Pacific Northwest
NATIONAL LABORATORY

PNNL-32664, Rev. 2

Recommendations for VSP Enhancements for Continuously Collected Survey Data

Fulfillment of Task 1a of 31310021F0022 and Task 2 of 31310023F0040

December 2024

Deb Fagan
Moses Obiri
Zachary Weller
Lisa Newburn
Jen (Huckett) Willis
Amoret Bunn
Jan Irvahn
Alejandro Heredia-Langner

Prepared for the U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Under Contract DE-AC05-76RL01830
Interagency Agreement: 31310019N0001
Task Order Number: 31310023F0040

DISCLAIMER

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;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <https://www.ntis.gov/about>
Online ordering: http://www.ntis.gov

Recommendations for VSP Enhancements for Continuously Collected Survey Data

Fulfillment of Task 1a of 31310021F0022 and Task 2 of 31310023F0040

December 2024

Deb Fagan
Moses Obiri
Zachary Weller
Lisa Newburn
Jen (Huckett) Willis
Amoret Bunn
Jan Irvahn
Alejandro Heredia-Langner

Prepared for the U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Under Contract DE-AC05-76RL01830
Interagency Agreement: 31310019N0001
Task Order: 31310023F0040

Pacific Northwest National Laboratory
Richland, Washington 99354

Acronyms and Abbreviations

AI	artificial intelligence
CFR	U.S. Code of Federal Regulations
CPM	counts per minute
DCGL	derived concentration guideline level
DCGLW	site-wide derived concentration guideline level
FRK	fixed rank kriging
FSS	final status survey
GIS	geographic information system
GLS	generalized least squares
GSLIB	Geostatistical Software Library
GPS	global positioning system
HPGe	high purity germanium
IL	investigation level
IL_{pp}	*a posteriori* investigation level
L_C	critical level
LISA	local indicator of spatial association
LOD	limit of detection
LOO	leave-one-out
MARSSIM	Multi-Agency Radiation Survey and Site Investigation Manual
MDA	minimum detectable amount
MDC	minimum detectable concentration
MDCR	minimum detectable count rate
ML	machine learning
NRC	U.S. Nuclear Regulatory Commission
OLS	ordinary least squares
PNNL	Pacific Northwest National Laboratory
SADA	Spatial Analysis and Decision Assistance
UAS	uncrewed aerial system
UAV	uncrewed aerial vehicle
UGS	uncrewed ground system
UGV	uncrewed ground vehicle
USL	upper simultaneous limit
UTL	upper tolerance limit
UXO	unexploded ordnance
VSP	Visual Sample Plan

Contents

Figures

Tables

1.0 Introduction

The U.S. Nuclear Regulatory Commission (NRC) has responsibility for regulating the safe decommissioning of facilities and sites to meet the License Termination Rule in Title 10 of the Code of Federal Regulations Part 20 (10 CFR Part 20), Standards for Protection Against Radiation, Subpart E, "Radiological Criteria for License Termination." Decommissioning is performed in accordance with 10 CFR Part 50, Domestic Licensing of Production and Utilization Facilities, as part of termination of license (10 CFR Part 50.82) and release of the facility or site for unrestricted use (10 CFR Part 50.83). Key guidance for demonstrating a facility or site meets these regulations, including radiological surveys, is provided in NUREG-1507, Revision 1, Minimum Detectable Concentrations with Typical Survey for Instruments for Various Contaminants and Field Conditions (NRC 2020a); NUREG-1757, Consolidated Decommissioning Guidance (NRC 2020b); and NUREG-1575, Revision 2, Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM/NUREG-1575) (NRC 2020c). The guidance currently demonstrates the minimum requirements and necessary conditions for conducting radiological surveys of surface soils and structures by a person carrying a radiation detector(s).

Radiological surveys to support decommissioning sites and facilities that may contain radioactive contamination involve both static measurements and scan measurements (scanning). Static measurements are aimed at characterizing an overall mean level of residual contamination but are unlikely to detect small areas of elevated activity because they are collected over a relatively limited area. Radiological scanning surveys, which allow for greater spatial coverage over a comparable study period, can guarantee a higher probability of detecting an area of elevated activity if it exists on a site.

Techniques for scanning surveys have traditionally involved surveyors moving instruments over surface or land areas and responding to audio output from the instrument as described in MARSSIM/NUREG-1575, NUREG-1507, and the Handbook for MARSSIM Users (NRC 2020c; NRC 2020a; Abelquist 2014). When the audio input changes from what is being observed under a "no residual contamination" paradigm, a surveyor changes the survey parameters (speed, height, path) in real time to gain a detailed characterization (size, shape, radioactivity) of the area that might contain residual contamination before resuming the survey under the planned parameters. This paradigm is referred to as "surveying with vigilance" (Fortin et al. 2023).

Licensees have modernized the methods used to perform radiological surveys since MARSSIM/NUREG-1575 (NRC 2020c) and NUREG-1507 (NRC 2020a) were published, by increasingly using additional survey instrumentation and data capture tools, including global positioning system (GPS) and geographic information system (GIS) technologies. GPS/GIS-based scan surveys deviate from the NUREG-1507 (NRC 2020a) approach in that they are now regularly conducted without surveyor vigilance, using autonomous vehicles, towing vehicles, or human surveyors that collect data continuously without identifying in real time areas where residual contamination might be present. When data are collected in a continuous and non-vigilant manner, additional planning and new statistical tools are needed to account for the differences in data collection, management, visualization, and analyses that might be required over traditional methods.

The Science Advisory Board states that current MARSSIM/NUREG-1575 (NRC 2020c) and NUREG-1507 (NRC 2020a) guidance does not adequately address modern GPS/GIS-based scanning surveys. Arising from significant technological advancement over the past two decades, newer scanning instruments and mobile systems represent attractive options for

consideration and assessment. The Science Advisory Board does not endorse specific detection systems or commercial equipment but does emphasize the importance of detection system calibration to yield measurement quantification with uncertainties that can support defensible final survey results collected using GPS/GIS-based scanning surveys.

This report identifies options for additional tools that could be required to facilitate data import, management, visualization, and analysis using continuous (without vigilance) methods for GPS/GIS-based surface (two-dimensional) radiological surveys. The discussions provided in each section are summarized as follows:

- Section 2: two types of correlation that could be present in spatial data, one is the result of how contamination occurred at a site and the other is due to the method of continuous data collection using GPS/GIS-based scan surveys.
- Section 3: two methods for accounting for spatial correlation in GPS/GIS-based final status survey (FSS) data analysis. Generalized least squares (GLS) is a method that could be implemented in Visual Sample Plan (VSP) prior to post-survey analysis. Machine learning (ML) is a method relevant to determining the boundaries of areas with elevated residual contamination that could be implemented to analyze data collected during the survey (i.e., post-survey analysis).
- Section 4: methods for determining potential areas of elevated contamination and boundaries of areas to revisit when GPS/GIS-based scan surveys are performed. Methods implemented *a posteriori* include upper tolerance limits (UTLs), upper simultaneous limits (USLs), and z-scores and methods implemented either *a priori* or *posteriori* include lag-k, and scan MDC. Methods that use observations to predict values at locations without associated observations include kriging methods and possibly ML.
- Section 5: considerations for importing data from continuously collected data from GPS/GIS-based scan surveys into VSP. In general, VSP can receive a large amount of spatial data without issue. Data management and pre-processing must generally be performed outside of VSP so that the analysis-ready dataset is imported.
- Section 6: visualization of continuously collected data. The amount of data from GPS/GIS-based surveys resulting in continuously collected data and the spatial area over which the data are collected would benefit from a few additional tools in VSP.
- Section 7: summarizes conclusion and recommendations made throughout the report in a central location.
- Section 8: provides references to citations made throughout the report.

Data quality assessment is vital for continuously collected data, and the methods discussed in this report assume that the data are adequate to support a decision. GPS/GIS-based survey data that are collected at a higher velocity or a higher altitude than what is planned are biased (tend to underestimate contamination) and could inhibit the ability to detect unacceptable levels of residual radioactivity. VSP has some capability for data quality assessment, such as posterior power curves for the quantile test implemented in MARSSIM/NUREG-1575 Scenario B (NRC 2020c). However, additional tools to assess the quality of continuously collected data are needed. Pacific Northwest National Laboratory (PNNL) implemented several data quality assessment methods in Bunn et al. (2022) and recommends additional work to identify methods to incorporate into VSP enhancements for continuously collected GPS/GIS-based survey data.

We have prepared two datasets to use as case studies and test datasets for data quality assessment analysis and new hot spot identification improvements, should we need them for testing and evaluating future VSP enhancements. The scenarios were provided by NRC for the Fortin et al. (2023) report. The datasets have been modified to better reflect an evaluation of a land area for the presence of small areas of elevated residual radioactivity. Appendix A discusses these datasets and their application to Scenarios A and B of MARSSIM/NUREG-1575 (NRC 2020c).

2.0 The Effects of Spatial Correlation on Statistical Tests

The statistical decision-making techniques (hypothesis testing, calculating statistical intervals) in MARSSIM/NUREG-1575 (NRC 2020c) rely on a set of assumptions regarding 1) the probability distribution of the population (or a population parameter, such as a mean concentration), 2) the spatial or temporal dependence present in the population distribution, and 3) the error structure of observations within a dataset. NUREG-1505 (NRC 1998) provides a nonparametric statistical methodology that loosens the requirements for underlying probability distributions, but assumptions about the spatial- and temporal-dependence, as well as sample assumptions, are still required to use these methods appropriately.

A general model for the relationship between observed data and a set of predictors is:

$$Y = f(X) + \varepsilon \qquad \text{Equation 1}$$

In Equation 1, Y is a quantity of interest, X is the set of predictors, f is a function that defines the relationship between Y and X, and ε is error (e.g., measurement error and/or model error). Often, statistical analysis assumes that observed data are sampled from a population whose members are independent of one another and randomly mixed, and that individual error terms, ε, are uncorrelated in such a way that the probability of a value taken on by one model error term has no effect on the probability of any of the remaining model error terms.

It is important to keep the distinction between the population characteristics (spatial, temporal independence) and the sample characteristics (distribution and independence of errors) in mind, as the toolbox of statistical analysis methods relies on assumptions about both. Sample characteristics can lead to exogenous correlation due to GPS/GIS-based survey collection factors, such as scanning speed, height, and detector specifications. Related terminology is introduced in Section 2.1. Endogenous correlation is related to the population characteristics, including distribution of possible residual contamination at a site, through processes such as deposition, transport, and decay of radionuclides, and applies to the concentrations or radiological activity at a site. Endogenous correlation is discussed in Section 2.2. Section 2.3 demonstrates the potentially deleterious effects that not accounting for these types of correlations can have on statistical inference through a case study.

2.1 Exogenous Correlation

Final status GPS/GIS-based surveys (conducted with or without vigilance) resulting in continuously collected data often induce spatial or temporal correlation due to the survey speed required to scan minimum detectable concentrations (MDCs) and the nature of radiation detection. As an example, Figure 1 shows a set of data collected along an outdoor transect from an experiment to compare human and UAV surveys (Bunn et al. 2022). The experiment used button check sources placed at known locations along a transect and a continuous scan survey with a sodium iodide scintillation detector along a survey path over the transect. The result is a dataset with the check sources evident as symmetric peaks—the data within a peak represent a single check source. Data collected in this way are clearly correlated, as points nearer to each other in time (and space) have radiation measurements more similar to each other than those further from each other.

The model in Equation 1 is modified to capture this autocorrelation by representing the radiation detected at each point in time as a function of previous detections. It uses a time index, t, associated with the movement of the detector over the transect as follows:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \qquad \text{Equation 2}$$

In Equation 2, Y_t is the observed value at time t, the ϕ's are the weights of the previous observations collected at times $(t-1), (t-2), \ldots, (t-p)$, and ε_t is the error at time t. In Equation 2, predictions of the current value are equal to a weighted combination of previously observed values. This model is called an autoregressive model of lag-k, where the correlation structure is identified through the ϕ's. The purpose of this model is to demonstrate how the dependency between observations can be captured. Although on its own it does not calculate an overall average or identify elevated areas, subsequent analysis could be done to these ends.

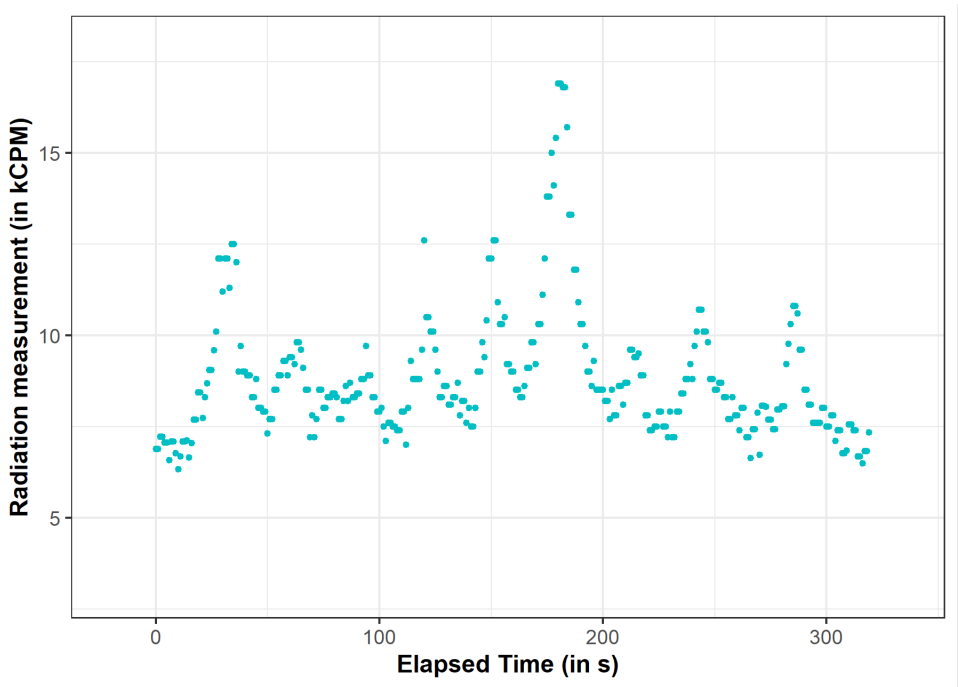

Figure 1. Example of a continuous scan survey transect over a set of button sources (see Bunn et al. 2022). Autocorrelation is evident as smooth peaks.

2.2 Endogenous Correlation

Endogenous correlation is introduced by the nature in which radioactivity is deposited and/or transported at a site. For surface contamination, endogenous correlation might be the result of air deposition from a stack, leakage from a tank, or transport of previously deposited radioactivity via surface water, animals, or weather, resulting in a spatial trend or pattern to the deposited contamination. As explained in Abelquist (2001), the result is a lack of stationarity in the mean or variance of the underlying contaminant distribution across the site. The model in Equation 1 can be modified in a few of different ways to reflect these circumstances, either by stating that the Xs are coordinates (x, y, z), similar to Equation 1, or through a spatial correlation model where the Xs are actually nearby (near in space and/or time) observed values (Ys), similar to Equation 2

Whether to use Equation 1 or Equation 2 as the statistical model for data evaluation depends on how such analysis will be used to make decisions about the site. If the goal is to compare a site parameter (mean or upper percentile) to an action limit, Equation 1 is the proper choice. If the goal is to predict values at unsampled locations or to estimate boundaries of elevated areas that might require further remedial action, Equation 2 is particularly useful.

2.3 The Effect of Spatial Autocorrelation

Spatial autocorrelation measures the spatial similarity of a set of geographically located variables and includes both exogenous and endogenous correlation (Beale et al. 2010; Cressie 2015). Model-based statistical inference is widely used in environmental analyses to account for spatial autocorrelation, which is represented by the assumptions made (or implied) about the model error terms. The validity of the model assumptions determines the model's reliability and performance.

Accounting for spatial autocorrelation accomplishes two goals. First, it assesses the degree and nature to which the spatial independence assumption is violated. Second, and perhaps more importantly, it determines how statistical conclusions are impacted when non-zero spatial autocorrelation is neglected. In MARSSIM/NUREG-1575 (NRC 2020c) Class I and Class II areas, for example, the working (null) hypothesis for FSSs is that the site is not acceptable for unconditional release unless FSS data are collected to refute (reject) this hypothesis and conclude the site is acceptable. The statistical hypotheses usually take the form:

$$H_o: \mu \geq action\ limit$$

$$H_a: \mu < action\ limit$$

where μ is a site parameter of interest, such as the mean Co-60 concentration in the top 6 inches of soil. Models that omit spatial autocorrelation when it is present tend to reject the null hypothesis more frequently than the nominal type I error rate (increased type I error rates), leading to conclusions that mean concentrations are below an action limit when they are not. Such tests are typically overly liberal in the presence of positive spatial autocorrelation because autocorrelation is a form of pseudo replication, resulting from observations being treated as statistically independent when they are not (Clifford et al. 1989; Dray et al. 2006; Hurlbert 1984). This effect is demonstrated by the following example, where a non-spatial model results in a p-value that leads to the conclusion that the mean concentration is below the action limit, but a spatial model does not.

The example data are shown in Figure 2, with zinc concentrations collected in a Meuse River flood plain near Stein, The Netherlands. The data are from the "meuse" dataset and a description can be found in the R sp package documentation (R Core Team 2020; Pebesma and Bivand 2005; Bivand et al. 2013; Pebesma et al. 2024). In general, higher zinc concentrations appear along the river shore and concentrations decrease as distance from the river increases, indicating spatial autocorrelation. Assume we are investigating whether there is an elevated level of zinc concentration compared to the historical average log zinc concentration value of 5.75 ppm. This can be expressed using the following statistical hypotheses testing framework:

$$H_o: \mu \geq 5.75\ (\text{mean concentration of zinc in the river exceeds the historical average})$$

$$H_a: \mu < 5.75\ (\text{mean concentration is less than the historical average})$$

Using the spatial model to account for spatial autocorrelation, the average log zinc concentration was found to have a 95% confidence interval of (5.27 ppm, 7.73 ppm). The null value of 5.75 ppm clearly lies within the 95% confidence interval and the p-value corresponding to the hypothesis test is not significant ($\alpha = 0.05$), as shown in Table 1. The non-spatial model, on the other hand, results in a significant p-value and the null hypothesis is rejected ($\alpha = 0.05$). The correct decision in this case would be made using the spatial model that accounts for spatial autocorrelation.

Table 1. Summary output from models highlighting the effect of spatial autocorrelation.

Model	Estimate	Standard Error	t-value	p-value	Result
Non-spatial model	0.1358	0.0580	2.3417	0.0205	False positive
Spatial model	0.7540	0.6296	1.1977	0.2329	True negative

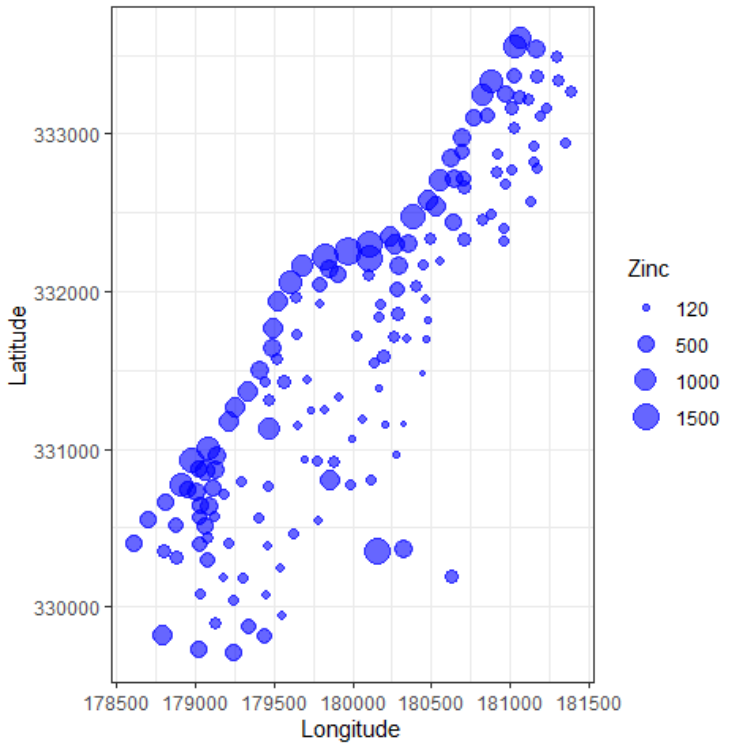

Figure 2. Zinc concentration (ppm) recorded in a flood plain of the Meuse River near the village of Stein, The Netherlands (see Pebesma et al. 2020).

A simulation of the effect of spatial autocorrelation on type I error rates is shown in Figure 3. Type I error rates from 1000 simulations of hypothesis testing were calculated using a model that uses non-spatial standard ordinary least squares (OLS) estimation that does not account for spatial autocorrelation and a model that accounts for spatial autocorrelation using a GLS model. The figure shows that in the OLS model, type I error rates increase consistently as the number of samples (pseudo replicates) increases, whereas error rates from the GLS model remain roughly constant, and at the nominal level ($\alpha = 0.05$), regardless of sample size. The GLS model is discussed further in Section 3.1.

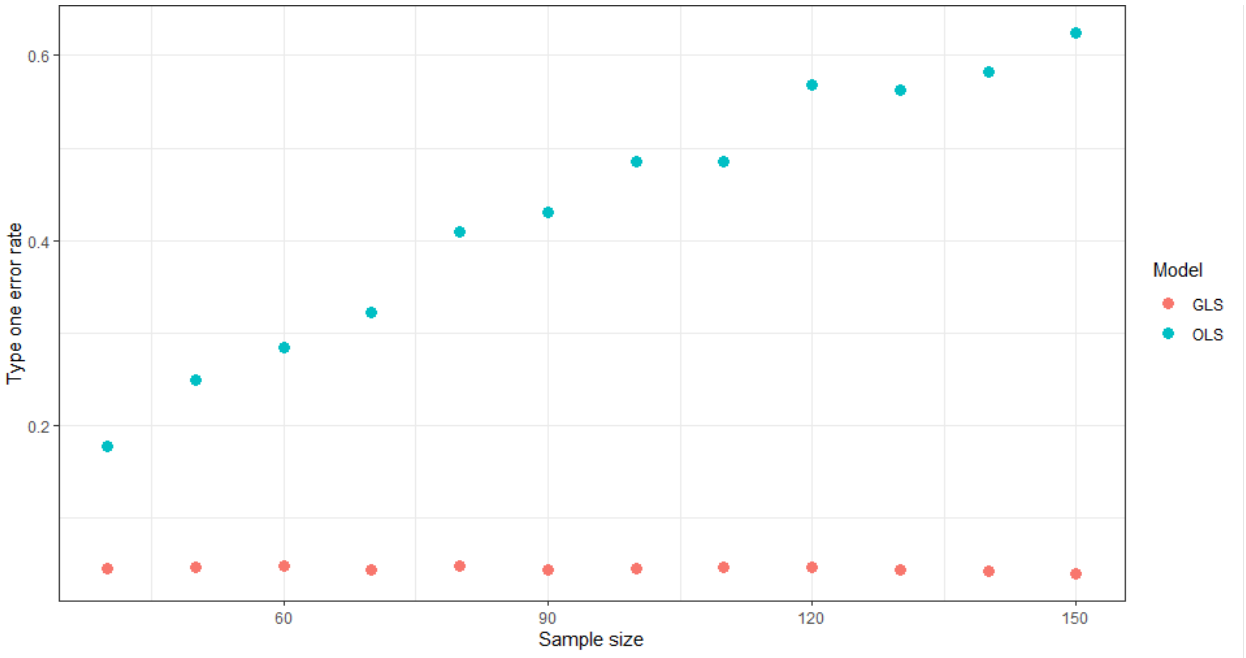

Figure 3. Type I error rates for spatial and non-spatial models.

Further, spatial autocorrelation also can compensate for unknown extrinsic and intrinsic factors that are missing from a model (Cressie 2015), where the proportion of variation explained for the dependent variable in a predictive model increases (Beale et al. 2010; Ver Hoef and Cressie 1993). As a result, regression parameter estimates are more precise, increasing statistical power for tests of these parameters. In a simulation study, Griffith and Layne (1999) reported an average 5% increase in the proportion of variability explained when spatial autocorrelation is accounted for in regression models.

3.0 Methods to Improve VSP Analysis of Continuously Collected FSS Data

In MARSSIM/NUREG-1575 (NRC 2020c) and NUREG-1507 (NRC 2020a), the integrated survey design is based on two objectives to demonstrate that a site meets or exceeds release criteria: 1) ensure that the site average activity does not exceed a dose-based threshold, and 2) ensure that there are no small areas of elevated activity that remain at a site. Objective 1 is usually determined by comparing an estimated mean concentration or activity to an action limit or by comparing a site mean to a reference mean. Per MARSSIM/NUREG-1575 (NRC 2020c) and NUREG-1507 (NRC 2020a), the background reference area is defined as an area that has physical, chemical, radiological, and biological characteristics similar to the survey unit(s) being investigated but has not been contaminated by site activities (i.e., non-impacted). A follow-on objective is that if the data indicate a small area of residual contamination may remain, the boundary of the area(s) that must be revisited to verify the extent of the elevated activity must be determined.

The methods discussed in this section are attempts to increase the statistical rigor associated with continuously collected FSS data that may contain endo- and exogenous correlation structure and are consistent with the MARSSIM/NUREG-1575 (NRC 2020c) and NUREG-1507 (NRC 2020a) hypothesis testing approach (NRC 2020c). Methods for determining whether areas must be revisited, and their boundaries, are discussed in Section 4.0.

Section 3.1 discusses the GLS model introduced in Section 2.0. By modeling the data using a GLS framework, spatial autocorrelation is accounted for without having to model a variogram, used in kriging. Section 3.2 discusses ML.

3.1 Generalized Least Squares Model

Section 2.0 showed that a GLS model that accounts for autocorrelation outperforms OLS that does not account for spatial structure when estimating a mean concentration or activity and performing hypothesis testing. Where a kriging model uses a variogram to model the spatial correlation (and results in the ability to predict values where no data are collected), kriging is not used in the hypothesis testing framework for comparing a mean to an action limit. The variogram can be used in the GLS framework to estimate the covariance matrix to account for spatial structure and perform hypothesis testing, create a confidence or tolerance interval, and calculate the number of samples required to achieve data quality objectives (EPA 2000).

3.1.1 Model Form

Consider the commonly used standard linear regression model:

$$Y = X\beta + \varepsilon \qquad \text{Equation 3}$$

In Equation 3, $Y = [Y_1, \dots, Y_N]^T$ is an $N \times 1$ column vector of response values given a sample of N observations, $X = [X_{ik};\ i = 1, \dots N, k = 1, \dots K\ ]$ is the design matrix of K predictor variables, $\beta = [\beta_1, \dots, \beta_K]^T$ is a column vector of unknown linear coefficients, and $\epsilon = [\epsilon_1, \dots, \epsilon_N]^T$ is an $N \times 1$ random error term. This model is a particular version of Equation 1 where the conditional mean of Y given X, $f(X)$, is a linear function of X, $X\beta$.

OLS is the most commonly used method for estimating the unknown coefficient β. One of the most important assumptions of OLS estimation is that the error terms are uncorrelated and have a constant variance and the covariance is a diagonal matrix $Var(\epsilon|X) = \sigma^2 I$, where I is an $N \times N$ identity matrix. However, in the presence of spatial autocorrelation, the random error terms are correlated, violating this assumption. The presence of this correlation causes the OLS estimators to no longer be efficient and has the potential to give misleading results from hypothesis tests or other inferences.

