An Evaluation of Nearest Neighbor Averaging's Improvements to the Measurement Sensitivity of Wide-Area Radiological Scanning Instruments -15560

Mark L. Miller*, Sean D. Fournier*, Robert P. Miltenberger*,
Patrick S. Beall*, Tyler J. Alecksen** and Mike J. Schierman**
* Sandia National Laboratories and ** ERG, Inc.

ABSTRACT

Through the Sandia National Laboratories (SNL) New Mexico Small Business Assistance (NMSBA) program, Sandia worked with the Environmental Restoration Group (ERG) Inc. to verify and validate a novel algorithm used to determine the scanning Critical Level (Lc) and Minimum Detectable Concentration (MDC) (or Minimum Detectable Areal Activity) for the 102F scanning system. Through the use of Monte Carlo statistical simulations the algorithm mathematically demonstrates accuracy in determining the Lc and MDC when a nearest-neighbor averaging (NNA) technique was used. To empirically validate this approach, SNL prepared several spiked sources and ran a test with the ERG 102F instrument on a bare concrete floor known to have no radiological contamination other than background naturally occurring radioactive material (NORM). The tests conclude that the NNA technique increases the sensitivity (decreases the Lc and MDC) for high-density data maps that are obtained by scanning radiological survey instruments. It is further described in Sandia Laboratories document number SAND2014-16522.

INTRODUCTION

The nearest-neighbor averaging technique is used to improve the sensitivity of field survey instruments and reduce the variance of the data. This paper will describe this averaging technique as well as its effects on the L_c and MDC calculations. The work herein does not attempt to describe the statistical sampling and analysis theory in any great detail.

Nearest Neighbor Averaging

Nearest-neighbor averaging is a technique used to improve the sensitivity and reduce the variance of spatially-correlated data maps. The raw number of counts in a given map is replaced by a new map of values that represent the average of the "N" nearest neighbors of the data. Figure 1 shows this approach graphically.

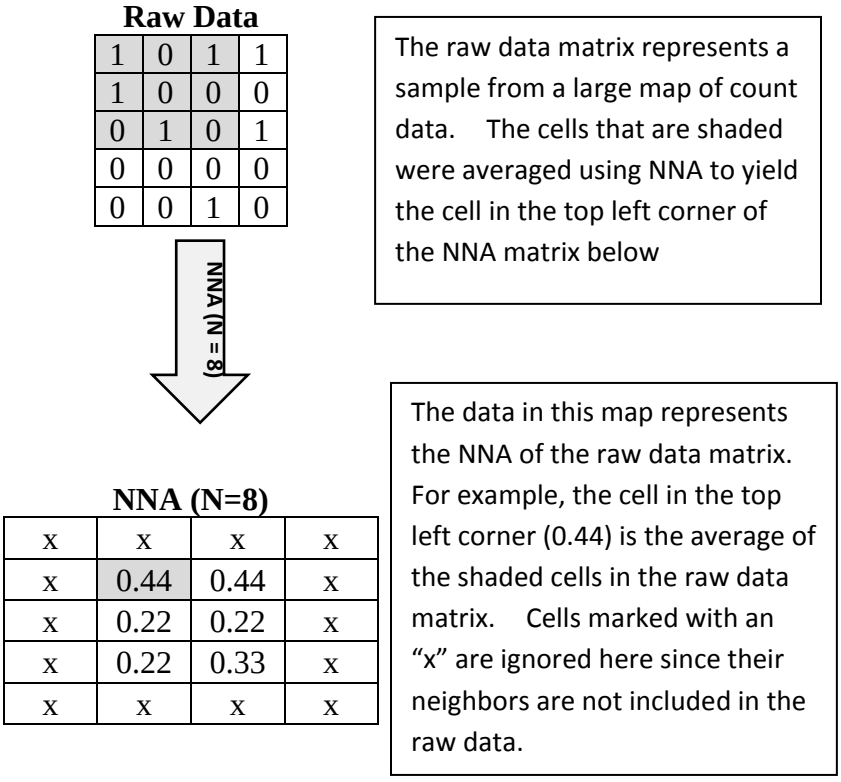

Figure 1. Demonstration of NNA in two dimensions.

