

A Geostatistical Assessment of the Extent of Radioisotope Contamination Under Building 324 in the Hanford 300 Area

September 2023

Moses Obiri
Alejandro Ojeda
Ben Jensen
Frederick Day-Lewis
Deborah Fagan

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062

www.osti.gov

ph: (865) 576-8401

fox: (865) 576-5728

email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312

ph: (800) 553-NTIS (6847)

or (703) 605-6000

email: info@ntis.gov

Online ordering: <http://www.ntis.gov>

A Geostatistical Assessment of the Extent of Radioisotope Contamination Under Building 324 in the Hanford 300 Area

September 2023

Moses Obiri
Alejandro Ojeda
Ben Jensen
Frederick Day-Lewis
Deborah Fagan

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

In October of 1986, a spill of highly radioactive cesium (^{137}Cs) and strontium (^{90}Sr) occurred in the B-Cell of the 324 Building in the 300 Area of the Hanford Site. An unknown volume of spilled material was lost to the subsurface through a leak in the sump in the B-Cell floor. The spill is estimated to have contained approximately 1.3 million curies of radioactivity. Understanding the spatial extent of contamination is critical to (1) the design of an effective remedy to address the contamination, and (2) the safe implementation of that remedy. In past work to characterize the spatial distribution of contamination, 45 borings were conducted at different orientations to access the subsurface under the B-Cell. Over 2200 measurements were acquired of dose rates, temperature, and moisture at different locations along the boreholes. In this study, a geostatistical approach is applied to the dose-rate measurements to assess the probability of exceeding the 1-R/hr dose-rate threshold in a 3-D volume underlying the B-Cell. Indicator variograms were developed based on the dose-rate data, and an indicator kriging (IK) algorithm was applied to the data using a 1-R/hr threshold. Post-processing of the IK results produces 3-D volumes of the probability of exceeding 1 R/hr throughout the subsurface under the B-Cell. 3-D visualizations and maps were generated to facilitate visualization of the region of the subsurface contaminated at or above the 1 R/hr threshold at 95% confidence. Three disconnected regions were identified. The areal extent of a box encapsulating the three regions with 1-R/hr contamination present at 5% probability and 95% confidence is estimated to be 34 ft x 34 ft and 20 ft deep, for a total volume of 23,120 ft³.

Acknowledgments

We thank Mark Rockhold, Amoret Bunn, and Meenu Mohankumar for technical feedback and reviews of aspects of this work. We also thank Opris (Vince) King for assistance with access to the borehole dataset used in this study.

Acronyms and Abbreviations

- 2-D – Two-dimensional
- 3-D – Three-dimensional
- ft – Feet, English distance units
- BLUE – Best Linear Unbiased Estimator
- CPCCo – Central Plateau Cleanup Company
- GSLIB – Geostatistical Software Library
- IK – Indicator Kriging
- m – meters, metric distance units
- OK – Ordinary Kriging
- R/hr – Rads per hour, dose rate

Contents

Abstract.....	ii
Acknowledgments.....	iii
Acronyms and Abbreviations	iv
1.0 Introduction	1
2.0 Dataset.....	2
2.1 Data Extraction.....	2
2.2 Data Processing	3
2.3 Data Summary	3
3.0 Approach.....	5
3.1 Kriging	5
3.2 Variography	5
3.3 Calculation of Kriging Weights.....	2
3.4 Anisotropy	3
3.5 Indicator Kriging	4
3.6 Accounting for Uncertainty	4
3.7 Delineating Regions of Exceedance.....	5
4.0 Results	6
4.1 Variogram estimation	6
4.2 Indicator kriging and volume estimation	7
5.0 Discussion and Conclusions	10
6.0 References.....	11
Appendix A – Dataset.....	A.1
Appendix B – Data Processing	B.6

Figures

Figure 1: Data curation workflow diagram	2
Figure 2: 3-D view of borehole observations	4
Figure 3: Northing, Easting, and Elevation dose rate distribution	4
Figure 4: Six commonly used variogram models, from Deutsch, 2003.....	2
Figure 5. Experimental variograms in various directions in the North – East plane.....	6
Figure 6: Variogram models at major (90 degrees) and minor (45 degrees) directions.....	7
Figure 7: Indicator kriging results.....	8
Figure 8: North and East view of extent.....	8
Figure 9: Northing and Easting by elevation extents	8

Tables

Table 1: Data summary.....	4
Table 2: Extent and volume estimates based on indicator kriging.....	9

1.0 Introduction

In October 1986, an accidental spill from a liquid waste stream of concentrated cesium-136 (^{137}Cs) and strontium-90 (^{90}Sr) occurred. Approximately 510 L of material was spilled onto the floor of the B-Cell in Building 324, located in the 300 Area of the U.S. Department of Energy's Hanford Site (WCH, 2011). Some quantity of the spilled material was lost to the subsurface through a sump in the B-Cell floor.

Understanding the spatial distribution of contamination in the subsurface is critical to selection and implementation of an effective remedy to address the contamination. For example, the ability to excavate the contaminated sediments under the B-Cell depends on the spatial extent of contamination (lateral and vertical) relative to the footprint of Building 324. Furthermore, the extent of high-level contamination has important implications for the planning of the work to ensure the safety of workers involved in implementing any site remedy.

Past work to assess the distribution of subsurface contamination below the B-Cell has focused on field measurements to characterize the extent of contamination (Rockhold, 2015) and numerical modeling to assess fate and transport of the radioactive contamination (Rockhold et al., 2012). Characterization efforts from 2010 – 2014 and more recently from 2022 - 2023 included installation of 45 boreholes at various orientations into the sediments under the B-Cell, with over 2200 measurements of dose rate, moisture, and temperature.

This report presents a geostatistical assessment of the spatial extent of radioactivity, as dose rate, present in the subsurface at levels exceeding 1 R/hr, a key decision threshold. An indicator kriging (IK) algorithm is used to calculate the spatially distributed probability of exceeding the 1-R/hr threshold. The report is organized as follows. Section 1 (this section) provides an introduction and overview of the problem. Section 2 describes the characterization data used as input (i.e., conditioning data) for the geostatistical analysis. Section 3 describes the geostatistical approach, including variogram inference, kriging, and IK. Section 4 presents the results. Section 5 closes with discussion of the results and conclusions drawn from the results. Appendices document the data and codes used for the analysis to enable reproducibility and facilitate follow on work. The appendices include the data files and computer codes used to perform the analysis and generate the figures in this report.

2.0 Dataset

The dataset used in this study comes from a total of 45 boreholes drilled at various orientations into the subsurface below the B-Cell. The dose-rate dataset comprises 2266 measurements acquired at locations along the boreholes. Organizing the data and putting it into a common spatial coordinate system entailed a substantial component of the overall effort.

Data curation is critical to enabling reproducibility of results and facilitating future work. **Figure 1** presents an overview of the data curation workflow, divided into two phases: (1) data extraction and (2) data processing. Data extraction entailed the compilation of unprocessed data files in various formats such as Excel, PDF, and SolidWorks files, followed by the development of codes to systematically extract and organize the pertinent information. The positions of the boreholes were derived from SolidWorks files, whereas the dose-rate data was acquired from multiple Excel spreadsheet files. The data were consolidated and transformed into a comma-separated values (CSV) file, which includes a tabular representation of the boreholes, along with their respective geographic coordinates, measures, and date of collection. An additional step was undertaken in the processing of data to create an analytical dataset. The processing entailed the conversion of dose rates into a standardized unit, the consolidation of adjusted and unadjusted data into a unified variable, and the implementation of corrections to account for measurement decay because data were collected over the span of several years. The subsequent sections and appendix provide a more detailed description of this method.

