RISK ASSESSMENT FOR CONTAMINATED SITES

by

Albert K. Essiam

B.S., University of Idaho
(1991)

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN CIVIL AND ENVIRONMENTAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

April 1995

Copyright © 1995 Massachusetts Institute of Technology

Signature of Author

Department of Civil and Environmental Engineering

April 28, 1995

Certified by

Professor Herbert H. Einstein

Department of Civil and Environmental Engineering

Thesis Supervisor

Accepted by

Dr. Joseph Sussman

Chairman, Graduate Thesis Committee

JUN 27 1995
RISK ASSESSMENT FOR CONTAMINATED SITES

by

ALBERT ESSIAM

Submitted to the Department of Civil and Environmental Engineering on April 28, 1995 in partial fulfillment of the requirement for the degree of Master of Science in Civil and Environmental Engineering.

ABSTRACT

The U.S. Environmental Protection Agency (EPA) has estimated that there are well over 20,000 solid and liquid wastes sites in the U.S. Many of the sites are the result of abandoned mining and manufacturing operations, and are known sources of chemical contamination. Costs for the cleanup of these sites has been estimated at up to $10-20 billion [Staples et. al, 1985]. Ideally, we would want every trace of contaminant cleaned from these polluted sites. This ideal situation can be realized with unlimited financial resources and the appropriate technology to clean all contaminated sites. Unfortunately, in real life we have limited resources (money, technology, and even time). Therefore, we must be satisfied with a clean-up that ensures safety from all damaging effects of pollutants without altering the beneficial products derived from processes which generate chemical waste; thus, there is a need for an accurate method of evaluating the risks contaminants pose to humans and the environment.

The first step in cleaning contaminated sites is to determine the physical characteristics, distribution and extent of the contaminants at a site. This process is known as hazard assessment. Hazard assessment begins by obtaining samples from the site and testing the samples in a laboratory to determine their chemical and physical properties. Samples are usually taken on a regular grid or in a random fashion over the contaminated area. The number of samples that are obtained initially depend on the amount of money available for the clean up project and the extent of contamination. When there are extensive resources available to take "enough samples," we can apply kriging or other interpolation and extrapolation techniques to estimate contaminant concentrations at locations from which we have no sample data. For classical geostatistical techniques, "enough samples" is defined as the number of samples which can help us derive a variogram. A variogram is a mathematical description of how the contaminant concentrations vary over the site in a specified direction. In cases where we have limited resources and cannot apply these geostatistical techniques, there are other tools such as the Bayesian updating technique which allows the engineer to determine an initial distribution of the contaminants and update his knowledge as more sample data is obtained from the site. The initial distribution the engineer uses to describe the contaminant distribution is either based on prior experiences from similar sites or on prior estimation exercises carried out at the site.

Determining the potential effects of these chemicals on humans and the environment is known as risk assessment. A number of mathematical models are available for assessing the probability of developing a deleterious health effect. The EPA has developed a set of equations which can be used in assessing the human and ecological risks from exposure to a particular contaminant. In addition to the EPA's equations, there exist other mathematical models which effectively evaluate human and ecological risk from chemical exposure.

This thesis discusses the details of hazard and risk measurement for a contaminated site and illustrates how information from risk assessment can be applied in making decisions about a contaminated site.

Professor Herbert H. Einstein
Department of Civil and Environmental Engineering
Thesis Supervisor
Acknowledgements and Dedication

I would like to express my sincere gratitude to Professor Herbert H. Einstein who advised and guided me through this thesis project, and to Professor Daniele Veneziano for his encouragement.

This thesis is dedicated to my mum, Mrs. Alice Essiam, who has been my source of encouragement and support throughout my life.
Table of Contents

CHAPTER 1 - INTRODUCTORY CONCEPTS AND DEFINITIONS
   FOR CONTAMINATED SITES ........................................... 7
   1.2 Definitions and Concepts ..................................... 11
   1.3 What should be regulated, hazard or risk? ............... 16

CHAPTER 2 - HAZARD ASSESSMENT FOR CONTAMINATED SITES
   ................................................................................. 18
   2.1 Obtaining Data for Site Hazard Assessment .............. 18
   2.2 Point and Global Estimates of Contaminant Concentrations
         ................................................................. 23
         2.2.1 Global Estimation Methods ............................. 23
         2.2.2 Point Estimation Methods ............................... 28
   2.3 The Bayesian Approach For Modeling Contaminant Extent
         ................................................................. 37
   2.4 Delineating Contaminant Boundaries Using Binary Logistic Regression
         ................................................................. 39
   2.4 Value of Perfect Information .................................... 44
   2.5 Value of Imperfect Information ................................. 47
   2.6 Value of Information and Decision Structuring ............ 52

CHAPTER 3 - RISK ASSESSMENT
   (Human and Environmental Risk From Exposure to Contaminants) ........................................... 53
   3.1 Introduction ......................................................... 53
   3.2 Quantification of Risk ............................................. 55
         3.2.1 Statistical or Empirical Approach ...................... 55
         3.2.2 Stochastic Models ......................................... 65
   3.3 Concluding Remarks .............................................. 71

CHAPTER 4 - MANAGING RISKS POSED BY A CONTAMINATED SITE ........................................... 72
   4.1 Introduction ......................................................... 72
       4.1.1 What Is an Acceptable Risk? ............................ 83
       4.1.2 Perception of Risk ........................................... 84
   4.2 Managing Risks ..................................................... 86
       4.2.1 Absolute Safety Approach ............................... 87
       4.2.2 'As Low as Reasonably Achievable' Approach (ALAR) 88
       4.2.3 Safety Standard Approach ............................... 88
   4.3 Risk Management in Practice ................................. 89

CHAPTER 5 - SUMMARY AND CONCLUSIONS ........................................... 92
   5.1 Summary ............................................................ 92
   5.2 Concluding Remarks ............................................ 98
List of Figures

| Figure 1.1:     | An influence diagram of the decision to clean a contaminated site | 10 |
| Figure 1.2:     | Schematic section through a valley containing stationary deposits | 13 |
| Figure 2.1:     | Data Posting For 42 Nitrate Sample Values                      | 20 |
| Figure 2.2:     | Contour Map For 42 Nitrate Samples                             | 21 |
| Figure 2.3:     | Symbol Map for 42 Nitrate Sample Data                          | 22 |
| Figure 2.4:     | An example showing the polygon of influence of a sample        | 24 |
| Figure 2.5:     | Construction of a polygon of influence using the method of perpendicular bisectors. Figures (a) to (f) show the steps in constructing a region within which the central sample is closer than any other sample. | 26 |
| Figure 2.6:     | Defining the polygon on the border of the global area. (a) shows a polygon that cannot be closed by the method of perpendicular bisectors between data points. Alternatively the polygon can be closed by a natural limit such as the lease boundary in (b) or by limiting the distance from a sample to the edge of a polygon as shown in (c). | 27 |
| Figure 2.7:     | An example of cell declustering                                  | 28 |
| Figure 2.8:     | The data configuration shown in this figure is used to illustrate several point estimation methods in the following sections. The goal is to estimate the value of V at the point 65E, 137N, located by the arrow, from the surrounding seven data values. | 31 |
| Figure 2.9:     | A perspective view showing the discontinuities inherent in polygonal estimates. | 31 |
| Figure 2.10:    | An example of an exponential variogram model                    | 35 |
| Figure 2.11:    | Using Sample Data to Update Uncertainty about parameter \( \theta \) via Baye's Theorem. | 38 |
| Figure 2.12:    | Random Contour of Sample Data Showing the 50% probability line of contaminants exceeding 550 ppm | 42 |
| Figure 2.13:    | 50% probability boundary obtained from spatial interaction model | 44 |
| Figure 2.14:    | Sampling Decision Tree With Perfect Information Alternative     | 46 |
| Figure 2.15:    | Decision Tree to Estimate the Value of Imperfect Sampling Information | 51 |
| Figure 3.1:     | Log-log plot of Risk versus Dose for three Response Models      | 69 |
| Figure 4.1:     | Three different shapes for utility functions                   | 75 |
| Figure 4.2a:    | A value tree for evaluating chemical contaminant risks          | 76 |
| Figure 4.2b:    | A value tree for evaluating chemical contaminant risks with relative weights assigned to each value and attribute | 81 |
| Figure 4.3:     | Decision cycle for Evaluating Potential Hazard Waste Site Problems | 91 |
List of Tables

Table 1: Distances to the sample values in the vicinity of 65E, 137N .................. 29
Table 2: Conditional Probabilities Characterizing Consultant's Prediction Ability .... 48
Table 3: Posterior Probabilities for State of Contamination at a Site Depending on a Consultant's Prediction .................................................. 50
Table 4: Equation for Estimating carcinogenic risk, R, or hazard index, HI from human exposure to contaminated soil ...................................................... 59
Table 5: Equation for estimating the human health risk by exposure to contaminated surface water ................................................................. 61
Table 6: Equation for calculating human health risk from exposure to contaminated groundwater ................................................................. 63
Table 7: Matrix of weights and scores for three chemical plant sites .................. 82
CHAPTER 1- INTRODUCTORY CONCEPTS AND DEFINITIONS
FOR CONTAMINATED SITES

1.1 Introduction

Before the cleanup of a contaminated site, the type of contaminant, the extent of contamination, and the spatial distribution of contaminants must be known in order to select an appropriate remedial method and estimate the cost of the cleanup project. The conventional method of determining pollutant concentrations is by direct sampling and laboratory analysis. However, due to the high cost of sampling and laboratory testing, it is impractical to use this method to determine the spatial distribution of pollutants over very large areas.

Suppose we need to determine the spatial distribution of contaminants over a large area. First, a small section of the contaminated site is selected and sampled. After determining the contaminant concentrations for that small region, a mathematical model is fit to the sample data to describe the spatial distribution of the contaminants. This mathematical model then can be used to estimate contaminant concentrations at locations where samples were not taken. Without the model, no inferences can be made about the unknown values at locations that were not sampled. Perhaps the most desirable information needed to derive an appropriate model is a description of how the contaminants were generated. In certain situations, the physical or chemical processes that generated the contaminants might be known in sufficient detail so that an accurate description of the entire contaminant profile can be made from only a few sample values. In such situations a deterministic model may be appropriate. Unfortunately very few earth science processes are understood well enough to allow one to represent contaminant distributions with deterministic models. Although the physics and chemistry of many fundamental processes about contaminants are known, the contaminant concentration at a particular location on a site is typically the end result of a vast number of processes whose complex interactions cannot yet be described quantitatively [Isaaks et. al., 1989]. Thus, our determination of the spatial distribution and extent of contamination at a site is governed by uncertainty. We are uncertain about: the processes that generated the
contaminants, the nature, type and amount of contaminants, and the local geology of the contaminated area. When we know the nature of the contaminants, the spatial distribution and concentration of the contaminants, and the local geology (or state of nature) of the contaminated area, we can properly evaluate the *hazard* of the contaminated site. In short, hazard evaluation is carried out to determine what the contaminants are, how concentrated the contaminants are at the site, how the contaminants are distributed, and the state of nature (local geology, the vegetational, wildlife, and human population) around the contaminated site.

Depending on the nature of the contaminants, their spatial distribution and the state of nature around the contaminated site, the contaminants can have different effects on the environment. The potential consequences of the contaminants on nature are what we term the *risk* due to contaminated sites. Since the same hazard can lead to entirely different consequences depending on the state of nature at a particular site we can define the relation between risk and hazard as follows:

\[
Risk = \text{hazard} \times \text{potential worth of loss}
\]

The loss can be the death of animals, the pollution of water bodies, devaluation of land, destruction of vegetation or other non monetary environmental effects.

The decision to clean up a contaminated site can be based upon either the hazard or the risk of the contaminated site. Let us say a regulating agency forbids the average concentration of chemical C exceeding x ppm at any site. If any company exceeds x ppm at a given site, then it has to automatically clean up the site. In this case, the decision to clean up the site depends on the level of x set by the regulating body. The regulating body might have arrived at the cutoff level x ppm by assessing the potential consequences of the contaminant C on the environment when that level is exceeded. However, the company might be at a place where those consequences have a very small or zero probability of occurring. In such a case, the decision to clean up the site is based purely on a hazard level set by the regulating agency. The question we ask is, will it be possible for the company to assess the effects the contaminated site would have on the environment, and
base their decision to clean on the outcome of that study? When a company has unlimited financial resources (which is never the case in reality) to clean up contaminated sites, then the ultimate goal becomes cleaning the site to pose zero risk to the environment. However, because of limited funds, we seek ways to clean up a site in the most economical fashion.

How can we make decisions to clean a site based on risk assessments? That is the main question this research project intends to answer. Our goal is to develop a framework by which decisions can be made about the best method to clean up a site. Figure 1.1a is a simplified influence diagram of the chance events that affect our decision to clean up a site. From this diagram we see that the process of assessing the risk posed by contaminants at a site begins with hazard assessment.

Hazard assessment involves the determination of physical and chemical characteristics of a site's contaminants and the media (soil, groundwater, surface water and air) possibly affected by these contaminants. With this information the potential effects on the environment can be estimated. To assess these effects or risks, we need to know the size and proximity of people, animals and vegetation who are potential victims of these contaminants. From the influence diagram we see that the decision to clean the site depends on the outcome of the risk assessment. At this point, it is important to make sure that we can distinguish between hazard and risk. Even the Webster's Dictionary uses cyclic references of these two terms. Hazard is defined as follows: "1) a foreseeable but unavoidable danger, 2) something causing this, 3) to venture or risk." Risk is defined as "1) the exposure to the chance of injury or loss, 2) danger, hazard, jeopardy, peril." In order to develop a conceptual approach in making decisions, we need to clearly distinguish these terms. This chapter provides a conceptual vocabulary including definitions of hazard and risk and several technical models for discussing and representing our knowledge about a contaminated site.
Figure 1: An influence diagram of the decision to clean a contaminated site.
1.2 Definitions and Concepts

**RCRA**

Resource Conservation and Recovery Act of 1976. This act is an amendment to the Solid Waste Disposal act of 1965. RCRA was amended in 1980 and most recently on November 8 1984 by HWSA.

**Solid Waste**

As defined in RCRA (Resource Conservation and Recovery Act) any garbage, refuse, sludge from a waste treatment plant, water supply treatment plant, or air pollution control facility and other discarded material including solid, liquid, semisolid, or contained gaseous material resulting from industrial commercial mining, from agricultural operations, and from community activities but does not include solid or dissolved materials in domestic sewage, or solid or dissolved material in irrigation return on or industrial discharges which are point sources subject to permits under a water act, or special nuclear or by-product material as defined by the Atomic Energy act of 1954.

**Hazardous Waste**

As defined in RCRA, a solid waste, or combination of solid wastes, which because of the quantity, concentration, or physical, chemical, or infectious characteristics may:

A. Cause, or significantly contribute to an increase in mortality or an increase in serious irreversible, or incapacitating reversible, illness.
B. Pose a substantial present or potential hazard to human health or the environment when improperly treated, stored, transported, disposed of, or otherwise managed.

As defined in the regulations, a solid waste is hazardous if it meets one of four conditions:

1. Exhibits a characteristic of a hazardous waste (40 CFR Sections 261.20 through 262.24).
2. Has been listed as hazardous (40 CFR Section 261.31 through 261.33).
3. Is a mixture containing a listed hazardous waste and a nonhazardous solid waste (unless the mixture is specifically excluded or no longer exhibits any of the characteristics of hazardous waste).
4. Is not excluded from regulation as a hazardous waste.
Site

The land or water area where any facility or activity is physically located or conducted, including adjacent land used in connection with the facility or activity.

