SANDIA REPORT

SAND2016-6676
Unlimited Release
Printed July 2016

Classification of Background Suppression Profiles for Low Background RPM Data

Isaac R. Shokair and Rossitza Homan

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2016-6676
Unlimited Release
Printed July 2016

Classification of Background Suppression Profiles for Low Background RPM Data

Isaac R. Shokair[1] and Rossitza Homan[2]

[1]Sandia National Laboratories, Livermore, CA 94551
[2]Sandia National Laboratories, Albuquerque, NM 87185

Mail Stop 9408
P. O. Box 0969
Livermore, CA 94551-0969

Abstract

Suppression of the ambient gamma background radiation by traffic structure and cargo is a well-understood and studied effect for deployed radiation portal monitors (RPM). For effective analysis of measured RPM profiles with the objective of inferring the spatial characteristics of radiation sources, it is important to account for the effects of background suppression. In this report we analyze background suppression for a test dataset from vehicle RPMs at a sample port and estimate the distributions of suppression amplitudes and shapes. Cluster analysis of standardized and normalized profiles is used to obtain the dominant suppression shapes in the sample field data. We determine that a large fraction of non-alarm RPM occupancies are represented by a small number of suppression shapes. This fraction increases when the signal-to-noise ratio of an occupancy profile is improved by the addition of signals for multiple RPM detectors located at the same height. The calculated suppression shapes from RPM data can be used along with source models in the process of spatial profile analysis both in the field or offline. This background suppression analysis is an important step in improving the effectiveness of the RPM profile analysis methodology which is currently being investigated and may lead to methods that reduce the number of secondary inspections as well as to decision support tools that aid operators in evaluating RPM data that do not contain spectral information.

Acknowledgements

Support of this work by the Department of Energy though the Second Line of Defense Program is gratefully acknowledged. The authors would also like to thank Alex Enders of Oak Ridge National Laboratory for providing the sample dataset, reviewing the report, and making insightful recommendations.

Contents

Executive Summary

Radiation Portal Monitors (RPMs) are deployed at many foreign ports as part of the U.S. Department of Energy Second Line of Defense (SLD) program for detection of illicit radioactive materials with emphasis on special nuclear materials (SNM). These RPMs measure gamma and neutron radiation as traffic passes through the portal producing a time series of gamma and neutron gross count measurements. The time profile of the measured signal can be analyzed to provide insight on the spatial distribution of the radiation source: point-like, or distributed. In the absence of spectral information, this characterization can be very helpful.

One important effect that needs to be accounted for in spatial profile analysis is that of background suppression. For relatively weak sources the background suppression can be of similar magnitude to source amplitudes and can result in significant distortion of the measured profile. Thus fitting the profile to a model source without accounting for the suppression can result in a statistically invalid fit with a large residual. Therefore, to properly assess the capability of profile analysis for inferring the spatial properties of radiation sources, it is important to account for the effect of background suppression in the analysis. If the suppression shape is known, this can be done by including it in addition to the source model in the optimization process of profile analysis. In radiation profile analysis the suppression amplitude needs to be constrained to values that are consistent with the amplitude distributions observed in the measured data. If there are multiple suppression shapes, as would be expected at a port with multiple traffic lanes, vehicles types, and cargo loading configurations, then a process of hypothesis testing can be carried out where the different suppression shapes are tested sequentially.

In this report we consider background suppression for a sample test dataset containing measured vehicle RPM profiles from a sample port. Analysis is conducted to study the suppression shape and amplitude, although the main thrust of the report is to determine the dominant suppression shapes in the test dataset. Only non-alarm data is used for the analysis. For the suppression amplitude for an occupancy, we use fractional suppression defined by:

Fractional Suppression: $F_S = \dfrac{B_0 - \langle P \rangle}{B_0}$,

where B_0 is the measured background counts for a detector prior to the start of the occupancy and $\langle P \rangle$ is the average of the profile over the occupancy samples. In general the pre- and post-occupancy samples are not used in finding the average $\langle P \rangle$. This definition, which is by no means unique, results in a positive value for a reduced background. A negative value indicates possible presence of a radiation source in the measured data. As expected, the test dataset contains rather wide distributions of suppression amplitudes, roughly in the range 0.0 to 0.25 for the lower RPM detector pair, and 0.1 to 0.35 for the upper RPM detector pair. Actual distributions are shown in section 2 of the report.

In order to estimate the suppression shape for non-alarm RPM occupancies, and to enable numerical comparison with suppression shapes for other RPM measurements, we perform a

number of pre-processing steps of the measured profile. Once the suppression shapes for a dataset are estimated, cluster analysis is used to find the dominant shapes. For some of the analysis shown in this report signals from two detectors of the same RPM were added (either the lower pair or the upper pair) to enhance the signal-to-noise ratio of the profiles. The following pre-processing steps are performed for either single gamma detectors or for the summed profile for two gamma detector panels:

1- Profile filtering: occupancy profiles (including pre- and post-samples) are filtered using 2nd degree centered polynomial smoothing with 21 samples (Savitzky-Golay filtering) to reduce the effect of Poisson noise while preserving the suppression shape.

2- Background subtraction: the measured background nearest in time to the start of the occupancy is subtracted from the filtered profile.

3- Standardization of the number of profile samples: each measured profile is represented as a vector with dimension equal to the total number of time samples. For cluster analysis, vectors of the same dimensionality are required. Interpolation is used to stretch/compress the profiles to arrive at a fixed number of samples. The pre- and post-samples are not included in the stretching/compressing process. For the analysis in this report the standardized number of samples is set at 60 (not including pre- and post-samples). The results are not expected to be sensitive to this value.

4- Normalization: the profile vectors are normalized to unit length, where length is defined in the Euclidean sense: $L = \left(\sum_{i=1}^{N_0} x_i^2 \right)^{1/2}$, where N_0 is the standardized number of samples. This ensures that the distance between two vectors in the same direction (i.e. $\mathbf{V}_1 = c\,\mathbf{V}_2$ where c is a scalar) is zero regardless of the distance measure used. These two vectors represent the same suppression shape even if the amplitudes are different.

After subtraction of the background, standardization of dimensionality, and normalization, measured RPM profiles with the same suppression shape, but different amplitudes, result in the same normalized profile vector, provided the noise effects are negligible.

The normalized vectors are clustered using a hierarchical clustering algorithm utilizing a simple Euclidean distance metric. This analysis produces the dominant background suppression shapes in the test dataset. Figures 1 and 2 show the three most dominant suppression shapes for a sample traffic lane that was chosen randomly (referred to as Lane-A) for detectors 1+3 (lower) and 2+4 (upper) along with sample normalized profiles that belong to the cluster. About 29% and 68% of the profiles belong to these clusters for the summed upper and lower detectors, respectively. Even though these percentages are significant, we expect that more suppression profiles physically belong to these clusters, but are not identified due to shape distortion by noise effects. For the spatial profile analysis we anticipate using the first three most populated clusters in the process of hypothesis testing for the top detectors. For the lower detectors a larger number of shapes is likely required. This will be further investigated during implementation of the suppression shapes in the profile analysis software.

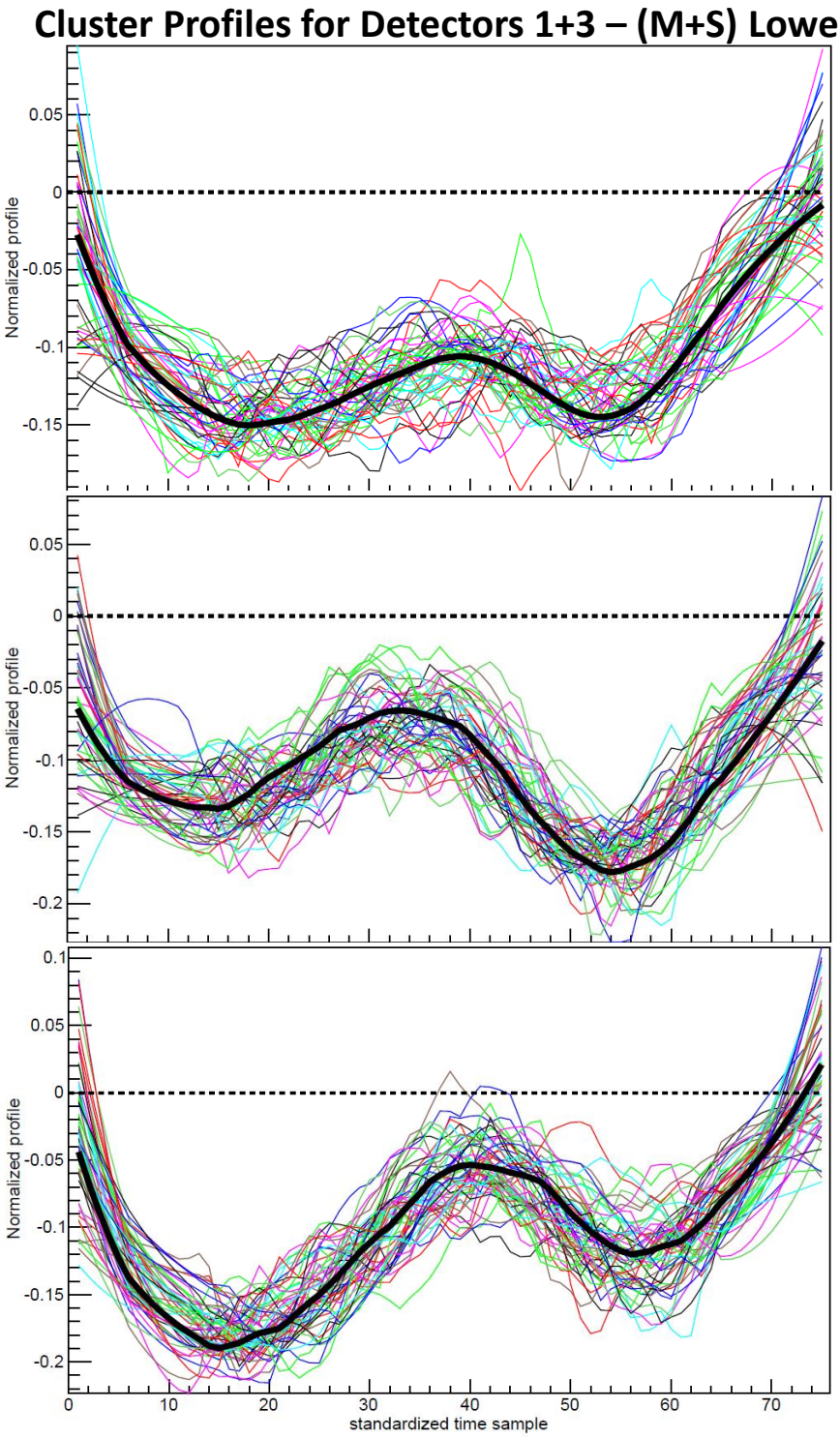

Figure 1. The three most-populated background suppression profile clusters for detectors 1+3 (lower) for Lane-A occupancies from the test dataset (cluster vectors shown as the dark curves). Also shown are randomly chosen normalized profiles that belong to each cluster. There are 1,495, 425, and 389 profiles belonging to these clusters respectively out of a total of 7,996 that are processed. These three clusters account for about 28.9% of the processed profiles.

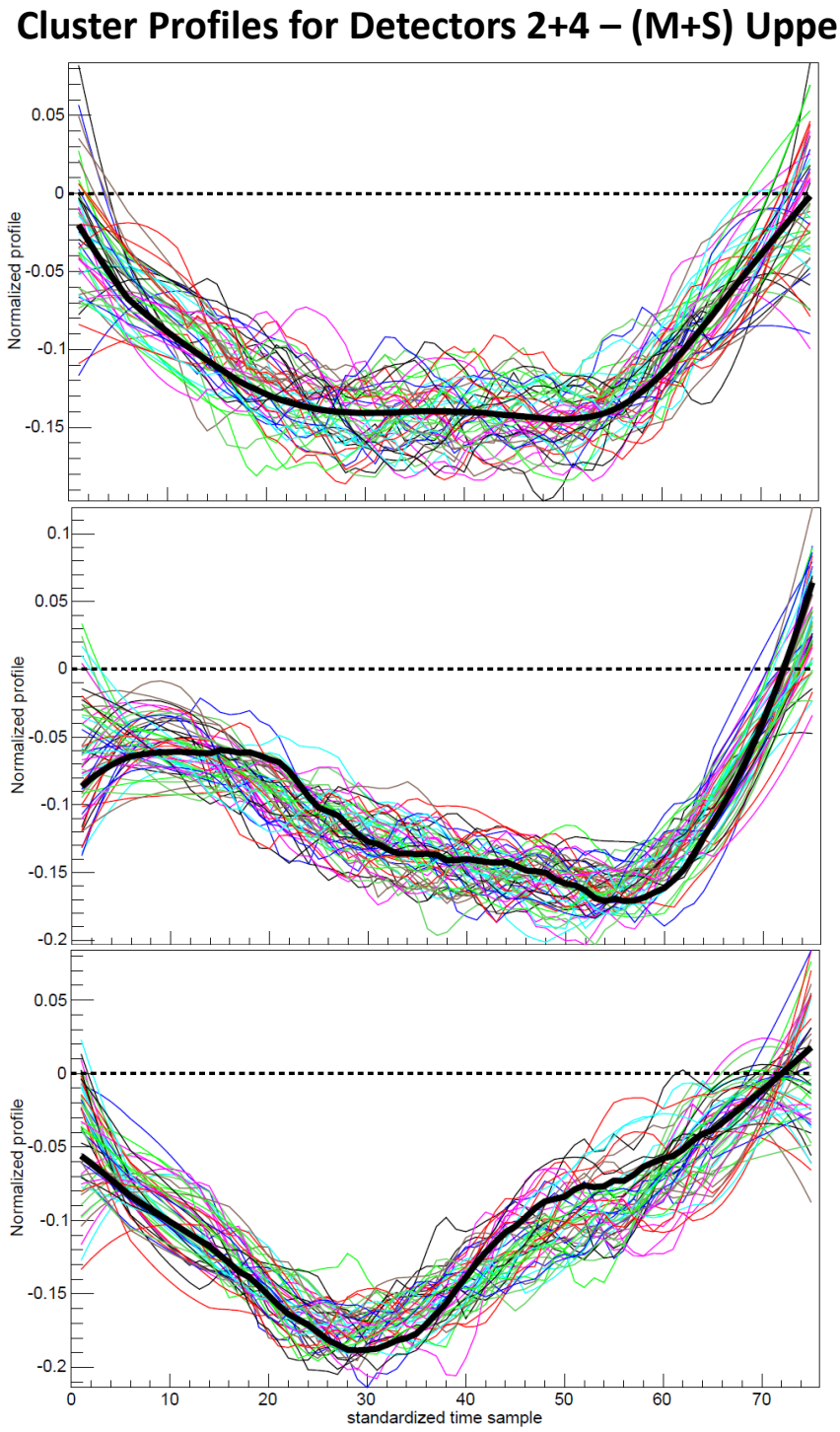

Figure 2. The three most-populated suppression profile clusters for detectors 2+4 (upper) for the same dataset. There are 5,951, 394, and 359 profiles belonging to these clusters respectively out of a total of 9,923 that are processed. These three clusters account for about 67.6% of the processed profiles. The noisy profiles are randomly chosen normalized profiles that belong to each cluster.