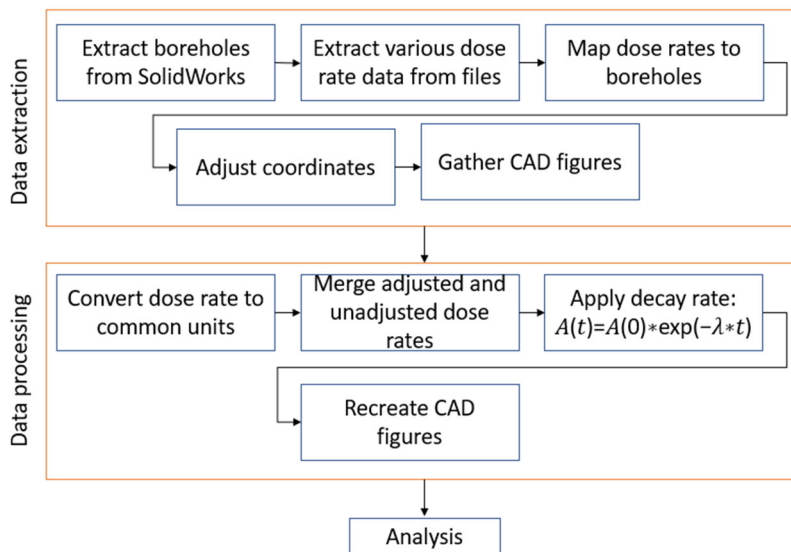


Figure 1: Data curation workflow diagram.

2.1 Data Extraction

The objective of the data-extraction effort was to associate spatial coordinates (North, East, elevation) to radiation dose-rate data collected via probes in boreholes extending under the 324 building. The local coordinate of the start and endpoints for the boreholes/probes were extracted from a SolidWorks 3-D model provided by the Site Contractor, Central Plateau Cleanup Company (CPCCo). In the SolidWorks model, measurement locations were in coordinate systems local to each probe—not in a common coordinate system as required for geostatistical analysis. The probe E-01 was used to align the local coordinates in SolidWorks; i.e., coordinates were originally specified for each starting point of the probes. A python script (see

Appendix A) was written to cross-reference the spatial data for probes with the measurement data to determine the coordinates of each dose-rate measurement. The details of this process and calculations are described in greater detail in *Appendix A -- Dataset* along with the content list of the data package accompanying this report.

2.2 Data Processing

The extracted dataset includes multiple boreholes, their corresponding spatial coordinates, dose-rate measurements, other data not used in this analysis (e.g., temperature), and the date of measurement collection. The dose-rate measurements varied in terms of units, collection dates, and adjusted vs. unadjusted values; thus, a data processing step was necessary to produce a consistent dataset for geostatistical analysis. The necessary transformations were performed using an R script (see *Appendix B*).

The measurements for different boreholes were modified to compensate for variations in the thickness of the casing wall, which varied between boreholes. If the adjusted measurement was accessible, it was used; otherwise, the unadjusted measurement was utilized. Subsequently, all readings obtained from the boreholes were standardized to a uniform dose-rate unit (R/hr). Data were checked for duplicate locations and errors corrected. The data-processing step involved accounting for the varying dates of collection and adjusting for radioactive decay between data collection and a common date. The following decay-rate formula was used to convert all measurements to a common date, selected to be 6/1/2023:

$$A(t) = A(0)e^{-\lambda t}, \text{ where}$$

$$\lambda = \frac{\ln(2)}{\text{half-life}}$$

$A(t)$ is the activity at time t and $A(0)$ is the initial activity. The fraction of the contamination coming from ^{137}Cs versus ^{90}Sr is unknown; hence an effective half-life of 29.5 years was assumed, corresponding to a simple average of the half lives for ^{137}Cs and ^{90}Sr .

2.3 Data Summary

Data extraction and processing resulted in a dataset in tabular form, with 2266 rows, where each row represents a distinct borehole observation. The dataset includes observations from a total of 45 distinct boreholes, each belonging to one of five borehole types denoted as C, E, N, S, and W. **Table 1** provides a summary of the quantity of boreholes, the quantity of observations available for each type, and the quantity of observations that exceed 1 R/hr. **Figure 2** presents a 3-D depiction of the boreholes, wherein observations over 1 R/hr are denoted by red spots, indicating their classification as high dose rate. Observations categorized as blue points are those with radiation levels that are equal to or below 1 R/hr, thus falling inside the low dose range. **Figure 3** illustrates the dosage rate distribution from the perspectives of Northing, Easting, and Elevation, thus collapsing the visualization to one-dimensional views. The observations made in the Easting direction covered a distance of roughly 52 feet, whereas the high dosage observations were limited to a span of around 30 feet. The observations made in the Northing direction covered a distance of roughly 99 feet, whereas the high dosage observations were limited to a span of around 32 feet. The recorded measurements in elevation covered an approximate range of 54 feet, whereas the observations with high dosage were limited to a range of around 13 feet. For reference, the elevation at the top of the floor of the B-Cell is 394.09 ft.

Table 1: Data summary

Borehole Type	Number of boreholes	Number of Observations	Observations > 1R/hr
C	16	1249	599
E	15	712	67
N	8	209	0
S	3	48	0
W	3	48	0
Total	45	2266	666

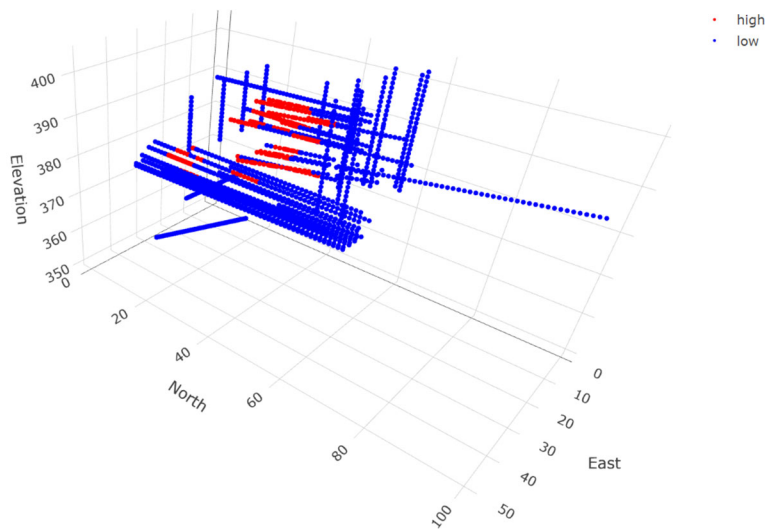


Figure 2: 3-D view of borehole observations

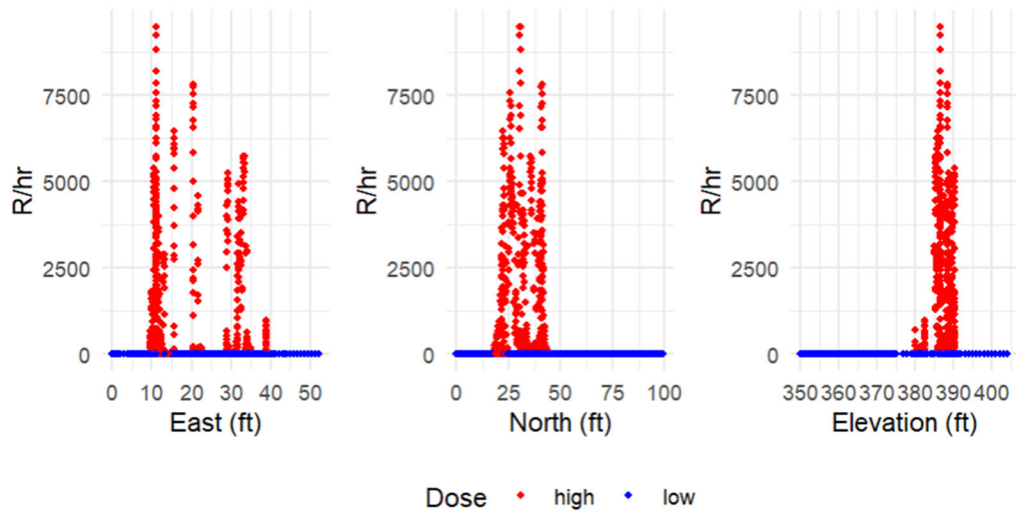


Figure 3: Northing, Easting, and Elevation dose rate distribution

3.0 Approach

This study utilized a geostatistical methodology to assess the extent of contamination in proximity of the B234 building. Geostatistics is a subfield of statistics for analyses of datasets with georeferenced or spatiotemporal attributes. Originally developed for the mining sector, geostatistics has since found applications in diverse fields such as environmental science, petroleum engineering, hydrogeology, and others (De Iaco et al., 2022; Msengwa, 2021; Rodrigues et al., 2020).