In the presence of spatial autocorrelation, GLS estimation can be used. GLS accounts for the spatial structure in the data by replacing the OLS covariance matrix with $Var(\epsilon|X) = \sigma^2 V$. The matrix V contains information about the form of the spatial structure in the data and/or can be specified using prior knowledge. It can also be estimated using the empirical variogram.

Variogram estimation is available in VSP as part of the kriging module. The equations for the GLS model are straightforward to program; Section 3.1.2 briefly discusses how the GLS model could be set up in the MARSSIM/NUREG-1575 (NRC 2020c) and NUREG-1507 (NRC 2020a) framework and included in VSP.

3.1.2 GLS in MARSSIM Framework

In the majority of MARSSIM/NUREG-1575 (NRC 2020c) environmental investigations, the statistical hypothesis testing approach is used for final decision-making to meet the first objective described above, ensuring the site average activity does not exceed a dose-based threshold. Typically, an approach must be developed for one of two cases: 1) comparing observations from the survey site to a threshold value (one-sample case), and 2) comparing observations from the survey site to observations from the reference site (two-sample case).

Case 1: One-Sample Case

The hypothesis test for the one-sample case is:

$$H_o: \mu \geq \mu_o$$

$$H_a: \mu < \mu_o$$

which can be expressed as the following linear model:

$$Y - \mu_o = X\beta + \epsilon$$

$$Y_{adj} = X\beta + \epsilon$$

where $Y_{adj} = Y - \mu_o$, $Y = [Y_1, \ldots, Y_N]^T$ is an $N \times 1$ column vector of response values given a sample of N spatially correlated observations from a survey area, and $X = [1_N]$ is $N \times 1$ design matrix of 1s. Additionally, μ is the mean concentration from the survey site, μ_o is a threshold value, and β is the average difference between the mean concentration from the survey area

and the null value (μ_o). In the two-sample case, the means from the survey area and reference area are compared such that:

$$H_o: \mu_s \geq \mu_r$$

$$H_a: \mu_s < \mu_r$$

These means can be expressed as the following linear model in which the term $Y = [Y_1^*, \ldots, Y_N^*, Y_{N+1}, \ldots, Y_{N+M}]^T$ is an $(N + M) \times 1$ column vector of response values given a sample of N adjusted or observed spatially correlated observations from a survey area and M observations from a reference area X, where

$$X = \begin{bmatrix} 1_N \\ 0_M \end{bmatrix}$$

is $(N + M) \times 1$ design matrix with 1s denoting measurements from the survey area and 0s denoting measurements from the reference area. Additionally, μ_s is the mean concentration from the survey area, μ_r is the mean concentration from the reference area, and β is the difference in mean concentration between observations from the survey and reference areas.

In both cases, the hypotheses reduce to:

$$H_o: \beta \geq 0$$

$$H_a: \beta < 0$$

As discussed in the previous section, inference using GLS estimation for β yields reliable results by accounting for any spatial structure that may exist in the data.

3.2 Machine Learning

The development and use of ML and artificial intelligence (AI) methods have dramatically increased over the last 15 years. One of the reasons for the popularity of AI/ML methods is their flexibility to discover and model complex and nonlinear relationships in large datasets, although recent advances like few-shot and zero-shot learning have also been developed for data-sparse applications. AI/ML includes a set of techniques that create analytical models by learning from data, recognizing patterns, and generating predictions and decisions based on those patterns. Contaminant fate and transport and physical characteristics of the surface could be incorporated into AI/ML methods as factors or potential predictors. For example, AI/ML could be used to predict contaminant levels at unmeasured locations by combining concentration measurements taken with and without vigilance, as well as with other surface measurements or model outputs (e.g., surface water transport, soil properties).

ML methods, such as tree-based methods, support vector machine regression, and deep learning methods, can be used to predict outcomes based on a set of independent variables. As one example, recent extensions to random forest algorithms have been developed for both global (Hengl et al. 2018) and local (Georganos et al. 2021; Ancell et al. 2021; Benito 2021) spatial regression problems. However, like non-spatial statistical methods, they may not adequately incorporate the properties of spatially correlated data. To remedy this, spatial coordinates could be used as predictor variables, although caution should be used to ensure overfitting does not result (Meyer et al. 2019). AI/ML methods have been applied to subsurface data to perform tasks such as delineating layers (Wohlberg et al. 2005), clustering observations

(Romary et al. 2015), and mapping contaminant plumes (Tao et al. 2019). Additional research and development could determine if similar approaches could be used for surface applications. AI/ML methods could also be useful for estimating the boundaries of an area with potential elevated residual contamination. A few ML methods are discussed in Section 4.2.5.

AI/ML methods can often lack interpretability, and quantifying uncertainty (to create confidence intervals or perform hypothesis tests) may require technical complexity exceeding that available to typical licensees. VSP would require research to identify which, if any, AI/ML methods are best suited to continuously collected data applications, and then subsequent algorithm development and coding to accommodate their use in characterization and/or FSSs.

4.0 Methods to Identify Areas of Potential Residual Contamination

As mentioned in Section 1.0, GPS/GIS-based scanning surveys do not benefit from real-time surveyor vigilance; thus, there is a need to identify areas that may contain elevated levels of residual elevated contamination, or hot spots, after the survey has been completed. Even in the absence of residual radioactivity, variability is expected in GPS/GIS-based survey results due to background levels, measurement uncertainty, and statistical counting errors. In combination, these sources of variability result in a distribution of measurements and possibly outliers, or observations that do not conform to the pattern established by other observations (Gilbert 1987). Gilbert (1987), Gibbons (1994), and the Environmental Protection Agency (EPA 2000) discuss the differences between outliers and elevated observations that represent true site conditions. Assuming anomalous values and outliers are corrected, there is a need for methods that can identify and delineate areas of potential concern. Identification of areas of potential concern is typically based on choosing an investigation level (IL) (NRC 2000). In this section, we focus on methods that can be used to develop an IL_{pp}, where the subscript pp indicates the IL is developed *a posteriori* using post-processed GPS/GIS-based survey data to identify areas of potential concern.

After data quality assessment has been performed to verify that the GPS/GIS-based survey was completed appropriately, two steps are necessary to determine whether post-survey areas should be revisited. The first is to identify the IL_{pp} comparison value such that any locations with observations greater than the IL_{pp} are flagged for further inspection. Once the data are compared with the IL_{pp}, boundaries of the area to revisit need to be identified.

This section discusses different options for choosing the IL_{pp}, including a UTL, USL, z-score, and local indicator of spatial association (LISA). With each method, an illustration of its implementation is provided using an example dataset.

Section 4.1 discusses methods to identify potential areas of elevated contamination (hot spots) when GPS/GIS-based surveys are completed using continuous data collection and are performed without surveyor vigilance.

Section 4.2 discusses methods to delineate areas to revisit, including a rule-based approach that is consistent with MARSSIM/NUREG-1575 (NRC 2020c) as well as geospatial methods.

4.1 Potential Areas of Elevated Contamination (Hot Spots)

4.1.1 UTL Method

The UTL method should be considered for identifying areas of elevated contamination (hot spots) if they are present. A UTL is favored over a confidence limit approach (typically used to determine if the site mean is below the site-wide DCGL or DCGLW), because it can be applied when nonuniform contamination is present. A UTL is used to estimate the p^{th} percentile of a population with a chosen level of confidence. For example, the ILpp can be estimated using the 95th percentile / 95% confidence UTL. The UTL can be based on observations from the reference area and the area of concern or just the area of concern if data from a reference area are not available.

There are several nonparametric, parametric, and quasi-parametric methods for developing a UTL (Krishnamoorthy and Mathew 2009; Davis and Wambach 2015). Parametric methods typically assume data from a population that can be described by a normal, lognormal, or gamma distribution. Nonparametric methods do not make distributional assumptions, but their degree of confidence can be limited by the number of samples. Parametric methods may enable a higher degree of confidence associated with a UTL for the same number of or fewer samples at the expense of distributional assumptions.

VSP currently implements calculation of parametric and nonparametric UTLs (Millard and Neerchal 2001; Hahn and Meeker 1991). Davis and Wambach (2015) propose an alternative quasi-parametric development of UTLs that assumes a lognormal distribution with log-scale deviation of no greater than two (2) and conservative heuristics about the number of samples exceeding certain fractions of the UTL. If distributional assumptions can be verified, parametric UTLs will typically be smaller than their nonparametric equivalent, being therefore less conservative as a comparison for finding elevated residual contamination (i.e., smaller UTL values will identify more areas with elevated contamination).

PNNL applied the UTL method as an illustration using an example dataset. The dataset was derived based on data previously provided by NRC for Fortin et al. (2023) research into lag-k scan minimum detectable concentration (MDC) research, including counts per minute (CPM) observed via scanning that included serial autocorrelation between observations. However, the CPM observations in the reference area were generally higher than in the area of concern. Therefore, PNNL injected the data with elevated observations to demonstrate the hot spot detection methods. The resulting synthetic dataset used in this example is shown in Figure 4, with derived observations shown on the left panel and hot spot locations identified in the right panel when observations were greater than a threshold (the threshold was set to 14,000 CPM for the purpose of demonstration). Setting the threshold provides a ground truth for the demonstration, where UTL, USL, and z-score results can be compared to the 23 points identified as hot spots (>14,000 CPM) and 333 points identified as cold spots (<14,000 CPM).

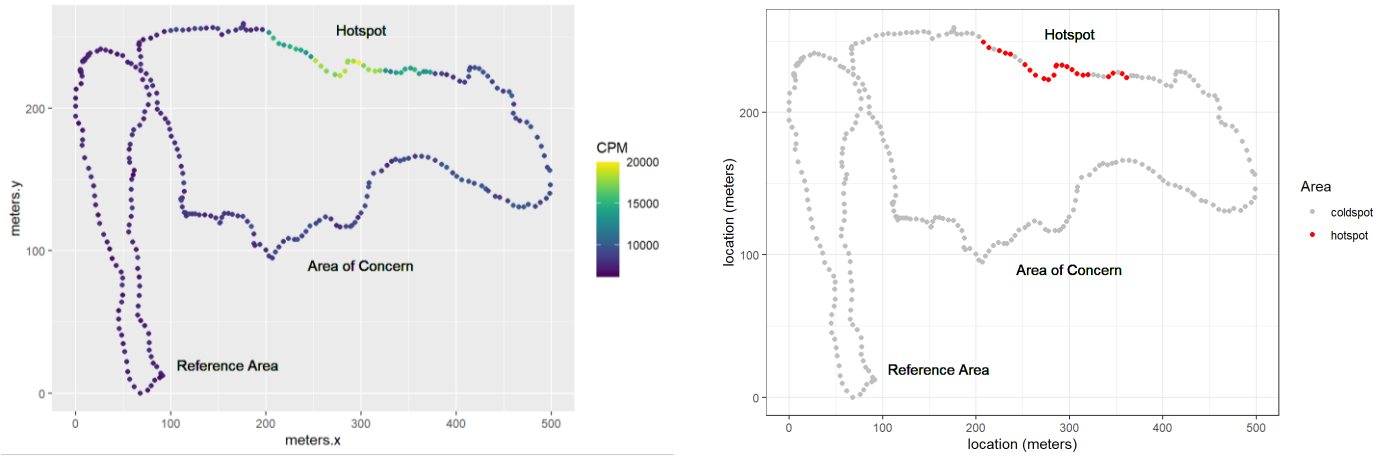

Figure 4. Example dataset where the left panel shows observed CPM, and the right panel indicates observations that are hot spots (>14,000 CPM points are red) and cold spots (<14,000 CPM points are grey).

The UTL method identifies the IL_{pp} comparison value based on a selected percentile of the data distribution. Any observations greater than that IL_{pp} are identified as potential elevated areas or hot spots. In this example, an IL95/95 is estimated based on the 95th percentile / 95% confidence UTL, shown by the dashed line in Figure 5, just less than 18,000 CPM.

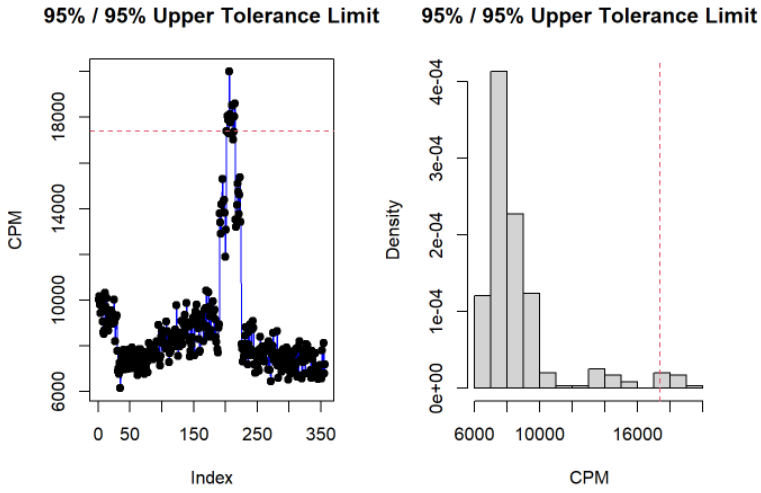

Figure 5. Example UTL and observations. The left panel shows CPM observations compared to the 95th percentile UTL, and the right panel shows the distribution of CPM observations.

The 95th percentile / 95% confidence UTL identified all cold spots correctly but only about 50% of the hot spots. It incorrectly identified 12 of the 23 hot spots as cold spots, as shown in Figure 6.

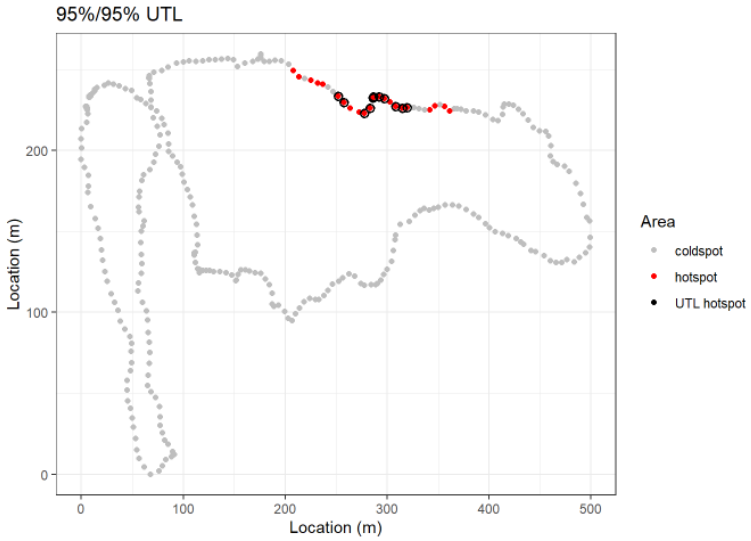

Figure 6. Example UTL results with grey dots correctly identified as cold spots, outlined red dots correctly identified as hot spots, and red dots without an outline incorrectly identified as cold spots.

A 99th percentile and 95% confidence UTL is shown compared to the observations in Figure 7.

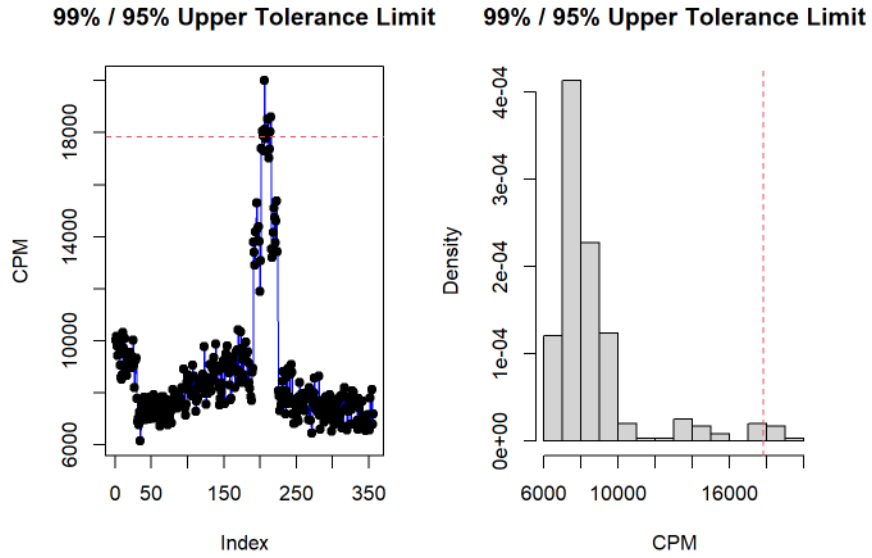

Figure 7. Example UTL diagnostic plots. The left panel shows the CPM observations compared to the 99th percentile threshold and the right panel shows the distribution of CPM observations.

The 99th percentile UTL identified all cold spots correctly but only 40% of the hot spots correctly. It incorrectly identified 14 of the 23 hot spots as cold spots, as shown in Figure 8.

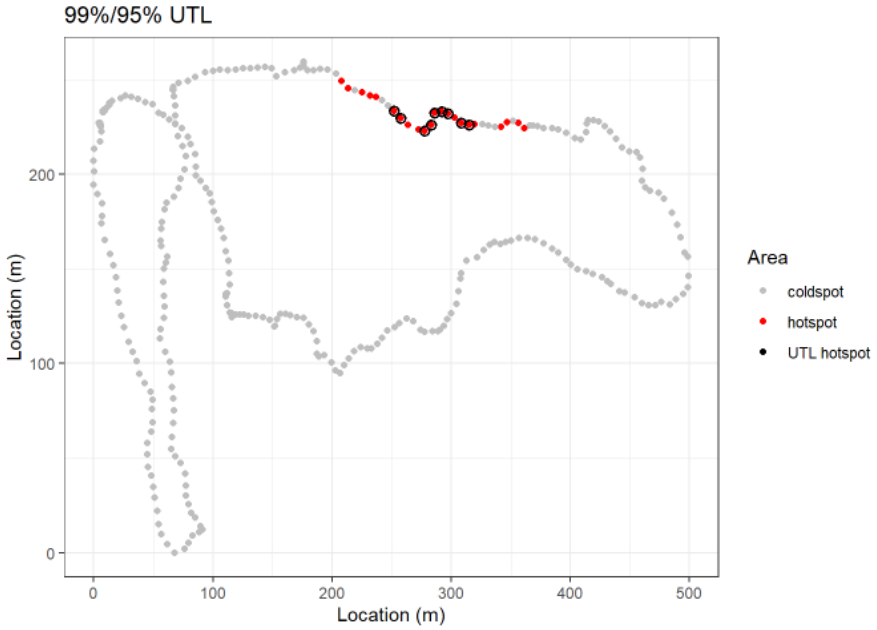

Figure 8. Example UTL results with grey dots correctly identified as cold spots, outlined red dots correctly identified as hot spots, and red dots without an outline incorrectly identified as cold spots.

In practice, incorrectly identifying hot spots as cold spots would lead to investigating fewer locations than necessary and potentially missing areas of concern.

4.1.2 USL Method

The USL method should also be considered for identifying areas of elevated contamination (hot spots) if they are present, and like the UTL can be applied when non-uniform contamination is present. The USL method derives the IL_{pp} from a reference area survey by estimating the maximum value of the background distribution, i.e., the value that all observations from the reference area are less than or equal to, with a given level of confidence. Like UTLs, parametric and nonparametric methods have been developed for deriving USLs (Hahn and Meeker 2011). A USL will always be greater than or equal to a UTL for the same site.

PNNL applied the USL method to the same dataset as described in Section 4.1.1. The USL method estimated the maximum of the reference area, and the IL_{pp} was just less than 9000 with 95% confidence. This is shown as the dashed line in the plots in Figure 9, in comparison to the observations (left panel) and distribution of observations (right panel) in both the reference area and area of concern.

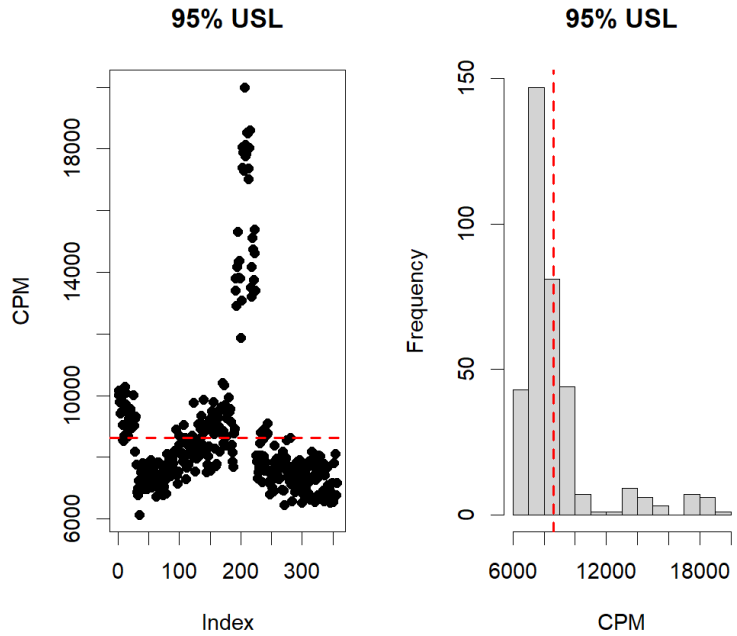

Figure 9. Example comparing IL_{pp} based on the USL method to observations from the reference area and the area of concern. The right panel shows the distribution of CPM observations.

Any observations in the area of concern that were greater than the IL_{pp} were identified as potential elevated areas (hot spots). The USL method identified all 23 hot spots correctly. It incorrectly identified 89 cold spots as hot spots, as shown in Figure 10. In practice, incorrectly identifying cold spots as hot spots would lead to investigating more locations than necessary.

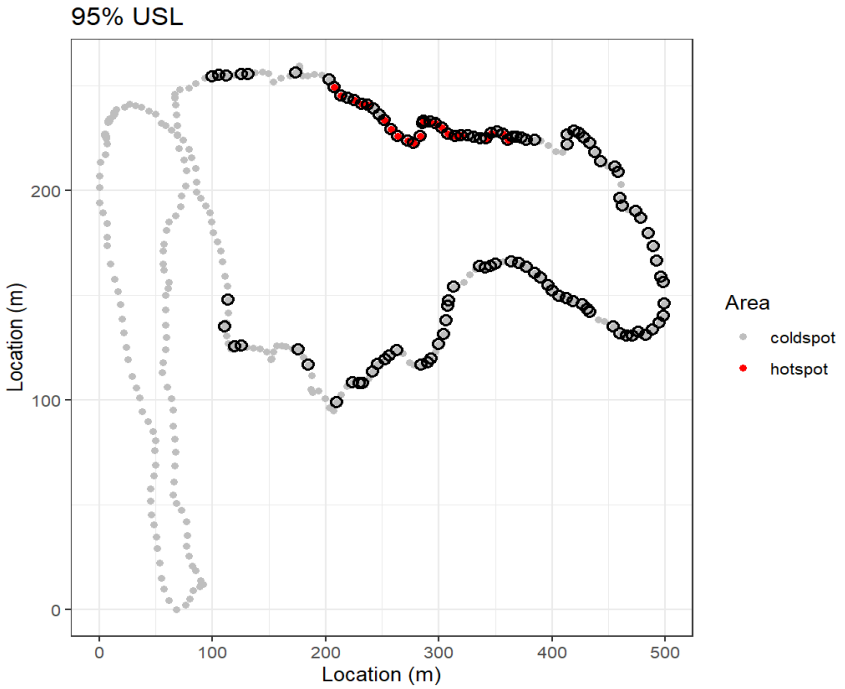

Figure 10. Example USL results with grey dots correctly identified as cold spots, outlined red dots correctly identified as hot spots, and outlined grey dots incorrectly identified as hot spots.

4.1.3 Z-score Method

A third method for developing an IL_{pp} uses z-scores, where the z-score quantifies the distance between each observation and the sample mean scaled by the sample standard deviation. The use of z-scores is commonly accompanied by an assumption that the data come from a population that can be described parametrically by a (log) normal distribution. Under the assumption of normality, it is easy to create an IL_{pp} based on a z-score corresponding to a chosen quantile of the standard normal distribution (e.g., for the 9^{9t}h percentile, Z = 2.3). The z-score is computed for each observation in the area of concern by comparing the observation to the sample mean of observations from reference area and area of concern. An IL_{pp} can be derived from z-scores nonparametrically using Chebyshev's inequality in cases where the lognormal distribution assumption does not hold. However, this can result in overly conservative investigation levels (Casella and Berger 1990).

An additional limitation of z-scores is the lack of uncertainty reflected in their determination, in contrast to UTLs and USLs. The latter are based on choosing values β and α such that $\alpha \times 100\%$ of the population is less than or equal to the UTL with $(1-\alpha) \times 100\%$ confidence and USLs as a special case of this, with $\beta = 1$. The confidence statements reflect sampling and population distribution uncertainty. While z-scores do reflect sampling uncertainty, they do not reflect the uncertainty about the underlying population distribution. As a result, there is no confidence statement associated with the IL_{pp} derived from a z-score.

PNNL applied the z-score method to the same dataset as described in Section 4.1.1. Two cases were examined for the z-score method, where one IL_{pp} was defined based on comparing z-scores to the 90th percentile of the standard normal and the other based on the 95th percentile. The results are summarized in Figure 11.

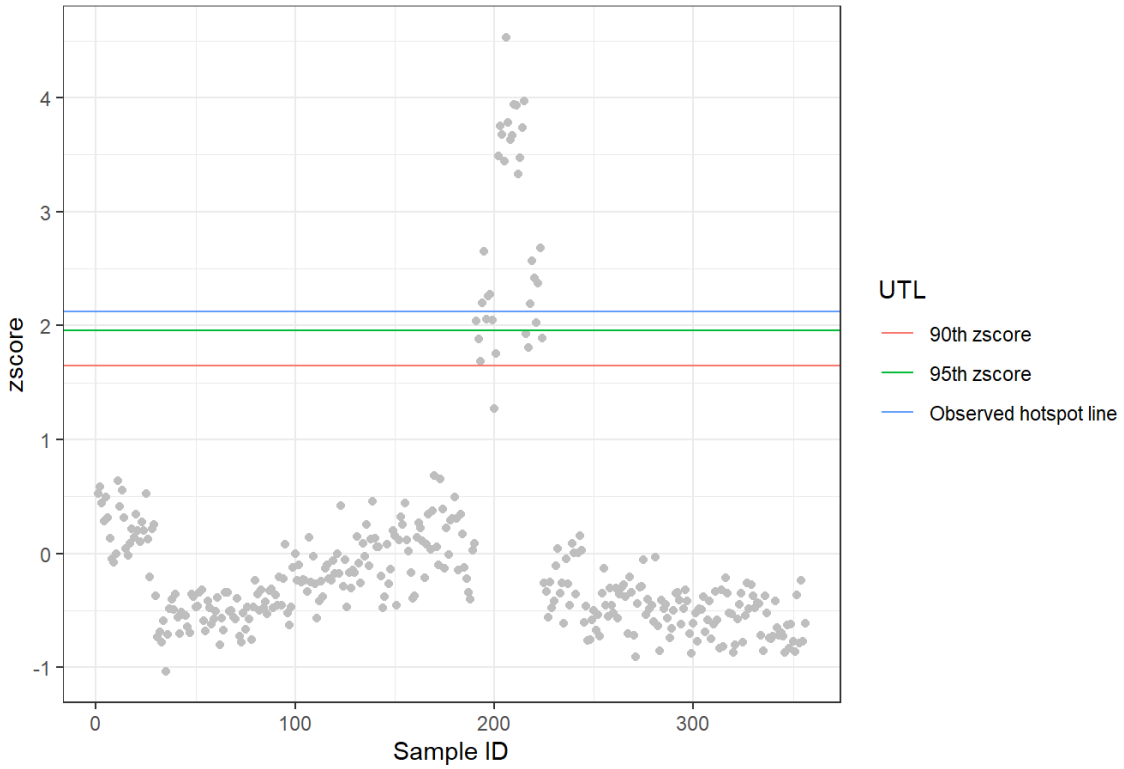

Figure 11. Example z-score method. The blue line is the observed hot spot threshold value, and the green and red lines are the 95th z-score and 90th percentile z-score from a standard normal distribution.