Contaminated Land/Site

Land that contains solid, semi-solid, liquid and gaseous wastes. We can also define a contaminated land/site as a land/site that contains substances that, when present in sufficient quantities or concentrations, are likely to cause harm, directly or indirectly, to man, to the environment, or on occasion to other targets. This definition embraces both old industrial sites that have become contaminated as a result of their former usage, and hazardous waste 'problem' sites or 'uncontrolled' hazardous waste sites (US Environmental Protection Agency).

Hazard

Hazard assessment is the process of determining the physical and chemical characteristics of the contaminants at a site and the possible media such as soil, surface water, groundwater and air that can be contaminated. For contaminated sites we define hazard as the average concentration of hazardous wastes at the site. Thus, in assessing hazard at a contaminated site, we will have to identify the site, determine the spatial distribution of the contaminants and then calculate the average concentration of hazardous waste. By measuring the average concentration of hazardous wastes, we determine the "severity" of the hazard at the particular time the contaminant distribution was determined. This raises the question then about sites where the concentration of the contaminants changes over a given time period. This could be a site located near a flowing river, or a site located on a hillside, or on a land with a significant slope such that the distribution of the wastes is altered by runoff events. For such sites we could define the hazard as the average concentration of contaminants over a given time period. If contaminants are dumped at regular intervals on such a site, then the time when the contaminant concentrations are measured will be crucial in determining the hazard level. Stationary
sites or sites where contaminant distribution remains fairly constant over a given time period are usually located on fairly flat lands, or in valleys as shown in Figure 1.2.

Figure 1.2 - Schematic section through a valley containing stationary deposits

Risk

Contaminants at a polluted site can result in a number of consequences including contamination of groundwater, threat to humanity, morbidity and mortality, and damage to activities and wealth both natural and man-made. Consequently the risk posed to the environment by a contaminated site depends on how widely distributed and toxic the pollutants are. Thus, the level of risk depends on the level of the hazard and the potential worth of loss. Hence we'll employ Einstein's [1988] definition of terrain risk in defining the risk of contaminated sites.

\[ \text{Risk} = \text{Hazard} \times \text{Potential Worth of Loss} \]

In estimating risk, we measure the threat potential of a hazard by answering the following questions: how great are the effects of the contaminated site on people and on the environment? We can measure how great the consequences are by placing a monetary value on the potential loss or by representing the potential loss with a utility value. Placing a monetary value on the loss of life remains an ethical question which is often resolved with the aid of insurance policies and the net worth of an individual person.

By assessing the risk due to a contaminated site, we are concerned in estimating the adverse effects of the pollutants on human lives and on the environment. It is sometimes difficult to find out, however, if the contaminants simply contribute to an
increase in a particular illness, or whether the contaminants can be identified as the cause or origin of a particular illness. For example, scientists have identified cigarette smoking as contributing significantly to lung cancer, chronic obstructive pulmonary disease, and heart disease. Notice that we do not say that cigarette smoking causes these health problems because we have not identified the causes (the etiology) of any of them, at least in the sense that scientists have identified poliomyelitis virus as the cause of polio. How then has cigarette smoking been identified as a factor in the increase of the above mentioned illnesses, if it cannot be identified as the cause? Medical researchers found the occurrence of these illnesses higher among people who smoked. In other words, a cigarette smoker has a higher likelihood of contracting any of the above mentioned illnesses than a non smoker.

The basic problem of assessing risk due to contaminated sites is, how to determine exactly what people die of, and how that death is related to pollutants at a contaminated site? Referring to the cigarette smoking example, the cause of death listed on death certificates, which is the source of much epidemiological information, does not specify cigarette smoking as the cause of death. In order to relate death from lung cancer to smoking, one must show that significantly more lung cancer deaths occur in smokers than in nonsmokers. Likewise for contaminated sites, we must show that significantly more people die from pollutants at a contaminated site than in locations with the same population density but without the contaminated site. Such a showing is called a standard mortality ratio (SMR) and is defined as:

\[
SMR = \frac{\text{observed deaths}}{\text{expected deaths}}
\]

The number of "expected deaths" in the above equation is the number of deaths from the particular pollutant. Determination of SMR tells us something about the epidemiology of the pollutants, but not about its etiology.

There are three characteristics of epidemiological reasoning about contaminated sites that are worth noting:
Not everyone exposed to a contaminated site will suffer from the adverse effects of the pollutants.

Some people in the population will die of causes which are not a result of the pollutants but might be difficult to distinguish from deaths caused by the pollutants.

Therefore one cannot unequivocally relate any given individual cause of death to the pollutants.

Risk also can be expressed in other ways. One commonly used statistic in risk analysis is the number of deaths from a given cause per 100,000 population. Another way representation of risk statistics is the number of deaths from a given cause per 1000 deaths. Thus, risk from contaminated sites can be expressed in the following ways:

1. Risk of death from being exposed to a given environmental pollutant (SMR)

   
   \[
   \text{Risk} = \frac{\text{number of deaths from a specific cause in a given population exposed to an environmental pollutant}}{\text{number of deaths from the same cause in a similar sized population not exposed to that pollutant}}
   \]

2. Risk of death from a given cause

   
   \[
   \text{Risk} = \frac{\text{number of deaths associated with the cause in a given time}}{\text{total population, all of whom will die due to some cause}}
   \]

3. Risk of death from a given cause

   
   \[
   \text{Risk} = \frac{\text{number of deaths associated with the cause}}{\text{total number of deaths}}
   \]

One of the problems in assessing the risks due to contaminated sites is that it is difficult to isolate the pollutants at the contaminated site as the main cause of death. The Environmental Protection Agency (EPA) has adopted the concept of unit risk in discussing potential risk. Unit risk is the risk to an individual from exposure to a concentration of 1 g/m³ of an airborne pollutant, or 10⁹ g/L of waterborne pollutant. This concept of unit risk basically asks the question: what are the chances of dying from a particular pollutant-related cause as compared to all other ways dying? Unit lifetime risk is the risk to an individual from exposure to the above concentration for 70 years, a lifetime, while unit occupational risk implies exposure for 40 hours per week for 50 years, a working lifetime. EPA defines unit lifetime risk as follows:

\[
\text{Risk} = \frac{(\text{concentration})(\text{unit risk})(\text{exposure time})}{70 \text{ years}}
\]
From the above equation, it is clear that if either the concentration or the time of exposure is zero the risk would also be zero. But any finite level of a pollutant and any exposure time would lead to a risk as long as the unit risk is greater than zero.

1.3 What should be regulated, hazard or risk?

From the former discussion it is clear that there is a distinction between hazards due to contaminated sites and the risks contaminated sites poses to people and the environment. Let's say we have two contaminated sites of equal pollutant concentration and spatial distribution, but one site is located in a desert (which is uninhabited and has almost insignificant plant growth) and the other is located in a city (with a dense human and vegetational population). The two contaminated sites would have the same hazard but very different risk levels. This raises an important question of what then should be regulated? Should the regulating agency regulate hazard, which implies defining a threshold level for each pollutant and regulating each contaminated site to have pollutant concentrations below the threshold level? The acceptable pollutant concentration or the threshold level will be determined from laboratory experiments conducted on the effects of the pollutant, or may be derived from studies conducted at sites contaminated with this particular pollutant. Or should the regulating agency concern itself solely with the risk posed by contaminated sites? We are not going to suggest the best method to regulate contaminated sites now, but after discussing the details of modeling hazards and risks due to contaminated sites, we hope to present the decision maker with tools which would make his/her decision easier.

The decision to regulate hazard or risk would also be helpful in determining the best method to remediate or clean the site. In addition to total excavation of contaminants, other methods such as:

- on site processing of contaminated soil
- in-situ treatment of contaminants
- macroencapsulation or containment of contaminated sites by using: covering systems, vertical barrier systems and/or underground (horizontal) lining systems.
- hydraulic measures, including groundwater treatment either in situ or after extraction.

can be used in cleaning up, mitigating or containing the pollutants at a contaminated site. These remedial measures render contaminants harmless, prevent release of contaminants, or decrease release of contaminants.
CHAPTER 2 - HAZARD ASSESSMENT FOR CONTAMINATED SITES

2.1 Obtaining Data for Site Hazard Assessment

Knowledge about the toxicity, extent and distribution of a contaminant are primary factors which have to be considered before selecting a remedial method for a contaminated site. Attempts at defining the boundaries of a contaminated site begin with a site investigation. Contaminated site investigation is a routine exercise designed to evaluate the site hazard - the average concentration of contaminants at a site, and the areal extent of contamination. Based on the areal extent of the contaminated site, and the possible source of contamination, a section of the contaminated site is selected for sampling. The site investigation also helps to provide a geologic interpretation of which samples are relevant for estimating the concentration of contaminants at each particular location (Isaaks et al., 1989)

Samples collected at the site are tested in laboratories to determine their material characteristics such as permeability and the concentration of contaminants in each sample. The laboratory analysis provides information about the types of chemical pollutants present at the site. Chemical pollutants which have the highest concentration and distribution are reported as target compounds. In addition to the target compounds, laboratories report trace values of other compounds in the samples (USEPA 1989f). Although these compounds are indicated by peaks on the chromatogram, they are not positively identified and their reported concentrations are highly uncertain. The concentrations of target compounds should be assessed before attempts are made in measuring concentrations of trace elements. Quantitative estimates of the effects of the target compounds are initially made. If these estimates indicate a significant effect that requires further action, incorporating the trace elements into the quantitative analysis would not add any useful information to the risk assessment and can hence be ignored.

In order to clearly distinguish between site and non-site related contaminant sources and to identify naturally occurring levels of metals in soils, we need to do background sampling. Background sampling is done by taking samples from
uncontaminated areas within the site and analyzing these samples in the laboratory to
determine the naturally occurring levels of metals in the soil. When the site has potential
pathways of contamination such as groundwater, surface and subsurface water then we
need to do background sampling for these pathways to determine the naturally occurring
levels of target chemicals in them. At some sites the background concentrations for some
organic compounds may exceed risk-based concentrations. For example, at agricultural
sites where fertilizers and pesticides are constantly applied, the levels of chemicals found
in the fertilizers and pesticides may exceed the risk-based concentrations. In such cases,
we do not have to eliminate these chemicals from the hazard assessment, but rather
evaluate them in light of knowledge about their use at the site.

After obtaining the data values from the field we develop a visual representation of
the sampled data. Data postings (Scatter plots, with data values displayed) give a good
overview of how the contaminants are distributed and can help identify erroneous samples
or irregularly high or low sample values. Figure 2.1 is an example of a data posting.
This figure displays the nitrate concentrations of samples taken near Ontario, Oregon
[Istok et. al, 1993].
Some features noticed on data postings become clearer when contoured. Overall trends in the sample data can be revealed by a contour map. Automatic contouring of irregularly gridded data usually requires the data values to be interpolated to a regular grid. The contour map for the nitrate data displayed in Figure 2.1 is shown in Figure 2.2.
Interpolated values are usually less variable than the original data values and make the contour surface appear smoother. Although this is an aesthetic asset, a smoother surface underestimates the variability in contaminant distribution and may be misleading. Sample data can be visually represented by other means such as indicator maps or symbol maps.
For symbol maps, the contaminant concentrations are grouped into classes. An example of a symbol map is shown in Figure 2.3. The chemical concentrations are grouped into classes and each class is represented by a different symbol on the map (see legend in Figure 2.3). This form of representation is especially useful when there is a known threshold value for the contaminant distribution. Such representations can be useful in making decisions about selectively or completely cleaning up a site.

These graphical representations of the data set enable us to gain an insight into the spatial distribution of the contaminant concentrations. The samples taken at a site are usually not exhaustive, hence special methods which are discussed below are used to determine the contaminant concentration at unsampled locations.
2.2 Point and Global Estimates of Contaminant Concentrations

For risk assessment purposes, we need to determine the average and maximum concentrations of target chemicals at the site. Since for large sites, it is not economically feasible to obtain samples at closely spaced intervals over the site, the usual practice is to obtain samples from a few points and use interpolatory and extrapolatory methods to determine the concentration of contaminants over the whole site. Estimates of the mean concentration of contaminants at the site are obtained by global estimation methods. The most widely used global estimation methods are polygonal estimation and cell declustering methods. Methods used to estimate contaminant concentrations at unsampled points at the site are known as point estimation methods. The most commonly used point estimation methods are: polygonal estimates, triangulation, local sample mean and inverse distance methods.

2.2.1 Global Estimation Methods

Both the polygonal and cell declustering methods use a weighted linear combination of all available data to estimate the global mean value of contaminant concentrations at the site. The equation used to calculate the weighted linear combinations has the form:

\[
\text{estimate} = \hat{\theta} = \sum_{i=1}^{n} w_i \times v_i
\]

(1)

where \(v_1, \ldots, v_n\) are the \(n\) available data values and \(w_i\) is a weight assigned to the value \(v_i\). These weights are standardized so that they sum to one, although this is not required of all estimation methods.
Polygonal Declustering

The polygonal method assigns a polygon of influence to each sample. The areas of these polygons are then used as declustering weights. To illustrate the polygon declustering method we'll use the arbitrary data samples shown in Figure 2.4. The shaded area shows the polygon of influence for the 328 ppm sample located near the center of this area. Any point within the shaded region is closer to the 328 ppm sample than to any other.

![Figure 2.4 - An example showing the polygon of influence of a sample](image)

Figure 2.4 shows how the boundaries of the polygon of influence are uniquely defined. The perpendicular bisectors between a sample and its neighbors form the boundaries of the polygon of influence. In polygonal estimation, the edges of the global area require special treatment. A sample located near the edge of the area of interest may not be completely surrounded by other samples and the perpendicular bisectors with its neighbors may not form a closed polygon. Figure 2.6a shows an example where the perpendicular bisectors between the 85 ppm sample and its three neighboring samples do not form a closed region. One way to solve this problem is to choose a natural limit such
as the lease boundary or a geologic contact, to serve as a boundary for the entire area; this can then be used to close the border polygons. In Figure 2.6b the rectangular boundaries of the map area are used as the natural limit of the area of contamination. An alternative in situations where a natural boundary is not easy to define is to limit the distance from a sample to any edge of its polygon of influence. This has the effect of closing the polygon with the arc of a circle. In Figure 2.6c we see how the polygon of influence is closed if it is not allowed to extend more than 10m from the 85 ppm sample.

By using the areas of the polygons of influence as weights in our linear combination, we accomplish the declustering we require. Clustered samples will tend to get small weights corresponding to their small polygons of influence. On the other hand, samples with large polygons of influence can be thought of as being representative of a larger area and are therefore entitled to a larger weight.
Figure 2.5 - Construction of a polygon of influence using the method of perpendicular bisectors. Figures (a) to (f) show the steps in constructing a region within which the central sample is closer than any other sample.
Figure 2.6 - Defining the polygon on the border of the global area. (a) shows a polygon that cannot be closed by the method of perpendicular bisectors between data points. Alternatively the polygon can be closed by a natural limit such as the lease boundary in (b) or by limiting the distance from a sample to the edge of a polygon as shown in (c).

Cell Declustering

In the cell declustering approach, the entire area is divided into rectangular regions called cells. Each sample receives a weight inversely proportional to the number of samples that fall within the same cell. Clustered samples will generally receive lower weights with this method because the cells in which they are located also contain several other samples.

Figure 2.7 shows a grid of such cells superimposed on a number of clustered samples; the dashed lines show the boundaries of the 20 x 20 m² cells. Each of the two northernmost cells contains only one sample, so both of these samples receive a weight of 1. The southwestern cell contains two samples, both of which receive a weight of 1/2. The southeastern cell contains eight samples, each of which receives a weight of 1/8. Since all the samples within a particular cell receive equal weights and each cell receives a weight of 1, the cell declustering method can be viewed as a two step procedure. First, we use our samples to calculate the mean value within moving windows, then we take these moving window means and use them to calculate the mean of the global area.
Figure 2.7 - An example of cell declustering

The mean concentration of contaminants we get from this cell declustering method will depend on the size of the cells we choose to use. If the cells are very small, then each sample will fall into a cell of its own and all samples will therefore receive equal weights of 1. If the cells are as large as the entire global area, all samples will fall into the same cell and will again receive equal weights. Somewhere between these two extremes we must find an appropriate medium.