Geostatistics provides a framework to account for spatial variability and make predictions at unobserved locations using the data from sampled locations (Hoef & Cressie, 1993). This is achieved by the assessment of statistical correlations among the observed data points in order to establish geographical patterns and trends (Pebesma, 2004). Compared to conventional interpolation techniques (e.g., nearest neighbor, splines, inverse distance, etc.), geostatistical approaches allow for (1) consideration of site-specific patterns of spatial correlation, based on observed data, (2) rigorous quantification of uncertainty, and (3) calculation of spatially variable probabilities of exceedance for thresholds on the quantity of interest.

3.1 Kriging

Kriging is a fundamental tool utilized in the field of geostatistics. Kriging encompasses a collection of statistical methodologies employed for the purpose of interpolating or predicting values at unobserved locations based on available data from sampled locations. The kriging method assumes that a portion of the spatial variability present in natural phenomena can be represented by stochastic processes exhibiting spatial autocorrelation. The primary aim of kriging is to provide estimates at unsampled locations that are both unbiased and possess the minimal variance, hence establishing it as a best linear unbiased estimator (BLUE) (Noel Cressie, 2015).

Kriging can be conceptualized as a two-step procedure. Initially, the spatial structure of the sampled points is inferred by fitting a variogram model to observed spatial variability. The variogram is a mathematical function that is the basis for estimating values in unsampled areas, similar to weighted least squares regression. The variogram describes how the quantity being estimated varies in space, as a function of separation distance and direction between observations. Subsequently, weights based on the variogram model are used to interpolate values for unsampled points or blocks within the spatial field of interest.

Kriging can be performed in one, two or three dimensions to produce estimates along profiles, over areas, or throughout volumes. In addition to the interpolated profile, surface, or volume, kriging also produces an assessment of the uncertainty associated with the estimates, in the form of a kriging variance. Based on the kriging variance, confidence thresholds can be calculated. In this work, kriging variance is used to establish confidence intervals on the estimated probabilities of exceeding the 1-R/hr level.

3.2 Variography

The first phase of kriging entails analysis of the data to infer a model variogram. The variogram is a statistical measure that summarizes the level of spatial dependency, i.e., correlation, between pairs of data points, taking into account their separation distance (h) and direction.

Commonly the semi-variogram, which is defined as one half the variogram, is considered. The experimental semi-variogram can be defined in its most general form as:

$$\gamma(h) = 0.5 * E[(Z(x + h) - Z(x))^2], \tag{1}$$

where $Z(x)$ is the observed value at location x , $Z(x + h)$ is the observed value at distance h from x , and $E[]$ is the expectation operator. The expected value is the sum of all possible values $Z(x + h) - Z(x)$ for available data pairs multiplied by their respective likelihoods. In Equation 1, $\gamma(h)$ represents half the mean-squared difference between the values of pairs of values separated by h . In the context of the sampled data, this gamma-value or semi-variance is plotted versus distance, or lag, between each pair of points; this “experimental variogram” is subsequently fitted with a theoretical model, i.e., the “model variogram,” yielding a continuous function that is used to represent the spatial correlation structure across the entirety of the study area. Kriging makes use of the model variogram to estimate values at unsampled locations while accounting for and capitalizing on observed spatial correlation. The selection of the theoretical model variogram is restricted to a finite set of functions which ensure a positive-definite variogram and non-singular kriging system. Commonly used models include the linear, spherical, exponential, and power models (see **Figure 4**).

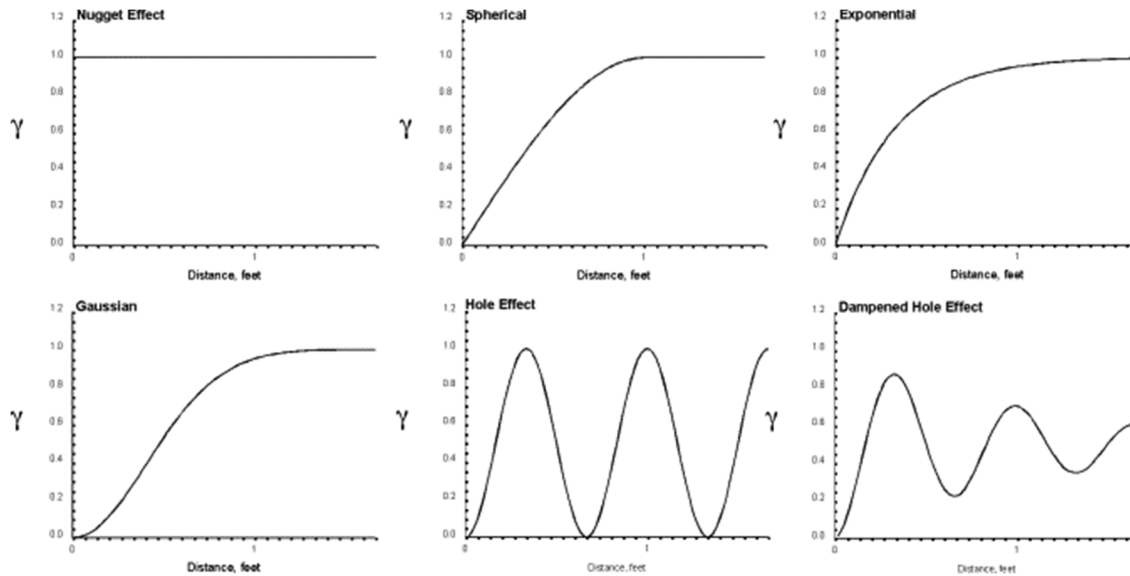


Figure 4: Six commonly used variogram models, from Deutsch, 2003.

3.3 Calculation of Kriging Weights

The second phase of the kriging procedure utilizes the variogram model to determine the weights for the purpose of interpolation. Sample locations are allocated weights in relation to the location to be predicted using the covariance matrix derived from the variogram model. As an illustration, in the case of an exponential variogram, the corresponding exponential covariance function can be represented as $C(h, a) = \sigma^2 \exp(-|h/a|)$. Here, the parameter a represents the estimated range parameter derived from the fitted model variogram, σ^2 denotes the nugget effect, and h is the distance between two points. The range parameter a is the distance where the variogram model reaches or approaches a plateau. For the spherical model, the range is where the plateau is reached, and beyond this value of h two observations are not correlated. In the context of the exponential covariance structure, the covariance tends towards zero and the

associated weight diminishes as the distance h from a given prediction location increases. Therefore, in general, sample points that are closer to a prediction location are allocated greater weights, and sample points that are farther from a prediction location are assigned lesser weights. In addition to the distances from measurements to prediction location, the spatial configuration of measurements also factors into the calculation of weights. If two measurements are located close together, the weights assigned to them will, in general, be less, as there is redundancy in the information the two measurements provide; thus, kriging serves to “decluster” measurements.

The general system of equations for kriging interpolation is expressed as

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i). \quad (2)$$

In Equation 2, $\hat{Z}(x_0)$ represents the predicted value at the location x_0 . It is obtained by summing the product of each sampled point value, $Z(x_i)$, and its corresponding unique weight, λ_i . Thus, through the variogram and the kriging system, kriging accomplishes interpolation by balancing local and global spatial information, accounting for the spatial correlation between sampled locations and the locations of estimates, and accounting for the spatial correlation between measurement locations.