1. Introduction

Radiation Portal Monitors (RPMs) are deployed at many foreign ports as part of the U.S. Department of Energy Second Line of Defense (SLD) program for detection of illicit radioactive materials with emphasis on special nuclear materials (SNM). These RPMs measure gamma and neutron radiation as traffic passes through the portal producing a time series of gamma and neutron gross count measurements. The time profile of the measured signal can be analyzed to provide insight on the spatial distribution of the radiation source: point-like, or distributed. This report is a first step in assessing the potential benefit of automated characterization of RPM measured profiles to attempt to identify the nature of the radiation sources that result in primary alarms. The characterization methodology is based on previous work for detection of embedded radiation sources[(1)] and uses simple source models to fit the measured profiles and the resulting fit parameters to infer the source spatial distribution. For RPMs that do not output any spectral information only the spatial information can be exploited in profile analysis.

For relatively weak sources, the vehicle-induced background suppression[(2-5)] can be of similar magnitude to relevant source amplitudes and thus using a source model without accounting for background suppression effects can result in large statistical residuals that would indicate that this model is not adequate to fit the measured data. Therefore, to properly assess the value of profile characterization analysis as an automated method for inferring the spatial properties of radiation sources, it is important to account for the effect of background suppression in the analysis. If the suppression shape is known, this can be done by including the suppression shape in addition to the source model in the optimization process of profile analysis. The suppression amplitude is to be constrained to values that are consistent with the amplitude distributions in the measured data. These distributions are likely location (port/lane) dependent and are also impacted by collimators installed on select RPMs.

If the assumption is made that a large fraction of commerce traffic involves relatively similar vehicle structures and cargo loading profiles, or can at least be aggregated into sets of similar profiles, then one might conclude that background suppression shapes are similar for these cases and that there might be a small number of suppression shapes that are representative of most of the background suppression profiles for a given port lane. This is borne out in cluster analysis of suppression shapes for a shipping port[(5)] where two shapes accounted for the majority of suppression profiles of non-alarm data for a large dataset. When multiple suppression shapes are present for a location, the profile characterization analysis is envisioned to involve a process of hypothesis testing where each of the suppression shapes is tested sequentially. The hypothesis with the most statistically meaningful fit or resulting in a threat-like source profile would be used for assessment of the RPM measurement. It is also possible that different suppression shapes are correlated with vehicle and cargo configurations, and when available, this information can be used to identify the proper suppression shape.

In this report we consider background suppression for a sample test dataset and use cluster analysis, following the same methodology used in reference (5), to extract the dominant suppression shapes that are present in the data. For this analysis it is important to consider data that doesn't contain any radiation sources and thus we only use non-alarm data that is subject to

an additional amplitude constraint discussed in the next section. Additional constraints on suppression amplitude are also imposed in an attempt to use only data that would result in usable signal-to-noise ratio (SNR).

In the next section of the report we consider the distributions of suppression amplitudes for a test dataset and discuss the SNR issues relevant to suppression shapes. In the following section hierarchical cluster analysis is used to determine the dominant suppression shapes[5]. The measured profiles are pre-processed using the following steps: filtering, subtraction of the background, standardizing the profiles to a fixed number of time samples, and normalization to unit length. These steps are necessary to extract the suppression shapes from large numbers of occupancies with different numbers of samples and suppression amplitudes. The cluster analysis is applied to the processed profiles and the results indicate that a significant fraction of suppression shapes are contained in a small number of clusters.

Because of signal-to-noise limitations, many of the suppression profiles are dominated by noise and no clear shapes can be extracted. To improve this situation we repeat the same analysis using the summed signals from two RPM detectors (left + right). The results of this analysis show a significantly larger fraction of the suppression shapes belonging to the dominant clusters. This is a significant improvement, but doesn't completely alleviate the noise effects which are large, especially when the background counts are low.

2. Distributions of Suppression Amplitudes

The measurements used to quantify background suppression are non-alarm RPM measurements from a sample test dataset from the stream-of-commerce at a specific location. The gross counts time profiles are generated as traffic passes through the RPM for time segments of 0.2 seconds extent. Each profile also includes a number of pre- and post-samples that provide the counts before a vehicle enters and after it exits the RPM. For the test dataset there are typically 5 pre-samples and 10 post-samples.

Definition of Fractional Suppression:

There are several definitions that can be used to quantify background suppression amplitudes and all are equally valid. In this report we use the average fractional suppression that was used in previous analysis[3,5]. This is defined for a detector and occupancy by:

Fractional Suppression: $F_S = \dfrac{B_0 - \langle P \rangle}{B_0}$, (1)

where B_0 is the measured background counts for the detector prior to the start of the occupancy and $\langle P \rangle$ is the average of the measured profile counts over the occupancy samples. In general, the pre- and post-samples are not used in finding the average $\langle P \rangle$. Note that the definition in Eq. (1) results in a positive value for true background suppression and a negative value if the average signal is larger than the background, possibly an indication that a source is present. For cases where a background measurement is not available, the average counts over post-samples are used to estimate B_0. For the test dataset considered in this report only 10 occupancies out of 125,903 that were processed for all data lanes (0.008%) had no reported background values. This fraction is so small that such occupancies can be included or excluded without affecting the results.

Constraints on Data Used for Analysis:

In addition to using non-alarm data for estimation of suppression amplitude distributions, we also limit the data used to have number of occupancy samples in the range: $30 \le N_{samples} \le 100$. This constraint is imposed to exclude extreme cases of short vehicles or traffic moving too fast through the RPM (the lower limit) or traffic stopping at the RPM (the upper limit). It is also useful to exclude occupancies with low numbers of samples because of SNR considerations discussed later in this report. Histograms of the number of occupancy samples (Figure 4 for example) demonstrate that imposing these constraints results in a reasonable range for the data included in the analysis but it does exclude a large fraction of the occupancies.

The data used for the suppression analysis is also limited to profiles that have the maximum of the filtered profile less than 1*sigma above background, where sigma is the standard deviation of the background or $\sqrt{B_0}$. This ensures that the probability of the presence of a radiation source in the measured data is low. Note that a significant fraction of non-alarm profiles, as identified in the data, are rejected because of this constraint and some of those could be free of radiation sources, but are rejected due to large noise effects. Figure 3 shows percentages of occupancy categories for data in all RPM lanes in the test dataset.

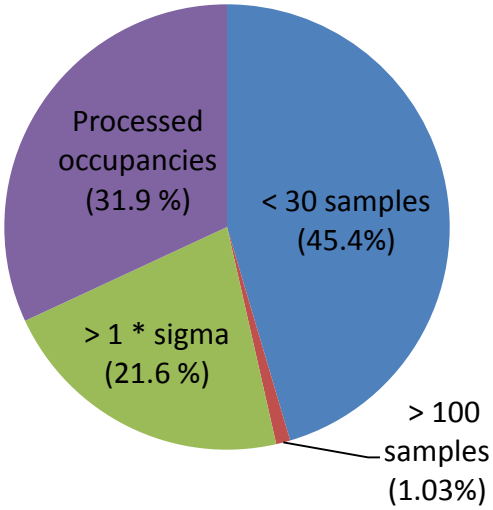

Figure 3. Percentages of different categories of occupancies for the test dataset with a total of 394,417 occupancies. The n*sigma limit is applied to all detectors for this chart.

Histograms of Fractional Suppression

Fractional suppression as defined by Eq. (1) was calculated for non-alarm occupancies for the test dataset, subject to the above constraints. Histograms of the number of occupancy samples are shown in Figure 4 for one sample lane, which will be denoted as Lane-A, and in Figure 6 for all lanes in the test dataset. These histograms show that a significant number of occupancies have less than 30 samples (note: cases with 0 samples are not real occupancies, but appear to be RPM triggered). These occupancies are not included in the analysis as discussed above. Histograms or distributions of the fractional suppression for all detectors are shown in Figure 5 for Lane-A and in Figure 7 for all lanes in the dataset.

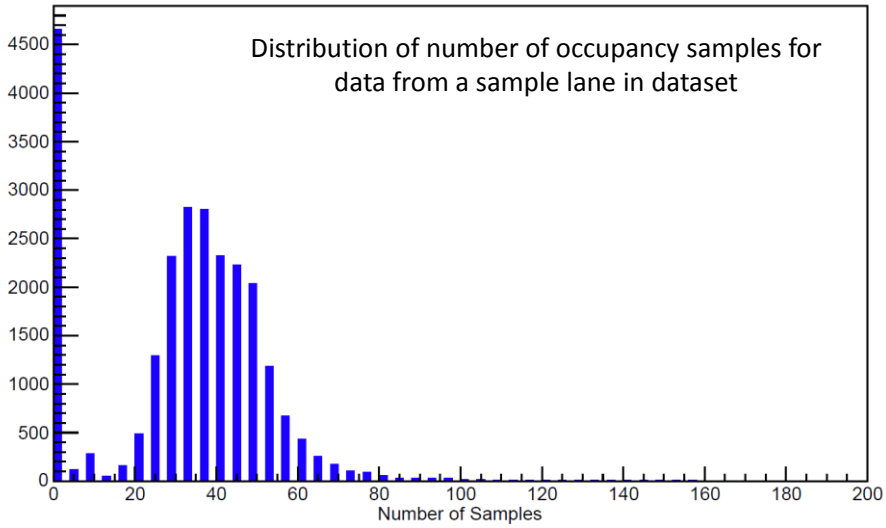

Figure 4. Histogram of the number of occupancy samples (not including pre- or post-samples) for one sample lane (Lane-A). For this lane there is a large number of null occupancies without samples. These appear to be self-triggered by the RPM.

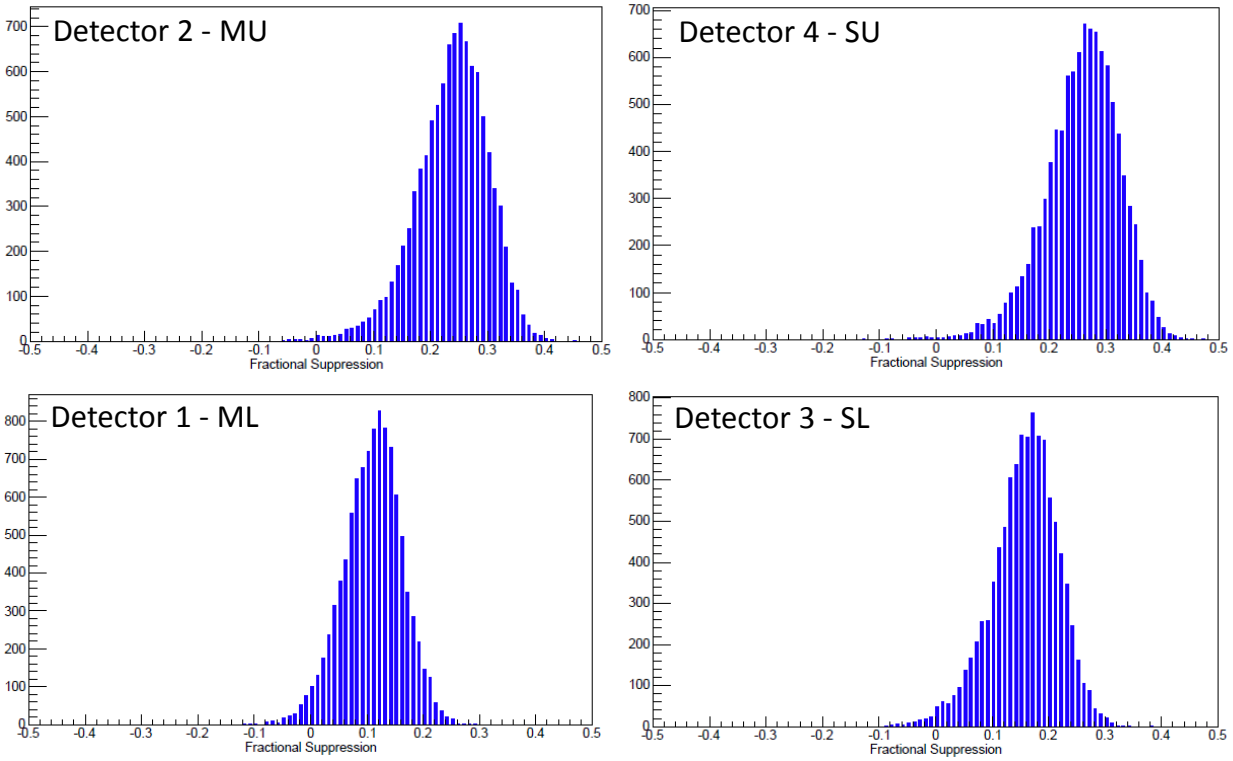

Figure 5. Distributions of fractional suppression for non-alarm traffic for Lane-A from the test dataset. Detector-1 is master-lower (ML), detector-2 is master-upper (MU), detector-3 is slave-lower (SL), and detector-4 is slave-upper (SU). Note that when the fractional suppression is below zero it implies that the average counts for the occupancy is larger than the ambient background. For this data, 25,409 occupancies are read from RPM data files, and 10,088 are processed for the histograms (7,985 had < 30 samples, 301 had > 100 samples, and 6,675 exceeded the 1 * sigma limit for at least one detector).