Both percentile choices led to correct identification of all 23 hot spot locations. The 95th percentile incorrectly identified 10 of 333 cold spots as hot spots and the 90th percentile incorrectly identified 4. In practice, incorrectly identifying cold spots as hot spots would lead to investigating more locations than necessary. The results are shown in Figure 12.

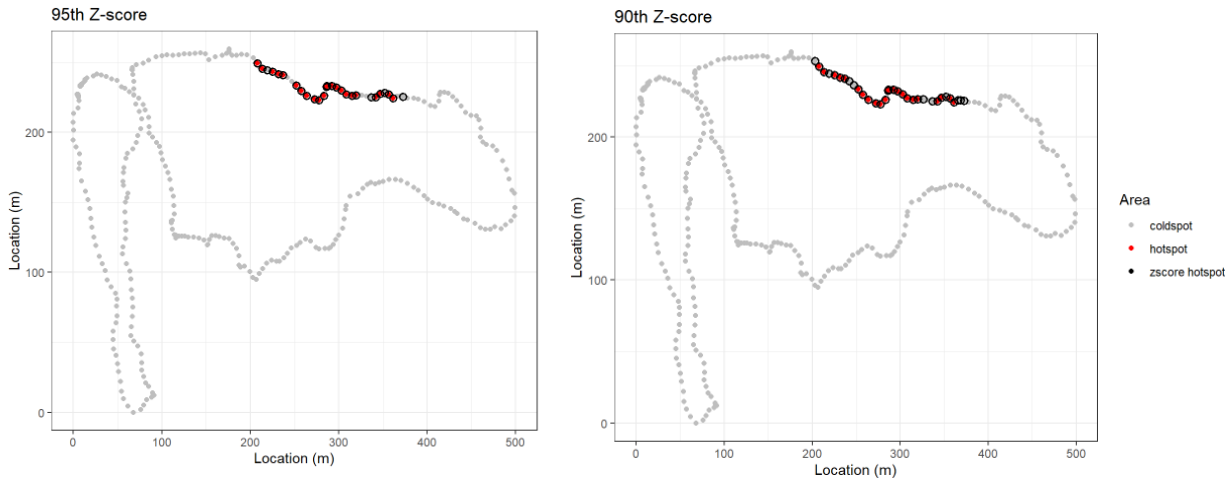

Figure 12. Example z-score results with grey dots correctly identified as cold spots, outlined grey dots incorrectly identified as hot spots, and outlined red dots correctly identified as hot spots.

4.1.4 Local Indicator of Spatial Association Method

A LISA, also known as the Local Moran's I statistic, is a method presented by Anselin (1995) to identify local clusters and local spatial outliers. The method can be applied to detect statistically significant points of high contamination within otherwise low-contamination areas and, conversely, can detect cool spots, or areas of significantly low contamination compared to surrounding high-contamination areas.

A LISA value is calculated for each observed location. Positive LISA values indicate neighboring locations have similarly high or low contamination. Negative LISA values indicate neighboring locations are higher/lower concentration. Each LISA value has a corresponding p-value, where lower p-values indicate that clusters of locations are statistically different than neighboring locations. Similar to the framework used in Macedo (2016), four types of clusters with statistical significance are considered:

1. Several locations with high values (HH)
2. Several locations with low values (LL)
3. Single high-value location surrounded by low-value locations (HL)
4. Single low-value location surrounded by high-value locations (LH)

Clusters of high values (HH and HL) indicate potential areas of high contamination (hot spots), while low values (LL and LH) indicate potential areas of comparatively low activity (i.e., cool spots).

The LISA method could also be useful for data quality assurance in addition to identifying high contamination areas in the context of UAS/UAV or human-based continuous data collection. By incorporating altitude, GPS, and velocity data into the analysis, it could help identify areas where scan velocity and altitude were not in compliance with established standards. For instance, areas with statistically significant low contamination values (LL and LH) that are outside the range of expected values could be flagged for further investigation. Additionally, NRC could use LISA to check whether scan velocity and altitude measurements collected by UAS/UAVs and humans comply with regulatory requirements. By analyzing LISA values for each observed location, the NRC could identify clusters of locations with similar velocity and altitude measurements that differ from neighboring locations, indicating potential noncompliance issues. This could help ensure that data collection is performed consistently and accurately, and that data quality is maintained over time. Additionally, the LISA method could be used to identify clusters of high or low values that may indicate specific sources of contamination or other environmental factors that are influencing the data.

PNNL applied the LISA method to the same dataset as described in Section 4.1.1. The results are shown in Figure 13. The LISA approach identifies 39 areas as having high (HH) contamination. It accurately recognizes all 23 induced hot spots, but mistakenly classifies 15 cold spots as hot spots. In practice, if cold spots were mistakenly identified as hot spots, this would result in more areas being investigated than are required.

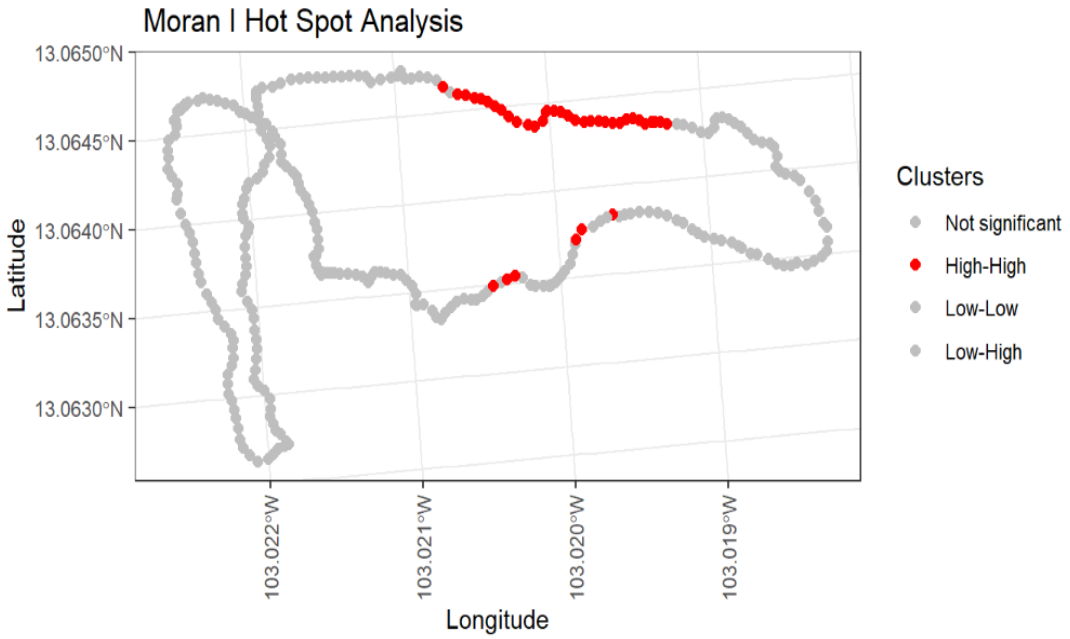

Figure 13. Example LISA method results with grey dots correctly identified as cold spots, and red dots as identified hot spots comprising 23 correctly identified hot spots and 15 cold spots incorrectly identified as hot spots.

4.1.5 Lag-k Method

Increased continuous data collection using automated data loggers and autonomous radiological GPS/GIS devices has introduced a need for corresponding guidance and statistical techniques for these data collected without surveyor vigilance. The *a priori* scan MDC calculations detailed in NUREG-1507 (NRC 2020a) assume surveyor vigilance (i.e., pausing or stopping to investigate further when audio click data indicate potential areas of concern). Fortin et al. (2023) provides *a priori* scan MDC calculations assuming GPS/GIS-based surveys will be completed without vigilance based on binned or integrated audio click data. These calculations provide an incremental advance in the *a priori* scan MDC methodology, moving from the with-vigilance assumption to a without-vigilance surveying paradigm.[1]

Varying background radiation levels pose a challenge when scanning to identify areas of contamination relative to background. When this variation is a nuisance and not the focus of an analysis, one mitigating approach is to use local differencing (computing differences between observations and their neighbor[s]) to identify elevated areas (hot spots).

King and Vitkus (2015) present an example of an applied scan MDC process starting with a minimum detectable count rate (MDCR) analysis through to radiological survey data. They demonstrated how deeply important the estimation of background is. Michaud (2021) explicitly account for a variable background in their Bayesian modeling work The lag-k method offers a local differencing approach. It computes net counts based on spatially localized average

[1] It is known that there are differences between binned or integrated audio click data and ratemeter data. However, to our knowledge the magnitude of such differences and their effect on scan MDC calculations remains unknown. Future research and development will be required to determine the method's efficacy and applicability to logged ratemeter display counts (Fortin et al. 2023).

observations rather than a single, overall background average. The distance between spatially localized neighbors is determined by the site conceptual model, expected hot spot sizes, observed background variation, and the distances between neighboring locations that are expected to be independent (uncorrelated). The parameter k indicates the distance between independent observations. Net counts are estimated by calculating the differences between observations k units or more apart. Details on the lag-k method and its derivation based on a familiar hypothesis testing framework are provided in Fortin et al. (2023).

The lag-k method can be used to calculate critical levels, detection limits, and MDCR values similar to those derived for static measurements (Currie 1968). The main inputs include desired false positive and false negative error rates, like in MARSSIM/NUREG-1575 (NRC 2020c) and NUREG-1507 (NRC 2020a). The MDCR values can be converted to *a priori* scan MDC values to ensure that survey parameters (e.g., scanning speed, scanning altitude, detector geometry) will lead to collecting data in which potentially contaminated areas can be detected with acceptable statistical error probabilities by the surveyor during a scanning process that meets regulatory requirements. The lag-k method can also be used after a GPS/GIS-based survey has been completed (*a posteriori*) to process continuously collected data and perform hypothesis testing to determine if the data indicate the presence of potential hot spots.

The following bullets provide a summary of assumptions and conditions under which the lag-k approach should be considered, with limitations noted as well:

- Lag-k is appropriate when the conceptual site model indicates that background radiation levels vary across the site with residual hot spots present in small areas. A site with non-localized contamination (i.e., residual contamination spread uniformly across the site or large areas of distributed residual contamination near the regulatory limit) is not a suitable candidate for the lag-k approach.
- The lag-k approach is a hypothesis testing procedure for localized sources intended to account for variability in background.
- The formulation of the lag-k method assumes that data is collected along a transect via a detector moving at a uniform speed recording counts in uniform time intervals.
- The primary advantages of the lag-k method include:
  - It is robust relative to fluctuation in background levels because the comparisons between differences automatically adjust for local variation. This ensures that areas where readings are slightly higher because of historical use or geological properties are not overly likely to trigger further investigation.
  - It provides the ability to detect and flag for follow-up the areas of elevated contamination within regions that feature lower than average background levels.
- Limitations of the lag-k approach include:
  - The introduction of the additional lag parameter (k) that must be understood and determined by the analyst. This parameter can significantly affect the detection performance of the lag-k method. However, the expected hot spot size can be used for determining a meaningful physical basis for the lag size.
  - Gradual increases in contamination levels across a site may not be detected during data analysis by this method (e.g., if a slowly increasing gradient of contamination across the site is present, local differencing will filter out that signal). However, this concern should not be overemphasized because scanning is aimed at identifying small areas of

contamination. Large area persistent trends can be identified by visual inspection as part of the data analysis process..

4.1.6 Scan MDC Methods

PNNL reviewed numerous articles and literature sources with scan MDC methods. This section summarizes the relevant literature and provides a timeline of when each article was published and what it contributes to the methods, applications, and calculation of scan MDC values.

4.1.6.1 Seminal Research Papers

The MDC (scanning or not) specifies a level of radioactivity beyond background that can be practically detected by a measurement process. For surface activity measurements, MDC often is reported in units of disintegrations per minute per 100 square centimeters (dpm/100 cm^2). Currie (1968) presented a statistical foundation for scan MDC that is in use today.

> *"The statistics of detection and determination apply directly to observations rather than to the underlying physical quantity and, therefore, the following discussions deal specifically with the observed (or observable) signal (meter reading) and its associated random fluctuations."*
> Currie (1968)

Some of the key ideas presented in Currie (1968) include the following:

- Currie focused on describing the statistical properties of observed counts and used Poisson counting statistics to model radioactive decay.
- For large counts a Poisson distribution can be approximated by a normal distribution.
- The mean and variance of a Poisson distribution are the same so distributions with larger means also have larger variances.
- Currie produced formulas for the limit of detection (LOD), based on α, the probability of a false positive, and β, the probability of a false negative.
- Currie produced formulas to determine a critical level (L_c), also called decision level for total counts or decision limit, which maps to the decision criteria yielding the desired error rate, α.

Currie (1968) also explained the distinction between gross signal (total counts) and net signal, arrived at by removing the mean in background which led to questions about the best way to estimate background to remove it from subsequent calculations. Currie investigated a background estimation approach where samples of interest were paired to blank samples that were "identical" to the sample of interest except missing the contaminant substance sought.

In a later publication, Currie (1984) summarized ideas surrounding the lower LOD for the NRC. This work included references to earlier work on LODs and presaged future publications by noting that the normal assumption must be replaced with an exact Poisson treatment for low count levels. Currie (1984) also considered differences between sample and blank (background) counting times with a model that allowed for measurements from multiple blanks.

4.1.6.2 Scaling and Efficiency

This section summarizes the scan MDC literature review under the scaling and efficiency topics.

- Brodsky and Gallaghar (1991) provided an examination of the minimum detectable amount (MDA) in the context of smear survey practices. During smear surveys, a designated area is selected, and samples are taken by wiping or swiping a specific surface using a suitable

collection medium, such as filter paper or a swab. The collection medium is analyzed to determine the presence, extent, and level of radioactive contamination. The purpose of smear surveys is to make sure radioactive materials are properly contained and do not pose a risk to workers, the public, or the environment. The standard assumption of the data generating process for a sample and its paired (and appropriate) blank being represented by two normal distributions with unequal variances was used to model net background counts and net sample counts.

- Gallaghar (1991) used a calibration constant (K) and sample counting time (T), assumed to be the same between a sample and its paired blank, to convert the decision level or decision limit to units of activity to determine the decision amount (DA). To determine the limit of detection (LOD), Brodsky and Gallaghar (1991) introduced, and Brodsky (1992) advanced, an update to Currie's (1984) formula for LOD. Where Currie's (1984) formula used mixed assumptions of Normal and Poisson distribution, the updated versions assumed Poisson distributions resulting in an update to the additive constant. Brodsky (1992) also included a formula for MDA, which is a scaled version of the LOD that includes K and T where K is a calibration factor that converts counts per minute to counts per minute per becquerel and where T is the total counting time of a sample assuming a paired blank.
- Building on Brodsky (1992), Strom and Stansbury (1992) adjusted the MDA for different counting times between sample and blank.
- MARSSIM/NUREG-1575 (NRC 2020c) was developed collaboratively among the U.S. Department of Defense, U.S. Department of Energy, Environmental Protection Agency, and NRC to provide a standardized guidance document for investigating radioactively contaminated sites.
- NUREG-1507 (NRC 2020a) is an NRC document that provides guidance to licensees for performing scanning surveys, including an introduction to MDC (under the unequal variances framework) and development of scan MDC using signal detection theory and an equal variance approximation obtained via the index of sensitivity, d′. In NUREG-1507 (NRC 2020a), MDCR is calculated first and then transformed into scan MDC via scaling factors for two stage with-vigilance survey planning, noting that scan MDC is generally computed in the first stage.
- NUREG/CR-6364 (NRC 1997) explored human performance factors, signal detection theory, detectability, and human surveyor efficiency. It also provides information for radiation exposure including dose exposure formulas.
- Hart et al. (2003) investigated the scan detection efficiency for detectors to inform corresponding scaling factors. Brandl and Herrera Jimenez (2008) elucidated continuously recorded measurements and explored alarm levels, which are conditional on known backgrounds, count rates, and sample counting times. King et al. (2012) accounted for the swing of a detector during walking surveys in minimum detectable concentration calculations.
- Brandl (2013) investigated low signal-to-background ratios to lower decision thresholds for continuous scanning measurements. This work assumed negative binomial distributions for the sequential count data, extending beyond the analysis of each data point individually.
- Abelquist (2014) explained and defined relevant statistics including LC, LOD, K, and T in the context of scan MDC via signal detection theory. This work included different collection times for background and samples surveyor efficiency and alpha scan MDC.
- Alecksen and Whicker (2016) focused on the conversion of MDCR to scan MDC via Monte Carlo N-particle Extended (MCNPX) to establish efficiency scaling factors. Note that MCNPX

is no longer being updated regularly. Starting with MCNP6, Los Alamos National Laboratory has put the alpha particle and heavy ion transport features into the MCNP6.x versions.

- Brogan and Brandl (2018) improved hypothesis testing and handled low count rate data when background is not perfectly known by adding data points together and extending collection times. This method does not account for varying background conditions. It uses a variation on the negative binomial distribution to calculate false positive rates when n out of N sequential measurements exceed a threshold level.
- Watson et al. (2018) performed an experiment for alpha/beta radiation using a Shiryaev-Roberts control chart methodology.
- Falkner and Marianno (2019) developed a well-defined relationship to predict MDA as a function of detector speed. Detector efficiency can be improved by slowing the speed of travel of the detector; this work quantifies the relationship between detector speed and the minimum detectable activity. The work is agnostic to the with-vigilance versus without-vigilance dichotomy.
- Justus (2019) developed decision levels and detection limits for audible scan surveys for the Poisson model without using a normal approximation and compared the approach to other methodologies and some available empirical results.
- Alecksen and Whicker (2023) showed that scan MDC can be calculated for without vigilance surveys using the existing NUREG-1507 approach (NRC 2020a), indicating that “whether based on the approximate MDCR calculation or the exact MDCR expression, the probabilistic method described by Alecksen and Whicker (2016) for determining scan MDCs for GPS/GIS-based gamma surveys is a statistically valid application of the MDCR concept, provided that the gamma instruments used for the scan system are operated in scaler counting mode rather than ratemeter mode.”
- Alecksen and Whicker (2023) performed field tests to compare unequal variance analysis to the simpler equal variance analysis, as presented in NUREG-1507 (NRC 2020a). They found that the more realistic, unequal variance, approach produced error rates closer to nominal but also stated that the equal variance approximation did not produce error rates that differ substantially from nominal and therefore could be preferred as its formulation is relatively more straightforward.

Alecksen and Whicker (2023) also consider further evaluation of GPS/GIS-based survey data that otherwise may remain unused using a traditional scan MDC determination. They investigated how grouping of individual observations, based on visualization or some other analysis of data may be able to improve sensitivity and specificity. We explore the methods described in Alecksen and Whicker (2016 and 2023) below.

MARSSIM Method Applied to Scan MDC for Surveys Without-Vigilance

The scan MDC methodology in MARSSIM/NUREG-1575 (NRC 2020c) and NUREG-1507 (NRC 2020a) was developed explicitly for a with-vigilance case, where a human surveyor listens for changes in clicks while scanning an area. Because of this, the procedure is not designed to be applicable for the case where a GPS/GIS-based survey is conducted without-vigilance; that is, without the presence of a human listening for a change in clicks or for a retrospective analysis of gathered data.

Alecksen and Whicker (2023) show empirical results of applying the NUREG-1507 MDCR development (NRC 2020a) to GPS/GIS-based surveys conducted without-vigilance in a

controlled environment. They found that the MDCR formulation in NUREG-1507 (NRC 2020a) produces results that are close to the expected false positive and true positive proportions. A simple modification of the NUREG-1507 MDCR methodology (NRC 2020a), allowing for different mean and variances for the background and background plus source distributions, results in false positive and true positive proportions closer to the expected (or target) values.

Probabilistic Method Developed by Alecksen and Whicker (2016)

The probabilistic method developed by Alecksen and Whicker (2016), as presented via the ERG calculator[1], uses the MCNPX transport code as part of a methodology to model and estimate scan MDC values. Results from the probabilistic method, available via the ERG calculator, provide scan MDC values for conditions defined by the following factors.

- Five contaminants: americium-241, cobalt-60, cesium-137, radium-226 and thorium-232
- Three thallium-doped sodium iodide NaI(Tl) scintillation instruments (NaI detectors): 3×3 in., 2×2 in., and 1.25×1.5 in.[2]
- Eight source diameter values: from 20 cm to 1,400 cm
- Four detector heights: between 10 cm and 100 cm
- Several scanning speeds: multiples of 0.1 m/s, from 0.1 to 1.5 m/s
- Numerous false positive and true positive target rates

The probabilistic method results in conservative scan MDC values because it estimates detector efficiency using the lowest potential maximum count rate as the detector travels across the source. Despite this, scan MDC values obtained using the probabilistic method are generally lower than those found using NUREG-1507 (NRC 2020a). This is due to the use of a detector sensitivity factor in NUREG-1507 (NRC 2020a), which is not included in the formulation by Alecksen and Whicker (2016).

An additional factor affecting the scan MDC values obtained with the probabilistic method is that NUREG-1507 (NRC 2020a) transforms the MDCR to an equivalent exposure rate by assuming a point source that sits directly below the center of the detector and by using a manufacturer-reported sensitivity value for the detector, which is limited to point sources and primary photons. Because scattered photons are not considered, this results in biased scan MDC values. The probabilistic method models detection efficiency directly and assumes a volume for the source and considers primary and scattered photons.

Alecksen and Whicker (2016) state that for smaller sources, NUREG-1507 (NRC 2020a) and probabilistic scan MDC values tend to be more similar than for larger sources. For larger sources, the probabilistic method generally produces lower scan MDC values than NUREG-1507 (NRC 2020a). For radionuclides with low photon energies, like americium-241, the probabilistic method will tend to produce ‘significantly higher’ scan MDC values than those from the NUREG-1507 (NRC 2020a) method because detector sensitivity in those cases is similar for volume and point sources. Alecksen and Whicker (2016) assert that NUREG-1507 (NRC 2020a) accuracy will decline as the photon energy of the radionuclide increases (sensitivity values for the two methods differ by less than 20% for volume or point sources below 100 keV).

[1] The ERG calculator and database can be accessed at https://ergoffice.com/erg-calculator.
[2] The ERG calculator includes the 3×3 in., 2×2 in., and 1.25×1.5 in. detector types while Alecksen and Whicker (2016) includes 7.6×7.6 cm and 5×5 cm, and Alecksen and Whicker (2023) includes 5×5 cm.

These observations indicate that different conditions may indicate which methodology should be used, rather than uniformly selecting one method to be applied in all cases.

More recent developments by Alecksen and Whicker (2023) are aimed at empirically validating the concept of using MDCR for without-vigilance GPS/GIS-based survey planning and considering the implication of which distributional model is assumed for background and background-plus-source observations.

Approximations to ERG Calculator of Scan MDC

The probabilistic method is implemented in the ERG Calculator, available online which provides scan MDC values for various background and background-plus-source distributions for conditions defined by the discrete set of factors and values listed above.[1] The ERG Calculator is freely available, but to our knowledge must be accessed for each individual combination of factors. It would therefore be helpful to be able to approximate the resulting scan MDC estimates without having to access the online calculator and to provide approximations for conditions not explicitly modeled in the ERG Calculator. To that end, PNNL fit two predictive models to scan MDC values to determine whether a lookup table and/or predictive model could provide reasonable approximations. We examined models and results for two radionuclides (americium-241 and radium-226).

PNNL applied a generalized linear model to the database values using detector type (3×3 in. or 2×2 in.; the 1.25×1.5 in. was omitted from this proof-of-concept demonstration), source diameter, detector height and scanning speed (between 0.5 and 1 m/s) for americium-241, with a gross background rate of 12,180 cpm. We set the false positive rate to 0.05 and the true positive rate to 0.95. We assumed a Poisson distribution and a log link. The generalized linear model is shown in Equation 4.

$$\begin{aligned}\log\left(\widehat{scan\ MDC}_{241\text{-}Am}\right) = \quad & \beta_0 + \beta_1 \cdot Type\ (3{\times}3\ \text{in.}) + \beta_2 \cdot Diameter + \beta_3 \cdot Height + \beta_4 \cdot Speed \\ & + \beta_{23} \cdot Diameter \cdot Height + \beta_{23} \cdot Diameter \cdot Speed \\ & + \beta_{23} \cdot Height \cdot Speed + \beta_{22} \cdot Diameter^2 + \beta_{33} \cdot Height^2 \end{aligned} \qquad \text{Equation 4}$$

Here, the β_i are model coefficients estimated from the database data. The variable $Type\ (3{\times}3\ \text{in.})$ is an indicator variable, equal to 1 for a 3×3 in. NaI detector and zero for a 2×2 in. NaI detector (the 1.25×1.5 in. NaI detector was omitted from this proof-of-concept demonstration). Parameter estimates and related statistics for Equation 4 are shown in Table 2.

Particular combinations of inputs result in biased predictions using Equation 4. The bias corrections in Equation 5 should be applied to the scan MDC predictions made from Equation 4.

[1] The database can be accessed through the following link: https://ergoffice.com/erg-calculator.

Table 2. Equation 4 model fit parameter estimates and related statistics.

Model	Estimate	Standard Error	t-value	p-value[(a)]
Intercept	6.0887	0.0377	161.50	0
Detector type (3×3 in.)	-0.7653	0.0076	-100.50	0
Diameter	-0.0288	1.7182×10^{-4}	-167.55	0
Height	0.0522	8.8414×10^{-4}	59.06	0
Speed	1.3055	0.0376	34.71	<0.001
Diameter · Height	-2.6528×10^{-5}	9.3918×10^{-7}	-28.25	<0.001
Diameter · Speed	-8.0763×10^{-5}	1.0450×10^{-4}	-7.73	<0.001
Height. · Speed	-0.0102	4.4670×10^{-4}	-22.86	<0.001
Diameter^2	1.9580×10^{-5}	9.5827×10^{-8}	204.32	0
Height^2	-1.5567×10^{-5}	6.0933×10^{-6}	-25.55	<0.001

(a) A p-value less than or equal to 0.05 indicates statistical significance. A value smaller than 0.001 indicates a highly significant term.

$$\widehat{scan\ MDC}_{241\text{-}Am} = \begin{cases} \frac{1}{0.8340} \cdot exp(Eq.1) \ \ if\ Diameter < 38\ cm \\ \frac{1}{2.1920} \cdot exp(Eq.1) \ \ if\ 38\ cm \leq Diameter < 125\ cm \\ exp(Eq.1) \ \ otherwise \end{cases} \quad \text{Equation 5}$$

Performance of the model in Equation 5 is shown graphically in Figure 14 and in Figure 15.

