PNNL-32664, Rev. 1

Recommendations for VSP Enhancements for Continuously Collected Survey Data

Fulfillment of Task 1a of TO 31310021F0022

July 2023

Deb Fagan
Moses Obiri
Zachary Weller
Lisa Newburn
Amoret Bunn
Jan Irvahn
Jen Huckett

Prepared for the U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Under Contract DE-AC05-76RL01830
Interagency Agreement: 31310019N0001
Task Order Number: 31310021F0022

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

PNNL-32664, Rev. 1

Recommendations for VSP Enhancements for Continuously Collected Survey Data

Fulfillment of Task 1a of TO 31310021F0022

July 2023

Deb Fagan
Moses Obiri
Zachary Weller
Lisa Newburn
Amoret Bunn
Jan Irvahn
Jen Huckett

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

Pacific Northwest National Laboratory
Richland, Washington 99354

Acronyms and Abbreviations

2D	two-dimensional
3D	three-dimensional
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
IL	investigation level
IL_{pp}	*a posteriori* investigation level
LISA	local indicator of spatial association
MARSSIM	Multi-Agency Radiation Survey and Site Investigation Manual
MDC	minimum detectable concentration
ML	machine learning
NL	The Netherlands
NRC	U.S. Nuclear Regulatory Commission
OLS	ordinary least squares
PNNL	Pacific Northwest National Laboratory
SADA	Spatial Analysis and Decision Assistance
UAV	unoccupied aerial vehicle
UAS	unoccupied aerial system
USL	upper simultaneous limit
UTL	upper tolerance limit
UXO	unexploded ordnance
VSP	Visual Sample Plan

Contents

Acronyms and Abbreviations iii
1.0 Introduction 1
2.0 The Effects of Spatial Correlation on Statistical Tests 3
2.2 Endogenous Correlation 4
2.3 The Effect of Spatial Autocorrelation 5
3.0 Methods to Improve VSP Analysis of Continuously Collected FSS Data 8
3.1 Generalized Least Squares Model 8
3.1.1 Model Form 8
3.1.2 GLS in MARSSIM Framework 9
3.2 Machine Learning 10
4.0 Methods to Identify Areas of Potential Residual Contamination 12
4.1 Potential Areas of Elevated Contamination (Hot Spots) 12
4.1.1 UTL Method 12
4.1.2 USL Method 16
4.1.3 Z-score Method 17
4.1.4 Local Indicator of Spatial Association Method 19
4.1.5 Lag-k Method 20
4.1.6 Summary 22
4.2 Boundaries of Areas to Revisit 23
4.2.1 VSP Kriging 24
4.2.2 Fixed Rank Kriging 26
4.2.3 Bayesian Ellipgrid and Markov Bayes 26
4.2.4 Searching for an Area of Elevated Contamination (Hotspot) via Transects 27
4.2.5 AI/ML Methods 27
5.0 VSP Data Import of Continuously Collected Data 29
6.0 VSP Data Visualization Considerations 30
7.0 Conclusions and Recommendations 31
7.1 Recommendations 31
7.2 Simulation Study for Lag-k 33
8.0 References 35
Appendix A – Data Sets for Demonstration of Methods for Continuously Collected Survey Data A.1

Figures

Figure 1. Example of a survey transect over a set of button sources (see Bunn et al. 2022). Autocorrelation is evident as smooth peaks. 4

Figure 2. Zinc concentration (ppm) recorded in a flood plain of the Meuse River near the village of Stein, NL (R Core Team 2020). 6

Figure 3. Type I error rates for spatial and non-spatial models. 7

Figure 4. Example data set 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). 13

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. 14

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. 14

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. 15

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. 15

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. 16

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. 17

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 18

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. 18

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. 20

Figure 14. 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. 25

Figure 15. 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. 25

Figure 16. Default VSP color sampling method (top image), improved clarity through use of sample symbols (middle image), and geostatistical analysis results (bottom image). 30

Tables

Table 1. Summary output from models highlighting the effect of spatial autocorrelation. 6

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 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 (§50.82) and release of the facility or site for unrestricted use (§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 (Abelquist et al. 2020); NUREG-1575, Revision 1, Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) (NRC 2000); and NUREG-1757, Consolidated Decommissioning Guidance (Banovac et al. 2006; Barr et al. 2020). 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 (NRC 2000; 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 was first published by increasingly using additional survey instrumentation and data capture tools, including global positioning system (GPS) and geographic information system (GIS) technologies (Abelquist et al. 2020). Scan surveys deviate from the MARSSIM 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.

This report identifies options for additional tools that could be needed to facilitate data importation and management, visualization, and analysis using continuous (without vigilance) methods for surface (2D) 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.
- Section 3: two methods for accounting for spatial correlation in 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: three methods for determining the boundaries of areas that may contain elevated residual contamination and must be revisited. Methods based on observations alone include upper tolerance limits (UTLs), upper simultaneous limits (USLs), and z-scores. 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 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 data set is imported.
- Section 6: visualization of continuously collected data. The amount of data from continuous surveys and the spatial area over which the data are collected would benefit from a few additional tools in VSP.

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. 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 Scenario B. 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 the draft drone report (Bunn et al. 2022) and recommends additional work to identify methods to incorporate into VSP enhancements for continuously collected survey data.

We have prepared two data sets to use as case studies and test data sets 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. They have been altered to better reflect the evaluation of a land area for the presence of small areas of elevated residual radioactivity. Appendix A discusses these data sets and their application to Scenarios A and B of MARSSIM.

2.0 The Effects of Spatial Correlation on Statistical Tests

The statistical decision-making techniques (hypothesis testing, calculating statistical intervals) in MARSSIM 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 data set. NUREG-1505 (Gogolak et al. 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:

$$Y = f(X) + \varepsilon \tag{1}$$

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 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 surveys conducted by continuous methods (with or without vigilance) 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 unoccupied aerial vehicle (UAV) surveys (Bunn et al. 2022). The experiment used button check sources placed at known locations along a transect and subsequent continuous survey with a NaI scintillation detector along a survey path over the transect. The result is a data set 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:

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

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-p, 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.

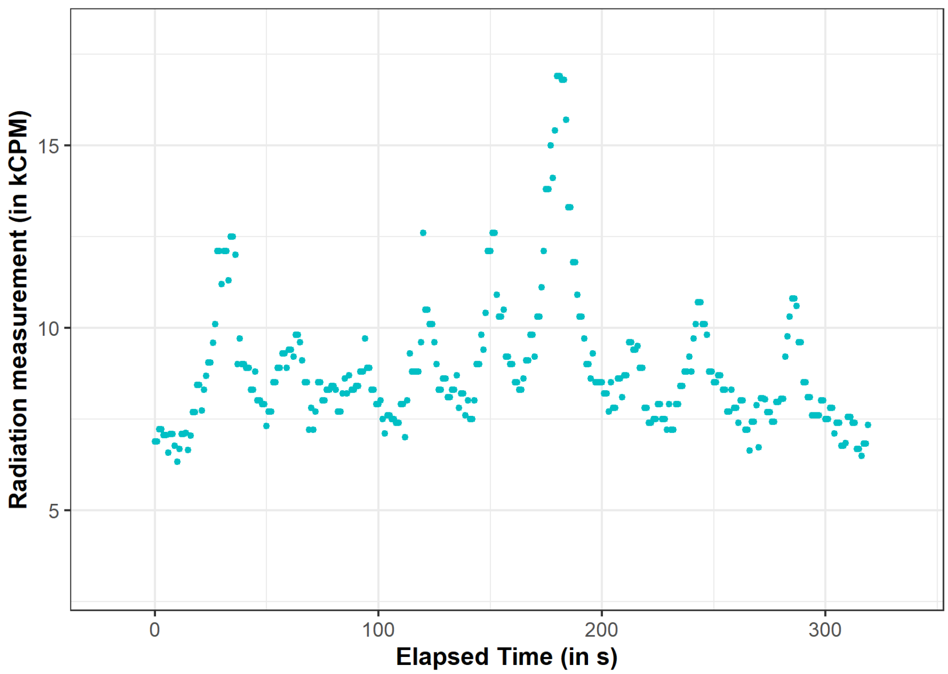


Figure 1. Example of a survey transect over a set of button sources (see Bunn et al. 2022). Autocorrelation is evident as smooth peaks.

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. The consequence 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 a couple 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 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:

$$H_o: \mu \geq \textit{action limit}$$

$$H_a: \mu < \textit{action limit}$$

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 (NL). The data are from the "meuse" data set and a description can be found in the R sp package documentation (R Core Team 2020; Pebesma and Bivand 2005; Bivand et al. 2013). 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:

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

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

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

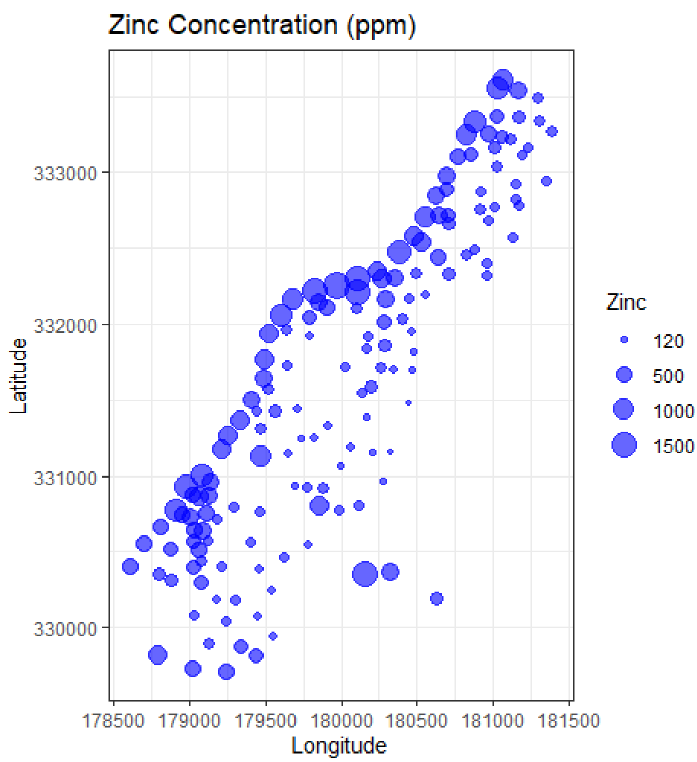


Figure 2. Zinc concentration (ppm) recorded in a flood plain of the Meuse River near the village of Stein, NL (R Core Team 2020).

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.

Further, spatial autocorrelation can also 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.

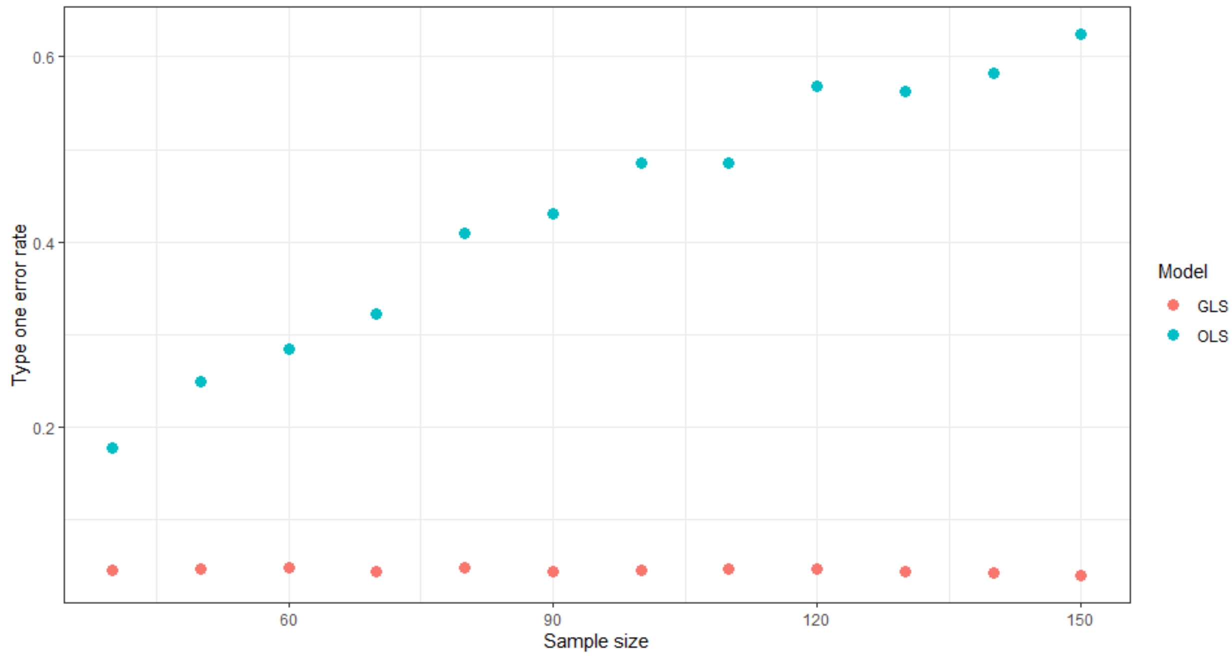


Figure 3. Type I error rates for spatial and non-spatial models.

3.0 Methods to Improve VSP Analysis of Continuously Collected FSS Data

In MARSSIM, 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, 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 hypothesis testing approach. 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:

$$Y = X\beta + \varepsilon \tag{3}$$

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 framework and included in VSP.

3.1.2 GLS in MARSSIM Framework

In the majority of MARSSIM 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).

Case 2: Two-Sample Case

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 linear model in Equation (3). Where $Y = [Y_1^*, \ldots, Y_N^*, Y_{N+1}, \ldots, Y_{N+M}]^T$ 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 $X = \begin{bmatrix} 1_N \\ 0_M \end{bmatrix}$ 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 site, 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 data sets, 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, continuous 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 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 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 p indicates the IL is developed *a posteriori* using post-processed survey data to identify areas of potential concern.

After data quality assessment has been performed to verify that the 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 implementing it is provided using an example data set.

Section 4.1 discusses methods to identify potential areas of elevated contamination (hot spots) when surveys are completed using continuous data collection and are0020performed without surveyor vigilance.

Section 4.2 discusses methods to delineate areas to revisit, including a rule-based approach that is consistent with MARSSIM as well as geospatial methods.

