of only 267.17 mL will be required (365.17 mL of water – 98.00 mL of stock solution).

Accordingly, the final total mass of soil required, based on wet weight, would be 1843.23 g [1478.06 g wet wt at the soil's initial moisture content (i.e., M_W) + 267.17 mL of water + 98.00 mL of stock solution], and the final moisture content of the soil, based on dry weight, would be 50.47 % {[(1843.23 – 1225.00)/1225.00] × 100}.

The final moisture content of this test treatment (i.e., 50.47% moisture) represents 70% of the test soil's water-holding capacity (50.47 ÷ 72.10 = 0.70).

final moisture content of each mixture (treatment) included in a test should be as similar as possible.

Investigators may choose to use natural control soil (Section 3.4.1) rather than artificial control soil (Section 3.4.2) as the negative control soil to be spiked with chemical(s) or chemical product(s) and for the corresponding replicates of control soil to be included in the test. Procedures described herein for artificial soil apply equally if natural soil is used. An exception is that the final moisture content of each batch of chemical-spiked soil (including control batches) prepared using field-collected soil should be adjusted to the optimal percentage of its WHC using guidance in Section 5.3. The volume of soil in each test vessel might also differ if natural soils are used, due to differences in bulk density of the various soils that might be used.

The procedure to be used for experimentally *spiking* soil is contingent on the study objectives and the nature of the test substance to be mixed with negative control soil or other soil. In many instances, a chemical/soil mixture is prepared by making up a *stock solution* of the test chemical(s) or chemical product(s) and then mixing one or more measured volumes into artificial or natural negative control soil (Section 3.4). The preferred solvent for preparing stock solutions is *test water* (i.e., de-ionized or distilled water); use of a solvent other than 100% test water should be avoided unless it is absolutely necessary. For test chemical(s) or chemical product(s) that do not dissolve readily in test water, a suitable water-miscible organic solvent of relatively low toxicity (e.g., acetone, methanol, or ethanol) may be used in small quantities to help disperse the test substance(s) in water. Surfactants should not be used.

If an organic solvent is used, the test must be conducted using a series of replicate test vessels containing only negative control soil (i.e., 100% artificial or natural *clean* soil containing no solvent and no test substance), as well as a series of replicate test vessels containing only *solvent control soil* (OECD, 1984a, 2000a; ISO, 1993a, 1995; ASTM, 1999b; EC, 2000). For this purpose, a *batch* of solvent control soil must be prepared containing the concentration of the solubilizing agent that is present in the highest concentration of the test chemical(s) or chemical product(s) in soil. Solvent from the same *batch* used to make the stock solution of test substance(s) must be used. Solvents should be used sparingly because they might contribute to the toxicity of the prepared *test soil*. The maximum concentration of solvent in the soil should be at a concentration that does not affect the *emergence* or *growth* of plants during the test. If this information is unknown, a preliminary *solvent only* test, using various concentrations of solvent in negative control soil, should be conducted to determine the threshold-effect concentration of the particular solvent being considered for use in the definitive test.

For tests involving the preparation of concentrations of chemical spiked in *artificial soil*, in which the chemical is insoluble in water but soluble in an organic solvent, the quantity of test substance needed to prepare a required volume of a particular test concentration should be dissolved in a small volume of a suitable organic solvent (e.g., acetone). This chemical-in-solvent mixture should then be sprayed onto or mixed into a small portion of the final quantity of fine quartz sand that is required when preparing each test concentration comprised of a measured amount of a particular chemical-in-solvent mixture spiked in artificial soil (see Section 3.4.2). The solvent is then removed by evaporation by placing the container under a fume hood for at least one hour, and until no residual odour of the solvent can be detected. Thereafter, the chemical-in-sand mixture (with solvent evaporated) is mixed thoroughly with the remaining quantity of pre-moistened sand and other ingredients required to make up artificial soil (Section 3.4.2). An amount of de-ionized water necessary to achieve a final moisture content of approximately 70% of the maximum water-holding capacity for this artificial soil is then added and mixed with the soil/sand/peat mixture. The chemical-spiked soil can then be added to the test vessels (OECD, 2000a).

The sample of *solvent control soil* to be included in the test must be prepared using the same procedure but without the addition of the test chemical. Additionally, the solvent control soil must contain a concentration of solvent that is as high as that in any of the concentrations of chemical-spiked soil included in a test.

If the test chemical to be spiked in artificial soil is insoluble in both water and any suitable (non-toxic) organic solvent, a mixture should be prepared which is comprised of 10 g of finely ground industrial quartz sand and the quantity of the test chemical necessary to achieve the desired test concentration in the soil. This mixture should then be mixed thoroughly with the remaining constituents of the pre-moistened artificial soil. An amount of de-ionized water necessary to achieve a final *moisture content* of ~70% of the maximum water-holding capacity is then added and mixed in. The resulting mixture of chemical-spiked soil can then be added to the test vessels (OECD, 2000a).

Concentrations of chemical(s) or chemical product(s) in soil are usually calculated, measured, and expressed as mg test substance/kg soil (or µg substance/g soil) on a dry-weight basis (OECD, 1984a; ISO, 1993a, 1995). The assessment endpoints (e.g., ICps) are similarly expressed on a dry-weight basis (Section 4.8).

Mixing conditions, including test solution:soil ratio, mixing and holding time, and mixing and holding temperature, must be standardized for each treatment included in a test. Time for mixing a *spiked soil* should be adequate to ensure homogeneous distribution of the chemical, and may be from minutes up to 24 h. During mixing, temperature should be kept low to minimize microbial activity and changes in the mixture's physicochemical characteristics. Analyses of subsamples of the mixture are advisable to determine the degree of mixing and homogeneity achieved.

For some studies, it might be necessary to prepare only one concentration of a particular mixture of

negative control (or other) soil and chemical(s) or chemical product(s), or a mixture of only one concentration of *contaminated* soil or particulate waste in negative control or other soil. For instance, a single-concentration test might be conducted to determine whether a specific concentration of chemical or chemical product in *clean* soil is *toxic* to the test organisms. Such an application could be used for research or regulatory purposes (e.g., "limit" test).

A multi-concentration test, using a range of concentrations of chemical added to *negative control soil* (or other soil) under standardized conditions, should be used to determine the desired endpoint(s) (i.e., EC50 and ICp; see Section 4.8) for the chemical/soil mixtures. A multi-concentration test using negative control soil spiked with a specific particulate waste might also be appropriate. At least nine test concentrations plus the appropriate control treatment(s) must be prepared for each multi-concentration test, and more (i.e., ≥11 plus controls) are recommended (see Sections 4.1 and 4.8). When selecting the test concentrations, an appropriate geometric dilution series may be used in which each successive concentration of chemical(s) or chemical product(s) in soil is at least 50% of the previous one (e.g., 160, 80, 40, 20, 10, 5, 2.5, 1.25, 0.63 mg/kg). Test concentrations may also be selected from other appropriate logarithmic dilution series (see Appendix H); or may be derived based on the findings of preliminary "range-finding" toxicity tests. The reader is referred to Section 4.1 for additional guidance when selecting test concentrations.

To select a suitable range of concentrations, a preliminary or range-finding test covering a broader range of test concentrations might prove worthwhile. The number of replicates per treatment (see Section 4.1) could be reduced or eliminated altogether for range-finding tests and, depending on the expected or demonstrated (based on earlier studies with the same or a similar test substance) variance among test vessels within a treatment, might also be reduced for nonregulatory screening *bioassays* or research studies.

Based on the objectives of the test, it might be desirable to determine the influence of substrate characteristics (e.g., particle size or organic matter content) on the toxicity of chemical/soil mixtures. For instance, the influence of soil particle size on chemical toxicity could be measured by conducting concurrent multi-concentration tests with a series of mixtures comprised of the test chemical(s) or chemical product(s) mixed in differing fractions (i.e., segregated particle sizes) or types of natural or artificial *negative control soil* (Section 3.4). Similarly, the degree to which the *total organic carbon* content (%) or *organic matter* content (%) of soil can modify chemical toxicity could be examined by performing concurrent multi-concentration tests using different chemical/soil mixtures prepared with a series of organically enriched negative control soils. Each fraction or formulation of natural or artificial negative control soil used to prepare these mixtures should be included as a separate control in the test.

Depending on the study objectives and design, certain soil toxicity tests using plants might be performed with samples of negative control soil or reference soil to which chemical(s) or chemical product(s) are applied to the soil surface, rather than mixing it with the soil. Surface applications can be applied in the field or the laboratory. Procedures for chemical application include the use of a calibrated track sprayer to achieve a uniform distribution of the chemical over a specific area. Concentration of chemical(s) or chemical product(s) in the soil can be determined based on the penetration depth, the surface area or swathe width, the nozzle size, the pressure, and the speed of coverage of the sprayer (G. L. Stephenson, personal communication, Aquaterra Environmental, Orton, ON, 2001). The OECD (2000a) provides some guidelines for applying test substances to the soil surface, in preparation for emergence-and-growth tests with terrestrial plants.

6.3 Test Observations and Measurements

A qualitative description of each mixture of *chemical-spiked soil* should be made when the test is being established. This might include observations of the colour, texture, and visual homogeneity of each mixture of chemical-spiked soil. Any change in appearance of the test mixture during the test, or upon its termination, should be recorded.

Section 4.6 provides guidance and requirements for the observations and measurements to be made at the beginning, during, and at the end of the test. These observations and measurements apply and must be made when performing the soil toxicity test described herein using one or more samples of chemical-spiked soil.

Depending on the test objectives and experimental design, additional test vessels might be set up at the beginning of the test (see Section 4.1) to monitor soil chemistry. These would be destructively sampled during or at the end of the test. Test organisms might or might not be added to these extra test vessels, depending on study objectives. Measurements of chemical concentrations in the soil within these test vessels could be made by removing aliquots of soil for the appropriate analyses, at the beginning of the test, as it progresses, and/or at its end, depending on the nature of the toxicant and the objectives of the test.

Measurements of the quality (including soil pH and moisture content) of each mixture of *spiked soil* being tested (including the *negative control soil*) must be made and recorded at the beginning and end of the test for pH and at the beginning of the test only for moisture content, as described in Section 4.6. If analytical capabilities permit, it is recommended that the *stock solution*(s) be analysed together with one or more subsamples of each spiked-soil mixture, to determine the chemical concentrations, and to assess whether the soil has been spiked satisfactorily. These should be preserved, stored, and analyzed according to suitable, validated procedures.

Unless there is good reason to believe that the chemical measurements are not accurate, toxicity results for any test in which concentrations are measured for each spiked-soil mixture included in the test should be calculated and expressed in terms of these measured values. As a minimum, sample aliquots should be taken from the high, medium, and low test concentrations at the beginning and end of the test[74]; in which instance, endpoint values calculated (Sections 4.8 and 6.4) would be based on nominal ones. Any measurements of concentrations of the test chemical(s) or chemical product(s) should be compared, reported, and discussed in terms of their degree of difference from nominal strengths. If nominal concentrations are used to express toxicity results, this must be explicitly stated in the test-specific report (see Section 7.1.6).

6.4 Test Endpoints and Calculations

Multi-concentration tests with mixtures of *spiked soil* are characterized by test-specific statistical endpoints (see Section 4.8). Guidance for calculating the EC50 for emergence data is provided in Section 4.8.2, whereas that for calculating an ICp (based on data showing growth inhibition; see Section 4.8) is presented in Section 4.8.3. Section 5.5 provides guidance for calculating and comparing endpoints for single-concentration tests using samples of field-collected soil, which applies equally to single-concentration tests performed with mixtures of spiked soil. For further information on these or other appropriate parametric (or nonparametric) statistics to apply to the endpoint data, the investigator should consult the Environment Canada report on statistics for the determination of toxicity endpoints (EC, 2004a).

For any test that includes *solvent control soil* (see Section 6.2), the test results for plants held in that soil must be compared statistically with that for test organisms held in *negative control soil*. If any of the endpoints for these two control soils used to establish test validity (see Section 4.4) differ significantly according to *Student's t-test*, only the *solvent control soil* may be used as the basis for comparison and calculation of results. If the results for the two controls are the same, the data from both controls should be combined before using it to calculate results or to assess test validity.

[74] Certain chemicals might be known to be stable under the defined test conditions, and unlikely to change their concentration over the test duration. In this instance, an investigator might choose to restrict their analyses to samples taken only at the beginning of the test.

Reporting Requirements

Each test-specific report must indicate if there has been any deviation from any of the *must* requirements delineated in Sections 2 to 6 and, if so, provide details of the deviation(s). The reader must be able to establish from the test-specific report whether the conditions and procedures preceding and during the test rendered the results valid and acceptable for the use intended.

Section 7.1 provides a list of items that must be included in each test-specific report. A list of items that must either be included in the test-specific report, provided separately in a general report, or held on file for a minimum of five years, is found in Section 7.2. Specific *monitoring* programs, related test *protocols*, or regulations might require selected test-specific items listed in Section 7.2 (e.g., details about the test material and/or explicit procedures and conditions during sample collection, handling, transport, and storage) to be included in the test-specific report, or might relegate certain test-specific information as *data to be held on file*.

Procedures and conditions common to a series of ongoing tests (e.g., routine toxicity tests for *monitoring* or *compliance* purposes) and consistent with specifications in this document, may be referred to by citation or by attachment of a general report that outlines standard laboratory practice.

Details on the procedures, conditions, and findings of the test, which are not conveyed by the test-specific report or general report, must be kept on file by the laboratory for a minimum of five years so that the appropriate information can be provided if an audit of the test is required. Filed information might include:

- a record of the chain-of-continuity for field-collected or other samples tested for regulatory or *monitoring* purposes;
- a copy of the record of acquisition for the sample(s);
- chemical analytical data on the sample(s) not included in the test-specific report;
- bench sheets for the observations and measurements recorded during the test;
- bench sheets and warning chart(s) for the reference toxicity tests;
- detailed records of the source of the test organisms, their taxonomic confirmation, and all pertinent information regarding their sorting, preparation, and storage; and
- information on the calibration of equipment and instruments.

Original data sheets must be signed or initialled, and dated by the laboratory personnel conducting the tests.

7.1 Minimum Requirements for a Test-Specific Report

The following items must be included in each test-specific report.

7.1.1 Test Substance or Material

- brief description of sample type (e.g., waste sludge, reference or contaminated field-collected soil, negative control soil) or coding, as provided to the laboratory personnel;
- information on labelling or coding of each sample; and
- date of sample collection; date and time sample(s) received at test facility.

7.1.2 Test Organisms

- species and source of test seeds;
- scientific name, seed variety, and lot number; and

- any unusual appearance or treatment of the seeds, before their use in the test.

7.1.3 Test Facilities

- name and address of test laboratory; and
- name of person(s) performing the test (or each component of the test).

7.1.4 Test Method

- citation of biological test method used (i.e., as per this document);
- design and description if specialized procedure(s) (e.g., preparation of mixtures of *spiked soil*; preparation and use of solvent and, if so, solvent control) or modification(s) of the standard test method described herein;
- brief description of frequency and type of all measurements and all observations made during test; and
- name and citation of program(s) and methods used for calculating statistical endpoints.

7.1.5 Test Conditions and Procedures

- design and description of any deviation(s) from, or exclusion of, any of the procedures and conditions specified in this document;
- number of discrete samples per treatment; number of replicate test vessels for each treatment; number and description of treatments in each test including the control(s); test concentrations (if applicable);
- volume and/or mass of soil in each test vessel;
- number of seeds per test vessel and treatment;
- dates when test was started and ended;
- measurements of light intensity adjacent to surface of soil in test vessels;
- for each soil sample—any measurements of soil particle size, moisture content, water-holding capacity, pH, and conductivity; and
- for each composite sample of subsamples taken at the same time from all replicates of each treatment—all measurements of temperature, pH, moisture content, and water-holding capacity.

7.1.6 Test Results

- number of seedlings and observations on seedling condition in each test vessel, as noted during each observation period over the test duration;
- mean (± SD) percent emergence in each treatment, including control(s), on Day 7 (if determined) and at test end (Day 14 or Day 21, depending on species of test organism);
- mean (± SD) shoot length of individual plants surviving in each treatment [(including the control(s)] at test end; mean (± SD) root length of individual plants surviving in each treatment at test end; mean (± SD) shoot dry weight of individual plants surviving in each treatment at test end; mean (± SD) root dry weight of individual plants surviving in each treatment at test end;
- mean (± SD) shoot and root wet weight of individual plants surviving in each treatment [(including the control(s)] at test end, if determined;
- any EC50 (including the associated 95% confidence limits, and if calculated, the slope) determined; any additional ECx (e.g., EC20) calculated;
- any ICp (together with its 95% confidence limits) determined for the data on growth (i.e., shoot and root lengths and shoot and root wet and dry weights of individual plants surviving at test end); details regarding any transformation of data, and indication of quantitative statistical method used or procedures applied to the data;
- for a multi-concentration test with *chemical-spiked soil*, indication as to whether results are

based on nominal or measured concentrations of chemical(s) or chemical product(s); all values for measured concentrations;

- results for any 7- or 10-day (depending on test species) ICp (including its 95% confidence limits) performed with the *reference toxicant* in conjunction with the definitive soil toxicity test, using the same lot of test seed; *geometric mean* value (± 2 SD) for the same reference toxicant and test species, as derived at the test facility in previous 7- or 10-day ICp tests using the procedures and conditions for reference toxicity tests described herein; and

- anything unusual about the test, any problems encountered, any remedial measures taken.

7.2 Additional Reporting Requirements

The following items must be either included in the test-specific report or the general report, or held on file for a minimum of five years.

7.2.1 Test Substance or Material

- identification of person(s) who collected and/or provided the sample;

- records of sample chain-of-continuity and log-entry sheets; and

- conditions (e.g., temperature, in darkness, in sealed container) of sample upon receipt and during storage.

7.2.2 Test Organisms

- name and address of seed supplier;

- year of collection (if applicable), packet size, lot number, percent germination rating, date of germination rating, date of purchase, name of supplier, and date seed package was opened;
- description of procedures used to sort seeds; and

- description of storage conditions and procedures, including temperature and duration of seed-lot storage.

7.2.3 Test Facilities and Apparatus

- all results for initial tests with negative control soil and reference toxicant, undertaken by the laboratory previously inexperienced with performing the biological test method described herein in advance of any reporting of definitive test results (see Section 3.2.1);

- description of systems for providing lighting and for regulating temperature within test facility;

- description of test vessels and covers; and

- description of procedures used to clean or rinse test apparatus.

7.2.4 Negative Control Soil or Reference Soil

- procedures for the preparation (if *artificial soil*) or *pretreatment* (if natural soil) of *negative control soil*;

- source of natural soil; history of past use and records of analyses for pesticides or other contaminants;

- formulation of *artificial soil*, including sources for the constituents and conditions and procedures for hydration and pH adjustment; and

- storage conditions and duration before use.

7.2.5 Test Method

- procedures used for mixing or otherwise manipulating *test soils* before use; time interval between preparation and testing;

- procedure used in preparing stock and/or test solutions of chemicals; description and concentration(s) of any solvent used;

- details concerning aliquot sampling, preparation, and storage before physicochemical analysis, together with available information regarding the analytical methods used (with citations); and

- use and description of preliminary or range-finding test.

