

PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

Standardized Down-Looking Ground-Penetrating Radar (DLGPR) data collections

Marie Talbott, Erik Rosen, Phil Koehn

Marie Talbott, Erik Rosen, Phil Koehn, "Standardized Down-Looking Ground-Penetrating Radar (DLGPR) data collections," Proc. SPIE 10628, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIII, 106280B (30 April 2018); doi: 10.1117/12.2304539

SPIE.

Event: SPIE Defense + Security, 2018, Orlando, Florida, United States

Standardized Down-Looking Ground-Penetrating Radar (DLGPR) Data Collections

Marie Talbott^a, Erik Rosen^a, Phil Koehn^a

^aInstitute for Defense Analyses, 4850 Mark Center Drive, Alexandria, VA 22311

ABSTRACT

Down-looking ground penetrating radar (DLGPR) has been used extensively for buried target detection. Performance of a DLGPR is typically measured by calculating the probability of detection (PD) and the false alarm rate (FAR) against a target set in a particular soil type. Variability in target sets, including target construction, size, layout, and burial depth, make comparing performance of a DLGPR across test sites and soil compositions a challenge. This paper describes a recent effort to collect data against a standardized set of target types, layouts, and depths. The goal of this effort is to have data sets collected in a uniform manner at various test sites in Australia and Canada for more meaningful comparisons of DLGPR performance in a range of soil types. The data is to be used to improve algorithms for the automatic detection of targets. This paper will describe test planning and execution, and discuss high-level DLGPR results and ongoing analyses from the Australian data collection.

Keywords: down-looking ground-penetrating radar, GPR, mine detection, IED detection, soil

1. INTRODUCTION

It is known that down-looking ground penetrating radar (DLGPR) detection performance depends on the soil in which targets are buried and on the conditions of the soil when data is collected¹. When soil conditions are the same, or similar, it is the characteristics of the targets themselves that matter most—with size, composition, and burial depth being the primary variables dictating performance. Supervised learning algorithms applied to DLGPR data have been shown to improve upon energy-based pre-screeners². But since targets and clutter manifest as similar-sized and similar-shaped hyperbolas in DLGPR data, it is anticipated that additional algorithm improvements will be marginal at best. So where do the challenges remain in improving a DLGPR target detection system?

There is some debate as to whether a DLGPR algorithm should be of a one-size-fits-all variety or a more adaptive one in which algorithm branch points are taken based upon characteristics of the data itself. Proponents of the more rigid one-size-fits-all approach fear that taking the more adaptive approach could result in degraded performance when the system finds itself in an environment that has not yet been experienced. Proponents of adaptive approaches are of course up to the challenge, but there is a dearth of data collected in varying soil conditions using the same DLGPR system. And when data is available, the targets typically differ from site to site.

To provide datasets that can be used to explore adaptive algorithm approaches, it is best to use a standardized target set so that performance differences and DLGPR data characteristics can be attributed to the soil and terrain, rather than to the particular target types and burial depths. In this paper, we describe an ongoing effort to collect and analyze DLGPR data over a standardized set of targets at sites in Australia and Canada. The effort is a cooperative one between the Night Vision and Electronic Sensors Directorate (NVESD) of the U.S. Army, the Defence Science and Technology Group (DSTG) of the Australian Department of Defence, and the Defence Research and Development Canada (DRDC) of the Canadian Department of Defence. While the primary objective was to provide diverse data for algorithm development, preliminary analysis has led to other avenues of exploration, including detection probability prediction based on DLGPR data characteristics, which can potentially be used to alert an operator that the system is in a degraded environment, or in an environment that is favorable to detection.

In section 2, we describe in detail the planning and execution of the data collections. Detection performance results are found in section 3, along with a primer on DLGPR data and various examples of how to view DLGPR data to understand detection performance. Section 4 provides a description of the dual-plate calibration target and how to calculate the real permittivity of the soil local to the calibration target, followed by ongoing analyses approaches in section 5.

Distribution Statement A: Approved for public release; distribution is unlimited

Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXIII
edited by Steven S. Bishop, Jason C. Isaacs, Proc. of SPIE Vol. 10628, 106280B
© 2018 SPIE · CCC code: 0277-786X/18/\$18 · doi: 10.1117/12.2304539

Proc. of SPIE Vol. 10628 106280B-1

2. DATA COLLECTION DESCRIPTION

2.1 Standardized Target Matrix Description

The standardized target set consists of relevant threats that are challenging to detect. Class 1, 2, and 3 targets are metallic, minimum metal, and non-metallic, respectively. Class 4 targets have little metallic content. The targets types selected were meant to be challenging to detect, hence only one target classified as metal was used. There are also dual plate calibration targets buried at the beginning, middle, and end of each lane. The calibration target consists of two circular metal plates that are 6 inches in diameter, non-overlapping, and are separated by exactly 4 inches in height by a plastic connector. The known separation in height between the two calibration plates is used to determine the velocity of the GPR wave in the soil, which can be used to calculate the real part of the permittivity of the soil. Figure 1a shows the target layout for each data collection lane and Figure 1b contains the target key.