In this example, the data set in the raw data map is "smoothed" by averaging each pixel with its 8 nearest neighbors and replacing its value with the averaged data. According to statistical sampling theory, when a distribution is sub-sampled and averaged, the variance is reduced by a factor equal to the number samples (in our case, N+1 where N is the number of neighbors). Furthermore, it can be shown by the Central Limit Theorem that averaging data in this way pushes the distribution to be more Gaussian (normal) as the number of neighbors increases. Both of these effects of averaging have positive consequences when applied to spatially-correlated radiological survey data.

By reducing the variance by a factor of N + 1 (The number of neighbors plus the cell itself), the uncertainty in any given measurement is reduced by a factor of $\sqrt{N+1}$. This reduced uncertainty is desirable when counting statistics are poor (such as measurements very near background). Also, by shifting the Poisson distribution to be more normal, conventional statistics are more applicable and intuitive for the data reviewers and decision makers.

The Critical Level (L_c) [2]

Conceptually, the L_c is derived from the distribution that results from an infinite number of

background measurements that have been subtracted by the true "well known" background value and must be centered about zero. The L_c is the activity (or number of counts) that result in a "false positive" rate that is decided upon for a given survey. Typically this is set to 5% (k = 1.645). Figure 2 demonstrates this concept graphically.

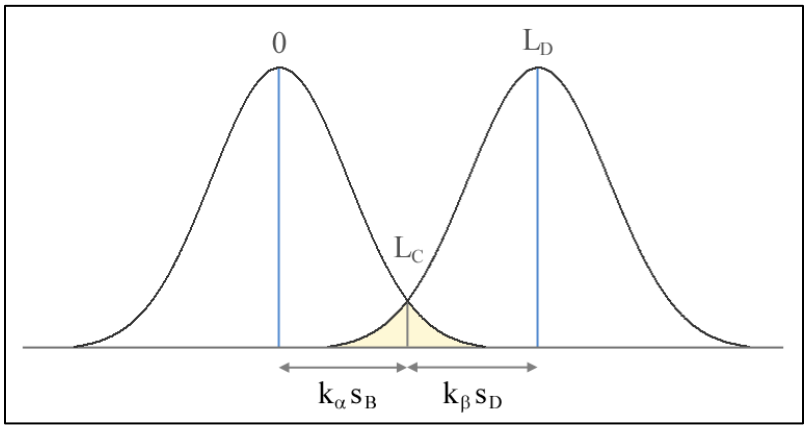

Figure 2. Graphical representation of the Critical Level and the Detection Level

Since by definition L_c must be centered about 0, the equation for L_c at the 95% confidence interval is as follows:

$$L_C = ks_B \qquad (1)$$

Where:

k – The statistical coverage factor for 95% confidence (equal to 1.645 in this case)

S_b – The standard deviation of the "well known" background distribution.

Operationally, the L_c is used to determine if a datum in a set is "different" than background (i.e. represents a measurement of radioactivity above background). In a large set of data, one would expect that only 5% of data that is truly at background will be quantified above the critical level.

Since data obtained with nearest-neighbor averaging has a reduced variance, one would expect the critical level to be reduced by the same factor. The novel concept of this algorithm is that the NNA L_c is calculated with the following equation:

$$L_C = \frac{ks_B}{\sqrt{N+1}} \qquad (2)$$

Where:

N - the number of nearest neighbors

One of the goals of this experiment was to assess the validity of this equation.

The Minimum Detectable Concentration (MDC) [2]

Similar to the Critical Level, the MDC is used as a metric for the sensitivity of a detection method. When the true activity of a source is at the theoretical MDC, only a pre-defined fraction (5% for the purpose of this study) of measurements should fall below the theoretical L_c.

The MDC can be defined as a function of the L_c and through algebra, a function of the standard deviation of the "well known" background. Typically, the Detection Level (L_D) is calculated and then converted to the MDC.

$$L_D = k^2 + 2ks_B \qquad (3)$$

Through the statistical theorems described in the section above, this equation becomes:

$$L_D = \frac{k^2+2kS_B\sqrt{(N+1)}}{(N+1)} \quad (4)$$

To convert L_D to a meaningful value of activity, the following equation is used:

$$MDC = \frac{L_D}{\epsilon\, t\, A} \quad (5)$$

Where,

A = detector area factor (unitless) for conversion to 100cm^2

ε = total detector efficiency in counts per disintegration. This is equal to the detector efficiency multiplied by the surface efficiency

t = counting time in minutes.

One of the goals of this experiment was to assess the validity of this equation.