4.1 Potential Areas of Elevated Contamination (Hot Spots)

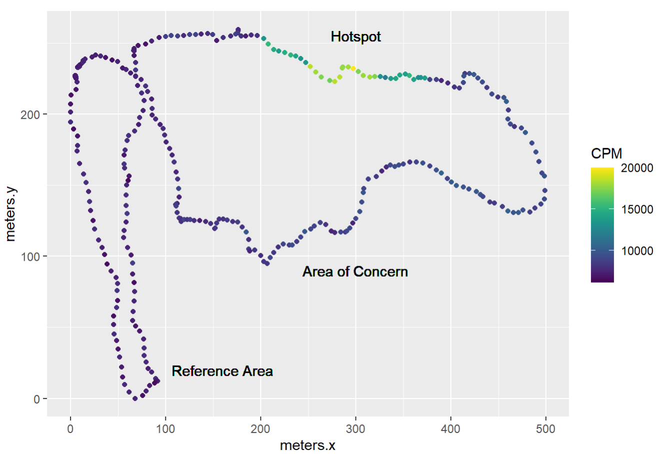
4.1.1 UTL Method

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-side DCGL (DCGLW) because it can be applied when non-uniform contamination is present. A UTL is used to estimate the pth percentile of a population with a chosen level of confidence. For example, the IL_{pp} can be estimated using the 95^{th} 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, log-normal, 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 log-normal 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, and so are 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 of implementing it using an example data set. The data set was derived based on data previously provided by NRC for lag-k scan MDC research (Fortin et al. 2023), 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 data set 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).



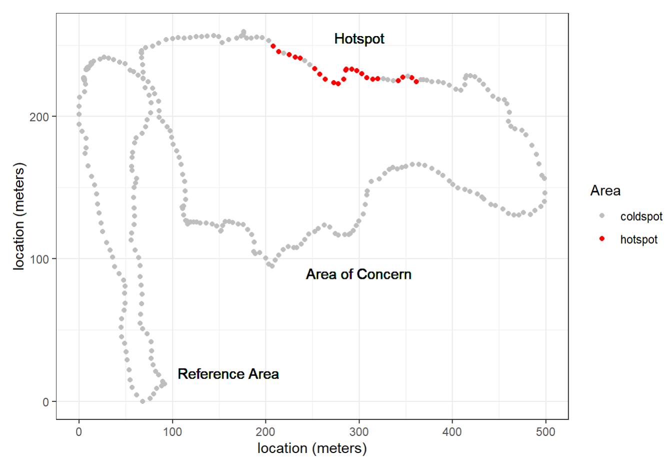


Figure 4. Example data set where the left panel shows observed CPM and the right panel indicates observations that are hot spots (>14,000 CPM points are red) and cold spots (<14,000 CPM points are grey).

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 then 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.

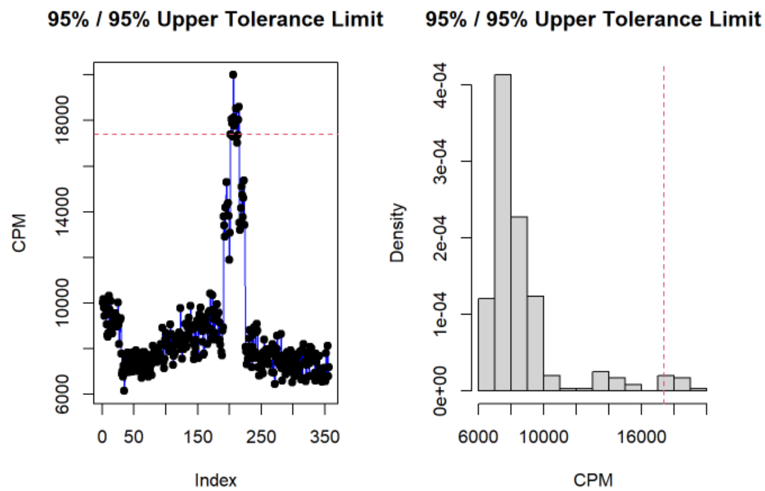


Figure 5. Example UTL and observations. The left panel shows CPM observations compared to the 95th percentile UTL and the right panel shows the distribution of CPM observations.

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.

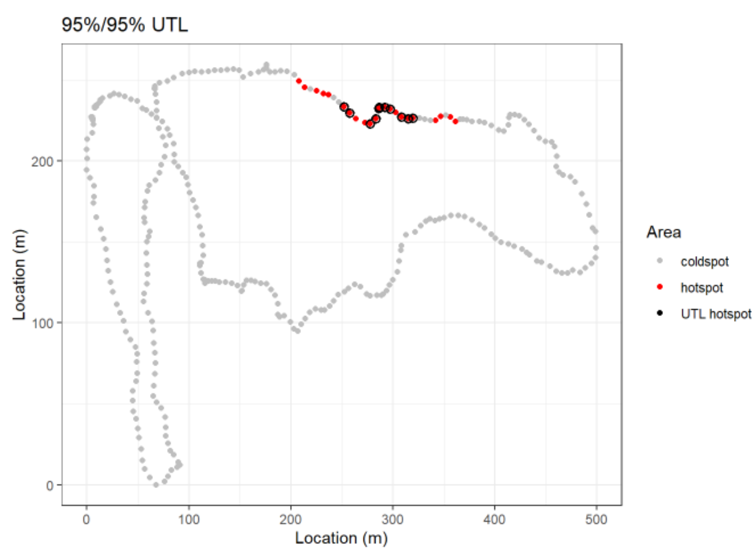


Figure 6. Example UTL results with grey dots correctly identified as cold spots, outlined red dots correctly identified as hot spots, and red dots without an outline incorrectly identified as cold spots.

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

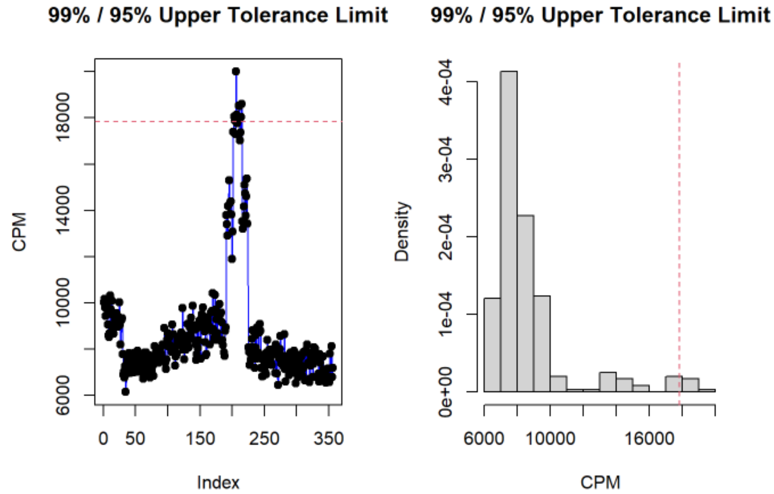


Figure 7. Example UTL diagnostic plots. The left panel shows the CPM observations compared to the 99th percentile threshold and the right panel shows the distribution of CPM observations.

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.

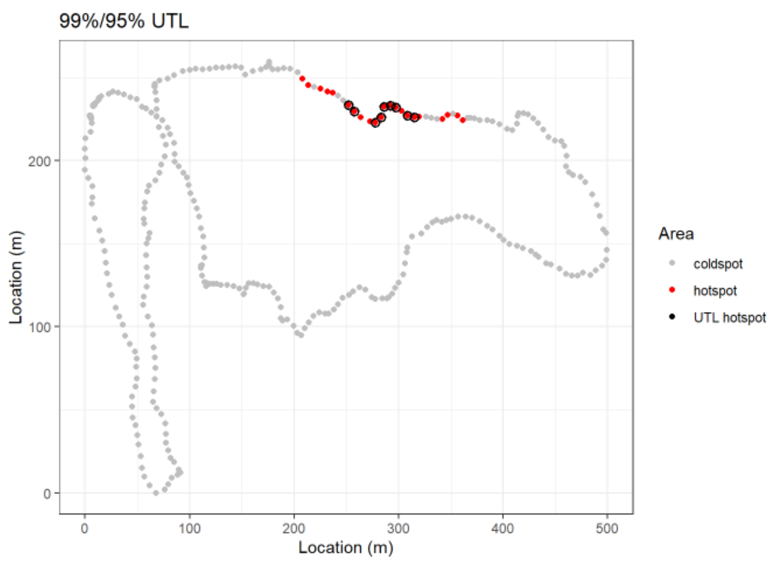


Figure 8. Example UTL results with grey dots correctly identified as cold spots, outlined red dots correctly identified as hot spots, and red dots without an outline incorrectly identified as cold spots.

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 data set 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 at 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.

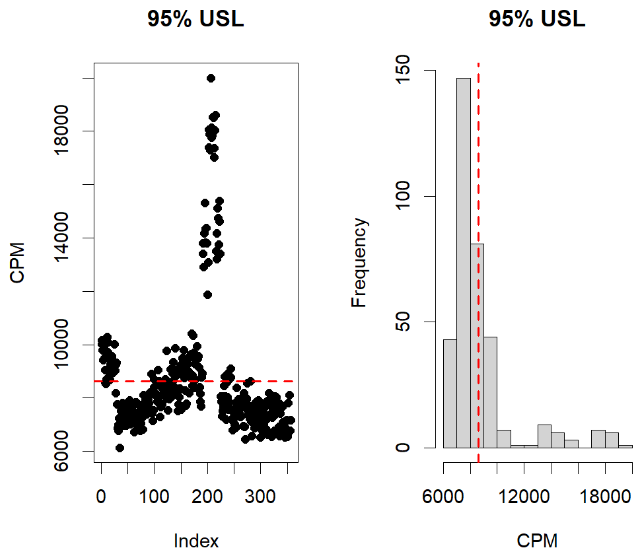


Figure 9. Example comparing IL_{pp} based on the USL method to observations from the reference area and the area of concern. The right panel shows the distribution of CPM observations.

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.

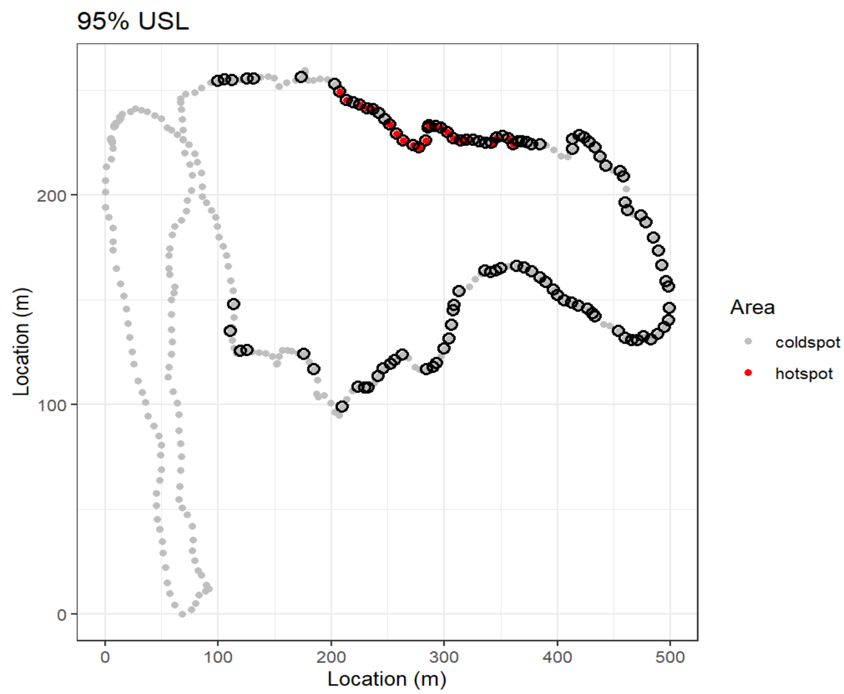


Figure 10. Example USL results with grey dots correctly identified as cold spots, outlined red dots correctly identified as hot spots, and outlined grey dots incorrectly identified as hot spots.

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 log-normal 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 data set 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 90^{th} percentile of the standard normal and the other based on the 95^{th} percentile.

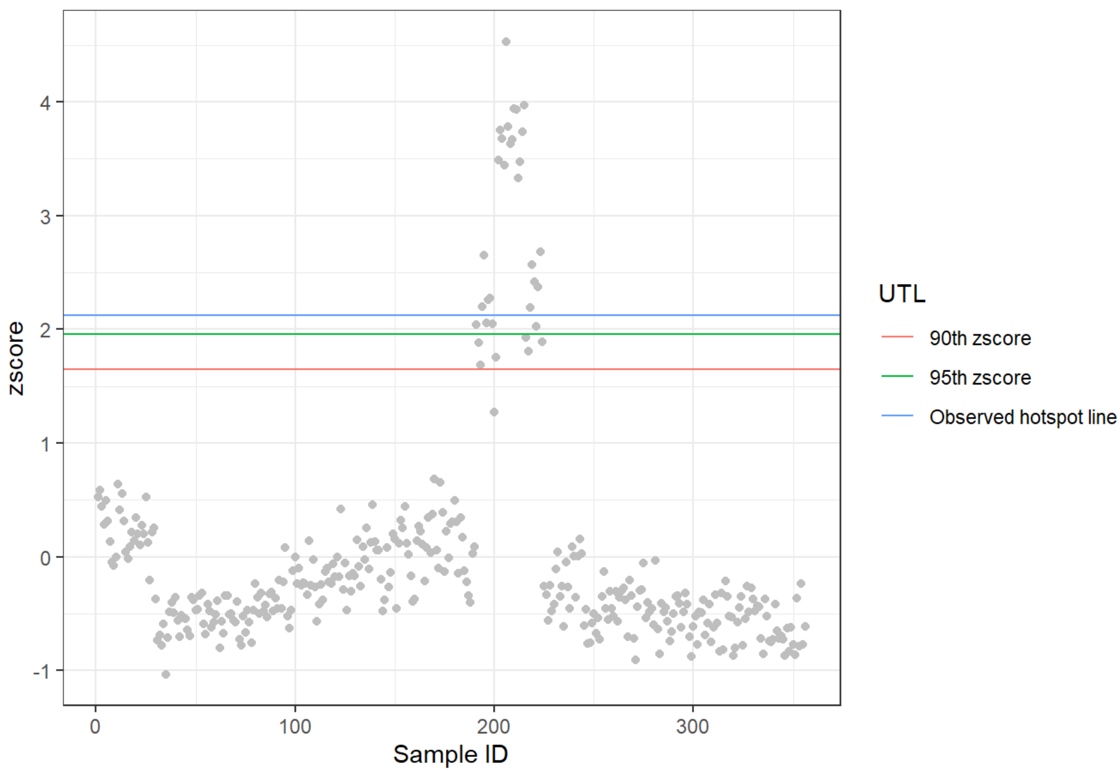


Figure 11. Example z-score method. The blue line is the observed hot spot threshold value, and the green and red lines are the 95th z-score and 90th percentile z-score from a standard normal distribution

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.