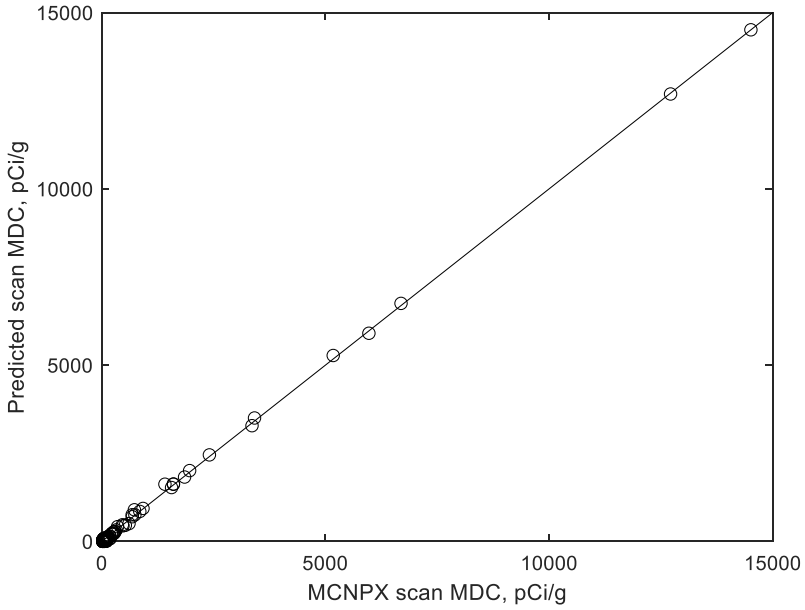

Figure 14. Approximate americium-241 scan MDC values (pCi/g) predictions from the model in Equation 4 (y-axis) as a function of americium-241 scan MDC values (pCi/g) from the online calculator (x-axis). The 45° line represents a hypothetical perfect approximation.

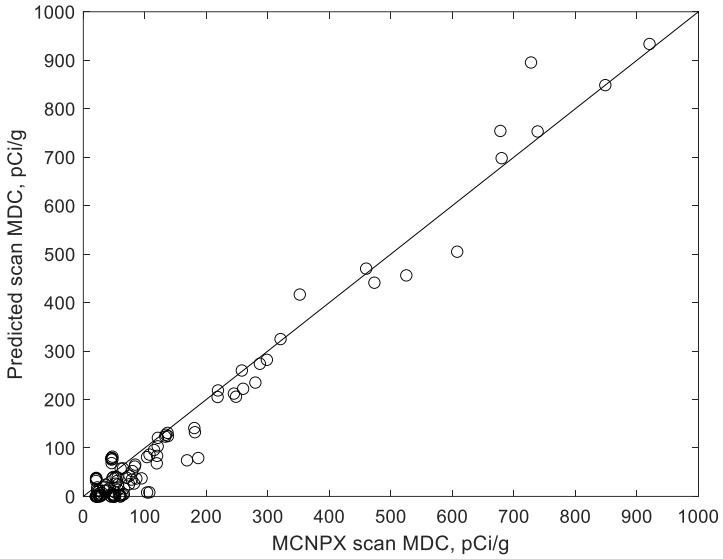

Figure 15. Approximate americium-241 scan MDC values (pCi/g) predictions from the model in Equation 4 (y-axis) as a function of americium-241 scan MDC values (pCi/g) from the online calculator (x-axis), limited to the range between 0 and 1000 pCi/g. The 45° line represents a hypothetical perfect approximation.

Figure 14 shows that the model in Equation 4 predicts online calculator scan MDC values well and that the methodology appears to be useful for approximating results by the probabilistic method of Alecksen and Whicker (2023) without having to access the online calculator for every set of conditions of interest.

Figure 15 shows the same results as Figure 14 but restricted to smaller scan MDC values, providing a more detailed look at model performance in this lower range. Here, there is more scatter and a tendency for underprediction when values are lower than 200 pCi/g, suggesting that the model in Equation 4 provides reasonable approximations for scan MDC values as low as roughly 200 pCi/g but that the bias should be considered more carefully for smaller values. We used a similar approach to fit scan MDC data for radium-266. This generalized linear model assumed a Poisson distribution and a reciprocal link for the response shown in Equation 6.

$$\begin{aligned} \widehat{scan\,MDC}^{-1}_{226\text{-}Ra} = \quad & \beta_0 + \beta_1 \cdot Type\ (3\times3\ \text{in.}) + \beta_2 \cdot Diameter + \beta_3 \cdot Height + \beta_4 \cdot Speed \\ & + \beta_{12} \cdot Type\ (3\times3\ \text{in.}) \cdot Diameter\ + \beta_{13} \cdot Type\ (3\times3\ \text{in.}) \cdot Height \\ & + \beta_{14} \cdot Type\ (3\times3\ \text{in.}) \cdot Speed + \beta_{23} \cdot Diameter \cdot Height \\ & + \beta_{34} \cdot Height \cdot Speed + \beta_{22} \cdot Diameter^2 + \beta_{33} \cdot Height^2 \end{aligned} \qquad \text{Equation 6}$$

Model parameter estimates and related statistics are shown in Table 3. As with the previous model, the variable Type is an indicator, where Type is equal to 1 if a 3×3 in. detector is used and zero when using a 2×2 in. detector type (the 1.25×1.5 in. was omitted from this proof-of-concept demonstration).

The radium-266 scan MDC model includes three terms with detector type interactions where the Type (3×3 in.) interactions with diameter, height, and speed are indicated by the "·" symbol. Statistical significance (p-values less than 0.05) indicates that the effect (slope) of each of these variables depends on the type of detector used.

Table 3. Equation 6 model fit parameter estimates and related statistics.

Model	Estimate	Standard Error	t-value	p-value[(a)]
Intercept	-0.0106	0.0048	-2.24	0.027
Type (3×3 in.)	0.0018	0.0021	0.84	0.400
Diameter	0.0018	2.4004×10^{-4}	7.34	<0.001
Height	-1.2597×10^{-4}	4.9742×10^{-5}	-2.53	0.013
Speed	-0.0091	6.6826×10^{-4}	-13.56	<0.001
Type (3×3 in.) · Diameter	5.6092×10^{-4}	9.8222×10^{-5}	5.71	<0.001
Type (3×3 in.) · Height	-1.0015×10^{-4}	7.5848×10^{-6}	-13.20	<0.001
Type (3×3 in.) · Speed	-7.1970×10^{-4}	2.2122×10^{-4}	-3.25	0.001
Diameter · Height	-1.6992×10^{-5}	2.2254×10^{-6}	-7.63	<0.001
Height · Speed	8.5160×10^{-5}	6.7094×10^{-6}	12.69	<0.001
Diameter^2	3.5410×10^{-6}	9.8687×10^{-7}	3.59	<0.001
Height^2	2.2799×10^{-6}	1.4763×10^{-7}	15.44	<0.001

(a) A p-value less than or equal to 0.05 indicates statistical significance. A value smaller than 0.001 indicates a highly significant term.

Figure 16 shows model performance for the entire range of scan MDC values and Figure 17 shows only the smaller scan MDC values (between 0 and 50 pCi/g). The plots in Figure 16 and Figure 17 show that the approximations from the model in Equation 6 are close to the hypothetical perfect approximation (45° line) for the whole the range of scan MDC values. There are two observations at the largest values (approximately 517 pCi/g and 602 pCi/g) that extend the range considerably. This could imply that those are quite influential for this model's parameter estimation. Figure 17 shows scattered predictions for the lower range of scan MDC values. Observations in the 0-5 pCi/g range show some systematic trends for subsets of data.

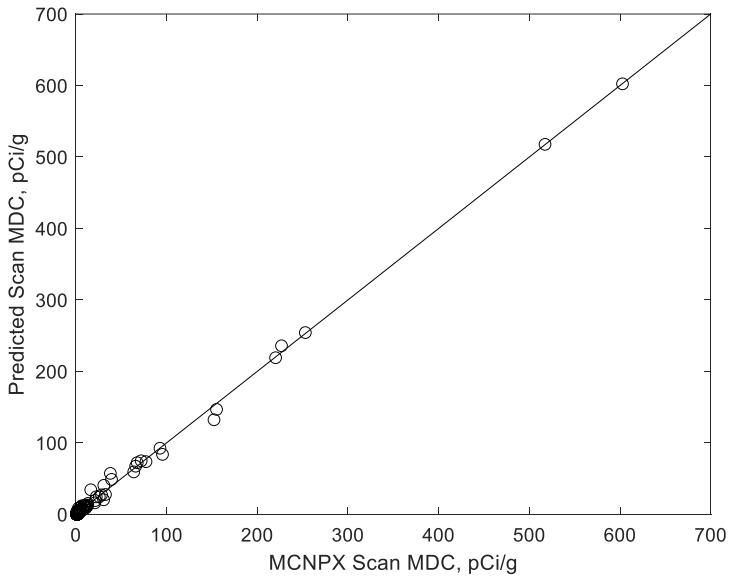

Figure 16. Approximate radium-266 scan MDC values (pCi/g) from the model in Equation 6 (y-axis) as a function of radium-266 scan MDC values (pCi/g) from the ERG Calculator (x-axis). The 45° line represents a hypothetical perfect approximation.

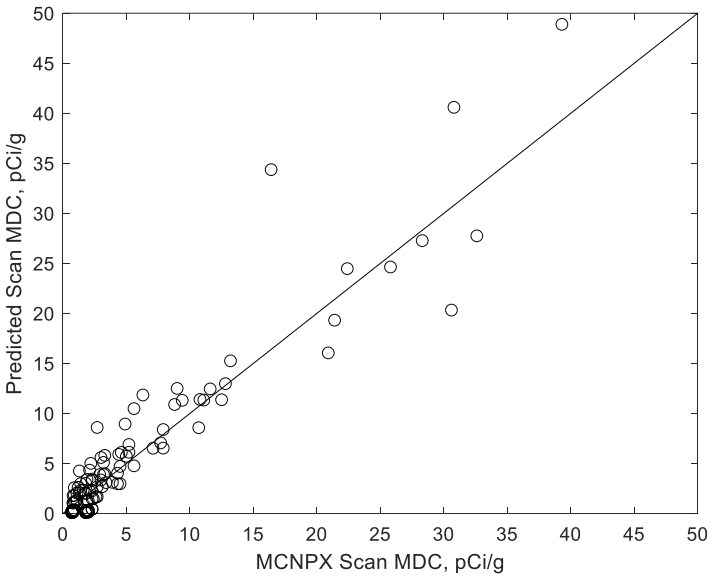

Figure 17. Approximate radium-266 scan MDC values (pCi/g) from Equation 6 (y-axis) as a function of radium-266 scan MDC values (pCi/g) from the online calculator (x-axis), limited to the range between 0 and 1,000 pCi/g, limited to the range between 0 and 50 pCi/g. The 45° line represents a hypothetical perfect approximation.

The models for both americium-241 and radium-226 were validated. In this procedure, referred to as a leave-one-out (LOO) method, one observation from the dataset is left out, the model is fitted to the remaining observations, and then the fitted model is to predict the “left out” observation. The LOO procedure is commonly used for such predictive model validation. In the two models above, implementing the LOO procedure indicated the following:

- The americium-241 model resulted in good prediction for the LOO procedure for data below 1500 pCi/g but poor predictive performance at larger values.
- The radium-226 model resulted in relatively good prediction for the LOO procedure for data below 100 pCi/g but poor predictive performance at larger values.

These LOO validation results are not unexpected, given that relatively few conditions result in scan MDC values above the above the higher thresholds. Fewer data with very large scan MDC values resulted in those data points having a disproportionately large influence on the model parameter estimates.

The models presented in this section provide information about the magnitude and direction of various factors on scan MDC. They demonstrate that such predictive models could be useful in GPS/GIS-based survey design in lieu of or to augment the results available from the ERG Calculator when determining scan MDC using the probabilistic method approach. The models shown in this document are not guaranteed to predict well when used for extrapolation outside the ranges identified and caution should be used as neither have been fully validated. Additional radionuclides and models should be considered to fully assess the feasibility of implementing such approximations in VSP.

Retrospective Analysis of Scan Data Collected Without Vigilance

Determination of a single scan MDC value may not be ideal for all survey objectives. Allowing for a large (around 60%) false positive rate when index of sensitivity, d′, in the MDCR calculation can be useful for detecting large and/or dispersed areas of contamination in a GPS/GIS-based survey without vigilance. On the other hand, selecting a low false positive rate (around 5%) will allow detection of small areas of contamination that may otherwise be confused with background noise in the previous case.

Conducting more than one GPS/GIS-based survey, with different false positive rates would allow for different patterns to be detected in the data analysis. Coding (color or indicator assignment) the detector readings could be used to distinguish areas with high contamination, visually or mathematically, allowing for recalculation of the MDCR and scan MDC to optimize cleanup efforts.

Because the retrospective analysis of continuously collected data aims to achieve the data quality objectives of the survey, techniques such as clustering or control charts (like CUSUM) are good choices for the retrospective analysis, allowing the possibility of more clearly defining areas of contamination. Control charts, originally created to monitor when an industrial process is behaving in a way that requires intervention, are an intuitive tool that can be used in environmental monitoring to determine when measurements have changed from background. These charts can be designed to achieve target false positive and false negative rates and may be a helpful addition for accurately identifying areas of contamination and minimize cleanup efforts. Control charts specifically designed for Poisson distributions may be particularly useful for identifying areas of contamination in the context of soil scanning.

The scan MDC calculations described above will need to account for many confounding factors, such as type of terrain and multiple radionuclides, in addition to sources of uncertainty in the instrument used and analysis of data that are not explicitly estimated in scan MDC calculations. These factors can obscure results and inflate uncertainty values, impacting decisions made based on regulatory thresholds.

4.1.6.3 Bayesian Methods

Klumpp (2013) explore a Bayesian approach for modeling low-count-rate radioactive sources via an always-on detection system. This approach used a gamma prior on the rate parameter, combined the prior with data, and created a posterior distribution for the rate. Their version of a hypothesis test involves extracting the probability that the true count rate was above a background from the posterior distribution.

Tandon et al. (2016) implement a Bayesian aggregation of data from mobile spectrometers. Their Bayesian approach to hypothesis testing extended beyond one observation per hypothesis test. Their method relies on spectrometry data in addition to count data to help separate background radiation from source radiation. They also leverage information in the spatial correlations of their observations to help localize sources.

Brogan and Brandl (2019) develop a Bayesian interaction model to analyze gross count measurements. They use the Bayesian linear regression model to study the relationship between gross count measurements and their standard deviation observed in the “current” and previous four measurements at fixed time intervals. They incorporate a predictor (explanatory variable) in the model to indicate whether data originated from background or measurements

and used its estimate to develop a decision rule to identify elevated areas. This method requires previously recorded background gross count data.

The following articles detail case studies that use Bayesian methods:

- Using a low-resolution thallium-activated sodium iodide, NAI(Tl), detector and Bayesian inference, Kim et al. (2019) find that the maximum detectable depth for weakly active radioactive sources was 21 cm.
- Arahmane et al. (2021) combine spectrometry results from high purity germanium (HPGe) detectors with Bayesian methodologies to estimate surface activities of low-activity uranium contamination.
- Michaud et al. (2021) describe a hierarchical Bayesian model to localize radiation sources amid a varying background.
- Arahmane et al. (2022) use Bayes factors to simulate the detection of weak uranium signals on concrete.

4.1.6.4 Counts to Dose Conversion Factors

To varying degrees NUREG-1507 (NRC 2020a), MARSSIM/NUREG-1575 (NRC 2020c), Abelquist (2014), Alecksen and Whicker (2016), and Brown et al. (2018) discuss components of the counts-to-dose conversion factors.

4.1.6.5 Varying Background

See Section 4.1.5 for details.

4.1.6.6 Additional Sources

The following research papers were identified as potentially applicable to scan MDC methodology for GPS/GIS-based surveys planned without vigilance. Upon review, we determined that they do not address scan MDC specifically, but do address uncertainty reduction, calibration, survey design, radiation monitoring and correlation between different measurement methods. These may prove useful to future research and development of scan MDC methodologies. We provide a brief summary of each paper below.

- Sanada and Torii (2015) describe the use of unmanned helicopters (UAS/UAV) equipped with radiation detector devices capable of discriminating between ^{134}Cs and ^{137}Cs. They present the use of an exponential equation to model dose above the ground and on the ground. The combined air- and ground-monitoring methodology is practical enough to conduct multiple monitoring determinations. In this way, effective characterization of radiation and background radiation can be carried out.
- Sanada et al. (2019) present results from conducting radiation measurement via manned and unmanned helicopters (UAS/UAV) used to monitor areas not easily accessible to humans. Modeling is carried out using exponential equations and time series analysis of data collected between 2012 to 2016. The method is useful for evaluating air dose rate via widespread use of manned and unmanned helicopters. Several other works have cited this paper, most of which present applications to Fukushima Daichi monitoring data.
- Peeva (2021) is an article from the International Atomic Energy Agency that discusses the potential for unmanned aerial vehicles to be used in radiation monitoring reports with

applications to the use of drones to monitor the area around the Fukushima Daichi nuclear plant.

- Lee and Kim (2019) review a variety of drones, radiation sensors and mission types (e.g., source-seeking, wide-range monitoring, etc.), and a variety of parameters (e.g., altitude, velocity, and sensor array geometries). They formulate an optimization problem to help design UAS/UAV-based systems and apply it to optimize UAS/UAV radiation sensor systems used in aerial surveys using an objective function formulated as a 'Figure of Merit.' Simulation results are compared to experimental data. Others have cited this work; however, other than some ML and neural networks on training UAS/UAVs, there does not seem to be much difference in the citing documents.
- Tanigaki et al. (2013) develop a UGS/UGV survey for "real time data accumulation of multiple mobile monitoring stations, such as monitoring cars." In Japan, cars with large-volume gamma-ray detectors drive prefectures collecting data. They state that the car fleet system did not work during the Fukushima Daichi disaster because infrastructure was damaged, and the cars could not be driven. Their solution was to use cloud computing to alleviate the communication problem. One recommendation was to use simple text files to store data, where the correct format can make collection, transmission, and analysis of data easier, which is especially important during emergencies.
- Sinclair et al. (2016) develop a Monte Carlo-based calibration process as an alternative to a uniform test line for calibration. The calibration was used to monitor three planned detonations of radiological dispersion devices in Alberta, Canada, where monitoring was carried out using large-volume gamma radiation detectors mounted on helicopters. They found that only 20% to 25% of the original activity from the device was deposited within 1.5 km of ground zero. The paper has information on pre-flight, during-flight, and post-flight preparations.
- Wilhelm et al. (2017) develop a method based on using narrow windows centered on and to each side of absorption peaks. Spectral denoising and background suppression are employed to diminish the effect of statistical fluctuations to isolate the signature of interest. Noise-adjusted singular value decomposition to remove non-correlated noise is used. The method is presented as a solution to challenges with airborne detection, including low counts due to short acquisition time and large distances, which can lead to unacceptably large uncertainties, and the variability in the features sampled (vegetation and soil). Results for airborne gamma-ray monitoring over extended sites are presented.
- Marques et al. (2021) compile and review information on radiation detection systems, sensors and platforms. This paper contains an extensive and very complete literature review that focuses on advancements made after the Fukushima Daichi disaster.
- Kock et al. (2014) develop a model containing a systematic component and a stochastic component that applies a moving average filter to the last *n* samples. The application is focused on UGS/UGV mobile gamma-ray spectrometry. They apply the method to data from Swedish national surveys to assess radiation background and correct radiation gradients that could otherwise lead to errors when monitoring for specific sources of radiation.
- Kock and Samuelsson (2011) compare results from geological surveys in Sweden with data obtained using a portable backpack system. The objective of their work is to correlate large-scale data (airborne) with information obtained using a car (UGS/UGV) or other small-scale system. Correlation between areas should be present for data from both types of sources but is not necessarily observed. For this reason, they create a transfer function to help correlate the data from large-scale and small-scale sources so that scaling between the two sets can be carried out.

- Ji et al. (2020) describe the use of a mobile gamma-ray spectrometer based on a lanthanum bromide scintillation radiation detector. They report good correlation between gross count rate and mobile results.
- Ji et al. (2019) present results of a multi-purpose system for environmental radiation surveys using two lanthanum bromide scintillation radiation detectors, mounted on a tripod, a backpack or aerial platform. The paper studies consistency across these various platforms and comparison to HPGe detector to compare survey results of ambient dose rates with those of natural radionuclides in a well-characterized site. The tripods and backpack surveys achieved consistency with the HPGe detector.
- Falciglia et al. (2018) describe results from an experiment to assess the effects of altitude, inclination, and detection on aerial detector efficiency and minimum detectable activity. They fit calibration models so that quantitative characterization can be performed.
- Azami et al. (2018) present results from surveying the Ukedo River, near Fukushima Daichi, with the objective of gathering data from multiple surveys of the same location to monitor radiation movement. They present results on how and where soil samples were obtained.
- Abd Rahman et al. (2020) present results of using a commercial mobile robot for radiation mapping with capabilities including generating a physical map of the targeted environment. They report that the first robots sent into Fukushima Daichi were lost due to lack of good remote connection and poor operability. They use a Turtlebot2 robot equipped with a Geiger Muller detector to perform UGS surveys. A map fed to the system was used by the robot to map the radiation in the region of interest. First, data were collected using the robot in a controlled environment with known radiation sources. Then, data were processed and analyzed, and results from the application were presented.

4.1.7 Summary

There are several advantages to using the methods above to identify elevated measurements for autonomously collected data:

- Many methods are readily available in open or free software (VSP, ProUCL, R).
- The USL is an estimate of the maximum background value, so we do not expect any observed measurements to exceed the corresponding IL_{pp} at a site that does not have residual radioactivity. Hence, the resulting elevated area identification is expected to be quite accurate under these conditions.
- The UTL and USL methods provide confidence statements and uncertainty quantification for the resulting estimates, although z-scores do not.
- The z-score approach, conversely, is particularly effective at identifying hot spots in normally distributed or lognormal distributed data.
- Nonparametric UTL and USL methods avoid potential misspecification of the underlying distribution.
- LISA is a local method that incorporates local spatial pattern or local spatial correlation to identify clusters of locations with high values (HH) or single high-value locations surrounded by low-value locations (HL), which indicate potential areas of high contamination (hot spots).
- The LISA approach can also find clusters of places with statistically significant low values (LL) or single low-value locations surrounded by high-value locations (LH), indicating possible low-

activity zones. These areas may be investigated for compliance as part of the data quality assurance process in the context of UAS/UAV or human-based continuous data surveys.

- The lag-k method provides an approach to determining *a priori* scan MDC values when residual contamination hot spots are expected to be present in small areas and background radiation levels vary across a site. It performs local differencing (computing differences between observations and their neighbors) to ensure that areas in which readings are slightly higher because of historical use or geological properties are not overly likely to trigger further investigation. It also improves the ability to detect and flag areas of elevated contamination for follow-up within regions with lower-than-average background levels.
- Alecksen and Whicker (2023) show that scan MDC can be calculated for GPS/GIS-based without vigilance surveys using the existing NUREG-1507 approach (NRC 2020a), expanding on Alecksen and Whicker (2016) where they develop a probabilistic method of calculating scan MDC. In 4.1.6.2, PNNL showed that the probabilistic-based scan MDC values available from the ERG Calculator can be approximated using predictive models that could be incorporated into VSP to use existing methods and results while also allowing some extrapolation to provide scan MDC values for conditions not included in the ERG Calculator, providing expediency for practitioners and making accessing the online calculator unnecessary.
- Scan MDC calculations will need to account for many confounding factors, such as type of terrain and multiple radionuclides, in addition to analyst variability and sources of uncertainty in the instrument used and analysis of data that are not explicitly estimated in scan MDC calculations. Depending on the survey objectives, conducting more than one GPS/GIS-based survey with unique survey-specific scan MDC values would allow multiple survey objectives to be met, like identifying large or small areas of residual contamination.

However, there are also several shortcomings:

- The UTL is an estimate of the upper bound of the selected percentile of a population (either reference area or site). This implies that there is some positive probability, although likely small, that the corresponding IL_{pp} would be exceeded in the absence of residual radioactivity. When comparing site data to a USL or UTL, any observation that exceeds the IL_{pp} is expected to originate from an area of residual radioactivity. However, when observations are compared to the IL_{pp} estimated from a reference area using the 95th percentile / 90% confidence UTL, it is expected that $10\% = 1 - 90\%$ of site observations that are actually background would be identified as elevated areas. This is the false positive rate α, where $\alpha = 1 - confidence$.
- The UTL, USL, and z-score methods rely on the assumption that observations are uncorrelated. Applying these methods to correlated observations but assuming they are uncorrelated could result in error rates that are higher than nominal levels (see Section 3.0). The literature search performed for this report did not identify statistical UTL and USL methods that account for (spatial) correlation; this could be an area for future research.
- UTLs, USLs, and z-scores are global methods that identify elevated observations relative to a single IL_{pp} value applied to an entire site. While these locations are useful for identifying where to begin a search for areas of concern, they do not provide an estimate of the potential spatial extent of the area of concern.
- Nonparametric calculations require large numbers of observations. An alternative (e.g., a lag-k approach) could significantly reduce the effective number of samples required. The lag-k method may not be suitable for sites with large areas of distributed residual contamination near the regulatory limit. The introduction of the additional lag parameter, k, required in the

lag-k method must be specified for the analysis. The expected hot spot size can be used to for determine a meaningful physical basis for the lag size.

- While the models and methods presented in Section 4.1.6 have been demonstrate as useful for calculating scan MDC values, they should be tested in further simulation studies and with field data to identify further strengths and weaknesses. The predictive models should be further validated and tested, and additional radionuclides should be considered to fully assess the feasibility of implementing such approximations in VSP.
- Of the methods described in Section 4.1, only the UTL method (both parametric and nonparametric) is implemented in VSP. The remaining methods (USL, z-score, and LISA) should be considered for future VSP updates. Simulation and studies using field data could employ multiple methods (e.g., all or select methods described in Section 4.1) to systematically compare and contrast their merits and shortcomings. Such results would facilitate prioritizing their implementation in VSP.

4.2 Boundaries of Areas to Revisit

Methods that account for spatial correlation and variability can address shortcomings of UTLs, USLs, and z-scores. Z-scores computed using a spatial moving window, similar to Fortin et al. (2023), can identify locally elevated observations. Geostatistical methods such as kriging can be implemented with or without assumptions about the data-generating distribution. Kriging can be used to identify the potential spatial extent of areas of concern. Zhang et al. (2008) and French and Hoeting (2015) developed Bayesian geostatistical approaches for identifying areas of elevated concentrations. Further communications with NRC are required before making recommendations.

Many traditional kriging methods rely on estimating variograms using least squares or likelihood methods. When using these estimation methods for kriging, it is typically assumed that the estimated covariance parameters are fixed. A shortcoming of this assumption is that uncertainty in estimated variogram parameters is not propagated to estimates of uncertainty in kriging predictions. Bayesian statistical methods, now commonly available in software and computationally feasible, offer an attractive means to overcome this shortcoming by incorporating variogram parameter uncertainty into kriging predictions. Furthermore, Bayesian uncertainty intervals offer intuitive interpretability to the user by providing probability statements about quantities of interest rather than probability statements about a method (e.g., about coverage rates of a confidence interval method). Most, if not all, kriging algorithms in VSP could be enhanced with a Bayesian approach.

