1-1-2010

Investigating an optimal decision point for probability bounds analysis models when used to estimate remedial soil volumes under uncertainty at hazardous waste sites

Charles O. Dankwah
Walden University

Walden University

COLLEGE OF MANAGEMENT AND TECHNOLOGY

This is to certify that the doctoral dissertation by

Charles O. Dankwah

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Aridaman Jain, Committee Chairperson,
Applied Management and Decision Sciences Faculty

Dr. Thomas Spencer, Committee Member,
Applied Management and Decision Sciences Faculty

Dr. Shannon Lynch, University Reviewer
Applied Management and Decision Sciences Faculty

Chief Academic Officer

David Clinefelter, Ph.D.

Walden University
2010

ABSTRACT

Investigating an Optimal Decision Point for Probability Bounds Analysis Models When Used to Estimate Remedial Soil Volumes under Uncertainty at Hazardous Waste Sites

by

Charles O. Dankwah

M.S., The George Washington University, 1986

B.S., University of Science and Technology, Ghana, 1974

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Applied Management and Decision Science

Walden University

May 2010

Abstract

Hazardous waste site remediation cost estimation requires a good estimate of the contaminated soil volume. The United States Environmental Protection Agency (U.S. EPA) currently uses deterministic point values to estimate soil volumes but the literature suggests that probability bounds analysis (PBA) is the more accurate method to make estimates under uncertainty. The underlying statistical theory is that they are more accurate than deterministic estimates because probabilistic estimates account for data uncertainties. However, the literature does not address the problem of selecting an optimal decision point from the interval-valued PBA estimates. The purpose of this study was to identify the optimal PBA decision point estimator and use it to demonstrate that because the PBA method also accounts for data uncertainties, PBA estimates of remedial soil volumes are more accurate than the U.S. EPA deterministic estimates. The research questions focused on determining whether the mean or the 95th percentile decision point is the optimal PBA estimator. A convenience sample of seven sites was selected from the U.S. EPA Superfund Database. The PBA method was used to estimate the remedial soil volumes for the sites. Correlation analyses were performed between the mean and 95th percentile PBA estimates and the actual excavated soil volumes. The study results suggest that the lower bound 95th percentile PBA estimate, which had the best R^2-value of 89%, is the optimal estimator. The R^2-value for a similar correlation analysis using the U.S. EPA deterministic estimates was only 59%. This confirms that PBA is the better estimator. The PBA estimates are less contestable than the current U.S. EPA deterministic point estimates. Thus, the PBA method will reduce litigation and speed up cleanup activities to the benefit of the U.S. EPA, corporations, the health and safety of nearby residents, and society in general.

Investigating an Optimal Decision Point for Probability Bounds Analysis Models When Used to Estimate Remedial Soil Volumes under Uncertainty at Hazardous Waste Sites

by

Charles O. Dankwah

M.S., The George Washington University, 1986

B.S., University of Science and Technology, Ghana, 1974

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Applied Management and Decision Science

Walden University

May 2010

Acknowledgements

It is my pleasure to thank those who have made this dissertation possible. I would like to express my deepest appreciation for my dissertation chair, Dr. Aridaman Jain, for making it a pleasure to walk through the long dissertation approval process. I appreciate his quick responses to my draft papers and his thought-provoking questions. I would also like to thank my committee members, Dr. Thomas Spencer and Dr. Shannon Lynch, for their constructive engagement throughout the dissertation process. My list will not be complete without thanking my family members. I thank my wife, Sally, and my children, Aileen and Broadwin, for their support and cooperation.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Background

Since the 1980s, a series of environmental regulations have been enacted in the United States and these have had financial impacts on companies that generate hazardous wastes in the course of their normal operations. One such regulation is the Comprehensive Environmental Response and Comprehensive Liability Act (CERCLA) of 1980.

As part of the due diligence requirements of CERCLA, the United States Environmental Protection Agency (U.S. EPA) requires all hazardous waste generating facilities in the USA to conduct a Preliminary Assessment (PA). The purpose of a PA is to investigate the potential for soil or groundwater contamination on a specific parcel of commercial or industrial real estate based on site hydrogeologic features, past site usage practices, and existing potential hazards such as underground storage tanks. If the potential for soil or groundwater contamination is identified, a Remedial Investigation/Feasibility Study (RI/FS) should be conducted to determine the full extent of contamination. If any contaminant is detected above the U.S. EPA's risk-based cleanup levels, then cleanup action is required.

At every hazardous waste site, there are one or several potentially responsible parties (PRPs). The PRPs are all the past and present owners and non-owners of the site who's past or present activities have contributed to the site soil and/or groundwater contamination. Typically, the cost of remediation is distributed among the PRPs using an agreed upon formula such as the length of time an owner occupied the site. In order to reduce litigation among the PRPs, a generally accepted cost estimation method must be used to estimate the site remediation cost.

The Sarbanes-Oxley Act, another regulation that requires an estimate of hazardous waste site remediation cost, was signed into U.S. law on 30th July 2002. The Sarbanes-Oxley Act requires companies to disclose in their annual financial reports any environmental damage or uncertainties that have had or may have an adverse effect on revenues. The accuracy of a company's annual financial report is very important for several reasons. Stock analysts use the information to rate stocks to help investors make educated decisions on whether or not to buy or sell certain stocks. Stock holders also need accurate information to ensure the value of their stocks are not inflated to fool them into believing the company is doing well when the contrary is true.

The insurance industry periodically estimates the cost of potential environmental liabilities at their insured sites. Prior to the enactment of CERCLA in 1980, the insurance industry had sold general liability policies to several polluting industries. After the enactment of CERCLA, the insurance companies were suddenly saddled with huge future debts associated with soil and groundwater remediation at several insured sites. Insurance companies are interested in getting out of these liabilities through negotiated buyouts of the existing policies. In a buyout transaction, the insured presents a claim for a negotiated settlement in the form of a future cost estimate for the site.

In all the aforementioned cases, the problem with estimating the future cost of environmental remediation is that the PA and RI/FS reports have too many data uncertainties to permit a deterministic cost estimate. Through this study, I seek to validate an alternate cost estimation method that requires the minimal data normally found in the PA and RI/FS reports.

Introduction

The problem with deterministic estimates is the implicit assumption that all the required site data values are known with precision. This is not the case with the data found in PA and RI/FS soil boring reports. A typical soil-boring report shows that for any given area, the depths at which contaminants are found above cleanup levels vary considerably across the site. The area impacted by the hazardous chemicals is also not constant because the contaminants tend to spread out, as they are slowly washed downward by rainwater through the soil into the groundwater aquifer. It is worth pointing out here that in a soil excavation activity, the soil samples are routinely checked at some pre-determined depths to ensure that the excavation does not go below the depths necessary to remove all soils that are contaminated above cleanup levels. This minimizes the soil excavation, treatment, and disposal costs. These uncertainties lead to the conclusion that selecting a single value for area and/or depth in a deterministic estimate entails the high risk of either grossly underestimating or grossly overestimating remedial soil volumes. The difficulties with deterministic estimates have made it necessary to find a remedial cost estimation methodology that does not ignore the uncertainties in the data values.

The soil remediation cost for a hazardous waste site is directly proportional to the excavated soil volume. Therefore, soil volume would be used as a surrogate for the remediation costs, and vice-versa, in the ensuing discussions. The American Society for Testing Materials (ASTM) has published a standard (ASTM E-2137) to designate standardized methods for estimating costs under uncertainty. The top three methods in ASTM's order of preference are: the expected value, the most likely value, and the range of values.

The most common method for calculating the expected value is to use a decision tree analysis to calculate, essentially, the weighted average of the product of the cost of all possible events and their respective probabilities. However, decision trees are oversimplified since events are assumed to have the exact probabilities used in the estimates. Also, the assignment of probability values to the events are done arbitrarily, based on either subjective opinion or expert opinion. For this reason, the expected values derived from the decision tree methodology are not readily accepted at contested settlements.

In order to overcome the deficiency of the decision tree methodology, some analysts use Monte Carlo (MC) simulation. In a MC simulation, the exact probabilities are replaced with probability distributions to account for all possible values the variable may assume. The most common method used to select the probability distributions is to fit the available data and constraining parameters to several distributions and then select the one that gives the best fit.

Unfortunately, the expected values derived from a Monte Carlo simulation are also open to challenges because of the assumptions required to implement the MC procedure. For example, there may be insufficient data to justify the assignment of a particular probability distribution to a variable. There is also the question of imprecision in the measurements that yielded the exact data values used to derive the parameters of the assigned probability distribution.

The data variability and imprecision in MC modeling may be reduced by using two-dimensional MC modeling. This modeling involves nesting one MC simulation within another. Typically, the inner simulation is selected to account for the natural variability of the underlying physical or biological process being modeled. The outer simulation is chosen to account for imprecision about the input parameters. The whole process is just like running a complete Monte Carlo simulation within a Monte Carlo simulation at each of the iterations. Besides the uncertainty in the selection of the correct probability distributions for the variables, two-dimensional MC presents a very large computational burden and is not easily accomplished.

The other future cost estimation methods, the most likely value and the range of values, are useful only when there is sufficient site-specific data to calculate costs to a reasonable degree of accuracy. The imprecision and variability of the site data in the RI/FS reports make it necessary that the future cost estimates be made under uncertainty. Therefore, there is a need for another cost estimation method that minimizes the use of subjective probabilities to calculate remediation costs under uncertainties.

In recent times, several authors have proposed methods based on imprecise probabilities that may be used for decision-making under uncertainty. The term imprecise probability is a generic term for all mathematical models that measure chance or uncertainty without sharp numerical probabilities. The perceived advantages of using imprecise probabilities as compared to the decision tree methodology or Monte Carlo simulation are:

- Modeling assumptions and inferences are more apparent and credible.
- Sensitivity analysis is built into the model.
- Information from different sources, such as interval estimates and probabilities could be coherently combined.

An extensive literature review did not reveal any instance when an imprecise probability method has been used to estimate the remedial soil volume for a hazardous waste site. Dankwah (2009) investigated the feasibility of using the following imprecise probability methods to estimate the remedial soil volume for a hazardous waste site: Interval Arithmetic (IA), Dempster-Shafer Theory (DST), and Probability Bounds Analysis (PBA). Dankwah (2009) concluded that both IA and DST are interval methods. As such, they have the following deficiencies that make them unsuitable for estimating remedial soil volumes under uncertainty:

- Difficulty of transforming measurements into intervals.
- Do not use additional information about the site data such as mean, median, standard deviation, and percentiles.
- Method is heavily dependent on expert opinion for probability assignments.

Dankwah (2009) found that PBA is suitable for estimating remedial soil volumes at hazardous waste sites because it does not require the assumption of exact probabilities or probability distributions. Instead, estimates could be made with summary statistics data, such as mean, median, and standard deviation. Dankwah (2009) used PBA to estimate the remedial soil volume for the Dover Manufactured Gas Light site in Dover, Delaware, using data from the PA and RI/FS reports for the site.

PBA has several advantages over both MC and deterministic models. However, the scanty literature available on the use of the PBA models for decision-making suggests that the method is not being widely used. The reason could lie with the difficulty of interpreting PBA modeling results for decision-making.

Problem Statement

PBA uses a combination of interval methods and classical probability theory to generate the calculation results. For this reason, the result for any selected percentile or any measure of central tendency is not a unique number but is rather an interval with lower and upper bounds. Also, since PBA uses minimal data, the probability bounds are not very tight and this leads to a wide span between the lower and upper bounds of the results. The problem, then, was to choose whether the lower bound, upper bound or mid-point of a selected percentile or that of a measure of central tendency is the optimal decision point.

The second problem with PBA results is the selection of a percentile that would be readily acceptable to both decision-makers and the impacted community. In a typical

MC, the scientific community generally recognizes the 95th percentile value as an acceptable decision point that minimizes risk in the estimates. However, the 95th percentile values for PBA results tend to be extremely high as compared with the MC results. For example, Bergback, Oberg, & Sanders (2006) compared the results of an exposure assessment model for Cadmium using deterministic estimates, MC, and PBA. The results showed that the 95th percentile values for the PBA result were higher than that of the MC result by several orders of magnitude. Ferson, Hope, & Regan (2002) also showed that PBA produced the highest upper bounds at the 95th percentile level for a food-web exposure model when compared with the result for a two-dimensional MC model. Dankwah (2009) compared the PBA and the probabilistic estimates of the remedial soil volume for the former Dover Manufactured Gas Light site. The result was that the upper bound of the 95th percentile PBA result was almost twice the mean value determined by probabilistic modeling. However, the mean remedial soil volume estimated by the probabilistic modeling was within the lower and upper bounds of the mean of the PBA estimate. There was no established selection criterion to choose any one of the upper or lower bound of the 95th percentile values or that of the mean values as the optimal decision point. This confirms the problem of selecting a decision point from the wide array of PBA results.

For PBA to gain wide acceptance as a method to estimate remedial soil volumes, there must be an acceptable solution to the problem of choosing a credible decision point from the wide array of the PBA results. The research questions addressed in this study would stimulate interest in the use of PBA models to estimate remedial soil volumes at hazardous waste sites under uncertainty.

Purpose

The purpose of this study was to determine whether the mean or the 95^{th} percentile PBA estimate would be the optimal decision point for PBA models when used to estimate remedial soil volumes under uncertainty. Since the PBA result for both the mean value and the 95^{th} percentile value are interval values, the investigation analyzed their lower bounds, upper bounds, and midpoint values separately.

Research Questions

The following research questions were addressed in the study:

- Is the 95^{th} percentile volume an appropriate PBA decision point for the remedial soil volume estimates? Separate analysis was performed for the lower bound, upper bound, and midpoint of the 95^{th} percentile interval values.
- Is the mean volume an appropriate PBA decision point for the remedial soil volume estimates? Separate analysis was performed for the lower bound, upper bound, and midpoint of the mean interval values.

Significance of the Study

The alternate method that is used to estimate remedial soil volumes and remediation cost under uncertainty is MC. However, MC requires probability distributions, which are either assumed from experience or established through curve-fitting techniques from the available data. The MC cost estimates are not readily accepted at remediation cost settlements because of the uncertainties in the assumption of probability distributions or the imprecision in the data that was used to derive the parameters of the probability distributions.

The advantage of the PBA methodology is that remedial soil volume estimates could be made with summary statistics data such as mean, median, standard deviation, and percentiles. PBA does not require the assumption of exact probabilities or probability distributions but permits the use of exact probabilities or probability distributions if needed. However, the problem with PBA is that the modeling result for any percentile is not a unique number but it is rather an interval with a wide span between the maximum and the minimum values. The lack of a unique decision point for PBA models makes it more difficult to use PBA for decision-making under uncertainty. A good resolution of the research questions would encourage greater use of the PBA methodology to estimate remedial soil volumes under uncertainty at hazardous waste sites.

Method of Inquiry

This is a quantitative study. I used site-specific data in PBA models to estimate remedial soil volumes at seven Superfund sites where soil excavation had actually occurred. I performed linear regression analysis to identify any linear statistical association that existed between the upper or lower bounds of the 95th percentile values or that of the mean of the PBA-estimated soil volumes and the actual excavated soil volumes. For the two strongest statistical associations identified in the study, I performed two separate hypothesis tests for the existence of a population coefficient of correlation to confirm the associations were not by chance but indeed, correlation did exist between the PBA-estimated soil volumes and the actual excavated soil volumes.

Next, I made an MC estimate of the soil volume for each site from the same site-specific data. I performed a series of regression analyses using the mean and the 95th percentile MC estimates as the independent variables and the actual excavated soil volumes as the dependent variables. I performed another regression analysis using the U.S. EPA's deterministic soil volume estimates as the independent variables and the actual excavated soil volumes as the dependent variables. Finally, I compared the R^2-values of the best PBA estimates, MC estimates, and the deterministic estimates to determine which one of them had the best correlation to the actual excavated soil volumes.

Operational Definitions

I obtained the data for the study from the U.S. EPA's public-domain reports. The U.S. EPA generates these reports as part of the hazardous waste site remedial investigation process. The operational definitions of the titles of the U.S. EPA reports and other terminologies used in this study are as follows:

Superfund: The name of the fund established by the United States congress to address abandoned hazardous waste sites that pose imminent threat to human health or environment.

Potentially responsible party (PRP): Any individual or company who at any time contributed to a spill or other contamination at a Superfund site.

Remedial investigation (RI): An investigation intended to gather the necessary data to: (a) determine the nature and extent of contamination, (b) establish the cleanup criteria for the site, (c) identify alternative remedial actions, and, (d) support the technical and cost analysis of the remedial alternatives.

Feasibility study (FS): A study of a hazardous waste site intended to: (a) evaluate alternative remedial actions in order to select the most cost-effective remedial action, and, (b) prepare cost estimates for budgetary purposes.

Contaminants of concern (COC): Chemicals identified during the RI/FS that need to be addressed by a cleanup action because they pose a potential threat to human health or environment.

