The Impact of Below Detection Limit Samples in (NPP) Decommissioning Residual Risk Assessment - 17033

Jung Hwan Kim*, Carol Hornibrook*, Man-Sung Yim*

Korea Advanced Institute of Science and Technology

ABSTRACT

There are some technical problems with finalizing regulations and standards that address nuclear power plant (NPP) decommissioning site soil analyses for license termination. It is difficult to know how to best include detection limit observations from site soil analysis, because typical soil collection data is reported above, at, and below the detection limit. In addition, multiple detection limits are required to adequately estimate the soil radionuclide concentrations. In the literature, researchers analyzing environmental data refer to below the detection limit observations as “left-censored data”. Left-censored data are less than the detection limit of the instrument(s) that made the measurements. Conventional approaches to managing left-censored data either ignore or simply replace the detection limits with zero, a fraction of the detection limit, or the detection limit itself. However, these approaches are statistically biased and limited in their usefulness. The Kaplan-Meier method, robust regression on order statistics method, and maximum likelihood estimation method are proposed to estimate the left-censored data more precisely. The proposed methods are applied to data from a Monazite powder manufacturing plant and the Colorado School of Mines Research Institute. Summary statistics such as the sample mean, sample standard deviation, sample percentiles are calculated from each data set and are used in risk and volume estimation calculations. Previous environmental analyses, such as the exposure assessment, recognize the uncertainty issue that is addressed in this research using the Latin hypercube sampling approach. Next, probabilistic distributions of the input parameters are developed for the uncertainty analyses.

Finally, a framework methodology depicting the decision-making process in the proposed method is presented. The framework defines the critical steps, the amount of radioactivity, types of distributions, censoring percentage, sample size, and number of detection limits; it also allows the user to select the appropriate approach based on their site-specific data and analyses. The goal of this research is to develop a more precise risk assessment, estimate of the volume of soil removal, examine potential cost savings.

KEYWORDS: Decommissioning, Below detection limit, Censored data, Kaplan-Meier method, Robust regression on order statistics method, maximum likelihood estimation method, Risk assessment, Uncertainty analysis, Decision-making framework

INTRODUCTION

Decommissioning is an emerging international issue in the nuclear industry. The term ‘decommissioning’ refers to the administrative and technical actions taken in order to allow the removal of some or all regulatory controls (e.g. operating license) from a nuclear facility [1]. In order to terminate an NPP license, the owner must demonstrate compliance with regulatory controls for restricted or unrestricted future use of the site. However, some technical issues associated with finalizing regulations related to the site release remain. For example, the representativeness of the contaminated soil samples related to hot spots, the lack of information on the radionuclide distribution, and observations reported as below detection limits (BDL). The techniques currently used for correctly assessing hot spots and the availability of radionuclide distribution are taken directly from the literature.

The regulatory guidelines for site reuse after decommissioning are commonly challenged because the majority of the activity in the soil is at or below the limit of detection [3]. Observations reported as below the detection limits are caused by the inherent limitations of the measurement methods, i.e. detectors have detection limits. For example, environmental data collected by the National Human Exposure Assessment Survey database, show that 30% to 70% of observations are below the detection limits for many pollutants [4]. Multiple detection limits arise from different sampling procedures or different sampling volumes. Although a detection limit (DL) might be insignificantly low, it is dangerous to ignore DLs because the dose is the result of functions such as dose conversion factors, daughter nuclides, type of radionuclide, etc. If the data are not treated correctly, there can be a significant affect, usually an overestimation of the health risk to the public and overestimation of the volume of soil removal and associated costs.

Conventional methods currently used for analyzing data below the detection limits are either ignored or simply replace the detection limit with zero, a fraction of the detection limit, or the detection limit itself. These approaches are statistically biased [5]. For example, ignoring or replacing the data with a DL overestimates the mean and replacing it with zero underestimates the mean. In order to resolve these issues, statistical techniques were evaluated.

The three methods used to estimate the summary statistics (e.g. mean, standard deviation) are the Kaplan-Meier method, robust regression on order statistics (ROS) method, and maximum likelihood estimation (MLE) method. These techniques are used by numerous researchers in environmental science and technology, but have not been widely used in nuclear evaluations of risk [6].

Analyzing the long term environmental health risk and costs can be significantly affected by relatively small variations in the mean value. The small variation in the mean value can affect the dose, cancer risk, and volume of soil removal. To identify the impact of including censored data (nondetects), it is necessary to calculate the dose, cancer risk, and volume of soil removal with and without censored data values.

To release a site from regulatory control, radioactive contamination should be removed, to the accepted regulatory level, reducing the potential for negative health effects. It is necessary to determine the amount of soil cleanup required to achieve the cleanup goals. The maximum total dose should be reduced to 0.15mSv/yr for unrestricted use in the future. The RESRAD code helps to estimate the volume of soil to be remediated. The goal of this research is to develop more precise methods for risk assessment, estimation of volume of soil removal, and examination of cost savings.

