ORNL/TM-13408

Environmental Sciences Division

A Framework for Assessing Ecological Risks of Petroleum-Derived Materials in Soil

Glenn W. Suter II

Environmental Sciences Division
Publication No. 4666

Date Published: May 1997

Prepared for
A. B. Crawley
U.S. Department of Energy
Bartlesville Project Office

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831-6285
managed by
LOCKHEED MARTIN ENERGY RESEARCH CORP.
for the
U.S. DEPARTMENT OF ENERGY
under contract number DE-AC05-96OR22464

CONTENTS

CONTENTS (continued)

Page

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND ABBREVIATIONS

AEC	ambient exposure concentration
ASTM	American Society for Testing and Materials
EPA	U.S. Environmental Protection Agency
CERCLA	Comprehensive Environmental Response, Compensation and Liability Act of 1980
COPEC	Chemical of Potential Ecological Concern
EDB	ethylene dibromide
EDC	ethylene dichloride
EqP	equilibrium partitioning
ERA	Ecological Risk Assessment
HQ	hazard quotient
NOEC	no-observed-effects concentration
RBCA	Risk-Based Corrective Action
PAH	polycyclic aromatic hydrocarbon
QA/QC	quality assurance /quality control
QSAR	Quantitative Structure Activity Relationship
RGO	Remedial Goal Option
TEC	toxicologically effective concentration
TIE	toxicity identification evaluation
TPH	total petroleum hydrocarbons

ACKNOWLEDGMENTS

This document benefited from discussions with Mike Harrass and Charles Menzie and from review comments by Rebecca Efroymson, Brad Sample, and Art Stewart. Research sponsored by Office of Fossil Energy, [DOE/FE AC 10 20 00 0], U.S. Department of Energy under contract number DE-AC05-96OR22464 with Lockheed Martin Energy Research Corp.

This project was administered through the Department of Energy's Bartlesville Project Office, by A. B. Crawley, targeting research needs of the Petroleum Environmental Research Forum.

ABSTRACT

Ecological risk assessment estimates the nature and likelihood of effects of human actions on nonhuman organisms, populations, and ecosystems. It is intended to be clearer and more rigorous in its approach to estimation of effects and uncertainties than previously employed methods of ecological assessment. Ecological risk assessment is characterized by a standard paradigm that includes problem formulation, analysis of exposure and effects, risk characterization, and communication with a risk manager. This report provides a framework that applies the paradigm to the specific problem of assessing the ecological risks of petroleum in soil. This type of approach requires that assessments be performed in phases: (1) a scoping assessment to determine whether there is a potential route of exposure for potentially significant ecological receptors; (2) a screening assessment to determine whether exposures could potentially reach toxic levels; and (3) a definitive assessment to estimate the nature, magnitude, and extent of risks. The principal technical issue addressed is the chemically complex nature of petroleum—a complexity that may be dealt with by assessing risks on the basis of properties of the whole material, properties of individual chemicals that are representative of chemical classes, distributions of properties of the constituents of chemical classes, properties of chemicals detected in the soil, and properties of indicator chemicals. The advantages and feasibility of these alternatives are discussed. The report concludes with research recommendations for improving each stage in the assessment process.

1. INTRODUCTION

1.1 BACKGROUND

Ecological risk assessment estimates the nature and likelihood of effects of human actions on nonhuman organisms, populations, and ecosystems. It is intended to be clearer and more rigorous in its approach to estimation of effects and uncertainties than previously employed methods of ecological assessment. It is characterized by a standard paradigm with the following characteristics (Barnthouse and Suter 1986; Suter 1993; EPA 1992):

- Separation into a problem-formulation phase, a phase analyzing exposure and effects, and a risk-characterization phase.
- Clear formulation of the problem, including clear assessment endpoints, well-defined relationships between measurements and the assessment endpoints, and an explicit conceptual model.
- Characterization of risk based on both the magnitude and the effects of the exposure. This ensures that decisions are not made of the basis of an exposure characteristic alone (e.g., detectable levels or levels above background) or on the basis of effects alone (e.g., banning all endocrine disrupters).
- Separation of risk assessment from risk management. The risk manager contributes to the problem formulation to ensure that issues relevant to the decision are addressed and then receives the results of the risk characterization. This clear division of roles serves to prevent two extremes. On the one hand, if the risk manager is intimately involved in the performance of the assessment, there is a tendency to bias the analysis to support a preferred alternative. On the other hand, if the risk manager is not sufficiently involved, the results of the risk assessment may be largely irrelevant to the decision.

This framework includes the standard paradigm and discusses alternative methods for carrying out each step in the assessment. The version of the ecological risk paradigm used here is more elaborate than the one in the U.S. Environmental Protection Agency (EPA) framework document (Fig. 1). Because the EPA framework was not designed primarily for assessment of contaminated sites, it does not specifically address issues such as determining the source of contamination or developing an assessment plan (EPA 1992).

In contrast, the American Society for Testing and Materials (ASTM) has developed a “streamlined” method for Risk-Based Corrective Action (RBCA) for petroleum-release sites that is said to protect human health and environmental resources (ASTM 1994). However, its treatment of ecological risks is limited to comparison of aqueous concentrations of petroleum compounds with water quality criteria. While RBCA’s focused approach to assessment efficiently addresses human health risks, the complexity of ecological risks requires a broader assessment approach. Ecological risk assessment must address a variety of receptors (e.g., plants and animals); levels of organization (individual, population, and ecosystem), responses (mortality, reproduction,

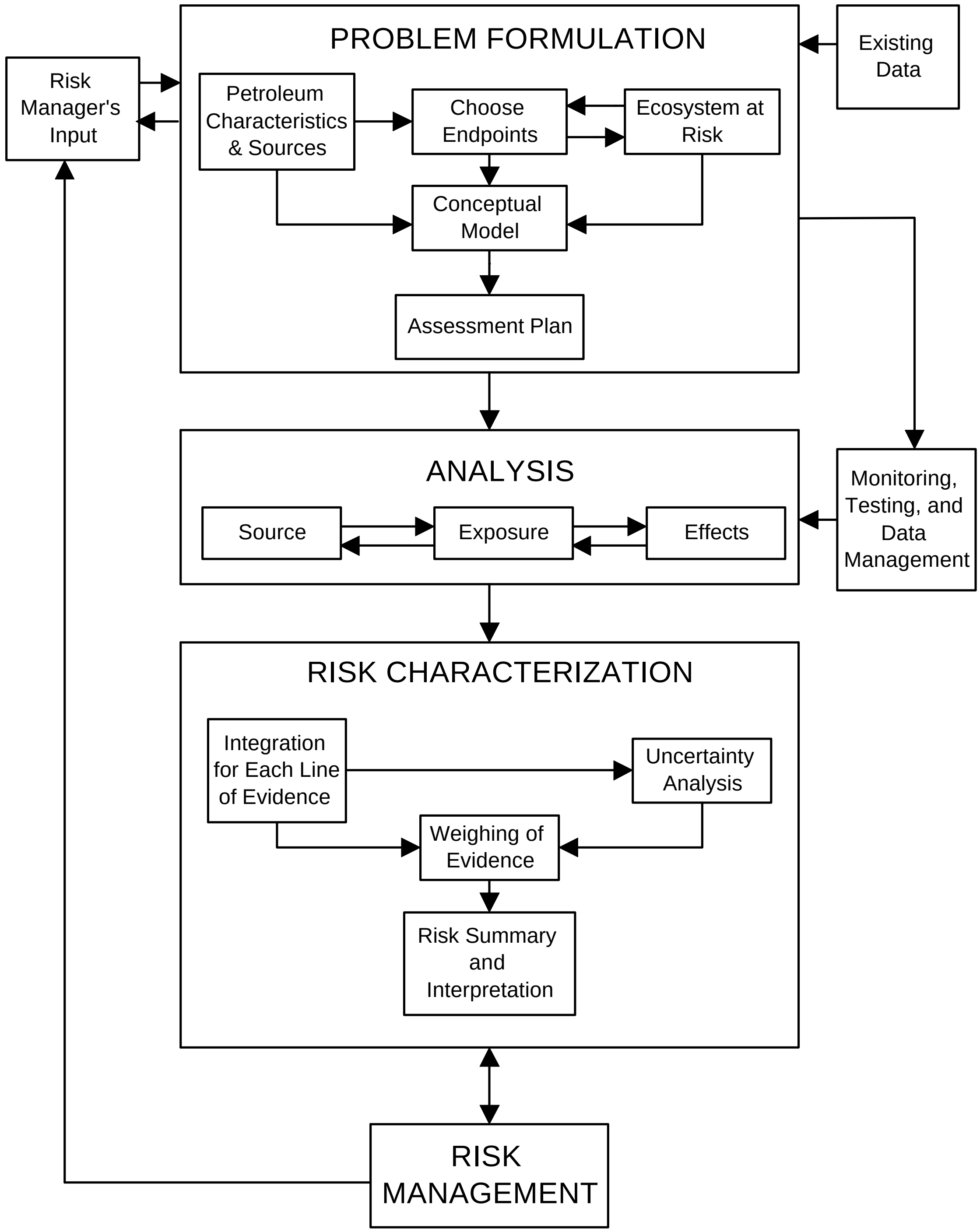

Fig. 1. Diagrammatic framework for ecological risk assessment of sites contaminated by petroleum materials.

and ecosystems processes); and routes of uptake (e.g., root uptake, grooming, soil ingestion, and food-web transfer). Standard ecological assessment endpoints, methods, and assumptions are not available to support a "streamlined" standard method like RBCA. However, where issues relevant to ecological risk assessment are addressed in RBCA, the RBCA guidance is cited.

Human health–based assessments like those conducted in RBCA might be sufficient if ecological endpoints were always protected by decisions based on risks to humans. However, that is not the case (Suter 1993). Ecological receptors are often more sensitive than humans because they are more exposed. They typically spend their entire lives on a particular site, obtaining all food, water, and air from the contaminated area. In addition, ecological receptors have modes of exposure that do not occur in human exposure models such as direct oiling followed by grooming or preening, drinking oily sump waters, or root uptake. Ecological receptors also have modes of effects induction that do not occur in humans such as loss of the insulating properties of fur or plumage. Finally, even when modes of exposure and effects are the same, some nonhuman receptors are likely to be inherently more sensitive than humans simply because there are so many nonhuman species.

Note that the discussions of multiple endpoints, methods of analysis of exposure and effects, and methods for risk characterization do not imply that all of them should be applied to every site. The multiplicity of alternatives results from the following facts: (1) different endpoints and methods are applicable to different sites, (2) sufficient research to determine the relative utility of alternative methods is not available, and (3) multiple methods may be used to create independent lines of evidence that can increase confidence in conclusions.

1.2 PURPOSE

This framework is intended to serve three purposes. First, it provides guidance on performing ecological risk assessments for petroleum-contaminated sites. These assessments may be performed to determine the need for remediation, the adequacy of completed remediation, or the nature and level of injuries to natural resource for which damages must be paid. Second, it provides a basis for the development of soil quality criteria based on ecological risks. The development of risk-based criteria is effectively a process of assessing the risks of contaminant levels to determine what level is, with adequate confidence, a threshold for significant effects on the specified assessment endpoints. Third, this framework identifies research needs by pointing out data sets and assessment tools that are needed but unavailable.

1.3 SCOPE

This framework is limited to risks to ecological endpoints of petroleum and petroleum products in soil. It does not include risks to humans that are mediated by ecological receptors, and it does not include aquatic ecological effects. Although oil spilled on soil may contaminate surface waters, relatively well developed techniques exist for assessing risks to aquatic systems. The materials considered include crude petroleum, petroleum fractions, refined and formulated products, and petroleum wastes such as sludges and tank bottoms.

1.4 ASSESSMENT PHASES/TIERS

In general, Ecological Risk Assessments (ERAs) are performed in phases or tiers including scoping assessments, screening assessments and definitive assessments. Each successive tier requires more time and effort than the preceding one. Scoping assessments determine whether an ecological risk assessment is needed, screening assessments determine what needs to be assessed, and definitive assessments determine the nature and magnitude of risks.

Scoping assessments ascertain whether a formal ecological risk assessment is needed by determining whether a potential for current or future exposure of ecological receptors exists. The first question to be answered is as follows: are there currently, or might there be in the future, ecological receptors on the contaminated site? In some cases, no complete pathway exists from the contaminants to an ecological receptor because the contamination is limited to an industrial facility or an inactive industrial site that will be returned to industrial use or converted to commercial use (i.e., a brownfield). In such cases the site currently has little ecological value and is not expected to have significant ecological value under future land uses. No complete pathway exists on the site because there are no significant receptors. The next question is this: could movement of components of the material result in significant exposure of ecological receptors off-site? The principal concern is with contamination of surface waters and wetlands through runoff or lateral groundwater movement. It is assumed that groundwater organisms are not potential endpoints, because protection of groundwater communities is not normally a basis for regulatory action in North America. Therefore, groundwater contamination is not normally a basis for performing an ecological risk assessment unless the groundwater intersects the surface.

Screening assessments go beyond scoping assessments by asking whether the identified pathways from the contaminants to ecological receptors could result in toxicologically significant exposures. Screening allows assessors to focus resources by first applying rapid and conservative assessment techniques to exclude sites, portions of sites, media, or chemicals that clearly pose minimal risks. More time and effort can then be devoted to definitive assessments that provide risk estimates by applying more realistic and site-specific analyses to those hazards that have not been excluded. Chemicals that are retained by the screening process are termed Chemicals of Potential Ecological Concern (COPECs). Equivalent screening could be applied to chemical fractions or classes rather than individual chemicals (Sect. 3).

The screening phase may itself be performed in tiers. For example, a screening assessment may be performed by using a relatively small set of preliminary data for the purpose of determining what type of sampling and analysis needs to be performed. The new data may then be used in a screening assessment to determine which hazards should be modeled and analyzed in detail in the definitive assessment.

In general, data gaps identified in screening assessments should be treated as a basis for including a hazard. For example, if a class of chemicals has not been measured in a medium and potentially occurs in the source, it should be included as a COPEC. Similarly, receptors that may occur on the site but are unconfirmed and routes of exposure that are credible but have not been investigated should be retained. Treating unknowns in this manner will result in a more credible assessment.

If screening assessments have identified credible hazards to ecological receptors, a definitive assessment must estimate risks and identify preliminary remedial goals. Definitive assessments should replace conservative assumptions with best estimates of exposures and effects and associated uncertainties. Because previous tiers of assessment reduce the scope of the assessment by identifying contaminants, media, and receptors that constitute credible hazards, it should be possible to devote sufficient time and resources to their assessment. In general, additional testing and analysis reduce uncertainties and increase realism, thereby reducing the need to overremediate in order to ensure protection. Definitive assessments are not normally performed in tiers unless the assessment process reveals potentially significant pathways and receptors that were missed by the screening assessments.