Record of decision (ROD): A public document that explains which cleanup alternative the U.S. EPA has selected for a Superfund site.

Explanation of Significant Differences (ESD): A public document that explains significant changes from the ROD.

Remedial design (RD): A design document detailing the technical specifications for the cleanup remedies and technologies.

Preliminary closeout report: A report prepared by the U.S. EPA remedial program manager to confirm that the physical construction of the remedy for a site is complete.

Five-year review: Reports required, following cleanup action, for sites where hazardous substances still remain onsite at high levels but below the cleanup levels. Five-year reviews are required to confirm the cleanup action remains protective of human health and the environment.

Limitations of the Study

The contaminated media at a hazardous waste site may include soil or groundwater contamination or both. This study is limited to the soil component of the remedial activity only. The groundwater contamination is very complex to model because the contaminants may come from multiple sources. For example, groundwater contaminants may migrate from an upstream facility to the groundwater underneath a downstream facility. All the sources that contribute to the groundwater contamination must be investigated thoroughly and included in the model.

Because of the need to select sites where soil excavation activities have been completed and where the actual excavated soil volumes are known with certainty, I did not select the sites randomly. This limits the generalization of the results to the general population of hazardous waste sites. However, the recommendations include suggestions for further studies and methodological enhancements to address this problem.

Social Change Implications

The adoption of the PBA methodology for estimating remedial soil volumes under uncertainty would lead to the following benefits:

- Reduce litigation expenses among the potentially responsible parties.
- Speed up remedial actions so contaminants do not continue to adversely impact the health and safety of nearby residents.
- Increase productivity in the area because nearby residents would be healthier and have less sick off-days.

Summary

The use of probability bounds analysis to estimate remedial soil volumes under uncertainty seems quite attractive because ordinary summary statistics such as mean, standard deviation, and percentiles could be used. PBA does not require the assumption of exact probabilities or probability distributions. However, the PBA estimate for any selected percentile or any measure of central tendency is not a unique number but is rather a wide interval. The problem, then, is to how to select an optimal decision point from the wide array of PBA estimates. In this study, I seek to address this problem. The other methods that could be used in place of PBA are MC and deterministic models.

However, both MC and deterministic models suffer from methodological problems. Chapter 2 presents a literature review that discusses the methodological problems associated with MC and deterministic models when compared to PBA estimates. The discussions and analyses in the chapter point to the advantages of PBA over both MC and deterministic estimates. Chapter 3 provides details of the analytical methods employed in the study. They are: probability bounds analysis, Monte Carlo simulation, regression analysis, and a hypothesis test for the existence of a population coefficient of correlation. I obtained the data for the study from the U.S. EPA Superfund Site Information Database. Chapter 4 presents the results of the PBA, MC, and deterministic models. Chapter 5 presents the interpretations, conclusions, and recommendations from the study.

Chapter 2: Literature Review

Introduction

I have discussed the following alternate remedial soil volume estimation methods in this literature review: Probability Bounds Analysis (PBA), Monte Carlo Simulation (MC), and deterministic modeling. Both PBA and MC are probabilistic models. The argument against the use of probabilistic models to make estimates in place of deterministic models is that probabilistic models inflate the outcome by including extreme data values with low probabilities in the estimates. The literature review in this chapter makes counter-arguments in favor of the use of probabilistic models to make estimates when uncertainties exist in the data. The literature review also shows that the PBA results consistently had the lowest lower bounds and the highest upper bounds when compared with both MC and deterministic estimates. This presents the problem of choosing a credible decision point for PBA results and I seek to provide an answer in this study.

Strategy Used for Searching Literature

I used the following sources for the literature search: WorldCat database and the Walden database. The WorldCat organization is a worldwide network of libraries. It makes it possible to search the catalogs of several libraries at once for an item and then locate it in a nearby library. For this study, I accessed the WorldCat database on the Internet at: http://www.worldcat.org/advancedsearch. The keywords I used for the search were: “probability bounds analysis”, “imprecise probabilities”, “deterministic probabilistic risk”, and “deterministic risk assessment”.

I also searched the following databases via the Walden database using the same key words: Academic Search Premier, Business Source Premier, Walden Dissertation Abstracts, and Sage Journals Online. I used the Walden document delivery system to obtain copies of all the literature related to this study.

.

Organization of the Literature Review

I have divided the literature review into two sections: literature supporting MC over deterministic modeling, and literature supporting PBA over MC modeling. Taken together, the literature review shows that PBA possesses several advantages over both MC and deterministic modeling.

Literature Supporting Monte Carlo Analysis over Deterministic Modeling

Mathematical models are generally used to represent physical systems. This enables the decision-maker to enter the parameters of the model and derive a result from which the decision-maker could make conclusions or decisions. Models may also be used to study cause and affect relationships between variables. The quality of the input data directly affects the reliability of the model result and thus deserves attention.

In quantitative risk assessment for environmental regulations, decision-makers prefer to use deterministic models to generate exact values. The argument against the use of deterministic models is that they ignore the imprecision and variability in the data. The imprecision in the data arises from the fact that measuring devices are not precise and have measurement errors. The variability in the data arises from natural randomness. Probabilistic models have been suggested as an alternative modeling approach that incorporates both imprecision and variability. Bergback & Oberg (2005) presented arguments to support the use of probabilistic models instead of deterministic models in quantitative risk assessment.

Bergback & Oberg (2005) argued that deterministic point estimate of risk ignores the variability and imprecision in the model parameters. In order to compensate for the uncertainty, risk managers use conservative values of the modeling parameters in the deterministic model. However, this compounding of worse case scenarios leads to an overestimate of risk. Also, a point estimate does not tell the percentage of the target population that would be exposed above the toxicity limits, or, the maximum possible exposure level. A probabilistic risk assessment is preferred because it accounts for both the variability and the imprecision in the exposure model. The output risk or exposure levels are presented as a probability distribution that could be analyzed further.

Bergback & Oberg (2005) supported the argument with a comparison of the results of the deterministic and probabilistic risk assessment for benzo[a]pyrene. The results of the probabilistic risk assessment indicated that the upper 95th percentile toxicological reference value for benzo[a]pyrene corresponded well with the deterministic reference value. The extra information provided by the MC result was that 5% of the target population would have an intake of benzo[a]pyrene that would be more than the toxicological reference value. This showed that probabilistic risk assessment could give similar results as point estimates with conservative parameter values. However the additional information provided by the MC could be used to quantify risk better and make better decisions.

Oberg & Sander (2006) also compared the results of deterministic and probabilistic risk assessment for an industrial site. The site chosen was a closed steel mill facility. Both deterministic and MC models were used to compute the exposure assessment for six elements in the site-specific study. Sixty- two soil samples had previously been analyzed and this provided data of the pollutant concentrations. The upper 95% confidence limit estimate of the pollutant concentrations and the other modeling parameters were used as the conservative values in the deterministic assessment. In the probabilistic assessment, the soil concentration for each constituent was represented by a lognormal distribution. The settings and probability distributions for the other parameters were selected from the USEPA Exposure Factors Handbook. The MC model was run for 10,000 iterations using the Crystal Ball software. The exposure assessment was done for both sensitive and less-sensitive land-use scenarios.

In the sensitive and less-sensitive land-use scenarios, the deterministic point estimates and the 95th percentile MC estimates for the various soil contaminants were approximately of the same order of magnitude. This case study also supported the conclusion of Bergback & Oberg (2005) that the probabilistic exposure estimates for a set of soil contaminants could be quite similar to deterministic estimates with conservative values. However, the MC modeling results histogram provided a database for further analysis of the modeling results.

Bergback, Oberg, & Sander (2006) continued the argument against the use of precise data in environmental models with another analysis to demonstrate how a change in the assumption of the probability distributions affected the results of a previously reported exposure assessment result for cadmium in soil. The investigation compared five modeling results based on five different sets of input values. The input values were: point estimates, point estimates with rounding errors, Monte Carlo (MC) with all probability distributions, probability boxes, and, probability boxes with rounding errors.

The results showed that the 5^{th} percentile value for the MC modeling result was of the same order of magnitude as both the point estimate, and, the point estimate with rounding errors result. However the 95^{th} percentile values for the MC, probability boxes, and the probability boxes with rounding errors differed by several orders of magnitude. The results for probability boxes had the widest range between the 5^{th} and 95^{th} percentiles, reflecting the fact that more of the uncertainties had been captured in the PBA analysis. This full disclosure of uncertainty may lead to better decisions.

Literature Supporting Probability Bounds Analysis over Monte Carlo Analysis

While MC models represent substantial improvement over the use of deterministic models, they are not free from uncertainties. There is imprecision in the exact values used to generate the probability distributions and there is uncertainty even in the choice of the probability distributions. For these reasons, Ferson, Regan, & Sample

(2002) argued that probability bounds analysis rather than MC should be used in environmental risk assessment.

Ferson , Regan, & Sample (2002) compared and contrasted the results of a wildlife ecological screening model for four scenarios. The scenarios used were: deterministic modeling using 90^{th} percentile values to represent conservative input values, deterministic modeling using median values, MC modeling, and, probability bounds analysis. The chemical constituents modeled were lead and DDT, and, the target organisms were meadow moles, and, short-tailed shrews. The wildlife screening levels were calculated by solving an U.S. EPA hazard quotient (HQ) model for the dose that gave an HQ value of one. The USEPA guidance document specifies that the calculations should be made using conservative values for the model parameters and should also be deterministic. The goal of the USEPA is that the screening levels, expressed as mg/kg of soil, should be as low as possible in order to protect the target organism.

Based on the 10^{th} percentile values, the PBA methodology consistently produced the lowest screening level estimates irrespective of the target chemical or organism. The highest screening levels were obtained for the deterministic estimates using median values and were up to 1000 times more than the 10^{th} percentile PBA estimates. The screening levels for the MC analysis, based on the 10^{th} percentile values, were the second highest. Additionally, the MC result had a problem where some of the results implied an HQ that was greater than one, which made them invalid solutions to the HQ model. In contrast, the PBA did not produce an HQ greater than one in any scenario. This showed that the PBA methodology produced screening levels that were more protective of ecological receptors than the current method of compounding conservative estimates in deterministic models.

One method to minimize the deficiencies of one-dimensional MC modeling is to use 2-dimensional Monte Carlo (2D-MC) modeling. Two-dimensional MC modeling involves nesting one MC simulation within another. Typically, the inner simulation is selected to account for the natural variability of the underlying physical or biological process being modeled. The outer simulation is chosen to account for imprecision about the input parameters. The whole process is just like running a complete Monte Carlo simulation within a Monte Carlo simulation at each of the iterations.

Ferson , Hope, & Regan (2002) compared and contrasted the results of two-dimensional MC and PBA models using existing data from a previously published probabilistic food-web exposure model. The original model was for a soil heavily contaminated with PCB and Aroclor-1254. The exposure to environmental receptors was previously modeled with a one-dimensional MC model. For the 2-D MC model, five parameters that were deemed most sensitive were modeled with inner and outer loops. The model simulations were performed with Crystal Ball software with 200 outer loop iterations to account for imprecision in the data. The inner loop consisted of 1000 iterations to account for variability. For the PBA, additional information such as means, standard deviations, upper and lower bounds, etc., were extracted from literature sources. The data was then used in the RAMAS RiskCalc software to generate p-boxes for the PBA analysis.

The output of the models was the toxicity reference value (TRV). The TRV for this study was defined as the lowest dose at which specified types of adverse effects occurred in mink. The 2-D MC result showed that there was a 95% chance that 95% of the mink population would exceed the maximum TRV. The PBA result showed there was a 95% chance that 100% of the mink population would exceed the maximum TRV. The PBA results had the lowest minimum bound and the highest upper bound. The maximum value of the PBA result exceeded the TRV by 12 orders of magnitude. The PBA methodology gave a much wider range of results than the 2D-MC because it captured the large number of uncertain parameters. In this study, there were over 40 uncertain parameters.

The PBA methodology finds application in other fields besides environmental risk assessments. Engineering design decisions and estimations are usually made under uncertainty with deterministic values. Without quantifying the extent of the uncertainty, the common practice to account for uncertainty is to add a safety factor, say 15% more, to the estimate. Another practice is to replace some of the point estimates with exact probability distributions in a MC analysis of uncertainty. Aughenbaugh & Paredis (2005) argued that assigning a random safety factor is inadequate and MC simulation does not fully account for imprecision and variability in the parameter estimates. Therefore, it is better to use PBA to capture the full extent of the uncertainty.

Aughenbaugh & Paredis (2005) illustrated the solution to the problem with a hypothetical design of a pressure vessel that was to contain air at a designated pressure. The vessel was to be used in a human occupied location and therefore the cost of failure or explosion weighed heavily in the estimate. The dimensions and the steel wall thickness of the vessel had to be selected in such a way as to minimize cost and also minimize the probability of failure. The criterion used to compare the pressure vessel designs was the utility of competing designs. Utility theory states that one design is preferred to another if, on average, it has a higher expected utility. The assumption was made that the steel material was new and that there were 50 yield stress measurements available. Instead of using all the 50 yield stress measurements to generate a probability distribution for the MC and a p-box for the PBA, various sample sizes were used to calculate the utilities for different samples sizes in order to compare the MC and PBA results for the different sample sizes. The selected data was fitted to a normal distribution for the MC analysis. For the PBA, the data was used to generate a p-box without making any assumptions about the underlying distribution.

Aughenbaugh & Paredis (2005) assumed an optimal design under precise information for the purpose of comparing the MC and PBA results. The 95th percentile value of the PBA result had a better utility than the 95th percentile MC result when 30 yield strength samples were used for the analysis. For smaller sample sizes than 30, the PBA consistently yielded better utility values. As the sample size approached 50, both MC and PBA yielded similar results. However, for sample size of 50 and possibly more than 50, the MC yielded better results. This result is similar to the result presented in Ferson, Hope, and Regan (2002). The PBA methodology gave better results than MC for small sample sizes because the imprecision was lower. However, for larger sample sizes, the MC gave better results because the imprecision was higher.

Because the PBA methodology is more efficient at capturing uncertainties than 2D-MC, it has been proposed as a method for conducting sensitivity analysis of MC results. Ferson & Troy (2006) argued that the PBA approach is a straightforward method for conducting sensitivity analysis of probabilistic models because it does not make unfounded assumptions about probability distributions. The PBA approach uses the available constraining data to derive bounds on the possible distributions in the form of a p-box. The p-box is distribution-free because it does not make any assumptions about the type of distributions that fit the constraining data. The PBA model then projects the uncertainty throughout the model to yield bounds on the model result.

Aughenbaugh and Paredis (2007) presented a similar argument to support the use of the PBA methodology for sensitivity analysis, especially in engineering designs. Aughenbaugh & Paredis (2007) argued that probability bounds analysis (PBA) is a better method for conducting sensitivity analysis because the PBA methodology preserves both the interval and probabilistic forms of uncertainty. The methodology does not make assumptions about distributions. The p-boxes are constructed only with the available constraining information. Also, the dependency bounds algorithm used to implement the PBA methodology uses only the bounds in the calculations and therefore includes all the family of distributions that meet the constraining data.

The extensive literature search conducted by this author did not reveal any existing study where PBA has been used to estimate the remedial soil volume at a hazardous waste facility. As a prelude to this study, this author used both PBA and probabilistic analysis to estimate the remedial soil volume at the Dover Gas Light site in Dover, Delaware (Dankwah, 2009). The estimated mean soil volume for PBA with summary statistics data was between 12219 and 12937 cubic yards. The expected soil volume for the probabilistic analysis was 12644 cubic yards, which fell within the range of the PBA mean results. On the other hand, the 95th percentile soil volume estimate for PBA with summary statistics data was between 20619 and 24917 cubic yards. This was almost two times the mean for the PBA with summary statistics data. The research problem, then, is to determine which bound of the PBA mean or the 95th percentile value represents the optimal decision point.

Summary

I have presented arguments and supporting studies in this literature review to show that PBA is a very versatile methodology that has found applications in environmental risk assessment, sensitivity analysis of MC results, engineering designs, and cost analysis. In the context of cost analysis, the remedial soil volume could be used as a surrogate for remediation costs since the two are directly related. The literature review also shows that PBA possesses several advantages over both MC and deterministic modeling.

I have provided details of the analytical methods employed for the study in Chapter 3. They are: probability bounds analysis, Monte Carlo simulation, regression analysis, and a hypothesis test for the existence of a population coefficient of correlation. I obtained the data for the study from the U.S. EPA Superfund Site Information Database.

Chapter 4 presents the results of the PBA, MC, and deterministic models.

Chapter 5 presents the interpretations, conclusions, and recommendations from the study.

Chapter 3: Research Method

Introduction

I employed the following research methods in the studies: probability bounds analysis (PBA), Monte Carlo (MC) modeling, linear regression analysis, and a hypothesis test for the existence of a population coefficient of correlation. Linear regression analysis and MC are well-established statistical methods and therefore do not need further introduction. I have briefly discussed the PBA method and the procedure for the test of hypothesis for the existence of a population coefficient of correlation in the subsequent paragraphs. This is followed by details of the study design, sampling population, data sources, data collection methods, and presentation of the data.