Without a basic understanding of uncertainty, it is difficult to appreciate how and why site specific soil data analysis techniques are required. Analyses of environmental issues such as exposure assessments are related to uncertainty [7]. Although only limited data is available for analysis, regulatory decisions should be based on the entire data set. When the data set is limited, some model parameter uncertainties can be represented using probability density functions [8]. The number of model input parameters are simulated simultaneously in order to determine their combined effect on the model outputs. Latin hypercube sampling has been used for this type of uncertainty analysis in probabilistic risk assessment.

In order to compare the strengths of the three statistical methods investigated in this research, errors between the actual mean and the statistical methods are estimated. These errors form the basis for selecting the appropriate statistical approaches. The smaller the error, the better matched the statistical approach is to the data. Each analysis addresses changes in the amount of radioactivity, types of distributions, censoring percentages, and numbers of detection limits used.

METHODS, RESULTS and DISCUSSION

Conventional methods

In order to illustrate the bias and limited usefulness of conventional methods, 24 arsenic concentrations (in µg/L) from the urban stream waters in the Oahu data set are analyzed [6]. The arsenic concentrations were 0.5, 0.5, 0.5, 0.6, 0.7, 0.7, <0.9, 0.9, <1.0, <1.0, <1.0, <1.0, 1.5, 1.7, <2.0, <2.0, <2.0, <2.0, <2.0, <2.0, <2.0, <2.0, 2.8, 3.2. The results of using 0, ½ DL, DL, and ignoring the BDL values are presented in Table I.

Table I. Results of ignoring or replacing the left--censored data in the Oahu data.

Value substituted	Mean	Standard deviation	Percentile 25	Median	Percentile 75
0	0.567	0.895	0.000	0.000	0.700
½ DL	1.002	0.699	0.500	0.950	1.000
DL	1.438	0.761	0.750	1.250	2.000
Ignoring	1.236	0.920	0.500	0.700	1.700

As seen in Table I, the mean values of the Oahu data do not appear to be significantly different, with the greatest difference being between using a 0 value and using the DL (0.567 vs. 1.438, respectively). However, analyzing the long term environmental health risk and costs can be significantly affected by this relatively small variation in the mean value.

In order to address these issues, it is important to identify alternative approaches that can more precisely represent the actual concentrations of activity in the soil and can be effectively verified. The following proposed methods were taken from research conducted in the environmental field specifically addressing the concentrations of hazardous materials in the soil and atmosphere.

Proposed methods

Kaplan-Meier (KM) method

The Kaplan-Meier method is a nonparametric technique. Nonparametric techniques describe data that do not follow a specific parametric distribution such as a normal, lognormal, or Weibull distributions. The Kaplan-Meier (KM) method calculates the probability distribution using censored data, to estimate the summary statistics such as the mean, the standard deviation, the percentiles, etc.,. Because it is a nonparametric technique, it is well-suited for many environmental data sets [10].

The following steps are recommended for applying KM method to estimate the summary statistics [11,17].

Step 1. From the sample of n=24 Oahu data points, sort the 11 detected values in decreasing order and compute the rank of each data point including nondetects.

Step 2. Compute the # at risk (*b*) for each detected values and find the # of detects. The # at risk *b* can be calculated with the total number of data points *n* and the rank *r*. It is expressed as *b = (n–r+1)*.

Step 3. Incremental survival probability is calculated using the # at risk *b* and the # of detects *d*. The incremental survival probability *P* can be expressed by *P = (b–d)/b*.

Step 4. The survival function probability *S* can be calculated using the incremental survival probabilities, working from high to low data for the k detected observations.

$$S = \prod_{j=1}^{k} \frac{b_j - d_j}{b_j} \qquad \text{(Eq. 1)}$$

Table II. Computation of the summary statistics using the Kaplan-Meier method for the Oahu data.

Mean	Standard deviation	Percentile 25	Median	Percentile 75
0.949	0.807	0.500	0.700	0.900

Robust regression on order statistics (ROS) method

The ROS method is a semi-parametric method that can be used to estimate summary statistics with censored data. The ROS method uses detected values to develop a probability plot and estimate the parameters using a regression line.

The following steps are recommended for applying ROS method to estimate the summary statistics [11,17].

Step 1. For a sample of n=24 Oahu data points, sort the 24 data points in decreasing order including nondetects.

Step 2. Calculate the probability of exceeding each detection limit.

In general, the probability of exceeding the j^{th} detection limit is,

$$pe_i = pe_{j+1} + \frac{A_j}{A_j + B_j}[1 - pe_{j+1}] \qquad \text{(Eq. 2)}$$

Where

A_j= the number of observations detected between the j^{th} and $(j+1)^{th}$ detection limits

B_j = the number of observations, censored and uncensored below the jth detection limit

The number of nondetects below the j^{th} detection limit is defined as C_j:

$$C_j = B_j - B_{j-1} - A_{j-1} \qquad \text{(Eq. 3)}$$

Step 3. Compute the plotting position percentiles to find z scores for each of the detected data points.