7.2.6 Test Conditions and Procedures

- procedure for adding seed to test vessels;
- relative humidity of test facility;
- appearance of each sample (or mixture thereof) in test vessels; changes in appearance noted during test;
- records of hydration of *test soils* in each test vessel throughout duration of test and qualitative description of drying of soils during the test (e.g., rate, appearance);
- any other physicochemical measurements (e.g., analyses of aliquots from the same *batch* to determine conductivity, homogeneity, contaminant concentration, total volatile solids, biochemical oxygen demand, chemical oxygen demand, total inorganic carbon, cation exchange capacity, oxidation-reduction potential, total nitrogen) made before and during the test on test material (including *negative control soil* and *reference soil*) and contents of test vessels, including analyses of whole soil and pore water;
- any other observations or analyses made on the test material (including samples of *negative control soil* or *reference soil*); e.g., qualitative and/or quantitative data regarding indigenous macrofauna or detritus, or results of geochemical analyses; and
- any chemical analyses of the concentration of chemical in *stock solution*(s) of *reference toxicant* and, if measured, in test concentrations.

7.2.7 Test Results

- results for any range-finding test(s) conducted;
- percent emergence of plants in each test vessel at test end (Day 14 or Day 21); and on Day 7 (if determined);
- number of surviving plants in each test vessel at test end (Day 14 or Day 21); mean individual shoot/root length; and replicate dry weight and mean individual dry weights of shoots and roots (and the same for shoot and root wet weight, if performed) of plants surviving in each test vessel at test end; for regression analyses, hold on file information indicating sample size (e.g., number of replicates per treatment), parameter estimates with variance, any ANOVA table(s) generated, plots of fitted and observed values of any models used, and the output provided by the statistical program (e.g., SYSTAT);
- warning charts (for ICps causing reduced shoot lengths) showing the most recent and historic results for toxicity tests with the *reference toxicant* and the selected species of test organism used in these tests;
- graphical presentation of data; and
- original bench sheets and other data sheets, signed and dated by the laboratory personnel performing the test and related analyses.

References

ACECSS (Agriculture Canada Expert Committee on Soil Survey), "The Canadian System of Soil Classification", Publication 1646, Second Edition, ISBN 0-660-12460-2, Research Branch, Agriculture Canada, Ottawa, ON (1987).

AESA (Alberta Environmentally Sustainable Agriculture Program), "AESA Soil Quality Benchmark Study — Soil Organic Matter", Fact Sheet FS2001-ISQ, 4 p., AESA Soil Quality Program, Conservation and Development Branch, Alberta Agriculture, Food and Rural Development, Edmonton, AB (2001).

AFBMI (Alberta Farm Business Management Initiative), "Durum Wheat", Available: http://members.shaw.ca/bethcandlish/dwht.htm (1998).

Anderson, M.S., "*Bouteloua gracilis*", In: *Fire Effects Information System* [Online]. United States Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (producer). Available: http://www.fs.fed.us/database/feis/ (2003).

Aquaterra Environmental, "Development of Plant Toxicity Tests for Assessment of Contaminated Soils", report prepared by Aquaterra Environmental, Ltd. (Orton, ON) for the Method Development and Applications Section, Environment Canada, Ottawa, ON (1998a).

———. "Development of a Reproduction Toxicity Test with *Onychiurus folsomi* for Assessment of Contaminated Soils", report prepared by Aquaterra Environmental, Ltd. (Orton, ON) for the Method Development and Applications Section, Environment Canada, Ottawa, ON (1998b).

———. "Development of Earthworm Toxicity Tests for Assessment of Contaminated Soils", report prepared by Aquaterra Environmental, Ltd. (Orton, ON) for the Method Development and Applications Section, Environment Canada, Ottawa, ON (1998c).

Aquaterra Environmental and ESG (ESG International Inc.), "Assessment of the Biological Test Methods for Terrestrial Plants and Soil Invertebrates: Metals", August 2000 Final Report, prepared by Aquaterra Environmental, Ltd. (Orton, ON) and ESG International Inc. (Guelph, ON) for the Method Development and Applications Section, Environment Canada, Ottawa, ON (2000).

ASTM (American Society for Testing and Materials), "Standard Guide for Use of Lighting in Laboratory Testing", E-1733-95, p. 1249–1259, In: *1999 Book of ASTM Standards—Biological Effects and Environmental Fate; Biotechnology; Pesticides Vol. 11.05*, ASTM, Philadelphia, PA (1999a).

———. "Standard Guide for Conducting Terrestrial Plant Toxicity Tests", E1963-98, p. 1481–1500, In: *1999 Annual Book of ASTM Standards — Biological Effects and Environmental Fate; Biotechnology; Pesticides, Vol. 11.05*, ASTM, Philadelphia, PA (1999b).

———. "Standard Guide for Conducting Laboratory Soil Toxicity or Bioaccumulation Tests With the Lumbricid Earthworm *Eisenia fetida*", E1676-97, p. 1062–1079, In: *1999 Annual Book of ASTM Standards — Biological Effects and Environmental Fate; Biotechnology; Pesticides, Vol. 11.05*, ASTM, Philadelphia, PA (1999c).

Becker-van Slooten, K., S. Campiche, and J. Tarradellas, "Research in Support of the Environment Canada Collembolan Toxicity Test Method with *Folsomia candida* for Assessment of Contaminated Soils", report prepared by the Laboratory of Environmental Chemistry and Ecotoxicology (CECOTOX), Ecole polytechnique fédérale de Lausanne (EPFL) ENAC-ISTE (Lausanne, Switzerland) for the Method Development and Applications Section, Environment Canada, Ottawa, ON (2003).

———. "Research in Support of the Environment Canada Soil Toxicity Test Methods: Consideration of Various Methods for

Measuring Soil pH, Water Holding Capacity, and Water Filled Pore Space", report prepared by Laboratory of Environmental Chemistry and Ecotoxicology (CECOTOX), Ecole polytechnique fédérale de Lausanne (EPFL), ENAC-ISTE (Lausanne, Switzerland) for the Method Development and Applications Section, Environment Canada, Ottawa, ON (2004).

Bonnell Environmental Consulting, "Assessment of Soil Toxicity Test Species for Canadian Representativeness", Technical Report TS-28, Prepared for the Method Development and Applications Section, Environment Canada, by Bonnell Environmental Consulting, Ottawa, ON (1994).

Boutin, C. and C.A. Rogers, "Pattern of Sensitivity of Plant Species to Various Herbicides - An Analysis with Two Databases" *Ecotoxicol., 9*:255–271 (2000).

Boutin, C., K.E. Freemark, and C.J. Keddy, "Overview and Rationale for Developing Regulatory Guidelines for Non-Target Plant Testing with Chemical Pesticides", *Environ. Toxicol. Chem., 14 (9):*1465–1475 (1995).

Boutin, C., H.B. Lee, E.T. Peart, P.S. Betchelor, and R.J. Maguire, "Effects of the Sulfonylurea Herbicide Metsulfuron Methyl on Growth and Reproduction of Five Wetland and Terrestrial Plant Species", *Environ. Toxicol. Chem., 19 (10):*2532–2541 (2000).

Boutin, C., N. Elmegaard, and C. Kjær, "Toxicity Testing of Fifteen Non-Crop Plant Species with Six Herbicides in a Greenhouse Experiment: Implications for Risk Assessment", *Ecotoxicol., 13 (4):*349‑369 (2004).

Brown, R.A. and D. Farmer, "Track-Sprayer and Glasshouse Techniques for Terrestrial Plant Bioassays with Pesticides", pp. 197–208, In: *Plants for Toxicity Assessment*: Second Volume, ASTM STP 1115, J.W. Gorsuch, W.R. Lower, W. Wang, and M.A. Lewis (eds.), American Society for Testing and Materials, Philadephia, PA (1991).

CCME (Canadian Council of Ministers of the Environment), "A Framework for Ecological Risk Assessment at Contaminated Sites in Canada: Review and Recommendations", The National Contaminated Sites Program, Scientific Series 1999, Ecosystem Conservation Directorate, Evaluation and Interpretation Branch, Ottawa, ON, 108 p. (1994).

———. "A Framework for Ecological Risk Assessment: General Guide", CCME Subcommittee on Environmental Quality Criteria for Contaminated Sites, The National Contaminated Sites Program, Winnipeg, MB, 32 p. (1996).

———. "A Framework for Ecological Risk Assessment: Technical Appendices", CCME Subcommittee on Environmental Quality Criteria for Contaminated Sites, The National Contaminated Sites Program, Winnipeg, MB, 67 p. (1997).

Callahan, C.A., C.A. Menzie, D.E. Burmaster, D.C. Wilborn, and T. Ernst, "On-Site Methods for Assessing Chemical Impact on the Soil Environment Using Earthworms: A Case Study at the Baird and McGuire Superfund Site, Holbrook, Massachusetts", *Environ. Toxicol. Chem., 10:*817–826 (1991).

Carter, M.R. (ed.), *Soil Sampling and Methods of Analysis*, Lewis Publishers, CRC Press Inc., Boca Raton, FL (1993).

Chaineau, C.H., J.L. Morel, and J. Oudot, "Phytotoxicity and Plant Uptake of Fuel Oil Hydrocarbons", *J. Environ. Qual., 26:*1478–1483 (1997).

Cole, J.F.H., L. Canning, and R.A. Brown, "Rationale for the Choice of Species in the Regulatory Testing of the Effects of Pesticides on Terrestrial Non-Target Plants", pp. 152–156, In: *Brighton Crop Protection Conference - Weeds*, United Kingdom (1993).

Courchesne, F., S. Savoie, and A. Dufresne, "Effects of Air-Drying on the Measurement of Soil pH in Acidic Forest Soils of Quebec, Canada", *Soil Science*, *160 (1):*56–68 (1995).

Crépin, J. and R.L. Johnson, "Soil Sampling for Environmental Assessment", p. 5–18, In: *Soil Sampling and Methods of Analysis*, M.R. Carter

(ed.), Lewis Publishers, CRC Press Inc., Boca Raton, FL (1993).

Deitze, G., "Spectral comparisons of sunlight and different lamps" *c74*, p. 197–199, In: *Proceedings of International Lighting in Controlled Environments Workshop*, T. W. Tibbits, (ed.), 1 Mar. 1994, Madison, WI (1994).

Duke, J. A., "Handbook of Energy Crops". Unpublished, exists only as an electronic publication on the New CROPS Web site (http://www.hort.purdue.edu/newcrop/duke_energy/dukeindex.html) (1983).

EC (Environment Canada), "Guidance Document on Measurement of Toxicity Test Precision Using Control Sediments Spiked with a Reference Toxicant", Environmental Protection Service, Ottawa, ON, Report EPS 1/RM/30, 56 p. (1995).

———. "Biological Test Method: Test for Survival and Growth in Sediment Using the Freshwater Amphipod *Hyalella azteca*", Environmental Protection Service, Ottawa, Ontario, Report EPS 1/RM/33, 123 p. (1997a).

———. "Biological Test Method: Test for Survival and Growth in Sediment Using Larvae of Freshwater Midges (*Chironomus tentans* or *Chironomus riparius*)", Environmental Protection Service, Ottawa, ON, Report EPS 1/RM/32, 131 p. (1997b).

———. "Guidance Document on Application and Interpretation of Single-Species Tests in Environmental Toxicology", Environmental Protection Service, Ottawa, ON, Report EPS 1/RM/34, 203 p. (1999a).

———. "Biological Test Method: Test for Measuring the Inhibition of Growth Using the Freshwater Macrophyte, *Lemna minor*", Environmental Protection Service, Ottawa, ON, Report EPS 1/RM/37, 98 p. (1999b).

———. "Standard Operating Procedure for Conducting Early Seedling Growth Bioassays", Standard Operating Procedure EC50V1.1, 21 p., April 2000, Environmental Toxicology Section, Pacific Environmental Science Centre, North Vancouver, BC (2000).

———. "Biological Test Method: Test for Survival and Growth in Sediment Using Spionid Polychaete Worms (*Polydora cornuta*)", Environmental Protection Service, Ottawa, ON, Report EPS 1/RM/41, 108 p. (2001).

———. "Guidance Document on Statistical Methods for Environmental Toxicity Tests", Method Development and Applications Section, Environment Canada, Ottawa, ON, Report EPS 1/RM/46 (2004a).

———. "Proceedings to the Workshop on Toxicity Test Methodologies for Assessing the Impacts of Contaminant Mixtures in Soil Using Terrestrial Species of Ecological Relevance to Canadian Soil Systems", Workshop held February 19–21, 2003 at the Pacific Environmental Science Centre, North Vancouver, BC, report published by the Method Development and Applications Section, Environment Canada, Ottawa, ON (2004b).

———. "Biological Test Method: Tests for Toxicity of Contaminated Soil to Earthworms (*Eisenia andrei*, *Eisenia fetida*, or *Lumbricus terrestris*)", Method Development and Applications Section, Environment Canada, Ottawa, ON, Report EPS 1/RM/43, 156 p. (2004c).

———. "Inter-laboratory Validation of Environment Canada's New Test Method for Measuring Soil Toxicity Using Plants", June, 2005, report prepared by J. Princz for the Biological Methods Division, Environment Canada, Ottawa, ON (2005a).

———. "Development of Performance Data for the Determination of Validity Criteria for Environment Canada's Biological Test Method for Assessing Soil Toxicity Using Plants", Biological Methods Division, Ottawa, ON, 50 p. (2005b).

———. "Biological Test Method: Test for Measuring Survival and Reproduction of Springtails Exposed to Contaminants in Soil", Biological Methods Division, Environment Canada, Ottawa, ON, Report EPS 1/RM/47, in preparation, xx p. (2005c).

Eisenhart, C., M.W. Hastay, and W.A. Wallace, *Techniques of Statistical Analysis*, Chapter 15, McGraw-Hill Book Co., New York, NY (1947).

ESG (ESG International Inc.), "Final Report on the Acute Screening and Definitive, Chronic Toxicity Tests with Motor Gasoline", unpublished report, prepared by ESG International Inc. (Guelph, ON) for Petroleum Technology Alliance Canada, Calgary, AB (2000).

———. "Toxicity of Petroleum Hydrocarbons to Soil Organisms and The Effects on Soil Quality: Phase 1 Fraction-Specific Toxicity of Crude Oil", report prepared by ESG International Inc. (Guelph, ON) for Petroleum Technology Alliance Canada, Calgary, AB (2001).

———. "Impacts of Metal-Contaminated Forest Soils From the Canadian Shield on Terrestrial Organisms", report prepared by ESG International Inc. (Guelph, ON) for the Metals in the Environment Research Network Canadian Network of Toxicology Centres, Guelph, ON and the Method Development and Applications Section, Environment Canada, Ottawa, ON (2002).

ESG (ESG International Inc.) and Aquaterra Environmental, "Assessment of the Biological Test Methods for Terrestrial Plants and Soil Invertebrates: Pesticides", report prepared by ESG International Ltd. (Guelph, ON) and Aquaterra Environmental, Ltd. (Orton, ON) for the Method Development and Applications Section, Environment Canada, Ottawa, ON (2002).

Fletcher, J.S., M.J. Muhitch, D.R. Vann, J.C. McFarlane, and F.E. Benenati, "PHYTOTOX Database Evaluation of Surrogate Plant Species Recommended by the U.S. Environmental Protection Agency and the Organization for Economic Cooperation and Development", *Environ. Toxicol. Chem., 4*:523–532 (1985).

Fletcher, J.S., F.L. Johnson, and J.C. McFarlane, "Database Assessment of Phytotoxicity Data Published on Terrestrial Vascular Plants", *Environ. Toxicol. Chem., 7*:615–622 (1988).

Fletcher, J.S., T.G. Pfleeger, and H.C. Ratsch, "Chlorsulfuron Influence on Garden Pea Reproduction", *Physiol. Plant, 94*:261–267 (1995).

Fletcher, J.S., T.G. Pfleeger, H.C. Ratsch, and R. Hayes, "Potential Impact of Low levels of Chlorsulfuron and Other Herbicides on Growth and Yield of Non-Target Plants", *Environ. Toxicol. Chem., 15* (7)*:*1189–1196 (1996).

Godbold, D.L. and A. Hüttermann, "Effect of Zinc, Cadmium and Mercury on Root Elongation of *Picea abies* (Karst.) Seedlings and the Significance of These Metals to Forest Die-Back", *Environ. Pollut. Ser. A (38):*375–381 (1985).

Hausenbuiller, R.L., *Soil Science — Principles & Practices*, 3rd Edition, W.C. Brown Publisher, Dubuque, IA (1985).

Hendershot, W.H., H. Lalonde, and M. Duquette, "Soil Reaction and Exchangeable Acidity", p. 141–145, In: *Soil Sampling and Methods of Analysis*, M.R. Carter (ed.), Canadian Society of Soil Science, Lewis Publishers, Boca Raton, FL (1993).

Hoagland, D.R. and D.I. Aaron, "The Water-Culture Method for Growing Plants Without Soil", Circular 347, College of Agriculture, University of California, Berkeley, CA (1950).

Holst, R.W. and T.C. Ellanger, "Pesticide Assessment Guidelines. Subdivision J. Hazard Evaluation: Non-Target Plants." EPA - 540/9-82-020, Office of Pesticides and Toxic Substances, United States Environmental Protection Agency, Washington, DC (1982).

Huxley, A., M. Griffiths, and M. Levy, *The New RHS Dictionary of Gardening*, MacMillan Press, London, United Kingdom, 3200 pp. (1992).

ISO (International Organization for Standardization), "Soil — Determination of the Effect of Chemical Substances on the Reproduction of Earthworms", Proposal from the Netherlands, 9 p., April 1991. Geneva, Switzerland (1991).

———. "Soil Quality —Determination of the Effects of Pollutants on Soil Flora. Part 1: Method for the Measurement of Inhibition of

Root Growth", ISO 11269-1, Geneva, Switzerland (1993a).

———. "Soil Quality — Effects of Pollutants on Earthworms (*Eisenia fetida*). Part 1: Determination of Acute Toxicity Using Artificial Soil Substrate", ISO 11268-1, Geneva, Switzerland (1993b).

———. "Soil Quality — Determination of the Effects of Pollutants on Soil Flora. Part 2: Effects of Chemicals on the Emergence and Growth of Higher Plants", ISO 11269-2, Geneva, Switzerland (1995).

———. "Soil Quality — Effects of Pollutants on Earthworms (*Eisenia fetida*). Part 2: Determination of Effects on Reproduction", ISO 11268-2, Geneva, Switzerland (1998).

———. "Soil Quality — Inhibition of Reproduction of Collembola (*Folsomia candida*) by Soil Pollutants", ISO 11267, Geneva, Switzerland (1999).

Kalsch, W. and J. Römbke, "Zur chronischen Wirkung von TNT auf die Stoppelrübe *Brassica rapa* im Labortest", p. 100–105, In: *Ökotoxikologie - Ökosystemare Ansätze und Methoden*, J. Oehlmann and B. Barkert (eds.), Berlag Urban & Fischer, Stuttgat, Jena (1999).

Kaputska, L.A., "Selection of Phytotoxicity Tests for Use in Ecological Risk Assessments", p. 515–548, In: *Plants for Environmental Studies*, W. Wang, J.W. Gorsuch, and J.S. Hughes (eds.), CRC Press, New York, NY (1997).

Kaputska L.A., J. Lipton, H. Galbraith, D. Cacela, and K. LeJeune, "Metal and Arsenic Impacts to Soils, Vegetation Communities and Wildlife Habitat in Southwest Montana Uplands Contaminated by Smelter Emissions: II. Laboratory Phytotoxicity Studies", *Environ. Toxicol. Chem., 14 (11):*1905–1912 (1995).

Keddy, C., J.C. Greene, and M.A. Bonnell, "Review of Whole-Organism Bioassays: Soil, Freshwater Sediment, and Freshwater Assessment in Canada", *Ecotoxicol. Environ. Safety, 30:*221–251 (1995).

Keith, L.H., *Environmental Sampling and Analysis: A Practical Guide,* 143 p., Lewis Publishers, Inc., Chelsea, MI (1992).

