Biodegradation **4**: 207–230, 1993.

Experimental linkage issues of petroleum site bioremediation

J.W. Blackburn,[1] E.J. Harner,[2] W.K. Robbins,[1] R.C. Prince,[1] J.R. Clark,[3] R.M. Atlas[4] & J.B. Wilkinson[5]
[1] *Corporate Research Laboratory, Exxon Research and Engineering Co., Clinton Township, Route 22 East, Annandale, NJ 08801–0998, USA;* [2] *Dept of Statistics and Computer Science, West Virginia University, Morgantown, WV;* [3] *Exxon Biomedical Sciences, Inc., East Millstone, NJ;* [4] *Dept of Biology, University of Louisville, Louisville, KY;* [5] *Exxon Engineering, Exxon Research and Engineering Co., Florham Park, NJ, USA*

Received 26 October 1992; accepted 15 August 1993

Key words: analytical chemistry, bioassay, biochemistry, composition, hydrocarbons, remediation

Abstract

Bioremediation of petroleum-contaminated sites is expected to be a cost-effective remediation technology. However, many potential users of the technology expect the reliability of this technology to be similar to other candidate technologies for widespread consideration. In particular, candidate technologies should possess the property of reliable experimental linkage – there should be reasonable confidence that experiments done at one scale can be reliably related to another. An important example is bench-scale treatability studies that should result in linkages with commercial-scale operations. In this respect comparison of bioremediation to other candidate technologies reveals that bioremediation is in an early stage of its evolution. It is being pursued at a variety of sites and scales with practitioners from a variety of disciplines. Integration of activities between disciplines and an ability to quantitatively compare results at different sites and scales is proceeding. This paper addresses a number of physical, chemical, biological, analytical, and statistical issues regarding the successful comparison of results between experiments.

Introduction

> . . . There is little doubt that it would be technically feasible to transfer any new process whatever from the laboratory directly to large-scale production provided that unlimited money were available, so that the designer could allow huge factors of safety and the operator could if necessary meet the cost of overcoming massive "teething" troubles.
>
> (Johnstone & Thring 1957)

Environmental remediation technology has evolved into one of the Nation's major enterprises in scarcely more than a decade. Cost projections for site remediation in this country have exceeded one trillion dollars distributed over decades into the future. Perhaps never before has there been such an explosion of activity of major economic and environmental impact over so brief a time.

Societal and political demands have driven this phenomenon as citizens have demanded a clean environment along with economic prosperity. Scientists, engineers and other technologists have risen to the challenge in much the same way as our society has faced major crises in the past--accelerating technology transfer and development with a significant number of successes at field-scale, but with a related number of activities where something less than complete success was achieved.

Under such pressure the need to identify and apply reliable remediation technology to these problems is severe. Linking results from lab- or pilot-scale experiments to achieve reliable predictions of

field-scale behavior becomes essential, yet has remained elusive. This paper explores a number of factors giving rise to limitations in linking experimental bioremediation results including:

- Impact of Multiple Disciplines on Experimental Requirements
- Impact of Physical System Structure on Experimental Linkage
- Impact of Chemical Complexity on Experimental Linkage
- Impact of Biological Factors on Experimental Linkage
- Impact of Diverse Analytical Methodologies on Experimental Linkage
- Impact of Statistical Design and Analysis on Experimental Linkage

Impact of multiple disciplines on experimental linkage

Bioremediation knowledge spans classical disciplines and requires interdisciplinary cooperation and interactions. The evolution of some of the essential interactions is now discussed for the sciences of chemistry, microbiology, microbial ecology and the engineering disciplines of chemical, biochemical and environmental engineering.

Sciences have long operated in the domain of basic and theoretical investigations while engineering disciplines have at their focus the development and application of the knowledge and tools ultimately needed to transfer the basic knowledge to productive operation. Historically chemists have developed reaction or separation knowledge of molecules and processes at small bench scales. The knowledge is then released or transferred to chemical or petrochemical engineers to develop and optimize the process, and design, construct and operate the full-scale production plants. These concepts have been implemented so efficiently in manufacturing that application of these approaches to environmental remediation systems is entirely logical. However, mechanistic knowledge gaps and the complexity of environmental remediation systems, along with the adaptive nature of biological systems, pose new challenges in commercialization that extend beyond those faced in our technological past.

Reliable and formal experimental linkage among behaviors at several scales has become the basis for collecting data at bench- or pilot-scales and designing plants at field-scale. In chemical engineering this linkage arose from the concept of similarity criteria (discussed later) and their application to carefully designed experiments at various scales. These experiments' results led to the development of equations for prediction of certain system parameters when others were known. Strong analogies were possible to relate fluid flow, heat transfer and mass transfer. Often called transport correlations, these permitted the prediction of mass transfer or heat transfer properties between similar systems if, say, fluid flow behavior was known.

The mass transfer rate is one of the rate processes important in experimental linkage. Mass transfer rate correlations can lead to the ability to develop mathematical models describing the concentrations of important components in various phases in the system. In addition to advection rates based on mass action (flux into and away from the system) other important rates include the rates of component reaction or conversion or chemical kinetics.

In these cases, the predictive equations to link experiments include mass transfer expressions, often containing information from correlations, as well as elementary or nonelementary chemical kinetic rate expressions. Sometimes the rates of compound removal or conversion from a phase or compartment can be mathematically complex and require calibration by experiment for validation. Validated expressions can be built into computer simulations providing the basis for predictive experimental linkage and scale-up.

Like chemistry, microbiology has emerged over the past century as a critical basic science. With petroleum hydrocarbon microbiology as an example, the microbiology of organisms capable of growth on petroleum hydrocarbons has been studied in some detail in the last 100 years. Much elegant work has been done with these isolates, but the perhaps justifiable obsession of microbiology with pure cultures, and the difficulty of analyzing such complex

materials as commercial hydrocarbon products, has ensured that the vast majority of this work has involved the growth of a single organism on a single hydrocarbon or class of hydrocarbons. We thus have a reasonably comprehensive view of the biochemistry and physiology of some of the microorganisms capable of metabolizing simple aliphatic and aromatic hydrocarbons (Hartman et al. 1989; Watkinson & Morgan 1990; Smith 1990). The development of the disciplines of biochemical engineering, modern environmental engineering, and microbial ecology provided a basis for projection of this knowledge into domains of mixed cultures and consortia, and environmentally complex and dynamic applications.

Biochemical engineering was born when, during the second World War, it became very important to manufacture antibiotics in bulk for treating casualties. This discipline has been defined broadly – 'Processing of biological materials and processing using biological agents such as cells, enzymes, or antibodies are the central domains of biochemical engineering' (Bailey & Ollis 1987), however emphasis seems to be on engineered reactor-based systems (Cooney 1983). With exceptions, this attention has also been limited to applications where:

- high-value products are desired,
- the microbiology is intentionally limited to pure culture,
- the feedstocks are defined and often manipulated, and
- the reactor conditions are controllable and carefully controlled.

The practitioners of the discipline of modern environmental engineering ventured into the domain where none of the above qualifications apply:

- the desired product (clean water, air or land) has low economic value, but high aesthetic value,
- the salient microbiology generally cannot be limited to pure cultures,
- it is generally a technical and economic impossibility to completely define the feedstocks – the waste streams, and
- environmentally-based reactors – natural or engineered – are subject to environmental disturbances and generally cannot be as closely controlled as fermentation systems.

Early environmental engineering directions and the regulations driving the activities emphasized analytical parameters that integrated responses to multiple specific chemicals in a feedstock. These included non-specific analytical parameters such as total organic carbon (TOC), chemical oxygen demand (COD), biochemical oxygen demand (BOD), total dissolved solids (TDS), and others. These non-specific parameters made analysis of the highly variable feedstocks technically and economically tractable and provided a basis for developing biological kinetic rate models and system design. Strategies for reliable linkage, design, and scale-up became possible.

As interest grew for regulating specific pollutants in biological process discharge, activities to evaluate the feasibility of using the non-specific parameter approach began to conclude that attention to specific pollutants demanded either knowledge the equivalent of the chemical or biochemical engineering model or acceptance of decreased predictive linkage (Blackburn 1987, 1988a). With increased interest in specific pollutants also came the need for site-specific biokinetic studies to evaluate the biological rate expression and the inevitable need to link these results with experiments at various scales.

Difficulty in linking biological behavior between experiments was identified as a potential problem area (Blackburn 1988a, 1988b) and recently, questions of standardization of testing protocols to minimize variability in bioactivity results has received attention (Linz et al. 1991; McFarland et al. 1991). The hierarchical nature of microbial systems studies has been well described (Needham 1935; Prokop 1983; Blackburn 1988) and has led to the recognition that biological activity experiments in general and treatability studies specifically address different questions at several scales often with different experimental designs (Grady 1985; EPA 1989; NETAC 1992). Currently, there is no single, general approach for running biological activity experiments, although efforts are proceeding to converge on a formalized approach, at least for oil spill biodegradation (NETAC 1992).

The discipline of microbial ecology has emerged in the autumn of this century partially in response to

the recognition of a need for an organized science to provide a framework within which to evaluate changes of states and activities in microbial systems arising from 'interrelationships between organisms and their living (biotic) and nonliving (abiotic) environments' – a definition of microbial ecology (Atlas & Bartha 1987).

While extensive effort has been applied to methods improvement for observing the state of microbial ecosystems, attention has also been given to the development of methods and strategies for measurement of microbial activity or rate of biological processing in environmental systems. Wiebe (1971) divides the issues in this discipline into the following:

- the numbers of microbes in environmental systems,
- the recognition of microbes in environmental systems,
- microbial performance, and
- microbial rate of function.

Many other scientific disciplines offer data and information that must be integrated in order to successfully develop methods to link results between petroleum site bioremediation experiments. Some of these may include analytical chemistry, agronomy, soil science, hydrology, toxicology, and so on. Information and contributions from these fields are critical and serve to highlight the need to view bioremediation as an effort that requires input from many disciplines in order to progress.

Perhaps it is the multitude of relevant disciplines and the historical methods adopted by each that has best highlighted the challenges in comparing experimental results in this area. These might be categorized into a spectrum that begins with fundamental mechanism elucidation and spans to phenomenological observation. Classical scientific philosophy has embraced the concept of the 'controlled experiment' in ever-increasing levels of reductionism. This trend has led to great insights at the mechanistic level but at the price of necessarily decoupling the experiment and its activity from its overall environment (Needham 1935).