In general, plotting positions for observed values are

$$pd_i = (1 - pe_j) + \frac{i}{A_j + 1}[pe_j - pe_{j+1}] \qquad \text{(Eq. 4)}$$

And for censored observations are

$$pc_i = \left(\frac{i}{C_j + 1}\right) * [1 - pe_j] \qquad \text{(Eq. 5)}$$

Step 4. Plot the z scores against the detected values on a probability plot and find the matching distributions.

Step 5. Compute a linear regression of the detected value and estimate the regression parameters such as slope and intercept for detected values.

Step 6. Compute the plotting position percentiles to find z scores for each of the censored data points.

Step 7. Using the regression parameters estimated in Step 5, estimate the regression parameters such as slope and intercept for censored data.

Step 8. Combine the censored data with the detected values and estimate the mean and standard deviation using estimated concentrations.

Table III. Computation of the summary statistics using the ROS method for the Oahu data.

Mean	Standard deviation	Percentile 25	Median	Percentile 75
0.972	0.718	0.518	0.700	1.103

Maximum likelihood estimation (MLE) method

The maximum likelihood estimation (MLE) method is a parametric, model-based method that can be used to estimate summary statistics with censored data. Probability plots and other goodness-of-fit techniques are used to determine the matching distributions. Nondetects are distributed similarly to the detected values. The parametric MLE method assumes a distribution that will closely fit the observed data.

The following equations are recommended for applying the MLE method to estimate the summary statistics [6].

$$\mathrm{L}(\theta_1,\theta_2,\dots,\theta_k)=\prod_{i=1}^{n} f(\mathrm{x}_i|\theta_1,\theta_2,\dots,\theta_k)\{\prod_{m=1}^{p}(\prod_{j=1}^{ND_m} F(\mathrm{DL}_m|\theta_1,\theta_2,\dots,\theta_k))\} \qquad \text{(Eq. 6)}$$

Where,

x_i= Detected data point, where, i=1, 2, ..., n

$\theta_1,\theta_2,\dots,\theta_k$= Parameters of the distribution

ND_m= Number of nondetects corresponding to DL_m, where, m=1, 2, ..., P.

P = Number of detection limits

f = Probability density function

F = Cumulative distribution function

Table IV. Computation of the summary statistics using the MLE method for the Oahu data.

Mean	Standard deviation	Percentile 25	Median	Percentile 75
0.9453	0.6559	0.5088	0.7766	1.1854

Bootstrap simulation

R code was used to generate 20 bootstrap samples, each of size 10. Each of the 20 columns in the following array is one bootstrap sample

```
43 36 46 30 43 43 43 37 42 42 43 37 36 42 43 43 42 43 42 43
43 41 37 37 43 43 46 36 41 43 43 42 41 43 46 36 43 43 43 42
42 43 37 43 46 37 36 41 36 43 41 36 37 30 46 46 42 36 36 43
37 42 43 41 41 42 36 42 42 43 42 43 41 43 36 43 43 41 42 46
42 36 43 43 42 37 42 42 42 46 30 43 36 43 43 42 37 36 42 30
36 36 42 42 36 36 43 41 30 42 37 43 41 41 43 43 42 46 43 37
43 37 41 43 41 42 43 46 46 36 43 42 43 30 41 46 43 46 30 43
41 42 30 42 37 43 43 42 43 43 46 43 30 42 30 42 30 43 43 42
46 42 42 43 41 42 30 37 30 42 43 42 43 37 37 37 42 43 43 46
42 43 43 41 42 36 43 30 37 43 42 43 41 36 37 41 43 42 43 43
```

Now, find $\bar{x}*$ (mean) for each bootstrap sample:

41.5 39.8 40.4 40.5 41.2 40.1 40.5 39.4 38.9 42.3 41.0 41.4 38.9 38.7 40.2 41.9

40.7 41.9 40.7 41.5. Next, compute $\delta^* = \bar{x}^* - \bar{x}$ for each bootstrap sample and arrange them from the smallest to the biggest:
-1.6 -1.4 -1.4 -0.9 -0.5 -0.2 -0.1 0.1 0.2 0.2 0.4 0.4 0.7 0.6 1.1 1.2 1.2 1.6 1.6 2.0
The critical values $\delta_{.1}$ and $\delta_{.9}$ can be approximated by $\delta_{.1}^*$ and $\delta_{.9}^*$. Since $\delta.1^*$ is at the 90th percentile, the 18th value in the list selected, i.e. 1.6. Similarly, sine $\delta_{.9}^*$ is at the 10th percentile the 2nd element in the list selected, i.e. -1.4.
Therefore, bootstrap 80% confidence interval for μ is represented by the following equation:
$[\bar{x} - \delta.1^*, \bar{x} - \delta.9^*] = [40.3 - 1.6, 40.3 + 1.4] = [38.7, 41.7]$

Software used for calculation of proposed methods

Statistical software is used to compute the estimates of the censored data. The three software packages used in this research are ProUCL, MATLAB, and R.