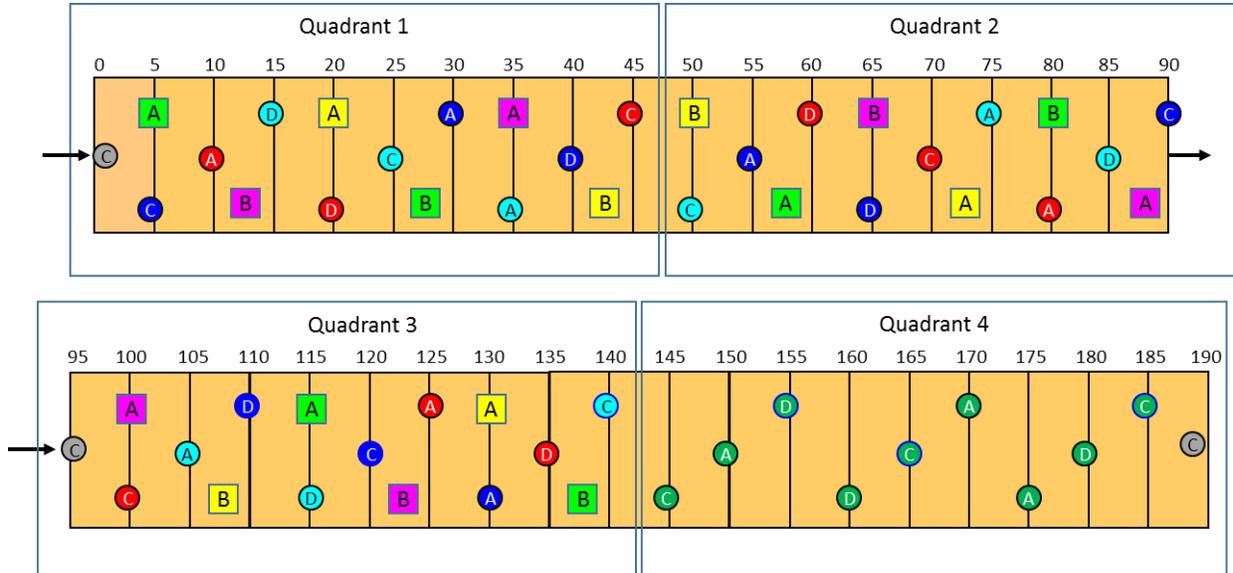


Figure 1a. Target layout utilized in Australia and Canada.

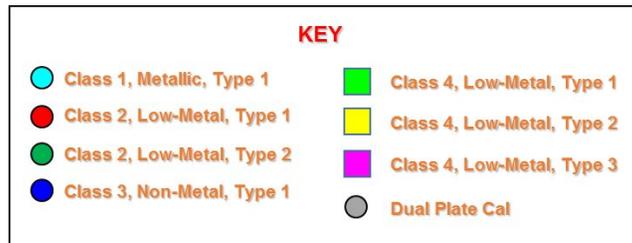


Figure 1b. Target key.

The target layout shown in Figure 1 was utilized for all data collection lanes in Australia and Canada, such that the target types, target depths, and target layouts are constant across sites, leaving soil, terrain, and climate as the independent variables. In each data collection lane, there are three instances of each target class and type buried at two or three depths. The depth of each target appears inside the square or circle that represents that target in the layout. Each 200-meter data collection lane is split into four 50-meter quadrants, where quadrants one, two, and three contain three instances of each target at each of two or three burial depths. Target classes 1, 2, and 3 were buried at three depths: A, C, and D, while target class 4 was buried at two depths: A and B. In a given quadrant, the three instances of a specific target class/type were buried on the left, center, and right sides of the lane. In the other two quadrants a given target type and burial depth would be found at the other two across-track locations, such that a given target/depth combination would ultimately be encountered on each side of the lane, as well as in the center. This was done to provide balance and to take into account lane trends as well as potential radar panel trends. Targets are buried at 5-meter intervals in each lane in different geometries relative to each other. The dual-plate calibration target for computing the real part of the soil permittivity is buried at three locations throughout the 200-meter lane.

In addition to the standardized target layout, the construction of all target classes and types are identical across lanes and sites to ensure there are no subtle differences or variations amongst targets of the same type.

2.2 Sites and Lanes

Thus far, the effort consists of a six-week data collection in Australia, a three-week data collection in Canada, and a two-week data collection at the same Canadian site, but in frozen ground and snowy conditions. The Australian data collection was executed in fall 2016 at four sites spread across the continent. Figure 2 is a map of Australia with the data collection sites identified. The weather conditions and climate varied from site to site and included temperate, cold and rainy, hot and sunny, and wet and humid. At site 1 in Australia, two of the three lanes were comprised of specific soil types that had been trucked in. The third lane at site 1 consisted of the indigenous soil with a grass covering. At each of the other three sites in Australia, two lanes were established by utilizing either open fields or secondary roads within the sites.