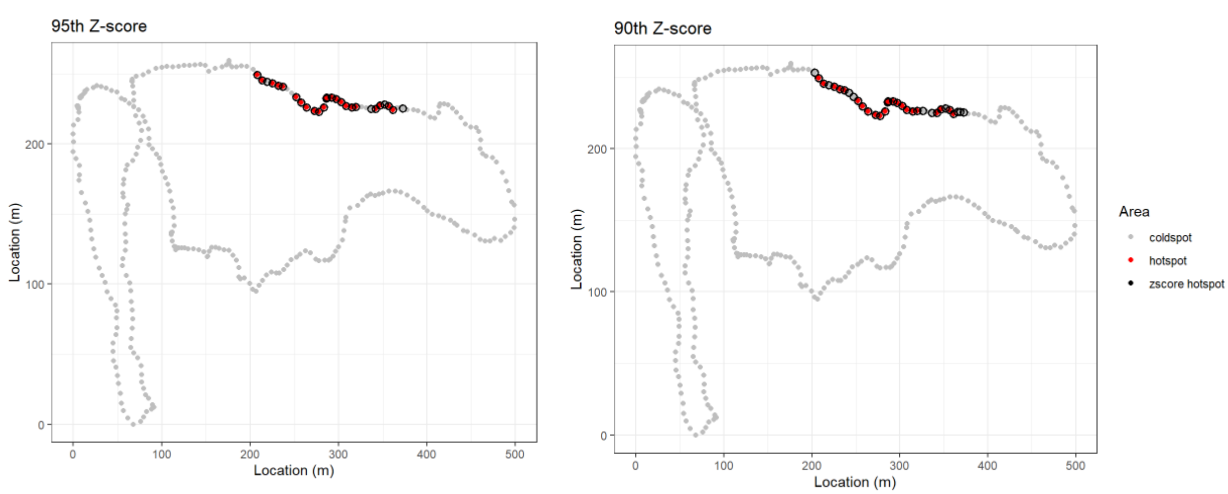


Figure 12. Example z-score results with grey dots correctly identified as cold spots, outlined grey dots incorrectly identified as hot spots, and outlined red dots correctly identified as hot spots.

4.1.4 Local Indicator of Spatial Association Method

A local indicator of spatial association (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. There are four types of clusters with statistical significance:

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 UAV or human-based continuous data collection. By incorporating altitude, GPS, and velocity data into the analysis, it could help to 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 UAVs and humans are in compliance 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 to 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 data set 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.

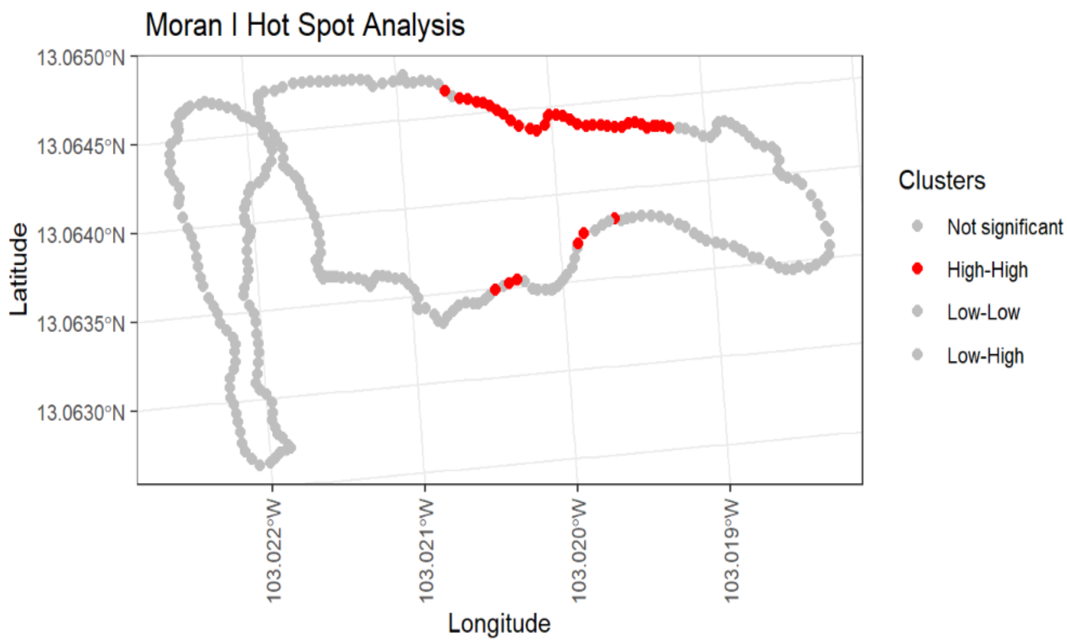


Figure 13. Example LISA method results with grey dots correctly identified as cold spots, and red dots as identified hot spots comprising 23 correctly identified hot spots and 15 cold spots incorrectly identified as hot spots.

4.1.5 Lag-k Method

Increased continuous data collection using automated data loggers and autonomous radiological survey 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 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 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 with 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 (hotspots). The lag-k method offers a local differencing approach. It computes net counts based on spatially localized average observations rather than a single, overall background average. The distance between spatially localized neighbors is determined by the site conceptual model, expected hotspot 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

[1] It is known that there are differences between binned or integrated audio click data and rate meter 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 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. 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 (within acceptable statistical error probabilities) by the surveyor during the scanning process that meets regulatory requirements. The lag-k method can also be used after the 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 hotspots.

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 hotspots 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 since the nature of its comparisons between differences allows the method to 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 also 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 hotspot 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 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 Summary

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

- They are readily available in statistical and data analysis 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 ILpp 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, on the other hand, 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 UAV or human-based continuous data surveys.
- The lag-k method provides an approach to determining *a priori* scan MDC values when residual contamination hotspots 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.

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 ILpp would be exceeded in the absence of residual radioactivity (see next bullet).
- When comparing site data to a USL or UTL, any observation that is exceeds the ILpp is expected to originate from an area of residual radioactivity. However, when observations are compared to the ILpp 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 ILpp 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 hotspot size can be used for determining a meaningful physical basis for the lag size.
- Of these methods, 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.

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

Section 4.3.2 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 continuous scanning surveys. FRK is also flexible in that it allows for multiple resolutions of data, so that, for instance, ground and aerial surveys 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 4.3.3 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 4.3.4 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 4.3.5 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 2D versions of these algorithms, but the extension to 3D 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 14. 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 15 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.

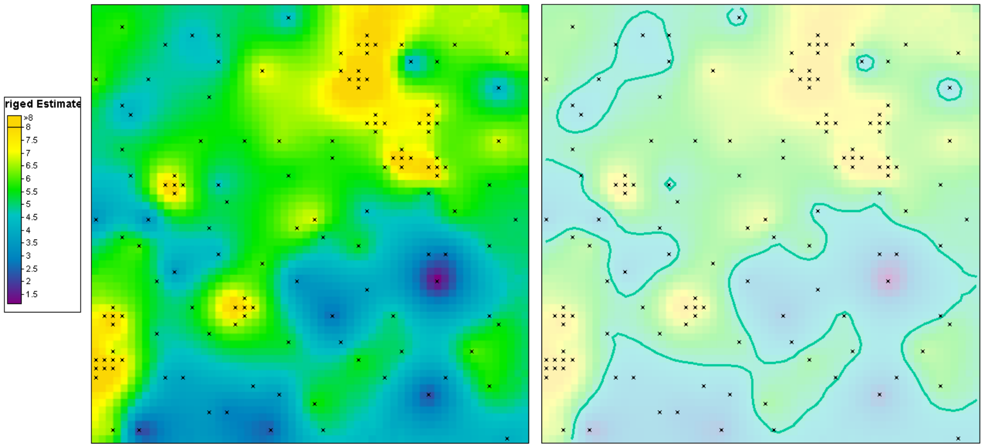


Figure 14. Example kriged surface output from VSP and associated contour. A black ‘x’ marks the sample locations. The left panel is the kriged surface; the right panel is a contour at the threshold level of 4.75.

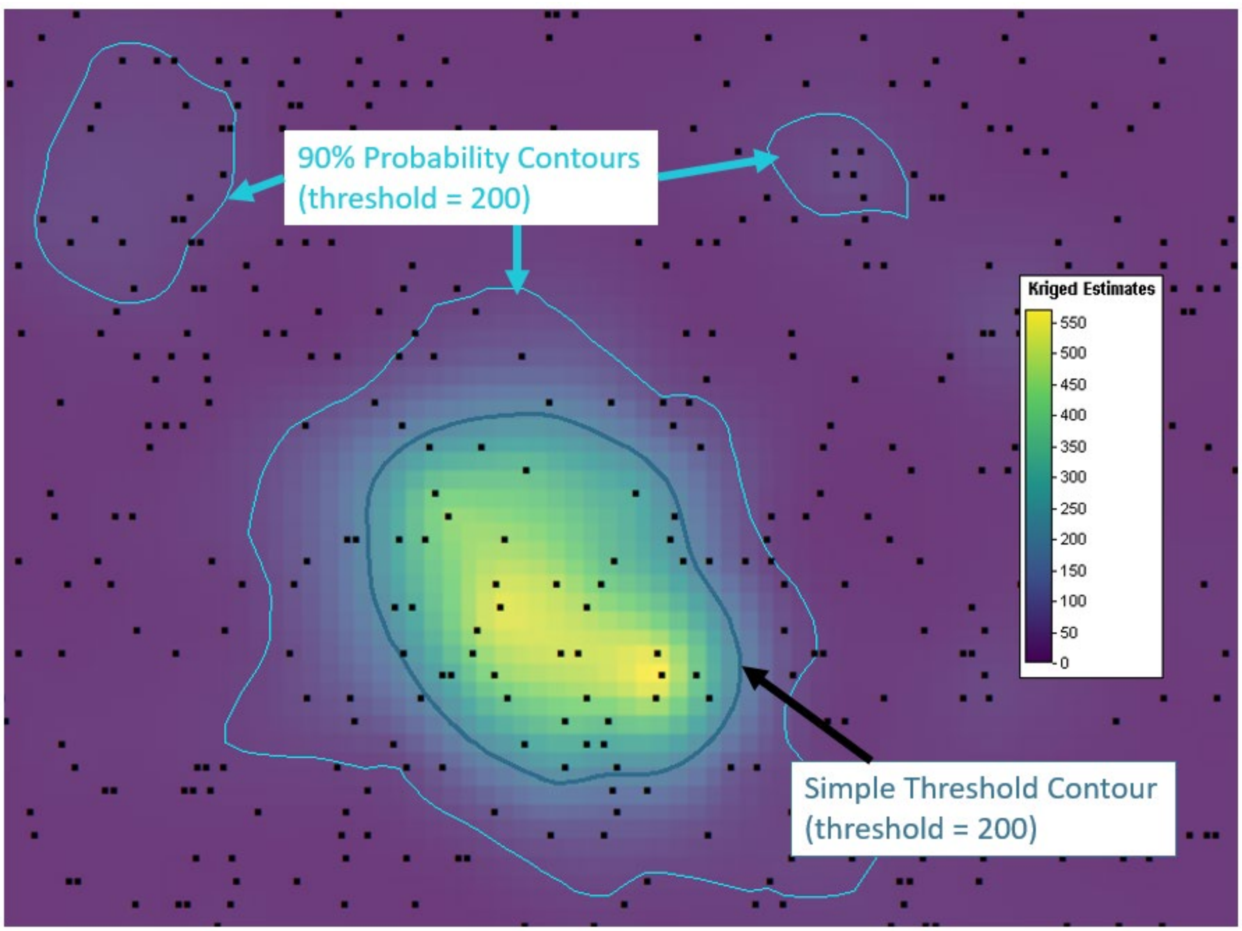


Figure 15. Example VSP output of a kriged region demonstrating how delineated probability contours better account for uncertainty in the spatial estimate and identify regions where there is less confidence of being below a specified threshold.

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., data sets 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 Elipgrid 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 Elipgrid 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 Elipgrid 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 Elipgrid is an extension of the (non-Bayesian) Elipgrid algorithm used to create sampling designs and is intended to detect regions of elevated contamination (SC&A 2022). Unlike Elipgrid, Bayesian Elipgrid 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 Elipgrid and Bayesian Elipgrid 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 Elipgrid 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 Elipgrid 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) Elipgrid algorithm but does not include Bayesian Elipgrid 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 (Hotspot) 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 on-the-ground or unoccupied aerial system (UAS) radiation surveys as well, 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 data set 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 data set. "Adequate" models can then be used to make predictions for future cases or for locations that are not in the training or test data sets. 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 data sets – 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. By applying deep learning to integrated multimodal sensing data, the performance of ML approaches is being defined, through training with large and small data sets, 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.

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 data sets much larger than that and may be desirable for usability improvements if data sets 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 is to 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 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 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) or survey execution that renders the data unreliable (e.g., driving too fast, 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 survey data presents a number of challenges and requires additional considerations compared to visualization of more sparsely collected data. Survey values 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 16, 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.