Kenaga, E.E., "Comparative Toxicity of 131,596 Chemicals to Plant Species", *Ecotoxicol. Environ. Safety, 5*:469–475 (1981).

Kjær, C. and N. Elmegaard, "Effects of Copper Sulfate on Black Bindweed", *Ecotoxicol. Environ. Safety, 33:*110–119 (1996).

Kjær, C., M.B. Pedersen, and N. Elmegaard, "Effects of Soil Copper on Black Bindweed (*Fallopia convolvulus*) in the Laboratory and in the Field", *Arch. Environ. Contam. Toxicol., 35:*14–19 (1998).

Klute, A. (ed.), *Methods of Soil Analysis, Part 1 – Physical and Mineralogical Methods*, 2nd Edition, American Society of Agronomy Inc. and Soil Science Society of America Inc., Madison, WI (1986).

Komex International, "Sampling and Shipping of Reference Soil for the Terrestrial Soil Toxicity Method Development Project", prepared for Environment Canada, Ottawa, ON, 17 p. (1995).

Lock, K. and C.R. Janssen, "Influence of Aging on Copper Bioavailability in Soils", *Environ. Toxicol. Chem., 22 (5):*1162–1166 (2003).

Magness, J.R., G.M. Markle, and C.C. Compton, "Food and Feed Crops of the United States." Interregional Research Project IR-4, IR Bul. 1 (Bul. 828 New Jersey Agr. Expt. Sta.), New Brunswick, NJ, 255 pp. (1971).

Markwiese, J.T., R.T. Ryti, M.M. Hooten, D.I. Michael, and I. Hlohowskyj, "Toxicity Bioassays for Ecological Risk Assessment in Arid and Semiarid Ecosystems", *Rev. Environ. Contam. Toxicol., 168:*43–98 (2000).

McLeod, E., "Feed the Soil", Organic Agricultural Research Institute, Graton, CA, 209 p. (1982).

Mearns, A.J., R.C. Swartz, J.M. Cummins, P.A. Dinnel, P. Plesha, and P.M. Chapman, "Interlaboratory Comparison of a Sediment Toxicity Test Using the Marine Amphipod,

Rhepoxynius abronius", *Marine Environ. Res., 19:*13–37 (1986).

Meier, J.R., L.W. Chang, S. Jacobs, J. Torsella, M.C. Meckes, and M.K. Smith, "Use of Plant and Earthworm Bioassays to Evaluate Remediation of Soil from a Site Contaminated with Polychlorinated Biphenyls", *Environ. Toxicol. Chem., 16 (5):*928–938 (1997).

Menzie, C.A., D.E. Burmaster, J.S. Freshman, and C.A. Callahan, "Assessment of Methods for Estimating Ecological Risk in the Terrestrial Component: A Case Study at the Baird & McGuire Superfund Site in Holbrook, Massachusetts", *Environ. Toxicol. Chem., 11:*245–260 (1992).

Miller, W.E., S.A. Peterson, J.C. Greene, and C.A. Callahan, "Comparative Toxicology of Laboratory Organisms for Assessing Hazardous Waste Sites", *J. Environ. Qual., 14:*569–574 (1985).

Munro, D.B. and E. Small, *Vegetables of Canada*, Agriculture and Agri-Food Canada, Research Branch, Ottawa, ON/National Research Council, Ottawa, ON, 417 pp. (1997).

NERI (National Environmental Research Institute), "Manual of SECOFASE — Development, Improvement and Standardization of Test Systems for Assessing Sublethal Effects on Chemicals on Fauna in the Soil Ecosystem", H. Lokke and C.A.M. van Gestel (eds.), Report from a Workshop held in Silkeborg January 18–19, 1993, Silkeborg, Denmark, 41 p. (1993).

Newman, M.C., *Quantitative Methods in Aquatic Ecotoxicology*, Lewis Pub., Boca Raton, FL, 426 p. (1995).

Norberg-King, T.J., "A Linear Interpolation Method for Sublethal Toxicity: The Inhibition Concentration (ICp) Approach (Version 2.0)", USEPA, Environ. Res. Lab.-Duluth, Duluth, Minnesota, Tech. Report 03-93 of National Effluent Toxicity Assessment Center (1993).

OECD (Organization for Economic Cooperation and Development), "OECD Guideline for Testing of Chemicals: Terrestrial Plants, Growth Test", No. 208, Paris, France (1984a).

———. "OECD Guideline for Testing of Chemicals: Earthworm, Acute Toxicity Tests", No. 207, Paris, France (1984b).

———. "OECD Guideline for the Testing of Chemicals: Proposal for Updating Guideline 208 — Terrestrial (Non-Target) Plant Test: 208 A: Seedling Emergence and Seedling Growth Test; and 208B: Vegetative Vigour Test, Draft Document, Paris, France, 17 p., July 2000 (2000a).

———. "OECD Guideline for the Testing of Chemicals: Proposal for a New Guideline — Earthworm Reproduction Test (*Eisenia fetida/andrei*), Draft Document, Paris, France, 17 p., January 2000 (2000b).

OMAFRA (Ontario Ministry of Agriculture, Food and Rural Affairs), "Field Crop Recommendations 1999–2000", Publication 296, 142 p., Queen's Printer, Toronto, ON (1999).

OMAF (Ontario Ministry of Agriculture and Food), "Agronomy Guide for Field Crops - Publication 811", T. Baute (ed.), [Available online] at www.gov.on.ca/omafra/english/crops/pub811/p811tox, (2002).

OMEE (Ontario Ministry of Environment and Energy), "Guidance on Sampling and Analytical Methods for Use at Contaminated Sites in Ontario — Section 8.3.7 Inorganics", Version 1.1, ISBN-0-7778-4056-1, December 1996, Queen's Printer, Toronto, ON (1996).

Paine, M.D. and C.A. McPherson, "Phase IV Studies of 10-Day Tests for Sediment Toxicity Using Marine or Estuarine Infaunal Amphipods", unpublished report prepared for the Marine Environment Division, Environment Canada, and IGATG by EVS Consultants Ltd., North Vancouver, BC (1991).

Petersen, R.G. and L.D. Calvin, "Sampling", p. 33–51, In: *Methods of Soil Analysis, Part 1 – Physical and Mineralogical Methods*, 2nd Edition, A. Klute (ed.), American Society of Agronomy Inc. and Soil Science Society of America Inc., Madison, WI (1986).

Pfleeger, T., C. McFarlane, R. Sherman, and G. Volk, "A Short-Term Bioassay for Whole Plant

Toxicity", p. 355–364, In: *Plants for Toxicity Assessment*: Second Volume, ASTM STP 1115, J.W. Gorsuch, W.R. Lower, W. Wang, and M.S. Lewis (eds.), American Society for Testing and Materials, Philadelphia, PA (1991).

Rader, B.R., D.W.R. Nimmo, and P.L. Chapman, "Phytotoxicity of Floodplain Soils Contaminated with Trace Metals Along the Clark Fork River, Grant-Kohrs Ranch National Historic Site, Deer Lodge, Montana, USA", *Environ. Toxicol. Chem., 16 (7):*1422–1432 (1997).

Redente, E.F., H. Zadeh, and M.W. Paschke, "Phytotoxicity of Smelter-Impacted Soils in Southwestern Montana, USA", *Environ. Toxicol. Chem., 21(2):*269–274 (2002).

Rocchini, R.J., M.J.R. Clark, A.J. Jordan, S. Horvath, D.J. McLeay, J.A. Servizi, A. Sholund, H.J. Singleton, R.G. Watts, and R.H. Young, "Provincial Guidelines and Laboratory Procedures for Measuring Acute Lethal Toxicity of Liquid Effluents to Fish", B.C. Ministry of Environment, Victoria, BC (1982).

Sager, J.C. and C. McFarlane, "Radiation", p. 1–30, In: *Plant Growth Chamber Handbook,* R.W. Langhans and T.W. Tibbits (eds.), North Central Regional Research Publication No. 340, Iowa Agriculture and Home Economics Experiment Station Special Report No. 99, Iowa State University of Science and Technology, Ames, IA (1997).

SAH (Soil Analysis Handbook), "Reference Methods for Soil Analysis", Soil and Plant Analysis Council Inc., Georgia University Station, Athens, GA, 202 p. (1992).

SAS (SAS Institute Inc.), "SAS Procedures Guide, Release 6.03, *and* Additional SAS/STAT Procedures, Release 6/03 (SAS Technical Report P-179). SAS Institute Inc., Cary, NC (1988).

Saterbak, A., R.J. Toy, D.C.L. Wong, B.J. McMain, M.P. Williams, P.B. Dorn, L.P. Brzuzy, E.Y. Chai, and J.P. Salanitro, "Ecotoxicological and Analytical Assessment of Hydrocarbon-contaminated Soils and Application to Ecological Risk Assessment", *Environ. Toxicol. Chem., 18 (7):*1591–1607 (1999).

Saterbak, A., R.J. Toy, B.J. McMain, M.P. Williams, and P.B. Dorn, "Ecotoxicological and Analytical Assessment of Effects of Bioremediation on Hydrocarbon-Containing Soils", *Environ. Toxicol. Chem., 19:*2643–2652 (2000).

Scher, J.S., "*Elymus lanceolatus*" In: *Fire Effects Information System* [Online]. United States Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (producer). Available: http://www.fs.fed.us/database/feis/ (2002).

Sheppard, S.C., "Development of Bioassay Protocols for Toxicants in Soil", Report PIBS 2838, ISBN 0-7778-1440-X, Environmental Research, Research and Technology Branch, Ontario Environment, Queen's Printer, Toronto, ON, February 1994 (1994).

Sheppard, S.C. and W.G. Evenden, "An Approach to Defining a Control or Diluent Soil for Ecotoxicity Assays", pp. 215–226, In: *Environmental Toxicology and Risk Assessment*, Seventh Volume, ASTM STP 1333, E.E. Little, A.J. DeLonay, and B.M. Greenberg (eds.), American Society for Testing and Materials, Philadelphia, PA (1998).

Siciliano, S.D., J.J. Germida, and J.V. Headley, "Evaluation of Prairie Grass Species as Bioindicators of Halogenated Aromatics in Soil", *Environ. Toxicol. Chem., 16 (3):*521–527 (1997).

Small, E., "New Crops for Canadian Agriculture", In: *Perspectives on New Crops and New Uses*, J. Janick (ed.), ASHS Press, Alexandria VA (1999).

Sokal, R.R. and F.J. Rohlf, *Biometry*, W.H. Freement and Co., San Francisco, CA (1969).

Steel, R.G.D. and J.H. Torrie, *Principles and Procedures of Statistics*, McGraw-Hill Book Co., New York, NY (1980).

Stephan, C.E., "Methods for Calculating an LC50", p. 65–84, In: *Aquatic Toxicology and Hazard*

Evaluation, F.L. Mayer and J.L. Hamelink (eds). ASTM STP 634, American Society For Testing and Materials, Philadelphia, PA (1977).

Stephenson, G.L., "Terrestrial Test Methods for Plants and Soil Invertebrates", Ph.D. Thesis, Department of Environmental Biology, University of Guelph, Guelph, ON, 230 p. (2003a).

———. Unpublished Data. Stantec Consulting, Ltd., Guelph, ON (2003b).

Stephenson, G.L., K.R. Solomon, B. Hale, B.M. Greenberg, and R.P. Scroggins, "Development of suitable test methods for evaluating the toxicity of contaminated soils to a battery of plant species relevant to soil environments in Canada", p. 474–489, In: *Environmental Toxicology and Risk Assessment: Modeling and Risk Assessment,* Sixth Volume, ASTM STP 1317, J. Dwyer, T.R. Doane, and M.L. Hinman (eds.), American Society for Testing and Materials, Philadelphia, PA (1997).

Stephenson, G.L., N. Koper, J. McCann, and Z. Wang, "Draft Report on the Acute Screening and Definitive, Chronic Toxicity Tests with Whole Federated Crude Oil", unpublished (draft) report, October 1999, prepared by ESG International Inc. (Guelph, ON) for Petroleum Technology Alliance Canada (Calgary, AB) and the Canadian Association of Petroleum Producers (Calgary, AB), 52 p. + app. (1999a).

Stephenson, G.L., N. Koper, J. McCann, and Z. Wang, "Draft Report on the Acute Screening and Definitive, Chronic Toxicity Tests with Fraction 3 Derived From Federated Crude Oil", unpublished (draft) report, December 1999, prepared by ESG International Inc. (Guelph, ON) for Petroleum Technology Alliance Canada (Calgary, AB) and the Canadian Association of Petroleum Producers (Calgary, AB), 48 p. + app. (1999b).

Stephenson, G.L., J. McCann, P. Jokuty, and Z. Wang, "Draft Report on the Acute Screening and Definitive, Chronic Toxicity Tests with Fraction 2 Derived From Federated Crude Oil", unpublished (draft) report, January 2000, prepared by ESG International Inc. (Guelph, ON) for Petroleum Technology Alliance Canada (Calgary, AB) and the Canadian Association of Petroleum Producers (Calgary, AB), 32 p. + app. (2000a).

Stephenson, G.L., N. Koper, G.F. Atkinson, K.R. Solomon, and R.P. Scroggins, "Use of Nonlinear Regression Techniques for Describing Concentration-Response Relationships of Plant Species Exposed to Contaminated Site Soils", *Environ. Toxicol. Chem., 19:*2968–2981 (2000b).

Stephenson, G.L., J.I. Princz, N. Koper, and P.G. Miasek, "Toxicity Testing with Volatile Substances", pp. 236–252, In: *Environmental Toxicology and Risk Assessment: Science, Policy, and Standardization - Implications for Environmental Decisions:* Tenth Volume ASTM STP 1403, B.M. Greenberg, R.N. Hall, M.R. Roberts Jr., and R.W. Gensemer (eds.), American Society for Testing and Materials, West Conshohocken, PA (2001a).

Stephenson, G.L., N.C. Feisthauer, N. Koper, J.H. McCann, and R.P. Scroggins, "The Influence of Four Types of Water on Seedling Emergence and Growth of Barley and the Toxic Interaction with Copper Sulfate", In: *Environmental Toxicology and Risk Assessment: Science, Policy, and Standardization - Implications for Environmental Decisions,* Tenth Volume ASTM STP 1403, B.M. Greenberg, R.N. Hall, M.R. Roberts Jr., and R.W. Gensemer (eds.), American Society for Testing and Materials, West Conshohocken, PA (2001b).

Stephenson, G.L., R.G. Kuperman, G.L. Linder, and S. Visser, "Toxicity Tests for Assessing Contaminated Soils and Ground Water", pp. 25–43, In: *Environmental Analysis of Contaminated Sites*, G.I. Sunahara, A.Y. Renoux, C. Thellen, C.L. Gaudet, and A. Pilon (eds.), Ecological and Environmental Toxicology Series, John Wiley & Sons Ltd., Chichester, United Kingdom (2002).

Stoskopf, N.D., "Barley and Oat", pp. 444–458, In: *Cereal Grain Crops,* N.C. Stoskopf (ed.), Reston Publishing Company, Inc., Reston, Virginia (1985).

Sullivan, J., "*Medicago sativa*". In Fire Effects Information System [Online]. U.S. Department

of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (producer). Available: http://www.fs.fed.us/database/feis/ (1992).

TOXSTAT, "Version 3.5", Lincoln Research Associates, Inc., PO Box 4276, Bisbee, Arizona 85603, USA, phone 520-432-4092, email danlra@msn.com. [Programs on disk, with printed user's manual.] (1996).

USDA-NRCS (United States Department of Agriculture - Natural Resources Conservation Services), The PLANTS Database, Version 3.5" [Available online]: http://plants.usda.gov, National Plant Data Center, Baton Rouge, LA (2000).

USEPA (United States Environmental Protection Agency), "Protocols for Short Term Toxicity Screening of Hazardous Waste Sites", Report EPA/600/3-88/029, prepared by J.C. Greene, C.L. Bartels, W.J. Warren-Hicks, B.R. Parkhurst, G.L. Linder, S.A. Peterson, and W.E. Miller, February 1989, Corvallis Environ. Research Lab., Corvallis, OR (1989).

———. "ASSESS Users Guide", Report EPA/600/8-91001, Environmental Monitoring Systems Laboratory, Las Vegas, NV (1991).

———. "Methods for Measuring the Toxicity and Bioaccumulation of Sediment-Associated Contaminants with Freshwater Invertebrates", report EPA 600/R-94/024, USEPA, Duluth, MN, 133 p. (1994a).

———. "Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Marine and Estuarine Organisms (Second Edition)", report EPA/600/4-91/003, Office of Research and Development, USEPA, Cincinnati, OH, 483 p. (1994b).

———. "Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to West Coast Marine and Estuarine Organisms", report EPA/600/R-95/136, Office of Research and Development, USEPA, Washington, DC, 661 p. (1995).

———. "Ecological Effects Test guidelines — Vegetative Vigor, Tier II", draft report EPA 712-C-96-364, April 1996, Office of Prevention, Pesticides, and Toxic Substances (OPPTS) 850.6200, Washington, DC (1996).

Valenzuela, H. and J. Smith, "Barley", Sustainable Agriculture Green Manure Crops, SA-GM-3, College of Tropical Agriculture and Human Resources, University of Hawaii, Mānoa, HI (2002).

van Ee, J.J., L.J. Blume, and T.H. Starks, "A Rationale for the Assessment of Errors in the Sampling of Soils", Report EPA/600/4-90/013, United States Environmental Protection Agency, Las Vegas, NV, 57 p. (1990).

van Ewijk, P.H. and H.A. Hoekstra, "Calculation of the EC50 and its confidence interval when subtoxic stimulus is present", *Ecotoxicol. Environ. Safety, 25:*25–32 (1993).

Vaughan, J.G. and C. Geissler, *The New Oxford Book of Food Plants*, Oxford University Press, New York, NY, 239 p. (1997).

Walsh, R.A., "*Festuca rubra*" In: *Fire Effects Information System* [Online]. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (producer). Available: http://www.fs.fed.us/database/feis/ (1995).

Wang, W., "Literature Review on Higher Plants for Toxicity Testing", *Water, Air and Soil Pollution, 59:*381–400 (1991).

———. "Use of Plants for the Assessment of Environmental Contaminants", *Rev. of Environ. Contam. Toxicol., 126:*87–127 (1992).

Wang, W. and K. Freemark, "The Use of Plants for Environmental Monitoring and Assessment", *Ecotoxicol. Environ. Safety, 30:*289–301 (1995).

Webster, R. and M.A. Oliver, *Statistical Methods in Soil and Land Resource Survey*, Oxford University Press, Oxford, United Kingdom, 315 p. (1990).

Wilke, B.-M., M. Mai, and F. Riepert, "The Use of Liquid Fertilizer in a Higher Plant Growth Test", Poster presentation at 13th Annual SETAC-Europe Meeting, Hamburg, Germany (2003).

Wong, D.C.L., E.Y. Chai, K.K. Chu, and P.B. Dorn, "Prediction of Ecotoxicity of Hydrocarbon-Contaminated Soils Using Physicochemical Parameters", *Environ. Toxicol. Chem., 18 (11):*2611–2621 (1999).