ProUCL is statistical software for environmental applications for data sets with and without nondetect observations; it is provided by the US Environmental Protection Agency (EPA) [11]. ProUCL assists in computing the upper confidence limit of the population mean based on left-censored data sets containing nondetect observations. It includes goodness-of-fit (GOF) tests and skewness for left-censored data sets. The nonparametric Kaplan-Meier (KM), robust regression on order statistics (ROS), and maximum likelihood estimation (MLE) methods are used to estimate the summary statistics including the mean, standard deviation, and percentiles. Whereas ProUCL is used to develop Box plots, histograms, and Q-Q plots.

MATLAB and R are programming languages and environments for statistical computing and generating graphics. These languages include effective data handling and storage. It is possible to estimate the parameters of each distribution using MATLAB. Using R’s “Nondetect And Data Analysis (NADA)” package, summary statistics for KM, ROS, and MLE methods are estimated.

Preliminary evaluation of proposed methods

Environmental data set benchmark

To benchmarking the environmental data set, each method is applied using the manganese groundwater concentrations data set. The manganese groundwater concentration data for each well were collected and are presented below [10].

Table V. Manganese groundwater concentrations data for each well.

Sample	Manganese concentrations (ppb) in background				
	Well 1	Well 2	Well 3	Well 4	Well 5
1	< 5.0	< 5.0	< 5.0	6.3	17.9
2	12.1	7.7	5.3	11.9	22.7
3	16.9	53.6	12.6	10.0	3.3
4	21.6	9.5	106.3	< 2.0	8.4
5	< 2.0	45.9	34.5	77.2	< 2.0

Two detection limits are listed (< 5.0 and< 2.0), along with detected observations. The manganese groundwater concentration data is analyzed and compared with the EPA reference A summary of the results using the conventional methods and the proposed methods are presented in Table VI .

Table VI. Benchmark results for comparing the concentration means between the conventional methods and the proposed methods.

Statistical methods	Concentration mean	
	This Research	Reference

Conventional methods	Ignore	2.800 log (ppb)	-
	Zero	N/A	N/A
	1/2 DL	2.238 log (ppb)	-
	DL	2.404 log (ppb)	-
Proposed methods	KM	2.309 log (ppb)	2.31 log (ppb)
	ROS	2.277 log (ppb)	2.28 log (ppb)
	MLE	2.264 log (ppb)	-

Case study of calculation of the summary statistics in monazite powder manufacturing plant

There was a monazite powder manufacturing plant in the Republic of Korea. Some plant facilities and soil were contaminated during the manufacturing process. The facility soil survey was conducted and site soil was decontaminated [12]. The monazite powder manufacturing plant data was analyzed to verify the proposed methods. After the decommissioning soil samples, the representative radioactivity was determined through sampling analyses and the properties of the residues or suspicious material from the factory. From Grid Box 1 and Grid Box 2, it is possible to obtain data points of U-238 and K-40, which includes data points below the detection limits.

The survey consisted of Grid Box 1 and Grid Box 2, each with 30 data points measuring the concentration of U-238 and K-40, in the same area. A single detection limit is listed along with detected observations for Grid Box 1 and Grid Box 2. In Grid Box 1, 8 data points were below the detection limit for U-238 and 17 for K-40. In Grid Box 2, 10 data points were below the detection limit for U-238 and 9 for K-40. The ORTEC HPGe detector is utilized. The counting time was 1800 sec for all samples. ORTEC Gamma Vision 6.01 was used to analyze the data. The minimum detectable activity (MDA) of each radionuclide was 0.0953 Bq/g for U-238 and 0.0118 Bq/g for K-40 for Grid Box 1, and 0.109 Bq/g for U-238 and 0.0126 Bq/g for K-40 for Grid Box 2.

Table VII. Mean radioactivity results using several estimation methods for U-238 and K-40 of each grid box.

Methods		Radioactivity mean (Bq/g)			
		Grid Box 1		Grid Box 2	
		U-238	K-40	U-238	K-40
Conventional methods	Ignore	0.339	0.092	0.492	0.092
	Zero	0.249	0.040	0.328	0.064
	½ DL	0.261	0.043	0.346	0.066
	DL	0.274	0.046	0.364	0.068
Proposed methods	MLE	0.215	0.042	0.425	0.063
	ROS	0.268	0.048	0.342	0.067
	KM	0.274	0.046	0.364	0.068

Table VIII. Computation of the confidence intervals for the mean using MLE/Bootstrap.