If the samples are taken on a pseudo regular grid, then the spacing of this grid usually provides a good cell size. If the sampling pattern does not suggest a natural cell size, a common practice is to try several cell sizes and to pick the one that gives the lowest estimate of the global areas with high values [Isaak, 1989].

2.2.2 Point Estimation Methods

In assessing the hazard posed by a contaminated site, we not only need to know the mean concentration of contaminants at the site, we also need to know the maximum concentration of contaminants and the areas in which they occur. For such location estimation problems we still use weighted linear combinations, but we now need to account not only for possible clustering but also for distance to the nearby samples. As
mentioned earlier, the most commonly used point estimation methods are the polygonal method, triangulation method, local sample mean method and inverse distance methods.

**Polygons**

The polygonal method of declustering that we discussed under global estimation can easily be applied to point estimation. We simply choose as an estimate the sample value that is closest to the point we are trying to estimate. To illustrate how the polygon method is used to estimate contaminant concentration at a specific point, we use the data set shown in Table 1. Let us say that we want to estimate the contaminant concentration within the vicinity of 65E, 137N. Table 1 shows the distances from the 65E, 137N to each of the sample locations shown in Figure 2.8.

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>X</th>
<th>Y</th>
<th>V</th>
<th>Distance from 65E, 137N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61</td>
<td>139</td>
<td>477</td>
<td>4.5</td>
</tr>
<tr>
<td>2</td>
<td>63</td>
<td>140</td>
<td>696</td>
<td>3.6</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>129</td>
<td>227</td>
<td>8.1</td>
</tr>
<tr>
<td>4</td>
<td>68</td>
<td>128</td>
<td>646</td>
<td>9.5</td>
</tr>
<tr>
<td>5</td>
<td>71</td>
<td>140</td>
<td>606</td>
<td>6.7</td>
</tr>
<tr>
<td>6</td>
<td>73</td>
<td>141</td>
<td>791</td>
<td>8.9</td>
</tr>
<tr>
<td>7</td>
<td>75</td>
<td>128</td>
<td>783</td>
<td>13.5</td>
</tr>
</tbody>
</table>
Figure 2.8 - The data configuration shown in this figure is used to illustrate several point estimation methods in the following sections. The goal is to estimate the value of V at the point 65E, 137N, located by the arrow, from the surrounding seven data values.

The sample at 63E, 140N is closest, so our polygonal estimate of the V value at 65E, 137N is 696 ppm. This polygonal estimator can be viewed as a weighted linear combination that gives all of the weight to the closest sample value.

Polygonal estimates of the V value at other points near 65E, 137N will also be 696 ppm. If the point we are estimating falls within the same polygon of influence, the polygonal estimate remains unchanged. As soon as we encounter a point in a different polygon of influence, the estimate jumps to a different value. Figure 2.9 shows how the polygonal estimates near 65E, 137N form a discontinuous surface of plateaus.
Figure 2.9 - A perspective view showing the discontinuities inherent in polygonal estimates

**Triangulation**

The discontinuities in the estimated values from the polygonal method are usually undesirable because in most cases, the contaminants are distributed in a continuous fashion. The triangulation method overcomes this problem of the polygonal method, removing possible discontinuities between adjacent points by fitting a plane through three samples that surround the point being estimated. The equation of the plane can be generally expressed as

\[ z = ax + by + c \]  \hspace{1cm} (2)

In our example, where we are trying to estimate V values using coordinate information, \( z \) is the V value, \( x \) is the easting, and \( y \) is the northing. Given the coordinates and the V value of three nearby samples, we can calculate the coefficients \( a, b \) and \( c \) by solving a system of equations:

\[ z_1 = ax_1 + by_1 + c \]
\[ z_2 = ax_2 + by_2 + c \]
\[ z_3 = ax_3 + by_3 + c \]  \hspace{1cm} (3)
which gives us the following equation as our triangulation estimator:

\[ \hat{\theta} = -11.25x + 41.614y - 4421.159 \]  

(4)

This is the equation of the plane that passes through the three nearby samples we have chosen. Using this equation we can now estimate the value at any location simply by substituting the appropriate easting and northing. Substituting the coordinates \( x = 65 \) and \( y = 137 \) into our equation gives us an estimate of 548.7 ppm at the location 65E, 137N.

**Inverse Distance Methods**

With the inverse distance method, samples which are closest to the point we are trying to estimate are given more weight and less weight are given to samples which are farthest away. One way of doing this is to make the weight for each sample inversely proportional to its distance from the point being estimated:

\[ \hat{\theta} = \frac{\sum_{i=1}^{n} \frac{v_i}{d_i}}{\sum_{i=1}^{n} \frac{1}{d_i}} \]  

(5)

\( d_1, \ldots, d_n \) are the distances from each of the \( n \) sample locations to the point being estimated and \( v_1, \ldots, v_n \) are the sample values. The inverse distance estimator given in the above equation can easily be adapted to include a broad range of estimates. Rather than using weights that are inversely proportional to the distance, we can make the weights inversely proportional to any power of the distance:

\[ \hat{\theta} = \frac{\sum_{i=1}^{n} \frac{v_i}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}} \]  

(6)
Different choices of the exponent $p$ will result in different estimates. As $p$ approaches 0 and the weights become more similar, the inverse distance estimate approaches the simple average of nearby sample values. As $p$ approaches $\infty$, the inverse distance estimate approaches the polygonal estimate, giving all of the weight to the closest sample. Traditionally the most common choice for the inverse distance exponent is 2. Inverse distance squared estimates are not necessarily better than estimates from inverse distance methods that use some exponent other than 2. The choice of $p$ is arbitrary, and the traditional popularity of 2 is due, in part, to the fact that this choice simplifies the calculations.

**Ordinary Kriging**

Ordinary kriging is the most widely used point and global estimation tool for spatial data. Ordinary kriging is often associated with the acronym B.L.U.E. for "best linear unbiased estimator." Ordinary kriging is "linear" because its estimates are weighted linear combinations of the available sample data; it is "unbiased" since it tries to have the mean residual error equal to 0; it is "best" because it aims at minimizing the variance of errors. Strictly speaking, the kriging variances that are minimized depend on the sample locations and not on the sample values. The theoretical principles of ordinary kriging are described in great detail in several references including Journel and Huijbregts, 1978, Isaaks and Srivastava, 1989 and David, 1977. The details of the geostatistical analysis carried out on sampled data can be summarized as follows:

The primary goal of ordinary kriging is to estimate contaminant concentration at every point where we do not have a sample, by using a weighted linear combination of the available samples:

$$\hat{u} = \sum_{j=1}^{n} w_j u_j$$

(7)

The set of weights is allowed to change as we estimate unknown values at different
locations. The set of weights that minimizes the error variance satisfies the following equations:

\[ \sum_{i=1}^{n} w_i y_i - \mu = \gamma_{io} \quad \forall \; i = 1, \ldots, n \quad (8) \]

\[ \sum_{i=1}^{n} w_i = 1 \quad (9) \]

where \( y_i \) expresses the spatial variation between the sample data points and \( \gamma_{io} \) expresses the spatial variation between the sample data points and the point at which we do not have a sample. The variation between the sample data points is assumed to be dependent only on the distance of separation between them. The mathematical relation which computes \( y_i \) or \( \gamma_{io} \) is known by geostatisticians as the variogram and is calculated as follows:

\[ \gamma(h) = \frac{1}{2N(h)} \sum_{(i,j) \neq (h)} (y_i - y_j)^2 \quad (10) \]

For a particular distance of separation \( h \) between the sample data points, the number of samples with that distance of separation \( N(h) \) is calculated. For any pair of samples separated by the distance \( h \), we compute the square of the difference between contaminant concentrations. The variogram, \( \gamma(h) \), is half the average squared difference between the paired data values. \( \mu \) is the mean value of sample concentrations. An example of a variogram is shown in Figure 2.10.
Figure 2.10 - An example of an exponential variogram model

The variogram shown in Figure 2.10 exhibits the main characteristics of most variograms:

- \( C_0 \), commonly called the *nugget effect*, which provides a discontinuity at the origin.
- \( a \), commonly called the *range*, which provides a distance beyond which the variogram or covariance value remains essentially constant.
- \( C_0 + C_1 \), commonly called the sill, which is a variogram value for very large distances, \( \gamma(\infty) \).

Details for estimating kriging weights are outlined in *Applied Geostatistics* by Isaaks E. H. and Srivastava R. M., 1989 (pages 280 to 290).

**Pitfalls of Geostatistics**

Although geostatistical methods are widely accepted and applied in estimating contaminant distributions, they have certain setbacks which have to be noted. Most of these setbacks are outlined in J. W. Merks' recent paper, "Geostatistics or Voodoo Statistics." The major pitfall of geostatistics as outlined in this paper is quoted as follows:

*Kriging attempts to make a few holes go a long way. However, it may create ore grade values in chaotic domains where grades and variances are unknown and discontinuities lurk.........Kriged variances inflate expectations for the continuity*
of mineralization between measured data points, and cannot possibly deliver the reliable and realistic precision estimates that ore reserves demand. However, a technique that replaces a bulldozer's crooked blade (kriged variances for global sets) with a surgeon's scalpel (variances for ordered data sets after variances for the measurement process are deducted, and sums of variances for contents of all elementary units in the set that constitutes an ore deposit, or any subset that constitutes a part thereof) has already been developed and published.

Kriging estimates are based on the assumption that the sample concentrations are correlated in space. From samples taken in the field a variogram is derived that measures the spatial correlation between the samples. When this assumption is blindly applied to regions where discontinuities exist, wrong estimates of unsampled areas would be obtained. Most mineral deposits form over a long period of time over which geologic events such as faulting, jointing or bedding may cause discontinuities in the mineral deposits. Thus, when the assumptions of geostatistics are applied to estimate the mineral concentrations of such deposits, erroneous results might be obtained.

In any earth science study, the size of each sample is an important consideration. If we went to a contaminated site and took samples with a tablespoon, there could be a lot of variability between sample values. One sample would contain almost pure contaminants, another might contain nothing. If we sampled the same site with large truck loads, there would likely be much less variability. The mixing of high and low values that is to be expected in large samples would give us less erratic values. In estimating the mean contaminant concentration of a site from samples taken on a large grid, we would have to average values over larger areas which generally has the effect of reducing the variance of the data. Though the size of the data has an effect on the spread and the symmetry of the distribution, it has no effect on the mean. Another effect of sample size can be explained by the Weibull's effect: the larger the sample, the more likely it is to have larger extreme values. Thus, in order to obtain accurate estimates of the contaminant distribution, the size or support of the samples taken should be approximately equal to the size of the support in the estimates we intend to make. This problem of the discrepancy
between the support of samples and the intended support of estimates is one of the most
difficult problems faced in geostatistical estimation.

Another pitfall of geostatistics is that a single variogram is used in calculating the
weights for estimating pollutant concentrations in unsampled regions. The variogram
model used is chosen from the major direction of anisotropy. Hence, estimates made for
regions which lie outside the major axis of anisotropy would either be over or
underestimated.

2.3 The Bayesian Approach For Modeling Contaminant Extent

A fairly recent approach which is not as widely used as the traditional geostatistical
approach combines probability assessments with geostatistical modeling to describe the
extent of contaminants in an area [Johnson, 1992]. An understanding of contaminant
distribution is obtained from prior sampling campaigns in the contaminated area. This
sample data is referred to as "hard data." Hard data is examined together with geological
and topographical maps, published reports (if available), well logs and results from prior
modeling exercises in the region. Information obtained from maps, reports, well logs and
prior estimation exercises is known as "soft data." An initial theoretical model \( P(X=x|\theta) \)
is selected to model our beliefs regarding an uncertain quantity \( X \), (for example, a binomial
distribution to model the presence or absence of contamination at a location \( X \)), and \( \theta \) is a
parameter about which we are uncertain (for example how are the contaminants
distributed in space?). Being uncertain about \( \theta \), we assess a prior probability distribution,
f(\( \theta \)), for the parameter. At this point, \( P(X=x|\theta) \) and f(\( \theta \)) together embody all of our
beliefs about the possible outcomes of the uncertain quantity \( X \). Thus, the next step in the
process is the acquisition of data and the incorporation of the new information into the
model with the help of Bayes' formula [Cornell and Benjamin, 1970],

\[
P(\theta|X=x) = f(\theta)P(X=x|\theta)
\]

where \( P(\theta|X=x) \) is the posterior probability distribution of \(\theta\).
Figure 2.11 - Using Sample Data to Update Uncertainty about parameter $\theta$ via Bayes' Theorem.

Figure 2.11 is a flow chart that shows how Bayesian updating is carried out to determine the contaminant value at an unsampled location.

Let us consider a physical process that can be described by a binomial distribution with parameters $n$ and $p$. That is the decision maker knows that the process generates many observations with either successes or failures, that in a group of $n$ observations, there will be some number $X$ of successes. Thus, $P(X=x|\theta)$ in this case is equal to the binomial probability $P(X=r|n,p)$ which is given by the expression

$$P(X=r|n,p) = \frac{n!}{r!(n-r)!}p^r(1-p)^{n-r}$$

We can calculate $P(X=r|n,p)$ for any value of $r$ as long as $p$ is known. In this case however we want to incorporate uncertainty about $p$. The exact value of $p$ can only be known when we have obtained an exhaustive set of the outcomes of the process. However, our initial estimate of $p$ is not from an exhaustive set; hence, we need to update.
our knowledge of \( p \) as we obtain more data. We model the uncertainty about \( p \) by representing its distribution with a beta distribution,
\[ f_B(p | r_0, n_0) \]. Based on this prior distribution for \( p \), a good estimate for it would be the mean of this distribution, \( r_0/n_0 \). Thus, when the decision maker observes \( n_1 \) independent trials, and \( r_1 \) of these are successes, then a good estimate for \( p \) would be \( r_1/n_1 \). The main problem is how to incorporate this information with the prior distribution. Combining the data with the prior distribution gives a posterior distribution for \( p \) that is also a beta distribution, \( f_p(p | r^*, n^*) \), where
\[ r^* = r_0 + r_1 \]
and
\[ n^* = n_0 + n_1 \]
That is, the decision maker only has to add the \( r \)'s and \( n \)'s to find the parameters of the posterior distribution for \( p \). This process could continue as more data is observed. The posterior \( f_p(p | r^*, n^*) \) becomes the prior distribution in the next round. Data \( r_2 \) and \( n_2 \) are observed. Thus, the parameters for the posterior distribution after observing \( r_2 \) and \( n_2 \) are \( r^{**} = r^* + r_2 \) and \( n^{**} = n^* + n_2 = n_0 + n_1 + n_2 \).

2.4 Delineating Contaminant Boundaries Using Binary Logistic Regression

Determining the extent of spatial distribution of contaminants at a site is a major step in making estimates of the hazard posed by the site and the cost of the clean up project. At the moment geostatistics offers tools to estimate the spatial distribution of contaminants over an area. Before an accurate mathematical model can be derived using geostatistical methods, the spatial variability of contaminants is represented by an equation known as the variogram. An accurate derivation of the variogram relies on a large data set. Thus, we lose some of the economic benefits of mathematical modeling because a large amount of money is spent obtaining samples to derive a variogram. How can we reliably obtain estimates of contaminant extent from very small datasets? We attempt to
answer this question by presenting a couple of methods that can be used to determine the initial extent of contamination. We show how a binary logistic regression model can be used to arrive at initial estimates of contaminant extent and to update our knowledge of contaminant extent when more samples are obtained from a site.