Biological Test Methods and Supporting Guidance Documents Published by Environment Canada's Method Development and Applications Section*

Title of Biological Test Method or Guidance Document	Report Number	Publication Date	Applicable Amendments
A. Generic (Universal) Biological Test Methods			
Acute Lethality Test Using Rainbow Trout	EPS 1/RM/9	July 1990	May 1996
Acute Lethality Test Using Threespine Stickleback (*Gasterosteus aculeatus)*	EPS 1/RM/10	July 1990	March 2000
Acute Lethality Test Using *Daphnia* spp.	EPS 1/RM/11	July 1990	May 1996
Test of Reproduction and Survival Using the Cladoceran *Ceriodaphnia dubia*	EPS 1/RM/21	February 1992	November 1997
Test of Larval Growth and Survival Using Fathead Minnows	EPS 1/RM/22	February 1992	November 1997
Toxicity Test Using Luminescent Bacteria (*Photobacterium phosphoreum*)	EPS 1/RM/24	November 1992	—
Growth Inhibition Test Using the Freshwater Alga *Selenastrum capricornutum*	EPS 1/RM/25	November 1992	November 1997
Acute Test for Sediment Toxicity Using Marine or Estuarine Amphipods	EPS 1/RM/26	December 1992	October 1998
Fertilization Assay Using Echinoids (Sea Urchins and Sand Dollars)	EPS 1/RM/27	December 1992	November 1997
Toxicity Tests Using Early Life Stages of Salmonid Fish (Rainbow Trout, Coho Salmon, or Atlantic Salmon)	EPS 1/RM/28 1st Edition	December 1992	January 1995
Toxicity Tests Using Early Life Stages of Salmonid Fish (Rainbow Trout)	EPS 1/RM/28 2nd Edition	July 1998	—
Test for Survival and Growth in Sediment Using the Larvae of Freshwater Midges (*Chironomus tentans* or *Chironomus riparius*)	EPS 1/RM/32	December 1997	—

* These documents are available for purchase from Environmental Protection Publications, Environment Canada, Ottawa, Ontario, K1A 0H3, Canada. Printed copies can also be requested by e-mail at: epspubs@ec.gc.ca. For further information or comments, contact the Chief, Biological Methods Division, Environmental Technology Centre, Environment Canada, Ottawa, Ontario K1A 0H3.

Title of Biological Test Method or Guidance Document	Report Number	Publication Date	Applicable Amendments
A. Generic (Universal) Biological Test Methods (cont'd.)			
Test for Survival and Growth in Sediment Using the Freshwater Amphipod *Hyalella azteca*	EPS 1/RM/33	December 1997	—
Test for Measuring the Inhibition of Growth Using the Freshwater Macrophyte, *Lemna minor*	EPS 1/RM/37	March 1999	—
Test for Survival and Growth in Sediment Using Spionid Polychaete Worms (*Polydora cornuta*)	EPS 1/RM/41	December 2001	—
Tests for Toxicity of Contaminated Soil to Earthworms (*Eisenia andrei, Eisenia fetida*, or *Lumbricus terrestris*)	EPS 1/RM/43	June 2004	—
Tests for Measuring Emergence and Growth of Terrestrial Plants Exposed to Contaminants in Soil	EPS 1/RM/45	February 2005	—
Test for Measuring Survival and Reproduction of Springtails Exposed to Contaminants in Soil	EPS 1/RM/47	2005	—
B. Reference Methods**			
Reference Method for Determining Acute Lethality of Effluents to Rainbow Trout	EPS 1/RM/13 1st Edition	July 1990	May 1996, December 2000
Reference Method for Determining Acute Lethality of Effluents to Rainbow Trout	EPS 1/RM/13 2nd Edition	December 2000	—
Reference Method for Determining Acute Lethality of Effluents to *Daphnia magna*	EPS 1/RM/14 1st Edition	July 1990	May 1996, December 2000
Reference Method for Determining Acute Lethality of Effluents to *Daphnia magna*	EPS 1/RM/14 2nd Edition	December 2000	—
Reference Method for Determining Acute Lethality of Sediment to Marine or Estuarine Amphipods	EPS 1/RM/35	December 1998	—
Reference Method for Determining the Toxicity of Sediment Using Luminescent Bacteria in a Solid-Phase Test	EPS 1/RM/42	April 2002	—

** For this series of documents, a *reference method* is defined as a specific biological test method for performing a toxicity test, i.e., a toxicity test method with an explicit set of test instructions and conditions which are described precisely in a written document. Unlike other generic (multi-purpose or "universal") biological test methods published by Environment Canada, the use of a *reference method* is frequently restricted to testing requirements associated with specific regulations.

Title of Biological Test Method or Guidance Document	Report Number	Publication Date	Applicable Amendments
C. Supporting Guidance Documents			
Guidance Document on Control of Toxicity Test Precision Using Reference Toxicants	EPS 1/RM/12	August 1990	—
Guidance Document on Collection and Preparation of Sediment for Physicochemical Characterization and Biological Testing	EPS 1/RM/29	December 1994	—
Guidance Document on Measurement of Toxicity Test Precision Using Control Sediments Spiked with a Reference Toxicant	EPS 1/RM/30	September 1995	—
Guidance Document on Application and Interpretation of Single-Species Tests in Environmental Toxicology	EPS 1/RM/34	December 1999	—
Guidance Document for Testing the Pathogenicity and Toxicity of New Microbial Substances to Aquatic and Terrestrial Organisms	EPS 1/RM/44	March 2004	—
Guidance Document on Statistical Methods to Determine Endpoints of Toxicity Tests	EPS 1/RM/46	December 2004	—

Environment Canada Regional and Headquarters Offices

Headquarters
351 St. Joseph Boulevard
Place Vincent Massey
Gatineau, Quebec
K1A 0H3

Ontario Region
4905 Dufferin St., 2nd Floor
Downsview, Ontario
M3H 5T4

Atlantic Region
15th Floor, Queen Square
45 Alderney Drive
Dartmouth, Nova Scotia
B2Y 2N6

Prairie and Northern Region
Room 210, Twin Atria No. 2
4999 - 98th Avenue
Edmonton, Alberta
T6B 2X3

Quebec Region
8th Floor, 105 McGill Street
Montreal, Quebec
H2Y 2E7

Pacific and Yukon Region*
401 Burrard Street
Vancouver, British Columbia
V6C 3S5

* A computer program for calculating LC50 is available from the Environmental Toxicology Section, Pacific Environmental Science Centre, 2645 Dollarton Highway, North Vancouver, BC, V7H 1B1, by providing a computer diskette.

Members of the Inter-Governmental Environmental Toxicity Group (as of December 2004)

Federal, Environment Canada

C. Blaise
Centre St. Laurent
Montreal, Quebec

U. Borgmann
National Water Research Institute
Burlington, Ontario

J. Bruno
Pacific Environmental Science Centre
North Vancouver, British Columbia

C. Buday
Pacific Environmental Science Centre
North Vancouver, British Columbia

K. Doe
Atlantic Environmental Science Centre
Moncton, New Brunswick

G. Elliott
Environmental Protection Service
Edmonton, Alberta

F. Gagné
Centre St. Laurent
Montreal, Quebec

M. Harwood
Environmental Protection Service
Montreal, Quebec

D. Hughes
Atlantic Environmental Science Centre
Moncton, New Brunswick

P. Jackman
Atlantic Environmental Science Centre
Moncton, New Brunswick

N. Kruper
Environmental Protection Service
Edmonton, Alberta

M. Linssen
Pacific Environmental Science Centre
Norrth Vancouver, British Columbia

L. Porebski
Marine Environment Branch
Gatineau, Quebec

J. Princz
Environmental Technology Centre
Ottawa, Ontario

G. Schroeder
Pacific Environmental Science Centre
North Vancouver, British Columbia

R. Scroggins
Environmental Technology Centre
Ottawa, Ontario

T. Steeves
Atlantic Environmental Science Centre
Moncton, New Brunswick

D. Taillefer
Marine Environment Branch
Gatineau, Quebec

S. Trottier
Centre St. Laurent
Montreal, Quebec

G. van Aggelen
Pacific Environmental Science Centre
North Vancouver, British Columbia

B. Walker
Centre St. Laurent
Montreal, Quebec

P. Wells
Environmental Conservation Service
Dartmouth, Nova Scotia

Federal, Fisheries & Oceans Canada

R. Roy
Institut Maurice Lamontagne
Mont-Joli, Quebec

Federal, Natural Resources Canada

J. McGeer
Mineral Sciences Laboratory, CANMET
Ottawa, Ontario

B. Vigneault
Mineral Sciences Laboratory, CANMET
Ottawa, Ontario

J. Beyak
Mineral Sciences Laboratory, CANMET
Ottawa, Ontario

Provincial

C. Bastien
Ministère de l'Environnement du Quebec
Ste. Foy, Quebec

B. Bayer
Manitoba Environment
Winnipeg, Manitoba

D. Poirier
Ontario Ministry of Environment
Rexdale, Ontario

J. Schroeder (Chairperson)
Ontario Ministry of Environment
Rexdale, Ontario

T. Watson-Leung
Ontario Ministry of Environment
Rexdale, Ontario

Members of the Scientific Advisory Group

SAG Members

Dr. Celine Boutin
National Wildlife Research Centre
Canadian Wildlife Service (CWS)
Environment Canada
1125 Colonel By Drive, Raven Road
Carleton University
Ottawa, ON K1A 0H3
Phone: 613-998-0493
Fax: 613-998-0458
e-mail: celine.boutin@ec.gc.ca

Dr. Murray Dixon
Ontario Ministry of Environment
Biomonitoring Section
125 Resources Road,
Etobicoke, Ontario M6P 3V6
Phone: 416-314-6258
Fax: 416-413-6270
e-mail: murray.dixon@ene.gov.on.ca

Mr. Joseph Gorsuch
Eastman Kodak Company
Health, Safety, and Environment
1100 Ridgeway Avenue
Rochester, NY, USA 14652-6278
Phone: 585-588-2140
Fax: 585-722-7561
e-mail: joseph.gorsuch@kodak.com

Dr. Greg Linder
Heron Works Field Office, Route 6
USGS/BDR/CERC
5400 Tacoma Street, N.E.
Brooks, Oregon, USA 97305
Phone: 503-393-5104
Fax: 503-390-3916
e-mail: linder2@usgs.gov

Mr. Christian Kjaer
Danmarks Miljøundersøgelser (DMU)
National Environment Research Institute
Dept. Of Terrestrial Ecology
Vejlsøvej 25
P.O. Box 314
8600 Silkeborg, Denmark
Phone: 45 89 20 1461
Fax: 45 89 20 1414
e-mail: cjk@dmu.dk

Dr. Agnès Renoux
Sanexen Services Environnementaux Inc.
579, rue Le Breton
Longueuil, Québec J4G 1R9
Phone: 450-646-7878
Fax: 450-646-5127
e-mail: arenoux@sanexen.com

Dr. Larry Kaputska
Ecological Planning and Toxicology Inc.
5010 S.W. Hout Street
Corvallis, OR, USA 97333-9540
Phone: 541-752-3707
Fax: 541-753-9010
e-mail: kaputska@aol.com

Mr. Richard Petrie
Office of Pesticide Programs
Antimicrobials Division (7510C)
Risk Assessment and Science Support Branch
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.,
Washington, D.C. USA 20460
Phone: 703-305-7358
Fax: 703-308-6281
e-mail: petrie.rick@epamail.epa.gov

Dr. Thomas Pfleeger
USEPA National Health and Environmental Effects Research Lab
ORD Western Ecology Division
200 S.W. 35th Street
Corvallis, OR, USA 97333-4902
Phone: 541-754-4374
Fax: 541-754-4567
e-mail: pfleeger.thomas@epa.gov

Dr. Roxana Roshon
Stantec Consulting Ltd.,
361 Southgate Drive
Guelph, Ontario N1G 3M5
Phone: 519-836-6050 ext 309
Fax: 519-763-4419
e-mail: rroshon@esg.net

Ms. Jane Staveley
ARCADIS Geraghty and Miller, Inc.
4915 Prospectus Dr., Suite F
Durham, NC, USA 27713
Phone: 919-544-4535
Fax: 919-544-5690
email: Jstaveley@arcadis-us.com

Mr. Barry Zajdlik
Zajdlik & Associates
R.R. # 3
Rockwood, Ontario N0B 2K0
Phone: 519-856-9440
Fax: 519-856-9446
e-mail: bzajdlik@sentex.net

Consultants

Ms. Jennifer Miller
Miller Environmental Sciences Inc.
1839 Lockhart Road
Innisfil, Ontario L9S 3E9
Phone: 705-431-9127
Fax: 705-431-9128
e-mail: miller.smith@sympatico.ca

Dr. Gladys Stephenson
Aquaterra Environmental
RR1, Site 5936
Orton, ON L0N 1N0
Phone: 519-836-6050
Fax: 519-836-2493
e-mail: gstephenson@stantec.com

Scientific Authority

Mr. Richard Scroggins
Environment Canada
Biological Methods Division
Environmental Technology Centre
335 River Road South
Ottawa, ON, K1A 0H3
Phone: 613-990-8569
Fax: 613-990-0173
e-mail: rick.scroggins@ec.gc.ca

Procedural Variations for Tests of Emergence and Growth in Soil Using Terrestrial Plants, as Described in International Methodology Documents

The following source documents are listed chronologically, by originating agency rather than by author(s).

OECD 1984a – the standard guideline for testing the effect of chemicals on the growth of terrestrial plants, published by the Organization for Economic Cooperation and Development (Paris, France) in 1984.

USEPA 1989 – the protocol for performing seed germination tests with the lettuce seed (*Lactuca sativa*), published in February 1989 by the United States Environmental Protection Agency (co-authors, J.C. Greene, C.L. Bartels, W.J. Warren-Hicks, B.R. Parkhurst, G.L. Linder, S.A. Peterson, and W.E. Miller) as one of several protocols for short-term toxicity screening of hazardous waste sites.

ISO 1993a – an international standard test method for determining soil toxicity using terrestrial plants and the inhibition of root growth, published in 1993 by the International Organization for Standardization in Geneva, Switzerland.

ISO 1995 – an international standard test method for testing the effects of chemicals on the emergence and growth of higher plants, published by the International Organization for Standardization in Geneva, Switzerland.

ASTM-94 is the standard practice (E 1598-94) for conducting early seedling growth tests to assess soil toxicity, written for the American Society for Testing and Materials (ASTM) under the jurisdiction of ASTM Subcommittee E47.11 on plant toxicity and published in February 1998. In 2003, this method was withdrawn as a separate standard and was included as an annex to E 1963-98. See ASTM 1999b in References.

ASTM-98 – the standard guide (E 1963-98) for conducting terrestrial plant toxicity tests (specifically, Annex A1: Seedling emergence) to assess soil toxicity, written for the American Society for Testing and Materials (ASTM) under the jurisdiction of ASTM Subcommittee E47.11 on plant toxicity and published in February 1998. See ASTM 1999b in References.

EC 2000 – the standard operating procedure for conducting early seedling growth toxicity tests using terrestrial plants, prepared in April 2000 by D. Moul for Environment Canada's Pacific Environmental Science Centre, North Vancouver, British Columbia.

1. Test Type, Duration, and Facility

Document[1]	Test Type	Test Duration	Test Facility
OECD 1984a	static	≥14 days after 50% of seedlings have emerged in controls	phytotrons, glasshouses, plant growth chambers
USEPA 1989	static	120 h	environmental chamber[5,6,7]
ISO 1993a	static	36 to 48 h seed pre-germination 5-day exposure[2,3]	growth cabinet
ISO 1995	static	14–21 days after 50% of seedlings have emerged in controls	phytotrons, greenhouse, plant growth room
ASTM-94	static	≥21 days after 50% of control plants have emerged; ≤28 days total[4]	greenhouse, growth chamber, phytotron[5,6,7]
ASTM-98	static	double the time required to achieve acceptable percentage germination levels, adjusted to the nearest whole week[4]	greenhouse, growth chamber, phtyotron[5,6,7]
EC 2000	static	7 days[8]	environmental chamber[5,6,7]

[1] See preceding pages for complete citation information.
[2] Test duration may be adjusted to accommodate other species.
[3] Test duration should be that which is known to produce roots no longer than 80% of the depth of the soil in the pot.
[4] For example: lettuce is 90% germinated in 4 days, therefore the test would be 7 days long.
[5] Free from toxic contamination and vapours.
[6] Maintains recommended temperature.
[7] Has reasonable humidity control and supplemental lighting.
[8] May be 10 days in length for some test species.

2. Test Organisms

Document	Description of Organisms at Start of Test	Species	Number of Species for Battery	Criteria for Selection of Battery	Other Species?[1]
OECD 1984a	seeds of same class, not imbibed	see Table 1, Appendix F	≥5	≥1 from each category[2]	Yes[3]
USEPA 1989	seeds of same size class, untreated	lettuce *(Latuca sativa)*	NA[4]	NA	No
ISO 1993a	pre-germinated seeds, undressed[5]	barley *(Hordeum vulgare L.)*[6]	NA	NA	Yes[7]
ISO 1995	seeds uniform in size, undressed, not imbibed	see Table 2, Appendix F	≥2	≥1 from each category[8]	NI[9]
ASTM-94	certified seeds uniform in size, from same batch/ lot, untreated	see Table 3, Appendix F	≥5	≥3 dicots[10] ≥2 monocots[11]	Yes[12]
ASTM-98	seeds uniform in size and colour, from same batch/ lot, preferably untreated, may be field-collected	any species; lists those commonly used in other test methods (FIFRA, TSCA, FDA, OECD, APHA/AWWA, ASTM)	NI	NI	Yes
EC 2000	certified seeds, uniform in size colour, and shape from same batch/lot, untreated	any species	5	3 dicots[13] 2 monocots[14]	Yes

1 Indicates whether species other than those specified in the method may be used in a test.
2 At least one species is selected from each of three categories; see Table 1 in Appendix F.
3 Other species may be used if the rationale for their selection is justified in the test report.
4 NA = not applicable.
5 Seeds are germinated in a petri dish on filter paper moistened with distilled water until the radicle has just emerged (radicle < 2 mm in length); for barley, seed germination takes 36–48 h at 20 °C in the dark.
6 Barley variety CV Triumph is recommended; however, other varieties may be used.
7 Method may be adapted for use with other dicotyledonous species with straight roots that are easily measurable.
8 At least one species is selected from each of 2 categories (dicotyledons and monocotyledons); see Table 2 in Appendix F.
9 NI = not indicated.
10 ≥3 species from ≥2 families (one legume and one root crop) from list of dicotyledons; see Table 3 in Appendix F.
11 ≥2 species from ≥1 family, including corn, from list of monocotyledons; see Table 3 in Appendix F.
12 If species selection guidance is followed.
13 3 species from ≥2 families (one legume and one root crop); recommend lettuce (*Lactuca sativa*).
14 2 species from ≥1 family, including corn; recommend barley (*Hordeum vulgare*).

3. Description of Seed/Seedling Handling

Document	Seed Sorting	Seed Planting	Seed Culling	Seed Storing
OECD 1984a	NI[1]	NI	NI	NI
USEPA 1989	size grading using wire mesh screens; visually inspected to remove trash, empty hulls, and damaged seeds	space seeds 1.27 cm (0.5 in) from edge of test vessel; press seeds into test soil with bottom of clean beaker; pour cover sand on top of test soil	None	in airtight waterproof containers at 4 °C
ISO 1993a	NI	10 mm beneath surface of test medium	None	NI
ISO 1995	NI	NI	after assessing emergence, thin seedlings to give a total of 5 evenly spaced, representative specimens	NI
ASTM-94	NI	use template for making holes; 1.0–1.5 cm deep for small seeds; 2.5–4.0 cm for large seeds[2]; tap pots lightly to cover seeds; use of microbial inocula is optional	None	at 4 ± 2 °C
ASTM-98	size grading may be done using sieves or screens; visually inspected to remove broken or damaged seeds	template or manual holes; 1.0–1.5 cm deep for small seeds; 2.5–4.0 cm for large seeds[2]; tap pots lightly to cover seeds; use of microbial inocula is optional	None	in desiccator at (4 ± 2 °C)
EC 2000	seeds are "hand-sorted" or screened to ensure uniformity of size, colour, and shape; damaged seeds are discarded	template for small seeds; manual for larger seeds; gently cover seeds with surrounding soil	None	4 ± 2 °C[3]

[1] NI = not indicated.
[2] Seeds should be planted at a soil depth 1.5 to 2 times the seed diameter.
[3] For storage ≥2 months, replicate germination tests are performed to check viability; seeds germinating < certified % germination rate are discarded.