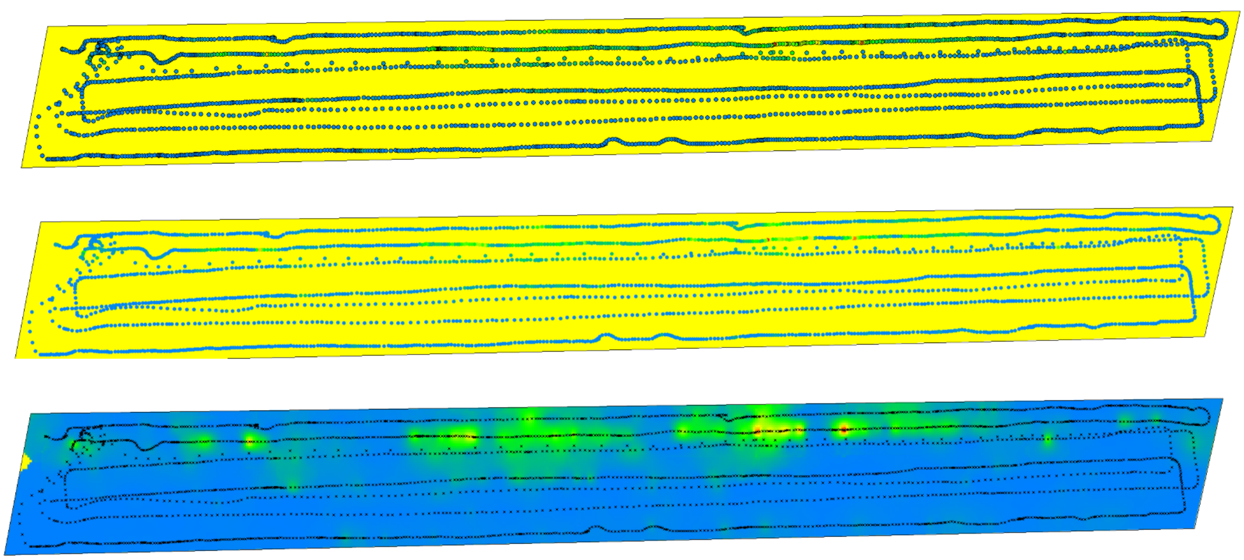


Figure 16. Default VSP color sampling method (top image), improved clarity through use of sample symbols (middle image), and geostatistical analysis results (bottom image).

Many types of sensors have spatial extent or spatial representativeness, such as a detector with a particular field of view or a 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 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 its capability and make it more useful. This report summarizes several methods that cover MARSSIM-type hypothesis testing, identifying boundaries for areas that need to be revisited, and data import and visualization. Additional review may be necessary to determine whether some of the newer methods available in the literature are consistent with the MARSSIM approach. Recommendations for added VSP capabilities are summarized below.

7.1 Recommendations

- Methods that account for spatial correlation, both endogenous and exogenous, should be implemented prior to hypothesis testing, as spatial correlation can result in reduced statistical power to verify a site meets release criteria. Data quality assessment is vital to ensure that the data are adequate to support a decision.
- Once spatial correlation is accounted for, one- or two-sample tests are appropriate. GLS is a method that accounts for spatial correlation as well as a straightforward comparison of a site mean to either an action limit or a reference area mean. ML methods are not appropriate in the hypothesis testing framework.
- Several methods to determine elevated values and subsequent boundaries of areas to revisit are already in VSP. Improvements could be made by allowing alternative formulations of UTLs and adding a USL method. Geospatial methods such as kriging are best used to determine boundaries of potential areas of elevated residual activity. Improvements to VSP geospatial capability that incorporate Bayesian and/or conditional probability methods would make the software more valuable by allowing users to incorporate prior knowledge into the determination of boundaries. Guidance for how to select prior distributions and verify method assumptions would need to be formulated. A thoughtful approach to that considers the ability of VSP users to identify needed inputs is a consideration when deciding whether to make these methods available.
- Pre-analysis data processing generally would need to occur outside of VSP, as multiple tools, such as spreadsheet software and database tools, are flexible, available, and widely used. VSP would benefit from the capability to do unit conversions and allow for data from disparate sensor platforms.
- Data visualization capability recommendations are mostly around the development of plots that make potential elevated areas easier to identify on maps. Improvements in the user interface to identify instrument fields of view would be useful for some kriging methods that allow data from multiple sensor platforms and/or sensors. Visualization that accommodates multiple analytes and/or multiple sample matrices using different symbols or color scales would also support practitioners in many site applications where more than one contaminant of concern may be present.

Fortin et al. (2023) recommended several areas for future research to advance the state of the art for scan MDC and post-processing of continuously collected data, including the following.

- The Science Advisory Board (SAB) agrees that current MARSSIM guidance does not adequately address modern 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. In addition to the literature reviewed throughout this report and Fortin et al. (2023), the following literature should be reviewed in the context of continuous data collection and statistical analysis.

  - Quantitative measurements with various example systems described in the scientific literature: Marques et al. 2021; Peeva, 2021; Ji et al. 2020; Rahman et al. 2020; Ji et al. 2019; Lee and Kim, 2019; Sanada et al. 2019; Azami et al. 2018; Falciglia et al. 2018; Wilhelm et al. 2017; Sinclair et al. 2016; Sanada and Torii, 2015; Kock et al. 2014; Sanderson, 2013; Tanigaki et al. 2013; Kock and Samuelsson, 2011.
  - Detection efficiency and minimum detectable activity for mobile scanning speed and signal processing: Falkner and Marianno, 2021; Marianno, 2015.
  - The SAB 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.

- Expand the library of available site and reference area datasets on which to demonstrate scan MDC and hypothesis testing (e.g., lag-k) methods.
- Evaluate performance of these methods compared to the traditional with-vigilance MDC calculation method.
- Consider the impact that omissions of “edge” or “fringe” locations would have when points in the scan would not have sufficient data on either side to calculate the lag-k background (e.g., in the lag-53 case), and whether (and how much) buffer should be added to mitigate such effects.
- Study the impacts of surveyor scan parameters (e.g., using sleds mounted both inside and outside) with various gross-counting instruments, sources, and source scenarios. Vary speed, distance, type of source (distributed vs. point source), and shielding to understand impacts of these variables on without vigilance survey results.
- Survey other existing signal detection frameworks for scan data (Brogan and Brandl 2019) and investigate applications to a priori MDCR calculation.
- Further investigate methods in Alecksen and Whicker (2016), Alecksen and Whicker (2023), and additional related resources to determine how MCNPX code could be used to convert a calculated MDCR to a scan MDC.
- Use simulation and field studies to evaluate the ability of hypothesis testing methods (e.g., lag-k) to detect elevated areas in data collected via without-vigilance surveys for various instrument configurations and radiological sources, concentrations, spatial contamination areas, and distributions.
  - As a result of the above studies, develop recommendations on the following:
    - Approaches to implementation, including software tools and needs for licensees and permit holders to implement the lag-k method.
    - Methods to select an optimal k value for the lag-k method.
    - 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.
  - Identify the limitations of the lag-k method by outlining site conceptual models for which this approach is/is not suitable.
  - Extend lag-k 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 method.

One of the key concerns with producing a scan MDC value for a without-vigilance survey is the difference between the audio click data stream and the logged ratemeter display data stream. NUREG-1507 warns against calculating a scan MDC if logged ratemeter observations will be used because there are concerns the ratemeter will not reach full scale. However, some technology could facilitate simultaneous and synchronized data collection (e.g., Aleckson and Whicker describe using scalar counting mode output to calculate scan MDC). Future work should review the data recommended in Alecksen and Whicker (2023) in greater detail to determine if it provides a viable solution; catalogue and review available technology; and address discrepancies between data streams, potentially building a model to translate the ratemeter display data stream to a binned audio click data stream. This work would require taking field data and pairing logged ratemeter counts with the true audio data stream (possibly via an audio recording with time stamps). After collection, the data would need to be processed, producing paired counts of the two data streams. The paired data streams would then be analyzed to identify and quantify biases between the data capture techniques. Further mathematical modeling work would then be required to determine whether a scan MDC could be calculated for without vigilance surveys that intend to use ratemeter display data.

7.2 Simulation Study for Lag-k

Considerable work needs to be done to characterize the conditions under which the lag-k method performs well and to determine when it performs poorly. The key metrics to evaluating performance include the true positive, true negative, false positive, and false negative counts and rates for lag-k hypothesis tests (often summarized using a confusion table, described in Section 5.2). Determining these metrics requires knowing the ground truth. This requirement leads us to make a recommendation that future work includes an extensive simulation study, in addition to field tests with known sources.

Simulation studies can cover many topics of interest including variation in the data generation process, variation in background estimation, variation in analysis methods, and even variation in the metrics of evaluation. The data generation process encompasses both background and source distributions. The background distributions could vary spatially and may extend beyond Poisson. The source distributions could range between a single point source to many point sources to elevated regions of varying shape and size.

The estimation of background is always a critical component in detection methodologies. Different background estimation approaches could be included in the simulation study, enabling the comparison of a variety of analysis methods, including lag-k, Currie's 1968 single paired measurement approach, and the signal detection theory approach given in NUREG-1507. The lag-k approach itself can be expanded to encompass a range of k values or weighted combinations of k values. The lag-k method could be extended to a two-dimensional distance metric expanding beyond the transect-based approach. Backgrounds could be estimated from the entire survey region, from subsets of the survey region, from reference areas that share the same background distribution as the simulated site of interest, or from reference areas that have different background distributions from the simulated site of interest.

Current testing methodology is built around individual location-based hypothesis tests with no multiple testing correction, so the natural scale for method evaluation is on the same individual location-based scale. However, some consideration should be given to a coarser grid for counting true positives and false negatives. A simulation study could provide helpful guidance concerning when a surveyor can expect to find sources of interest and when analysis methods are likely to fail.

The data from these and/or similar simulation studies could be used to investigate and evaluate the other continuous data methods considered in this report for the purposes of determining sensitivities to site conditions, parameter specifications, and for validation and verification once implemented in VSP or other software.

8.0 References

10 CFR Part 20, Subpart E. *Code of Federal Regulations*, Title 10, Energy, Part 20, "Standards for Protection Against Radiation," Subpart E, *Radiological Criteria for License Termination. U.S. Nuclear Regulatory Commission*.

10 CFR Part 50. *Code of Federal Regulations*, Title 10, Energy, Part 50, "Domestic Licensing of Production and Utilization Facilities."

Abelquist, EW. 2014. *Decommissioning Health Physics, A Handbook for MARSSIM Users*, 2nd edition. CRC Press, Boca Raton, FL.

Abelquist, EW, JP Clements, AM Huffert, DA King, TJ Vitkus, and BA Watson. 2020. *Minimum Detectable Concentrations with Typical Radiation Survey for Instruments for Various Contaminant and Field Conditions*. NUREG-1507, Revision 1. U.S. Nuclear Regulatory Commission, Washington, D.C.

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

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

Ancell, E. & Bean, B. (2021). *Autocart—spatially-aware regression trees for ecological and spatial modeling.* Accessed July 15, 2022. arXiv preprint arXiv:2101.08258

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

Azami, K., Ootagaki, T., Ishida, M. and Sanada, Y. 2018. “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.

Banovac, KL, JT Buckley, RL Johnson, GM McCann, JD Parrott, DW Schmidt, JC Shepherd, TB Smith, PA Sobel, BA Watson, DA Widmayer, and TH Youngblood. 2006. *Consolidated Decommissioning Guidance Decommissioning Process for Materials Licensees. Final Report*. NUREG-1757, Volume 1, Revision 2. U.S. Nuclear Regulatory Commission, Washington, D.C.

Barr, CS, S Clark, GC Chapman, JP Clements, DW Esh, RW Fedors, AM Huffert, LA Kauffman, MM LaFranzo, DB Lowman, CA McKenney, LL Parks, DW Schmidt, AL Schwartzman, and BA Watson. 2020. *Consolidated Decommissioning Guidance Characterization, Survey, and Determination of Radiological Criteria. Draft Report for Comment*. NUREG-1757, Volume 2, Revision 2. U.S. Nuclear Regulatory Commission, Washington, D.C.

Beale, CM, JJ Lennon, JM Yearsley, MJ Brewer, and DA Elston. 2010. Regression analysis of spatial data. *Ecology Letters*, 13(2):246–264. https://doi.org/10.1111/j.1461-0248.2009.01422.x

Benito, M.B. (2021). spatialRF: Easy spatial regression with random forest. *R package version* 1, No. 0. https://cran.r-project.org/web/packages/spatialRF/spatialRF.pdf

Bivand, RS, E Pebesma, and V Gomez-Rubio. 2013. *Applied Spatial Data Analysis with R*, 2nd edition. Springer, NY. https://asdar-book.org/

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

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

Bunn, A, K Wagner, DK Fagan, H Gadey, T Ikenberry, K Markham, and MY Obiri. 2022. *Draft Drones for Decommissioning*. PNNL-32519. Pacific Northwest National Laboratory, Richland, WA.

Casella, G, and RL Berger. 1990. *Statistical Inference*. Wadsworth & Brooks/Cole, Pacific Grove, CA.

Clifford, P, S Richardson, and D Hemon. 1989. Assessing the Significance of the Correlation between Two Spatial Processes. *Biometrics*, 45(1):125-134. https://www.jstor.org/stable/2532039

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

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

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

Davidson, JR. 1995. *ELIPGRID-PC: Upgraded Version*. ORNL/TM-13103. Oak Ridge National Laboratory, Oak Ridge, TN.

Davis, CB, and PF Wambach. 2015. "Quasi Nonparametric" Upper Tolerance Limits for Occupational Exposure Evaluations. *Journal of Occupational and Environmental Hygiene*, 12: 342-349. DOI:10.1080/15459624.2014.995301.

Deutsch, CV, and AG Journel. 1998. *GSLIB Geostatistical Software Library and User's Guide*, 2nd Edition, Applied Geostatistics Series, Oxford University Press, Inc. New York, NY.

Dray, S, P Legendre, and PR Peres-Neto. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 196(3-4): 483-493. https://doi.org/10.1016/j.ecolmodel.2006.02.015

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

Falciglia, P., L. Biondi, R. Catalano, G. Immè, S. Romano, and F. Vagliasindi. 2018. "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.

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

French, JP, and JA Hoeting. 2015. Credible regions for exceedance sets of geostatistical data. *Environmetrics*, 27(1). DOI:10.1002/env.2317.

Fortin, D, Irvahn, J, LL Newburn, DK Fagan, and JC Huckett. 2023. *Overview of a Methodology to Calculating the Scanning Minimum Detectable Concentration for Post-Processed Radiological Surveys*. PNNL-34211, Rev.1. Pacific Northwest National Laboratory, Richland, WA.

Georganos, S., Grippa, T., Gadiaga, A.N., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., & Kalogirou, S. (2021). 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.

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

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

Gilbert, RO, JE Wilson, RF O'Brien, DK Carlson, BA Pulsipher, and DJ Bates. 2003. *Version 2.0 Visual Sample Plan (VSP): UXO Module Code Description and Verification*. PNNL-14267. Pacific Northwest National Laboratory, Richland, WA.