Cases	90% confidence interval for the mean (unit: Bq/g)	95% confidence interval for the mean (unit: Bq/g)
U-238 in Grid Box 1	[0.106, 0.336]	[0.068, 0.348]

(Mean:0.263 Bq/g)		
U-238 in Grid Box 2 (Mean:0.353 Bq/g)	[0.140, 0.448]	[0.086, 0.465]

MLE/Bootstrap method is used to quantify uncertainty for the mean of censored data sets. Empirical bootstrap simulation is used to get censored bootstrap samples from the original data. Maximum likelihood estimation (MLE) is used to fit parametric probability distributions to each bootstrap sample [9].

Case Study for Analyzing Risk, Volume of Soil and Cost

Risk assessment

The RESRAD code is used to determine regulatory compliance. The US Nuclear Regulatory Commission (USNRC) and US Department of Energy (USDOE) use a 0.25 mSv/yr general limit or constraint for soil cleanup and site decontamination. The RESRAD code has basic models and parameters, but it can be modified to meet specific needs. The RESRAD code can be used to compute the potential annual doses or lifetime risks to workers or the public resulting from exposure to residual radioactive material in the soil. In addition it can support an as low as reasonably achievable (ALARA) analysis or cost-benefit analyses that can assist in the cleanup decision-making process [13].

Originally, the farmer scenario was the most conservative case. Because there is a low possibility for the farmer case in the Republic of Korea, the recreationist is a reasonably conservative scenario that can be used. The International Atomic Energy Agency (IAEA) proposes 0.01–0.30 mSv/yr, USDOE and USNRC propose 0.25 mSv/yr, and USEPA proposes 0.15 mSv/yr as the general limit for soil cleanup for 1000 years [13]. Korea uses 0.10–0.20 mSv/yr; therefore, 0.15mSv/yr was used as the general limit in this research. It was assumed that the radionuclides were homogeneously distributed in the area. Analysis of the long term environmental health risks and costs can be significantly affected by a relatively small variation in the mean value. Therefore, the best site specific data is required. The monazite powder manufacturing plant has limited data; thus, CSMRI was used to analyze the risk because it provides the best option for this analysis.

CSMRI stands for Colorado School of Mines Research Institute site in Golden, Colorado [14]. The S.M. Stoller Corporation conducted soil characterization and remediation activities necessary for the termination of the radioactive materials license and free release of the site.

Table IX. Comparison of the mean of radioactivity between ignoring the ND and including the ND.

Radionuclides	Mean of radioactivity Ignore ND (pCi/g)	Mean of radioactivity Include ND using KM (pCi/g)
Ra-226	27.96	25.98
Ra-228	3.66	3.43
Th-232	3.47	3.36
Th-230	21.45	20.58
Th-228	3.56	3.33
U-234	19.45	18.63
U-235	1.13	1.12
U-238	19.86	18.95

Table X. The maximum total dose for ignoring ND and including ND.

Case	Maximum total dose (mSv/yr)
Ignore ND	0.2592
Include ND	0.2414

Table X shows the maximum total dose for CSMRI. Again, the recreationist scenario which assumes no consumption of plant, milk, and water was used. Site-specific data from CSMRI was prepared for use in the RESRAD code. It was possible to calculate the maximum total dose. Differences in the radioactivity mean, directly affect the maximum total dose. According to Table X, the maximum total dose for both cases (ignoring and including the ND), are higher than the 0.15 mSv/yr the general limit for soil cleanup or site decontamination. In this case, site remediation is required to meet the cleanup criteria. It is necessary to reduce the maximum total dose to 0.15 mSv/yr in order to release the site for either restricted or unrestricted use in the future. The maximum total dose was 0.2592 mSv/yr when ignoring the NDs and 0.2414 mSv/yr when NDs are included. The difference in maximum total dose is not large. However, the maximum total dose is higher whenNDs are ignored vs including NDs. As a result, the cost of remediation could be reduced using the proposed methods, which include NDs.

Soil volume and cost savings estimates using ROK fees

Removal of contaminated soil includes excavation, transportation, and disposal [15]. By calculating and comparing the difference in total soil volume requiring removal = to meet the cleanup criteria, based on ignoring vs including nondetects, it is possible to estimate the number of 200-liter drums and the associated costs. If all contaminated soil is assumed to be

low level waste (LLW), the cost of managing the contaminated soil, in ROK, is all included in the cost of each 200-liter drum used. Assume that the cover thickness of 0.1 m will be used and some contaminated soil will be excavated for remediation. The total volume of soil removal can be estimated using the site geometry.

Table XI. The maximum total dose for ignoring ND and including ND.