Section 4.2.1 summarizes the use of kriging to identify boundaries of elevated residual contamination in VSP.

Section 4.2.2 discusses fixed rank kriging (FRK), which is a method incorporated into VSP in 2022. FRK is more computationally efficient than other kriging algorithms, which could become an issue with large amounts of data generated from GPS/GIS-based scanning surveys. FRK is also flexible in that it allows for multiple resolutions of data, so that, for instance, ground and aerial GPS/GIS-based (UGS/UGV and/or UAS/UAV) survey data can be combined into a single kriged surface. The approach to identifying boundaries of potentially elevated areas using FRK would be similar to that of the other kriging methods.

Section 4.2.3 discusses Markov Bayes as presented in the Spatial Analysis and Decision Assistance (SADA) Version 5 User's Guide (Stewart et al. 2009; Goovaerts 1997). It is not a

strict Bayesian method but does appear to have some usefulness in predicting values at locations without data, accounting for prior knowledge about the probability of residual contamination.

Section 4.2.4 discusses a method for searching for the presence of elevated values that might be clustered, as when searching for unexploded ordnance. If a site is transect-surveyed with less than 100% coverage, this module can be used to delineate boundaries of areas to revisit if elevated values are found.

Section 4.2.5 discusses ML methods.

4.2.1 VSP Kriging

VSP uses the kriging algorithms available through the Geostatistical Software Library (GSLIB) (Deutsch and Journel 1998), which includes ordinary, simple, and indicator kriging. As of the date of this report, VSP implements two-dimensional versions of these algorithms, but the extension to three dimensions is a matter of “turning on” those sections of the code, software testing, and updating help files and documentation. Additionally, the VSP implementation assumes isotropy but could be extended to include anisotropy as well.

VSP can produce maps of kriged surfaces, along with maps of the conditional variance, interquartile range of predictions, and a reference uncertainty index. VSP can produce contours on kriged surfaces with auto- or user-defined values for contour, like those shown in Figure 18. Additionally, it can delineate boundaries that account for uncertainty in the kriged estimates but using a method that is based on the probability of exceeding a specified threshold or based on an upper confidence limit. Regions based on an upper confidence limit have the interpretation that for a given location *outside* the boundary, there is 95% confidence that the values do not exceed a specified threshold.

Figure 19 shows how probability contours in VSP can be used to account for uncertainty in kriged estimates. In this figure, the black squares represent measured data points, the colored background raster is the kriged spatial estimate map, and the dark blue line shows the contour created from a simple threshold of 200 on the spatial estimate map (delineating the kriged estimates with values greater than 200). The light blue contours are generated from delineating the locations where the probability of exceeding 200 is 10% or more (that is, areas outside the boundary have a 90% probability of not exceeding 200). In addition to creating a larger boundary around the main hot spot, several other areas are identified as potentially exceeding the threshold level of 200, even though the kriged estimates in those areas are all below 200.

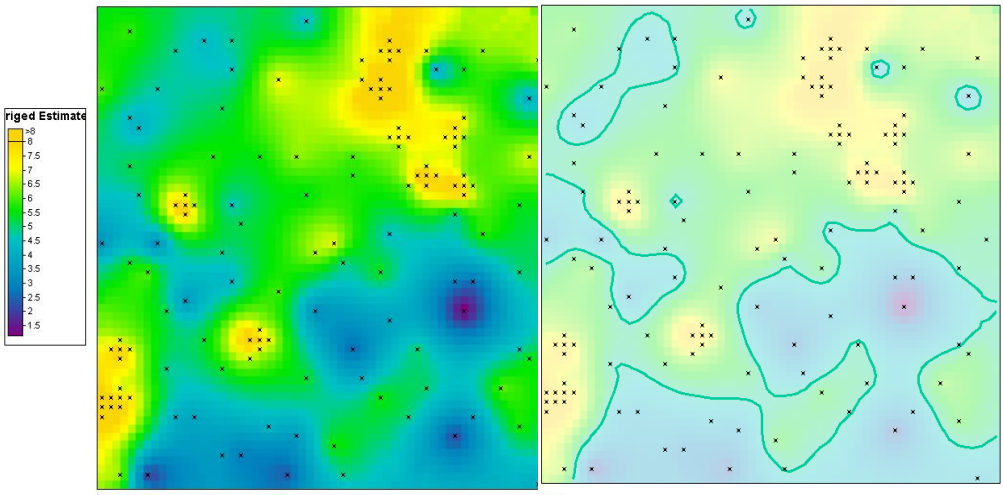

Figure 18. Example kriged surface output from VSP and associated contour. A black ‘x’ marks the sample locations. The left panel is the kriged surface; the right panel is a contour at the threshold level of 4.75.

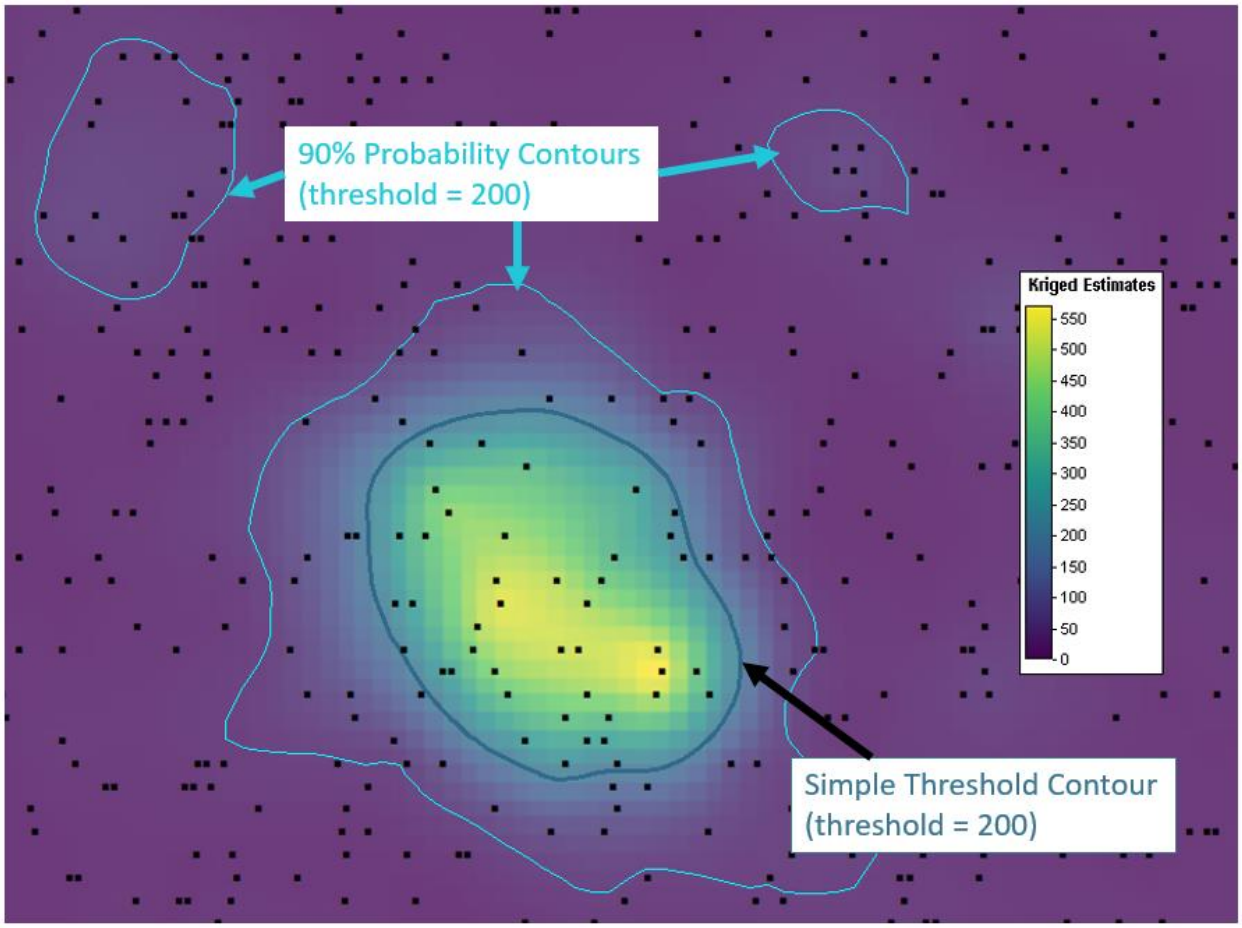

Figure 19. Example VSP output of a kriged region demonstrating how delineated probability contours better account for uncertainty in the spatial estimate and identify regions where there is less confidence of being below a specified threshold.

4.2.2 Fixed Rank Kriging

The FRK algorithm being implemented in VSP is based on the approach of Cressie and Johannesson (2008 and 2016) and the R package FRK (R Core Team 2020; Zammit-Mangion and Cressie 2021). This method uses the concepts of a spatial random effects model and a basic areal unit over which observations are averaged in the kriging equations. The spatial random effects model includes terms for large (site wide, endogenous correlation) and small (nearby, exogenous correlation) scale spatial structure. FRK is a possible solution when large amounts of data make standard kriging calculations slow or intractable, although other solutions are available for the large data problem (Heaton et al. 2019). FRK allows data with different fields of view (i.e., datasets with different supports such as ground scans and aerial scans) to be combined into a single model (Cressie and Johannesson 2008 and 2016; Zammit-Mangion and Cressie 2021). An additional benefit of FRK is that it employs a flexible covariance structure that does not assume stationarity or isotropy. As with other kriging methods, uncertainty in the kriged estimates can be incorporated into the identification of boundaries of a potential area of elevated residual activity.

Application of this method to continuously collected data should be explored, especially with respect to its flexible covariance assumptions. PNNL released a new version of VSP that includes FRK in September 2022.

4.2.3 Bayesian Ellipgrid and Markov Bayes

SC&A (2022) recommends Bayesian Ellipgrid for initial survey design, before a detailed contamination of concern map is available, and Markov Bayes for secondary survey design to indicate where additional data need to be collected. Markov Bayes and Bayesian Ellipgrid are not strictly Bayesian methods but use a conditional probability framework to incorporate “soft data” described by Stewart et al. (2006) as well as “hard data” (observations). SC&A proposed using Bayesian Ellipgrid as the initial survey design, based on knowledge about potential hot spot sizes, and then after the data are collected according to this design, using Markov Bayes for secondary survey design.

Bayesian Ellipgrid is an extension of the (non-Bayesian) Ellipgrid algorithm used to create sampling designs and is intended to detect regions of elevated contamination (SC&A 2022). Unlike Ellipgrid, Bayesian Ellipgrid assumes elevated zones exist with user-specified probabilities, determined based on knowledge about the site and/or results of previous radiological surveys and site investigations. Both Ellipgrid and Bayesian Ellipgrid can be used to determine the number and location of samples. They assume a uniform probability that contamination exists. In cases where anisotropy is expected, calculations assuming isotropy will likely overestimate the number of samples required, resulting in too many samples (Stewart and Powers 2009). Assuming isotropy could also lead to sampling that does not capture, for example, contamination concentrated along transport pathways. Such considerations should be examined when using this method for survey design.

Once data are collected using the initial Bayesian Ellipgrid survey design, SC&A (2022) recommends using Markov Bayes to create a probability map (e.g., probability of contamination exceeding a threshold) by combining the “prior belief” (soft data) with observations collected through the survey (hard data). Markov Bayes is a distribution-free method that honors the observations. It can be used to create point estimates of the probability of elevated contamination in areas or volumes of the subsurface (SC&A 2022; Zhu & Journel 1993; Goovaerts 1997). Note that it is not a fully Bayesian method – it does not rely on a statistical

prior distribution or likelihood function to derive a posterior distribution accounting for distributional assumptions. Markov Bayes does not provide a measure of uncertainty in resulting probability maps. Further, information is lost when hard data are converted to zeros and ones via thresholding suggested in SC&A (2022), where there is no distinction between, for example, a measurement well below the threshold and a measurement slightly below the threshold, because both are coded as zero. The Markov assumption also needs to be checked when applying this method (Goovaerts 1997; Goovaerts and Journel 1995), a critical and non-trivial assumption.

SC&A recommends using the combination of Bayesian Ellipgrid and Markov Bayes for the purpose of characterization or initial scoping survey designs early in the radiological survey and site investigation. We agree that these methods could be appropriate at early stages and in isotropic conditions but potentially less appropriate in the compliance phase, primarily because the Markov Bayes results do not include a measure of uncertainty and therefore could not be used to support statistically-based decisions. VSP's current module for detecting hot spots implements the Davidson (1995) Ellipgrid algorithm but does not include Bayesian Ellipgrid or Markov Bayes modules. We recommend adding both, which would require new code development. Resources may exist in GSLIB and (possibly) in a SADA development version that has not been released; either or both could be leveraged for VSP development purposes.

4.2.4 Searching for an Area of Elevated Contamination (Hot Spot) via Transects

VSP also includes a method for planning transect sampling schemes to search for areas of elevated contamination in the UXO (unexploded ordnance) module (Gilbert et al. 2003; O'Brien et al. 2005). This module was designed to plan aerial geophysical surveys but could apply to uncrewed ground or surface vehicles (UGS/UGV) and/or uncrewed aerial systems or vehicles (UAS/UAV), whenever detectors are deployed in an array configuration so that a transect has some specified width. It is particularly useful when the size of a hot spot is small relative to the size of the site and 100% survey coverage is either not feasible or not required. As with other continuously collected data, the exogenous correlation structure should be accounted for prior to data analysis.

When data are collected along transects consistent with the UXO module methods, the area to be revisited would be based on implementing a moving average data adjustment (Fortin et al. 2023) and delineating areas that exceed IL_{pp} based on survey path orientation.

4.2.5 AI/ML Methods

There have been recent advances in AI/ML approaches for geographic data analysis. Typically, AI/ML methods divide a dataset into two subsets: 1) the training set, and 2) the test set. A model is formulated using the spatial attribute in the training data to create models that predict, categorize, or cluster data. The performance of these models (prediction capability) is then tested using the test dataset. "Adequate" models can then be used to make predictions for future cases or for locations that are not in the training or test datasets. Large amounts of data are required to train and test traditional AI/ML models, which may not be a constraint for the potentially large amounts of data generated through a continuous FSS but is more likely to be a factor for fixed location data or large sites where the proportion of sampled area is relatively small.

Few-shot ML is a method that requires relatively small amounts of data and is being advanced in conjunction with remote sensing techniques and high-performance forward prediction. This

approach is being developed to reliably estimate subsurface property distributions, including (but not limited to) permeability, porosity, and hydraulic conductivity, that control fate and transport of radioactive material, thereby addressing the paucity of characterization data and complexities of heterogeneous subsurface systems. Research is required to determine if and how similar approaches could be used in surface applications.

One such method includes hot spot analysis, which creates clusters of polygons to highlight where the high and low values are likely concentrated (Wang et al. 2020). Geographically weighted random forests allocate weights to data points based on their distance from one another and create predictions by combining the findings of numerous decision trees (Nikparvar and Thill 2021). Boundaries based on such predictions could be determined (as with geospatial methods), but because prediction uncertainty is not typically quantified with ML methods, confidence and/or probability statements could not be made with respect to those boundaries. Many ML methods are “black boxes” that are not easy to interpret and/or do not offer insight into underlying processes. While such inference capabilities may be lacking, AI/ML results could be used to guide data collection in conjunction to traditional geostatistical methods.

Applying ML approaches to a combination of datasets—those collected with vigilance and those collected without vigilance—may lead to predictive models and predictions that capture relationships between multimodal measurements and variables of interest that can be used for compliance survey design for GPS/GIS-based scanning surveys. By applying deep learning to integrated multimodal sensing data, the performance of ML approaches is being defined, through training with large and small datasets, to estimate governing system-scale subsurface parameters and their spatiotemporal evolution. These advancements may reduce the uncertainty of system-scale characterization and radiation dose assessments, minimize costs, and increase worker safety and protection of human health and the environment.

Additional literature review and research should be undertaken to identify feasible AI/ML methods for application to continuously collected data, both a priori and a posteriori. The reviews and research should include both simulation and field data studies to assess methods and develop a decision tree to guide practitioners in selecting the appropriate method depending on the objectives of their GPS/GIS-based scanning survey.

5.0 VSP Data Import of Continuously Collected Data

This section discusses the file size constraints for the current version of VSP, as well as other factors that could affect users' ability to analyze continuously collected data in VSP.

Basic data import of continuously collected data is available in VSP for any format of .csv or .xlsx data, with flexibility as to column order and mappings. The primary distinct challenges for import and management of continuously collected radiological survey data are as follows:

- Data size. This type of data can be extremely dense and, depending on the area covered, can include a large number of values. The ability to import and manage upwards of 70,000 data points using its standard data storage methods has been demonstrated at PNNL. Improved memory management and processing may be necessary for datasets much larger than that and may be desirable for usability improvements if datasets of that size will regularly be processed.
- Disparate sensor platforms. These can be managed as separate analytes in VSP, but methods to track disparate sensor platforms while still combining analysis from all data sources would be valuable.
- Unit conversion and decay correction. Currently, processing to account for unit conversions and/or decay corrections needs to be completed outside of VSP, since VSP operates under the assumption that data values for a particular analyte are all directly comparable, e.g., they are drawn from the same statistical population.
- Multiple disparate sample matrices, for example, groundwater and soils data. These can also be managed as separate analytes in VSP, even if they are measuring the same radionuclide. However, this limits the ability to visualize data since visualization of data colored by value is limited to one analyte at a time. Adding the capability to visualize multiple analytes or sample matrices on the same map by using different sample symbols or multiple color scales would provide a better comprehensive picture of sample results.

Data cleaning and pre-processing is typically necessary before import into VSP. However, common issues that arise can be addressed in VSP's import process. For example, in earlier versions of VSP, a comma in a field value such as a name in the format "Lastname, Firstname" would disrupt the data import and cause errors. Subsequent updates to the VSP data import enabled VSP to correctly handle commas in input fields, making the data import process more efficient. While the VSP strategy does not attempt to fully replicate the data processing functionality available through Excel, simple changes should be made whenever possible to simplify data cleaning activities necessary for import.

Additionally, as new technologies and GPS/GIS-based survey methods evolve and are utilized to collect continuously collected data, there will likely be increasing needs for data aggregation and data quality assurance. For example, data from multiple streams (e.g., sensor, GIS, and GPS) will need to be aligned to ensure spatial and temporal correspondence between the two instruments. Data will need to be evaluated for potential instrument failure (e.g., dust contamination, overheating) and/or GPS/GIS-based survey execution that renders the data unreliable (e.g., UGS/UGV driving too fast, UAS/UAV flying too high). To the extent that these activities and decisions about the data impact statistical analysis and decisions, functionality to execute them should be included in VSP and captured in the generated report for regulatory review.

6.0 VSP Data Visualization Considerations

Visualization of continuously collected GPS/GIS-based survey data presents a number of challenges and requires additional considerations compared to visualization of more sparsely collected data. GPS/GIS-based survey data can certainly be displayed as values at discrete locations spaced along the collection time interval, but depending on the interval and the scale, it can be difficult to distinguish individual values and how they vary in distance and time.

In Figure 20, using the default VSP method to color map location values results in a cluttered image since the values are displayed with a black outline that enhances clarity for sparse data but was not designed for dense data. The survey values can be made much easier to see using a custom VSP sample symbol, as shown in the middle image, which eliminates the unnecessary outline. This type of visualization would be a superior default option for data imported and flagged as “survey data”. However, the density of the data value points can still mask fluctuations in the survey values, especially when there is small-scale variability.

The localized elevated regions are more clearly seen in the lowest image, which shows the results of kriging the survey values. It would be possible to develop an intermediate visualization method, between the middle and the lowest image, that performs some averaging and smoothing of the densely spaced survey data, without needing to develop geostatistical estimates for the entire site. Such an option could be useful for more accurately visualizing survey data.

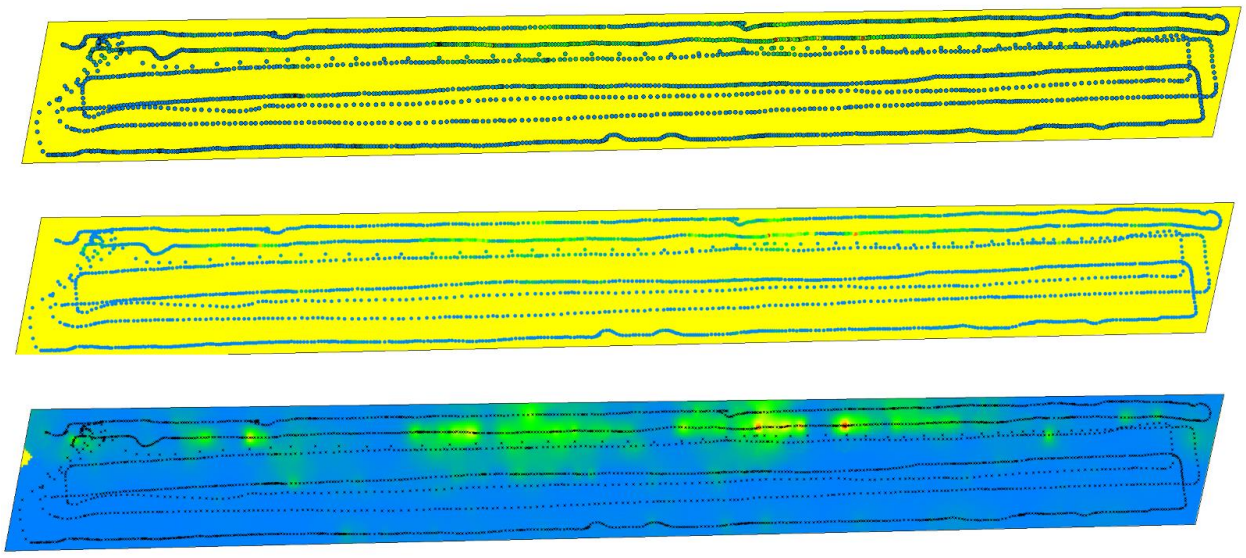

Figure 20. Default VSP color sampling method (top image), improved clarity through use of sample symbols (middle image), and geostatistical analysis results (bottom image).

Many types of sensors have spatial extent or spatial representativeness, such as a detector with a particular field of view or a UAS/UAV that represents different areas of coverage depending on the altitude. Accounting for the field of view in the data values in visualization would allow for more accurate and effective visual communication of what the GPS/GIS-based survey data values represent. This capability would be particularly useful when implementing FRK, as that module is designed to use data from multiple instruments with different fields of view.

7.0 Conclusions and Recommendations

VSP includes several geospatial methods that are useful for analyzing continuously collected data. Additional methods and tools would increase VSP support of survey design for continuously collected data and subsequent analyses. This report summarizes several methods that cover testing MARSSIM-type hypotheses, identifying boundaries for areas that need to be revisited, and importing and visualizing data. Recommendations for new VSP capabilities are summarized in Section 7.1.

Methods for calculating *a priori* scan MDC are in beginning stages of development and performance evaluation and assessment in the literature. Recommendations for further research, development, testing, and evaluation of candidate methods are summarized in Section 7.2, including key evaluation metrics, conditions under which evaluation should be considered, and how simulation and field studies could fill current gaps in understanding the strengths and shortcoming of each method.

7.1 VSP Recommendations

Below, we provide a list of recommended additions to VSP, based on the material presented in this report. These recommendations include methods and functions that are sufficiently understood and expected to perform well for continuously collected data. Additional recommendations are provided in the Section 7.2 for those methods that we believe further research and development is needed.

- **Datasets**. Access to continuously collected datasets is integral for research, development, testing, evaluation, and demonstration of methods but limited at present. It is recommended that the library of simulated and real site and reference area datasets be expanded upon and that such datasets be included in VSP, used to demonstrate functionality of new VSP capabilities in the user guide and help documentation.
- **Spatial Correlation**. Currently, VSP has limited ability to account for spatial correlation and not doing so can result in reduced statistical power of hypothesis tests (e.g., to verify a site meets release criteria).
  - Methods that test for spatial correlation, both endogenous and exogenous, should be provided as options that users can apply prior to hypothesis testing.
  - GLS should be added to VSP, providing a method to account for spatial correlation in hypothesis testing. When such correlation exists, GLS functionality in VSP should include one- or two-sample hypothesis testing, including comparison of a site mean to an action limit or a reference area mean. (Note that ML methods may not be appropriate in the hypothesis testing framework.)
- **Nonuniform Contamination**. VSP provides several methods to determine whether sampling reveals elevated values (hypothesis testing) and subsequently determine the boundary of an area of concern under conditions of uniform contamination but not under conditions of nonuniform contamination.
  - VSP improvements could be made by adding hypothesis testing formulations of UTL and the USL method described in Section 4.1.1 and 4.1.2, both of which perform well under conditions of nonuniform contamination.
  - Currently, geospatial methods in VSP such as simple kriging are used to determine boundaries of potential areas of elevated residual activity. Improvements to VSP

geospatial capability that incorporate anisotropy, Bayesian, and/or conditional probability methods that would allow users to incorporate prior knowledge into the determination of boundaries. Guidance for users about the selection of prior distributions for Bayesian methods and verifying assumptions of each method would need to be included.

- **Unit Conversions**. Currently, most data processing occurs outside of VSP. VSP users would benefit from the capability to do some data processing within VSP, specifically unit conversions and combining data from disparate sensor platforms, especially as the complexity of datasets increase with the complexity of technology options.
- **Data Visualization**. Data visualization in VSP is limited. Capabilities could be added to enhance the user interface and functionality for the development of plots for data, analysis, and result visualization.
  - Improvements in the user interface to identify instrument fields of view would be useful for kriging methods that allow data from multiple sensors and/or sensor platforms.
  - The ability to display multiple analytes or multiple sample matrices using distinct symbols or color scales would allow VSP users to better visualize site characteristics in many site applications where more than one contaminant of concern may be present.

7.2 Simulation and Field Study Recommendations

Considerable work remains to assess performance of lag-k, probabilistic (Alecksen and Whicker 2016), and/or the approximation to the probabilistic method for calculating *a priori* scan MDC described in Section 4.1.5 and 4.1.6. Simulation studies and field tests with known sources are recommended to do such performance assessments. Simulation studies are useful to examine a wider range of factors than field studies, due to savings in time (effort), schedule, and costs, and they allow for controlled environments and the ability to manipulate variables with precision which causal relationships easier to isolate. However, simulation studies are limited by shortcomings in capturing site and contaminant complexities. Findings from simulation studies can help narrow the focus for field studies though and should be used to identify which factors are expected to be most impactful in the field so that they can be examined more closely. Further, datasets generated by both simulation studies and field studies can be collected into a library for wider distribution for research and development related to continuously collected data (i.e., through VSP or its website and/or via the NRC website).

Generally, performance of various methods can be assessed by applying each to calculate an *a priori* scan MDC value for a particular site investigation (simulated or real) where the ground truth is known, then completing surveys (simulated or real) to collect data (ideally, both with and without-vigilance), and then applying one or more *a posteriori* methods to the collected survey data to identify hot spots and boundaries of areas of concern. Finally, performance metrics summarized using a confusion matrix (typically including true positive, false positive, true negative, and false negative rates) can be determined by comparing the *a posteriori* results to the ground truth.