Gogolak, CV, GE Powers, and AM Huffert. 1998. *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. U.S. Nuclear Regulatory Commission, Washington, D.C.

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

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

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

Hahn, GJ, and WQ Meeker. 1991. *Statistical Intervals: A Guide for Practitioners*. John Wiley & Sons, Inc. DOI:10.1002/9780470316771.

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

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

Meeker, WQ and GJ Hahn. (2011). Statistical Intervals: A Guide for Practitioners. Germany. Wiley.

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 & A. Zammit-Mangion. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, 24, 398-425.

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

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

Ji, Y.Y., Lim, T., Choi, H.Y., Chung, K.H. and Kang, M.J. 2019. “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.

Ji, Y.Y., Lim, T., Hitomi, K. and Yajima, T. 2020. “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.

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

Kock, P., Rääf, C. and Samuelsson, C. 2014. “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.

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

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

Meyer, H, C Reudenbach, S Wöllauer, and T Nauss. 2019. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. Ecological Modelling, Volume 411, 108815, ISSN 0304-3800, https://doi.org/10.1016/j.ecolmodel.2019.108815.

Millard, SP and NK Neerchal. 2001. Environmental Statistics with S-Plus. CRC Press, NY.

Nikparvar, B and J-C Thill. 2021. Machine Learning of Spatial Data. ISPRS International Journal of Geo-Information 10, no. 9: 600.

NRC. 2020. *Multi-agency Radiation Survey and Site Investigation Manual (MARSSIM)*. NUREG-1575, Revision 2. EPA-402-)-20-001. DOE/AU-0002. U.S. Nuclear Regulatory Commission, Washington, D.C.

O'Brien, RF, DK Carlson, RO Gilbert, JE Wilson, DJ Bates, and BA Pulsipher. 2005. Statistical algorithms for designing geophysical survey to detect UXO target areas. *Subsurface Sensing Technologies and Applications*, 6(3). DOI:10.1007/s11220-005-0012-1

Pebesma, EJ, and RS Bivand. 2005. Classes and methods for spatial data in R. *R News* 5(2).

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

R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

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

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

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

Sanada, Y., Urabe, Y., Sasaki, M., Ochi, K. and Torii, T. 2019. "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.

Sanderson, D. 2013. "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.

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

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. 2006. *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, TN.

Tanigaki, M., Okumura, R., Takamiya, K., Sato, N., Yoshino, H. and Yamana, H. 2013. "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.

Tao, H., Liao, X., Zhao, D., Gong, X., & Cassidy, D. (2019). Delineation of soil contaminant plumes at a co-contaminated site using BP neural networks and geostatistics. *Geoderma, 354,* 113878. https://doi.org/10.1016/j.geoderma.2019.07.036

Ver Hoef, JM, and N Cressie. 1993. Multivariable Spatial Prediction. *Mathematical Geology*, 25(2).

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

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

Wohlberg, B., Tartakovsky, D.M., & Guadagnini, A. (2005). Subsurface characterization with support vector machines. *IEEE transactions on geoscience and remote sensing 44(1),* 47—57.

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

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

Zhu, H, and A Journel. 1993. 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. DOI:10.1007/978-94-011-1739-5_1.

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

Additional data sets representing continuously collected survey data of land areas were prepared for future studies to demonstrate methods of statistical analysis and visualization. The data sets 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 (MARSSIM) (NRC 2020). 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 data sets support surveys classified by MARSSIM as Class 1 and Class 3.

<u>Class 1 Areas</u>

Areas that have, or had before remediation, a potential for residual radioactive material (based on site 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 on the site 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.

<u>Class 3 Areas</u>

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 an area’s classification from Class 1 or Class 2 to Class 3, the existing data (from the historical site assessment (HAS), 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 Data Set 1

Data set 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 data set, 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, PNNL altered the data so that 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.

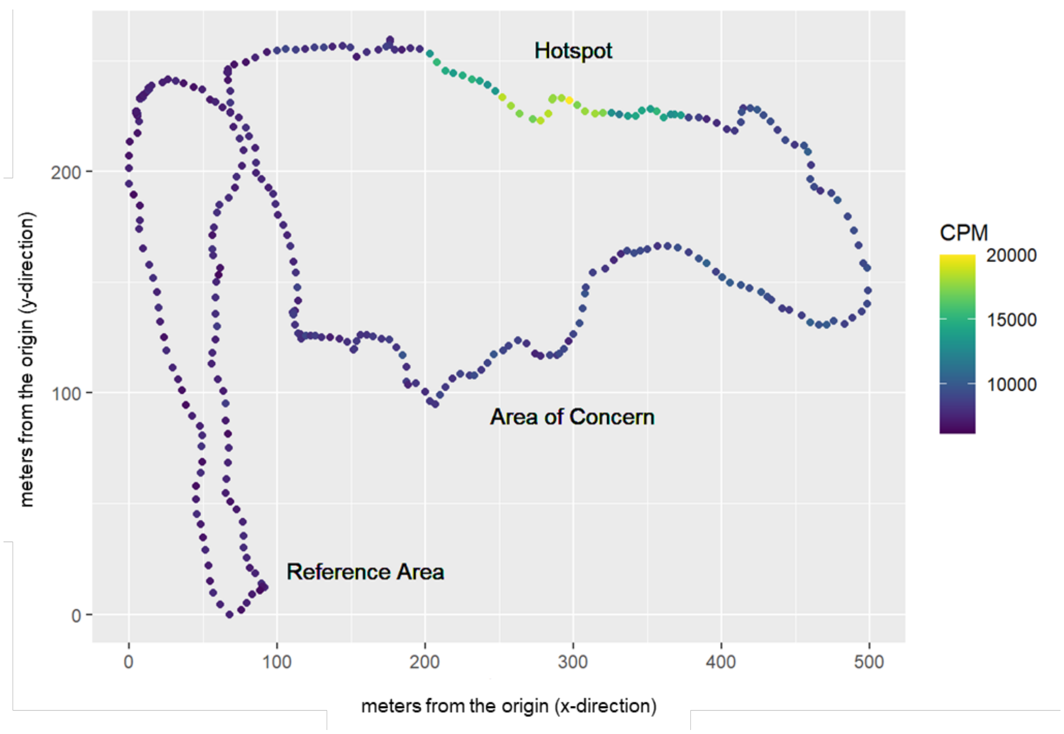


Figure A.1. Data set 1 observed radiation measurements (CPM) across a land area with a reference area and an area of concern.

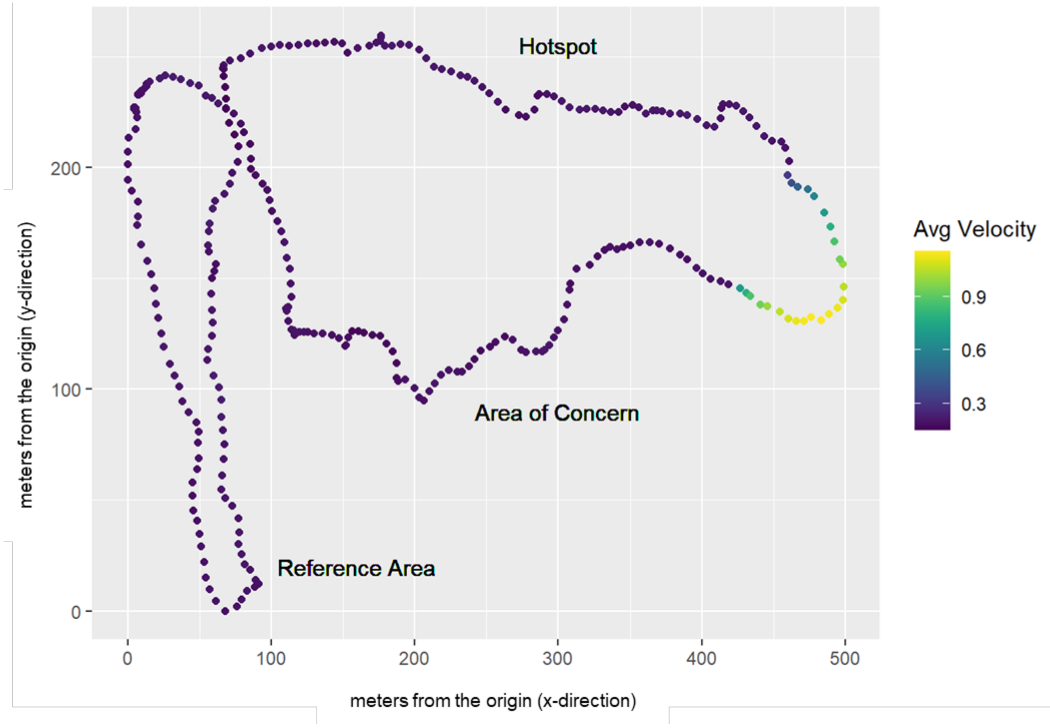


Figure A.2. Data set 1 observed average velocity (m/s) within reference area and area of concern.

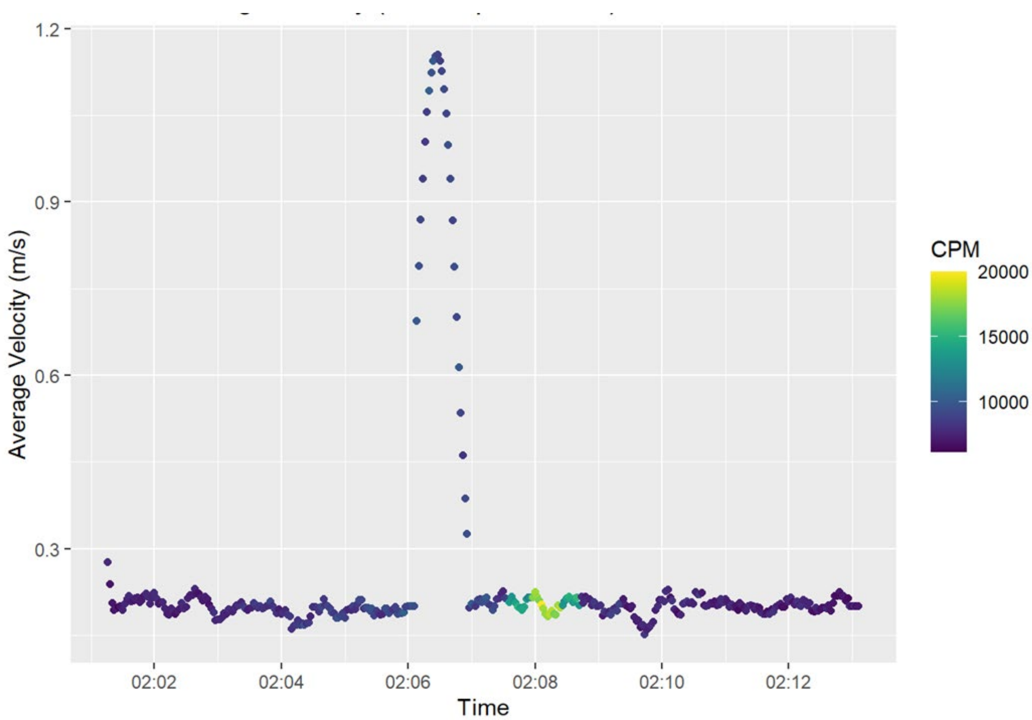


Figure A.3. Data set 1 observed average velocity (in m/s) over the elapsed scanning time (in hours and minutes) with coloration showing the radiation measurements (CPM).

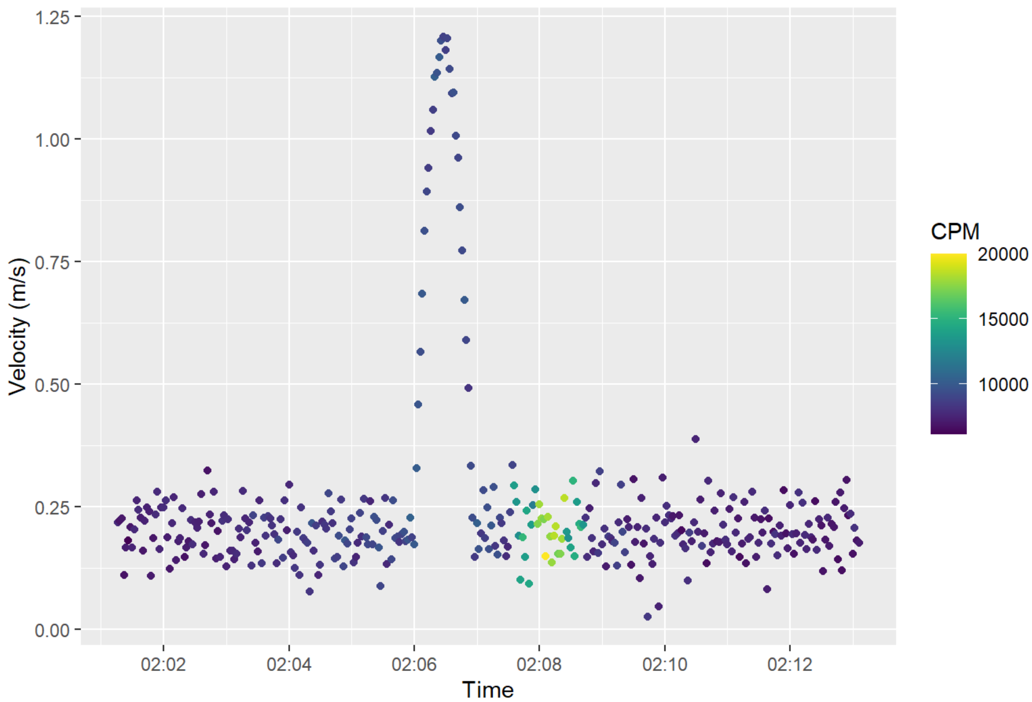


Figure A.4. Data set 1 observed instantaneous velocity (m/s) vs. elapsed time (in hours and minutes) with coloration reflecting the radiation measurements (CPM).

A.2 Data Set 2

Data set 2 includes the same data from a reference area and an area of concern, scanned using a continuous surveying technique without vigilance. As for Data set 1, Figure A.5 through Figure A.8 show observations of each variable in the data set. 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.