**Model Description**

The contaminant concentration at an unsampled location is analyzed as dependent on observed history in the following way: contaminant samples are labeled 1, 2, ..., in the order drilled. Associated with each contaminant sample is a description of the state of that sample: its spatial location and whether the contaminant concentration in the sample exceeds or is below a certain threshold limit.

The contaminant states are defined as follows:

Let \( \{x_i, y_i\} \) denote the coordinates of location of a contaminant sample \( i \) and define \( v_i = 1 \) if sample \( i \) exceeds the threshold value or \( v_i = 0 \) otherwise. Then

\[
s_i = \{x_i, y_i, v_i\}
\]

is the state description for sample \( i \), and after taking \( n \) samples, the observed history is

\[
s^{(n)} = (s_1, \ldots, s_{n})
\]

Sampling outcomes \( v_i \) are made depending on their history in a fashion to be shortly described. We assume that sampling locations are not subject to uncertainty but are chosen after a careful study of the geology of the contaminated area and history of deposition of contaminants.

The data-generating process is of the form:

\[
p(x_i, y_i | s^{(i-1)}) = \frac{\exp(h(x_i, y_i) + \beta_o + \beta_1 x_i + \beta_2 y_i)}{1 + \exp(h(x_i, y_i) + \beta_o + \beta_1 x_i + \beta_2 y_i)}
\]

That is, given the history \( s^{(i-1)} \) and a sample to be taken at location \( (x_i, y_i) \), the probability that this sample contains contaminants exceeding the given threshold is \( p(x_i, y_i | s^{(i-1)}) \). This is a particular case of a logit model in which \( h(x_i, y_i) \) represents the effect of the location of
the \(i\)th sample on its probability of exceeding the given threshold. Let us consider a hypothetical scenario in which the probability of finding contaminants at a location \(i\) depends on the spatial location of the sample but is independent of outcomes of earlier samples. Such a case represents a Bernoulli process with varying probabilities and constitutes a particular form of trend surface analysis. This particular model may be taken as a null hypothesis against which we wish to test the alternative that the \((n+1)\)st sample outcome depends on both the locations and outcomes of the samples \(1, 2, \ldots, n\) within a spatial window around the \((n+1)\)st sample.

For this hypothetical Bernoulli case, we could define the joint probability of realizing \((v_1, v_2, \ldots, v_n)\) as

\[
\prod_{i=1}^{n} \theta_i^{v_i}(1-\theta_i)^{1-v_i} \quad (4)
\]

where \(\theta_i = p(x_i, y_i | s^{(i)})\). To solve equation (3) we have to find values of \(\beta_o, \beta_1, \) and \(\beta_2\) that will maximize equation (4). We can obtain this by maximizing the natural logarithm of the equation (4). By taking the natural logarithms of equation (4), we obtain

\[
\sum_{i=1}^{n} v_i(\beta_o + \beta_1 x_i + \beta_2 y_i) - \sum_{i=1}^{n} \ln(1 + \exp(\beta_o + \beta_1 x_i + \beta_2 y_i)) \quad (5)
\]

Since maximizing a log-likelihood function is equivalent to minimizing its negative, we can find values of \(\beta_o, \beta_1\) and \(\beta_2\) that minimize

\[
\sum_{i=1}^{n} v_i \frac{\exp(\beta_o + \beta_1 x_i + \beta_2 y_i)}{1 + \exp(\beta_o + \beta_1 x_i + \beta_2 y_i)} \quad (6)
\]

\[
\sum_{i=1}^{n} v_i x_i \frac{\exp(\beta_o + \beta_1 x_i + \beta_2 y_i)}{1 + \exp(\beta_o + \beta_1 x_i + \beta_2 y_i)} \quad (7)
\]

\[
\sum_{i=1}^{n} v_i y_i \frac{\exp(\beta_o + \beta_1 x_i + \beta_2 y_i)}{1 + \exp(\beta_o + \beta_1 x_i + \beta_2 y_i)} \quad (8)
\]
The appropriate choice of form for $h(x_i, y_i)$ depends very much on the geological setting. To simplify exposition, momentarily let us assume that no trend effect is present [$h(x_i, y_i) = 0$]. The simultaneous solution of equations 6, 7, and 8 gives the values of $\beta_0$, $\beta_1$ and $\beta_2$.

Let us apply this methodology to define the contaminant extent for a data set taken from page 291 of Applied Geostatistics by Isaaks et. al, 1989. Data posting of the locations where samples have been taken and analyzed is shown in Figure 2.8. For this particular example, let us assume that the threshold value for the contaminants is 550 ppm. This means any sample with concentrations equal to or exceeding 550 ppm will be assigned a value of 1 and a value of 0 otherwise. Using only the sample data, we obtain $\beta_0 = -2.190$, $\beta_1 = 18.05$ and $\beta_2 = 7.674$ as the values which would maximize equations 6, 7 and 8. Substituting these values into equation (3),

\[ h(x_i, y_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i \]

![Figure 2.12](image.png)  
Figure 2.12 - Random Contour of Sample Data Showing the 50% Boundary for a Threshold Value of 550 ppm.

42
we can find the $x$ or $y$ values which will give us a 50% probability of finding contaminant concentrations of 550ppm. The 50% probability line of contamination is shown on Figure 2.12.

**Model with Spatial Interaction**

Assuming that there exists spatial interaction between the sample data points, we can formulate the probability of having contamination at a point: $(x_i, y_i)$ as

$$p(x_i, y_i|\mu^{-1}) = \frac{\exp(h(x_i, y_i) + \beta_0 + \beta_1x_i + \beta_2y_i + \beta_3x_iy_i)}{1 + \exp(h(x_i, y_i) + \beta_0 + \beta_1x_i + \beta_2y_i + \beta_3x_iy_i)}$$  \hspace{1cm} (9)$$

Using this model, the computed parameters for the dataset shown in Figure 2.8 are $\beta_0 = -2524$, $\beta_1 = 23.05$, $\beta_2 = 10.05$ and $\beta_3 = -0.035$. The 50% probability boundary produced by these parameters is shown in Figure 2.13. For the spatial interaction model, we observe a slight shift of the 50% probability contamination boundary to the left, which is a result of the negative coefficient of the interactive term $\beta_3$. Thus for this data set, the spatial interaction model yields a larger contaminated area than the model without spatial interaction.
Figure 2.13 - 50% probability boundary obtained from spatial interaction model

2.4 Value of Perfect Information

Whenever a mathematical model is used to describe the contaminant distribution at a site, the problem of determining its accuracy arises. It is clear that in the case of deterministic models, for example the models obtained through kriging of sampled data, one should just compare deterministic predictions with actual observations. The accuracy of probabilistic models of contaminant distributions can be assessed in a similar fashion. However, accuracy of a procedure for the estimation of contaminant hazard is not by itself a sufficient measure of engineering value. What is of greater interest for engineering purposes is "contaminant risk," which has already been defined in the introductory chapter. This raises an important question, how much value can we place on information about contaminant distributions in making decisions on how best to clean up a contaminated site? Is it possible that we can clean up a contaminated site effectively
without going through an expensive sampling program? Does our decision to clean up a site depend on what an expert says about the contaminant distribution?

To try and answer the questions posed above, let us consider a hypothetical site S, that needs to be cleaned. For site S, let us say there are only two decisions that have to be made about the clean up project: to completely clean up the site, which would cost $1 million dollars, or to partially clean the site which would cost $0.4 million dollars. If the site is completely polluted and we decide to completely clean it, there would be no detrimental effects on the environment. However, if we decide to partially clean the site and the site is completely polluted, then the contaminants we left uncleaned can have detrimental effects on human health, wildlife, plants and the environment in general. Let us assume that the cost of curing people and wildlife who get sick from exposure to these contaminants plus the costs of revegetating is approximately $0.8 million dollars. For this site let us consider two outcomes from our sampling program: in the first outcome, our sampling program gives us perfect information about the contaminant distribution (which is rarely the case in real world situations). In this case we consult a clairvoyant who can tell us that there is an 85% chance that the site is completely polluted and a 15% chance that it is partially polluted.

The decision tree below represents the case where we have perfect information about the contaminant distribution.
From the decision tree we have three options:

**Decision 1** Decide not to take any samples and go ahead and completely clean the contaminated site.

**Decision 2** Decide not to take any samples and go ahead and selectively clean the contaminated site. A site is selectively cleaned by selecting and cleaning areas with suspected high chemical concentrations.

**Decision 3** Decide to take samples from the contaminated site which will provide us with perfect information about the distribution of contaminants at the site. Since the information is perfect, if it indicates that the site is completely polluted, we will go ahead and completely clean the site. On the other hand, if the samples indicate that the site is partially contaminated, then we would selectively clean those contaminated areas.

We calculate the expected monetary value (EMV) of decisions 1, 2 and 3 respectively as follows:
\[ EMV_1 = 0.85 \times -1,000,000 + 0.15 \times -1,000,000 = -$1,000,000 \]
\[ EMV_2 = 0.85 \times -1,200,000 + 0.15 \times -400,000 = -$1,080,000 \]
\[ EMV_3 = 0.85 \times -1,000,000 + 0.15 \times -400,000 = -$910,000 \]

The least expected monetary loss from either decision 1 or 2 is -$1,000,000. When we obtain perfect sampling information, the expected monetary loss is -$910,000. Thus the expected value of the sampling information is calculated as,

\[ \text{Expected Value of Perfect Sampling Information} = -910,000 - (-1,000,000) = $90,000 \]

Hence we should not pay more than $90,000 to get sample information for the contaminated site. This is a very simple example to illustrate how we can incorporate uncertainty about sampled information to make decisions about a contaminated site. It is worth noting that the way we are thinking about the value of the sampled information is in a strictly \textit{a priori} sense. The decision-tree representation reinforces this notion because we actually include the decision branch that represents the possibility of obtaining information about contaminant distributions by sampling a site. Different decisions would be made for different clean up costs and for different assessments about the accuracy of the sampling campaign. Also, if different probabilities are assigned to the outcomes of the sampling campaign, the optimal decision would be different.

2.5 Value of Imperfect Information

We rarely have access to perfect information. In fact, our information sources are usually subject to considerable error. Thus, we must extend our analysis to deal with imperfect information. The analysis of imperfect information parallels that of perfect
information. We still consider the value of the imperfect information before obtaining it, and we still call it the expected value of sample information.

In the example presented about the cleanup of site S, suppose that we hire a sampling consultant who analyzes sample data from contaminated sites. Because he can make mistakes, his information is not perfect. In the previous example, we assume that the data analyst's (clairvoyant's) assessment about the probabilities that the site is either partially or completely polluted is perfect, that is they are unconditional. In this case however, what the consultant says is not perfect. For example the track record of the consultant is such that when he says the site is partially polluted, then there is an 80% chance that the site is partially polluted. The probabilities are conditional; for example, \( P(\text{consultant says "partially polluted" | partially polluted}) = 0.80 \). We can construct the following table characterizing the consultant's prediction ability.

Table 2. Conditional Probabilities Characterizing Consultant's Prediction Ability

<table>
<thead>
<tr>
<th>Consultant's Prediction</th>
<th>True State of Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Partially Polluted</td>
</tr>
<tr>
<td>Partially Polluted</td>
<td>0.8</td>
</tr>
<tr>
<td>Completely Polluted</td>
<td>0.2</td>
</tr>
</tbody>
</table>

How should we use the consultant's information? We must use Baye's theorem to calculate the posterior probabilities. In order to do this, we need to assess the prior probability that a site would be completely polluted or partially polluted. Let us say that from our prior experience with such sites, \( P(\text{completely polluted}) = 0.7 \) and \( P(\text{partially polluted}) = 0.3 \). Then we can assess the following posterior probabilities.

- \( P(\text{Site is completely polluted | Consultant says " site is completely polluted")} = P(\text{CP | "CP"}) \)
- \( P(\text{Site is partially polluted | Consultant says "site is partially polluted")} = P(\text{CP | "PP"}) \)
- \( P(\text{Site is partially polluted | Consultant says "site is completely polluted")} = P(\text{PP | "CP"}) \)
- \( P(\text{Site is completely polluted | Consultant says "site is partially polluted")} = P(\text{CP | "PP"}) \)

where

CP is a denotation for a completely polluted site.
PP is a denotation for a partially polluted site,
"CP" denotes the consultant says "the site is completely polluted",
"PP" denotes the consultant says "the site is partially polluted".

\[
P(CP \mid "CP") = \frac{P("CP" \mid CP) P(CP)}{P("CP" \mid CP) P(CP) + P("CP" \mid PP) P(PP)} = \frac{0.85 \times 0.7}{0.85 \times 0.7 + 0.2 \times 0.3} = \frac{0.595}{0.655} = 0.9084
\]

The denominator of this equation represents the probability that the consultant says "the site is completely polluted." Thus \( P(\text{Consultant says "site is completely polluted"}) = 0.655 \).

\[
P(PP \mid "PP") = \frac{P("PP" \mid PP) P(PP)}{P("PP" \mid PP) P(PP) + P("PP" \mid CP) P(CP)} = \frac{0.8 \times 0.3}{0.8 \times 0.3 + 0.15 \times 0.7} = \frac{0.24}{0.345} = 0.6957
\]

The denominator of this equation represents the probability that the consultant says "the site is partially polluted." Thus \( P(\text{Consultant says "site is partially polluted"}) = 0.345 \).

We can calculate the remaining posterior probabilities as follows:

\[
P(\text{Site is partially polluted} \mid \text{Consultant says "site is completely polluted"}) = 1 - P(CP \mid "CP")
\]
\[= 1 - 0.9084 = 0.0916\]

\[
P(\text{Site is completely polluted} \mid \text{Consultant says "site is partially polluted"}) = 1 - P(PP \mid "PP")
\]
\[= 1 - 0.6957 = 0.3043\]

The posterior probabilities are summarized in Table 3.
Table 3 - Posterior Probabilities for State of Contamination at a Site Depending on a Consultant's Prediction

<table>
<thead>
<tr>
<th>Consultant's Prediction</th>
<th>Posterior Probability for:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Partially Polluted</td>
<td>0.6957</td>
<td>0.0916</td>
</tr>
<tr>
<td></td>
<td>Completely Polluted</td>
<td>0.3043</td>
<td>0.9084</td>
</tr>
</tbody>
</table>

The value of imperfect sampling information can be calculated from the decision tree shown in Figure 2.15. The value of the imperfect sampling information is $88,249. Thus, we should not pay more than $88,249 for the sample information.
The Value of Imperfect Sampling Information = $1,000,000 - $953,510 = $46,490

Figure 2.15 - Decision Tree to Estimate the Value of Imperfect Information
2.6 Value of Information and Decision Structuring

Our goal for this chapter has been to examine situations in which information is available and to show how decisions can be made systematically regarding what source of information to select and how much an expert's information might be worth. When the expected value of perfect information in a decision problem is low, then there is little sense in spending a lot of effort in reducing uncertainty by collecting information and doing sophisticated analysis with the data. But if the expected value of the perfect is high, then it may be worthwhile to put considerable effort into collecting sample data and doing geostatistical or similar analysis to understand the distribution of contaminants at the site. Such information can have a relatively high payoff by reducing uncertainty and improving the decision maker's expected monetary value.
CHAPTER 3 - RISK ASSESSMENT

(Human and Environmental Risk From Exposure to Contaminants)

3.1 Introduction

This chapter deals with the measurement of risk posed by contaminated sites. The goal of risk measurement is to determine in either monetary terms or in other quantifiable ways the potentially negative consequences of pollutants on human beings and on the environment in general. For a contaminated site, we are interested in determining the toxicity of the chemicals, how the chemicals are distributed in the environment and the possible negative consequences the chemicals can have on the environment. In chapter 2 we discussed how the distribution and amount of contaminants at the site are quantified. We defined the quantification of the toxicity and distribution of the contaminants as the hazard. Thus for a given contaminant, the risk associated with it is directly proportional to the amount of hazard times the potential worth of loss.

\[ \text{Risk} \propto \text{Hazard} \times \text{Potential Worth of Loss} \]

This definition of risk was posited by Einstein in 1988. For a contaminated site, the presence of toxic chemicals does not necessarily mean that the site will pose a considerable risk to the environment. For a contaminated site to pose a considerable risk to the environment, the following elements have to be present: 1) Exposure Routes, 2) Biological Receptors and/or infrastructure which can be damaged from exposure to the contaminants and 3) the contaminants with a substantial toxicity.