Measure	Ignore ND (0.2592 mSv/yr)	Include ND (0.2414 mSv/yr)
Cover thickness needed (m)	0.1	0.1
CZ thickness to be removed (m)	0.611	0.5385
Total volume of soil needed (m^3)	724.4308	650.5613
Volume difference (m^3)	73.8695	

When the analysis ignores ND, it is necessary to cover non contaminated soil of 0.1m thickness; the removal of contaminated soil of 0.611m Contaminated Zone (CZ) thickness is required. Similarly, when the analysis includes ND, it is necessary to cover non contaminated soil of 0.1m thickness; the removal of contaminated soil of 0.5385m CZ thickness is required. Meeting the 0.15 mSv/yr cleanup criteria for unrestricted use, the total volume of soil removal was calculated for both cases of ignoring and including the nondetects. The amounts of soil removal for ignoring the nondetects and including the nondetects were 724.4308 m^3 and 650.5613 m^3, respectively.

In order to dispose 73.8695 m^3 of contaminated LLW soil, it is necessary to prepare

370 200-liter drums. A 200-liter drum can accommodate 200–500 kg of waste. In Korea, the cost of a 200-liter drum is 12,190,000 won [16]. Therefore, a more precise estimate of the activity in the soil results in a lower volume of soil removal and significant cost savings are achieved through including the BDL data.

Uncertainty Analysis

There are 44 types of distributions for the Latin hypercube sampling (LHS) technique, each with their own parameters [13]. The peak total dose, peak pathway doses, and peak nuclide doses that result from the set of input variables is analyzed, and the cumulative density function (CDF) is presented graphically. The probabilistic distributions for most of the parameters are developed from the RESRAD recommended probabilistic distributions and CSMRI site-specific data. Specifically, the uniform distribution is used for soil concentration parameters for uncertainty analyses to consider the effect of the including ND. The results provide the correlations and regression coefficients for the doses.

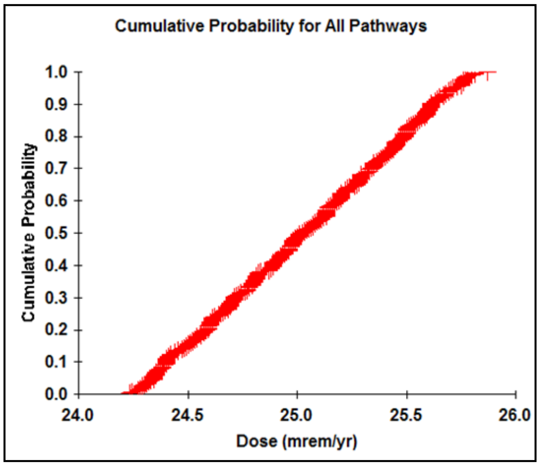

Fig. 1. Sensitivity analysis of soil concentrations CDF changes of the mean of peak dose.

The dose can be significantly affected by soil concentration, contaminated zone, and ingestion non-dietary. In contrast, saturated zone, unsaturated zone, occupancy, and ingestion dietary do not significantly affect the dose. It is proven that analyzing the long term environmental health risk can be significantly affected by the relatively small variation in the soil concentration mean value.

The contribution ranking for the dose was evaluated for 41 input parameters using Partial Correlation Coefficient (PCC) and Partial Ranked Correlation Coefficient (PRCC) approaches. As seen in Table XII, the radioactivity of a select number of radionuclides can contribute significantly to the maximum total dose (i.e., rank 1,3,4). Although the mean of radioactivity and the detection limits are insignificantly small in case of Ra-228, it ranks 3rd in contribution for the dose. When the detection limits are small, data below the detection limits is usually ignored. However, it is ill advised to ignore censored data, since the dose and cancer risk can be significantly affected by not only the size of detection limits but also the type of radionuclides, daughter radionuclides, and amount of radioactivity.

Table XII. Correlation analysis results using the PCC and PRCC approaches.

Rank	PCC	PRCC
1	Ra-226	Ra-226

2	Soil ingestion	Soil ingestion
3	Ra-228	Ra-228
4	Th-228	Th-228
5	Runoff coefficient	Runoff coefficient
6	Evapotranspiration coefficient	Evapotranspiration coefficient
7	CZ density	CZ density
8	CZ area	U-238
9	U-238	CZ area, Depth of soil mixing layer
10	Mass loading for inhalation	
11	Th-230	Th-232
12	Th-232, Well pumping rate, SZ effective porosity	Mass loading for inhalation
13		U-235, Th-230, SZ hydraulic gradient, Well pumping rate, Length parallel to aquifer flow, USZ thickness
14		
15	CZ thickness	
16	USZ total porosity	
17	SZ density, Depth of soil mixing layer, SZ hydraulic conductivity, Well pump intake depth, SZ total porosity	
18		
19		Well pump intake depth, USZ density, Precipitation
20		
21		
22	b parameter of USZ, USZ hydraulic conductivity, SZ b parameter, SZ hydraulic gradient, CZ total porosity	b parameter of USZ, USZ hydraulic conductivity, SZ density, CZ b parameter
23		
24		
25		
26		CZ hydraulic conductivity, SZ b parameter, External gamma shielding factor
27	U-235, CZ hydraulic conductivity, USZ density, External gamma shielding factor	
28		
29		Indoor dust filtration factor, SZ hydraulic conductivity, Aquatic food
30		
31	Indoor dust filtration factor, Length parallel to aquifer flow	
32		SZ effective porosity
33	CZ erosion rate, Aquatic food, USZ effective porosity, CZ b parameter, USZ thickness	CZ total porosity, SZ total porosity, CZ erosion rate, U-234
34		
35		
36		
37		USZ total porosity
38	U-234	CZ thickness
39	Precipitation	USZ effective porosity
40	Wind speed	Wind speed
41	Depth of roots	Depth of roots