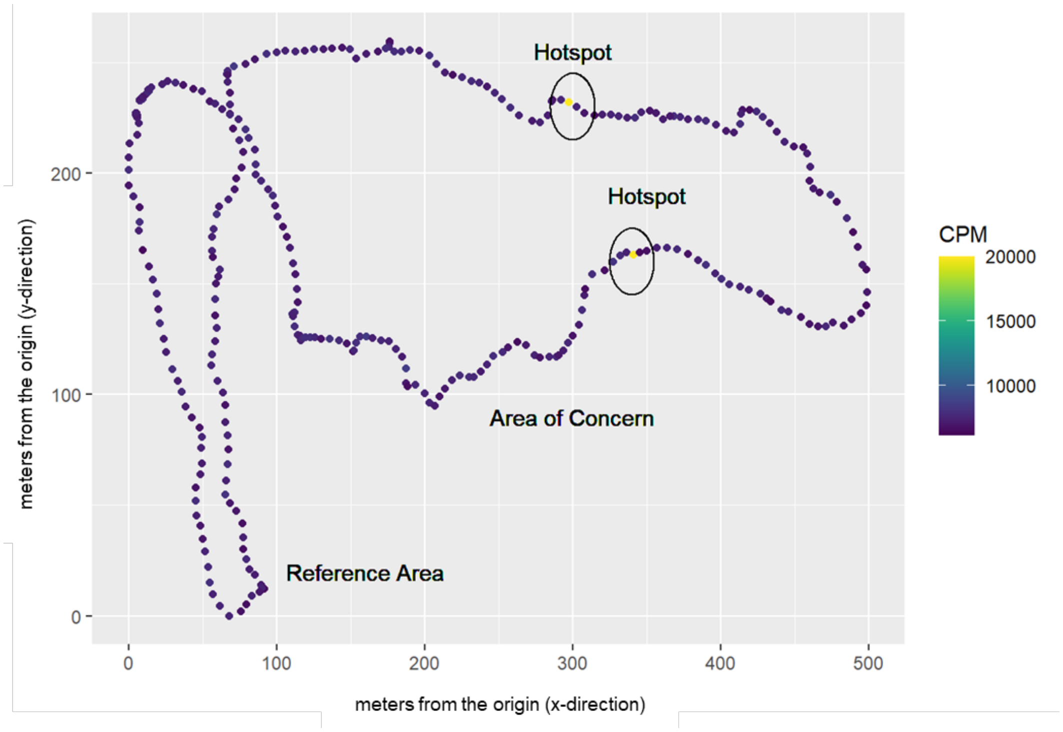


Figure A.5. Data set 2 visualization of radiation measurements (in CPM) across a land area with a reference area and area of concern with two hot spots.

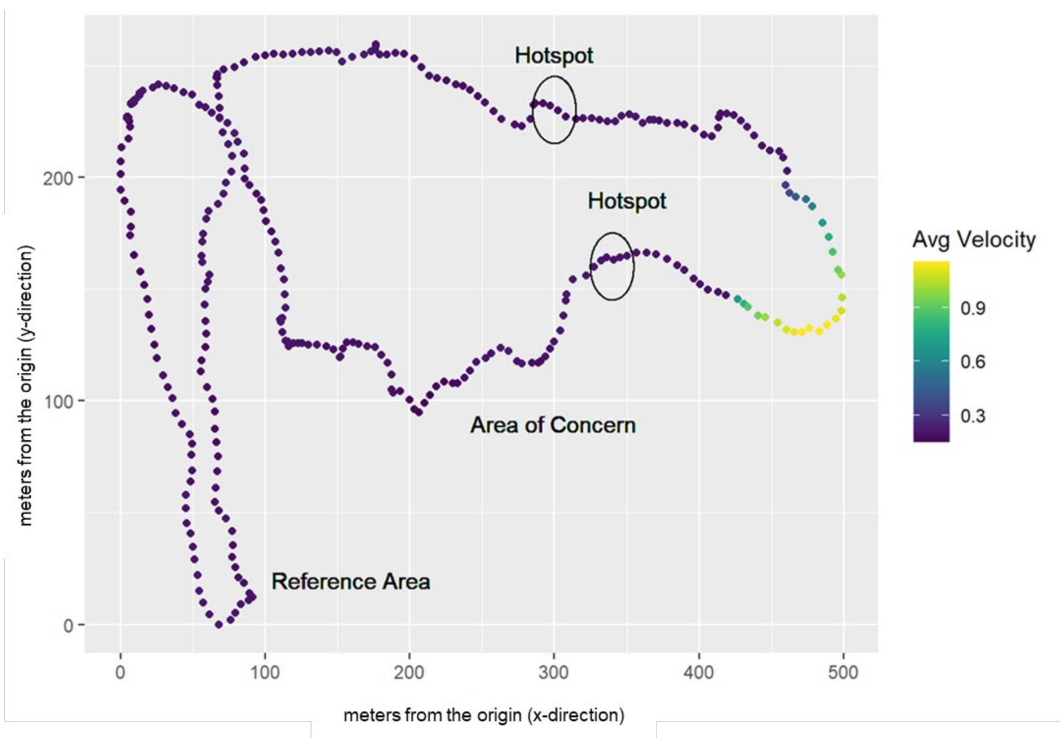


Figure A.6. Data set 2 visualization of average velocity (in m/s) across a land area with a reference area and area of concern with two hot spots.

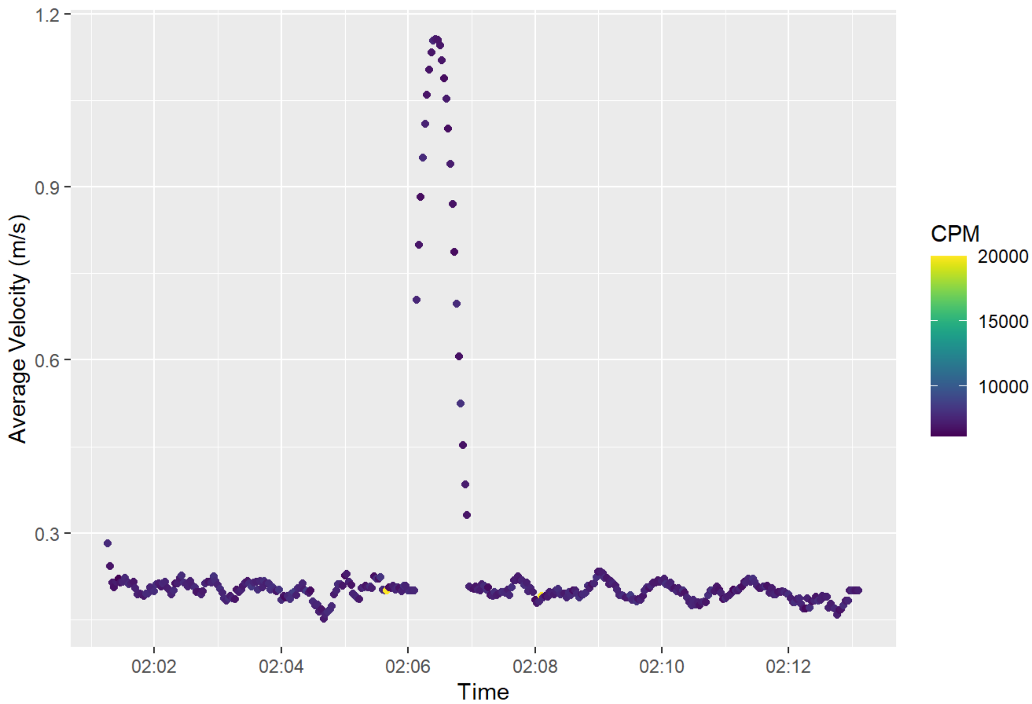


Figure A.7. Data set 2 visualization of average velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for the land areas in Figure A.5.

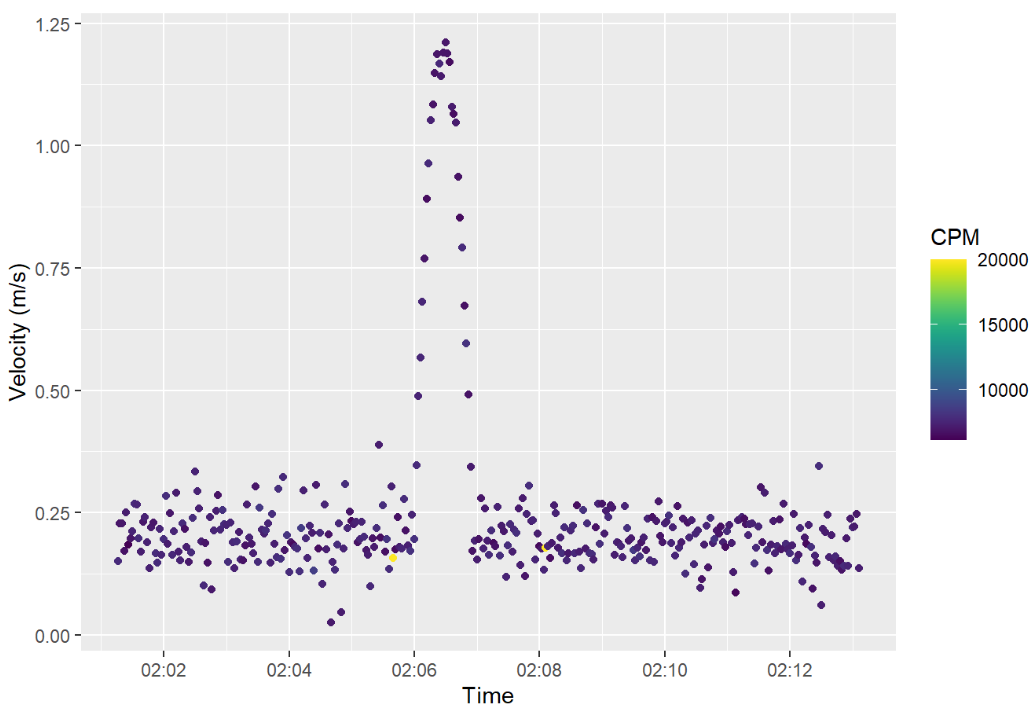


Figure A.8. Data set 2 visualization of instantaneous velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for the land areas in Figure A.5.

A.3 Data Set 3

Data set 3 includes the same data from a reference area and an area of concern, scanned using a continuous surveying technique without vigilance. As for Data sets 1 and 2, Figure A.9 through Figure A.12 show observations of each variable in the data set. 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.

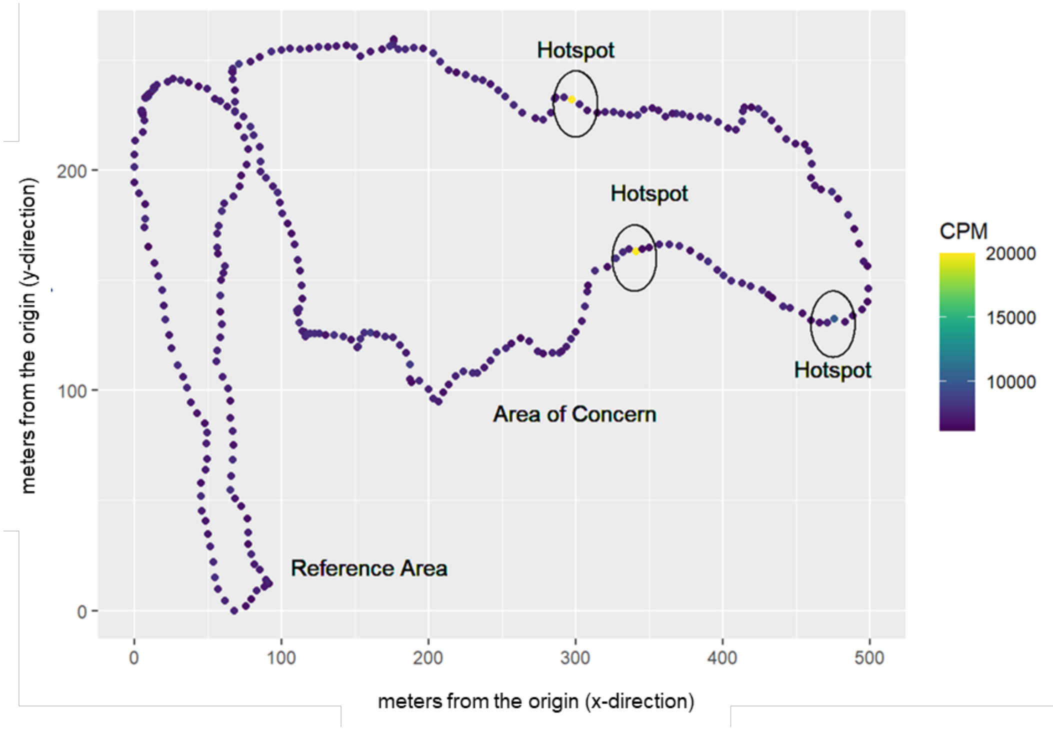


Figure A.9. Data set 3 visualization of radiation measurements (in CPM) across a land area with a reference area and area of concern with three hot spots.

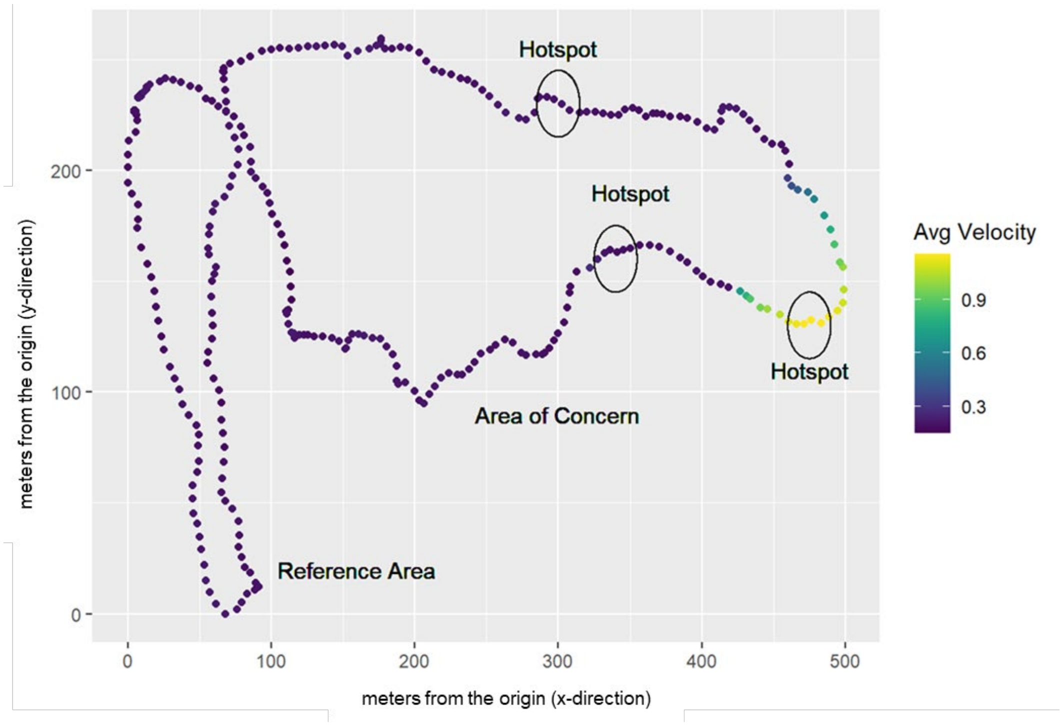


Figure A.10. Data set 3 visualization of average velocity (in m/s) across a land area with a reference area and area of concern with three hot spots.