The ERG 102F Scanning Instrument

The ERG 102F floor scanning instrument (see Figure 3) is configured with six 100 cm^2 zinc-sulfide plastic scintillator probes running through a dual-channel analyzer that records scalar detection events. All events detected within a pre-defined scanning period are integrated and process through custom software on a laptop. Each data point is correlated with an (X,Y) position determined by a laser positioning system. The data is processed and displayed on a laptop computer in real time as the instrument is being run. Averaging and data flagging can be done onboard or the data can then be exported to a shape file or .CSV for further analysis.

Figure 3. The ERG 102F Floor Scanning Instrument

The ERG 102F instrument was the primary instrument used for this study. However, the methods described and tested during this study for NNA and how it affects the L_C and MDC can be applied to any spatially-correlated data set that is processed through NNA techniques.

STATISTICAL SIMULATIONS

Description of Experiment

Several Monte-Carlo simulations of radiation counting were run through the nearest-neighbor averaging algorithm and the true critical level values were compared to theory. For simplicity's sake, a 1D approximation was made in that data was organized in a column and not on a 2-dimentional map. However, the same averaging approach was applied along the column of data.

A Poisson distribution representing the "background" was generated randomly with a mean equal to a typical number of background counts. The distribution of number of counts was scaled for a 1, 3, and 7 second simulated count time. For example, the distribution for a 1 second count time may have a mean of 0.5. When scaled to 3 seconds, the new mean has a value of 1.5. Both of these distributions are for the same true background but the number of expected counts increases with the increased count time. The three distributions were averaged using NNA at various numbers of neighbors. The elements that exceeded the theoretical critical level were tallied as "false positives".

The purpose of this simulation is to demonstrate that the variance reduction caused by the averaging will lower the L_c proportionally.

Results

The predicted L_c and observed false positive rate were tallied for each of the nine trials for a typical alpha background and a typical beta background. The results are summarized in Table I below.

Table I. Results of L_c Monte Carlo Simulations

	Alpha Background = 4 cpm								
	t = 1 sec			**t = 3 sec**			**t = 7 sec**		
	N = 0	**N = 2**	**N = 4**	**N = 0**	**N = 2**	**N = 4**	**N = 0**	**N = 2**	**N = 4**
Predicted Lc @ k = 1.645	0.49	0.31	0.26	0.94	0.62	0.53	1.59	1.12	0.97
Observed false positive rate	5.50%	15.53%	2.61%	19.90%	15.33%	10.34%	8.00%	6.21%	9.04%

Note: N is the number of neighbors averaged. Lc values in this table are dimensionless; they represent the number that results after the raw counts are NN-averaged.

Table I (continued)

	Beta Background = 250 cpm								
	t = 1 sec			**t = 3 sec**			**t = 7 sec**		
	N = 0	**N = 2**	**N = 4**	**N = 0**	**N = 2**	**N = 4**	**N = 0**	**N = 2**	**N = 4**
Predicted Lc @ k = 1.645	7.52	6.11	5.67	18.32	15.86	15.10	38.05	34.30	33.14
Observed false positive rate	5.70%	6.21%	5.91%	6.30%	5.91%	5.42%	4.50%	6.31%	6.02%

Note: N is the number of neighbors averaged. Lc values in this table are dimensionless; they represent the number that results after the raw counts are NN-averaged.

Conclusions

The results for alpha are somewhat inconclusive in that the statistics are not reliable for the distributions that are obtained. The results indicate that for distributions that are strongly Poisson

(i.e. alpha with a short count time) the false positive rate for the predicted L_c can vary wildly. This is likely due to the fact that the distributions have very few bins which lead to a largely unpredictable number of ones and zeroes. Distributions with means farther away from zero tend to yield more predictable statistics. Note that the false positive rates for the alpha 7 second trial are more uniform. The reader is encouraged to reference the Multi-Agency Radiological Laboratory Analytical Protocols Manual (MARLAP)[3] Chapter 20, Appendix A "Low-Background Detection Issues" which states that the Currie method of calculating the Lc may "produce a high rate of Type I errors" when the number of counts is low.

The beta results are much more promising. The false positive rates are all in the neighborhood of 5% and are uniform across NNA and count time settings. This verifies that in theory the NNA technique will lower the L_c by a factor proportional to the square root of N+1. However, the field test is needed to showcase this technique in a realistic scenario when all uncertainties will be observed.