Decision-Making Framework for Selecting the Best Statistical Technique

A decision-making framework was developed because the selection of the best statistical technique is very site/data specific. Therefore selection of a technique is accomplished on a case-by-case basis using the evaluation steps defined in this framework.

The first step in the process is to develop a known distribution using the parameters, e.g. the mean and the standard deviation, to set the detection limit and the sample size, and to delete the BDL data. The best-fit statistical technique can be recommended through calculating the bias of the mean from the known distribution [17]. Three types of distributions were considered in order to evaluate the effect of the different types of distributions. The second step is to remove or delete the data below detection limits. The detection limits for 10%, 30%, and 60% cumulative probability of censoring were calculated using the parameters of the distributions and 107 samples. The Kaplan-Meier method, the robust regression on order statistics method, and the maximum likelihood estimation methods were applied to each specified population distribution for each type of distribution, censoring percentage, and sample size to estimate data below the detection limits. It was conducted by simulating 1000 bootstrap samples. Not only the censoring percentage and sample size, but also the amount of radioactivity, the number of detection limits, and the types of distributions can affect the selection of the best statistical technique. However, the study of these influences have not been done.

Table XIII. Results of decision-making framework.

> 10 pCi/g (big size)			
Lognormal			
Detection limit	CP	Sample size	
		≤40	> 40
Single	≤30%	ROS or KM	ROS
	>30%	ROS or MLE	MLE (no KM)
Multiple	No >30%	KM	MLE
	Include >30%	KM or MLE	KM or MLE
Gamma			
Detection limit	CP	Sample size	
		≤40	> 40
Single	≤30%	ROS	ROS
	>30%	ROS or MLE	ROS (no KM)
Multiple	No >30%	KM	MLE
	Include >30%	KM or MLE	MLE
Weibull			
Detection limit	CP	Sample size	
		≤40	> 40
Single	≤30%	MLE	KM
	>30%	ROS or MLE	MLE (no KM)
Multiple	No >30%	ROS	MLE
	Include >30%	ROS	MLE
≤ 10 pCi/g (small size)			
Lognormal			
Detection limit	CP	Sample size	
		≤40	> 40
Single	≤30%	MLE	ROS
	>30%	ROS or MLE	MLE (no KM)
Multiple	No >30%	MLE	MLE

	Include >30%	MLE	KM or MLE
Gamma			
Detection limit	CP	Sample size	
		≤40	> 40
Single	≤30%	ROS	ROS
	>30%	ROS or MLE	ROS (no KM)
Multiple	No >30%	MLE	MLE
	Include >30%	ROS or MLE	ROS or MLE
Weibull			
Detection limit	CP	Sample size	
		≤40	> 40
Single	≤30%	MLE	MLE
	>30%	ROS or MLE	ROS (no KM)
Multiple	No >30%	ROS or MLE	MLE
	Include >30%	ROS or MLE	MLE

CONCLUSION

The KM, ROS, and MLE methods were verified to give more precise estimates of the mean compared to conventional methods where censored data sets are ignored or replaced with the detection limit, half of the detection limit, or zero.

The proposed methods of the Kaplan-Meier, ROS, and MLE methods were performed using the soil samples from the monazite powder manufacturing plant and CSMRI. The KM, ROS, and MLE are flexible and robust methods for analyzing data below the detection limits. The concept of the bootstrap simulation to estimate confidence intervals for the mean was introduced, and the MLE/Bootstrap method was implemented in respect to the various percent confidence intervals for the mean of monazite powder manufacturing plant data set. The preliminary evaluation demonstrated that the proposed methods can be effectively used to provide the best estimated radioactivity levels at a decommissioned NPP site, and it can also estimate the uncertainty in the mean values. The RESRAD code was used to estimate the radiation doses and cancer risks in each case. The risk assessment and volume estimation was performed using the proposed decision-making framework. The amount of remediation of the contaminated soil was estimated and compared with the results of the conventional method. Furthermore, the cost saving difference was analyzed between the conventional method and the proposed methods.