Probability Bounds Analysis

The development of the PBA methodology originated from the probability bounding methods that have been accumulated in the course of the history of probability theory. For example, Chebychev’s inequalities provide a method to place an upper bound on a probability distribution when only the mean and variance of the distribution are known. Similarly, Markov’s inequality provides a method for finding the bounds on the probability that a positive variable is greater than or equal to a certain positive number. However, the idea of probability bounds analysis did not emerge until Williams and Downs (1990) presented numerical methods for computing the results of the bounds on a given set of input distributions after the mathematical operations of addition, subtraction, multiplication, and division of the random variables. The numerical methods are referred to as the "dependency bounds and convolutions" (DBC). Scott Ferson (2002) later extended the dependency bounds algorithms developed by Williams and Downs (1990) to cover mathematical transformations such as: logarithms, square roots, and convolutions such as maximum, minimum, and powers. This was implemented in a commercial software called RiskCalc (Ferson, 2002).

Probability bounds analysis is a discrete analytical method that requires minimal data, which may or may not include probability distributions. The input data may be expressed as an interval or as an imprecisely known probability distribution. For example, the upper and lower bounds on an imprecisely known normal distribution with mean between μ_1 and μ_2, and standard deviation between σ_1 and σ_2, could be determined from the envelope formed by the normal distribution functions corresponding to each parameter set:

$$(\mu_1,\sigma_1),\ (\mu_1,\sigma_2),\ (\mu_2,\sigma_1),\ (\mu_2,\sigma_2)$$

I determined the parameter set by interval arithmetic operations. I have illustrated this distribution with numeric data in Figure 1. The resulting graph is called a probability box or a p-box. A p-box is not a single probability distribution but rather represents all the classes of distributions that fit within the box. Any particular individual member from the class can be thought of as representing variability in a Monte Carlo model. The class as a whole represents the associated imprecision in the parameter estimates. A p-box enables the model to propagate both variability and imprecision through the calculations simultaneously. The p-box yields a lot of information about the family of

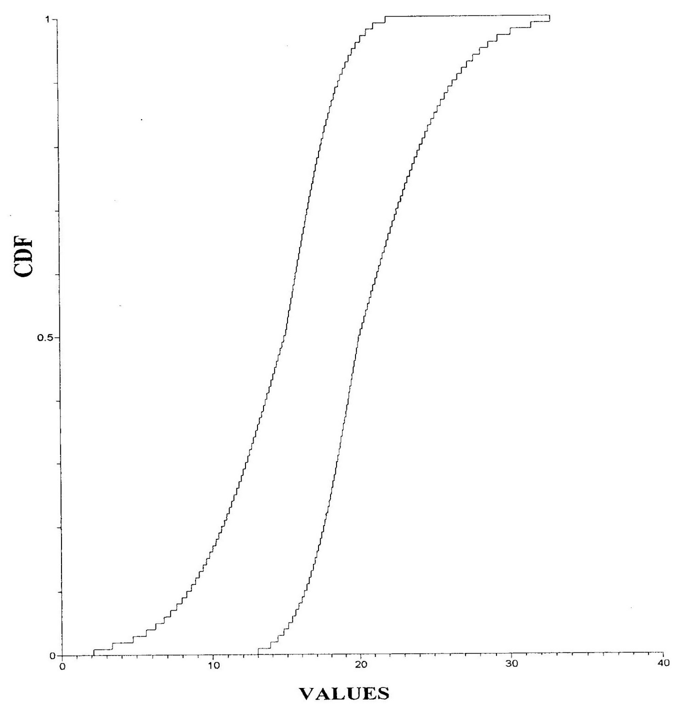

Figure 1. P-Box for an imprecisely known Normal Distribution bounded by N(15, 3) and N(20, 5).

distributions. For example, a horizontal slice at the median (CDF = 0.5) gives the interval bounds on the median as [12, 17].

The advantage of PBA over MC is that a distribution-free analysis could be performed with summary statistics data such as mean, median, mode, standard deviation, and percentiles. The RiskCalc software contains transformation algorithms that use these constraining data to generate p-boxes for further analysis. I have illustrated this with a numeric example in Figure 2 for the minimum, maximum, and mean values.

RiskCalc is a calculator. The PBA calculations proceed first by using the constraining data to generate the p-boxes, discretizing the bounds of the p-boxes and then, following the modeling equations, using the dependency bounds convolution algorithms to calculate the distribution of the modeling results. The program allows the user to specify a desired output such as the 95th percentile value or a measure of central tendency such as the mean or median.

Test of Hypothesis for a Population Coefficient of Correlation

The r-value from the regression analysis is the sample coefficient of correlation. A high r-value implies a high degree of association between the variables. However, because the r-value is a sample variable and would be different for a different set of seven sites, a test of hypothesis for the existence of a population coefficient of correlation is needed to confirm if any association that is found is not by chance but indeed, correlation does exist between the PBA-estimated and the actual excavated soil volumes. The test of

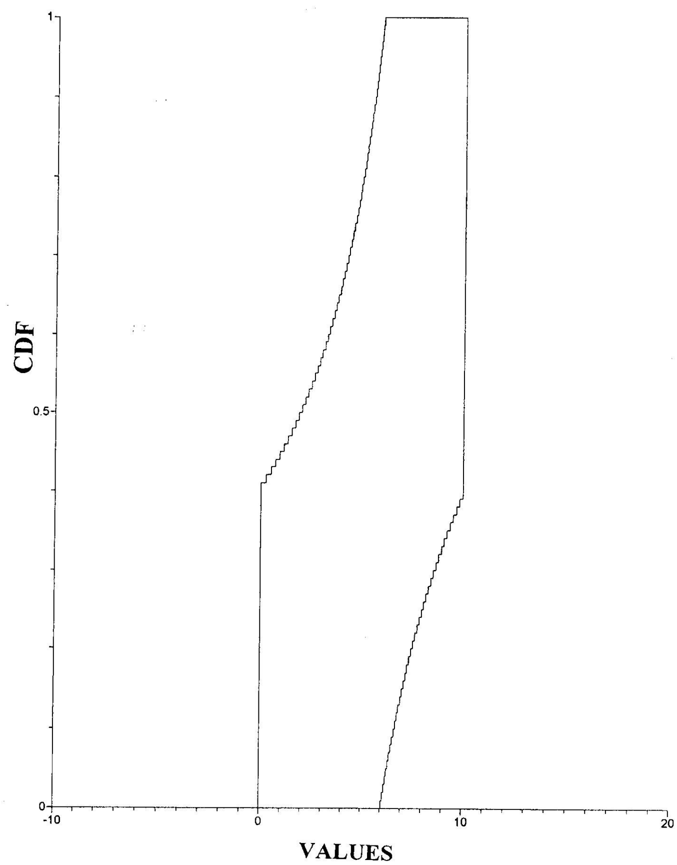

Figure 2. P-Box for a summary statistics with minimum=0, mean = 6, and, maximum = 10

hypothesis for the existence of a population coefficient of correlation follows the following procedure in Guzelian (1979, p511-515):

"Let ρ = population coefficient of correlation

The hypothesis to be tested is:

$H_0 : \rho = 0$ (Null hypothesis, no correlation,)

$H_1 : \rho \neq 0$ (Alternative hypothesis, correlation exists,)

Use the r and r^2 values from the regression analysis to compute the t-statistic

$t = r\sqrt{(n-2)/(1-r^2)}$ where, n = number of samples

Based on a 5% significance level and (n-2) degrees of freedom, use t-distribution tables to obtain the critical values corresponding to:

$CV_1 = -t_{.025}$

$CV_2 = t_{.025}$

The following will be the decision rule for the test of correlation of ρ:

Accept H_0 when $CV_1 < t < CV_2$

Reject H_0 when $t \leq CV_1$ or $t \geq CV_2$"

Design of the Study

This is a quantitative study. I employed the following research methods in the study: probability bounds analysis (PBA), Monte Carlo (MC) analysis, linear regression analysis, and a hypothesis test for the existence of a population coefficient of correlation. I used the commercial PBA software, RiskCalc Version 4.0, in a distribution-free analysis using two different sets of distribution-free data for the depths as the constraining variables to generate p-boxes to calculate the soil volume estimates for seven Superfund sites. In the first set of the PBA soil volume estimates, I used the following constraining parameters to represent the excavation depths: minimum depth, average depth, maximum depth, and the standard deviation. In the second set of the PBA soil volume estimates, I used the following constraining parameters to represent the excavation depths: minimum depth, 25th percentile depth, 50th percentile (median) depth, 75th percentile depth, and the maximum depth. The purpose of using the two different sets of constraining data to generate two different sets of PBA estimates was to determine which of the two data sets would lead to a better correlation between the PBA estimated soil volumes and the actual excavated soil volumes.

I used the Excel spreadsheet software to perform the linear regression analysis between the actual excavated soil volumes as the dependent variables and the PBA-estimated soil volumes as the independent variables. I used the R^2-values as a measure of the degree of statistical association between the variables. A high R^2-value implied a high degree of statistical association. For the two best statistical associations identified, I performed a test of hypothesis to confirm the existence of a population coefficient of correlation.

Next, I performed an MC estimate of the excavated soil volumes for the seven sites using the Oracle Crystal Ball (OCB) Software, Release 11.1.1.3.0. OCB software is an Excel add-in that enables simulations to be performed in spreadsheet calculations. Each simulation run consisted of 1000 iterations. I performed a second set of regression analysis using the mean and the 95th percentile MC estimates, respectively, as the independent variables and the actual excavated soil volumes as the dependent variables.

I performed a third set of regression analysis using the U.S. EPA's deterministic soil volume estimates as the independent variables and the actual excavated soil volumes as the dependent variables. Finally, I compared the R^2-values of the best PBA estimates, MC estimates, and the U.S. EPA deterministic estimate to determine which of them had the best correlation to the actual excavated soil volumes.

Sample and Population

The U.S. EPA's Superfund Site Information Database (SSIB) contains data on all the abandoned hazardous waste sites in the USA where the U.S. EPA has determined that there is an uncontrolled release of hazardous chemicals to the soil or groundwater and this poses an immediate danger to the health and safety of nearby communities. The sites listed in the SSIB are called "Superfund Sites" and they were the universe population for the study.

Sample Selection Procedure

I could not select the study sites randomly from the U.S. EPA universe of hazardous waste sites due to the following reasons:

- The sites in the SSIB were at different stages in the remediation process. Some sites had been cleaned up completely while others were yet to commence the cleanup activities. Therefore, it was likely that a random sample would include sites where there is no data or where there is incomplete data.
- The sites listed in the SSIB did not all require soil remediation. Some sites were listed only because of groundwater contamination. Therefore, it was likely that a random sample would include sites where there is no soil data.
- Even at sites where soil remediation was needed, soil excavation was not always part of the remedy. For example, in-situ technologies could be employed to remove highly volatile organic compounds from soils without excavation. In the cases where the threat of soil contaminants to groundwater was not severe, the U.S. EPA only recommended that an asphalt cap be placed on top of the soil to prevent further ingress of rainwater to carry contaminants further downward into the groundwater aquifer.

Because of the reasons stated above, the sample population for this study was reduced to the subset of the SSIB universe where soil excavation had been completed and where the actual excavated soil volume was known with certainty. This condition was imposed by the data needs for the regression analysis.

Number of Samples

I found from the U.S. EPA's database that only U.S. EPA Region 3 covering Pennsylvania, Delaware, Maryland, Virginia, and West Virginia had complete site reports on the website. Of the sites I reviewed, only five sites met the criteria for inclusion in the study. U.S EPA Region 7 and U.S. EPA Region 10 had limited reports for a few sites on their website. However, only one site from U.S. EPA Region 7 and one site from U.S. EPA Region 10 met all the conditions for inclusion in the study. For these reasons, I selected the seven sites with readily available data for the study.

Seven sites are adequate for this study because the study seeks to explore a linear statistical association between the PBA-estimated soil volumes and the actual excavated soil volumes. While three data points are enough to define a linear relationship or straight line, I used seven sites to help define the straight line better and also provide enough data for the hypothesis test for the existence of a population coefficient of correlation.

I could have requested additional site data from the U.S. EPA under the Freedom of Information Act (FOIA). Unfortunately, the U.S. EPA views such requests for data on 'closed-cases' with suspicion and requests could take several months to be fulfilled with no guarantee that the site reports would have adequate data for inclusion in the study.

Data Collection

Introduction

The United States Congress provides funds to the U. S. EPA annually for the cleanup of the Superfund sites. The U.S. EPA is, however, authorized to use legal processes to recover the cleanup costs from the past and present owners and non-owners of the site who's past or present activities have contributed to the soil and/or groundwater contamination. Since the U.S. EPA uses public funds for the cleanup activities at the Superfund sites, all reports generated in the course of the site investigations and remedial activities are public records and can be accessed directly from the SSIB at: http://www.epa.gov/superfund/cleanup/index.htm. I used the search capacity of the SSIB to select only sites where the soil remediation has been completed. Next, I used the Closeout Reports for the sites to identify the sites where soil excavation was part of the remedy and also where the actual excavated soil volume was reported in cubic yards. In the majority of the Closeout Reports, the excavated soil volume was reported in tons with no further information on the soil density to permit the conversion from tons to cubic yards. The models in the study calculate soil volumes and not weights. Therefore, those sites could not be included in the study because of the lack of this critical data. In the end, seven sites met all the data requirements for the study and all such data were available in the public domain and freely accessible on the SSIB.

The next sections on data collection address the data collection methods and the assumptions made for the PBA models and the MC models separately.

Data Collection for PBA Models

The data required for the PBA models to estimate the soil excavation volume at each site are the expected excavation depths at each site, and the expected contaminated area at each site. These data were not available in ready-to-use forms in the U.S. EPA site study reports. Therefore, I made assumptions in order to translate the available data into a form that could be used for the PBA models. The procedures I used to obtain the data for the PBA models are discussed separately for the expected excavation depths and the expected excavation areas.

Assumptions for PBA excavation depths.

The soil-boring data from the RI/FS reports show, at each soil-boring location, the concentration of the contaminant of concern (COC) at several depths below ground level. The mere presence of a COC at a soil-boring location does not automatically mean that the area must be excavated. The U.S. EPA rules demand soil excavation or soil treatment only when the COC concentration exceeds the risk-based site-specific cleanup level that has been established by the U.S. EPA for that particular site. The U.S EPA establishes the site-specific cleanup levels from risk analysis using the RI/FS data on the hydrogeology of the site, the nature of the COC, and the potential threats to both the ecology and groundwater resources. The concentrations of a COC at the various soil-boring depths when compared with the site-specific cleanup level give an indication of the expected depths to which each soil- boring location area must be excavated in order to reduce the COC concentration remaining in the soil to a level below the site-specific cleanup level.

I made the following assumptions in order to translate the available RI/FS soil-boring chemical analysis data into the raw data of expected excavation depths:

- If the COC concentration does not exceed the site-specific cleanup level at all the depths at which the COC concentration was measured in any single bore location then the expected excavation depth at that bore location is zero. In other words, no soil excavation is required at such a location.
- If the COC concentration exceeds the site-specific cleanup level at any one of the depths at which the COC concentration was measured in a bore location, then the expected excavation depth is the next lower depth at which the measured COC concentration does not exceed the site-specific cleanup level. For example, if a COC was found above the site-specific cleanup level at a depth of 5 feet below grade but not at the next lower depth of 6 feet below grade, then I assumed the expected excavation depth for that soil-boring location is 6 feet below grade.
- The site reports indicate the U.S. EPA did not select the soil-boring locations statistically. Instead, the U.S. EPA selected the soil-boring locations according to the past site uses. The areas most likely to have a COC received the most soil bores. Even though this may bias the results towards higher soil volumes, I assumed that the soil-boring locations were acceptable for this study.

Table 1 shows the sites I selected from the U.S. EPA Superfund Database for this study, a brief description of the previous onsite activities, and the soil conditions before the cleanup activities began. Table 2 presents a summary of the references for the U.S. EPA documents from which I obtained the soil-boring data for each site in the study. Table 3 shows the COC at each site, the site-specific cleanup level for the COC at each site, and the corresponding reference for the source of that information.