Factors that should be considered when evaluating *a priori* scan MDC methods include the following. The impact of numerous conditions on *a priori* scan MDC calculation performance should be studied, either by independently or simultaneously varying each factor.

- **Scanning parameters**. Performance assessments should be conducted for various instrument configurations and radiological sources to understand and identify any strengths or shortcomings in methodology, relative to technology and source configurations. Different

vehicle configurations (e.g., detection technology mounted both inside and outside of sleds, four-wheelers, remotely controlled vehicles, autonomous vehicles, etc.) paired with various gross-counting instruments, sources, and source configurations should be considered. Speed, distance, and shielding should be varied to understand impacts of these variables on surveys and resulting analysis of continuously collected data.

- **Source distribution**. Source distributions may adhere to Poisson or normal distribution assumptions, or those of another statistical distribution. They may exhibit independence or spatial correlation, including isotropy or anisotropy. They could be uniformly distributed or occur with variation and could comprise a single point source, many point sources, and/or elevated regions of varying shape and size.
- **Background distribution and estimation**. The source distribution factors should be considered for background distributions, in addition to whether the background tends to be high and/or low background compared to source. Background distributions can be estimated from an entire survey region or subsets of the survey region, or from reference areas where the background distribution is similar or different from the source distribution.
- ***A priori* scan MDC method**. The purpose of the recommended studies is to determine the performance of each *a priori* scan MDC method under consideration. Based on the literature review and research done in this report, it is recommended that the lag-k, probabilistic, and/or approximation to the probabilistic methods are included in simulation and/or field studies to calculate *a priori* scan MDC values. Each method has additional factors that must be considered to assess its performance.
  - Lag-k factors to consider:
    - Selection of an optimal k value.
    - Flagging values based on integration over a reasonable range of values for k to remove the requirement for a user-specified k and/or reduce the false positive rate.
    - Extend lag-k and/or other method(s) to two dimensions when scanning transects are close to one another
    - Generate test datasets that would be available for licensees and permit holders to learn how to post-process scan data collected without vigilance and apply the lag-k and/or other method(s)
  - Probabilistic method (Alecksen and Wicker 2016 and 2023) and approximation
    - Compare results from the ERG online calculator with other scan MDC methods to establish conditions that affect differences in estimated values.
    - Implement the distribution assumptions from Alecksen and Whicker (2023) for GPS-based scan MDC calculations and compare with results obtained following the NUREG-1507 (NRC 2020a) approach using normal distributions with equal variance.
- ***A posteriori* analysis methods**. *A priori* scan MDC values could be assessed on their own using comparisons to known ground truth. However, it may be more valuable to understand the performance of the suite of tools that will be used in a site investigation. It is possible that, the performance depends on the subsequent *a posteriori* analysis method used. If so, such information would provide insight into method selection beyond *a priori* scan MDC calculations.
  - The methods outlined in Section 4.1 and Section 4.2 should be applied to perform *a posteriori* analysis of scanning survey (simulated or real). Further, lag-k, probabilistic,

and the approximation to the probabilistic methods can be applied again, to determine the *a posteriori* scan MDC and compare it to the *a priori* value.

  - Further, current hypothesis testing methodology is built around individual location-based tests and does not consider correcting for the occurrence of multiple testing which is inherently done with continuously collected data. Additional consideration should be given to multiple testing concerns and possible corrections should be investigated.

- **Method recommendations**.
  - It is recommended the factors above be examined, ranked, and down-selected according to priority for initial studies. Limited initial assessments of each *a priori* scan MDC method using select key factors (and levels) of interest can direct which factors to consider in subsequent simulation and/or studies.
  - Subsequent simulation and/or field studies should use design of experiments, essential when multiple factors are present and resources are limited because it helps to determine the optimal levels of each factor that maximize or minimize the response variable, efficiently and effectively analyze the impact of the factors (and their interactions) and ensure all this is done in a statistically sound manner.
  - The collection of study results should be used to identify strengths and shortcomings of each *a priori* scan MDC method, provide a full picture of its performance under different conditions, and provide a basis from which assumptions can be developed (assumptions that must be met for each method to perform as nominally specified). The study findings and conditions will be useful in outlining one or more site conceptual models for which each approach is and is not suitable and/or a decision tree to help practitioners determine which method(s) to use given their site characteristics and investigation objectives. The use of design of experiments throughout will provide rigorous study of each *a priori* scan MDC method, leading to defensible guidance and documentation of implemented selected methods under regulatory review.

8.0 References

10 CFR Part 20 (2024). Subpart E. *Radiological Criteria for License Termination.*

10 CFR Part 50. (2024). *Domestic Licensing of Production and Utilization Facilities.*

Abd Rahman, N.A., K.S.M. Sahari, M.F.A. Jalal, A.A. Rahman, M.I. Abd Adziz, and M.Z. Hassan. “Mobile robot for radiation mapping in indoor environment,” *IOP Conference Series: Materials Science and Engineering* 785, 012021, 2020.

Abelquist, E.W. *Decommissioning Health Physics—A Handbook for MARSSIM Users*, Second Edition. Taylor and Francis, Boca Raton, FL, 2014.

Alecksen, T. and R. Whicker, “Scan MDCs for GPS-Based Gamma Radiation Surveys,” *Health Physics* 111(2): S123–S132, 2016.

Aleksen, T. and R. Whicker. “Retrospective detection sensitivity for GPS-based gamma radiation surveys,” *Health Physics* 124(6): 451–461, 2023.

Ancell, E. and B. Bean. “Autocart—spatially-aware regression trees for ecological and spatial modeling.” *arXiv 2021*, arXiv:2101.08258. 2021.

Anselin, L. “Local indicators of spatial association—LISA,” *GEOGRAPHICAL ANALYSIS* 27(2): 93–115, 1995.

Arahmane, H., Dumazert, J., Barat, E., Dautremer, T., Dufour, N., Carrel, F., Michel, M. and Lainé, F. “A reliable absolute and relative Bayesian method for nuclear decommissioning: low-level radioactivity detection with gamma-ray spectrometry,” *IEEE Transactions on Instrumentation and Measurement* 70: 1-18, 2021.

Arahmane, H., Dumazert, J., Barat, E., Dautremer, T., Carrel, F., Dufour, N. and Michel, M. “Bayesian inference based on a bivariate gamma distribution of Kibble for low-level radioactivity detection in nuclear decommissioning operations,” *Process Safety and Environmental Protection* 163: 727–742, 2022.

Azami, K., Ootagaki, T., Ishida, M., and Sanada, Y. “Characteristics of radiocesium contamination of dry riverbeds due to the Fukushima Daiichi Nuclear Power Plant accident assessed by airborne radiation monitoring,” *Landscape and Ecological Engineering* 14: 3-15, 2018.

Beale, C.M., J.J. Lennon, J.M. Yearsley, M.J. Brewer, and D.A. Elston. “Regression analysis of spatial data,” *Ecology Letters* 13(2): 246–264, 2010.

Benito, M.B. “spatialRF: Easy spatial regression with random forest.” *R package version* 1.1.0, 2021.

Bivand, R.S., E.J. Pebesma, and V. Gomez-Rubio. *Applied Spatial Data Analysis with R*, Second Edition, Springer, New York, New York, 2013.

Brandl, A., “Statistical considerations for improved signal identification from repeated measurements at low signal-to-background ratios.” *Health Physics* 104(3): 256-263, 2013.

Brandl, A., and A.D. Herrera Jimenez. “Statistical criteria to set alarm levels for continuous measurements of ground contamination.” *Health Physics* 95(2): S128-S132, 2008.

Brodsky, A., and R.G. Gallaghar. “Statistical considerations in practical contamination monitoring.” *Radiation Protection Management* 8(4), 1991.

Brodsky, A. “Exact calculation of probabilities of false positives and false negatives for low background counting.” *Health Physics* 63(2): 198-204, 1992.

Brogan, J. and A. Brandl. “Enhancing test statistics by utilizing data patterns in sequential measurement strings in radiation detection,” *Health Physics* 115(6): 698–704, 2018.

Brogan, J. and A. Brandl. “Developing Detection DecisioNS on the Absence or Presence of a Radiological Source Using a Bayesian Interaction Model,” *Health Physics* 117(6): 637–647, 2019.

Brown, S.H., R. Edge, J. Elmer, and M. McDonald. “Establishing radiological screening levels for defense-related uranium mine (drum) sites on BLM land using a recreational future-use scenario,” *Health Physics* 114(6): 588–601, 2018.

Bunn, A., K. Wagner, D.K. Fagan, H. Gadey, T. Ikenberry, K. Markham, and M.Y. Obiri. “*Draft Drones for Decommissioning*,” PNNL-32519 Revision 1, Pacific Northwest National Laboratory, Richland, Washington, 2022.

Casella, G. and R.L. Berger. *Statistical Inference*, Thompson Learning, Pacific Grove, California, 1990.

Clifford, P., S. Richardson, and D. Hemon. “Assessing the Significance of the Correlation between Two Spatial Processes,” *Biometrics* 45(1): 125-134, 1989.

Cressie, N. *Statistics for Spatial Data,* Revised Edition, John Wiley & Sons, Inc., New York, New York, 2015.

Cressie, N. and G. Johannesson. “Spatial prediction for massive data sets,” *Mastering the Data Explosion in the Earth and Environmental Sciences: Proceedings of the Australian Academy of Science Elizabeth and Frederick White Conference* 1–11, Australian Academy of Science, Canberra, Australia, 2006.

Cressie, N. and G. Johannesson. “Fixed rank kriging for very large spatial data sets,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70(1): 209–226, 2008.

Currie, L.A. “Limits for qualitative detection and quantitative determination. Application to radiochemistry.” *Analytical Chemistry* 40(3): 586-593, 1968.

Currie, L.A. “Lower limit of detection: definition and elaboration of a proposed position for radiological effluent and environmental measurements (No. NUREG/CR-4407).” National Bureau of Standards, Washington, DC (United Stated), 1984.

Davidson, J.R. *ELLIPGRID-PC: Upgraded Version*. ORNL/TM-13103. Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1995.

Davis, C.B. and P.F. Wambach. "Quasi Nonparametric Upper Tolerance Limits for Occupational Exposure Evaluations," *Journal of Occupational and Environmental Hygiene* 12: 342–349, 2015.

Deutsch, C.V., and A.G. Journel. *GSLIB Geostatistical Software Library and User's Guide*, Second Edition, Applied Geostatistics Series, Oxford University Press, Inc. New York, New York, 1998.

Dray, S., P. Legendre, and P.R. Peres-Neto. "Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM)," *Ecological Modelling,* 196(3-4): 483–493, 2006.

Falciglia, P., L. Biondi, R. Catalano, G. Immè, S. Romano, and F. Vagliasindi. "Preliminary investigation for a quasi-quantitative characterization of soils contaminated with 241Am and 152Eu by low-altitude unmanned aerial vehicles (UAVs) equipped with small size γ-ray spectrometer: detection efficiency and minimum detectable activity (MDA) concentration assessment." *Journal of Soils and Sediments 18,* 2399–2409, 2018.

Falkner, J. and C. Marianno. "Modeling Minimum Detectable Activity as a Function of Detector Speed," *Radiation Detection Technology and Methods, 3,* 1–8, 2019.

French, J.P. and J.A. Hoeting. "Credible regions for exceedance sets of geostatistical data," *Environmetrics* 27(1), 2015.

Fortin, D., J. Irvahn, L.L. Newburn, D.K. Fagan, and J.C. Huckett. "*Overview of a Methodology to Calculate the Scanning Minimum Detectable Concentration for Post-Processed Radiological Surveys*," PNNL-34211 Rev.1, Pacific Northwest National Laboratory, Richland, Washington, 2023.

Georganos, S., T. Grippa, A.N. Gadiaga, C. Linard, M. Lennert, S. Vanhuysse, N. Mboga, E. Wolff, and S. Kalogirou. "Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling," *Geocarto International 36(2):* 121–136, 2021.

Gibbons, R.D. *Statistical Methods for Groundwater Monitoring*, John Wiley & Sons, Inc., New York, New York, 1994.

Gilbert, R.O. *Statistical Methods for Environmental Pollution Monitoring*, Wiley & Sons, Inc., New York, New York, 1987.

Gilbert, R.O., J.E. Wilson, R.F. O'Brien, D.K. Carlson, B.A. Pulsipher, and D.J. Bates. "Version 2.0 Visual Sample Plan (VSP): UXO Module Code Description and Verification," PNNL-14267, Pacific Northwest National Laboratory, Richland, Washington, 2003.

Goovaerts, P. *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, New York, 1997.

Goovaerts, P. and A.G. Journel. "Integrating soil map information in modelling the spatial variation of continuous soil properties," *European Journal of Soil Science* 46(3): 397–414, 1995.

Griffith, D.A. and L.J. Layne. *A Casebook for Spatial Statistical Data Analysis: A Compilation of Analyses of Different Thematic Data Sets*, Oxford University Press, New York, New York, 1999.

Hahn, G.J. and W.Q. Meeker. *Statistical Intervals: A Guide for Practitioners*. John Wiley & Sons, Inc., New York, New York, 1991.

Hart, K., W. Duffy, K. Higley, C. Marianno, and C. Moss. “Predicting instrument detection efficiency when scanning point and small area detection sources.” *Health Physics* 84(5): 616-625, 2003.

Heaton, M.J., A. Datta, A.O. Finley, R. Furrer, J. Guinness, R. Guhaniyogi, F. Gerber, R.B. Gramacy, D. Hammerling, M. Katzfuss, F. Lindgren, D.W. Nychka, F. Sun, and A. Zammit-Mangion. “A case study competition among methods for analyzing large spatial data,” *Journal of Agricultural, Biological and Environmental Statistics* 24, 398–425, (2019).

Hengl, T., M. Nussbaum, M.N. Wright, G. Heuvelink, and B. Gräler. “Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables,” *PeerJ* 6, e5518, 2018.

Hurlbert, SH. “Pseudoreplication and the design of ecological field experiments.” *Ecological Monographs*, 54(2):187–211, 1984.

Ji, Y.Y., T. Lim, H.Y. Choi, K.H. Chung, and M.J. Kang. “Development and Performance of a Multipurpose System for the Environmental Radiation Survey Based on a LaBr 3 (Ce) Detector,” *IEEE Transactions on Nuclear Science* 66(12): 2422–2429, 2019.

Ji, Y.Y., T. Lim, K. Hitomiand T. Yajima. “Spectrometric Estimation of Dose Rate Induced from Radioactive Cesium in the Ground Using a Mobile Gamma-Ray Spectrometry Based on a LaBr3 (Ce) Detector,” *Health Physics* 118(2): 215–225, 2020.

Justus, A.L. “A Purely Poisson-based Approach to Estimating Audible Scan Survey Sensitivities,” *Health Physics* 116(1): 27–41, 2019.

Kim, J., K.T. Lim, K. Park, and G. Cho. “A Bayesian Approach for Remote Depth Estimation of Buried Low-level Radioactive Waste with a NaI(TI) Detector,” *Sensors* 19(24): 5365, 2019.

King, D., A., N. Altic, and C. Greer. “Minimum detectable concentration as a function of gamma walkover survey technique.” *Health Physics*, 102(2): S22-S27, 2012.

King, D.A., Vitkus, T. “Lessons learned on the presentation of scan data”, *Health Physics*, 109(3): S212-S218, 2015.

Klumpp, J. “Bayesian analysis of energy and count rate data for detection of low count rate radioactive sources.” ASME 2013 15th International Conference on Environmental Remediation and Radioactive Waste Management 2013.

Kock, P. and C. Samuelsson. “Comparison of airborne and terrestrial gamma spectrometry measurements-evaluation of three areas in southern Sweden,” *Journal of Environmental Radioactivity* 102(6): 605–613, 2011.

Kock, P., Rääf, C., and C. Samuelsson “On background radiation gradients–the use of airborne surveys when searching for orphan sources using mobile gamma-ray spectrometry,” *Journal of Environmental Radioactivity* 128: 84–90, 2014.

Krishnamoorthy, K. and T. Mathew. *Statistical Tolerance Regions: Theory, Applications, and Computation*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2009.

Lee, C. and H.R. Kim. “Optimizing UAV-based radiation sensor systems for aerial surveys,” *Journal of Environmental Radioactivity* 204: 76–85, 2019.

Macedo, J. and Haddad, M.A. “Equitable distribution of open space: Using spatial analysis to evaluate urban parks in Curitiba, Brazil,” *Environment and Planning B: Planning and Design* 43(6): 1096-1117, 2016.

Marianno, C.M. “Signal processing and its effect on scanning efficiencies for a field instrument for detecting low-energy radiation,” *Health Physics* 109, 78–83, 2015.

Marques, L., A. Vale, and P. Vaz. “State-of-the-Art Mobile Detection Systems for Different Scenarios,” *Sensors* 21(4): 1051, 2021.

Meeker, W.Q., G.J. Hahn, and L.A. Escobar. *Statistical Intervals: A Guide for Practitioners and Researchers,* Second Edition, Wiley, Germany, 2017.

Michaud, I.J., Schmidt, K., Smith, R.C. and Mattingly, J. “A hierarchical Bayesian model for background variation in radiation source localization,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1002: 165288, 2021.

Meyer, H., C. Reudenbach, S. Wöllauer, and T. Nauss. “Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction,” *Ecological Modelling* 411, 108815, 2019.

Millard, S.P. and N.K. Neerchal. *Environmental Statistics with S-Plus*, CRC Press, NY, 2001.

Nikparvar, B. and J.C. Thill. “Machine Learning of Spatial Data,” *ISPRS International Journal of Geo-Information* 10(9): 600, 2021.

O’Brien, R.F., D.K. Carlson, R.O. Gilbert, J.E. Wilson, D.J. Bates, and B.A. Pulsipher. “Statistical algorithms for designing geophysical survey to detect UXO target areas,” *Subsurface Sensing Technologies and Applications* 6(3), 2005.

Pebesma, E.J., and R.S. Bivand. “Classes and methods for spatial data in R,” *R News* 5(2), 2005.

Pebesma, E., Bivand, R., Rowlingson, B., Gomez-Rubio, V., Hijmans, R. J., Sumner, M., MacQueen, D., Lemon, J., O'Brien, J., and O'Rourke, J. “Package “sp”: Classes and methods for spatial data.” *R package version* 2.1-4, 2024.

Peeva, A. “Now Available: New Drone Technology for Radiological Monitoring in Emergency Situations.” *International Atomic Energy Agency (IAEA) News (website)*, February 2021.

R Core Team. *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020.

Rahman, N.A.A., K.S.M. Sahari, M.F.A. Jalal, A.A. Rahman, M.I.A. Adziz, and M.Z. Hassan. "Mobile robot for radiation mapping in indoor environment," *IOP Conference Series: Materials Science and Engineering* 785, 012021, 2020.

Romary, T., F. Ors, J. Rivoirard, and J. Deraisme. "Unsupervised classification of multivariate geostatistical data: two algorithms," *Computers & Geosciences, 85,* 96–103, 2015.

Sanada, Y. and T. Torii. "Aerial radiation monitoring around the Fukushima Dai-ichi nuclear power plant using an unmanned helicopter," *Journal of Environmental Radioactivity* 139: 294–299, 2015.

Sanada, Y., Y. Urabe, M. Sasaki, K. Ochi, and T. Torii. "Evaluation of ecological half-life of dose rate based on airborne radiation monitoring following the Fukushima Dai-ichi nuclear power plant accident," *Journal of Environmental Radioactivity* 210: 105816, 2019.

Sanderson, D. "Measuring regional scale distribution of radiocaesium." Presented at the Caesium workshop: Fukushima recovery–understanding, modelling and managing radiocaesium decontamination, COEASSE," Fukushima, Japan. September 30–October 3, 2013.

SC&A, Inc. 2022. *Guidance on Surveys for Subsurface Radiological Contaminants. White Paper.* Prepared for U.S. Nuclear Regulatory Commission Office of Nuclear Regulatory Research.

Sinclair, L.E., R. Fortin, J.L. Buckle, M.J. Coyle, R.A. Van Brabant, B.J. Harvey, H.C. Seywerd, and M.W. McCurdy. "Aerial mobile radiation survey following detonation of a radiological dispersal device." *Health Physics* 110(5): 458–470, 2016.

Stewart, R. and Powers, G. 2009. *A Subsurface Decision Model for Supporting Environmental Compliance. Report for U.S. Nuclear Regulatory Commission.* NUREG/CR-7021. U.S. Nuclear Regulatory Commission, Washington, D.C.

Stewart, R., C. Welsh, T. Purucker, and D. Stewart. *An Introduction to Spatial Analysis and Decision Assistance (SADA), Environmental Applications for Version 5, User Guide*. The Institute for Environmental Modeling, University of Tennessee, Knoxville, Tennessee, 2006.

Strom, D.J., and P.S. Stansbury. "Minimum detectable activity when background is counted longer than the sample." *Health Physics* 63(3): 360-361, 1992.

Tandon, P., P. Huggins, R. Maclachlan, A. Dubrawski, K. Nelson, and S. Labov. "Detection of radioactive sources in urban scenes using Bayesian Aggregation of data from mobile spectrometers." Information Systems 57: 195-206, 2016.

Tanigaki, M., R. Okumura, K. Takamiya, N. Sato, H. Yoshino, and H. Yamana. "Development of a car-borne γ-ray survey system, KURAMA." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 726: 162–168, 2013.

Tao, H., X. Liao, D. Zhao, X. Gong, and D. Cassidy. “Delineation of soil contaminant plumes at a co-contaminated site using BP neural networks and geostatistics,” *Geoderma, 354,* 113878, 2019.

U.S. Environmental Protection Agency (EPA). *Guidance for Data Quality Assessment: Practical Methods for Data Analysis*. EPA QA/G-9, QA00 update, EPA/600/R-96/084, Washington, D.C, 2000.

U.S. Nuclear Regulatory Commission (NRC). “Human Performance in Radiological Survey Scanning,” NUREG/CR–6364, Washington, D.C, 1997.

U.S. Nuclear Regulatory Commission (NRC). “A Nonparametric Statistical Methodology for the Design and Analysis of Final Status Decommissioning Surveys, Interim Draft Report for Use and Comment,” NUREG–1505, Revision 1, Washington, D.C, 1998.

U.S. Nuclear Regulatory Commission (NRC). “Minimum Detectable Concentrations with Typical Radiation Survey for Instruments for Various Contaminant and Field Conditions,” NUREG–1507, Revision 1, Washington, D.C, 2020a.

U.S. Nuclear Regulatory Commission (NRC). “Consolidated Decommissioning Guidance Characterization, Survey, and Determination of Radiological Criteria. Draft Report for Comment,” NUREG–1757, Volume 2, Revision 2, Washington, D.C, 2020b.

U.S. Nuclear Regulatory Commission (NRC). “Multi-agency Radiation Survey and Site Investigation Manual (MARSSIM),” NUREG–1575, Revision 2, Washington, D.C, 2020c.

Ver Hoef, J.M. and N. Cressie. “Multivariable Spatial Prediction,” *Mathematical Geology* 25(2), 1993.

Wang, Y., L. Feng, S. Li, F. Ren, and Q. Du. “A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China,” *CATENA* 188, 104425, 2020.

Watson, M.M., A.F. Seliman, V.N. Bliznyuk, and T.A. DeVol. “Evaluation of Shiryaev-Roberts procedure for on-line environmental radiation monitoring.” *Journal of Environmental Radioactivity* 192: 587-591, 2018.

Wilhelm, E., S. Gutierrez, S. Ménard, and A.M. Nourreddine. “A method for determining Am-241 activity for large area contamination,” *Applied Radiation and Isotopes* 119: 86–93, 2017.

Wohlberg, B., D.M. Tartakovsky, and A. Guadagnini. Subsurface characterization with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 44(1): 47–57, 2005.

Zammit-Mangion, A. and N. Cressie. “FRK: An R package for spatial and spatio-temporal prediction with large datasets,” *Journal of Statistical Software* 98:1–48, 2021.

Zhang, J., P.F. Craigmile, and N. Cressie. “Loss function approaches to predict a spatial quantile and its exceedance region,” *Technometrics* 50(2):216–227, 2008.

Zhu, H. and A. Journel. “Formatting and integrating soft data: stochastic imaging via the Markov-Bayes algorithm,” in: Soares, A (ed.), Geostatistics Tróia ’92. Quantitative Geology and Geostatistics, Volume 5. Springer, Dordrecht, 1993.

Appendix A – Datasets for Demonstration of Methods for Continuously Collected Survey Data

Additional datasets representing continuously collected survey data of land areas were prepared for future studies to demonstrate methods of statistical analysis and visualization. The datasets were provided by the U.S. Nuclear Regulatory Commission to Pacific Northwest National Laboratory as case studies for the case studies in Fortin et al. (2023). They were altered to demonstrate the methods therein as well as in this report and the survey elements associated with the Multi-Agency Radiation Survey and Site Investigation Manual (NRC 2020c) and NUREG-1507 (NRC 2020a). This appendix describes the datasets only and not data analysis. Any analysis performed using these data are included in Fortin et a. (2023) and in the body of this report.

These datasets support surveys classified by MARSSIM/NUREG-1575 as Class 1 and Class 3.

Class 1 Areas

Areas that have, or had before remediation, a potential for residual radioactive material (based onsite operating history) or known residual radioactive material (based on previous radiation surveys) above the DCGLW. Examples of Class 1 areas include:

- Site areas previously subjected to remedial actions.
- Locations where leaks or spills are known to have occurred.
- Former burial or disposal sites.
- Waste storage sites.
- Areas with residual radioactive material in discrete solid pieces of material and high specific activity.

Remediated areas are identified as Class 1 areas because the remediation process often results in less than 100% removal of the radioactive material. The residual radioactive material that remains onsite after remediation is often associated with relatively small areas with elevated levels of radioactive material. This results in a non-uniform distribution of the radionuclide and a Class 1 classification. If an area is expected to have no potential to exceed the DCGLW and was remediated to demonstrate the residual radioactive material is as low as reasonably achievable, the remediated area might be classified as Class 2 for the FSS.

Class 3 Areas

Class 3 areas are any impacted areas that are not expected to contain any residual radioactive material or are expected to contain levels of residual radioactive material at a small fraction of the DCGLW, based on site operating history and previous radiation surveys. To justify changing the classification of an area from Class 1 or Class 2 to Class 3, existing data (i.e., from the historical site assessment (HSA), scoping surveys, or characterization surveys) should provide a high degree of confidence that either there is no residual radioactive material or any levels of residual radioactive material are a small fraction of the DCGLW. Other justifications for this change in an area's classification may be appropriate based on the outcome of the data quality objective process. Examples of areas that might be classified as Class 3 include buffer zones around Class 1 or Class 2 areas and areas that have very low potential for residual radioactive material but insufficient information to justify a non-impacted classification.

A.1 Dataset 1

Dataset 1 includes data from a reference area and an area of concern, scanned using a continuous surveying technique without vigilance. Figure A.1 through Figure A.4 show observations of each variable in the dataset, including radiological measurements (reported in counts per minute [CPM]), geospatial information (reported in meters from the origin, similar to latitude and longitude coordinates), and speed of the detector (reported in meters per second [m/s]). In the provided data, observed CPM in the reference area were generally larger than in the area of concern and the area of concern did not have elevated measurements. Thus, Pacific Northwest National Laboratory (PNNL) altered the data so that reference area CPM observations were less than or equal to majority of the CPM observations in the area of concern. PNNL also injected elevated measurements into the northern part of the area of concern to represent a hot spot. These data can be used to evaluate hot spot detection methods.