The phased assessments performed at contaminated sites require a different logical approach from the tiered assessment schemes developed to assess risks from new chemicals. The latter schemes were based on the performance of brief and inexpensive tests and simple exposure models in the first tier and performance of more expensive and realistic tests in higher tiers if assessments based on lower tiers were inconclusive (Cairns et al. 1979; Urban and Cook 1986). However, the acute lethality tests used in the early tiers of those assessment schemes cannot be used to determine acceptability of risks at a contaminated site. For example, if a soil from a contaminated site is acutely lethal to plants, one can conclude that there are significant risks. However, if the soil is not acutely lethal, one cannot conclude that no potentially significant risks exist. The soil may in fact be lethal in longer exposures or may reduce growth or seed production. In hazard assessment schemes, the insensitivity of traditional early-tier tests is compensated for by applying safety factors to the test endpoints [(e.g., median lethal concentrations (LC_{50} values)] and by using conservative assumptions in the transport and fate models. However, when testing contaminated media, those techniques are not applicable. Therefore, one must begin with a sensitive test in order to avoid falsely concluding that there are no significant risks.

Tiered assessment schemes for ERA are more complex than those for human health risk because of the multiple lines of evidence that are available to ecological assessors. For example, RBCA assumes that the assessment is performed by comparing modeled or measured exposure levels for individual chemicals in air, soil, and groundwater to levels that constitute thresholds for significant human risk. In RBCA, three tiers are defined by use of (1) generic values, (2) easily derived site-specific values, and (3) values derived by complex and extensive site-specific analyses to estimate exposure. Ecological risk assessors perform similar analyses but also employ a variety of modeled or measured exposure levels for individual chemicals in air, soil, and groundwater to levels that constitute thresholds for significant human risk. In RBCA, three tiers are defined by use of (1) generic values, (2) easily derived site-specific values, and (3) values derived by complex and extensive site-specific analyses to estimate exposure. Ecological risk assessors perform similar analyses but also employ a variety of toxicity tests of contaminated media from the site and various surveys of biota at the site. Surveys range from analyses of body burdens and biomarkers to surveys of species composition and abundance. Any of these lines of evidence may be employed in any tier of the assessment, depending on the availability of data, the availability of time and resources to obtain data, or the appropriateness to the source and receiving environment. This makes development of standard tiers, like those in RBCA, a difficult proposition.

2. PROBLEM FORMULATION

Problem formulation is a critical phase of ERA that is often neglected. Assessment scientists are often tempted to begin generating data without carefully thinking through the assessment problem and defining how the data will be used in decision making. Problem formulation organizes existing information concerning the characteristics of the source and site, defines the endpoints for the assessment, defines the temporal and spatial scope of the assessment, and organizes the information into a conceptual model. This is one of the two points of interaction between the assessment scientists and the risk manager described in the EPA framework (EPA 1992). The risk manager's role is to ensure that the assessment problem is defined in such a way as to ensure that the results of the assessment will be useful for decision making. If a standard assessment scheme like RBCA were available for ERA of petroleum- contaminated soils, adoption of the scheme would resolve many issues a priori, but even in RBCA there are many detailed and specific issues that would need to be resolved by consultation with the risk manager. In the absence of such a scheme, every effort should be made to intensively involve risk managers in the problem formulation and to elicit their concerns and needs as clearly as possible.

2.1 STRESSOR CHARACTERISTICS—PETROLEUM CHARACTERISTICS AND SOURCES

In general, petroleum and its constituents have low toxicity relative to other classes of chemicals that raise environmental concerns such as pesticides, heavy metals, and chlorinated diaromatic hydrocarbons. In addition, most petroleum constituents are relatively nonpersistent, and petroleum hydrocarbons do not tend to biomagnify through food webs. However, significant ecological effects of petroleum and related materials occur because the mass of material that is produced, transported, and used can result in very high levels of exposure due to spillage, leakage, and disposal. At high exposure levels, conventional toxic effects may be less important than the effects of the physical properties of oils coating plants and animals and coating and saturating the soil. In addition, some components of petroleum such as polycyclic aromatic hydrocarbons (PAHs) and heterocyclic compounds are moderately toxic and persistent and can accumulate to hazardous levels at locations where heavy fractions of petroleum have been deposited.

The characterization of petroleum products is complicated because they exist as highly complex mixtures. A preliminary characterization should describe the petroleum product in terms of the type and form that entered the environment, the manner in which it entered, the amount released, the spatial and temporal characteristics of the release, and treatment of the site (Table 1). Even if considerable time has elapsed since the release, this information provides important background for the assessment because it allows the assessor to determine the expected composition of the released material and even estimate its concentrations. In higher-tier assessments, results of chemical analyses permit characterization of the composition of the contaminated soil (Sect. 3).

Table 1. Initial characterization of the contamination of soil by petroleum products in terms of the characteristics called for by EPA[a] plus treatment of the site

(Scenarios are (A) wreck of a tank truck carrying gasoline and (B) an oily waste land farm)

Characteristic	Tank truck sceranio	Land farm scenario
Type	Gasoline, unleaded	Oily sludges and tank bottoms
Duration	Instantaneous	8 years
Frequency	Single event	Irregular, approximately monthly
Timing	April 20, 1995, ground thawed and vegetation emerging	Year-round, from September 1985 to August 1990
Scale	Surface flow covered 700 m^2	4000 m^2
Treatment	None, gasoline volatilized or was absorbed by the soil before reaching surface water	Area was tilled after each addition of waste

[a]U.S. EPA 1992.

Modes of release must also be specified. Cases of oil contamination of soil can be classified as acute or chronic. Acute releases are isolated events that occur in a short time period, such as spills from tanks or pipes or blowout of a well. Chronic releases occur over an extended period due to repeated events—such as land disposal of oily wastes (e.g., land farms, oiling of dirt roads, or dumping of used crankcase oil)—or to an event that releases oil over an extended period such as leaks from tanks or pipes. Releases may be direct to surface soils, or soils may be indirectly contaminated by groundwater or by surface runoff to seeps, wetlands, or floodplains.

2.2 DEFINING THE RECEIVING ENVIRONMENT—ECOSYSTEM POTENTIALLY AT RISK

The description of the receiving environment should provide the information needed to complete the problem formulation, justify the conceptual model, and support the selection of techniques for assessing exposure and effects. It should not be a compilation of everything known about the site. The environmental description for the first-tier assessment is based on a preliminary site survey. Later-tier descriptions are based on information gathered in the preceding tiers.

The site description should include the following:

Location and scale—Particular attention should be paid to the spatial extent of the assessment. This area may be larger than the immediately contaminated area because of spreading of the material or its components, movement of animals contaminated on the site, or concern for the broader consequences of effects on the site (e.g., population-level consequences of effects on individuals in the contaminated area).

Physical features—Physical features of the site that are relevant to the assessment should be described. These include slope, locations and hydrologic characteristics of surface water bodies, standing water or other evidence of wetland status, or evidence of past flooding.

Biota – The biotic community of the site should be described in general terms (e.g., woody shrub stage of old field succession, approximately 8 years following agricultural use). The description should also describe the dominant plant species of the site. Any species of special concern such as federal- or state-listed threatened or endangered species that are observed or that could occur on the site based on their range and habitat requirements should be noted.

Soil—Soil properties including texture, porosity, organic matter content, nutrient content, and depth that are relevant to the fate and effects of petroleum should be described as completely as possible.

Land use and disturbance—The risks from contaminants in soil depend on the current and future state of the ecosystem. The uses of the land and planned or projected future uses should be specified (e.g., pasture, oil field, residential, or wildlife refuge). If the land is physically disturbed by the release, by actions taken to control the release, or by unrelated activities (e.g., agricultural tillage), the disturbance should be described.

2.3 CHOOSING ASSESSMENT ENDPOINTS

Assessment endpoints are the explicit expression of the environmental value to be protected (EPA 1992; Suter 1989, 1993). It is the ecological equivalent of the lifetime cancer risk to a reasonable maximally exposed individual in human health risk assessments. Therefore, it must be something that is important and can be estimated, not a vague goal, such as achieving healthy ecosystems. The selection of the assessment endpoints depends on a knowledge of the receiving environment and of the contaminants, provided by the assessment scientists, as well as the values that will drive the decision, provided by the risk manager. A completely specified assessment endpoint for any ecological risk assessment that measures effects includes an entity such as a vascular plant community, a property of that entity such as net production, a level of effects to be detected such as 15% reduction relative to reference communities, and a desired degree of statistical confidence such as 20%. All assessment endpoints should at least specify the entity and property. Criteria for selection of endpoint entities and properties (EPA 1992; Suter 1989; Suter 1990a) are found in the numbered list that follows. Classes of potential assessment endpoints are then discussed.

1. Policy Goals and Societal Values—Because the risks to the assessment endpoint are the basis for decision making, it is important that they reflect the policy goals and societal values that the risk manager is expected to protect.

2. Ecological Relevance—Entities and properties that are significant determinants of the properties of the system of which they are a part are more worthy of consideration than those that could be added or removed without significant system-level consequences. Examples include a keystone predator species or the process of primary production.

3. Susceptibility—Susceptible entities are those that are potentially highly exposed and responsive to the exposure.

4. Operationally Definable—Without an unambiguous operational definition, it is not possible to determine what must be measured and modeled in the assessment and the results of the assessment are too vague to be balanced against costs of regulatory action or against countervailing risks.

Soil ecosystem properties—Given the importance of soil as a biogeochemical system supporting all terrestrial life, it would seem obvious that assessment endpoints for contaminated soils should include appropriate soil properties. However, it is not self-evident which properties are appropriate. Many of the properties that change in soils following contamination with petroleum, such as reduced nutrient availability and changes in the relative abundance of microbial taxa, are results of biodegradation, a desirable process. In other words, many of the changes occur because the oil acts as an organic substrate as well as a toxicant. As a result, many of the soil processes and properties that have been proposed as test endpoints would not be appropriate (Health Council of the Netherlands 1991; Suter 1981). For example, soil respiration increases as petroleum degrades and net nitrogen mineralization is reduced due to immobilization. These effects can mask any toxic effects on mineralization of native organic carbon and nitrogen. In addition, to most decision makers and stakeholders, the soil is a “black box” that is acceptable if it supports plants and animals. Therefore, soil properties are less likely to be drivers for decision making than other potential assessment endpoints.

Plant properties — Plant production is one of the clearest and most generally accepted assessment endpoints for contaminated soils. The biological and societal importance of plant production is clear. Moreover, plants have a scale of exposure that is appropriate to contaminated sites: plants do not wander out of the contaminated area, and many contaminated sites are large enough to encompass a population of herbaceous plants. Although plants do not appear to be particularly sensitive to soil contaminants on average, their sensitivity is not well predicted by other receptors and they are highly sensitive to some chemicals. Various other properties might be used for the assessment endpoint (e.g., mortality or species richness); however, the common use of tests of plant growth suggests that production is the endpoint property.

Properties of soil fauna—Soil invertebrates are ecologically important in terms of soil structure and nutrient cycling and as food for wildlife. They are sensitive to soil contaminants due to their intimate contact with and consumption of the contaminated soil, and, like plants, they have an appropriate scale of exposure. Their societal significance is less clear. A review of bases for regulatory decisions by the EPA found that aquatic and benthic invertebrates, fish, birds, mammals, reptiles, amphibians, and plants were considered but that soil invertebrates and microorganisms were not (Troyer and Brody 1994). Therefore, if risk managers are willing to make remedial decisions on the basis of effects on soil invertebrates, they are appropriate assessment endpoint organisms. The appropriate property is less clear. The common use in the United States of earthworm survival, growth, and reproduction as test endpoints suggests that the assessment endpoint is population abundance or production of earthworms, or of all invertebrates as represented by earthworms. The Dutch have used protection of 95% of species of soil invertebrates as an endpoint (van Straalen and Denneman 1989) as well as survival, production, and abundance of earthworms and collembolans (Health Council of the Netherlands 1991).

Properties of vertebrates—Mammals and birds are commonly used assessment endpoints for contaminated terrestrial sites, and effects on these organisms are linked in the public mind with oil spills. However, vertebrates in general are less ecologically important than plants, invertebrates, and microbes, and they typically have an inappropriate scale for contaminated sites: that is, all bird populations and many other vertebrate populations have much larger ranges than contaminated sites. Even individual vertebrates often have ranges that are much larger than the contaminated areas. As a result, the susceptibility of vertebrates is often low if risks are realistically assessed because the exposure is diluted over the entire range of organisms and the effects are diluted over the range of the population. Shrews and moles are potentially important exceptions because they have relatively small ranges and high dietary and direct exposures. Terrestrial salamanders and burrowing anurans and reptiles are also potentially highly sensitive, but their responses to chemical exposures are poorly known and there are no standard toxicity tests for them. Commonly used endpoint properties for terrestrial vertebrates include survival of individuals and abundance or production of populations.

2.4 CONCEPTUAL MODEL

Conceptual models summarize the results of the problem formulation and guide the analytical phase of the assessment. They are working hypotheses about how the hazardous agent or action may affect the endpoint entities (Barnthouse and Brown 1994; EPA 1992; Suter et al. 1994). Typically a conceptual model includes a graphical representation of the entities involved and the processes that link them. It also includes a narrative that describes those entities and processes plus aspects of the problem formulation that are not included in the graphic such as the spatial and temporal limits.

A conceptual model should be developed for each distinct risk scenario. There are three classes of scenarios that may require assessment. First is the acute scenario, which involves the immediate effects of direct exposure to released oil: smothering of the soil community, oiling of wildlife and subsequent ingestion during grooming, etc. This scenario is generally not the subject of assessment at particular terrestrial sites, because such releases are usually followed by time-critical attempts to recover the oil or minimize its spread. By the time an assessment is performed, the acute risks are past. However, assessments of this type of scenario should be included in the planning of facilities (e.g., the location of tank farms or routing of pipe lines), in the design of response procedures (e.g., to ensure that damage from the remedial actions does not exceed the benefits), and in the assessment of injury for Natural Resource Damage Assessment under the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA). The second scenario is the current baseline case for a remedial investigation. Assessments of this scenario are performed after the emergency response (if one has occurred) and are intended to determine whether the nature and magnitude of current risks are sufficient to justify remedial actions. Finally, future scenarios may be assessed. They are assessed (1) if risks could possibly increase in the future due to changes in exposure (transport) or in the receptors (occupation of the site by species not currently present) or (2) if the decision hinges on the expected rate of recovery.