The goal of bioremediation experimental linkage requires that the mechanistic knowledge must be ordered into a framework or network of relationships that depict the interrelationships of the basic mechanisms. As in the analogous development of chemical engineering correlations discussed earlier, these relationships come from comparing the observed phenomena in various environmental experiments while formulating and testing hypotheses to discover the nature of the interactions.

An example of an approach to compare the responses of perhaps complex experiments is the chemical engineering scale-up approach. This approach is rooted in the concept that quantitative comparison becomes reliable if the experiments are 'similar enough' to each other.

> Material objects and physical systems in general are characterized by three qualities: size, shape, and composition. All three are independently variable, so that two objects may differ in size while having the same shape and composition... the spatial and temporal configuration of a physical system is determined by ratios of magnitudes within the system itself and does not depend upon the size or nature of the units in which these magnitudes are measured. (Johnstone & Thring 1957)

Similarity criteria important to chemical engineering (Table 1) are: geometric similarity, mechanical (static and kinematic) similarity, thermal similarity and chemical similarity. These are also interrelated where thermal and chemical criteria also require a simultaneous geometric and/or mechanical similarity.

If the goal is to relate the biological responses between experiments in terms of observable quantities, then the concept of biological similarity, conspicuously absent from the historical chemical engineering approach (Table 1), is seminal. There are perhaps two basic philosophies that can be considered:

- Attempts to create experiments that are geometrically, mechanically, thermally, chemically, and biologically similar and where differences in the responses can be directly attributed only to comparative effects, or
- Work with systems that are dissimilar in some way, but where the dissimilarities can be described by predictive models.

Here the question is – Can biological activity be compared between systems if sufficient knowledge

of the physical, chemical, mechanical, and thermal processes are known?

For example, it is the first approach that aquatic ecologists have sought where highly similar test systems have been used to observe biological responses with the hope that high levels of linkage with the environmental cases can be expected. On the issue of necessary and sufficient conditions for 'similarity' in aquatic microcosms: 'To fully realize their potential as tools for ecological research, microcosms must be designed which optimize replicability, baseline stability, persistence, and functional similarity to natural ecosystems. Evidence [in the referenced paper] is presented that complex aquatic microcosms, containing whole biotic communities taken directly from a pond, satisfy these criteria (Giddings & Eddlemon 1979).

Table 1. Chemical Engineering Similarity Criteria (Johnstone & Thring 1957).

Similarity criterion	Definition
Geometric	To every point in one body there exists a corresponding point in the other
Mechanical (Static)	Geometrically similar bodies are statically similar when under constant stress, their relative deformations are such that they remain geometrically similar
Mechanical (Kinematic)	Geometrically similar bodies are kinematically similar when corresponding particles trace out geometrically similar paths in corresponding intervals of time
Mechanical (Dynamic)	Geometrically similar moving systems are dynamically similar when the ratios of all corresponding forces are equal
Thermal	Geometrically similar systems are thermally similar when corresponding temperature differences bear a constant ratio to one another and when the systems, if moving, are kinematically similar
Chemical	Geometrically and thermally similar systems are chemically similar when corresponding concentration differences bear a constant ratio to one another and when, if moving, are kinematically similar

The notions that only complex microcosms are sufficient for biological similarity and linkage implies that linkage-based studies necessarily lead to very complicated, expensive, and time-consuming efforts or alternatively the conclusion that only observation in the field is useful. The implications of either of these suppositions have significant impact upon the economies and response times of using bioremediation and the competition of this technology with alternative technologies (Rao, et al. 1992; Woodyard 1991; Litchfield 1991).

Impact of physical structure on experimental linkage

That contaminated environments are physically diverse is obvious. Oil spilled on the open ocean would be expected to exhibit widely different characteristics from oil spilled on and seeped into the ground. Comprehensive discussion of structural diversity is beyond the scope of this paper, but several concepts will be noted that are expected to have an impact on physical similarity and subsequent experimental linkage.

Compartmentalized models to envision the phases in which a contaminant may reside are important in the development of predictive models that incorporate advection, mass-transfer and reaction mechanisms. Bulk and specific components of oil partition between air, water, sediment, soil, and related biotic compartments as time proceeds after oil is introduced into the environment. Physical properties of the oil (i.e., viscosity, density, solubility, octanol-water partition coefficient, Henry's law constant, etc.) determine the equilibrium concentrations of the oil in the compartments – the maximum potential concentrations. The impact of rate processes, mass-transfer, chemical or biological reactions or conversions, and advection result in concentrations often much lower than those calculated assuming thermodynamic equilibrium.

'Open systems' are those that have no physical boundaries for advection in one or more compartment. Oil spilled on the surface of an ocean is 'open' in the atmospheric compartment as well as the aqueous phase compartment. The 'bulk' air phase

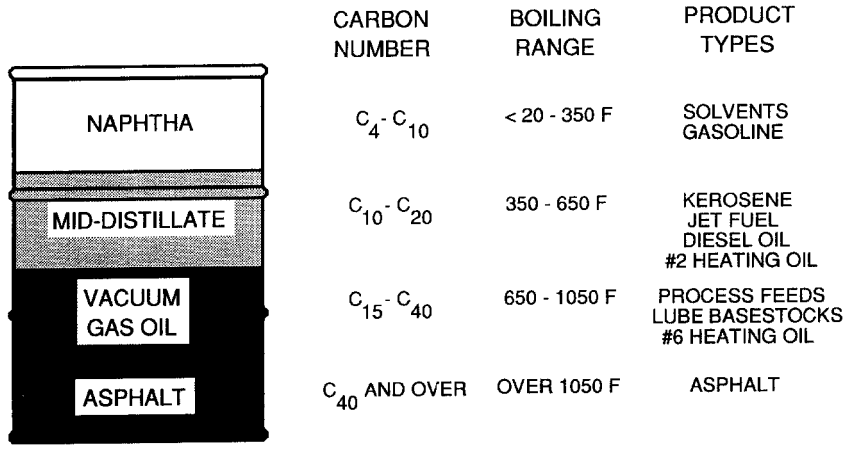

Fig. 1. Crude oil fractions and their properties.

above the spill never reaches equilibrium partition concentrations with the oil spill because the wind replaces the air faster than the components can evaporate. The same is true for the bulk aqueous phase. Atmospheric and aqueous compartmental concentrations of the components are relatively low in these cases because the advection rates are large relative to the mass-transfer rates. Note that in terms of the oil phase compartment, the composition and resulting physical properties are controlled by advection of the atmospheric and aqueous compartments and change with time, with viscosity increasing and mass transfer processes lessening.

In a terrestrial spill, another compartment, soil, is present, and soon after the spill much of the oil has moved into the subsurface. The soil is 'closed' in that no additional soil 'enters, leaves, or accumulates' in the system, but air and water may move through the system. Limited by the flow restrictions of soil pore and capillary flow, advection rates are low and the concentrations of components might more closely approach equilibrium partitioning concentrations.

An interesting question arises when considering whether biological conversion rates from these two cases could be linked. Overall biological conversion would obviously be very different in each case. But what would happen if the biological rate expressions were 'normalized' for concentrations of component and biomass, conditions, and advection and mass transfer rates? Could general bioremediation rate expressions be developed between dissimilar systems? Unfortunately, past limits in methods of chemical and biological analysis have generally prevented calculation of the biological conversion rate in multi-compartment, mixed culture microcosms.

Environmental systems are driven by varying conditions and have dynamic responses. Experiments to simulate the environment are often controlled to reduce or eliminate variations in conditions. Temporal variations in system parameters can have a major effect upon biological kinetics and system stability (Blackburn 1989; DiGrazia 1991). Experimental strategies that include disturbances and perturbations may increase variability and make statistical comparisons more difficult, but may offer a more realistic linkage to field-scale behavior.

Impact of chemical complexity on experimental linkage

Petroleum oils may be discussed very generally in terms of the four primary distillation fractions obtained from crude oil (Fig. 1). Each of these fractions, in turn, consist of a limited number of compound types:

Saturates

These include the straight chain normal paraffins, the branched isoparaffins (including pristane and phytane) and the cycloalkanes or napthenes with one or more saturated rings. Among the latter, the

four-ring steranes and five-ring hopanes are frequently used as conserved quantitative internal standards because they are resistant to biodegradation.

Aromatics

These include compounds with one or more aromatic rings. Depending on the type of oil, compounds with one to four or more fused aromatic rings dominate this class. Each of these ring types may have attached multiple saturated rings and/or saturated side chains.

Polars

These include asphaltenes (the high molecular weight or highly functionalized materials that fall out of solution when exposed to a low carbon number saturated solvent such as pentane or heptane) and weaker polars that may be removed by adsorbents such as clay. For crude oils, these weaker polars consist mainly of *n*-heterocyclics, acids, and sulfoxides.

Refining processes separate and interconvert different compound types to obtain products with high market value. A typical distribution of oil fractions and related products is illustrated in Fig. 1. Broader ranges of the compound types may be found in specialty products. Figure 2 presents the composition of the oil fractions in terms of paraffins, napthenes, aromatics and polars. The relative concentrations of paraffins and napthenes decrease while aromatics and polars increase with higher-boiling fractions.

Frequently, the contaminated site has been exposed to a wide number of petroleum products over a number of years. This can lead to very complex residual compositions of petroleum materials in the soils. Terminals, storage tanks, even refinery units are terminated and demolished for different purposes, and over decades wholesale changes in facilities may have occurred. Consequently, at each specific location within a site, there can be a range of different contaminant species at widely different levels. The oil has had an opportunity to be weathered and transformed by a variety of physical, chemical, and biological processes over time.

When oil is spilled onto soil, it is subject to a number of mechanisms that affect its composition: evaporation, permeation, dissolution, dispersion, chemical- and photo-oxidation and biodegradation. Each of these mechanisms has a characteristic time scale and a set of physical parameters which influence the mechanism.

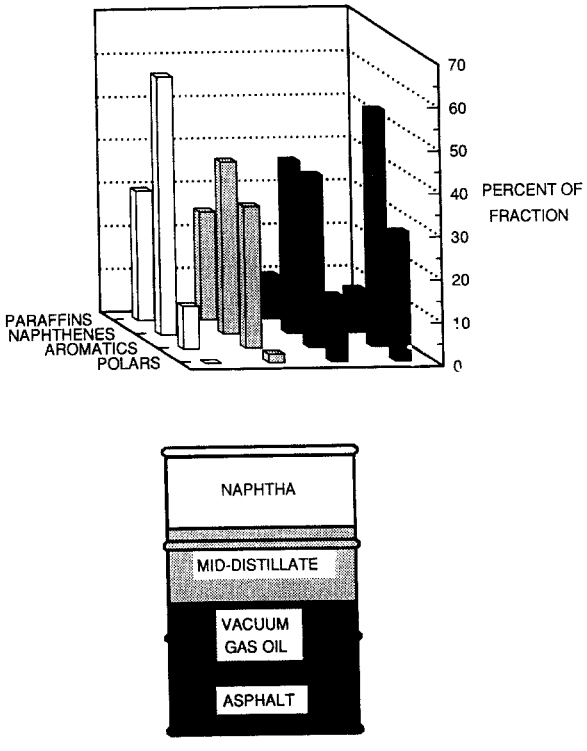

Fig. 2. Composition of crude oil.