For the contaminant to have any negative effect on the environment, there has to be a medium or route through which the contaminants can come into contact with living or non-living things in the environment. The possible type of media are soil, groundwater, subsurface water, surface water and air. In the risk assessment literature, these media are known as exposure routes. Depending on the nature of the contaminant and the location of the site, the pollutants can contact the environment through all or some of the media. Thus, volatile contaminants may have air as their primary exposure route. In addition to these primary exposure routes, animals and plants can also serve as media through which the pollutants can be distributed in the environment. This is especially true for
contaminated sites located on or near agricultural stations. Contaminants in the soil or water can be absorbed in plants or consumed by animals or other living organisms. The gradual absorption of the pollutants into the tissue of plants or animals leads to bioaccumulation, the process by which organisms retain chemical pollutants in their tissues at levels greater than in the ambient environment. Animals, plants or fish with high concentrations of pollutants in their tissues can be consumed by humans, thereby, serving as media for transporting the contaminants. If the contaminated site serves as the feeding ground for birds, for example ducks, and these are consumed by humans, then the birds serve as a transporting route for the contaminants. Thus, to completely understand how the contaminants are transferred to humans and other living organisms, we need to understand the complete food chain existent at the site.

From the above discussion we realize that in order to develop a complete scenario for assessing exposure to contaminants, we have to consider the land use, exposure pathway, the amount of contaminants taken in by an organism on a daily basis, how often the organism are exposed to the contaminants (exposure frequency), how long they are exposed to the contaminants (exposure duration), and the body weight or size of the organism.

Before discussing the details of risk measurement, we need to identify two approaches of risk measurement: the empirical approach of risk measurement and the stochastic or probabilistic approach of risk measurement. The empirical approach of risk measurement involves a physical measurement of the toxicity, identification of all the exposure routes at the contaminated site, and a measurement of the daily intake of chemicals by all identifiable organisms at the site. These measurements are then used to calculate risk factors. This empirical approach has been adopted by the US EPA in quantifying risks due to contaminated sites. This methodology somehow standardizes risk measurement and aims at reducing the subjective aspects of risk measurement. As we discuss the details of the empirical approach, we will discuss some of its flaws.

The stochastic or probabilistic approach of risk measurement has as its underlying philosophy that risk measurement is inherently uncertain; hence, given a contaminated site,
we can estimate the likelihood of a human or any living organism experiencing negative effects from the contaminants, without necessarily making detailed measurements of the exposure routes. This approach views the extrapolation of toxicity studies on animals to human beings as not exact and misleading. It also realizes that even though exposure routes may exist at a site, they may not necessarily be used to transport the pollutants in the deterministic manner posited by the empirical models.

3.2 Quantification of Risk

3.2.1 Empirical Approach

Human health risks can be categorized as cancerous and non-cancerous effects. The statistical approach of estimating the risk of obtaining cancer and non-cancerous effects from a contaminated site are calculated as follows:

a) The carcinogenic (likelihood of contracting cancer) risk associated with a contaminated site, \( R \), is the summation of the individual carcinogenic risks associated with each exposure pathway (soil, air, groundwater, surface water etc.). Thus

\[
R = \sum_{i=1}^{n} R_i
\]

where \( R_i \) is the carcinogenic risk associated with exposure to the contaminated soil,

\( R_2 \) is the carcinogenic risk associated with exposure to contaminants in the air

......

\( R_n \) is the carcinogenic risk associated with the exposure to contaminants in the nth exposure pathway.

The risk associated with each pathway, \( R_p \), is calculated as follows:

\[
R_i = CDI_i \times SF_i
\]

where \( CDI_i \) is the chronic daily intake and is defined as the exposure expressed as mass of a contaminant contacted per unit body weight per unit time, averaged over
a long period of time. The EPA uses as a guideline for Superfund sites seven years to seventy years as the period of intake.

$SF_i$ is the cancer slope factor and represents the maximum daily intake level (mg/kg/day)$^{-1}$ for a particular chemical contaminant that is likely to be without an appreciable risk of developing cancerous cells within the human body.

b) Non-carcinogenic risks associated with a contaminated site are called hazard index to differentiate it from carcinogenic risks. The hazard index for a contaminated site is computed as the summation of the individual hazard indices associated with each exposure pathway.

$$HI = \sum_{i=1}^{n} HI_i$$

where $HI_i$ is the hazard index associated with groundwater,

$HI_i$ is the hazard index associated with air,

......

$HI_n$ is the hazard index associated with the nth exposure pathway.

The hazard index for each individual exposure pathway is computed as the ratio of the amount of exposure through that pathway to the reference dose. Thus,

$$HI_i = \frac{I}{RfD}$$

where $I$ is the amount of chemical taken in through the exposure pathway. This amount of intake is usually expressed as the mass of contaminant contacted per unit body weight per unit time (mg/kg/day).

$RfD$ is the reference dose and is based on the EPA's identification of the threshold effects level with an added margin of safety. The $RfD$ represents an estimate of daily intake (mg/kg/day) for a particular contaminant that is likely to be without any appreciable risk of deleterious health effects when exposure occurs over a given period (usually 70 years).
Both carcinogenic risk, $R$, and non-carcinogenic risk, $HI$, are unitless probabilities of an individual either developing cancer or a deleterious health effect as a result of exposure to the contaminant. Risks, $R$, and hazard indices, $HI$, for each pathway are summed to estimate a single risk or hazard index for an individual exposed to the contaminants via more than one pathway at site. By summing the individual risks, $R_i$, or hazard indices, $HI_i$, from each pathway, we are assuming that there is no interaction between the different exposure pathways, or that the interaction between the different exposure routes is assumed negligible.

For a site contaminated with both carcinogenic and non-carcinogenic chemicals, both carcinogenic risks and non-carcinogenic risks will have to be accessed for each exposure pathway. The risks and hazard index computed for each site are compared with target risk levels established by the USEPA or by state agencies. The USEPA’s guideline states that the total incremental carcinogenic risk, $R$, for an individual should not exceed a range of $1 \times 10^{-4}$ to $1 \times 10^{-6}$ and that the hazard index, $HI$, should not exceed 1 (USEPA 1990a). The incremental carcinogenic risk, $R$, is defined as the sum of the risks from all the possible exposure pathways over a period of 70 years.

**Slope Factors and Reference Dose Factors**

The slope factor, $SF$, and reference dose factor, $RfD$, used in calculating the carcinogenic risk and the non-carcinogenic risk respectively are established through toxicological studies. For each chemical contaminant, there is an established slope factor or reference dose factor depending on whether it is a carcinogenic contaminant or not. Values for the slope factors, $SF$, or reference dose factors, $RfD$, can be obtained through special databases established by the EPA such as the Integrated Risk Information Systems (IRIS 1993). However, for chemicals which do not have any cancer slope factor or reference dose factor developed, the guidelines outlined by the EPA (USEPA 1993a) can be followed to calculate them.

The rationale for calculating cancer slope factors is that for every finite amount of exposure to a carcinogenic chemical, there is a finite probability, however small, of cancer
developing in an individual (USEPA 1989e). Reference dose factors are used to evaluate the effects of being exposed to non-carcinogenic contaminants. Reference dose factors are based on the assumption that protective mechanisms exist in human beings and these have to be overcome before any deleterious health effects occur. In other words, there is a threshold value for non-carcinogenic contaminants below which no adverse health effects might result.

Calculating the Amount of Intake or Exposure to Contaminants

The EPA has developed several equations to calculate the amount of contaminants which would be taken in by an individual through surface water, groundwater, soil or any of the other pathways. The general equation used for calculating chemical intake into the human body is presented below, and represents the idea that the amount of contaminant that enters the human body is dependent on the concentration of that chemical contaminant and the conditions of exposure.

\[ I = \frac{C \times IR \times EF \times ED}{BW \times AT} \]

where

\( I \) = Mass of chemical that enters the body per kilogram of body weight per day (mg/kg/day)

\( C \) = concentration of the chemical (expressed as mg chemical per liter of water, i.e. mg/L, or mg chemical per kg soil, mg/kg)

\( IR \) = Daily ingestion rate on days exposed during the exposure period (measured in mg/day)

\( EF \) = Exposure frequency or the number of exposure events during the exposure period divided by the number of days in the exposure period (usually assumed to be 1 event/day)

\( ED \) = Duration of exposure period (usually assumed to be 7 years)
BW = Average body weight of the receptor of concern during the averaging period (measured in kg)

AT = Averaging Time (measured in years)

The conditions of exposure are expressed by the following factors: the rate at which the contaminants are taken in [intake rate (IR)], how often the person is exposed [exposure frequency (EF)], and the period over which exposure occurs [exposure duration (ecd)]. Body weight occurs in the denominator because intake is always expressed in terms of unit body weight. The averaging time in the denominator averages the mass of contaminant taken in over an average lifetime for carcinogenic effects (usually estimated as 70 years) and over the duration of exposure for non-carcinogenic effects.

The EPA has derived specific equations to estimate carcinogenic risks or non-carcinogenic risks from chemical contamination of soil, surface water, groundwater and other media. In Tables 4 through 6 we present three examples of these equations and discuss how they are used in estimating human health risk.

Table 4 - Equation for Estimating carcinogenic risk, R, or hazard index, HI from human exposure to contaminated soil.

\[
R \text{ or } HI = \frac{C \times ED \times A_1 + A_2 + A_3 + A_4}{BW \times AT \times 365 \text{ days/year}}
\]

where

\[
A_1 = TF_0 \times 10^{-6} \text{ kg/mg} \times IR_{soil} \times ABS_{mg} \times EF
\]

\[
A_2 = TF_1 \times IR_{air} \times EF \times \left(\frac{1}{VF} + \frac{1}{PEF}\right)
\]

\[
A_3 = TF_0 \times 10^{-6} \text{ kg/mg} \times SA \times AF \times ABS_{dern} \times EF
\]

\[
A_4 = TF_0 \times BAF_{DEER} \times IR_{MEAT} \times EF_{MEAT} \times FI_{MEAT}
\]
and

\[ A_1 = \text{the amount of contaminants that enters the human body by direct ingestion of soil or by ingesting of vegetables grown in the contaminated area} \]

\[ A_2 = \text{the amount of contaminants that enters the human body by inhaling the chemicals which vaporize from the soil} \]

\[ A_3 = \text{the amount of contaminant that enters the human body through the dermal tissues or by direct contact with the human body} \]

\[ A_4 = \text{the contaminants that enter the human body by ingesting meat from animals which graze in the contaminated area.} \]

\[ C = \text{chemical concentration is soil (mg/kg)} \]

\[ TF_1 = \text{inhalation slope factor ((mg/kg-day)}^{-1} \text{) for carcinogens; 1/inhalation reference dose (mg/kg-day) for noncarcinogens} \]

\[ TF_0 = \text{ingestion slope factor ((mg/kg-day)}^{-1} \text{) for carcinogens; 1/ingestion reference dose (mg/kg-day) for noncarcinogens} \]

\[ BW = \text{body weight (kg)} \]

\[ AT = \text{averaging time (year) (for noncarcinogens, this is equal to the exposure duration, ED)} \]

\[ EF = \text{exposure frequency (days/year)} \]

\[ ED = \text{exposure duration (year)} \]

\[ IR_{\text{soil}} = \text{soil ingestion rate (mg/day)} \]

\[ IR_{\text{air}} = \text{inhalation rate (m}^3/\text{day)} \]

\[ VF = \text{soil-to-air volatilization factor (m}^3/\text{kg)} \]

\[ PEF = \text{particulate emission factor (m}^3/\text{kg)} \]

\[ SA = \text{skin surface area available for contact (cm}^2/\text{event)} \]

\[ AF = \text{soil to skin adherence factor (mg/cm}^2) \]
\[ \text{ABS}_{\text{derm}} = \text{dermal absorption factor (unitless)} \]
\[ \text{ABS}_{\text{ing}} = \text{ingestion absorption factor (unitless)} \]
\[ \text{BAF}_{\text{Deer}} = \text{meat bioaccumulation factor (kg soil/kg meat)} \]
\[ \text{IR}_{\text{Meat}} = \text{meat ingestion rate (kg/day)} \]
\[ \text{EF}_{\text{Meat}} = \text{meat ingestion frequency (days/year)} \]
\[ \text{FI}_{\text{Meat}} = \text{fraction of meat from contaminated source (unitless)} \]

This equation calculates the amount of chemical contaminant that can be taken into
the human body through the soil. It estimates the amount of chemical which can be
inhaled when the chemical contaminated vaporizes from the soil, the amount of chemical
contaminant which can be ingested through plant or animals which graze on the
contaminated land and the amount of chemical which can be absorbed into the skin by
direct contact with the contaminated soil.

A similar equation which is presented below is used in estimating the human risk
through exposure to contaminated surface water.

Table 5 - Equation for estimating the human health risk by exposure to
contaminated surface water

\[ R \ or \ HI = \frac{ED \times C \times [2 \times \lambda_1 + \lambda_2 + \lambda_3]}{BW \times AT \times 365 \ days/\text{year}} \]

and

\[ \lambda_1 = EF \times TF_o \times SA \times PC \times CF_1 \times CF_2 \]

\[ \lambda_2 = TF_o \times EF \times IR_w \]

where

\[ \lambda_3 = TF_o \times IR_{\text{fish}} \times BAF_{\text{fish}} \times FI \times EF_{\text{fish}} \]
\( \lambda_1 \) provides an estimate of the contaminants inhaled or taken into the human body through direct contact of the skin with water,

\( \lambda_2 \) provides an estimate of the amount of contaminants taken into the human body by drinking the contaminated water,

\( \lambda_3 \) provides an estimate of the contaminants taken into the human body by eating contaminated fish, and where

\[
\begin{align*}
C &= \text{mg of chemical contaminants in liter of water (mg/L)} \\
TF_0 &= \text{ingestion slope factor } (\text{mg/kg-day})^{-1} \text{ for carcinogens;} \\
&\quad \text{1/ingestion reference dose (mg/kg-day) for noncarcinogens} \\
BW &= \text{body weight (kg)} \\
AT &= \text{averaging time (year)} \\
EF &= \text{exposure frequency (days/year)} \\
ED &= \text{exposure duration (year)} \\
IR_{FISH} &= \text{fish ingestion rate (kg/meal)} \\
IR_w &= \text{daily water ingestion rate (L/day)} \\
SA &= \text{skin surface area available for contact (cm}^2\text{)} \\
CF_1 &= \text{volumetric conversion factor for water (1L / 1000cm}^3\text{)} \\
CF_2 &= \text{conversion factor (24hrs/day)} \\
PC &= \text{dermal permeability constant (cm/hr)} \\
BAF_{FISH} &= \text{bioaccumulation factor for fish (L/kg)} \\
FI_{FISH} &= \text{fraction of fish ingested from contaminated source} \\
EF_{FISH} &= \text{fish exposure frequency (meals/year)}
\end{align*}
\]
Table 6 - Equation for calculating human health risk from exposure to contaminated groundwater

\[ R \text{ or } HI = \frac{C \times ED \times [\gamma_1 + \gamma_2 + \gamma_3]}{BW \times AT \times 365 \text{ days/year}} \]

where

\[ \gamma_1 = EF \times TF_1 \times K \times IR_{air} \]
\[ \gamma_2 = EF \times TF_0 \times SA \times PC \times CF_1 \times CF_2 \]
\[ \gamma_3 = TF_0 \times EF \times IR_w \]