FIELD TEST

Description of Experiment

An experiment was designed to test the validity of the NNA technique when applied to determining the L_C and MDC for radiological surveys that utilize scanning instruments. Two separate tests were done with the ERG 102F instrument. The first test was to determine the validity of the L_c calculation. The second part of the experiment was to determine the validity of the MDC calculation.

Critical Level (L_C) Test

The L_c test was performed by scanning a concrete floor at typical scan settings that was known to be free of man-made contamination (or not detectable by standard survey methods). The data map obtained was put through the NNA technique and the results that fell above the critical level were tallied as "false positives". To perform this test, the ERG 102F needed to be calibrated and a background had to be established (see Figure 4).

The instrument was calibrated for detector efficiency with NIST-traceable standards. These anodized aluminum sources have an active area of 100 cm^2. The pure-alpha source was Pu-238, the pure-beta source was Sr/Y-90. Each source has approximately 10,000 dpm certified to +/- 3% uncertainty at 1-sigma. The efficiency was determined by performing 10 1-minute counts of each source and determining the 2pi efficiency using the sources' certified emission rate. The average 2pi efficiency from these ten runs for each detector was then applied as the detector efficiency. The "well known" background was determined by performing 10 1-minute static counts of the bare concrete floor. The average background was then applied as the instrument background for each detector. The recommended surface efficiency correction factors (as described in MARSSIM [1]) for concrete were used in the determination of the total instrument efficiency. These values are 0.5 and 0.25 for beta and alpha respectively. Note that the surveyor efficiency is

assumed to be 1.00 with this instrument since it does not rely on a surveyor's attention during measurement acquisition.

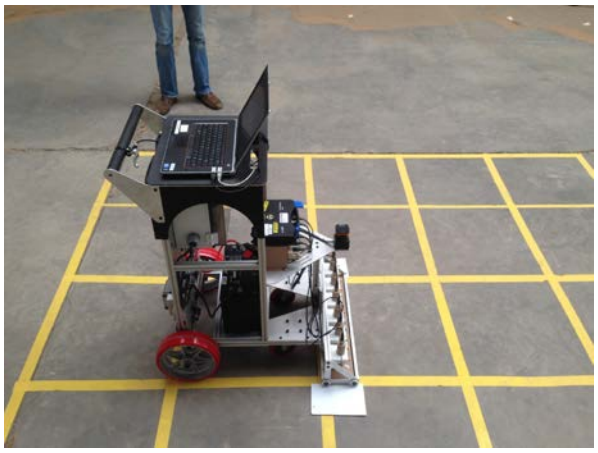

Figure 4. Calibration and Background Determination with the ERG 102F

Two measurements of the bare concrete floor were done at different scan settings. One test was done with scan speeds and count times similar to a typical decommissioning survey performed by ERG and the other was at a slower speed for the same count time to obtain more data. A summary of the scan settings for the two tests are described below in Tables II and III. The results from this test are shown in Tables VI and VII in the following section.

Table II. Measurement Settings for the Low Density L_c Test

Background Media	Bare Concrete
Alpha Surface Efficiency	0.25
Beta Surface Efficiency	0.5
Scan Speed	2 inches/sec
Count Time	10 seconds
# of Measurements	390

Table III. Measurement Settings for the High Density L_c Test

Background Media	Bare Concrete
Alpha Surface Efficiency	0.25
Beta Surface Efficiency	0.5
Scan Speed	0.5 inches/sec
Count Time	10 seconds
# of Measurements	588

Minimum Detectable Concentration (MDC) Test

The MDC test was conducted by scanning a row of sources spiked to a typical MDC for given scan settings. Two sets of 9 identical sources each were constructed by SNL. The source material was NIST-traceable depleted uranium (DU). Depleted uranium was chosen since it is commonly encountered in decommissioning work and is a mixed alpha/beta source material.

The source material was diluted and gravimetrically stippled onto the surfaces of 4-inch diameter (~ 100 cm^2) stainless steel planchets. (See Figure 5) The material was allowed to dry onto the planchet before the test. The sources were spiked to levels typically encountered as required minimum detectable concentrations. The alpha set was spiked to 147 +/- 1.5 dpm alpha. The beta set was spiked to 505 +/- 5 dpm beta. Note: Errors quoted here are at 1-sigma.