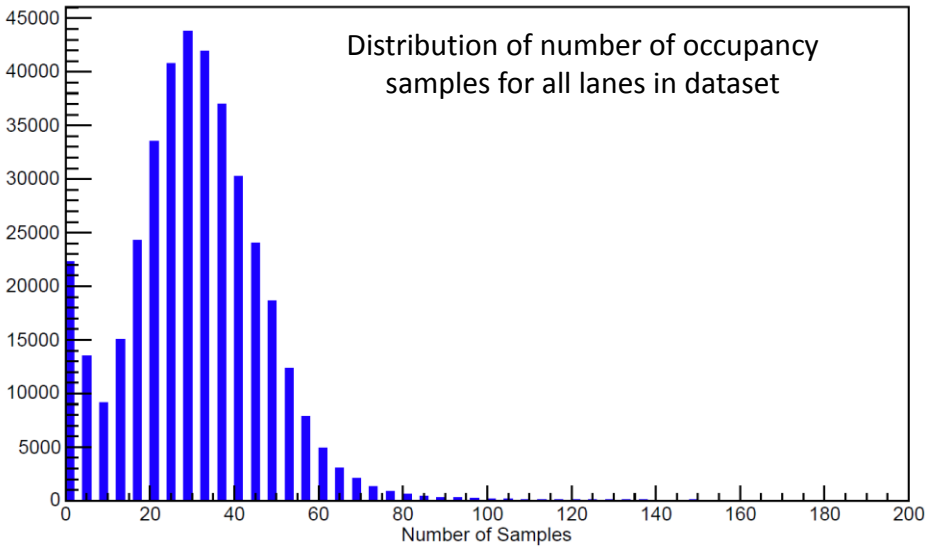

Figure 6. Histogram of the number of occupancy samples (not including pre- or post-samples) for all lanes in the test dataset. The distribution of samples is similar but not identical to that in Figure 4.

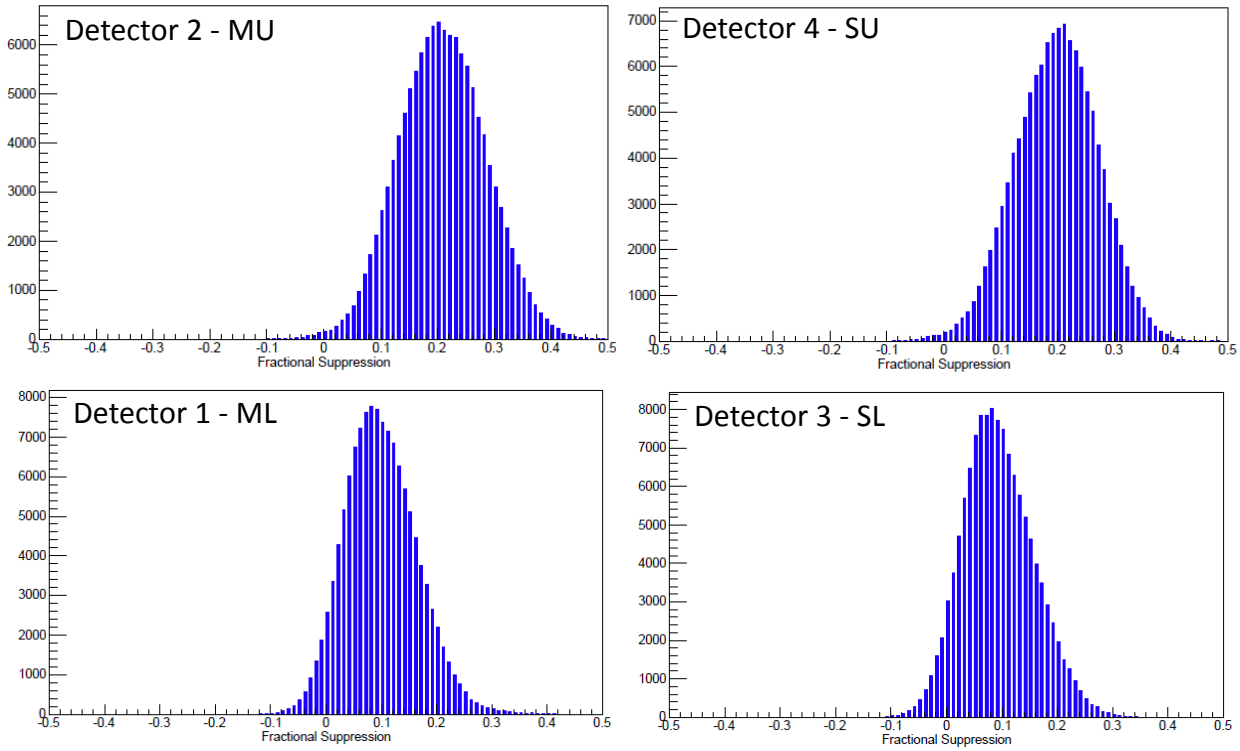

Figure 7. Distributions of fractional suppression for non-alarm traffic for all lanes in the test dataset. These histograms represent data from 125,903 occupancies that are processed out of a total of 394,417 that are read from RPM data files. Out of this data, 179,129 had less than 30 samples, 4,072 had more than 100 samples, and 85,313 exceeded the specified maximum signal limit of 1 * sigma for at least one of the detectors.

From these figures the distributions of suppression amplitudes appear to be similar for all lanes and the upper detectors show higher suppression than the lower ones. The mean fractional suppression is about 10% for the lower panels and about 20% for the upper panels. The mean values for Lane-A data are somewhat higher as seen in Figure 5.

For the purpose of checking any correlation between background suppression and the number of profile samples, we considered fractional suppression for cases with less than 30 samples to test whether or not these cases have similar distributions. In Figure 8 the distributions of fractional suppression are shown for occupancies with number of samples restricted to the range 20-30 for the same dataset used in Figure 7. These distributions appear very similar to those in Figure 7 and thus it seems there is no correlation between suppression amplitudes and number of occupancy samples for this data. It is possible that such occupancies represent shorter vehicles. This information can be investigated by considering the distribution of speeds along with numbers of samples.

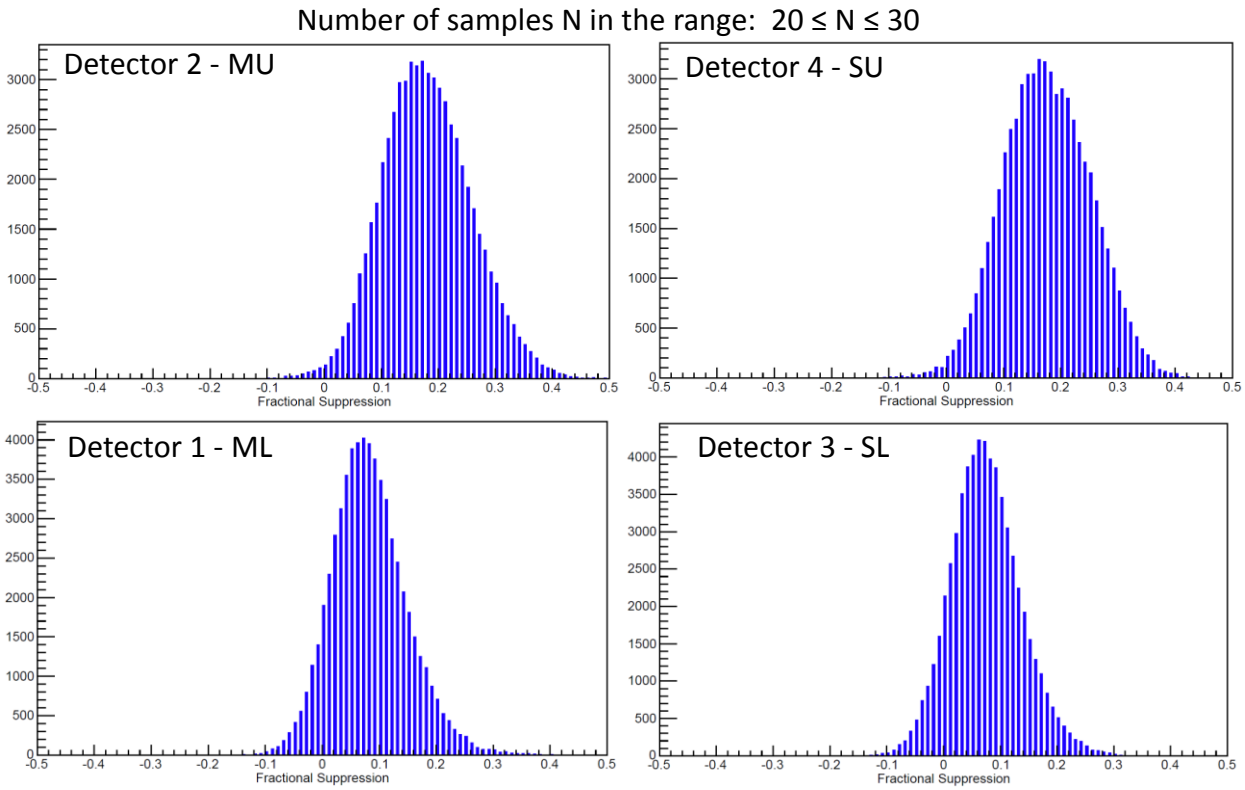

Figure 8. Distributions of fractional suppression for non-alarm traffic for all lanes in the test dataset. These histograms are for 62,175 occupancies with **number of samples in the range 20-30**. Occupancies exceeding the maximum signal limit of 1 * sigma for any individual detector are not used in the histograms.

Width of Suppression Distributions and Effect of Noise

The spread in fractional suppression, or the widths of the distributions in Figures 5 and 7, are due to physical differences between occupancies and the effect of Poisson noise. The contribution from Poisson noise can be estimated based on Eq. (1) by:

$$\sigma_f^2 = \frac{1}{B_0^2}\left[\sigma_{<p>}^2 + \sigma_{<B_0>}^2 \frac{\langle P \rangle^2}{B_0^2}\right] \qquad (2)$$

where σ_f^2 is the variance of the fractional suppression, $\sigma_{<p>}^2$ is the variance of the profile average counts, and $\sigma_{<B_0>}^2$ is the background variance. Since the averaging time for the background measurements in the data is not known, the background variance is estimated directly from the data. Figure 9 shows background variation for a sample data file and estimates the noise in background at about 0.4 counts for a time sample, that is, $\sigma_{<B_0>} \approx 0.4$. To estimate $\sigma_{<p>}$ consider the histogram of background values read from the data files in Figure 10. The mean background (used to approximate the occupancy counts) is 20.3 counts per 0.2 second interval. At the minimum number of 30 samples used in the analysis, the expected noise for the

average occupancy counts is: $\sigma_{<p>} \approx \sqrt{(20.3)(30)}/30 = 0.82$ counts. Using these values, the mean deviation in fractional suppression given by Eq. (2) is: $\sigma_f \approx \frac{1}{B_0}\sqrt{\sigma^2_{<p>} + \sigma^2_{B_0}\frac{\langle P\rangle^2}{B_0^2}}$ $\approx 0.91/20.3$ or about 0.04 which is less than the minimum half-width of the distributions for the different detectors. This implies that the widths of the fractional suppression distributions are dominated by physical differences between occupancies. Nevertheless, noise is a significant contributor to the observed spread.

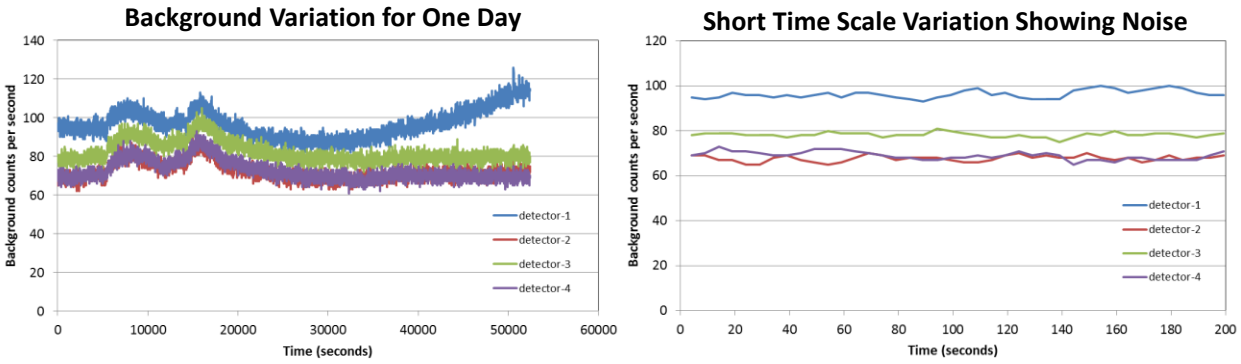

Figure 9. Background variation for one data file from Lane-A used above. The left plot shows variation throughout the day. The right plot shows short time scale variation due to Poisson noise. For detector-1 this results in a standard deviation of 2.0 counts per second which is equivalent to 0.4 counts for a 0.2 second time sample.

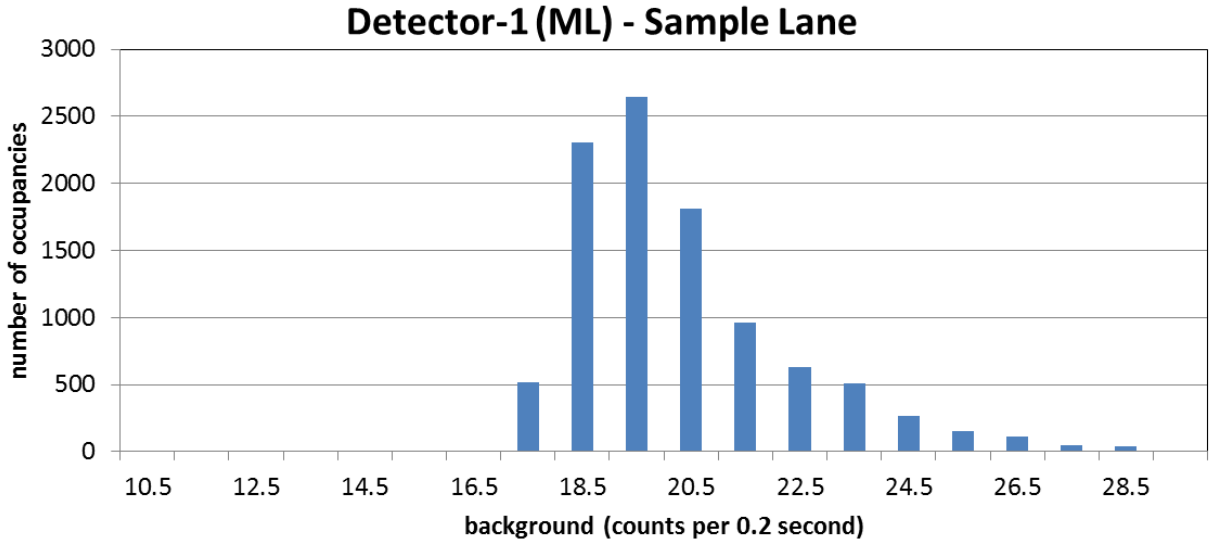

Figure 10. Histogram of background values for detector 1 for the data used in Figure 5 (Lane-A only). The background is in units of counts per time sample (each sample is 0.2 seconds).

3. Cluster Analysis of Background Suppression Profiles

In this section we investigate the shapes of the suppression profiles observed in the data and attempt to assess the likelihood that these shapes are dominated by a small number of shapes that can be used in profile analysis. We use a hierarchical clustering method that was developed in a previous study of background suppression[5]. The objective of this analysis is to analyze a large number of measured profiles, for occupancies without radiation sources, separate out the dependence on suppression amplitude and isolate the suppression profile shapes that are present in the test dataset.

The details of the previous cluster analysis work are given in reference (5). Here we briefly outline the required steps for data preparation and the subsequent cluster analysis. Only occupancies that meet all the constraints outlined in the previous section are used for cluster analysis.

Data Preparation:

1- Profile filtering: occupancy profiles (including pre- and post-samples) are filtered using 2nd degree centered polynomial smoothing with 21 samples (Savitzky-Golay filtering) to reduce the effect of Poisson noise while preserving the suppression shape. Using 2nd degree polynomial filtering is likely better at preserving the suppression shape than a simple rolling sum average which is equivalent to 0-degree polynomial filtering.

2- Background subtraction: the measured background nearest in time to the start of the occupancy is subtracted from the filtered profile.

3- Standardization of the number of profile samples: each profile is represented as a vector with dimension equal to the total number of time samples. For cluster analysis, vectors of the same dimensionality are required. Interpolation is used to stretch/compress the profiles to have a fixed number of samples. The pre- and post-samples are not included in the stretching/compressing process. For the analysis in this report the standardized number of samples is set at 60 (not including pre- and post-samples). The results are not expected to be sensitive to this choice.

4- Normalization: the profile vectors are normalized to unit length where length is defined in the Euclidean sense: $L = \left(\sum_{i=1}^{N_0} x_i^2 \right)^{1/2}$, where N_0 is the standardized number of samples. This is done so that the distance between vectors in the same direction (i.e. $\mathbf{V}_1 = c\,\mathbf{V}_2$ where c is a scalar) is zero regardless of the distance measure used. These two vectors represent the same suppression shape even when the amplitudes are different.

After subtraction of the background, standardization of dimensionality, and normalization, profiles with the same suppression shape, but different amplitudes result in the same normalized profile vector, provided noise effects are negligible. However, it is expected that noise effects are large for the test dataset due to the low background counts.

Cluster Analysis:

1- Defining clusters: Use a relatively large subset of the pre-processed suppression profiles to define a set of cluster vectors using hierarchical clustering. This is done by calculation of a distance matrix for all the vectors and then combining the nearest two vectors into one. The distance matrix is updated and the process is repeated until a maximum distance criterion is met, that is, no further combining of vectors is executed when the distance between the nearest two vectors exceeds a specified value D_0. At this point the remaining vectors define the clusters. For the analysis in this report, the first 1,000 profiles are used for this step (assumed representative of the whole dataset). The distance limit is set as $D_0 = 0.28$. This distance limit is relatively large, and had to be used because of large noise effects. Note that the maximum possible distance is $\sqrt{2}$ since the vectors are normalized to unit length.

2- Clustering of all profiles: After defining the cluster vectors in the previous step, for each pre-processed profile in the full dataset, we calculate the distance to all clusters and assign the profile to the nearest cluster. If the minimum distance is larger than D_0, the profile is said to be not clustered. It should be noted here that the distance between a processed profile and multiple cluster vectors can be less than D_0. Thus, even though we assign the profile to the cluster resulting in the minimum distance, it is possible that the profile belongs to other clusters as well.

Effect of Speed:

In this analysis we assume that the speed is nearly constant as a vehicle traverses the RPM and no correction for changing speed is made to the suppression profiles. The suppression profiles should be roughly independent of the speed (for reasonable speeds) for the same vehicle since they are slowly varying profiles in most cases aside from noise effects. This is in contrast to a point source (especially if the detector is collimated) where distance and angle between source and detector can change significantly during a time sample. Both the source profile shape and amplitude depend on the speed for such cases.

In the appendix of this report we show histograms of the reported speeds for the analyzed data from the test dataset, and when multiple speeds are reported, histograms of the speed ratios are shown. These histograms indicate that traffic speeds are nearly constant within the RPMs for most of the analyzed data. However, as is discussed in the appendix, there are issues with some of the speed measurements. If speed measurements were reliable and accurate, it would be beneficial to apply and test the effect of a correction due to changing speed. However, information regarding speed reliability (e.g. ground truth) is not currently available for the test dataset and thus, no speed correction is made in our current analysis.

Noise Considerations:

To meet the objectives of this analysis it is very difficult to overcome noise effects because of the low SNR of the background suppression profiles, especially for low suppression amplitudes. If we consider a time sample and a background value for a single detector of 20 counts per 0.2 second sample, for fractional suppression of 0.1, the average suppression is of order 2 counts, whereas the noise level is roughly $\sqrt{20}$ or about 4.5 counts. Thus the SNR of the suppression for a time sample is < 1 which is extremely low. Filtering of the profiles helps, but it remains very

challenging to extract accurate suppression profiles for the data being considered. Larger background values improve the SNR and can be attained for the present data by adding detector signals. However, one has to be careful to only add detectors that are expected to have similar suppression profile shapes. We expect that the left and right detector signals can be added, but the top and bottom detectors are not likely to have the same suppression profile shapes. Note that the profile normalization described above scales the signal and noise equally and therefore doesn't affect the SNR. However, changing the number of samples to standardize the dimensionality using interpolation will have an effect on noise.

Another constraint that is applied to help reduce the effect of noise is to limit the profiles used for clustering to those whose suppression amplitude is larger than a user-specified value. For the analysis in this report the fractional suppression is limited to the range of 0.1 to 0.5 for a profile to be used in clustering. The upper limit is rarely encountered and thus has no effect on the results.

Distance Calculation:

The Euclidean distance is used as the metric for distance between two normalized vectors. For two vectors $\mathbf{V}_i$ and $\mathbf{V}_j$ this distance is:

$$D_{ij} = \sqrt{\sum_{k=1}^{N_0} \left(\mathbf{V}_i(k) - \mathbf{V}_j(k)\right)^2} \qquad (2)$$

A variance-weighted distance, such as the Mahalanobis distance, can be also be used. In this analysis this complication in not necessary since all the components of a profile vector have similar variance. Note that normalization of the profile vectors results in different variances for different suppression amplitudes, but this is not expected to have any significant impact on the overall clustering results.

Cluster Analysis Results:

Several clustering analyses are conducted for the test dataset considered in section 2. Only profiles that met all the constraints outlined previously and have suppression amplitudes in the range 0.1 to 0.5 are used. A distance limit of 0.28 was used for forming the cluster vectors.

Before showing results of the cluster analysis, to illustrate the effect of noise on the normalized profiles, Figure 11 shows the first two processed profiles for data from Lane-A. Consistent with the discussion of noise effects above, these plots also show that the filtered profiles can be significantly affected by count noise resulting in distortion of the normalized suppression profiles. This limits the effectiveness of cluster analysis for many of the profiles, especially for those with low suppression amplitudes where the effect of noise is dominant. Note that in the limit of zero suppression, after subtraction of the background, the resulting normalized profile is purely due to noise. For finite suppression and when a large number of profiles are clustered to form a cluster profile, we expect that the effect of noise will be reduced due to averaging and this cluster profile closely represents the true suppression profile for the clustered data.

Below we discuss results for several cases to show differences between suppression profiles for different detectors, the effects of adding multiple detectors, and results for a large dataset. The following cases are considered:

1- Lane-A for detector 1 (M) Lower: 9% of occupancies belong to the top 3 clusters

2- Lane-A for detectors 1+3 (M+S) Lower: 29% of occupancies belong to the top 3 clusters

3- Lane-A for detectors 2+4 (M+S) Upper: 68% of occupancies belong to the top 3 clusters

4- All lanes for detectors 1+3 (M+S) Lower: 21% of occupancies belong to the top 3 clusters

5- All lanes for detectors 2+4 (M+S) Upper: 43% of occupancies belong to the top 3 clusters

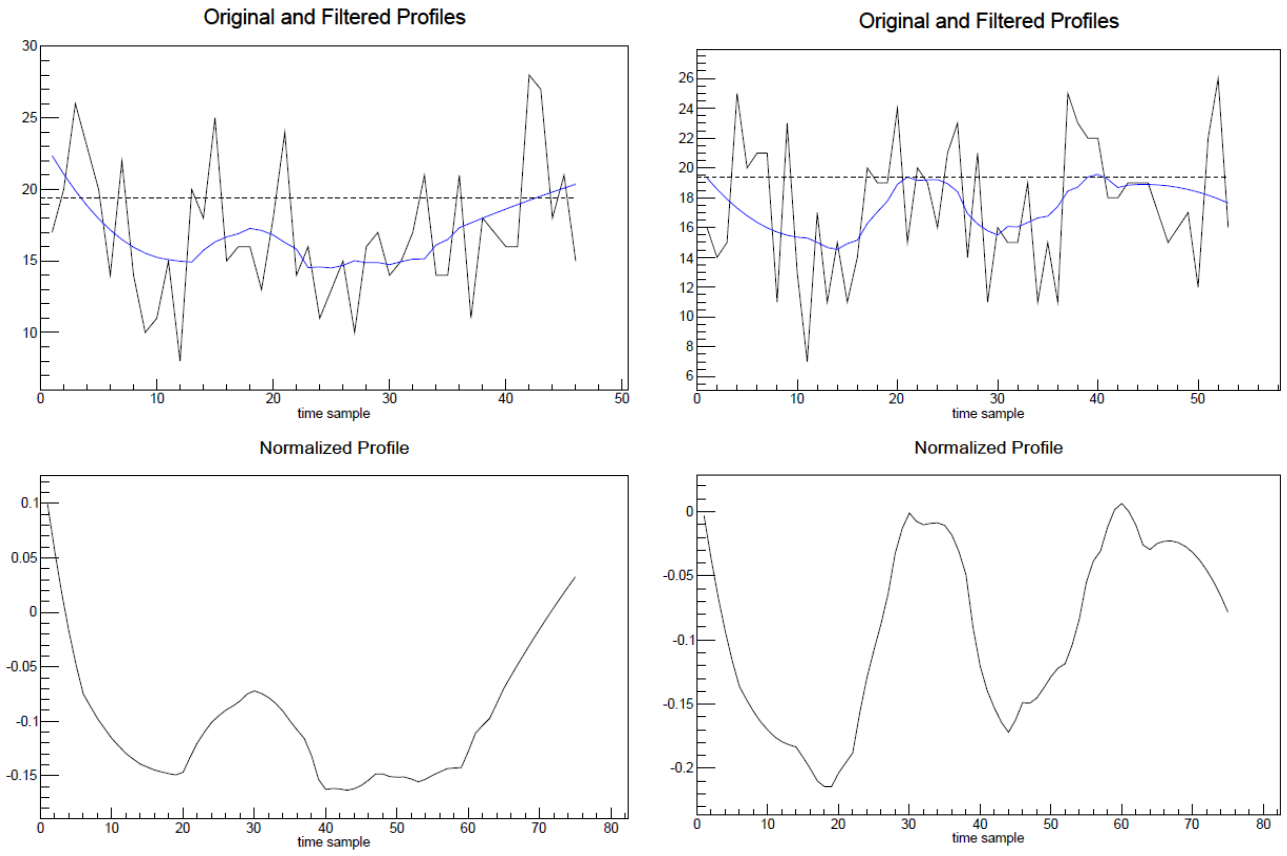

Figure 11. First two profiles that are processed from Lane-A data for detector 1. The top plots shows the original and filtered profiles along with the background value (dashed line). The bottom plots for the normalized profiles show the shapes of the suppression profiles.

Case 1. Lane-A – Detector 1:

Figure 12 shows the three most-populated clusters along with sample normalized profiles that are closest in distance to these cluster profiles. The numbers of profiles closest to these clusters and within a Euclidian distance of 0.28 from the overall cluster vectors are 326, 128, and 110 respectively out of a total of 6,210 occupancies that are processed. This amounts to about 9% of processed profiles for the top three clusters. Note that the test dataset for the selected lane includes a total of 25,049 occupancies read from RPM data files. Of these 7,985 had less than 30 samples, 301 had more than 100 samples, 6,675 had signal amplitudes larger than 1*sigma **for any of the four detectors**, and 3,878 are outside the fractional background suppression specified limits of 0.1 to 0.5.

It should be noted that a profile is assigned to the cluster that results in the minimum distance metric between the profile and all the cluster vectors. This distance has to be less than the specified maximum distance ($D_0 = 0.28$ was used). It is therefore possible for a profile to be close to halfway between two clusters such that the distance from the profile to the two clusters is less than D_0. For such cases one can reasonably assign the profile to multiple clusters. If this process is followed the number of profiles for the most-populated clusters will increase.

Case 2. Lane-A – Detectors 1+3:

Figure 13 shows the three most-populated clusters along with sample normalized profiles. The numbers of profiles closest to these clusters and within a Euclidian distance of 0.28 from the overall cluster vectors are 1495, 425, and 389 out of a total of 7,996 occupancies that are processed. When the signals from detectors 1 and 3 are added, only 2,092 occupancies are outside the specified fractional suppression limits. Note that the multiple of sigma limit is checked before the addition of detector signals and thus the same number, as for case 1, are rejected because of this condition.

Even though the fraction of occupancies belonging to the top three clusters is still relatively small (29%), that is still a large increase above the 9% obtained in the previous case when only detector 1 was used. This is consistent with our hypothesis that the low SNR of the suppression profiles results in distortion of these profiles and thus our expectation is that the top three clusters represent a larger fraction than 29% of the true suppression profiles in the data.

Case 3. Lane-A – Detectors 2+4:

Detectors 2 and 4 have larger fractional suppression than 1 and 3 based on the distributions in Figures 5 and 7. Because of this the effect of noise-caused distortion is expected to be smaller for these two detectors. Figure 14 shows the three most-populated clusters along with sample normalized profiles. These clusters have 5,951, 394, and 359 profiles respectively out of a total of 9,923 that are processed. For this case only 165 profiles are rejected for being outside the specified suppression limit. Note that the numbers of rejected profiles for the other conditions are identical to those for cases 1 and 2 above. As expected the higher SNR of the suppression profiles results in about 68% of occupancies belonging to the three most-populated clusters.

Case 4. All lanes – Detectors 1+3:

This is similar to case 2 except the analysis is done for all lanes in the test dataset. The profiles for the three most-populated clusters are similar to those obtained for Lane-A and are shown in Figure 15. Note that one difference from case 2 is that the clusters are derived from the first 1,000 processed profiles which in this case came from other lanes. For this case the three most-populated clusters had 5,062, 4,928, and 2,566 profiles respectively out of a total of 59,067 that are processed. For this data there are 394,417 occupancies read from RPM data files. Out of these 179,129 had less than 30 samples, 4,072 had more than 100 samples, 85,313 exceeded the sigma limit, and 66,836 are outside the specified suppression bounds.

Case 5. All lanes – Detectors 2+4:

This is similar to case 3 except the analysis is done for all lanes in the test dataset. Again for this case the profiles for the three most-populated clusters are similar to those obtained for Lane-A. These profiles are shown in Figure 16. For this case the three most-populated clusters had 31,425, 11,914, and 7,992 profiles respectively out of 118,976 that are processed.

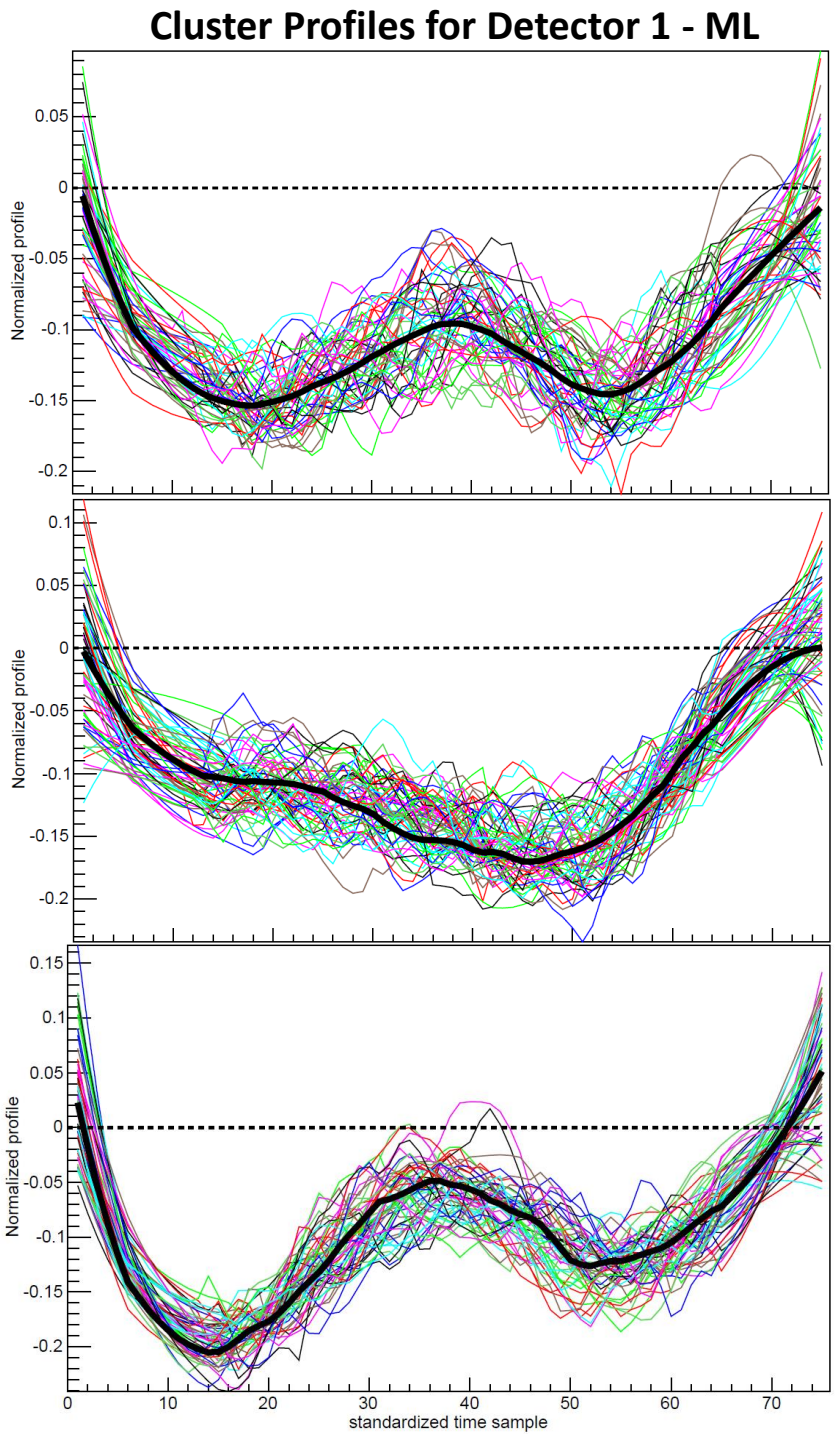

Figure 12. The three most-populated suppression profile clusters for detector 1 for data in Lane-A (shown as the dark curves). Also shown are randomly chosen normalized profiles that belong to each cluster. There are 326, 128, and 110 profiles belonging to these clusters respectively out of a total of 6,210 that are processed.

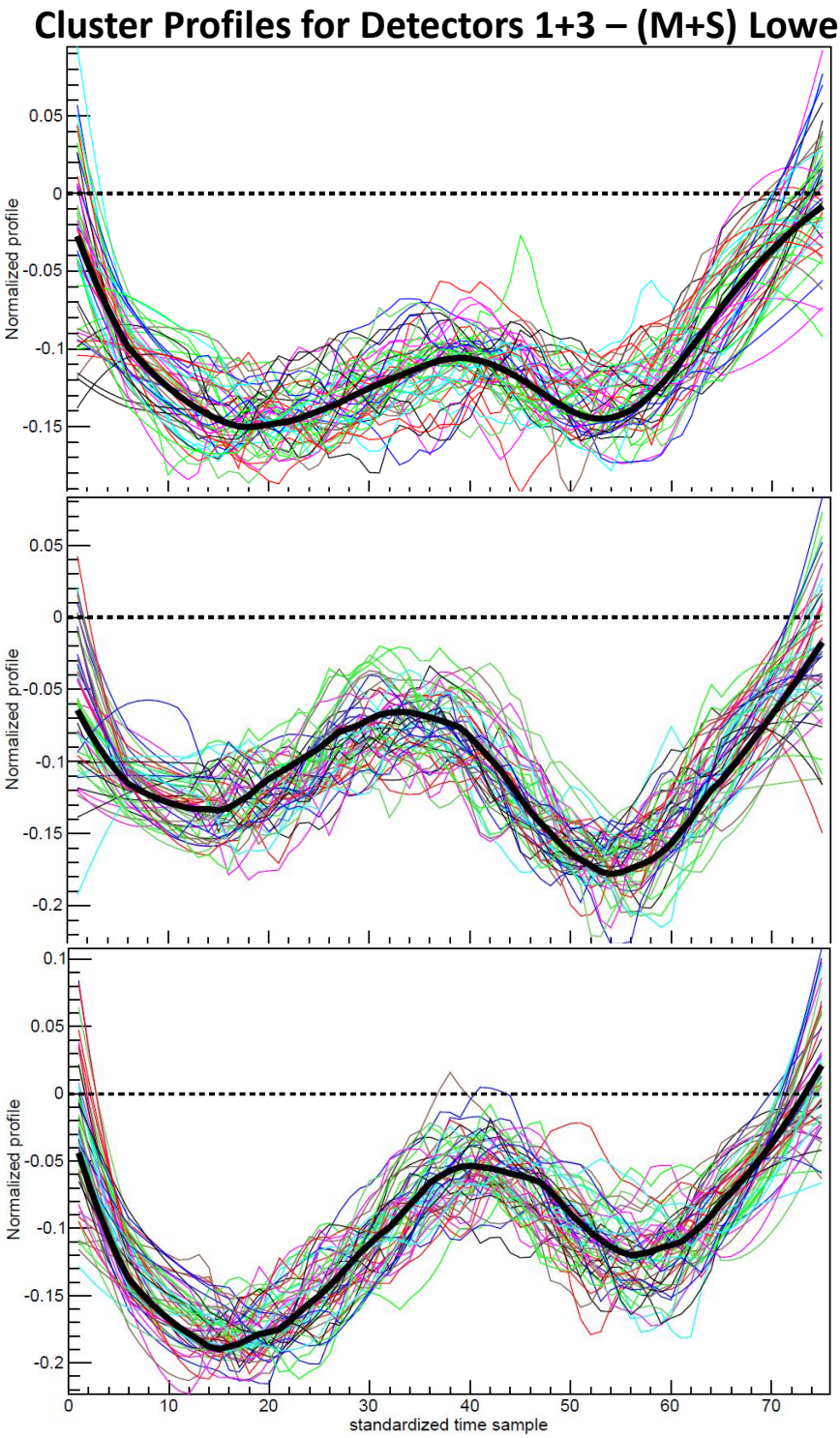

Figure 13. The three most-populated suppression profile clusters for detectors 1+3 for Lane-A occupancies. There are 1,495, 425, and 389 profiles belonging to these clusters respectively out of a total of 7,996 that are processed. These three clusters account for about 28.9% of the processed profiles. The noisy profiles are randomly chosen normalized profiles that belong to each cluster.

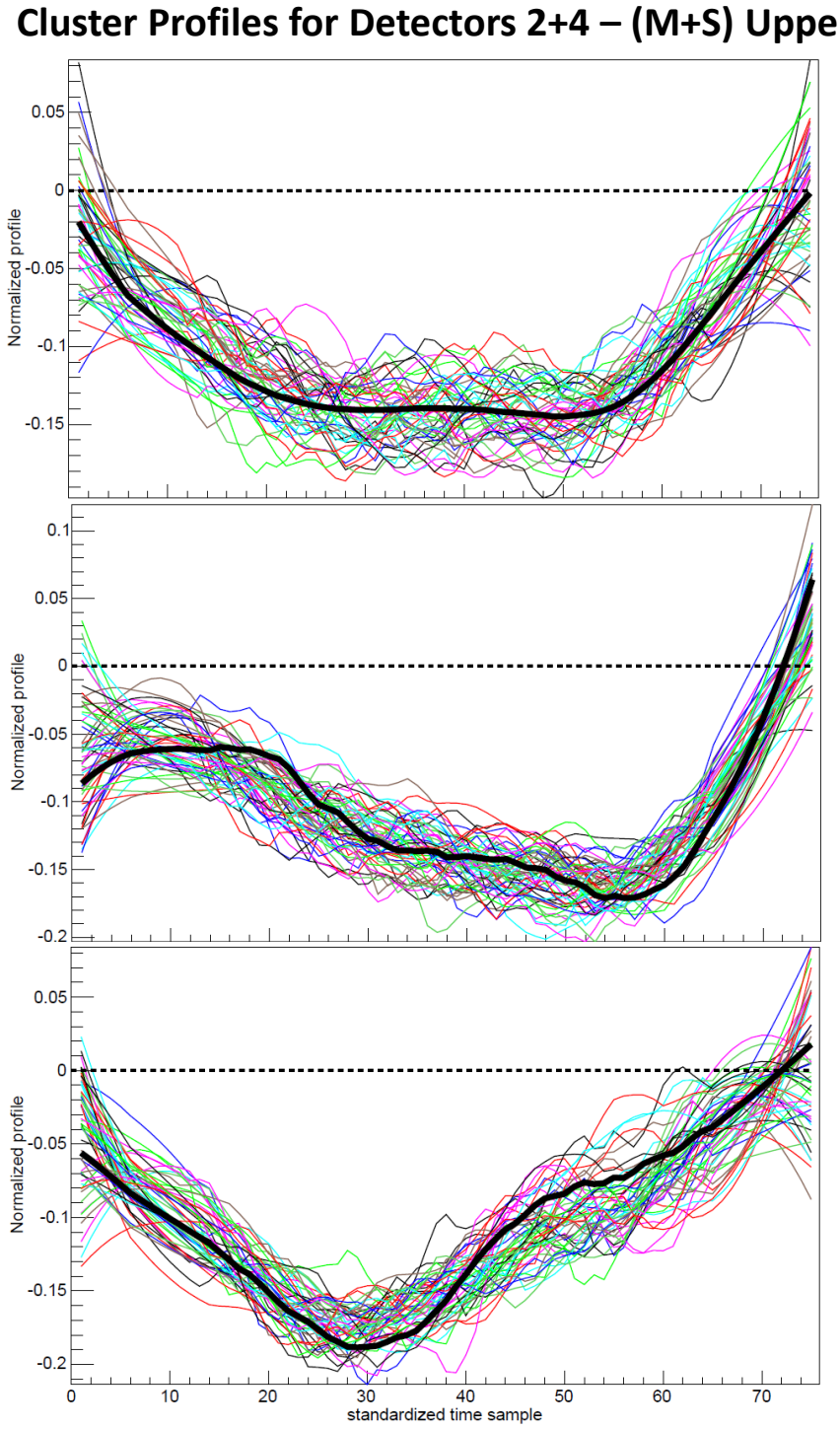

Figure 14. The three most-populated suppression profile clusters for detectors 2+4 for Lane-A occupancies. There are 5,951, 394, and 359 profiles belonging to these clusters respectively out of a total of 9,923 that are processed. These three clusters account for about 67.6% of the processed profiles. The noisy profiles are randomly chosen normalized profiles that belong to each cluster.

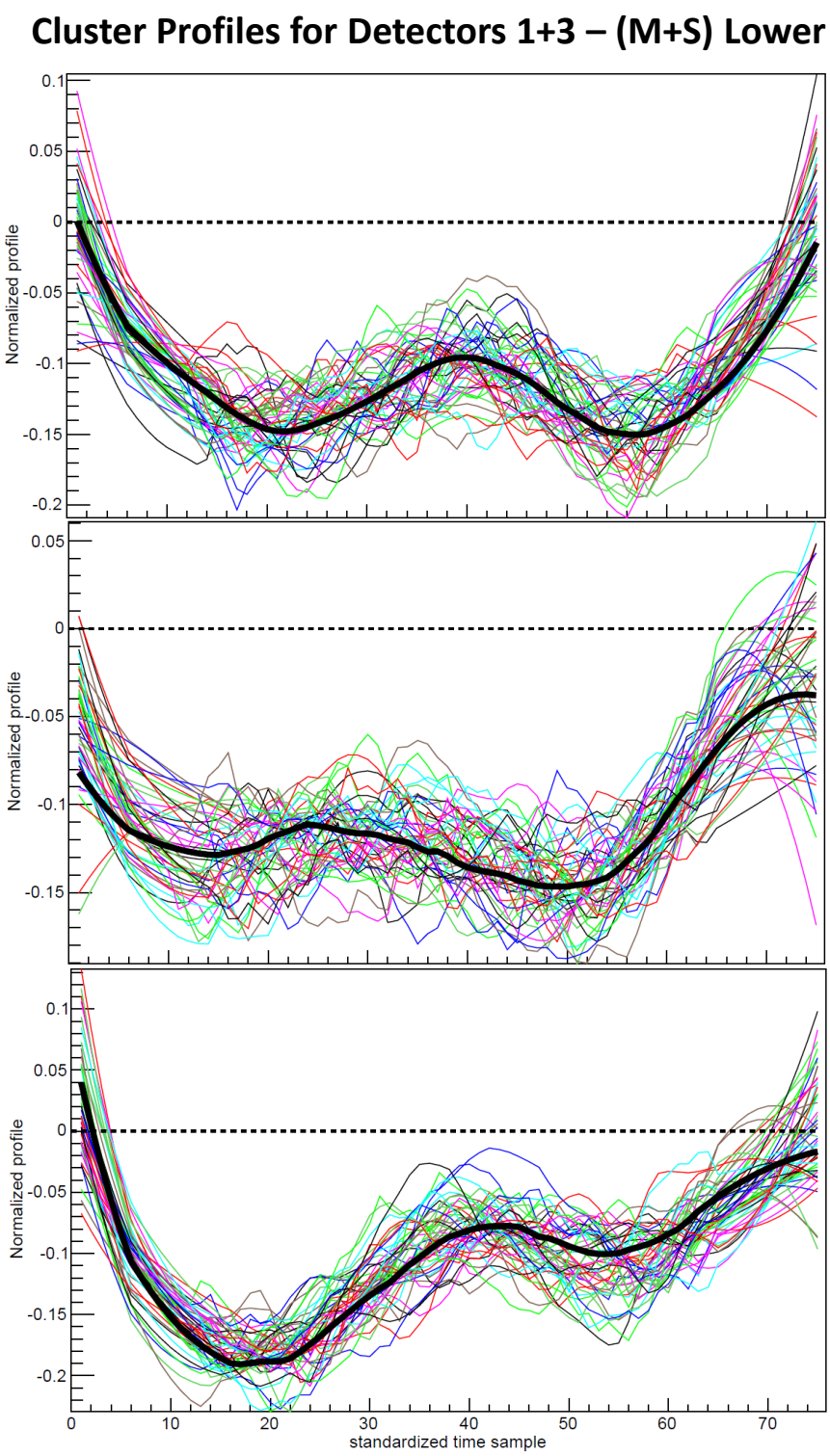

Figure 15. The three most-populated suppression profile clusters for detectors 1+3 for data for all lanes of the test dataset. There are 5,062, 4,928, and 2,566 profiles belonging to these clusters respectively out of a total of 59,067 that are processed. These three clusters account for about 21% of the processed profiles. Note that the cluster vectors are obtained from the first 1,000 processed profiles and those came from lanes that are different from Lane-A.

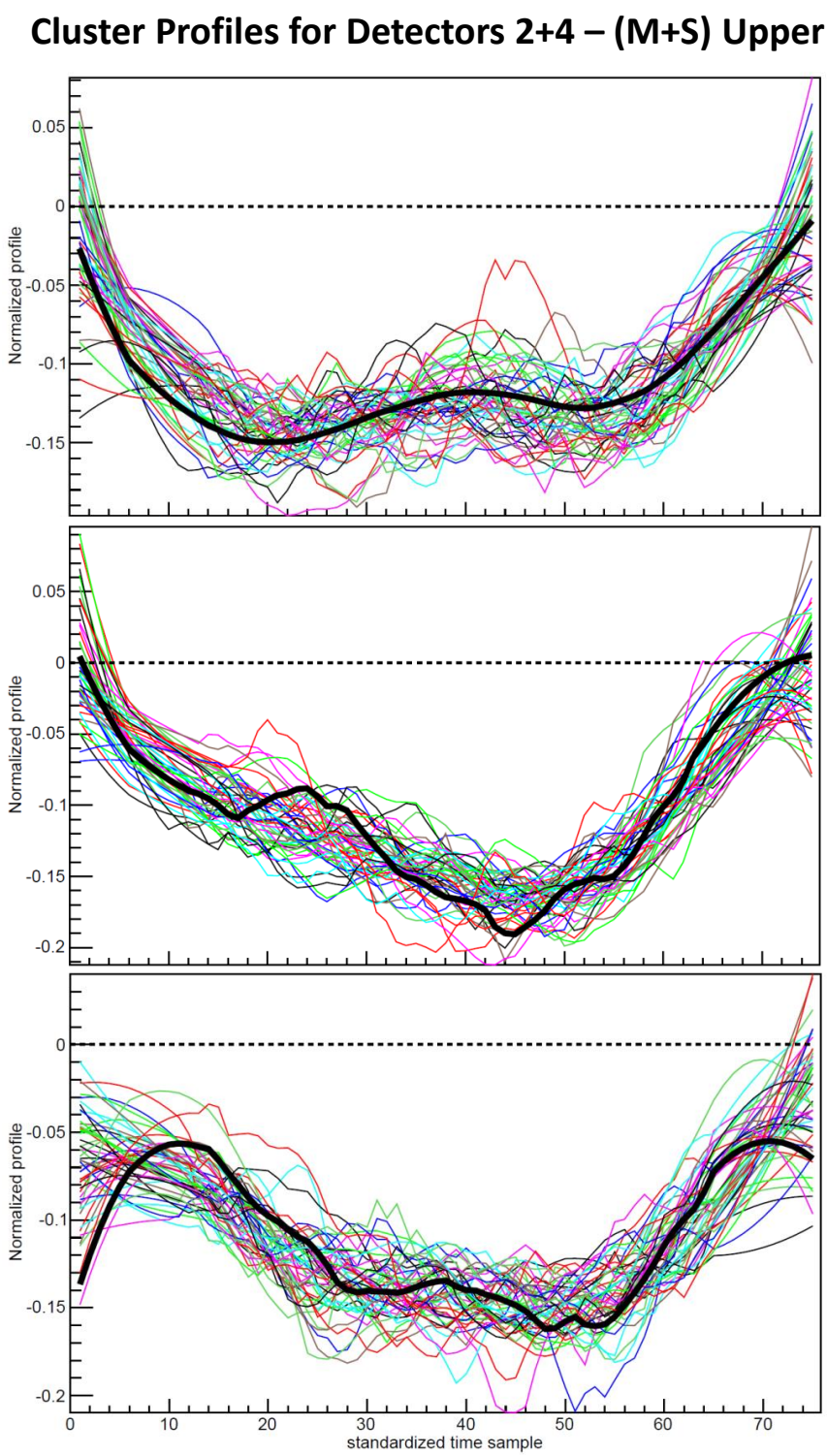

Figure 16. The three most-populated suppression profile clusters for detectors 2+4 for data for all lanes of test dataset. There are 31,425, 11,914, and 7992 profiles belonging to these clusters respectively out of a total of 118,976 that are processed. These three clusters account for about 43% of the processed profiles. Note that the cluster vectors are obtained from the first 1,000 processed profiles and those came from lanes that are different from Lane-A.

4. Summary

This report characterizes an important phenomenon in Radiation Portal Monitors (RPM) deployed as part of radiation detection systems for interdicting the illicit trafficking of nuclear materials under programs such as NSDD (formerly Second Line of Defense) and the domestic counterpart under Customs and Border Protection though the Domestic Nuclear Detection Office. RPMs exhibit detection characteristics that make them susceptible to a wide range of radiation sources from common naturally occurring radioactive materials (NORM) to threat materials. The large number of NORM alarms caused by legitimate commercial goods creates an undue burden on responders to understand the radiation alarm profiles and adjudicate the RPM alarms. Often, a lengthy secondary inspection using handheld instruments is required to determine the cause of the alarm thus slowing commerce and requiring increased manpower and resources.

This report analyzes data from deployed RPM systems at a sample port to characterize the shape and amplitude of background suppression which occurs when background is shielded by a large conveyance that enters the detection zone. The analysis found that a large fraction of RPM occupancies are represented by a small number of suppression shapes and through this analysis established a common set of background suppression profiles for the test dataset that can be used in future profile characterization analysis.

Currently, RPM operators are required to evaluate RPM data profiles and make subjective determination whether to hold the conveyance for secondary inspection or release. This is difficult and somewhat ineffective process even for an experienced operator or an expert with extensive training. This background suppression analysis is an important step in improving the effectiveness of the RPM profile analysis methodology which is currently being investigated and may lead to methods to reduce the number of secondary inspections or decision support tools to aid operators in evaluating RPM data.

This report investigated the shapes of the background suppression profiles for a test dataset from a sample port. The analysis was done for the purpose of incorporating the dominant background suppression shapes in automated profile analysis for radiation sources. The end goal is to assess the capability of such analysis to detect the presence of point-like sources in the stream of commerce. The suppression shapes will be used along with source models in the optimization process of spatial profile analysis to fit measured data.

A Hierarchical clustering analysis algorithm was used to cluster together similar background suppression shapes in a large dataset. Before the cluster analysis, the data was pre-processed as follows: 1- filter the profile, 2- subtract the background, 3- standardize to a constant number of time samples, and 4- normalize the standardized profiles to unit length. The clustering was then applied using a Euclidean distance measure between the normalized profile vectors.

Statistical noise had a significant effect on the analysis and limited the effectiveness of the clustering. For many of the profiles, even after filtering, the suppression shape is dominated by noise rather than by the physics of the suppression. This is especially the case for low suppression amplitudes (SNR $\rightarrow$ 0 as suppression amplitude $\rightarrow$ 0). Addition of detector signals was used to reduce noise effects and resulted in significant improvement in the clustering

efficiency. For a given occupancy either the top or bottom two detectors were added since these detector pairs are expected to have similar suppression profiles for most cargo.

The results of the cluster analysis showed that a significant percentage of processed profiles can be represented by a relatively small number of suppression shapes. For the upper detectors the analysis shows a larger fraction of occupancies are well-represented by a smaller number of shapes than for the lower detectors. This is partly because the upper detectors have larger suppression amplitudes and thus the suppression profiles have better signal-to-noise ratio than the lower ones. Other factors such as vehicle structure also contribute to the differences between the upper and lower detector panels.

The cluster analysis presented in this report resulted in a number of cluster profiles or shapes that represent background suppression profiles for a significant fraction of the data. These profiles can be stretched/compressed to match the number of occupancy samples for direct use in spatial profile analysis. An alternative would be to use representative analytical forms for these shapes. Different shapes need to be used for the lower and upper detector panel in profile analysis.

A significant amount of work remains to determine if the profile analysis approach will work with sufficient robustness to be useful. The background suppression analysis is an important step in the evaluation of the effectiveness of the profile analysis methodology which is currently being tested.

Appendix: Distributions of Speeds

For the test data used in this report traffic speeds are reported in RPM data files for most occupancies. In most instances multiple speeds are reported for the same occupancy. If accurate multiple speeds are available then it is possible to include a background suppression shape correction due to speed changes of traffic passing through the RPM. In this appendix we examine the reported speeds for the same test dataset used for the analysis in the body of the report and consider implications on background suppression analysis.

For the examined data there are many instances where the reported speed is unrealistically large (such as over 50 MPH) and in some other instances the maximum possible value of 99 MPH is reported, which occurs when the two photo-beams are broken nearly instantaneously. However, considering the distributions for a large amount of data provides some insight regarding reliability of the overall speed data. Here we consider the distributions of speeds including these extreme values and then attempt to remove such values using a vehicle length constraint.

Analysis of Data from One Lane:

In the figures below histograms of the reported speeds for occupancies are shown for two cases: 1- Lane-A data and 2- data for all lanes in the test dataset. Only occupancies that have numbers of samples in the range 20 – 200 are included in the histograms in order to avoid problem occupancies. For Lane-A there are 19,570 processed occupancies, of those 81 had one reported speed, 2,339 had two reported speeds, 6,210 had three reported speeds and 10,940 had more than three reported speeds. There are 106 occupancies with no reported speed. Figure A.1 shows the distributions of speeds for Lane-A data with more than three reported speeds. This figure also shows histograms of the ratios v2/v-last and v3/v-last and of the acceleration as determined from the speed and time values for v3 and v-last.

Application of Vehicle Length Constraint:

The results in Figure A.1 show that a large fraction of occupancies have significant variance in reported speeds, especially for the second reported speed. An attempt can be made to filter out erroneous speeds by constraining the vehicle length to expected values where the length is estimated by: $\text{Length} \approx \text{speed} * \text{number of samples} * \delta t$, where δt is the width of the time sample or 0.2 seconds. Since this length estimate assumes constant speed, the length constraint is not expected to be precise. With this limitation in mind we constrained the length to be in the range of 20 – 100 feet. Figure A.2 shows the same results as in Figure A.1 with application of this constraint. The results show significant improvement in that most of the occupancies have nearly constant speeds between the third and last reported speeds. However, if we consider the second reported speed there appears to be significant deceleration to the third reported speed.
Consideration of the acceleration plot from v2 to v3 indicates that the values with large negative acceleration are not likely realistic and thus it is probable that acceleration/deceleration larger than 1 MPH/sec is an indication of a faulty speed measurement. In this case it is suspected that values of v2 are at fault. Information on vehicle length would allow further determination of accuracy of reported speeds, but such information is not available in the data files.

Analysis of Data from all Lanes:

Next we consider speed analysis for data from all lanes in the test dataset. Here also we limit the analysis to occupancies that have 20 – 200 samples. For this dataset there are 305,241

occupancies that are processed. Of those occupancies, 9,958 had one reported speed, 152,556 had two reported speeds, 64,330 had three reported speeds, and 78,397 had more than three reported speeds. There are 1,379 occupancies with no reported speeds and these are not processed. Thus the largest number of occupancies had two reported speeds. Since the previous analysis considered one lane of data and this lane is part of this large dataset, it appears that the number of reported speeds strongly depends on the traffic lane. It is not clear if this is caused by the RPM itself or by the type of traffic that passes through the lane.

Since the largest subset of the data includes two reported speeds, this is the case we concentrate on here. Figure A.3 shows similar histograms to what was shown for the one lane analysis. This set of histograms does not include any vehicle length constraint. Figure A.4 shows the same histograms with the length constraint applied. For this data the length constraint removed most but not all of the cases with high acceleration. Most of the data with two reported speeds appears to be well-behaved with only a small fraction showing higher than expected acceleration between v1 and v2. The small bump to the left of the main distribution in the ratio v1/v2 is likely due to incorrect speed measurements, mainly in v2.

One final question that we address is if there is any correlation between the number of reported speeds and the number of occupancy samples. Figure A.5 shows histograms of the number of occupancy samples for the cases of 2, 3, and > 3 reported speeds for data from Lane-A. There appears to be a weak correlation but it is clear that the whole range of occupancy samples is covered for all three cases. The same conclusion holds when occupancies in all lanes in the dataset are considered. Thus the number of occupancy samples does not dictate the number of reported speeds.

Conclusions:

The analysis of reported occupancy speeds shows that most reported speeds are reasonable and likely accurate. However, there are many speeds that are clearly incorrect. Using a vehicle length constraint (20 – 100 feet was used) resulted in removal of a large fraction of these speeds. The following general conclusions can be drawn from this analysis:

- Number of reported speeds varies between 1 and ~6 and this number depends on the traffic lane. There are occupancies with as many as 8 reported speeds. The number of reported speeds does not depend on the number of occupancy samples.
- The last reported speed often has the least number of problem speed values in the sense of the distribution. However, this is not a general rule and there are cases where it doesn't hold.
- The vehicle length constraint removes a significant fraction of problem speed values. If approximate vehicle length information is available, this constraint can be used to remove most if not all problem cases.
- With removal of the problem cases, the distributions of speed ratios and accelerations show that the speed variation for most occupancies is not large. The maximum variation is approximately $\pm 25\%$ near the edges of the distributions which are highly peaked.
- Given the small change in occupancy speed within the RPM for most data, a correction for the background suppression profile does not appear to be necessary.

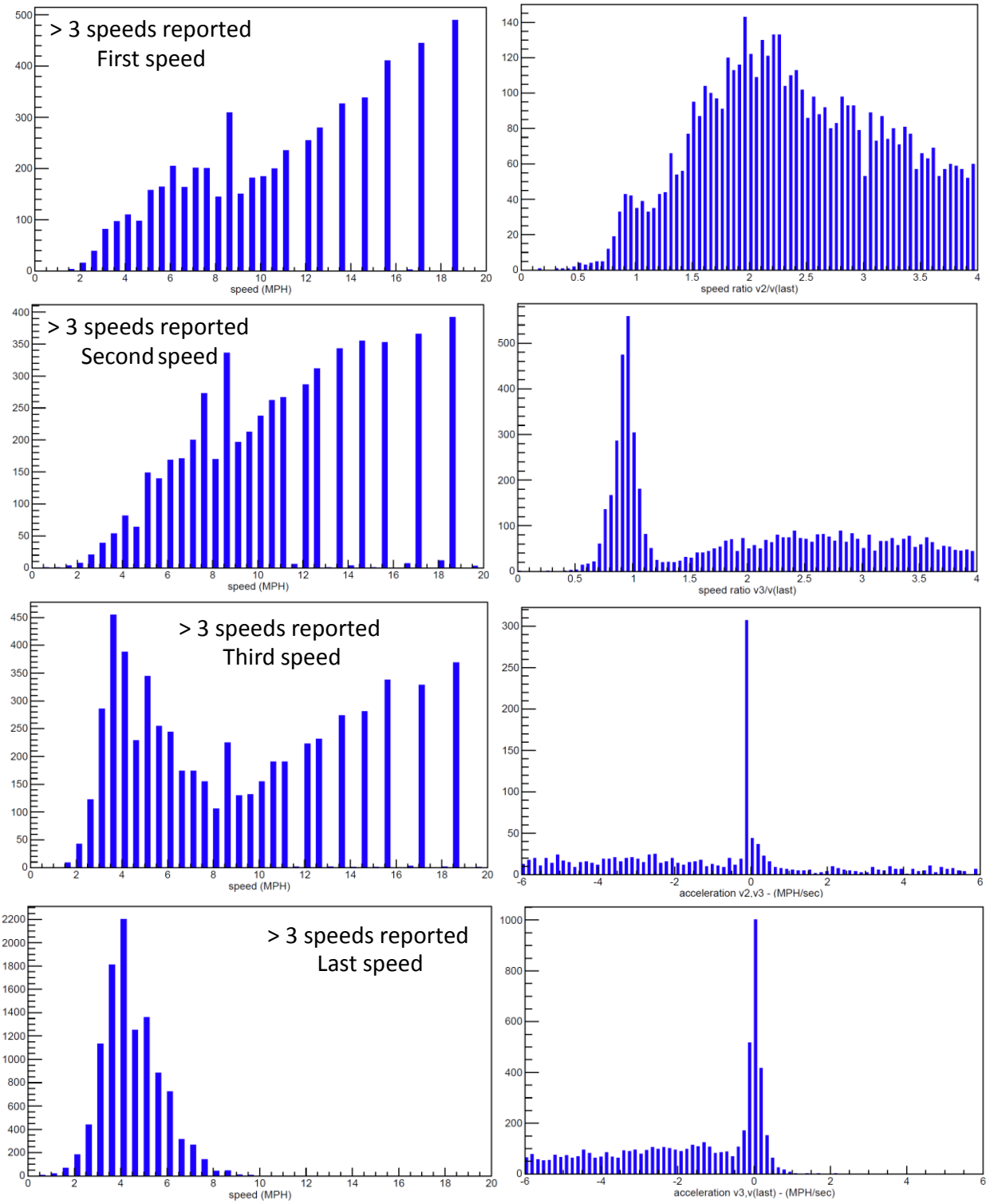

Figure A-1. Histograms of speeds for Lane-A occupancies with more than three reported speeds. The data used for generation of the histograms is restricted to occupancies with number of samples in the range 20 to 200 without any vehicle length constraint. The histograms to the right are of the ratio v2/v-last (top), v3/v-last (second), acceleration from v2 to v3 (third) and acceleration from v3 to v-last (bottom). The time values for acceleration estimates are obtained directly from the time stamps in data lines and the unit of acceleration is in MPH/sec.

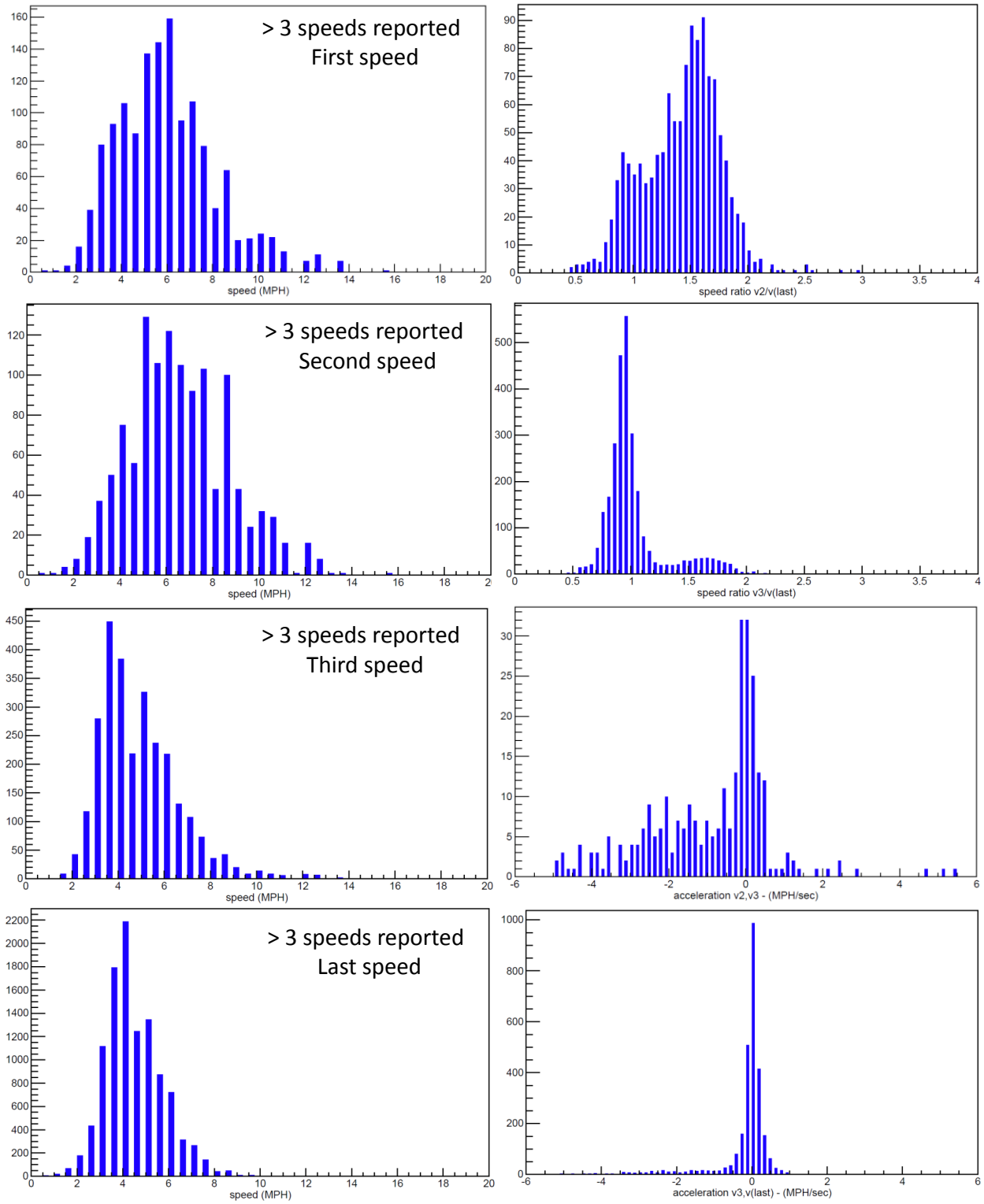

Figure A-2. Same as Figure A.1 with the additional application of the length constraint. Occupancies with estimated length outside the range of 20 – 100 feet are not included in the histograms.

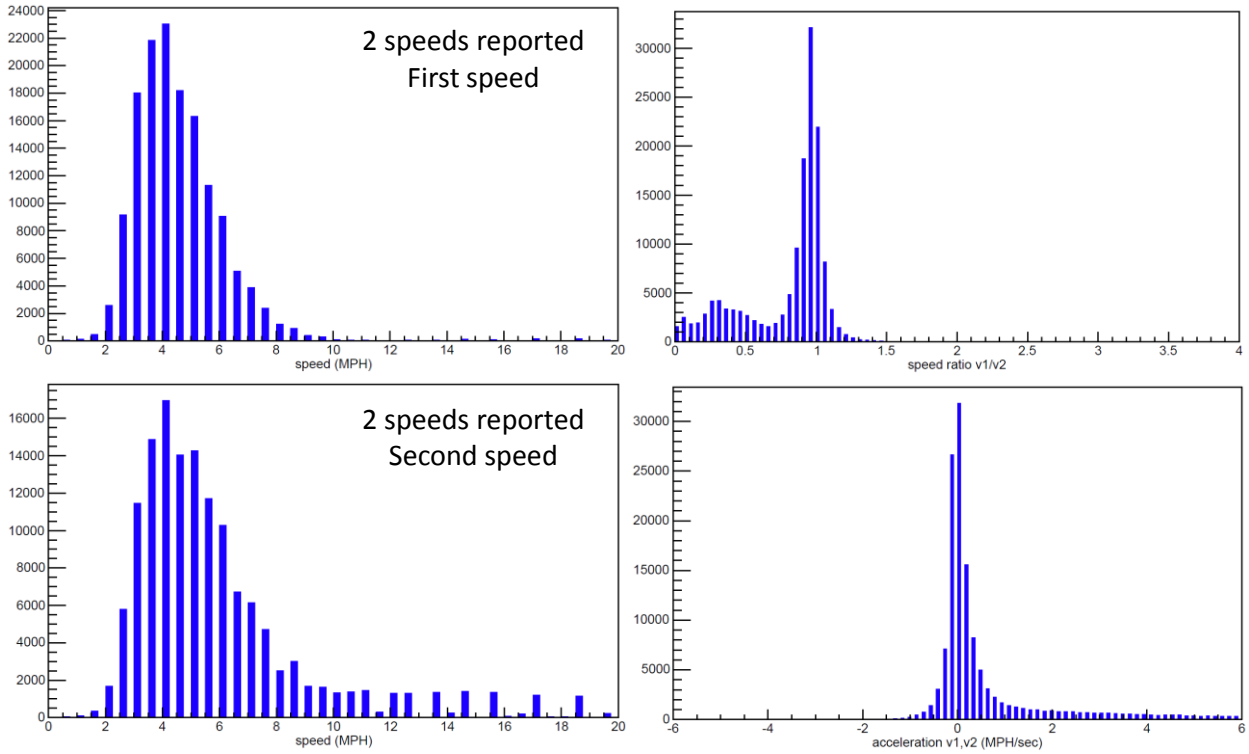

Figure A-3. Left: Histograms of speeds for all lanes for occupancies with two reported speeds. The data used for generation of the histograms is restricted to occupancies with number of samples in the range 20 to 200 without any vehicle length constraint. The histograms to the right are of the ratio v1/v2 (top) and acceleration from v1 to v2 (bottom).

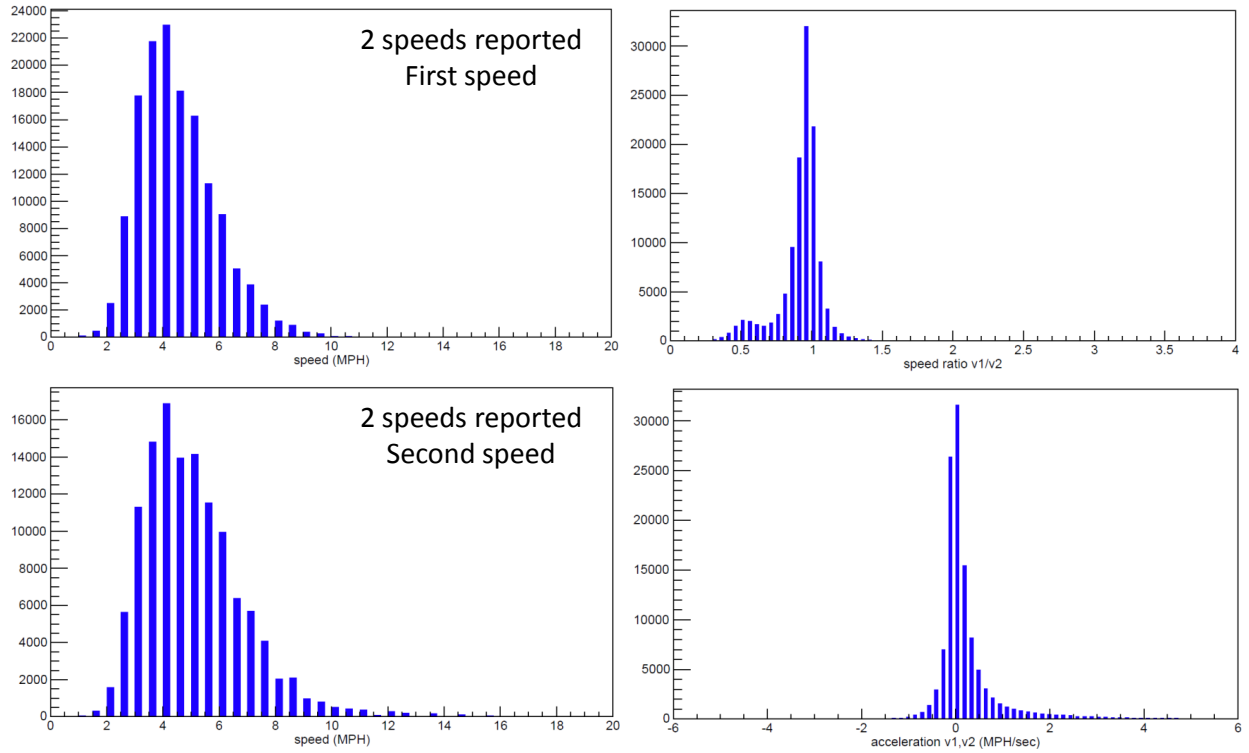

Figure A-4. Same as Figure A.3 with the length constraint (length in range 20 – 100 feet).

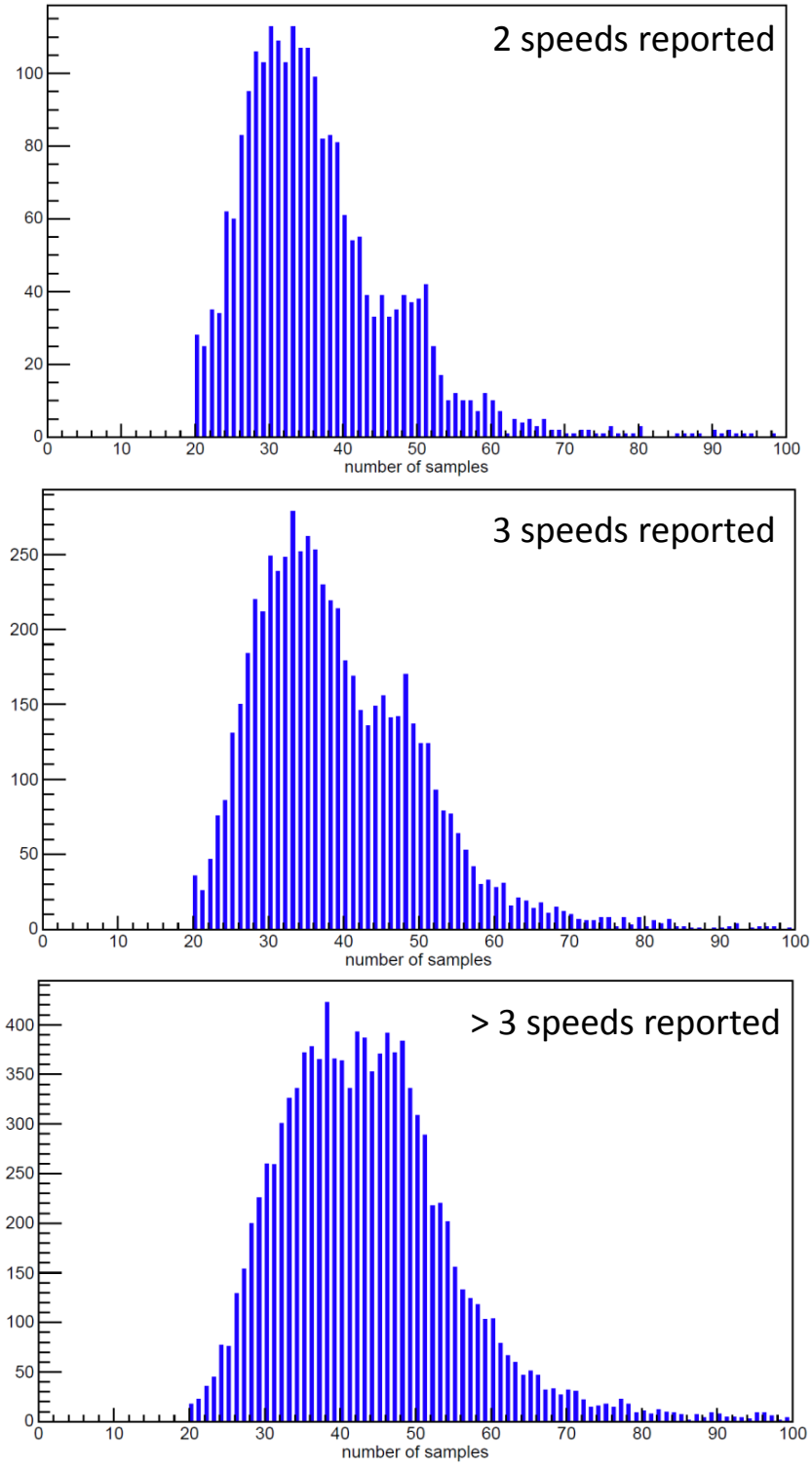

Figure A-5. Histograms of the number of occupancy samples for different numbers of reported speeds for the data used in Figure A-1. The distributions are somewhat different, but the correlation is not very strong.

References

(1) I. R. Shokair and D. Cohen, "Detection of Embedded Localized Radiation Sources Using RPM Data", Sandia National Laboratories report SAND2008-3469, June 2008.

(2) T. Burr, J. Gattiker, and G. Tompkins, "Background Suppression Issues", Los Alamos National Laboratory Report LAUR0503486, 2005.

(3) I. R. Shokair and John K. Estrada, "Analysis of Background Radiation Suppression for RPM Data", Sandia National Laboratories report SAND2006-0035, January 2006.

(4) C. A. Presti, D. R. Weier, R. T. Kouzes, and J. E. Schweppe, "Baseline Suppression of Vehicle Portal Monitor Gamma Count Profiles: A characterization study", Nuclear Instruments & Methods In Physics Research A, 562 (2006) 281-297.

(5) I. R. Shokair and C. L. Kunz, "Hierarchical Cluster Analysis of Background Radiation Suppression for RPM Data", Sandia National Laboratory Report SAND2007-8132, December 2007.

Distribution

Alexander L. Enders, Oak Ridge National Laboratory (electronic copy)

1	MS 1138	Rossitza Homan	6921	(electronic copy)
1	MS 1359	Holly A. Dockery	6810	(electronic copy)
1	MS 1377	Kimberlyn C. Mousseau	6813	(electronic copy)
1	MS 1377	Gregory C. Stihel	6813	(electronic copy)
1	MS 9001	Craig R. Tewell	8005	(electronic copy)
1	MS 9004	Sheryl L. Hingorani	8110	(electronic copy)
1	MS 9004	James C. Lund	8130	(electronic copy)
1	MS 9159	Noel M. Nachtigal	8958	(electronic copy)
1	MS 9402	Kristin L. Hertz	8126	(electronic copy)
1	MS 9402	William C. Johnson	8126	(electronic copy)
1	MS 9402	Christopher L. Kunz	8126	(electronic copy)
1	MS 9402	Alfredo M. Morales	8126	(electronic copy)
1	MS 9408	Willard R. Bolton	8137	(electronic copy)
1	MS 9408	Isaac R. Shokair	8137	(electronic copy)
1	MS 0899	Technical Library	9536	(electronic copy)


Sandia National Laboratories