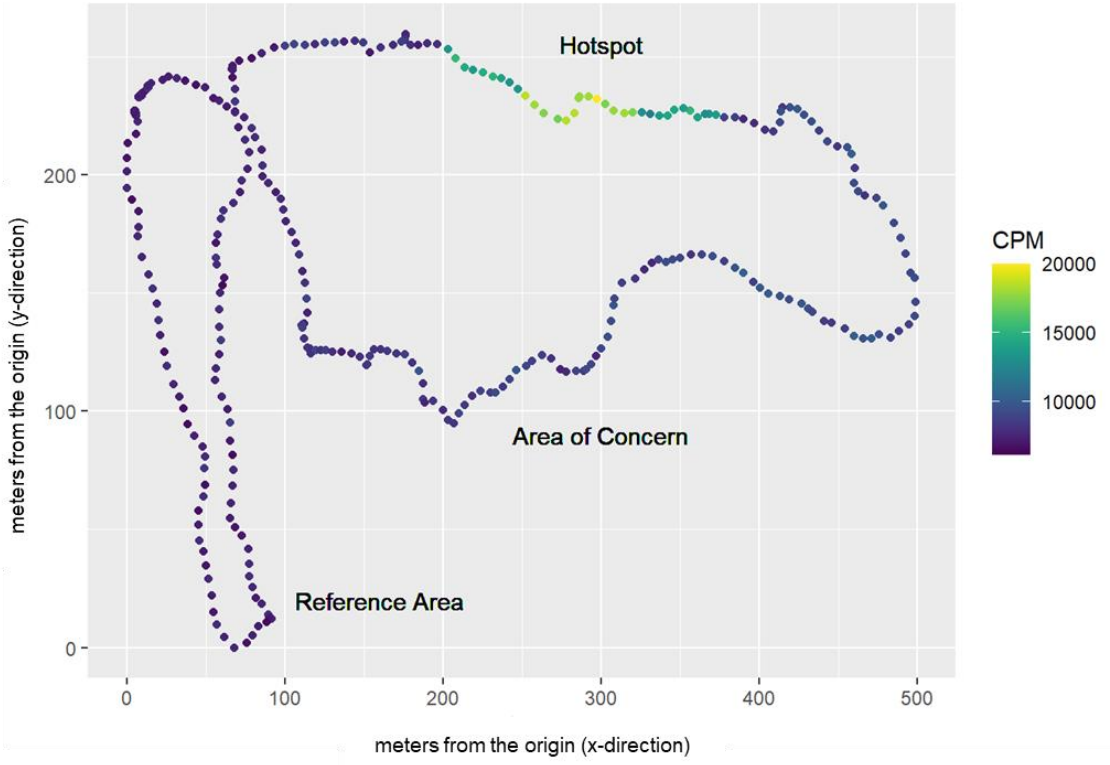

Figure A.1. Dataset 1 observed radiation measurements (CPM) across a land area with a reference area and an area of concern.

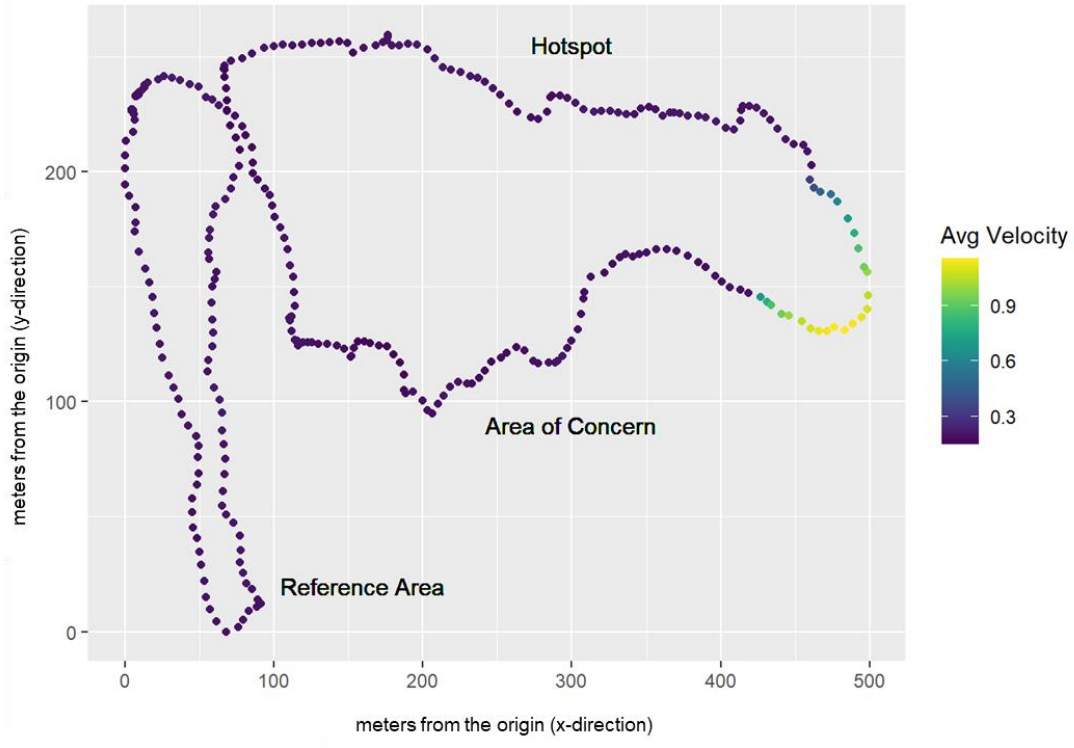

Figure A.2. Dataset 1 observed average velocity (m/s) within reference area and area of concern.

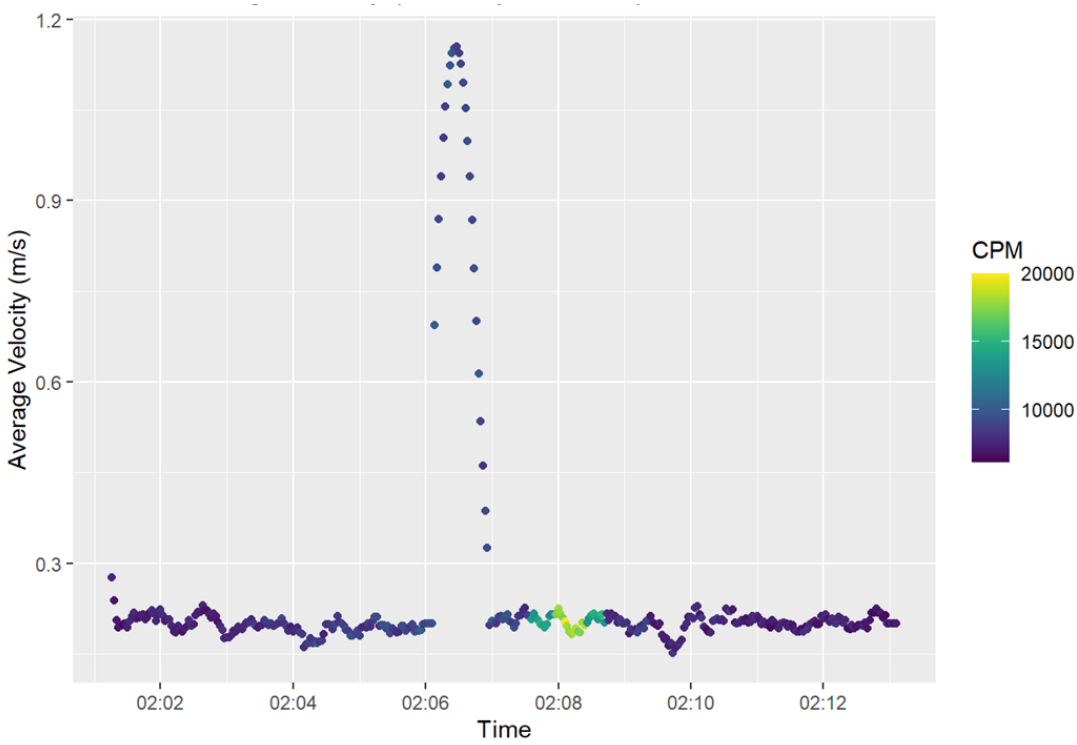

Figure A.3. Dataset 1 observed average velocity (in m/s) over the elapsed scanning time (in hours and minutes) with coloration showing the radiation measurements (CPM).

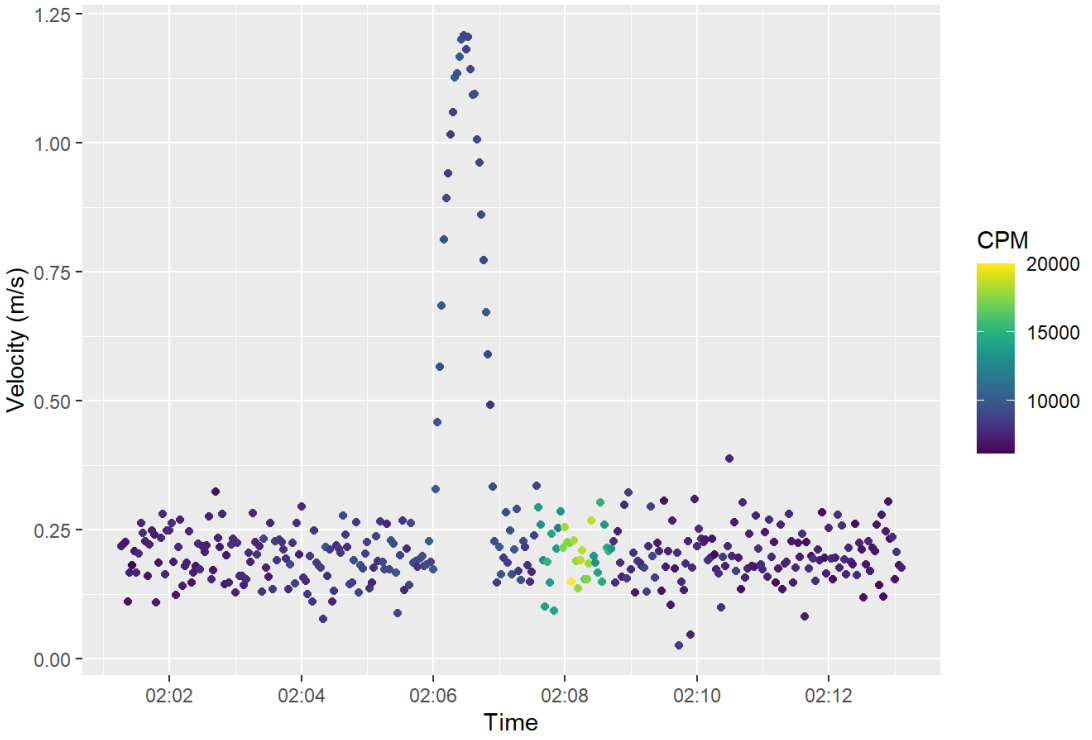

Figure A.4. Dataset 1 observed instantaneous velocity (m/s) vs. elapsed time (in hours and minutes) with coloration reflecting the radiation measurements (CPM).

A.2 Dataset 2

Dataset 2 includes the same data from a reference area and an area of concern, scanned using a continuous surveying technique without vigilance. Figure A.5 through Figure A.8 show observations of each variable in the dataset. PNNL injected an additional elevated measurement into the northern part of the area of concern to represent a second hot spot. These data can be used to evaluate hot spot detection methods.

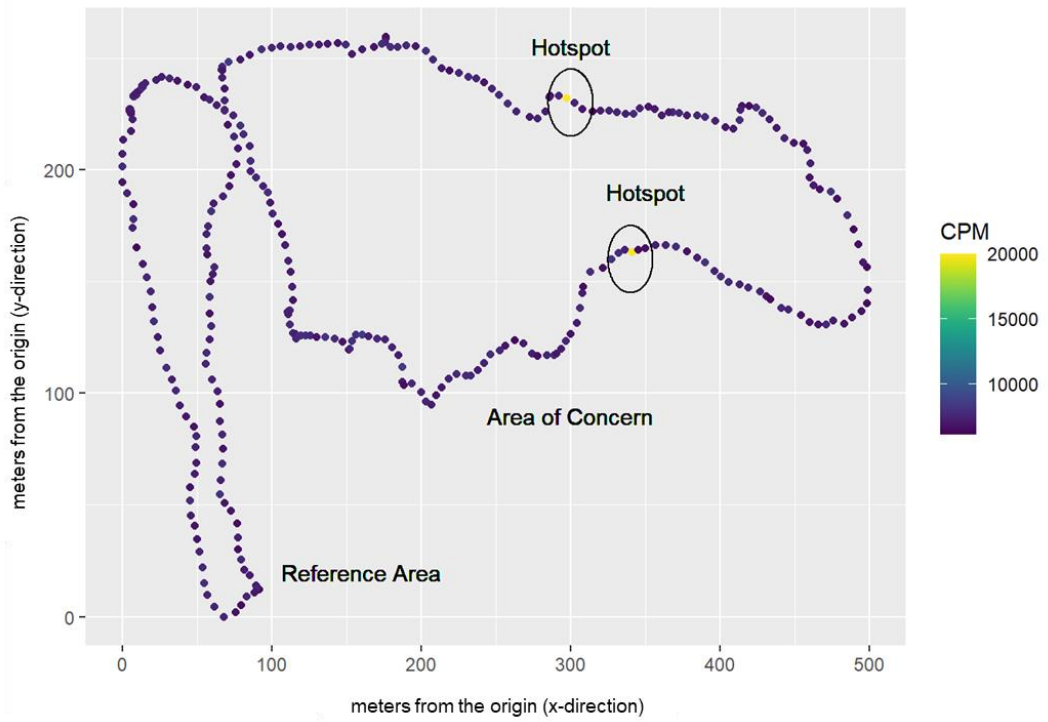

Figure A.5. Dataset 2 visualization of radiation measurements (in CPM) across a land area with a reference area and area of concern with two hot spots.

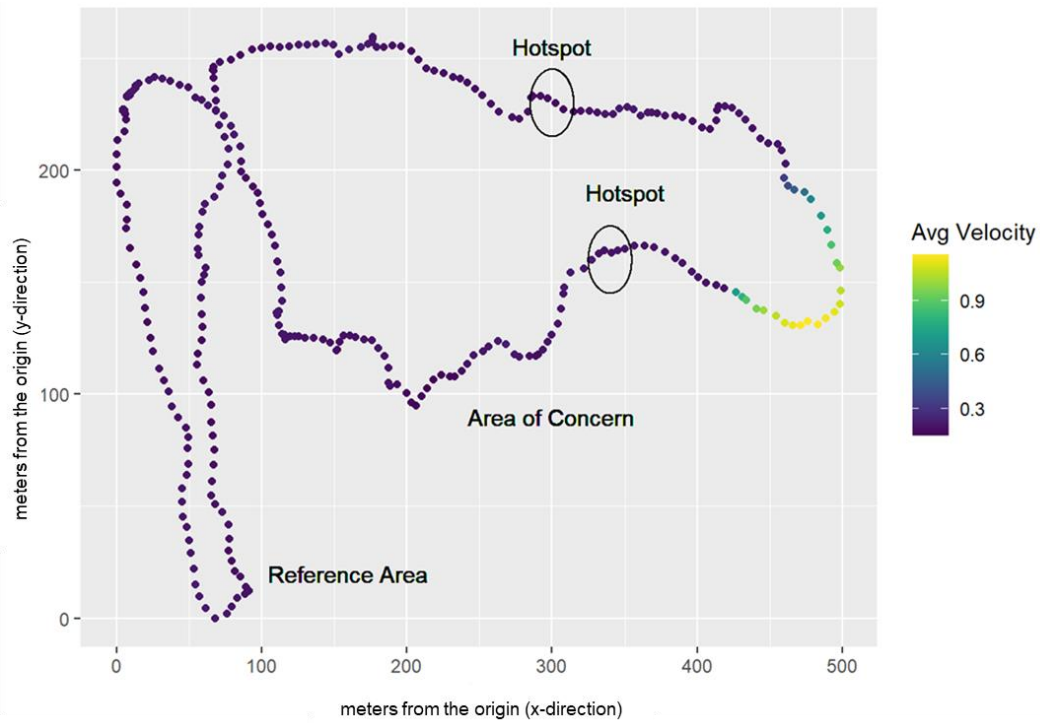

Figure A.6. Dataset 2 visualization of average velocity (in m/s) across a land area with a reference area and area of concern with two hot spots.

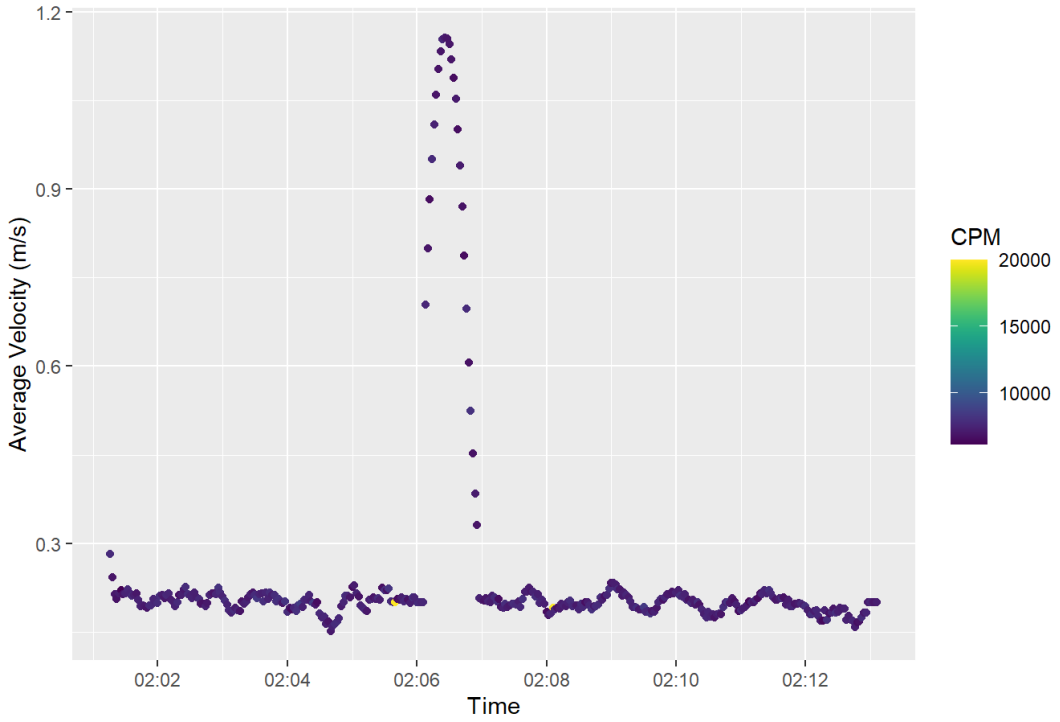

Figure A.7. Dataset 2 visualization of average velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for the land areas in Figure A.5.

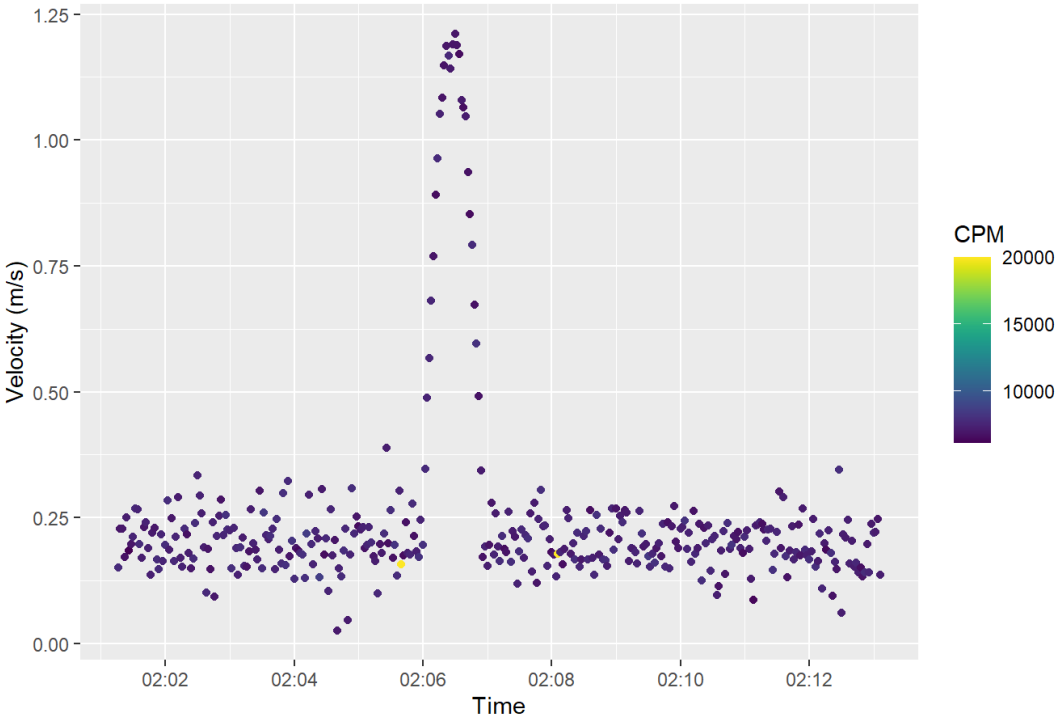

Figure A.8. Dataset 2 visualization of instantaneous velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for the land areas in Figure A.5.

A.3 Dataset 3

Dataset 3 includes the same data from a reference area and an area of concern, scanned using a continuous surveying technique without vigilance. Like described in the previous sections, Figure A.9 through Figure A.12 show observations of each variable in the dataset. PNNL injected an additional elevated measurement into the northern part of the area of concern to represent a third hot spot. These data can be used to evaluate hot spot detection methods.

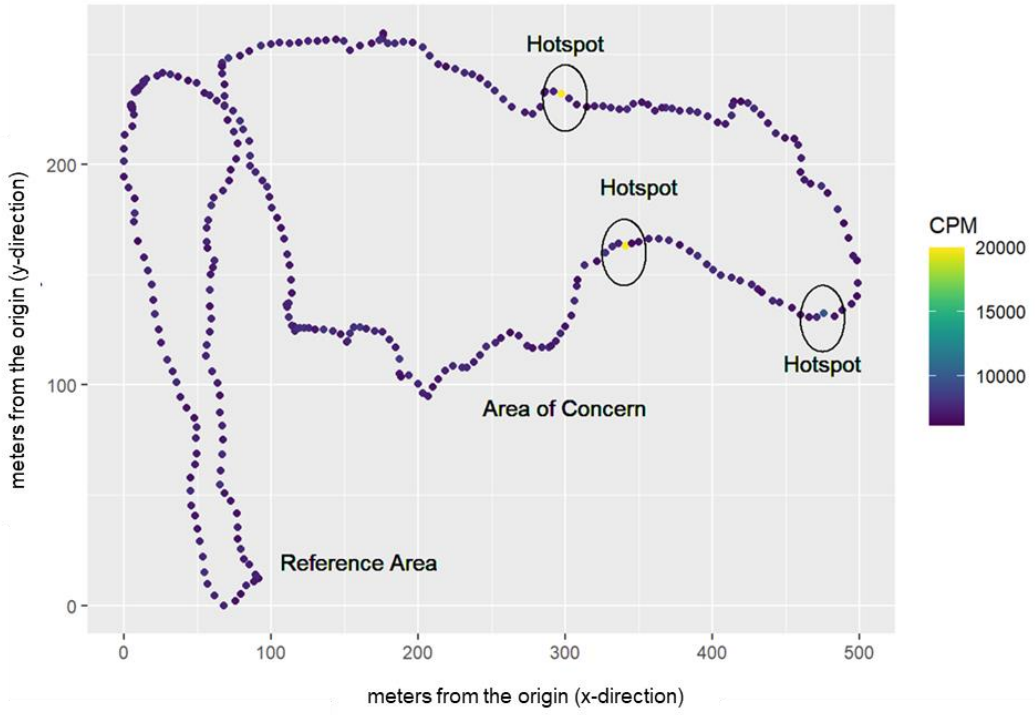

Figure A.9. Dataset 3 visualization of radiation measurements (in CPM) across a land area with a reference area and area of concern with three hot spots.

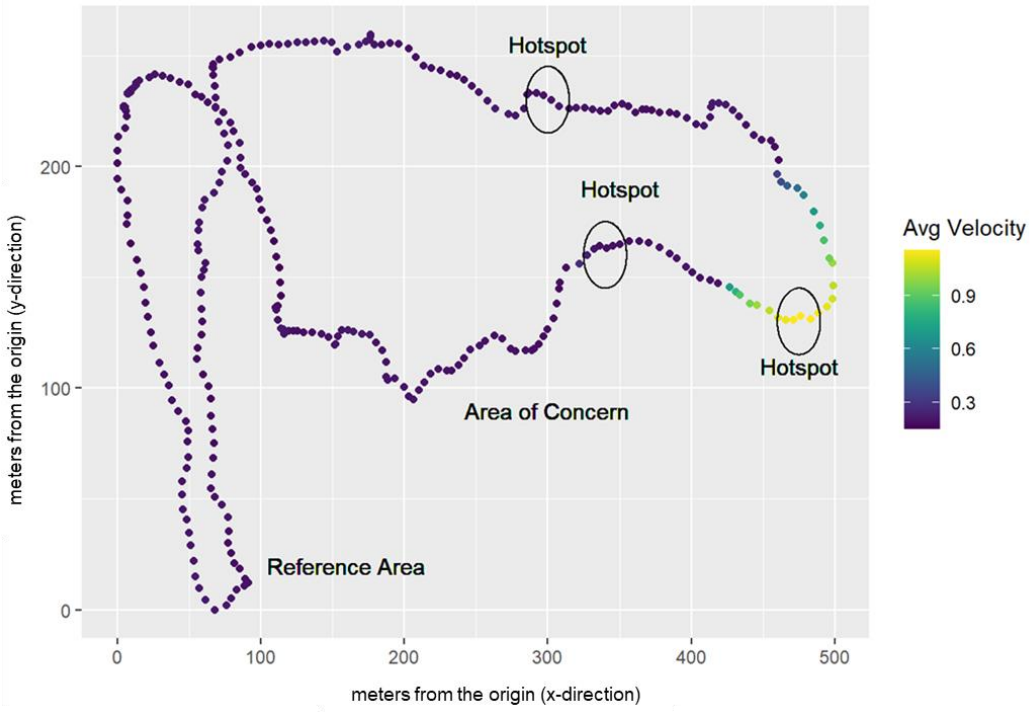

Figure A.10. Dataset 3 visualization of average velocity (in m/s) across a land area with a reference area and area of concern with three hot spots.