Both evaporation and permeation tend to deplete the bulk oil phase of the lower boiling solvent molecules leaving increasingly complex viscous films that are subject to continued physical stress and chemical attack. While photo-oxidation at the soil surface may convert the two-to-four ring aromatics, aerobic and anaerobic biodegradation in the subsurface convert saturates and simpler aromatics. Higher molecular weight aromatics and polars remain relatively conserved. Thus, oil spilled at a site progressively increases in aromaticity and polarity approaching a more 'asphalt-like' composition as time proceeds. The depth of analytical characterization of the oil present at a site is key to understanding its response to treatment.

Impact of biological factors on experimental linkage

Detection of biodegradation

The objective of bioremediation is to remove hydrocarbons and restore contaminated media to a prior state, thereby reducing risk of adverse human or ecological health effects. Large intrinsic variability in oil load may require a large reduction before detection of reduction can occur with statistical confidence. Such changes may be slow, so there may be value in tracking indirect measures of success. Since the basic rationale for bioremediation is that microbial degradation will be stimulated by providing the right environmental conditions (e.g. oxygen, nitrogenous nutrients, temperature etc.) or perhaps the right organisms, an early detectable measure of success is likely to be an increase in microbial metabolism. This can be detected by radiorespirometry, and perhaps by an increase in microbial numbers (Brown et al. 1991; Lindstrom et al. 1991; Brown & Braddock 1990).

Unfortunately radiorespirometric techniques mandate that they be carried out on the bench, and so the rates obtained in such experiments may be difficult to extrapolate to the field-scale. Nevertheless, comparing activities on bioremediated and control (not sterile control) experiments, should provide indications of the success or failure of a bioremediation strategy. Similarly, enumeration assays might compare bioremediated with unamended plots to allow control and analysis of general environmental factors, such as the weather.

A promising new approach, which may be useful even in the early stages of bioremediation, is to take advantage of the fact that petroleum hydrocarbons typically have a different stable isotope composition from more modern biomass, so that CO_2 from hydrocarbon degradation can be separated from other carbon dioxide based on its isotopic composition (Aggarwal & Hinchee 1991). An increase in CO_2 evolution from hydrocarbons in treated plots would be strong evidence for the success of a bioremediation strategy, providing the isotopic composition of any organic material added as part of the treatment is taken into account.

Study at the bench, of samples with a well-characterized history collected from the field, offer another approach to the assessment of biodegradative responses. For such an approach to work, the source of the oil must be known, and some of the initial material must be available for analysis. The spill from the Amoco Cadiz was perhaps the first opportunity for this type of study (Gundlach 1983), and it has been used with success in determining the rate and extent of biodegradation following the spill from the Exxon Valdez (Prince 1992). In order for this approach to be quantitative, relatively inert conserved internal markers must be identified; hopanes seem to serve this role in crude oils (Butler et al. 1991) and phenanthrenes with three pendant carbon atoms fulfill this role in diesel fuels (Douglas 1992). This approach cannot dissect the processes contributing to the disappearance of hydrocarbons into, for example primary metabolism and cometabolism, but it does provide a way of assessing the ultimate limits of biodegradation in a particular environment. Unfortunately it can only be used if a contaminant is relatively uniform in chemical composition throughout the contaminated area as is true in a scenario of a spill from a single source, and for many sites this may not be a realistic assumption.

Diversity of species and activity

Söhngen (1913) isolated a number of bacteria from soil, water and compost that could grow on gasoline, kerosene, paraffin oil and paraffin wax (Sohngen 1913). Many have followed in these footsteps so that there are now thousands of isolates capable of growth on such compounds (Atlas 1981; Austin et al. 1977a; Austin et al. 1977b; Austin et al. 1979; Bartha & Atlas 1977; Buckley et al. 1976; Foght et al 1990; Leahy & Colwell 1990; Venkateswaran et al. 1991). Many fungi and yeast have also been shown to grow on petroleum hydrocarbons (Atlas 1981; Blasig et al. 1989; Leahy & Colwell 1990; Snellnam et al. 1988). Such oil-degrading microbes have been isolated from essentially every kind of environment where they have been searched for, and it seems unlikely that any environment on earth would be with-

out at least a small population of such organisms (Aeckersberg et al. 1991; Bazylinski et al. 1989; Bertrand et al. 1990).

Several generalizations can be made:

- For the relatively simple hydrocarbons, it is the insertion of oxygen into the hydrocarbon that seems to be the rate-limiting step for degradation.
- The inserted oxygen is molecular oxygen, although there is no doubt that anaerobic degradation, which must use some other oxygen source (water, sulfate, nitrate?), can occur (Bertrand et al. 1989; Aeckersberg et al. 1991).
- There is evidence that some organisms that can degrade aliphatics do not seem to be able to degrade aromatics, and vice versa (Foght et al. 1990).
- Biodegradation of alkanes, alkenes and simple aromatics (up to at least four-ring compounds) can be ultimately complete, generating only microbial biomass, water and carbon dioxide (Walter et al. 1991).

Biodegradability of less degradable components

The biodegradation of the more slowly-degrading compounds has been much less studied, and our knowledge is at best fragmentary. Several different approaches are providing important insights. One is to analyze the molecular composition of petroleum products, and address the specific degradation of individual components, usually by pure microbial strains. The other is to attempt to classify the molecules in the petroleum into broad classes, and study the degradation of these classes, usually by mixed microbial populations. As examples of the former approach, Gough and Rowland (1990) have described techniques for analyzing the previously 'unresolved complex mixture' of hydrocarbons in petroleum and many refined products, and have examined the degradation of these compounds by a pure strain of *Pseudomonas fluorescens* (Gough et al. 1991). In a similar vein, but on a quite different class of compounds, Chosson et al. found that only seven Gram-positive strains, from a library of 73 aerobic organisms, were able to degrade steranes (Chosson et al. 1991). Such approaches promise to provide important insights into the specific metabolism of well characterized microorganisms at higher levels of microbiological rigor. They may also shed light on the phenomenon of cometabolism, the process whereby some compounds are degraded essentially by accident, providing no useful energy for the cell. Cometabolism may be very important for the removal of some compounds from the biosphere.

An alternative approach is to study the biodegradation of complex mixtures of hydrocarbons by correspondingly complex mixtures of organisms. Such an approach is relatively new for microbiologists, perhaps because of an initially well-founded belief that such work was irreproducible; 'work with impure cultures yields nothing but nonsense and *Penicillium album*' (Brefeld, 1881, quoted in Atlas and Bartha 1991).

The current recognition of the importance of what is occurring in the environment rather than only in controlled experiments has stimulated a reevaluation of such views. For example, Oudot (1984) has studied the biodegradation of Arabian light crude oil by a marine mixed culture, and analyzed degradation by gas chromatography and mass spectrometry. He found that the different classes of compounds he could resolve could be divided into five classes based on their susceptibility to degradation. The n- and iso-alkanes were highly susceptible to degradation, the 1-, 2-, 5- and 6-ring cycloalkanes, 1-ring and sulfur aromatics a little less so, the 3- and 4-ring cycloalkanes and 2- and 3-ring aromatics less again, the tetra-aromatics, steranes, triterpanes, naphtheno-aromatics very resistant, and the penta-aromatics, asphaltenes and resins highly resistant. Similar findings have been reported by Kennicutt (1988) and Chianelli et al. (1991).

Another factor impacting biodegradation rates of less-degradable components is their availability to microorganisms. Most petroleum hydrocarbons have only limited solubilities (Eastcott et al. 1988), so microbes must produce biosurfactants in order to interact with them (Hommel 1990). The production of surfactants during bioremediation (Oberbremer & Müller-Hurtig 1989) clearly influences the process of biodegradation, but the phenomenon, especially as to how it affects processes in

mixed cultures degrading complex hydrocarbon mixtures, is poorly understood.

Impact of microbial ecology

Microbial ecology addresses the development of methods and techniques to characterize the microbial populations in active environmental samples. These include determination of community structure and interactions among populations that are necessary for quantitative predictive models of the system's 'biological catalysts.' Currently the methods available for application to environmental samples lag the demand for this information.

One important question in performing bench tests with the aim of later linkage to field treatments is whether the diversities of the microbial communities in samples used in replicate tests are representative of the overall microbial diversity of the impacted area. Culturing of microorganisms yields only a low percentage of the total microbial community; generally 99 percent of the organisms in a sample are not culturable. Analysis of the DNA from soils and sediments indicates that the overall diversity of the microbial community is thousands of times higher than the diversity estimated based upon the culturable organisms. However, those same molecular analyses of DNA from soils and sediments suggests that just a few grams of soil contains populations representing over 99 percent of the total diversity of a given site. This applies to both impacted and pristine soils and sediments. Thus, experiments with only a few grams of soils or sediments should provide an adequate basis for determining the range of compounds that can be degraded at a given site by the indigenous populations.

Significant differences, though, occur within and between sites, especially between impacted and pristine sites. Adaptation is a critical factor in determining the ability of a microbial community to degrade pollutants. Often weeks to months are necessary before the succession of populations are able to attack the compounds in a complete mixture. The range of compounds that can be attacked, thus, cannot be readily predicted except by using samples from the site of interest. In the sediments impacted by the Amoco Cadiz spill it was found that the adapted populations could rapidly degrade low molecular weight hydrocarbons before they evaporated; the populations in those sediments were adapted to continuous exposure to ballast washings. The populations in the sediments impacted by the Exxon Valdez spill were able to degrade branched isoprenoid hydrocarbons; the populations in those sediments were adapted to continuous exposure to terpenes from pine trees.

The populations of hydrocarbon degraders increase from less than one percent of the total community in pristine areas to a higher percentage of the total community in impacted areas. One might expect a direct relationship between the growth rate of microorganisms and the rate of hydrocarbon degradation. Generally only the standing biomass and not the biomass turnover rate is measured and the rates of pollutant degradation cannot be determined from such measurements.