Table 1

Selected Sites and Description of Past Site Uses

Site name and location	Site description and soil conditions
Aladdin Plating Company, Lackawanna County, PA (1947 to 1982)	The site was used for chromium electroplating. As a result of the waste-handling activities, the site soil was contaminated with chromium from the electroplating activities.
C&R Battery Company, Chesterfield County, VA (1973 to 1985)	The company operated a battery breaker for the purpose of separating and recovering lead from discarded automobile and truck batteries. The site soil was contaminated with lead from battery breaking and leads recovery operations.
Paoli Rail Yard, Paoli, PA (1915 to present)	The site was used to repair electric-powered rail cars. In the 1950's, polychlorinated biphenyls (PCBs) were used to cool the transformers in the train cars. The site soil was contaminated with polychlorinated biphenyls (PCBs) from the waste handling activities.
Peoples Natural Gas Company, Dubuque, IA (1930's to 1954)	The site was the location of a former manufactured gas plant. The byproducts of the operation were coal tar and cyanide-bearing wood chips. These were buried onsite. As a result, the site soil was contaminated with polynuclear aromatic hydrocarbons (PAHs).
Taylor Lumber and Treating Company, Sheridan, OR (1946 to 2001)	The company operated a sawmill and wood treating operations onsite. The site soil was contaminated with arsenic from the chemicals used for the wood treating operations.
Tonolli Corporation, Nesquehoning County, PA (1974 to 1986)	The corporation operated a lead-acid battery recycling and secondary lead reclamation facility at the site. The site soil was contaminated with lead from the smelter operations.
U.S. Titanium Company, Nelson County, VA (1931 to 1971)	The company produced titanium dioxide pigment at the site. The byproduct from the operation was hydrated ferrous sulfate, also called "copperas". The copperas was buried onsite at several locations. This was the source of acidic waters that contaminated nearby rivers

Table 2

Soil Boring Data Source References

Site name and location	Source(s) of soil boring data
Aladdin Plating Company, Lackawanna County, PA (1947 to 1982)	Roy F. Weston (1988a, p AR300324 – AR300325)
C&R Battery Company, Chesterfield County, VA (1973 to 1985)	NUS Corporation (1990a, p AR302030)
Paoli Rail Yard, Paoli, PA (1915 to present)	Groundwater Technology Inc. (1990, p AR300721 – AR300739)
Peoples Natural Gas Company, Dubuque, IA (1930's to 1954)	Barr Engineering Company (1994, Table 2-3)
Taylor Lumber and Treating Company, Sheridan, OR (1946 to 2001)	CH2MHILL (2004, Table A-1) CH2MHILL (2006, Figure A-1)
Tonolli Corporation, Nesquehoning County, PA (1974 to 1986)	Paul C. Rizzo Associates Inc. (1991, p AR301613 – AR301650)
U.S. Titanium Company, Nelson County, VA (1931 to 1971)	Hydrosystems Inc. (1987b, p 301345 - 301349)

Table 3

Site-Specific Soil Risk-Based Cleanup Standards and References

Site name and location	Risk-based cleanup standards
Aladdin Plating Company, Lackawanna County, PA (1947 to 1982)	Excavate all soils containing more than 50 mg/kg chromium Ref: U.S. EPA (1998)
C&R Battery Company, Chesterfield County, VA (1973 to 1985)	Excavate all soils containing more than 1000 mg/kg lead. Ref: U.S. EPA (1990)
Paoli Rail Yard, Paoli, PA (1915 to present)	Excavate all soils containing more than 25 mg/kg PCBs Ref: U.S. EPA (1992a)
Peoples Natural Gas Company, Dubuque, IA (1930's to 1954)	Excavate all soils containing more than 100 mg/kg carcinogenic PAHs or 500 mg/kg total PAHs up to 6 ft. Excavate up to upper confining unit if oil sheen is visible after 6 ft. Ref: U.S. EPA (1991)
Taylor Lumber and Treating Company, Sheridan, OR (1946 to 2001)	Excavate all soils containing more than 159 mg/kg arsenic in the West Facility area. Ref: U.S. EPA (2005)
Tonolli Corporation, Nesquehoning County, PA (1974 to 1986)	Excavate all soils containing more than 1000 mg/kg lead Ref: U.S. EPA (1992b)
U.S. Titanium Company, Nelson County, VA (1931 to 1971)	Excavate copperas and all visibly impacted soils in Area 1. Ref: U.S. EPA (1989)

I used the site-specific cleanup levels of the COCs in Table 3, and the assumptions for the excavation depths, as decision rules to translate the soil-boring results from the sources listed in Table 2 into a raw data of expected excavation depths for each site. Tables 4 to 10 present the raw data of expected excavation depths at each site and the corresponding summary statistics.

With the exception of the Paoli Rail Yard site (Table 6), all the sites in the study had one contiguous area to be excavated. As a result, there is only one summary statistics data on the expected excavation depths for those sites. The soil investigation at the Paoli site covered four separate areas. Two areas were investigated with soil bores at several depths. Table 6 presents the raw data on the expected excavation depths and the summary statistics for those two areas. At the other two remaining areas, only surface soils at six inches below grade were analyzed for COCs. A COC was found above the site-specific cleanup level but there was no further soil-boring data at depths below six inches. Therefore, I assumed the expected excavation depths for those two areas were interval values that range from zero inches to one foot below grade.

In the case of Peoples Natural Gas, the ROD limited the excavation depth to six feet below grade. However, if oil sheen was visible after the first six feet of excavation then excavation should continue down to the upper confining unit which is between six feet and 10 feet below ground level. For this reason, the default maximum excavation depth for the Peoples Natural Gas site was 10 feet. This is shown in Table 7.

Table 4

Excavation Depth Data for the Aladdin Plating Company Site

Soil bore location	Expected excavation depth (ft)
H-1	2.5
H-2	2.5
H-3	2.5
G-1	2.5
G-2	2.5
G-3	2.5
L-0	1.5
M-0	2.5
1	3.0
2	3.0
3	2.0
J-1	2.0
N-0	2.5
I-1	3.0
I-2	2.5
I-3	2.5
K-1	3.0
M-1	3.0
A-0	6.0
I-0	6.0
J-0	6.0

Summary statistics for excavation depth (ft)

Mean	3.0238
Standard Error	0.2834
Median	2.5
Mode	2.5
Standard Deviation	1.2988
Sample Variance	1.6869
25^{th} Percentile	2.5
75^{th} Percentile	3.0
Range	4.5
Minimum	1.5
Maximum	6

Table 5

Excavation Depth Data for the C&R Battery Company Site

Soil bore location	Expected excavation depth (ft)	Soil bore location	Expected excavation depth (ft)
SO-01	6	SO-31	0
SO-02	3	SO-15	9
S0-05	6	SO-16	9
SO-06	3	SO-17	3
SO-07	3	SO-18	9
SO-08	9	SO-20	0
SO-09	3	SO-21	0
SO-10	9	SO-22	6
SO-11	9	SO-23	0
SO-12	6	20-24	0
SO-13	9	S0-25	6
SO-14	6	SO-28	6
SO-19	3	SO-32	0
SO-28	6	SO-37	0

Summary statistics for excavation depth (ft)	
Mean	4.6071
Standard Error	0.6452
Median	6
Mode	6
Standard Deviation	3.4139
Sample Variance	11.655
25th Percentile	1.5
75th Percentile	7.5
Range	9
Minimum	0
Maximum	9

Table 6

Excavation Depth Data for the Paoli Rail Yard Site

A. Throat & East Car Shop Area

Soil bore location	Expected excavation depth (ft)
2-3	0
4-5	8
3-4	4
1-2	2
B-west	4.5
B-N.E.	0
B-SW	0
C-SE	0
C-SW	0
Pile-1	16.5

B. South & West Car Shop Area

Soil bore location	Expected excavation depth (ft)
A-North	4.5
A-south	0
East	30.5
SB-1	2.5
SB-2	14.5
SB-3	14.5
West	21.5

Summary statistics for excavation depth Throat & East Car Shop Areas (ft)

Mean	3.5
Standard Error	1.68
Median	1
Mode	0
Standard Deviation	5.3125
Sample Variance	28.222
25th Percentile	0
75th Percentile	4.5
Range	16.5
Minimum	0
Maximum	16.5

Summary statistics for excavation depth South and West Car Shop (ft)

Mean	12.571
Standard Error	4.1782
Median	14.5
Mode	14.5
Standard Deviation	11.054
Sample Variance	122.20
25th Percentile	2.5
75th Percentile	21.5
Range	30.5
Minimum	0
Maximum	30.5

C. Expected excavation depth for East Storage Yard (ft) = [0, 1]

D. Expected excavation depth for Turnaround Track (ft) = [0, 1]

Table 7

Excavation Depth Data for the Peoples Natural Gas Site

Soil bore location	Expected excavation depth (ft)
SB-101	0
SB-102	2
SB-103	10
SB-104	10
SB-105	0
SB-106	0
SB-106A	6
SB-107A	0
SB-108	0
SB-109	0
SB-110	10
SB-111	0
SB-112	0
SB-113	0
SB-113A	0
SB-113X	0
SB-116	0

Summary statistics for excavation depth (ft)

Mean	2.2353
Standard Error	0.9684
Median	0
Mode	0
Standard Deviation	3.9926
Sample Variance	15.94
25th Percentile	0
75th Percentile	2
Range	10
Minimum	0
Maximum	10

Table 8

Excavation Depth Data for Taylor Lumber & Treating Site

Soil bore location	Expected excavation depth (ft)
OS-12	2
OS-13	2
OS-14	2
GP-01	5
PS-01	6
WF-4	0
WF-05	0
WF-06	0
WF-07	2
WF-09	0
WF-12	2

Summary statistics for excavation depth (ft)

Mean	1.9091
Standard Error	0.6098
Median	2
Mode	2
Standard Deviation	2.0226
Sample Variance	4.0909
25th Percentile	0
75th Percentile	2
Range	6
Minimum	0
Maximum	6

Table 9

Excavation Depth Data for the Tonolli Corporation Site

Soil bore location	Expected excavation depth (ft)	Soil bore location	Expected excavation depth (ft)
SO-S1	0	SO-S30	0
SO-S2	0	SO-S31	0
SO-S3	0	SO-S32	0
SO-S4	5	SO-S33	0
SO-S5	0	SO-S34	0
SO-S6	0	SO-S35	0
SO-S7	0	SO-S36	0
SO-S8	5.5	SO-S37	0
SO-S9	0	SO-S38	0
SO-S10	3	SO-S39	0
SO-S11	5	SO-S40	10
SO-S12	0	SO-S41	0
SO-S13	0	SO-S42	0
SO-S14	10	SO-S43	0
SO-S15	9	SO-S44	0
SO-S16	0	SO-S50	1.5
SO-S17	5	SO-S51	0
SO-S18	8	SO-S52	0
SO-S19	0	SO-S53	0
SO-S20	10	SO-S54	2
SO-S21	10	SO-S55	0
SO-S22	0	SO-S56	0
SO-S24	0	SO-S57	0
SO-S25	0	SO-S58	3
SO-S26	5	SO-S59	6
SO-S28	5	SO-S60	5
SO-S29	10	SO-S61	5
		SO-S62	5

Summary statistics for excavation depth (ft)		Summary statistics for excavation depth (ft)	
Mean	2.3273	25th Percentile	0
Standard Error	0.4633	75th Percentile	5
Median	0	Range	10
Mode	0	Minimum	0
Standard Deviation	3.4362	Maximum	10
Sample Variance	11.808		

Table 10

Excavation Depth Data for the U. S. Titanium Corporation Site

Soil bore location	Expected excavation depth (ft)
3-86	0
4'86	13
5-86	0
6-86	9.5
7-86	0
8-86	0
9-86	12
10-86	15.5
11-86	0
12-86	7
13-86	0
14-86	23
15-86	6
16-86	0
17-86	13.5
18-86	13
19-86	20
20-86	24

Summary statistics for excavation depth (ft)

Mean	8.6944
Standard Error	2.0000
Median	8.25
Mode	0
Standard Deviation	8.4855
Sample Variance	72.004
25th Percentile	0
75th Percentile	8.25
Range	24
Minimum	0
Maximum	24

Assumptions for PBA excavation areas.

For some of the sites, the U.S. EPA Remedial Design Reports contained data on the expected excavation areas for the sites. The U.S. EPA estimated those areas after further site delineation studies or by computer programs designed for that purpose. When provided, I assumed the U.S. EPA reported excavation areas to be the minimum expected excavation areas in the PBA models. I estimated the other areas that were not available in the U.S. EPA reports from available scaled drawings that showed both the soil boring locations and the contaminated areas. To be consistent, I only used the soil-boring results within the designated excavation areas in the PBA estimates.

When rainwater washes a COC down into the soil, the COC does not go straight down but tends to spread out laterally as it percolates slowly through the soil. For this reason, I assumed the contaminants in the expected excavation area had spread laterally increasing the area by up to 10%. Therefore, I assigned a 10% halo to the delineated areas to account for the contaminant spread. For example, given an expected excavation area of k square feet, I represented the area in the PBA model as an interval ranging from k to 1.1k square feet. Table 11 shows the expected excavation areas used in the PBA models for the selected sites, and the sources of the data.

Data Collection for Monte Carlo Models

The data required for the MC models to calculate the expected soil volumes are the probability distributions for the areas and the excavation depths. I have discussed

Table 11

Expected Site Excavation Areas and Source References

Site name and location	Minimum expected excavation area (ft^2)	Source of area data
Aladdin Plating Company, Lackawanna County, PA	363,500	Estimated from scaled drawing. Ref: U.S. EPA 1993, Figure 4.
C&R Battery Company, Chesterfield County, VA	142,700	NUS Corporation (1990, p AR302144)
Paoli Rail Yard, Paoli, PA	Throat area & east car shop = 177,500 East storage yard = 189,300 South & west of car shop yard = 119,800 Turnaround track = 91,520	Groundwater Technology (1991, p AR301686 - AR301687)
Peoples Natural Gas Company, Dubuque, IA	127,602	Barr Engineering Company (1994, Appendix B)
Taylor Lumber and Treating Company, Sheridan, OR	171,191	CH2MHILL (2006, p 3-6)
Tonolli Corporation, Nesquehoning County, PA	540,000	Paul C Rizzo Associates (1992, p AR304023)
U.S. Titanium Company, Nelson County, VA	87,120	Hydrosystems Inc. (1987a, p AR301081)

the assumptions used to obtain the distributions for the areas and the expected excavation depths separately in this section.

Assumptions for MC excavation areas.

I assumed the expected excavation areas were uniformly distributed with the areas reported in Table 11 as the minimum areas. I assumed in this case also that there has been a 10% increase in area due to the COC spread during its downward migration. Therefore, the maximum area is 110% of the area in Table 11.

Assumptions for MC excavation depths.

I performed two sets of MC estimates using the same uniform distributions for the expected excavation areas but different distributions for the expected excavation depths. In the first set of MC runs, I assumed that the excavation depths were normally distributed. The mean and standard deviation of the excavation depths for each site were the same as those from the summary distributions in Tables 4 to 10. The only exception was the two Paoli areas where only surface soils were analyzed. In this case, I assumed the excavation depths were uniformly distributed between zero inches and 12 inches.

In the second set of MC runs, I represented the expected excavation depth for each site by percentile distributions corresponding to the values in Tables 4 to 10. For the two Paoli sites where only surface soils were analyzed, I assumed again that the excavation depths were uniformly distributed between zero inches and 12 inches. Tables

12 and 13 present a summary of the data used for the two sets of MC models, respectively.

Data for Deterministic Models

I compiled the U.S. EPA deterministic estimate for each site from the U.S. EPA Feasibility Study reports for the purpose of comparing the outcomes of MC, deterministic, and PBA models. Table 14 presents the U.S. EPA deterministic soil volume estimate for each site and the corresponding reference source for the data.

Data for Regression Models

For this study, the data required for the regression models are the mean and the 95th percentile estimated soil volumes from the PBA and MC models, and the U.S. EPA deterministic soil volumes. These are the independent variables in the regression models. The dependent variables in the regression models are the actual excavated soil volumes. Table 15 presents the actual soil volumes excavated during site remedial activities and the references for the data.