3.4 Anisotropy

In Equation 1, the semi-variance is defined only as a function of distance, h , and not direction; however, in many geological and environmental contexts, the spatial continuity of attributes differs as a function of direction. For example, soil properties may exhibit greater correlation (e.g., longer ranges) in the horizontal direction versus the vertical direction, consistent with soil formation processes and development of soil horizons or layers. This phenomenon of directional dependence is called anisotropy. In the context of kriging, anisotropy means that the influence of measurements in predicting unsampled locations may need to change depending on the direction between the location to be estimated and the locations of observations. When anisotropy is present, the usual variogram formula in Equation 1 is modified by adjusting the distances between points based on the anisotropy ratios and angles. For a 3-D spatial dataset, the anisotropic distance, h' , between two locations separated by h can be expressed as:

$$h' = \sqrt{(h_{major}/a_{major})^2 + (h_{minor}/a_{minor})^2 + (h_{tertiary}/a_{tertiary})^2}, \text{ where} \quad (3)$$

a_{major} , a_{minor} and $a_{tertiary}$ are range parameters along the principal directions of anisotropy, and h_{major} , h_{minor} , $h_{tertiary}$ are lag separations rotated to align with the principal directions of anisotropy. The system of equations to rotate the lags is

$$\begin{bmatrix} h_{major} \\ h_{minor} \\ h_{tertiary} \end{bmatrix} = [R] \begin{bmatrix} x_1 - x_2 \\ y_1 - y_2 \\ z_1 - z_2 \end{bmatrix}, \text{ where} \quad (4)$$

$[R]$ is the rotation matrix calculated from the three angles that characterize the orientation of an ellipsoid of range values.

3.5 Indicator Kriging

Indicator kriging (IK) is a specialized form of kriging that deals with categorical or thresholded data. When applied to continuous data, it transforms such data into a binary format, indicating whether a value is above (1) or below (0) a particular threshold. The resulting kriged values are probabilities of exceeding the selected threshold. In essence, IK amounts to kriging data that have been transformed into probabilities of exceedance for a specified threshold, and IK produces estimates of exceedance probabilities.

The IK procedure has the following steps:

1. Convert the continuous data into binary (indicator) data for a predetermined threshold of interest. That is, given a threshold t , the indicator transformation for a data value $Z(x_i)$ is:

$$I(x_i, t) = \begin{cases} 1 & \text{if } Z(x_i) > t \\ 0 & \text{otherwise} \end{cases}, \text{ where} \quad (5)$$

$I(x_i, t)$ is the indicator value at location x_i for threshold t , $Z(x_i)$ is the observed data value or concentration at location x_i . In effect, $I(x_i, t)$ is the probability of exceeding t at location x_i . Given that a measurement available at x_i , $I(x_i, t)$ will be either 1 or 0.

2. Calculate and plot the experimental variogram value for the indicator data.
3. Select a model variogram and fit the model to the experimental variogram.
4. Apply kriging to the indicator-transformed data using the model variogram to estimate probabilities of exceedance at unsampled locations.

The kriging system for the indicator variable at location x_0 is:

$$\hat{I}(x_0, t) = \sum_{i=1}^n \lambda_i I(x_i, t), \text{ where} \quad (6)$$

λ_i are the kriging weights assigned to the observed indicator values, and the result $\hat{I}(x_0, t)$ is the estimated probability that the value at location x_0 exceeds the threshold t . When considering environmental characterization and/or remediation, Equation 5 can model various criteria, depending on regulatory or health constraints; i.e., selection of the threshold t can be guided by decision criteria. In this study, $t = 1 \text{ R/hr}$, therefore $\hat{I}(x_0, t = 1) = 0.95$ would mean there is 95% chance that the contamination at location x_0 exceed 1 R/hr , given the observed data and modeled variogram. The outcomes from Equation 6 are a probability map (for 2D) or volume (for 3-D) that demarcates regions that are probable to surpass the desired contamination thresholds. A separate calculation yields the kriging variance for IK which can be used to quantify the uncertainty associated with the IK estimate.

3.6 Accounting for Uncertainty

A principal advantage of kriging over non-statistically based interpolation is that kriging provides a quantification of the uncertainty inherent in the estimation process. Although the estimate $\hat{I}(x_0, t)$ is a probability, there is uncertainty associated with this estimate. Confidence intervals provide a range of probable values, allowing more insight into risks and thus enabling risk-based decisions. For each estimated probability $\hat{I}(x_0, t)$ at location x_0 , a confidence interval (CI) can be derived. The interval indicates the range within which the true probability lies with a

certain level of confidence (typically 95%). The interval is centered around the kriged estimate $\hat{I}(x_0, t)$ and accounts for the estimation variance. Formally, the CI can be defined as the interval:

$$CI(x_0, t) = (\hat{I}(x_0, t) - d \times \sqrt{\text{var}[I(x_0, t)]}, \hat{I}(x_0, t) + d \times \sqrt{\text{var}[I(x_0, t)]}). \quad (8)$$

In Equation 8, $CI(x_0, t)$ is the confidence interval for the estimated probability at location x_0 , $t = 1 \text{ R/hr}$ threshold, $\text{var}[I(x_0, t)]$ represents the kriging variance for location x_0 at threshold $t = 1 \text{ R/hr}$, and d is the z -value corresponding to the desired confidence level. In this study, $d = 1.96$ for a confidence level of 95% was used.

3.7 Delineating Regions of Exceedance

In this study, we sought to delineate the regions where the 1 R/hr may be exceeded with at least a 5% probability with 95% confidence. Remediation (e.g., excavation) of the regions thus delineated would address all contamination above the 1 R/hr threshold with 95% probability and 95% confidence (95%/95%). The IK results were post-processed by applying another threshold to ensure 95% confidence in the result. To convert the continuous probability map into a binary map, a decision threshold of 5% is employed:

- If the lower bound of the CI for estimated probability $\hat{I}(x_0, t)$ is greater than 5%, then the location x_0 is designated as contaminated at the 1-R/hr level or greater with at least 5% probability and 95% confidence.
- Conversely, if the upper bound of the CI is less than or equal to 5%, the area is considered not to be contaminated at the 1-R/hr level with 95% probability and 95% confidence.

This approach ensures that areas are only designated as being below the decision level (1 R/hr) if there is high probability and high confidence (at least 95% with 95% confidence) that contamination levels are below the 1-R/hr threshold. We apply the confidence threshold in part to address sources of errors not explicitly accounted for in the IK, e.g., positioning errors and measurement noise. This approach provides a conservative safety buffer, erring on the side of caution to prioritize health and environmental considerations and striking a prudent balance between accuracy and safety.

In examining and interpreting the results that follow, it is important to understand that contamination exists outside the delineated regions. As explained above, the focus of this study is to identify the regions of the subsurface where action would be required to address the presence of high-level ($> 1 \text{ R/hr}$) contamination; however, lower-level contamination extends beyond the volumes delineated. If lower thresholds on dose rate (e.g., 0.5 R/hr) were considered, the regions where the probability of exceedance is greater than 0.95 would expand, but that is beyond the scope of the current effort.

4.0 Results

This section presents results for inference of the model variogram based on the experimental variogram, IK results, and visualizations of regions where the 1-R/hr threshold may be exceeded. The dataset, as described in Section 2, is highly amenable to variogram inference and IK, owing to the wealth of closely spaced measurements collected along multiple 3-D orientations. Based on site knowledge, geologic principles, and inspection of the dataset, anisotropy was expected, and therefore analyzed first. IK was performed using a threshold of 1 R/hr, the decision level specified by DOE. The resulting probability and uncertainty estimates were used to calculate 95%/95% coverage intervals to provide a conservative estimate of plume boundaries. 3-D and 2D visualizations were developed using input from CAD drawings provided by CPCCo.

4.1 Variogram estimation

Anisotropy in spatial correlation of contaminant levels is common given the nature of geologic heterogeneity, spatial patterns of contaminant release, and the physics of contaminant transport. Preliminary inspection and visualization of the dose-rate dataset showed indications of anisotropy, particularly in the north-east direction. To rigorously assess anisotropy, directional variograms were calculated and plotted within the north-east plane for angles of 0, 30, 60, 90, 120, and 150 degrees, with 0 aligned with north, 90 east, 180 south, and 270 west. This series of experimental variograms is presented in **Figure 5**. Among these, the variogram oriented at 90 degrees (east direction) exhibited the longest range, indicating pronounced anisotropy and more spatial correlation in this orientation.