In sediments impacted by the Amoco Cadiz oil spill the growth rates of oil degrading microorganisms were found to be 5-8 hours. Based on the growth rates the rates of oil degradation could be estimated based upon the assumption that 30-50 percent of the degraded carbon was incorporated into microbial biomass. These estimated rates of hydrocarbon degradation were within 15 percent of those estimated based upon measurements of residual hydrocarbons in field sediments. Measurements of degradation rates of specific representative hydrocarbons, using radiolabelled substrates and one gram sediment samples also correlated very well with the changes in composition of the residual oil in the sediments. Disappearance of oil from sediments on high energy beaches were accurately predicted by the in vitro rate determinations in low energy high intertidal sediments that were heavily impacted.

Clearly interactions among microorganisms in a community have a significant impact on the rates of pollutant degradation. Using actual field samples (rocks, water, soils, sediments) and the associated mixed communities gives an appropriate community that permits interactions. Studies with isolated

microorganisms fail to account for this community complexity.

In a balanced ecosystem it is likely that oil degradation is stimulated by the presence of predators (Vestal et al. 1984); for example Rogerson and Berger (1983) found that the presence of the predatory ciliate *Colpidium colpoda* stimulated the degradation of Norman Wells crude oil by a mixture of oil-degrading bacteria, and that the 'equilibrium' bacterial cell densities were essentially unaffected by the presence of the predators, even though these were present at relatively high numbers. On the other hand, predators may also inhibit biodegradation, especially by non-indigenous organisms that are not well adapted to their new environment. For example Ramadan et al. (1990) found that microbial predation by protozoa could remove all of a small inoculum of a *Pseudomonas cepacia* capable of degrading p-nitrophenol, although larger inocula survived and thrived.

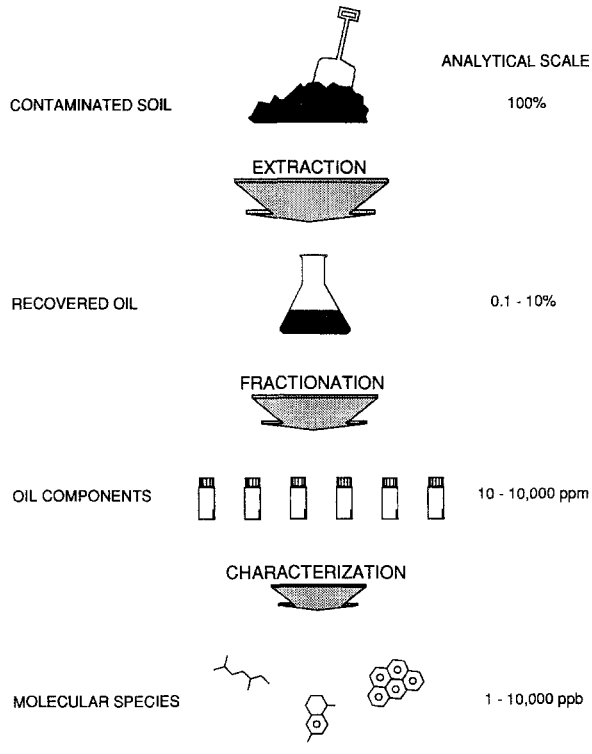

Fig. 3. Procedures in petroleum analysis.

Impact of analytical methodology on experimental linkage

Because of the chemical complexity of petroleum, selection of the type of analytical procedure has major impacts upon the quantitative results of experiments. Apart from the costs of various procedures and the impact on project economy, results from the same type of analytical detector, but on soil extracts from different solvents or employing different fractionation/purification methods will lead to different results and difficulty in comparison of experimental data. Further, a number of analytical detectors are possible, each offering its own set of strengths and limitations. Increasingly, responses of living organisms to petroleum and its fractions in environmental matrices are being considered for analytical methods that integrate effects across various compound types.

Selection of the analytical techniques for monitoring petroleum hydrocarbon bioremediation is critical for several reasons. Non-specific analysis of the bulk oil phase is of interest for identification of location and extent of contamination. Conversely, specific compound information is of interest because it can be modeled in kinetic and structure-function relationships important in developing experimental comparisons.

The choice of analytical method is based on a balance between the desired information content and cost required for a particular application. At one extreme, visible indicators of contamination such as oil sheens on water or black oily-looking soils have been used. At the other extreme, the understanding of mechanistic fate pathways may require measurement of the concentration of a specific compound present at the part per billion level in a relative 'ocean' of other compounds. While the former analysis requires only human observation, the latter may require weeks of sample preparation combined with sophisticated measurement and data interpretation.

In most studies, a middle ground is taken. Measurements are made on some component groups without detailed speciation. While simple observation requires no treatment (Fig. 3), the specific compound analysis requires at least three steps: extraction, fractionation and characterization. By using

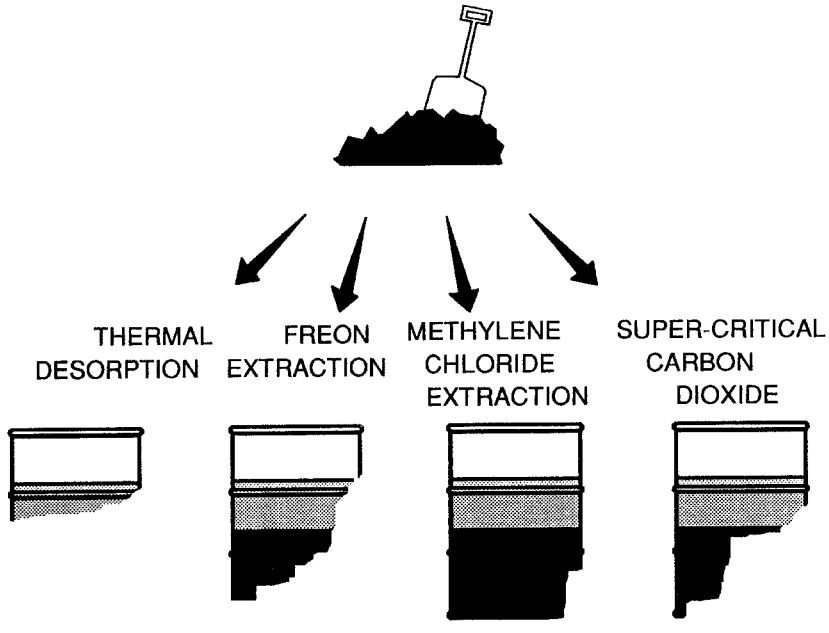

Fig. 4. Impact of extraction on analytical response.

subsets of these steps, a number of intermediate measurements of 'group-type' behavior can be monitored.

Each monitoring technique provides a partial 'snapshot' of the barrel of oil shown in Fig. 2. The choice of method or 'filter' affects the scale and focus of that snapshot. As the analysis proceeds through the different levels of filtration, some portion of the sample is sacrificed to allow better focus on the remainder. Thus snapshots taken with different filters will present a different image of the barrel. The challenge to the scientist is to take snapshots that focus on portions that are appropriate for the experiment and to allow for comparison between experiments using diverse analytical methods.

The following sections discuss the effects of the extraction, fractionation and characterization on measurements of the different portions of the oil. All measurements involve some combination of these steps so it is important to understand their interdependence.

Extraction processes

The recovery of oil from a soil matrix can be achieved by a number of techniques that are dependent upon the fraction of the barrel involved (Fig. 4).

While volatile components ($<C_{15}$) can be removed by thermal desorption, heavier components are generally recovered by some type of solvent extraction. Single equilibrations in separatory funnels or shake flasks are limited to qualitative applications. More thorough extraction of the oil from the soil is required for quantitative measurements. In most cases, samples are acidified and dried (by addition of anhydrous sulfates) and subjected to either triplicate solvent treatments in ultrasonic baths or to extended extraction in a Soxhlet apparatus.

Two commonly used solvents that illustrate the extraction differences are Freon 113 and methylene chloride. The Freon is effective in extracting the nonpolar, lower molecular weight portions of the barrel, but gives incomplete extraction of the larger aromatics and the polars. Methylene chloride is less selective and effectively extracts all the oil components. However, the methylene chloride is more likely to extract non-hydrocarbon organics from the soil.

Alternatively, supercritical fluid extraction (SFE) with carbon dioxide has been proposed for

removing oil from contaminated soil. Initial work has demonstrated that carbon dioxide alone only extracts the nonpolars portion of the oil but suggests that 'modifier solvents' may extend its application.

Fractionation methods

Once an oil has been extracted from a soil, it is often fractionated to remove materials that interfere in the following analytical steps. Fractionation schemes range from simple solvent rejection to complex multi-step chromatography. For this discussion, we will consider the common techniques as they apply to a methylene chloride extract (Fig. 5). Similar results will arise from the use of other extraction solvents.

The initial extraction condition often influences the extraction selectivity. Acidification of the samples converts organic bases to salts that resist solvent extraction, especially if a weak solvent like Freon is used. If the solvent is removed by evaporation in order to concentrate the extract for subsequent analysis, components below C_{10}-C_{15} may be lost.

One common practice in the preparation of heavier oils is to 'remove the asphaltenes". This is a nonselective technique that utilizes solvents to reject high molecular weight components that are difficult to characterize. Both waxes ($>C_{30}$) and polars are rejected by low molecular weight paraffins such as pentane, hexane or heptane. The elution of heavy oil through clay with hexane will effectively remove the polar molecules, which can be recovered from the clay with more polar solvents such as toluene, acetone, or methanol.

Adsorption onto silica gel can be used to isolate the saturates from the aromatics. The extent of resolution depends upon the solvent used. For example, moist silica gel is used to treat Freon extracts to remove larger aromatics and polars. With hexane and dry silica gel, the saturates and aromatics can be completely resolved. Adsorption onto alumina on the other hand is better suited for the separation of aromatics by ring type. It is particularly useful in isolating the 3-to-5 aromatic ring types.

High performance liquid chromatography (HPLC) and gel permeation chromatography are more sophisticated approaches for similar types of separations. Chromatographers use a range of adsorbents and solvents to fractionate the oil by functionality or another property.

Characterization techniques

The simplest characterization technique is gravimetric analysis – the evaporation of a fraction to constant weight. The evaporation of the extraction solvent limits this technique to the heavier portion of mid-distillates and higher-boiling-fraction materials with carbon numbers greater than 10-15 and gives no information on the chemical character of the oil.

Infra-red spectroscopy, now commonly practiced in the Fourier Transform Infra-red Spectrometer (FTIR), provides a quantitative measure of dipole-active functionalities. While capable of detecting a number of chemical bond functionalities, FTIR is not a universal detector. FTIR detects the equivalent of 'bright spots' in the crude oil. If one looks at, for instance, a CH stretch in the 2830 cm^{-1} region, FTIR can be used to estimate the hydrocarbon content of a Freon extract. However, the values are only approximate because the response depends on compound structure. This application is also limited to analysis of extracts of the few extraction solvents that are free of CH bonds of their own.