Data Protection Measures

The process of cleaning up a hazardous waste site is a joint venture between the U.S. EPA and the impacted community. The U.S. EPA, through community meetings and public hearings, shares all site investigation reports and discusses the most effective

Table 12

Data for Monte Carlo Models Assuming Normal Probability Distributions for the Excavation Depth

Site name and location	Excavation area (ft^2)	Excavation depth (ft)
Aladdin Plating Company, Lackawanna County, PA	Uniform (363500, 399850)	Normal (3.02, 1.30) Max /Min = 6 / 1.5
C&R Battery Company, Chesterfield County, VA	Uniform (142700, 156970)	Normal (4.61, 3.41) Max / Min = 9 / 0
Paoli Rail Yard, Paoli, PA	Throat area & east car shop = Uniform (177500, 195250)	Normal (3.5, 5.31) Max / Min = 16.5 / 0
	East storage yard = Uniform (189300, 208230)	Uniform (0, 1)
	South & west of car shop yard = Uniform (119800, 131780)	Normal (12.57, 11.05) Max / Min = 30.5 / 0
	Turnaround track = Uniform (91520, 100672)	Uniform (0, 1)
Peoples Natural Gas Company, Dubuque, IA	Uniform (127602, 140362)	Normal (2.235, 3.993) Max / Min = 10 / 0
Taylor Lumber and Treating Company, Sheridan, OR	Uniform (171191, 188310)	Normal (1.91, 2.02) Max / Min = 6 / 0
Tonolli Corporation, Nesquehoning County, PA	Uniform (540000, 594000)	Normal (2.33, 3.44) Max / Min = 10 / 0
U.S. Titanium Company, Nelson County, VA	Uniform (87120, 95832)	Normal (8.69, 8.49) Max / Min = 24 / 0

Table 13

Data for Monte Carlo Models Assuming Percentile Distributions for the Excavation Depths

Site name and location	Excavation area (ft^2)	Excavation depth (ft)
Aladdin Plating Company, Lackawanna County, PA	Uniform (363500, 399850)	Percentile distribution from Table 4.
C&R Battery Company, Chesterfield County, VA	Uniform (142700, 156970)	Percentile distribution from Table 5.
Paoli Rail Yard, Paoli, PA	Throat area & east car shop Uniform (177500, 195250)	Percentile distribution from Table 6.
	East storage yard = Uniform (189300, 208230)	Uniform (0, 1)
	South & west of car shop yard Uniform (119800, 131780)	Percentile distribution from Table 6.
	Turnaround track = Uniform (91520, 100672)	Uniform (0, 1)
Peoples Natural Gas Company, Dubuque, IA	Uniform (127602, 140362)	Percentile distribution from Table 7.
Taylor Lumber and Treating Company, Sheridan, OR	Uniform (171191, 188310)	Percentile distribution from Table 8.
Tonolli Corporation, Nesquehoning County, PA	Uniform (540000, 594000)	Percentile distribution from Table 9.
U.S. Titanium Company, Nelson County, VA	Uniform (87120, 95832)	Percentile distribution from Table 10.

Table 14

U.S. EPA Deterministic Soil Excavation Volume Estimates

Site name and location	U.S. EPA deterministic soil volume (cu. yds)	Source of area data
Aladdin Plating Company, Lackawanna County, PA	11,000	Roy F. Weston (1988a, p AR300326)
C&R Battery Company, Chesterfield County, VA	36,000	NUS Corporation (1990b, p AR302142)
Paoli Rail Yard, Paoli, PA	25,219	Groundwater Technology (1991, p AR301686 - AR301687)
Peoples Natural Gas Company, Dubuque, IA	9,500	Barr Engineering Company (1994, Appendix B)
Taylor Lumber and Treating Company, Sheridan, OR	12,824	CH2MHILL (2006, p 3-6)
Tonolli Corporation, Nesquehoning County, PA	39,300	Paul C Rizzo Associates (1992, p AR304023)
U.S. Titanium Company, Nelson County, VA	32,000	Hydrosystems Inc. (1987a, p AR301167)

Table 15

Actual Excavated Soil Volumes Obtained from U.S. EPA Sources

Site name and location	Actual excavated soil volume (cu. yds)	Source of volume data
Aladdin Plating Company, Lackawanna County, PA	28,600	U.S. EPA (1999a, p2)
C&R Battery Company, Chesterfield County, VA	38,600	U.S. EPA (2003, p6)
Paoli Rail Yard, Paoli, PA	83,000	U.S. EPA (2005, p5)
Peoples Natural Gas Company, Dubuque, IA	17,350	U.S. EPA (2000a, p3)
Taylor Lumber and Treating Company, Sheridan, OR	15,700	U.S. EPA (2008, p3)
Tonolli Corporation, Nesquehoning County, PA	114,300	U.S. EPA (1999b, p3)
U.S. Titanium Company, Nelson County, VA	65,000	U.S.EPA (2000b, p6)

cleanup methods with the impacted community throughout the remediation process. The U.S. EPA has created the SSIB as the database from which the general public can obtain information about the cleanup progress for a site and/or obtain copies of the site investigation reports.

The reports from which I obtained the data for the study are still in the public domain. They could be freely accessed from U.S. EPA's website at: http://www.epa.gov/superfund/cleanup/index.htm. Therefore, no individual, organization, or company would be injured as a result of the publication of the data in this study and data protection measures are not needed.

Summary

I have presented the analytical methods employed for the study in this chapter. They are: probability bounds analysis, MC analysis, linear regression analysis, and a hypothesis test for the existence of a population coefficient of correlation. I have presented details of the study design, data sources, data collection methods, the U.S. EPA deterministic soil volume estimates for each site, and finally, the data used for the PBA and MC models. Chapter 4 presents the results of the PBA and MC models, and analysis of the regression results. Chapter 5 presents the interpretations of the results, conclusions, and recommendations from the study.

Chapter 4: Results

Introduction

Chapter 4 presents the results of the PBA and MC modeling and analyses conducted for the study. I have reported the results and findings separately for PBA soil-volume estimates, MC soil-volume estimates, and the U.S. EPA deterministic soil volume estimates. Each section contains the results of the modeling soil volume estimates followed by a series of two-variable linear regression analysis to determine the strengths of the correlations between the various modeling estimates of soil volumes and the actual excavated soil volumes. Finally, I have compared and contrasted the best regression results for the PBA models, MC models, and the U.S. EPA deterministic estimates in this chapter.

PBA Modeling Results

I performed two sets of PBA estimates using the data from Tables 4 to 10. In both sets, I represented the expected site excavation areas as intervals using the corresponding area data in Table 11 for each site as the lower bound, and 110% of the same area as the upper bound.

PBA offers the choice of using either summary statistics data or percentile distributions to represent the excavation depths. For the first set of PBA estimates, I represented the excavation depths by summary distribution data. For the second set of

PBA estimates, I represented the excavation depths by percentile data. The purpose was to determine which of them would lead to a better correlation with the actual soil volumes. I have discussed these below.

PBA Using Summary Statistics Data for Excavation Depths

The Risk Calc software has a function called "minmaxmeanstddev(minimum, maximum, mean, standard deviation)" that uses the following summary statistics data to generate p-boxes for the PBA estimates: minimum, maximum, mean, standard deviation. In the first set of PBA estimates, I used this function to represent the expected excavation depths at each site, using the data from Tables 4 to 10. The PBA programs that I used to calculate the expected soil excavation volume at each site and the results generated are presented in Appendices A to G. Table 16 presents the mean soil volume estimates based on PBA using summary statistics data and Table 17 presents the results for the 95th percentile soil volume estimates based on PBA using summary statistics data. I have included the actual soil volumes excavated during site remediation in the tables to facilitate the comparisons.

As expected, PBA does not give a unique result for the mean (Table 16). The model output is the minimum and maximum values for the mean. I calculated the mid-point values in the tables as the average of the maximum and minimum soil volume estimates. With the exception of the Aladdin and Paoli sites, all the maximum mean values underestimate the actual excavated soil volumes. The maximum underestimate was for the Tonolli site where it was only 48% of the actual excavated soil volume.

Table 16

Results for PBA-Estimated Mean Soil Volumes Using Summary Statistics Data for Excavation Depths

Site name and location	Actual soil volume (cubic yards)	PBA mean results with summary statistics data: Minimum	Maximum	Mid-point
Aladdin Plating Company Lackawanna County, PA	28,600	39,783	45,599	42,691
C & R Battery Company Chesterfield County, PA	38,600	23,463	27,703	25,583
Paoli Rail Yard Paoli, PA	83,000	74,585	102,299	88,442
Peoples Natural Gas Dubuque, IA	17,350	9,643	12,588	11,116
Taylor Lumber & Treating Company, Sheridan, PA	15,700	11,469	13,962	12,716
Tonolli Corporation Nesquehoning County, PA	114,300	43,160	54,700	48,930
U. S. Titanium Company Nelson County, VA	65,000	26,670	32,214	29,442

Table 17

Results of PBA-Estimated 95th Percentile Soil Volumes Using Summary Statistics Data for Excavation Depths

Site name and location	Actual soil volume (cubic yards)	95th Percentile PBA results with summary statistics data		
		Minimum	Maximum	Mid-point
Aladdin Plating Company Lackawanna County, PA	28,600	50,194	88,856	69,525
C & R Battery Company Chesterfield County, PA	38,600	36,386	52,323	44,355
Paoli Rail Yard Paoli, PA	83,000	92,662	279,623	186,143
Peoples Natural Gas Dubuque, IA	17,350	35,977	51,986	43,982
Taylor Lumber & Treating Company, Sheridan, PA	15,700	22,778	41,847	32,313
Tonolli Corporation Nesquehoning County, PA	114,300	123,548	220,000	171,774
U. S. Titanium Company Nelson County, VA	65,000	50,308	85,184	67,746

The 95^{th} percentile estimates for the PBA modeling using summary statistics data for depths present a different contrast (Table 17). The upper bound 95^{th} percentile results overestimate the actual excavated soil volumes at each site. In the case of the Paoli site, the PBA-estimated soil volume was 337% of the actual excavated soil volume. The lower bound results of the 95^{th} percentile PBA estimates appear to be within a more reasonable range of the actual excavated soil volumes.

PBA Using Percentile Data for Excavation Depths

The Risk Calc software provides the option to use percentile distributions as the constraining data to generate p-boxes for the estimates. In the second set of PBA estimates, I used the Risk Calc function, "fivenumbers(minimum, 25^{th} percentile value, median, 75^{th} percentile, maximum)" to represent the expected excavation depths at each site, using the percentile data from Tables 4 to 10. The PBA programs that I used to calculate the expected soil volume at each site and the results generated are presented in Appendices H to N. Table 18 presents the mean PBA soil volume estimates using percentile data for depths, and Table 19 presents the 95^{th} percentile PBA soil volume estimates using percentile data for depths. I have included the actual soil volumes excavated during site remediation to facilitate the comparisons.

The upper bound of the mean soil volume estimates using percentile data for depths (Table 18) are higher than the upper bound of the mean soil volume estimates using summary statistics data for depths (Table 16). The upper bounds of the mean soil volume estimates using percentile data for depths (Table 18) appear to be within a more

Table 18

Results of PBA-Estimated Mean Soil Volumes Using Percentile Data for Excavation Depths

Site name and location	Actual soil volume (cubic yards)	PBA mean results from percentile data		
		Minimum	Maximum	Mid-point
Alladin Plating Company Lackawanna County, PA	28,600	31,772	51,833	41,803
C & R Battery Company Chesterfield County, PA	38,600	19,423	34,883	27,153
Paoli Rail Yard Paoli, PA	83,000	50,496	135,407	92,952
Peoples Natural Gas Dubuque, IA	17,350	2268	15,596	8,932
Taylor Lumber & Treating Company, Sheridan, PA	15,700	6,213	17,437	11,825
Tonolli Corporation Nesquehoning County, PA	114,300	24,000	82,500	53,250
U. S. Titanium Company Nelson County, VA	65,000	17,109	40,596	28,853

Table 19

Results of PBA-Estimated 95th Percentile Soil Volumes Using Percentile Data for Excavation Depths

Site name and location	Actual soil volume (cubic yards)	95th Percentile PBA results from percentile data		
		Minimum	Maximum	Mid-point
Aladdin Plating Company Lackawanna County, PA	28,600	40,389	88,856	64,623
C & R Battery Company Chesterfield County, PA	38,600	39,639	52,323	45,981
Paoli Rail Yard Paoli, PA	83,000	95,396	279,623	187,510
Peoples Natural Gas Dubuque, IA	17,350	9,452	51,986	30,719
Taylor Lumber & Treating Company, Sheridan, PA	15,700	12,681	41,847	27,264
Tonolli Corporation Nesquehoning County, PA	114,300	100,000	220,000	160,000
U. S. Titanium Company Nelson County, VA	65,000	43,560	85,184	64,372

reasonable range of the actual soil volumes than the soil volume estimates using summary statistics data for depths (Table 16).

In Table 19, the upper bounds of the 95th percentile estimates exceed the actual excavated soil volumes at all sites. The extent of the overestimate is quite significant, ranging from 337% of the actual excavated soil volume for the Paoli site to 31% for the U.S. Titanium site. This eliminates the possibility that the upper bound of the 95th percentile PBA soil volume estimate could be used as a quick estimate of the upper bound of the expected soil volume for budgetary purposes.

Regression Results between PBA Estimates and Actual Soil Volumes

I performed a series of two-variable linear regression analyses using the actual excavated soil volumes as the dependent variables and each of the lower bound, upper bound, and mid-point PBA estimates from Tables 16, 17, 18, and, 19 as the independent variables. Table 20 presents the regression results for PBA estimates using summary statistics data for depths, and Table 21 presents the regression results for PBA estimates using percentile data for depths. From Table 20, the best correlation is the lower bounds of the 95th percentile results for the PBA estimates using summary statistics data for excavation depths. The coefficient of correlation, R-value, is 0.943. The coefficient of determination, R^2-value, is 0.89 which implies that about 89% of the variation in the actual excavated soil volumes from site to site is explained by the corresponding linear statistical relationship in Table 20.

Table 20

Regression Results between PBA-Estimated Soil Volumes Using Summary Statistics Data for Excavation Depths and the Actual Excavated Soil Volumes

Excavation depth data source for PBA analysis	Decision point	R^2 -Value	R-Value	Slope	Intercept
PBA using summary statistics data	mean lower bound	0.464	0.681	1.126	14,982
PBA using summary statistics data	mean upper bound	0.460	0.678	0.812	18,255
PBA using summary statistics data	mean mid-point	0.463	0.680	0.946	16,800
PBA using summary statistics data	95^{th} percentile lower bound	0.890	0.943	0.970	-5,290
PBA using summary statistics data	95^{th} percentile upper bound	0.705	0.840	0.332	12,949
PBA using summary statistics data	95^{th} percentile mid-point	0.785	0.886	0.516	6,397

Table 21

Regression Results between PBA-Estimated Soil Volumes Using Percentile Distribution Data for Excavation Depths and the Actual Excavated Soil Volumes

Excavation depth data source for PBA analysis	Decision point	R^2 -Value	R-Value	Slope	Intercept
PBA using percentile data	mean lower bound	0.309	0.556	1.270	24,341
PBA using percentile data	mean upper bound	0.564	0.751	0.656	16,323
PBA using percentile data	mean mid-point	0.501	0.708	0.910	17,373
PBA using percentile data	95^{th} percentile lower bound	0.884	0.940	0.966	4,727
PBA using percentile data	95^{th} percentile upper bound	0.705	0.840	0.332	12,949
PBA using percentile data	95^{th} percentile mid-point	0.772	0.878	0.507	9,720

The second best correlation result is for the lower bounds of the 95th percentile PBA soil volume estimates using percentile data for excavation depths. The R^2-value is 0.884 and the R-value is 0.940. The closeness between the two best PBA correlation results suggest that there is no significant difference between the lower bound 95th percentile PBA results for PBA estimates using summary statistics data for excavation depths and PBA estimates using percentile data for excavation depths.

Figure 3 presents an ordered bar graph summary of the correlation results for all the PBA estimates. The graph indicates that the 95th percentile estimates do, in general, give better correlation to the actual soil volumes than the mean estimates.

Results for the Monte Carlo Estimates

I performed two sets of soil volume estimates using Monte Carlo simulation. I used the Crystal Ball software and each simulation was set to 1000 iterations. In both sets of MC estimates, I represented the expected excavation area by uniform distributions. I used the same maximum and minimum area values in the PBA estimates for each site for the MC estimates. However, in the MC estimates, I assumed the area was uniformly distributed between the maximum and minimum values.

The main difference between the two sets of the MC estimates was the type of distribution used to represent the expected excavation depths. In the first set of MC estimates, I represented the expected excavation depth at each site by a normal distribution with the same mean and standard deviation as the data in Tables 4 to 10. In

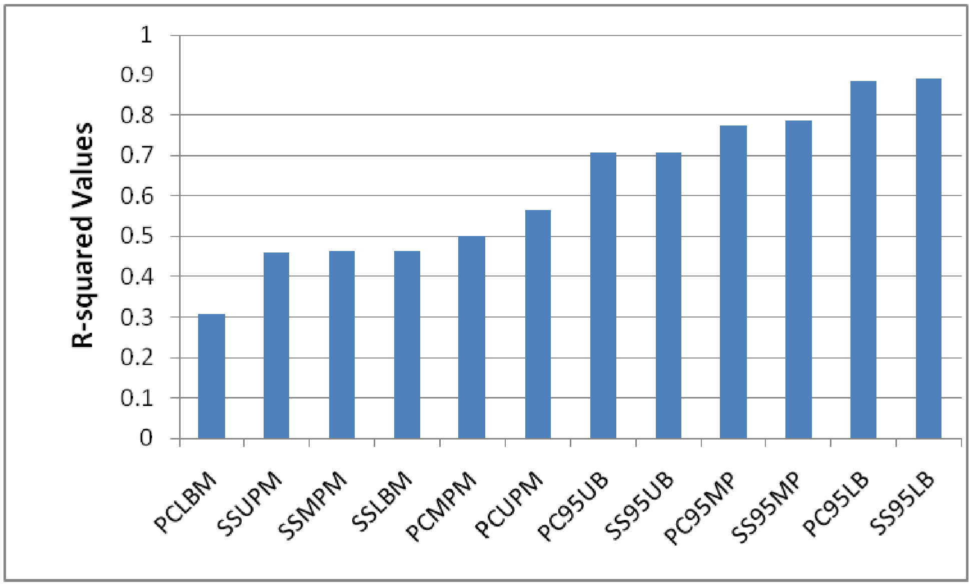

Figure 3. Ordered bar graph summary of R^2-values for all PBA estimates

PCLBM = Lower bound mean estimate, PBA using percentile distribution for depths.