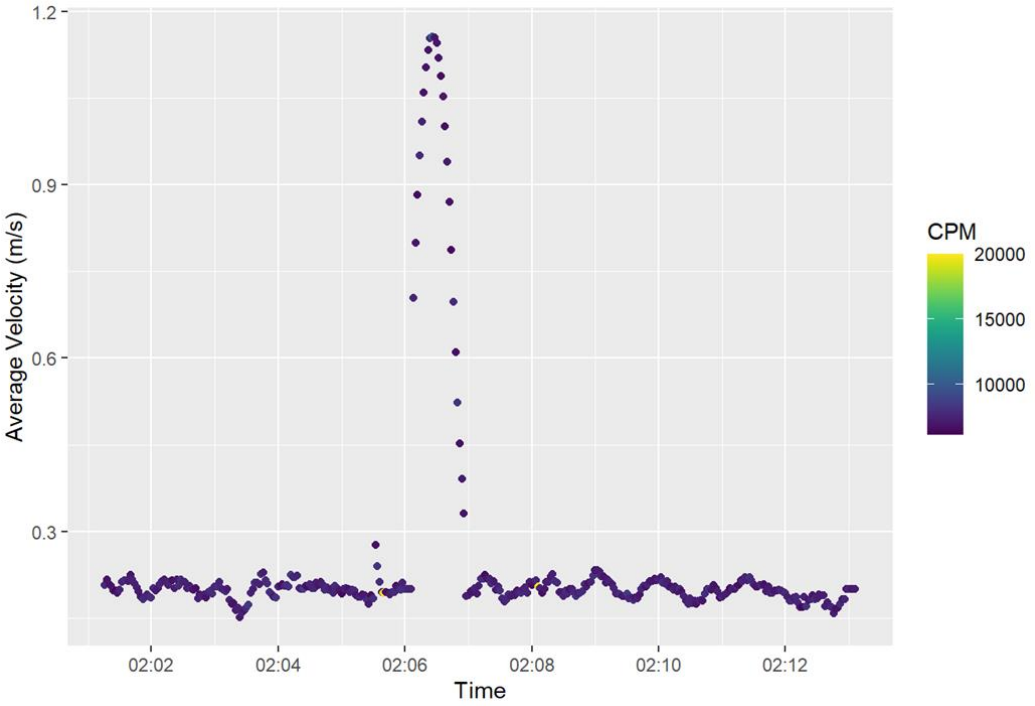

Figure A.11. Dataset 3 visualization of average velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for the land area in Figure A.9.

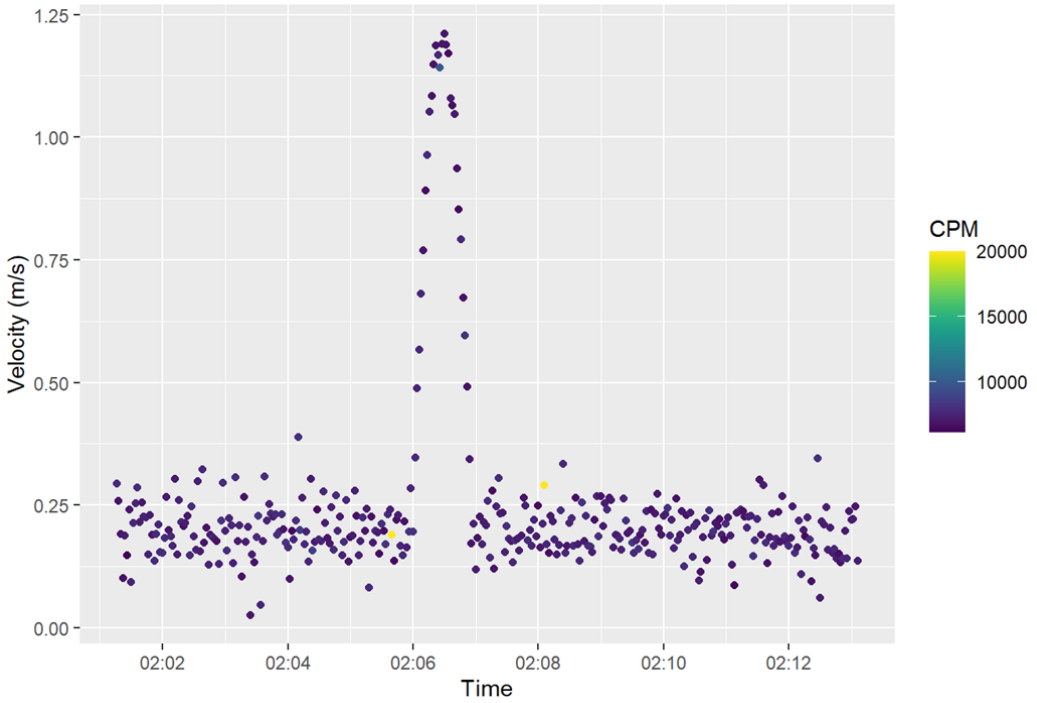

Figure A.12. Dataset 1, Test 1a, visualization of instantaneous velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for the land areas in Figure A.9.

A.4 Dataset 4

Dataset 4 includes radiological measurements (reported in CPM) as well as exogenous information, including geospatial information (reported in meters from an origin, similar to latitude and longitude coordinates) and speed of the detector (reported in meters per second [m/s]).

Dataset 4 represents four areas of concern (A, B, C, and D) and six reference areas (E, F, G, H, I, J) that were scanned using a continuous surveying technique without vigilance. Like for Dataset 1, A.13 through Figure A.21 show observations of each variable in the dataset. PNNL injected an additional elevated measurement into the northern part of the area of concern to represent a second hot spot. These data can be used to evaluate methods associated with variable background.

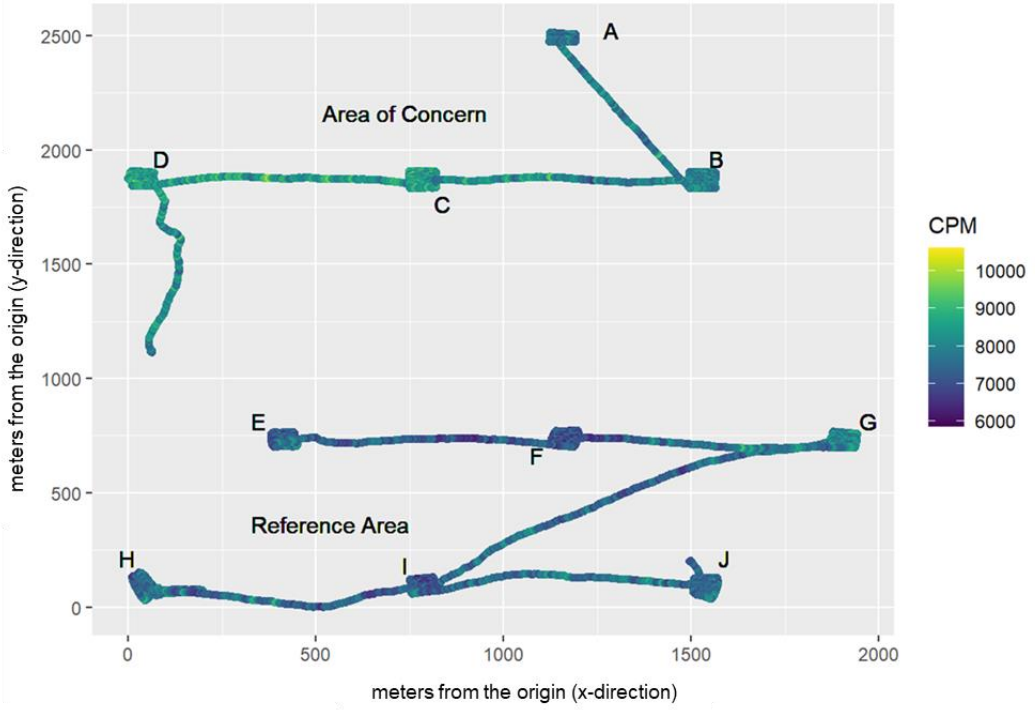

Figure A.13. Dataset 4 radiation measurements (in CPM) across a land area including several reference areas and several areas of concern.

Suppose that regions “D” has been identified as the area of concern for a specific analysis and that region “J” was identified as an appropriate reference area for region “D”. The data in Figures A.14 through A.17 and Figures A.18 through A.21 provide the data associated with each, respectively.

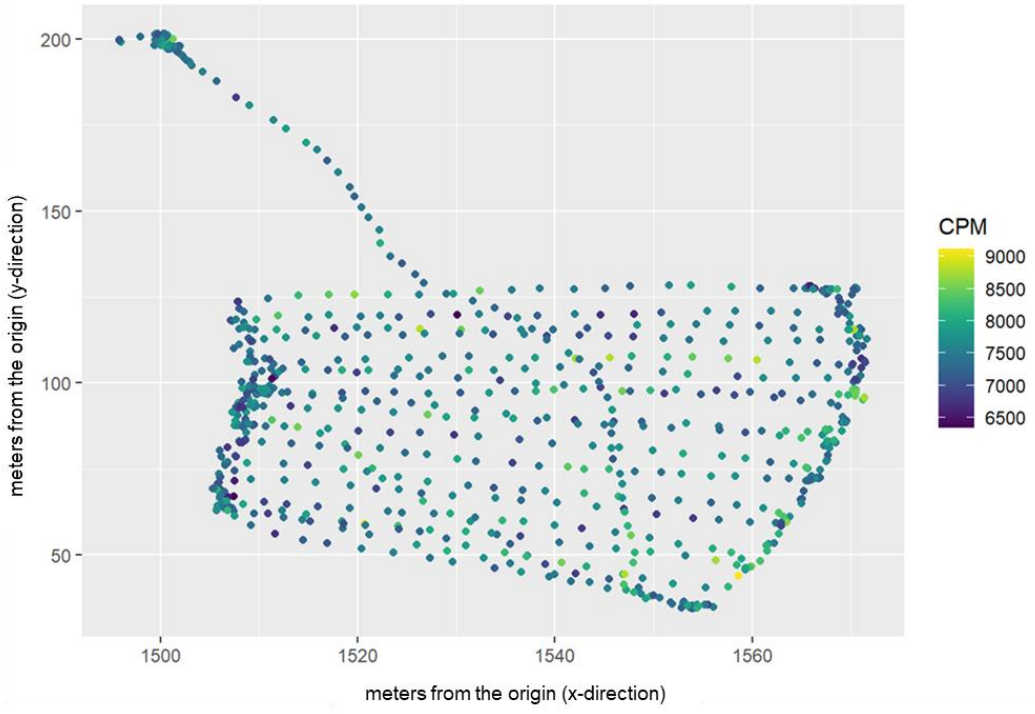

Figure A.14. Dataset 4 radiation measurements (in CPM) across subsite “J” reference area from Figure A.13.

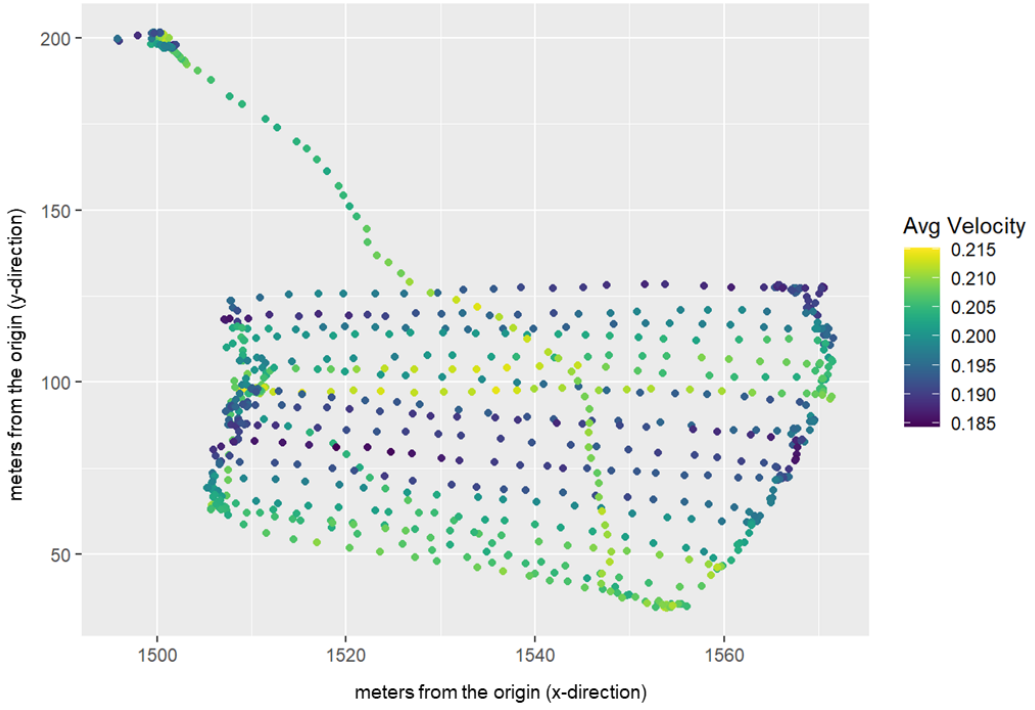

Figure A.15. Dataset 4 average velocity (in m/s) across reference area “J” from Figure A.13.

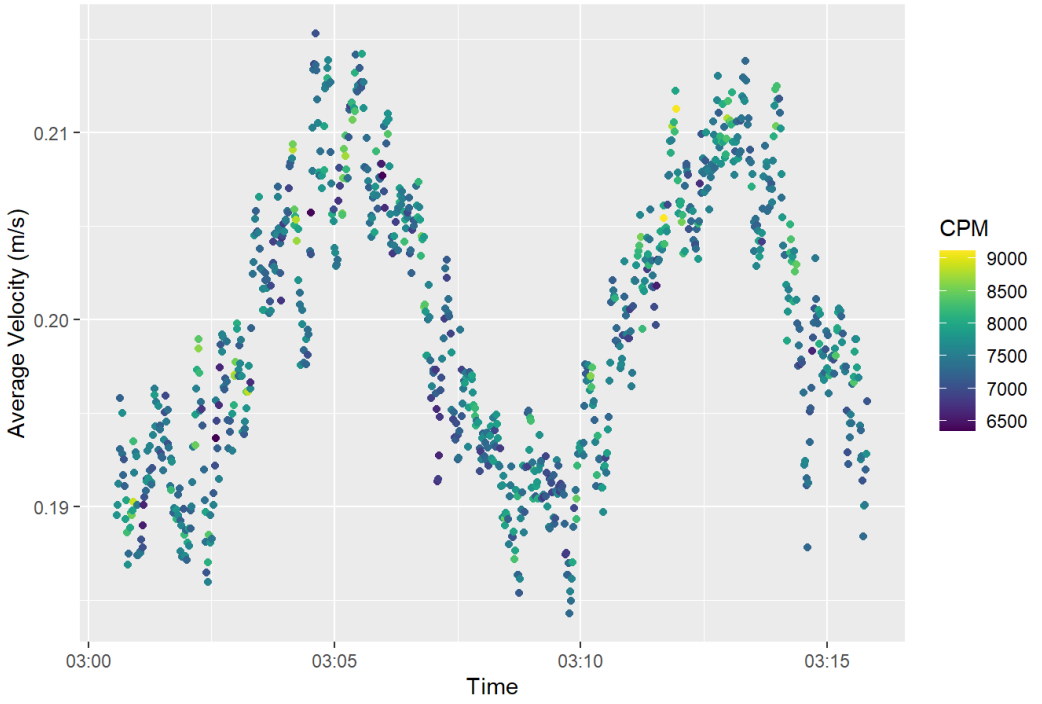

Figure A.16. Dataset 4 average velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for reference area “J” in Figure A.14.

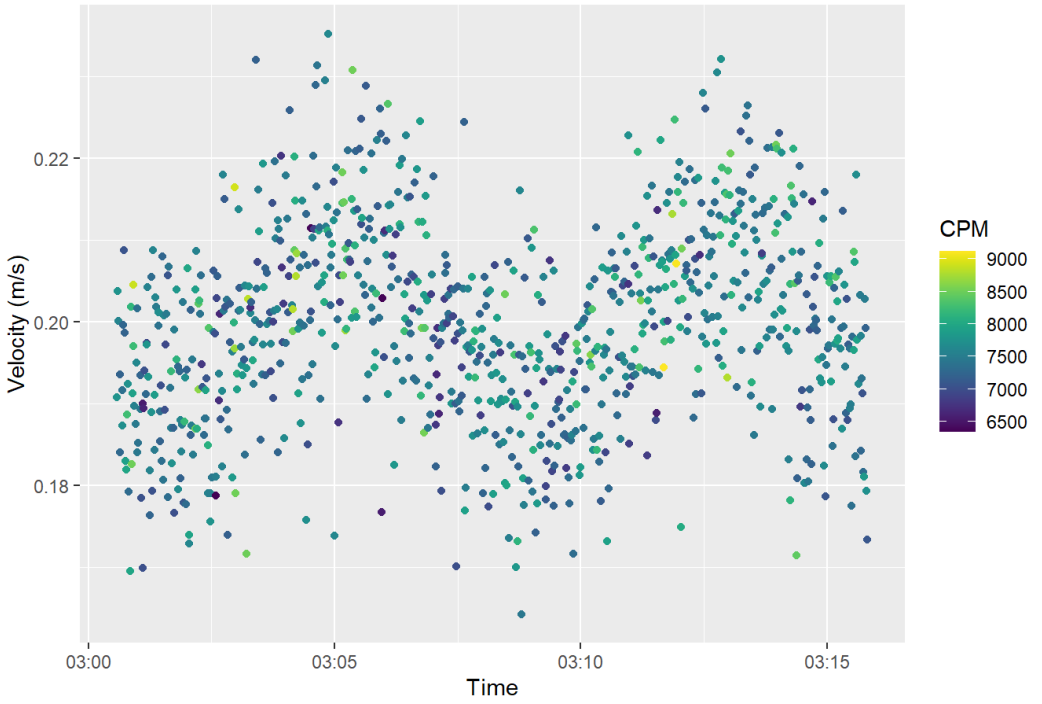

Figure A.17. Dataset 4 instantaneous velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for reference area “J” in Figure A.14.

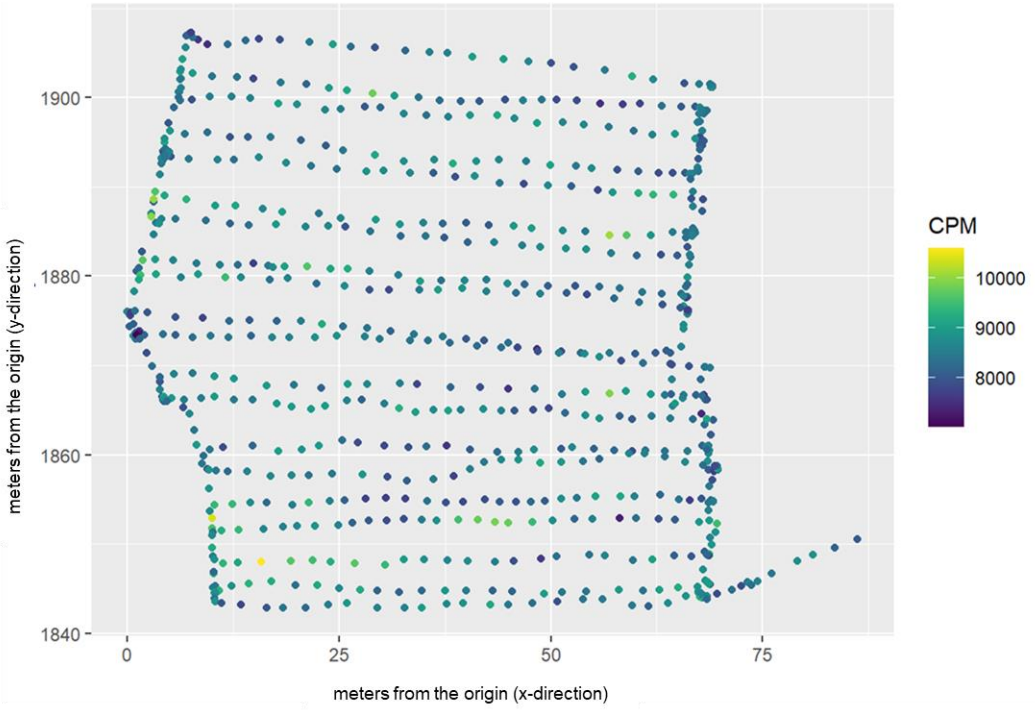

Figure A.18. Dataset 2 radiation measurements (in CPM) across area of concern “D”.

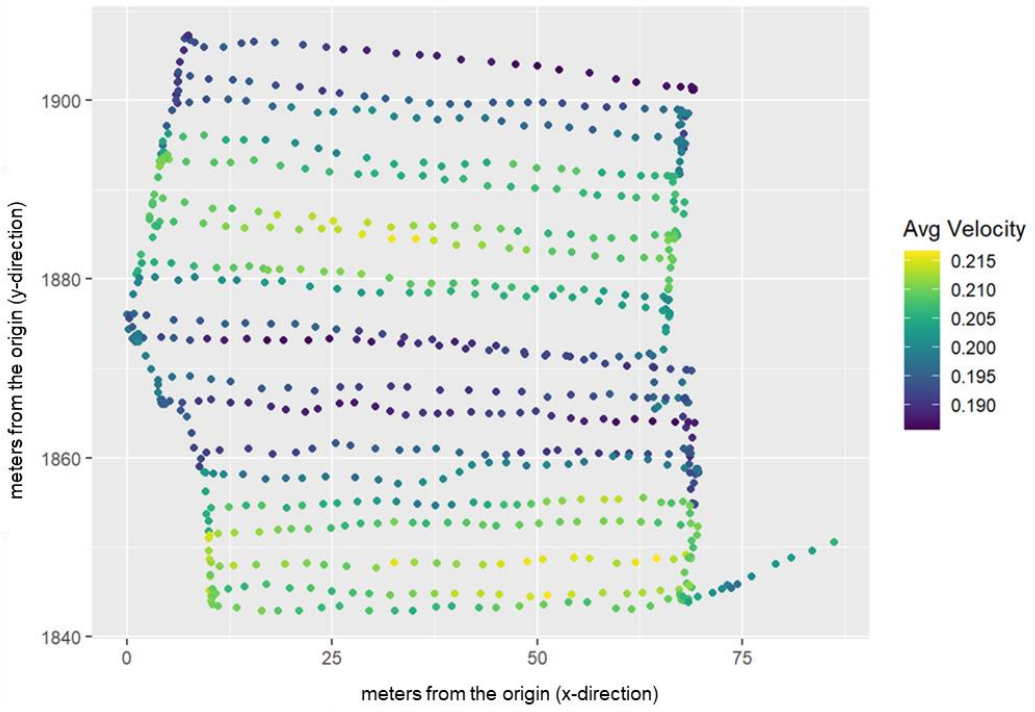

Figure A.19. Dataset 2 average velocity (in m/s) across area of concern “D” in Figure A.18.

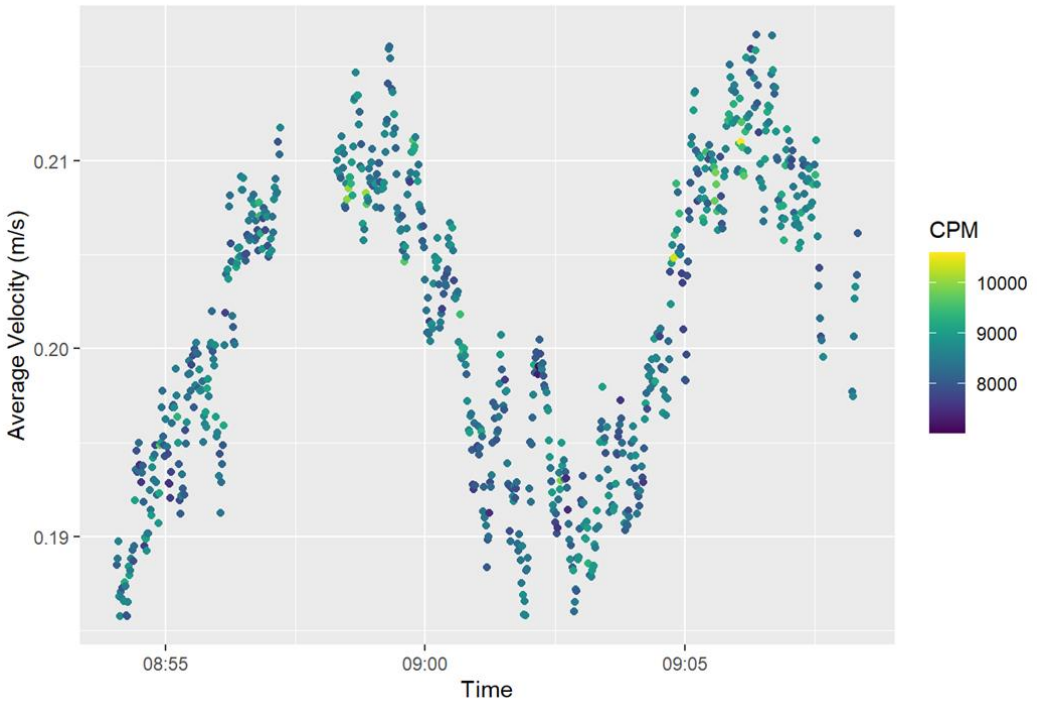

Figure A.20. Dataset 2 average velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for area of concern “D” in Figure A.18.

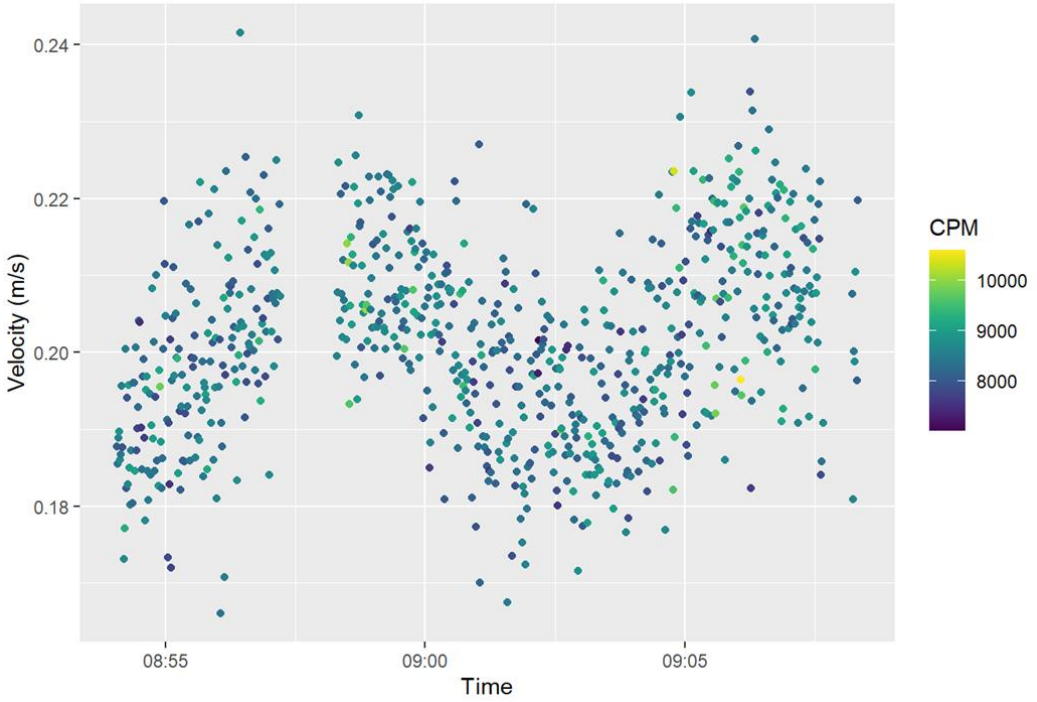

Figure A.21. Dataset 2 visualization of instantaneous velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for subsite “D” area of concern evaluated in Figure A.19.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354
1-888-375-PNNL (7665)

www.pnnl.gov* | *www.nrc.gov