Gas chromatography (GC) can be used to quantify individual components in the naphtha and the uniquely resolved n-paraffins up through the vacuum gas oil. Although GC can quantify the mass of unresolved components on the basis of boiling point, it does not distinguish them in functionality. GC fails to provide data on asphalts.

Magnetic resonance for protons (H-NMR) and carbon (C-NMR) probes the molecular structure of oil fractions. Since this technique groups carbon and hydrogen functionalities independently of boiling point, it is valuable in identifying the functional forms of carbon and hydrogen, but gives limited insight into how the functionalities are connected into

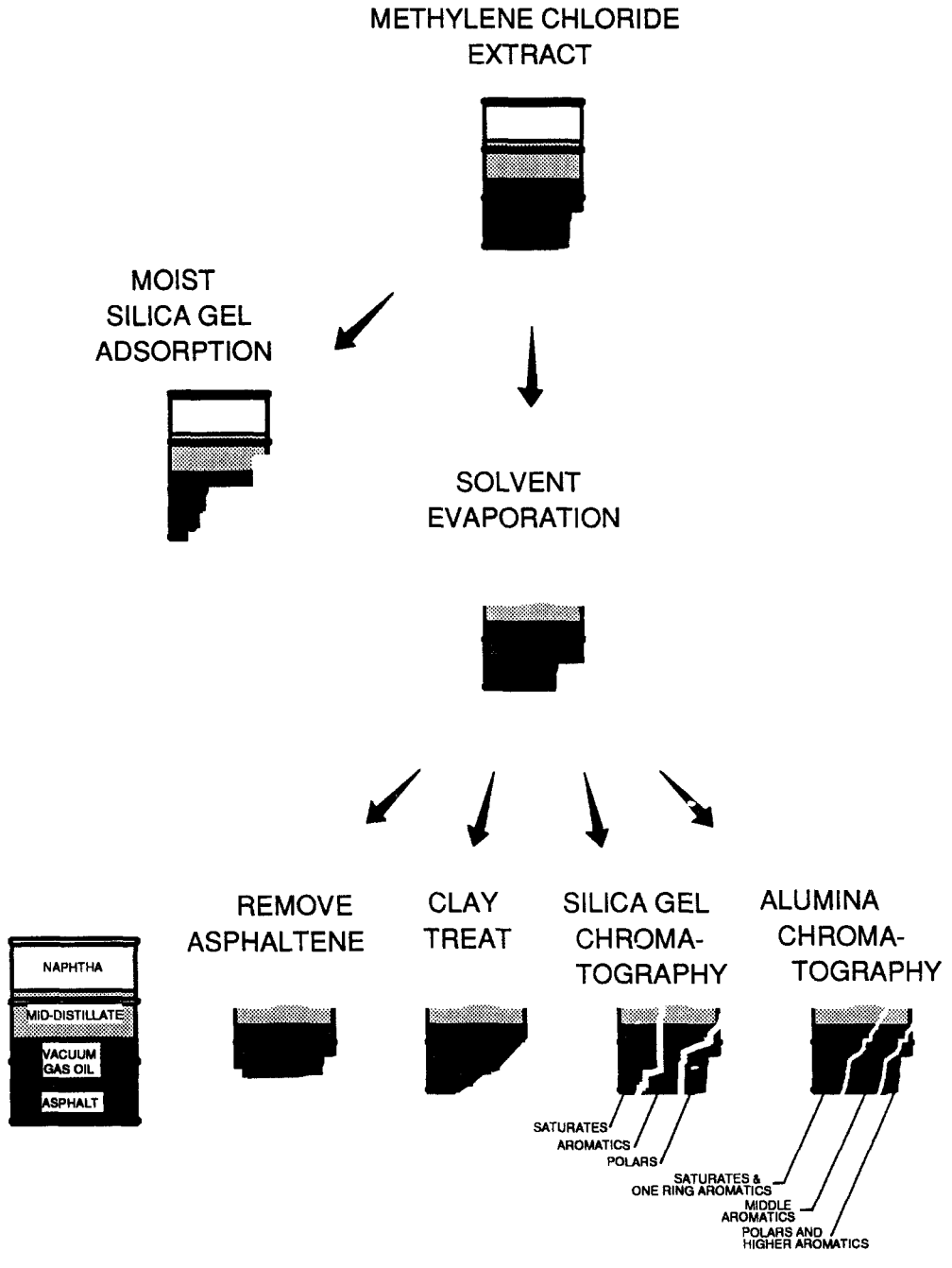

Fig. 5. Impact of fractionation on analytical response.

molecules. Thus, carbon or hydrogen in naphtha and residuum molecules cannot be discriminated.

With appropriately calibrated detectors, HPLC can also be used to quantify the mass and aromaticity in fractions of the barrel with component analysis similar to that presented in the bar chart of Fig. 2.

Mass spectrometry (MS) is most effective when combined with some type of chromatographic separation. MS alone characterizes fractions volatilized into its detector on the basis of molecular weight. However, aromatics and multi-ring napthenes are not resolved so saturates, aromatics, and

polars are generally separated before MS analysis. Low voltage, high resolution MS uses molecular ions to determine the paraffin and napthene content of saturates. Advanced forms of MS are being developed that allow the characterization of vacuum residuum, not sufficiently volatile for detection in current MS methods.

The combination technique GC/MS is perhaps the most powerful method for quantifying individual species within a contaminant oil. It combines the boiling point resolution of GC with the molecular type information of MS. It is often used after extensive sample preparation for two reasons, first to minimize interferences, and second to sufficiently concentrate the trace amounts for quantification.

One direct analysis of the soil contamination is thermogravimetric analysis combined with mass spectrometry (TGA/MS). The weight loss is determined as a function of temperature and the characteristics of the volatilized material are determined by GC/MS. Commercial instruments are now available for characterization of contaminated soil. This is a powerful qualitative probe. Because of large variabilities experienced in small site samples and the very small sample sizes used in this approach, replicate analyses are needed to determine the statistics of the soil samples.

Biological responses can be used to quantify the amounts of certain chemicals or elements in soil or aqueous media. These have been defined as bioassays, in the truest sense of the word, and have historical use for determination of vitamins or minerals in growth media for plants and animals (Butler 1978; Rand & Petrocelli 1985). Bioassays were widely used in quantitative analytical applications before chemical analytical tools could quantify concentrations at the part per million concentrations and below. Today, applications of bioassays have been redirected to assist in characterizing concentrations of contaminants that cause an adverse effect on the survival, growth or reproduction of biota. Such applications are called toxicity tests today (Butler 1978; Rand & Petrocelli 1985).

Analytical chemistry has advanced detection and quantification of contaminant residues to such low levels that it is often difficult to evaluate the biological and ecological significance of exposures at these concentrations. As an analytical tool, the response of plants and animals to contaminant concentrations can be used to define acceptable cleanup targets for contaminated soils or waters. Some or all of the contaminant extracted by analytical chemical approaches may actually be bound in an environmental matrix and not available for uptake by plants and animals, e.g., not bioavailable (Hassett & Banwart 1989; Weber et al. 1983; Knezovich et al. 1987). Reduction of contaminant concentrations to a level acceptable for human and ecological safety is becoming a much more prominent goal than cleanup activities driven by analytical quantification and detection limits.

Biological analytical approaches use quantitative dose-response relationships, fundamental to the field of toxicology, to define safe contaminant concentrations for various conditions. Using survival, growth, reproduction, or physiological health as measures of biological response, safe contaminant concentrations can be defined as concentrations below which no biological response is elicited.

Mammalian and environmental toxicology are applying existing approaches and rapidly developing new techniques for quantifying biological responses to contaminants in environmental media. In particular, toxicity tests with microbes, plants or animals can be applied to determine the rate at which bioremediation treatments can restore soils, sludges or water to conditions that reduce human health concerns and support healthy plant and animal life (Butler 1978; Rand & Petrocelli 1985; Pritchard & Bourquin 1985; Wang et al.1990; Gorsuch et al.1991; Warren-Hicks et al.1989; NRC 1981). Because the biological systems integrate effects of all constituents in the growth media, effects of degradation intermediates, by-products, and residual contaminants can be assessed.

A number of standardized tests utilizing bacterial, plant, and animal responses are being developed and applied to assess the concentrations and bioavailability of contaminants. Alterations in microbial fluorescent activity has become a common, rapid test of contaminant effects in soil and aqueous media (Munkittrick et al. 1991), although the ecological significance of test results may be difficult to extrapolate beyond the screening level of assess-

ment Survival, growth and reproduction of plants and animals are evaluated in a number of standardized test methods (Wang et al.1990; Gorsuch et al.1991; Warren-Hicks et al.1989; NRC 1981), with additional methods continually under development. Tests with terrestrial and aquatic microcosms, which are complex biological arrays in small scale systems, have been used to determine more complex relationships between degradation, persistence, and bioavailability of contaminants in complex environmental matrices. Test results from such scaled test systems often are more easily extrapolated to full scale environmental assessments (NRC 1981). These biological tools are being applied more frequently as part of a comprehensive bioremediation program in order to evaluate the efficacy and safety of waste treatment options.

Impact of statistical design and analysis on experimental linkage

Sound design and the ability to statistically compare experimental results are central to the issue of experimental linkage. Once disciplinary impact, physical structure, chemical composition, biological characteristics, and analytical methodology are considered, statistical design and analysis are necessary for conducting successful comparative studies. This section deals with methods and learnings derived from several Exxon bioremediation studies.

The development of methods for statistical design and analysis of bioremediation experiments should be viewed as an evolving process. First, in bioremediation studies too little is generally known to focus on just a few biodegradation enhancement treatments. Secondly, little is known in situ about the nature of the time-concentration profile of these treatments. What is needed is a design strategy consisting of a series of experiments to find the most effective treatments as well as the best method of application of these treatments in the field.

The strategy depends on the experimental objectives. One possible strategy is:

- to conduct screening experiments to narrow down a relatively large number of treatment types;
- to perform follow-up experiments to identify and perhaps to quantify the effectiveness of a limited number of treatments;
- to conduct other follow-up experiments to determine the application rate and timing for the best treatment types.

Statistical design consists of three interrelated components: treatment design, experimental design, and sampling design. Treatment design concerns the selection of treatments as well as the arrangement of treatments across experimental units. Experimental design consists of methods to control variability. Sampling design concerns the specification of the sampling unit, the determination of the number of samples needed to attain adequate power (defined later), and the determination of sampling times.

Experiments are designed for different purposes. Two types of studies will be considered here: bench studies (*B experiments*) and field studies (*F experiments*). A long-term design strategy typically involves both types of experiments.

Bench experiments are generally conducted to determine the feasibility of certain treatments. A treatment consists of the type of application (e.g., fertilizer), the amount applied per unit of oil, and the times of application. Experimental units for soil bioremediation applications generally can be categorized as completely-mixed reactors (e.g., slurry reactors) or alternatively fixed-bed reactors (e.g., soil pan reactors).

Field experiments are of two major types: large versions of the bench experiments (large-scale experiments) and *in situ* experiments. Large reactors or plots are the analogs of the bench reactor and pan experimental units. If treatments are applied directly to the contaminated areas, the experiment is said to be *in situ*.

The sampling design determines whether the samples are collected at a few time positions or are obtained more frequently. As a consequence, the two types of studies can be further identified according to whether the objective is to estimate absolute or relative changes among treatments at one or more fixed time periods (Δ-modeling) or to assess these differences in terms of rate equations (R-modeling).

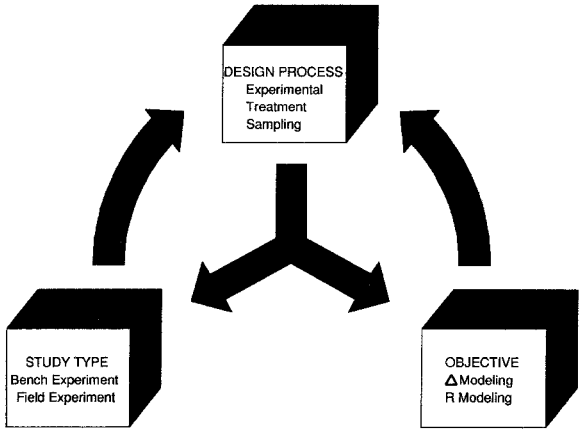

Fig. 6. Statistical design for experimental linkage.

The following figure (Fig. 6) shows the relationships among the design elements, the types of experiments, and the modeling process. Although modeling has not yet been discussed, it will be constrained by the design elements and the modeling objectives will in turn influence the design process.

The success of the statistical design should be quantified. The signal-to-noise ratio and the power to detect meaningful differences will be used as criteria for design characteristics. The signal is the estimated treatment effect, which depends principally on the treatment design, and the noise depends primarily on the experimental design. The signal-to-noise ratio is expressed in terms of a test statistic such as the t or F statistic. The power of a test quantifies the plausibility of the alternative hypothesis, i.e., it gives the probability of accepting the alternative hypothesis. Power can be computed *a priori* or *a posteriori.* If power is computed as part of the design process, estimates of the signal-to-noise ratio are needed.

The P-value (the probability of obtaining a value of the test statistic as large or larger than the observed test statistic under the null hypothesis) quantifies the plausibility of the null hypothesis. If the P-value is small, the null hypothesis is likely to be false.

Experimental design issues

Experimental design consists of methods to reduce the variability or noise in the response data. Bench experiments are discussed first. The sources of variability, besides the treatment effect, are the experimental unit-to-unit variability and the variability within units. Generally both sources of variability will be present when treatment comparisons are made.

The unit-to-unit variability is controlled by thoroughly mixing the substrate prior to dividing it among reactors or pans. If among-unit variability is high, it will almost certainly mask (or perhaps confound) treatment effects of interest. The second source of variability is that due to variation within a reactor or pan. This may be due to an edge effect in pan or reactor samples or due to vertical zones within a reactor.

Because of the internal variability, it may be difficult to obtain a representative sample of the reactor or pan. These samples are obtained in one of two ways: by grab samples or by composite samples. Grab samples are obtained and analyzed separately for the components of interest. Grab sampling is generally expensive and thus a more useful approach is to do composite sampling. Individual grabs, determined randomly or systematically, are mixed prior to analysis resulting in a physical averaging process.

Experimental design is also directly associated with establishing valid error terms. The experimental unit is the entity to which the treatment is applied. For bioremediation bench studies it is generally either a slurry reactor or a pan. Although many sources of variability may be present in an experiment, the experimental unit variability (among units) is the one most appropriately used as the error term.

Large-scale experiments generally have relatively few experimental units (i.e., plots or reactors) and are limited by costs associated with planning and implementation. Nonetheless, the experimental units may incorporate sufficient diversity and complexity to provide adequate representation of the experimental site.

Experimental design issues can be illustrated by considering an experiment for testing the feasibility of various treatments on contaminated soil. The sampling materials for these plots can be obtained

in one of several ways. The region can be subdivided with random samples obtained from the grid of available sites. Each sample would then be used to fill a plot. The difficulty with this approach is that a large number of plots would be needed to be representative of the region. An alternative is to composite the samples prior to applying the treatment. In this case, only one or a few composited plots might be used for each treatment.

In situ experiments can address issues regarding site-specific problems related to the disposition of the contaminant or its clean-up. Spatial and temporal variability at the site must be controlled. If the region has gradients, e.g., systematic moisture or substrate changes, then agricultural blocking-type sampling restrictions should be used. By using appropriate blocking strategies, variation due to gradients in one or more directions can be removed from the error term, thus increasing the precision of the experiment. Another problem with *in situ* experiment is the treatment transfer phenomenon. If sampling sites are not separated sufficiently, treatments can cross over into other treated or control sites.

What are the scale-up implications of these experiments? The major statistical change is that both the between and within experimental unit variability increase in moving from pan to plot to site (or slurry to large-scale reactor) experimental units. Furthermore, the success of the scale-up depends on the similarity of the physical, chemical, and biological properties and processes at different scales. Thus it may become more difficult to assess whether or not a treatment works in the field.

Treatment design

Initially, a study may focus on screening a large number of treatments. The objective is to identify which treatments are worthy of further investigation. Once poor contenders have been eliminated from consideration, the remaining treatments can be studied in depth.

A second experimental objective is to determine the optimum settings for certain process variables. For example, a fertilizer may be known to be an effective bioremediation agent. The objective then is to establish optimal values for the rate and frequency of application of the fertilizer. Traditionally, response-surface designs have been used to find the optimum operating values. Alternatively, orthogonal array designs are useful in identifying optimal conditions (Phadke 1989).

Sampling design

The sampling design, as considered here, concerns the number of samples required and the frequency of sampling. It is directly related to the power of the statistical tests as explained below. Bioremediation studies are nearly always longitudinal, i.e., the effect of the treatment occurs over time. How the sampling is done depends on whether 'difference' or 'rate' modeling is of principal interest.

R-Modeling – experiments designed to model rate behavior

Any measurable component in a bioremediation experiment has a response curve over time. The baseline response curve is that obtained by natural processes without the addition of any treatments. This section examines the sampling approaches used to compare the response curves of one or more treatments to that of a 'control.'

The sampling design for comparing some parameter (e.g., the rate constant of a first-order rate equation) of the response curves for two treatments consists of two parts. First, for a given sample size for each treatment, determine the optimum allocation among sampling times and replications at those times to minimize the standard error of the estimated parameter difference. The second part concerns how the power of this test is related to the overall sample size.

If the researcher has a rough idea of the functional form of the model, then a program such as JMP Design (Sall & Lehman 1992) could be used to find the D-optimal design. On the other hand, if little is known about the response function, an approach is to space the samples over the length of the experiment, but replicate at every other or every third sampling time. This allows many functional types to

be fit with reasonable, but not optimal, efficiency. The sample spacing is generally at equal time periods, but if rapid change occurs at the beginning of the experiment, then a geometric sequence might be preferable.

Once the sampling pattern has been established, a power analysis should be done to determine if differences of interest can be detected with a high probability. Power studies can be done *a priori* with programs such as JMP (SAS Institute,1989) in a 'what if' scenario. The overall sample size can be increased, if necessary, without substantially changing the sampling pattern by increasing the number of replications, by filling in the unsampled time intervals, or by reducing the number of treatments.

Δ-Experiments – experiments designed to estimate differences at given times

Certain experiments are conducted to answer the following type of question: Is treatment A more effective than the control or perhaps treatment B? This section will examine sampling schemes to maximize the power of statistically detecting a difference.

The difficulty with this type of experiment is that the differences change with time. This raises the question of what is meant by a difference, since natural biodegradation processes may eventually cause response curves to converge. Therefore, a target time must be identified.

In Δ-modeling experiments all of the sampling is done at time zero and at one or more target times. Generally only one or two target times are selected; otherwise, the experiment essentially becomes an R-modeling experiment. It might seem that comparisons should be made at each of the target times. This typically will not work.

A better analysis is to consider the treatment design to be crossed with time. The interaction effects are then of principal interest since they test relative changes between the treatment and controls. Adjustment at time zero is automatic and the degrees of freedom associated with the error term are larger than an analysis at each sampling time.

Hybrids

Putting all of the sampling effort at time 0 and at a few target times is efficient for detecting a change at these target times. Two objections may arise. First, the target times may be misplaced. An incorrect conclusion may be drawn that no differences exist when in fact they exist at other time periods. Second, little information is available about rate equations unless relatively simple response functions are true.

A hybrid sampling scheme can overcome these objections, at least partially. The strategy is to take intermediate samples, but to maintain adequate numbers of replicates at the target times. This will result in reasonable power for treatment comparisons and it will allow more complex rate equations to be investigated.

Model building

This section discusses model building issues for both Δ- and R-type modeling experiments. Both linear and nonlinear models will be considered.

Δ-Modeling

Δ-modeling studies are designed to ask whether or not treatments differ at one or more time periods after the initiation of an experiment. Linear statistical models, often on a log transformed scale, are the most common approach for testing treatment effects. For simplicity, assume a one-way treatment design with 'repeated' measurements over time. Two cases must be distinguished: measurements made over time on different experimental units and repeated measures made on the same experimental unit.

Sacrificed experimental units

The reactors in bench studies are often sacrificed when the sample material is collected in order to acquire complete recovery of the analytes. Likewise, plot experiments are tilled between sampling times, which in effect creates new experimental units, and different locations may be chosen in site studies. In all of these cases, each observation for a variable of interest is measured on a different experimental unit.

The model is then:

$$C_{ijk} = m + q_i + t_j + g_{ij} + e_{ijk}$$

where C_{ijk} is the component response (perhaps transformed) for the k^{th} replicate on the i^{th} treatment at time j. The treatment-by-time interaction, g_{ij}, is the principal effect of interest. It measures the differential effects of the treatments over time. In other words, if the treatments increase biodegradation, but at the same rate, then the interaction effect will be zero.