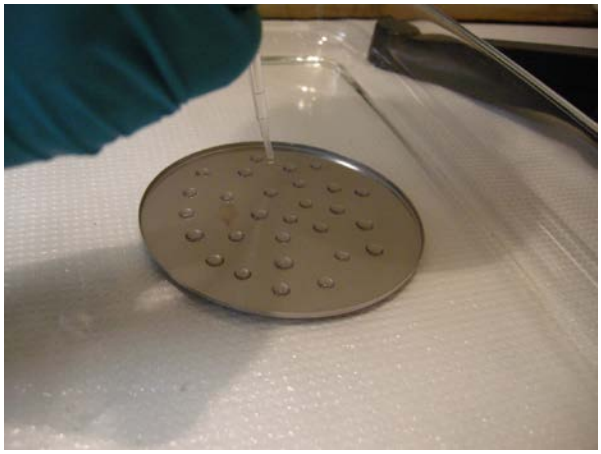

Figure 5. DU Source Preparation

Since the source matrix differs from concrete, the appropriate surface efficiency factors had to be calculated using one of the prepared sources. The prepared source was counted for 10 min and the resulting counts per minute was divided by the true spiked activity to obtain the total efficiency. This total efficiency was then divided by the detection efficiency to obtain the surface efficiency. These were determined to be 0.59 for Beta and 0.25 for Alpha.

The stainless steel was found to have significantly reduced the background from the NORM in the concrete. To account for this difference, backgrounds were determined by 10 sequential 1-minute counts of a blank stainless steel planchet. Only data from the detector that the row was positioned under were considered in the calculation of the “false negative” rate. These sources represent a surface contaminated uniformly at the MDC. The detector was placed over the first source and the scan speed was chosen such that all measurements were acquired while the detector was over the sources (See Figure 6).

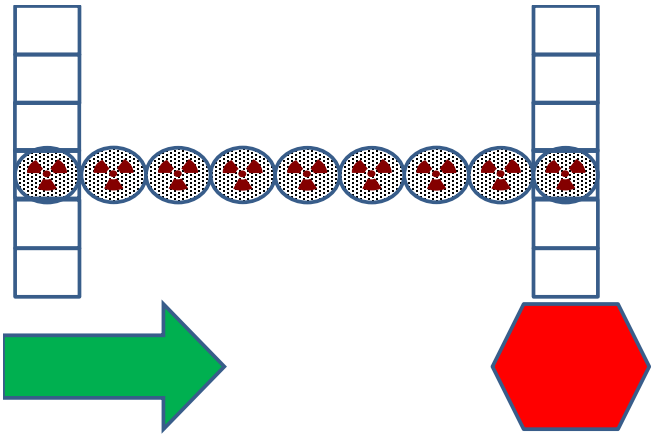

Figure 6. Schematic of MDC Test Scanning Scenario

Tables IV and V below summarize the measurement settings for the alpha and beta MDC tests.

Table IV. Measurement Settings for the Alpha MDC Test

Background Media	Planchet
Alpha Background (cpm)	1.1
Alpha Efficiency (2pi)	0.314
Alpha Surface Efficiency	0.25
# of neighbors	4
Scan Speed	1 inch/sec
Count Time	5 seconds
Alpha L_c (dpm)	34
Alpha MDC (dpm)	151
# of Measurements/trial	7
# of trials	30

Table V. Measurement Settings for the Beta MDC Test

Background Media	Planchet
Beta Background (cpm)	289.1
Beta Efficiency (2pi)	0.387
Beta Surface Efficiency	0.589
# of neighbors	5
Scan Speed	1.5 inches/sec
Count Time	3 seconds
Beta L_c (dpm)	224
Beta MDC (dpm)	487
# of Measurements/trial	8
# of trials	31

RESULTS

Critical Level (L_C) Test

Tables VI and VII (below) show the results from the Critical Level test. A discussion of these results is found in the next section.

Table VI. Low-Density L_c Test Results

	% False Positive Rate	
NNA #	**Alpha**	**Beta**
0	7.18%	6.67%
2	5.38%	5.64%
4	1.54%	5.38%
6	4.10%	4.10%

Table VII. High-Density L_c Test Results

	% False Positive Rate	
NNA #	**Alpha**	**Beta**
0	9.52%	4.25%
2	5.78%	3.91%
4	4.08%	4.76%
6	4.59%	0.00%
8	5.27%	5.10%

Minimum Detectable Concentration (MDC) Test

For each MDC test (Alpha and Beta), two metrics were calculated to assess the validity of the methods. The first metric is the overall false negative rate for each datum in the trial's set. For this metric, the averaged data that was below the theoretical critical level was tallied as a false negative. In theory, this metric should always be below 5% but fluctuations can be expected since counting statistics are very poor at very near background levels in a scanning scenario and trials

that fail entirely (all data below L_c) has the potential to skew the results. For this metric, the false negative rate is a measure of the inability to detect a source at the MDC for a single-point measurement.

The second metric is the false negative rate for the entire trial. For this metric, the trials that contain averaged values that all fall below the critical level are tallied as a false negative. More weight should be placed on this metric since most decommissioning surveys involve assessing the activity over a large area (e.g. 1 square meter) and the counting statistics over the whole trial are more reliable. For this metric, the false negative rate is a measure of the inability to detect a source at the MDC over a large area.

The false negative rate on a per-datum basis (metric 1) for the Alpha test was found to be 11.4%. The number of failed trials (metric 2) was 2 out of 30 (or 6.7%). It is clear that the failed trials are skewing the results of metric 1. The false negative rate on a per-datum basis (metric 1) for the Beta test was found to be 6.5%. The number of failed trials (metric 2) was 0 out of 31 (or 0%). Certainly, this method (or any method for that matter) shows more predictable results when counting statistics are more reliable. A summary of the results for these tests is presented in Table VIII.

Table VIII. MDC Test Results

			% False Negative Rate	
Test	**NNA #**	**# of Trials**	**per datum (metric 1)**	**per trial (metric 2)**
Alpha	4	30	11.40%	6.70%
Beta	5	30	6.50%	0%

DISCUSSION

Critical Level (L_C) Test

Overall, the empirical field test of the L_C calculation proved to be valid for the NNA method. All false positive rates for alpha and beta for both the low-density and high-density tests were in the neighborhood of 5% (with the exception of alpha with no NNA used). The larger false positive rate when no NNA is used is evidence of poor counting statistics playing a significant role in the unpredictability of the results.

Minimum Detectable Concentration (MDC) Test

Several factors in this experiment contributed to the uncertainty of the measurements taken. Since scanning measurements have short count times over large areas, counting statistics are often

poor and unpredictable. This is clearly observed in the alpha measurements since the number of counts in any measurement ranged between 0 and 4. While NNA has shown to "smooth" the data out and make it more statistically predictable, the counting error certainly still plays a major role in the unpredictability of the results.

The Multi-Agency Radiological Laboratory Analytical Protocols Manual (MARLAP)[3] Chapter 20, Appendix A "Low-Background Detection Issues" states that the Currie method of calculating the Lc may "produce a high rate of Type I errors" when the number of counts is low. The figure below (from MARLAP Figure 20.4) shows clearly that for the counting scenario in this experiment, false-positive rates of up to 25% can be expected. Note that P is the Type I (False Positive) error rate and $R_B t_S$ is the number of counts observed. In all of the trials run in this experiment, the number of counts in each measurement ranged from 0 to 4. It would follow that since the MDC is a function of the L_C that similar false negative error rates would be observed.

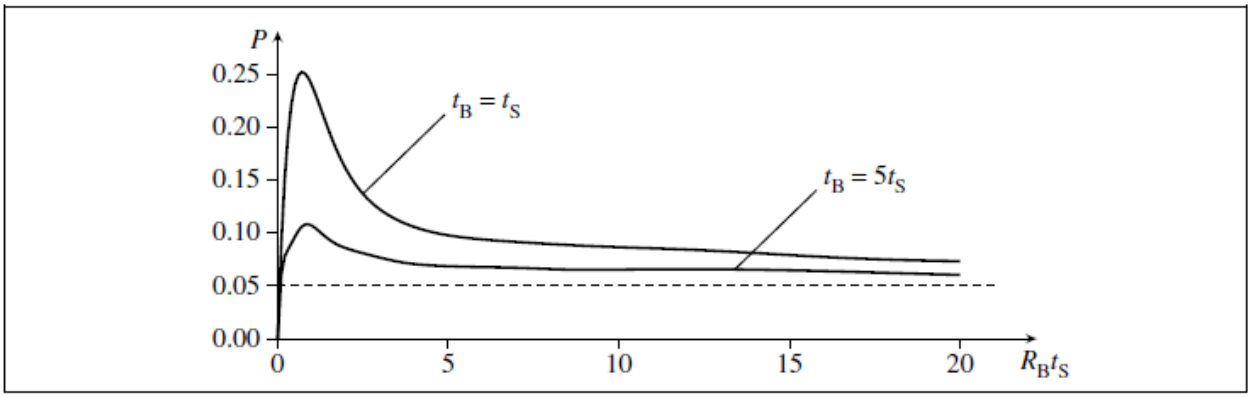

Figure 7. Excerpt from MARLAP Chapter 20, Appendix A [3], demonstrating true false positive rate for a Currie approach to determining the Lc