\( \gamma_1 \) represents the amount of chemical contaminant which is inhaled,
\( \gamma_2 \) represents the amount of chemical taken into the human body by direct contact with the contaminated groundwater, and
\( \gamma_3 \) represents the amount of chemical contaminant taken into the human body by direct ingestion of contaminated water and

\( C \) = the amount of chemical contained in a liter of contaminated groundwater (mg/L)

\( TF_1 \) = inhalation slope factor ((mg/kg-day)^-1) for carcinogens; 1/inhalation reference dose (mg/kg-day) for noncarcinogens

\( TF_0 \) = ingestion slope factor ((mg/kg-day)^-1) for carcinogens; 1/inhalation reference dose (mg/kg-day) for noncarcinogens

\( BW \) = body weight

\( AT \) = averaging time (year)

\( EF \) = exposure frequency (days/year)
\[ EF_s = \text{number of days an individual showers per year (days/year)} \]

\[ IR_{\text{AIR}} = \text{daily indoor inhalation rate (m}^3/\text{day)} \]

\[ IR_w = \text{daily water ingestion rate (L/day)} \]

\[ K = \text{volatilization factor (L/m}^3) \]

\[ SA = \text{skin surface area available for contact (cm}^2) \]

\[ CF_1 = \text{volumetric conversion factor for water (1 L/1000 cm}^3) \]

\[ CF_2 = \text{conversion factor (24 hrs/day)} \]

\[ PC = \text{dermal permeability constant (cm/hr)} \]

From Tables 4, 5 and 6 we see that many parameters need to be estimated before the risk from exposure to contaminated soil, water or groundwater can be estimated. If \( n \) parameters are to be estimated, and there is a probability of \( \alpha \) associated with making an error in estimating each parameter, then the probability of making at least one error in estimating the parameters is

\[ 1 - (1-\alpha)^n = n\alpha \]

Thus as the number of parameters increases, the probability of making at least one error also increases. In addition to the errors associated with estimating the individual parameters, there are other uncertainties associated with this empirical approach in estimating human health risks. First, there is uncertainty associated with how human populations can be exposed to contaminants. Sometimes the current and future use of contaminated lands is unclear. This makes it difficult to determine which animals use the contaminated land as grazing grounds. Even when the land use is identified, there is uncertainty associated with how much contaminants ingested by animals are transferred to human beings. Moreover, when the geology of the contaminated site is not well understood, there is a high uncertainty associated with which transport pathways...
(especially with groundwater) can be used to transport the chemical contaminants to humans.

In estimating risks from chemical contaminants, the health risk from each chemical contaminant is calculated independently. Hence overlooking the potential joint action of chemicals, that when ingested together can increase or decrease each other's effectiveness. Another uncertainty in estimating the human risk involves the use of reference dose factors or cancer slope factors. These factors are derived from toxicological studies conducted on animals; hence, using these factors derived from animal studies to estimate human health risks involves some uncertainty.

The equations presented in Tables 5, 6 and 7 are modified and applied in quantifying the risks of exposure of wildlife and animals to chemical contaminants. Procedures for assessing ecological risks parallel those for human health risk assessment. However, ecological risk assessment has not yet achieved the formalized approach human health risk has; thus, there are additional uncertainties associated with assessing ecological risks. For example, in estimating the amount of contaminants taken in through dermal contact for wildlife, absorption and adherence factors taken from human health risk assessment are used. Furthermore in estimating the health risk for wildlife, it is assumed that 3% of their diet is soil. This percentage is considered conservative, however it gives as an idea of the risks posed to wildlife by contaminants.

3.2.2 Stochastic Models

The process of estimating human cancer risk from animal-based experiments is fraught with a variety of complex problems that are often difficult or impossible to totally resolve. For example, the appropriateness of any given animal species for assessing the potential of that chemical to cause cancer in humans is obviously dependent on the difference between the species' ability to absorb or uptake, to distribute and store, metabolize and excrete that particular chemical. Furthermore, animal species used to conduct these chemical experiments are highly inbred, homogenous strains, while the human population is generally genetically heterogenous. Also in many instances, the
subgroup at the highest risk; therefore, of greatest interest may represent only a small portion of the exposed population.

Most mathematical models that have been developed to estimate the risk from long-term exposure to a potential carcinogen are concerned only with the problem of low-dose extrapolation. Animal data from experiments conducted at dose levels high enough to produce tumors in an appreciable percentage of the test animals are used to estimate the corresponding risk at ambient levels for the human population. The only consideration generally given to the problems with animal-to-man extrapolation is to express the data on the basis of dose per unit of surface area.

The magnitude of the risk estimate for any given low-dose level will obviously depend on the specific form that is assumed for the dose-response curve in making the estimate. Models that are customarily proposed include the threshold model, the multistage model, and the probit-type models (Hoel et al. 1975).

The threshold model assumes some critical level of exposure below which the carcinogenic process will not be initiated. The arguments which support this model generally center around the concept that when potential carcinogenic chemicals are absorbed into the body at very low levels, the body is able to detoxify them and repair any damage done by the low concentrations of the chemicals. Also there is a possibility that activation of the processes which result in cancer are not initiated at very low dose levels. These threshold values are used in establishing slope factors for carcinogens or reference dose factors for noncarcinogens. The opposing views argue that since human populations are genetically heterogenous, separate thresholds would have to be established for various susceptibility levels, and that no single threshold level would be adequate for a heterogenous population.

The multistage model for carcinogenesis assumes that cancer is a single cell in origin and that the cell undergoes a series of mutations before cancer develops in the human body. This multistage model is represented mathematically as follows:

\[ P(d) = 1 - \exp \left[ -(\lambda_0 + \lambda_1d + \lambda_2d^2 + \ldots + \lambda_kd^k) \right] \]
where $P(d)$ represents the risk (which is a unitless probability number) from taking in a dose, $d$, of chemical contaminants.

$k$ represents the number of transitional events in the carcinogenic process for the particular carcinogen in question and the $\lambda_i$ ($i = 0, 1, ..., k$) are unknown parameters which need to be estimated. When $k=1$, then the cancerous risk from a dose, $d$, of a chemical contaminant is given by the equation

$$P(d) = 1 - \exp \left[-(\lambda_0 + \lambda_1 d)\right]$$

This model is known as the one-hit model, implying that as soon as the chemical contaminant enters the human body, the cancer forming process is spontaneous and does not go through any special stages of mutations before the final cancerous cells develop.

Low dose risk estimates are obtained in the following manner. First, tests of potential carcinogens are conducted on animals and administered doses at which cancerous effects are noticed in the animals recorded. These data are fit to the multistage model and estimated parameter values are obtained. Using the estimated model, the risk of developing cancerous cells at low dose levels is available along with statistical upper bounds. Next, it is necessary to extrapolate the low dose risk estimates for the experimental animal to humans. This is the point where it is felt that the greatest errors are made (Hoel et al. 1975).

The third class of models includes probability distribution models such as the Weibull, logit, and probit models. These models, although they estimate the probability of developing a cancerous effect from exposure to a chemical contaminant, tend to lack an obvious biological basis. For example, the Weibull model used in estimating the probability, $P(d)$, of developing a cancerous effect from a dose, $d$, of a toxic chemical is given by the equation:

$$P(d) = 1 - \exp[-\exp(\alpha d)^{\beta}]$$

where $\alpha, \beta$ are constants.
The logit model for estimating the probability of developing a negative response from a chemical intake of $d$ is given by the expression

$$ P(d) = \frac{1}{\left[1 + \exp(-\alpha - \beta \log d)\right]} $$

where $\beta > 0$ and $\alpha$ is a constant.

The probit model like the logit and Weibull models can also be viewed as a tolerance model and is based on the notion that each animal in a population has its own tolerance to the test compound. Any dose, $d$, not exceeding the tolerance for an individual will have no effect on that individual while any dose exceeding the tolerance will result in a positive response. The risk $P(d)$ of experiencing a response from a dose is estimated as follows:

$$ P(d) = (2\pi)^{-1/2} \int_{-\infty}^{\beta \log(d)} \exp\left(-\frac{u^2}{2}\right) du $$

where $\alpha$ and $\beta > 0$ are unknown parameters. The probit model assumes that the individual tolerances follow a lognormal distribution. Specific steps in the complex chain of events that lead to carcinogenesis are likely to have lognormal distributions.

Figure 3.1 is a log-log plot of risk $P(d)$ from dosage $d$ estimated from the Probit, Logit and Weibull models. For illustrative purposes, the following parameters were used to estimate the risks for the various models:

- **Probit Model** - $\alpha=0$, $\beta=1$
- **Logit Model** - $\alpha=0$, $\beta=1.81$
- **Weibull Model** - $\alpha=-0.58$, $\beta=1.28$

The log-log plot of $P(d)$ versus $d$ for the logit and Weibull models is almost linear at low doses. Risk $P(d)$ estimated from the Probit model decreases more rapidly than that for the logit and Weibull models.
Figure 3.1 - Log-log plot of risk $P(d)$ versus dose for three Response Models

The stochastic models have been found [Ryzin and Krewski, 1981] to adequately model many types of biological dose-response data, but it is an overly simplistic expectation to represent the entire carcinogenic process by one tolerance distribution. One major problem of the probabilistic models is the large statistical uncertainty in extrapolating data orders of magnitude below the observable range. Two sources of variation that contribute to the uncertainty inherent in tumor frequencies are statistical variability which implies a large variation in tumor counts and experiment-to-experiment variability.

The variation observed in tumor frequency is a function of the uncertainty involved in classifying lesions into categories of "tumors" and "non-tumors." This classification of "tumors" and "non-tumors" is especially difficult to make at low doses because a dose-response may exist in the severity of lesion as well as the frequency of lesion, and is mostly made subjectively and may not be consistent from pathologist to pathologist (Squire, 1981a). At high doses malignancy is often clearly defined; however, at low doses the
responses are often equivocal and may be classified as tumors or nontumors, depending on the opinion of the pathologist.

3.2.3 What needs to be done?

The empirical equations adopted by the EPA to estimate human health risk of contamination from chemicals lack parsimony. Due to the high number of parameters that have to be estimated and the uncertainty associated in estimating each parameter, the engineer or risk assessor tends to use the most conservative value for each parameter and ends up with a conservative upper bound on the true potential. The equations do not account for the variance in each parameter; hence, the risk equations do not account for the uncertainty in the parameters. A first step in remedying this problem will be the development of more parsimonious models for assessing human health risk. The stochastic models presented in the previous section are suitable for laboratory analysis but may be too simplistic for assessing human health risk from exposure to contaminants. We will not suggest the exact form these newer models should have but we will describe some key attributes these models should have.

1 The model should identify the inherent variability in estimating human health risks. That is, the main parameters used in estimating human health should not only include their average or expected values but should also account for the variability of each parameter. In other words the models should account for the dispersion of the variables around their mean or central values.

2 The models should be parsimonious in their number of parameters. A fewer number of parameters will lead to a better estimation of the model's parameters and, hence, lead to better estimates of human health risks. Some of the key parameters this new model could address are: the proximity of the contaminated site to humans, animals and vegetation, the chemical properties of the pollutant such as their volatility or solubility, the total amount of chemical contaminant present and the toxicity of the chemical.
3.3 Concluding Remarks

Risk assessment has an obvious appeal to the user of mathematical methods. It is important, though, to recognize both the strengths and weaknesses of a reliance on reducing qualitative assessments such as human health and general environmental impacts to numerical values alone. Mathematical statements or descriptions initially appear to remove the 'complexities' which inevitably surround terms such as reasonable or acceptable risk. There are dangers, though, in making the assessment superficially over-precise, incomprehensible to the less numerate, and alienating to those who believe that it is extremely difficult or even impossible to quantify or put in money-terms some of the risks involved in having the environment exposed to chemical contaminants.

Mathematical approaches to risk assessment expose a problem to logical analysis and help us to identify areas of uncertainty and the issues where relative merits of different actions are clearly seen. Such an approach may not necessarily ease the acceptance of such an appraisal by groups perceiving a risk but will provide the intellectual ground for the appreciation of why such a decision was made. Unfortunately, mathematical analysis can be used for hiding inconvenient or muddled thinking behind a facade of apparent technical and scientific expertise. It is important to realize that mathematical assumptions are still assumptions and require estimates of the errors implicit in them. Using ranges of values rather than only a central estimate is a necessary adjunct to risk assessment and forms the basis of sensitivity analysis which tests how general findings of an assessment may be.
CHAPTER 4 - MANAGING RISKSPOSED BY A CONTAMINATED SITE

4.1 Introduction

Over the past century life expectancy has increased and the main causes of death have changed. Both are due to technological advances in developed countries. Advances in medicine have reduced death rates while other advances such as the harnessing of nuclear energy has created potentially hazardous by products. Doll and Peto's [1981] finding that 80% of cancers are environmental in origin has concentrated the public's attention on the chemical additions to the environment which can cause potential health hazards. Their worry is in no small part due to the publicity surrounding events such as accidents at nuclear facilities (Chernobyl and Three Mile Island) and the abandonment of chemically contaminated sites by major manufacturing companies. This awareness has increased concern that activities such as mining, chemical production and industrial manufacturing produce large amounts of chemical wastes which have risks associated with them. The risks associated with chemical contaminants were not initially perceived or assessed when they were originally introduced, so the public's questioning of the benefits derived from these industries has increased. The anxiety about the benefits of the aforementioned industries increased as old mining companies abandoned tailing ponds and large chemical manufacturing companies dumped their wastes into natural rivers and streams leaving vast contaminated tracts of land. Most of these contaminated sites were completely abandoned by the companies responsible for their contamination and thus became the responsibility of the government. It is estimated that there are over 20,000 solid and liquid waste sites in the U.S [Staples, 1985]. Costs for the clean-up of these sites were estimated in 1985 to be $10-20 billion. Using an inflation rate of 10% the 1995 estimated costs for the clean-up of these sites is approximately $26 - $52 billion dollars, implying that the clean-up of a single site may cost between $1,297,000 and $2,594,000.

A consequence of the perceived risks of chemical contaminants has resulted in a demand by some for absolute safety either by the provision of totally safe chemicals or by the complete removal of hazardous compounds from the environment. Anything less than absolute implies a level of risk or the concept that a chemical in the environment is unsafe
to some extent. The word unsafe may be synonymous with dangerous, implying to some, a degree of recklessness. Hence government or local regulating agencies are faced with a dilemma: should they simply regulate the presence of chemicals in the environment, or do they have to estimate the effects that chemicals can have on the environment and by so doing decide on a permissible amount which can be allowed in the environment. The first method, that is, regulating the presence or hazard of chemicals is easier, but at what economic cost? The second method, which involves regulating the risk a chemical contaminant poses to the environment is much more difficult and involves a lot of uncertainty but may provide some benefits to society as a whole.

The necessity of balancing risk against benefits was emphasized in the clearest manner in Section 2 of the American Chemicals Law passed by Congress in 1976, the 'Toxic Substances Control Act.' This reads:

Authority over chemical substances and mixtures should be exercised in such manner as not to impede unduly or create unnecessary economic barriers to technological innovation while fulfilling the primary purpose of this Act to assure that such innovation and commerce in such chemical substances and mixtures do not present an unreasonable risk of injury to health or the environment.

The problem then becomes twofold: the problem of defining what is an unreasonable risk and how risk can be measured and regulated in the environment. To clearly define what an unreasonable risk is, we will begin by clarifying what an acceptable risk is. How much risk a decision maker is willing to assume can be determined by modeling his/her preferences. The preferences of a decision maker are represented mathematically by utility functions.