4. Test Vessels and Materials

Document	Test Vessel	Cover	Type of Test Soil [1,2]	Amount of Soil/Container
OECD 1984a	non-porous plastic or glazed pots of adequate size to allow unrestricted growth	NI[3]	NI	NI
USEPA 1989	bottom halves of 150 × 15 mm plastic petri dishes	35 × 35 cm (12 × 12 in) polyethylene resealable bags	AS, SWM, and mixtures thereof; SAS	100 g, dry weight aritificial or test soil; 90 g cover sand
ISO 1993a	cylindrical pots, 8 cm diameter × 11 cm high; parallel sides (not tapered); base of pot perforated (lined with filter paper)	watch-glass	AS, SS, RS, SWM, and mixtures thereof; SAS, SRS	500 g, dry weight
ISO 1995	non-porous plastic or glazed pots with top internal diameter of 85–95 mm	NI	AS, SAS, SRS, RS[4]	500 g, dry weight
ASTM-94	glass, stainless steel, or paper containers with drainage holes recommended; polyethylene or other material may be used if free from toxic materials[5]; large enough so as not to restrict seedling growth for test duration[6]	NI	AS, SWM, SAS, SRS	NI
ASTM-98	glass, stainless steel, or paper containers with drainage holes recommended; polyethylene or other material may be used if free from toxic materials[5]; large enough so as not to restrict seedling growth for test duration[6]	vessels covered during pre-germination period, but removed upon emergence	AS, RS, SWM, SS and mixtures thereof: SL, EL, SAS, SRS	100–300 g nominal dry weight
EC 2000	uncovered plastic petri dish (100 × 15 mm), placed in an 18 × 20 cm, plastic zip lock freezer bag (Glad™) that is placed inside a sealed 15 × 23 cm (6 × 9 in) 1-L all-glass parfait jar	Yes	AS, SS, SL, RS and mixtures thereof; SAS, SRS	50 g dry weight

[1] See Table 3 in this appendix for a description.
[2] AS = artificial soil; SWM = solid waste material; SS = site soil; RS = reference soil; SL = sludge/slurry; EL = eluates; SAS = spiked artificial soil; SRS = spiked reference soil.
[3] NI = not indicated.
[4] Method can be adapted for use with solid waste material, site soils, and spiked site soils.
[5] Suitability of soil medium for particular test species and conditions determined before testing.
[6] Test vessels (e.g., plant pots) are inert to test and control substances (e.g., test substance does not adhere to or react with vessel).

5. Description of Test Soils, Including Composition of Artificial Soil

Document	Description of Test Soil(s)	Description of Artificial Soil[1]
OECD 1984a	solids incorporated into soils; aqueous chemical substances mixed into soil; not necessarily sterile; <1.5% carbon content (3% organic matter); 10–20% fine particle (<20 μm)	NI[2]
USEPA 1989	solid hazardous waste (contaminated soil) or aqueous chemical substances mixed in artificial soil	20-mesh washed silica sand; cover-sand is 16 mesh sand (sieved to remove fines; 20 mesh)
ISO 1993a	reference or potentially toxic site soil; solids incorporated into soils, waste residues, or aqueous chemical substances mixed into soil; alternatively, soil diluted with reference or artificial soil	washed industrial sand or similar; particle size distribution: 10% > 0.6 mm, 80% 0.2–0.6 mm, 10% < 0.2 mm
ISO 1995	reference or potentially toxic site soil; solids incorporated into soils, waste residues, or aqueous chemical substances mixed into soil; alternatively, soil diluted with reference or artificial soil	sterile or non-sterile sieved (4–5 mm sieve) artificial soil; carbon content ≤1.5% (3% organic content); fine particles (<0.02 mm) ≤20% or dry mass[3]
ASTM-94	solids incorporated into soils; aqueous chemical substances applied to or mixed into soil; "standard soil" with <5% organic matter recommended	synthetic soil mixes (sieved, 2.0 mm), glass beads, or washed quartz sand
ASTM-98	reference or potentially toxic site soil; solids incorporated into soils; aqueous chemical substances or sludge applied to or mixed into soil; alternatively, soil diluted with reference or artificial soil	synthetic soil mixes or washed quartz sand
EC 2000	reference or potentially toxic site soil; domestic or industrial sludge; soil spiked with chemicals or soil diluted with reference or artificial soil	10% sieved (2.36-mm) sphagnum peat, 20% kaolinite clay, and 70% "grade 70" silica sand; adjust pH to 7.0 with $CaCO_3$

[1] Percentages are expressed on a dry-mass basis.

[2] NI = not indicated.

[3] Sand should be added to natural soils to bring the organic or fine particle content to within approved limits.

6. Description of Negative Control Soil and Reference Soil

Document	Description of Control Soil	Description of Reference Soil
OECD 1984a	NI[1]	NA[2]
USEPA 1989	100% artificial soil[3]	NI
ISO 1993a	reference soil and/or artificial soil, where applicable[3]	soil of same textural class and as similar as possible (without toxicants) to the test soil
ISO 1995	reference soil and/or artificial soil, where applicable[3]	soil of similar textural class as test soil (without toxicants)
ASTM-94	reference soil and/or artificial soil, where applicable[3]	natural soil (free of chemical contaminants), sieved (e.g., 2.0 mm); modified to specific soil characteristics (% clay, silt, sand, and organic matter), if necessary
ASTM-98	reference soil and/or artificial soil, where applicable[3]	natural soil (free of chemical contaminants)
EC 2000	reference soil and/or artificial soil, where applicable[3]	field-collected soil from an area that has not been cultivated or treated with pesticides or fertilizers in the past 25 years

[1] NI = not indicated.
[2] NA = not applicable.
[3] See Table 5, this appendix.

7. Storage and Characterization of Test Soil

Document	Storage Conditions	Soil Characterization
OECD 1984a	NI[1]	NI
USEPA 1989	seal in plastic (twice) and then in a pail; chill to 4 °C, ship on ice, store at 4 °C; initiate test within 24 h of collection	moisture content of site soils; water-holding capacity of artificial and site soils; pH at start and end of test
ISO 1993a	NI	NI
ISO 1995	if non-sterile, store in accordance with ISO 10381-6	NI
ASTM-94	NI	standard soil characterized: organic matter, pH_w[2], soil texture and type, cation exchange capacity, and major nutrients
ASTM-98	seal in plastic (twice) and then in a pail	water-holding capacity
EC 2000	in the dark at <8 °C	moisture content and pH

[1] NI = not indicated.
[2] pH_w = pH in water.

8. Manipulation of Soil Before Use in Test

Document	Mixing	Sample Holding Time	Hydration	pH adjustment
OECD 1984a	screened (0.5-cm mesh); use any mixing method resulting in even dispersion of test substance throughout soil; surfactants should not be used; solvent may be used[1]	test to begin <24 h after mixing test substance into soil	NI[2]	NI
USEPA 1989	homogenize solid test material with artificial soil using a blender; or hydrate artificial soil with aqueous test samples	not to exceed 36 h; test should begin ≤24 h after sample collection	hydrate to 85% of WHC with de-ionized water[3]	if pH <4 or >10[4]
ISO 1993a	artificial soil/test soil dried at 30 ± 2 °C for 16 h; sieved (4-mm sieve); homogenize solid test material with artificial soil or reference soil or hydrate test soil with aqueous test samples; solvents may be used[5,6]	NI	maintain at 70 ± 5% WHC with de-ionized water	NI
ISO 1995	any method ensuring even distribution of chemical throughout soil; homogenize solid test material with artificial soil or reference soil, or hydrate test soil with aqueous test samples; solvents may be used[5,6]	test to begin <24 h after mixing test substance into soil; if nutrients added to soil or solvent used, allow soil to equilibrate before starting test (24 h if solvent used)	as required with de-ionized water[7]	NI
ASTM-94	test substance added to test medium by mixing, spraying soil surface, or by sub-irrigation; solvents may be used[8]	NI	initially, to less than saturation	optional, if pH outside 6.0–7.5 range[9]
ASTM-98	preferable to mix test substance or or contaminated soil directly with test medium; stock solution[10] may be prepared and added to test medium; solvents may be used[11,12]	NI	hydrate to WHC of test soil with de-ionized water, at beginning of test	optional, if pH outside 6.0–7.5 range[9]
EC 2000	screen (4–9 mm) if required, in which case dry to 10–20% moisture; mix; hydrate; solvents may be used	NI	hydrate to ~35% of dry weight for each test soil, only while mixing or preparing test soils; once seeds are planted, hydrate to saturation	NI

[1] For solvent use: dissolve chemical in a volatile solvent; mix the solution with sand; let the solvent evaporate; mix the sand with soil; maintain a constant sand-soil ratio for all treatments including control.

[2] NI = not indicated.

[3] WHC = water-holding capacity.

[4] If pH range outside 4–10, results might reflect pH toxicity; altering the pH of the soil can increase or decrease (depending on contaminant) the toxicity of contaminants therein.

[5] For chemicals with low water solubility, dissolve chemical in water; mix with sand; mix treated sand with soil.

[6] If solvent required, dissolve chemical in a volatile solvent and mix with sand; dry sand with air flow and continuous mixing; mix the sand with the soil; ensure same quantity of solvent and sand are used for all treatments including control.

[7] The appropriate water-holding capacity should be predetermined and maintained throughout test (e.g., 80% for *Avena sativa* and 60% for *Brassica rapa*).

[8] Test substances with low aqueous solubility might require being dissolved in an organic solvent such as acetone. The solvent/chemical substance stock solution can be added to quartz sand or glass beads and allowed to dry. The sand and/or glass beads can then be mixed with soil for testing, or seeds can be placed in the sand or glass beads with nutrient solution.

[9] pH raised with calcium carbonate; pH lowered with sulphuric acid, gypsum, ammonium sulphate.

[10] If a stock solution is used, the concentration and stability of the test substance in the stock should be determined before the beginning of the test.

[11] The concentration of solvent in test solutions should be kept to ≤1% volume-to-volume or weight-to-volume (this does not apply to any ingredients or a formulated mixture or a commercial product).

[12] If solvent concentration is not the same in all test solutions, then a solvent test must be run or results of a previous solvent test must be available.

9. Number of Organisms per Vessel, Number of Replicates per Treatment, and, for a Multi-Concentration Test, Number of Concentrations per Sample and Recommended Dilution Factor or Application Rate

Document	Number of Seeds per Vessel	Number of Replicates per Treatment or Concentration	Number of Concentrations per Sample or Test Material	Recommended Dilution Factor/ Application Rate
OECD 1984a	≥5	≥4	3, plus control	0,1.0,10.0, and 100 mg/kg soil d.w. [1]
USEPA 1989	40	3	≥5, plus control	0.5 (e.g., 100%, 50%, 25%); d.w. hazardous waste/d.w. artificial soils, plus control (100% artificial soil)
ISO 1993a	6	3	highest concentration for test substance ≤1000 mg/kg d.w.	geometric series, ≥0.5
ISO 1995	20 (culled to 5 seedlings) [2]	4	highest concentration for test substance ≤1000 mg/kg d.w.	geometric series, ≥ 0.5
ASTM-94	≥15 per concentration	NI	≥5, plus control	NI[3]
ASTM-98	5–20[4]	5	number based on goal of study	NI
EC 2000	5–10[4]	5	≥9, plus control	NI

[1] d.w. = dry weight.
[2] See Table 3, this appendix.
[3] NI = not indicated.
[4] Number of seeds per vessel depends on size of seeds and seedling, and on test requirements.

10. Temperature, Lighting, Humidity, pH, and Watering During Test

Document	Temperature (°C)	Lighting Conditions	Humidity	pH Range	Watering
OECD 1984a	suitable for test species	suitable for test species	suitable for test species	5.0–7.5	as needed
USEPA 1989	24 ± 2 °C	dark for 48 h; then 16 h light: 8 h dark, 4300 ± 430 lux; fluorescent	NI [1]	4.0–10.0	none
ISO 1993a	20 ± 2 °C, day; 16 ± 2 °C, night	12–16 h light: 8–12 h dark; 25 000 lm/m^2	60 ± 5%	NI	maintain at 70% WHC [2]
ISO 1995	suitable for test species	suitable for test species[3]	suitable for test species	5.0–7.5	daily adjustment to pre-determined WHC
ASTM-94	air temp 20–30 °C	≥14 h light; fluorescent/ incandescent or sun; ≥300 µmol/(m^2 · s) [recommend 300– 400µmol/(m^2 · s)]	>30%; ≥50% recommended	6.0– 7.5	as required; nutrient solution used weekly if quartz sand, glass beads, or soil low in nutrients are used as soil medium
ASTM-98	suitable for test species; air temp 20–30 °C	16 h light:8 h dark; incandescent; 100–200 µmol/(m^2 · s) (PAR 400–700 ηm)[4]	>30%; ≥50% recommended	6.0–7.5	once covers are removed, water as required (at least daily) to saturation or less (e.g., 85% WHC)
EC 2000	24 ± 2 °C	16 h light: 8 h dark[5]; full spectrum (Duro-test™); 4300 ± 430 lux [765 µmol/(m^2 · s)]	NI	>4, <10	hydrated to saturation; de-ionized water sprayed onto soil surface

[1] NI = not indicated.
[2] WHC = water-holding capacity.
[3] Method recommends 16 h daylight and a minimum of 7000 lux light intensity in the photosynthetic wavelength.
[4] PAR = photosynthetically active radiation.
[5] Small seeds planted at the surface should be kept in the dark for the first 48 h.

11. Measurements and Biological Observations During Test

Document	Measurements[1]	Biological Observations
OECD 1984a	NI[2]	percent emergence per replicate[3]; wet or dry weight per replicate, expressed on per- plant basis[4]
USEPA 1989	pH of soil at start and end; soil temperature at beginning of each 24-h exposure period in each test concentration and control (1 replicate)	percent germination in each replicate[5]
ISO 1993a	NI	length of longest root for each plant at test end[6]
ISO 1995	confirmation of concentration of test chemical at test end [7]	percent emergence in each replicate[3]; mean wet or dry weight of seedling shoots per replicate, at test end [8]
ASTM-94	photoperiod; light intensity daily; continuous measurements for air temperature and relative humidity (soil temp. of representative pot); pH (pH_w)[9] when test medium is prepared and at test end	number of seedlings that emerge[10]; time to emergence during 1st week; percent survival; plant height; radicle (root) length; dry weight of above-ground vegetation and roots; severity of phytotoxicity (qualitative observations)
ASTM-98	light irradiance level at start and end of test; continuous (or at least once daily) measurements for air temperature, relative humidity, and barometric pressure; soil temperature of representative pot; soil pH or pH_w[9] when test soil medium is prepared and at test end	number of emerged seedlings[11]; qualitative abnormalities in growth, development, or morphology at test end; optional for both shoot/root length[12], and shoot/root dry weight[13]
EC 2000	continuous temperature and light; pH and conductivity at start and end, each treatment; percent moisture at start	number of seedlings emerged[14]; shoot length, and length of longest root; quantitative observations of phytotoxic effects

1 Measurements include pH (hydrogen-ion concentration), temperature, light, humidity, etc.
2 NI = not indicated.
3 Emergence = appearance of seedling above soil surface.
4 Measure wet weight of plant immediately after harvest, or dry weight after oven drying at 70 °C.
5 Germination = seedling protrudes above soil surface.
6 Shoot length may also be measured.
7 Record of temperature and humidity recommended.
8 Fresh mass weighed immediately after cutting shoots above soil surface or dry mass after oven drying at 70– 80 °C for 16 h.
9 pH_w = pH in water.
10 Emergence = the hypocotyl hook or first true leaves (coleoptile) are observed above the surface of the soil medium.
11 Emergence = epicotyl above soil surface.
12 Shoot measurements are made from the transition point between the hypocotyl and root to the tallest point of the shoot; root measurements are made from the transition point between the hypocotyl and root to the tip of the root.
13 Oven dried at 70 °C until constant weight achieved (recommend 24 h).
14 Emergence = shoot height of 3 mm above soil surface.

12. Terminating Test, Biological Endpoints, Statistical Endpoints, and Other Observations

Document	Terminating Test	Biological Endpoints	Statistical Endpoints
OECD 1984a	count the number of plants that emerge per replicate, and determine average weight of plants	percent emergence; percent growth inhibition at test end	LC50 for emergence EC50 for growth
USEPA 1989	count seedlings protruding above soil surface	percent mortality (lack of seed germination)	LC50[1]
ISO 1993a	lay pot on its side in a trough of water 5-cm deep; wash soil out of pots and wash each plant; measure the longest root to the nearest 0.5 mm	mean root length	NOEC/LOEC
ISO 1995	count the number of plants that emerge per replicate, and determine the total mass of shoots of seedlings per replicate	percent emergence; mean mass [2]	EC50
ASTM-94	count number of plants emerged; yield for each plant species is determined by harvesting the portion of each seedling above ground or below ground for roots and then oven-drying	percent emergence; mean time to emergence; mean heights and/or root lengths; mass; scores for qualitative phytotoxic effects	mean, 95% CL, and SD for each quantitative data set; NOEC/LOEC, EC50[1]
ASTM-98	count seedlings above soil surface; conduct qualitative observations and optional quantitative measurements	percent emergence; optional mean shoot/root length; mean dry weight	mean, 95% CL, and SD for each quantitative data set; NOEC/LOEC, EC50, ECx, and ICp[1]
EC 2000	count number of emerged seedlings; photograph pots to show above ground phytomass; conduct qualitative observations; separate plants from soil and wash roots to dislodge soil; measure shoot length and longest root length	percent emergence; mean shoot/root length; scores for qualitative phytotoxic effects	LC50[1], IC50, IC25, NOEC/LOEC

[1] Including the 95% confidence limits (CL).
[2] Dry mass preferred; see Table 11, this appendix.

13. Requirements for Valid Test–Use of Reference Toxicity Test

Document	Requirements for Valid Test	Reference Toxicant(s)	Procedures and Conditions for Reference Toxicity Test
OECD 1984a	≥80% of control seeds produce healthy seedlings; control seedlings exhibit normal growth throughout test	no reference substance recommended	if reference substance tested, results should be given
USEPA 1989	mean control survival ≥90%	SDS, NaPCP, or $CdCl_2$ [1]	using 100% artificial soil and test concentrations of reference toxicant diluted in the de-ionized water used to hydrate the soil, determine 120-h LC50 on each batch of seed; plot results on control chart; invalid if mean control survival <90%
ISO 1993a	NI [2]	NI	NI
ISO 1995	5 healthy seedlings per control pot	sodium trichloroacetate	reference toxicant test conducted if any major changes in procedure (e.g., test chamber, soil watering regime)
ASTM-94	mean control seedling growth does not exhibit phytotoxicity or developmental effects; ≥90% control survival through the exposure duration	no reference chemical required or recommended	NI
ASTM-98	mean control seedling growth does not exhibit phytotoxicity or developmental effects; survival during exposure period meets minimum standards for that species (≥80% mean control survival unless a lower criterion is established for the species)	boron as boric acid	a watering solution of boric acid at desired concentration is added to the test soil (0.5 dilution series with 7 concentrations from 10–640 mg/kg soil dry weight)[3]; alternate substances may be used; test invalid if mean control survival is <80%
EC 2000	≥90% control survival (pooled replicates); ≥90% control plants showing no stress (e.g., chlorosis, deformity)	potassium chloride	conduct every 2 months, with seed lots being used in tests and after the acquisition of any new seeds; plot results on warning chart

[1] SDS = sodium dodecylsulphate; NaPCP = sodium pentachlorophenate; $CdCl_2$ = cadmium chloride.
[2] NI = not indicated.
[3] Fewer concentrations may be used once the range of sensitivity for a given test species is established.