Furthermore, several circular sources may not have simulated a completely uniform and consistent source material. An ideal experiment would be to perform this test on a NIST-traceable large area source that is guaranteed homogeneous. However, in practice one typically encounters very heterogeneous residual contamination distribution, making such an "idealized" experiment academic. Furthermore, such a source would be unreasonably expensive to create and subsequently dispose of.

With that said, the results still show that the method of NNA does indeed decrease the L_C and MDC by a factor proportional to the number of neighbors. The results also show that the Currie method for determining L_C and MDA may not provide the most reliable results due to poor counting statistics. MARLAP[3] states that the Stapleton Approximation "appears to out-perform" the other more commonly used approaches when it comes to the true error rates.

CONCLUSIONS

The results obtained in this study clearly show that NNA is a valid approach and that it impacts the

measurement threshold and sensitivity in an advantageous way. This study also helped to identify areas where measurement settings can be adjusted to meet the needs of the survey work. As a result of this study, several important lessons were learned.

1. Alpha counting statistics are very poor (due to very low background rates) for most scanning counting scenarios. It is advantageous to slow the instrument down or increase the count time (or both) to obtain more statistically-reliable results and lower critical and detection levels.
2. It is very important to characterize a background as well as possible since this is the primary piece of information that is used to determine L_C and MDC. Concrete is particularly important to characterize since the NORM concentrations that are intrinsically present can vary widely from specimen to specimen.
3. The default source efficiency factors for concrete worked well during the L_C test.
4. The automated scanning-type survey instruments provide a very fast and efficient way to survey a large area. The density of the data obtained can be used in novel computerized algorithms (such as NNA) to significantly improve measurement sensitivity while expediting the survey and data review process.
5. Alternative approaches to determining the Critical Level and MDA (such as the Stapleton Approximation[3]) should be evaluated for use when dealing with alpha contamination and poor counting statistics.

This experiment offered opportunities to recognize future testing that could be done to further justify the method and to determine the optimum scanning settings for various counting scenarios. If time and funding become available, several improvements can be made to the experiment presented here.

1. To minimize the interference due to alpha and beta radiations in the incorrect channel, pure emitters (such as Sr-90 for Beta, and Pu-239 for alpha) can be used for the MDC testing.
2. The instrument being used should obtain a background for the area in which the experiment is being conducted prior to source creation so that the MDC for typical scan settings can be more precisely spiked onto the source matrix (create a check source that accounts for the test area background).
3. To minimize the uncertainties and interferences involved in using several circular sources, a large area mat source can be used as the testing matrix.
4. Once appropriate backgrounds, efficiencies, and source geometries are devised, the MDA should be chosen such that more neighbors can be used (5 or 6 as opposed to 4 which was used in this study). The more sub-samples taken from the raw data set, the more normal the resultant data set would be and the statistics would be more predictable.
5. An experiment to determine the scan setting bounds (highest scan speed, lowest count time, and the bounds for the # of neighbors for averaging) for the detection of alpha-emitting contamination is needed. It was clear in this experiment that the scan settings used and the check source spike activity yielded unpredictable results due to very poor counting statistics.

6. For some applications, the use of sources that were spiked with higher activity levels that "represented" nominal clearance levels for target radionuclides might be used.

REFERENCES

[1] Schierman, M. J.; Farr, C. P.; Wrubel, N.; Baker, K. R.; "Performance Characteristics of the 3-DISS Surface Contamination Monitor" http://www.ergoffice.com/Download%20Files/HPS_2007_3DISS_PerformanceCharacteristics.pdf
[2] U.S. EPA, Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM), http://www.epa.gov/rpdweb00/marssim/
[3] U.S. EPA, Multi-Agency Radiological Laboratory Analytical Protocols Manual (MARLAP), http://www.epa.gov/rpdweb00/marlap/links.html