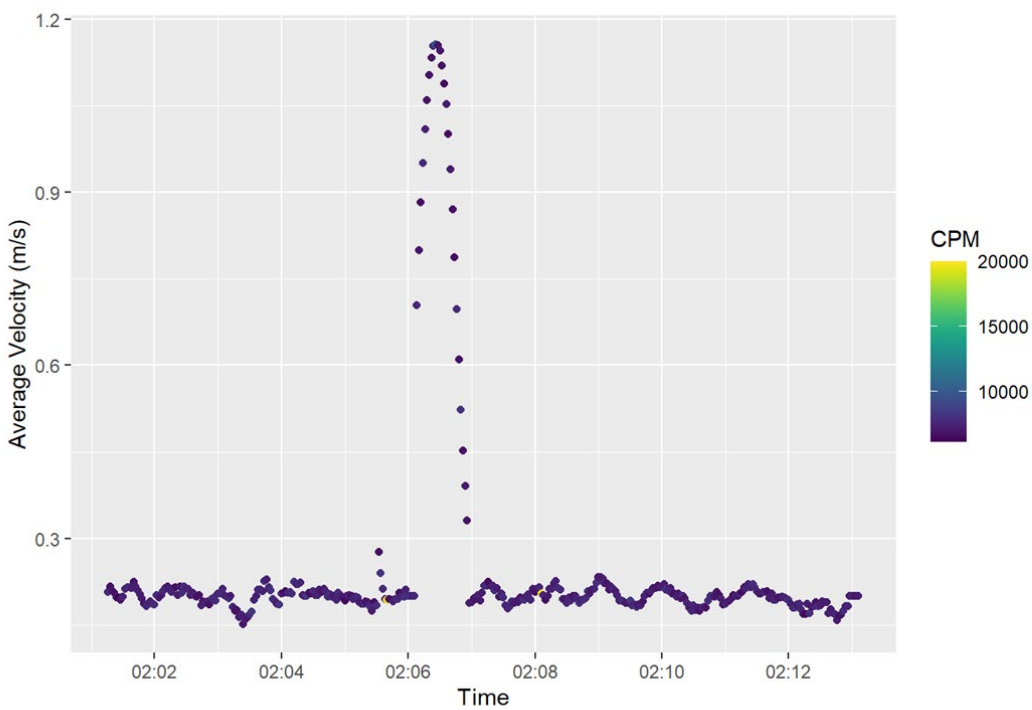


Figure A.11. Data set 3 visualization of average velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for the land area in Figure A.9.

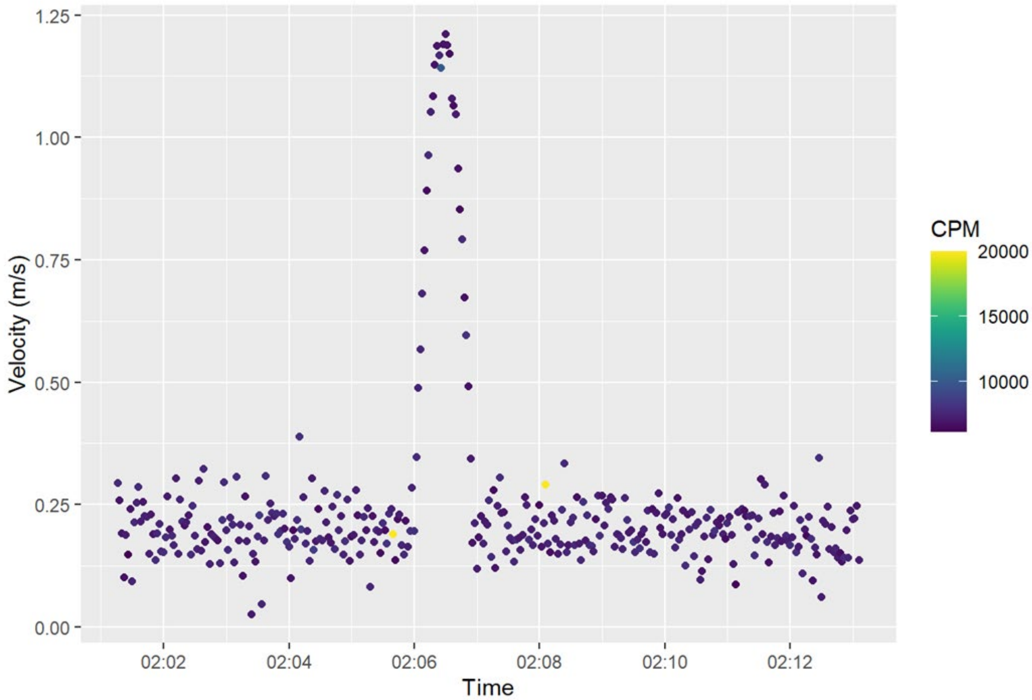


Figure A.12. Data set 1, Test 1a, visualization of instantaneous velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for the land areas in Figure A.9.

A.4 Data Set 4

Data set 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 m/s).

Data set 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. As for Data set 1, Figure A.5 through Figure A.8 show observations of each variable in the data set. 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.

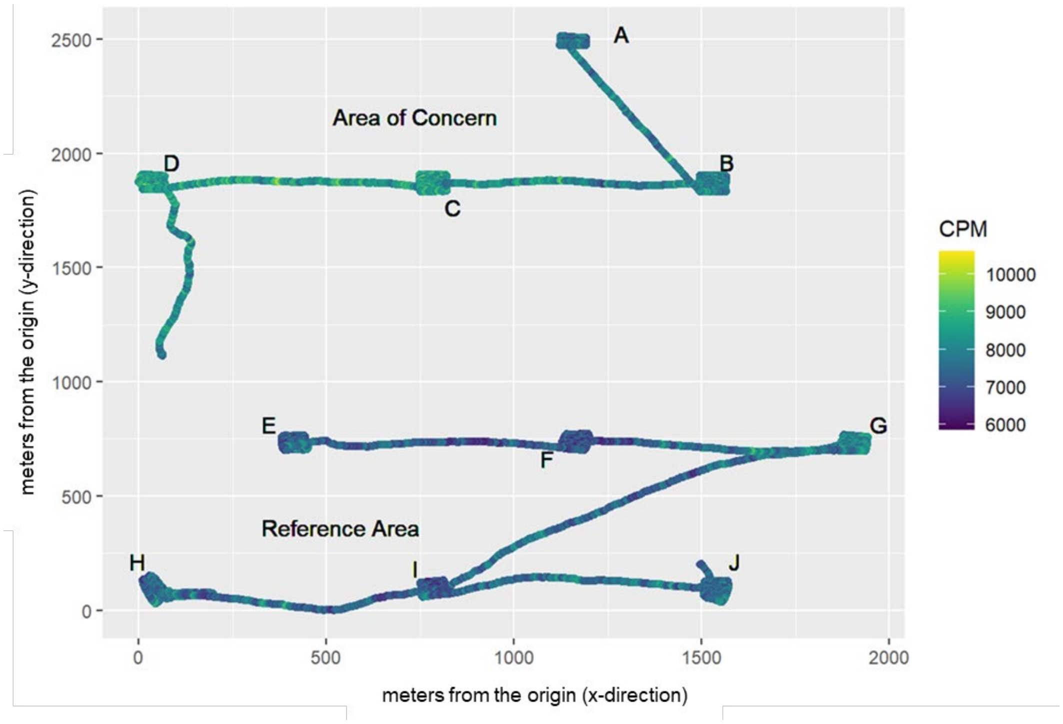


Figure A.13. Data set 4 radiation measurements (in CPM) across a land area including several reference areas and several areas of concern.

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.

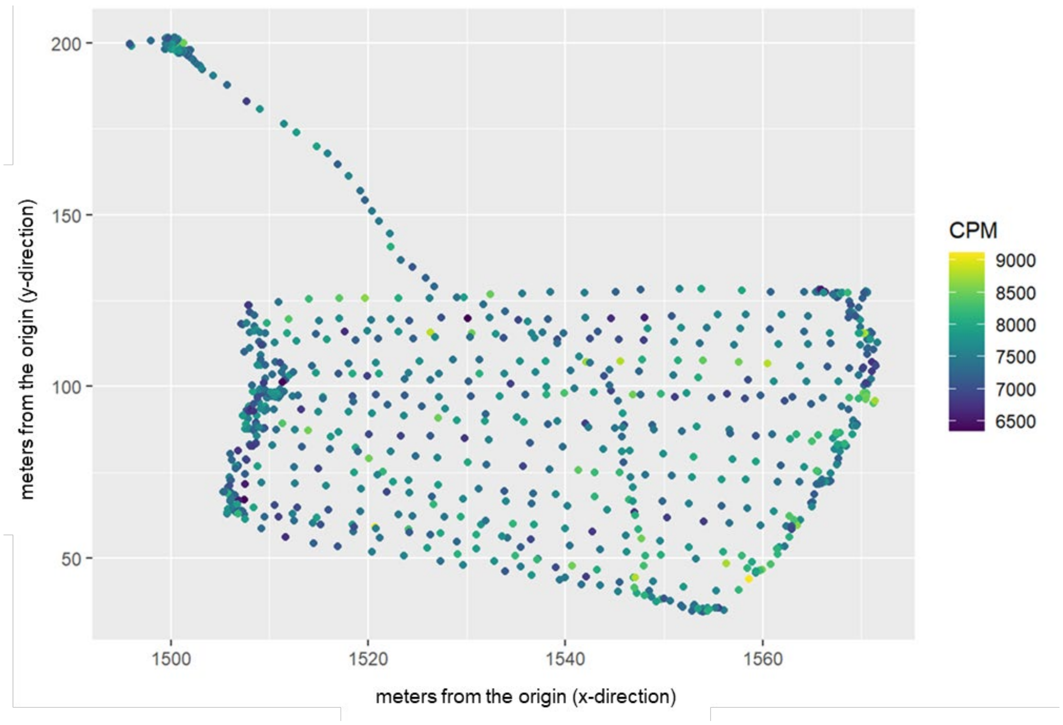


Figure A.14. Data set 4 radiation measurements (in CPM) across subsite "J" reference area from Figure A.13.

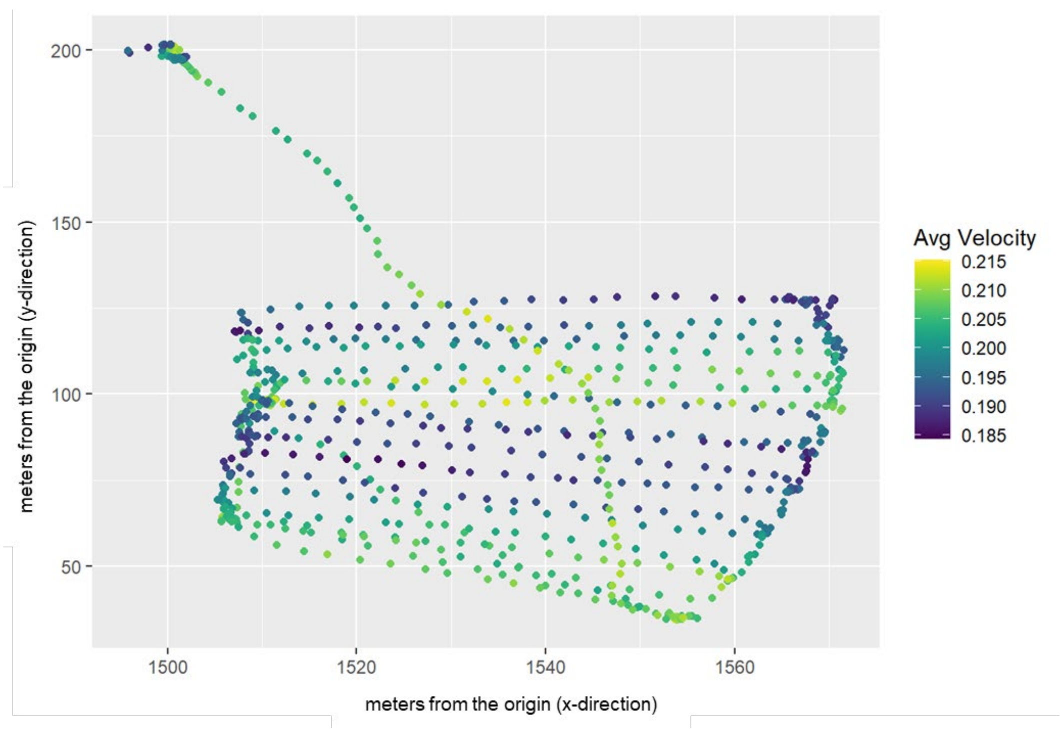


Figure A.15. Data set 4 average velocity (in m/s) across reference area "J" from Figure A.13.