A generic conceptual model that is appropriate for current baseline or future scenarios is shown in Fig. 2. The model would be implemented at a particular site by identifying specific contents of the cells such as appropriate receptor taxa, by adding processes or entities that are not generally appropriate but may be applicable to particular circumstances (e.g., predators), and by eliminating processes and states that are not relevant to the site or the decision (e.g., contamination of soil by waterborne petroleum components where that route is unlikely to be significant). The model is initiated by a release of petroleum to soil and is structured in terms of transport and uptake of the petroleum either by direct exposure to the soil or indirectly through air, food chains, or aqueous transport. The only cell of the model that is not self-explanatory is “Soil Properties.” This includes all of the changes in the state of the soil due to interactions of the petroleum with the microbial community and the nonliving components of the soil. Hence it includes loss of petroleum components due to microbial metabolism, generation of intermediate metabolites, immobilization of nutrients, reduction in soil oxygen levels, etc. All of these changes have an impact on the state of the soil, which affects future petroleum transport and exposure as well as the suitability of the soil for other processes.

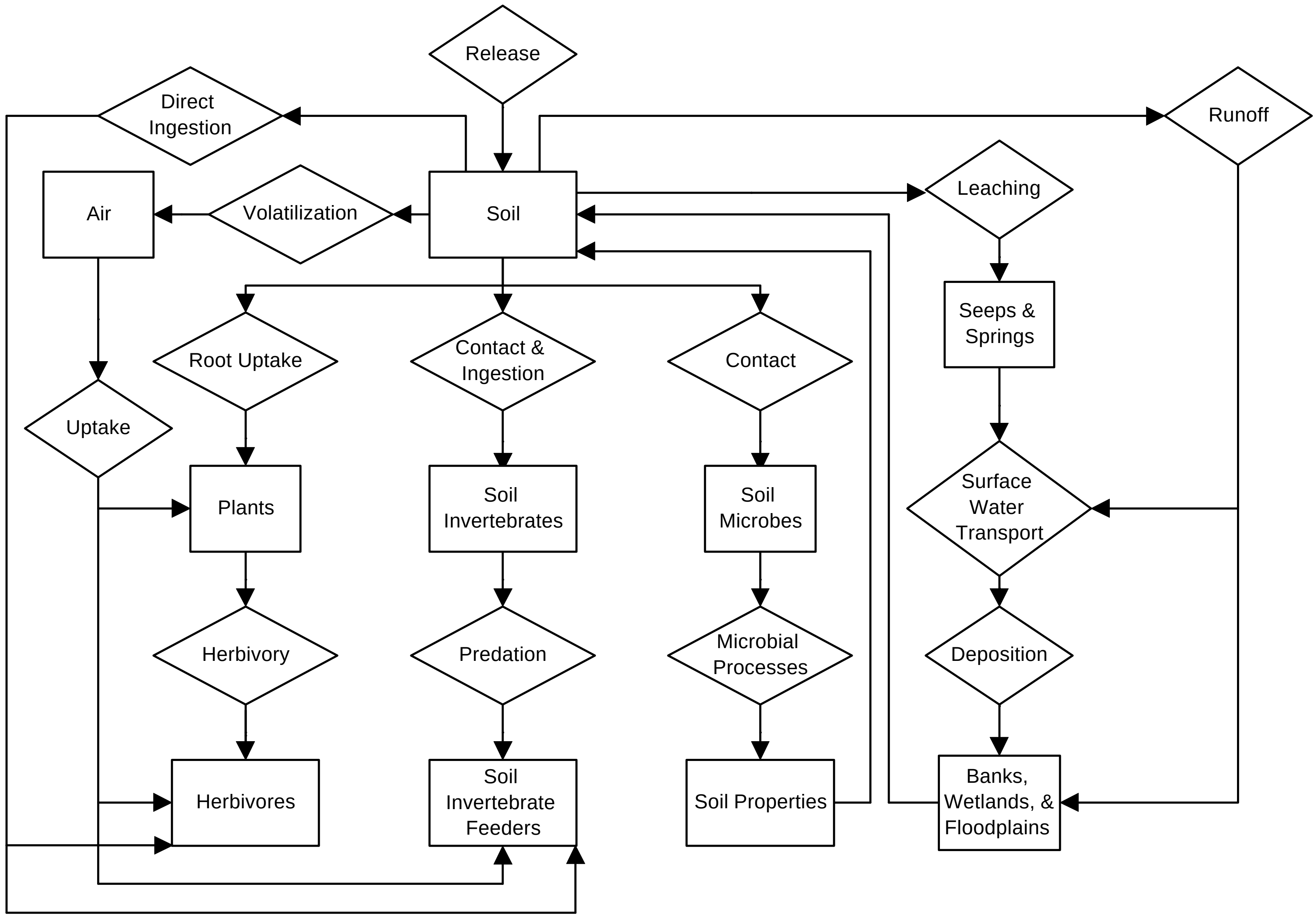

Fig. 2. A conceptual model of the direct and indirect contamination of soil by petroleum materials and subsequent exposure of ecological receptors. Rectangles represent entities, and diamonds represent processes.

2.5 ASSESSMENT PLAN

The product of the problem formulation is a plan for conducting the assessment, including what sampling, analysis, testing, and measurement will be conducted and how the data will be used to estimate risks. It should include specific sampling, analysis, and measurement methods that will be used to characterize effects, sources, and exposure as well as the models that will be used to relate those measures to each other and to estimate risks. The assessment plan should include a technique for ensuring the quality of the data, including plans for data management. It should also specify why the measurements are needed and how they will be used in the assessment. Therefore, the development of the assessment plan requires that the assessors plan the analytical and risk-characterization phases of the assessment so that data needs are specified.

Measures of effects or measurement endpoints are statistical or arithmetic summaries of observations used to estimate the effects of exposure on the assessment endpoint (EPA 1992; Suter 1989; Suter 1993). They include test endpoints such as median lethal dose (LD_{50}) values or dose-response functions and summaries of field measurements such as catch per unit effort or mean density. The distinction between assessment and measurement endpoints is needed because the endpoint for a set of measurements should not simply be adopted as the endpoint of the assessment. The property measured is at best an estimate of the property to be protected (e.g., a mean aboveground standing biomass) and it is often a related effect that must be extrapolated to the assessment endpoint (e.g., median effective concentration (EC_{50}) for lettuce seedling growth).

Measures of exposure must also be specified during the problem formulation. Most commonly, in studies of contaminated soils, measures of exposure are in the form of summaries of concentrations of contaminants such as hectares of soil with concentrations greater than some prescribed value. However, alternative expressions are available (Sect. 3). As with measures of effects, there may be policy considerations as well as technical restraints on the measures of exposure. For example, the best estimate of the exposure of soil biota may be concentrations in an aqueous extract, but regulators often prefer the conservatism of using total extractable concentrations. Specification of the measures of exposure should include the media, constituents, limits of detection, and enough information about the needed spatial and temporal coverage and desired level of precision to allow the statistical design of the sampling and analysis plan.

3. ANALYSIS OF EXPOSURE

Analyses of exposure should be carried out in such a way as to allow risk characterization: that is, the exposure estimates should be appropriate for characterizing risks by paramaterizing the exposure variables in the exposure-response models. This requires that the exposure estimates address the same forms or components of the petroleum as the effects assessment and also have concordant dimensions. For example, estimation of effects on plants may require that concentrations of chemicals in the aqueous phase of the soil be estimated, that concentrations be averaged over the rooting depth of the plants on the site, and that the results be expressed as a median concentration and other percentiles of the distribution of point concentrations. In contrast, estimation of risks to wildlife due to ingestion may require total concentrations in surface soil, averaged over the foraging range of the species, expressed as the mean and standard deviation.

The degree of detail and conservatism in the analysis of exposure depends on the tier of the assessment. Scoping assessments need only determine qualitatively that an exposure may occur by a prescribed pathway. Screening assessments must quantify exposure but should use conservative assumptions to reduce the likelihood that a hazardous exposure is inadvertently excluded. Definitive assessments should treat the estimation of exposure and uncertainty separately by estimating distributions of exposure or by estimating both most-likely exposure and upper-bound exposure.

Analysis of exposure for petroleum-contaminated soil can be divided into a set of distinct questions. First, how can chemical mixtures be analyzed so as to generate useful fate and exposure properties (Sect. 3.1)? Second, how can soils be analyzed to provide useful estimates of exposure (Sect. 3.2)? Third, might analyses of biota be used to estimate food-web exposure or internal exposure (Sect. 3.3)? Fourth, might bioassays substitute for chemical analyses (Sect. 3.4)? Fifth, how might exposure and uptake be modeled from current concentrations (Sect. 3.5), and how might future exposures be modeled (Sect. 3.6)?

3.1 CHARACTERIZATION OF PETROLEUM-DERIVED CHEMICAL MIXTURES

The preliminary characterization of the released material should be followed by analysis of the petroleum-derived chemicals in soil. Both the composition and concentration of the material should be determined. While the composition of the released material may be known, the weathering process may be quite rapid; as a result, the composition is likely to have changed considerably by the time an assessment is conducted. The problem is deciding how to analyze the mixture in a way that both adequately characterizes the mixture and provides a basis for characterizing the fate, transport, and toxicological properties of the mixture. Three approaches are proposed: analysis of the whole material, analysis of chemical classes, and analysis of individual chemicals (Table 2).

Table 2. Methods for analyzing chemical mixtures and characterizing their physical, chemical, and toxicological properties.

Mixture analysis	Property characterization
Whole material	Whole-material properties
Chemical classes	Properties of representative chemicals
	Distribution of properties of class
Individual chemicals	Properties of detected chemicals Properties of indicator chemicals

The simplest approach is to analyze the whole material by determining total petroleum hydrocarbons (TPH), oil and grease, or some equivalent metric of gross contamination. TPH provides little basis for performing a risk assessment because it supplies limited information about the composition and, therefore, the properties that determine potential fate and toxicity of the material. However, it may be useful for determining the extent of contamination or the locations of the greatest contamination.

Rather than attempt to characterize the whole material, one may characterize constituent classes of compounds. Hydrocarbons are divided into aliphatics and aromatics; long- and short- chain aliphatics; one-, two-, three- and more-than-three ring aromatics, etc. In addition to hydrocarbons, petroleum contains metals; nitrogen-, sulfur-, and oxygen-containing organics; and other compounds. Finally, petroleum products may contain several classes of additives such as oxidants; scavengers; and, in old wastes and spills, organolead compounds. Analyses of these classes provides considerably more information than TPH or indicator chemicals. However, in order to model transport and transformation of these chemical classes or to determine their toxicity, they must be associated with concentrations of particular chemicals. This can be done by identifying representative compounds for each class. The representative compounds should be selected based on the following criteria:

- They are an abundant member of the class of chemicals in the petroleum product being assessed.
- Data are available concerning their environmental fate and effects.
- They have greater-than-average (screening assessments) or average (definitive assessments) toxicity for their class and greater-than-average (screening assessments) or average (definitive assessments) persistence and bioavailability.

Once representative chemicals are selected, the assessment can be performed by assuming that the entire mass of each class of chemicals is made up of that representative chemical.

If fate or effects data are available or can be estimated for several chemicals in a class, a more sophisticated approach can be used. The statistical distribution of the fate and effects properties of the members of a chemical class can be used in the assessment to represent the distribution of the properties in the entire class. For example, if water solubility values for several short-chain aliphatic hydrocarbons are found in the literature, a distribution fit to those data is an estimate of the distribution of water solubility for the class. Alternatively, if Quantitative Structure Activity Relationships (QSARs) are available to estimate fate and effects properties, they could be used to estimate the parameter distributions for the class from the physical properties or structural characteristics of the individual chemicals in the class. For example, the water solubility of hydrocarbons can be estimated from their structure (Lyman et al. 1982). Therefore, by specifying the structures of short-chain aliphatic hydrocarbons, one can estimate the solubilities of all members of the class, and the distribution of those individual solubilities is the solubility distribution for the class. If the relative abundances of the class constituents can be estimated, the distribution can be refined by weighting the observations (e.g., the individual solubility estimates).

Finally, a total analysis of the petroleum materials can be performed. This would include the hydrocarbons and other organic and inorganic compounds found in petroleum, as well as organic and inorganic additives (e.g., oxygenates) and the various chemicals that may occur in mixed wastes (e.g., drilling fluid components). This analytical approach offers the greatest flexibility in that all of the various exposure metrics previously discussed can be reconstructed from a total analysis. In addition, if the petroleum-contaminated site has been contaminated by other materials, a thorough chemical analysis may identify nonpetroleum causes of toxicity. However, it is likely that the majority of individual chemicals in any particular petroleum material will have unknown toxicity. Therefore, the risk characterization must be based on a subset of the detected chemicals.

One approach to characterizing the properties of the mixture from analysis of individual chemicals is to identify indicator chemicals (ASTM 1994). These chemicals are assumed to account for the major risks from the mixture: that is, if risks from those chemicals are acceptable, then the risks from the whole mixture are acceptable. The ASTM recommends using benzene and benzo(a)pyrene [and, in cases where they are present, ethylene dibromide (EDB) and ethylene dichloride (EDC)] because of their carcinogenicity but also presents other potential indicator chemicals. Clearly, other criteria would need to be used for ERA.

Another sort of indicator chemical approach is the analysis of chemicals that are characteristic of the petroleum product. For example, the exposure of kangaroo rats to petroleum from a well blowout was confirmed by analyzing liver tissue for a set of PAHs that are characteristic of oil and coal (Kaplan et al. 1996). Vanadium, which is typically concentrated in petroleum, was found to be elevated in San Joaquin kit foxes from oil fields relative to values for foxes from other sites, suggesting that the foxes from oil fields were exposed to petroleum in some form (Suter et al. 1992). Chemical fingerprinting, or comparison of the distributions of abundance of PAHs, was used to determine whether hydrocarbons found in biological samples from the vicinity of the *Exxon Valdez* spill were from the spilled oil, diesel fuel, or analytical artifacts (Bence and Burns 1995). These approaches are useful for confirming that exposure has occurred, that exposure is associated with a particular source, or that petroleum is being transported by a particular route. However, they are not adequate exposure metrics for ERA because they cannot be related to exposure-response relationships to predict toxicity.

3.2 CHEMICAL ANALYSIS OF SOIL

In addition to determining the method for characterizing the composition of the mixture, one must decide how to characterize the forms in which the components occur within the complex soil matrix. The choice of method determines in turn the methods that are used to characterize exposure at the site and the appropriate expressions of exposure in the toxicity tests (Table 3). Characterization of exposure requires choice of a method of characterizing the mixture from Table 2 and a method of characterizing its occurrence in soil from Table 3.