Variations in Recommended Test Species for Tests of Emergence and Growth in Soil Using Terrestrial Plants, as Described in International Methodology Documents

The following source documents are listed chronologically, by originating agency rather than by author(s).

OECD 1984a – the standard guideline for testing the effect of chemicals on the growth of terrestrial plants, published by the Organization for Economic Cooperation and Development (Paris, France) in 1984.

ISO 1995 – an international standard test method for testing the effects of chemicals on the emergence and growth of higher plants, published by the International Organization for Standardization in Geneva, Switzerland.

ASTM-94 – the standard practice (E 1598-94) for conducting early seedling growth tests to assess soil toxicity, written for the American Society for Testing and Materials (ASTM) under the jurisdiction of ASTM Subcommittee E47.11 on plant toxicity and published in February 1998. In 2003, this method was withdrawn as a separate standard and was included as an annex to E 1963-98.

1. OECD 1984a

Category	Common Name	Scientific Name
Category 1	ryegrass	*Lolium perenne*
	rice	*Oryza sativa*
	oat	*Avena sativa*
	wheat	*Triticum aestivum*
	sorghum	*Sorghum bicolor*
Category 2	mustard	*Brassica alba*
	rape	*Brassica napus*
	radish	*Raphanus sativus*
	turnip	*Brassica rapa*
	chinese cabbage	*Brassica campestris* var. Chinensis
Category 3	vetch	*Vicia sativa*
	mung bean	*Phaseolus aureus*
	red clover	*Trifolium pratense*
	fenugreek	*Trifolium ornithopodioides*
	lettuce	*Lactuca sativa*
	cress	*Lepidium sativum*

2. ISO 1995

Category	Common Name	Scientific Name
Category 1 (Monocotyledons)	rye	*Secale cereale* L.
	ryegrass (perennial)	*Lolium perenne* L.
	rice	*Oryza sativa* L.
	oat (common or winter)	*Avena sativa* L.
	wheat (soft)	*Triticum aestivum* L.
	barley (spring or winter)	*Hordeum vulgare* L.
	sorghum, common (or shattercane or durra, white or millet, great)	*Sorghum bicolor* L. Moench
	sweetcorn	*Zea mays* L.
Category 2 (Dicotyledons)	mustard, white	*Sinapis alba*
	rape [rape (summer) or rape (winter)]	*Brassica napus* L. ssp. *napus*
	radish, wild	*Raphanus sativus* L.
	turnip, wild	*Brassica rapa* ssp. (DC.) Metzg.
	chinese cabbage	*Brassica campestris* L. var. Chinensis
	birdsfoot fenugreek	*Trifolium ornithopodioides* L.
	lettuce	*Lactuca sativa* L.
	cress, garden	*Lepidium sativum* L.
	tomato	*Lycopersicon esculentum* Miller
	bean	*Phaseolus aureus* Roxb.

3. ASTM-94

Dicotyledons		
Family	**Species**	**Common Name**
Compositae	*Lactuca sativa*	lettuce
Cruciferae	*Brassica alba*	mustard
Cruciferae	*Brassica campestris* var. Chinensis	chinese cabbage
Cruciferae	*Brassica napus*	rape
Cruciferae	*Brassica oleracea*	cabbage
Cruciferae	*Brassica rapa*	turnip
Cruciferae	*Lepidium sativum*	garden cress
Cruciferae	*Raphanus sativus*	radish
Cucurbitaceae	*Cucumis sativa*	cucumber
Leguminosae	*Glycine max*	soybean
Leguminosae	*Phaseolus vulgaris*	pinto bean
Leguminosae	*Phaseolus aureus*	mung bean
Leguminosae	*Trifolium pratense*	red clover
Leguminosae	*Trifolium ornithopodioides*	fenugreek
Leguminosae	*Vicia sativa*	vetch
Solanaceae	*Lycopersicon esculentum*	tomato
Umbelliferae	*Daucus carota*	carrot
Monocotyledons		
Amaryllidaceae	*Allium cepa*	onion
Gramineae	*Avena sativa*	oat
Gramineae	*Lolium perenne*	perennial ryegrass
Gramineae	*Zea mays*	corn
Gramineae	*Oryza sativa*	rice
Gramineae	*Triticum aestivum*	wheat
Gramineae	*Sorghum bicolor*	sorghum

Natural and Artificial Negative Control Soils Used for Method Development and the Establishment of Test Validity Criteria

Negative control soil must be included as one of the experimental treatments in each soil toxicity test. This treatment requires a soil which is essentially free of any contaminants that could adversely affect the performance of plants during the test (see Section 3.4). Before applying the test method described in this document as a standardized test to be conducted according to Environment Canada, it was necessary to first assess the performance of test organisms in different types of negative control soil representative of an array of *clean* soils found within Canada. Five types of negative control soils were used to develop the biological test method described herein and to further assess the robustness of the test method with samples of soil that varied considerably in their physical and chemical characteristics. These soils were also used to establish reasonable criteria for valid test results, based on control performance. The five soils tested include an *artificial soil* (see Section 3.4.2) and four natural soils (see Section 3.4.1) (Aquaterra Environmental, 1998a; Stephenson *et al.*, 1999a, b, 2000a, b; Aquaterra Environmental and ESG, 2000; ESG, 2001, 2002; ESG and Aquaterra Environmental, 2002; EC, 2005b). The *artificial soil* was formulated in the laboratory from natural ingredients. The four natural soils included two agricultural soils from southern Ontario, a prairie soil from Alberta, and a forest soil from northern Ontario. The physicochemical characteristics of all five soils are summarized in Table G-1.

The artificial control soil (AS) used in this series of performance evaluation studies with diverse soil types was the same as that recommended for use herein (see Section 3.4.2). It consists of 70% silica sand, 20% kaolin clay, 10% *Sphagnum* sp. peat, and calcium carbonate (10–30 g $CaCO_3$/kg peat). The soil was formulated by mixing the ingredients in their dry form thoroughly, then gradually hydrating with de-ionized water, and mixing further until the soil was visibly uniform in colour, texture, and degree of wetness.

The four natural soils used as negative control soil while developing this biological test method and establishing the test validity criteria herein (see Section 4.4) do not represent all Canadian soil types. However, they do vary greatly in their physicochemical characteristics and include agricultural soils with diverse textures as well as a forest soil (see Table G-1). The soils originated from areas that had not been subjected to any direct application of pesticides in recent years. They were collected with either a shovel or a backhoe, depending on the location and the amount of soil collected. Sampling depth depended on the nature of the soil and the site itself.

The sample of clay loam soil, classified as a Delacour Orthic Black Chernozem, was collected in May 1995 from an undeveloped road allowance east of Calgary, Alberta. The soil beneath the sod was air dried to about 10–20% moisture content, sieved (4 or 9 mm), placed into 20-L plastic pails, and shipped to the University of Guelph (Guelph, ON) where it was kept in cold storage (4 °C) until needed. The soil was determined to be virtually free of any contaminants (Komex International, 1995). The physicochemical characteristics of the soil show that it is a moderate-to-fine clay loam, with a relatively high organic content and cation exchange capacity compared to the other *clean* soils used during the development of this biological test method and the establishment of test validity criteria (see Table G-1).

Table G-1 Physicochemical Characteristics of Candidate Artificial and Natural Negative Control Soils [1]

Parameter	Artificial Soil	Clay Loam	Sandy Loam	Silt Loam	Forest Soil	Analytical Method
Source	formulated from constituents	field-collected from Alberta	field-collected from Ontario	field-collected from Ontario	field-collected from Ontario	—
Soil Texture	Fine Sandy Loam	Clay Loam	Fine Sandy Loam	Silt Loam	Loam	as per Hausenbuiller (1985); based on grain size distribution
Sand (%)	77.3	26.6	60.8	36.6	48.6	gravimetric grain size distribution
Silt (%)	7.8	43.3	27.8	50.1	36.9	gravimetric grain size distribution
Clay (%)	14.9	30.1	11.4	13.3	14.5	gravimetric grain size distribution
Gravel (%)	—[2]	—	0	0	0	gravimetric grain size distribution
Very Coarse Sand (%)	—	—	1.5	1.2	0.6	gravimetric grain size distribution
Coarse Sand (%)	—	—	3.2	2.3	2.2	gravimetric grain size distribution
Medium Sand (%)	—	—	10.1	5.4	9	gravimetric grain size distribution
Fine Sand (%)	—	—	25.9	13.4	20.4	gravimetric grain size distribution
Very Fine Sand (%)	—	—	20.2	14.3	16.4	gravimetric grain size distribution
Water-holding capacity (%)	71.5	80.3	44	56.5	75.6	gravimetric analysis[3]
pH (units)	6	5.9	7.3	7.4	4.2	0.01 *M* $CaCl_2$ method[4]
Conductivity (mS/cm)	0.3	1.52	0.092	0.373	0.39	saturated paste method
Bulk Density (g/cm^3)	0.98	0.83	—	—	0.51	clod method
Total Carbon (%)	4.46	6.83	1.88	2.57	11.9	Leco furnace method
Inorganic Carbon (%)	—	—	0.18	0.58	< 0.05	Leco furnace method
Organic Carbon (%)	—	—	1.7	1.99	11.9	Leco furnace method
Organic Matter (%)	9	12.8	2.9	3.5	19.9	dichromate oxidation

Parameter	Artificial Soil	Clay Loam	Sandy Loam	Silt Loam	Forest Soil	Analytical Method
Cation Exchange Capacity ($Cmol^+/kg$)	18.5	34.5	16.1	21.9	20	barium chloride method
Total Nitrogen (%)	0.05	0.59	0.115	0.166	0.74	Kjeldahl method
NH_4-N (mg/kg)	—	—	0.53	10.25	260	Kjeldahl method
NO_3-N (mg/kg)	—	—	6.94	5.44	2.26	Kjeldahl method
NO_2-N (mg/kg)	—	—	0.94	< 0.1	< 0.1	Kjeldahl method
Phosphorus (mg/kg)	23	12	6	10	35	nitric/perchloric acid digestion
Potassium (mg/kg)	22	748	61	75	250	NH_4 acetate extraction, colourimetric analysis
Magnesium (mg/kg)	149	553	261	256	192	NH_4 acetate extraction, colourimetric analysis
Calcium (mg/kg)	1848	5127	1846	4380	963	NH_4 acetate extraction, colourimetric analysis
Chloride (mg/kg)	—	—	69	42	113	H_2O extraction, colourimetric analysis
Sodium (mg/kg)	67	57	33	19	38	NH_4 acetate extraction, colourimetric analysis

[1] Characteristics of the artificial and various negative control soils that have been used to develop the definitive biological test method and associated criteria for test validity described herein in this test method document (Aquaterra Environmental, 1998a; Stephenson *et al.*, 1999a, b, 2000a, b; Aquaterra Environmental and ESG, 2000; ESG, 2001, 2002; ESG and Aquaterra Environmental, 2002; and EC, 2005b).

[2] Not determined.

[3] Determined according to USEPA (1989) using a Fisherbrand P8 creped filter paper (see Section 5.3).

[4] Determined by Becker-van Slooten *et al.* (2004) according to Hendershot *et al.* (1993) (see Section 4.6).

A large (~3000 L) sample of sandy loam soil was collected in June 1999 from Beauchamp Farms, Eramosa, Ontario, from a site that had been cultivated regularly for crop production but not subjected to pesticide application. The soil was air-dried and sieved (2 or 5 mm), placed into 20-L plastic buckets, and kept in cold storage (4 °C) until needed. This soil was analyzed for common organic and inorganic contaminants, and its physicochemical characteristics established to determine if any unusual soil characteristics (e.g., high conductivity or anomalous nutrient levels) were present. The sample was found to be virtually free of both contaminants and anomalies. This soil is a fine sandy loam with a moderate organic content and a moderate cation exchange capacity compared to the other *clean* soils included in these studies (see Table G-1).

The sample of silt loam soil was collected in June 1999 from the University of Guelph Elora Research Station, in Nichol Township, Ontario. The topsoil had been removed several years ago when the research facility was built, and had been stockpiled beside a field. Soil collected for these method development studies was removed from the interior of the pile to avoid collecting soil that might have been inadvertently contaminated with pesticide or fertilizer spray drift from the adjacent field. The soil was air-dried and sieved (2 or 5 mm), placed into 20-L plastic buckets, and kept in cold storage (4 °C) until needed. The soil was also analyzed and found to be free of both organic and inorganic contaminants and anomalies. The measured physicochemical characteristics of this silt loam soil showed that it had a moderate organic content and a moderate cation exchange capacity, compared to the other four soils included in these method development studies (see Table G-1).

A 400-L sample of forest soil, classified as Orthic Humo-Ferric Podzols, was collected in June 2001 from a forested area located on the Canadian Shield, approximately 40 km east of Sudbury, Ontario. The leaf litter was gently raked away and a hand trowel was used to remove soil to a depth ranging from 5–10 cm. The soil was placed without sieving into 20-L plastic-lined buckets, and transported to ESG International at Guelph, Ontario. It was air-dried for 48 hours to no less than ~10% moisture content, homogenized, and then sieved through 6-mm mesh. Once the sample was sieved, it was thoroughly homogenized and stored in the same 20-L plastic buckets until used. This soil was stored at room temperature (20 °C) until used. The physicochemical characteristics of the forest soil show that it is a loam with a moderate cation exchange capacity, and the highest total organic carbon content (11.9%) and highest percentage of organic matter (19.9%) of the five soils used in the method development studies (see Table G-1).

Logarithmic Series of Concentrations Suitable for Toxicity Tests*

Column (Number of concentrations between 10.0 and 1.00, or between 1.00 and 0.10)**

1	2	3	4	5	6	7
10.0	10.0	10.0	10.0	10.0	10.0	10.0
3.2	4.6	5.6	6.3	6.8	7.2	7.5
1.00	2.2	3.2	4.0	4.6	5.2	5.6
0.32	1.00	1.8	2.5	3.2	3.7	4.2
0.10	0.46	1.00	1.6	2.2	2.7	3.2
	0.22	0.56	1.00	1.5	1.9	2.4
	0.10	0.32	0.63	1.00	1.4	1.8
		0.18	0.40	0.68	1.00	1.3
		0.10	0.25	0.46	0.72	1.00
			0.16	0.32	0.52	0.75
			0.10	0.22	0.37	0.56
				0.15	0.27	0.42
				0.10	0.19	0.32
					0.14	0.24
					0.10	0.18
						0.13
						0.10

* Modified from Rocchini *et al.* (1982).

** A series of five (or more) successive concentrations should be chosen from a column. Midpoints between concentrations in column (x) are found in column (2x + 1). The values listed can represent concentrations expressed on a weight-to-weight (e.g., mg/kg) or weight-to-volume (e.g., mg/L) basis. As necessary, values can be multiplied or divided by any power of 10. Column 1 might be used if there was considerable uncertainty about the degree of toxicity. More widely spaced concentrations should not be used, since such will provide poor resolution regarding the confidence limits surrounding any threshold-effect value calculated.

Instruction on the Derivation of ICps Using Linear and Nonlinear Regression Analyses

I.1 Introduction

This appendix provides instruction for the use of linear and nonlinear regression analyses to derive, based on the concentration-response relationships for quantitative endpoint data (in this instance, the mean length and dry mass of seedling shoots and roots), the most appropriate ICps. It represents an adaptation and modification of the approach described by Stephenson *et al.* (2000b). Instructions herein are provided using Version 11.0 of SYSTAT*; however, any suitable software may be used. The regression techniques described in this appendix are most appropriately applied to continuous data from tests designed with ten or more concentrations or treatment levels (including the negative control treatment). The test design for measuring the effects of prolonged exposure on various plant species is summarized in Table I.1.

An overview of the general process used to select the most appropriate regression model for each data set under consideration is presented in Figure 3 within the main text (see Section 4.8.3.1).

The reader is encouraged to refer to the appropriate sections within this biological test method document, as well as the sections on regression analyses within the "Guidance Document on Statistical Methods for Environmental Toxicity Tests" (EC, 2004a) before data analyses. Environment Canada (2004a) also contains several additional references for the statistical analysis of quantitative test data using linear and nonlinear regression procedures. Some of the related guidance from these documents has been provided in this appendix, where appropriate.

* The latest (e.g., Version 11.0) version of SYSTAT™ is available for purchase by contacting SYSTAT Software, Inc., 501 Canal Boulevard, Suite C, Point Richmond, CA 94804-2028, USA, phone no. 1-800-797-7401; see Web site www.systat.com/products/Systat/.

Table I.1 Summary of Test Design for Environment Canada's Biological Test Method for Measuring Effects of Exposure on the Emergence and Growth of Various Plant Species

Parameter	Description
Test type	whole soil toxicity test; no renewal (static test)
Test duration	14 days for barley, cucumber, durum wheat, lettuce, radish, red clover, and tomato; 21 days for alfalfa, blue grama grass, carrot, northern wheatgrass, and red fescue
Test species	for monocotyledons: choose barley *(Hordeum vulgare)*, blue grama grass *(Bouteloua gracilis)*, durum wheat *(Triticum durum)*, northern wheatgrass *(Elymus lanceolatus*; formerly named *Agropyron dasystachyum)*, or red fescue *(Festuca rubra)*; for dicotyledons: choose alfalfa *(Medicago sativa)*, carrot *(Daucus carota)*, cucumber *(Cucumis sativus)*, lettuce *(Lactuca sativa)*, radish *(Raphanus sativus)*, red clover *(Trifolium pratense)*, or tomato *(Lycopersicon esculentum)*
Number of replicates	≥4 replicates/treatment if equal replicate test design; if unequal replicates among test treatments, use: - 6 replicates per control treatment - 4 replicates for each of the lowest 4 to 6 test concentrations - 3 replicates for the highest 5 test concentrations
Number of treatments	negative control soil and ≥9 test concentrations as a minimum; however, ≥11 concentrations plus a negative control are strongly recommended
Statistical endpoints	Quantal: • mean percent emergence in each treatment • 14- or 21-day EC50 calculated, data permitting (using appropriate statistical procedures for quantal data; the procedures outlined in this appendix are not appropriate for quantal data) Quantitative: • mean shoot and root length and dry mass in each treatment, on Day 14 or 21 • ICp (e.g., IC25 and/or IC20) for the mean shoot and root length and shoot and root dry mass

I.2 Linear and Nonlinear Regression Analyses

I.2.1 Creating Data Tables

Note: The statistical analysis must encompass the transformation of the concentrations logarithmically (e.g., $\log_{10}$ or $\log_e$). If the concentrations fall below one (1) (e.g., 0.25), then the data can be transformed by transforming the units (e.g., from mg/kg to µg/g) with a multiplication factor (e.g., 1000); the modified data are then transformed logarithmically. The transformation can be done either in the original electronic spreadsheet, or when the original data are transferred to the SYSTAT data file.

(1) Open the appropriate file containing the data set in an electronic spreadsheet.