Utility Functions

Decisions based on expected monetary values are convenient but can lead to decisions that are not intuitively appealing. For example, consider the following two games. Let us say we have the opportunity to play either game, but only one time.
Game 1
Win $40 with probability 0.5
Lose $2 with probability 0.5

Game 2
Win $2000 with probability 0.5
Lose $1500 with probability 0.5

Game 1 has an expected value = 40 \times 0.5 - 2 \times 0.5 = $19
Game 2 has an expected value = 2000 \times 0.5 - 1500 \times 0.5 = $250

If our decision to play is based on expected value, then we would choose Game 2.
However, most people will consider Game 2 to be riskier than Game 1, and hence it seems more reasonable that most people will actually prefer Game 1.

Using only expected values to make decisions means that the decision maker is considering only the expected payoff. The expected payoff according to probability theory is the amount we would be likely to win over an infinite number of plays of the game. The expected payoff does not take into account the possible range of values we could obtain by playing the game a number of times. After all if we play each game 10 times, the worst we could do in Game 1 is to lose $20. But in Game 2 we could expect to lose $15,000! Thus we need formulations which can capture the riskiness in making such decisions. Utility functions are mathematical formulations that enable us to model a decision maker's risk attitude or model his/her preferences. Utility functions are a way of translating dollars $X$ into utility units $u(X)$.

Three basic risk attitudes are modeled by utility functions: risk-averse, risk-seeking and risk-neutral attitudes. A risk averse person is an individual who is afraid of risk or sensitive to risk. Such a person will most likely not play Game 2. The risk-seeker on the other hand is someone who is willing or eager to enter into a gamble; for example, such a person will be willing to play Game 2. Finally, a risk neutral person does not care about risk and can ignore risk aspects of the alternatives he or she faces. Thus a risk-neutral person can base their decisions solely on expected monetary values, because it ignores
risk. The utility curves for a risk averse, risk-seeking and risk-neutral person are shown in Figure 4.1.

![Utility curves for different risk attitudes](image)

Figure 4.1 - Three different shapes for utility functions

The risk-averse curve shown in Figure 4.1 can be modeled by utility functions such as \( \log(X) \), \( 1-\exp(-X/R) \) and \( \sqrt{X} \). In all these functions \( X \) is the monetary value and \( R \) is a constant known as the risk tolerance. Linear functions of the form \( A+BX \) model risk neutral attitudes and functions such as \(-AX^2+BX+C\) can be used to model risk seeking attitudes, where \( A, B \) and \( C \) are constants.

The utility functions embody an important fundamental tradeoff: monetary return versus riskiness. The basic reason for using a utility function as a preference model in decision making is to capture our attitudes about risk and return. Obtaining high returns and minimizing risk are two conflicting objectives. In managing the risk posed by chemical contaminants, we aim at achieving a host of economic, environmental and health related objectives. But these objectives conflict within the decision making framework. Consequently, one might not be able to achieve an optimal outcome for all the objectives simultaneously. Hence, we need a framework which will allow us to make compromises and consider the tradeoffs necessary to achieve the best outcome. The theory of multiattribute utility (Keeney & Raiffa, 1976) helps us to deal with such conflicting objectives.
The first step in dealing with the management of chemicals in the environment is to construct a value tree. A possible value tree for evaluating contaminant risks is presented in Figure 4.2.

![Figure 4.2a - A value tree for evaluating chemical contaminant risks](image)

A value tree begins with several fundamental objectives as main branches. Each fundamental objective is then expanded and explained with more specific objectives and values. For example, the fundamental objectives for evaluating chemical contaminants are: minimize cost, maximize direct benefits, minimize risks and address other impacts. The
minimize risk objective can be broken down further into environmental and health risks. Environmental risks can be further divided into risks to flora and fauna, contamination of air, water, and soil, and deterioration of aesthetics in the contaminated area. Health risks can be subdivided into accidents and human and animal health risks from exposure to the chemical contaminants. Although we would like to minimize such risks, choosing an alternative that does so will probably imply a reduction in benefits gained in other dimensions. Completing the value tree requires identification of attributes that correspond to the detailed objectives. That is at the end of each branch there is a subdivision of the detailed objective into operational attributes. According to Clemen (1990), an attribute is operational if it has the following characteristics:

1. One can explain to someone else what to measure and why.
2. The measurement of the attribute is a reasonable chore.
3. Given the measurement by someone else, one would be able to tell how well the alternative in question achieved the objective.

In some cases, however, the attribute corresponding to an objective may not be so obvious. Take the "minimized risks" branch in Figure 4.2a, for example. The detailed objective is to minimize the human and animal health risks from exposure to the contaminants. But how can these risks be measured? We have discussed means of quantifying human and animal health risks in Chapter 3 but converting these probability values to monetary values is not easy. One way to obtain monetary values for human life is to quantify our willingness to pay to save a human life. This approach prevents us from dealing with the controversial issue of quantifying the value of human lives.

The essential problem in multi-objective decision making is deciding how best to tradeoff an increased value on one objective for a lower value on another. Making these tradeoffs is a subjective matter and requires the decision maker's judgement.
The next step after constructing a value tree is to assess numerical weights for each fundamental objective. This weighting procedure asks how the objectives compare in terms of importance. For the value tree presented in Figure 4.2a, we could assume that the fundamental objectives carry the same weight. Thus, on a scale of 0 to 100% each objective will be assigned a weight of 25%. Next, we would assign weights to each attribute. We could assign the weights by considering the relative importance of each attribute. Let us take the "minimize risks" branch for example. We could say that the health risks were twice as important as the environmental risks. Furthermore, for the sake of elicitation we could assume that risks due to accidents and human and animal exposure carried the same weight. A similar assumption can be made for the environmental risk branch, meaning we will assign the same weights to contamination of air, water, soil, flora and fauna and deterioration of aesthetics. The completed value tree is presented in Figure 4.2b. A fundamental property of the value tree is that the weights assigned to the fundamental objectives should sum to 1 or 100%, and the weights assigned to the attributes for each fundamental objective should also sum to 1 or 100%.

One can use the value tree in making decisions about the most suitable place to locate a chemical manufacturing plant. For a number of possible sites we could use this method of assessment to tabulate weights for each attribute at each site. An example of such a matrix of weights and scores is presented in Table 7. For the purpose of illustration we assigned arbitrary scores to the various attributes. Scores are numerical values between 0 and 1 assigned to a particular attribute for a particular site. Unlike weights, the numerical score values assigned to the attributes of a fundamental objective do not need to
sum to 1 or 100%, but the score assigned to each attribute cannot exceed the maximum value of 1. Scores are assigned to particular attributes as a measure of how well an option (which in our example is a site) will achieve the specific attribute and the fundamental objective in general. Thus a decision maker assigns numerical scores between 0 and 1 or 0 to 100% as a measure of his/her belief about how an option (site) will achieve an attribute. An analogy can be made between test scores and scores assigned to attributes in a value tree. We can view the attributes as particular subjects (say Mathematics, English, History etc.) and the score as a measure how well a particular student performs on a subject. For test scores we obtain a measure of a student's performance after he/she has written the test, however scores for attributes are assigned a priori, that is before the specific facility is located at a proposed site. Consequently, the assigned scores express the decision maker's belief about how well the site will perform on a particular attribute if it were located there. After assigning scores to the various attributes we obtain a weighted score of each fundamental objective by multiplying the score for each attribute by its weight. For example, the weighted score for the "minimize costs" objective for site 1 was obtained as follows:

\[ \text{Weighted score for Site 1 ("Minimize Costs" Attribute)} = 25\% \times (50\% \times 0.2 + 25\% \times 1 + 25\% \times 1) = 15\% \]

The weighted scores for each fundamental objective for a particular site were then summed to obtain the total weighted score for the site. For example the total weighted score for site 1 was obtained as follows:

\[ \text{Total Weighted Score for Site 1} = 22.5\% + 15\% + 13.61\% + 16.25\% = 67.36\% \]
After obtaining the total weighted scores for each site, we select the site with the highest total weighted score as the most favorable site to locate the chemical producing plant. For our example, the site with the highest score is site 1. Thus for this hypothetical case, one would select site 1 as the most favorable site for a chemical plant. It is noteworthy that there are a number of methods to assign weights to attributes. A detailed treatment of assigning weights to attributes is presented by Clemen (1990).
Figure 4.2b - A value tree for evaluating chemical contaminant risks with relative weights assigned to each value and attribute.
Table 7 - Matrix of Weights and Scores for three Chemical Plant Sites

<table>
<thead>
<tr>
<th>Attributes</th>
<th>(%) (Weights)</th>
<th>Site 1 (Scores)</th>
<th>Site 2 (Scores)</th>
<th>Site 3 (Scores)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Maximize Direct Benefits (25%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Company Profitability</td>
<td>100</td>
<td>0.9</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>Weighted Score</td>
<td>22.5</td>
<td>22.5</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td><strong>Minimize Costs (25%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleanup Costs</td>
<td>50</td>
<td>0.2</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>Transportation of Contaminants</td>
<td>25</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Other Structures Built to Contain Contaminants</td>
<td>25</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Weighted Score</td>
<td>15</td>
<td>11.25</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td><strong>Minimize Risks (25%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accidents</td>
<td>33.35</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Human and Animal Exposure</td>
<td>33.35</td>
<td>0.3</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>Air, Water and Soil</td>
<td>11.1</td>
<td>0.5</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>Flora and Fauna</td>
<td>11.1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>11.1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Weighted Score</td>
<td>13.61</td>
<td>14.16</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td><strong>Other Impacts (25%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimize Resident's Concern</td>
<td>25</td>
<td>0.6</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>Maximize Acceptability</td>
<td>25</td>
<td>0.5</td>
<td>0.4</td>
<td>0.2</td>
</tr>
<tr>
<td>Maximize Employment</td>
<td>25</td>
<td>0.6</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Maximize Business Growth</td>
<td>25</td>
<td>0.9</td>
<td>0.6</td>
<td>1</td>
</tr>
<tr>
<td>Weighted Score</td>
<td>16.25</td>
<td>10</td>
<td>9.38</td>
<td></td>
</tr>
<tr>
<td>Total Weighted Score</td>
<td>67.36</td>
<td>57.91</td>
<td>46.88</td>
<td></td>
</tr>
</tbody>
</table>

The multiattribute approach is beneficial for a company trying to site a chemical producing plant to attain the goals similar to those outlined in Figure 4.2b. However, it does not help us in determining what the acceptable level of pollution is. A community
concerned about minimizing risks posed chemical contaminants might not be willing to even consider the benefits to be derived from the chemical production. Their main concern will be to minimize the risks posed by the contaminants or to reduce the chemical concentrations in the environment to an acceptable level. In order to determine what the acceptable level of contamination is we have to be clear about the definition of acceptable risk.

4.1.1 What Is an Acceptable Risk?

The levels of risk that an individual finds acceptable depends upon a number of factors. These factors may include the person’s prior experiences, the benefits the person expects to derive from the risky activity, and may even depend on what the person perceives as a risky event. What might be safe for one person may cause fear in another person. However as we look at society as a whole, we find that people consistently take part in activities which have risks associated with them. Pochin [1991] has pointed out that the risk of a child dying from complications arising from common vaccinations is 1 in $10^5$ cases, similarly the regular train traveller accepts an annual risk $10^{-5}$ of death without too much concern, implying that people who take part in these activities knowingly accept those risks and are not necessarily deterred from those events just because certain risk levels are associated with them. It is difficult though to define what is an acceptable risk for a society as a whole. Certain people in society are prone to higher risks. For example more than 1 out of every 1000 males between the ages of 20 and 30 years are routinely exposed to risk of death in any one year. Various occupations and participants in some sports have higher rates of death, in the region of $3 - 6 \times 10^{-3}$ per year. One occupation - President of the United States carries an annual risk of death by assassination of 2%, meaning that in every 100 years we should expect about 2 U.S. presidents to be assassinated.

At the other end of the scale are risks which are viewed by some as trivial. The EPA considers this to be an annual risk of death of $10^{-6}$ or less (that is, one death per year per million exposed to the hazard). This level of $10^{-6}$ annual risk is used conventionally by
most regulating agencies such as the OSM (Office of Surface Mining) and OSHA (Occupational Safety and Health Association). Individuals may also regard as serious risks which society as a whole may regard as trivial. This is because risk involved in an activity is perceived differently by different people. In conclusion, there is no right or wrong definition of acceptable risk. However, in defining acceptable risks posed by contaminants we accept the EPA's definition, as a chemical risk which can result in the death of 1 in a million people per year.

One needs to distinguish between individual and collective risks. Risk quantification aims at estimating the probability that an individual may experience a detrimental health effect from exposure to a chemical contaminant. Although estimates from the risk quantification may indicate low levels of risk to an individual, the overall risk to society may be high. The societal risk can be defined as the probability that at least one individual may experience detrimental health effects from exposure to the contaminants. If the risk posed to an individual from exposure to the contaminants is \( p \), then for a society of population size \( n \), the probability that at least one individual experiences a detrimental effect is

\[
\text{Probability that at least an individual experiences a detrimental effect} = 1 - P(\text{no one experiences a detrimental health effect})
\]

but

\[
P(\text{no one in the society experiences a detrimental health effect}) = (1-p)^n
\]

hence

\[
\text{Probability that at least an individual in the society experiences a detrimental effect} = 1 - (1-p)^n \approx np
\]

Thus the societal risk can be approximated by \( np \). Consequently for areas with large populations we realize that although individual risks may be extremely low, the societal risks may be unacceptable.

4.1.2 Perception of Risk

Perceptions of the risks involved in an activity vary amongst individuals. People use different methods to evaluate their own individual risks. People might evaluate an activity as risky depending on whether they gain their livelihood from the activity, whether
they enjoy the activity and ignore the risks, whether they feel that 'naturally occurring
risks' are better than risks from man-made sources, or may perceive events with immediate
extreme effects as more risky than slow events with the same effects. Risks which are
'new' seem to arouse more concern than those that are 'routine.' Outcomes which are rare,
unpredictable, and catastrophic, such as chemical plant explosions, are viewed as more
disturbing than those that are common, predictable, and small in size, such as road
accidents, even if the overall cost in human life and suffering may be similar. There also
seems to be a 'dread' of the unknown in people's perception of certain types of risk.

The failure of people to conform their perceptions and their choice of actions to
reduce risk according to the statistical and mathematical analyses of a risky event should
not be dismissed. For example science has presented statistical and medical evidence
about the riskiness of smoking. However, these facts do not deter some people from
smoking. Aspects of human behavior may be of considerable importance in evaluating
risks, as the persistence of cigarette smoking illustrates.

'Natural' contaminants and toxic compounds in food may be considered
acceptable even though such fungal and bacterial contamination may cause illness, while
food additives whose introduction (or identification) in foodstuffs is to assist in
preservation may or may not be acceptable to some people. Like most issues of risk-
perception other dimensions are relevant, such as who obtains benefit from such 'improved
storage' or use of previously unusable food.'

The benefits derived from a risk producing source can change the way in which
people perceive the risk. For example, mine workers living near a tailings pond may find
the risks of contamination from the pond more acceptable than non-mine workers because
they derive their livelihood from the mining activities.

The question that remains to be answered is 'do the benefits warrant the risk?'
Attempts can be made to relate risk and benefit to some common scale, which if this is a
monetary base, converts the procedure into a form of cost-benefit analysis. Assigning
monetary values to some procedures may be relatively easy. For instance, costs associated
with reducing risks due to contaminated sites can be easily calculated, but assigning a
monetary value to the benefits derived from a chemical contaminant process can be difficult. For example how do we assess the benefits of a drug? What is the monetary value of reduced mortality, increased life span, and improved quality of life? Similarly what are the cost of adverse effects of chemicals which may result in accidents, injuries, deaths, birth defects, cancer or mutations?