The linear model as expressed above must satisfy certain assumptions. Besides additivity, the error term is assumed to be normally and independently distributed with constant variance. Independence may not be satisfied due to autocorrelation over the sampled time periods. In addition, the variance may be a function of the mean response which violates the constancy of variance assumptions. These assumptions need to be verified by statistical diagnostics during the model building process.

Conserved experimental unit with repeated measures

Pan experiments or plot experiments in which the material is not substantially changed between sampling times have a fundamentally different analysis strategy. This type of model is called a repeated measures model.

The principal difference is that two error terms are used for testing purposes. The model is given by:

$$C_{ijk} = m + q_i + f_{ik} + t_j + g_{ij} + e_{ijk}$$

where the terms are defined as above except f_{ik} is the variability of experimental units within treatments and e_{ijk} is the interaction between experimental units and time. These are assumed to be random effects in the model with the standard assumptions of normality, independence, and constancy of variance.

The repeated measures model may provide a more precise estimate of the interaction effects (using the second error term). However, the assumptions of this model must be verified. Autocorrelation is more likely to be a problem for this model. Adjustments for this autocorrelation should be made, which has the effect of reducing the sample size and thus decreasing the precision of the estimates or the power of the tests.

R-Modeling

Sampling over time allows more complex rate models to be fit. First-order models are most commonly fit in which the rate constant is often generalized to be a function of other variables such as the nitrogen-to-contaminant ratio. The simplest form of this model in differential form is:

$$d\,[c(t)]/dt = (\Theta_1 + \hat{\Theta}_2\, x)\, c(t)$$

where x is a dummy variable representing the presence or absence of a single treatment. The treatment effect is estimated by $\hat{\Theta}_2$, which characterizes the difference in the rate constants.

The nature of the error term determines whether a linear or nonlinear model should be fit. If the error term is multiplicative in the integral form of the model, then a linear model on the log-transformed scale may be possible. If the error term is additive, nonlinear models must be fit (Bates & Watts 1988). Constancy of variance is a critical assumption in these models. If the variance is a function of the mean, then these models can be adjusted by iteratively weighting the regression models.

The models described so far assume that the underlying form of the model does not depend on the treatment. This is what allows the simplification using dummy variables. If the form of the underlying model does not fit all treatments well, then it may be possible to improve the fit by increasing the complexity of the underlying model, e.g., by using multiple rate constants. This may create a difficult or intractable estimation problem. Furthermore, it may mask simpler underlying models which depend on the treatment, e.g., one treatment may be first order and another may be logistic.

A further modification is to incorporate coordinates into the model. As an example, bioremediation of oil may also depend on the physical properties of the oil, temperature, or related adapting microbial communities. These variables, which may be time dependent, need to be integrated into the

model. The likely result is that the rate parameter is no longer a constant but now a time-dependent variable.

Summary

Reliable comparison of bioremediation experiments between scales leading to predictive capability and related commercialization of petroleum site bioremediation is a challenge yet remaining in this fast-paced field. Issues have been identified at a number of levels in this paper and no claim has been made that this analysis has been comprehensive. Specific issues include:

- The commercialization process is inherently interdisciplinary and experimental methods borrowed from specific disciplines may limit comparison across disciplines.
- Two broad strategies exist on which to base experimental comparisons. One strategy strives to create experiments that have exact geometric, mechanical, thermal, chemical and biological similarity between experiments and interpret differences in responses to 'scale-specifc' relationships. The second strategy recognizes the difficulties, perhaps impossibilities creating biologically similar experiments and uses modeling to establish the relationships between geometric, mechanical and thermal characteristics with best efforts to maintain chemical similarity. Differences in response are considered phenomenologically with future advances in microbial ecology needed to establish rate relationships for improved experimental linkage.
- Petroleum oils and products are compositionally complex and quantitative analytical results and conclusions depend strongly not only on the general methods chosen but on relatively slight variations accepted within methods. Because of the possible impact of apparently minor variations in methods on results, meaningful comparison of results among investigators may be difficult.
- Petroleum-contaminated sites may have great variability in composition owing to diverse products spilled at random times over perhaps a century. Weathering processes lead to compositional changes and pollutant compositions on the contaminated soils may be site-specific on scales as small as a few feet. The impact of this chemical variability on the bioremediation response is currently unknown.
- Bench-scale methods exist to assay for biodegradation potential and rates on site samples but linkage of these results back to the field experiment is limited. This is influenced by problems of changing the environment and community structure upon sampling as well as in the biodegradation assay. Techniques that make use of biodegradation-resistant internal chemical markers promise to aid in the measurement of bioremediation effects, but apply best to cases where the contaminating oil comes from a single source at a single time.
- Analysis of petroleum contamination using bioassays or the responses of biological organisms is an emerging area that may help link experiments based upon their biological responses.
- Statistical designs are central to the success of both bench and field experiments as well as to the comparison among them. A balance must be found between the number of treatments studied and the precision of estimating treatment effects and rate models. Since bench experiments are generally less variable than field experiments, treatment screening should be done at the bench level (assuming an acceptable level of similarity), whereas field studies should be confirmatory and more narrowly focused. The treatment design must be complemented by a sound experimental design in order to structure the experiment to minimize the error variability.
- The dual objectives of a bioremediation study: 1) to estimate or test treatment effects or 2) to estimate a rate model, are both the result of and the guidance for the sampling design. If both objectives are important to the investigator, hybrid sampling plans are appropriate in which the time-series samples are supplemented by replications at targeted comparison times. In any case, whatever the experimental objective, after selecting the most important treatment comparisons (i.e., the treatment design) and optimizing

the precision of the comparisons (i.e., the experimental design), the power to establish differences among treatment effects or rate models must be acceptable (the sampling design).

Reliable predictive linkage of bioremediation experiments is a difficult and challenging endeavor that will require a relatively large-scale effort and may also require the development of new methods before the ultimate goal is achieved.

References

Aeckersberg F, Bak F & Widdel F (1991) Anaerobic oxidation of saturated hydrocarbons to CO_2 by a new type of sulfate-reducing bacterium. Arch. Microbiol.156: 5-14

Aggarwal PK & Hinchee RE (1991) Monitoring in situ biodegradation of hydrocarbons by using stable carbon isotopes. Environ. Sci. Technol. 25: 1178-1180

Atlas RM (1981) Microbial degradation of petroleum hydrocarbons: an environmental perspective. Microbiol. Rev. 45: 180-209

Atlas RM & Bartha R (1987) Microbial Ecology: Fundamentals and Applications. Benjamin/Cummings, Menlo Park, California

Austin BJ, Calomiris JJ, Walker JD & Colwell RR (1977a) Numerical taxonomy and ecology of petroleum-degrading bacteria. Appl. Environ. Microbiol. 34: 60-68

Austin BJ, Colwell RR, Walker JD & Calomiris JJ (1977b) The application of numerical taxonomy to the study of petroleum-degrading bacteria isolated from the aquatic environment. Dev. Ind. Microbiol. 18, 685-695

Austin B, Garges S, Conrad B, Harding EE, Colwell RR, Simidu U & Taga N (1979) Comparative study of the aerobic heterotrophic bacterial flora of Chesapeake Bay and Tokyo Bay. Appl. Environ. Microbiol. 37: 704-714

Bailey JE & Ollis DF (1987) Biochemical Engineering Fundamentals (p. xix). McGrawHill. New York

Bartha R & Atlas RM (1977) The microbiology of aquatic oil spills. Adv. Appl. Microbiol. 22: 225-266

Bates DM & Watts DG (1988) Nonlinear Regression Analysis and its Applications, Wiley, New York

Bazylinski DA, Wirsen CO & Jannasch HW (1989) Microbial utilization of naturally occurring hydrocarbons at the Guaymas Basin hydrothermal vent site. Appl. Environ. Microbiol. 55: 2832-2836

Bertrand JC, Caumette P, Mille G, Gilewicz M & Denis M (1989) Anaerobic biodegradation of hydrocarbons. Sci. Progr. (Oxford) 73: 333-350

Bertrand JC, Almallah M, Acquaviva M & Mille G (1990) Biodegradation of hydrocarbons by an extremely halophilic archaebacterium. Lett. Appl. Microbiol. 11: 260-263

Blasig R, Huth J, Franke P, Borneleit P, Schunck W-H & Müller H-G (1989) Degradation of long-chain n-alkanes by the yeast *Candida maltosa*. Appl. Microbiol. Biotechnol. 31: 571-576

Blackburn JW (1987) Prediction of organic chemical fates in biological treatment systems. Environ. Progress 6: 217-223

Blackburn JW (1988a) Problems in and potential for biological treatment of hazardous wastes. Proc. 81 st Annual APCA (now Air and Waste Management Association) Meeting. Pittsburgh

Blackburn JW (1988b) Is there an 'uncertainty principle' in microbial waste treatment? In: Huntley ME (Ed) Biotreatment of Agricultural Wastewater (pp 149-161). CRC Press, Inc. Boca Raton, Florida

Blackburn JW (1989) Improved understanding and application of hazardous waste biological treatment processes using microbial systems analysis techniques. Haz. Waste & Haz. Materials 6: 173-193

Brown EJ & Braddock JF (1990) Sheen screen, a miniaturized most-probable number method for enumeration of oil-degrading microorganisms. Appl. Environ. Microbiol. 56: 3895-3896

Brown EJ, Resnick SM, Rebstock C, Luong HV & Lindstrom JE (1991) UAF radiorespirometric protocol for assessing hydrocarbon mineralization potential in environmental samples. Biodegradation 2: 121-127.

Buckley EN, Jonas RB & Pfaender FR (1976) Characterization of microbial isolates from an estuarine ecosystem: relationship of hydrocarbon utilization to ambient hydrocarbon concentrations. Appl. Environ. Microbiol. 32: 232-237

Butler GC (Ed) (1978) Principles of Ecotoxicology. Wiley, New York

Butler EL, Douglas GS, Steinhauer WS, Prince RC, Aczel T, Hsu CS, Bronson MT, Clark JR & Lindstrom JE (1991) Hopane, a new chemical tool for measuring oil biodegradation. In: Hinchee RE & Olfenbuttel RF (Eds) On-site Reclamation. Processes for Xenobiotic and Hydrocarbon Treatment (pp 515-521). Butterworth-Heinemann, Boston

Cooney CL (1983) Bioreactors: design and operation. Science 219: 728-733

Chianelli RR, Aczel T, Bare RE, George GN, Genowitz MW, Grossman MJ, Haith CE, Kaiser FJ, Lessard RR, Liotta R, Mastracchio RL, Minak-Bernero V, Prince RC, Robbins WK, Stiefel EI, Hinton SM, Bragg JR, McMillen SJ & Atlas RM (1991) Bioremediation technology development and application to the Alaskan spill. Proceedings of the 1991 International Oil Spill Conference (pp 549-558). American Petroleum Institute, Washington DC

Chosson P, Lanau C, Connan J & Dessort D (1991) Biodegradation of refractory hydrocarbon biomarkers from petroleum under laboratory conditions. Nature 351: 640-642

DiGrazia PM, King JMH, Blackburn JW, Applegate BA, Bienkowski PR, Hilton BL & Sayler GS (1991) Dynamic response of naphthalene biodegradation in a continuous flow soil slurry reactor. Biodegradation 2: 81-91

Douglas GS, McCarthy KJ, Dahlen DT, Seavey JA, Steinhauer WG, Prince RC & Elmendorf DL (1992) The use of hydrocarbon analyses for environmental assessment and remediation.