(2) Open the SYSTAT program. In the main screen, go to **File**, **New**, and then **Data**. This will open up an empty data table. Insert the variable names into the column heading by double-clicking on a variable name,

which opens the '**Variable Properties**' window. Insert an appropriate name for the variable of interest within the '**Variable name**' box, and select the variable type; additional comments can be inserted within the '**Comments:**' box. For example, the following variable names might be used:

conc	=	concentration or treatment level
logconc	=	$\log_{10}$ transformation of concentration or treatment level
rep	=	replicate within a treatment level
mnlengths	=	mean length of shoots
mnlengthr	=	mean length of roots
drywts	=	dry weight of shoots
drywtr	=	dry weight of roots

(3) The data can now be transferred. To transfer the data, copy and paste each column from the electronic spreadsheet containing the concentrations, the replicates, and associated mean values, to the SYSTAT data table.

(4) Save the data by going to **File**, then **Save As**; a **'Save As**' window will appear. Use appropriate coding to save the data file. Select **Save** when the file name has been entered.

(5) Record the file name of the SYSTAT data file in the electronic spreadsheet containing the original data.

(6) If the data (i.e., the test concentrations) require transformation, the data can be transformed by selecting **Data**, **Transform**, and then **Let...**. Once in the **Let...** function, select the column heading containing the appropriate header for the transformed data (e.g., logconc), and then select **Variable** within the '**Add to**' box to insert the variable into the '**Variable:**' box. Select the appropriate transformation (e.g., L10 for $\log_{10}$ transformation or LOG for the natural logarithm) in the '**Functions:**' box (the '**Function Type:**' box should be **Mathematical**), and then select **Add** to insert the function into the '**Expression:**' box. Select the column heading containing the original untransformed data (i.e., 'conc' for concentration or treatment level), followed by **Expression** within the '**Add to**' box to insert the variable into the '**Expression:**' box. If a multiplication factor is required to adjust the concentration before the log-transformation, this step can be completed within the '**Expression:**' box (e.g., L10[conc*1000]). Select **OK** when all desired transformations have been completed. The transformed data will appear in the appropriate column. *Save the data* (i.e., select **File**, followed by **Save**).

Note: The $\log_{10}$ of the negative control treatment cannot be determined (i.e., the $\log_{10}$ of zero is undefined); therefore, assign the negative control treatment level a very small number (e.g., 0.001) known or assumed to be a no-effect level, to include this treatment in the analysis and differentiate it from the other transformed treatment levels.

(7) From the data table, calculate and record the mean of the negative controls for the variable under study; each measurement endpoint is statistically analyzed independently. The mean value of these control data will be required when estimating the model parameters. In addition, determine the maximum value within the data set for that particular variable and round up to the nearest whole number. This number is used as the maximum value of the y-axis (i.e., 'ymax') when creating a graph of the regressed data.

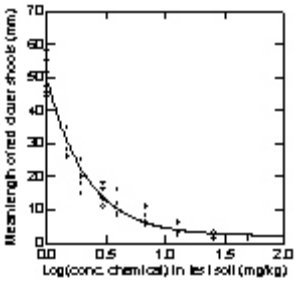
I.2.2 Creating a Scatter Plot or Line Graph

The scatter plots and line graphs provide an indication of the shape of the concentration-response curve for the data set. The shape of the concentration-response curve can then be compared to each model (Figure I.1) so that the appropriate model(s) likely to best suit the data is (are) selected. Each of the selected models should be used to analyze the data. Subsequently, each model is reviewed, and the model that demonstrates the best fit is selected.

Exponential Model

IC50: mnlengths = a*exp(log((a-a*0.5-b*0.5)/a)*(logconc/x))+b
IC25: mnlengths = a*exp(log((a-a*0.25-b*0.75)/a)*(logconc/x))+b

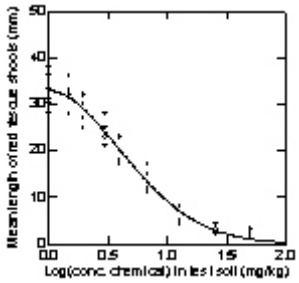
Where:
a = the y-intercept (the control response)
x = ICp for the data set
logconc = the logarithmic value of the exposure concentration
b = a scale parameter (estimated between 1 and 4)



Gompertz Model

IC50: mnlengths = g*exp((log(0.5))*(logconc/x)^b)
IC25: mnlengths = g*exp((log(0.75))*(logconc/x)^b)

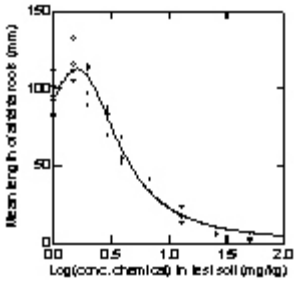
Where:
g = the y-intercept (the control response)
x = ICp for the data set
logconc = the logarithmic value of the exposure concentration
b = a scale parameter (estimated between 1 and 4)



Hormesis Model

IC50: mnlengthr = (t*(1+h*logconc))/(1+((0.5+h*logconc)/0.5)*(logconc/x)^b)
IC25: mnlengthr = (t*(1+h*logconc))/(1+((0.25+h*logconc)/0.75)*(logconc/x)^b)

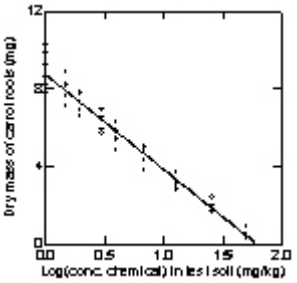
Where:
t = the y-intercept (the control response)
h = the hormetic effect (estimated between 0.1 and 1)
x = ICp for the data set
logconc = the logarithmic value of the exposure concentration
b = a scale parameter (estimated between 1 and 4)



Linear Model

IC50: drywtr = ((-b*0.5)/x)*logconc+b
IC25: drywtr = ((-b*0.25)/x)*logconc+b

Where:
b = the y-intercept (the control response)
x = ICp for the data set
logconc = the logarithmic value of the exposure concentration



Logistic Model

IC50: drywts = t/(1+(logconc/x)^b)
IC25: drywts = t/(1+(0.25/0.75)*(logconc/x)^b)

Where:
t = the y-intercept (the control response)
x = ICp for the data set
logconc = the logarithmic value of the exposure concentration
b = a scale parameter (estimated between 1 and 4)

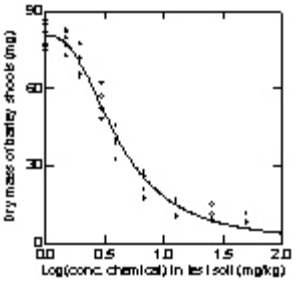


Figure I.1 SYSTAT Version 11.0 Equations for Linear and Nonlinear Regression Models and Example Graphs of the Observed Trends for Each Model

"mnlengths" and "mnlengthr" refer to the mean length of shoots or roots, respectively, and "drywts" and "drywtr" refer to the individual mean shoot or root dry weight, respectively.

(1) Select **Graph**, **Summary Charts**, and then **Line...**. Select the independent variable (e.g., logconc), followed by **Add** to insert the variable into the '**X-variable(s):**' box. Select the dependent variable under examination, followed by **Add** to insert the variable into the '**Y-variable(s):**' box. Select **OK**. A graph will be displayed within the '**Output Pane**' of the main SYSTAT screen containing the mean values for every treatment level; to view a larger version of the graph, simply select the '**Graph Editor**' tab located below the central window. A scatter plot of the data can also be viewed by selecting **Graph**, **Plots**, and then **Scatterplot...** and following the same instructions for inserting the x- and y-variables. The graphs will provide an indication as to the general concentration-response trend allowing the selection of the potential model(s) of best fit to be chosen, in addition to an estimation of the ICp of interest.

Note: The main SYSTAT screen is divided into three parts. The left-hand side of the screen (i.e., '**Output Organizer**' tab) provides a list of all of the functions completed (e.g., graphs) – each function can be viewed by simply selecting the desired icon. The right-hand side of the screen forms the central window in which the general output of all of the functions completed (e.g., regression, graphs) can be viewed. The tabs below this central window allow the user to toggle between the data file (i.e., '**Data Editor**'), individual graphs (i.e., '**Graph Editor**') and the output (i.e., '**Output Pane**'). The various graphs produced can be viewed individually within the '**Graph Editor**' tab by selecting the graph of interest within the left-hand side of the screen (i.e., '**Output Organizer**' tab). The bottom portion of the screen displays the command codes used to derive the desired functions (e.g., regression and graphing codes). The '**Log**' tab within this command screen displays a history of all of the functions that have been completed.

(2) Visually estimate and record an estimate of the ICp of interest (e.g., IC25) for the data set. For example, for an IC25, divide the average of the controls by four, and find this value on the y-axis. Estimate a horizontal line from the y-axis until the line intercepts the data points. At this intersection point, extend a vertical line down towards the x-axis and record this concentration value as an estimate of the IC25.

(3) Using the scatter plots or line graphs, select the potential model(s) that will best describe the concentration-response trend (refer to Figure I.1 for an example of each model).

I.2.3 Estimating the Model Parameters

(1) Select **File**, **Open**, and then **Command**.

(2) Open the file containing the command codes for the particular model chosen from Section I.2.2 (i.e., select the appropriate file, followed by **Open**):

nonline.syc	=	exponential model
nonling.syc	=	gompertz model
nonlinh.syc	=	logistic with hormesis model
linear.syc	=	linear model
nonlinl.syc	=	logistic model

The file will provide the command codes for the selected model within the appropriate tab of the command editor box at the bottom of the main screen. All of the command codes for deriving IC50s and IC25s are provided in Table I.2; however, the equations can be formatted to derive any ICp. For example, the command codes for the logistic model to derive an IC25 would be:

```
nonlin
print          = long
model drywts   = t/(1+(0.25/0.75)*(logconc/x)^b)
```

Table I.2 SYSTAT Command Codes for Linear and Nonlinear Regression Models

Model	Command Codes	
Exponential	nonlin print = long model mnlengths = a*exp(log((a-a*0.25-b*0.75)/a)*(logconc/x))+b save resid1/ resid estimate/ start = 25[a], 1[b], 0.3[c] iter = 200 use resid1 pplot residual plot residual*logconc plot residual*estimate	where: [a]Represents the estimate of the y-intercept (i.e., 'a') (the control response) [b]Represents the scale parameter (i.e., 'b') (estimated between 1 and 4) [c]Represents the estimate of the ICp for the data set (i.e., 'x')
Gompertz	nonlin print = long model mnlengths = g*exp((log(0.75))*(logconc/x)^b) save resid2/ resid estimate/ start = 16[a], 0.8[b], 1[c] iter = 200 use resid2 pplot residual plot residual*logconc plot residual*estimate	where: [a]Represents the estimate of the y-intercept (i.e., 'g') (the control response) [b]Represents the estimate of the ICp for the data set (i.e., 'x') [c]Represents the scale parameter (i.e., 'b') (estimated between 1 and 4)
Hormesis	nonlin print = long model mnlengthr = (t*(1+h*logconc))/(1+((0.25+h*logconc)/ 0.75)*(logconc/x)^b) save resid3/ resid estimate/start = 48[a], 0.1[b], 0.7[c], 1[d] iter = 200 use resid3 pplot residual plot residual*logconc plot residual*estimate	where: [a]Represents the estimate of the y-intercept (i.e., 't') (the control response) [b]Represents the hormetic effect (i.e., 'h') (estimated between 0.1 and 1) [c]Represents the estimate of the ICp for the data set (i.e., 'x') [d]Represents the scale parameter (i.e., 'b') (estimated between 1 and 4)
Linear	nonlin print = long model drywtr = ((-b*0.25)/x)*logconc+b save resid4/ resid estimate/start = 5[a], 0.7[b] iter = 200 use resid4 pplot residual plot residual*logconc plot residual*estimate	where: [a]Represents the estimate of the y-intercept (i.e., 'b') (the control response) [b]Represents the estimate of the ICp for the data set (i.e., 'x')
Logistic	nonlin print = long model drywts = t/(1+(0.25/0.75)*(logconc/x)^b) save resid5/resid estimate/start = 85[a], 0.6[b], 2[c] iter = 200 use resid5 pplot residual plot residual*logconc plot residual*estimate	where: [a]Represents the estimate of the y-intercept (i.e., 't') (the control response) [b]Represents the estimate of the ICp for the data set (i.e., 'x') [c]Represents the scale parameter (i.e., 'b') (estimated between 1 and 4)

Note: "mnlengths" and "mnlengthr" refer to mean length of shoots or roots, respectively; "drywts" and "drywtr" refer to mean individual shoot or root dry weight, respectively; "pplot" refers the creation of a probability plot based on the residuals derived from the regression model under study.

```
save resid1/ resid
estimate/ start      =   85, 0.6, 2 iter  =  200
use resid1
pplot residual
plot residual*logconc
plot residual*estimate
```

(3) Type in the header of the column in the data table containing the variable of interest to be analyzed within the line entitled 'model y=' (where 'y' is the dependent variable, e.g., drywts).

(4) The 4th line of the text should read 'save resida/ resid', where 'a' indicates a number to which the residual file is assigned. Substitute this same number into the 6th line (i.e., 'use resida') so that the same file is used to generate a normal probability plot and graphs of the residuals. The command lines that follow provide instruction for the generation of a probability plot (i.e., 'pplot residual'), the generation of a graph of residuals against the concentration or treatment level (i.e., 'plot residual*logconc'), and a graph of the residuals against the predicted and fitted values (i.e., 'plot residual*estimate'). These graphs are used to aid in the assessment of the assumptions of normality (e.g., probability plot) and homogeneity of the residuals (e.g., graphs of the residuals) when evaluating for the model of best fit (Section I.2.4).

(5) Substitute the mean of the controls and the estimated ICp (e.g., IC25) within the fifth line entitled 'estimate/start=' (refer to Table I.2 for details on the substitution for each model). These values were initially derived from examination of the scatter plot or line graph. The model, once it converges, will provide a set of parameters from which the ICp, and its 95% confidence limits, are reported (i.e., parameter 'x'). It is essential that accurate estimates for each parameter be provided before running the model, or the iterative procedure used to derive the reported parameters might not converge. The scale parameter (Table I.2) is typically estimated to range from one to four. The number of iterations can be changed, but for this example, has been set to 200 (i.e., 'iter = 200'). Typically, 200 iterations are sufficient for a model to converge; if more iterations are required, it is likely that the most appropriate model is not being applied.

(6) Select **File**, and then **Submit Window** to run the commands; alternatively, right-click the mouse and select **Submit Window**. This will generate a printout of the iterations, the estimated parameters, and a list of the actual data points with the corresponding predicted values and residuals. A preliminary graph of the estimated regression line will also be presented; this preliminary graph should be deleted. The graph can be deleted by selecting the graph in the left-hand window within the main screen. A normal probability plot and graphs of the residuals will also be presented.

I.2.4 Examining the Residuals and Test Assumptions

An examination of the residuals for each model tested helps to determine whether assumptions of normality and homoscedasticity have been met. If any of the assumptions cannot be met, regardless of the model examined, a statistician should be consulted for further guidance on assessing additional models or the data should be re-analyzed using the less desirable linear interpolation method of analysis (using ICPIN; see Section 4.8.3.2).

I.2.4.1 Assumptions of normality. Normality should be assessed using *Shapiro-Wilk's test* as described in EC (2004a); Section I.2.4.3 provides instructions for conducting this test. The normal probability plot, displayed in the '**Output Pane**', can also be used to evaluate whether the assumption of normality is met. The residuals should form a fairly straight line diagonally across the graph; the presence of a curved line represents deviation from normality. The normal probability plot should not, however, be used as a stand-alone test for normality, since the detection of a 'normal' (e.g., straight) or 'non-normal' (e.g., curved) line depends on the subjective assessment of the user. If the data are not normally distributed, then the user should try another model, or

should consult a statistician for further guidance or the data should be analyzed using the less desirable linear interpolation method of analysis.

I.2.4.2 Homogeneity of residuals. Homoscedasticity (or homogeneity) of the residuals should be assessed using *Levene's test* as described in EC (2004a) (Section I.2.4.3 provides instructions for conducting this test), and by examining the graphs of residuals against the actual and predicted (estimated) values. Homogeneity of the residuals is described by an equal distribution of the variance of the residuals across the independent variable (i.e., concentration or treatment level) (Figure I.2A). Levene's test, if significant, will indicate that the data are not homogeneous. If the data (as indicated by Levene's test) are heteroscedastic (i.e., not homogeneous), then the graphs of the residuals should be examined. If there is a significant change in the variance and the graphs of the residuals produce a distinct fan or 'V' pattern (refer to Figure I.2B for a plot of the 'residual*estmate'; a corresponding 'V' pattern in the opposite direction also occurs in the plot of the 'residual*logconc'), then the data analysis should be repeated using weighted regression. Alternatively, a divergent pattern suggestive of a systematic lack of fit (Figure I.2C) will indicate that an inappropriate or incorrect model was selected.

I.2.4.3 Assessing assumptions of normality and homogeneity of residuals. SYSTAT Version 11.0 can perform both Shapiro-Wilk's and Levene's tests to assess the assumptions of normality and homogeneity of residuals. Levene's test can only be performed by conducting an analysis of variance (ANOVA) on the absolute values of the residuals derived in Section I.2.3.

(1) Select **File**, **Open**, and then **Data** to open the data file containing the residuals created in Section I.2.3 (e.g., resid1.syd).

(2) Insert a new variable name into an empty column by double-clicking on the variable name, which opens the '**Variable Properties**' window. In this window, insert an appropriate name for the transformed residuals (e.g., absresiduals) into the '**Variable name:**' box. Transform the residuals by selecting **Data**, **Transform**, and then **Let...**. Once in the **Let...** function, select the column heading containing the appropriate header for the transformed data (e.g., absresiduals), and then select **Variable** within the '**Add to**' box to insert the variable into the '**Variable:**' box. Select the appropriate transformation (e.g., ABS for the transformation of data into its absolute form) in the '**Functions:**' box (the '**Function Type:**' box should be **Mathematical**), and then select **Add** to insert the function into the '**Expression:**' box. Select the column heading containing the original untransformed data (i.e., residuals), followed by **Expression** within the '**Add to**' box to insert the variable into the '**Expression:**' box. Select **OK**; the transformed data will appear in the appropriate column. Save the data.

(3) To perform Shapiro-Wilk's test, select **Analysis**, **Descriptive Statistics**, and then **Basic Statistics...**. A '**Column Statistics**' window will appear. Select the residuals from the '**Available variable(s):**' box, followed by **Add** to insert this variable into the '**Selected variable(s):**' box. Within the '**Options**' box, select the **Shapiro-Wilk normality test**, followed by **OK**. A small table will appear within the SYSTAT Output Organizer window, where the Shapiro-Wilk critical value (i.e., 'SW Statistic') and probability value (i.e., SW P-Value') will be displayed. A probability value greater than the usual criterion of $p > 0.05$ indicates that the data are normally distributed.

(4) To perform Levene's test, select **Analysis**, **Analysis of Variance (ANOVA)**, and then **Estimate Model...**, an '**Analysis of Variance: Estimate Model**' window will appear.