Probabilistic distributions were developed for the RESRAD input parameters to analyze uncertainty. The uncertainty in the maximum total dose for different parameters was analyzed and the contribution rankings were estimated. The number of model input parameters were simulated simultaneously in order to determine their combined effect on the model outputs. Although only limited data was available for analysis, a regulatory decision can be made based on the uncertainty analysis. A sensitivity analysis of the RESRAD input parameters was performed for the CSMRI data set. The key sources that contribute to the maximum total dose were identified. If the MDAs are less than 10% of the Derived Concentration Guideline Levels (DCGL), it is possible to ignore data below the detection limits [2]. Although the mean of radioactivity and detection limits are insignificantly small, it is ill advised to ignore data below the detection limits. Since the dose and cancer risk can be significantly affected by not only the size of detection limits but also the type of radionuclides, daughter radionuclides, and amount of radioactivity. The key advantages of the proposed methods are that they are statistically unbiased estimates and can be used for a variety of situations such as different types of distributions, censoring percentages, sample sizes and the number of detection limits.

Through changing the amount of radioactivity, types of distributions, censoring percentages, sample sizes, and number of detection limits, the recommended methods are defined for estimating the summary statistics. Recommended methods are defined to estimate summary statistics, based upon simulations that address lognormal, gamma, and Weibull distributions for different sample sizes of 20, 40, and 100 and censoring percentages of 10%, 30%, 60%, (10,30)%, (30,60)%, (10,60)%, (10,30,60)%. The development of a nondetects analysis framework for decision-making will be provided to the regulators.

Using additional statistical analyses of the contaminated soil before or after decommissioning is expected to provide better and more reliable probabilistic exposure assessments, better economics, and improved communication with the public. Efforts to include nondetects in order to assess risk, estimate volume of soil removal, and examine cost savings more precisely should be made.

REFERENCE

1. INTERNATIONAL ATOMIC ENERGY AGENCY, Decommissioning of Facilities Using Radioactive Material, IAEA Safety Standards Series No. WS-R-5, IAEA, Vienna (2006).
2. US Nuclear Regulatory Commission. Multi-agency radiation survey and site investigation manual (MARSSIM). NUREG-1575, Rev. 2000 Aug;1.
3. Zheng J, Frey HC. Treatment of Censored Data and Dependency among Sampling Distributions in Probabilistic Human Health Exposure Assessment.
4. Gordon SM, Callahan PJ, Nishioka MG, Brinkman MC, O'Rourke MK, Lebowitz MD, Moschandreas DJ. Residential environmental measurements in the national human exposure assessment survey (NHEXAS) pilot study in Arizona: preliminary results for pesticides and VOCs. Journal of Exposure Analysis and Environmental Epidemiology. 1999 Oct 1;9(5):456-70.
5. Zhao Y, Frey HC. Quantification of variability and uncertainty for censored data sets and application to air toxic emission factors. Risk Analysis. 2004 Aug 1;24(4):1019-34.
6. Helsel DR. Nondetects and data analysis. Statistics for censored environmental data. Wiley-Interscience; 2005.
7. Baker S, Driver J, McCallum D. Residential Exposure Assessment. In Residential Exposure Assessment 2000 (pp. 9-15). Springer US.
8. Frey HC. Quantitative analysis of uncertainty and variability in environmental policy making. Fellowship Program for Environmental Science and Engineering, American Association for the Advancement of Science, Washington, DC. 1992 Sep.
9. Zhao Y. Probabilistic Modeling of Variability and Uncertainty in Urban Air Toxics Emissions
10. versus Time TC. STATISTICAL ANALYSIS OF GROUNDWATER MONITORING DATA AT RCRA FACILITIES.
11. Schumacher B, Barnett F, Singh A, Singh AK. ProUCL Version 4.00. 04

Technical Guide.

12. Orbitech, Development of Radioactivity Inference Method and Acceptance Criteria Based on Statistical Approach for Treatment or Disposal of NORM Waste. 2016.
13. Yu C, Zielen AJ, Cheng JJ, LePoire DJ, Gnanapragasam E, Kamboj S, Arnish J, Wallo III A, Williams WA, Peterson H. User’s manual for RESRAD version 6. Environmental Assessment Division, Argonne National Laboratory, Argonne. 2001 Jul 23.
14. New Horizons Environmental Consultants, Inc. 6585 S. Wright Street Littleton, CO 80127, CSMRI CHARACTERIZATION SUMMARY, 2003 Aug 21.
15. EPRI Report # 1023464, Groundwater and Soil Remediation Guidelines for Nuclear Power Plants, 2011 July.
16. Ministry of Trade, Industry and Energy, Ministry of Trade, Industry and Energy Law No. 2015-132, 2015 June 30.
17. Annan SY, Liu P, Zhang Y. Comparison of the Kaplan-Meier, Maxilllum Likelihood, and ROS Estimators for Left-Censored Data Using Simulation Studies.