SSUPM = Upper bound mean estimate, PBA using summary data for depths

SSMPM = Mid-point mean estimate, PBA using summary data for depths

SSLBM = Lower bound mean estimate, PBA using summary data for depths

PCMPM = Mid-point mean estimate, PBA using percentile data for depths

PCUPM = Upper bound mean estimate, PBA using percentile data for depths

SS95UB = Upper bound 95^{th} percentile estimate, PBA using summary data for depths

PC95UB = Upper bound 95^{th} percentile estimate, PBA using percentile data for depths

PC95MP = Mid-point 95^{th} percentile estimate, PBA using percentile data for depths

SS95MP = Mid-point 95^{th} percentile estimate, PBA using summary data for depths

PC95LB = Lower bound 95^{th} percentile estimate, PBA using percentile data for depths

SS95LB = Lower bound 95^{th} percentile estimate, PBA using summary data for depths.

the second set of MC estimates, I represented the expected excavation depths by percentile distributions corresponding to the data in Tables 4 to 10.

Results for MC Estimates Assuming Normal Distributions for Excavation Depths

Table 22 presents the results of the MC soil volume estimates assuming normal distributions for the excavation depths. The 95th percentile results overestimate the actual soil volumes for all sites. The results for the mean estimates are mixed, with some very high overestimates and some very low underestimates.

Results for MC Estimates Assuming Percentile Distributions for Excavation Depths

Table 23 presents the results of the MC soil volume estimates assuming percentile distributions for the expected excavation depths. The 95th percentile results again overestimate the actual soil volumes for all sites. In this case also, the results for the mean estimate are mixed, with some very high overestimates and some very low underestimates.

Regression Results between MC Estimates and Actual Soil Volumes

Table 24 presents the regression results between the MC estimated soil volumes as the independent variables and the actual excavated soil volumes as the dependent variables. The best correlation was obtained for the 95th percentile soil volume estimate in the case where I represented the areas by uniform distributions and the expected

Table 22

Results of MC-Estimated Soil Volumes Assuming Normal Distribution for Excavation Depths

Site name and location	Actual soil volume (cubic yards)	Mean soil volume (cubic yards)	95th Percentile soil volume
Aladdin Plating Company Lackawanna County, PA	28,600	45,517	71,512
C & R Battery Company Chesterfield County, PA	38,600	25,261	46,354
Paoli Rail Yard Paoli, PA	83,000	108,898	185,668
Peoples Natural Gas Dubuque, IA	17,350	19,261	40,858
Taylor Lumber & Treating Company, Sheridan, PA	15,700	16,048	34,383
Tonolli Corporation Nesquehoning County, PA	114,300	79,623	170,958
U. S. Titanium Company Nelson County, VA	65,000	34,861	69,961

Table 23

Results of MC-Estimated Soil Volumes Assuming Percentile Distribution for Excavation Depths

Site name and location	Actual soil volume (cubic yards)	Mean soil volume (cubic yards)	95^{th} Percentile soil volume
Aladdin Plating Company Lackawanna County, PA	28,600	57,768	87,547
C & R Battery Company Chesterfield County, PA	38,600	31,294	51,511
Paoli Rail Yard Paoli, PA	83,000	28,163	81,062
Peoples Natural Gas Dubuque, IA	17,350	26,404	50,310
Taylor Lumber & Treating Company, Sheridan, PA	15,700	35,966	61,038
Tonolli Corporation Nesquehoning County, PA	114,300	115, 431	212,627
U. S. Titanium Company Nelson County, VA	65,000	28,051	80,987

Table 24

Regression Results between MC-Estimated Soil Volumes and the Actual Excavated Soil Volumes

Excavation depth assumption for MC analysis	Decision point	R^2-Value	R-Value	Slope	Intercept
MC assuming normal distribution	Mean	0.643	0.802	0.856	11,488
MC assuming normal distribution	95th Percentile	0.792	0.890	0.524	5,443
MC assuming percentile distribution	Mean	0.390	0.625	0.715	18,795
MC assuming percentile distribution	95th Percentile	0.666	0.816	0.536	3,890

excavation depths by normal distributions. The R^2-value is 0.792 which implies about 79% of the variation in the actual excavated soil volumes are explained by the corresponding linear relationship in Table 24.

Figure 4 presents an ordered bar graph summary of all the correlation results for the MC estimates. Figure 4 also shows that the 95th percentile MC estimates correlate better with the actual soil volumes than the mean estimates.

Regression Results between U.S. EPA Deterministic Estimates and Actual Volumes

Table 25 presents the results of the regression analysis between the U.S. EPA deterministic soil-volume estimates as the independent variables and the actual excavated soil volumes as the dependent variables. The R^2-value is only 58.9%.

Comparison of Best Correlation Results for PBA, MC, and Deterministic Estimates

Table 26 presents a summary of the best correlation results for the PBA estimates, the MC estimates, and the U.S. EPA deterministic estimate. The summary is also presented in an ordered bar graph in Figure 5. The results indicate that the lower bound of the 95th percentile PBA estimate is a better estimator of remedial soil volumes than either the mean or 95th percentile MC estimates. The result with the poorest correlation was the U.S. EPA deterministic estimates. This shows that the current U.S. EPA deterministic methods are inadequate for estimating soil volumes and the probabilistic methods, PBA and MC, give much better soil volume estimates.

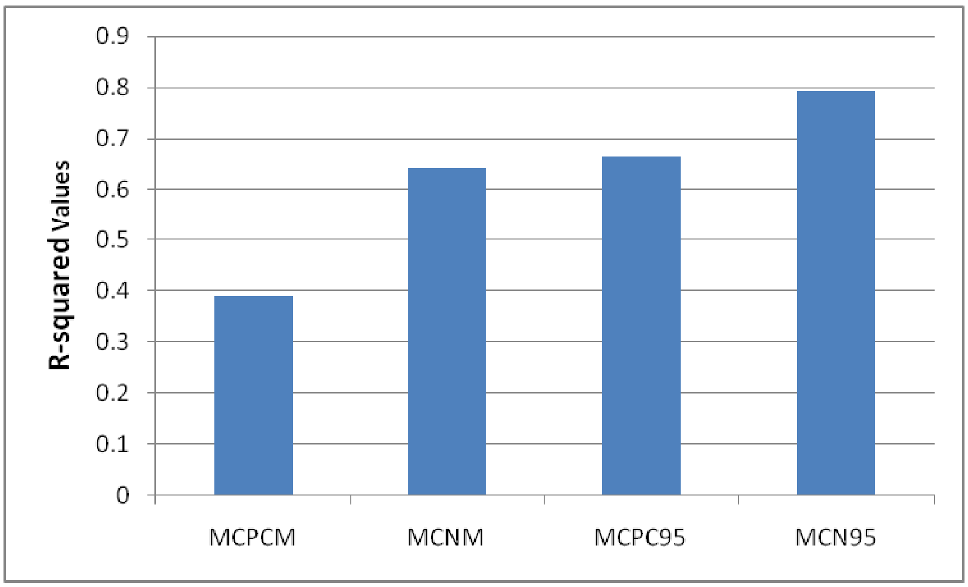

Figure 4. Ordered bar graph summary of R^2-values for all MC estimates

MCPCM = Mean estimate, MC using percentile distributions for depths

MCNM = Mean estimate, MC assuming normal distributions for depths

MCPC95 = 95th percentile estimate, MC using percentile distributions for depths

MCN95 = 95th percentile estimate, MC assuming normal distributions for depths

Table 25

Regression Results between U.S. EPA Deterministic Soil Volumes and the Actual Excavated Soil Volumes.

Source of soil volume estimates	Decision point	R^2-Value	R-Value	Slope	Intercept
U.S. EPA Deterministic Estimates	N/A	0.589	0.7675	2.266	-1,892

Table 26

Summary of Best Soil Volume Estimates from the PBA, MC, and Deterministic Models

Excavation depth assumption	Decision point	R^2-Value	R-Value	Slope	Intercept
PBA using summary statistics data	95th Percentile Lower Bound	0.890	0.943	0.970	-5,290
PBA using percentile data	95th Percentile Lower Bound	0.884	0.940	0.966	4,727
MC using summary statistics data	95th Percentile	0.792	0.890	0.524	5,443
U.S. EPA deterministic estimates	Calculated Result	0.589	0.7675	2.266	-1,892

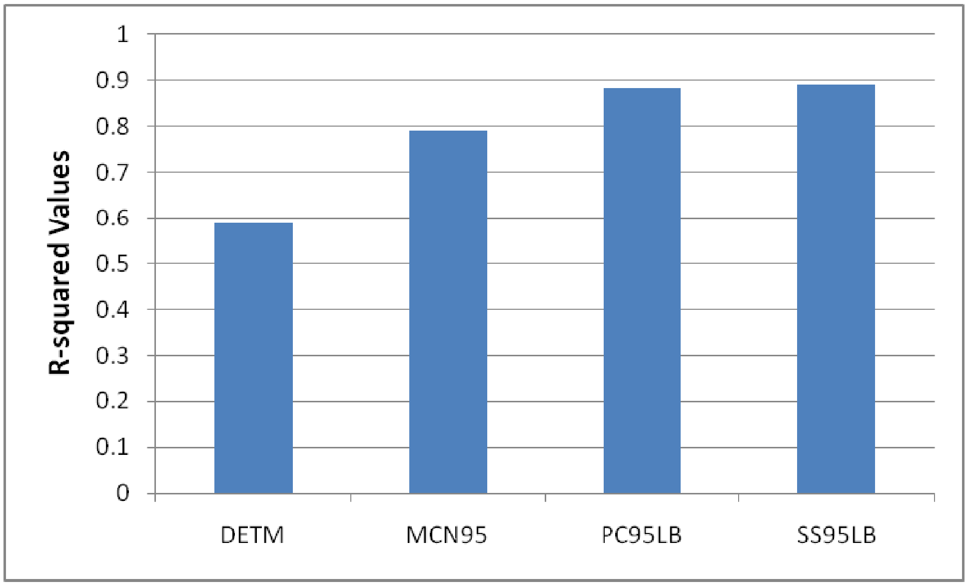

Figure 5. Ordered bar graph summary of best R^2-values

DETM = U.S. EPA deterministic estimates

MCN95 = 95th percentile estimate, MC assuming normal distributions for depths

PC95LB = Lower bound 95th percentile estimate, PBA using percentile data for depths

SS95LB = Lower bound 95th percentile estimate, PBA using summary data for depths.

Hypothesis Test for a Population Coefficient of Correlation

I performed a test of hypothesis on the two best PBA results in Table 26 to confirm the existence of a population coefficient of correlation. The null hypothesis was that the population coefficient of correlation equaled zero. The alternate hypothesis was that the population coefficient of correlation was not zero. I have shown the calculations in Appendices O and P. The results do confirm the existence of correlation between the lower bound 95th percentile PBA results and the actual excavated soil volumes.

Summary

This chapter presented the soil volume estimates obtained by using PBA and MC, and the results of the correlation analysis. The goal of the study was to find an optimal decision point for PBA models when used to estimate remedial soil volumes at hazardous waste sites. Table 26 presents a summary of the best correlations obtained from the PBA-estimates, MC-estimates, and the U.S. EPA deterministic estimates. The results of this study indicate that the optimal decision point for PBA models when used to estimate remedial soil volumes is the lower bound value of the 95th percentile PBA estimate. The results also show that both the lower bound 95th percentile PBA estimates and the 95th percentile MC estimates are better predictors of remedial soil volumes than the present U.S. EPA deterministic estimates.

There was no significant difference in the correlation results for the lower bound 95th percentile PBA estimates in the case where I used summary distribution data for the excavation depths or the case where I used percentile data for the excavation depths.

This indicates that it is optimal to use either the summary statistics or the percentile distributions in making PBA estimates. Chapter 5 presents the interpretations of the results, conclusions, and recommendations.

Chapter 5: Discussion, Conclusions, and Recommendations

Introduction

The United States Environmental Protection Agency (U.S. EPA) currently uses deterministic methods to estimate remedial soil volumes at hazardous waste sites. I have presented data in this study to show that the U.S. EPA deterministic estimates widely underestimate the actual excavated soil volumes. Probability Bounds Analysis (PBA) is an alternate method to estimate remedial soil volumes that also accounts for the uncertainties and the imprecision in the soil volume estimates. However, the PBA result for any selected percentile or any measure of central tendency is not a unique number but is rather a wide interval. The research questions, then, were to determine whether the lower bound or upper bound of the 95th percentile PBA estimate or that of the mean estimate is the optimal decision point. The identification of an optimal decision point would encourage a greater use of the PBA method to estimate remedial soil volumes at hazardous waste sites under uncertainty.

In the course of the study, additional questions arose regarding how the best PBA decision point estimate would compare with a similar MC estimate or the current U.S. EPA deterministic estimate. In other words, it is not only enough that the best PBA decision point estimate should correlate well with the actual excavated soil volumes but it should also be a better estimator of the remedial soil volumes than both MC estimates and the current U.S. EPA deterministic estimates.

I addressed the above questions in the study by comparing the strengths of the correlations between the actual excavated soil volumes and the PBA, MC, and U.S. EPA deterministic estimates. I have summarized the results in Table 26 and this will be the reference material for the discussion of the results of the study, and the conclusions.

Discussion of Results

The study shows that the PBA methodology can be applied to estimate remedial soil volumes using data that is normally collected during hazardous waste site investigation studies. The two best correlations in Table 26 are the lower bound 95^{th} percentile PBA soil volume estimates using summary statistics data for depths, and the lower bound 95^{th} percentile PBA soil volume estimates using percentile distributions for depths. Table 26 also indicates that the correlation for the lower bound 95^{th} percentile PBA estimates is about the same for both PBA with summary statistics data for excavation depths and PBA with percentile distributions for excavation depths.

Table 26 also indicates that PBA gives a higher correlation with the actual excavated soil volumes than both MC and U.S. EPA deterministic approach. The reason could be that only two variables, area and depth, are needed for the analysis. Aughenbaugh & Paredis (2005) have shown that PBA gives better results than MC when the number of variables in the model is less than 30.

Although the MC soil volume estimates give better correlations than the U.S. EPA deterministic estimates, the arbitrary selection of the probability distributions for the excavation areas and depths in the MC estimates may require justification in a contested settlement. This is not the case with PBA models. For example, I assumed the excavation areas in the MC estimates were uniformly distributed. The shapes of the cumulative density functions for the uniform distributions are the same as that of the p-boxes derived from the same interval limits for the PBA estimates. However, the assumption of interval limits for areas in the PBA estimates instead of uniform distributions is intuitive enough to be accepted without justification. Another problem with the MC estimates in this study is that it is hard to justify the assumption of normal distributions for the expected excavation depths given the highly skewed data for excavation depths in Tables 4 to 10.

The worst correlation result in Table 26 is for the U.S. EPA deterministic estimates using conservative values. Table 14 confirms that The U.S. EPA deterministic methods consistently underestimate the remedial soil volumes. The best deterministic underestimate is for the C&R Battery site where it is 93% of the actual excavated soil volume. The worst deterministic underestimate is for the Paoli Rail Yard site where it is only 30% of the actual excavated soil volume. The median deterministic underestimate is for the U.S. Titanium site where it is 49% of the actual excavated soil volume.

In the course of soil excavation activities, soil samples are routinely taken at some pre-determined depths and analyzed for the presence of COCs. The purpose is to minimize cost by ensuring that the excavation does not proceed beyond what is necessary to remove only the soils that have COC concentrations above the cleanup level. However, because the deterministic models underestimate the remedial soil volumes, situations arise where additional soil volumes must be excavated in order to remove all soils that have COC concentrations above the cleanup level. When this occurs and the additional excavation work appears to be significant, the U.S. EPA issues a document called, “Explanation of Significant Differences”, (ESD), to authorize the additional soil excavation. For example, the original U.S. EPA deterministic soil volume estimate for the Tonolli site was 39,300 cubic yards. However, as the excavation progressed, the US E.P.A. issued an ESD (U.S. EPA 1999c) to explain that the excavation would need to be expanded because more contaminated soils above the cleanup level had been found in the course of the excavation. The ESD did not provide a revised estimate of the expected total soil volume. In the end, a total of 114,300 cubic yards of soil was excavated at the Tonolli site. The financial impact of such a dramatic increase in the total soil volume could have a crippling effect on the business operations of the PRPs who would eventually have to pay for this additional excavation cost on short notice. Since the study results show that both PBA and MC give much higher remedial soil volume estimates than the U.S. EPA deterministic estimates, the use of these probabilistic methods to estimate remedial soil volumes under uncertainty should be encouraged. This is better for the PRPs because the additional soil volumes that may have to be excavated in the event there is an underestimate would be much less in this case than in the case of the deterministic estimates.