Estimating the monetary value of a life, will always be controversial. We perceive our own lives as priceless or of inestimable value as it is the only one we have. Others may attempt to place a value on it. Various approaches have been taken. They include the value of life as determined by insurance companies, court awards (which may reflect adversarial litigation), loss of future earnings (which might discriminate against certain groups such as pensioners who might have zero or even negative values), or questionnaire evaluation. In the last case respondents are asked how much they would pay to reduce risks, such as how much increase in air fare in return for improved aircraft safety. But there can clearly be no 'correct' approach to this. The problem is that people do not really behave in the way they indicate in surveys. One way to circumvent this problem with the value of human lives is to quantify human lives as our willingness to pay to save a life.

Another question that arise in cost-benefit analysis is whether all the risks and benefits spread evenly throughout the population or one group is taking the risk and another receiving the benefits. Food additives, for instance, improve the storage of food, prevent spoilage and improve the appearance of the original foodstuffs, but do the benefits which result from the more efficient use of the food result in lower costs to the consumer, a greater profit to the manufacturer, or both? Are the small risks that the consumer may be exposed to compensated by the benefits they receive? Would the consumer be prepared to pay more or accept food in a different form to remove a perceived risk to themselves from food additives?

4.2 Managing Risks

After the risks from contaminated sites have been estimated and evaluated, there remains the problem of how to determine an acceptable level and ensure that acceptable
risks are achieved. The concept that society as a whole determines an acceptable level of risk is, of course, an oversimplification. There may not be a consensus of view in the evaluation of risk, but at some point in time a decision is made through some process of public opinion, expert guidance or governmental action. The process of getting and keeping risks below the set level is called risk management. Such processes may result in an apparent reordering of priorities different from those arising from a strictly mathematical risk-benefit analysis. It is doubtful though how any other approach can provide outlets for concerns which to some may appear irrational yet represent genuine human worries. Three approaches to risk management can be identified: 1) the absolute safety approach, 2) the 'as low as reasonably achievable' approach and 3) the safety standard approach. and 4) the cost-effective approach.

4.2.1 Absolute Safety Approach

One method of risk management attempts to develop a strategy which demands absolute safety. Any evidence of harm from a process or chemical contaminant would be sufficient to suspend its use. The most obvious case is the so-called 'Delaney Amendment' which was enacted by the US Congress in 1958 as an amendment to the Food and Drug Act. This requires that

no [food] additive shall be deemed to be safe if it is found .....after tests which are appropriate for the evaluation of the safety of food additives to induce cancer in man or animals.

Such an approach gives no weight to the potential benefit of the chemical. It also provides little guidance on how toxicological studies which determine the cancerous potential of a chemical are to be evaluated (Issues arising from the Delaney amendment are discussed in more detail by Kessler [1984]).

An issue which arises is how does one determine that a food is free from a particular proscribed constituent or chemical contaminant? The level of detection of many chemicals such as fungal contaminants or pesticide residues was until recently at about 1 part in a million

87
Technical developments have made it possible to detect traces of chemicals at 1 part in a billion \((1 \times 10^{-9})\). The costs of preventing such levels of chemical contamination can be prohibitive.

4.2.2 'As Low as Reasonably Achievable' Approach (ALAR)

A second approach is the development of regulations and procedures which place a responsibility on the operator of a plant to decrease the hazard level as low as is reasonably achievable. This approach is dependent on the definition of the word 'reasonably'. For example, this term is applied in the field of radiological protection. The International Committee on Radiological Protection suggests that use of radiation should be justified by the benefits of any procedure exceeding the sum of the costs of protective measures and radiation-induced detriments to health. Furthermore the protection from the hazard should be increased until the costs of further improvements clearly exceed the corresponding reduction in risk. At this point exposures are described as ALAR. The ALAR approach is mostly applied in cases when it is difficult to determine the safe hazard level as in the case of radiology. The ALAR approach can be interpreted as the point when the costs are seriously out of proportion to the benefits achieved by a reduction in risk.

4.2.3 Safety Standard Approach

Another approach to risk management is to set safety standards. Specific upper limits of exposure called threshold limit values have been set for over 500 common industrial chemicals in the workplace [American Conference of Governmental Industrial Hygienists, 1986]. These values are based on airborne concentrations of chemicals believed to produce no adverse effects during routine work exposure. Threshold limit values are usually derived from exposure in industry, from volunteer studies, or from experiments using animals. Unfortunately, threshold values do not provide assurance of safety; in fact some workers may suffer illness or have existing conditions aggravated. Threshold limit values simply represent control levels at which exposure is at present
considered safe. Exceeding such levels is considered reprehensible and may result in legal consequences. Threshold limit values are, therefore, not 'safe' doses.

For many chemicals no evidence of a 'safe' dose has or can be attained. The use of threshold limit values is a means of limiting the concentrated exposure to as few individuals as possible using existing procedures without exposing any single individual to unreasonably high concentrations. Threshold limit values can be altered, usually by lowering them, either as a result of fresh evidence of risks and potential hazards or because technical advances can reduce the exposures.

4.2.4 Cost-Effective Approach

The cost-effective approach aims at managing contaminant levels which will pose levels of risk to society such that the benefits derived from the chemical producing process will outweigh the risks. The detailed method of applying such an approach will begin by listing all the benefits and risks to be derived from the chemical. An example of such a listing is provided in Figure 4.2. A level of risk which is acceptable to society but that also helps optimize the benefits to be derived from the chemical production is set. There is no example to illustrate where this approach has been applied but it is presented for the sake of completeness.

4.3 Risk Management in Practice

In practice, managing the risks associated with contaminated sites involves a mix of safety standards and the 'as low as reasonably achievable' approach. A key factor in cleaning a contaminated site is establishing a justifiable endpoint, by answering 'how clean is clean?'. The EPA considers a contaminated site clean when the contaminants do not pose a substantial risk to the environment. These substantial risks are defined as follows: a) for carcinogenic compounds the estimated risk should not exceed a range of 1 x 10^-4 to 1 x 10^-6 per year and b) for a noncarcinogenic compound the hazard index should not exceed 1 per year [USEPA 1990a]. We have defined hazard index in chapter 3.
In order to achieve these acceptable levels of risk at a contaminated site, the decision cycle, shown in Figure 4.3, is applied. The procedure presented in Figure 4.3 provides us with a stepwise approach to correctly ascertain the necessary extent of remediation for an identified waste site.

Three main assumptions are considered here. First, the existence of contamination substantial enough to be of potential concern has been established. Second, the source is considered sufficiently large such that the problem will not rapidly correct itself. Third, an acceptable safe concentration has been or can be established for the chemicals of concern.

Beginning with Figure 4.3, the first steps for the site evaluation include defining the scope of the problem. The main emphases are on 'what chemicals are found where and in what concentrations?' This is followed by a determination of approximate rates of continued chemical input from the source. Rapid or catastrophic release of chemicals to groundwater would call for a more immediate remediation than the slow leaching of a source. These data provide a definition of the contaminants at the site.

Then the following question is asked: Does the preliminary data suggest any potential problems? If not, one ends the investigation. If potential problems are indicated, a first stage or tier 1 risk assessment is carried out to estimate the potential consequences on human health or the environment in general. The tier 1 risk assessment is conducted to estimate the worst possible effects on people and on the environment in general.

If the tier 1 risk assessment reveals a high level of risk, then the chemical contaminants at the site are remediated with the appropriate technique. After the initial remediation of contaminants at the site, chemical concentrations are monitored and a tier 2 risk assessment is performed using the average concentration of contaminants at the site. If the tier 2 risk assessment yields an acceptable risk level, we continue to monitor the site to ensure that the chemical concentrations are at an acceptable level. If not, we remediate again and go through another cycle of risk assessment.

This cyclic approach is necessary because the current technology for cleaning most contaminants does not guarantee a complete clean-up of all contaminants at the site. Thus
Figure 4.3 - Decision Cycle For Evaluating Potential Waste Site Problems
the clean-up of contaminated waste is possible in a manner that is scientifically adequate, technologically feasible and cost effective.

In prioritizing the cleanup of a number of contaminated sites, we would begin the cleanup campaign with the site that poses the highest societal risk. Using societal risks instead of individual risks would ensure that the risk posed to society as a whole would be minimized.
CHAPTER 5 - SUMMARY AND CONCLUSIONS

5.1 Summary

Before making an optimal decision about the cleanup of a contaminated site, we have to identify the types of contaminants present, their spatial distribution, and the possible effects the contaminants can have on man and the environment in general. Our determination of the physical attributes and distribution of the contaminants at the site is governed by uncertainty. At most abandoned waste sites, we are uncertain about the processes that generated the pollutants, the local geology, and the possible effects the contaminants can have on the environment. In our work, we identified a systematic method for determining the effects chemical pollutants can have on the environment.

The first stage in assessing the risks chemical pollutants can have on the environment is to identify the physical and chemical characteristics of the contaminants. This process known as hazard identification involves identifying how widely the contaminants are distributed, the concentration(s) of contaminant(s), and their chemical characteristics. Conventionally, hazard identification begins by taking samples either on a regular grade or in a random fashion at the site. These samples are then analyzed in the laboratory to determine their chemical properties and the amount of contaminants contained within each sample.

Besides finding contaminant concentrations, the samples are sometimes used in toxicological studies to determine the threshold limit value at which laboratory species will develop cancerous or other detrimental health conditions when exposed to a contaminant. The threshold limit values for most widely used chemicals have already been determined and can be found in data bases such as IRIS which was established by the EPA. Whenever a new contaminant is identified whose threshold limit value and/or toxicological properties are unknown, extrapolations are made from chemicals with similar toxicological properties.

To determine the amount of money to be spent in a sampling campaign, we have to make a rough estimate of how much it would cost if the contaminants are improperly
cleaned. We illustrated how to determine the amount of money to be spent in a sampling campaign with two examples in Chapter 2. In the first, we determined the value of perfect information. Perfect information in our case is data which would help us predict with 100 percent certainty the distribution of contaminants at a site. In the second example, we estimated the value of imperfect information. Imperfect information is also known as noisy data or data which can help us determine the contaminant distribution but with an associated probability less than one and greater than zero. In both cases, we showed that it was advantageous to obtain samples from the site before a decision can be made about how to clean it.

After developing a budget for sampling, we can then determine which approach is most appropriate in estimating the contaminant distribution: a classical geostatistical approach or a Bayesian approach. The classical geostatistical approach is suitable in cases where we have a budget large enough to acquire an exhaustive data set from closely spaced grid or random sampling. This large and exhaustive data set will enable us to derive variograms which describe the spatial pattern of contaminant concentrations at the site. Having obtained accurate descriptions of the spatial patterns, we can apply kriging to obtain the mean and maximum concentrations of contaminants which would be used in estimating human and ecological risks from exposure to the pollutants. The Bayesian approach is a very useful tool in determining contaminant distributions at sites with limited sampling budgets. It is flexible and helps us to refine our knowledge about contaminant distributions as more data becomes available.

Another approach which helps us to determine contaminant distribution from sparse data sets is the binary logistic regression approach discussed in Chapter 2. This method does not give us exact contaminant concentrations for a site location. However, it determines the probability of exceeding a specific concentration level at a given point. Thus, when given a sparse data set, this approach enables us to estimate the expected contaminant concentration at any given location of the site.

The next step after hazard assessment is to calculate the health risks from exposure to the chemical contaminants. The EPA has developed a set of equations that are used to
estimate the probability of developing cancer or experiencing deleterious health effects from exposure to chemical contaminants. These equations are based on the assumption that exposure to these contaminants occurs through specific pathways: groundwater, air, surface water and soil. They consider these media as independent exposure pathways. In other words, the detrimental effects a person or the environment experiences will be the sum of the individual effects of exposure through air, surface water, soil and/or groundwater. One assumption made by the EPA in estimating the health and environmental risks from chemical exposure is that when two or more chemicals are present they act independently on a biological receptor (that can be a human being or any living organism). By making this assumption the EPA ignores the joint action of chemicals, that when taken together increase or decrease each other's effectiveness.

Another problem with the EPA's method of human and ecological risk assessment is that the equations used in estimating the risks involve too many parameters. As the number of parameters in an equation increases, the probability of making at least one error in estimating these parameters increases as well.

Another class of models that is used in estimating human and ecological risks from exposure to chemical contaminants is generally known as stochastic or probabilistic models. Three main types of models can be identified in this group: threshold models, multistage models and general probability models (Weibull, logit and probit models). The threshold models assume that there is a certain amount of chemical contaminants that can be safely detoxified by the human body. Hence, for each chemical contaminant we can identify a safe level of exposure above which humans and other organisms will experience detrimental health effects. Arguments against this type of modeling suggest that since human populations exhibit heterogeneity in genetical characteristics, each person will have a specific tolerance level of exposure to chemical contaminants. Thus, a singular value cannot be applied as the safe threshold level of exposure for a human population. The multistage models posit that when a person is exposed to a pollutant, a single cancerous cell develops and undergoes a series of mutations in the body before the cancer is detectable. Multistage models assume that the development of any detrimental health
effect from exposure is only dependent on the concentration and amount of pollutants to which an individual is exposed. A special case of the multistage model is the single-hit model which assumes the development of cancer in the body is a spontaneous incident that occurs after a person is exposed to chemical pollutants. The general probability models tend to lack an obvious biological basis but recognize risk as uncertain or probabilistic. Thus, when a person is exposed, there is a certain probability that the person will develop a deleterious health effect. These mathematical models all assume that as the amount of chemical an individual takes into his/her body increases, the probability of developing health problems increases.

After estimating human health and ecological risks from exposure to chemical contaminants, how do we incorporate this information in our decision to clean up a contaminated site? To answer this question, we described a systematic scheme of incorporating risk factors into the decision cycle for contaminated site clean-up in Chapter 4. This decision cycle for cleaning up a site begins with an investigation of the nature and distribution of contaminants. Then the following question is asked: does the preliminary data suggest any potential problems? If they do not pose any potential problems, then we do not need to take any immediate action to clean up the site. However, if potential problems are indicated, an initial risk assessment is carried out to estimate the potential consequences on human health or the environment in general. This initial risk assessment is performed using the maximum concentration of chemicals at the site. This is done in order to estimate the worst case scenario for the effects of the contaminants.

If the initial risk assessment reveals a high level of risk, then the chemicals at the site should be remediated immediately with the appropriate technology. A second risk assessment is then performed with the average concentration of contaminants at the site. If the results of this second assessment still indicate an unacceptable risk level, we remediate again continuing this assessment-remediation cycle until an acceptable risk level is obtained. Finally, we continue to monitor the site to ensure that the chemical concentrations remain acceptable.
5.2 Concluding Remarks

Cleanup of contaminated sites is possible in a manner that is scientifically adequate, technologically feasible and cost effective. Remedial activities must be guided by a complete comprehension of identified problems through hazard and risk assessment. The techniques that are used in defining the hazards at a site have been discussed in Chapter 2. These techniques are being refined and newer techniques are being developed to provide us with more cost effective methods of determining the physical and chemical characteristics of contaminated sites. The science of risk assessment is relatively new, and better estimation methods have to be developed to simplify the process of assessing human health and ecological risks. These newer methods should involve fewer parameters than the models posited in the EPA equations presented in Tables 4, 5 and 6 in Chapter 3. As newer models are developed and applied, we should understand that risk assessment will always be controversial. There will never be the 'correct' answers and 'absolute' solutions. To assume otherwise would be to accept that there will be no further changes in the views, values, rights, and duties accepted by society and its members over time.
REFERENCES

American Conference of Governmental Industrial Hygienists, 'Documentation of the Threshold Limit Values', 5th edn., ACGIH Inc., Cincinnati, Ohio, 1986.


98