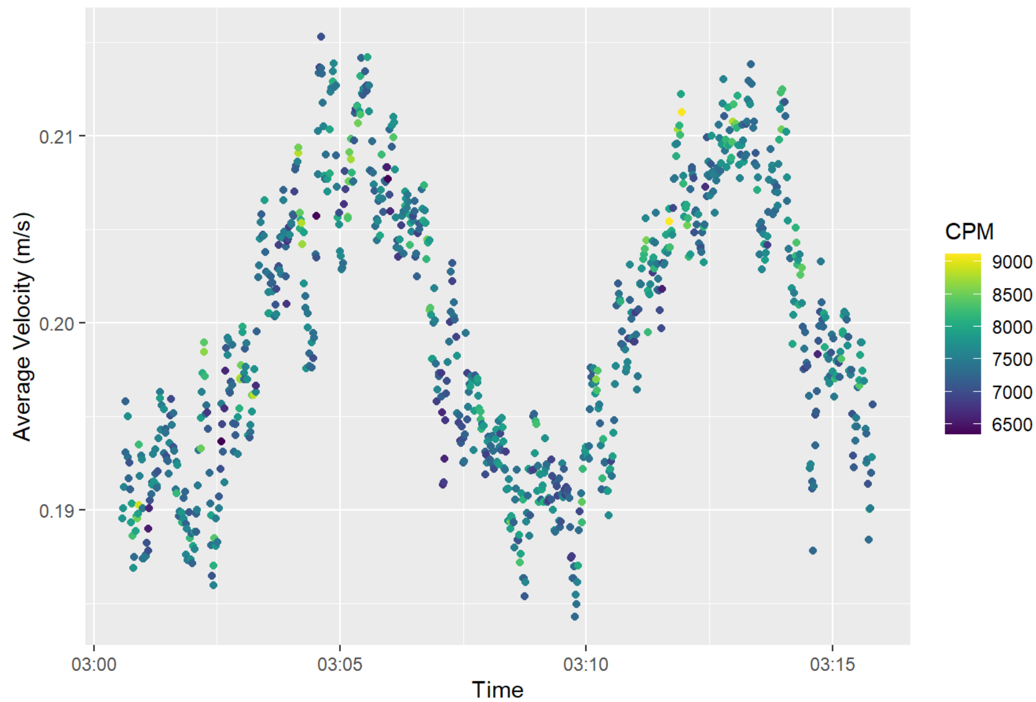


Figure A.16. Data set 4 average velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for reference area “J” in Figure A.14.

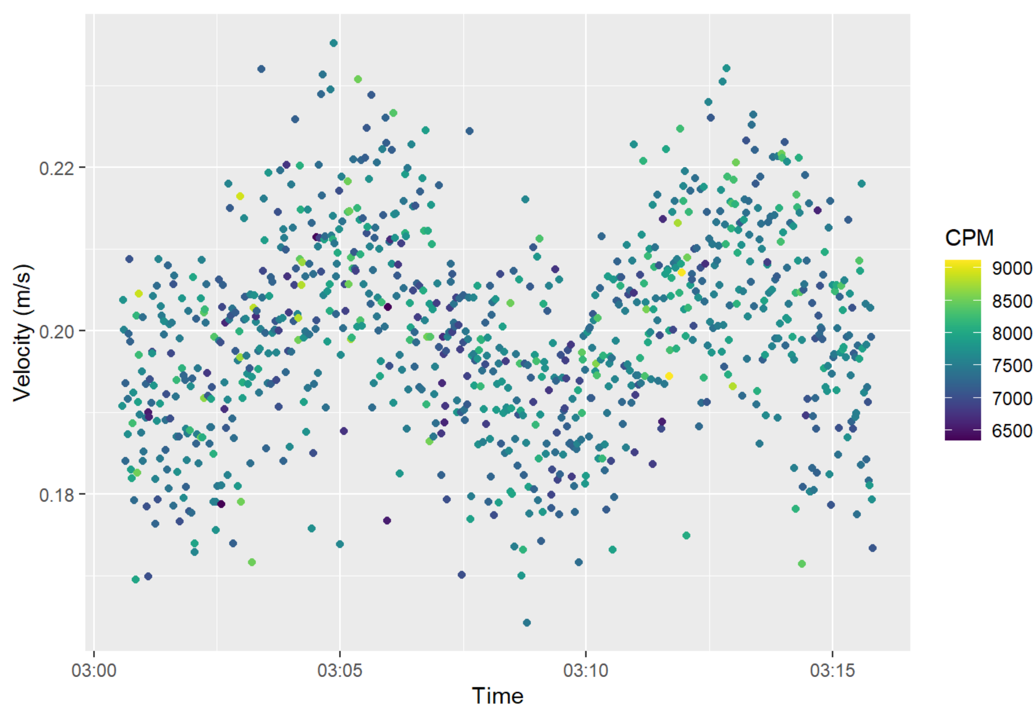


Figure A.17. Data set 4 instantaneous velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for reference area “J” in Figure A.14.

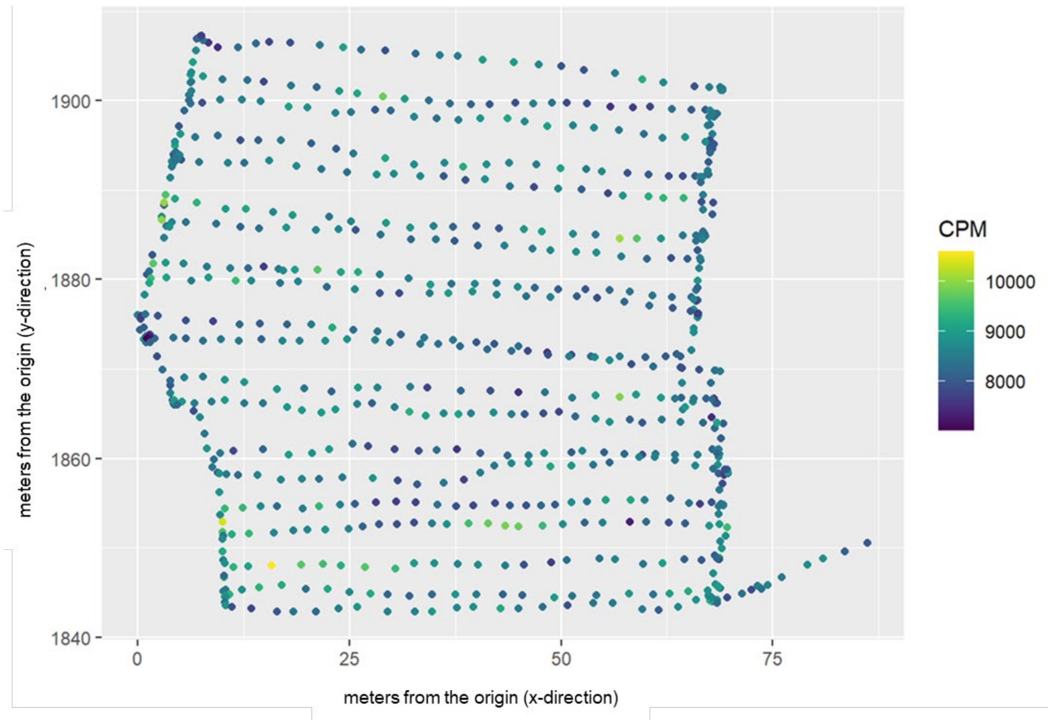


Figure A.18. Data set 2 radiation measurements (in CPM) across area of concern “D”.

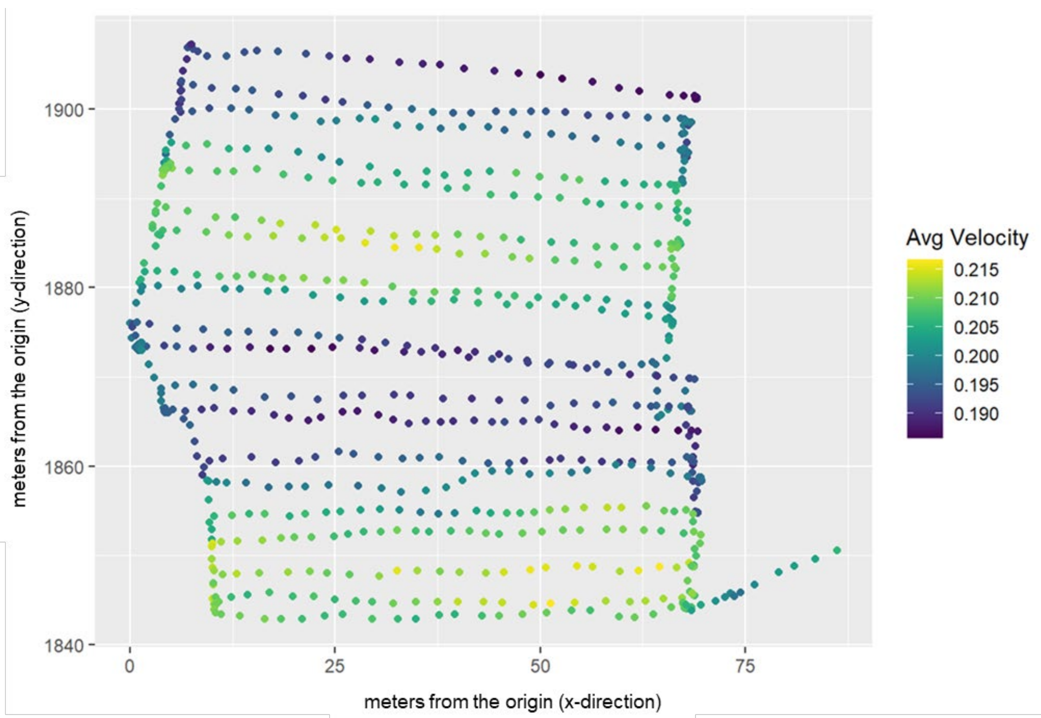


Figure A.19. Data set 2 average velocity (in m/s) across area of concern “D” in Figure A.18.

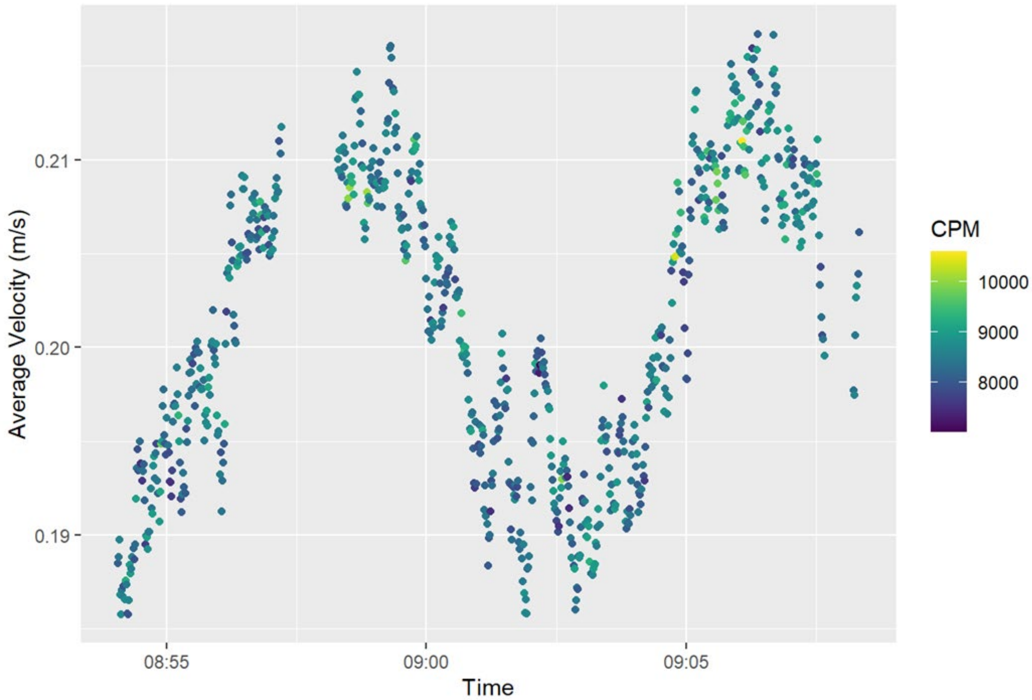


Figure A.20. Data set 2 average velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for area of concern "D" in Figure A.18.

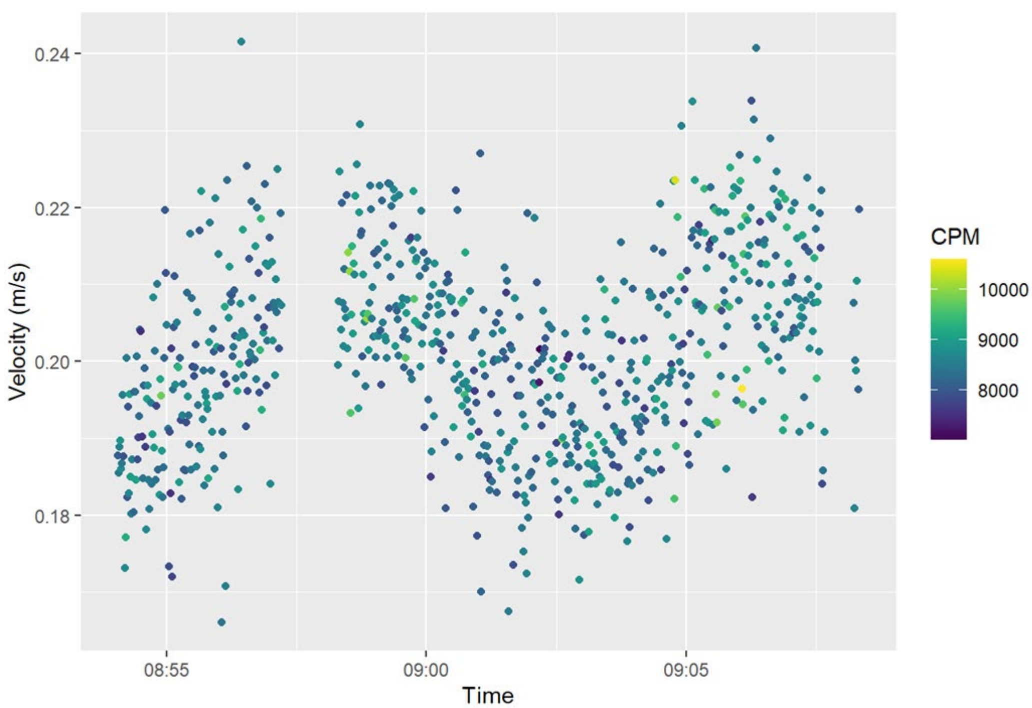


Figure A.21. Data set 2 visualization of instantaneous velocity (in m/s) vs. elapsed time (in hours and minutes) showing the radiation measurements (in CPM) for subsite "D" area of concern evaluated in Figure A.19.

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