This study is very relevant to the work of the U.S. UPA and I would share the results with the agency. I hope to get the co-operation of the U.S. EPA to carry out a larger study that overcomes the limitation of this study.

The limitation of the study is that I did not select the sites randomly from the U.S. EPA universe of hazardous waste sites. Therefore, the results and conclusions apply to this population only and may not be extended to include the general population of hazardous waste sites.

Conclusion

The conclusion from the study is that the optimal decision point for PBA models when used to estimate remedial soil volumes at hazardous waste sites under uncertainty is the lower bound 95th percentile estimate. If we let PBA95TH represent the lower bound 95th percentile PBA soil volume estimate, then the statistical relationship between the actual excavated soil volume and the lower bound 95th percentile PBA soil volume estimate is:

$$\text{Actual Soil Volume} = \text{PBA95TH} * 0.970 - 5290$$

The R^2 value is 89% which means about 89% of the variation in the actual excavated soil volumes in the study are explained by the above equation. There is no significant difference in the estimate whether the expected excavation depths are represented by summary statistics data or percentile distributions in the PBA soil volume estimate.

Social Change Implications

The adoption of the PBA methodology for estimating remedial soil volumes under uncertainty would lead to the following benefits:

- Reduce litigation expenses among the potentially responsible parties.
- Speed up remedial actions so contaminants do not continue to adversely impact the health and safety of nearby residents.
- Increase productivity in the area because nearby residents would be healthier and have less sick off-days.

Recommendations

Although the study results look promising, the fact that I did not select the study sites randomly limits the general application of the regression equation. One recommendation to improve the study is to validate the above regression equation by applying it to a different set of sites. The procedure would be to select other sites where soil excavation has been completed from the U.S. EPA universe, use PBA to estimate the remedial soil volumes, select the lower bound 95th percentile estimates, and then, use the regression equation as a correction factor to estimate the actual soil volumes. The 'corrected' soil volumes could then be compared with the actual excavated soil volumes for correlations.

The sample size limitation for this study was imposed by the availability of data on the U.S. EPA's website. The study could also be improved by selecting a greater number of sites randomly from the subset of the U.S. EPA's universe of hazardous waste sites where soil excavation is complete for another study. However, this can only be feasible when the U.S. EPA is an active participant in the study. In this way, the information that is not available on U.S. EPA's website could be recovered from the U.S. EPA files or from contractor records.

References

Aughenbaugh, J. M., & Paredis, C. J. J. (2007). Probability bounds analysis as a general approach to sensitivity analysis in decision making under uncertainty. *SAE World Congress & Exhibition, April 2007,* 1-15.

Aughenbaugh, J. M. & Paredis, C. J .J. (2005). The value of using imprecise probabilities in engineering design. *Proceedings of ASME 2005 Design Engineering Technical Conference, Long Beach, CA,* 1-13.

Barr Engineering Company (1991). *Remedial investigation /feasibility study – Peoples Natural Gas Company*. Kansas City, KS: United States Environmental Protection Agency Region 7.

Barr Engineering Company (1994). *Final remedial design submittal – Peoples Natural Gas site.* Kansas City, KS: United States Environmental Protection Agency Region 7.

Bergback, B. & Oberg, T. (2005). A review of probabilistic risk assessment of contaminated land. *Journal of Soils and Sediments, 5,* 213-224.

Bergback, B., Oberg, T., & Sander, P. (2006). Uncertain numbers and uncertainty in the selection of input distributions: Consequences for a probabilistic risk assessment of contaminated land. *Risk Analysis: An International Journal, 26 (5),* 1363-1375.

CH2MHILL (2004). *Remedial investigation report – Taylor Lumber and Treating Superfund site.* Seattle, WA: United States Environmental Protection Agency Region 10.

CH2MHILL (2006). *Taylor Lumber and Treating Superfund site final design and design basis report.* Seattle, WA: United States Environmental Protection Agency Region 10.

Dankwah, C. O. (2009). KAM-7 Application Section. *Investigating the Possibility of Using an Imprecise Probability Methodology to Estimate Future Remediation Costs for Hazardous Waste Sites under Uncertainty*. Minneapolis, MN: Walden University.

Ferson, S. (2002). *RAMAs Risk Calc 4.0 Software: Risk Assessment with uncertain numbers*. Boca Raton, Florida: Lewis Publishers.

Ferson, S., Hope, B. K., & Regan, H. M. (2002). Analysis and Portrayal of uncertainty in a food-web exposure model. *Human and Ecological Risk Assessment, 8(7),* 1757-1777.

Ferson, S., Regan, H. M., & Sample, B. E. (2002). Comparison of deterministic and probabilistic calculation of ecological soil screening levels. *Environmental Toxicology and Chemistry, 21,* 882-890.

Ferson, S., & Troy, W. T. (2006). Sensitivity analysis using probability bounding. *Reliability Engineering & System Safety, 91(10-11),* 1435-1442.

Groundwater Technology Inc. (1990). *Final remedial investigation and risk Assessment (RI/RA) report for the remedial investigation and feasibility study – Paoli Rail Yard*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

Guzelian, R. C. (1979). Correlation Analysis. *Statistics for Decision Making (pp. 511-515)*. Philadelphia, Pennsylvania: W. B. Saunders Company.

Hydrosystems, Inc. (1987a). *Supplemental remedial investigation, volume 2 of 5 – U.S. Titanium site*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

Hydrosystems, Inc. (1987b). *Supplemental remedial investigation, volume 4 of 5 – U.S. Titanium site*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

NUS Corporation. (1990a). *Final remedial investigation report – C&R Battery site*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

NUS Corporation. (1990b). *Final feasibility study – C&R Battery site*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

Oberg, T & Sander, P. (2006). Comparison of deterministic and probabilistic risk assessments: A case study at a closed mill in southern Sweden. *Journal of Soils and Sediments, 6,* 55-61.

Paul C. Rizzo Associates Inc. (1991). *Remedial investigation – Tonolli Corporation Superfund site*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

Paul C. Rizzo Associates Inc. (1992). *Feasibility study – Tonolli Corporation Superfund site*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

Roy F. Weston Company (1988). *Field study of the Aladdin Plating site.* Philadelphia, PA: United States Environmental Protection Agency Region 3.

Roy F. Weston Company (1988). *Engineering evaluation/cost analysis: Field study of the Aladdin Plating site.* Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (1989). *Record of decision: U.S. Titanium Company*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (1990). *Record of decision: C & R Battery Co., Inc*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (1991). *Record of decision: Peoples Natural Gas Co.* Kansas City, KS: United States Environmental Protection Agency Region 7.

U.S. EPA (1992a). *Record of decision: Paoli Rail Yard*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (1992b). *Record of decision: Tonolli Corporation*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (1993). *Declaration for record of decision: Aladdin Plating site operable unit*

#2. Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (1998). *Record of decision: Aladdin Plating Company*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (1999a). *Five-year review report: Aladdin Plating Superfund Site*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (1999b). *Superfund closeout report: Tonolli Corporation*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (1999c). *Explanation of significant differences: Tonolli Corporation*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (2000a). *Preliminary closeout report: Peoples Natural Gas Site*. Kansas City, KS: United States Environmental Protection Agency Region 7.

U.S. EPA (2000b). *Five-year review report: U.S. Titanium Superfund Site*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (2003). *Second five-year review report: C & R Battery Superfund Site*.

Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (2005). *Record of decision: Aladdin Plating Company*. Seattle, WA: United States Environmental Protection Agency Region 10.

U.S. EPA (2005). *Preliminary closeout report: Paoli Rail Yard Superfund Site*. Philadelphia, PA: United States Environmental Protection Agency Region 3.

U.S. EPA (2008). *Preliminary closeout report: Taylor Lumber and Treating Superfund Site*. Seattle, WA: United States Environmental Protection Agency Region 10.

Williams, R. C. & Downs, T. (1990). Probabilistic Arithmetic I: Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning. 4,* 89-158.

Appendix A: PBA Soil Volume Estimate Using Summary Statistics Data – Aladdin Plating Site

```
// DISTRIBUTION-FREE MODELING FOR THE ALADDIN PLATING COMPANY SITE
// USING p-BOXES CONSTRUCTED FROM THE SUMMARY STATISTICS DATA IN TABLE 4

// Expected soil excavation area range, assuming 10% halo
        Area = [363500, 399850]  // interval estimate of area, ft^2
// Expected soil excavation depth from the summary statistics data in Table 4
        Dmin = 1.5      // minimum depth, ft
        Dmax = 6        // maximum depth, ft
        Dmean = 3.02    // mean depth, ft
        Stddev = 1.30 // standard deviation, ft
// Probability bounds on depth is constructed from the above summary statistics data using the
//Risk Calc  Version 4.0 function: minmaxmeanstddev (minimum, maximum, mean, standard
//deviation)
        Depth = minmaxmeanstddev(Dmin, Dmax, Dmean, Stddev)

// Calculation of the expected soil volume
        Volume = (Area * Depth)/27    //   soil excavation area in cu. yds.

// Displaying results
        Volume      // displays min, max, mean, and, variance for soil volume
  ~(range=[20194.4,88855.6],  mean=[39783,45599],  var=[1e+07,9e+08])
        cut(Volume, 95%)  // displays the 95th percentile soil volumes
  [ 50194.23, 88855.56]
```

Appendix B: PBA Soil Volume Estimate Using Summary Statistics Data – C&R Battery Plating

```
// DISTRIBUTION-FREE MODELING FOR THE C & R BATTERY
// USING p-BOXES CONSTRUCTED FROM THE SUMMARY STATISTICS DATA IN TABLE 5
// Expected soil excavation area range, assuming 10% halo
        Area = [142700, 156970] // interval estimate of area, ft^2
// Expected soil excavation depth from the summary statistics data in Table 5
        Dmin = 0        // minimum depth, ft
        Dmax = 9        // maximum depth, ft
        Dmean = 4.61    // mean depth, ft
        Stddev = 3.41 // standard deviation, ft
// Probability bounds on depth is constructed from the above summary statistics data using the
//Risk Calc Version 4.0 function: minmaxmeanstddev (minimum, maximum, mean, standard
//deviation)
        Depth = minmaxmeanstddev(Dmin, Dmax, Dmean, Stddev)

// Calculation of the expected soil volume
        Volume = (Area * Depth)/27    // soil excavation area in cu. yds.

// Displaying results
        Volume      // displays min, max, mean, and, variance for soil volume
  ~(range=[0,52323.3],  mean=[23463,27703],  var=[4e+07,6e+08])

        cut(Volume, 95%)  // displays the 95th percentile soil volumes
  [ 36386.22, 52323.34]
```

Appendix C: PBA Soil Volume Estimate Using Summary Statistics Data – Paoli Rail Yard

```
// DISTRIBUTION-FREE MODELING FOR THE PAOLI RAIL YARD SITE USING
// p-BOXES CONSTRUCTED FROM THE SUMMARY STATISTICS IN TABLE 6

// Throat & East Car Shop area soil volume calculation
        Area1 =[177500, 195250]   // expected excavation area at this location, ft^2
        Dmin1 = 0          // minimum excavation depth, ft
        Dmax1 = 16.5    // maximum excavation depth, ft
        Dmean1 = 3.5    // mean excavation depth, ft
        Dstddev1 = 5.31       // standard deviation of excavation depth, ft
// Probability bounds on depth at each location is constructed from the above summary statistics
//data using the Risk Calc function: minmaxmeanstddev(minimum, maximum, mean, standard
//deviation)
        Depth1 = minmaxmeanstddev(Dmin1, Dmax1, Dmean1, Dstddev1)

// Calculation of soil volume in the Throat and East Car Shop area
        Volume1 = (Area1 * Depth1)/27      // volume in cubic yards

// South & West Car Shop area soil volume calculation
        Area2 = [119800, 131780]     // expected excavation area at this location, ft^2
        Dmin2 = 0         // minimum excavation depth, ft
        Dmax2 = 30.5   // maximum excavation depth, ft
        Dmean2 = 12.57            // mean excavation depth, ft
        Dstddev2 = 11.05          // Standard deviation of excavation depth, ft
// Calculation of excavation depth in the South & West Car Shop area
```

```
        Depth2 = minmaxmeanstddev(Dmin2, Dmax2, Dmean2, Dstddev2)


// Calculation of soil volume in the South & West Car Shop area

        Volume2 = (Area2 * Depth2)/27  // volume in cubic yards


// East Storage Yard soil volume calculation

        Area3 = [189300, 208230]          // expected excavation area, ft^2

        Depth3 = [0, 1]              // expected excavation depth, ft


        Volume3 = (Area3 * Depth3)/27  // volume in cubic yards


// Turnaround Track soil volume calculation

        Area4 = [91250, 100672]           // expected excavation area, ft^2

        Depth4 = [0, 1]              // expected excavation depth, ft


        Volume4 = (Area4 * Depth4)/27  // volume in cubic yards


// Calculation of total site soil excavation volume

        Volume = Volume1 + Volume2 + Volume3 + Volume4      // Total site volume, cu. yds.


// Displaying results

        Volume             // displays min, max, mean, variance for soil volume

  ~(range=[0,279623],  mean=[74585,102299],  var=[0,1e+10])

        cut(Volume, 95%)  // displays the 95th percentile soil volumes

  [ 92662.35, 279622.9]
```

Appendix D: PBA Soil Volume Estimate Using Summary Statistics Data – Peoples Natural Gas

```
// DISTRIBUTION-FREE MODELING FOR PEOPLES NATURAL GAS SITE USING
// p-BOXES CONSTRUCTED FROM THE SUMMARY STATISTICS IN TABLE 7

// Expected soil excavation area range, assuming 10% halo
        Area = [127602, 140362]     // interval estimate, ft^2
// Expected soil excavation depth summary statistics data from Table 7
        Dmin = 0     // minimum depth, ft
        Dmax = 10   // maximum depth, ft
        Dmean = 2.24   // mean depth, ft
        Dstddev = 3.99   // standard deviation, ft

// Probability bounds on depth is constructed from the above summary statistics data using the
//Risk Calc function: minmaxmeanstddev(minimum, maximum, mean, stddev)
        Depth = minmaxmeanstddev(Dmin, Dmax, Dmean, Dstddev)

// Calculation of expected soil volume
        Volume = (Area * Depth)/27   // volume in cubic yards

// Displaying results
        Volume      // Displays min, max, mean, variance for soil volume
  ~(range=[0,51985.9],  mean=[9643,12588],  var=[1e+08,4e+08])
        cut(Volume, 95%)     // displays the 95th percentile soil volumes
  [ 35976.76, 51985.93]
```

Appendix E: PBA Soil Volume Estimate Using Summary Statistics Data – Taylor Lumber & Treating

```
// DISTRIBUTION-FREE MODELING FOR THE TAYLOR LUMBER & TREATING SITE

// USING p-BOXES CONSTRUCTED FROM THE SUMMARY STATISTICS DATA IN TABLE 8

// Expected soil excavation area range, assuming 10% halo

        Area = [171191, 188310] // interval estimate of area, ft^2

// Expected soil excavation depth from the summary statistics data in Table 8

        Dmin = 0        // minimum depth, ft

        Dmax = 6      // maximum depth, ft

        Dmean = 1.91   // mean depth, ft

        Stddev = 2.02   // standard deviation, ft


// Probability bounds on depth is constructed from the above summary statistics data using the

//Risk Calc Version 4.0 function: minmaxmeanstddev (minimum, maximum, mean, standard

//deviation)

        Depth = minmaxmeanstddev(Dmin, Dmax, Dmean, Stddev)


// Calculation of the expected soil volume

        Volume = (Area * Depth)/27    // soil excavation area in cu. yds.


// Displaying results

        Volume      // displays min, max, mean, and, variance for soil volume

  ~(range=[0,41846.7],  mean=[11469,13962],  var=[1e+07,3e+08])

        cut(Volume, 95%)  // displays the 95th percentile soil volumes

  [ 22778.42, 41846.67]
```

Appendix F: PBA Soil Volume Estimate Using Summary Statistics Data – Tonolli Corporation

```
// DISTRIBUTION-FREE MODELING FOR THE TONOLLI CORPORATION SITE

// USING p-BOXES CONSTRUCTED FROM THE SUMMARY STATISTICS DATA IN TABLE 9

// Expected soil excavation area range, assuming 10% halo

        Area = [540000, 594000] // interval estimate of area, ft^2

// Expected soil excavation depth from the summary statistics data in Table 9

        Dmin = 0        // minimum depth, ft

        Dmax = 10     //   maximum depth, ft

        Dmean = 2.33   // mean depth, ft

        Stddev = 3.44   // standard deviation, ft


// Probability bounds on depth is constructed from the above summary statistics data using the

//Risk Calc Version 4.0 function: minmaxmeanstddev (minimum, maximum, mean, standard

//deviation)

        Depth = minmaxmeanstddev(Dmin, Dmax, Dmean, Stddev)


// Calculation of the expected soil volume

        Volume = (Area * Depth)/27    //   soil excavation area in cu. yds.


// Displaying results

        Volume      // displays min, max, mean, and, variance for soil volume
```