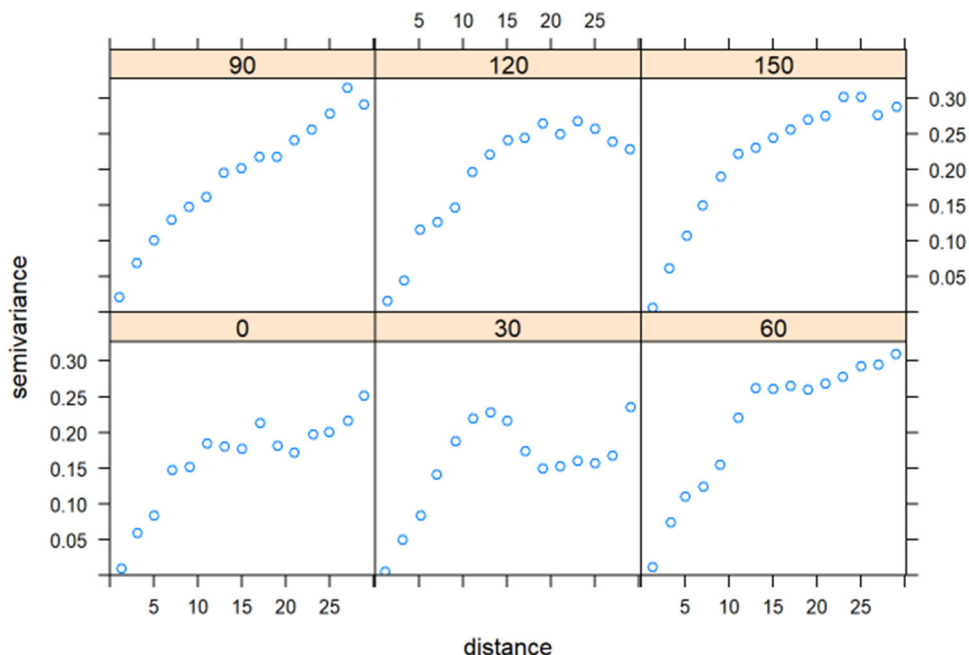


Figure 5. Experimental variograms in various directions in the North – East plane.

Further validation was sought using moment of inertia tensor calculations. After determining the moment of inertia (MOI) tensor from the covariance volume of the data, eigen vector decomposition of the MOI provided the principal directions. The resulting major principal direction was observed at 90 degrees and the dip angle in the direction of the major principal

angle measured from the horizontal plane was 45 degrees. These findings agree with the initial assessment of anisotropy. The fitted exponential variogram models tailored to these principal directions are shown graphically in **Figure 6**. The range for the major direction was determined to be approximately 20 ft. The range in the minor direction (45 degrees) was inferred to be approximately 60% of the range in the major direction, or 12 ft.

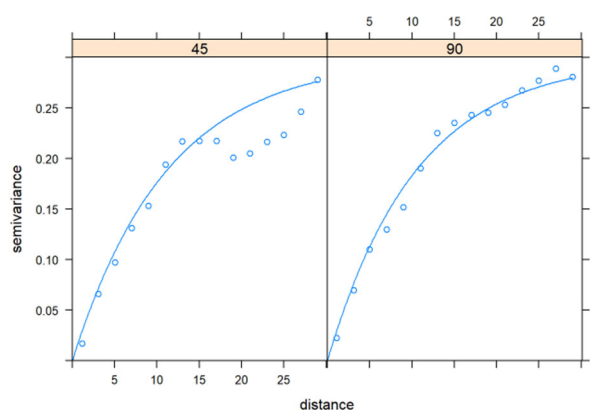


Figure 6: Variogram models at major (90 degrees) and minor (45 degrees) directions

4.2 Indicator kriging and volume estimation

IK was performed using the 3-D anisotropic indicator variogram described in the previous section and the dataset described in Section 2. *R* scripts were written to implement IK using the *vgm* function in the *Gstat* package (Pebesma, 2004; Gräler et al., 2016). The model domain spans the 3-D volume encompassing the data available and was discretized into cubic voxels with dimensions of 2 ft in each direction.

To address the uncertainty inherent in the IK-estimated probabilities of exceedance, a 95% confidence interval was calculated for each predicted probability as described in Section 3.6. Voxels with a lower limit of the confidence interval exceeds 0.05 were identified as potentially contaminated. This demarcation process resulted in the formation of three disconnected regions of internally contiguous voxels with a total volume of 5760 ft³. **Figure 7** provides a 3-D visualization of the regions where dose rate is predicted to exceed 1 R/hr with at least 5% probability and 95% confidence. Similar results were obtained when considering a probability threshold of 1%. Whereas **Figure 7** shows a 3-D view of the regions without any building references, **Figure 8** shows the plumes projected to a bird's eye view (northing as x-axis, easting a y-axis) with B-cell inner and outer wall boundaries identified. **Figure 9** shows a projected side view facing East on the left panel and North on the right panel. Inner and outer B-cell walls are shown, with a rectangular excavation volume identified. In **Figures 5 – 7**, red spheres are used to label the 2-ft³ voxels within the regions delineated. **Table 2** summarizes the estimated contaminant extent and volume, as represented by the red rectangles shown in **Figure 8** and **Figure 9**.

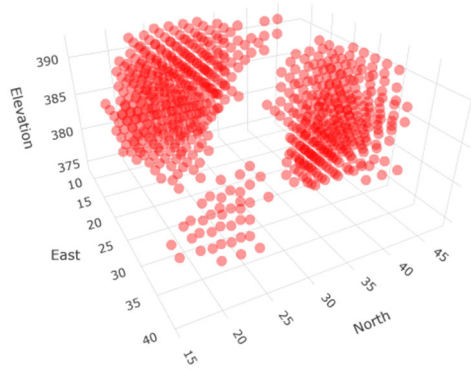


Figure 7: Indicator kriging results for the regions where contaminant levels are predicted to exceed 1 R/hr with 5% probability and 95% confidence. Voxels occupying these regions are labelled with red spheres. Elevations are ft above mean sea level. Northing and easting values are in feet relative to a local origin. Note that contamination is expected outside the volumes labelled with red spheres, which only indicate where levels exceed 1 R/hr with the probability and confidence described above.

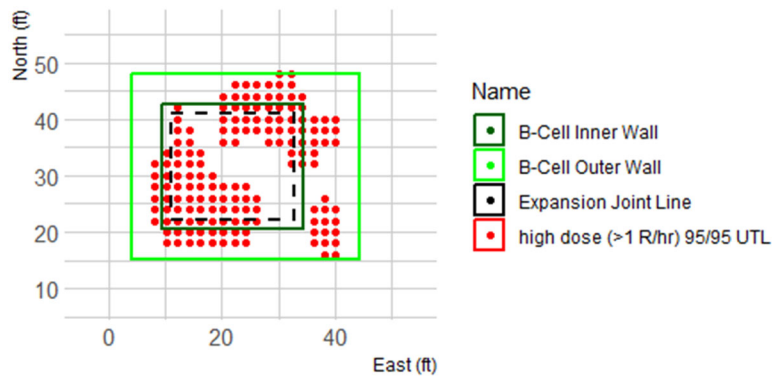


Figure 8: North and East view of extent. Northing and easting coordinates are relative to a local origin. Note that contamination is expected outside the volumes labelled with red spheres, which only indicate where levels exceed 1 R/hr with the probability and confidence described above.