Table 3. Alternative methods of soil analysis and associated methods for estimation of exposure and toxicity.

Soil analyses	Estimate of exposure	Exposure in effects test
Total extractable analyses	Total extractable concentration	Total extractable concentration in test soil
	Modeled solution-phase concentration	Modeled solution-phase concentration in soil test
	Normalized concentration	Normalized test soil concentration
Aqueous extract analyses	Extract concentration	Tests in solution culture
		Modeled solution concentration

The most direct and commonly used approach for estimation of exposure to soil contaminants is collection and analysis of bulk soil. Rigorous extraction techniques permit the estimation of total concentrations. These analyses have the reassuring feature of including everything. However, because organisms do not extract chemicals so thoroughly, results of rigorous extractions tend to overestimate exposure. An advantage of these exposure estimates is that the results can be compared with those obtained in similar analyses of soils used in toxicity tests or, with much less accuracy, to nominal soil concentrations from tests of spiked soils.

The major limitation of total analyses of soil is that it is not very predictive of toxicity. Because of variation in soil properties that control the availability of chemicals to organisms, total concentrations in different soils or even the same soil at different times may result in very different levels of effects. This issue is termed “bioavailability.” The following approaches to exposure estimation are different ways of dealing with bioavailability—either by estimating the bioavailable component of soil contaminant concentrations or by producing an estimate of exposure that is better correlated with toxicity.

One approach is to measure total concentrations in bulk soil and then estimate concentrations in a bioavailable compartment. This approach is employed by the EPA in the derivation of sediment quality criteria (EPA 1993). Neutral organic compounds (which include nearly all organic components of petroleum) are assumed to be in equilibrium between the aqueous phase (pore water) and the organic component of the solid phase. This same approach has been proposed for soils (Lokke 1994). If one assumes that exposure occurs solely to the aqueous phase, the estimated pore water concentrations can then be used with toxicity data based on aqueous toxicity tests (plants in hydroponic solutions, invertebrates on blotter paper, or even aquatic organisms in water). This equilibrium partitioning (EqP) approach remains controversial when applied to sediments and is largely hypothetical for soils. The situation with soils is more complicated than that with sediments: variation occurs in water content of soils, which leads to saturation and other nonequilibrium dynamics. Fewer assumptions would be required if the EqP models were used simply to normalize soil concentrations: that is, responses in soil toxicity tests expressed as a function of estimated pore water concentrations could be used with estimated pore water concentrations from contaminated site soils to generate more accurate estimates of effects.

Some method of normalization other than EqP between aqueous and organic phases of soils may be more effective. For example, the Dutch reference values for various chemicals in soil were derived by normalizing to organic matter and clay content using linear regression (van Straalen and Denneman 1989). For example, the reference value for cadmium (R_{Cd}) was

$$R_{Cd} = 0.4 + 0.007(c + 3o) ,$$

where *c* is percentage of clay and *o* is percentage of organic matter. Some organic chemicals could be normalized with organic matter alone. If it can be shown that effective exposure concentrations for petroleum constituents are a function of a set of soil properties, it would be possible to normalize soil concentrations across test soils and site soils.

Another approach is to perform aqueous extractions of soil that are designed to simulate the extraction processes of organisms: that is, the extract concentrations approximate bioavailable concentrations. Appropriate procedures would depend on the organisms for which exposure is being estimated—for example, relatively mild extractions for root uptake and stronger or sequential extractions for uptake by a mammalian gastrointestinal system. Although many extraction procedures have been proposed, none have been demonstrated to be reliable for a variety of soils and contaminants. These exposure estimates could be compared with similar estimates of exposure from extraction of the soil from toxicity tests. Alternatively, extractions with dilute aqueous salts or acids could be assumed to estimate the concentrations in soil pore water. This would allow comparison with aqueous toxicity test results as in the EqP approach.

3.3 CHEMICAL ANALYSIS OF BIOTA AND BIOMARKERS

Analysis of soil provides a measure of external exposure to contaminants in soil but not internal exposure or exposure through trophic transfers. These require estimates of uptake from soil and transfer between biotic compartments. In the absence of reliable models of uptake and transfer, internal exposures and trophic transfers can be estimated by collecting and analyzing biota from the contaminated site or from laboratory exposures. This approach has the advantage of avoiding the use of highly variable empirical models or unvalidated mechanistic models. However, it can be expensive, and many petroleum components are rapidly metabolized and may not accumulate to detectable levels. One may analyze indicator chemicals, as discussed previously for soils. Indicator chemicals that are persistent are not only more likely to be detectable but are also more likely to accumulate to toxic levels. Finally, one may analyze biochemical biomarkers such as hepatic mixed-function oxidases as surrogates for internal exposure (Huggett et al. 1992). Biomarkers may be detected when the contaminant cannot, but they tend to be nonspecific and to vary with extraneous variables such as the animal's breeding cycle or nutritional state.

Care must be taken to ensure that the analysis is relevant to the assessment. For example, if earthworms are not depurated, the analysis may be dominated by chemicals in the gut contents that have not been incorporated. This may either overestimate or underestimate internal exposure of the worms and dietary exposure by vermivores, depending on whether the uptake factor (organism concentration:soil concentration) is less than or greater than one. Problems arise with estimation of dietary exposure because soil ingestion is often included in wildlife exposure models as a separate route. Consequently, if soil is included in the food compartment, it is counted twice.

3.4 BIOASSAY

Bioassays are measures of biological responses that may be used to estimate the concentration or determine the presence of some chemical or material. It has been proposed that activity of petroleum-degrading microbes be used as a bioassay for bioavailable petroleum constituents (Alexander et al. 1995). A weak form of this hypothesis is as follows: if biodegradation has stopped, no more bioavailable chemical exists to cause toxicity. This idea requires the assumption that biodegradation has stopped because the residue is unavailable rather than because it is resistant. A stronger hypothesis would be that bioavailable concentration is a function of biodegradation rate, therefore, one could estimate exposure from measures of degradation. This idea requires the assumption that availability for degradation by microbes is proportional to availability for uptake by endpoint plants and animals.

3.5 EXPOSURE MODELS

Uptake may be modeled empirically (e.g, uptake factors) or mechanistically (i.e., toxicokinetic models). Although empirical and mechanistic approaches have been developed for uptake of organic chemicals including petroleum components in water by aquatic organisms, uptake from soil has been relatively poorly characterized. In general, development of empirical factors is hindered by the problem of variance in bioavailability discussed previously. Uptake factors developed for soil are highly variable because of the large variance among soils in factors controlling bioavailability. Other measures of soil concentration such as concentrations in aqueous extracts may be more useful for calculating uptake factors, but such measures are rarely used.

Mechanistic modeling depends on an understanding of mechanisms. Vascular plants take up hydrocarbons from the soil and metabolize them (Lytle and Lytle 1987; Trapp and McFarlane 1995). However, large PAHs and other compounds with low solubility and high Henry's law constants are more likely to be taken up from the air than from soil, and material taken up by roots will be poorly transported to aboveground parts (Bromilow and Chamberlain 1995; Wild et al. 1992). This suggests that a mechanistic model of plant uptake and accumulation of petroleum wastes should be multimedia. Such models have been developed for herbicides and could be adapted for petroleum chemicals, but research is needed. Mechanistic uptake models are not available for soil invertebrates, but quasi-mechanistic models involving exposure to soil pore water and partitioning between the pore water and the organism have been used to estimate earthworm bioaccumulation (Menzie et al. 1992; van Gestel and Ma 1988). Studies of mammalian toxicokinetics for human health risk assessment could be applied to wildlife species. Because hydrocarbons are not greatly accumulated in food webs, direct ingestion of contaminated soil is a particularly important pathway for wildlife accumulation of petroleum.

Wildlife exposures, like human exposures, may have multiple significant routes. These may include direct ingestion of soil, ingestion of food items and liquids, and respiratory uptake. Wildlife exposure models and appropriate exposure parameters such as consumption rates are available and potentially applicable to petroleum-contaminated sites (McVey et al. 1993; Sample and Suter 1994).

3.6 MODELED FUTURE EXPOSURE

The prior discussions are based on the assumption that, because the current state of the receiving environment is being assessed, one has the option of choosing whether to measure or model exposure parameters. However, for future scenarios, one must model exposures. The primary demand for modeling is to predict transfer from the contaminated soil to other media, including soils at other locations. A basic set of models for the transport of petroleum constituents from contaminated terrestrial sites is presented as part of the RBCA (ASTM 1994). Modeling may also be required to predict uptake and exposure of organisms that are not currently present. These might include plants if the site is currently devegetated or wildlife if the site is currently industrialized. Finally, modeling is required to predict future states of the site without remediation and under various remediation options. This requires prediction of biodegradation rates as well as the physical transport processes predicted by RBCA and the uptake processes included in the biotic exposure models. Such predictions are highly uncertain, particularly for the more recalcitrant fractions of petroleum (Larson and Cowan 1995; Lyman et al. 1982). It has been suggested that resistance to degradation is due more to physical sequestering of chemicals in soil micropores than to inherent resistance of the chemicals to degradation (Alexander et al. 1995).

3.7 SUMMARY OF EXPOSURE CHARACTERIZATION

Table 4 summarizes the options for exposure characterization. The rows of the table represent optional means of characterizing the petroleum material in terms of the amount of material, constituent classes, constituent chemicals, bioassays, or biomarkers of exposure. Columns list materials in which the measures may be analyzed and models in which the analyses may be parameters. The plus and minus signs are indicative of the practicality and utility of each particular combination. Note that the number of plus signs in a row or column is not indicative of relative utility or importance. That is determined by the ability of each combination of measure and medium to characterize risks to an endpoint at a site (Sect. 5).

Table 4. The utility of combinations of methods of characterizing the material (measures) and media for which they may be characterized as well as their utility in exposure models

Measures	Whole soil	Aqueous extracts	Biota	Modeled exposure
Whole material	+	+	−	−
Chemical class	+	+	+	+
Individual chemicals	+	+	+	+
Bioassays	+	−	−	−
Biomarkers	−	−	+	−

4. ANALYSIS OF EFFECTS

Effects data may be obtained from field monitoring, from toxicity testing of the contaminated soil, and from traditional single-chemical laboratory toxicity tests. The analysis of effects evaluates and summarizes the relevant data in such a way that they can be related to the exposure estimates, thereby allowing characterization of the risks to each assessment endpoint during the risk-characterization phase.

The analysis of effects must determine which of the available data are relevant to each assessment endpoint and reanalyze and summarize the information as appropriate to make it useful for risk characterization. This requires consideration of two issues.

The first issue, what form of measurement endpoint best approximates the assessment endpoint, should have been considered during the problem formulation. However, the availability of unanticipated data and better understanding of the situation after data collection will often require reconsideration of this issue.

The second issue in analysis of effects is expression of the effects data in a form that is consistent with expressions of exposure. Integration of exposure and effects defines the nature and magnitude of effects given the spatial and temporal pattern of exposure levels. Therefore, the relevant spatial and temporal dimensions of effects must be defined and used in the expression of effects. For example, if the exposure is to a material such as unleaded gasoline that persists at toxic levels only briefly in soil, then effects that are induced in that time period must be extracted from the effects data for the chemicals of concern and analysis of effects monitoring data should focus on biological responses such as mass mortalities that could occur rapidly, rather than on long-term average properties.

The degree of detail and conservatism in the analysis of effects depends on the tier of the assessment. Scoping assessments need only determine qualitatively that an effect may occur because a receptor is potentially exposed. Screening assessments must quantify effects but typically define the exposure-effects relationship in terms of a benchmark value, a concentration that is conservatively defined to be a threshold for toxic effects. Definitive assessments should treat the estimation of effects and uncertainty separately by estimating distributions of effects or by estimating both most-likely and upper-bound effects.

4.1 SINGLE-CHEMICAL OR PURE-MATERIAL TOXICITY TESTS

4.1.1 Selection of Toxicity Test Data

In ERAs for contaminated sites, single-chemical or pure-material (e.g., gasoline) toxicity data are usually obtained from the literature or from data bases rather than generated ad hoc. Therefore, it is necessary to select data that are most relevant to the assessment endpoints and that can be used with the exposure estimates. As far as possible, data should be selected to correspond to the assessment endpoint in terms of taxonomy, life stages, responses, exposure duration, and exposure conditions. However, because the variance among chemicals is greater than the variance among species and life stages, any toxicity information concerning the chemicals of interest is potentially useful.

In general, tests in soil and in solution may be useful for assessing risks from soil contaminants. The relevance of tests in soil seems self-evident, but, unless the properties of the test soil are similar to those of the site soil, the toxicity observed in the test soil concentration may be poorly correlated with effects at the site. This variance may be reduced by normalizing the test soil concentrations to match normalized site soil concentrations (Sect. 3.2). Tests conducted in solution are potentially more consistent than those conducted in soil. They may be related to concentrations in soil extracts or estimated pore water concentrations. It has even been proposed that aquatic toxicity test results could be used to estimate the effects of exposure of plants and animals to contaminants in soil solution (Lokke 1994; van de Meent and Toet 1992).

4.1.2 Analysis of Toxicity Test Data

Most of the work of effects analysis is devoted to determination of the relationship between exposure and effects for each chemical or material of concern. In conventional risk assessments, this involves deriving an exposure-response model from laboratory toxicity tests. This requires analysis of the test data to derive a test endpoint and extrapolation from the test endpoint to the assessment endpoint. The extrapolation may be performed in various ways including the following (Suter 1993).

Selection—It may be assumed that the endpoint species, life stages, and responses are equal to those in the most sensitive reported test or in the test that is most similar in terms of taxonomy or other factors.

Safety factors—A test endpoint can be divided by 10, 100, or 1000 to estimate a safe level, as in the EPA review of new industrial chemicals (Zeeman 1995).

Species sensitivity distributions—A percentile of the distribution of test endpoint values for various species can be used to represent a level that would be protective of that percentage of the exposed community.

Regression models—Regressions of one taxon on another, one life stage on another, one test duration on another, etc., can be used to extrapolate among taxa, life stages, durations, etc.

Mathematical models—Toxicodynamic models can be used to estimate effects on organisms from physiological responses, population or ecosystem models, to estimate effects on populations or ecosystems from organism responses.