~(range=[0,220000], mean=[43160,54700], var=[8e+08,9e+09])

```
        cut(Volume, 95%)  // displays the 95th percentile soil volumes
```

[123547.7, 220000]

Appendix G: PBA Soil Volume Estimate Using Summary Statistics Data – U. S. Titanium Corporation

```
// DISTRIBUTION-FREE MODELING FOR THE U. S. TITANIUM CORPORATION SITE

// USING p-BOXES CONSTRUCTED FROM THE SUMMARY STATISTICS DATA IN TABLE 10

// Expected soil excavation area range, assuming 10% halo

        Area = [87120, 95832] // interval estimate of area, ft^2

// Expected soil excavation depth from the summary statistics data in Table 10

        Dmin = 0        // minimum depth, ft

        Dmax = 24      //   maximum depth, ft

        Dmean = 8.69   // mean depth, ft

        Stddev = 8.49   // standard deviation, ft


// Probability bounds on depth is constructed from the above summary statistics data using the

//Risk Calc Version 4.0 function: minmaxmeanstddev (minimum, maximum, mean, standard

//deviation)

        Depth = minmaxmeanstddev(Dmin, Dmax, Dmean, Stddev)


// Calculation of the expected soil volume

        Volume = (Area * Depth)/27    //   soil excavation area in cu. yds.


// Displaying results

        Volume      // displays min, max, mean, and, variance for soil volume

  ~(range=[0,85184],  mean=[26670,32214],  var=[8e+07,1e+09])

        cut(Volume, 95%)  // displays the 95th percentile soil volumes

  [ 50307.6, 85184]
```

Appendix H: PBA Soil Volume Estimate Using Percentile Distributions – Aladdin Plating Site

```
// DISTRIBUTION-FREE MODELING FOR THE ALADDIN PLATING COMPANY SITE

// USING p-BOXES CONSTRUCTED FROM THE PERCENTILE DATA IN TABLE 4

// Expected soil excavation area range, assuming 10% halo

        Area = [363500, 399850]   // interval estimate of area, ft^2

// Expected soil excavation depth from the summary statistics data in Table 4

        Dmin = 1.5        // minimum depth, ft

        D25 = 2.5         // 25th percentile depth, ft

        D50 = 2.5         // median (50th percentile) depth, ft

        D75 = 3           // 75th percentile depth, ft

        Dmax = 6          // maximum depth, ft.


// Probability bounds on depth is constructed from the above percentiles data using the Risk Calc

// Version 4.0 function: fivenumbers(minimum, 25th percentile, median, 75th percentile,

//maximum)

        Depth = fivenumbers(Dmin, D25, D50, D75, Dmax)


// Calculation of the expected soil volume

        Volume = (Area * Depth)/27    //  soil excavation area in cu. yds.


// Displaying results

        Volume     // displays min, max, mean, and, variance for soil volume

  ~(range=[20194.4,88855.6],  mean=[31772,51833],  var=[2832052,7e+08])

        cut(Volume, 95%)          // Displays 95th percentile result.

  [ 40388.88, 88855.56]
```

Appendix I: PBA Soil Volume Estimate Using Percentile Distributions – C&R Battery

```
// DISTRIBUTION-FREE MODELING FOR THE C & R BATTERY COMPANY SITE

// USING p-BOXES CONSTRUCTED FROM THE PERCENTILE DATA IN TABLE 5

// Expected soil excavation area range, assuming 10% halo

        Area = [142700, 156970]   // interval estimate of area, ft^2

// Expected soil excavation depth from the summary statistics data in Table 4

        Dmin = 0        // minimum depth, ft

        D25 = 1.5       // 25th percentile depth, ft

        D50 = 6         // median (50th percentile) depth, ft

        D75 = 7.5       // 75th percentile depth, ft

        Dmax = 9        // maximum depth, ft.


// Probability bounds on depth is constructed from the above percentiles data using the Risk Calc

// Version 4.0 function: fivenumbers(minimum, 25th percentile, median, 75th percentile,

//maximum)

        Depth = fivenumbers(Dmin, D25, D50, D75, Dmax)


// Calculation of the expected soil volume

        Volume = (Area * Depth)/27    //  soil excavation area in cu. yds.


// Displaying results

        Volume     // displays min, max, mean, and, variance for soil volume

  ~(range=[0,52323.3],  mean=[19423,34883],  var=[1e+08,5e+08])

        cut(Volume, 95%)        // displays the 95th percentile soil volumes

  [ 39638.88, 52323.34]
```

Appendix J: PBA Soil Volume Estimate Using Percentile Distributions – Paoli Rail Yard

```
// DISTRIBUTION-FREE MODELING FOR THE PAOLI RAIL YARD SITE USING
// p-BOXES CONSTRUCTED FROM THE PERCENTILES IN TABLE 6


// Throat & East Car Shop area soil volume calculation
        Area1 = [177500, 195250]   // expected excavation area at this location, ft^2
        Dmin1 = 0          // minimum excavation depth, ft
        D25A = 0           // 25th percentile depth, ft
        D50A = 1           // median (50th percentile depth), ft
        D75A = 4.5         // 75th percentile depth, ft
        Dmax1 = 16.5       // maximum depth, ft
// Probability bounds on depth at each location is constructed from the above summary statistics
//data using  the Risk Calc function: fivenumbers(minimum, 25th percentile, median, 50th
//percentile, maximum)
        Depth1 = fivenumbers(Dmin1, D25A, D50A, D75A, Dmax1)


// Calculation of soil volume in the Throat and East Car Shop area
        Volume1 = (Area1 * Depth1)/27      // volume in cubic yards


// South & West Car Shop area soil volume calculation
        Area2 = [119800, 131780]     // expected excavation area at this location, ft^2
        Dmin2 = 0          // minimum excavation depth, ft
        D25B = 2.5         // 25th percentile depth, ft
        D50B = 14.5        // median (50th percentile depth), ft
        D75B = 21.5        // 75th percentile depth, ft
```

```
        Dmax2 = 30.5    // maximum depth, ft

// Calculation of excavation depth in the South & West Car Shop area

        Depth2 = fivenumbers(Dmin2, D25B, D50B, D75B, Dmax2)


// Calculation of soil volume in the South & West Car Shop area

        Volume2 = (Area2 * Depth2)/27  // volume in cubic yards


// East Storage Yard soil volume calculation

        Area3 = [189300, 208230]       // expected excavation area, ft^2

        Depth3 = [0, 1]                // expected excavation depth, ft

        Volume3 = (Area3 * Depth3)/27  // volume in cubic yards


// Turnaround Track soil volume calculation

        Area4 = [91250, 100672]        // expected excavation area, ft^2

        Depth4 = [0, 1]                // expected excavation depth, ft

        Volume4 = (Area4 * Depth4)/27  // volume in cubic yards


// Calculation of total site soil excavation volume

        Volume = Volume1 + Volume2 + Volume3 + Volume4      // Total site volume, cu. yds.


// Displaying results

        Volume          // displays min, max, mean, variance for soil volume

  ~(range=[0,279623],  mean=[50496,135407],  var=[0,1e+10])

        cut(Volume, 95%)  // // displays the 95th percentile soil volumes

  [ 95396.29, 279622.9]
```

Appendix K: PBA Soil Volume Estimate Using Percentile Distributions – Peoples Natural Gas

```
// DISTRIBUTION-FREE MODELING FOR THE PEOPLES NATURAL GAS SITE

// USING p-BOXES CONSTRUCTED FROM THE PERCENTILE DATA IN TABLE 7

// Expected soil excavation area range, assuming 10% halo

        Area = [127602, 140362]   // interval estimate of area, ft^2

// Expected soil excavation depth from the summary statistics data in Table 4

        Dmin = 0          // minimum depth, ft

        D25 = 0           // 25th percentile depth, ft

        D50 = 0           // median (50th percentile) depth, ft

        D75 = 2           // 75th percentile depth, ft

        Dmax = 10         // maximum depth, ft.

// Probability bounds on depth is constructed from the above percentiles data using the Risk Calc

// Version 4.0 function: fivenumbers(minimum, 25th percentile, median, 75th percentile,

//maximum)

        Depth = fivenumbers(Dmin, D25, D50, D75, Dmax)


// Calculation of the expected soil volume

        Volume = (Area * Depth)/27    //  soil excavation area in cu. yds.


// Displaying results

        Volume      // displays min, max, mean, and, variance for soil volume

  ~(range=[0,51985.9],  mean=[2268,15596],  var=[1e+07,5e+08])

        cut(Volume, 95%)   // displays the 95th percentile soil volumes

  [ 9452, 51985.93]
```

Appendix L: PBA Soil Volume Estimate Using Percentile Distributions – Taylor Lumber & Treating

```
// DISTRIBUTION-FREE MODELING FOR THE TAYLOR LUMBER & TREATING SITE

// USING p-BOXES CONSTRUCTED FROM THE PERCENTILE DATA IN TABLE 8


// Expected soil excavation area range, assuming 10% halo

        Area = [171191, 188310]   // interval estimate of area, ft^2

// Expected soil excavation depth from the summary statistics data in Table 4

        Dmin = 0          // minimum depth, ft

        D25 = 0           // 25th percentile depth, ft

        D50 = 2           // median (50th percentile) depth, ft

        D75 = 2           // 75th percentile depth, ft

        Dmax = 6          // maximum depth, ft.

// Probability bounds on depth is constructed from the above percentiles data using the Risk Calc

// Version 4.0 function: fivenumbers(minimum, 25th percentile, median, 75th percentile,

//maximum)

        Depth = fivenumbers(Dmin, D25, D50, D75, Dmax)


// Calculation of the expected soil volume

        Volume = (Area * Depth)/27    //   soil excavation area in cu. yds.


// Displaying results

        Volume      // displays min, max, mean, and, variance for soil volume

   ~(range=[0,41846.7],  mean=[6213,17437],  var=[3e+07,2e+08])

        cut(Volume, 95%)   // displays the 95th percentile soil volumes

   [ 12680.81, 41846.67]
```

Appendix M: PBA Soil Volume Estimate Using Percentile Distributions – Tonolli Corporation

```
// DISTRIBUTION-FREE MODELING FOR THE TONOLLI CORPORATION SITE

// USING p-BOXES CONSTRUCTED FROM THE PERCENTILE DATA IN TABLE 9

// Expected soil excavation area range, assuming 10% halo

        Area = [540000, 594000]   // interval estimate of area, ft^2


// Expected soil excavation depth from the summary statistics data in Table 4

        Dmin = 0        // minimum depth, ft

        D25 = 0         // 25th percentile depth, ft

        D50 = 0         // median (50th percentile) depth, ft

        D75 = 5         // 75th percentile depth, ft

        Dmax = 10       // maximum depth, ft.

// Probability bounds on depth is constructed from the above percentiles data using the Risk Calc

// Version 4.0 function: fivenumbers(minimum, 25th percentile, median, 75th percentile,

//maximum)

        Depth = fivenumbers(Dmin, D25, D50, D75, Dmax)


// Calculation of the expected soil volume

        Volume = (Area * Depth)/27    //   soil excavation area in cu. yds.


// Displaying results

        Volume     // displays min, max, mean, and, variance for soil volume

  ~(range=[0,220000],  mean=[24000,82500],  var=[1e+09,9e+09])

        cut(Volume, 95%)   // displays the 95th percentile soil volumes

  [ 100000, 220000]
```

Appendix N: PBA Soil Volume Estimate Using Percentile Distributions – U.S. Titanium Corporation

```
// DISTRIBUTION-FREE MODELING FOR THE U. S. TITANIUM SITE

// USING p-BOXES CONSTRUCTED FROM THE PERCENTILE DATA IN TABLE 10

// Expected soil excavation area range, assuming 10% halo

        Area = [87120, 95832]   // interval estimate of area, ft^2


// Expected soil excavation depth from the summary statistics data in Table 4

        Dmin = 0          // minimum depth, ft

        D25 = 0           // 25th percentile depth, ft

        D50 = 8.25        // median (50th percentile) depth, ft

        D75 = 13.5        // 75th percentile depth, ft

        Dmax = 24         // maximum depth, ft.

// Probability bounds on depth is constructed from the above percentiles data using the Risk Calc

// Version 4.0 function: fivenumbers(minimum, 25th percentile, median, 75th percentile,

//maximum)

        Depth = fivenumbers(Dmin, D25, D50, D75, Dmax)


// Calculation of the expected soil volume

        Volume = (Area * Depth)/27    //  soil excavation area in cu. yds.


// Displaying results

        Volume      // displays min, max, mean, and, variance for soil volume

  ~(range=[0,85184],  mean=[17109,40596],  var=[2e+08,1e+09])

        cut(Volume, 95%)   // displays the 95th percentile soil volumes

  [ 43560, 85184]
```

Appendix O: Hypothesis Test for Correlation - Best Regression Outcome

The best regression outcome was obtained using the lower bound 95th percentile results of PBA with summary statistics data for depths.

Let Ω = population coefficient of correlation

The hypothesis to be tested is as follows:

$H_0 : \Omega = 0$ (no correlation)

$H_1 : \Omega \neq 0$ (correlation exists)

Based on a two-tailed probability test, a 5% significance level, 5 degrees of freedom, and the Student's t-distribution, the decision rules are as follows:

Accept H_0: if $-2.571 < t < 2.571$

Reject H_0: if $t \leq -2.571$ or if $t \geq 2.571$

The data requirements for this hypothesis test were obtained from Table 25 as follows:

$r = 0.943$

$r^2 = 0.890$

$n = 7$

$t = 0.943\sqrt{[(7-2)/(1-.890)]}$

$t = 6.358$

Based on the decision rules the null hypothesis is rejected leading to the conclusion that correlation exists between the 95th percentile lower bound PBA estimates with summary statistics data for depths and the actual excavated soil volumes.

Appendix P: Hypothesis Test for Correlation – Second Best Regression Outcome

The second-best regression outcome was obtained using the lower bound 95th percentile results of PBA with percentile distribution data for excavation depths.

Let Ω = population coefficient of correlation

The hypothesis to be tested is as follows:

$H_0 : \Omega = 0$ (no correlation)

$H_1 : \Omega \neq 0$ (correlation exists)

Based on a two-tailed probability test, a 5% significance level, 5 degrees of freedom, and the Student's t-distribution, the decision rules are as follows:

Accept $H_{0:}$ if $-2.571 < t < 2.571$

Reject H_0: if $t \leq -2.571$ or if $t \geq 2.571$

The data requirements for this hypothesis test were obtained from Table 25 as follows:

$r = 0.940$

$r^2 = 0.884$

$n = 7$

$t = 0.940\sqrt{[(7-2)/(1-.884)]}$

$t = 6.171$

Based on the decision rules the null hypothesis is rejected leading to the conclusion that correlation exists between the 95th percentile lower bound PBA estimates with summary statistics data for depths and the actual excavated soil volumes.

Curriculum Vitae

Charles O. Dankwah
6193 Deer Ridge Trail
Springfield, VA 22150
703-922-6796

EDUCATION:

B.S. Chemical Technology, University of Science and Technology, Kumasi, Ghana, 1974.
M.S. Energy & Environmental Systems, The George Washington University, Washington, DC, 1986.

EXPERIENCE:

Sept. 1974 – July 1975 **Ashanti Goldfields Corporation, Obuasi, Ghana**. Production Supervisor

- Controlled production processes to ensure maximum extraction of gold from ores.
- Supervised production workers.

Aug. 1975 – June 1982 **Shell Oil Ghana Limited, Accra, Ghana**. Technical Sales Executive

- Provided advice to customers on the choice of industrial lubricants for various machines.
- Monitored quality of fuels and lubricants from supply sources to ensure they were on specification.

Aug. 1982 – Dec. 1992 **Technology Applications, Inc., Alexandria, VA**. Senior Chemical Engineer

- Provided environmental modeling support to the United States Environmental Protection Agency for the Land Disposal Restrictions Rule.
- Provided environmental modeling support to the United States Environmental Protection Agency for the Toxicity Characteristics Rule.
- Provided support to the U.S.EPA in developing the U.S. EPA Composite Model for Landfills (EPACML)

Mar. 1993 – Nov. 2001 **DPRA, Inc., Rosslyn, VA.** Senior Analyst

- Provided statistical support to attorneys in evaluating hazardous waste site remediation cost claims.
- Provided EPACML modeling support to the U.S. EPA

Dec. 2001 – Present **Hawknad Manufacturing Industries, Inc., Alexandria, VA**. President

- Supervises compounding of skin care products
- Manages daily activities of the company
- Manages the research and development activities of the company.