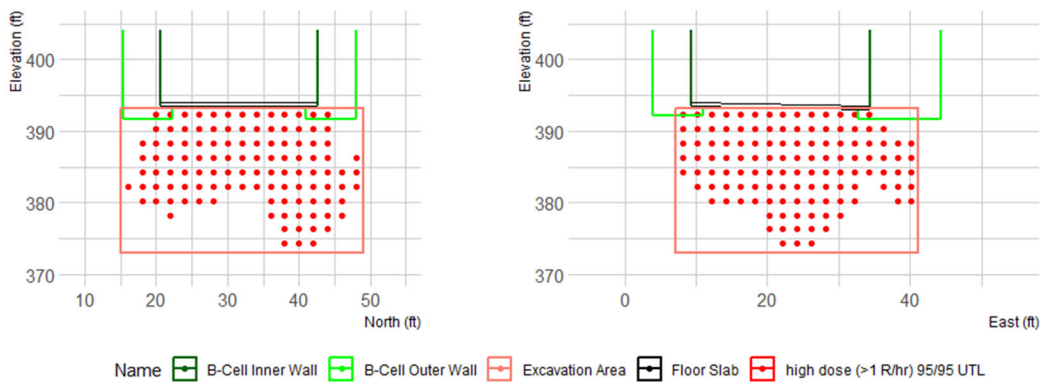


Figure 9: Northing and Easting by elevation extents. Elevation units are ft above mean sea level. Note that contamination is expected outside the volumes labelled with red spheres, which only indicate where levels exceed 1 R/hr with the probability and confidence described above.

Table 2: Extent and volume estimates based on indicator kriging 95%/95% tolerance regions. Note that contamination is expected outside this volume, which encompasses where levels exceed 1 R/hr with the probability and confidence described in the text.

Elevation (ft)	North (ft)	East (ft)	Volume (ft ³)
20	34	34	23120

5.0 Discussion and Conclusions

Indicator kriging was used to calculate the probability of exceeding the 1-R/hr threshold, an important decision level, throughout the model domain using data available from oriented boreholes and a model of spatial variability inferred from the data. The dataset proved amenable to variogram inference and IK, as closely spaced measurements were available along a variety of 3-D orientations. The IK results were used to identify where the 1-R/hr threshold was exceeded with 5% probability and 95% confidence. Three disconnected regions were identified. An effective remedy for these regions would ensure that contaminated regions exceeding 1-R/hr level would be addressed with 95% probability and 95% confidence (95%/95%).

Any mitigation strategy aiming to target the three identified regions, which are irregularly shaped, through excavation and removal of each individually would be challenging and impractical. Conventionally, contamination removal involves employing an excavator to remove a regular volume bounding contaminated sediments. To define a regular volume that encapsulates the >1 R/hr contamination, a bounding box is shown in **Figure 7**. Accounting for the 20-ft vertical (elevation) extension and 34-ft lateral extent in northing and easting directions (**Table 2**), the total volume that would need to be excavated to remove the box thus defined is 23,120 ft³.

The volume of sediment identified at potentially exceeding the 1-R/hr threshold is similar to the recent volume estimate made by CPCCO through deterministic interpretations of new soil dose-rate data from 2022-2023. Compared to conventional, non-statistically based interpretation, however, the IK approach provides a framework to assess and quantify uncertainty. Although this work focused on the 1-R/hr threshold, other thresholds could be considered. Whereas the focus here was assessing probabilities of exceedance, the approach could be used to estimate dose rates throughout the model domain and with associated quantifications of uncertainty.

6.0 References

- De Iaco, S., Hristopulos, D.T., and Lin, G. 2022. Special Issue: Geostatistics and Machine Learning. *Mathematical Geosciences* 54 (3) <https://doi.org/10.1007/s11004-022-09998-6>.
- Gräler, B., Pebesma, E., Heuvelink, G. 2016. Spatio-Temporal Interpolation using gstat. *The R Journal* 8, 204-218 <https://journal.r-project.org/archive/2016/RJ-2016-014/index.html>.
- Hoef, J.M. Ver, and Cressie, N. 1993. Multivariable Spatial Prediction. In *Mathematical Geology* 25 (2).
- Msengwa, A.S. 2021. Geostatistics for Environmental Scientists. *JOURNAL OF THE GEOGRAPHICAL ASSOCIATION OF TANZANIA* 41 (1) <https://doi.org/10.56279/jgat.v41i1.13>.
- Noel Cressie. 2015. *Statistics for Spatial Data*. Wiley.
- Pebesma, E.J. 2004. Multivariable geostatistics in S: The gstat package. *Computers and Geosciences* 30 (7). <https://doi.org/10.1016/j.cageo.2004.03.012>
- Rockhold, M., Lindberg, M.J., Bacon, D.H., Clayton, C.E., and Freedman, D.L. 2012. Numerical Modeling of 90Sr and 137Cs Transport from a Spill in the B-Cell of the 324 Building, Hanford Site 300 Area, PNNL-21214.
- Rockhold, M. 2015 Neutron Moisture Logging, Detector Testing, and Assessment of Contaminant Stability under the 324 Building, Hanford Site 300 Area, PNNL-24664.
- Rodrigues, M.S., Castrignanò, A., Belmonte, A., da Silva, K.A., and da Trindade Lessa, B.F. 2020. Geostatistics and its potential in Agriculture 4.0. *Revista Ciencia Agronomica* 51 (5) <https://doi.org/10.5935/1806-6690.20200095>.
- WCH. 2011. Characterization of the Soil Contamination Under 324 B Cell, Calculation No. 0300X-CA-N0140, Rev. 1., Washington Closure Hanford, Richland, Washington.

Appendix A – Dataset

A .zip file accompanies this report and contains a Data Package with the contents described below.

Document Package Contents:

- Python_probe_data_list (folder)
 - csv_files_full_list (folder)
 - probe_id_data_list.csv
 - xyz_start_end_empty.csv
 - probe_id_data_list_maker.py
 - requirements.txt
- Python_xyz_plus_data (folder)
 - csv_files_full_list (folder)
 - 324_probe_points_and_data_out.csv
 - probe_id_data.csv
 - xyz_start_end.csv
 - xyz_pts.py
 - requirements.txt
- SolidWorks (folder, contains probe geometry)
- B-Cell Soil GeoProbe Dose Rates & Temperatures (Oct 2014).xls (data for probes C8207 to C9438)
- ECR-22-000607-00 (With RLDs 1-5).pdf
- Probe_Coordinates_from_SolidWorks.xlsx
- Stabilization Casing Dose Profile Data Graphs.xlsx (data for probes E-01 to N-39)

Methods:

The first step in the extraction process is to extract and compile the positional data from the provided SolidWorks file for the probes of interest. For each probe in the SolidWorks geometry file, the coordinates of the start and end of the probes are extracted and put into the first sheet in the spreadsheet titled “Probe_Coordinates_from_SolidWorks.xlsx.” Some of the probes contained a bend, i.e., a change in orientation, for which the coordinates of the start, bend, and end are recorded. For this initial extraction of coordinates, an arbitrary origin local to the SolidWorks model is selected. This first set of coordinates are transformed so that the origin aligns with the start of probe E-01. A coordinate system available for probe E-01 is found in “ECR-22-000607-00 (With RLDs 1-5).pdf”, H-3-318562, Rev 2, Sheet 9. From there the coordinates are transformed from feet to meters and aligned with the coordinate specified for E-01 in “ECR-22-000607-00 (With RLDs 1-5).pdf”, H-3-318562, Rev 2, Sheet 9. The final start and ending coordinates are arranged into a list with the Probe ID, Starting, and Ending coordinates which is then converted to a CSV file to be read and processed in late steps by a python script.