4.1.3 Quantitative Structure Activity Relationships

Because of the large number of chemicals in petroleum materials, it is unlikely that all components will be individually tested for toxicity to all potential endpoint receptors. Therefore, ecotoxicological QSARs for petroleum-related chemicals are potentially highly useful. As with exposure-related properties (Sect. 3.1), ecotoxicological QSARs could be used not only to predict the properties of individual chemicals but also to predict the distribution of properties for a class of chemicals. Many hydrocarbons have a baseline narcosis mode of action in vertebrates, which is relatively predictable (Hansch and Leo 1995). However, modes of action are less well defined for invertebrates and plants. In addition, chronic effects may result from a different mode of action than the acute lethality, the response that has been used to establish nearly all ecotoxicological QSARs.

4.2 CONTAMINATED-SOIL TOXICITY TESTS

A considerable increase in realism can be obtained by testing the contaminated soil rather than individual chemicals in laboratory media. This can be done in at least three ways. The most direct approach is to cage, pen, or plant organisms along a gradient of contamination or at contaminated and reference sites. This approach, termed “field testing,” is relatively easy for immobile organisms such as plants and more difficult for organisms that are mobile and that forage for food. It is highly realistic in that the organisms are subject to realistic conditions and variation in exposure. However, such studies are subject to the effects of variation among sites in conditions other than contamination and to loss of the study due to vandalism, predation, or extreme conditions. In addition, cage effects may modify the sensitivity of the organisms. An example of field testing is the placement of worms for 7 days in contaminated soil in plastic buckets buried at the locations where the soil was collected (Menzie et al. 1992). Carabid beetles have been tested in field pens on pesticide-contaminated soils (Heimbach et al. 1994). Rodents have been placed in cages or pens at contaminated sites, but pens must be large unless the investigator feeds the animals, thereby eliminating or diminishing effects of dietary exposure (Barrett 1968). Several proposed field testing methods are presented by Linder et al. (1992).

A more common approach is to bring contaminated and reference soil into the laboratory for toxicity testing. This is a very active area of ecotoxicology, and tests have been developed for ambient waters, sediments, soils, and biota. Testing of contaminated soils in the United States is largely limited to earthworms and seedlings of vascular plants. However, various tests have been conducted with soils that might be adapted to use with soils from petroleum contaminated fields (Donker et al. 1994; Linder et al. 1992; van Straalen and van Gestel 1993). In particular, recent research has expanded the range of soil invertebrates used in toxicity testing (Donkin and Dusenbery 1993; Kammenga et al. 1996; van Gestel and van Straalen 1994). Because these tests are performed on the site soil, there is generally no need to consider normalization of the concentrations. However, care must be taken to match reference soils to contaminated soils in terms of chemistry, texture, and nutrient status. Particularly for growth and reproduction endpoints, tests may be highly sensitive to soil properties. Therefore, it is desirable to test soils from multiple reference locations in order to estimate the natural variance.

Finally, the least-used technique is to bring contaminated biota into the laboratory and test them. This technique is appropriate if the contaminant is persistent and bioaccumulated, or if it is known to cause persistent injury. For example, herring eggs from areas exposed to spilled oil and from unexposed areas were brought into the laboratory and their hatching rates and frequencies of abnormalities recorded (Pearson et al. 1995). We know of no use of this technique with organisms exposed to contaminated soil.

As discussed in Sect. 1.3, assessment schemes based on tiers of toxicity testing begin with rapid and inexpensive screening tests. However, the common practice of using acute lethality tests for this purpose is not appropriate for contaminated media. Therefore, it is necessary to develop rapid and sensitive tests for effects of petroleum materials on soil-exposed organisms. Biochemical biomarkers are potentially useful for that purpose but require development (Huggett et al. 1992).

Because the exposure component of the exposure-response analysis for these tests is not different from the exposure analysis for the assessment, the discussion of exposure-response analysis for these tests is deferred to the risk characterization. However, as part of the analysis of effects it is important to consider whether some qualitative or quantitative extrapolation model should be applied to the ambient soil toxicity tests to make them relevant to the assessment endpoint. The types of extrapolation models used with single-chemical toxicity tests are potentially useful for these tests as well.

4.3 BIOSURVEYS

Biological surveys of effects include a wide variety of techniques for enumeration and characterization of biological populations, communities, and ecosystems. In the simplest case, the measurement endpoint for the biological survey is an estimate of the assessment endpoint. In such cases, the effects analysis consists of summarizing the data in such a way as to reveal the relationship of effects to exposure. Examples include plotting the species richness of the soil microinvertebrate assemblage on exposure axes such as kilometers from a source, TPH, or concentrations of a particular chemical. Biosurvey techniques are used less frequently for contaminated soils than for waters or sediments, even though there are fewer inherent difficulties in obtaining samples. Methods for determining injury to soils due to spills of petroleum and other substances were reviewed for use in Natural Resource Damage Assessments (Van Voris et al. 1987).

If the measurement endpoints do not directly estimate the assessment endpoint, then the relationship between them must be characterized. For example, if data are available for stream macroinvertebrates and the assessment endpoint is some property of the fish community, then the relationship between them must be characterized in terms of the trophic dependence of fish on invertebrates, the relative sensitivity of fish and invertebrates, the similarity of their exposure, and other relevant properties.

4.4 INDIRECT EFFECTS

Ecological risk assessments have followed human health risk assessments in emphasizing direct toxic effects. However, because nonhuman organisms are much more subject than humans to indirect effects such as habitat modification and reductions in the abundance of food species, competitors, or predators, indirect effects cannot always be ignored. Indirect effects of petroleum in soil include the usual effects on trophic and competitive relationships as well as the peculiar effects of adding oils to soil including asphyxiation of soil animals due to rapid decomposition, immobilization of nutrients, and filling of soil pores with a nonaqueous liquid. After decomposition is largely completed, plant production may actually be greater due to effects on soil structure, nitrogen availability, or other factors (Bossert and Bartha 1984; McKay and Singleton 1974). These indirect effects, which should have been identified in the conceptual model, should be quantified as far as possible in this component of the assessment. Biological surveys of contaminated areas can potentially reveal indirect effects, but, because the exposures are uncontrolled and unreplicated, indirect effects are difficult to distinguish in such studies. When they are available, the results of microcosm, mesocosm, or field tests can be used to empirically estimate the indirect effects or, for less selective chemicals, the combined direct and indirect effects. Alternatively, simple assumptions can be made: for example, *x*% loss of riparian wetlands due to oiling will result in an *x*% reduction in the abundance of species that depend on that community for any of their life stages. Less common, but more rigorously, ecosystem models may be used to estimate the consequences for all endpoint taxa of toxic effects on all modeled components of the exposed ecosystem (Bartell et al. 1992 ; Emlen 1989; O'Neill et al. 1982; Suter 1993).

4.5 EXPOSURE-RESPONSE PROFILE

The output of the analysis of effects is the exposure-response profile. For individual contaminants of concern, this should indicate how the effects increase with increasing duration and concentration of exposure. It should also, to the extent that such information is available and relevant, indicate the effects of environmental variables such as soil organic matter content and pH on toxic effects. It should indicate the mode of action and the variation in sensitivity among taxa, life stages, and processes.

For ambient soil toxicity tests, the exposure-response profile should summarize the results in terms of the spatial and temporal distribution, the nature and magnitude, and the consistency of toxicity. If more than one test is performed on a contaminated medium, the relative sensitivities of the tests should be explained as far as possible in terms of the relative sensitivities of the species and life stages involved, in terms of the nature and duration of the exposure in the test system, or other relevant factors.

5. ANALYSIS OF SOURCES

For many assessments of petroleum-contaminated soils, the source will have been adequately characterized in the problem formulation. However, in some cases the contaminant may not be characterized. In such cases, it may be appropriate to obtain and analyze samples of the material at the source. In other cases, the source may be unknown and characterization of the source may not only serve in the analysis of risks but also aid in determining responsibility. In these cases, the assessors should seek out potential sources and characterize them. If indicator chemicals or fingerprinting techniques are to be used to associate ambient contamination with the source (Sect. 3.1), then analyses of the sources and the contaminated soil must be coordinated.

6. RISK CHARACTERIZATION

Risk characterization consists of integration of the available information about exposure and effects, analysis of uncertainty, weighing of evidence, and presentation of conclusions in a form that is appropriate to the risk manager and stakeholders (Fig. 1). The integration process should be carried out for each line of evidence independently so that the implications of each are explicitly presented. This makes the logic of the assessment clear and allows independent weighing of the evidence. For each line of evidence, it is necessary to evaluate the relationship of the measurement endpoint to the assessment endpoint, the quality of the data, and the relationship of the exposure metrics in the exposure-response data to the exposure metrics for the site.

6.1 SCREENING ASSESSMENT

Risk characterization for screening assessments consists of using available information to narrow the scope of the assessment by eliminating contaminants that clearly do not constitute a significant risk, soils that are clearly not significantly contaminated, or receptors that are clearly not at risk (Suter 1995). Screening can be performed on a number of bases. First, concentrations in soil can be compared with ecotoxicological screening benchmarks (EPA 1996; Jones et al. 1996; Sample et al. 1996; Suter and Tsao 1996; Will and Suter 1995a; Will and Suter 1995b). Because conservative estimates of exposure concentrations and conservative benchmarks are used, chemicals with exposure concentrations below benchmarks are considered to pose minimal risk.

A second type of screen compares concentrations in nominally contaminated media with background concentrations. Chemicals that do not occur at concentrations exceeding the range of concentrations at uncontaminated sites may be screened out. Although, regulators are often reluctant to use background screening with organic chemicals, natural plant and microbial hydrocarbons and some other constituents of petroleum materials do occur in soils (Alexander et al. 1995; Kaplan et al. 1996). However, in regions where petroleum, coal, or oil shales occur in soils, the occurrence of those hydrocarbons cannot be used to screen out hydrocarbons as background.

A third type of screen compares the chemicals of potential concern against the characterization of the source. For example, investigations of crude oil spills need not assess risks from lead or lead scavengers found in some gasoline formulations and refinery wastes.

These techniques are conventionally used to screen individual chemicals. For petroleum and other complex materials, screening could be performed on representative chemicals or indicator chemicals (Sect. 3.1). In theory, it could be applied to measures of whole material concentration. For example, although the toxicity of petroleum residues is quite variable, it should be possible to establish a conservatively defined safe TPH level in soil that could be used to exclude areas that do not need further investigation or assessment. Because of the greater variability in toxicity, such benchmarks would simply be more conservative than screening benchmarks for more narrowly defined contaminants.

Finally, a logical screen may be performed to eliminate routes of exposure that were part of the preliminary conceptual model but were found not to be appropriate for a site because no complete pathway exists from source to receptor.

6.2 DEFINITIVE ASSESSMENT

A definitive assessment is one that is intended to provide the basis for making a decision concerning remediation. Therefore, it is intended to estimate the nature and magnitude of effects on the assessment endpoints or to determine the likelihood that a prescribed magnitude of effect has been or will be exceeded. Risk characterization for definitive assessments should proceed by integrating exposure with effects to produce a risk estimate for each line of evidence and then weighing the results for all lines of evidence to produce an overall risk estimate.

6.2.1 Integration of Environmental Monitoring Data

The first line of evidence is supplied by integrating the biological survey results with soil analyses and other exposure information. This line of evidence is the most realistic in that it represents the actual state of the environment, but it provides relatively poor evidence of causation. In many cases, the measurement endpoint (e.g., number of earthworms extracted per square meter) will be an estimate of the assessment endpoint property (e.g., earthworm abundance), so no extrapolation will be necessary. It is necessary only to estimate the uncertainty associated with the measurement endpoint. However, in other cases it is necessary to consider the relationship of the measurement endpoint (e.g., fledging success) to the assessment endpoint (e.g., likelihood of population extinction). In this case, the extrapolation could be performed using a demographic model, but in many cases the extrapolation is performed simply by exercising professional judgment. In some cases, monitoring data do not provide a basis for estimating risks, but they provide supporting evidence. For example, small mammal trapping is seldom sufficient to estimate effects on populations, however, if all animals trapped at a site were young of the year or if many animals had pathologies, those findings would tend to support other evidence suggesting the occurrence of toxic effects.

The quality of monitoring data must be evaluated. This includes not only the usual quality assurance and quality control (QA/QC) issues such as whether procedures were followed and whether detection limits were adequate but also more fundamental questions about the appropriateness of the procedures. The quality of biological survey data strictly limits their interpretation. However, even minimal or qualitative information is potentially useful because it may constrain the judgments that can be made concerning the state of the system. For example, a visual survey of a terrestrial site can serve to indicate that it is vegetated, that the dominant species are the same as those on a nearby uncontaminated site, that the density is similar, and that the plants are not visibly injured. Such a finding does not mean that significant phytotoxic effects have not occurred, but they do limit the potential severity of the effects. If such conditions exist on a site where an analysis of the soil indicates that chemicals occur at severely phytotoxic concentrations, then one might reexamine the relevance of the phytotoxicity data or conduct additional studies to determine whether species differences or some other factor in the toxicity tests is applicable to the site.

The risk characterization for the biological survey data must estimate the level of apparent effects and evaluate whether they are real and associated with the contaminants rather than with other environmental factors. If the exposure is treated as categorical (i.e., uncontaminated vs contaminated or discrete classes of contamination), the estimate of effects is the differences in the levels of the endpoint properties between or among the categories (e.g., the difference in the number of species between oiled and unoiled sites). If the exposure is treated as continuous (e.g., a gradient away from a leak), then the effects can be derived from an exposure-response curve as in toxicity tests. Results may be expressed as the estimated level of effects or the likelihood that effects exceed some prescribed threshold. In any case, the issue of the reality of apparent effects must be considered in terms of the possibility that apparent effects are due to sampling error or confounding variables such as habitat difference. The association of the effects with the contaminant consists of defining contaminant concentrations and temporal dynamics for each exposure category or for the exposure gradient. The relationship of the exposure metrics to the effects must not be taken for granted.

The utility of biological survey data is determined by the rate of recovery. If the rate of reduction in toxicity is greater than the rate of recovery of the endpoint biota, then the survey results will reflect prior exposures. For example, a gasoline spill is likely to be acutely lethal to soil invertebrates but dissipates and degrades relatively rapidly. Recolonization of the soil at the spill site by invertebrates, though, is likely to be slow. Therefore, a survey of soil invertebrates 6 months after the spill is likely to find a depauperate community even if the soil is not currently toxic. This outcome is acceptable and even desirable if the goal is to document injury for the sake of determining damages (Van Voris et al. 1987). However, it is misleading if the goal is to determine whether additional remedial actions are needed.

6.2.2 Integration of Soil Toxicity Data

The second line of evidence is the testing of the contaminated soil for toxicity. Most of these are tests of soils that have been brought into the laboratory. The relationship of the test endpoints to assessment endpoints depends on the relevance of modes of exposure in the tests to the field exposures and on the relevance of the test organisms, life stages, and responses to the assessment endpoint. These questions are answered in a generic sense by the validation of media tests against field surveys. For standard tests of ambient waters used in the United States, it has been shown that where toxicity is detected, the species richness of aquatic communities is low (Dickson et al.

1992; Hartwell et al.1995). For tests that have not been validated against field surveys, inferences must be made: for example, reduction in seedling growth in the laboratory is equivalent to reductions in primary production of the plant community or in growth of particular plant species (depending on the form of the assessment endpoint).

The quality of these tests is often limited by the performance of test organisms in reference media. Diseases, contamination from prior releases, low nutrient levels, or unsuitable physical-chemical properties of the soil may cause organisms to die or perform poorly in the tests. In general, it is advisable to use control soil (e.g., potting soil) to determine that the test procedures are adequate and to use reference soil (e.g., soil from adjacent areas) to determine the incremental toxicity of the contaminants of concern.

The relationship of effects to exposure is relatively straightforward if analyses are performed on the tested soil; that is, the effects are caused by the constituents of the tested soil, and those soils are the ones to which organisms in the field are exposed. The principal issue to be addressed is the degree to which the test soils represent the variance over space and time of contamination levels in the field. In general, temporal variance is not a significant issue for soil as it is for water, but spatial variance is a serious problem that can be solved only by good sampling designs.

Interpretation of soil test results is relatively simple: if toxic effects are observed, the soil is toxic to the test endpoint property and species. For most currently used tests, the responses measured are clearly relevant to population- or community-level endpoints. If toxicity is not observed, it cannot be concluded that no significant toxic effects have occurred in the field. As discussed previously, the sensitivity and relevance of the test must be evaluated to avoid false negative conclusions. If toxicity is observed in any tests, the distribution of the effects relative to sources or to contamination levels should be demonstrated: that is, the relationship of exposure to the frequency or intensity of toxic effects should be tabulated and plotted. Exposure may be expressed as categories (e.g., areas of a site with different levels or types of soil contamination), as concentrations of a contaminant marker (i.e., an indicator chemical, a representative chemical, or a whole material measure like TPH), or as spatial gradients (e.g., distance from a source). These exposure-response relationships serve to support the contention that the differences in response among the tests are due to the contaminants. In addition, if contaminant levels are known for more locations than toxicity levels are, the exposure-response relationships can be used to estimate which areas are toxic. This not only provides a more complete description of the magnitude and extent of ecological effects but also can help to guide the design of remedial actions.

It would seem logical that if apparent toxic effects are not correlated with contaminant concentrations, then those contaminants are not responsible for the effects. However, various extraneous variables may mask an exposure-response relationship, as shown in the numbered list that follows. This problem is greatest when the ranges of concentration or toxicity are small, when correlating across sites that are distant and therefore likely to have extraneous differences, and when the response is naturally highly variable with respect to soil properties (e.g., growth).

1. Variation in bioavailability
   - Due to variance in soil characteristics
   - Due to variance in contaminant age among locations
   - Due to variance in transformation or sequestration rates among locations
2. Variation in the form of the chemical (e.g., ionization state)
3. Variation in concentration over time or space (i.e., samples for analysis may not be the same as those tested)
   - Spatial heterogeneity
   - Temporal variability (e.g., rapid changes in composition after release)
4. Variation in composition of the mixture
5. Variation in co-occurring contaminants
6. Inadequate detection limits (even if the chemicals are detected when toxic, no correlation will be found if they are not detected when where there is low or no toxicity)
7. Inherent variation in toxicity tests
8. Variation in toxicity test due to variance in medium characteristics

Soil toxicity tests may also provide evidence concerning which chemicals or chemical classes are responsible for toxicity. If the distributions of contaminants are not too strongly correlated, then relative strengths of correlation of toxicity with concentrations can suggest which components of the ambient contaminant mixture are responsible for toxicity. This may be the case at industrial sites where various petroleum materials and other wastes have been released in various patterns. Better evidence would be provided by toxicity identification evaluation (TIE), a process of selectively removing chemicals from the contaminated medium and retesting to determine what treatments eliminate toxicity. TIE methods have not yet been developed for soil. However, such techniques are well developed for water and are under development for sediments (Ankley et al. 1992).

6.2.3 Integration of Chemical Toxicity Data

Single-chemical toxicity data provide the third major line of evidence. Estimates of effects based on assessment of single-chemical data are in general much more tenuously related to the assessment endpoint and to events in the field than the other lines of evidence. The relation of the standard test endpoints to assessment endpoints should have been addressed in the analysis of effects by applying extrapolation models (Sect. 4.1.2).

The relationship to exposure is commonly expressed as a quotient of the ambient exposure concentration (AEC) divided by the toxicologically effective concentration (TEC), termed the hazard quotient (HQ):

$$HQ = AEC/TEC .$$

The TEC may be a test endpoint, a test endpoint corrected by a factor or other extrapolation model, a regulatory criterion, or some other benchmark value. This type of analysis is commonly used for screening purposes to determine whether particular chemicals are credible contributors to risk and therefore worthy of further assessment. In that case, conservative AEC values are used and an HQ greater than one is treated as evidence that the chemical is worthy of concern (Sect. 6.1).

If exposure, effects, or both are expressed as distributions, the results of risk integration can be expressed as the probability that HQ ≥ 1 (Suter et al. 1983). Distributions of TEC can be derived from the variance on regression models that relate test species and life stages to species and stages of interest, distributions of test endpoints, or probabilistic population or ecosystem models (Barnthouse et al. 1987; Bartell et al. 1992; O'Neill et al. 1982; Sloof et al. 1986; Suter et al. 1983).

6.2.4 Integration of Exposure and Sources

The question of where the contaminants came from may be peripheral to an ERA or a central issue. For example, immediately following the *Exxon Valdez* oil spill, there was little concern for determining the source of oil on marine birds and mammals, because the source was obvious. However, years after the spill it was necessary to consider whether ongoing exposures to petroleum hydrocarbons were due to mobilization of old spilled oil or to ongoing relatively small releases of crude oil from tankers or fuel from other boats (Bence and Burns 1995). The principal approaches to determining sources of exposure are empirical and modeling based.

Empirical approaches to determining sources may be qualitative or quantitative. Qualitative approaches may be as simple as identifying the sole source that could be responsible for a particular exposure, given the distribution of contamination. For complex materials such as petroleum, the contribution of different sources can be quantitatively estimated by fingerprinting. For example, following the *Exxon Valdez* spill, the contributions of that spill, spills of diesel fuel, and other sources were estimated from the relative proportions of PAHs in biological samples (Bence and Burns 1995).

Modeling for source identification and characterization, termed “receptor modeling,” is the logical inverse of modeling to predict the transport and fate of releases of chemicals (Gordon 1988): that is, rather than using release rates to predict exposures, concentrations in biota and media are used to estimate the contributions of potential sources. Such modeling can both identify sources and apportion exposures among multiple sources.

6.2.5 Weight of Evidence

Inferences in ecological risk assessments are made by weight of evidence rather than traditional scientific standards of proof (EPA 1992; Suter and Loar 1992). The traditional standard for inference in science is, in effect, proof beyond a reasonable doubt in a decisive experiment. That standard is embodied in the use of a 95% confidence requirement before a hypothesized phenomenon is deemed to be demonstrated and in scepticism concerning results that have not been replicated in an independent study. Such a standard is appropriate for pure science, which is engaged in adding to the body of reliable knowledge concerning the nature of the world. However, risk assessors do not have the luxury of suspending judgment until a scientific standard of confidence can be met. Decisions are made on schedules that are not within the control of scientists and will be made on other bases if scientific input is not available. Suspension of judgment constitutes an abrogation of responsibility.

Given that one has estimated risks based on each line of evidence, the process of weighing the evidence amounts to determining what estimate of risks is most consistent with those results. If the assessment endpoint is defined in terms of some threshold for significance, then the process can be conducted in two steps. First, for each line of evidence one should determine whether it is consistent with exceeding of the threshold, inconsistent with exceeding, or ambiguous. Second, one should determine whether the results as a whole indicate that it is likely or unlikely that the threshold is exceeded. If the results for all lines of evidence are consistent or inconsistent, the result of the weighing of evidence is clear. Assuming that there is no consistent bias in the assessment, agreement among multiple lines of evidence is strong evidence to support a conclusion. However, if there are inconsistencies, the true weighing of evidence must occur. The weights are determined based on the following considerations (Menzie et al. 1996; Suter 1993).

Relevance — Evidence is given more weight if the measurement endpoint is more directly related to the assessment endpoint. Evidence is relevant if the measures of effects, mode of exposure, and contaminant are relevant.

- Effects are relevant if the measurement endpoint is a direct estimate of the assessment endpoint or if validation studies have demonstrated that the measurement endpoint is predictive of the assessment endpoint. Note that a measurement endpoint based on statistical significance [e.g., a no-observed-effects concentration (NOEC)] is less likely to bear a consistent relationship to an assessment endpoint than one that is based on biological significance (e.g., an EC_x).
- The mode of exposure may not be relevant if the soil used in a test is not similar to the site soil. Normalization of soil concentrations may increase the relevance of a test if the normalization method has been validated. Similarly, the relevance of tests in solution to soil exposures may be low unless the models or extraction techniques used to estimate aqueous-phase exposures have been validated.
- When the measurement endpoints are derived from the literature rather than from site-specific studies, it is necessary to consider whether the material used in the test is relevant to the field contaminant. For example, is it the same petroleum material or a similar material (different unleaded gasolines crude oil and gasoline), and has the weathering of the field contaminant changed its composition in ways that are not reflected in the test?

When the relationship is unclear, relevance may be evaluated by listing the ways in which the results could be wrong because they are fundamentally inappropriate or so inaccurate as to nullify the results. One can then evaluate the likelihood that they are occurring in the case being assessed. For example, single-chemical toxicity tests could be performed with the wrong form of the chemical or in soils that differ from the site soil in ways that significantly affect toxicity; moreover, the tests may be insensitive due to short duration, a resistant species, or the lack of measures of sublethal effects.

Exposure/Response – As in all toxicological studies, a line of evidence that demonstrates a relationship between the magnitude of exposure and the effects is more convincing than one that does not. For example, apparent effects in soil toxicity tests may be attributed to petroleum, but unless the tested soil is analyzed and an exposure-response relationship demonstrated, it may be suspected that effects are a result of other contaminants, nutrient levels, soil texture, or other properties. If an exposure-response relationship has not been demonstrated, then consideration should be given to the magnitude of the observed differences. For example, if soil test data include only comparisons of contaminated and uncontaminated soils, the observed differences are less likely to be due to extraneous factors if they are large (e.g., 100% mortality rather than 25% less growth).

Temporal Scope – One should determine whether the data encompass the relevant range of conditions. For example, if contaminated and reference soils are surveyed during a period of drought, few earthworms will be found at any site and toxic effects will not be apparent. Temporal scope may also be inadequate if the remedial decision is being made long after the collection of data for a particular line of evidence, because exposure levels or other relevant conditions may have significantly changed.

Spatial Scope – It is necessary to determine whether the data adequately represent the area to be assessed—not only the directly contaminated area but also indirectly contaminated and indirectly affected areas. In some cases the most contaminated or most susceptible areas may not have been sampled because of access problems or because of the sampling design (e.g., random sampling with few samples).

Quality – The quality of the data should be evaluated on the following bases: the protocols for sampling, analysis, and testing; the expertise of the individuals involved in the data collection; the adequacy of the quality control during sampling, sample processing, analysis, and recording of results; and any other issues that are known to affect the quality of the data for purposes of risk assessment. Although use of standard methods tends to increase the likelihood of high-quality results, they are no guarantee. Standard methods may be poorly implemented or may be inappropriate to a particular site. In contrast a well-designed and well-performed site-specific measurement or testing protocol can give very high-quality results.

Quantity – The adequacy of the data should be evaluated in terms of the number of observations taken. Results based on small sample sizes are given less weight. The adequacy of the number of observations must be evaluated relative to the variance as in any analysis of a sampling design, but it is also important in studies of this sort to consider their adequacy relative to potential biases in the sampling (see preceding discussions of spatial and temporal scope).

These and other considerations can be used as points to consider in forming an expert judgment or consensus about which way the weight of evidence tips the balance. Table 5 presents an example of a simple summary of the results of weighing evidence based on this sort of process. Alternatively, the considerations can be used to assign a grade to each line of evidence (e.g., high, moderate, or low weight). This still leaves the inference to a process of expert judgment or consensus but makes the bases clearer to readers and reviewers. Finally, a scoring system could be developed that would formalize the weighing of evidence. For example, a numerical weight could simply be assigned to each line of evidence based on quality, relevance, and other factors; a plus or minus sign assigned, depending on whether the evidence is consistent or inconsistent with the hypothesized risk; and the weights summed across lines of evidence. A quantitative system has been developed by a group consisting of representatives of the state of Massachusetts, the private sector, and U.S. government agencies (Menzie et al. 1996). Such systems have the advantage of being open, consistent, and less subject to hidden biases, but they may not give as reasonable a result in every case as a careful ad hoc weighing of the evidence would. Regardless of how the evidence is weighed, it is incumbent on the assessment scientist to make the basis for the judgment clear to readers and reviewers.

The use of quantitative weighing of evidence and equivalent expert judgment about which lines of evidence are most reliable is based on an implicit assumption that the lines of evidence are logically independent. Another approach to weighing multiple lines of evidence is to determine whether there are logical relationships among the lines of evidence. Based on knowledge of site conditions and of environmental chemistry and toxicology, one may be able to explain why inconsistencies occur among the lines of evidence. For example, one may know that spiked soil tests tend to overestimate the availability and hence the toxicity of contaminants and may even be able to say whether the bias associated with this factor is sufficient to account for discrepancies with tests of site soils. As another example, one may know that a chemical form used in toxicity tests is unlikely to occur at the site (e.g., hexavalent chromium is rapidly converted to trivalent chromium in humid soils). Because it is mechanistic the process of developing a logical explanation for differences among lines of evidence is potentially more convincing that simple weighing of the evidence. However, it is important to remember that such explanations can degenerate into “just-so” stories unless the relevance of the proposed mechanisms is well supported.

In general, a logical analysis of the data should proceed from most realistic (i.e., site specific) to most precise and controlled (e.g., single-chemical and species toxicity tests). Field surveys indicate the actual state of the receiving environment; therefore, other lines of evidence that contradict the field surveys (after allowing for limitations of the field data) are clearly incorrect. For example, the presence of plants that are growing and not visibly injured indicates that lethal and gross pathological effects are not occurring but does not preclude reductions in reproduction or growth rates. Those other effects could be addressed by more detailed field studies of growth rates and seed production and viability. The presence of individuals of highly mobile species such as birds indicates almost nothing about risks because dispersal can replace losses of individuals and obscure effects on reproduction. Soil toxicity tests indicate whether toxicity could be responsible for differences in the state of the receiving environment, including differences that may not be detectable in the field. However, field effects are usually more credible than negative test results, because field exposures are longer and otherwise more realistic and site species and life stages may be more sensitive than test species and life stages. Single-chemical toxicity tests indicate which components of the soil contaminants could be

Table 5. A hypothetical summary of a risk characterization by weight of evidence for a soil invertebrate community in petroleum-contaminated soil at an industrial site

Evidence	Result[a]	Explanation
Biological surveys	−	Soil microarthropod taxonomic richness is within the range of reference soils of the same type and is not correlated with concentrations of petroleum components
Toxicity tests	−	Soil did not reduce survivorship of the earthworm *Eisenia foetida*. Sublethal effects were not determined.
Organism analyses	$\underline{+}$	Concentrations of PAHs in depurated earthworms was elevated relative to worms from reference sites but toxic body burdens are unknown
Soil analyses	+	If the total hydrocarbon content of the soil is assumed to be composed of benzene, then deaths of earthworms would be expected. Toxicity data for other detected contaminants are unavailable
Weight of evidence	−	Although earthworm tests may not be sensitive, they and the biological surveys are both negative and are more reliable than the single-chemical toxicity data used with the analytical results for soil

[a]+ indicates that the evidence is consistent with the occurrence of a 20% reduction in species richness or abundance or the invertebrate community; − indicates that the evidence is inconsistent with the occurrence of a 20% reduction in species richness or abundance of the invertebrate community; $\underline{+}$ indicates that the evidence is too ambiguous to interpret.

responsible for observed effects. Because they are less realistic than other lines of evidence, single-chemical toxicity tests are usually less credible than the other lines of evidence for determining whether effects have occurred. They do not include combined toxic effects, the test medium may not represent the site soil or solutions, the exposure may be unrealistic, and the chemicals may be in a different form than those at the site. However, because these studies are more controlled than those from other lines of evidence, they are more likely to detect sublethal effects. In addition, single-chemical toxicity tests may include longer exposures, more sensitive responses, and more sensitive species ad hoc, because they depend on the characteristics of the data and the site.

After the lines of evidence have been weighed to reach a conclusion about the significance of risks to an assessment endpoint, it is usually appropriate to proceed to estimate the nature, magnitude, and distribution of effects. A significant risk is sufficient to prompt consideration of remedial actions, but the nature, magnitude, and distribution of effects determine whether remediation is justified given remedial costs and countervailing risks (Sects. 7.2 and 7.3). In general, it will be clear that one line of evidence provides the best estimate of effects and that is likely to be the most site specific.

6.2.6 Uncertainty in the Risk Characterization

Uncertainties should be listed and, as far as possible, quantified during all phases of the risk assessment. Although the concept of uncertainty is central to risk assessment and much has been written on the subject, risk assessments in the regulatory arena have made little use of formal quantitative uncertainty analysis methods (Morgan and Henrion 1990; National Research Council 1994). The uncertainty in the various components of risk assessments must be estimated using techniques that are appropriate to the data and models. The types and sources of uncertainties have been cataloged in various ways (Suter 1990b). They include natural environmental variability, sampling variance, measurement error, extrapolation error, and model uncertainty.

The analysis of uncertainty should present the uncertainties associated with each line of evidence and the uncertainty associated with the final risk estimate. The estimation of uncertainty associated with a weight-of-evidence analysis is not straightforward. In general, the uncertainty should be less than that of any individual line of evidence. For example, for a case with an assessment endpoint of a 10% reduction in soil invertebrate species richness, the uncertainty concerning the conclusion that effects on the soil invertebrate community are significant would be very small if the following conditions are met: biological surveys show an 80% reduction in the number of soil microarthropod taxa relative to the average of reference sites, more than half of the earthworms die in all toxicity tests of site soils, and five chemicals are found in site soils at concentrations exceeding test endpoints for earthworms. Given that consistency and the fact that effects are estimated to be well above the threshold for significance, it matters little if some component uncertainties are large. For example, the sensitivity of other invertebrates may differ by more than an order of magnitude from that of *Eisenia foetida* (the standard earthworm test species). However, preceding in the example, the biological survey data indicate that the invertebrates in this community were not much less sensitive than *E. foetida*. Therefore, the biological survey data in effect limit the uncertainty concerning the relevance of the earthworm test.

In less ideal cases, the evidence is inconsistent and indicated effects are near the threshold for significance. In such cases, the estimation of uncertainty must be based on the logic used to weigh the evidence. If the conclusion about risk is based on one line of evidence that is judged to be much stronger than the others, then the uncertainty associated with that analysis is the overall uncertainty. If there are conflicting lines of evidence, if there is no logical explanation for the apparent conflicts, and if no one line is strong enough to provide a clear basis for the conclusion, then the overall uncertainty may be higher than the uncertainty associated with the individual lines of evidence. In any case, assessors should at least provide an estimate of uncertainty on a qualitative scale (low, moderate, high, etc.), and identify the most important sources of uncertainty.

6.2.7 Risk Summary and Interpretation

The risk assessor's task does not end with the estimation of risks in term of effects and uncertainties. The process of informing the risk manager and stakeholders is the next critical step. The first step is to prepare an interpretive summary of the risk assessment and its results. This process requires more creativity and effort than simply condensing the risk assessment into an executive summary. Maps, graphs, graphical conceptual models, and other presentations can be helpful.

The second step is interaction with the risk manager. Depending on the relationship between the risk assessors and manager, this may be an informal and friendly process, a formal process of exchanging written comments and responses, or a series of formal and potentially adversarial meetings. However, it is always desirable to view this interaction as an educational one in which the assessors ensure that the risk manager understands the results including the uncertainties and the risk manager is sufficiently open about his level of understanding and his technical and policy concerns to allow the assessors to expand or clarify aspects of the risk characterization to make it more useful.

Presentation of the results of the risk assessment to stakeholder groups or the public is an extension of the risk assessment–risk management interchange in that a democratic public is the ultimate risk manager. Although there is an extensive body of literature concerning communication of risks to the public (National Research Council 1989), this literature does not specifically address the presentation of ecological risks. However, the same advice about an open process of mutual education is applicable to communicating with the public as well as with the designated risk manger.

7. RISK MANAGEMENT

7.1 REMEDIAL GOALS

The risk assessor's primary input to risk management is proposed cleanup criteria termed "preliminary remediation goals" (EPA 1991), "treatment endpoints" (Alexander et al. 1995), "corrective action goals" (ASTM 1994), and similar terms. The term "Remedial Goal Options" (RGOs) used by EPA Region IV is preferable because it emphasizes the fact that reducing risks from contamination to minimal levels is only one of the risk manager's options to consider when setting a remedial goal (Sects. 7.2 and 7.3). The term RGO is also appropriate in that it emphasizes that it may be appropriate for assessors to present multiple options based on different levels of risk, different endpoints, or different definitions of the goal.

Different definitions of RGOs are possible because of the multiple lines of evidence that are available in ERA. Conventionally, RGOs are defined as concentrations of a particular chemical that constitute a threshold for unacceptable risk. Soils with concentrations below the RGO are assumed to be acceptable, but those with concentrations above the RGO may be remediated. Alternatively, RGOs may be defined in terms of soil toxicity tests: that is, one may specify that areas in which some test endpoint is exceeded (e.g., > 20% mortality of earthworms) are candidates for remediation or that areas in which any one of a set of test endpoints is exceeded are candidates for remediation. Finally, one may specify that areas where biological surveys find levels of effects exceeding some measurement endpoint (e.g., dead plants or fewer than five earthworms per meter square) are candidates for remediation. Combinations of these types of RGOs may also be used. For example, to confirm that apparent effects are due to petroleum, one may require that areas to be remediated have both some level of toxicity and some level of an indicator chemical.

7.2 RISK BALANCING

Although regulatory decision making tends to emphasize reduction of risks to some prescribed level, the goal of the risk manager should be broadened to selection of the option that results in the least net injury. This balancing is done at multiple levels.

First, the risks of the remedial actions must be balanced against the risks from the contaminants without remediation. This balancing is needed because remediation may involve tillage; removing the contaminated soil and associated plants, animals, and microbes; and other injurious activities. Conceptually this balancing can be thought of in terms of the time integral of effects. For example, both petroleum and removal of contaminated surface soil may kill all plants on the site; consequently remediation is preferable if succession on the exposed subsoil is more rapid than degradation of the petroleum to nontoxic levels plus succession on the surface soil. Clearly, this depends on the type of ecosystem.

Second, risks to different ecological assessment endpoints may need to be balanced. One endpoint, such as earthworm abundance, may be significantly affected but not others, such as plant or soil microarthropod diversity. Since remedial actions often damage all components of the ecosystem, benefits to some endpoints must be balanced against injury to others.

Third, ecological risks may need to be balanced against human health risks. Remedial actions such as removal of contaminated soil may reduce risks to humans but increase ecological risks. In such cases remedial alternatives such as control of land use until adequate degradation has occurred may be preferable to more ecologically injurious actions.

It is important to recognize that this process of risk balancing requires good risk characterizations. It is not sufficient to state that concentrations exceed a criterion, that risks are significant, or that the soil is toxic. To balance ecological risks against countervailing ecological or human health risks, it is necessary to estimate the nature and magnitude of effects on each endpoint and their temporal and spatial extent. It is also necessary to consider the full life cycle of the remedial actions. For example, removal of contaminated soil, disposal in a landfill, and replacement of the soil with new surface soil requires injury to ecosystems at the contaminated site, the disposal site, and the borrow area.

7.3 COST-BENEFIT

Cost-benefit analysis adds an additional level of complexity to risk management. It is based on the assumption that the best decision is one that, rather than simply reducing risks to an acceptable level or choosing the alternative with the least total risk, ensures that the economic benefits of remediation exceed their cost. Although quantitative cost-benefit analysis is seldom applied to remedial decisions, risk managers make qualitative judgements concerning the cost effectiveness of proposed remediation. Therefore, it is incumbent upon risk assessors to include in the risk characterization a description of the importance and implications of changes in the condition of the endpoint properties.

8. RESEARCH NEEDS

The following is a list of research needs for ecological risk assessment of petroleum contaminated soils. The list is based on published results and therefore does not reflect ongoing research.

Problem formulation—Although assessment endpoints must be defined on a site-specific basis, the potential utility of alternative ecological assessment endpoints should be researched in terms of their ecological importance, societal value, susceptibility to petroleum materials, and practicality. This could lead to development of a set of standard or default assessment endpoints.

Exposure—Potential representative chemicals or indicator chemicals need to be assessed to determine whether they can adequately represent the ecological risk of whole-petroleum materials.

Methods to estimate degradation rates are needed. These need to incorporate both natural processes for unremediated sites and remedial practices intended to enhance degradation.

Methods are needed for estimation of uptake of constituents of petroleum materials from soil and transfer through food webs.

The validity of simple multimedia fate models for predicting the fate of constituents of petroleum materials needs to be tested.

Efficient and risk-relevant methods are needed for analysis of chemicals and chemical classes in soil.

Effects—Safe levels of petroleum-related chemicals or chemical classes (i.e., ecotoxicological screening benchmarks) are need for screening soil contaminants.

Toxicity profiles for petroleum-related chemicals, chemical classes, or materials need to be developed, including relative sensitivities of taxa and life stages, modes of action, and toxicokinetics.

If tiered assessments are to be performed on the basis of tiers of soil toxicity tests, then tests must be developed for the early tiers that are sensitive without being expensive or requiring long exposures. These requirements suggest that the tests need to be based on physiological responses of small organisms.

The relationship between proposed standard soil toxicity tests and effects on potential endpoint taxa and ecosystem processes needs to be investigated and appropriate extrapolation models developed.

New tests need to be developed or adapted for testing petroleum-contaminated soils to better represent effects of long-term exposures on ecological assessment endpoints.

QSARs for effects of components of petroleum materials on various ecological endpoint taxa are needed.

TIE techniques need to be developed to identify the toxic components of soils.

If the large data set for toxicity of chemicals to aquatic biota are to be used to estimate risks to organisms exposed to chemicals in soil solution, the validity of that practice must be determined.

Standard biosurvey techniques need to be developed for contaminated soils to determine whether the receiving community is being affected. These may range from counts of microarthropods ([potentially equivalent to the benthic invertebrate surveys used in streams (Plafkin et al. 1989)]) to measures of biomarkers or histopathologies.

Risk characterization—Proposed procedures for weighing evidence need to be validated.

Methods for presenting the results of weight-of-evidence analysis to risk managers and stakeholders need to be developed.

Appropriate methods for estimating and expressing the uncertainties for different lines of evidence need to be developed.

9. REFERENCES

Alexander, M., L. Goldstein, S. Pauwels, D. Edwards, O. Zaborsky, C. Menzie, W. Heiger–Bernays, C. Montgomery, D. Nakles, R. Loehr, and others. 1995. Environmentally acceptable endpoints in soil: Risk-based approach to contaminated site management based on availability of chemicals in soil. Report GRI-95/0000. Gas Research Institute, Chicago.

Ankley, G. T., M. K. Schubauer–Berigan, and R. A. Hoke. 1992. Use of toxicity identification techniques to identify dredged material disposal options: A proposed approach. Environ. Manage. 16:1–6.

ASTM (American Society for Testing and Materials). 1994. Emergency standard guide for risk-based corrective action applied to petroleum release sites. Report ES 38–94. American Society for Testing and Materials, Philadelphia.

Barnthouse, L. W., and J. Brown. 1994. Chap. 3. In Conceptual model development. Ecological Risk Assessment Issue Papers, EPA/630/R-94/009. U.S. Environmental Protection Agency, Washington, D.C.

Barnthouse, L. W., and G. W. Suter II. 1986. User's manual for ecological risk assessment: Oak Ridge National Laboratory. ORNL-6251. Oak Ridge National Laboratory, Oak Ridge, Tenn.

.

Barnthouse, L. W., G. W. Suter II, A. E. Rosen, and J. J. Beauchamp. 1987. Estimating responses of fish populations to toxic contaminants. Environ. Toxicol. Chem. 6:811–24.

Barrett, G. W. 1968. The effect of an acute insecticide stress on a semi-enclosed grassland ecosystem. Ecology 49:1019–35.

Bartell, S. M., R. H. Gardner, and R. V. O'Neill. 1992. Ecological Risk Estimation. Lewis Publishers, Ann Arbor, Mich.