In: Kostecki PT & Calabrese EJ (Eds) Contaminated Soils; Diesel Fuel Contamination (pp 1-21). Lewis Publishers, Chelsea, MI

Eastcott L, Shiu WY & Mackay D (1988) Environmentally relevant physical-chemical properties of hydrocarbons: a review of data and development of simple correlations. Oil & Chem. Pollut. 4: 191-216

Efron B & Tibshirani R (1991) Statistical data analysis in the computer age. Science 253: 390-395

Elder RS, Thompson WO & Myers RH (1980) Properties of composite sampling procedures. Technometrics 22: 179-186

Foght JM, Fedorak PM & Westlake DWS (1990) Mineralization of (^{14}C)hexadecane and (^{14}C)phenanthrene in crude oil; specificity among bacterial isolates. Can. J. Microbiol. 36: 169-175

Giddings JM & Eddlemon GK (1979) Some ecological and experimental properties of complex aquatic microcosms. Intern. J. Environmental Studies 13: 1-5

Gorsuch JW, Lower WR, Lewis MA & Wang W (1991) Plants for Toxicity Assessment: Second Volume. ASTM STP 1115. Amer. Soc. Test. Mater., Philadelphia

Gough MA & Rowland SJ (1990) Characterization of unresolved complex mixtures of hydrocarbons in petroleum. Nature 344: 648-650

Gough MA, Rhead MM & Rowland SJ (1991) Biodegradation studies of unresolved complex mixtures of hydrocarbons; model UCM hydrocarbons and the aliphatic UCM. Org. Geochem.18: 17-22

Grady CPL, Jr. (1985) Biodegradation: its measurement and microbiological basis. Biotechnol. Bioengineer., 27(5): 660-674

Gundlach ER, Boehm PD, Marchand M, Atlas RM, Ward DM & Wolfe DA (1983) The fate of Amoco Cadiz oil. Science 221: 122-129

Hartman S, de Bont JAM & Harder W (1989) Microbial metabolism of short-chain unsaturated hydrocarbons. FEMS Microbiol. Rev. 63, 235-264.

Hassett JJ & Banwart WL (1989) The sorption of nonpolar organics by soils and sediments. In: Sawhney BL & Brown K (Eds) Reactions and Movement of Organic Chemicals in Soils (pp. 31-44). American Society of Agronomy, Madison, WE

Hommel RK (1990) Formation and physiological role of biosurfactants produced by hydrocarbon-utilizing microorganisms. Biodegradation 1: 107-119

Johnstone RE & Thring MW (1957) Pilot Plants, Models, and Scale-up Methods in Chemical Engineering, McGraw-Hill, New York

Kennicutt MC (1988) The effect of biodegradation on crude oil bulk and molecular composition. Oil & Chem. Pollut. 4: 89-112

Knezovich JP, Harrison FL & Wilhelm RG (1987) The bioavailability of sedimentsorbed organic chemicals: A review. Water Air Soil Pollut. 32: 233-245

Linz DG, Neuhauser EF & Middleton AC (1991) Perspectives on bioremediation in the gas industry. In: Sayler GS, Fox R & Blackburn JW (Eds) Environmental Biotechnology for Waste Treatment (pp 25-36). Plenum Press, New York

Lindstrom JE, Prince RC, Clark JR, Grossman MJ, Yeager TR, Braddock JF & Brown EJ (1991) Microbial populations and hydrocarbon biodegradation potentials in fertilized shoreline sediments affected by the T/V Exxon Valdez oil spill. Appl. Environ. Microbiol. 57: 2514-2522

Litchfield CD (1991) Practices, potential, and pitfalls in the application of biotechnology to environmental problems. In: Sayler GS, Fox R & Blackburn JW (Eds) Environmental Biotechnology for Waste Treatment (pp 147-157). Plenum Press, New York

Leahy JG & Colwell RR (1990) Microbial degradation of hydrocarbons in the environment. Microbiol. Rev. 54: 303-315

McFarland MJ, Sims RC & Blackburn JW (1991) Use of treatability studies in developing remediation strategies for contaminated soils. In: Sayler GS, Fox R & Blackburn JW (Eds) Environmental Biotechnology for Waste Treatment (pp 163–174). Plenum Press, NewYork

Munkittrick KR, Power EA & Sergey GA (1991) The relative sensitivity of Microtox, daphnid, rainbow trout, and fathead minnow acute lethality tests. Environ. Toxicol. Water Quality 6: 35-62

Needham J (1935) Order and Life. The MIT Press, Cambridge, Massachusetts

NETAC (1992) Interim Protocol for Bioremediation Product Testing. National Environmental Technology Applications Corporation, University of Pittsburgh Applied Research Center, Pittsburgh

NRC (1981) Testing For Effects of Chemicals on Ecosystems. National Research Council. National Academy Press, Washington, DC.

Oberbremer A & Müller-Hurtig R (1989) Aerobic stepwise hydrocarbon degradation and formation of biosurfactants by an original soil population in a stirred reactor. Appl. Microbiol. Technol. 31: 582-586

Oudot J (1984) Rates of microbial degradation of petroleum components as determined by computerized capillary gas chromatography and computerized mass spectrometry. Mar. Env. Res.13: 277-302

Phadke MS (1989) Quality Engineering Using Robust Design. Prentice Hall, Englewood Cliffs, New Jersey.

Prince RC (1992) Bioremediation of oil spills, with particular reference to the spill from the Exxon Valdez. In: Fry JC, Gadd GM, Herbert RA, Jones CW & Watson-Craik IA (Eds) Microbial Control of Pollution (pp.19-34). Society for General Microbiology Symposium 48, Cambridge University Press

Pritchard PH & AW Bourquin (1985) Microbial toxicity studies, In: Rand GM & SR Petrocelli (Eds) Fundamentals of Aquatic Toxicology (pp 177-217). Hemisphere Publishing, New York

Prokop A (1983) Reactor design fundamentals, hydrodynamics, mass transfer, heat exchange, control and scale-up. Biochemical Engineering (pp 355-376). American Chemical Society, Washington DC

Rao PSC, Bellin CA & Brusseau ML (1992) Coupling of sorption, biodegradation, and transport of organic contaminants in soils and aquifers: paradigms and paradoxes. In: Linn D et al. (Ed) Sorption and Degradation of Agrichemicals in Soils. Special Publication Soil Sci. Soc. Amer., Madison, WI

Rand GM & Petrocelli SR (Eds) (1985) Fundamentals of Aquatic Toxicology (pp 1 -30). Hemisphere Publishing, New York

Ramadan MA, El-Tayeb OM & Alexander M (1990) Inoculum size as a factor limiting success of inoculation for biodegradation. Appl . Environ. Micro. 56: 1392-1396

Rogerson A & Berger J (1983) Enhancement of the microbial degradation of crude oil by the ciliate *Colpidium colpoda.* J. Gen. Appl. Microbiol. 29: 41-50

Sall J & Lehman A (1992) JMP Design User's Guide: Version 2, SAS Institute, Inc., Cary, NC

SAS Institute (1989) JMP Users Guide, Cary, NC

Smith MR (1990) The biodegradation of aromatic hydrocarbons by bacteria. Biodegradation 1: 191-206

Snellnam EA, Collins RP & Cooke JC (1988) Utilization of fuel oils by fungi isolated from oceanic tarballs. Lett. Appl. Microbiol. 6: 105-107

Sohngen NL (1913) Benzin, Petroleum Paraffinöl, und Paraffin als Kohlenstoff und Energiequelle fur Mikroben. Zentr. Bakt. Parasitenk., Abt. II, 37: 595-605

US Environmental Protection Agency (1989) Guide for Conducting Treatability Studies under CERCLA. Interim Final Report EPA/540/2-89/058. Office of Research and Development and Office of Emergency and Remedial Response. Washington, DC

Venkateswaran K, Iwabuchi T, Matsui Y, Toki H, Hamada E & Tanaka H (1991) Distribution and biodegradation potential of oil-degrading bacteria in northeastern Japanese coastal waters. FEMS Microbiol. Ecol. 86: 113-122

Vestal R, Cooney JJ, Crow S & Berger J (1984) The effects of hydrocarbons on aquatic microorganisms. In: Atlas RM (Ed) Petroleum Microbiology (pp. 475–505). Macmillan, New York

Wang W, Gorsuch JW & WR Lower (1990) Plants for Toxicity Assessment. ASTM STP 1091. Amer. Soc. Test. Mater., Philadelphia.

Walter U, Beyer M, Klein J & Rehm H-J (1991) Degradation of pyrene by *Rhodococcus sp.* UW1. Appl. Microbiol. Biotechnol. 34: 671-676

Warren-Hicks W, Parkhurst BR & Baker SS Jr (Eds) (1989) Ecological Assessments of Hazardous Waste Sites: A field and laboratory reference document. EPA 600/3-89/013, Washington, DC

Watkinson RJ & Morgan P (1990) Physiology of aliphatic hydrocarbon-degrading microorganisms. Biodegradation 1: 79-92

Weber WJ Jr, Voice TC, Pirbazari M, Hunt GE & Ulanoff NM (1983) Sorption of hydrophobic compounds by sediments, soils and suspended solids. II. Sorbent evaluation studies. Water Res.17: 1443-1452

Weibe WJ (1971), Perspectives in Microbial Ecology, In: Odum G. (Ed.) Fundamentals of Ecology, W. B. Sanders, Chapter 19.

Woodyard PT (1991) Considerations in the selection of environmental biotechnology as viable in field-scale waste treatment applications. In: Sayler GS, Fox R & Blackburn JW (Eds) Environmental Biotechnology for Waste Treatment (pp 37–45). Plenum Press, New York