The first python script in the folder “Python_probe_data_list” takes the list of probes from the “xyz_start_end_empty.csv” file to create a list of data from the probes. This is an empty set of coordinate data because the only thing that is of interest are the probe IDs for which coordinate data is available. The title of the python script is “probe_id_data_list_maker.py”. The output from the python script for two probes is shown in Figure A1, which is titled “probe_id_data_list.csv”. The probe length data goes into the first row of the probe ID.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	C9437															
2	C9437_mr															
3	C9437_mr115															
4	C9437_r															
5	C9437_AMP-50															
6	C9437_AMP-100															
7	C9437_AMP-200															
8	C9437_ML_Gross_mr															
9	C9437_ML_Net_mr															
10	C9437_ML_Net_r															
11	C9437_B_Gross_r															
12	C9437_B_Net_r															
13	C9437_Temp															
14	C9437_Date															
15	C9435															
16	C9435_mr															
17	C9435_mr115															
18	C9435_r															
19	C9435_AMP-50															
20	C9435_AMP-100															
21	C9435_AMP-200															
22	C9435_ML_Gross_mr															
23	C9435_ML_Net_mr															
24	C9435_ML_Net_r															
25	C9435_B_Gross_r															
26	C9435_B_Net_r															
27	C9435_Temp															
28	C9435_Date															

Figure A1: Output of python script “probe_id_data_list_maker.py”, List of Probe ID and associated data

From this output, the data associated with the probes are manually copied and pasted from the spreadsheets “B-Cell Soil GeoProbe Dose Rates & Temperatures (Oct 2014).xls” and “Stabilization Casing Dose Profile Data Graphs.xlsx.” This manual preprocessing puts the data into a uniform format that the python script can read and converted to a common coordinate system (north, east, elevation). Raw data from the spread sheets is organized by specifying data for points along the lengths of the probes. An example of the manually processed data is shown in Figure A2. The numbers in the row “C9437” and “C9435” correspond to the distance in feet from the north (or top) end of the probe.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Probe ID															
2	C9437	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
3	C9437_mr															
4	C9437_mr115															
5	C9437_r															
6	C9437_AM	0.016	0.017	0.016	0.016	0.014	0.015	0.013	0.012	0.012	0.012	0.013	0.016	0.016	0.016	0.018
7	C9437_AMP-100															
8	C9437_AMP-200															
9	C9437_ML	0.016	0.017	0.016	0.016	0.014	0.015	0.013	0.012	0.012	0.012	0.013	0.016	0.016	0.016	0.018
10	C9437_ML	-0.002	-0.001	-0.002	-0.002	-0.004	-0.003	-0.005	-0.006	-0.006	-0.006	-0.005	-0.002	-0.002	-0.002	0
11	C9437_ML_Net_r															
12	C9437_B_Gross_r															
13	C9437_B_Net_r															
14	C9437_Temp						75.2		75.3		76.7		77.7		79.9	
15	C9437_Dat	2014	2014	2014	2014	2014	2014	2014	2014	2014	2014	2014	2014	2014	2014	2014
16	C9435	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
17	C9435_mr															
18	C9435_mr115															
19	C9435_r															
20	C9435_AM	0.419	0.525	0.784	1.16	1.75	2.45	4.08	5.69	7.67	9.71	12.2	17	26.7	92.3	1170
21	C9435_AM	0	0	1	1	2	2	4	6	11	14	16	22	37	116	1121
22	C9435_AM	0	0	0	0	0	0	0	0	0	0	0	0	0	100	1400
23	C9435_ML	0.419	0.525	0.784	1.16	1.75	2.45	4.08	5.69	7.67	9.71	12.2	17	26.7	92.3	1170
24	C9435_ML	0.405	0.511	0.77	1.146	1.736	2.436	4.066	5.676	7.656	9.696	12.186	16.986	26.686	92.286	1169.986
25	C9435_ML	0.000405	0.000511	0.00077	0.001146	0.001736	0.002436	0.004066	0.005676	0.007656	0.009696	0.012186	0.016986	0.026686	0.092286	1.169986
26	C9435_B_C	0.000419	0.000525	0.001	0.00116	0.002	0.00245	0.00408	0.006	0.011	0.014	0.016	0.022	0.037	0.116	1.4
27	C9435_B_N	0.000405	0.000511	0.000986	0.001146	0.001986	0.002436	0.004066	0.005986	0.010986	0.013986	0.015986	0.021986	0.036986	0.115986	1.399986
28	C9435_Temp															
29	C9435_Dat	2014	2014	2014	2014	2014	2014	2014	2014	2014	2014	2014	2014	2014	2014	2014

Figure A2: Probe data manually extracted from spread sheets “B-Cell Soil GeoProbe Dose Rates & Temperatures (Oct 2014).xls” and “Stabilization Casing Dose Profile Data Graphs.xlsx”

This list “probe_id_data.csv” of position, data for the probes, and probe start and end coordinates “xyz_start_end.csv” are fed into the python script “xyz_pts.py” in the “Python_xyz_plus_data” folder. Each position point in “probe_id_data.csv” is converted into the coordinate system from “ECR-22-000607-00 (With RLDs 1-5).pdf”, H-3-318562, Rev 2, Sheet 9. These positional coordinate points are then associated with the corresponding data. If data is missing for a particular located, a flagging value of -999 is used.

To calculate the coordinates along the probe the following procedure is implemented. First, the length of the probe is calculated from the start and end points using equation (A1),

$$dist = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{A1}$$

Where $x_1, x_2, y_1, y_2, z_1, z_2$ are the starting (1) and ending (2) coordinates. Next the unit vector is calculated from the start, end, and distance,

$$i = (x_2 - x_1)/dist \tag{A2}$$

$$j = (y_2 - y_1)/dist \tag{A3}$$

$$k = (z_2 - z_1)/dist \tag{A4}$$

Then the unit vector is used to calculate the change in x, y, and z coordinates from the position along the probe,

$$dx = l/i \tag{A5}$$

$$dy = l/j \tag{A6}$$

$$dz = l/k \tag{A7}$$

With l being the length along the probe. Finally the x, y, and z coordinates are calculated from the change in each component,

$$x = x_1 + dx \tag{A8}$$

$$y = y_1 + dy \tag{A9}$$

$$z = z_1 + dz \tag{A10}$$

From this the data can be associated with the coordinates for each specified location in the data spreadsheets. The output from the python script “xyz_pts.py” is titled “324_probe_points_and_data_out.csv” and contains all of the data associated with the