Bence, A. E., and W. A. Burns. 1995. Fingerprinting hydrocarbons in the biological resources of the *Exxon Valdez* spill area. pp. 84–140. In P. G.Wells, J. N. Butler, and J. S. Hughes (eds.), *Exxon Valdez* Oil Spill: Fate and Effects in Alaskan Waters. American Society for Testing and Materials, Philadelphia.

Bossert, I., and R. Bartha. 1984. The fate of petroleum in soil ecosystems. In R. Atlas (ed.), Petroleum Microbiology. pp. 435–473. Macmillan Pub., New York.

Bromilow, R. H., and K. Chamberlain. 1995. Principles governing uptake and transport of chemicals. pp. 37–68. In F. Trapp and J. C. McFarlane (eds.). Plant Contamination: Modeling and Simulation of Organic Chemical Processes. Lewis Pub., Boca Raton, Fla.

Cairns, J., Jr., K. L. Dickson, and A. W. Maki. 1979. Estimating the hazard of chemical substances to aquatic life. Hydrobiologia 64:157–66.

Dickson, K. L., W. T. Waller, J. H. Kennedy, and L. P. Ammann. 1992. Assessing the relationship between ambient toxicity and instream biological response. Environ. Toxicol. Chem. 11:1307–22.

Donker, M. H., H. Eijsackers, and F. Heimbach (eds). 1994. Ecotoxicology of Soil Organisms. Lewis Pub., Boca Raton, Fla.

Donkin, S. G., and D. B. Dusenbery. 1993. A soil toxicity test using the nematode *Caenorhabditis elegans* and an effective method of recovery. Environ. Contam. Toxicol. 25:145–51.

Emlen, J. M. 1989. Terrestrial population models for ecological risk assessment: A state–of–the–art review. Environ. Toxicol. Chem. 8:831–42.

EPA (U.S. Environmental Protection Agency). 1991. Risk assessment guidance for superfund, Vol. 1, Human health evaluation manual, Part B: Development of risk based preliminary remediation goals. OSWER Directive No. 9285.7-1B, EPA, Washington, D.C.

EPA (U.S. Environmental Protection Agency). 1992. Framework for ecological risk assessment. EPA/630/R-92/001. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, D.C.

EPA (U.S. Environmental Protection Agency). 1993. Technical basis for deriving sediment quality criteria for nonionic organic contaminants for the protection of benthic organisms by using equilibrium partitioning. EPA–822–R–3–001. Office of Water, U.S. Environmental Protection Agency, Washington, D.C.

EPA (U.S. Environmental Protection Agency). 1996. Ecotox thresholds. EPA 540/F-95/038. Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency, Washington, D.C.

Gordon, G. E. 1988. Receptor models. Environ. Sci. Technol. 22:1132–42.

Hansch, C., and A. Leo. 1995. Exploring QSAR: Fundamentals and Applications in Chemistry and Biology. American Chemical Society, Washington, D.C.

Hartwell, S. I., C. E. Dawson, D. H. Jordahl, and E. Q. Durell. 1995. Demonstrating a method to correlate measures of ambient toxicity and fish community diversity. CBRM-TX-95-1. Maryland Department of Natural Resources, Chesapeake Bay Research and Monitoring Division, Annapolis, Md.

Health Council of the Netherlands. 1991. Quality parameters for terrestrial ecosystems and sediments. Report No. 91/17E. Health Council of the Netherlands, The Hague.

Heimbach, U., P. Leonard, R. Miyakawa, and C. Able. 1994. Assessment of pesticide safety to the carabid beetle, *Poecilus cupreus*, using two different semifield enclosures. pp. 205–40. In M. H. Donker, H. Eijsackers, and F. Heimbackers (eds.), Ecotoxicology of Soil Organisms. Lewis Publishers, Boca Raton, Fla.

Huggett, R. J., R. A. Kinerle, P. M. Mehrle, and H. L. Bergman (eds.). 1992. Biochemical, Physiological, and Histological Markers of Anthropogenic Stress. Lewis Publishers, Boca Raton, Fla.

Jones, D. S., R. N. Hull, and G. W. Suter II. 1996. Toxicological benchmarks for screening potential contaminants of concern for effects on sediment-associated biota. ES/ER/TM-95/R2. Oak Ridge National Laboratory, Oak Ridge, Tenn.

Kammenga, J. E., P.H.G. Van Koert, J.A.G. Riksen, G. W. Korthals, and J. Bakker. 1996. A toxicity test in artificial soil based on the life history strategy of the nematode *Plectus acuminatus*. Environ. Toxicol. Chem. 15(5):722–27.

Kaplan I., S.-T. Lu, R.-P. Lee, and G. Warrick. 1996. Polycyclic hydrocarbon biomarkers confirm selective incorporation of petroleum in soil and kangaroo rat liver samples near an oil well blowout site in the western San Joaquin Valley. Calif. Environ. Toxicol. Chem. 15:(5)696–707.

Larson, R. J., and C. E. Cowan. 1995. Quantitative application of biodegradation data to environmental risk and exposure assessments. Environ. Toxicol. Chem. 14(8):1433–42.

Linder, G., E. Ingham, C. J. Brandt, and G. Henderson. 1992. Evaluation of terrestrial indicators for use in ecological assessments at hazardous waste sites. EPA/600/R-92/183. U.S. Environmental Protection Agency, Corvallis, Oreg.

Lokke, H. 1994. Ecotoxicological extrapolation: Tool or toy. pp. 411–25. In M. Donker, H. Eijsackers, F. Heimbach (eds.), Ecotoxicology of Soil Organisms. Lewis Publishers, Boca Raton, Fla.

Lyman, W. J., W. F. Reehl, and D. H. Rosenblatt. 1982. Handbook of Chemical Property Estimation Methods. McGraw-Hill Book Company, New York.

Lytle, J. S., and T. F. Lytle. 1987. The role of *Juncus roemerianus* in cleanup of oil polluted sediments. American Petroleum Institute, Washington, D.C.

McKay, D., and P. C. Singleton. 1974. Time required to reclaim land contaminated with crude oil—An approximation. Report B–612. Agricultural Extension Service, University of Wyoming, Larmie, Wyo.

McVey, M., K. Hall, P. Trenham, A. Soast, L. Frymier, and L. A. Hirst. 1993. Wildlife exposure factors handbook. EPA/600/R-93/187. Office of Health and Environmental Assessment, U.S. Environmental Protection Agency, Washington, D.C.

Menzie, C. A., D. E. Burmaster, D. S. Freshman, and C. Callahan. 1992. Assessment of methods for estimating ecological risk in the terrestrial component: A case study at the Baird and McGuire Superfund site in Holbrook, Massachusetts. Environ. Toxicol. Chem. 11:245–60.

Menzie, C., M. H. Henning, J. Cura, K. Finkelstein, J. Gentile, J. Maughan, D. Mitchell, S. Petron, B. Potocki, S. Svirsky, and others. 1996. A weight–of–evidence approach for evaluating ecological risks: Report of the Massachusetts Weight–of–Evidence Work Group. Human Ecol. Risk Assess. 2(2):277–304.

Morgan, M. G., and M. Henrion. 1990. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, New York.

National Research Council. 1989. Improving Risk Communication. National Academy Press, Washington, D.C.

National Research Council. 1994. Science and Judgement in Risk Assessment. National Academy Press, Washington, D.C.

O'Neill, R. V., R. H. Gardner, L. W. Barnthouse, G. W. Suter II, S. G. Hildebrand, and C. W. Gehrs. 1982. Ecosystem risk analysis: A new methodology. Environ. Toxicol. Chem. 1:167-77.

Pearson, W. H., E. Moksness, and J. R. Skalski. 1995. A field and laboratory assessment of oil spill effects on survival and reproduction of Pacific herring following the *Exxon Valdez* spill. pp. 626–61. In P. G. Wells, J. N. Butler, and J. S. Hughes (eds.), ASTM, *Exxon Valdez* Oil Spill: Fate and Effects in Alaskan Waters. Philadelphia.

Plafkin, J. L., M. T. Barbour, K. D. Porter, S. K. Gross, and R. M. Hughes. 1989. Rapid bioassessment protocols for use in streams and rivers: Benthic macroinvertebrates and fish. EPA/444/4-89-001. U.S. Environmental Protection Agency, Washington, D.C.

Sample, B. E., and G. W. Suter II. 1994. Estimating exposure of terrestrial wildlife to contaminants. ES/ER/TM-125. Oak Ridge National Laboratory, Oak Ridge, Tenn.

Sample, B. E., M. Opresko, and G. W. Suter II. 1996. Toxicological benchmarks for wildlife. ES/ER/TM-86/R3. Oak Ridge National Laboratory, Oak Ridge, Tenn.

Sloof, W., J. A. M. van Oers, and D. de Zwart. 1986. Margins of uncertainty in ecotoxicological hazard assessment. Environ. Toxicol. Chem. 5:841–52.

Suter, G. W., II. 1981. Laboratory tests for effects of chemicals on terrestrial population interactions and ecosystem properties. pp. 93–154. In A. S. Hammons (ed.), Methods for Ecological Toxicology. Ann Arbor Science, Ann Arbor, Mich.

Suter, G.W., II. 1989. Ecological endpoints. pp. 2-1–2-28. In W. Warren-Hicks, B. R. Parkhurst, S. S. Baker Jr. (eds.), Ecological Assessment of Hazardous Waste Sites: A Field and Laboratory Reference Document. EPA 600/3-89/013. Corvallis Environmental Research Laboratory, Corvallis, Oreg.

Suter, G. W., II. 1990a. Endpoints for regional ecological risk assessments. Environ. Manage. 14:9–23.

Suter, G. W., II. 1990b. Uncertainty in environmental risk assessment, Chap. 9, pp. 203–30. In G. M. von Furstenberg (ed.), Acting Under Uncertainty: Multidisciplinary Conceptions. Kluwer Academic Publishers, Boston.

Suter, G. W., II. 1993. Ecological Risk Assessment. Lewis Publishers, Boca Raton, Fla.

Suter, G. W., II. 1995. Guide for performing screening ecological risk assessments at DOE facilities. ES/ER/TM–153. Lockheed Martin Energy Systems, Inc., Oak Ridge, Tenn.

Suter, G. W., II, and J. M. Loar. 1992. Weighing the ecological risks of hazardous waste sites: The Oak Ridge case. Environ. Sci. Technol. 26(3):432–38.

Suter, G. W., II, and C. L. Tsao. 1996. Toxicological benchmarks for screening potential contaminants of concern for effects on aquatic biota. ES/ER/TM-96/R2. Oak Ridge National Laboratory, Oak Ridge, Tenn.

Suter, G. W., II, D. S. Vaughan, and R. H. Gardner. 1983. Risk assessment by analysis of extrapolation error, a demonstration for effects of pollutants on fish. Environ. Toxicol. Chem. 2:369–78.

Suter, G. W., II, A. E. Rosen, J. J. Beauchamp, and T. T. Kato. 1992. Results of analysis of fur samples from the San Joaquin kit fox and associated water and soil samples from the Naval Petroleum Reserve No. 1, Tupman, California. ORNL/TM-12244. Oak Ridge National Laboratory, Oak Ridge, Tenn.

Suter, G. W., II, J. W. Gillett, and S. Norton. 1994. Chap. 3 In Characterization of exposure. Ecological Risk Assessment Issue Papers, EPA/630/R-94/009. U.S. Environmental Protection Agency, Washington, D.C.

Trapp, F., and J. C. McFarlane. 1995. Introduction. pp. 37–68. In F. Trapp and J. C. McFarlane (eds.), Plant Contamination: Modeling and Simulation of Organic Chemical Processes. Lewis Pub., Boca Raton, Fla.

Troyer, M. E., and M. S. Brody. 1994. Managing ecological risks at EPA: Issues and recommendations for progress. EPA/600/R-94/183. U.S. Environmental Protection Agency, Washington, D.C.

Urban, D. J., and N. J. Cook. 1986. Hazard evaluation, standard evaluation procedure, ecological risk assessment. EPA-540/9-85-001. U.S. Environmental Protection Agency, Washington, D.C.

van de Meent, D., and D. Toet. 1992. Dutch priority setting system for existing chemicals. Report No. 679120001. National Institute for Public Health and Environmental Protection, Bilthoven, The Netherlands.

Van gestel C. A. M., and W. C. Ma. 1988. Toxicity and bioaccumulation of chlorophenols in earthworms in relation to bioavailability in soil. Ecotoxicol. Environ. Saf. 15:289–97.

Van gestel C. A. M., N. M. van Straalen. 1994. Ecotoxicological test systems for terrestrial invertebrates. pp. 205–40. In M. H. Donker, H. Eijsackers, F. Heimbackers (eds.), Ecotoxicology of Soil Organisms. Lewis Publishers, Boca Raton, Fla.

van Straalen, N. M., G. A. J. Denneman. 1989. Ecological evaluation of soil quality criteria. Ecotoxicol. Environ. Saf. 18:241–45.

van Straalen, N. M., and C. A. M. van Gestel. 1993. Chap. 15, Soil. In P. Calow (ed.), Handbook of Ecotoxicology. Blackwell Scientific, Oxford, England.

Van Voris, P., G. W. Dawson, J. K. Fredrickson, D. A. Cataldo, L. E. Rogers, C. M. Novich, and J. Meuser. 1987. Type B—Technical Information Document: Approaches for assessment of injuries to soil arising from discharge of hazardous substances and oil. PNL-5751. Pacific Northwest Laboratory, Richland, Wash.

Wild, S. R., M. L. Berrow, S. P. McGrath, and K. C. Jones. 1992. Polynuclear aromatic hydrocarbons in crops from long-term field experiments amended with sewage sludge. Environ. Pollut. 76:25–32.

Will, M. E., and G. W Suter II. 1995a. Toxicological benchmarks for screening potential contaminants of concern for effects on soil and litter invertebrates and heterotrophic processes. ES/ER/TM-126. Oak Ridge National Laboratory, Oak Ridge, Tenn.

Will, M. E., and G. W. Suter II. 1995b. Toxicological benchmarks for screening potential contaminants of concern for effects on terrestrial plants. ES/ER/TM-85/R1. Oak Ridge National Laboratory, Oak Ridge, Tenn.

Zeeman M. G. 1995. Ecotoxicity testing and estimation methods developed under Section 5 of the Toxic Substances Control Act (TSCA). pp. 703-15. In G. Rand (ed.), Fundamentals of Aquatic Toxicology: Effects, Environmental Fate, and Risk Assessment. Taylor and Francis, Washington, D.C.