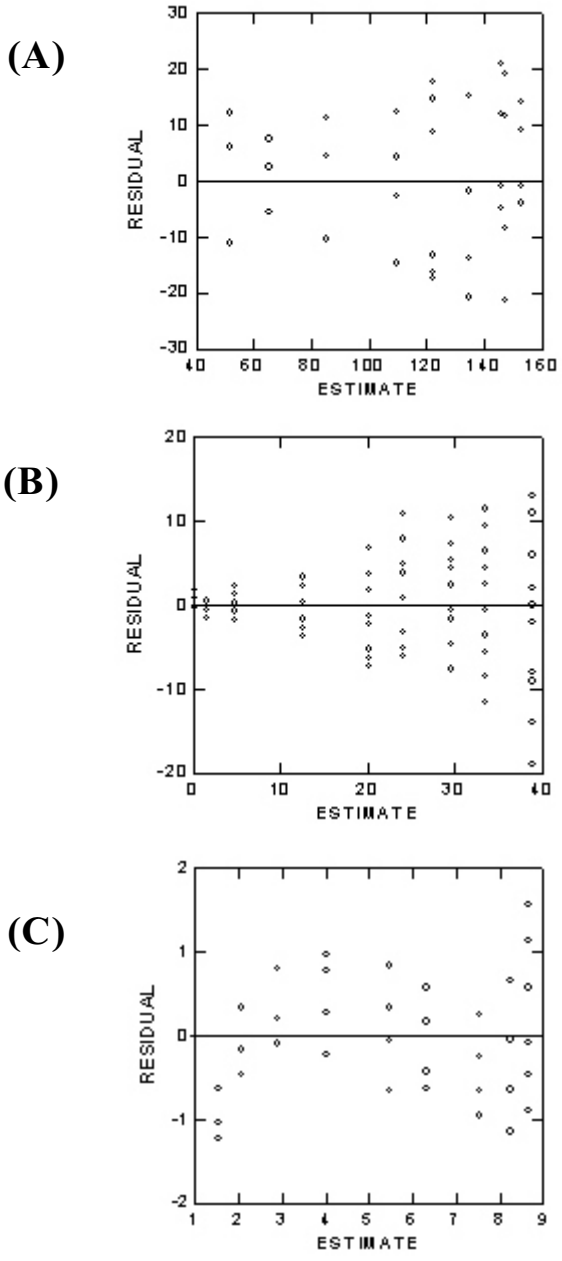


Figure I.2 Graph of the Residuals Against the Predicted (Estimated) Values (i.e., 'residuals*estimate') Indicating Homoscedasticity (A), and Two Types of Heteroscedasticity; One Demonstrating a Fan or 'V' Shape (B) Requiring Further Examination Using Weighted Regression, and a Second Demonstrating a Systematic Lack of Fit (C) as a Result of the Selection of an Incorrect Model

(5) Select the variable within which the data are to be grouped (e.g., logconc), and place this variable into the '**Factor(s):**' box by selecting **Add**.

(6) Select the transformed residuals (i.e., absresiduals), followed by **Add**, to insert the variable into the '**Dependent(s):**' box. Select **OK**. A graph of the data and a printout of the output will appear within the '**Output Pane**' tab. A probability value greater than the usual criterion of $p > 0.05$ indicates that the data are homogeneous.

I.2.5 Weighting the Data

If the residuals are heteroscedastic, as indicated by Levene's test, and there is a significant change in variance across treatment levels (i.e., the presence of a distinct fan or 'V' shape; refer to Figure I.2B), the data should be re-analyzed using weighted regression. Weighted regression involves using the inverse of the variance of observations within each concentration or treatment level as the weights. When performing the weighted regression, the standard error for the ICp (presented in SYSTAT as the asymptotic standard error ('A.S.E.'; refer to Figure I.3) is compared to that derived from the unweighted regression. If there is a difference of greater than 10% between the two standard errors, then the weighted regression is selected as the regression of best choice. However, if there is a significant change in variance across all treatment levels, and there is less than a 10% difference in the standard error between the weighted and unweighted regressions**, then the user should consult a statistician for further guidance and the application of additional models, or the data could be re-analyzed using the less desirable linear interpolation method of analysis. The comparison between weighted and unweighted regression is completed for each of the selected models while proceeding through the process of final model selection (i.e., model and regression of best choice). Alternatively, if Levene's test demonstrates that the data are not homogeneous, and the graphs of the residuals demonstrate a non-divergent pattern (e.g., Figure I.2C), an inappropriate or incorrect model might have been selected. The user is then advised to consult a statistician for further guidance on the use and application of alternate models.

(1) Select **File**, **Open**, and then **Data**. Select the file containing the data set to be weighted. Insert the two new variable names into the column heading by double-clicking on a variable name, which opens the '**Variable Properties**' window. In this window, insert an appropriate name for the variable of interest, select the variable type, and specify comments if desired. The two new column headings should indicate the variance of a particular variable (e.g., vardrywts), and the inverse of the variance for that variable (e.g., varinvsdrywts). Save the data file by selecting **File**, and then **Save**.

(2) Select **Data**, followed by **By Groups...**. Select the independent variable (i.e., logconc), followed by **Add**, to insert this variable into the '**Selected variable(s):**' box; this will enable the determination of the variance of the variable of interest by concentration or treatment level (i.e., "group"). Select **OK**.

(3) Select **Analysis**, **Descriptive Statistics**, and then **Basic Statistics...**. Select the variable of interest to be weighted (e.g., drywts), followed by **Add** to insert this variable into the '**Selected variable(s):**' box. Select **Variance** within the '**Options**' box, followed by **OK**. This function will display the variance for the variable of interest, grouped by concentration or treatment level within the '**Output Pane**' tab of the main screen.

** The value of 10% is only a "rule-of-thumb" based upon experience. Objective tests for the improvement due to weighting are available, but beyond the scope of this document. Weighting should be used only when necessary, since the procedure can introduce additional complications to the modeling procedure. A statistician should be consulted when weighting is necessary, but the parameter estimates are nonsensical.

```
SYSTAT Rectangular file C:\SYSTAT\STATAPP.SYS,
created Tue June 11, 2004 at 13:46:14, contains variables:
           CONC         REP          LOGCONC      MNLENGTHS      MNLENGTHR      DRYWTS
           DRYWTR

 Iteration
 No.      Loss      T           X           B
   0 .366675D+04 .340000D+02 .600000D+00 .200000D+01
   1 .957667D+04 .804856D+02 .641058D+00 .307486D+01
   2 .719588D+04 .805783D+02 .612443D+00 .249296D+01
   3 .717477D+04 .807441D+02 .610962D+00 .252952D+01
   4 .717464D+04 .807415D+02 .610862D+00 .253311D+01
   5 .717464D+04 .807413D+02 .610849D+00 .253341D+01
   6 .717464D+04 .807413D+02 .610848D+00 .253344D+01

Dependent variable is DRYWTS

    Source   Sum-of-Squares    df  Mean-Square
 Regression      103462.776     3    34487.592
   Residual         717.464    32       22.421  ------->  residual mean square error

      Total      104180.240    35
Mean corrected    27222.027    34

       Raw  R-square (1-Residual/Total)        =
   0.993
Mean corrected R-square (1-Residual/Corrected) =        0.974
          R(observed vs predicted) square      =        0.974

                                                      Wald Confidence Interval
Parameter         Estimate       A.S.E.    Param/ASE        Lower < 95%> Upper
 T                  80.741        1.641       49.216       77.400       84.083
 X                   0.611        0.021       29.652        0.569        0.653
 B                   2.533        0.187       13.554        2.153        2.914

             DRYWTS       DRYWTS
  Case     Observed    Predicted      Residual
      1      76.300       80.741        -4.441
      2      84.300       80.741         3.559
      3      74.800       80.741        -5.941
      4      81.500       80.741         0.759
      5      76.800       80.741        -3.941
      6      86.100       80.741         5.539
      7      76.500       77.427        -0.927
      .      ......       ......        ......
      .      ......       ......        ......
      .      ......       ......        ......
      .      ......       ......        ......
     31      14.800        8.590         6.210
     32       8.700        8.590         0.110
     33      11.200        5.560         5.640
     34       8.400        5.560         2.840
     35       7.800        5.560         2.240

Asymptotic Correlation Matrix of Parameters
```

ICp, asymptotic standard error, and lower and upper 95% confidence limits

	T	X	B
T	1.000		
X	-0.641	1.000	
B	-0.466	0.359	1

Figure I.3 Example of the Initial Output Derived using the Logistic Model in SYSTAT Version 11.0. The initial output provides the residual mean square error used to select the model of best choice, as well as the ICps, the standard error for the estimate, and the upper and lower 95% confidence limits. The number of cases displayed has been shortened for the purpose of this diagram; however, the output within SYSTAT displays all cases including the actual variable measurement and the corresponding predicted estimate and residual.

(4) Select **Data**, **By Groups...**, and then click on the box beside **Turn off**, and select **OK** so that any analysis that follow will not be analyzed according to each individual concentration or treatment level; the analysis should consider the entire data set as a whole.

(5) Return to the data file by selecting the '**Data Editor**' tab within the main screen. Transfer the variances for each concentration or treatment level to the corresponding concentration within the variance column (e.g., vardrywts). Note that the variance is the same among replicates within a treatment.

(6) Select **Data**, **Transform**, and then **Let...**, and select the column heading containing the inverse of the variance (e.g., varinvsdrywts) for the variable of interest, followed by **Variable** within the **'Add to**' box to insert the variable into the '**Variable:**' box. Select the '**Expression:**' box and type in '1/', and then select the column heading containing the variances (e.g., vardrywts) of the variable of interest for each replicate and concentration, followed by **Expression** within the '**Add to**' box to insert the variable into the '**Expression:**' box. Select **OK**. The inverse of the variance for each replicate and concentration will be displayed in the appropriate column. Save the data by selecting **File**, and then **Save**.

(7) Select **File**, **Open**, and then **Command**; open the file containing the command codes for estimating the equation parameters (e.g., Section I.2.3, step 2) for the same model selected for the *unweighted* analysis.

(8) Insert an additional row after the third line by typing 'weight=varinvsy', where 'y' is the dependent variable to be weighted (e.g., weight=varinvsdrywts), as per the shaded area below:

```
nonlin
print=long
model drywts = t/(1+(0.25/0.75)*(logconc/x)^b)
weight=varinvsdrywts
save resid2/ resid
estimate/ start = 85, 0.6, 2 iter=200
use resid2
pplot residual
plot residual*logconc
plot residual*estimate
```

(9) Assign a new number for the residuals within the line entitled 'save resida' (where 'a' represents the assigned number).

(10) Substitute the mean of the controls and the estimated ICp within the line entitled 'estimate/ start . . .' (refer to Table I.2 for details on the substitution for each model). These estimates will be the same as those used for the unweighted analysis.

(11) Select **File**, and then **Submit Window** to run the commands. This will generate output of the iterations, the estimated parameters, and a list of the data points with the corresponding predicted data points and residuals within the '**Output Pane**' tab of the main screen. A preliminary graph of the estimated regression line will also be presented; this should be deleted. A normal probability plot and graphs of the residuals will also be presented.

(12) Proceed with the analysis as described in Section I.2.4 to ensure that all model assumptions have been met.

(13) Compare the weighted regression analysis with the unweighted regression analysis. Select the weighted regression if weighting reduced the standard error for the ICp by 10%, relative to the unweighted regression analysis.

I.2.6 The Presence of Outlier(s) and Unusual Observations

Outliers are indicative of a measurement that does not seem to fit the other values derived from the test. Outliers and unusual observations can be identified by examining the fit of the concentration-response curve relative to all data points, and by examining the graphs of the residuals. If an outlier has been observed, the test records (e.g., hand-recorded and electronic data sheets and experimental conditions) should be scrutinized for human error. If the outlier is a data point that has been obtained through a transcription error that cannot be corrected, or through a faulty procedure, then the data point should be removed from the analysis. If an outlier has been identified, the analysis should be completed with and without the presence of the outlier. The decision on whether or not to remove the outlier should also take into consideration natural biological variation, and biological reasons that might have caused the apparent anomaly. Regardless of whether or not the outlier is removed, a description of the data, outliers, analyses with and without the outlier, and interpretive conclusions, must accompany the final analysis. If it appears as if there is more than one outlier present, the selected model should be re-assessed for appropriateness and alternative models considered. Additional guidance on the presence of outliers and unusual observations is provided in EC (2004a) and should be consulted for further details.

The Analysis of Variance (ANOVA) function within SYSTAT can be performed to determine whether or not the data contain outliers. However, ANOVA assumes that the residuals are normally distributed, and therefore, assumptions of normality must be met before to using the ANOVA to detect outliers. The presence of outliers can also be determined from the graphs of residuals.

(1) Perform an Analysis of Variance (ANOVA) as described in Section I.4 of this appendix, to determine whether any outliers exist. Any outlier(s) will be identified as a case number that corresponds with the row number in the SYSTAT data file. The program uses the studentized residuals as an indication of outliers; values >3 indicate the possibility of an outlier. This should be confirmed with the graphs of the residuals.

(2) If a decision is made to remove the outlier(s), delete the value from the original data table (file), and re-save the file under a *new* name (i.e., select **File**, and then **Save As...**). For example, the new file name might contain the letter 'o' (for outlier(s) removed) at the end of the file's original name.

(3) Repeat the regression analysis with the outlier(s) removed, using the same model and estimated parameters that were used before the outlier(s) were removed. Alternatively, additional models may be used for analysis if the alternative model results in a better fit and smaller residual mean square error. If the removal of the outlier(s) does not result in a significant change to both the residual mean square error and the ICp (including its corresponding confidence intervals), then the individual performing the analysis must make a subjective decision (i.e., professional judgement) as to whether or not to include the outlier(s). Justification for the removal or inclusion of the outlier(s) must be recorded along with the final analysis.

I.2.7 Selection of the Most Appropriate Model

Once all of the contending models have been fit, each one should be assessed for normality, homogeneity of the residuals, and the residual mean square error. The model which meets all of the assumptions and has the smallest residual mean square error (refer to Figure I.3) should be selected as the most appropriate model. However, in the case where more than one model has the same residual mean square error, and all other factors are equivalent, the simplest model should be selected as the model of best choice. If a weighted regression was performed, the weighted and unweighted analyses should be compared and the weighted analysis selected if weighting reduced the standard error for the ICp by more than 10%. The residual mean square error is presented in the '**Output Pane**' tab just following the iterations, and preceding the parameter estimates. However, if none of the models adequately fit the data, then the user is advised to consult a statistician for the application of additional models, or the data should be re-analyzed using the less desirable linear interpolation method of analysis (see Section 4.8.3.2).

Note: Since the concentration or treatment levels were logarithms in the calculations, the ICps and their confidence limits should be transformed to arithmetic values for the purpose of reporting them.

I.2.8 Creating the Concentration-Response Curve

Once an appropriate model has been selected, the concentration-response curve for that particular model must be generated.

(1) Within the command editor window at the bottom of the screen, copy the model equation (i.e., the equation after the '=' sign, third line of the command codes depicted in Table I.2) from the command codes used to derive the estimates for the selected model; the equation should consist of the original alphabetic characters (e.g., t, b, h, etc.). The equation can be copied by highlighting the equation and selecting **Edit**, followed by **Copy** (or right-clicking the mouse and selecting **Copy**).

(2) Select **File**, **Open**, and then **Command** and open an existing graph command file (i.e., any file with '*.cmd') similar to the following example (or, if and as necessary, create a new one), using the logistic model. The first plot (i.e., 'plot') is a scatter plot of the dependent variable against the log concentration series. The second plot (i.e., 'fplot') is the regression equation, which is superimposed upon the scatter plot.

```
graph
begin
plot drywts*logconc/ title = 'Dry Mass of Barley Shoots', xlab = 'Log(mg boric acid/kg soil d.wt)',
ylab = 'Mass (mg)',
xmax = 2, xmin = 0, ymax = 90, ymin = 0
fplot y = 80.741/(1+(0.25/0.75)*(logconc/0.611)^2.533); xmin = 0,
xmax = 2, xlab = '' ymin = 0, ylab = '', ymax = 90
end
```

(3) Paste the previously copied equation in place of the pre-existing equation (as seen in the shaded area above) by highlighting the previous equation, and then selecting **Edit**, followed by **Paste** (or right-clicking the mouse and selecting **Paste**). Replace all of the alphabetical characters (e.g., t, b, h, x, a, etc.), together with the respective estimates, provided in the '**Output Pane**' tab generated by the application of the selected model.

(4) Type in the correct information within the line entitled 'plot y*logconc . . . ', where 'y' is the dependent variable under study (e.g., drywts). Adjust the 'xmax' (i.e., the maximum log-concentration used) and 'ymax' (refer to Section I.2.1, Step 7) numerical values accordingly. Ensure that all 'xlab' and 'ylab' (i.e., axis labels) entries are correct, if not, then adjust accordingly. Ensure that all quotation marks and commas are placed within the command program as depicted in the previous example; SYSTAT is case- and space-insensitive.

Note:

'title'	refers to the title of the graph
'xlab'	refers to the x-axis label
'xmin'	refers to the minimum value requested for the x-axis
'xmax'	refers to the maximum value requested for the x-axis
'ylab'	refers the y-axis label
'ymax'	refers to the maximum value requested for the y-axis
'ymin'	refers to the minimum value requested for the y-axis

The 'xmin', 'xmax', 'ymin', and 'ymax' must be the same for both plots to superimpose the regression line accurately on the scatter plot of the data. An example of the final regression graph is provided in Figure I.1 for each of the five proposed models.

(5) Select **File**, then **Save As** to save the graph command codes in an appropriate working folder using the same coding used to generate the data file, with indication as to which model the regression corresponds to. Select **Save** to save the file.

(6) Select **File**, then **Submit Window** to process the command codes. A graph of the regression, using the model estimate parameters for the selected model, will appear.

I.3 Determining Additional ICps

In some cases, it might be desirable to estimate another value for 'p' (besides or instead of an IC25). The models proposed by Stephenson *et al*. (2000b) enable the selection and determination of any ICp. The following section provides guidance on determining an IC20, however, the models can be changed to suit any 'p' value.

(1) Select **File**, **Open**, and then **Command** and open the file corresponding to the command codes used to generate the estimate parameters (refer to Table I.2 for the command codes for each model). Change the model equation such that it will calculate the desired ICp (e.g., IC20) by modifying the fractions used in each model. For example, to calculate an IC20 using the logistic model, the equation would change from 't/(1+(0.25/0.75)*(logconc/x)^b)' (for calculating an IC25) to 't/(1(0.20/0.80)*(logconc/x)^b)' (for calculating an IC20).

(2) Once the equation has been adjusted for the ICp of interest, follow each step outlined in Section I.2.3 of this appendix. However, substitute the estimated ICp (e.g., IC20) within the fifth line entitled 'estimate/ start=' (refer to Figure I.1 for details on the substitution for each model). These values were initially derived from an examination of the scatter plot or line graph. The model, once it converges, will provide a set of parameters from which the ICp, and its corresponding 95% confidence limits, are reported (i.e., parameter 'x').

(3) Proceed with the analysis as described in Sections I.2.4 to I.2.8 herein.

I.4 Analysis of Variance (ANOVA)

(1) Select **File**, **Open**, and then **Data** to open the data file containing all of the observations for the data set under examination.

(2) Select **Analysis**, **Analysis of Variance (ANOVA)**, and then **Estimate Model...**.

(3) Select the variable within which the data are to be grouped (e.g., logconc), and place this variable into the '**Factor(s):**' box by selecting **Add**.

(4) Select the variable of interest (e.g.,drywts), followed by **Add**, to insert the variable into the '**Dependent(s):**' box.

(5) Select the box beside '**Save**' (bottom left-hand corner of the 'Analysis of Variance: Estimate Model' window) and scroll down the accompanying selections to choose **Residuals/Data**. Type in an appropriate file name within the adjacent empty box to save the residuals (e.g., anova1). Select **OK**. A graph of the data and the generate output will appear within the '**Output Pane**' tab. At this point, any outlier(s), based on the studentized residuals, will also be identified (refer to Section I.2.6 of this appendix for guidance on assessing outlier(s)).

(6) Assess the assumptions of normality and homogeneity of the residuals as per Section I.2.4 using the data file that was created to save the Residuals/Data prior to conducting the ANOVA (i.e., anova1). After assessing normality and homogeneity of the residuals using Shapiro-Wilk's and Levene's tests, respectively, the following coding may be used to examine the graphs of the residuals:

```
graph
use anova1
plot residual*logconc
plot residual*estimate
```