coordinate points. Figure A3 shows selected set of outputs from “324_probe_points_and_data_out.csv”.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Probe ID	Depth	North	East	Elevation	mRem/hr	mRem/hr*	Rem/hr	AMP-50	AMP-100	AMP-200	Most Likely	Most Likely	Most Likely	Bounding	Bounding	Temp [F]	Date
2	C9437	50	379921.1	1949600	389.5236	-999	-999	-999	0.016	-999	-999	0.016	-0.002	-999	-999	-999	-999	2014
3	C9437	51	379920.1	1949601	389.4985	-999	-999	-999	0.017	-999	-999	0.017	-0.001	-999	-999	-999	-999	2014
4	C9437	52	379919.1	1949601	389.4735	-999	-999	-999	0.016	-999	-999	0.016	-0.002	-999	-999	-999	-999	2014
5	C9437	53	379918.1	1949601	389.4485	-999	-999	-999	0.016	-999	-999	0.016	-0.002	-999	-999	-999	-999	2014
6	C9437	54	379917.1	1949601	389.4235	-999	-999	-999	0.014	-999	-999	0.014	-0.004	-999	-999	-999	-999	2014
7	C9437	55	379916.1	1949601	389.3984	-999	-999	-999	0.015	-999	-999	0.015	-0.003	-999	-999	-999	-999	2014
8	C9437	56	379915.1	1949601	389.3734	-999	-999	-999	0.013	-999	-999	0.013	-0.005	-999	-999	-999	-999	2014
9	C9437	57	379914.1	1949601	389.3484	-999	-999	-999	0.012	-999	-999	0.012	-0.006	-999	-999	-999	-999	2014
10	C9437	58	379913.1	1949601	389.3233	-999	-999	-999	0.012	-999	-999	0.012	-0.006	-999	-999	-999	-999	2014
11	C9437	59	379912.1	1949601	389.2983	-999	-999	-999	0.012	-999	-999	0.012	-0.006	-999	-999	-999	-999	2014
12	C9437	60	379911.1	1949601	389.2733	-999	-999	-999	0.013	-999	-999	0.013	-0.005	-999	-999	-999	-999	2014
13	C9437	61	379910.1	1949601	389.2482	-999	-999	-999	0.016	-999	-999	0.016	-0.002	-999	-999	-999	-999	2014
14	C9437	62	379909.1	1949601	389.2232	-999	-999	-999	0.016	-999	-999	0.016	-0.002	-999	-999	-999	-999	2014
15	C9437	63	379908.1	1949601	389.1982	-999	-999	-999	0.016	-999	-999	0.016	-0.002	-999	-999	-999	-999	2014
16	C9437	64	379907.1	1949601	389.1731	-999	-999	-999	0.018	-999	-999	0.018	0	-999	-999	-999	-999	2014
17	C9437	65	379906.1	1949601	389.1481	-999	-999	-999	0.021	-999	-999	0.021	0.003	-999	-999	-999	-999	2014
18	C9437	66	379905.1	1949601	389.1231	-999	-999	-999	0.022	-999	-999	0.022	0.004	-999	-999	-999	-999	2014
19	C9437	67	379904.1	1949601	389.098	-999	-999	-999	0.022	-999	-999	0.022	0.004	-999	-999	-999	-999	2014
20	C9437	68	379903.1	1949601	389.073	-999	-999	-999	0.022	-999	-999	0.022	0.004	-999	-999	-999	-999	2014
21	C9437	69	379902.1	1949601	389.048	-999	-999	-999	0.022	-999	-999	0.022	0.004	-999	-999	-999	-999	2014
22	C9437	70	379901.1	1949601	389.0229	-999	-999	-999	0.024	-999	-999	0.024	0.006	-999	-999	-999	-999	2014
23	C9437	71	379900.1	1949601	388.9979	-999	-999	-999	0.024	-999	-999	0.024	0.006	-999	-999	-999	-999	2014
24	C9437	72	379899.1	1949601	388.9729	-999	-999	-999	0.023	-999	-999	0.023	0.005	-999	-999	-999	-999	2014
25	C9437	73	379898.1	1949601	388.9478	-999	-999	-999	0.023	-999	-999	0.023	0.005	-999	-999	-999	-999	2014
26	C9437	74	379897.1	1949601	388.9228	-999	-999	-999	0.024	-999	-999	0.024	0.006	-999	-999	-999	-999	2014
27	C9437	75	379896.1	1949601	388.8978	-999	-999	-999	0.023	-999	-999	0.023	0.005	-999	-999	-999	-999	2014
28	C9437	76	379895.1	1949601	388.8727	-999	-999	-999	0.022	-999	-999	0.022	0.004	-999	-999	-999	-999	2014
29	C9437	77	379894.1	1949601	388.8477	-999	-999	-999	0.021	-999	-999	0.021	0.003	-999	-999	-999	-999	2014
30	C9437	78	379893.1	1949601	388.8227	-999	-999	-999	0.022	-999	-999	0.022	0.004	-999	-999	-999	-999	2014
31	C9437	79	379892.1	1949601	388.7976	-999	-999	-999	0.021	-999	-999	0.021	0.003	-999	-999	-999	-999	2014
32	C9437	80	379891.1	1949601	388.7726	-999	-999	-999	0.023	-999	-999	0.023	0.005	-999	-999	-999	-999	2014
33	C9437	81	379890.1	1949601	388.7476	-999	-999	-999	0.024	-999	-999	0.024	0.006	-999	-999	-999	-999	2014
34	C9437	82	379889.1	1949601	388.7225	-999	-999	-999	0.024	-999	-999	0.024	0.006	-999	-999	-999	-999	2014
35	C9437	83	379888.1	1949602	388.6975	-999	-999	-999	0.023	-999	-999	0.023	0.005	-999	-999	-999	-999	2014
36	C9437	84	379887.1	1949602	388.6725	-999	-999	-999	0.022	-999	-999	0.022	0.004	-999	-999	-999	-999	2014
37	C9437	85	379886.1	1949602	388.6474	-999	-999	-999	0.021	-999	-999	0.021	0.003	-999	-999	-999	-999	2014
38	C9437	86	379885.1	1949602	388.6224	-999	-999	-999	0.02	-999	-999	0.02	0.002	-999	-999	-999	-999	2014
39	C9437	87	379884.1	1949602	388.5974	-999	-999	-999	0.018	-999	-999	0.018	0	-999	-999	-999	-999	2014

Figure A3: Output from the python script “xyz_pts.py” with data associated with the coordinates.

In compiling the data for geostatistical analysis, several issues were identified and addressed as follows:

1. A slight mismatch was found between SolidWorks spatial data and data in “ECR-22-000607-00 (With RLDs 1-5).pdf”. In the North coordinate this mismatch is approximately 0.8 feet maximum, and 0.3 feet in the East coordinate. In the SolidWorks files it appears that all E-XX probes are drawn on a single plane normal to the North direction. In the coordinate data from “ECR-22-000607-00 (With RLDs 1-5).pdf” it appears that not all E-XX probes start on the same North coordinate. The cause of the discrepancy in the East coordinate is unknown.
2. The C8215 and C8216 probes have bends in the middle of the boreholes. The starting xyz coordinate is at the bend, and the ending is at the end of the probe. It is assumed that the length of the probe is measured from the start of the bend.
3. No coordinate system reference data was available for the CXXXX probes, as is available for the E-01 probes.
4. There is spatial data for probes C8411, 8410, E-20, and E-25, but there is no measurement probe data for these probes. Also, there are probes that have measured rad data but no spatial data. Data was not used if it could not be geo-located.
5. There are inconsistencies between data from different probes, in particular between E-XX and CXXXX probes, and between 2010 CXXXX and 2014 CXXXX probes. Between the different probes/boreholes, different data sets are reported. There is some

uncertainty as to which the most important data points are, so the data header from E-01, and the 2014 data for C8207 is used.

In the Python Script headers are AMP-50, AMP-100, and AMP-200 from the 2014 data. The corresponding 2010 header are RO-7 Lo, RO-7 Mid, and RO-7 Hi. The data for RO-7 Lo was put in the AMP-50 column and so forth. The 2010 gross data was used and only one unit range was used even though both units are specified (mR/hr vs R/hr). In the 2010 data “Most Likely Net” and “Bounding Net” are absent

C8207_2014 data header:

Depth (ft-in)	Depth (ft)	Detector			Most Likely			Bounding	
		AMP-50	AMP-100	AMP-200	mR/hr Gross	R/hr Net	R/hr Net	Gross	Net

C8207_2010 data header:

Depth (ft-in)	Depth (ft)	Detector			Most Likely		Bounding	
		RO-7 Lo	RO-7 Mid	RO-7 Hi	mR/hr Gross	R/hr Gross	mR/hr Gross	R/hr Gross

Probe C9433 has “Bounding Net x 1.09”, this was left out of the data

Probe C9437 has no “Bounding” or “Most Likely” headers. Placed “Gross” and “Net” into the “Most Likely” columns. Also in the header for probe C9437 are “Adjusted Net” and “Adjusted Net x 1.09”. The data for these heading is left off.

Probe C9437 headers:

Depth (ft-in)	Depth (ft)	Detector			mR/hr			R/hr		
		AMP-50	AMP-100	AMP-200	Gross	Net	Adjusted Net	Adj. Net x 1.09		°F

Appendix B – Data Processing

Data cleaning process steps (pseudo code)

Go through the process flow diagram in more detail

Pseudo code is provided to describe the data processing steps:

Import dependencies dplyr, tidyr, lubridate, readr, readxl

Import 324_probe_points_and_data_out.csv as probe_data

Import Stabilization Casing Dose Dates.xlsx as dose_dates

In probe_data replace -999 values with NA

From probe_data extract the first letter from Probe.ID to get the borehole type as ProbeLetter

Left join ShortDate from dose_dates to probe_data on Probe.ID

If mRem.hr.1.15 is not NA

 Extract mRem.hr.1.15 as ExposureRate_mRem.hr.

Else if mRem.hr is not NA

 Extract mRem.hr as ExposureRate_mRem.hr.

Else if Rem.hr is not NA

 Extract Rem.hr

 Calculate ExposureRate_mRem.hr. as Rem.hr*1000

Else

 Extract Most.Likely.Gross..mRem.hr. as ExposureRate_mRem.hr.

If spatial coordinates are duplicated

 Jitter coordinate

Calculate ExposureRate_Rem.hr. as ExposureRate_mRem.hr./1000

Calculate t as time difference from Date to 6/1/2023

Calculate Corrected_Rem.hr. as ExposureRate_Rem.hr.*exp(-(log(2)/29.5)*t)

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov