www.environment-agency.gov.uk

Biological Test Methods for Assessing Contaminated Land

Stage 2 - A demonstration of the use of a framework for the ecological risk assessment of land contamination

ENVIRONMENT AGENCY

www.environment-agency.gov.uk

Biological Test Methods for Assessing Contaminated Land

Stage 2 - A demonstration of the use of a framework for the ecological risk assessment of land contamination

Science Group Report P5-069/TR1

[1]Weeks J.M., [1]Sorokin N., [1]Johnson I.J, [1]Whitehouse P., [2]Ashton D., [3]Spurgeon D., [3]Hankard P. and [3]Svendsen C. [4]Hart A.

1: WRc-NSF Ltd 2: Environment Agency Science Group 3: CEH, Monks Wood 4. CSL, York

Commissioning Organisation:

Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury BRISTOL BS32 4UD.
Tel: 01454 624400 Fax: 01454 624409
Website: www.environment-agency.gov.uk

ISBN : 1844322963

Dissemination status:
Internal: Released to Regions
External: Public Domain

Statement of use
This report describes the evaluation and subsequent modification of a previously developed tiered risk assessment framework for assessing risks to ecosystems from contaminants present in soil. It also describes the results of a suite of standardised acute and sublethal biological tests for assessing ecological health in terrestrial systems. The tests have been applied in the laboratory using soils from two potentially contaminated sites in the UK. The report highlights those tests that generated useful information and recommends where test modifications are warranted. It also addresses the issue of how data arising form the tests can be used to inform decisions about ecological risk. In addition, it proposes modifications to the decision-making framework to achieve this more efficiently.

The tests may be used to determine the health of an individual organism, a population or whole soil ecosystem. The predictive application of tests will also be useful in determining the future health of soils and their ability to respond to stress resulting from point source or diffuse pollution events, including agrochemical use.

Keywords
Ecological Risk Assessment, Contaminated land, Part IIA, Biological testing, Ecotoxicity, Tiered framework, Soil bioassay, Soil processes, Invertebrates, Microbes, Soil pollution, Laboratory assays.

Research contractor
This document was produced under Science Project P5-069 by:
WRc-NSF Ltd, Henley Road, Medmenham, Marlow, SL7 2HD

Tel: 01491 636500 Fax: 01491 636501
Website: www.wrcnsf.com

WRc Report No.: EA6339/2

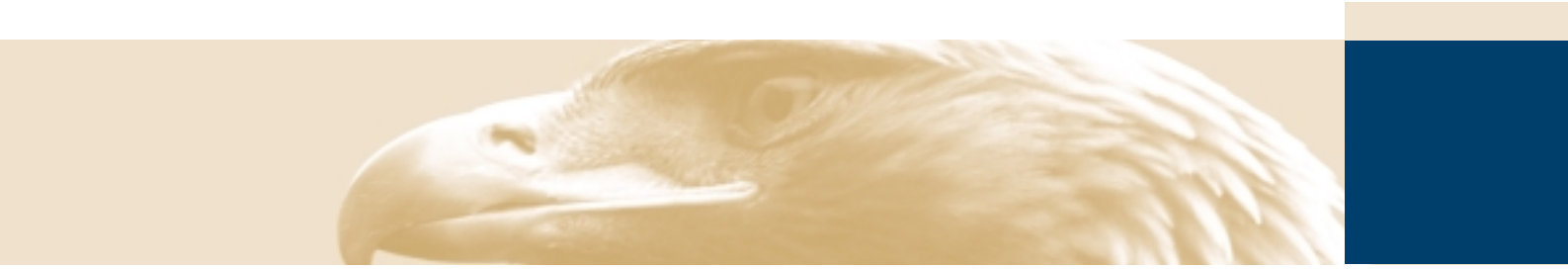
Environment Agency's Project Manager
The Environment Agency's Project Manager for R&D Project P5-069 was:
Dr Danielle Ashton, Science Group, Wallingford

Further copies of this report are available from our National Customer Contact Centre by emailing enquiries@environment-agency.gov.uk or telephoning 08708 506 506

Contents

List of Tables

List of Figures



Executive Summary

Ecological Risk Assessment (ERA) is an increasingly important part of the decision-making process for managing environmental problems. Under Part IIA of the Environmental Protection Act 1990, ERA clearly has a key role in the ecological assessment of contaminated land. This report describes a decision-making framework to identify sites where ecoreceptors may be subject to risk from contaminants. It also reports the use of biological tests to determine whether ecoreceptors are at risk of harm from soil contaminants.

In this report, we address the performance of a series of biological tests and a proposed decision-making framework gained from practical experience at two study sites in the UK. We highlight those tests that generated useful information and recommend where test modifications are warranted. In addition, we also address how data arising from the tests may be used to inform decisions about ecological risk, again proposing modifications to the decision-making framework to achieve this more efficiently.

The following seven test methods were performed on soil samples from the study sites:

- Microbial bioluminescence tests with *Vibrio fischeri* for example, Microtox™ and Mutatox;
- Microbial carbon mineralisation test;
- Microbial nitrogen mineralisation test;
- Germination and root growth test with a monocotyledonous plant species (for example, wheat) and dicotyledonous plant species (for example, cabbage);
- Acute lethal earthworm test with *Eisenia andrei*;
- Acute springtail test with *Folsomia candida*.

All these methods have undergone standardisation and are published by the International Organisation for Standardisation (ISO) or Organisation for Economic Co-operation and Development (OECD), with defined test validity criteria. Together with ecological and chemical data, the results of these tests have been used in a weight-of-evidence approach to make decisions about the risks to ecoreceptors at the two study sites (Sites A and B). One of the selected trial sites within the project has considerable contextual environmental data (Site A). This was useful because it provided a benchmark against which we could judge the value of the data from laboratory testing.

A significant feature of the decision-making framework is that it adopts a tiered approach. It starts with the development of a conceptual site model (CSM), reviewing all known information, defining the boundaries of the site under investigation, and identifying plausible contaminant-pathway-receptor (C-P-R) linkages for chemical contaminants. This CSM is termed 'Tier 0' and was an important addition made by this project to the proposal made by Byrns & Crane (2002). This tier can provide an opportunity to terminate the process

early, when this is warranted, that is, when no significant C-P-R linkages are identified in the CSM.

The next stage is a Tier 1 screen. This identifies whether a site can be excluded because it is unlikely to pose a risk to ecoreceptors, or whether further investigation is needed. This decision is made by comparing chemical residue data with Soil Quality Guideline Values (SQGVs) in a simple deterministic risk assessment. This is supplemented with toxicity screening as a way of assessing soils containing contaminants for which SQGVs do not exist, have not been analysed for, or whose toxicity is greater than expected when they occur in combination (synergism). Experience from the study shows that the Microtox™ test exhibits many of the characteristics required of such a screening test.

Biological tests are used at Tier 2. They assess whether exceedances of SQGVs are translated into harmful effects. This study has shown that higher plant tests may usefully be adopted, with the recommended modifications. In addition, a sublethal variant of the earthworm test is likely to yield useful information, though the acute test should not be adopted. While a nutrient cycling test is considered useful, the nitrogen mineralisation test would first need significant development to make it a practical proposition. Problems of fungal contamination in the Collembolan tests would also need to be resolved before this test could routinely be adopted at Tier 2.

We did not undertake a final tier of assessment (Tier 3) in this project. This tier may be carried out when uncertainties remain about the significance of chemical residues found at a site, or their biological effects as revealed by ecotoxicity testing. In particular, food chain modelling could be applied to test C-P-R linkages of particular concern (for example, that apply to rare or endangered species) or to establish the level of risk from bioaccumulative substances that cannot readily be addressed at Tier 2. Tier 3 investigations may also be carried out to better understand the extent of areas that pose risk and might require remediation. The type of investigation carried out at Tier 3 is therefore much more site-specific than hitherto, and is guided strongly by the CSM, for example, improved assessment of the exposure route of ecoreceptors to the contaminants present, that is, measures of ingestion rates, sampling of food sources and tissue analysis.

ACKNOWLEDGEMENTS

This study has benefited from the input of all members of the Project Board and collaborators. We have welcomed the advice and input from the Environment Agency, and we are particularly indebted to our industrial partner for organising access to one of the study sites and associated historical data.

Introduction

The ability of a soil to support functioning ecosystems depends, in part, on the extent of any chemical contamination that could have direct (toxic) effects on soil biota, or indirect ones for example, through altering nutrient cycling or litter decomposition. Many industrial sites throughout the UK are subject to contamination, often by a 'cocktail' of different substances that may be present at levels high enough to cause adverse effects.

1.1 Background

Part IIA of the Environmental Protection Act 1990 introduced a new statutory regime for the identification and control of land potentially affected by contamination. The statutory definition of contaminated land (DETR, 2000) requires:

1. a contaminant source to be present, a pathway along which it can move and for the contaminant to affect - or potentially affect - a specified receptor;
2. a significant possibility of significant harm;
3. pollution of controlled waters is occurring, or is likely to occur.

Related statutory guidance (Section 57 Part IIA, EPA 1990) identifies 'ecosystems' as one of the receptors that may define a site as 'contaminated land' for the purposes of the regime. Currently, only risks to controlled waters and certain protected habitats, defined as specified receptors in Part IIA, are covered. Nevertheless, there is clearly an ecological imperative that prompts a requirement for a methodology that will identify ecosystems potentially at risk and assess the level of any risk.

Local authorities have sole responsibility for designating land as contaminated under Part IIA, and may act as the enforcing authority. In addition, the Environment Agency has certain responsibilities both for Special Sites (where the Agency will take the lead) and in situations outside Part IIA, such as land redevelopment under planning regulations. Therefore, it is necessary to describe a process that can be used to predict or estimate risks to ecosystems from land that is, or might be, contaminated land under the Act.

1.2 Introduction to Ecological Risk Assessment (ERA)

Ecological risk assessment is an increasingly important part of the decision-making process for managing environmental problems. In the past, ERA has been associated mostly with risks to the aquatic or marine environments. But attention is now turning to impacts on soil quality. Under Part IIA, ERA clearly has a key role to play there, too.

Risk assessments may be used to evaluate environmental problems arising from historical and ongoing activities (retrospective risk assessment) and, in some cases, those associated with future activities (prospective risk assessment). In the case of the contaminated land regime, we are largely concerned with retrospective risk assessment.

Principles for the conduct of risk assessments are to be found in *"Greenleaves"* (DETR, Environment

Agency and Institute of Environmental Health, 2000). Typically, an ERA for toxicants relies on comparing some exposure estimate for each chemical of interest with a corresponding toxicity threshold for individual organism endpoints such as survival, growth or reproductive potential. This comparison is typically accomplished by the derivation of a hazard quotient. This is simply the exposure estimate divided by the toxicity threshold. In the derivation of these thresholds, uncertainty factors or other toxicity extrapolation methods are often applied to translate an endpoint of interest into a toxicity threshold.

There are, though, limitations of screening approaches, especially where exposure and toxicity remain unquantified (that is, exposure and effects data are lacking) or where biota are exposed to a complex mixture of chemicals. As well as making assessments on the basis of exposure and toxicity of individual substances, the biological testing of the whole soil can also be useful, as can evidence gained from ecological surveys (sometimes referred to as the *triad approach*).

Biological tests may include the use of ecotoxicity tests or biosensors in the laboratory or *in situ*. Biological assessment can be undertaken at varying levels of biological organisation, including ecological function or structure, where effects may be assessed at the sub-cellular level, through effects on individual species, up to effects at the population, community or ecosystem levels (see Section 3). The use of biological tests in ERA is attracting research interest. They are a useful supplement to chemical analysis in decisions related to contaminated land (Ferguson *et al.*, 1998; CIRIA, 2002). To date in the UK, biological tests have been used only in a regulatory role in the aquatic environment. This project investigates the potential for their use within risk assessments of contaminated land. This application in the terrestrial environment represents a significant development of the use of biological testing for regulatory purposes.

Ecological assessment can also play an important role in helping judge the biological status of a site at higher levels of biological organisation. Typically, such surveys address the biological diversity at a site and the abundance of species. They may entail assessments of a wide range of taxa, or they may be confined to abundance of certain taxa known to be sensitive to change or have a key ecological or conservation role.

1.3 Why biological testing?

Standardised ecotoxicity tests (e.g., described as formal OECD or ISO test guidelines) were originally designed to assess the toxic effects of new and existing chemicals such as plant protection products (that is, pesticides) added to soils, to enable a regulatory risk assessment. These are usually used as part of a prospective assessment of risk. Such tests are also used in the derivation of toxicity data that form the basis for chemical thresholds, for example, Environmental Quality Standards (EQS).

Biological tests in this study use soil and water samples directly from potentially contaminated sites to generate site-specific risk assessments. This 'Direct Toxicity Assessment' (DTA) approach has been the subject of considerable research in the UK and elsewhere in the context of controlling emissions to controlled waters. In some countries, it has a clear regulatory role (for example, Wharfe, 1996). As a first step to implementation in the UK, DTA approaches are specified in certain IPPC impact assessments in the Agency's Horizontal Guidance Note on IPPC (H1) *Environmental Assessment and Appraisal of BAT* (Environment Agency, 2002a).

DTA offers a number of advantages over chemical-specific approaches. But is also prone to some limitations, which we explore in detail in Section 3.

To date, the DTA approach has not been applied in the UK in the context of contaminated land. The principles involved in biological testing of whole soils are, though, identical to those for testing water or effluent samples. Numerous single-species toxicity tests (bioassays) are available to measure contaminant exposure and its effects on different biological components of soil ecosystems. We describe whole soil tests using terrestrial invertebrates (earthworms, springtails etc) plants and microorganisms later (see Section 3). Tests can be conducted either *ex situ* (in the laboratory) or *in situ* (in the field) and the results of either may be used to assess the risk of effects of contaminated soil on ecological receptors.

Biological testing is not without drawbacks. Biological tests are as prone to problems of heterogeneity in contaminant levels and soil properties across a site as is chemical analysis. A more fundamental difficulty is identifying chemical(s) responsible for any toxicity. While the identity of toxic contaminants may not matter when judging whether adverse impacts are likely, it does become important when assessing remediation options. An additional criticism commonly levelled at biological testing schemes is that the variability within and between testing laboratories can give rise to differences in interpretation about the risks posed at a particular site. In the UK approach to DTA for wastewater controls, quality control measures have been proposed to address this issue (Environment Agency, 1999). The key principle here is to ensure that any variability is understood and accounted for in decision-making, that is, that the levels of any

measures to constrain variability are sufficient to ensure fitness-for-purpose, without being over-restrictive.

1.4 Previous and current research

Previous Agency research (Byrns and Crane, 2002) recommended an ERA framework for use in the UK based on a review of the strengths and limitations of ERA frameworks in the US, Australia, Canada and the Netherlands.

The key recommendation from this review was for a three-tier framework in which the assessments became increasingly site-specific in response to indications of possible risk, and the use of a weight-of-evidence approach to assess risk. This framework emphasises the importance of correctly defining the problem under review and the need to screen and prioritise risks prior to more detailed risk assessment. Section 2 explains in more detail the various steps involved in the framework. It starts with a conceptual model to identify sources of contaminants and receptors at risk, an assessment of risks based on chemical data and, if uncertainties remain after this step, the use of biological tests to assess directly the impacts of soil on biota.

A separate report (Crane and Byrns, 2002) reviewed biological methods for assessing the toxicity of soils to living organisms and their suitability for use in ERA. This focused strongly on standardised laboratory tests on whole organisms (i.e. tests that were subject to standardisation by OECD or ISO). A separate review (Spurgeon *et al.*, 2002) extended this appraisal to cover established and novel laboratory and *in situ* tests, particularly those covering sub-lethal endpoints (the 'SubAssess' project, Environment Agency, P5-063, Spurgeon *et al.*, in press). This project set out (a) to investigate new techniques that might enable assessment of the sublethal effects of chemicals on terrestrial organisms and (b) to trial selected tests and review their usefulness in practice. As explained below (Section 1.8), we evaluated biological tests in this project in a way that would be harmonised with parallel evaluations of tests arising from Project P5-063. Fishwick (2004) reviews methods for deriving Soil Screening Values (SSVs) for use in Tier 1 and recommends an approach for the UK. This report is a review of methods and lists priority contaminants. The Environment Agency consultation document (Environment Agency, 2003) seeks opinion on proposed methods and the first batch of SSVs.

A key point about the proposals for the ERA framework is that they place emphasis on the triad approach outlined above. They adopt biological testing of soils and ecological survey data to assess soil quality, in addition to assessments of risks based on individual chemical contaminants.

1.5 Purpose of the project

The purpose of this project is to:

1. Critically test the performance of a series of ecotoxicological tests that were identified in the earlier project (Crane and Byrns, 2002), for assessing risks to soil biota. Are they capable of discriminating contaminated soils?
2. Assess the value of these test methods for risk assessment purposes and recommend a strategy for implementing these methods using two case studies. Do they generate useful information that can be used to inform regulatory decisions?
3. Suggest modifications to the risk assessment framework in terms of how the tests are deployed and how the resulting test data are used in decision-making. How can the risk assessment process be modified to generate more sound decisions about soil quality?

The project did not set out to evaluate the general effectiveness of the entire proposed ERA framework to assess contaminated sites. Nor does it address possible impacts arising from contaminated soil on groundwater or surface waters. Rather, it evaluates the use of biological tests within the proposed framework and makes recommendations, from which technical guidance on ERA may be developed. Experience gained has, though, led to some alterations in the framework, notably the inclusion of a further initial tier. These alterations are described in Section 2.

Two sites with known chemical contamination were used to trial the tests. One (A) has a well-documented history of metal contamination. The other (B) may be contaminated by petroleum products. We subjected the sites to Tiers 0, 1 and 2 of the framework (see Section 2.3 for an overview of the framework). Significantly, we applied the activities involved in each of these Tiers to both sites, irrespective of whether this would normally be required by the framework. This reflects the project aim to trial the biological tests in a range of scenarios (Objectives 1 and 2 above) and to ascertain the value they add to decision-making. For the same reason, we deployed the entire suite of biological tests, recognising that, in practice, only certain tests would be applied.

The outcomes from this project will be compatible with developed generic guideline procedures for the

protection of human health (e.g. DEFRA/Agency CLR 10, 2002c), in that they also take a tiered approach that is fundamentally risk-based. There are also wider implications of any ERA framework applied to land that is potentially contaminated. Ideally, it should provide an assessment of risks that relate to:

- land redevelopment (a change of use);
- sites used for activities authorised under or subject to the Groundwater Regulations 1998, Integrated Pollution and Prevention Control Regulations 2000 or the Control of Major Accident Hazards Regulations 1999;
- establishing remediation liabilities for historic contamination on-sites that are not being actively redeveloped and which, in their current or likely use, present unacceptable risks to the wider environment.

1.6 Selection of trial sites

We selected two trial sites for the tests. One (Site A) is an area adjacent to a smelter that is subject to metal contamination. The other (Site B) has soil contaminated with petroleum products capped with uncontaminated soil. Though there is evidence of historical contamination at both sites, neither is formally designated as an ecosystem that would fall under the guidance for Part IIA sites. In order to demonstrate the ERA framework, we have overlain some hypothetical ecoreceptors that would make the sites eligible for investigation under Part IIA.

Further details of Sites A and B are to be found in Sections 5 and 6.

1.7 Harmonisation with evaluation of sublethal tests (P5-063)

For the trials of the selected biological tests, we needed to identify sites that were suitable for detailed investigation. To harmonise the work being conducted in this project with that concerning the sublethal assays being undertaken in Project P5-063, we used the same sites and sampling patches for both projects. Thus the two projects complement one another. This project focuses on risk assessment using standardised test methods that measure predominantly acute effects of contaminants in soils on primary producers (plants), decomposers (earthworms, springtails) and on the functional status of soil microbial communities (nitrogen and carbon mineralisation). Project P5-063 focuses on the use of sublethal bioassays, biomarkers and functional assessments.

1.8 This report

In this report, we address the performance of a series of biological tests and the proposed decision-making framework gained from practical experience. We highlight those tests that generate useful information and recommend where test modifications are warranted. Where appropriate, we also recommend modifications to the proposed decision-making framework. Much of our report is concerned with the performance characteristics of a series of biological tests, including test standardisation, sensitivity, and definition of response thresholds denoting harm to ecological receptors. We also consider how data arising from the tests may inform decisions about ecological risk.

It should be noted that this report is primarily an appraisal based on research experience. It is not a guidance document on the use of ecotoxicity tests for ERA purposes, or on the implementation of the ERA framework. Both the framework and the suite of biological tests are undergoing further road-testing by a group of industrial partners. The Agency's Process teams will develop guidance based on the findings of this extended trialling.

Section 2 describes the tiered ERA framework used in this project. Section 3 reviews current and past uses of biological tests, including those selected for further evaluation here. We cover the performance criteria against which these tests might be judged in Section 4. Sections 5-6 describe the performance of the suite of biological tests deployed at the two trial sites. We review the performance of tests in Section 7, and propose revisions to the risk assessment framework, in the light of the experience gained, in Section 8. Finally, Section 9 sets out our key recommendations.

The ERA Framework

In this Section, we describe the use of the tiered risk assessment tool for deciding whether defined ecoreceptors are at risk from contaminants present in soil. In later Sections (5 and 6), we present and use data in conjunction with these recommendations to assess whether plausible decisions can be reached using this framework.

2.1 Introduction

The proposed framework for carrying out an ERA is based on that recommended by Byrns and Crane (2002) with some modifications, notably the addition of a new tier, Tier 0, and the use of toxicity screening at Tier 1. In Section 2.2, we outline the key principles in the framework. Section 2.3 describes in some detail the activities involved at each tier and how to decide whether to continue investigation or to allow a site to exit from the process.

2.2 Key principles of framework

The framework is based on a review of schemes used in other countries, notably the US, Australia, Canada and the Netherlands. A key element of these schemes is the reliance placed on a tiered framework, that is, one in which progression through a series of tiers reflects greater refinement in the quality and quantity of information gathered and progressive reduction in uncertainty. Initial decision-making is based on conservative assumptions, so that sites that are truly benign may be eliminated from further investigation with a high degree of confidence. In other words, early tiers are a way of screening out sites that pose little or no risk to ecoreceptors. Subsequent data gathering is intended to make more realistic assessments of the risk of harm to ecoreceptors. This tiered approach allows a transparent approach to decision-making. It also ensures that resources are allocated where they are most likely to be needed.

This framework is designed for assessing risks to ecosystems only from contaminants present in soil. It does not address the effects of other stressors such as physical soil structure, climate change or changes in land use that may impact ecosystems.

An important feature of the tiered framework is that it doesn't rely only on chemical or biological data. Rather, it favours a triad approach in which chemical data is used in conjunction with biological data (for example, toxicity testing) and ecological survey data to inform a decision about risks to ecoreceptors.

Finally, while there is much in common between the assessment of risks to human health and to ecoreceptors, there are also important distinctions. In particular, human health risk assessments focus on individuals. Evaluations of ecological risk typically focus on risks to populations, communities and ecosystems. At least for plants and soil organisms, biological data to support an ERA are more amenable to direct field observation than human exposure and epidemiological data.

2.3 Overview of the framework

Figure 2.1 illustrates the various tiers in the framework and the decisions that may need to be taken at each. Essentially, decision points at each tier

determine whether to stop the work (because the risk of adverse effects is judged to be acceptably low) and allow the site to exit from the process, or to seek further refinement through data generation. In cases where there is overwhelming evidence of an existing adverse impact, for example, from ecological surveys, a decision may be made to take management action without progressing through all the tiers. We give detailed descriptions of the activities and decisions made at each of these tiers in Figures 2.2 and 2.3.

2.3.1 Tier 0

The first step of the ERA process **(Tier 0)** aims to determine whether a site falls under the Part IIA considerations. It involves the development of a **conceptual site model** (CSM). This describes what is currently known about the site, its geographical limits, and identifies potential contaminants, pathways and receptors. The level of detail that is required in the risk assessment will be influenced by many factors, but particularly by:

- the sensitivity of the site (the receptors present);
- the inherent hazardous properties of the contaminants (toxicity);
- risks of exposure of receptors to contaminants (the existence of pathways between contaminants and receptors, and chemical properties such as persistence);
- the potential for contaminants to bioaccumulate through food chains.

Appendix A gives details of Tier 0. The appendix describes the tier in some detail because this is a modification to the framework proposed by Byrns and Crane (2002).

In summary, a key feature of Tier 0 is that it emphasises collating existing data on geology, chemical analytical data (where available), biological surveys and a site walkover rather than de novo data generation. As well as establishing a firm basis for further evaluation, the Tier 0 assessment can also provide an opportunity for early termination of the process when this is warranted. For example, it may become clear that there is no ecological risk because one or more of the critical elements (contaminant-pathway-receptor) is absent. In this case, the ERA process would be terminated before any practical work starts, ensuring a more efficient use of resources than would be the case without this step.

Tier 0 is important because it underpins the entire process. The development of a CSM allows information generated in subsequent tiers to be properly evaluated and helps identify what further

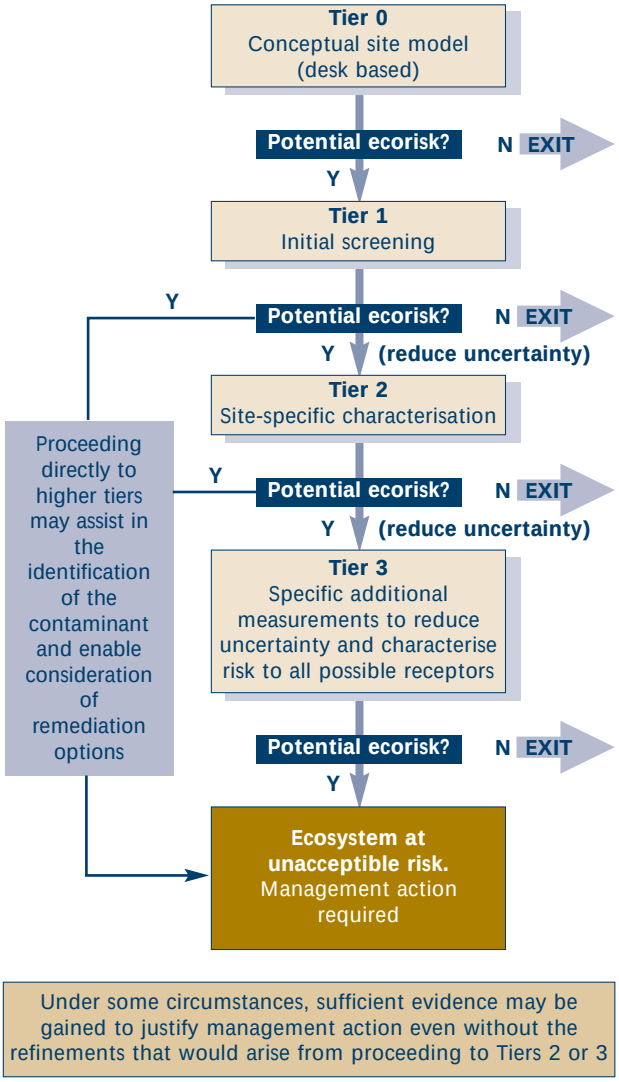


Figure 2.1 Overview of tiered risk assessment

information is needed. We would expect the complexity of the problem and the level of uncertainty in the conceptual model to impact on the progression through the tiers and determine where the risk assessor may halt or exit from the ERA framework. If a site does progress beyond Tier 0, the CSM should be revised in the light of new knowledge collected in any subsequent tiers as part of an iterative process.

2.3.2 Tier 1

Having identified at Tier 0 that a site could fall under the Part IIA conditions, and also that there are indeed at least theoretical contaminant-pathway-receptor linkages, the next step is to evaluate whether these are realised and to determine whether further evaluation is warranted.

Tier 1 is thus essentially a step to screen out those

sites where an unacceptable risk is unlikely to be realised. This decision is based on conservative assumptions, effectively according benefit of doubt to the environment. Two consequences follow from this. First, the primary concern is to minimise the risk of false negatives (that is, failing to detect sites that truly pose a risk to ecological receptors). In doing so, though, it is important not to increase the incidence of false positives (that is, progressing a site that does not actually pose an unacceptable risk to ecoreceptors) to a level that the process becomes inefficient because it fails to exclude genuinely benign sites. Second, the assessment is generic in nature, since work of direct relevance to assessment endpoints at a particular site is addressed in higher tiers.

At Tier 1, the emphasis is on a chemical-specific risk assessment. Essentially, concentrations of known or potential contaminants present in soil are compared against thresholds for individual chemicals. These thresholds are referred to generically as soil quality guideline values (SQGVs), concentrations of chemicals below which no adverse effects on the specified receptor are expected. They are sometimes referred to as soil screening guidelines or soil screening values. In addition, some toxicity screening may usefully be incorporated as a means of reducing the chances of missing contaminants that are not covered by SQGVs.

Sampling strategy

Soils act as a final sink for many contaminants by sequestering these compounds. The extent to which contaminants are sequestered and rendered unavailable to biota depends on the following factors:

- Chemical characteristics and concentration of pollutant(s);
- Composition of the soil matrix, in particular organic matter and fine particles, pH;
- Residence time of pollutants in soil (ageing);
- Presence of co-contaminants.

Due to the heterogeneity of soils, the sorption potential varies considerably. This results in changing contaminant availability, sometimes within a small area. With time, sequestration processes become even more pronounced, a phenomenon generally referred to as 'ageing'. We must therefore ensure that procedures for sampling and handling soils prior to (and during) toxicity testing minimise changes in the bioavailability of any potential contaminants present in the soil. In practice, this is achieved by adhering to guidelines for soil sampling (for example, BSI, 2001; Environment Agency, 2001; Nathanail *et al.*, 2002).

Targeted sampling regimes (where sampling focuses on locations of particular interest) or non-targeted sampling regimes may be adopted. Within the latter, several different approaches can be applied, for example random designs, square grid or herringbone (DoE, 1994). Very few site investigations, though, adopt a truly random sampling regime. Sampling is typically guided by the CSM. The number of samples taken depends on the level of resolution that is needed and should generally be greater in situations where contamination is likely to be particularly heterogeneous (Nathanail *et al.,* 2002).

Finally, the nature of likely sources of contamination (for example, aerial deposition or incorporation of solid waste) will inform the depth of sampling, as will the receptors of concern, again highlighted during formulation of the CSM. For most ecoreceptors, relatively shallow sampling (that is, within the root zone dependent upon the plant species of concern) is in order.

Screening using SQGVs

Chemical analysis produces chemical contaminant data (i.e., contaminants of potential concern emerging from the Tier 0 assessment). For each contaminant, a reasonable worst-case estimate of the predicted environmental concentration (PEC)[1] is made. In some cases, existing chemical data will be available and should be used unless there is reason to believe it is no longer relevant (for example, residues data for volatile compounds). The PEC is compared with the predicted no-effect concentration (PNEC). Where SQGVs are available, these are used as the PNEC[2] but where they are not, PNECs can be derived ab initio from ecotoxicity data. The Environment Agency is proposing to use the EC Technical Guidance Document on Risk Assessment methodologies for deriving PNECs, as the basis of SSVs (Fishwick, 2004).

This is therefore a simple quotient approach to risk assessment, similar to that adopted in many other regulatory regimes. If at this step no adverse risk is indicated for all the chemicals identified within potential pollutant linkages at Tier 0 (that is, the PEC/SQGV ratio is < 1), then no further work is necessary - unless toxicity screening (see below) reveals evidence of toxicity.

The toxicity of a chemical is strongly governed by the physical and chemical properties of a soil. Though soil contaminant concentrations might exceed SQGVs, resulting in prediction of negative effects on the soil ecosystem, it is not possible to

[1] The PEC should be based on the upper 95 percent C.L. of the mean based on the available data, thereby minimising the effects of outliers but representing a reasonable worst case scenario (see description of mean value test in CLR7). Where data are available, this may be estimated from measured data

[2] Different SQGVs for the same substance can apply. In this situation, the PNEC is the lowest SQGV available, thereby basing the assessment on the most conservative interpretation

evaluate the true extent of the toxic effects, without carrying out specific toxicity assessments of the soils at the site. The biological significance of chemical residues is considered further at Tier 2.

In the Agency consultation document (December 2003) proposals are made at Tier 1 further to investigate sites where it is suspected that elevated concentrations of naturally occurring elements may be influencing the outcome of the PEC/SQGV ratio. We support the proposals to develop practical and useable screening values, and we await the outcome of the consultation. For the purposes of this screening assessment, background levels are assumed to be zero and all the substance detected is assumed to be bioavailable. While this is likely to overestimate the actual exposure of ecoreceptors, it is a precautionary approach at this stage. It may be refined in later tiers.

A judgement must be made about the significance of any exceedances of SQGVs. An isolated or very small exceedance would carry less weight than large exceedances of several SQGVs in a large number of samples. Uncertainty at this level of assessment is, though, high and it would not be prudent automatically to exclude such sites. It is inappropriate to provide hard and fast rules about the magnitude and number of exceedances required before progressing to Tier 2. Rather, we advocate using this information to prioritise resources. So, where several sites compete for resources, attention would be focussed on those where exceedances were large, where residues in a large number of samples exceeded SQGVs, and/or where there are exceedances for bioaccumulative compounds. An exception to this would be when only a few samples are available (that is, spatial coverage of the site is poor) or only few determinands are represented. Under these circumstances, there may be a need for further sampling and analysis. If this is not practically possible (for example, the site is a nature reserve where sampling might itself cause harm), any exceedance would be sufficient to trigger progression to the next tier.

Sources of SQGVs

For this project, the Project Board and Study Team decided to use SQGVs from three sources: CCME soil quality guidelines (Canada) (CCME, 1997), VROM (Revised Dutch List) soil remediation guidelines (Netherlands) (VROM, 2000) and US Department of Energy screening benchmarks (US) (US Department of Energy, 1998). ICRCL trigger values (UK) have not been used because they do not explicitly address the protection of ecoreceptors and have been formally withdrawn by Defra (DEFRA, CLAN 3/02; Withdrawal of ICRCL Guidance Note 59/83 (2nd Edition).

While some of these SQGVs cover all ecoreceptors, others are concerned with effects on particular biota. US Department of Energy benchmarks are designed to afford protection to a fixed proportion (90 per cent) of species of plants, earthworms and microbes. Different land uses are specified in CCME guidelines (agricultural, parkland, commercial and industrial). VROM makes no distinction between different possible land uses. The VROM soil remediation guidelines are based on both human health and ecotoxicological data, with the more sensitive receptor determining the final outcome. Appendix C provides further detail about the provenance of these SQGVs and the basis for their derivation.

In this project, all these SQGVs have been used to illustrate the risk assessment process. It is important to point out that the Environment Agency is deriving soil screening values (SSVs) that will be appropriate for use in the ERA framework (Environment Agency, 2003 and work in progress R&D Project P6-020/5, Development of soil screening values).

Toxicity screening at Tier 1

The framework proposed by Byrns and Crane (2002) suggested that sites thought to be contaminated with complex mixtures of substances would 'side step' Tier 1 and advance immediately into Tier 2. The logic behind such an approach is that relying on SQGVs may be inadequate and would fail to deal with interactions between substances that might give rise to more-than-additive effects (that is, there is a risk of false negatives). On the other hand, most sites would fall into this category, and it is difficult to develop criteria as to what constitutes a 'complex mixture'. Under this rule, Tier 1 could become redundant for most sites.

Though the emphasis at Tier 1 is on chemical data, some rapid biological screening e.g., based on bacterial bioluminescence testing such as Microtox™, could usefully be introduced at this stage. By incorporating Microtox™ testing into Tier 1, the risk of false negatives should decline because toxicologically significant mixtures that are not detected using conventional chemical-specific SQGVs may well be detected using this bioassay. Toxicity screening would therefore serve as a backstop to reduce the incidence of false negatives rather than as a means of understanding the biological significance of any chemical contamination. Where the outcome of any Microtox™ testing is negative (that is, the results indicate no toxicity), but SQGVs indicate exceedance, then progression to Tier 2 is warranted.

Though Microtox™ appears a useful addition to SQGVs in Tier 1, concerns over its robustness and sensitivity mean that, at present, the Agency will not

accept decisions on negative effects based *solely* on Microtox™ information. Thus in the next section on decision making, where SQGV is not available for Tier 1, even though Microtox™ data may have been obtained and indicate no effects, the risk assessment should proceed to Tier 2. Further details of the Microtox™ method in relation to testing of soil samples appear in Appendix B and in the Environment Canada report EPS 1/RM/42.

The potential application of the bait lamina test as a toxicity screen was trialled as part of project P5-063 (Spurgeon *et al.*, in press). This technique could prove a useful complement or replacement to Microtox™ testing, particularly in circumstances where Microtox™ is known to perform poorly that is, in the presence of PAHs (Doherty, 2001).

Decision-making

The main decision to be made at the end of Tier 1 is whether to advance a site to Tier 2. Progression and consequently more detailed evaluation would be required if either chemical or toxicity screening tests indicated a possible risk to biota. Any of the following circumstances would trigger progression to Tier 2:

- Absence of a soil screening or appropriate guideline value (e.g. No SQGV) for any chemical present;
- When a soil screening value is available but soil screening testing (for example, using Microtox™) is negative despite chemical concentrations exceeding the guideline;
- The PEC/SQGV ratio is >1 for one or more contaminants;
- There are insufficient data to assess the risk;
- Screening ecotoxicity tests (Microtox™) suggest that toxic contaminants are present but the chemical analyses have as yet failed to detect them.

Where chemical data indicate a potential risk, then Microtox™ testing is not essential because a decision to progress to Tier 2 will already have been made on the basis of the chemical data alone. In these circumstances, though, some useful guidance about spatial variability in toxicity may still be gained from using Microtox™ testing as a screen to help locate contaminant hotspots for subsequent soil remediation. It follows that a site would exit the ERA framework only if none of these criteria are met.

At this point, the risk assessors can refine their understanding of the CSM in the light of the information gained at Tier 1. For example, it may be possible to confirm some contaminant-pathway-receptor linkages while at the same time, excluding others. The outcome of the chemical assessments and any biological screening should be reported, along with any assumptions that have been made in reaching a decision about whether or not a site should be progressed to Tier 2. The risk assessors should, though, also document areas of outstanding uncertainty and gaps in the available data. For example, contamination at a single location might exceed a SQGV while others all fall below it. If a very large number of samples have been taken, the chances of identifying an extensive 'hotspot' (or several 'hotspots') are higher than if few locations have been sampled. Subsequent work may then focus on more intense sampling but concentrate on chemical characterisation for only a small number of contaminants.

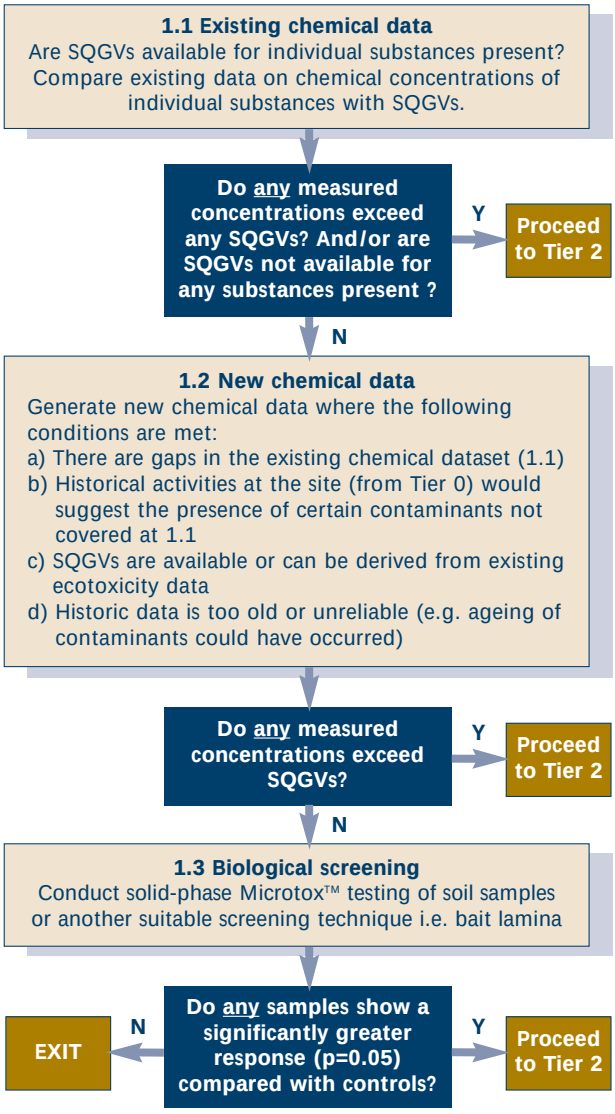


Figure 2.2 Activities and decision-making in Tier 1

Though they are not explicitly addressed at Tier 1, it is important to review the assessment endpoints defined at Tier 0 before entering Tier 2. This is because the measurement endpoints employed in any further biological testing should, as far as possible be related to those defined at Tier 0 as being particularly important. This is explored in more detail in Box 1.

2.3.3 Tier 2

The aim of Tier 2 is to enable a decision about whether receptors of concern are actually at risk of harm, now or under the proposed use of the site. This tier is therefore concerned primarily with the biological significance of contaminants present. This will be the most detailed level of assessment applied at the majority of contaminated sites.

At Tier 1, decision-making was predominantly based on chemical-specific data. At Tier 2, the emphasis shifts toward biologically based decision-making. This is because we are now more explicitly concerned with the assessment endpoints defined at Tier 0, and also with the biological significance of chemical residues present at the site. This biological information may be obtained through toxicity testing, ecological assessments or a combination of both.

Chemical analysis

Further chemical analysis may or may not be merited at Tier 2. As explained above, there are circumstances in which a better understanding of the bioavailability of contaminants identified at Tier 1 would improve our understanding of the level of risk faced by ecoreceptors.

For example, particularly high levels of certain substances may have been measured at Tier 1 without being reflected in biological impacts from ecological surveys or toxicity testing. This could be explained by sequestration or speciation of the chemical(s) at the site of concern to an extent that they are not bioavailable or do not occur in a toxic form at sufficient levels to cause adverse effects. Such cases could be addressed by site-specific chemical analysis as part of activity 2.6 (Fig 2.3), probably on soil extracts that vary in extraction efficiency (ranging from extraction techniques to remove all residues to aqueous extracts, intended to assess levels of readily bioavailable residues or weight-of-evidence (Section 4).

Conversely, Tier 1 assessment might have revealed evidence of modest exceedances of SQGVs at only a small number of sampling locations (perhaps only one). In this case, it would be reasonable to concentrate efforts on better understanding the extent of any 'hotspots' through more intensive sampling but against a limited range of determinands (determinands dependent on the former use of the site and the type of contamination likely to occur on such sites). It is a moot point whether this would be a Tier 2 activity or iteration within Tier 1. On balance, we argue it is best regarded as an intensification of Tier 1 activity.

Examples of further chemical issues that might be addressed at Tier 2 could include:

- Are the chemical residues bioavailable?
- Is the site subject to elevated natural background levels of contaminants? (see Section 2.3.2)
- Are the receptors being exposed at levels that cause harm?
- If a complex mixture of chemicals is present, what is the biological significance of this combined toxicity?
- What is the spatial extent of contamination (relative to distributions or home ranges of key receptors), and is there evidence of any particular 'hot spots'?
- Can the contaminant(s) that are responsible for adverse effects be identified, thereby focusing any remediation toward the critical sources of contaminants or pathways?

Where toxicity may be clearly evident but cannot be linked directly to available chemical evidence, more detailed studies at Tier 3 could be warranted.

Biological testing

Because the main emphasis is to understand the biological significance of the contaminants that were found to be present at Tier 1, emphasis is placed on the use of biological tests and, where they are available, on the results of ecological surveys.

A sensible starting point is to review existing ecological data. This will determine whether, on the basis of this alone, there is sufficient evidence of adverse impacts to conclude that there is a risk to ecoreceptors (that is, obvious signs of damage, decrease in species diversity and abundance). This data could be collected as part of the site walk-over. If so, further testing may be circumvented and risk management options considered (Figure 2.3). In many cases, such data will not be available, or they do not yield compelling evidence of impact. It is at this point that soil ecotoxicity tests are employed.

The suite of biological tests highlighted by Crane and Byrns (2002) and evaluated by this research is modest. These tests can therefore be regarded only

as a subset of the potential measures that could be used at a contaminated site. They do, though, include measurement endpoints that could act as surrogates for assessment endpoints of direct relevance to soil ecosystems. Some further steps can be taken to improve the relevance of decisions made on the basis of biological tests and the assessment endpoints identified at Tier 0. For a discussion of the significance of the distinction between assessment and measurement endpoints, see Box 1.

If, for example, ecoreceptors are defined in terms of particular species of concern, such as a protected plant species, then it would be prudent to ensure that plant tests are incorporated into the suite of toxicity tests. Further refinement may be possible for example, if the species of concern are families such as Orchidae or Liliaceae, the range of monocotyledonous species might be extended at the expense of dicotyledonous species. Similarly, if a key assessment endpoint arises from the fact that the site supports a breeding area for a butterfly species, emphasis would be placed on tests with insects, or on tests using potential prey/food species. If a contaminant-pathway-receptor linkage can be made to agricultural land, tests with earthworms, plants and indicators of soil functioning may be more useful. As a default, the battery of biological tests should include at least tests of soil function, plant growth and earthworm mortality. In Sections 5 and 6 reporting the field trial results, all the test methods have been employed, for the purposes of this research on their suitability. Usually, only a subset of such tests would be appropriate. (Appendix B contains further details of the tests used in the present project).

Box 1

Assessment and measurement endpoints - deciding what we want to protect and what we can measure

The ultimate aim of the ERA process is to enable the risk of contaminants present at a site to be measured and described but this first requires the protection goal to be identified. The ERA process uses surrogate measures termed **assessment** and **measurement** endpoints (e.g. Suter, 1993) to focus the process.

An **assessment endpoint** is an explicit expression of the environmental resource that is to be protected. It is defined operationally in structural terms (e.g. a population of a particular species) or functionally (e.g. supporting processes that are typical of a particular habitat). If the protection goal is a population of a particular species of bird within close proximity to a contaminated site and we wish to preserve a viable population, we must stipulate an assessment endpoint (e.g. that the population size should not decline by more than a certain percentage). In reality, it will rarely be possible to carry out experiments on species defined within the assessment endpoint because it is likely these species will be endangered or protected. The ERA process therefore uses surrogate measures termed **measurement endpoints**. Measurement endpoints are quantifiable indicators that are related as closely as possible to the defined assessment endpoints. Therefore, in our example, the measurement endpoint may be the number of viable offspring per female of a species of bird for which test data have already been generated, or can be generated. Ideally, test endpoints should relate to changes in population numbers but this will rarely be the case when relying on existing data. Both assessment and measurement endpoints need to be established with input from statutory consultees (e.g. English Nature).

The range of standard test methods that may be used to assess chemical effects on soil biota is limited (Crane and Byrns, 2002). Therefore, in practice, it will rarely be possible for the assessment and measurement endpoints to be the same. This discrepancy may be dealt with through extrapolation (e.g. the use of safety factors) to allow for the uncertainties involved in making predictions about effects on one species/endpoint using data from another species/endpoint. Typically, larger safety factors are employed when data are limited to just a few species and endpoints i.e. when uncertainty is greatest. An alternative approach involves the use of species sensitivity distribution models to predict chemical concentrations that should protect a given proportion of species (Posthuma *et al.,* 2002).

Decision making

At this point, assessors need to make key decisions about the potential for harm to ecoreceptors. They may decide for example that there is no harm and hence no need for management action. Or they may decide that monitoring is needed to ensure conditions do not deteriorate, or that there is evidence of adverse effects for which management action is required. In order to come to such a decision, though, they may need first to take steps to reduce any outstanding uncertainties (Tier 3).

As we explain in Section 4, we favour a weight-of-evidence approach in which the significance of the available ecological data and measured responses in a series of toxicity tests are weighted according to their relevance to defined assessment endpoints, reliability and the magnitude of effects seen. At Tier 2, this principle is extended to consider the available evidence in its entirety, including not only biological effects data but also chemical data and, where they are available, ecological survey data as well.

Several outcomes are possible at Tier 2:

1. If no biological effects are evident from either ecological surveys or toxicity testing, and if chemical residues are not readily bioavailable under present or anticipated future conditions, then the **ERA process may be terminated**.
2. Where biological effects are evident and there is a clear cause-effect link with certain chemical contaminants[2], then **risk management** of the contaminated areas is merited. At the very least, this would involve a requirement to monitor to ensure conditions do not deteriorate. But it might also set out possible remediation options, ideally with an assessment of expected benefits and costs. Where there is clear evidence of harm and adequate information for risk management, **no further investigative work is required**.
3. It is possible that biological effects are evident but cannot reliably be linked to the presence of chemical contaminants. While this may be sufficient to classify the site as one in which ecoreceptors are at risk, the lack of a link with chemical data makes it difficult to develop a suitable management plan, or at least, would limit remediation options. Under these circumstances, **further site-specific investigation is warranted (Tier 3)**
4. Remaining uncertainties in the evidence for chemical contamination or biological effects (or both) prevent decision-makers from deciding with adequate certainty whether risk management action is appropriate. Often, the use of conservative assumptions in the face of uncertainty results in inflated predictions of risk, so further work to explore key uncertainties is merited. Again, under these circumstances, **further site-specific investigation is warranted (Tier 3)**.
5. For sites where **bioaccumulative compounds** are an issue, unless specific work has been done to explore their risks to the assessment endpoints, **more advanced work is needed in Tier 3 to look at this pathway.**

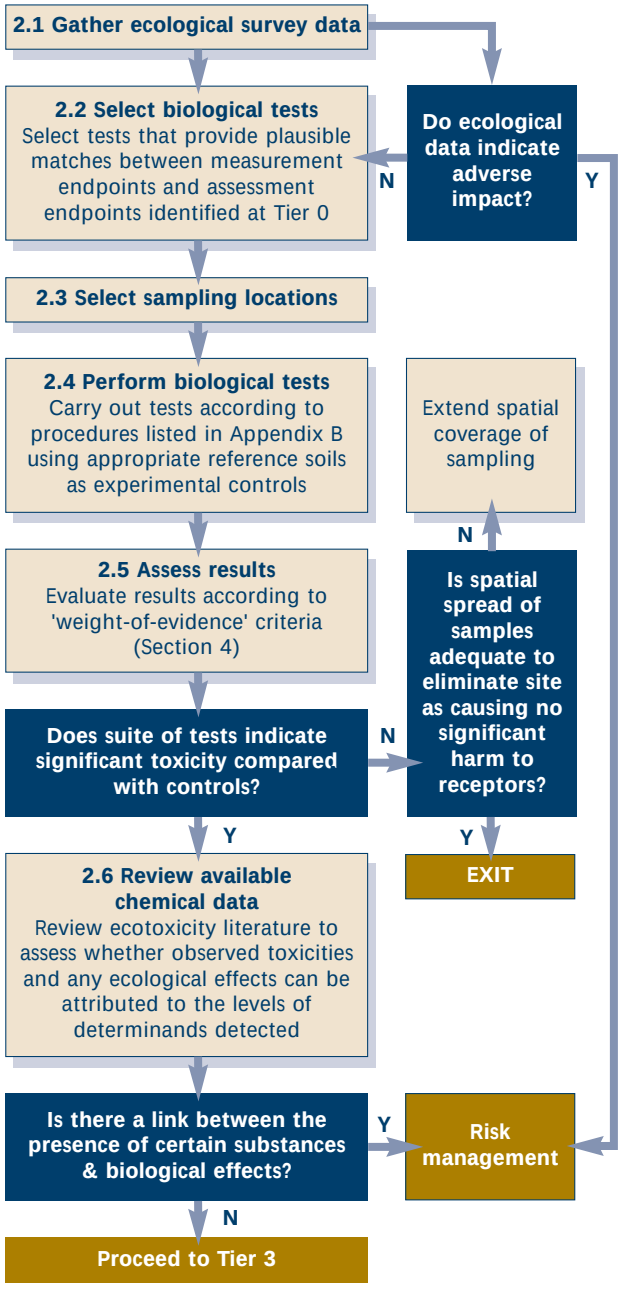


Figure 2.3 Activities and decision-making in Tier 2

[2] A plausible relationship between the contaminant and species sensitivity is evident, residues of the chemicals of concern are coincident in space and there is a clear relationship between residue levels and toxicity or some relevant ecological parameter

The report of work done at Tier 2 should include the risk assessor's scientific assessment of the risk of adverse effects. Any assumptions that have been made should also be reported. Specifically, the risk assessor should revisit the stated assessment endpoints (Tier 0) and characterise the nature and magnitude of the potential risks identified in the light of the evidence gained from the measurement endpoints used in Tier 2. Again, it is important to document areas of uncertainty and gaps in the data. It may also be useful to refine the contaminant-pathway-receptor linkages identified at Tier 0 if new data suggest that this is warranted.

2.3.4 Tier 3

The main reason for considering a progression to Tier 3 would be when the results of Tier 2 studies are equivocal. That is, whether there is a risk to biota at a particular site (that is, whether or not the 'harm' criteria for classification under Part IIA are met), and whether, therefore, management action is required.

This does not exclude the possibility of further studies to refine our understanding of ecological risks. That, though, is a different question from the one about site classification. Tier 3 helps determine the **magnitude** of the risks to ecoreceptors over and above there being a 'significant risk of significant harm'. This can be useful when it is necessary to prioritise resources, or to help select remediation options. In practice, though, few sites are likely to advance to Tier 3.

Refining effects estimates

Risk assessors frequently highlight the need for ecological risk assessments to consider risks to populations, not simply to individual organisms. For some species (e.g. protected species or valued species such as songbirds), individual organisms are highly valued. For the great majority of species, though, deaths of individuals probably go unnoticed and management action may only be taken in the face of perceived risks of local extinction, greatly reduced abundance or food chain issues with bioaccumulative chemicals.

While the ERA framework considers risks to populations to be the assessment endpoint, single species toxicity tests tend to describe effects at the level of the individual. The same is true of the biological tests used directly to assess soil quality. Though test endpoints are ecologically relevant, in the sense that they describe effects on survival, growth or some ecologically relevant function, they do not relate directly to measures of population sustainability. Tier 3 is therefore more likely to require the application of ecological theory and modelling, either to refine estimates of exposure or better to assess impacts at different levels of biological organisation.

Ecological models may be used to translate the results of fecundity and mortality measured in toxicity tests to estimate effects on populations and recovery times (for example, Klok *et al*., 1997; Kammenga *et al*., 2003; Pastorok *et al*., 2002). They are, though, demanding in terms of the data required. Typically, only long-term studies in which survivorship, fecundity and the timing of reproduction are all monitored generate useful data. Where key questions remain about the ecological significance of short-term ecotoxicity tests for effects at the population level, the generation of such test data may be warranted. Alternatively, direct measurement of population changes, for example, numbers of breeding pairs of a particular raptor species, or number of flowering spikes of a rare or sentinel plant species, could be made. Population changes can, though, often be explained by multiple factors, often unrelated to chemical exposure (e.g., changes in habitat, changes in food supply, natural variability). This makes it more difficult to prove causality.

An alternative approach is to move from reliance on *ex situ* testing to *in situ* testing. This is particularly useful where questions of contaminant complexity or bioavailability are important but remain intractable. An example of this is the test systems for soils, such as mini-container tests for organic decomposition (Spurgeon *et al*., 2002).

Tier 3 would require the application of more ecological theory and would probably need modelling. Many ecologists recognise the value of population and ecosystem modelling as applied to ERA for toxic chemicals in soils. Ecological models are used to translate the results of fecundity and mortality (measured in toxicity tests) to estimate effects on population, ecosystem and landscape endpoints. Endpoints that could be considered in risk modelling include species richness and population age, structure and productivity. Ecological models can be used to address two critical questions in site restoration and assessment: 1) how does population growth rate change as a function of toxic chemical concentration? And 2) how rapidly can a population recover from an impact due to transient exposure to a toxic chemical? Further development and use of ecological models with population, ecosystem and landscale endpoints are clearly needed to increase the value of the ERA process for environmental managers. The proposed risk assessment framework has the potential to involve ecological modelling

within Tier 3. It is beyond the scope of the current project to examine this in great detail. It is, though, clearly an important stage. Pastorok *et al*., (2002) outline this in much detail in their recent book *Ecological Modelling in Risk Assessment*. The Society of Environmental Toxicology and Chemistry's Ecological Risk Assessment Advisory Group has a subcommittee working on population modelling for risk assessment, and its work will be another resource in the near future. The Environment Agency has commissioned an R&D project (SC030003 'Review of ecological models for use in Tier 3 of the Agency's ERA framework') to identify models of potential use in Tier 3 assessments.

Food web modelling is another tool that is commonly used to support ERA at higher tiers of assessment. Simple food chains or webs can be built using spreadsheets, using a combination of measured data and assumptions to predict loading of bioaccumulative chemicals to higher trophic levels. To evaluate effects (harm), doses and predicted tissue concentrations are then compared to literature values for surrogate species.

Refining exposure estimates

Byrns and Crane (2002) draw attention to the fact that the primary focus of most assessments will be organisms at higher trophic levels (birds and mammals) and that exposure from a variety of routes, including via the food chain, may be important. An estimation of the total dose received by such organisms can be based on knowledge of diet, feeding behaviour, and contaminant concentrations in food, soil, sediment and water. The levels of exposure for these species can be estimated and compared with (oral) toxicity data for relevant receptors using exposure models (mostly not commercially available), probably in conjunction with newly generated data on contaminant residues in prey items. Beyer *et al*., (1996) provide guidance on how to interpret body residue data.

Probabilistic methods

So far, we have concerned ourselves entirely with deterministic assessments of risk, in which a single value representing the whole of the exposure dataset is compared with a single value representing the entire effects (toxicity) dataset. Often, we use conservative values from each dataset, for example, toxicity for the most sensitive species or endpoint, and an estimate of worst-case exposure. This approach is current standard practice and should be somewhat precautionary, that is, biased towards over-estimating risk. A major disadvantage of this approach, though, is that the *degree* of protection it provides is difficult to assess: conservative assessments might be over-conservative, leading to a waste of resources assessing or managing risks that are actually small. But they may sometimes be under-conservative. A second (and related) disadvantage is that by producing a single value for the risk, the deterministic approach fails to communicate both the variability of real outcomes and also the degree of scientific uncertainty about those outcomes. Both these disadvantages can potentially be addressed by using probabilistic methods.

Deterministic methods use single numbers to represent each component of a risk assessment. Probabilistic methods, by contrast, use distributions to quantify variability and/or uncertainty in the inputs and output of the assessment. Variability refers to natural variation that exists in the factors that determine whether effects will occur. For example, the variation of environmental concentrations in space and time, or the variation in toxicity between or within species. Uncertainty refers to the limitations in our knowledge about these factors. For example, measurement error in chemical analysis or toxicity testing, or sampling uncertainty due to limitations in the numbers of samples collected or species tested. Research reduces uncertainty but it does not reduce variability. Probabilistic methods quantify variability and/or uncertainty for the various factors that influence risk and propagate them through the risk assessment calculations. Probabilistic methods thus quantify the variability and uncertainty of predicted effects. A probabilistic assessment that quantifies both variability and uncertainty can estimate both the frequency and magnitude of effects. It can also put confidence bounds on those estimates to show the degree of scientific certainty attached to them. This has important advantages for the decision-makers. First, it enables them to choose what frequency and magnitude of effects they are willing to tolerate, and how certain they want to be that the effects are within acceptable limits (that is, control the degree of precaution). Second, it enables them to identify which sources of variation and uncertainty have most influence on the risk estimate. This, in turn, enables decision-makers to target risk reduction measures on the most important control variables (if the risk is too high), or target additional investigation on the most important uncertainties (if the risk is too uncertain).

The best-known probabilistic method is Monte Carlo simulation (for example, Vose, 2000), and in particular the simple 'one-dimensional' Monte Carlo (1D MC). This can propagate both variability and uncertainty, but it does not separate them. This is a significant limitation, as the distinction between

variability and uncertainty can be important to decision-makers (for example, a 10 per cent chance of a 90 per cent effect may have different implications from a 90 per cent chance of a 10 per cent effect). It may therefore be preferable to separate variability and uncertainty, using two-dimensional Monte Carlo (2D MC). Whether uncertainty and variability are separated or not, it is important to include them both as far as is possible, to minimise underestimation of the range of possible outcomes. Most examples of probabilistic ecological risk assessment to date have, though, used 1D MC and concentrated primarily on quantifying variability (for example, Moore *et al*., 1999).

There is a widespread but incorrect perception that probabilistic methods are feasible only with very large datasets. Monte Carlo methods require that the assessor specifies the shapes and parameters of each distribution, and the dependencies (for example, correlations) between variables. Very large datasets would be required to do this precisely. It may, though, be possible to conduct a valid analysis with more limited data if the uncertainties resulting from the data limitations can be incorporated in a 2D MC. This is straightforward for some types of uncertainty, for example, sampling uncertainty for normal distributions, though it is rarely done. Other types of uncertainty are harder to quantify. An alternative is to adopt non-Monte Carlo approaches such as probability bounds (P-bounds, Ferson, 2002). P-bounds methodology requires no assumptions whatsoever about distribution shape or dependencies, and is designed specifically for use with minimal data. But it provides only conservative bounds on the risk distribution (no central estimate). A promising new approach is to combine these methods, using 1D MC to provide a 'best estimate' of the risk distribution, and overlay on it the conservative bounds provided by P-bounds. Regan *et al*., (2002) provide an example of this for a food-web model involving contaminated land. Similarly, P-bounds could be overlaid on a 2D-MC. These combinations are new and require further development. But they are beginning to be used alongside more conventional methods in some regulatory assessments in the USA (for example, a Superfund assessment for the Housatonic River, to be published shortly, see http://www.epa.gov/ne/ge/).

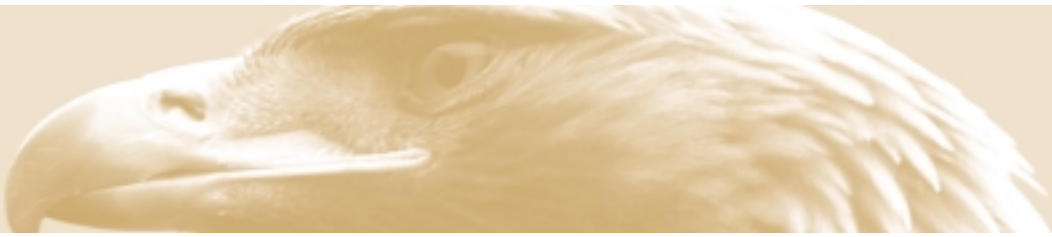
A number of other concerns about probabilistic methods are frequently raised. One is that if uncertainty is included, the confidence bounds may often be too wide to be useful for regulatory decision-making. This is somewhat perverse, because the bounds represent real uncertainty that decision-makers should take into account: if the science is very uncertain, this should surely be communicated to decision-makers, not hidden in deterministic assessments. A contrasting concern is that the bounds may be misleadingly narrow, because it is never possible to quantify all uncertainty (you don't know what you don't know). This is a serious limitation, but a practical solution may be to quantify uncertainty as far as possible and present the resulting bounds as minimum estimates, accompanied by a qualitative discussion of unquantified uncertainties. This seems preferable to relying entirely on a qualitative discussion of uncertainty, or on arbitrarily chosen 'uncertainty factors' whose degree of precaution is unknown. Another important concern is the difficulty of communicating unfamiliar and potentially complex probabilistic assessments to decision-makers and the public. This is to be expected at the current stage of development. Effective communication methods must be found if probabilistic methods are to find wide application.

Finally, there is some concern about the lack of guidance on probabilistic methods, especially given the diversity and complexity of options available. Again, this is an issue that will need to be addressed if probabilistic methods enter routine use. The US Environmental Protection Agency (2001) has already published an extensive guidance document on conducting probabilistic assessments for contaminated land (Superfund sites). This concentrates mainly on 1D and 2D MC approaches. But it also includes a brief discussion of Bayesian approaches (as does Vose, 2000). It contains much useful information, and it could be very helpful in developing guidance for other jurisdictions.

2.4 Summary

In this section, we have outlined the proposed decision-making framework to identify sites where ecoreceptors may be subject to risk from contaminants. We advocate a tiered approach based largely on the recommendations of Byrns and Crane (2002) but with some modifications, to ensure that sites where there is acceptable risk of adverse impacts exit from the process at the earliest opportunity. At the same time, the framework seeks to identify those receptors that are at risk and where some form of risk management will need to be considered. We emphasise the use of biological testing, especially at Tier 2, to assess the risk of harm to ecoreceptors. In Section 3, we review recent experience with the use of biological testing, with particular reference to contaminated land.

3



Experience in the use of biological testing with particular reference to contaminated land

The regulation of environmental quality has conventionally been based on the use of chemical-specific thresholds. The soil environment, where the use of SQGVs has been paramount, is no exception. It is, though, important to rememeber that these are ultimately biologically based, i.e., they are based on assessing the biological effects of a chemical by estimating (in laboratory or field studies) a concentration that causes no effects.

3.1 Introduction

Before the late 1980s, the assessment of soils was limited to physical and chemical characterisations. Though SQGVs have been developed by some jurisdictions (see Section 2.3.2 for examples) they are not available for all the contaminants that could be encountered. Furthermore, quantifying chemical concentrations alone may not always be adequate to assess potential adverse environmental effects. This is either because of possible interactions that underestimate the effects of chemicals in combination, or because there are simply not the thresholds available for all the chemicals that might be encountered. Also, the relationship between total chemical concentrations and biological availability is strongly site-dependant and is not always predictable. So the actual biological impacts associated with an exceedance of regulatory chemical thresholds can be difficult to interpret. Under these circumstances, direct biological testing may be valuable for assessing the actual risk of environmental harm.

To date, the use of biological testing in a regulatory context has been confined to surface waters, aquatic sediments and groundwater. In some countries, notably the US and Canada, Whole Effluent Toxicity testing is an important component of regulatory schemes to control point source emissions to surface waters (for example, Power and Boumphrey, 2004). In the UK, a programme of research commissioned by the Environment Agency has led to proposals for a similar approach to be adopted for effluent control (Direct Toxicity Testing) within certain regulatory regimes.

In contrast, experience of the use of soil toxicity testing as a means of assessing, classifying or otherwise regulating soils, is limited, particularly in the UK. Recent international examples include investigations in Canada by Carter *et al.,* (1998). These examined the usefulness of a battery of soil ecotoxicity tests at three sites. Sunahara *et al.,* (2002) have summarised recent international efforts combining chemical and biological approaches to

dealing with contaminated land sites. In the UK, a report published by the Construction Industry Research and Information Association (CIRIA, 2002) examined the suitability of biological methods for the assessment and remediation of contaminated land. It suggests that biological methods for the assessment of contaminated land are increasingly being used to assess risks to ecological receptors. The report examined a battery of standardised soil ecotoxicity tests. These indicated that testing of soils from a site could take less than 24 hours (but average durations were typically much longer, that is, 14-28 days). The authors were keen to see biological test methods developed further, pointing out that the tests potentially offer a more direct appraisal of risk than is possible using chemical methods alone. The report also included a series of case studies demonstrating examples where biological testing has been used to assess a wide range of contaminants under scenarios that were representative of many industrial sites in the UK.

Commercial applications of biological testing in the UK have been restricted mostly to monitoring the progress of bioremediation processes (using microbial measures) and directly measuring bioavailability of soil contaminants for risk assessment purposes. CIRIA (2002) suggested the outputs from such studies are less conservative than purely chemical-based approaches to exposure assessment, because they address only the bioavailable fraction. The same report concluded that ecotoxicological testing in soils may offer the scope for cost savings on contaminated land projects by using relatively simple biological tests to screen for bioavailable contamination. Additionally, it concluded that a contaminated site might not need remediation if the bioavailability of contaminants to relevant receptors is shown to be low.

A recent review of the Contaminated Land Rehabilitation Network For Environmental Technologies in Europe (CLARINET) (Bardos, 2002) summarised the findings of seven various working groups. Working Group 5 (WG5) 'Ecological requirements for Land Reuse' is particularly relevant. It identified several deficiencies in the use of ecological methods, including how to interpret the heterogeneity of a site, variation in expert opinions, and the extrapolation of laboratory ecotoxicity data to the field. WG5 suggested that a tiered framework would be helpful, with increasing levels of sophistication and effort being expended, but only when circumstances demanded it.

3.2 Advantages and limitations of biological testing

3.2.1 Advantages

In Table 3.1, we highlight some of the advantages of adopting biological testing approaches. It is also important, though, to recognise their limitations, and we consider these in Section 3.2.

Table 3.1 Summary of advantages of biological testing over chemical-specific approaches to the assessment of environmental quality

Limitations of the chemical-specific approach	Significance to assessment of contaminated land	Advantages of biological testing
Thresholds can be established only for substances for which a national or international guideline value exists;	Interdepartmental Committee on the Redevelopment of Contaminated Land (ICRCL) guidelines were removed, there are no UK guidelines for protecting ecological receptors. There are, though, proposals for deriving Soil Screening Values (Environment Agency, 2003; Environment Agency, in press);	The effects of all substances can be accounted for, including those for which no guidelines or thresholds have been derived;
Standards can be derived only where sufficient (eco)toxicological information is available;	There are many more aquatic toxicity data than there are terrestrial toxicity data. For example, the USEPA ECOTOX databases contain aquatic toxicity data for 7,300 chemicals and 4,190 species. The corresponding TERRATOX database contains data for 2,950 chemicals and 1,840 species*. Significantly, much of the available soil toxicity data is simply anecdotal observation, which cannot be used in defining thresholds;	Determining effects does not depend on existing ecotoxicological data;
Substances are rarely found in isolation in soils and, where complex mixtures are present, there is the potential for interactions between substances;	It is difficult to predict the effects of antagonistic, additive or synergistic interactions of substances in a complex mixture in terms of deriving acceptable concentrations for each substance;	Interactions between substances are effectively integrated in the overall observed response;
Only total concentrations are generally measured, including the non-bioavailable fraction.	Basing decisions solely on total concentrations of a substance may overestimate the actual risks to resident biota.	Only bioavailable substances contribute to the observed responses.

* A search based on 150 different commonly occurring hydrocarbons yielded 4,700 separate toxicity values for aquatic species and only 1,500 for soil dwelling species. Aquatic data could be located for 66/150 chemicals, but soil toxicity data could be located for only 35/150 chemicals.

3.2.2 Limitations

The use of biological testing in a retrospective assessment poses a number of challenges that are not evident when they are used in prospective assessments for example, for chemical testing. The main limitations are:

1. The concentration of chemical contaminants may be too low to elicit a biological effect in a short-term (acute) test, even though it could cause an effect after a long period of (chronic) exposure. This means that there is a risk of underestimating toxicity when only acute tests are used. At their present stage of development, soil tests are predominantly acute tests, though some chronic methods have recently been reviewed and evaluated (Spurgeon *et al*., 2002; in press).

   Practitioners can overcome this limitation to some extent when testing pure chemicals simply by increasing the concentration of the substance they are evaluating. This is, though, simply not possible in the direct biological testing of environmental samples such as soils or sediments, where they are limited by the concentration already present. In addition, chronic low-level exposures can sometimes cause greater effects than acute, high-level exposures. In practice, this means that ideally longer-term (chronic) tests would be employed, though this can pose practical difficulties with environmental samples because test animals must be fed during the exposure period.

2. The volatility of some potential contaminants will cause concentrations to alter significantly due to biodegradation during the test period, particularly in tests that take longer (for example, 28 days).

3. Immobilisation of nitrogen by microbes degrading organic contaminants in the soil could cause nitrogen limitation for plant growth; this could be misinterpreted as an effect of the contaminants.

4. Reported poor plant growth during testing could be due to competition for oxygen in the root zone when there are large amounts of biodegradable organic compounds present in the sample.

5. Direct toxicity testing can tell us whether contaminants are present at toxic levels (and where their effect is greatest on a site). But the identity of the contaminants responsible may be unknown. Identification can require a good deal of diagnostic work for example, through Toxicity Identification Evaluations (TIE). While TIE protocols have been developed for wastewaters, receiving waters and sediments, none have yet been developed for soils.

6. Biological tests of environmental samples are prone to interferences that are unrelated to contaminants and toxicity. For example, differences in water content, nutritional status, physical structure and organic matter content of soils could give rise to differences in biological response, even when there are no differences in the levels of contamination, or when no contaminants are present. Recent evidence from Spurgeon *et al.,* (in press) shows that bioavailability and toxicity can vary seasonally.

There is consequently a real difficulty in identifying the control condition - should it be based on a pristine soil that is known to be free of contaminants (but which could differ from a soil from a site of concern in the way suggested above)? Or should soil from a location close to the site of concern, but which is believed to be contaminant-free, be used (but there remains uncertainty about this assumption and also the similarity in physical soil characteristics)?

In practice, a combination of negative, reference and positive controls may be used:

- Negative control - an uncontaminated soil in which the test organisms can thrive and which is expected to produce low (or no) adverse responses. Limits for acceptable responses in controls have been defined for some tests (Section 4) and may be used to evaluate the performance of tests prior to analysis of the test data.
- Reference soil - a soil with grain and other physicochemical properties that match the soils from the site of concern. Unfortunately, such a set of standardised soils does not exist and assessors of contaminated sites must therefore rely on 'clean' reference sites collected from locations near the site of concern.
- Positive control - a soil to which a toxic chemical has been added at a level that is expected to elicit a given level of response (for example, 50 per cent response over the normal test duration). In practice, a positive control is used more crudely, simply to check that exposed organisms respond (that is, a pass/fail assessment). Under more sophisticated regimes, these data may be used to establish warning charts with a reference toxicant, as has been developed for aquatic toxicity tests used in DTA (Environment Agency, 1999).

Conventional methods of toxicity data analysis involve comparing response levels in treatment groups with those in controls and reference sites (e.g., Analysis of Variance, ANOVA and other hypothesis testing methods for normally distributed data) or inferring point estimates e.g., EC50 from regression. In both cases, the response levels in the controls/reference samples are critical. The problem remains about which of the negative and reference

soils to choose as the basis for the control response when analysing the test data, especially when there is a significant difference between their response levels. This could have a marked influence on the judgement of the level of toxicity associated with a particular site and on progress through subsequent tiers of the framework. We return to this issue in Section 7.

3.3 The selection of biological tests for contaminated site evaluation

As outlined in Section 1, biological testing may be performed *in situ* (that is, biological receptors are deployed and monitored in the field) or *ex situ*, (i.e. soil samples are collected and tested under relatively controlled laboratory conditions). Though test methods and decision-making frameworks have been developed for biological testing in the aquatic environment, progress in the use of biological tests for assessing soil quality and reaching regulatory decisions lags behind. Currently, risk assessors or operators are not formally required to generate biological data as part of a risk assessment for contaminated soil.

OECD and ISO have developed and standardised a number of soil toxicity tests. The approach usually adopted is to expose representative soil species to contaminated soils for relatively short periods of time. This is typically up to 14 days to assess acute toxicity, up to 28 days to assess potential chronic effects, and longer periods to assess bioaccumulation potential. These are, though, intended primarily for hazard assessment of chemicals via the soil medium for possible use in a *prospective* risk assessment rather than to assess soil quality directly. Many methods used were adapted from standard tests for chemical testing. Even aquatic tests have been investigated (see review in Crane and Byrns, 2002). Aquaterra (1998) and Riepert and Wilke (1998) have identified tests that they considered useful in assessing contaminated soils. Carter *et al*., (1998) compared a suite of tests and found that acute tests with earthworms *(Eisenia andrei)* were less sensitive than those with lettuce *(Lactuca sativa)* which, in turn, were less sensitive than the Microtox™ test.

To our knowledge, only one biological test method designed specifically for assessing soil quality *retrospectively* has been standardised and published. This is a growth and reproduction test based on responses of the plants, *Brassica napus* (oilseed rape) and *Avena sativa* (oat) (in press). As far as we are aware, OECD is concerned only with the development of methods for testing single substances.

Recently, two studies have sought to identify *ex situ* test methods that may be useful for assessing soil quality:

(i) Crane and Byrns (2002) reviewed tests with biomarkers, microbes, plants, invertebrates, vertebrates and microcosms. They also recommended approaches for sampling soil for biological testing and the subsequent statistical analysis of toxicity data from such tests.

(ii) A broadly similar approach was adopted by Spurgeon *et al*., (2002) (SubAssess project) who reviewed a wide range of biological tests with sublethal endpoints that could be used both *ex situ* and i*n situ*. They summarised each method and appraised performance characteristics such as reproducibility, responsiveness, robustness and relevance. This review was not confined to whole organism tests with conventional survival, growth or reproduction endpoints. It also included tests assessing effects on soil function, life-cycle bioassays and biomarker responses. These authors also considered a range of potential ecological and genetic measures that could be applied to communities, individual organisms or tissues.

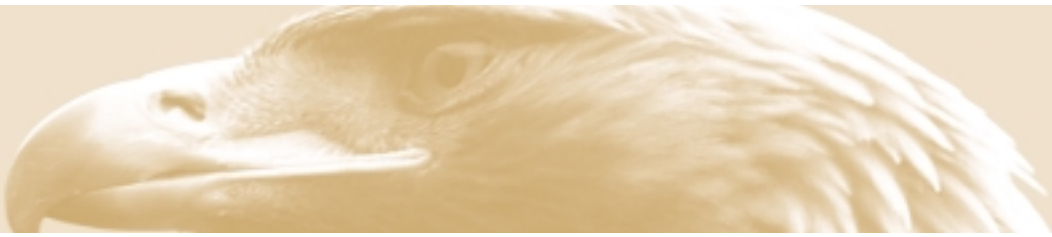
The following seven test methods were highlighted in the context of assessing soil quality from the Crane and Byrns review (2002) (see Appendix B for details):

- Microbial bioluminescence tests with *Vibrio fischeri* for example, Microtox™ and Mutatox;
- Microbial carbon mineralisation test;
- Microbial nitrogen mineralisation test;
- Germination and root growth test with a monocotyledonous plant species for example, wheat and dicotyledonous plant species for example, cabbage;
- Acute lethal earthworm test with *Eisenia fetida* or *E. andrei;*
- Acute springtail test with *Folsomia candida.*

ISO or OECD have standardised and published all these methods, with defined test validity criteria. The tests were designed for testing soils into which known contaminants had been added. But it was felt that the tests could also be applied to retrospective assessments where the level and type of contamination is usually unknown. The recommendation to test a range of species at different trophic levels can provide useful information about possible interspecies differences in sensitivity.

Section 4 considers how to evaluate the performance of these tests and use the results in making decisions about whether to progress a site through the tiers of the ERA framework.

4



Performance assessment of biological tests

4.1 Introduction

In this section we discuss how to judge the performance of biological tests in order to help reach decisions about the level of risk posed to soil organisms at contaminated sites. There are two aspects to this:

1. The performance of the test methods themselves.
2. The usefulness of data generated in determining whether a soil is subject to significant harm that is, in decision-making.

4.2 Test methods

4.2.1 Test validity criteria

The test guidelines describing the seven test methods referred to in Section 3 all include test validity criteria. These are features that must be met for the test to be regarded as satisfactory (summarised in Table 4.1). If these criteria are met, it may be reasonably assumed that any resulting conclusion about toxicity is reliable.

Typically, these validity criteria specify the maximum acceptable level of effect in control groups and also a minimum level of response in treatment groups (that is, samples from the site of concern) above which an effect may be judged as 'real'. In other words, they predefine the minimum difference between controls and treatments[3] that would be regarded as an 'effect'. This is in contrast to the conventional approach, in which the level of variability between replicates within a treatment or within the controls determines the least significant difference.

4.2.2 Variability in test outcomes

Like any other measurement technique, biological tests can give different responses to the same level of a contaminant. This depends on a range of factors, for example, where the test organisms are sourced, the environmental conditions under which they are held, method of data analysis and even the skill of the test operator. As a result, the same test method and toxicant can give markedly different outcomes between laboratories and even within laboratories (Whitehouse *et al.,* 1996). The use of standardised test methods such as those published by ISO and OECD goes a significant way to restricting the latitude available to an experimenter in carrying out a particular procedure. The intention is usually to avoid physiological stress on test organisms and ensure analyses are performed in a robust fashion. This standardisation also serves to restrict the variability in outcomes that can result. Some OECD test methods, though, still give a wide degree of latitude in the test species that may be used, test containers and soil types. All of these can result in differences in measured toxicity, even with the same toxicant.

Reference toxicants (positive controls) can be used as a means of judging consistency within and between laboratories (Environment Agency, 2000). But specific measures for soil ecotoxicity tests have not yet been developed. Though ring-tests of some methods have been performed, the resulting data have not generally been used to define Quality Control criteria. Some tests periodically test stock animals with a reference toxicant (for example, Betanal Plus for springtails (ISO 11267)) as a means of checking for any 'drift' in sensitivity over time.

These sometimes specify the concentration range within which specific responses are to be expected. The accuracy and precision of tests requires repeat testing in a number of laboratories and research based on aquatic toxicity tests (Whitehouse *et al.,* 1996) suggests that at least six such tests are needed to generate plausible estimates of precision and accuracy. This is an important performance characteristic of testing, especially if the data are to be used for regulatory decision-making, but not an area of investigation covered by this project.

[3] Actually the mean of the responses in replicates from the control and treatment groups

Table 4.1 | Test validity and assessment criteria for selected tests

Test method	Test validity criteria	Assessment criteria*		Reference
		acute	chronic	
Microtox™	Phenol reference test results needs to be within 16 and 20 mg/l for a 15 min EC50	None	None reported	Environment Agency (2000)
Carbon Mineralisation Test	Variation between replicates in the control must be <10%	N/a	End of test difference in variation between treated and control soil samples must be >15% at any point after day 28 to indicate an effect	OECD, 2000a
Nitrogen Mineralisation test	Variation between replicates in the control must be <10%	N/a	End of test difference in variation between treated and control soil samples must be >15% at any point after day 28 to indicate an effect	OECD, 2000b
Terrestrial Plant Test 208a Seedling emergence and seedling growth test	Seedling emergence and seedling growth test a minimum of 65% emergence The mean seedling growth does not exhibit visible phytotoxic effects	None	None	OECD, 2000d
OECD Earthworm test	Percentage mortality of adults in the control must be <10%	>20% adult mortality	Reduction >50% of the number of offspring compared to control. A reduction of 20% of the biomass after 28d compared to the start of the test	OECD, 2000c
ISO Collembola Test	Mortality of adults in the controls should not exceed 20% by end of the test Coefficient of variation in the control should not > 30%		Reproduction rate must be a minimum of 100 instars per control vessel Coefficient of variation in the control must not exceed 30%	ISO 11267: 1999

*minimum levels of response below which effects are not considered to be 'real' effects. N/a = Not applicable.

4.3 Decision-making

4.3.1 Weight-of-evidence

Most people intuitively understand the concept of 'harm'. But this concept presents difficulties in a regulatory ERA because it is laden with value judgements and perceptions that make it open to interpretation. The problem is exacerbated further by the fact that we must combine information from a variety of sources (chemical, ecotoxicological and ecological) in a way that enables the risk assessor to make an overall judgement about the risk of significant harm at a particular site. If acceptability were determined simply by compliance with a single threshold, this would be a relatively trivial matter, perhaps as simple as a pass/fail decision. Where a triad approach is adopted - as advocated here - the determination of harm will not usually be straightforward. Instead, it requires a judgement that integrates all the available information. This leads to a decision based on the weight of all the available evidence.

Under the proposed ERA framework, a decision about the acceptability (or otherwise) of a site is

based on an assessment of the risks identified at Tier 0. In this risk assessment, predicted or measured levels of contaminant exposure are compared to those deemed to be acceptable (that is, SQGV) or to direct effects being measured through testing and comparison with controls. When biological testing is used, the biological tests effectively integrate both exposure and effects in a single assessment. This contrasts with the chemical-specific approach, where they are derived independently. Therefore, decision-making based on biological tests is simply a question of whether or not an adverse effect has reliably been detected, and whether this is of sufficient magnitude to warrant further investigation. When both sorts of data are available, possibly with ecological data as well, there is a real challenge in coming to a decision that is transparent and auditable. Below, we outline an approach based on that advanced by Suter (1993) that sets out to meet this objective.

In reality, a site is assumed not to pose a risk of significant harm unless evidence is available to suggest otherwise. As we explain in Section 3, though, there is no guarantee that effects would be seen from conducting standard tests at a site that was truly contaminated. The tests may not be sensitive to the toxicants present, or test durations may be too short to elicit an effect. The relatively late discovery of the adverse effects of Tri-butyl-tin (TBT) to marine molluscs and endocrine disrupting effects of a range of synthetic substances are examples where true risks remained undetected because of reliance on a small suite of standard tests. It follows that risk of such a false negative is reduced by deploying a large range of species representing as wide a range of potential receptors as possible. A key aspect of this project is to judge how many tests are necessary, which are the most useful tests in order to reduce the risks of false negatives to an acceptable level, and how these data should be integrated with chemical and ecological information.

4.3.2 The weighting of evidence

In the tiered approach outlined in Section 2, significant linkages (or the possibility of significant harm occurring) will have been identified in the CSM and the problem formulation stage. The subsequent risk characterisation determines whether these risks are significant for each identified receptor. It then attempts to determine the magnitude of the risk (that is, the extent of any effects) and the associated uncertainties. In a weight-of-evidence approach, all available data (e.g. from chemical analyses, toxicity testing and other available data) are used to estimate the likelihood that significant effects are occurring or are likely to occur, and to describe the extent of these effects (Suter, 1993).

The process of weighting the evidence effectively estimates the level of risk that is most likely, given all the available data. If the assessment endpoint is defined in terms of a threshold, such as a difference between control and treatment group responses of >20 per cent, then the process can be performed in two steps:

1. Examine the outcome of each individual test result independently and draw a preliminary conclusion. Has the measured response exceeded the minimum level of response in treatment groups above which a significant effect may be concluded (Table 4.1)?
2. Determine whether the results, taken together, indicate that it is likely that a risk of significant harm will arise. If there is no bias in the assessment that affects all lines of evidence, and all tests yield consistent outcomes, then it is reasonable to draw a clear conclusion (of an adverse effect or no adverse effect). If, though, there are inconsistencies between tests, then a process of weighting must take place.

Suter *et al*., (2000) describe this process in detail and considers a number of factors that will affect this assessment, such as:

- Relevance of the test. More weight is given to measures of effect that are more directly related to the assessment endpoint (Section 2.2). In other words, where there is concern about a particular plant species, then a test with a closely related species would be more relevant than, say, testing using Collembolans.
- Exposure-Response. More weight is given to data that demonstrate a clear relationship between magnitude of exposure (concentration) and effects for example, from dose-response tests.
- Temporal scope. The test should consider a range of temporal variances relevant to the site and its future intended use.
- Spatial scope. Testing is performed on samples that are representative of the area of concern.
- Quality. More weight is given to data generated using standardised protocols and to studies that are properly replicated, executed and interpreted.
- Quantity. More weight is given to a large quantity of data than to a small body of data.

- Uncertainty. A line of evidence that estimates the assessment endpoint with low uncertainty should be given more weight for example, test species and routes of exposure are relevant to assessment endpoints of concern (this also relates to relevance and quantity of data).

Suter *et al*., (2000) recommend a simple scoring system of + or - to summarise test results. A '+' is assigned if test data are consistent with significant adverse effects, and a '-' if test data do not reveal significant effects. If the data are ambiguous (for example, a positive response was seen but test validity criteria were not met, or no response was seen in the positive control), a +/- notation is assigned. This is the system we will adopt to assess the test results in the present study.

The final conclusion is not based simply on the relative number of + or - signs, but also on the reliability of the conclusions drawn from various lines of evidence. This still leaves the final decision to a process of expert judgement, though it attempts to make the reasoning transparent.

Table 4.2 illustrates this type of weight-of-evidence approach for a fictitious site. The same approach will be used in Sections 5 and 6 to assess the outcomes of the data generated in this project. Essentially, each type of data is scored, and the reasoning for the score explained. In Table 4.2, the outcome is that the soil in this instance is not sufficiently contaminated to warrant further action.

4.4 Decision-making in this project

In the following sections, we describe how a weight-of-evidence approach to decision-making may be employed when data from a variety of sources are available. This has been undertaken by detailed investigation at two different sites (Sites A and B). The ERA scenarios employed on these two sites are hypothetical. They demonstrate the usefulness of the ERA framework and proposed test suite. To evaluate the seven selected biological tests, the incidence of false negative outcomes can be judged from the responses obtained in tests on soil from a contaminated site where ecological receptors are

Table 4.2 Application of weight-of-evidence approach to interpreting data generated at a fictitious contaminated site (Suter *et al.*, 2000)

Test option (evidence)	Result (test outcome)[a]	Explanation
Soil analyses/ single chemical tests (Tier 1)	+	High Total Petroleum Hydrocarbon content reported, literature data reports such levels are likely to be toxic to soil organisms. Significant adverse effects on earthworms would be expected. Relevant toxicity data for other detected compounds are unavailable
Earthworm acute toxicity test (Tier 2)	-	Soil did not reduce survivorship of the earthworm *Eisenia fetida*; sublethal effects were not determined
Body residue data (Tier 3)	+/-	Concentrations of PAHs in earthworms were seen to be elevated relative to worms from control sites
Biological survey data (Tier 2)	-	Soil microarthropod taxonomic richness is within the range of reference soils of the same type, and not correlated with the range of petroleum components
Final Decision	+/-	Though the earthworms tests may not be sensitive, they and the biological surveys are both negative. These are considered to be more reliable than the single chemical toxicity data estimated from the analytical results of the soil. Risks to higher trophic levels as a result of chemical uptake via the food chain have not been established.

[a] Results of the risk characterization for each line of evidence and for the weight of evidence approach
\+ indicates that the evidence is compelling and consistent with a significant biological effect (according to defined test criteria)
\- indicates that the evidence is inconsistent with the occurrence of a significant biological effect;
+/- indicates that the evidence is too ambiguous to interpret

known to be impacted. One of the selected trial sites within the project has considerable contextual data (Site A). This is useful, because the utility of each test in the decision-making process will be identifiable. Of course, most sites will not have such a wealth of background data available.

The incidence of false positives (that is, indicating adverse effects when none actually exist) is more difficult to assess. Data derived from reference and negative control soils may be used to address this. Of course, such prior knowledge will not be available for many sites, and it will thus not be possible to ascertain the incidence of false negatives or positives.

Testing the ERA framework: Site A

To test the application of the framework and the suitability of the tests for use within it, we selected two potentially contaminated sites for practical investigation.

5.1 Introduction

We took a decision early on that at least one of the test sites should be well characterised both chemically and biologically, and that both could act as surrogates for Part IIA sites. The sites selected were:

Site A: an area contaminated by aerial deposition from a primary cadmium/lead/zinc smelter located at Avonmouth in the South-west of England.

Site B: a former (demolished) tank farm area where crude oil and refined petroleum products used to be stored.

It is important to recognise that these sites were selected for illustrative purposes. For this reason, they were subjected to evaluation at Tiers 0, 1 and 2, even if they would normally exit the framework at an earlier stage. Because neither of these sites has direct relevance for Part IIA, it was necessary to superimpose potential Part IIA receptor scenarios onto existing conditions at each of the sites.

This section provides a brief overview of Site A, along with the superimposed scenarios mentioned above. It then describes the chemical and biological testing performed at Site A using Tiers 0, 1 and 2 of the tiered framework. Corresponding data for Site B are provided in Section 6.

5.2 Site description

Site A is close to the city of Bristol. It is a large area of mixed urban development and intensive heavy industry, with a patchwork of small, privately owned areas of cultivation. The area is impacted by the workings of one of the world's largest zinc smelters. Considerable inputs of copper (Cu), cadmium (Cd), lead (Pb) and zinc (Zn) are deposited as particulates over an area up to 20 kilometres from the site. The area of deposition has been modelled (Coglan *et al.*, 2002), which confirms a gradient resulting from the interplay between particle size and wind direction. The local soil characteristics have been described, most recently by Filzeck *et al.,* (in press).

This site is not regulated under Part IIA, but it is a useful demonstration-site for several reasons. There is more than 25 years' worth of published research about the smelter, its emissions and their ecological effects. Furthermore, those working on this project have personal experiences of the site. From this understanding of local biology, the value of the biological tests used here can more readily be assessed. For example, if biological testing failed to yield toxicity at locations where we would expect effects to be seen, the data or test methods may further be interrogated to establish the reason for this failure. This site therefore allows us to address the ecological relevance and risks of false negatives using the proposed framework.

Further detail about the history and geography of Site A is to be found in Spurgeon *et al.,* (in press).

5.3 Hypothetical ERA scenario

This scenario exploited a known gradient in metal contamination onto which a hypothetical ecological receptor - a Site of Special Scientific Interest (SSSI) - was superimposed, occupying an area under the deposition plume. This provides a background against which to assess the performance of the framework and the associated biological tests.

The SSSI used for the hypothetical ERA is based on the combined characteristics of two SSSIs that occur in south Gloucestershire and North Somerset (within 60 kilometres of Site A). This fictitious SSSI is an area of ancient broad-leaf woodland. It overlays the selected 10 patches (that is, distinct areas at different distances from the point source, subjected as a result to different levels of aerial deposition) as shown in Figure 5.1.

Figure 5.1 Sampling patches at site A in relation to primary cadmium/lead/zinc smelter. (Axes give Ordnance Survey grid reference values in kilometres. Shaded areas indicate the extent of major urban zone of Bristol. Major motorways (M4, M5, M48 and M49) are indicated.)

The 10 patches were along two transects from the smelter and are consistent with the sampling conducted under project P5-063 at patches A1-A5. We used these because they were consistent with previous research conducted at the site, thus increasing the applicability of available contextual data. Existing evidence indicates that the patches contain a range of metal contamination levels. They thus provide the different scenarios for the location of the ecological receptor and the required contaminant concentration gradient to evaluate biological tests. One transect, running in a north-easterly direction, is based on a series of previous studies that focused on the assessment of the biochemical basis of species responses to metals (Kammenga, 1997). Selected examples of these sites have also been used in assessment of community level responses of various soil dwelling invertebrate groups (Sandifer, 1996; Read *et al.,* 1998; Hopkin, 1989; Spurgeon and Hopkin, 1999) and small mammals (Read and Martin, 1993).

5.4 Tier 0 Assessment

5.4.1 Conceptual model development

The ecological receptor has been a designated SSSI since 1967. The site is ancient broadleaved woodland of approximately 153 hectares, located in the deposition plume of a metal refinery. The reserve is on neutral, heavy-clay soils, and an area of common land divides it into two distinct areas. One side of the site is low-lying and prone to flooding. This is dominated by marshland species. Wildlife surveys have only recently started and have recorded more than 1,000 species. This diversity represents the principal reason for the designation of the site. The woodland itself consists of oak, ash, field maple and yew trees. But it also contains small-leaved limes and wild service tree, which are found only in ancient woodland, along with many rare whitebeams. There is a luxuriant growth of ferns. Ground flowers include a fine spring display of primroses, bluebells, dog-violets, wood-anemones and early purple orchids as well as rare woodland plants such as green hellebores and bird's-nest orchids. If contaminants were deposited at phytotoxic levels, all of these species could be compromised. A range of small mammals feeds on the plants. They include field and bank voles as well as a variety of seed-eating birds. If these animals were to feed on contaminated plants, they could all be subject to contaminant exposure. As well as plant-eating birds, invertebrate feeders also live here. They include woodcocks, song-thrushes and snipe with duck species (on the wetland area). Woodland passerines are also found in abundance. For these invertebrate-feeding birds, an obvious potential contaminant-pathway-receptor linkage exists. Previous work in areas contaminated by metals from mining, industrial activities and the use of lead shot have indicated that metals are present in the kidney tissues of invertebrate-feeding species. These metals, the work has shown, occur at concentrations likely to cause pathological damage in human kidneys (Hopkin, 1989; Stansley and Roscoe, 1996; Read *et al.,* 1998). As an additional step in the food chain, predatory birds have also been recorded at the site. Buzzards, sparrowhawks and tawny owls have all bred here in the past 10 years. The red kite population is beginning to spread across the country, and this woodland is a potential breeding site.

5.4.2 *Site information and initial observations following a site walkover*

We visited each patch and conducted an initial site walkover. The aim was to establish suitable locations for soil collection, the distance of these locations from the point source, the presence of other local sources, the nature of the ecosystem present, and the suitability of the area as a potential location of the ecological receptor. On the basis of this information, we identified sampling areas and drew up a list of priority target chemical analytes for each patch. These are detailed in Table 5.1 along with brief summaries of the principal characteristics of the Site A area.

Geographical information: The area surrounding the smelter source at Site A is characterised by flat estuarine alluvium less than 10 metres above sea-level. An elevated zone adjoins this plain in the East. This area reaches up to 80 metres within the investigation area. Each of the 10 selected patches are located along two transects from the smelter. These run to the north and north-east of the smelter site.

Contamination sources: The smelter is the main-source of metal contamination in the Avonmouth area. The region is also home to a complex mixture of light and heavy industry, domestic sewage works, local waste management sites (including landfill sites and a disused refuse incinerator) and minor and major roads (see www.environment-agency.gov.uk for permitted activities). These all represent potential sources of local contamination that could influence the results of bioassays conducted in soils at these sites. The severity and length of the gradient of metal pollution from the smelter, though, represents the dominant contamination influence in the area (see R&D Technical Report P5-063/TR2, Spurgeon *et al.*, in press).

Sampling locations: Existing chemical contamination data reveal a gradient of metal contamination, and this was confirmed in the present study (Figure 5.2) (Spurgeon *et al.*, in press). For the purposes of this study, these data have been used to locate the designated ecoreceptor. The patches for chemical analysis and biological testing were selected to reflect this gradient.

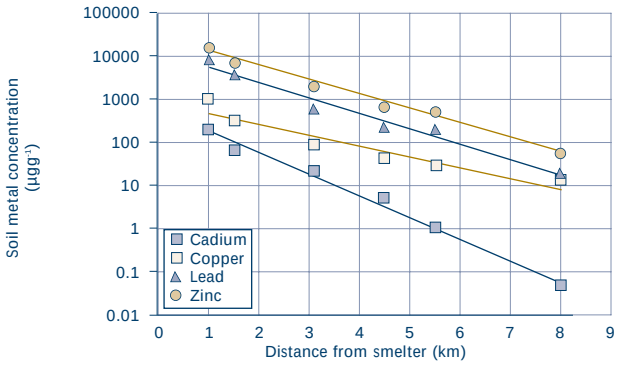


Figure 5.2 Relationship between concentrations of metals and distance from smelter

All patches are on grassland. Five patches (2 and 4, 7, 9 and 10) are on unmanaged grassland adjacent to minor roads, and two patches (1 and 8) are on unmanaged grassland adjacent to a public footpath. Patch 3 is a managed grassland area used in the past for grazing. Patch 5 is grassland overgrown by a sparse oak plantation, and patch 6 is in a grassland area within an abandoned apple orchard. Geologically, eight patches (2, 4, 5, 6, 7, 8, 9 and 10) are on low-lying land primarily of alluvium or mixed alluvium (head material) origin. Two patches (1, 3) are on elevated areas of Jurassic origin, and both are on or close to a slope.

5.4.3 *Conceptual Site Model Summary*

We identified a number of contaminant-pathway-receptor linkages for the site based on our initial investigation of the history of operations at Site A and the site walkover. The principal source of

Table 5.2 Summary of the contaminant-pathway-receptor linkages for Site A

Source	Contaminant Pathway	Contaminant Source	Receptors
Principal Primary metal (zinc/lead/ cadmium) smelter **Secondary** Hydrocarbon black plant Fertilizer plant Chemical plant	Aerial deposition	Metal contaminated soils	**Site of Special Scientific Interest** Primary production Soil ecosystem function Plant diversity Invertebrate diversity Population size of feature bird species Food chain: plant/ herbivorous invertebrate/ vertebrate Food chain: Invertebrate Bird/mammal/ bird of prey

Table 5.1 Walkover information gathered for ERA of the hypothetical receptor and soil collection for trialing biological tests.

Site-Sample	OS GR	Brief description	Underlying geology	Distance from smelter (km)	Altitude (m above sea level)	Other local contamination sources and disturbances	Proposed analytes for analysis	Evidence of ecological effect in walkover
A1	ST 593827	Grass verge of footpath on (west) crest of north-west facing slope bordered by oak woodland and managed pasture.	Jurassic/ Keuper marl	8.1	80	Approximately 100m from M5 motorway.	A metal suite including arsenic and mercury. PAH compounds.	None observed
A2	ST 568825	Grass verge (20m wide) of minor road bordered by road and field hedgerow. Heavily rutted in places.	Head	5.8	7	Cattle disturbance. Possible agrochemical input and motor vehicle emissions.	A metal suite including arsenic and mercury. PAH compounds.	Site shows possible evidence of disturbance.
A3	ST554802	Steep south-west facing slope of managed pasture with adjacent oak woodland.	Jurassic	3.3	60	Horse disturbance. Possible agrochemical input. 200m from M5 motorway.	A metal suite including arsenic and mercury. PAH compounds.	Initial survey noticed some typical soil and plant species absent.
A4	ST 537799	Densely grassed verge of minor road (20m wide) with occasional hawthorn. Bordered by pasture, road and a ditch.	Alluvium	1.7	6	Close to site of disused refuse incinerator. Subject to fly tipping.	Metals including arsenic and mercury, PAH compounds, possibly dioxins.	Litter accumulation noted. Some typical species absent. Evidence of pathological changes in some species.
A5	ST 534790	Sparsely planted oak stand with dense grass cover as field layer, surrounded by abandoned pasture. Approximately 20 m from a well used local road.	Alluvium	0.7	7	Sulphuric acid input from nearby plant. Motor vehicle emissions from road.	Metals, sulphates, soil pH, PAH compounds.	No detritiverous soil fauna (snails, worm, woodlice) found. Limited plant diversity. Extensive litter build-up.
A6	ST 536800	Abandoned small apple orchard (approximately 10 well spaced trees), now heavily grassed with a deep litter.	Alluvium	1.5	10	Potential past input of persistent organic and metal (copper, arsenic) based pesticides.	A metal suite including arsenic and mercury. Particularly note copper levels.	Extensive litter build up. Evidence of limited soil and plant fauna.

Table 5.1 (cont.) Walkover information gathered for ERA of the hypothetical receptor and soil collection for trialing biological tests.

Site-Sample	OS GR	Brief description	Underlying geology	Distance from smelter (km)	Altitude (m above sea level)	Other local contamination sources and disturbances	Proposed analytes for analysis	Evidence of ecological effect in walkover
A7	ST 528797	Grassed roadside verge area with sparse vegetation. Within 5m from local road.	Alluvium	0.7	5	Sulphuric acid from nearby plant. Motor vehicle emissions. Located next to smelter slag deposition area.	Metals, sulphates, soil pH, PAH compounds.	No detritiverous soil fauna found. Limited plant diversity and poor tree growth. Extensive litter build-up.
A8	ST 540817	Grassed area adjacent to public footpath. Dominated by dense grass and herb layer, with nearby hawthorn.	Alluvium	3.3	9	Located within 200m of a carbon black manufacture site. Suggests possible hydrocarbon contamination.	Metals, full suite of PAHs, other hydrocarbons.	Some expected plant and invertebrate species not found. Possible litter accumulation.
A9	ST 552853	Managed grass area situated approximately 30 metres from major road. Samples taken immediately adjacent to oak/hawthorn hedgerow.	Alluvium	5.8	12	Located close to busy road junction and possible pesticide input for grass management.	A metal suite including arsenic and mercury, PAH compounds.	Low plant biodiversity. No evidence of litter accumulation.
A10	ST 561880	Grass area close to (5m) minor road that runs along the tidal boundary (within 10m) of the Severn Estuary.	Alluvium	10.0	3	No obvious sources.	A metal suite including arsenic and mercury.	None observed

potential exposure to the ecoreceptor (the hypothetical SSSI) is from the smelter itself, as well as other minor sources (which vary depending on where the SSSI is placed). In all cases, the principal pathway from the source to the ecological receptor is through aerial deposition and possible solid waste transport (for example, as wind-blow particulates) (See Table 5.2). The receptor is the SSSI itself.

We prepared a conceptual site model (CSM) based on the initial identification of the principal contaminants and pathways and the characteristic of the ecological receptor (Figure 5.3). The CSM used was based on the simplest (and most demonstrable) set of linkages existing for the site. The primary source was identified as particulate metal material derived principally from the smelter stack and possibly from wind-blown solid waste (dust) kept on-site. The principal pathway is aerial deposition onto the surrounding countryside. The principal contaminant source is the contaminated soil. The pathway to the patches was by way of dissolution of the metals from particulate matter into soil porewater. Once present in soil solution, metal contaminants are available to plants, soil and soil-surface dwelling invertebrates and the various component groups that comprise the soil microbial community. These plants and animals are the receptors. They are potentially at risk of direct toxicity, with these contaminants causing either acute or chronic effects.

In addition to the primary receptors, species that could experience secondary exposure to metal contaminants can also be identified. These are linked to contaminants through the terrestrial food chain. They include:

- Seed-eating birds and small mammals, such as voles and many invertebrates (including phytophagus species and also species feeding on leaf-litter after leaf-fall) that feed on plants in which there is significant accumulation of metals.
- Small mammals, birds, amphibians and reptiles feeding on groups of soil invertebrates such as earthworms (Hopkin, 1993), woodlice (Hopkin, 1993; Ashton, 1998), molluscs (Hopkin, 1993) and Collembola (Smit and Van Gestel, 1996) that are all known to accumulate substantial amounts of metals from contaminated soils.
- Raptors feeding on small mammals, birds, amphibians and reptiles that have accumulated metals in body tissues as a direct result of exposure to elevated metal concentration in their diet.

With this information, coupled with decisions on the assessment endpoints of interest, we have enough information to enter Tier 1 in order to investigate the potential pollutant linkages described in Figure 5.3 and to identify potential biological tests at Tier 2.

As described in Section 2, the ERA process uses surrogate measures to identify and reach an ecological protection goal. These measurements are called assessment and measurement endpoints (see Box 1). In table 5.3, the desired assessment endpoint is ecosystem function and structure. Measurement endpoints provide quantifiable endpoints that relate directly to the assessment endpoints (suggestions are detailed in Table 5.3 below).

Using the contaminant-pathway-receptor conceptual model, we showed that an ecoreceptor (SSSI) was at potential risk from the aerial deposition of heavy metals. The management goal of this SSSI is to protect features and sub-features of interest, and to maintain or achieve favorable conditions within the designated ecoreceptor with no adverse effect on its integrity. The

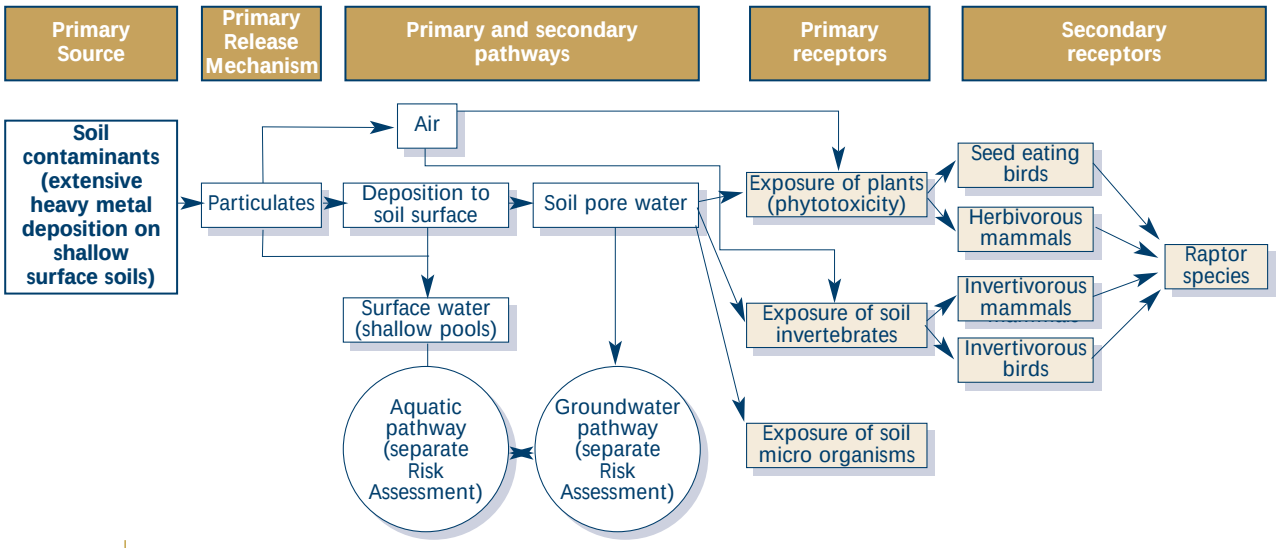


Figure 5.3 The initial conceptual site model for Site A showing main contaminants, pathways and receptors. (Boxes for primary and secondary receptors have light brown backgrounds)

Table 5.3 Summary of receptors, their relevance to the hypothetical ecoreceptor and assessment and measurement endpoints

Receptor	Relevance to the hypothetical ecoreceptor	Assessment endpoint	Measurement endpoints
Plant community	General biodiversity and feature plants species such as whitebeam, green hellebore and birds nest orchid. Food supply to animal species.	Ecosystem function in order to sustain plant germination and growth.	Plant toxicity test with monocotyledonous species. Plant toxicity test with dicotyledonous species.
Soil invertebrate community	Food supply to animal species.	Ecosystem function in order to sustain diverse invertebrate populations.	Earthworm acute toxicity test. Chronic springtail toxicity test.
Soil microbial community	Nutrient supply to support plant growth	Ecosystem function in order to sustain microbial populations.	Soil nitrification rate. Soil carbon mineralisation rate.
Seed eating birds	Variety of woodland passerine and other species.	Ecosystem diversity. In order to sustain suitable seed plants.	Chemical measurement of metal residues in plant material collected at termination of the plant toxicity test.
Vegetarian small mammals	Small mammals such as voles.	Ecosystem diversity in order to sustain suitable food plants.	Chemical measurement of metal residues in plant material collected at termination of the plant toxicity test.
Insectivorous mammals	Small mammals such as shrew.	Ecosystem diversity in order to sustain suitable insect prey.	Analysis of earthworms collected on termination of the earthworm acute toxicity test.
Insectivorous birds	Protected bird species such as thrush, woodcock and snipe.	Ecosystem diversity in order to sustain suitable insect prey.	Analysis of earthworms collected on termination of the earthworm acute toxicity test.
Raptor species	Raptors such as buzzard, sparrow hawk, owls and red kite.	Ecosystem diversity in order to sustain prey.	Food chain modelling.

protection of soil ecosystem function is a priority, because it maintains the integrity of a habitat where rare birds of prey live. The conservation objective of the SSSI is the continued maintenance or enhancement of population abundance and assemblage structure of features of interest, and of structure or function of supporting sub-features. Appropriate measurement endpoints need to be identified at Tier 2 in order to validate the protection goal.

5.5 Tier 1 assessment

5.5.1 Introduction

We:

1) Visited sites to collect soils for chemical analysis and to use in the later soil bioassays conducted as part of a Tier 2 assessment;

2) Carried out a suite of chemical analyses on sub-samples of the collected soils;

3) Compared measured soil chemical concentrations to SQGVs;

4) Undertook bioassays with the Microtox™ biosensor system to support the chemical analyses;

5) Decided whether any of the pathways within the conceptual site model were relevant and whether the site should proceed further through the risk assessment process to Tier 2. We based this decision on the outcomes of the chemical analyses, the comparison with SQGVs, and on Microtox™ biosensor studies.

Sample collection

Sampling was identical to that carried out for the evaluations carried out as part of the parallel Project P5-063. That is, the samples were split and used in both projects.

We outline the collection of soil samples and subsequent chemical analysis below. Details of sampling design and collection and storage of soil samples are covered in the report from the parallel Project P5-063 (Spurgeon *et al.,* in press). It is worth emphasising that sampling positions were determined in a highly targeted fashion, to create the two linear transects described in Section 5.4.2, based on previous experience of the site and the understanding gained of contaminant residues.

We evaluated contaminant concentrations at each patch analysing a sample of soil collected from the larger bulk soil sample that was later used for the laboratory toxicity test trials. Analysis was conducted on two pseudoreplicated samples (separate sub-samples from a single batch). It was possible, on the basis of the Tier 0 screening, to identify a series of priority analytes. For the purposes of the project, though, we based our choice of compounds selected for chemical analysis on the ICRCL (ICRCL, 1997) suite of determinands (which at the initiation of this project were still valid), as decided by the project board and study team.

Sampling details

At each patch, we marked out areas and excavated soil using a spade from the four corners of a marked central square. In each corner, 0.5m^2 of soil was excavated to a maximum depth of 25 centimetres (depth of sampling will be site-specific - though deep sub-surface soil is of little value for biological testing). This provided four samples of 20 l volume (we removed large stones etc. by hand). If turf was present at the sampling point, this was removed (we collected the soils from the root-mat) and the soil was excavated below the root-mat. We then mixed collected samples on-site to ensure homogeneity. Finally, we bagged soils and individually marked them with unique sample point codes, ready to be taken to the laboratory in refrigerated containers.

Upon our return, sub-samples for chemical analysis were immediately stored at -20°C. This kills the indigenous soil macroinvertebrate fauna. Because site A is contaminated principally by metals that are unlikely to be volatilised, soils were air-dried, and aggregates broken-up, before use. After drying, large soil aggregates were further broken up and the whole material passed through a 10 millimetre and subsequently a two millimetre mesh. Soils to be tested with the earthworms and Collembola were then crushed in a mechanical soils crusher (the equipment was carefully cleaned after each patch soil), and sieved through a two millimetre mesh. At this point, we took soil samples for analysis of pH, percentage loss on ignition (%LOI), and maximum water-holding capacity and field capacity. Finally, soils that were for use in the bioassays were placed in the correct volume into the experimental containers and re-wetted to the relevant percentage of their moisture retention capacity (Spurgeon & Hopkin, 1995), as determined using an established method (Kalra & Maynard, 1991). It should be noted that minimal treatment of soils is advocated in order to retain as many of the original in-situ characteristics as possible. This does, though, need to be balanced against the requirements of the biological test procedures.

Chemical analysis

We analysed replicated soils samples for a range of metals and inorganic determinands following extraction using aqua regia (3:1 hydrochloric to Nitric Acids). Boron analysis, though, was carried out on a water-soluble extract. This involved boiling the soil in water for 10 minutes and filtering the sample for analysis. The choice of extraction procedure may have led to an underestimate of the concentrations of trace metals at the site. The use of more ecologically relevant measurements, such as extractions in dilute salts, dilute acids or complexing agents, may have provided a more reliable indication of the potential risk arising from trace metals present in the soil. We analysed most metals using ICPMS, apart from arsenic which we measured using FI-HGAAS (Flow Injected Hydride Generation Atomic Absorption Spectroscopy) and mercury using CV-AAS (Cold Vapour Atomic Absorption Spectroscopy). Total petroleum hydrocarbons were extracted from soil samples using soxhlet extraction with freon. We analysed these using infrared techniques.

5.5.2 Comparison of soil contaminant concentrations with Soil Quality Guideline Values

Background to SQGVs

Canadian CCME

The CCME (Canadian Council of Ministers of the Environment) soil quality guidelines are derived specifically for the protection of ecological receptors in the environment and for the protection of human health associated with four land uses: agricultural, residential/parkland, commercial and industrial. They approximate a 'no or low' effect level for the protection of human health and ecological receptors.

The environmental soil quality guidelines are derived using toxicological data to determine the threshold level of effects for key ecological receptors. The primary exposure route used in the derivation procedure for environmental quality guidelines is exposure to soil. The agricultural land guideline, though, is based on soil exposure and food ingestion. The lower of the two values is considered to be the environmental soil quality guideline for agriculture.

Human health guidelines are derived using a process similar to site-specific risk assessment. Assumptions are made about the sensitive receptor and the chemical exposure for each land use to establish the soil quality guidelines.

The lowest of either the ecological or human health value is taken as the recommended soil quality guideline value for each of the four land uses.

Netherlands soil quality guidelines (VROM)

The Dutch ministry of Housing, Spatial Planning and the Environment (VROM) has set a number of intervention values and target values for the assessment of soil and groundwater contamination (VROM, 2000).

The soil remediation intervention values indicate when the functional properties of soil for humans, and for plant and animal life, are seriously impaired or threatened. The values represent the level of contamination above which there is serious soil contamination.

The intervention values are based on extensive studies of both human and ecotoxicological effects of soil contaminants. Human toxicological effects have been quantified in the form of concentrations in the soil above which the so-called maximum permissible risk (MPR) for humans may be exceeded. For non-carcinogenic substances, this corresponds to the Tolerable Daily Intake (TDI). For carcinogenic substances, it is based on an additional chance of tumour incidence of 10^{-4} for lifetime exposure. Ecotoxicological effects are quantified by concentrations in soil above which 50 per cent of species and processes may experience negative effects. The ultimate intervention values for soil are based on a harmonisation of the human and ecotoxicological effects. In principle, the most critical effects are definitive.

Soil target values indicate the level at which there is a sustainable soil quality. In terms of contaminated land, the target values thus indicate the level that has to be achieved to recover fully the functional properties of the soil for humans, plants and animal life. The target values also give an indication of the benchmark for environmental quality in the long term, based on the assumption of negligible risks to the ecosystem. Again, the final target values for soil and sediment are based on an integration of the human and ecotoxicological effects, with the lowest value being adopted.

US DoE Standards

The United States Department of Energy (DoE) set a number of screening level benchmarks for earthworms, plants and soil microbes. The screening levels are based on laboratory toxicity data for these organisms and are based on the 10^{th} percentile of a distribution of toxicity data. That is, they aim to protect 90 per cent of the earthworm, plant or microbe populations. The benchmarks do not have a protection goal as such, but are proposed for general contaminant screening purposes. If a chemical concentration exceeds the screening benchmark, then the contaminant is highlighted as of 'potential concern' and will require further analysis. Concentrations that fall below the screening level may be ignored, unless public concern or ancillary evidence suggests a chemical should be investigated further.

Harmonised standards such as the CCME soil guideline values and the Dutch target and intervention values, where the lowest of either the ecological or human health value is taken as the recommended guideline, should be used with caution when used in ecological risk assessment. Final guidelines based on human health will not be directly relevant to an ecological endpoint. They will, by default, be more stringent than the corresponding ecological guideline. Caution should therefore be taken when comparing soil concentrations with harmonised soil quality guidelines for ERA. It is advisable to consult the original documentation for each of the guidelines to obtain the basis of the specific SQGVs and hence their relevance to ERA.

In relation to this project, we consulted the original documentation supporting the guidelines. Where a SQGV has been based on human health values, this has been highlighted in the soil comparison tables. We should point out that all of the Dutch guidelines for the chemicals in question were based on ecological data. A number of the CCME guidelines have, though, been based on human heath values. But all were in the same order of magnitude as the ecological guideline, and so are not overly stringent. The arsenic guideline, for example, is set at 12 mg/kg for all land uses, based on human health effects, whereas the corresponding ecological guidelines were 17 and 26 for the agricultural and residential and the commercial and industrial land uses respectively.

Soil Comparison

For each measured contaminant, we compared the higher of the two measured soil concentrations with three international SQGVs from the United States (Department of Energy DoE), the Netherlands (Dutch Target and Intervention values) and Canada (CCME) (Tables 5.4 - 5.6, respectively). We used maximum measured chemical concentrations, so that we could illustrate our decisions based on individual patches. In normal practice, though, the upper 95th percent C.L. of the mean value should be taken as the PEC for comparison with SQGVs. Values for the SQGVs used are given at the top of each table and in Table C1 in Appendix C.

The data confirm widespread metal contamination at Site A. Arsenic, cadmium and zinc are the major contaminants, with 60-100 per cent of soil samples exceeding all international SQGVs. Soils are also heavily contaminated with lead and mercury, but to a lesser degree than with arsenic, cadmium and zinc, as shown by fewer exceedances of SQGVs.

Minor contaminants at the site were chromium, copper and nickel, with on average, four patches exceeding SQGVs (A3, A4, A5 and A7) for these metals. Concentrations of all other determinands were generally below guideline concentrations.

Based simply on the number of SQGVs exceeded, patches A4, A5 and A7 appear to be the most contaminated, with approximately seven contaminants exceeding international soil guidelines. The magnitude of the exceedances is also substantial, for example the highest guideline for zinc (Dutch Intervention level of 720 mg/kg) is exceeded by 28 and 32 times for patches A4 and A7, respectively. Levels of lead and cadmium are also very high, with concentrations up to 16 times higher than the highest guideline value at the most contaminated patches (A4, A5 and A7). There is a tendency for patches A3-A7 to be the most contaminated and patches A1, A2 and A8-A10 to be the least contaminated. This reflects the geography of the patches on the site. Sites A5-A6 are closest to the smelter, and Sites A1, A9 and A10 are the furthest away. Though patch A6 is close to the smelter, measured concentrations of metals were lower than those found at patch A7, as a consequence of the aerial deposition patterns of the smelter emissions (Coglan *et al.,* 2003).

5.5.3 Toxicity screening

We prepared aqueous leachates from a control soil (Kettering loam) and the 10 soil samples from Site A (A1-10). We then exposed populations of the bioluminescent marine bacterium *Vibrio fischeri* to them for a period ranging from five to 30 minutes. The results for these are shown in Figure 5.4.

In the control soil (Kettering Loam), the 0.01M $CaCl_2$ control and Site A soil samples A10, A1, A9, A2 and A8, the results were comparable to the control (with no significant difference in light inhibition, and an IC_{50} value could therefore not be determined for any of these soil samples. In leachates of soils from A5, A7, A4, A6 and A3, though, we observed inhibition of light output (relative to the control). These responses were sufficient to allow an IC_{50} value to be calculated. These exhibited a narrow range of IC_{50} values, between 0.78 per cent for soil sample A5 (most toxic) up to 13.1 per cent for soil sample A3 (Figure 5.4).

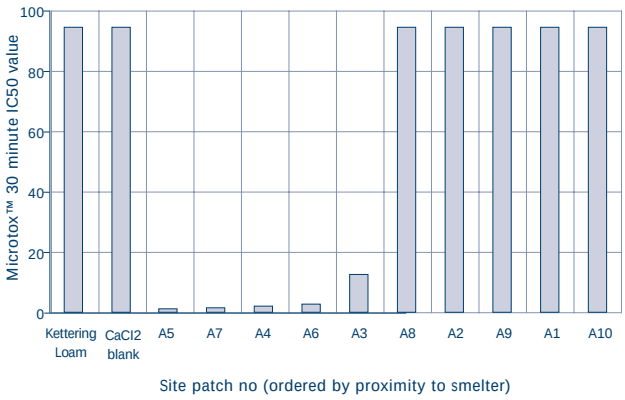


Figure 5.4 Site A Microtox™ test results for aqueous extractants from the 10 site patches compared with a reference soil and a procedural blank.

Table 5.4 Comparison of Site A soils with US DoE SQGVs

SQGV	T-CN	F-CN	S	As	B	Cd	T-Cr	Cu	Hg	Ni	Pb	Se	SO4	TDYS %	Zn
Earthworm	N/G	N/G	N/G	60	N/S	20	0.4	50	0.1	200	500	70	N/G	N/G	200
Microbial	N/G	N/G	N/G	100	3000	20	10	100	30	90	900	100	N/G	N/G	100
Plant	N/G	N/G	N/G	10	500	4	1	100	0.3	30	50	1	N/G	N/G	50
Site A	**T-CN**	**F-CN**	**S**	**As**	**B**	**Cd**	**T-Cr**	**Cu**	**Hg**	**Ni**	**Pb**	**Se**	**SO4**	**TDYS %**	**Zn**
A1	<2.5	<1	0.4	43.2	1.29	7.39	34.2	40.1	0.2	40.7	246	<0.9	49.2	65.3	841
A2	<2.5	<1	0.5	7.81	0.92	4.2	28.1	30.3	0.2	23.4	167	<0.9	53.8	77.3	423
A3	<2.5	<1	0.5	46.3	0.84	64.3	43.2	156	1.34	43.7	212	3.8	62.9	68.4	4123
A4	<2.5	1.25	5.7	237	0.89	187	76.3	1250	7.32	56.5	15765	27.5	267	71.9	20500
A5	<2.5	<1	0.4	224	1.01	254	35.6	1510	8.23	42.3	14543	37.5	203	56.7	8050
A6	<2.5	<1	0.5	34	0.76	28.7	35.3	87.4	0.58	32.4	857	<0.9	43.8	60.9	2012
A7	<2.5	<1	5.9	256	0.53	234	70.3	1400	8.04	67.8	15000	35.5	247	61.6	23400
A8	<2.5	<1	2.1	11.2	0.72	2.03	33.6	16.6	0.05	27.6	93.2	<0.9	51.6	67.6	257
A9	<2.5	<1	<0.1	14.5	1	1.56	45.3	23.4	0.1	34.2	83.4	<0.9	49.3	71.9	256
A10	<2.5	<1	0.4	17.3	1.1	1.67	25.6	25.7	0.08	20.2	234	<0.9	51.2	65.4	841

N/G = No Guideline available

Exceedance of Earthworm benchmark (mg/kg)

Exceedance of Plant benchmark (mg/kg)

Concentration higher than two or more Guidelines

Table 5.5 Comparison of Site A soils with Dutch Target and Intervention Levels

SQGV	T-CN	F-CN	S	As	B	Cd	T-Cr	Cu	Hg	Ni	Pb	Se	SO4	TDYS %	Zn
Target	5	1	N/G	29	N/G	0.8	100	36	0.3	35	85	N/G	N/G	N/G	140
Intervention	650/50†	20	N/G	55	N/G	12	380	190	10	210	530	N/G	N/G	N/G	720
Site A	**T-CN**	**F-CN**	**S**	**As**	**B**	**Cd**	**T-Cr**	**Cu**	**Hg**	**Ni**	**Pb**	**Se**	**SO4**	**TDYS %**	**Zn**
A1	<2.5	<1	0.4	43.2	1.29	7.39	34.2	40.1	0.2	40.7	246	<0.9	49.2	65.3	841
A2	<2.5	<1	0.5	7.81	0.92	4.2	28.1	30.3	0.2	23.4	167	<0.9	53.8	77.3	423
A3	<2.5	<1	0.5	46.3	0.84	64.3	43.2	156	1.34	43.7	212	3.8	62.9	68.4	4123
A4	<2.5	1.25	5.7	237	0.89	187	76.3	1250	7.32	56.5	15765	27.5	267	71.9	20500
A5	<2.5	<1	0.4	224	1.01	254	35.6	1510	8.23	42.3	14543	37.5	203	56.7	8050
A6	<2.5	<1	0.5	34	0.76	28.7	35.3	87.4	0.58	32.4	857	<0.9	43.8	60.9	2012
A7	<2.5	<1	5.9	256	0.53	234	70.3	1400	8.04	67.8	15000	35.5	247	61.6	23400
A8	<2.5	<1	2.1	11.2	0.72	2.03	33.6	16.6	0.05	27.6	93.2	<0.9	51.6	67.6	257
A9	<2.5	<1	<0.1	14.5	1	1.56	45.3	23.4	0.1	34.2	83.4	<0.9	49.3	71.9	256
A10	<2.5	<1	0.4	17.3	1.1	1.67	25.6	25.7	0.08	20.2	234	<0.9	51.2	65.4	841

N/G = No Guideline available
† = Values relate to a pH<5 and pH≥5 respectively

Exceedance of Target Level (mg/kg dry weight)

Exceedance of Intervention Level (mg/kg dry weight)

Table 5.6 | Comparison of Site A soils with CCME SQGVs

SQGV	T-CN	F-CN	S	As	B	Cd	T-Cr	Cu	Hg	Ni	Pb	Se	SO4	TDYS %	Zn
Agricultural	N/G	N/G	500	12*	750	1.4*	64	N/G	6.6*	50	70	N/G	N/G	N/G	200
Res./Park	N/G	N/G	N/G	12*	500	10	64	N/G	6.6*	50	140*	N/G	N/G	N/G	200
Comm.	N/G	N/G	N/G	12*	2000	22	87	N/G	24*	50	260*	N/G	N/G	N/G	360
Industrial	5	5	N/G	12*	2000	22	87	N/G	50	50	600	N/G	N/G	N/G	360
Site A	**T-CN**	**F-CN**	**S**	**As**	**B**	**Cd**	**T-Cr**	**Cu**	**Hg**	**Ni**	**Pb**	**Se**	**SO4**	**TDYS %**	**Zn**
A1	<2.5	<1	0.4	43.2	1.29	7.39	34.2	40.1	0.2	40.7	246	<0.9	49.2	65.3	841
A2	<2.5	<1	0.5	7.81	0.92	4.2	28.1	30.3	0.2	23.4	167	<0.9	53.8	77.3	423
A3	<2.5	<1	0.5	46.3	0.84	64.3	43.2	156	1.34	43.7	212	3.8	62.9	68.4	4123
A4	<2.5	1.25	5.7	237	0.89	187	76.3	1250	7.32	56.5	15765	27.5	267	71.9	20500
A5	<2.5	<1	0.4	224	1.01	254	35.6	1510	8.23	42.3	14543	37.5	203	56.7	8050
A6	<2.5	<1	0.5	34	0.76	28.7	35.3	87.4	0.58	32.4	857	<0.9	43.8	60.9	2012
A7	<2.5	<1	5.9	256	0.53	234	70.3	1400	8.04	67.8	15000	35.5	247	61.6	23400
A8	<2.5	<1	2.1	11.2	0.72	2.03	33.6	16.6	0.05	27.6	93.2	<0.9	51.6	67.6	257
A9	<2.5	<1	<0.1	14.5	1	1.56	45.3	23.4	0.1	34.2	83.4	<0.9	49.3	71.9	256
A10	<2.5	<1	0.4	17.3	1.1	1.67	25.6	25.7	0.08	20.2	234	<0.9	51.2	65.4	841

N/G = No Guideline available
* = Guideline based on human health effects

Exceedance of Agricultural Guideline(mg/kg)

Value higher than two or more Guidelines

The estimation of an IC_{50} merely allows us to rank samples in terms of toxicity. Comparison with the control is a more meaningful comparison, because it enables us to make decisions about the acceptability of a soil. Though not subject to statistical analysis, it is clear that samples from patches A3-A7 (all within 3.3 kilometres of the known point source) are appreciably more inhibitory than both samples from the other sites further away from the point source, and samples from the controls. Indeed, toxicity in samples from patches A1, A2 A8-10 is indistinguishable from that in the reference site or test controls. This trend in toxicity correlates well with the measured contaminant profile. The most toxic soil samples tested have heavy metal concentrations above threshold values.

The tests did not produce any unexpected positive results (that is, samples showing toxicity but with low levels of contamination), but exceedance of SQGVs did not always correspond to toxicity in the Microtox™ test. This may be attributable to contaminants present in chemical forms that are non-toxic, or not being bioavailable (or both). To make such a judgement on the basis of Microtox™ alone would be unwise. But if confirmed in other biological tests, it highlights the possibility of false positives when judgements are made on the basis of exceedances of SQGVs alone.

5.5.4 Ecological survey data

Spurgeon *et al.,* (in press) describe available ecological survey data for Site A in some detail. The key points are that abundances of detritivore groups such as earthworms, molluscs, millipedes, woodlice and springtails are impacted, particularly at patches A3, A4, and A5. Spurgeon *et al.,* (in press) conclude that the presence of high to very high metal contamination at patches A3, A4 and A5 is sufficient to cause significant harm to these ecosystems.

5.6 Decision-making

Site A would clearly progress to Tier 2 testing, because at least one contaminant in soil samples from all the sampling locations exceeded its respective SQGV. This would apply also if the upper 95th percent C.L. of mean measured concentrations were taken as the PEC for Site A as a whole. Furthermore, the nature of the contaminant, a persistent and accumulative heavy metal, will drive the assessment to tier 2 for further evaluation.

Where major exceedances are found, especially if they are coupled with evidence of gross ecological impact, there is little additional value to be gained from Microtox™ testing at Tier 1. Patches A3, A4 and A5 are three such locations. Indeed, if these were to be considered as separate sites, it would be possible to exit the framework at this point, omitting Tier 2 entirely and considering risk management options without further assessment. The extent and type of any remediation could, though, be influenced by further evaluation at Tier 2. These patches may therefore be progressed into Tier 2, in order to increase the information available on which to make risk management decisions.

Modifications to the CSM in light of the Tier 1 assessment

The fact that at least one Dutch list SQGV was exceeded in each patch suggests that the potential exists in all patches for direct toxicity. This is because the Dutch list values are based on soil invertebrate, plant and microbial toxicity. They are thus designed to protect these receptors from the direct toxic effects of contaminants.

Assessing the significance of possible secondary effect pathways is more difficult, though the Dutch List standards employ a simple food chain model in deriving SQGVs. Food chain effects are addressed in the CCME values for agricultural land use only. The ultimate driver is the protection of human health from contaminated food crops and dairy products. The fact that there were exceedances of cadmium SQGVs at all the patches, though, and because cadmium is one of the few metals (with mercury and methyl-mercury) that accumulates in terrestrial food-chains, it is reasonable to suppose there is potential for secondary poisoning effects from this metal.

The Tier 1 assessment for site A suggests that all pathways within the CSM remain valid and should be investigated further as part of the Tier 2 assessment. Receptors at the periphery of the contamination gradient, that is, those most remote from the point source, still remain within those patches to be taken forward into the Tier 2 assessment. Furthermore, for some patches, the evidence of ecological impact is already so compelling that further study would merely refine the decision-makers' understanding and possibly assist in identifying the causal feature.

5.7 Tier 2 assessment

5.7.1 Introduction

No further chemical analysis was performed on soil samples from Site A. Biological testing was, though, performed on soil samples from all patches, irrespective of the level of contamination detected.

On the following pages, the results of biological tests are presented as a summary figure, with biological response (expressed as a percentage of the site reference control) plotted against the distance of the patch from the smelter point source.

The CSM, summarised in Figure 5.3 helps inform the selection of biological tests to be used in Tier 2 as measurement endpoints. For Site A, the following are relevant:

- Plant tests can act as surrogates for assessing phytotoxin hazards to native plants;
- Earthworm acute toxicity test allows us to evaluate effects on soil fauna that can alter biodiversity and food supply;
- Nitrification and carbon mineralisation tests may be used as surrogates to assess possible effects on nutrient supply to the ecosystem;
- For the protection of the food chain, chemical residues can be quantified in plant and earthworm material (collected on termination of each of these two tests). This information may be used for modelling the potential movement of each contaminant through food chains.

5.8 Results of biological tests

The following sections summarise the key findings of the biological testing undertaken for Site A. The key findings of these tests are compiled and used in a weight-of-evidence table to reach a decision about further action.

5.8.1 Carbon and nitrogen mineralisation

We prepared duplicate samples for each patch within each site. We used two control soils for comparative and quality control purposes: the Kettering loam as used in all the previous studies, and a local pasture soil collected from the vicinity of the laboratory at Medmenham. A further duplicated set of controls was spiked with sodium azide (2% w/w), to inhibit microbial processes, with a further non-spiked replicate control to produce positive controls for comparison. Ground glucose (1.2g) was added to the carbon mineralisation test soils accordingly. It was mixed homogeneously by hand, to induce the respiration rate and elicit a maximum respiratory response. OxiTop®-C sensors were used to monitor changes in pressure and hence rates of mineralisation.

Carbon mineralisation

The experimental test data to determine soil carbon mineralisation for Site A were inconclusive, with wide variation. They failed to demonstrate any clear trends in carbon mineralisation activities. The most likely reason for this huge variability is that heterotrophic soil microorganisms acquire carbon for maintenance and growth by decomposing plant residues and other organic materials added to soils. The presence of certain contaminants in soils may interfere with these mineralisation processes, and carbon mineralisation rates may go up or down accordingly. Soils, though, tend to be heterogeneous, varying spatially in a number of chemical, physical, and biological properties. Therefore, measurements of carbon mineralisation rates determined in isolation of other data tend to show large degrees of spatial heterogeneity, mimicking this variability. Variability in soil mineralisation rates is therefore a combination of the interplay of the physical, chemical, and biological components of the soil matrix at the microscale. The results of this work highlight the fact that an unacceptable level of uncertainty pervades the test methodology.

Nitrogen mineralisation

The soils were tested according to the OECD guideline described in Appendix B. Powdered lucerne was added to the test vessels (this has a high carbonto-nitrogen ratio) to prevent carbon starvation of the microbes. Sufficient sample containers were provided to allow the study to run for 100 days if necessary. Duplicate test samples and four controls were submitted for nitrate analysis every 14 days.

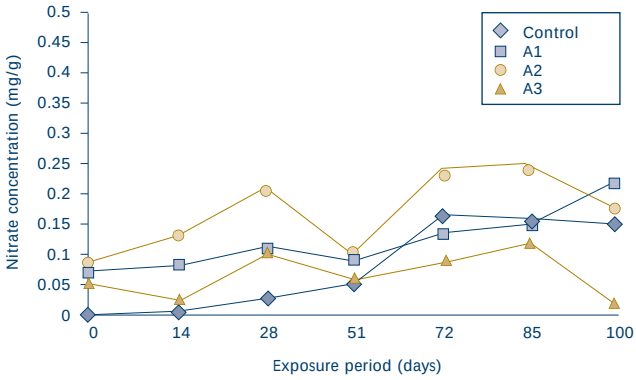


Figure 5.5 Net nitrate production for soils sampled from Site A

Figure 5.5 shows the concentration of nitrate with time generated during the nitrogen mineralisation test for site A soils. Net nitrate production is a complicated biological process. It is the product of production via mineralisation and nitrification and consumption by denitrification. Figure 5.5 shows no clear relationship between proximity to the point source (that is, the smelter, and by inference, levels

of metal contamination) and nitrate production. To aid interpretation, values are shown in Figure 5.5 for control soil and soil from three patches only, these results were typical of those reported with soil from the other patches.

5.8.2 Plant tests

We assessed the effects of the Site A soils on terrestrial plants using OECD guideline 208A - seedling emergence and growth test (Appendix B). We made a number of modifications to the guideline, though, to make the procedure more relevant to contaminated land assessment. These modifications concerned changes to the protocol on seeding density and pot sizes. Details appear in Appendix B.

We carried out definitive tests on the effects of Site A soils on cabbage, pea, tomato (dicotyledons), oat and wheat (monocotyledons). We recorded the effects on emergence, wet weight and dry weight. Emergence was found not to be a sensitive endpoint with which to assess effects (Appendix B); plant wet weight and dry weight are more sensitive measures of effect. On balance, dry weight is the preferred endpoint because it reflects changes in biomass and can be estimated more accurately than wet weight. Below, we report only effects on the dry weight of plants.

Results

We present dry weights per emerged plant for each of the five species in Figures 5.6 a-e. Results have been plotted against patch distance from the smelter. Cabbage and tomato are the only species for which a discernible effect gradient can be seen, with lowest dry weight at the patches closest to the smelter (most contaminated) (Figures 5.6 a and d). For the other species, effect gradients are less clear. Reduced growth in the most contaminated soils (A5, A7, A6, A4 and A3) is, though, evident.

Statistical analysis of the data indicate that soil samples A5 and A7 had the greatest effect on plant biomass. All species showed significantly lower dry weights ($P < 0.05$) at these sites when compared with the other soil samples. Soils from patch A4 also had a significant effect on the biomass of three of the plant species (cabbage, oat and wheat). There was no significant difference ($P < 0.05$) in biomass between the three site controls (A1, A9 and A10) and the test control for 4 of the 5 species. For cabbage there was, though, a significant difference ($P > 0.05$) between site A1 and the Kettering loam control, but not for site A10 and the test control. These data suggest that, in general, the soil physico-chemical properties at Site A had little or no effect on plant dry weight. Therefore it is reasonable to conclude that effects on dry weight of plants were most likely due to the presence of chemical (heavy metal) contamination overriding any disparity in the mineral composition or nutritional status of the soils themselves.

In summary, plants appear to be sensitive to the contaminants at Site A. Though it is not possible to discern an effect gradient with most species, the most contaminated soils have a significant effect on plant biomass (dry weight). The data also suggest that soil quality at the site did not affect plant growth. Therefore, based on this data, the plant growth test does provide a broad measure of bioavailable contamination at the site. It is also possible to separate effects related to chemical contamination from effects of soil quality.

Positive controls

We also tested positive controls using a concentration range of zinc (0, 10, 32, 100, 320 and 1000 mg/kg). EC50s, NOECs and LOECs based on the exposures are presented in Table 5.7. The data indicate the higher sensitivity of tomato and cabbage compared with the other species tested. It is not possible, though, to calculate an EC50. The only way to compare sensitivities of the different species to the Site A soils is on the basis of differences in response rates (that is, the magnitude of reductions in dry weight). This is confounded by the fact that as the results are presented as dry weight, larger species (pea and cabbage) will always have higher dry weights than smaller species (tomato). Consequently, differences in sensitivity to contaminated soils cannot readily be distinguished in this way.

Table 5.7 Results of toxicity tests with positive controls (zinc, ugl^{-1})

Species	EC50	NOEC	LOEC
Pea	>1000	1000	>1000
Tomato	655	320	1000
Oat	>1000	1000	>1000
Wheat	>1000	1000	>1000
Cabbage	740	320	1000

Figure 5.6 Dry weights of plants germinated in soils from site A (average of 5 replicates) (Kettering loam represents a clean soil).

5.8.3 Earthworm survival and growth

Survival

Survival of the earthworm *Eisenia andrei* in the negative and site controls, and indeed most of the other site soils was, on average, above 90 per cent after 14 days, and so complied with the test validity criteria (Figure 5.7). Comparison of survival for the worms incubated in Site A soils indicated significant effects after both seven and 14 days (7 days, $F = 29.04$, $p > 0.001$; 14 days $F = 13.84$, $p > 0.001$). Post-hoc comparison (Tukey, $p < 0.05$) showed that the only soil causing significantly decreased survival of worms was soil from patch A5 (the patch located closest to the point source). There were no significant differences in the survival between the worms incubated in any of the other Site A soils or the negative control soil. Given the known levels of metals in these soils and the effects reported for other species of earthworms in soils from these locations, this suggests that earthworm survival over short exposure periods is a relatively insensitive parameter.

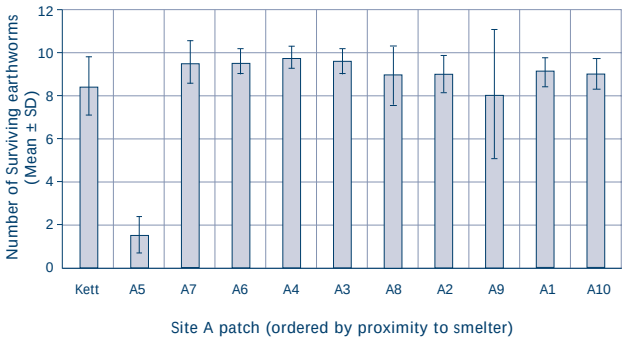


Figure 5.7 Survival (Mean ± SD, n = 4) of *Eisenia andrei* incubated for 14 days in soils collected at all Site A patches. (Kett represents a clean (Kettering loam) soil)

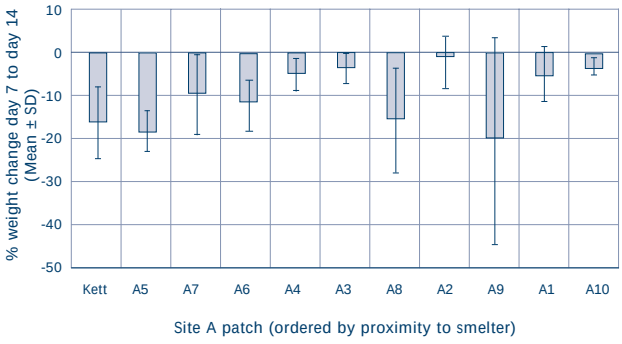


Figure 5.8 Earthworm weight change (Mean ± SD, n = 4) between day 7 and 14 in *Eisenia andrei* incubated for 14 days in soils collected at all Site A patches. (Kett represents a clean Kettering loam soil)

Growth

Individual weight changes of earthworms between day 7 and day 14 of the exposure indicated that animals in all treatments lost weight (Figure 5.8). This is unsurprising since, in accordance with the OECD (1984) guidelines, the worms were not supplied with a source of food throughout the test. For worms incubated in Site A patch soils, weight change showed no trend that could be related to patch proximity to the point source. The absence of a significant effect of soil on earthworm mean individual weight change was confirmed using ANOVA ($F = 1.69$, $p = 0.152$).

In some of the Site A patch soils, a high variance in weight change between replicates was apparent (as can be seen from SD-error bars on Figure 5.8). This variance can be attributed to a number of causes. First, the variability of survival in the assay may affect stocking density both in terms of numbers of individuals and biomass. The three Site A patches where the highest variability in weight change occurred, (A9, A8 and A7), are also the ones with the highest variability in survival (see Figure 5.7). Earthworm growth is known to be density-dependent (Kammenga *et al.*, 2003). Another potential reason may be heterogeneity between replicates, despite the extensive mixing of the soil prior to the start of the test. Again, as for survival, this indicates that earthworm weight change is a relatively insensitive endpoint when measured following the acute exposure protocol.

5.8.4 Springtails survival and juvenile production

Survival

Unfortunately, the springtail tests performed poorly, with high adult mortality and low reproduction found in many of the patch soils. Both adult survival and juvenile production were frequently below the validity criteria applied for studies conducted with the ISO (1999) test: adult mortality was greater than 20 per cent in all but two replicates, and less than 100 juveniles were produced in all but one replicate. This is most likely due to the standard test guidelines being developed for testing of compounds in standard soils, and hence not accommodating the use of field collected soils very well. Despite this, there was still a significant effect of Site A patches on survival (ANOVA ($F = 4.69$, $p < 0.001$) (Figure 5.9). Post-hoc Tukey ($p < 0.05$) analysis revealed that the only significant difference was that survival in patch A6 soils was higher than in all other soils, except those from patch A3.

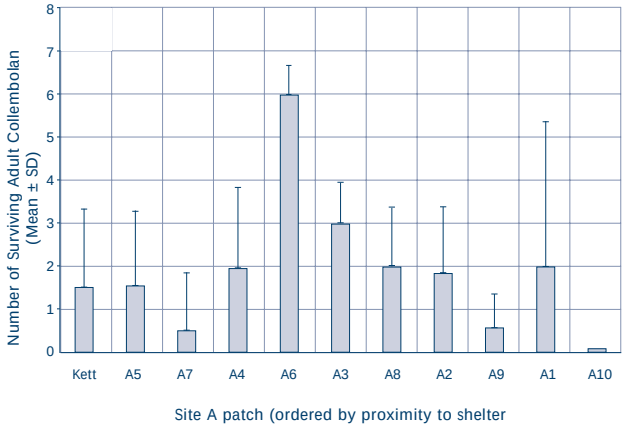


Figure 5.9 Survival (Mean ± SD, n = 5) of *Folsomia candida* incubated in soils collected at all Site A patches. (Kett represents clean Kettering loam soil)

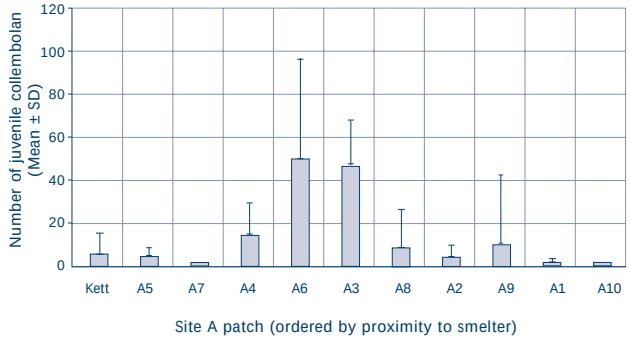


Figure 5.10 Number of juveniles (Mean ± SD, n = 5) produced from an initial population of 10 adult *Folsomia candida* incubated in soils collected at all Site A patches. (Kett represents a clean Kettering loam soil)

Juvenile production

Given that the production of juvenile Collembolans is measured in the same assay as adult survival, it suffers from the same poor test performance, as discussed above. Again, there are some significant effects (ANOVA ($F = 4.90$, $p < 0.001$) (Figure 5.10), with the post-hoc Tukey test ($p < 0.05$), showing that juvenile production in soils from Patch A3 and A6 was significantly higher than that observed in soils from patch A10, A1, A2, A7 and A5, where juveniles were totally or nearly absent.

Given the poor test performance, it is difficult to comment on the results beyond the point that the standard protocol will need some modification in terms of pre-test treatment of soils in line with those undertaken by Fountain and Hopkin (2001) (See Section 7.2.5).

5.9 Decision - making (Tier 2)

There are clearly instances where soil samples exhibited toxicity, and these occur at sites subject to the greatest metal contamination (notably patch A5). At the most contaminated sites, where marked impacts on community structure were also evident (A3, A4 and A5), biological testing confirmed adverse effects on survival of earthworms and phytotoxicity. Toxicity testing did not, though, reveal effects that could not have been concluded from chemical data and ecological survey data.

The situation at other patches, where chemical contamination and toxicity are lower, is more equivocal. At patches A1, A2 and A8 - A10, though, there is no discernible evidence of toxicity. So any exceedances of SQGVs here appear to be of little biological significance. These patches might therefore exit the framework at this point. According to Section 4, we need to apply the weight-of-evidence principles to judge the significance of the toxicity data and the extent to which the chemical data accounts for observed biological effects in toxicity tests and from biological survey. This would then be used to decide whether there is no evidence of significant harm, a need to pursue risk management, or a need to carry out further assessments (Tier 3).

5.9.1 Modifications to the CSM in light of the Tier 2 assessment

The potential for significant harm has been clearly demonstrated at Site A, at least in patches close to the direct source of contaminants. The area of greatest concern could therefore be delineated more closely than was previously the case.

Uncertainties remain about the significance of metal residues at patches further away from the smelter than patch A3, and also about possible secondary effect pathways. As suggested in Section 5.7.1, though, it is reasonable to suppose there is potential for secondary poisoning effects, at least from cadmium. To understand whether these risks actually apply, these patches should progress to Tier 3.

The Tier 2 assessment does not suggest any new contaminant -pathway-exposure linkages.

5.10 Summary

- C-P-R linkages were identified - metals - soil pore water - soil-dwelling species and potential for secondary poisoning identified that is, cadmium.
- The CSM (Tier 0) identified linkages that resulted in progression from Tier 0 to Tier 1.
- Significant contamination at patches close to smelter meant that SQGVs were exceeded; there was some confirmation of this in Microtox™. These results indicated progression to Tier 2 to investigate potential risk further.
- Tier 2 biological testing left some uncertainties regarding potential secondary poisoning, therefore further investigation within Tier 3 is justified.

Testing the ERA framework: Site B

This section describes the ERA for Site B. Site B is a demolished tank farm area where crude oil and refined petroleum products used to be stored. It comprises an area of approximately 150 metres by 150 metres, currently covered by rough grassland and situated at the edge of an operational manufacturing facility.

6.1 Site B description

Between 1967 and 1972, six oil storage tanks were built on the site and used for storing crude oil or refined petroleum products. By 1986, the storage tanks had become surplus to requirement and were removed. Following demolition of the concrete tank bases and removal of the larger concrete fractions, the resulting smaller pieces of rubble from the tank bases were covered with 25-50 centimetres of topsoil. A fairly uniform grass cover, interspersed with a range of shrub and herb species, has since appeared. The site has been investigated twice since the tanks were demolished and the soil covered. A plan of the site with the locations of these sampling points is presented in Fig. 6.1.

A programme of window sampling in 1998 and 1999 (Figure 6.1) indicated that soils from a limited number of areas within the site contained elevated concentrations of either fresh or weathered crude oil. Most of these areas of hydrocarbon contamination were associated either with the previous location of tank number '1986' or with the drainage channels that surrounded the site. Based on total petroleum hydrocarbon (TPH) data from the window-bore sampling, we selected locations where a range of TPH concentrations in the top one metre of soil should be found. The location of these patches and the TPH concentrations found during the window sampling are shown in Figure 6.2.

6.1.1 Justification for using Site B

Site B was selected for use in the project because the contamination there was different to that at Site A, and because of its previous use and its availability. Site B was considered likely to be typical of sites that will be encountered during enforcement of the Part IIA contaminated land regime in terms of types and concentrations of contaminants present. While the site is located within the boundary of a working industrial facility, it is close to the perimeter. This means there would be at least a theoretical possibility for horizontal movement of contaminants (by association of contaminants with soil particles mobilised by rainfall, or by minor flooding events). Indeed, a hypothetical scenario can be created in which the site is placed immediately adjacent to a Part IIA ecoreceptor such as a SSSI. It could then be subject to formal investigation under the Part IIA regime, if concerns regarding potential contamination of the ecoreceptor as a result of former activities on the site were raised.

6.1.2 Site walkover

We made an initial site visit at the start of the project. The purpose of this was twofold: to conduct a visual inspection of the site, and also to collect soil samples for analysis in order to confirm the presence of the TPH gradient between patches indicated in the window-sampling data of 1998-9.

During this initial visit, we noted several spots of crude oil on the soil surface. These seemed to be emanating from sources below 0.5 metres. In any case, it seemed that pools of hydrocarbon remained at some depth below the current overfills. We collected soil samples for TPH analysis. We also collected soil from a separate patch at which surface

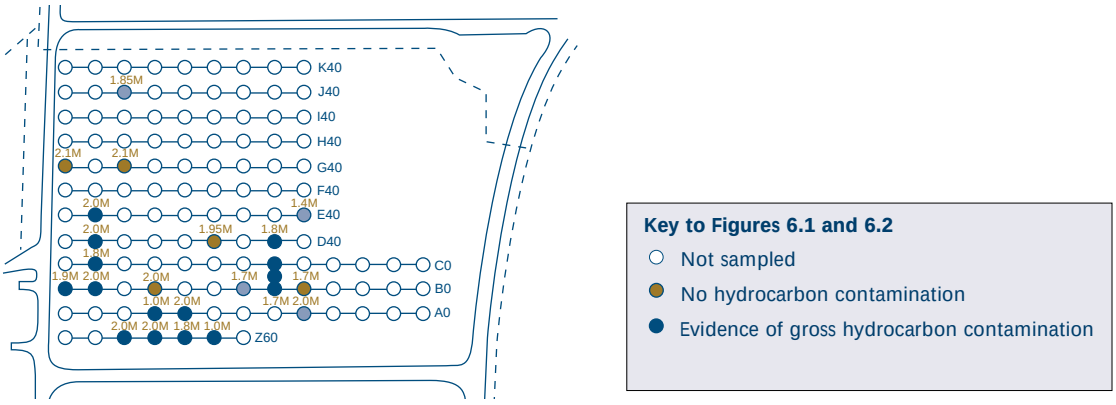


Figure 6.1 Location of window sampling in 1998/99 and total petroleum hydrocarbon concentrations (µg TPH g^{-1} soil). (Site map and TPH data provided by site owners)

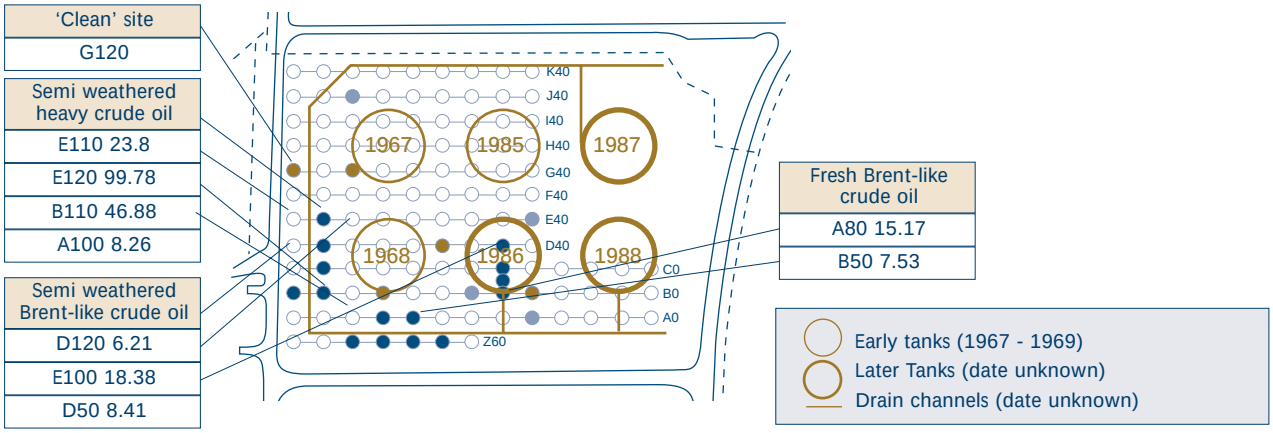


Figure 6.2 Sampling points where bulk soil was removed for chemical analyses and biological testing. (TPH concentrations are those from historical data)

oil contamination was observed. The samples were returned and stored at 3°C, and a sub-sample submitted for TPH analysis. At the same time, we also collected large soil volumes for use in more detailed chemical analysis and bioassays.

6.2 Theoretical ERA scenario

A hypothetical ecoreceptor has been superimposed on-site B by placing a SSSI bordering the site (Fig 6.3). This comprises grassland that is home to several species of plants and birds of high conservation value.

Visual inspection of the industrial site from the fenced boundary of the SSSI suggests that, in the recent past, hydrocarbon material has leaked from one of the oil storage tanks (probably Tank 3) located about 40 metres from the boundary fence. This leakage was probably due to a minor rupture of one of the supply pipelines. The extent of the visible hydrocarbon contamination (black crude oil) extends close to the boundary of the SSSI. This prompts the concern that there may have been leakage of lighter fractions from the area that has extended over the boundary into the SSSI, or that there has been some horizontal flow of the lighter hydrocarbons into the SSSI by surface flow during periods of intense rainfall.

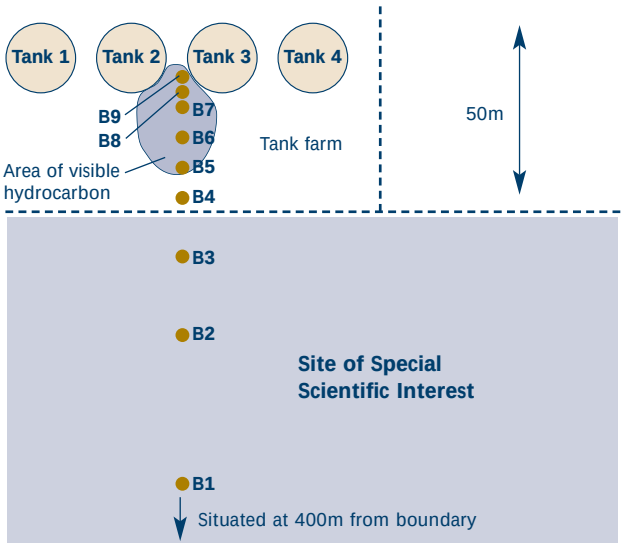


Figure 6.3 Schematic representation of the hypothetical ecological receptor used for Site B showing the location of the patch soils collected from the site.

6.3 Tier 0 assessment

The hypothetical SSSI adjacent to Site B is an area of unimproved wet grassland, one of the most threatened wildlife habitats in the UK. The rich flora includes species such as the southern marsh orchid, greater burnet, quaking grass and lady's smock. There is also a great diversity of butterflies and damselflies. Dense, overgrown hedges support abundant bird life at all times of the year. One of the main features of the site is the abundance of raptors feeding on the site; these have attracted bird-watchers. Kestrels and barn owls are regularly seen hunting over the fields. Probably the rarest biological feature of the site is the colony of breeding hobbies. Though now increasing in number, this species remains one of Britain's rarest breeding bird. A stream that meanders through the meadows provides good breeding ground potential for redshanks, which currently breed only in a handful of places in this area.

6.3.1 Site information and initial observation from a hypothetical site walkover

Geographical information: The site is approximately 20 metres above sea-level, with a minimal slope from North to South (in the direction of the adjacent SSSI).

Contamination sources: Staining and odours provide evidence of hydrocarbon contamination at the soil surface. This contamination is apparently most severe to the western side of the former Tank 3 on the site.

Ecosystem characteristics: The site itself is a grassed area with a range of shrub and herb species present. This vegetation is patchy and bare in the area where hydrocarbon staining is evident. The vegetation on the site extends to the boundary fence; beyond which the grassland area of the SSSI begins.

6.3.2 Summary

The initial investigation of the history of operations at Site B and the outcomes of the site walkover revealed a potential contaminant-pathway-receptor linkage (Table 6.1). The principal historic source of contamination appears to be a spill or leak of crude petroleum product from the now removed Tank 3 on a partially remediated tank farm area. Visual inspection of the site indicated the presence of hydrocarbon contamination in an area that spread some 20 metres away from the likely source. This is to within 20 metres of the boundary fence of the adjacent SSSI. Inspection of the area beyond the visual spill suggests that there may have been some movement of the crude oil (or some derived product) further towards the boundary. This is indicated by hydrocarbon odour, which can be detected in soil up to, and possibly beyond, the boundary fence. The presence of visual contamination and odour close to the SSSI suggests there may have been a pathway for the oil to the ecoreceptor following a period of flooding when free phase and dissolved phase oil was carried over the boundary. The SSSI receptor itself has high biological value. Specific features can be recognised as being vulnerable both to direct hydrocarbon exposure and possibly exposure through the food chain.

We prepared a conceptual site model (CSM) for Site B based on the initial identification of the principal contaminants and pathways and the characteristics of the ecological receptor (Fig 6.4). The primary contaminant source identified was of hydrocarbon leakage from storage tanks. The main pathway was horizontal movement at the soil surface on the boundary of the ecoreceptor. Once present in these soils, the individual hydrocarbon contaminants could

Table 6.1 Summary of the contaminant-pathway-receptor linkage for Site B

Source	Contaminant Pathway	Contaminant Source	Receptors
Leakage of petroleum products from site	Horizontal surface flow of hydrocarbons attached to soil particles during storm events. Flow to groundwater is also probable as is movement to controlled surface waters (however these remain beyond the remit of this study)	TPH associated with soil particles or waters	**Site of Special Scientific Interest** Soil Ecosystem Plants Invertebrates Population size (maintenance) of recognised features (e.g. birds) Food chain: Plant/herbivore Food chain: Invertebrate - bird/small mammal - bird

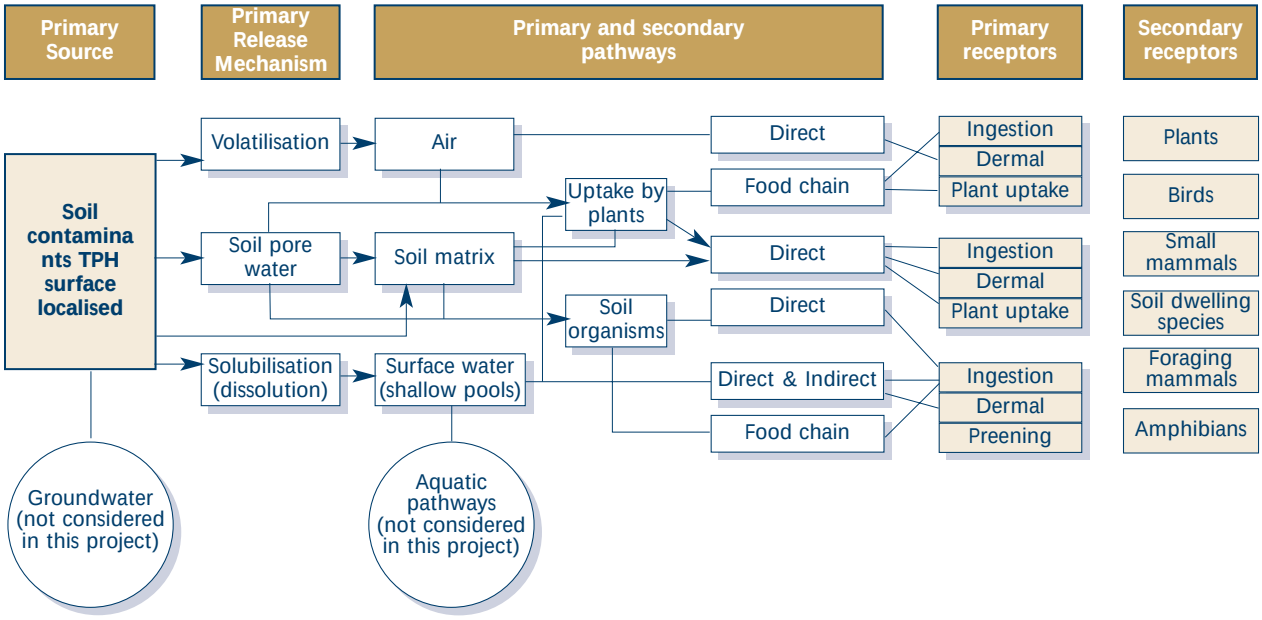


Figure 6.4 The initial CSM for Site B. The model shows main sources, pathways and receptors

potentially be available to plants, soil and soil surface-dwelling invertebrates, and the various component groups that comprise the soil microbial community. As was the case for all three of these receptors at Site A, the principal potential effects of exposure for these groups would be through direct exposure, causing either acute or chronic toxicity.

In addition to the primary receptors, other species present on the hypothetical SSSI could experience secondary exposure to hydrocarbon contaminants through the terrestrial food chain. Examples include:

- Seed-eating birds, vegetarian small mammal species such as watervoles, and many species of phytophagous and detritivorous invertebrates;
- Small mammals, birds (including raptors such as kestrels and hobbies), amphibians and reptiles feeding on soil invertebrates.
- Raptor species (kestrels, barn owls, hobbies) feeding on small mammals, birds, amphibians and reptiles that have accumulated TPH in body tissues as a direct result of exposure to elevated hydrocarbon concentration in their diet.

In fact, a comparison of the conceptual models for Site A and B (see Figure 6.4) indicated many similarities. In both cases, there are two designations of receptor:

- Primary receptors are all principally soil-dwelling species. These supply essential services such as primary production (plants), catalysts of nutrient turnover (soil invertebrates) and nutrient turnover itself (microbes) to the ecosystem. For these groups, protecting diversity and/or function activity is the principal concern, with protection of specific species of high conservation value a concern only for plants;
- Secondary receptors are exposed through the food chain rather than by direct contact with contaminated soil. The exact nature of the secondary receptors to be considered will vary depending on what species are present within the ecoreceptor. These species are often of high conservation value, so targeted assessment may be necessary in the later stages of ERA.

From the CSM (Figure 6.4), we identified the primary assessment endpoint (preservation of ecosystem function) and selected surrogate measurement endpoints (see Table 6.2 for details). Following screening at Tier 1, soils samples were tested with all the selected biological tests.

6.4 Tier 1 assessment

We carried out the following activities at site B in order to meet the requirements of measuring appropriate surrogate measurement endpoints to assess the assessment endpoint (as listed in Table 6.2).

We visited patches within site B to collect soils for use in chemical analysis and subsequent soil bioassays conducted as part of the Tier 2 assessment[4].

[4] Normally biological testing would be undertaken only if the outcome of any Tier 1 assessment suggested there was a requirement to advance to Tier 2

Table 6.2 Summary of receptors, their relevance to the hypothetical ecoreceptor and assessment and measurement endpoints.

Receptor	Relevance to the hypothetical ecoreceptor	Assessment endpoint	Measurment endpoints
Plant community	General plant species. Food supply to animal species.	Ecosystem function in order to sustain plant germination, growth and biodiversity.	Plant toxicity test with monocotyledenous species. Plant toxicity test with dicotyledenous species.
Soil invertebrate community	Food supply to animal species.	Ecosystem function in order to sustain invertebrate populations.	Earthworm acute toxicity test. Chronic springtail toxicity test.
Soil microbial community	Nutrient supply to support plant growth.	Ecosystem function in order to sustain microbial populations.	Soil nitrification rate. Soil carbon mineralisation rate.
Seed eating birds	Variety of woodland passerine and other species.	Ecosystem diversity in order to sustain suitable seed plants.	Chemical measurement of TPH residues in plant material collected at termination of the plant toxicity test.
Vegetarian small mammals	Small mammals such as voles.	Ecosystem diversity in order to sustain suitable food plants.	Chemical measurement of TPH residues in plant material collected at termination of the plant toxicity test.
Small mammals	Small mammals such as shrews	Ecosystem diversity in order to sustain suitable insect prey.	Analysis of TPH residue in earthworms at end of earthworm acute toxicity test.
Raptors	Protected raptor species such as kestrel and hobby.	Ecosystem diversity in order to sustain suitable insect prey.	Analysis of earthworms collected on termination of the earthworm acute toxicity test.
Amphibians and Reptiles	Frog and toad populations in local ponds, sand lizards.	Ecosystem diversity in order to sustain prey.	Food chain modelling

1. We undertook chemical analyses based on the ICRCL analytical suite, supplemented by BTEX and detailed PAH (polycyclic aromatic hydrocarbon) analyses.
2. We compared measured soil chemical concentrations with SQGVs.
3. We undertook toxicity screening using Microtox™.

6.4.1 Sample collection

We sampled soil as described in Appendix D. Upon return, sub-samples for use in chemical analysis were immediately collected for storage at -20°C to prevent compound volatilisation and degradation. As more volatile organic compounds (for example, gasoline range hydrocarbons) were expected to be present in the Site B soils than at site A, soil was screened thorough a 10 millimetre mesh while still damp. They were then remixed and used (after passing through a two millimetre mesh where possible) directly for the chemical analyses and biological tests. This approach ensured the comparability of biological and chemical tests conducted on sub-samples of the soil from each sampling point. Even with this preparation method, a proportion of some of the more volatile fractions would (if present) have been lost from the samples. This approach does, though, represent the most practical option to testing soils where existing information suggests the presence of volatile compounds.

6.4.2 Chemical analyses

We conducted analyses twice during the study. First, we analysed TPH and metals on replicated soil samples. Subsequently, we measured concentrations of 54 separate PAHs and BTEX by the partner project P5-063/TR2. We conducted these latter analyses on an unreplicated sample of soil from each sampling position. Analytical methods were the same as those described in Section 5.5.1. Chemistry data for the 9 sampling points showed that the concentration gradient anticipated from the historical analytical data was not present in samples collected for this study. This is partly attributable to biodegradation. The historical samples analysed in 1998-9 would have undergone some degradation by microbial action and volatilisation in the intervening four years. More significantly, only the shallow (biologically active) surface soils were collected for testing when the majority of the contamination was below 0.5 metres.

We needed to determine the ability of the toxicity tests to resolve differences in contamination (one of the project objectives). So we tested soils containing a range of contaminant levels from the patches at site B. The collected samples had proved to have either very low or very high concentrations of TPH, so we decided instead to manipulate the soils to create a range of TPH concentrations. We therefore prepared a dilution series by mixing soils containing the highest concentrations (34,400 µg TPH g^{-1}) with a blend of the least contaminated soils (see Appendix D for details). In combination with five of the original soils, a dilution series of site B patches yielded soils with the following nominal range of TPH concentrations: 8 (control), 12, 160, 320, 700, 1600, 3500, 7250, 14,300 µg TPH g^{-1} wet wt soil. These patches were then designated patch B1 through to patch B9.

Comparison of site soil concentrations with SQGVs

TPH could not be detected in the B1 soil, indicating that this soil might be suitable as an on-site reference (control). Sites B2, B3 and B4 also contained very low levels of TPH (5.0-50.0µg TPH g^{-1} wet weight soil). In patch B9 (where there was visible oil contamination) we found a very high concentration of TPH. This exceeded the expected nominal concentration by almost an order of magnitude, suggesting that exposure to this soil would be above the nominal concentration predicted from the dilution series manipulation (Appendix D). It is, though, important to point out that this analysis is based on a single sample only. The soil sent for analysis was notably heterogeneous, with distinct lumps of oily material. As noted above, measured TPH concentrations were all much lower than anticipated in the dilution series (B5-B8). This is probably due to volatilisation of the short chain hydrocarbon during the mixing process, problems in the dilution series mixing and/or sample heterogeneity. Despite this, there was a clear increase in the concentration of TPH increasing in the order B5 < B6 < B7 < B8, with concentration roughly doubling between each treatment.

BTEX concentrations (data not shown in tables) were close to, or below, the detection limit in all soils except B7 and B9 (which were within a factor of two of the detection limit). As the detection limit is the same as the Dutch target values, these compounds may pose a minor risk in these soil samples. No soil contained BTEX concentrations above the New Dutch List intervention value. Measurement of BTEX levels in soils at the end of the earthworm bioassay found no BTEX compounds above the detection level.

Tables 6.3-6.5 summarise the results from the detailed analysis of the concentrations of 54 PAH compounds in all the Site B patch soils.

The PAH concentrations measured in each patch indicated the presence of a clear concentration series within the selected Site B patch soils. This ranged from very low levels in patches B1 and B2 to moderate concentrations in the most contaminated soils within the dilution series (B7, B8). High concentrations in the field soil were found where visible oil contamination was present (B9). Though there were differences between individual compounds, the trend for increases in concentration from patch B1 - B9 was reflected for all 54 of the

measured compounds (data not shown) as well as the sum of PAHs. The presence of a concentration series of PAH within the Site B patches was confirmed in the more detailed analyses of soil taken at the end of the 42-day earthworm bioassay. Again, the concentration increased from low in B1 soil to very high in the B9 soil for individual compounds and also for the sum of total PAHs.

As with Site A (Section 5), measured soil concentrations at Site B were compared with three international Soil Quality Guideline Values (SQGVs) from the United States (Department of Energy DoE), the Netherlands (Dutch Target and Intervention values) and Canada (CCME) (Tables 6.3-6.5, respectively). Again, due to the harmonisation of the CCME and Dutch guidelines, a number of the SQGVs may be based on human health endpoints as opposed to ecological values. Where this is the case, the guidelines in question have been highlighted within the tables. In contrast with Site A, far fewer chemicals exceeded the selected guidelines. As expected with a petroleum-contaminated site, the levels of most metals were below guideline threshold values. Only lead, zinc, chromium and mercury concentrations exceeded SQGVs, and generally only in the most TPH/PAH contaminated samples (B7-B9).

Only the Dutch have guideline values for total PAH. All samples except B1 exceeded the Dutch Target Values. Therefore, based on the exceedences of these particular PAH guidelines, sites B2-B9 may not provide sustainable soil conditions.

Interestingly, there are no international SQGVs available for TPH. It is difficult, therefore, to establish the level of contamination due to the presence of these chemicals at the site. The Dutch guidelines, though, provide target and intervention levels for mineral oil. In the absence of TPH SQGVs, these are the most relevant guidelines available. Comparison of the mineral oil standard with TPH concentrations showed that all patches (except B1 and B2) exceeded the listed 'optimum' level (below which no toxicity has been observed) as (50 mg/kg). They also showed that patches B8 and B9 (Table 6.4) exceeded the intervention level (5000 mg/kg).

One interesting point arose as a result of analysis of the extended suite of PAHs. This is that the usual practice of analysing soils for only the 16 individual PAHs recommended by the US EPA for site investigations gave a rather poor picture of the PAH contamination present. The US EPA suite was defined with reference to gasworks sites. The percentage of total PAHs represented by the US EPA standard suite was 50 per cent in soil B1-B4, but decreased steadily through the dilution series up to only approximately 10 per cent in soil B9. Analysis of the restricted US EPA PAH suite in these more contaminated soils would thus have underestimated PAH exposure by a factor of 10.

Patches B1-B6 showed the least contamination. Patches B7-B9 showed the greatest contamination. Based on the exceedences of the PAH guidelines, all sample patches, except B1, would go through to Tier 2 of the ERA.

6.4.3 Toxicity screening

There are two possible approaches to testing soil samples in the Microtox™ test. In the conventional Basic and 100 per cent Toxicity Test Procedures, populations of the bioluminescent marine bacterium *Vibrio fischeri* are exposed, for a period ranging from five to 30 minutes, to an aqueous extract of the sample under test. In the Solid Phase Test (SPT), bacterial populations are exposed for a period of 30 minutes in the presence of the soil slurry. This allows the testing of a solid sample using serial dilutions of the sample, where the test organisms come into direct contact with the solid sample in an aqueous suspension of the test sample. It is thus possible to detect the toxicity of compounds that are non-extractable in the conventional process.

We prepared leachates of Kettering loam and 10 soil samples taken from Site B for definitive Microtox™ testing with the 100 per cent Toxicity Test Procedure. To ensure the test material was leached from the soils, the control soil and five of the Site B soils (B5, B6, B7, B8 and B10) were leached for 2, 8 and 24 hours. Five of the soils (B1, B2, B3, B4, and B9) were leached for 8 and 24 hours. The SPT procedure was conducted on a control soil (Kettering loam), a positive control soil (Kettering loam spiked with zinc) and on the same 10 soil samples.

Table 6.6 shows the Microtox™ toxicity test results for the 10 Site B soil samples conducted on 2, 8 and 24 hour leachates and the Solid Phase Tests carried out on the soils.

The tests with the leachates showed that, in the control soil (Kettering Loam) and in the samples from patches with low levels of hydrocarbon contamination (B1, B2, B3 and B4), leaching over eight hours or 24 hours resulted in a leachate that stimulated light output in the Microtox™ test. These effects were probably due to the extraction of nutrients from the soils into the leachates, which resulted in increased bacterial metabolism and increased light output.

Table 6.3 Comparison of Site B soils with US DoE SQGVs

SQGV	T-CN	F-CN	S	As	B	Cd	T-Cr	Cu	Hg	Ni	Pb	Se	SO4	TDYS %	Zn	Total	TPH
Earthworm	N/G	N/G	N/G	60	N/S	20	0.4	50	0.1	200	500	70	N/G	N/G	200	N/S	N/S
Microbial	N/G	N/G	N/G	100	3000	20	10	100	30	90	900	100	N/G	N/G	100	N/S	N/S
Plant	N/G	N/G	N/G	10	500	4	1	100	0.3	30	50	1	N/G	N/G	50	N/S	N/S
SQGV	**T-CN**	**F-CN**	**S**	**As**	**B**	**Cd**	**T-Cr**	**Cu**	**Hg**	**Ni**	**Pb**	**Se**	**SO4**	**TDYS %**	**Zn**	**Total**	**TPH**
B1	<2.5	<1	<0.1	6	0.12	0.13	13.2	7.5	<0.01	10.4	28.2	<0.8	17.6	84.1	27	0.6	8.7
B2	<2.5	<1	<0.1	6.7	<0.1	0.13	12.8	10.7	<0.01	12.4	38.2	<0.8	17.7	83.1	40.2	1.6	11.3
B3	<2.5	<1	0.2	5.72	0.2	0.14	12.4	13.2	0.04	11.6	45.6	<0.8	16.4	83.7	45.6	3.3	160
B4	<2.5	<1	0.2	6.3	<0.1	0.13	13.7	15.6	0.04	12	48.9	<0.8	17.3	82.5	38.7	1.5	320
B5	<2.5	<1	0.3	6.73	0.13	0.2	14.6	17.5	0.05	11.1	50.5	<0.8	83.2	83.2	45	5.8	1075
B6	<2.5	<1	0.5	6.7	<0.1	0.22	14.5	16	0.05	12.5	55.7	<0.8	17.2	82.7	52.9	6.1	2150
B7	<2.5	<1	1	6.06	0.13	0.17	12.4	14.3	0.12	11.5	52.4	<0.8	17.4	84.2	45.1	15.9	4300
B8	<2.5	<1	1.3	6.13	0.12	0.2	12.3	12.3	0.11	12.1	54.1	<0.8	17.2	83.2	42.3	25.3	8600
B9	<2.5	<1	3.5	6.45	0.13	0.21	13.7	23	0.13	15.7	86.7	<0.8	16.3	84.2	69.7	36.9	14300

N/G = No Guideline available

Exceedance of Earthworm benchmark (mg/kg)

Exceedance of Plant benchmark (mg/kg)

Concentration higher than two or more Guidelines

Table 6.4 Comparison of Site B soils with Dutch Target and Intervention Values

SQGV	T-CN	F-CN	S	As	B	Cd	T-Cr	Cu	Hg	Ni	Pb	Se	SO4	TDYS %	Zn	Total	TPH
Target	5	1	N/G	29	N/G	0.8	100	36	0.3	35	85	N/G	N/G	N/G	140	1	50*
Intervention	650/50†	20	N/G	55	N/G	12	380	190	10	210	530	N/G	N/G	N/G	720	40	5000*
SQGV	**T-CN**	**F-CN**	**S**	**As**	**B**	**Cd**	**T-Cr**	**Cu**	**Hg**	**Ni**	**Pb**	**Se**	**SO4**	**TDYS %**	**Zn**	**Total**	**TPH**
B1	<2.5	<1	<0.1	6	0.12	0.13	13.2	7.5	<0.01	10.4	28.2	<0.8	17.6	84.1	27	0.6	8.7
B2	<2.5	<1	<0.1	6.7	<0.1	0.13	12.8	10.7	<0.01	12.4	38.2	<0.8	17.7	83.1	40.2	1.6	11.3
B3	<2.5	<1	0.2	5.72	0.2	0.14	12.4	13.2	0.04	11.6	45.6	<0.8	16.4	83.7	45.6	3.3	160
B4	<2.5	<1	0.2	6.3	<0.1	0.13	13.7	15.6	0.04	12	48.9	<0.8	17.3	82.5	38.7	1.5	320
B5	<2.5	<1	0.3	6.73	0.13	0.2	14.6	17.5	0.05	11.1	50.5	<0.8	83.2	83.2	45	5.8	1075
B6	<2.5	<1	0.5	6.7	<0.1	0.22	14.5	16	0.05	12.5	55.7	<0.8	17.2	82.7	52.9	6.1	2150
B7	<2.5	<1	1	6.06	0.13	0.17	12.4	14.3	0.12	11.5	52.4	<0.8	17.4	84.2	45.1	15.9	4300
B8	<2.5	<1	1.3	6.13	0.12	0.2	12.3	12.3	0.11	12.1	54.1	<0.8	17.2	83.2	42.3	25.3	8600
B9	<2.5	<1	3.5	6.45	0.13	0.21	13.7	23	0.13	15.7	86.7	<0.8	16.3	84.2	69.7	36.9	14300

N/G = No Guideline Available
† = Values relate to a pH<5 and pH≥5, respectively.
* = Guideline relates to mineral oil

Exceedance of Target Level (mg/kg dry weight)

Exceedance of Intervention Level (mg/kg dry weight)

Table 6.5 Comparison of Site B soils with CCME SQGVs

SQGV	T-CN	F-CN	S	As	B	Cd	T-Cr	Cu	Hg	Ni	Pb	Se	SO4	TDYS %	Zn	Total	TPH
Agricultural	N/G	N/G	500	12*	750	1.4*	64	N/G	6.6*	50	70	N/G	N/G	N/G	200	N/S	N/S
Res./Park	N/G	N/G	N/G	12*	500	10	64	N/G	6.6*	50	140*	N/G	N/G	N/G	200	N/S	N/S
Comm.	N/G	N/G	N/G	12*	2000	22	87	N/G	24*	50	260*	N/G	N/G	N/G	360	N/S	N/S
Industrial	5	5	N/G	12*	2000	22	87	N/G	50	50	600	N/G	N/G	N/G	360	N/S	N/S
SQGV	**T-CN**	**F-CN**	**S**	**As**	**B**	**Cd**	**T-Cr**	**Cu**	**Hg**	**Ni**	**Pb**	**Se**	**SO4**	**TDYS %**	**Zn**	**Total**	**TPH**
B1	<2.5	<1	<0.1	6	0.12	0.13	13.2	7.5	<0.01	10.4	28.2	<0.8	17.6	84.1	27	0.6	8.7
B2	<2.5	<1	<0.1	6.7	<0.1	0.13	12.8	10.7	<0.01	12.4	38.2	<0.8	17.7	83.1	40.2	1.6	11.3
B3	<2.5	<1	0.2	5.72	0.2	0.14	12.4	13.2	0.04	11.6	45.6	<0.8	16.4	83.7	45.6	3.3	160
B4	<2.5	<1	0.2	6.3	<0.1	0.13	13.7	15.6	0.04	12	48.9	<0.8	17.3	82.5	38.7	1.5	320
B5	<2.5	<1	0.3	6.73	0.13	0.2	14.6	17.5	0.05	11.1	50.5	<0.8	83.2	83.2	45	5.8	1075
B6	<2.5	<1	0.5	6.7	<0.1	0.22	14.5	16	0.05	12.5	55.7	<0.8	17.2	82.7	52.9	6.1	2150
B7	<2.5	<1	1	6.06	0.13	0.17	12.4	14.3	0.12	11.5	52.4	<0.8	17.4	84.2	45.1	15.9	4300
B8	<2.5	<1	1.3	6.13	0.12	0.2	12.3	12.3	0.11	12.1	54.1	<0.8	17.2	83.2	42.3	25.3	8600
B9	<2.5	<1	3.5	6.45	0.13	0.21	13.7	23	0.13	15.7	86.7	<0.8	16.3	84.2	69.7	36.9	14300

N/G = No Guideline available
* = Guideline based on human health effects

Exceedance of Target Level (mg/kg dry weight)

In the leachates of soils containing higher levels of hydrocarbon contamination (>2000 mg kg^{-1}), there was evidence that leaching for 8 hr or 24 hr resulted in the leachates causing inhibition of light output (relative to the control). Light output inhibition was generally below 50 per cent and was not sufficient to allow an IC_{50} value to be calculated (except for B10, where a 30 minute IC_{50} value of 87.9% was calculated (an excessively oil contaminated soil collected from Site B with very high hydrocarbon concentrations). These data suggest that either only low levels of hydrocarbons were being leached from the soils, that the test was not very sensitive to hydrocarbons, or that the soils were not toxic.

In the leachate of sample B10 (containing 34,400 mg/kg TPH), inhibition of light output (relative to the control) was observed after leaching for 8 and 24 hr. The responses after 30 minutes exposure to the 8 and 24 hr leachates were sufficient to allow IC_{50} values of 44.5 per cent and 47.1 per cent respectively to be calculated. In the solid phase tests, inhibition of light output (relative to the control) was observed for all the soils. The responses (30 min IC_{50} values) measured for soil samples were greater than those in the control soil (20.3 per cent) for all nine test soil samples, with IC_{50} values ranging from 11.3 (B8) to 0.3 per cent for B10. Table 6.6 shows that sample B10 (not used in the biological testing for soils at site B as it contained very high concentrations of TPH (>34,000 mg/Kg)) shows the full Microtox™ data for this additional soil sample.

These data indicate the relatively low toxicity of site B soils except where TPH/PAH concentrations were highest when using the standard extraction/elution Microtox™ assay. When tested using the solid-phase methodology, though, more of the available TPH may have come into direct contact with the bacteria, resulting in higher toxicities and providing a better measure of TPH bioavailability - and hence toxicity.

6.4.4 *Decision-making and modification of the CSM in light of the Tier 1 assessment*

Based on compliance with SQGVs, the Revised Dutch List PAH target value was exceeded at all patches except B1. None of these, though, exceeded the higher total PAH 'intervention' value (Table 6.7). Similarly, the mineral oil target value was exceeded at all patches except B1 and B2. In this case, samples from B8 and B9 also exceeded the 'intervention' value. The Microtox™ data showed no inhibition of luminescence in soil B2 to B5 when compared to B1. This suggests that there are no unmeasured contaminants present at toxicologically significant concentrations in these soils. Luminescence was, though, lower than the reference site for soil from patches B6 to B9. This provides broad agreement with chemical data, though some exceedances of SQGVs are indicated that were not detected by the Microtox™ test (for example, exceedances of the mineral oil target value in samples B3 to B5).

Exceedance of the Dutch List target standard at B3 for both PAHs and for mineral oil, and at B2 for PAHs, indicates the potential for a significant contaminant exposure at these locations. For this reason it would be necessary to proceed to a Tier 2 test regime for these patches. Though patches B4-B9 all contain elevated concentrations of PAHs and mineral oils, they are outside the SSSI and, as a result, will not need formal investigation. Toxicity data may, though, be useful, because it would then be possible to compare responses across a range of PAH and oil contamination concentrations.

The Tier 1 assessment suggests that there is some potential for exposure at patch B1. But the fact that the Revised Dutch List 'intervention' values are not exceeded suggests, though, that the extent of any impact could be limited. For this reason, it may be best for any further biological assessment to focus on chronic exposures. This includes the plant tests and the chronic springtail test, which should be used to gauge effects on soil invertebrates.

Because of the complex chemical nature of crude and refined oils, the potential for food-chain transfer of contaminants is not immediately apparent. Analysis of plant material at the end of the plant test is advised to account for the possible pathway between plants and invertebrates. Species such as earthworms can either be exposed to soil *ex situ* as part of an acute bioassay, or they (and other species such as woodlice and snails) can be collected directly from the field. Neither the direct toxicity pathways or the food chain route can be discounted on the basis of the assessment, so all pathways within the initial CSM remain viable.

6.5 Tier 2 Assessment

6.5.1 *Expression of biological effects data*

The Tier 1 assessment suggests that the extent of contamination of the SSSI as a result of the visible spill may be limited. But because SQGVs for two groups of compounds have been exceeded, we would have made the decision to go forward to biological testing, at least for patch B3.

For the purposes of this project, the concentration-response relationships for each of the previously recommended bioassays have been determined for

Table 6.6 Site B soils Microtox™ leachate and solid phase results. (Numbers in parentheses after the B1, B2 etc are the total sum of determined PAHs (54) for each representative patch soil)

	Microtox™ test IC_{50} results									
	2hr leachate % effect			8hr leachate % effect			24hr leachate % effect			Solid-Phase AC_{50}
	5 min	15 min	30 min	5 min	15 min	30 min	5 min	15 min	30 min	30 min
Kettering Loam (Control)	Enhancement	Enhancement	Enhancement	Enhancement	0	26.7	Enhancement	Enhancement	Enhancement	20.27 % (16.75 - 24.54)
B1 (0.6	-	-	-	Enhancement	Enhancement	Enhancement	Enhancement	Enhancement	No effect	5.30 % (4.55 - 6.16)
B2 (1.6)	-	-	-	Enhancement	Enhancement	Enhancement	Enhancement	Enhancement	No effect	3.13 % (2.83 - 3.47)
B3 (3.3)	-	-	-	Enhancement	Enhancement	No effect	Enhancement	Enhancement	20.4% Inhibition	6.32 % (5.70 - 7.01)
B4 (1.5)	-	-	-	Enhancement	Enhancement	Enhancement	Enhancement	Enhancement	No effect	2.81 % (2.21 - 3.57)
B5 (5.8)	No effect	No effect	No effect	No effect	No effect	No effect	Enhancement	Enhancement	Enhancement	8.9 % (5.61 - 14.09)
B6 (6.1)	27.5% Inhibition	No effect	No effect	27.2% Inhibition	No effect	23.5% Inhibition	No effect	No effect	No effect	7.43 % (5.96 - 9.27)
B7 (15.9)	Enhancement	No effect	No effect	24.8% Inhibition	No effect	38.1% Inhibition	17.8% Inhibition	25.4% Inhibition	23.0% Inhibition	2.34 % (2.07 - 2.65)
B8 (25.3)	Enhancement	Enhancement	Enhancement	17.0% Inhibition	No effect	32.8% Inhibition	0	24.4% Inhibition	30.6% Inhibition	11.25 % (10.79 - 11.72)
B9 (36.9)	-	-	-	21.6% Inhibition	37.1 % Inhibition	46.5% Inhibition	37.5% Inhibition	38.3% Inhibition	52.7% Inhibition IC_{50} = 87.9% (82.2 - 94.0)	0.387 % (0.32 - 0.47)
B10 (soil containing >34,000 mg/kg TPH; not used for biological testing other than Microtox™)	Enhancement	Enhancement	IC_{50} = 86.7% (64.7 - 116.3)	IC_{50} = 55.6% (42.8 - 72.4)	IC_{50} = 59.3% (54.2 - 64.9)	IC_{50} = 44.5% (39.3 - 50.4)	IC_{50} = 51.3% (48.2 - 54.7)	IC_{50} = 47.3% (40.9 - 54.7)	IC_{50} = 47.1% (41.1 - 53.9)	0.306 % (0.25 - 0.37)

Enhancement - Stimulation of light output above that observed in the controls
No effect - No difference from light output in control.

Table 6.7 Outcomes of the Tier 1 assessment for the nine patches at Site B used as potential location for the hypothetical ecoreceptor

Site sample	SQGV exceedance at the site - yes/no, and major chemical or group in exceedance (in parentheses)	Assessment endpoint	Measurement endpoints
B1	N	No evidence	No obvious ecological effects
B2	Y (PAH)	No evidence	No obvious ecological effects
B3	Y (PAH, mineral oil)	Weak odour	No obvious ecological effects
B4	Y (PAH, mineral oil)	By odour	No obvious ecological effects
B5	Y (PAH, mineral oil)	By odour/visual surface staining	Reduced plant biodiversity but outside SSSI
B6	Y (PAH, mineral oil)	By strong odour/visual surface staining	Low plant biodiversity
B7	Y (PAH, mineral oil)	By strong odour/visual surface staining	Bare soil patch
B8	Y (PAH, mineral oil)	By strong odour/pronounced visual surface staining	Large area of bare stained soil
B9	Y (PAH, mineral oil)	By strong odour/pronounced visual surface staining	Large area of bare soil with presence of pooled oil on surface

all Site B soils. At Site A, distance from contaminant source was a convenient x variable. At site B, though, we decided to plot responses against the concentration of an 'indicator compound'. We chose total PAH for this purpose. We used this in preference to TPH because the TPH measurement was based only on a single value measurement made in two pseudoreplicates. The total PAH value was, though, based on the sum measurement of 54 separate compounds. Because more measurements were made for the PAH, the sum value is likely to be a more accurate comparative representation of the concentrations of these compounds in each soil.

As described in Appendix B, we developed a number of modifications of the OECD plant tests. Data presented here are again confined to dry weight measurements, because this endpoint was felt to be the most useful measure of toxicity.

6.5.2 Carbon and nitrogen mineralisation

Results from the nitrogen mineralisation test at site B were inconclusive. All the sample locations showed lower nitrate production than the control. There was no clear relation between hydrocarbon concentration and nitrate production.

The results recorded in the carbon mineralisation were also inconclusive, very varied and difficult to interpret. There were no obvious trends in response that would assist in decision-making within the ERA.

6.5.3 Plant tests

We carried out tests on the effects of Site B soils on higher plant growth using the same procedure as that used to assess Site A soils (Section 5). Again, we tested five species: three dicotyledons (cabbage, tomato and pea) and two monocotyledons (wheat and oat). Plants were exposed to nine Site B soils. Two of these (B1-PAH 0.6 mg/kg and B2-PAH 1.6 mg/kg) were classed as site controls, a test control of Kettering loam and a positive control using zinc. We analysed results using one-way ANOVA and multiple comparisons tests (Tukey's HSD) (See Figure 6.5 a-e for data)

Results are plotted against total PAH (sum of 54 individual PAHs) in the soils. There is no discernible effect of PAH gradient for any of the species tested. Statistical analysis of the data indicates no significant difference ($P < 0.05$) in dry weight of plants exposed to the highest PAH concentration when compared with plants exposed to soils with lower concentrations of PAH. The soil quality at Site B also appears to have had a significant effect on plant growth. All species (apart from pea) showed significantly lower growth ($P < 0.05$) in the site control soils (B1-PAH 0.6 mg/kg and B2-PAH 1.6 mg/kg) when compared with the test control (Kettering loam). Consequently, effects of soil quality (nutrient status, physical characteristics) could mask any effects of soil contamination.

The results from the plant tests suggest that contaminants are either present at levels insufficient to elicit biological effects, or that they are not biologically available, or that they are not toxic.

Figure 6.5 Plots representing the total dry weight for each of the test plant species (average of five replicates) after germination and subsequent growth in soil collected from Site B.

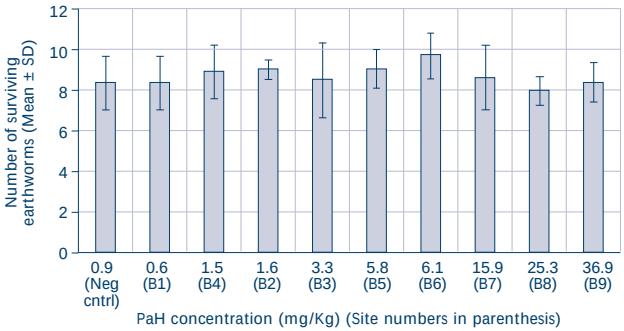


Figure 6.6 Survival (Mean ± SD, *n* = 4) of *Eisenia andrei* incubated for 14 days in soils collected at all Site B patches plotted on the basis of total PAH concentrations.

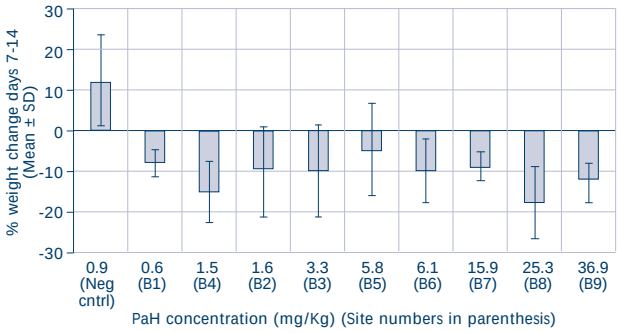


Figure 6.7 Earthworm weight change (Mean ± SD, *n* = 4) of *Eisenia andrei* incubated for 14 days in soils collected at all Site B patches plotted on the basis of total PAH concentrations.

6.5.4 Earthworm survival and growth

Survival

As found at Site A, the earthworm test performed within the normal acceptance thresholds for survival in the procedural and site controls. In addition, there were no significant effects of any of the soils on earthworm survival at either 7 or 14 days (ANOVA; 7 days, $F = 0.6$, $p = 0.78$; 14 days $F = 0.59$, $p = 0.79$). When earthworm survival is plotted against the total PAH concentrations for each of the patch soils, no overall trend was apparent. Within the four soils stemming from the dilution series from soil collected from patch B10 (see Appendix D), though, there was an indication that survival decreases with increasing total PAH concentration (Figure 6.6). The lack of an overall effect was probably due to the mixtures of organic compounds in soils from Site B being highly varied between patches, while the dilution series of the initial TPH soil (patch B10) used to create the soil gradient provided a mixture of similar compounds.

Growth

There was no significant effect of any site B patches on earthworm weight change (ANOVA, $F = 1.06$, $p = 0.423$) (Figure 6.7), nor any overall trends based on total PAH concentration.

One other significant observation was that earthworms experienced a very high weight loss of 15 per cent at patch B4, even though the total PAH concentration was 1.5 mg/kg. This supported the observation of pronounced sub-lethal effects in this patch soil made in associated work looking at earthworm reproduction in Site B soils (R&D project P5-063, Spurgeon *et al.*, in press). The fact that these two studies both indicated sub-lethal stress for earthworms in the B4 soil suggests that some part of the mixture of compounds present at B4 is unique to this patch and may influence overall earthworm performance.

6.5.5 Springtails survival and juvenile production

Survival

The springtail tests we conducted performed poorly, frequently failing to meet the validation criteria for the ISO (1999) test (see Table 4.1). For Site B soils, all but one replicate had adult mortality greater than the 20 per cent allowed for controls. No replicate produced more than the 100 juveniles specified for controls. There was no significant effect of any site B patch on the adult survival (ANOVA ($F = 1.86$, $p = 0.098$) (Figure 6.8). Also, the level of variability observed was much higher than that experienced with the soils from Site A, which means that there are no indications of a trend in this data set. The data should not be used for risk predictions.

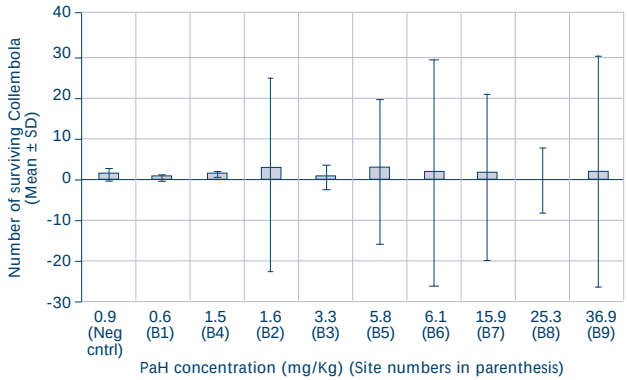


Figure 6.8 Survival (mean ± SD, *n* = 5) of *Folsomia candida* incubated in soils collected at all Site B patches plotted on the basis of total PAH concentrations.

Juvenile production

Effects on juvenile production reflect those on adult survival, as had been seen for Site A. There were no significant effects of the Site B patches on juvenile numbers produced (ANOVA (F = 1.48, p = 0.201) (Figure 6.9). Again, the level of variability observed was much higher than that experienced with the soils from Site A. Consequently, we cannot comment on any trend in the data set

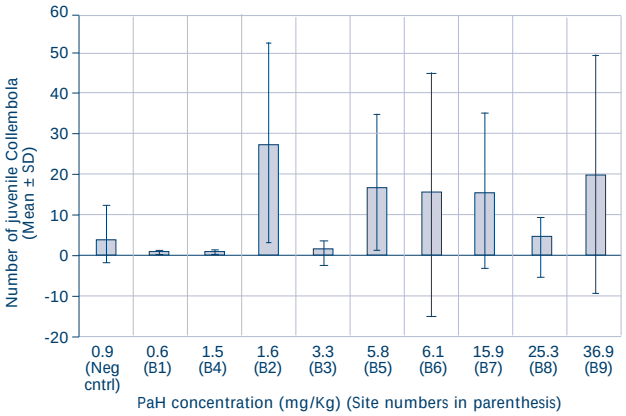


Figure 6.9 Number of juveniles (Mean ± SD, n = 5) produced from an initial population of 10 adult *Folsomia candida* incubated in soils collected at all Site B patches plotted on the basis of total PAH concentrations.

6.6 Decision-making (Tier 2)

Section 8 covers the decision-making process on the basis of the test data gathered. The following summarises the major outcome of the biological testing for Site B.

- We identified C-P-R linkages: TPH - soil pore water/soil particles - soil-dwelling species, and we identified the potential for secondary poisoning through ingestion.
- The CSM (Tier 0) identified contaminant pathways and linkages that resulted in progression from Tier 0 to Tier 1.
- Significant contamination at several patches close to the source meant that SQGVs (or surrogate measures) were exceeded. There was some confirmation of this by solid phase Microtox™. These results indicated progression to Tier 2 to investigate potential risk further.
- Tier 2 biological testing left some uncertainties regarding potential secondary poisoning. Further investigation within Tier 3 would therefore be justified.

Performance of biological tests

In Sections 5 and 6, we described the results obtained from the battery of standardised and sub-lethal biological tests applied at Sites A and B, along with a summary of the interpretation of available chemical data from the same locations.

7.1 Introduction

In this section, we evaluate the performance of individual test methods and the information they have generated in terms of the value they contribute to decision-making. This assessment meets objectives 1 and 2 of the project (Section 1.5).

We review each of these tests, discuss their major strengths and limitations and suggest how they might usefully be incorporated within an ERA framework. It should be noted that some of the performance criteria suggested by Hopkin (1993) cannot actually be evaluated (reproducibility between and within laboratories) because the tests were neither repeated with reference materials nor tested in different laboratories. This aspect of test performance therefore remains unknown. The standardised tests can, though, be conducted to Good Laboratory Practice (GLP) standard, and this quality assurance process should ensure high reproducibility.

7.2 Evaluation of individual biological tests

For each test method, we provide a brief introduction to the test and describe its performance at study sites A and B. We then comment on the practicalities of performing the test, their responsiveness and robustness, and comment on their ecological relevance. Finally, we make a recommendation about further inclusion in the ERA framework and the role to which each biological test method is best suited.

7.2.1 MicrotoxTM

Introduction

The Microtox™ test has been used extensively as a screening tool, in Toxicity Reduction Evaluations (TREs) and in Toxicity Identification Evaluations (TIEs) in connection with wastewater discharges to the aquatic environment (for example, USEPA 1989a,b; USEPA, 1995) and in some cases, it has also been used to investigate contaminated soils (Dorn *et al*., 1998; Doherty, 2001). Its main attraction is that it can generate toxicity data very rapidly (within minutes), using well-defined procedures and equipment. Some regulatory jurisdictions, including the Environment Agency, have developed their own test protocols.

Performance

The test protocols are well defined. In tests with aquatic samples, at least, this contributes to a level of repeatability and reproducibility that is higher than that of many other biological tests (Whitehouse *et al*., 1996).

The results in Section 5 show a clear relationship between levels of metals at site A and response levels in the Microtox™ test, with inhibition of light output being greater at locations closest to the smelter. At site B, though, we saw responses to aqueous leachates only from soils containing the highest levels of hydrocarbon contamination (Table 6.6). Indeed, in most samples, there is evidence of a stimulation of light output, probably as a result of extraction of nutrients during the leaching process.

The Microtox™ procedure permits both a liquid-phase exposure of (dissolved) contaminants and a solid-phase version that involves exposure of bound contaminant residues. Experience of hydrocarbon-contaminated soil from Site B shows that the solid phase test with aqueous suspensions of soil is significantly more sensitive than tests on aqueous leachates. All the soil samples yielded modest levels of toxicity, but with no apparent relationship between PAH or TPH contamination and toxicity, that is, low discrimination. It is possible that these rather uniform responses were due to the effects of interferences (impaired light transmission and reductions in measured differences in light output as a result) rather than true differences in toxicity. Doherty (2001) highlights a possible link between silt: clay: ratio in determining the extent of interference by colour. Without further investigation, it is unclear whether this is a general trend or merely a characteristic of this site.

Practicalities

In terms of the practicalities of testing, speed and understanding about sensitivity to a wide range of contaminants, the Microtox™ test is probably unequalled. In all these respects, the Microtox™ test performs very well. There is also evidence (Whitehouse *et al*., 1996) that, compared to other standard aquatic toxicity tests, it is both more repeatable and reproducible.

Responsiveness

There is a substantial published database of chemical toxicity commercially available (see website of the manufacturer www.azurenv.com/mtox.htm). One known weakness is its low sensitivity to some commonly occurring contaminants such as metals, pesticides with specific modes of action and PAHs (Doherty, 2001).

Robustness

Apart from certain classes of compound, the test responds to a wide spectrum of contaminants and has been adapted to be used for both liquid and solid media. The method of presenting soil samples in the test (that is, aqueous leachates or solid phase testing) can, though, have a marked impact on the conclusions drawn, especially for soils contaminated with substances that have low solubility in water. This mirrors conclusions drawn by Doherty (2001) in a study of the usefulness of Microtox™ for assessing contaminated sediments and soils. The toxicity of such substances in aqueous leachates could be underestimated, and there is therefore a risk of failing to detect the presence of toxic substances that could affect other biota (as shown by the greater sensitivity of the plant tests, for example). It is clear that further investigation of methods of sample preparation would be beneficial.

Relevance

Questions remain about the ecological relevance of a bioluminescent marine bacterium as an indicator of effects that contaminants have on soil organisms. A practical consequence of this is that salt must be added to the test medium, and this could affect speciation of some substances. Nevertheless, because we advocate its use as a screening tool at Tier 1, ecological relevance should not be over-emphasised. Issues of relevance are more important at higher tiers of the framework. As far as screening purposes are concerned, we argue that the practical advantages outweigh this concern.

There is considerable research interest in incorporating the *lux* gene responsible for bioluminescence into other organisms (See Spurgeon *et al*., 2002 for a full summary). If this were achieved for soil organisms, or at least for soil bacteria, it would improve the ecological relevance of the test, but is not a prerequisite for inclusion in the ERA framework. A more important aspect of alternative bioluminescene tests is the adequacy of QA and QC procedures and the standardisation of the test method. For a screening test, the key question is whether the levels of false negatives and false positives are acceptable. Suter *et al*., (2000) cover issues surrounding screening particularly from a chemical point of view; they discuss screening against ecotoxicological benchmarks. They do not, though, examine the issues surrounding the standardisation of test methods to reduce false reporting. Environment Canada has initiated a programme to develop a series of standardised biological test methods and guidance documents. EPS1/RM/42 covers the standardisation of one method using luminescent bacteria, including issues pertaining to normalising data for moisture, turbidity and colour.

Proposed use in ERA

Microtox™ is recommended for use in the ERA framework as a screening test within Tier 1, though it may also have some applications at higher tiers, for example for mapping toxicity. Similar conclusions were drawn in the review of this method for testing of soils by Spurgeon *et al*., (2002). Tier 1 is essentially an assessment of the concentrations of contaminants present at the site against suitable SQGVs. It is thus chemistry-driven, and Microtox™ may be used in support of this or to reduce uncertainty.

7.2.2 Carbon and nitrogen mineralisation

Introduction

Soil microbial processes, which include carbon and nitrogen mineralisation, are mediated by specific components of the soil microbial biomass. Nitrogen and carbon cycling are vital processes in soils, and changes in soil mineralisation rates may be indicative of the effect of a pollutant. Soil-testing programmes include the mineralisation of organic matter and other functions of carbon and nitrogen cycles because they are thought to be sensitive to pollutants and to represent processes that determine the availability of nutrients through the food chain. Bulk soil respiration methods, though, have been challenged (for example, Domsch, *et al.,* 1983) for being insensitive to pollutants. Nevertheless, nitrogen mineralisation tests have been used to assess the effects of soil contaminants in a variety of case studies (see Spurgeon *et al*., 2002 for examples).

Performance

Tests for nitrogen mineralisation in this study yielded highly variable data. No correlation between levels of chemical contamination and nitrate or CO_2 production could be discerned at Site A. At Site B nitrate (NO_3) production showed no correlation with hydrocarbon contamination. Our experience with carbon mineralisation mirrors this situation.

Practicalities

Tests for C and N mineralisation are documented in standardised (ISO) guidelines and are widely used. Though the tests can be carried out by experienced operators without undue complications, the concerns about their performance override any practical facility.

Responsiveness

The sensitivity of the tests to different contaminants is not well documented. As tests with a functional endpoint, it has been suggested that the test may not be as sensitive as bioassays with single species (Spurgeon *et al*., 2002).

Robustness

Soils are heterogeneous: they vary spatially in a number of chemical, physical, and biological properties. Measurements of carbon and nitrogen mineralisation rates, determined in isolation of other data, tend therefore to show large degrees of spatial heterogeneity, reflecting their natural variability. A further possible explanation of this variability, and also of the lack of correlation between measurements of N mineralisation and concentrations of soil contaminants in this study, is the recommended depth of soil-sampling for such mineralisation measurements. This is typically 0 to 15 cm. Mobile nutrients such as NO_3 may accumulate below this depth, rendering the near-surface values useless.

Issues of sample heterogeneity apply equally to other test methods, and so the greater variability seen with C and N mineralisation tests appears to be a particular feature of these tests. It raises concerns about the tests' sensitivity to contaminants, and also about their reproducibility and repeatability.

Relevance

A microbial measurement within the bioassay suite is useful to address important functional endpoints of nutrient cycling which directly impact on soil fertility.

Proposed use in ERA

It is important that a microbial parameter be included in the ERA process. But our experience shows that further research is needed in order to develop more robust and reliable test methodologies that deal with the level of variability associated with both these tests. Priority for such investigation should be given to the N mineralisation test, because a direct measure remains a useful indicator of a key functional process. If this could be achieved, the test would probably be deployed at later tiers of the ERA framework (Tier 2 and beyond).

7.2.3 Plant tests

Introduction

Plant tests are important within a battery of tests for assessing soil quality. They fulfil a key ecological role and, in a commercial sense, relate directly to soil fertility and productivity. A standard OECD test guideline (OECD 208) is available. It was originally designed for testing chemicals added to soil, but it has also been used in contaminated site assessment (for example, Linder *et al*., 1990; Van Gestel *et al*., 2001). The test may be performed with a range of species, including both monocotyledonous and dicotyledonous species.

Performance

At Site A, responses in terms of dry weight (and fresh weight) gave rise to plausible differences between locations, varying according to the level of metal contamination. This relationship between metal burden and response was clearer with the plant tests than with any other biological test. The most sensitive species were tomato and cabbage. These species also yielded the most useful information in terms of the relationship between the magnitude of

responses and level of contamination. In other words, greater growth inhibition was seen in the more contaminated soils.

The same tests at Site B generated information that was less useful. The relationship between biological response and contaminant levels (as indicated by measured PAH concentrations) was more erratic and showed much less discrimination between patches (Figure 6.5). It is possible, though, that contaminants other than PAHs may have dominated the toxic responses exhibited by the plants in tests made on soil samples from this site.

Cabbage was the most sensitive species, and pea was consistently the least sensitive species. This may be due to the fact that, as a legume, pea has certain biochemical pathways that are not present in other plants (for example, the capacity for β-oxidation is the basis for legume tolerance to certain auxin herbicides). Individual plants are also larger than the other test species; plants in pots where emergence is low may benefit from less competition and grow larger, thereby skewing the effect measures (dry or wet weight per plant). These factors would suggest that pea is less useful as a test species than oat or wheat.

Plant emergence was not a suitable endpoint with which to assess effects, because it was similar over all contaminant levels. This is most likely because the seeds contain all the nutrients required for emergence, so plants are not exposed to external contaminants until their roots are formed and contribute to nutrient uptake. We found that foliage wet weight and dry weight were better indicators of effect. This is shown by the results of tests with Site A soils and those in Appendix B. We felt, though, that dry weight provided the better measure of effect, as it gave a more accurate measure of biomass.

The effect data were not normally distributed, so it was difficult to satisfy the criteria for *post-hoc* statistical analysis of results.

Practicalities

The test guideline is simple to perform, requiring only basic botanical appreciation. The guideline does not stipulate operating temperatures. Since all comparisons are done using an internal control, valid data may nevertheless be generated under different conditions. Absolute comparisons of responses (such as dry weight) cannot, though, be made between experiments carried out in different laboratories, or even in the same laboratory at different times of year. In principle, the test could be carried out *in situ* as well as under controlled conditions.

In this study, we adapated the OECD method to use smaller pots. Statistical analysis of the effects on plant biomass of using smaller pots indicated that there was no negative effect on plant growth of using smaller pots, providing the seeding density was altered accordingly.

Compliance with test validity criteria

The validity criteria for the OECD plant test are based on control emergence (>65% emergence required) and visual condition of the emerged plants. The plant experiments carried out for this project employed two types of control: a test control, consisting of Kettering loam, and site-specific controls consisting of two uncontaminated soils from the test sites.

The Kettering loam controls for both sites satisfied the validity criteria, as did the site controls at Site A. Some of the site-specific controls at Site B, though, failed the validity criteria. Pea and tomato emergence was below 65 per cent in both control soils at Site B. Cabbage emergence was below 65 per cent at patch A8. In the case of pea, fungal infection undoubtedly played a part, leading to seed death even before germination. Though there was no evidence of fungal infection in tomatoes, this was the smallest seeded species and took longest to germinate. It is possible that tomato was consequently more sensitive to poor physical structure of soil samples (for example, large air voids) or to periods of water stress. Technically, this would invalidate the results obtained at these particular sites, though the reliance on-site-specific controls is open to debate (see Section 8).

Data interpretation

For four of the five species tested at Site A, there was no significant difference between the test control (Kettering loam) and the site controls. A dose response was evident, and the effects at the most contaminated sites were significantly different from those at less contaminated sites. This indicates that differences in soil structure were not having a negative effect on plant growth and could be discounted as a significant interfering factor. This was not, though, the case at Site B. There, no dose response was evident. Site-specific controls were significantly different from the test control, indicating that soil structure or nutrient status was having an effect on plant biomass. This means that we cannot necessarily distinguish between growth inhibition due to the presence of contaminants (toxicity), and physical or nutritional effects at this site.

One possible solution would be to acquire prior knowledge of the effects of soil structure and

conditions. This could entail generating a database on the effects of different soil types on plant growth in the absence of contaminants, thereby providing a baseline with which to establish background effects of soil type. A further option may therefore be to add nutrients to all contaminated test soils, in order to minimise the effects of this potential interference.

Responsiveness

Compared to investigations of test with macro-invertebrates, the performance characteristics of plant tests are only poorly understood. We would expect them to be sensitive to persistent herbicides, but there is little information on the spectrum and levels of industrial contaminants that can be detected. Spurgeon *et al*. (2002) review examples of the application of these tests to contaminated land. Plant tests have been recommended as a component in a battery of ecotoxicological tests by other jurisdictions (for example, CCME, 1996).

Robustness

The OECD Test Guideline gives considerable latitude in its operation. Experimenter judgement is required to run plant tests successfully. This is particularly evident in the lack of information on experimental conditions. As a result, there may be questions of reproducibility between plant tests performed by different laboratories. Without inter-laboratory testing of reference materials, the levels of repeatability and reproducibility remain unknown. Nevertheless, the conduct of the plant tests is conceptually simple. Measurements of wet and dry weights and plant emergence were also easy to make and interpret. In this study, we successfully addressed some procedural issues (pot size, seeding density) to make the test more suitable for contaminated land assessment.

As noted above, plant growth is sensitive to the nutrient status of soil. Differences in biomass can arise simply as a result of differences in nutrient status of soil samples, rather than toxicity. One way of overcoming this potential interference may be to supplement all soil samples with nutrients, thereby removing nutrient limitation as a factor.

Relevance

Plant growth is clearly essential to normal ecosystem functioning and, in this respect, the plant test is highly relevant. The flexibility to select test species that represent those present at a site under investigation (either the species itself or a close relative) is a particularly appealing aspect of the test, because it allows the gap between measurement and assessment endpoints to be narrowed, thereby reducing uncertainty in extrapolation.

Proposed use in ERA

We propose that higher plant tests be used in the ERA framework at Tier 2 and beyond.

7.2.4 Acute earthworm tests

Introduction

OECD (2000c) and ISO (1998) test guidelines are available for assessing chemical impacts on the survival, growth and reproduction of earthworms. They have been used mainly for assessing potential impacts of plant protection products as part of the registration process. In this study, our investigations were confined to assessments on survival. Reproductive effects (cocoon production) have been investigated at the same sites by Spurgeon *et al.* (in press).

Performance

Compliance with test validity criteria

At both sites, the test met the published validity criteria for control survival (>90%), though we should point out that the test period was 14 days, whereas this validity criterion actually applies to survival after four weeks. Strictly, we do not know whether this criterion would have been attained. The test appears to adapt easily to the use of field soils.

Sensitivity

The survival and growth endpoints were of low sensitivity. At Site A, the test identified only the most contaminated area that was dominated by a single compound (zinc), where several SQGVs were exceeded. At Site B, there was no significant effect of any of the soils, and no overall trend with PAH or total hydrocarbon concentration. Only in a separate study using soil samples that had been manipulated to achieve a range of PAH and TPH concentrations was it possible to discern any biological effects.

At both sites, only small changes in body weight were evident. Responses were unclear and did not correlate well with contaminant concentrations. Based on the evidence generated at sites A and B, this test could be used only to confirm toxicity at highly contaminated sites. The test would not be sensitive enough to identify hotspots of contamination or to monitor remediation.

We identified a possible loss of light fractions of petroleum hydrocarbons at Site B during soil

collection and preparation (see Appendix D). Furthermore, no food is added to the experimental units during the exposure period, which may cause additional stresses to the animals, though clearly not to an extent that would compromise the validity of the test after 14 days.

Practicalities

The survival test is simple to perform. A variant of the test is available for in situ testing (Hankard *et al.,* pers. comm.). Comments on the reproductive aspects of the test are covered by Spurgeon *et al.* (in press).

Responsiveness

Evidence from other researchers (e.g., Heimbach, 1998; Spurgeon *et al*., 2000; Van Gestel *et al*., 2001, Spurgeon *et al*., in press) suggests the Draft OECD Earthworm Reproduction Test generates more useful data than the acute test. The longer exposure period contributes to greater sensitivity, and the reproduction endpoint is arguably more demographically relevant than survival alone. We therefore recommend adopting the reproduction test in place of the acute test, but to monitor survival at the 14-day period. If effects are seen at this point, the test may be terminated. If no effects are reported at this stage, the test should continue, and effects on reproduction should be monitored at 28 days.

Robustness

Survivorship is influenced by soil conditions such as pH and organic matter content, and by the presence of shards that can injure test animals. In other respects, the test may be regarded as robust.

Relevance

Earthworms fulfil an important ecological function in physically cycling material in the soil and contributing to breakdown of organic matter. There are suggestions that tests using the reproduction endpoint correlate with changes in population size in the field (Spurgeon and Hopkin, 1999). There is also some understanding of the role played by earthworms in food-chain transfer of chemical residues to higher trophic levels (Spurgeon and Hopkin, 1996).

Proposed use in ERA

The acute earthworm test is recommended for use in Tier 2 and beyond, though greater sensitivity is expected from use of the reproductive endpoint (Spurgeon *et al*., in press). For this reason, the earthworm reproduction test may be a preferable option.

7.2.5 Springtail adult survival and reproduction tests

Introduction

An ISO (1998) guideline describes a method based on survival and reproduction of the springtail, *Folsomia candida*. It is described in detail by Spurgeon *et al*., (2002).

Performance

The test did not meet the ISO validity criteria at either site. This was due to high adult mortality and low reproduction. Consequently, no useful information was generated for either the adult survival or reproduction endpoints.

In all experimental units, large amounts of fungal growth were produced during the 28-day period of the test (Fig 7.1). We believe this to be due to the non-sterile field soils being left undisturbed under optimum conditions for this period of time. Fungal growth could not be removed without loss of individuals. This will have two possible influences on the data generated: first, the growth will out-compete the Collembolan; second, the amount of fungal biomass obscures the adult and juveniles during counting.

These fungal hyphae covered the whole surface of the soil, preventing the springtails from moving at the soil-air interface. Blanketing of soil by fungi during toxicity tests has been observed previously by Fountain (pers. comm.), who initially found that fungal growth inhibited springtail survival and reproduction in field collected soils. By modifying the standard protocol, he was able later successfully to complete the springtail test in soils from two sites

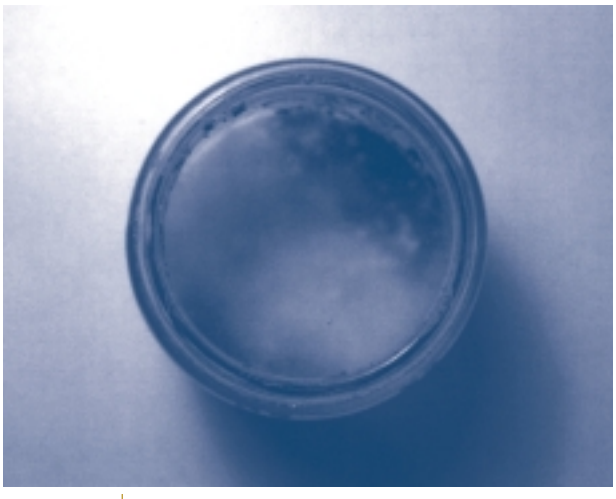


Figure 7.1 Fungal growth in Collembolan exposure jar at the end of exposure

(including Site A), using oven-drying and long-term freezing to sterilise the soils prior to testing (Fountain, and Hopkin, 2001). We therefore expect that a similar suitable modification to the standard protocol would allow completion of the springtail tests within the test validation criteria.

Practicalities

For the reasons outlined above, the published method would need to be adapted along the lines recommended by Fountain and Hopkin (2001). Methods of sterilisation techniques either before exposure (but which do not cause loss of volatiles) or at the end of exposure (that do not destroy carcasses) should be investigated. It would also be prudent to check that such interventions did not alter the speciation or bioavailability of contaminants (at least through Microtox™ testing) before and after freeze-drying. Alternative extraction techniques such as the use of Tullgren funnels may usefully replace the ISO stain and rinse technique.

Responsiveness

There is comparatively little existing data on sensitivity to single chemicals. Some studies do, though, describe the use of the test for assessing contaminated soils (Spurgeon *et al*., 2002).

Robustness

Our experience suggests that, when used with field-collected soils, the test is prone to contamination by fungi. Spurgeon *et al*., (2002) also highlight sensitivity to pH and organic matter content that means control sites must be selected carefully to aid interpretation of observed effects.

Relevance

Collembolans occur widely in natural soils, and the test may be used to assess ecologically relevant endpoints that directly influence increases in population size (numbers of individuals).

Proposed use in ERA

There is previous experience of using this test successfully in contaminated land applications (see Spurgeon *et al*., 2002 for examples). The measures outlined above should allow the Collembola tests to be performed within the test validity criteria. This test is therefore recommended for use in the framework from Tier 2 and beyond, but not until this investigation has been done and the modifications tested, ideally through inter-laboratory ring-testing.

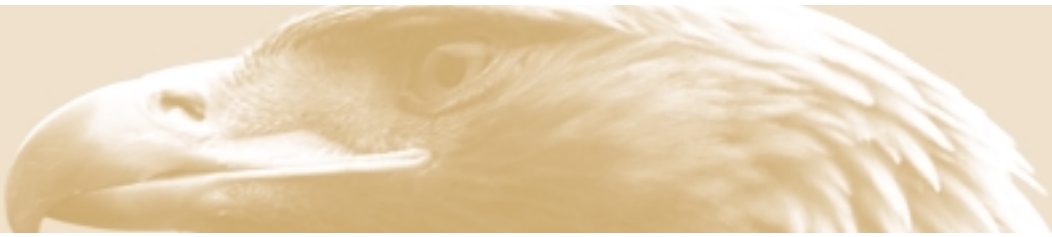
7.3 Summary

Based on the preceding reviews, we offer the following summary of the suitability of biological tests for use in the ERA framework.

Table 7.1 Recommendations for biological tests investigated in this study

Test method	Proposed point of application	Comments
Microtox™	Tier 1	Recommended
N mineralisation	Possibly Tier 2	Requires modification prior to inclusion
C mineralisation	No further use	There is no evidence for a useful role in the ERA framework
Plant tests	Tier 2 and beyond	Recommended
Earthworm tests	Possibly Tier 2 and beyond	A longer term exposure period and reproduction endpoint should be more sensitive than the acute/survival test investigated here
Collembola tests	Possibly Tier 2 and beyond	If the practical problems can be overcome and the test validated, this test could be useful in the ERA framework

8



Decision-making in the ERA framework

Decision-making criteria for progression between Tier 1 and 2 and beyond Tier 2 are outlined in Figure 8.1. The key point is that progression to Tier 2 is not automatic, but is confined to situations where a potential risk is identified and needs further investigation to demonstrate either that (a) it does not apply or (b) to confirm there is a risk of significant harm.

8.1 Introduction

For instances where Tier 2 assessment is needed, we turn our attention to the way in which decisions about whether there is a risk of significant harm can be made, using a weight-of-evidence approach as advanced in Section 4. This takes into account information generated by the battery of biological tests, as well as information arising from ecological survey and chemical data. This meets objective 3 of the project (Section 1.5).

Before we address decision-making in the ERA framework, we must first be clear about what we mean by *significant harm*. In Section 8.2, we offer an approach to defining an adverse biological effect. While this is open to challenge from theoretical viewpoints, it offers a pragmatic basis for decision-making in the context of the ERA framework and could be extended to decision-making in other regulatory regimes. In Section 8.3, we used data generated from studies at Sites A and B to show how decisions would be made about the progress of these sites through the ERA framework, that is, integrating information from all the patches that were analysed or tested. In Section 8.4, for the purposes of illustration, we also treat individual patches as separate sites. This allows us to show how different conclusions would be reached if sites had shown the characteristics of individual patches.

8.2 What are the criteria for *significant harm?*

8.2.1 Definitions of harm

In common with many other standards, SQGVs are extrapolated from laboratory or field data to define a concentration that should give rise to no adverse effects in the field. In many circumstances, they will therefore provide a precautionary measure of harm because of the varying sensitivities of organisms present in the ecosystem under study. Conventional chemical analyses do not measure the concentration of a substance that is actually bioavailable or in a toxicologically active form. Taken together, these factors mean that exceedance of the SQGV does not necessarily mean that biological impacts will result. It is against this background that direct biological testing has an important role to play, because it will respond only to the bioavailable fraction. Biological testing is therefore of direct relevance to the principle of *significant harm*, as stated in the Part IIA Regulations. Table 4.1 gives measures of significance for biological testing for each of the biological tests undertaken in this project. These can be used to indicate harm to the test organisms or function of concern in that specific test. By inference, if the test outcome is positive, that is, there is harm to test organisms (typically, death) during or at the end of

the test duration, then this demonstrates harm resulting from the soil contamination.

Definitions of harm are inevitably linked to the protection objectives of ERA. There is a continuing scientific and philosophical debate about what the protection goals of ERA should be. The statutory guidance for Part IIA of the Environmental Protection Act 1990 is concerned with situations where there is a significant possibility of significant harm. The Environment Agency proposes that significant harm is when growth, reproduction, or mortality are adversely affected, such that the survival of the population/community/species is threatened. There is a significant possibility of significant harm when the indicators of significant harm (outlined above) differ from reference or control values at an agreed statistical confidence level.

When interpreting ecotoxicological data obtained from environmental samples, including soils, the only practical approach is to compare mean responses in a treatment group (a soil that is suspected of contamination) with those in an uncontaminated control soil. Significant harm may be concluded when a statistically significant difference is evident, based on hypothesis testing, as described in the mean value test in CLR7 (Defra and Environment Agency, 2002). Normally, a *post hoc* test of significance would be performed at the 95 per cent level of confidence. This is a level conventionally used in science. Different levels of confidence may, though, be appropriate, depending on other factors such as cost and benefit. For example where a rare species is at risk, you might increase the level of protection by lowering the significance level to 90 per cent. This is a risk-management decision. Hypothesis testing in biological testing is not without limitations though, and we outline these below.

In addition, the risk assessor might consider questions of the biological significance of statistically significant differences. By this we mean whether a difference that is detected statistically is likely to be of biological importance. In practice, where levels of replication are low and background variability is high (Section 8.2.2), it is reasonable to suppose that any statistically significant difference will also be of biological importance.

8.2.2 Power of analysis in hypothesis testing

The ability to resolve differences between groups obviously depends on the size of any difference between means, but also on the variance (differences between replicates) within each treatment group. If variance is high, larger differences are required before two samples can be declared different.

It is well known that the level of variance can be managed in the design and execution of toxicity tests. Indeed, this is now regarded as a flaw in the use of NOECs (that are estimated in the same way) for regulatory purposes (Crane and Newman, 2000). When comparing samples taken from the field, reliance on hypothesis testing is probably unavoidable. Statistical advice is therefore helpful, to ensure that adequate power is retained in these analyses, in practice through determining the level of replication required and through the choice of *post hoc* tests of significance.

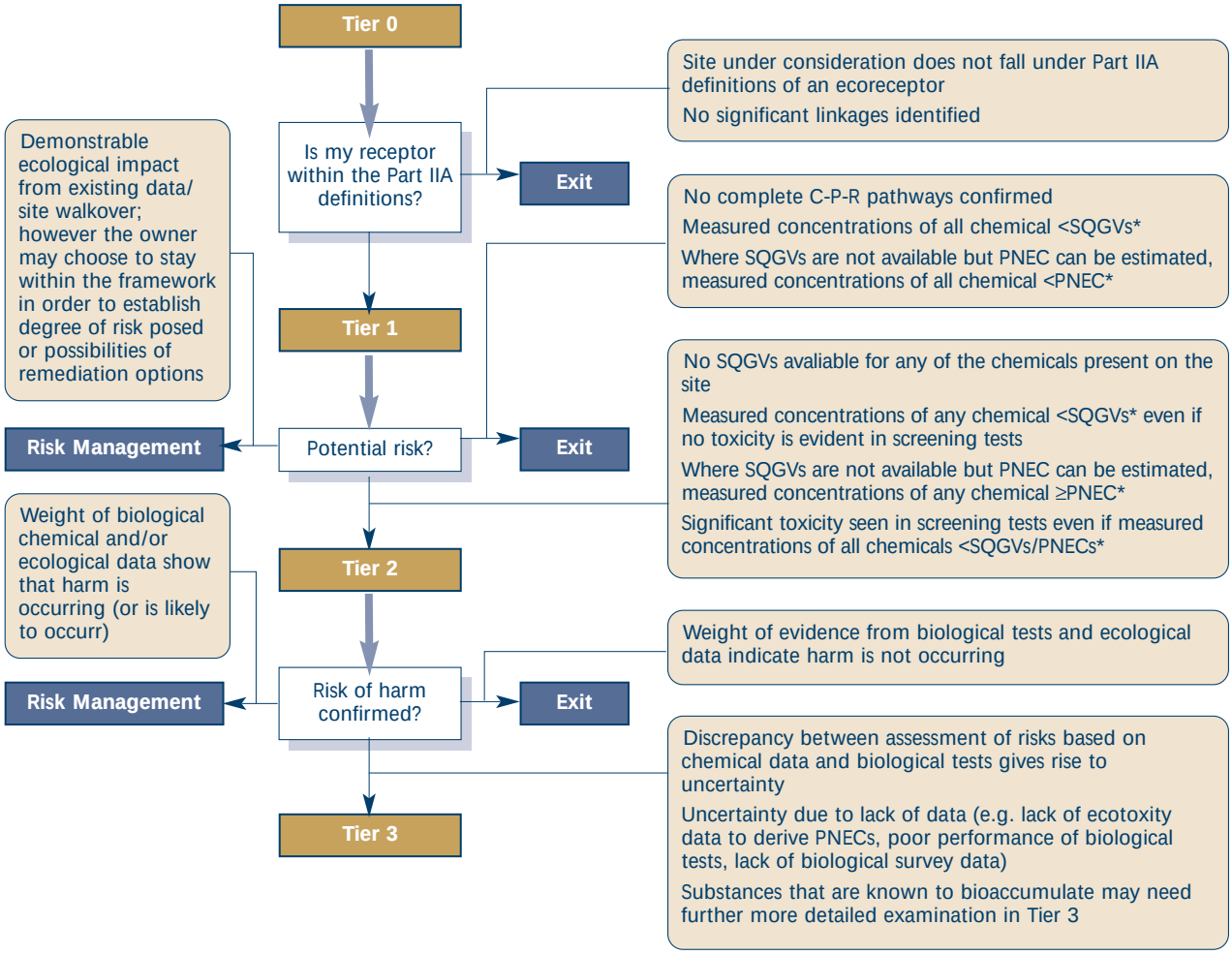
The skill of the experimenter can also affect within-treatment variance (inexperienced operators tend to give rise to higher variance). Consideration should therefore also be given to the level of experience expected of laboratories performing these tests. Adoption of formal QA/QC procedures such as Good Laboratory Practice (GLP) or United Kingdom Accreditation Service (UKAS) accreditation can be helpful in this regard.

8.2.3 Choice of controls

In Section 3.2, we raised the issue of positive and negative controls, and in particular whether to base comparisons on (a) negative controls obtained from a pristine but remote location, or (b) on negative controls from a site close to the study site that is known not to be contaminated.

Evidence from plant tests where both were used shows that any differences can be insignificant (for example, tomato, oat, wheat and pea at Site A). But in other cases, statistically significant differences between these two sets of controls can occur (for example, cabbage at Site A and all species at Site B). Such differences can have a major impact on whether we declare there to be an adverse effect or not with a particular soil sample. In the case of Site B, we cannot be confident that differences in plant biomass between test samples and controls are a result of toxic contaminants or differences in soil nutrient/physical status. This issue applies particularly to plant and Microtox™ tests, where a growth-related endpoint is used and where nutrients can play an important role in apparently compensating for toxic effects. This problem can be overcome by ensuring the test medium is not nutrient limited. Because we didn't compare controls in the other tests, though, we cannot determine the importance of choice of controls in these tests.

From a practical viewpoint, it is much simpler to adopt approach (a). A series of different controls (including both (a) and (b)) can, though, be helpful



* The upper 95% confidence interval around the mean concentration is compared with the SQGV or PNEC. This has the effect of 'rewarding' extensive surveys of chemical contamination because uncertainty is diminished and the upper 95% confidence limit is consequently lower than if the comparison is based on only a few samples (CLR 7)

Figure 8.1 Decision making in the ERA framework

in understanding the background condition. An appropriate volume of control field soil with similar physico-chemical properties to the contaminated soil should be collected (Saterbak *et al*., 1999). If higher tier assessment takes place, a standard substrate should also be used, to afford an assessment of quality control for the bioassays (Van Gestel *et al.*, 2001).

8.2.4 Test sensitivity

Relationship between sensitivity and decision-making

Biological tests at Tier 2 need to be sensitive enough to minimise the risks of false negatives (that is, failing to detect a real effect when chemical residues occur at toxicologically significant levels). Currently, our understanding of the sensitivity of most of these tests to all but a few reference toxicants is poor. At the very least, the sensitivity of selected biological tests to common contaminants at concentrations corresponding to their SQGVs needs to be better understood. This may be based in part on literature data, but it could also require *de novo* investigation as well.

Chronic and sublethal tests

As suggested by Spurgeon *et al*., (in press), the reliance placed on acute tests with lethal test endpoints means that they may not be as sensitive as chronic tests with sublethal endpoints.

Numerous analyses of aquatic toxicity data have shown that acute tests are generally less sensitive than chronic tests using the same species, and even the same endpoint. Sublethal endpoints

(reproduction and growth, for example) also tend to be more sensitive than the survival endpoint over equivalent exposure periods, as shown by studies with earthworms and zinc (Spurgeon *et. al.,* 2000), springtails and cadmium (Crommentuijn *et al*., 1997) and copper, pyrene and chlorpyriphos (Herbert *et al*., 2004). In the future development of methods for use at Tier 2, these observations would encourage the use of longer-term tests with sublethal endpoints. Practical issues of costs, response time and physical or chemical changes in soil during testing would also, though, need to be addressed.

8.3 Role of toxicity screening at Tier 1

In Section 2.3.2, we explain that Microtox™ testing is used at Tier 1 to reduce the risk of failing to detect contaminated soils that are not flagged as potential risks by chemical risk assessment alone. This can arise when:

- SQGVs are unavailable;
- The analytical suite does not cover substances for which SQGVs are available;
- Substances interact in a way to generate higher-than-expected toxicity.

A characteristic of the Microtox™ results at Site A was the 'all-or-nothing' responses seen (Figure 5.4). While this diminishes the usefulness of the test for ranking purposes, it simplifies interpretation into a categorical (pass/fail) assessment of soils. In terms of the ability of Microtox™ to discriminate contaminated sites, it was clearly one of the more informative assays, because it yielded responses where other assays did not (that is, it was more sensitive). This has to be balanced against the ecological relevance of decisions for soil quality made on the basis of responses by the test organism *Vibrio fisherii*, a marine bacterium.

Without further examples, we cannot be sure to what extent the incidence of false negatives would be reduced by incorporating Microtox™ into Tier 1. It would seem to be a practical means, though, of ensuring that Tier 1 remains valid for all sites, thereby potentially reducing the need for additional chemical sampling.

8.4 Applying a weight-of-evidence approach to Sites A and B

8.4.1 Introduction

As we explained in Section 4, the information generated from a range of measurement/assessment techniques may be integrated through the use of a Weight of Evidence (WOE) approach. The methodology requires a number of questions to be addressed, each contributing to an overall conclusion.

Chapman *et al*., (2002) discuss a WOE approach in relation to sediments, but the issues are equally applicable to soils. Essentially, they challenge practitioners to answer a series of questions that are used in combination to form an overall view about the likely risks to ecoreceptors. Clearly, evidence for most or all of these issues would lead to a more compelling conclusion about a risk of significant harm than would positive responses to just one or two:

1. Are contaminants present at levels of concern (based on chemistry data at Tier 1 or the CSM)?
2. Are the contaminants capable of causing harm (based on laboratory toxicity testing at Tier 2)?
3. Are resident biotic communities adversely affected (community structure analysis at Tier 1 or 2)?
4. Is there evidence for the contaminants causing the observed toxicity and/or community alterations (Tier 2 and investigations at Tier 3)?
5. Are any contaminants of concern likely to biomagnify (food-chain modelling and tissue analysis at Tier 3, but potential risk should be flagged at Tier 0)?

It may also be possible to assign a rank to measured effects or observations, rating each measurement endpoint as indicating high, moderate, or negligible/low ecological risk. Alternatively, some scale may be applied to the measured endpoint - a severity index - and such indices may be summed to provide an integration of multiple toxicological endpoints and support a decision, possibly linked to professional judgement. The combination of endpoints raises further questions, though; not least, about how many endpoints are required to demonstrate a response.

There is a further question about the weight accorded to different endpoints. To some extent, this is addressed under the debate about significant harm (Section 8.2.1). When faced with different lines of evidence, though, the risk assessor should relate available information to the stated assessment endpoints. The effect of this might be to accord greater significance to studies that estimate effects at the population level (estimates of the intrinsic rate of population increase(λ) or risk of extinction, for example) than individual traits such as survival or reproduction. Burton *et al.,* (2002) argue that the weighting or integration of multiple lines of evidence

into one summary does not remove the uncertainty associated with that evidence, but it does provide a mechanism for making the best use of all available scientific information. They suggested that WOE approaches provide the most useful information for decision-making.

Below, we adopt a WOE approach advocated by Suter *et al.,* (2000) to assess the combined datasets obtained from the chemical and biological tests data at sites A and B. The final conclusion is based not only on the frequency of + or - signs. It is also based on the reliability of the conclusions drawn from the various lines of evidence (that is, chemical analyses, ecological surveys or biological tests). This still leaves the final decision to a process of expert judgement, though it attempts to make the reasoning transparent.

Tables 8.1 and 8.2 illustrate this type of WOE

Table 8.1 Application of weight-of-evidence approach to interpreting data generated at Site A following Suter *et al.*, 2000

Test option (evidence)	Result (test outcome)[a]	Explanation
Tier 0	+	Using the contaminant pathway receptor conceptual model, an eco-receptor (SSSI) has been shown to be at potential risk from the aerial deposition of heavy metals. The management goal of this SSSI is to protect features and sub-features of interest and maintain or achieve favourable condition within the designated ecoreceptor with no adverse effect on its integrity. The maintenance of soil ecosystem function is seen as a priority in order to maintain the naturalness of the habitat on which rare birds of prey populations are resident. The conservation objective of the SSSI is the continued maintenance or enhancement of population abundance and assemblage structure of features of interest, and of structure or function of supporting sub-features. Appropriate measurement endpoints need to be utilised at Tier 2 as surrogate measures of ecosystem function. (see Section 5.4.3)
Soil analyses/ single chemical tests (Tier 1)	+	A gradient of heavy metal contamination including Cu, Cd, Pb, Zn, As and Hg has been determined from a single point source. Soil total concentrations of heavy metals within a few kilometres of the point source are exceptionally high. Comparisons with literature data and the three sets of SQGVs demonstrate clear exceedances, especially in soils closest to the source. Based on the numbers of SQGVs exceeded, patches A3, A4 and A5 are the most heavily contaminated, with at least seven contaminants exceeding SQGVs. Based on existing toxicity data, many soil-dwelling organisms would not be expected to survive at locations in close proximity to the site.
Microtox™ (Tier 1)	+	Microtox™ data generated for water elutriates of soils taken from the 10 patches along the Site A gradient indicated high levels of inhibition at sites A3 - A7 (all within 3.3 km of the known point source). The other patches: A1, A2, A8, A9 and A10, were indistinguishable from control reference patches and external controls. This observed trend in toxicity corresponds to the observed heavy metal concentration gradient prevailing at the site.
Plant tests (Tier 2)	+	In general, the plant species were sensitive to the presence of heavy metals at the site, though a clear gradient could not be demonstrated with most species. Two species though (cabbage and tomato) showed clear, dose response effects.

Table 8.1 (cont) Application of weight-of-evidence approach to interpreting data generated at Site A following Suter *et al.*, 2000

Test option (evidence)	Result (test outcome)[a]	Explanation
Earthworm acute toxicity test (Tier 2)	+	Earthworm survival was reduced at only one patch (A5) for the worm species *Eisenia fetida*. This patch is in close proximity to the single point source and is consistent with both the chemistry data and other biological results. Sensitivity was low, though, since earthworm survival was not significantly different from relative control data at other patches, despite exceedances of SQGVs and biological responses using other tests.
Collembolan reproduction test (Tier 2)	+/-	The data from this soil-dwelling microarthropod reproduction test were inconclusive, due to poor test performance and the failure to meet test validity and acceptance criteria. No correlation or causation can be usefully inferred from the data.
Carbon Mineralisation (Tier 2)	+/-	The data from this assay were highly variable and the results inconclusive
Nitrogen Mineralisation (Tier 2)	+/-	The data from this assay were highly variable and the results inconclusive.
Final decision **Go on to Tier three, as there is still uncertainty and possibility of secondary poisoning.**	+	Clear exceedences of SQGVs by several of the heavy metals coincide with ecological effects observed in previous studies conducted at the site. Metal residues are evidently bioavailable as indicated by toxicity of the patches in plant and Microtox™ tests in a manner consistent with metal residues. It is reasonable to conclude that the presence and bioavailability of metal residues are sufficient to give rise to adverse biological effects at patches A3, A4, A5, A6 and A7. This is at odds with our protection goal identified at the Tier 0 level of protecting the assessment endpoint. The surrogate measurement endpoints (biological testing at Tier 2) have been useful in determining the significance of these impacts.

[a] Results of the risk characterization for each line of evidence and for the weight of evidence approach
\+ indicates that the evidence is compelling and consistent with a significant biological effect (according to defined test criteria)
\- indicates that the evidence is inconsistent with the occurrence of a significant biological effect;
+/- indicates that the evidence is too ambiguous to interpret

Table 8.2 | Application of weight-of-evidence approach to interpreting data generated at Site B following Suter *et al.*, 2000

Test option (evidence)	Result (test outcome)[a]	Explanation
Tier 0	+	Using the contaminant pathway receptor conceptual model, an eco-receptor (SSSI) has been shown to be at potential risk from the horizontal movement of hydrocarbons from a site with hydrocarbon contamination. The management goal of this SSSI is to protect features and sub-features of interest and maintain or achieve favorable condition within the designated ecoreceptor with no adverse effect on its integrity. The maintenance of soil ecosystem function is seen as a priority in order to maintain the integrity of the habitat on which rare birds of prey populations are resident. The conservation objective of the SSSI is the continued maintenance or enhancement of population abundance and assemblage structure of features of interest, and of structure or function of supporting sub-features. Appropriate measurement endpoints need to be utilised at Tier 2 as surrogate measures of ecosystem function.
Soil analyses/ single chemical tests (Tier 1)	+	Historical data from the site owners suggested the presence of petroleum-related products as a result of the prior activities on-site. A further, targeted sampling campaign showed that most concentrations (based on the use of TPH as a measure of oil contamination) were very patchy with two localised hotspots of 14,300 and 34,400µg TPH g^{-1}, respectively. There are no international SQGVs for TPH. When total PAH concentrations (based on the sum of concentrations of 54 individual congeners) were measured, a clear exceedance of the revised Dutch List soil guideline target value was seen for all soil patches with the exception of B1. Consequently, patches B2-B9 could pose a risk to soil quality.
Microtox™ (Tier 1)	+/-	Microtox™ data generated for water elutriates of soils taken from the nine patches at Site B using the 100 per cent Toxicity Test procedure indicated no significant toxicity. It is unclear whether this is due to: (a) poor aqueous extraction of hydrocarbons from the soils due to the low water solubilities of these compounds; (b) the test being insufficiently sensitive to hydrocarbons; (c) toxicants not being present at bioavailable concentrations (d) contaminants not being toxic to this species Solid Phase Microtox™ testing facilitated direct contact between the test bacteria and available contaminants. This test resulted in low toxicities for Site B soils, except at the two patches where TPH concentrations were highest.
Plant tests (Tier 2)	+/-	A range of five plant species (including both monocotyledon and dicotyledonous species) were grown in soil collected from each of the nine patches at Site B. Plant growth was impaired compared to external controls at all patches. But a clear gradient could not be demonstrated for either PAHs or TPH. It is not possible to tell whether the observed effects were due to toxic effects or the effects of soil physical structure or low nutrient status.

Table 8.2 Application of weight-of-evidence approach to interpreting data generated at Site B following Suter *et al.*, 2000

Test option (evidence)	Result (test outcome)[a]	Explanation
Earthworm acute toxicity test (Tier 2)	-	Earthworm survival was not reduced at any patch from Site B.
Collembolan reproduction test (Tier 2)	+/-	The data from this soil-dwelling microarthropod reproduction test were inconclusive due to poor test performance and the failure to meet test validity and acceptance criteria. No correlation or causation can be usefully inferred from the data.
Carbon Mineralisation (Tier 2)	+/-	The data from this assay were highly variable and the results inconclusive.
Nitrogen Mineralisation (Tier 2)	+/-	The data from this assay were highly variable and the results inconclusive.
Final Decision	+/-	There were clear exceedences of international SQGVs for total PAH (and a measure (mineral oil) of TPH) at all patches (apart from one). The results of Microtox™ tests suggest these contaminants were bioavailable, but not to an extent that caused ecotoxicity in other biological tests (or these tests were of low sensitivity to the contaminants involved). The extracts of soils with high TPH were probably coloured due to the polar compounds extracted in the aqueous extract. The depressed luminescence in Microtox™ may have been due to quenching by this colour. In part, any biological effects may have been masked by high variability, but it is clear the biological test data do not confirm a risk of significant harm. Therefore the management protection goal identified at the Tier 0 level of protecting the assessment endpoint (ecosystem function) is met. The surrogate measurement endpoints (biological testing at Tier 2) have been useful in determining the significance of any ecological effects from the contaminant impacts. On this basis, no further work is warranted as the scale of any contamination is small. Thus in this instance, Site B would exit the ERA framework at Tier 2, with the appropriate risk determination outlined.

[a] Results of the risk characterization for each line of evidence and for the weight of evidence approach
\+ indicates that the evidence is compelling and consistent with a significant biological effect (according to defined test criteria)
\- indicates that the evidence is inconsistent with the occurrence of a significant biological effect;
+/- indicates that the evidence is too ambiguous to interpret

8.5 Decisions based on individual patches - an illustration

Taking Site A as an example, it is clear that the patches investigated yielded different levels of chemical contamination, toxicity in laboratory tests, and were subject to varying levels of ecological degradation. In Figure 8.2, we have treated each patch as an individual site to illustrate the range of decisions that may arise.

It is clear from Figure 8.2 that, if we were to treat individual patches as sites, all the patches within Site A would be progressed to Tier 2 testing, and ultimately to Tier 3. This is because of the nature of the contamination: heavy metals are persistent and potentially bioaccumulative. The advantage of adopting this approach is that suggestions for appropriate remediation strategies may be forthcoming, but it also has a greater understanding of the chemicals concerned and their impacts (if any) on the ecoreceptor. The majority of sites we envisage would be investigated at Tier 2, rarely proceeding to Tier 3. This tier is helpful in discriminating between those where a risk of harm is confirmed and those where any chemical contamination is not judged to be of biological significance. Of course, there are limitations in the performance of a number of the biological tests, and this may compromise their ability to detect adverse effects. In an evidence-based approach, though, we can make decisions only on the basis of the data that have been generated.

The evidence for impacted ecological functioning is so compelling for patches A3, A4, A5 having completed only the lower tiers of the framework that further investigation is unlikely to yield extra information that would alter this decision to invoke risk management options (obviously dependent on the level of risk anticipated). By progressing these patches through to Tier 2 and Tier 3, the exact nature of the contamination could be determined, and appropriate remediation steps taken or processes used.

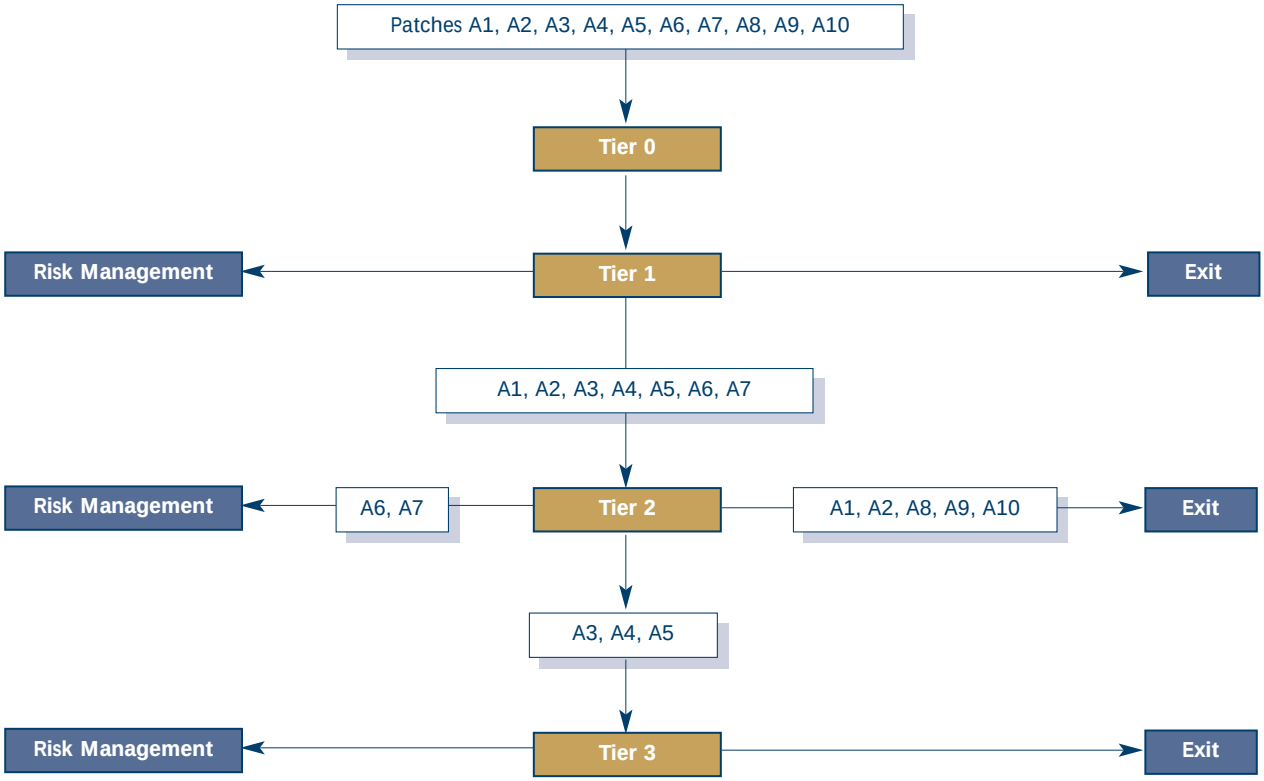


Figure 8.2 Decisions based on data for individual patches at Site A

Conclusions and recommendations

Figure 8.1 summarises the decision-making process in terms of the criteria for moving between tiers, for reaching decisions about the need for risk management or for exiting the framework. We propose a four-tier scheme. Each tier fulfils a different purpose, and each places different technical demands on the investigator.

9.1 Overview

The purpose of the tiered approach is to maximise the benefit gained from the information collected in each tier. As the investigator progresses through the tiers, the level of information increases. Confidence in the decisions made increases. The costs of site assessment also increase as more site-specific investigation is conducted, but the costs of risk determination (action taken) can decrease as the extent and severity of contamination present is elucidated. Progression through all the tiers is, though, by no means automatic. At each tier, the assessor decides whether to exit the framework, to consider risk management options or to refine the available evidence by progressing to the next tier.

When information about chemical residues is available, along with ecological processes and biological testing, assessors may make decisions about risks of significant harm to ecoreceptors using a weight of evidence approach that integrates these different types of information. This becomes increasingly important at higher tiers of assessment, when it is more likely that this range of data has been generated.

The following sections summarise the key points of the proposed framework.

9.2 Tier 0

The first step (Tier 0) takes the form of a conceptual site model (CSM). It reviews all known information, defines the spatial boundaries of the site under investigation and identifies potential contaminant-pathway-receptor (C-P-R) linkages for chemical contaminants.

Through a combination of reviewing chemical residue data, ecological data and biological testing, subsequent tiers investigate potential C-P-R linkages to verify whether they apply. This identifies whether a site requires risk management, further evaluation, or is deemed not to pose a significant risk of harm to ecoreceptors. Sites may exit at Tier 0 if they are found not to fall within Part IIA definitions, or if no significant linkages are identified.

Throughout the process, the CSM is used to inform the design of practical investigations. At the same time, the CSM is refined in the light of the information gained through investigation.

9.3 Tier 1

Tier 1 is essentially a screen to identify whether a site can be excluded because it is unlikely to pose a risk to ecoreceptors, or whether further investigation is needed. This decision is made by the comparison of chemical residue data with Soil Quality Guideline Values (SQGVs), in a simple deterministic risk assessment. This is supplemented with toxicity screening as a way of detecting soils containing contaminants for which SQGVs do not exist, have not been analysed for, or whose toxicity is greater than expected (synergism) when they occur in combination.

Two points about Tier 1 are worth highlighting. First, sites may be judged to require risk management

even at this early stage if there is compelling evidence of adverse impacts on local ecology. Even these sites, though, may progress to Tier 2 and Tier 3 to guide any potential risk mitigation (if warranted). This was indeed the case at a number of patches within Site A in this study. Second, the incorporation of a toxicity-screening step (for example, Microtox™) allows sites that are contaminated with complex mixtures of chemicals to be evaluated at Tier 1 rather than to progress automatically to Tier 2, as suggested by Byrns and Crane (2002). Any site that has persistent compounds present will, though, need to progress through the tiers to ensure that bioaccumulation is not occurring. This toxicity-screening step reduces the numbers of chemical analyses undertaken at a site and further refines the sampling strategy. This allows resources to be focused on those areas within a site where they are warranted. It follows that the screening test should be responsive to a wide range of substances, so that the chances of failing to detect truly toxic samples are reduced. Microtox™ has been reported to have low sensitivity to some commonly occurring contaminants (such as metals, pesticides with specific modes of action, and PAHs). This weakness should be taken into consideration when interpreting data provided by Microtox™ (Doherty, 2001).

As indicated in Figure 8.1, either a positive response in a toxicity-screening test or exceedance of SQGVs is sufficient to warrant progression to Tier 2. At present, though, the Environment Agency will not accept decisions on negative effects based solely on Microtox™ information.

9.3.1 Toxicity screening

Based on the experience gained in this project, the Microtox™ test currently fulfils the criteria for a toxicity-screening test that can be applied routinely. The potential value of this test was identified previously (Crane and Byrns, 2002; Spurgeon *et al*., 2002), and this has been borne out in this project. This does not, though, exclude the possibility that other screening tests, perhaps also based on bioluminescence in bacteria or higher organisms, may not offer advantages over Microtox™ in the future. It is also evident from the research described here that Microtox™ testing could play a role in identifying 'hotspots', thereby refining any future investigative or remediation work at a site.

Samples may be presented to the Microtox™ test as either aqueous leachates or in the solid phase. The choice of sample preparation is important, because it can affect availability of toxicants and therefore the ability of the Microtox™ test to detect contaminants at toxic levels. The limited experience gained in this project suggests that the solid phase variant of the test may be more appropriate for screening. Testing of aqueous leachates may be more suitable for polar contaminants, but it is reasonable to suppose these will not persist in historically contaminated sites, because these are likely to have dissipated through leaching. Water-soluble contaminants can survive in surficial soils if they are protected from leaching (for example, under hard standing).

9.3.2 Risk assessment using SQGVs

Clearly, assessors would accord a high priority to sites where many chemicals exceed SQGVs by large margins (or where high levels of toxicity are found) and over a wide area (especially where they coincide with key ecological receptors). Assessors would accord a lower priority to sites where exceedances were for only one or two chemicals, were small, and limited to a small area. Using an upper % C.L. of the mean (e.g. 95th) as the PEC in the chemical risk assessment helps overcome the risk of progressing sites on the basis of outliers. It is also useful to adopt a weight-of-evidence approach in making transparent the decision about whether to advance a site to Tier 2.

9.4 Tier 2

Ecotoxicity tests can demonstrate biological effects at sites that are subject to chemical contamination. They therefore have a useful role to play in demonstrating whether chemical residues found at Tier 1 are of biological significance. This is important because it relates directly to the concept of significant harm required under the Part IIA Regulations. It may not be possible to draw such conclusions on the basis of chemical data alone, because the bioavailability of chemical residues is not usually known.

We have already highlighted the merits of the Microtox™ test as a screening test at Tier 1. In addition, a number of biological tests have been highlighted as worthy of inclusion at Tier 2 (Table 7.1). Higher plant tests may usefully be adopted, with the modifications referred to in Section 7.2.3. In addition, a sublethal variant of the earthworm test is likely to yield useful information, though the acute test should not be adopted. While we consider a nutrient cycling test useful, the nitrogen mineralisation test would first need significant development to make it a practical proposition. Problems of fungal contamination in the Collembolan tests would also need to be resolved before this could be adopted routinely at Tier 2.

Again, the weight of evidence approach illustrated in Section 8 provides a useful way of integrating

biological data generated at Tier 2 with other types of data, and for making transparent the decisions drawn as a result.

9.5 Tier 3

Assessors will carry out investigations at Tier 3 when uncertainties remain either about the significance of chemical residues found at a site, or about the biological effects of residues as revealed by ecotoxicity testing. In particular, specific studies may be designed to test C-P-R linkages of particular concern (for example, that apply to rare or endangered species) or to establish the level of risk from bioaccumulative substances that cannot readily be addressed at Tier 2. For example, at Site A, effects on plant growth were clearly evident. If further understanding of the risks to defined receptors such as herbivores were required, this would be done by progression to Tier 3. Tier 3 investigations may also be carried out to understand better the size of areas that require remediation, or to help define levels of contamination that would be acceptable (remediation targets). The type of investigation carried out at Tier 3 is therefore much more site-specific than hitherto, and is strongly influenced by the CSM.

9.6 Recommendations

This report is only part of a wider investigation by the Environment Agency into assessing the potential risk posed to ecosystems from land contamination. The recommendations in this report are concerned with research and data gaps identified in the present project, but also those identified as part of the wider study. These are aimed primarily at further informing and refining the Environment Agency's risk assessment procedures.

- ***Consultation***

There is a need to test the practicability and reasonableness of the framework and the suite of suggested tests in the 'real world'. The Environment Agency is therefore seeking feedback through a formal consultation process, which has two strands. The Agency is collaborating with partners who will road-test the framework by performing ERAs on their own sites. Alongside the road-testing, the Agency has invited feedback from peer reviewers. Ideally, these will be experienced practitioners who have performed ERAs in the past. The consultation will allow the Agency to draw on the experience of those practitioners best placed to comment on both the usefulness of the framework and the appropriateness and applicability of the tests. The Agency will use the knowledge acquired through this consultation to refine and improve the framework.

The consultation allows those who use ERAs and those who will be affected by the outcomes of ERAs to comment on, and where appropriate, influence the framework's development.

- ***Further development of tests***

The performance of some of the tests in the present project and in project P5-063 (Spurgeon *et al.*, in press) with contaminated soil samples raised the need for modifications and refinement of the protocols to increase their robustness and repeatability, i.e. Collembola, and nitrogen mineralisation. In addition to specific modifications to individual tests, baseline data is required on the performance of the tests when using different soil types. This information will help risk assessors assign cause and effect when examining contaminated soils.

- ***Increasing the number of tests in the suite***

The performance of the microbial tests in the present study (carbon and nitrogen mineralisation) was disappointing. The tests proved to be inconclusive and did not therefore provide any information that was useful to the risk assessment process. There is a gap, therefore, for robust microbial measures that could be used as part of the suite. In addition, the number of tests available for inclusion in the suite with appropriate validation has increased since the original report of Crane & Byrns (2002), and their suitability needs to be determined. New developments in biological testing will in future present further tests that could usefully play a role in the ERA framework, and horizon-scanning by the Agency needs to highlight such tests.

The tests trialled in the present project and P5-063 were recommended in previous R&D projects. They were limited by the budget available for field validation to a suite of 13 tests. In practice, ERA assessors may want to use tests that are not presently part of the recommended suite, if such tests fulfil the same criteria against which the tests in the present study and P5-063 were assessed - that is, the five R's (responsive, robust, reproducible, representative and relevant).

- ***Tier 3 development***

The methodology required for Tier 3 assessments needs to be defined and included within the ERA framework. The activity within this tier is likely to include food-chain modelling, and improved assessment of the exposure route of ecoreceptors to the contaminants present (that is, measures of ingestion rates, sampling of food sources and tissue analysis.)

- ***Training***

The successful application and uptake of the proposed ERA framework depends on risk assessors, laboratory, Local Authority and regulatory staff

References cited

Aquaterra Environmental. (1998) Development of Plant Toxicity Tests for Assessment of Contaminated Soils. *Report for Method Development and Application Section, Environmental Technology Centre, Environment Canada, Gloucester, Ontario.*

Ashton D. (1998) The effects of heavy metals on the physiological energetics and reproductive biology of Oniscus asellus. *PhD Thesis University of Sheffield, Sheffield, UK.*

Bardos R.P. (2002) Report of the NICOLE Workshop: Cost-effective Site Characterisation - Dealing with uncertainties, innovation, legislation constraints, 18-19 April 2002, Land Contamination and Reclamation, 10 (3) 189-219.

Beyer W.N., Heinz, G.H, and Remond-Norwood A.W. (1996). Environmental contaminants in Wildlife, Interpreting Tissue Concentrations *(A SETAC special publication), CRC Press, New York.*

British Standards Institute (2001). BS 10175 Investigation of potentially Contaminated site Code of Practice. *BSI, London.*

British Standards Institute (2002). Draft British Standard BS ISO 10381-5 Soil Quality - Sampling Part 5 Guidance on investigation of soil contamination of urban and industrial sites. *BSI, London.*

Byrns G., and Crane, M. (2002). Assessing risks to ecosystems from land contamination. *R&D Technical Report P299. Environment Agency and SNIFFER, Bristol, UK.*

Burton G.A., Chapman, P.M., and Smith E.P. (2002). Weight of evidence approaches for assessing ecosystem impairment. *Human and Ecological Risk Assessment Vol. 8: 1657-73.*

Canadian Council of Ministers of the Environment (CCME) (1996). An assessment of the risk assessment paradigm for ecological risk assessment. *Commission on risk assessment and risk management Prepared by Menzie Cura & Associates. CCME Documents, c/o Manitoba Statutory Publications, Manitoba, Canada.*

Canadian Council of Ministers of the Environment (CCME) (1997). Recommended Canadian Soil Quality Guidelines. *CCME Documents, c/o Manitoba Statutory Publications, Manitoba, Canada.*

Carter J.A., Mroz, R.E., Tay, K.L., and Doe, K.G. (1998). An evaluation of the use of soil and sediment bioassays in the assessment of three contaminated sites in Atlantic Canada. *Water Quality Research Journal of Canada, 33(2): 295-317.*

Chapman P.M., McDonald, B.G., and G.S. Lawrence (2002). Weight-of-evidence issues and frameworks for sediment quality (and other) assessments. *Human and Ecological Risk Assessment 8 (7): 1489-1515.*

CIRIA (2002). Biological methods for assessment and remediation of contaminated land - case studies. *Ciria research project 625. CIRIA, London.*

Coglan A., Hankard, P. K., Spurgeon, D.J., Svendsen, C., Wadsworth, R., and Weeks, J.M. (2003) Closing the loop: A spatial analysis to link observed environmental damage to predicted heavy metal emissions. *Environmental Toxicology and Chemistry, vol. 22, (5): 970-76.*

Crane M., and Byrns, G. (2002). Review of ecotoxicological and biological test methods for the assessment of contaminated land. *R&D Technical Report P300. Environment Agency, Bristol, UK.*

Crane M., and Newman, M.C. (2000). What level of effect is a no observed effect? *Environmental Toxicology and Chemistry 19: 516-19.*

Crommentuijn T., Doornekamp, A., and van Gestel, C.A.M. (1997). Bioavailability and ecological effects of cadmium on Folsomia candida (Willem) in an artificial soil substrate as influenced by pH and organic matter. *Applied Soil Ecology, 5: 261-71.*

DETR, Environment Agency and MRc Institute for Environment and Health. Guidelines for environmental risk assessment and management. *Revised Departmental Guidance ISBN 0-11-753551-6*

Department of the Environment. (1994). Contaminated Land Research report 2, Guidance on preliminary site inspection of contaminated land *(Two volumes). Department of the Environment, London.*

Department of the Environment, Transport and the Regions (2000). Contaminated Land Environmental Protection Act 1990: Part IIA. *DETR Circular 02/2000.*

Department for Environment, Food and Rural Affairs and the Environment Agency (2002a). Assessment of risks to human health from Land contamination: An overview of the development of soil guideline values and related research. *Report CLR7.*

Department for Environment, Food and Rural Affairs (2002b) Note on the withdrawal of ICRCL trigger Values CLAN.

Department for Environment, Food and Rural Affairs and the Environment Agency (2002c). Contaminated land exposure assessment model (CLEA). Technical Basis and Algorithms, *Report CLR10.*

Domsch K.H., Jagnow, G., and Anderson T.H. (1983). An ecological concept for the assessment of side-effects of agrochemicals on soil microorganisms. *Residue Reviews, 86: 65-105.*

Dorn P. B., Vipond, T. E., Salantrio, J.P., and Wisniewski, H.L. (1998). Assessment of the acute toxicity of crude oils in soils using earthworms, Microtox™ (R) and plants. *Chemosphere, 37: 845-60.*

Doherty F. G. (2001). A review of the Microtox (R) toxicity test system for assessing the toxicity of sediments and soils. *Water Quality Research Journal of Canada 36: 475-518.*

Environment Agency (1999). A proposed scheme to ensure the quality of data generated by laboratories undertaking regulatory ecotoxicological testing. *R&D Technical Report P166. Bristol, UK.*

Environment Agency (2000). Secondary Model Procedure for the Development of Appropriate Soil Sampling Strategies for Land Contamination. *Bristol, UK.*

Environment Agency (2001) Secondary model procedure for the development of appropriate soil sampling strategies for land contamination. *R&D Technical Report P5-066/TR. Bristol, UK.*

Environment Agency (2002). Guidance note on IPPC (H1) Environmental Assessment and Appraisal of BAT.

Environment Agency (2003). Ecological Risk Assessment. A public consultation on a framework and methods for assessing harm to ecosystems from contaminants in soil. *Bristol, UK.*

Ferguson C., Darmendrail, D., Freier, K., Jensen, B.K., Jensen, J., Kasamas, H., Urzelai, A., and Vegter, J. eds., (1998). Risk assessment for contaminated sites in Europe. Volume 1: *Scientific basis Vol. 1. 165 LQM press, Nottingham.*

Ferson S. (2002). Risk assessment with uncertain numbers. *Lewis Publishers, CRC Press, Boca Raton.*

Filzeck P. D. B., Spurgeon, D. J., Broll, G., Svendesen, C., Hankard, P.K., Parekh, N., Stubberub, H., and Weeks, J.M. Metal effects on soil invertebrate feeding: measurements using the bait lamina method. *Ecotoxicology (in press).*

Fishwick S. (2004). Soil screening values for use in ecological risk assessment. *Environment Agency R&D Report P5-091/TR, Bristol, UK.*

Fountain M.T., Hopkin, S.P. (2001) Continuous monitoring of Folsomia candida (Insecta: Collembola) in a metal exposure test. *Ecotoxicology and Environmental Safety Mar. 2001.*

Heimbach F. (1998). Comparison of the sensitivities of an earthworm (Eisenia fetida) reproduction test and a standardised field test on grassland. *In advances in earthworm ecotoxicology. Proceedings from the second international workshop on earthworm ecotoxicology 2-5th April 1997 (Eds SC Sheppard, JD bembride, M. Holmstrup and L Posthuma). Society of Environmental Toxicology and Chemistry, Amsterdam.*

Herbert I.N., C. Svendsen, Spurgeon D.J. and Hankard P.K. (2004). Comparison of instantaneous rate of population increase and critical-effect estimates in Folsomia candida exposed to four toxicants. *Ecotoxicology and Environmental Safety 57(2): 175-183.*

Hopkin S. P. (1989). Ecophysiology of metals in terrestrial invertebrates. *London, Elsevier Applied Science.*

Hopkin S. P. (1993). Deficiency and excess of copper in terrestrial invertebrates. Ecotoxicology of Metals in Invertebrates. *R. Dallinger and P. S. Rainbow. Chelsea, USA, Lewis Publishers.*

ICRCL (1987). Guidance on the Assessment and Redevelopment of Contaminated Land, *Report ICRCL 59/83, 2nd edn.*

Institute of Petroleum (1993). Code of Practice for the Investigation and Mitigation of Possible Petroleum-Based Land Contamination. *Institute of Petroleum, London.*

ISO (1998). Soil quality - effects of pollutants on earthworms (Eisenia fetida), Part 1: Determination of acute toxicity using artificial substrate. *International Standards Organisation, Geneva.*

ISO (1999). Soil Quality. Terminology and Classification. Terms and definitions relating to sampling. *International Standards Organisation, Geneva.*

Kalra Y.P., and Maynard, D.G. (1991). Methods manual for forest soil and plant analysis, Rep. No. NOR-X-319. *Forestry Canada, Northwest Region, Northern Forestry Centre.*

Kammenga J. E., and Simonsen, V. (1997). Manual of BIOPRINT-II, First technical report. *Ministry of Environment and Energy, National Environmental Research Institute, Silkeborg, Denmark.*

Kammenga J.E., Spurgeon, D.J., Svendsen, C., and Weeks, J.M. (2003) Explaining density-dependent regulation in earthworm populations using life-history analysis. *Oikos, 100: 89-95.*

Klok C., De Roos, AM., Marinissen, JCY., Baveco, H.M. and Ma, W.C. (1997). Assessing the effects of abiotic environmental stress on population growth in Lumbricus rubellus (Lumbricidae Oligochaeta). *Soil Biology and Biochemistry. 29, (4): 287-93.*

Linder G., Greene, J. C., Ratsch, H., Nwosu, J., Smith, S., and Wilborn, D. (1990). Seed germination and root elongation toxicity tests in hazardous waste site evaluation: methods development and applications. *In Plants for Toxicity Assessment. ASTM STP 1091. (eds W.W., J.W. Gorsuch and W.R. Lower), pp 177-87. American Society for Testing and Materials, Philadelphia.*

McMahon, A., Heathcote, J., Carey, M. and Erskine, A. (2001). Guide to good practice for the development of conceptual models and the selection and application of mathematical models of contaminant transport processes in the subsurface. *Environment Agency R&D Technical Report: NC/99/38/2. Bristol, UK.*

MEF (1998). Politique de protection des sols et de réhabilitation des terrains contamines. Service des lieux contamines, Ministère de l'Environnement du Québec.

Moore D. R. J., Sample, B. E., Suter, G.W., Parkhurst, B.R., and Teed, S.R. (1999). A probabilistic risk assessment of the effects of methyl mercury and PCBs on mink and kingfishers along East Fork Poplar Creek, Oak Ridge, Tennessee, USA. *Environmental Toxicology and Chemistry, 18: 2941-53.*

Nathanail J., P., Bardos, P., and Nathanail P. (2002). Contaminated Land Ready Reference. Land Quality Press, EPP Publications.

OECD (2000a). OECD Guidelines for the testing of chemicals 217 Soil microorganisms, carbon transformation test. *Organisation for Economic Co-operation and Development, Paris, France.*

OECD (2000b). OECD Guidelines for the testing of chemicals. 216 Soil microorganisms, nitrogen transformation test. *Organisation for Economic Co-operation and Development, Paris, France.*

OECD (2000c). OECD Guidelines for the testing of chemicals. Draft new guidance: earthworm reproduction test (Eisenia fetida/Andrei). *Organisation for Economic Co-operation and Development, Paris, France.*

OECD (2000d). Proposal for updating Guideline 208 A: Seedling Emergence and seedling growth test 208B Vegetative vigour test. *Organisation for Economic Co-operation and Development, Paris, France.*

Pastorok R.A., Bartell, S.M., Ferson, S. and Ginzburg, L.R. Ecological Modelling in risk assessment; Chemical effects on populations, ecosystems and landscapes. (2002). *Lewis Publishers, Boca Raton FL.*

Posthuma L., Suter, G.W., and Traas, T.P. (2002). Species sensitivity distributions in ecotoxicology. *Lewis Publishers, Boca Raton FL*

Power E.A., and Boumphrey, R.S. (2004). International trends in bioassay use for effluent management. *Ecotoxicology, Volume 13, Number 5: 377-98.*

Riepert F., and Wilke B.M (1998) Guidance on the ecotoxicological characterisation of soils and soil materials 9th Draft Report to the international Standards Organisation ISO/CD 15 799, *ISO/TC 190/SC7/WG 3*

Read H.J., and Martin, M.H. (1993). The effects of heavy metals on populations of small mammals from woodlands in Avon (England); with particular emphasis on metal concentrations in Sorex araneus L. *Chemosphere 27: 2197-2211.*

Read H. J., Martin, M. H., and Rayner, J. M. V. (1998). Invertebrates In woodlands polluted by heavy metals - An evaluation using canonical correspondence analysis. *Water Air and Soil Pollution 106 (1-2): 17-42.*

Regan H. M., Hope, B. K., and Ferson, S. (2002). Analysis and portrayal of uncertainty in a food web exposure model. *Human and Ecological Risk Assessment, 8: 1757-77.*

Salanitro J. P., Dorn, P. B., Huesemann, M.H., Moore, K.O., Rhodes, I.A., Jackson, L.M.R., Vipond, T.E., Western, M.M., and Wisniewski, H.L. (1997). Crude oil hydrocarbon bioremediation and soil ecotoxicity assessment. *Journal of Environmental Science and Technology, 31, (6): 1769-76.*

Sandifer R. D. (1996). The effects of cadmium, copper, lead and zinc contamination on arthropod communities in the vicinity of a primary smelting works. *University of Reading, Reading, UK.*

Saterbak A., Toy, R. J., Wong, D.C.L., McMain, B.J., Williams, M.P., Dorn, P.B., Brzuzy, L.P., Chai, E.Y., and Salanitro, J.P. (1999). Ecotoxicological and analytical assessment of hydrocarbon-contaminated soils and application to ecological risk assessment. *Environmental Toxicology and Chemistry, 18: 1591-607.*

Schaefer M. (2001). Earthworms in Crude Oil Contaminated Soils: Toxicity Tests and effects on crude oil degradation. *Contaminated Soil and Water 35-37 .*

Smit C. E. and Van Gestel, C. A. M. (1996). Comparison of the toxicity of zinc for the springtail Folsomia candida in artificially contaminated and polluted field soils (vol 3, p. 127, 1996). *Applied Soil Ecology, 4: 93-4.*

Spurgeon D. and Hopkin, S. P. (1995). Extrapolation of the laboratory based OECD earthworm toxicity test to metal-contaminated field sites. *Ecotoxicology, 4: 190-205.*

Spurgeon D. and Hopkin, S. P. (1996). The effects of metal contamination on earthworm populations around a metal smelting works - quantifying species effects. *Applied Soil Ecology, 4: 147-60.*

Spurgeon D. and Hopkin, S. P. (1999). Seasonal variation in the abundance, biomass and biodiversity of earthworms in soils contaminated with metal emissions from a primary smelting works. *Journal of Applied Ecology, 36: 173-83.*

Spurgeon D., Svendsen, C., Hankard, P., Weeks, J., Kille, P., and Fishwick, S. (2002). Review of sublethal ecotoxicological tests for measuring harm in terrestrial ecosystems. *Technical Report No. P5-063/TR1. Environment Agency, Bristol, UK.*

Spurgeon D., Svendsen, C., Rimmer, V.R., Hopkin, S.P., and Weeks, J.M. (2000). Relative sensitivity of life cycle and biomarker responses in four earthworm species exposed to zinc. *Environmental Toxicology and Chemistry, 19: 1800-8.*

Spurgeon D.J., Svendsen, C., Hankard, P.K., Toal, M., McLennan, D., Wright, J., Walker, L., Ainsworth, J., Wienberg, C., Fishwick, S.K., (in press) Application of sub lethal ecotoxicological tests for measuring harm in terrestrial ecosystems. *Technical Report No. P5-063/TR2, Environment Agency, Bristol, UK.*

Stansley W. and Roscoe, D. E. (1996). The uptake and effects of lead in small mammals and frogs at a trap and skeet range. *Archive of Environmental Contamination and Toxicology. 30: 220-226.*

Sunahara G., Renoux A., Pilon, A., and Gudet, C. (Eds) (2002). Environmental Analysis of contaminated sites. *John Wiley and Sons, Chichester & New York.*

Suter G.W. (1993). Ecological risk assessment. *Lewis Publishers, Boca raton FL.*

Suter G.W., Efroymson, R.A., Sample, B.E., and Jones, D.S. (2000). Ecological risk assessment for contaminated sites. *Lewis Publishers, Boca Raton, FL.*

U.S. Department of Energy (1998). A Guide to the ORNL Ecotoxicological Screening Benchmarks: Background, Development, and Application. *(Available to the public from the National Technical Information Service, U.S. Department of Commerce, Springfield, VA, United States).*

U.S. Environmental Protection Agency (US EPA). (1989a) Guidelines for ecological risk assessment USEPS.

U.S. Environmental Protection Agency (US EPA). (1989b) Ecological risk management principles for superfund sites. *Directive 92857-28P, Office of solid waste and emergency response, Washington DC.*

U.S. Environmental Protection Agency (US EPA). (1995) Elements to consider when drafting EPA risk characterisations, *USEPA Washington DC.*

U.S. Environmental Protection Agency (US EPA). (2000). Ecological Soil Screening Level Guidance (draft). July 2000. *US EPA, Office of Emergency and Remedial Response, Washington DC, United States.*

U.S. Environmental Protection Agency (US EPA). (2001). Risk Assessment Guidance for Superfund (RAGS), Volume III - Part A. Process for Conducting Probabilistic Risk Assessment. *US EPA Office of Emergency and Remedial Response. Washington DC, United States.*

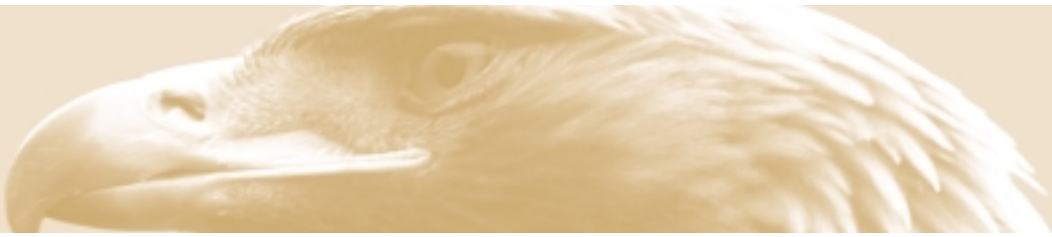
Van Gestel C. A. M., van der Waarde, J.J., Derksen, J.G.M., van der Hoek, E.E., Veul, M., Bouwens, S., Rusch, B., Kronenburg, R., and Stokman, G.N.M. (2001). The use of acute and chronic bioassays to determine the ecological risk and bioremediation efficiency of oil-polluted soils. *Environmental Toxicology and Chemistry, 20: 1438-49.*

Vose D. (2000). Risk analysis: a quantitative guide. *2nd edition John Wiley & Sons Ltd, Chichester.*

VROM - Ministry for Housing, Spatial Planning and Environmental Protection. (2000). Circular of target values and intervention values for soil clean up. No 39.

Wharfe J. (1996) Toxicity based criteria for the regulatory control of waste discharges and for environmental monitoring and assessment in the United Kingdom. In Toxic Impacts of Wastes on the Aquatic Environment (eds., Tapp, J.f., Hunt, S.M., and Wharfe, J.R.) *The Royal Society of Chemistry, Cambridge, UK.*

Whitehouse P., Crane, M., Redshaw, C.J., and Turner, C. (1996). Aquatic toxicity tests for the control of effluent discharges in the UK - the influence of test precision. Ecotoxicology, 5 155-168.



Appendix A:

Description of Tier 0 - Conceptual Site Model

Summary

This section describes the Conceptual Site Model (CSM) and its role in the ERA process. We provide examples to illustrate the step-by-step development of a CSM.

Key points:

- The CSM is a narrative summarising existing site data;
- The CSM is the essential building block of the ERA tiered process and is used to designate, implement, or complete an ecoreceptor site investigation;
- The CSM is used to evaluate site risks (contaminants, pathways receptors), data gaps, and cost-effective solutions for remediation;
- As new information becomes available during the site investigation, the CSM is updated;
- The CSM is a tool for effectively communicating complex site investigations issues to stakeholders and providing them with an opportunity to contribute to the process with their knowledge about the site under investigation;
- The level of detail in a CSM is consistent with objectives of site investigation;
- It is often useful to provide both textual and graphical representations of the CSM.

Introduction

This section describes the type of information that should be included when developing a CSM for assessing the impacts of contaminated land to an ecoreceptor.

Early on in the project, we recognised the need for another tier ('Tier 0') to be added to the proposed ERA framework. This tier provides an early opportunity to determine whether the site under investigation does actually fall within Part IIA considerations, and to develop the CSM before progressing onto further tiers.

A logical approach is required to understand the complexities of any site. This begins with a combination of desk studies and subsequent site visits and explorations. A CSM is a systematic documentation and planning procedure that supports a consistent approach to site investigation. The CSM is the essential first step. It underpins the ERA framework for evaluating the importance of information and planning the activities that make up the subsequent investigative effort. All activities undertaken as part of the ERA, including measurement, analysis and integration of data, should have a direct and logical connection to the CSM.

Byrns and Crane (2002) discuss the use and role of CSMs within the risk assessment process. In summary, the purpose of the CSM is to identify potential contaminants, pathways and receptors with a view to identifying (initially) potential (and eventual) significant pollutant linkages. The CSM should highlight the degree of uncertainty in the data and provide all interested parties with a vision of the site. It underpins each subsequent stage of contaminated land management and is tested and reviewed at subsequent higher stages (tiers) of the ERA framework.

Answers to three simple questions about the hazard form the basis of the CSM:

- What is already known about the site and its historic use?
- What is unknown and relevant to the planned re-use and its environmental impact?
- What needs to be known now?

A CSM can be presented in a table, diagram, pictorial or flowchart form (such as Figure A1), but will also include a narrative description. This describes the contents of the diagram in enough detail for an educated layperson to understand it. The narrative conceptual model will not, though, duplicate information that exists in other sections of the document, such as the site description. An example CSM for exposure to hydrocarbon fuels at petroleum retail sites is presented in DETR (2000). This example is not restricted to ecosystem receptors.

The first objective of the CSM procedure within Tier 0 is to review all available published and unpublished literature, reports, documents and data for each site. This review includes details of samples analysed to date, local meteorological conditions, other studies ongoing in the region, and other documents available from all parties. Generic technical guidance on conceptual models is provided in McMahon *et al.,* (2001), which focuses on the risks to the subsurface environment, but includes the essential components of CSMs. The consideration of Part IIA ecosystem receptors will require more emphasis on certain areas of the CSM, pathways of exposure for example. The following list highlights aspects that should be considered during the CSM development when Part IIA ecosystem receptors are of potential concern or are the recognised focus of the study:

In summary the CSM steers the site investigation process and should identify:

- potential contaminants, pathways, and receptors for chemicals of concern;
- actual versus perceived site risks and urgency for response;
- relevant fate and transport mechanisms for chemicals of concern;
- data gaps and their significance.

An initial CSM may be sufficient to exit the ERA process if no linkage to an eco-receptor can be justified. The CSM should, though, identify the following as a minimum.

Contaminants

Each distinct type of ultimate source will be identified separately in the CSM. Types of sources will be distinguished when they contain chemicals that are distinctly different in form or composition or disposed of in different manners (ponds versus tanks, for example) or in situations that would result in different modes of transport (floodplains versus uplands, for example). These may be in the form of fluxes of surface water, groundwater, eroded soil, or suspended sediments or as dusts. They will be identified in terms of their nature and source.

At this stage, assessors should highlight contaminants of concern which have the potential to bioaccumulate and/or biomagnify, and consider them further within Tier 1 (see Section 2).

Routes of transport

The assessor must describe all pathways by which contaminants are transferred from the sources to ambient media (to which organisms may be exposed). Fate processes (such as degradation and sequestration) should be included in the CSM if deemed of value. This may be particularly useful if suspected breakdown products are more hazardous than the parent compound.

Pathways of exposure

Identify all plausible exposure pathways to ecological receptors, specifically:

- Ingestion by food and/or water, soil ingestion may be excluded for species that have little exposure to soil;
- Incidental ingestion (via consumption of soil-covered plant roots, for example);
- Dermal contact - amphibians, for example, are likely to experience significant dermal uptake. Feathers and fur can exclude most dermal exposures. They can, though, create another route of exposure: grooming and preening, which contribute to incidental soil ingestion;
- Plants, soil invertebrates, and soil microbes are assumed to be directly exposed to whole soil;
- Wind-blown dust;
- Inhalation of vapours;
- In cases where shallow groundwater is contaminated, plants are exposed to that water;
- Fish, aquatic invertebrates, and aquatic plants are assumed to be exposed to contaminants in water and aquatic sediments;
- For those chemicals known to bioaccumulate the pathways between the Part IIA receptor and their food supply need to be reviewed (for example, if a specific bird is a Part IIA receptor, then identify the potential pathway of chemical uptake via its food supply, such as via soil to soil invertebrates). If merited, this connection can be explored further in higher tiers of the framework as the investigation progresses.

Exposure characterisation and assessment is discussed in Byrns & Crane (2002). We discuss it for the receptors of concern identified in this project for sites A and B in Sections 5 and 6.

Receptors

The assessor should examine the ecological receptors identified to establish that they satisfy the requirements listed in Statutory Guidance for Part IIA (DETR, 2000; Byrns & Crane, 2002). Receptors should be broken down into separate categories where appropriate (for example, separate identification of different life stages, or separation into individuals versus communities).

Spatial and temporal boundaries

The spatial boundaries of the CSM need to be defined. These should take into account the legal boundaries specified by the legislation. In early tiers of ERA, one often assumes that all organisms under consideration have 100 per cent spatial and temporal overlap with contamination sources. This may result in an overestimation of risk. It is, though, appropriate in the early (screening) phases of the ERA. These risk predictions can then be refined in subsequent tiers of the ERA framework or rejected if they remain unsubstantiated. The issue of designated receptors that use a contaminated site for a proportion of their life cycle (moving on and off-site foraging, etc.) needs to be highlighted within the CSM. The same applies to pathways that transport contaminants off-site, subsequently posing a risk to designated species.

The temporal overlap of receptors by their seasonal and diurnal patterns should be highlighted. This ensures that exposure that is coincident with sensitive periods (for example, breeding seasons or raising young) can be identified and evaluated in more detail at a higher tier. Byrns & Crane (2002) and references therein discuss the issue of spatial and temporal factors in attempting to assess exposure of potential receptors to contaminants. They also provide information on available modelling and statistical techniques.

Preliminary assessment of a pollutant linkage for each contaminant, pathway and receptor combination

Once the assessor has collated data to identify potential pollutant linkages of concern, the following must be identified; the (potential) pathways, contamination hazards that can affect the likely receptors, the scale of the contaminant exposure and which receptors (if any) could be affected. If receptors or routes of exposure are omitted due to lack of information, the assessor will acknowledge that omission. It will be included in the analysis of uncertainty that will be taken up within Tier 1 of the framework and re-examined following the collection of additional data and subsequent refinement of the CSM. It may be that the decision is taken to exit at Tier 0 if no (or little) risk is perceived to be acting on an ecoreceptor of concern.

Relationship to other conceptual models

The conceptual model for ecological risks must be consistent with any CSM derived to assess risks to human receptors at the same study site. That is, it should identify the same contaminants, routes of transport of contaminants, and contaminated media. The different routes of exposure and receptors will, though, be highlighted. This commonality will permit the assessor to extract maximum information from the literature search conducted during the development of the CSM and about any site-specific chemical or structural data collected as the investigation progresses, saving both time and expenditure.

Development of the CSM in subsequent tiers

The CSM will be refined and developed in the subsequent tiers of the framework. This happens as more information becomes available from activities such as site-visits and walk-overs. The plausibility of the potential pollutant linkages identified in Tier 0 can be assessed in this tier, but may also benefit from the additional data collected at higher tiers. The ERA process will refine information about:

1. receptors that do not occur, or are not important at the site;
2. routes of exposure that are not credible or important;
3. routes of exposure that do not lead to endpoint receptors;
4. potential sources that are not deemed credible or important.

In addition, the process makes the conceptual model more specific by identifying particular endpoints and by defining the spatial and temporal scale of the assessment.

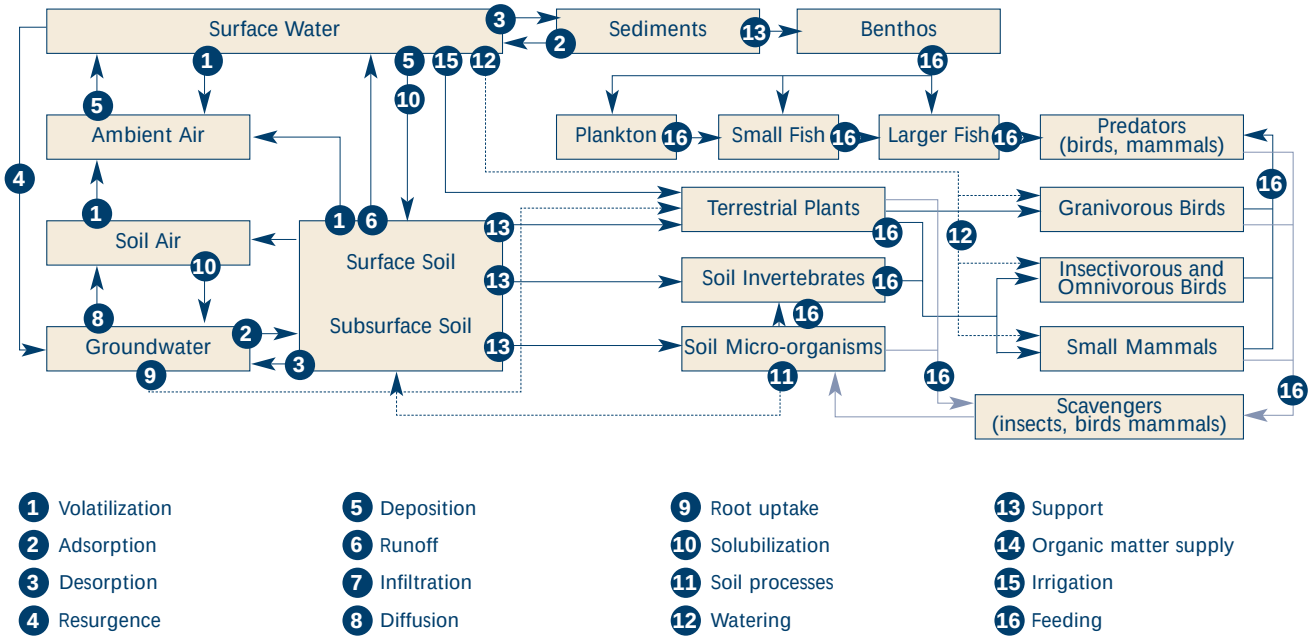
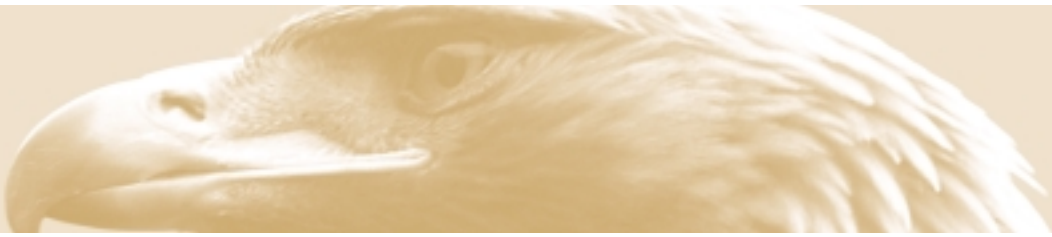


Figure A1 | Generic conceptual site model for ERA - example of flow chart



Appendix B:

Details of the selected methods

Microtox™ bacterial bioluminescence test

Test procedure

The toxicity of environmental samples (leachates, treated or untreated industrial and sewage effluents, sediment and soil pore waters or leachates) can be assessed using the Microtox™ bacterial bioluminescence test, which measures the effects of contaminants on Vibrio fischeri. In the Microtox™ acute toxicity test procedure, bacterial populations are exposed for periods of 5, 15 or 30 minutes to the test substance either at a single concentration or diluted to a range of concentrations. Under appropriate conditions, the bacteria produce light as a by-product of metabolism, and this can be measured by a photomultiplier. Contaminants can act to reduce the light produced. The effects range from an absence of effect on light output at lower test concentrations to a complete inhibition of light output at higher test concentrations. Changes in light output are used to determine, where possible, the median inhibitory concentration of the sample, that is the concentration that reduces light output by 50 per cent, after 5, 15 or 30 minutes, relative to a control. The Microtox™ Basic Toxicity Test Procedure and the Microtox™ 100% Toxicity Test Procedure are an Agency Standard Procedure. The technique uses small volumes of material. It is, therefore, useful for testing sediment or soil pore waters where the amount of available sample may be limited.

Mutatox is a further variant of the Microtox™ method. It determines the genotoxicity of samples in liquid phase (from aqueous soil extracts) by measuring the changes of light produced by bioluminescent bacteria. The test is a screening tool to detect the presence of genotoxins (DNA-damaging substances) in complex leachates/ extracts. The same analytical unit is used as for Microtox™, but with an extended light sensitivity. The relative genotoxicity of the sample is determined by comparing the stress-induced increase in light emissions with controls.

Sample preparation

Leachates from soils are prepared by initially drying aliquots of the soil for 24h at 60°C before extracting the contaminants. The extraction procedure involves adding 0.01M $CaCl_2$ to dried soil at a ratio of 1 part soil to 4 parts solution. The soil/$CaCl_2$ suspension is then shaken at 100 rpm on an orbital shaker for 2 hours[5]. The resulting solution is then centrifuged for 10 minutes, and the supernatant used as the test solution.

The Agency Standard Procedure includes a Solid Phase Test in which effects are measured on a defined number of bacterial cells exposed to soil suspensions over a 30-minute period. Typically, the bacteria are exposed to a concentration series of aqueous suspensions of each soil. Experience shows that, over a concentration range of 0.005 to 19.7 per cent Kettering Loam, there is no reduction in light measurements due to the turbidity of the suspended sediment. Any effects observed can be ascribed to leaching of contaminants from the soil particles.

Quality assurance/quality control

All tests are accompanied by reference toxicant tests with zinc and phenol for quality control purposes.

Nitrogen mineralisation test

The nitrogen mineralisation test assesses the ability of the autochthonous populations of microorganisms in a soil to convert organic nitrogen to inorganic nitrogen. The extent of this mineralisation process in soil from each contaminated site is compared to a control soil of good quality. Soils are freshly collected from the field and their physicochemical properties characterised. Large objects are removed, and the soil sieved to a particle size less than two millimetres. The moisture content of the soil is adjusted to a value 40-60 per cent of its maximum water-holding capacity.

[5] This duration has been shown to be as effective at removing inorganic contaminants as using shaking periods of 24h

The processed soils from the contaminated sites and the control are amended with a suitable source of organic nitrogen, such as powdered lucerne-grass-green meal. After thorough mixing, the soils are incubated in the dark at 20 +/- 2°C for 28 days. The concentration of nitrate is measured after 14 and 28 days incubation.

Carbon mineralisation test

This test is based on the OECD test guidelines. It is designed to assess the effects of toxic substances on the carbon mineralisation activity of soil micro-organisms. The test is sensitive to changes in size and activity of the microbial populations that are responsible for carbon mineralisation.

The contaminated test soil and the healthy control soils are cleared of large objects, sieved to a defined particle size (2 mm) and their moisture content adjusted to 40-60 per cent. The soils are then incubated at 20°C for 14 and 28 days. At these time intervals, the soils are supplemented with glucose. Following the nutrient addition, the respiration rates of the soils are monitored. This is through the evolution of carbon dioxide released or consumption of oxygen, for a period of 12 hours. The deviation of the test soil respiration rate from the control soil respiration rate is calculated. If, at the end of the 28-day period, the respiration rate of the contaminated test soil differs by more than 15 per cent from that of the control, the test duration is extended. It continues until a difference equal to or less than 15 per cent is obtained, or for a maximum of 100 days, whichever is shorter.

The individual data (that is, per replicate) on carbon dioxide evolution or oxygen consumption is analysed. The results are evaluated by appropriate statistical methods, for example the F-test at the five per cent significance level.

OECD germination and shoot-growth tests

Seedling emergence and growth tests would be carried out according to a draft OECD test guideline (OECD 208A). This method entails monitoring the biological responses (shoot emergence and shoot growth) of plants exposed to soil to which the test substance has been added. The substance is added either directly to the surface (when the aim is to simulate pre-emergence application of an agrochemical) or is incorporated into the growing medium. In the context of this project, the latter route of administration is the more relevant.

Tests may be carried out using a single application rate ('limit test') or with a range of exposure levels from which an EC50 or NOEC may be estimated.

Whether a 'limit test' or 'dose response' design is adopted could depend on the point within the risk assessment process. Both have a place; the former in an initial screening role, and the latter where data indicate that a 'refined' assessment of risks is warranted. In both cases, replication is important in ensuring adequate statistical power and to avoid bias. OECD 208A recommends the use of four true replicates in a dose-response design. Our experience, though, shows that a higher level of replication is advisable for wild species. They have a greater inherent variability in emergence and growth, and possibly sensitivity, compared with the greater genotypic and phenotypic uniformity that characterises cultivated species.

Test species

OECD 208A recommends a range of test species (including both monocotyledonous and dicotyledonous species). But they are all cultivated species. Only two (red clover *Trifolium pratense* and perennial ryegrass *Lolium perenne*) occur indigenously in the UK. Consequently, there remains an issue about standardisation in the choice of test species versus relevance of the chosen species.

Relying on cultivated species may lack field relevance, but it is appropriate to regard the species used as indicator species only. This parallels the situation in aquatic toxicity testing where, for example, Daphnia magna is often used as a surrogate for a particular trophic level. We suggest that the responses obtained with cultivated species are no more or less relevant than those obtained using indigenous species (except if we were concerned with risks to a particular species). For this reason, we would take a single dicot and monocot species from Annex 2 of OECD 208A, ensuring that they are species that can be obtained and grown throughout the year under glasshouse conditions. Previous experience of such studies, and also of the sensitivity of many of these species to a wide range of chemicals, suggests that suitable species may be drawn from the following species:

Dicotyledons	**Monocotyledons**
Mustard (Brassica napus)	Oat (Avena sativa)
Sugarbeet (Beta vulgaris)	Wheat (Triticum aestivum)

Modifications to plant test protocols

We made a number of modifications to the guideline to make the procedure more relevant to

Testing contaminated soil using OECD test 208a

Approach	Strengths	Limitations
• Contaminated soil incorporated into growing medium (including 100% contaminated soil)*.	• Provides assessment of the biological effects of 'whole' soil sample • Realistic bioavailability	• There is an upper 'ceiling' on level of contaminated soil that can be incorporated (would reduce sensitivity of 'limit test' option) • Presence of nutrients in contaminated soil may give rise to hormesis • Physical effects of test sample can give rise to spurious biological responses (risk of 'false positive')
• Aqueous extract of contaminated soil prepared and incorporated into growing medium	• Closer adherence to OECD 208A • Comparison with control is more robust (only variable is level of extract)	• Selective extraction of contaminants • Difficult to relate results to phytotoxicity of whole soil • May be differences in availability of contaminants between extract-amended soil and contaminated soil

*Controls = plants grown in pure growing medium or in soil from a reference site where contamination is absent but whose physical properties are similar to those of the contaminated soil

contaminated land assessment.

The major method deviations involved the size of the test pot and the seeding density. Guideline 208A suggests the use of 15 centimetre (diameter) pots for testing. Such large test pots, though, require considerable volumes of soil (~1 kilogramme per pot), making them unsuitable for contaminated soil experiments. Consequently, we carried out comparative experiments to assess the effects of using smaller test pots (10 centimetres diameter) on emergence and growth of pea and wheat (comparisons were made on identical sowing densities in each pot size). Results indicated that there was no significant difference (T-test P> 0.05) in emergence or dry weight of either species grown in 10 centimetre pots when compared with 15 centimetre pots (based on two sowing densities of three and five pea seeds/pot and six and 10 wheat seeds/pot). So we decided to use 10 centimetre pots.

In addition to the pot-size experiments, we carried out tests to assess the optimum seeding density for 10 centimetre pots. Pea (representing large seeded species) and wheat (representing small seeded species) were sown in 10 centimetre pots at sowing densities of three, five and 10, and at six, 10 and 20 seeds per pot, respectively. Analysis of the results using oneway ANOVA and multiple comparisons (Tukey's HSD) indicated a significant difference (P <0.05) in emergence and dry weight of plants at seeding densities greater than 10 and 20 seeds/pot for pea and wheat, respectively. That is, the seeding densities were too high. Therefore, the optimum seeding densities were between three and five seeds/10 centimetre pot for pea, and six to 10 seeds/10 centimetre pot for wheat. So we used seeding densities of four seeds/pot for pea, and six seeds/pot for wheat, cabbage, oat and tomato for the contaminated soil experiments.

The OECD acute earthworm test with Eisenia fetida/Eisenia andrei

The OECD Earthworm, Acute Toxicity Tests guideline describes two kinds of tests: a paper contact toxicity test and an artificial soil test. The simple paper contact toxicity test is described as an optional initial screen to indicate those substances likely to be toxic to earthworms in soil and which will require further more detailed testing in an artificial soil. The artificial soil test involves keeping earthworms in samples of a precisely defined artificial soil to which a range of concentrations of the test substance has been applied. Mortality is assessed seven and 14 days after application. The mortality in the controls should not exceed 10 per cent at the end of either test.

The recommended test species is *Eisenia fetida*. Though this is not a typical soil species, it occurs in soil rich in organic matter, and it has a number of advantages. It has a short life cycle, hatching from cocoons in three to four weeks, and reaching

maturity in seven to eight weeks at 20°C. It is very prolific; each worm produces two to five cocoons per week from each of which several worms emerge. It is available commercially, and it can be bred readily in a wide range of organic waste materials. *Eisenia fetida* exists in two races, which some taxonomists have separated into species. These are morphologically similar, but one, *E. fetida fetida*, has a typically transverse striping or banding on the segments, and the other, *E. fetida andrei*, lacks this and has a variegated reddish colour. Where possible, *E. fetida fetida* should be used. Other species may be used if the necessary methodology is available. Worms should be adult (at least two months old with clitellum), with an individual weight of 300 to 600 milligrammes.

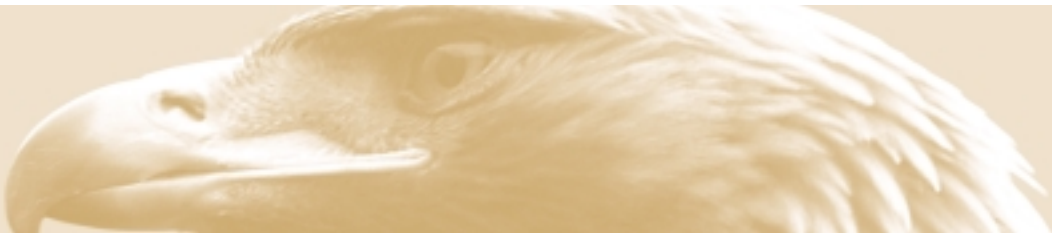
For each test, 750 grammes of the test medium is placed into each glass container and 10 earthworms, which have been conditioned for 24 hours in an artificial soil and then washed quickly before use, are placed on the test medium surface. The containers are covered with perforated plastic film to prevent the test medium from drying and kept under the test conditions for 14 days. Four replicates for each treatment are recommended. For each test, four control dishes, treated with the same solvent as that used in the test and containing 10 worms, are used. The test duration is 14 days (assessment of mortality at seven and 14 days), and the test temperature is 20° ± 2°C. Testing is done in continuous light (to ensure that worms remain in the test medium throughout the duration of test). Mortality is assessed by emptying test medium onto a glass tray or plate, sorting worms from the medium and testing their reaction to a mechanical stimulus at the front end. After the seven-day assessment worms and medium are replaced in the test container. Any behavioural or pathological symptoms noted should be reported. At the end of the test, the moisture content of the test medium should be assessed and reported.

Mortality data can be analysed using log probit and ANOVA methods. Results can be expressed as an LC50 (specifying exposure period), the highest concentration causing no mortality and the lowest concentration causing 100 per cent mortality.

The ISO springtail test with Folsomia candida

This test uses reproduction rates of the soil-dwelling springtail *Folsomia candida* to assess the toxicity of compounds or contaminated soils. *F. candida* is parthenogenetic, reaches an adult length of between one-and-a-half and three millimetres and is easy to breed, with low expenditure on time and equipment. The development period is short (two to three weeks at 20°C) and the reproductive rate is high. Springtails are also suitable for ecotoxicological testing because they occur in all types of soil and are important to soil biology. They are affected by both mechanical and chemical insults and are exposed to toxins via the epidermis, ventral tube (water uptake) or gut via food. It is not clear, though, which uptake routes are the most important. Fungi are a major component of the diet of most springtails. This is an important exposure route for toxicants, as most fungal species accumulate metals in their hyphae.

The ISO guideline (ISO, 1998) describes the recommended protocol for conducting a standard soil exposure test for *F. candida*. This involves adding 10 *F. candida* to each of at least four replicates of an artificial soil made from a mixture of peat, clay and sand. Collembola are then incubated for four weeks at 20°C. At the end of the experiment, the animals are separated from the soil by flooding with water, and adults and offspring are counted. The test is inexpensive and relatively easy to conduct given sufficient experimenter experience. Standard test procedures begin with 10 to 12-day-old *F. candida* (that is, virtually adult).



Appendix C:

Soil quality guideline values

Soil quality guidelines can have a number of specific protection goals (that is, humans, earthworms, microbes, land-uses). But they all provide information on soil concentrations that are regarded as 'safe' with respect to a specific receptor. Therefore, any exceedance of the guidelines would imply that the chemical in question might have a negative effect on the receptor/protection goal in question.

Below, we outline a number of international soil quality guidelines (the Canadian CCME, Revised Dutch (VROM) and US Department of Energy (DOE) values. We provide the values used in the assessments of risk based on chemical residue data at Sites A and B in Table C1.

Canadian CCME

The CCME (Canadian Council of Ministers of the Environment) recommended soil quality guidelines are based on four land uses: agricultural, residential/parkland, commercial and industrial. They aim to protect soil and, at the guideline levels, will provide a healthy functioning ecosystem capable of sustaining the current and likely future uses of the site by receptors.

These soil quality guidelines (environment) are derived using toxicological data to determine the threshold level of effects for key ecological receptors. Exposure from soil is the primary derivation procedure for environmental quality guidelines regarding all but the agricultural land use. For that, another derivation procedure based on soil and food ingestion is applied, with the lower of the two values considered as the environmental soil quality guidelines for this land use. The SQG (environment) and the SQG (human health) are then compared before the final SQG is derived.

Netherlands (Revised Dutch list) soil remediation guidelines (VROM)

The Netherlands set two guidelines for soil remediation: target or optimum values, and intervention or action values. Target/optimum values indicate the level at which there is a sustainable soil quality, that is, the level that has to be achieved to recover fully the functional properties of the soil for humans and plant and animal life. In addition, the target values give an indication of the benchmark for environmental quality in the long term, based on the assumption of negligible risks to the ecosystem. Intervention/action values are the concentration above which there is serious contamination. If the intervention values are exceeded, the functional properties of the soil for humans, flora or fauna have been seriously diminished or are in danger of being seriously diminished. An exceedance of the intervention values is measured as the mean concentration of at least one substance in at least 25 m^3 of soil volume.

US DoE Standards

The United States Department of Energy (DOE) has set a number of screening level benchmarks for earthworms, plants and soil microbes. The screening levels are based on laboratory toxicity data for these organisms and are based on the 10th percentile of a distribution of toxicity data. The aim, therefore, is to protect 90 per cent of earthworm, plant or microbe species. The benchmarks do not have a protection goal per se, but are proposed for general contaminant screening purposes. If a chemical concentration exceeds the screening benchmark, then the contaminant is highlighted as being of 'potential concern' and will require further analysis. Concentrations that fall below the screening level may be ignored, unless public concern or ancillary evidence suggests that a chemical should be investigated further.

Table C1 | Summary Soil Quality Guideline Values

Contaminant	CCME				New Dutch list				US DoE		
	Soil				Soil sediment		Groundwater		Earthworm	Microbial	Plant
	Agri	Res/Park	Comm	Indu	(mg/kgdry weight)		(µg/l)		benchmark	benchmark	benchmark
	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	Target value	Intervention value	Target value Shallow/deep	Intervention value	(mg/kg)	(mg/kg)	(mg/kg)
Inorganic chemicals											
Aluminium										600	50
Antimony					3	15	No data/0.15	20			
Arsenic	12	12	12	12	29	55	10/7.2	60	60	100	10
Barium	750	500	2000	2000	160	625	50/200	625		3000	500
Boron										20	0.5
Bromide					20	No data	0.3mg/l[2]	No data			
Cadmium	1.4	10	22	22	0.8	12	0.4/0.06	6	20	20	4
Chromium (total CCME)	64	64	87	87	100	380	1/2.5	30	0.4	10	1
Chromium Hexavalent(Cr(VI))	0.4	0.4	1.4	1.4							
Cobalt					9	240	20/0.7	100		1000	20
Copper					36	190	15/1.3	75	50	100	100
Cyanides (New Dutch-free)	0.9	0.9	8	8	1	20	5	1500			
Cyanides-complex (pH<5)1				5	5	650	10	1500			
Cyanides-complex (pH<5)				5	5	50	10	1500			
Fluoride					500[3]	No data	0.5mg/l[2]	No data			
Fluorine									30	30	200
Iron										200	
Lanthanum										50	
Lead	70	140	260	600	85	530	15/1.7	75	500	900	50
Lithium										10	2
Manganese											500
Mercury	6.6	6.6	24	50	0.3	10	0.05/0.01	0.3	0.1	30	0.3
Molybdenum					3	200	5/3.6	300		200	2
Nickel	50	50	50	50	35	210	15/2.1	75	200	90	30

Table C1 (cont.) | Summary Soil Quality Guideline Values

Contaminant	CCME				New Dutch list				US DoE		
	Soil				Soil sediment		Groundwater		Earthworm	Microbial	Plant
	Agri	Res/Park	Comm	Indu	(mg/kgdry weight)		(µg/l)		benchmark	benchmark	benchmark
	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	Target value	Intervention value	Target value Shallow/deep	Intervention value	(mg/kg)	(mg/kg)	(mg/kg)
Inorganic chemicals											
Selenium									70	100	1
Silver										50	2
Thalium	1	1	1	1							1
Tin										2000	50
Titanium										1000	
Tungsten										400	
Vanadium	130	130	130	130						20	2
Zinc	200	200	360	360	140	720	65/24	800	200	100	50
Organic chemicals											
1,1-Dichloroethene					0.02	15	7	900			
1,1-Dichloroethene					0.1	0.3	0.01	10			
1,1,1-Trichloroethene					0.07	15	0.01	300			
1,1,2-Trichloroethene					0.4	10	0.01	130			
1,2-Dichloroethene					0.02	4	7	400			
1,2-Dichloroethene					0.2	1	0.01	10			
1,1,2-Trichloroethene (Trichloroethylene, TCE)	0.1	3	31	31							
1,1,2,2-Trichloroethene (Tetrachloroethylene, PCE)	0.1	0.2	0.5	0.6							
Anthracene					No data	No data	0.0007	5			
Benzene	0.05	0.5	5	5	0.01	1	0.2	30			
Benzo(a)anthracene					No data	No data	0.0001	0.5			
Benzo(a)fluoranthene					No data	No data	0.003	0.05			
Benzo(a)pyrene	0.1	0.7	0.7	0.7	No data	No data	0.0005	0.05			

Table C1 (cont.) | Summary Soil Quality Guideline Values

Contaminant	CCME				New Dutch list				US DoE		
	Soil				Soil sediment		Groundwater		Earthworm	Microbial	Plant
	Agri	Res/Park	Comm	Indu	(mg/kgdry weight)		(µg/l)		benchmark	benchmark	benchmark
	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	Target value	Intervention value	Target value Shallow/deep	Intervention value	(mg/kg)	(mg/kg)	(mg/kg)
Organic chemicals											
Benzo(g,h,i)perylene					No data	No data	0.003	0.05			
Benzo(k)fluoroanthrene					No data	No data	0.0004	0.05			
Catechin					No data	20	No data	1200			
Catechol(o-dihydroxybenzene)					0.05	20	0.2	1250			
Chloride					No data	No data	100mg/l[2]	No data			
Chrysene					No data	No data	0.003	0.2			
Cresol					0.05	5[d]	0.2	200			
Cyclohexanone					0.1	45	0.5	15000			
DDT (2,2-Bis(p-chlorophenyl) -1,1,1-Dichloro disphenyl trichloroethane	0.7	0.7	12	12							
Dichlorobenzene					No data	No data	3	50			
Dichloropropane					0.002	2	0.8	80			
Endosulfan					0.00001	4	0.2ng/l	5			
EOX					0.3	No data	No data	No data			
Ethylbenzene	0.1	1.2	20	20	0.03	50	4	150			
Ethylene glycol	960	960	960	960							
Fluoroanthrene					No data	No data	0.003	1			
Heptachlor					0.0007	4	0.005ng/l	0.3			
Heptachloro-epoxide					0.0000002	4	0.005ng/l	3			
Hexachlorobenzene					No data	No data	0.00009	0.5			
Hexavalent Chromium (Cr(VI))	0.4	0.4	1.4	1.4							
Hydroquinone (p-hydroxybenzene)					0.05	10	0.2	800			

Table C1 (cont.) | Summary Soil Quality Guideline Values

Contaminant	CCME				New Dutch list				US DoE		
	Soil				Soil sediment		Groundwater		Earthworm	Microbial	Plant
	Agri	Res/Park	Comm	Indu	(mg/kgdry weight)		(µg/l)		benchmark	benchmark	benchmark
	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	Target value	Intervention value	Target value Shallow/deep	Intervention value	(mg/kg)	(mg/kg)	(mg/kg)
Organic chemicals											
Indenol (1,2,3-c,d)pyrene					No data	No data	0.0004	0.05			
MCPA					0.00005	4	0.02	50			
Monochloroaniline					0.005	50	No data	30			
Naphthalene	0.1	0.6	22	22	No data	No data	0.01	70			
Organotin compounds					0.001	2.5	0.05 - 16ng/l	0.7			
PAH Total					1	40	No data	No data			
PAH (sum 10) [4,14]					1	40	No data	No data			
Pentachlorobenzene					No data	No data	0.003	1			
Pentachlorophenol (PCP)	7.6	7.6	7.6	7.6							3
Phenathrene					No data	No data	0.003	5			
Phenol(s)	3.8	3.8	3.8	3.8	0.05[d]	40	0.2	2000	30	100	70
Polychlorinated biphenyls (PCBs)	0.5	1.3	33	33							40
Polychlorobiphenyls					0.02	1	0.01	0.01			
Pyridine					0.1	0.5	0.5	30			
Resorein					No data	10	No data	600			
Resorcinol (m-dihydroxybenzene)					0.05	10	0.2	600			
Tetrachlorobenzene					No data	No data	0.01	2.5			
Thiocynates (sum)					1	20	No data	1500			
Toluene	0.1	0.8	0.8	0.8	0.01	130	7	1000			200
Trichlorobenzene					No data	No data	0.01	2.5			
Trichlorobenzene (Tri)					0.1	60	24	500			
Xylene	0.1	1	17	20	0.1	25	0.2	70			

Table C1 (cont.) | Summary Soil Quality Guideline Values

Contaminant	CCME				New Dutch list				US DoE		
	Soil				Soil sediment		Groundwater		Earthworm	Microbial	Plant
	Agri	Res/Park	Comm	Indu	(mg/kgdry weight)		(µg/l)		benchmark	benchmark	benchmark
	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	Target value	Intervention value	Target value Shallow/deep	Intervention value	(mg/kg)	(mg/kg)	(mg/kg)
Chlorinated Hydrocarbons											
1,2-Dichloroethane					No data	4	0.01[d]	400			
1,2,3,4-tetrachlorobenzene									10	No data	
1,2,3-trichlorobenzene									20	No data	
1,2,4-trichlorobenzene									20	No data	
1,4-dichlorobenzene									20	No data	
2,3,4,5-tetrachlorophenol									20	No data	
2,4,5-trichlorophenol									9	No data	4
2,4,6-trichlorophenol									10	No data	
3,4-dichlorophenol									20	No data	20
3-chlorophenol									10	No data	7
4-nitrophenol									7	No data	
Acrylonitrile										1000	
Carbon tetrachloride										1000	
Chlorobenzenes					0.03	30	No data	No data	40		
Chlorodane					0.00003	4	0.02ng/l	0.2			
Chloronapthylene					No data	10	No data	6			
Chlorophenols (total)					0.01	10	No data	No data			
Cis-1,4-dichloro-2-butene										1000	
Dichlorobenzol (total)					0.01	No data	0.01[d]	50			
Dichloromethane					0.4	10	0.01	1000			
Dichlorophenol					No data	No data	0.2	30			
Hexachlorobenzene					0.0025	No data	0.01[d]	0.5		1000	

Table C1 (cont.) | Summary Soil Quality Guideline Values

Contaminant	CCME				New Dutch list				US DoE		
	Soil				Soil sediment		Groundwater		Earthworm	Microbial	Plant
	Agri	Res/Park	Comm	Indu	(mg/kgdry weight)		(µg/l)		benchmark	benchmark	benchmark
	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	Target value	Intervention value	Target value Shallow/deep	Intervention value	(mg/kg)	(mg/kg)	(mg/kg)
Chlorinated Hydrocarbons											
Monochlorobenzene					No data	No data	7	180			
Monochlorophenol					No data	No data	0.3	100			
Nitrobenzene									40	1000	
PCBs (PolyChloro-Biphenyls) total					0.02	1	0.01	0.01[d]			
Pentachlorobenzene					0.0035	No data	0.01[d]	1	20		
Pentachlorophenol					No data	No data	0.04	3	6	400	
Tetrachlorobenzol (total)					0.01	No data	0.01[d]	2.5			
Tetrachloroethane					0.01	4	0.01[d]	40			
Tetrachloroethene					0.002	4	0.01	40			
Tetrachloromethane					0.4	1	0.01[d]	10			
Tetrachlorophenol					No data	No data	0.01	10			
Trans-1,4dichloro-2-butene										1000	
Trichlorobenzol (total)					0.01	No data	0.01[d]	10			
Trichloroethane					0.001	60	0.01[d]	500			
Trichloromethane					0.001	10	0.01[d]	400			
Trichlorophenol					No data	No data	0.03	10			
Vinylchloride (Styrene)					0.01	0.1	0.01	5			
Pesticides											
Aldrin					0.00006	No data	0.009 ng/l	No data			
Atrazine					0.0002	6	29 ng/l	150			
Carbaryl					0.00003	5	2 ng/l	50			
Carbofuran					0.00002	2	9 ng/l	100			

Table C1 (cont.) | Summary Soil Quality Guideline Values

Contaminant	CCME				New Dutch list				US DoE		
	Soil				Soil sediment		Groundwater		Earthworm	Microbial	Plant
	Agri	Res/Park	Comm	Indu	(mg/kgdry weight)		(µg/l)		benchmark	benchmark	benchmark
	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	(mg/kg-1)	Target value	Intervention value	Target value Shallow/deep	Intervention value	(mg/kg)	(mg/kg)	(mg/kg)
Pesticides											
DDT/DDD/DDE (total)					0.01	4	0.004 ng/l	0.01			
Dieldrin					0.0005	no data	0.1 ng/l	No data			
Drins (total)					0.005	4	No data	0.1			
Endrin					0.00004	No data	0.0 4ng/l	No data			
HCH combined					No data	2	No data	1			
HCH-compounds					0.01	2	0.05	1			
α-HCH					0.003	No data	33 ng/l	No data			
β-HCH					0.009	No data	8 ng/l	No data			
γ-HCH					0.00005	No data	9 ng/l	No data			
Maneb					0.002	35	0.05 ng/l	0.1			
Miscellaneous											
1,2-dichloropropane									700		
2,3,5,6-tetrachloroaniline									20		20
2,4,5-trichloroaniline									20		20
2,4-dichloroaniline									100		
3,4-dichloroaniline									20		
3-chloroaniline									30		20
Chloroacetimide									2		
Chloroform					0.02	10	6	400			
Cyclohexanone					0.1	270	0.5	15000			
Dimethylphthalate									200		
Mineral Oil					50	5000	50	600			

Table C1 (cont.) | Summary Soil Quality Guideline Values

Contaminant	CCME				New Dutch list				US DoE		
	Soil				Soil sediment		Groundwater		Earthworm	Microbial	Plant
	Agri	Res/Park	Comm	Indu	(mg/kgdry weight)		(µg/l)		benchmark	benchmark	benchmark
	(mg/kg^{-1})	(mg/kg^{-1})	(mg/kg^{-1})	(mg/kg^{-1})	Target value	Intervention value	Target value Shallow/deep	Intervention value	(mg/kg)	(mg/kg)	(mg/kg)
Miscellaneous											
N-nitrosodiphenylamine									20		
Pentachloroaniline									100		
Phthalates (total)					0.1	60	0.5	5			
Pyridine					0.1	0.5	0.5	30			
Styrene					0.3	100	6	300			300
Tetrahydrofuran					0.1	2	0.5	300			
Tetrahydrothiophene					0.1	90	0.5	5000			
Tribromomethane					No data	75	No data	630			

Appendix D:

Soil sampling and treatment

This appendix describes the sampling design and protocol by which we sampled and prepared soils from the two selected sites (A and B) and carried out the bioassays. These methods were also used in the parallel Environment Agency R&D project 'Review and Application of Sublethal Ecotoxicological Tests for Measuring Harm in Terrestrial Ecosystems' (P5-063), which considers other biological test approaches such as biomarkers and functional endpoints determined *in situ* (Spurgeon *et al*., in press).

We co-ordinated all site activities so as to avoid duplication and maximize the information and understanding across the two projects. Much of the initial scoping activity was desk-based, using historical records (including photographs and maps) of the site(s), along with a site visit. Most importantly, the review included details on the type and quantity of chemical contamination, including soil profile and depth information and any biological effects data available. On the basis of the level of information available for each site, we developed the following work programme.

Background

An Environment Agency R&D report on the development of appropriate sampling regimes (Environment Agency, 2000) highlighted the relative contributions to uncertainty from sampling and field-testing. It acknowledged that the correct use of an appropriate sampling technique reduced uncertainty the most, while the misuse of an appropriate technique and correct use of an inappropriate technique could both increase uncertainty. The greatest uncertainty, though, is likely to be caused by the true variability in soil properties.

Countless studies have looked at the statistical requirements of sampling regimes. Many of these have tried to define the number of sample points necessary to identify hotspots with confidence. This is so as to maximise the likelihood of identifying areas within the site that may contain any of a prescribed list of contaminants at above threshold trigger concentrations, for example the CLEA values or any of the other benchmarks selected for comparison. It is true that the extent of investigation of a site is very much a matter for expert judgement, based on an understanding of the history of site usage, the availability of supporting information, and the sensitivity of future site use. Nonetheless, in order to ensure robustness of approach and results, and to allow comparison (and therefore prioritisation) of sites investigated using ecotoxicological risk assessment tests, some degree of standardisation should be sought.

Criteria relating to sampling strategies (design type, number of samples and replicates, etc.) are much the same for ecotoxicity testing as for traditional analytical chemistry sampling. In general, the greater the number of samples (and the larger the number of replicates used for each sample point), the greater the accuracy (and therefore the robustness) of the conclusions drawn. Often, it is possible to distinguish between the advantages of maximising the number of sample points used (assisting the establishment of the extent of the contamination threshold), and maximising replication at a single sample point to allow more robust (and therefore defendable) links between contamination profile and receptor response.

The degree of confidence that can be associated with the results of a sampling regime rely on several factors. These include the number of sampling points tested within a site, the number of samples tested within defined sampling point, the spatial layout of the samples and the frequency and duration of any subsequent monitoring.

To identify a circular hotspot with an area of one square metre with 95 per cent accuracy, at a site with a total area of 100 square metres, more than 100 samples would need to be taken, using a grid approach. Therefore, the sample numbers required to characterise a site fully would be great and

potentially resource-intensive. This is obviously one major caveat with this project, because we have the resources for only 10 biological sampling points. This makes the project conform very much with issues about the suitability of the biological tests rather than issues of site heterogeneity which, due to the constrained sampling size, are impossible to address.

Soil sampling

The sampling strategy adopted in the current project depended on many features, not least the nature of the selected sites, the prior history of site use and the likely contaminants of concern.

Soils are subject to chemical, physical and biological changes as soon as they are collected (for example, fermentation, oxidation or carbonation changes occur, volatile substances are lost). Possible changes will be considered. Sampling conditions have been designed accordingly to limit the effects of such changes on the results of the ecotoxicity tests. Where possible, all sampling will adopt and use these draft protocols (outlined above). This is how one might collect soil from a site, depending on the nature of the soil.

Project soil sampling protocol

We carried out site-sampling design with reference to relevant British and International Standards (BS 10175:2001 - BSI 2001 and ISO/DIS 10381-5 - ISO 2002) and informed by the Tier 0 and Conceptual models developed early in the project.

At each sampling point, we removed turf to the base of the grass root area. The turf was placed to one side and replaced once the excavation had been filled with uncontaminated topsoil. We dug soil out of the exposed surface to a maximum depth of 25 centimetres, using a spade, discarding large bits of debris and root material. Soil taken from four one-square-metre sample locations within one metre of a central foci was mixed thoroughly on-site, then placed into labelled, double-lined plastic bags and taken to the laboratory within 12 hours. All collected material was stored in the dark at 4°C (Institute of Petroleum, 1993).

We examined the soils visually and recorded any observations. We took general photographs of the field sites. Sample sites were located relative to known points (position recorded) and marked so that we could re-locate them later.

Historical (1998) chemical sampling data were available for Site B. The heterogeneity of contaminant concentrations on this site was likely to be small due to the importation of a clean topsoil cap to encourage natural recolonisation. Furthermore, a period of time had elapsed since the chemistry data were generated. The subsequent land disturbances during remediation probably resulted in a more homogeneous distribution of surface contaminants (at least in the surface layers). Therefore chemistry data was provided showing the nature, extent and degree of soil contamination at that time, but there was no information on the current levels of contaminants present. The sampling locations were therefore selected on the basis of targeted or judgmental sampling patterns (that is, not in a regular pattern, but where a specific source of contamination is known (or at least suspected) and confirmation is required). Hence a set of 11 trial areas were sampled, soil collected and TPH determined. The premise for selecting these sites was the anticipated gradient of low, intermediate and high soil TPH concentrations predicted to be present.

Soil preparation followed the recommendations of BS 10175:2001, ISO/DIS 10381-5 and the requirements of the requisite analysis. Note that no OECD, ISO or BSI guidelines have been located for the preparation of contaminated field soils for use in OECD or ISO laboratory-based tests. In summary, the soil collection details adopted are described below for the two sites A & B (for site B the drying stage for collected material was omitted).

Collection and on-site treatment

At each designated sampling point, we marked out areas of one square meter and dug the soil (with a spade) from the four corners of a marked central square (reserved for *in situ* testing in the SubAssess project). In each corner, a 0.5 m^2 was excavated to a maximum depth of 25 cm, providing four samples of 20-litre volume (we removed large stones etc. at this point). Collected samples were mixed on-site, then double bagged and individually marked with unique sample point codes both within and on the outside of the bags. If a turf layer was present at the sampling point, this was removed (collecting the soil from the root-mat) and the soil excavated to 16 centimetres below the root-mat. Soils were taken to the laboratory in refrigerated containers. This on-site mixing of the four soil sub-samples is permitted under the BSI standard, which indicates that mixed samples may be used, depending on the nature of the investigation. Mixed samples can be used to enhance the representativeness of the samples and may be considered in a number of specific situations.

Samples taken for future chemical analysis received minimum treatment: no drying, and storage at -20°C to prevent the loss of volatile organic compounds through volatilisation, degradation etc.

Pre-experimental treatment (off-site)

The pooled collected samples from each sampling point were sieved through a 10 millimetre sieve and mixed until homogenous. For Site A, any soil aggregates were broken up while still damp, resident fauna removed and the soil placed in an oven at 60°C until dry (where there was concern about the loss of volatile organic compounds, freeze sterilisation techniques were used). Dried soils (from Site A) were then sieved through a two millimetre mesh and re-wetted for bioassay use. After sieving, a sample of soil was taken for analysis of pH, percentage loss on ignition (%LOI), and maximum water-holding capacity and field capacity.

For all macro-species tests (Collembola, earthworms and plants) it was necessary to prepare the soils before they could be added to the experimental units and test organisms (such as plant seeds and earthworms) introduced. This was because soils in the experimental units must be a) homogenous with regard to particle size and b) at a standardised moisture content.

To achieve standard moisture content, it was necessary partially to air-dry the soils, prior to re-wetting. This enabled the sample to be subsequently brought up to a biologically acceptable proportion of total WHC (Water Holding Capacity), usually 60 per cent.

Finally, soils were left to stabilise under ambient conditions one week before testing.

This approach ensured the comparability of soils for the biological and chemical tests conducted on sub-samples of the soil from each sampling point at a specific time period. Soils apportioned for each test were prepared and stored in the dark at 4°C (further sieved, re-moistened, etc.) for experimentation, according to the respective test standard guideline.

We acknowledged that, during these processes, a large proportion of some of the more volatile (that is, below C_{10}) fractions may have been lost from the samples. We analysed soils from each sampling point from the two sites at the time of collection. We took a further set of analyses at the start of the longer-term tests (for example, earthworm, Collembolan, plant test) (or in the case of the artificially mixed TPH soil range, when it was prepared (see below). A further set of soils (not treated) but held in similar conditions to those of the biotests was analysed after 28 days - the average duration of all our tests combined.

Site B

After receiving the chemistry data for the 11 sampled points at Site B, we discovered that the actual Total Petroleum Hydrocarbon (TPH) concentrations were low to intermediate, with two very high values. The Ministry for Environment (MEF) (1998) gives some benchmarks for human health. In fact, these are the only soil screening values available for TPH (mg/kg) as <100, (Background) 700 (residential/ commercial uses) and 3600 (industrial/ commercial uses). The TPH soil values measured included two very high TPH concentrations, 14,300 and 34,400 mg/kg. It also included two low intermediates (including one that was presumed to have very low levels of contamination and thus had been designated as a control), 160 and 320mg/kg, and a range of low to very low values, from 8.7 to 36.3 for total hydrocarbons. The adoption of these test soil concentrations would have prevented us testing a useful range of soil test concentrations.

We therefore mixed the soil from the highest concentration (34,400 mg/kg) with a blend of the low concentration range soils to prepare a range of soils of very similar comparative composition. Such approaches were used successfully by Schaefer (2000). By creating a dilution series (keeping five of the original soils collected), we prepared a range from high to low TPH (mg/Kg wet wt) values of; Viz. 14,300; 7250, 3500, 1,600, 700, 320, 160, 12 (control) and 8 (control). (Viz. B9 to B1).

This produced a full range of soil exposures in a fairly consistent soil matrix. It resulted in a range comparable to the various published values for soil ecotoxicity effects data (for example, Salanitro *et al*., 1997, Dorn *et al*., 1998).

Appendix E:

Costs of biological testing

Table E1 | Estimated costs of commercial biological testing

Test	Duration	Range of potential commercial test costs (GBP)	Comment
Microtox™	Can be less than 30 minutes	500-1000 (may be considerably lower in cost if large numbers of samples are processed simultaneously)	Additional costs likely to be encountered for solid-phase tests (as development/ training in many laboratories anticipated
Nitrogen Mineralisation test	Can be up to 100 days - typically 28 days	500-3500	Costs vary according to the test duration and its variability
Carbon Mineralisation test	Can be up to 100 days - typically 28 days	500-4000	Costs vary according to the test duration and its variability
OECD germination and shoot growth test	Timing dependent on season and time taken for seeds to germinate	2500-4000	Costs may vary with duration
OECD Earthworm test (acute)	14 days (plus setting up time)	1000-3500	Same effort required as for longer term test
OECD Earthworm Test (chronic)	Two stages each 28 days in duration	2000-3500	Routine test undertaken by many testing laboratories - but using a standardised soil. Few laboratories have experience of natural non - standard soils
ISO Springtail test	28 days but requires sufficient numbers of synchronised aged animals at the start	2500-3500	Very few commercial laboratories in a position to undertake this test

CONTACTS:

THE ENVIRONMENT AGENCY HEAD OFFICE

Rio House, Waterside Drive, Aztec West, Almondsbury, Bristol BS32 4UD.
Tel: 01454 624 400 Fax: 01454 624 409

www.environment-agency.gov.uk
www.environment-agency.wales.gov.uk

enquiries@environment-agency.gov.uk

ENVIRONMENT AGENCY REGIONAL OFFICES

ANGLIAN
Kingfisher House
Goldhay Way
Orton Goldhay
Peterborough PE2 5ZR

SOUTHERN
Guildbourne House
Chatsworth Road
Worthing
West Sussex BN11 1LD

MIDLANDS
Sapphire East
550 Streetsbrook Road
Solihull B91 1QT

SOUTH WEST
Manley House
Kestrel Way
Exeter EX2 7LQ

NORTH EAST
Rivers House
21 Park Square South
Leeds LS1 2QG

THAMES
Kings Meadow House
Kings Meadow Road
Reading RG1 8DQ

NORTH WEST
PO Box 12
Richard Fairclough House
Knutsford Road
Warrington WA4 1HG

WALES
Cambria House
29 Newport Road
Cardiff
CF24 0TP

ENVIRONMENT AGENCY
CUSTOMER SERVICES LINE
08708506506

ENVIRONMENT AGENCY
F L O O D L I N E
0845 988 1188

ENVIRONMENT AGENCY
EMERGENCY HOTLINE
0800 80 70 60

ENVIRONMENT AGENCY

EP261-TLV Printed on Cyclus Print

www.environment-agency.gov.uk

We welcome feedback including comments about the content and presentation of this report.

If you are happy with our service please tell us. It helps us to identify good practice and rewards our staff. If you are unhappy with our service, please let us know how we can improve it.

For further copies of this report or other reports published by the Environment Agency, contact general enquiries on 0878 506 506 or email us on enquiries@environment-agency.gov.uk

ENVIRONMENT AGENCY

SCHO0804BICW-E-P