Figure 2. Australia Data Collection Locations.

The first Canadian data collection was executed in fall 2017 at two sites located in New Brunswick. Figure 3 is a map of Canada with the data collection location identified as a red star. The two Canadian data collection sites are located about 30 minutes apart in a temperate environment. Two data collection lanes were established at each site. At site 1, two somewhat sparse grassy fields were used, while at site 2, a grader was used to scrape off the indigenous tall grass and shrubs. In the winter of 2018, data was collected under frozen ground conditions, as well as with 4 to 6 inches of fresh snow on 2 days.



Figure 3. Canada Data Collection Location.

At site 1 in Australia and at both sites in Canada, soil was collected for later analyses, and some in situ measurements of permittivity, permeability, and moisture content were made, but the efforts were not extensive when compared to the collection of DLGPR sensor data. Future data collections will include a more significant effort to collect soil information at numerous locations in and around the lanes at each site. The real part of the permittivity in each of the lanes was

computed using the GPR responses from the top and bottom plates of the dual plate calibration targets that were buried in each lane at each site in Australia and Canada. Results are shown in section 4.

2.3 Execution

The data collection platform used was a small four wheeled off-road vehicle platform that was modified to accept a DLGPR payload and a Differential Global Positioning Satellite (DGPS) so that all GPR sensor data would be tagged with accurate coordinates. The platform was chosen for ease of transportation and shipping. The platform has robotic control that allows the operator to drive the system remotely or tethered via Bluetooth. All data collection runs were executed with the operator controlling the platform in tethered via Bluetooth mode.

At three of the four sites in Australia, the data collection was executed in 5 days and generally took the following form:

- GPR data was collected on Day 1, prior to targets being buried so that the efficacy of change detection could be assessed.
- On Days 1 to 2, targets were surveyed, photographed, and buried in the lanes.
- On Days 2 to 4, GPR data was collected over the buried targets.
- On Days 4 and 5, targets were excavated, cleaned, stored, and then shipped to the next site.

More than 5 days were spent at Site 1 since there were three lanes and only two complete sets of targets. At site 2, there was not enough time to bury all the targets, so quadrant 3 is missing from lanes 1 and 2 at this site. It typically took several days to drive the truck containing the system and targets from one site to the next.

In Canada, enough targets were acquired such that lanes 1 and 2 at each of the two sites could be populated without having to excavate targets from one lane for use in another. Also, since the areas on the two sites could be secured at the end of the fall 2017 data collection, the targets were left in the ground for the early 2018 data collection, as well as for a final June 2018 data collection and target retrieval and forensics effort.

Table 1 identifies the number of GPR passes executed at each data collection site. An excursion is one traversal in a given direction of travel on a lane. A run is defined as a pass in each direction of travel on the same lane.

Table 1: GPR runs collected at each OCONUS location

Country	Site #	Dates	Number of Lanes	Number of GPR passes
Australia	1	4-17 Oct 2016	3	32
	2	21-25 Oct 2016	2	20
	3	1-3 Nov 2016	2	16
	4	9-11 Nov 2016	2	22
Canada	1	16-22 Sep 2017	2	32
	2	25-29 Sep 2017	2	40

3. DETECTION RESULTS

3.1 Receiver Operator Characteristic Curves

Top-level detection performance results are given in the form of receiver operator characteristic (ROC) curves by site and lane. ROC curves are formed by first matching system declarations, also referred to as alarms, to surveyed target locations, and then computing probability of detection (PD) and false alarm rate (FAR) as a threshold is applied to all the alarms. Mapping out the ROC curve from low PD to high PD is achieved by rank ordering the alarms by the magnitude of the primary decision statistic, and then continuously lowering a threshold such that more and more alarms are considered. The maximum PD and maximum FAR occur when all the alarms are used. When the PD does not reach 1.0, it means the sensor and algorithm did not generate an alarm near some percentage of the targets.

Figure 4 shows ROC curves for each lane at each site for the six-week Australia data collection effort. We have removed the labels on the x-axis since for this paper since it is the investigation into the causes of detection performance variability that are of more interest than the raw false alarm rates themselves. Note the significant variability in detection performance,

where we have divided the lanes into three regimes—good (green), average (yellow), and poor (red). In some cases, the two extremes were experienced at the same site. Results were good in lane 1 at site 1 (black solid line), while they were poor in lane 3 at site 1 (dotted black line). At site 2, results were good in lane 1 (solid red line), but poor in lane 2 (dashed red line). The right edge of the colored boxes intersects and separates the ROC curves into *Good* lanes in which the PD was about 0.95 to 1.0, *Average* lanes in which the PD was between 0.60 and 0.75, and *Poor* lanes where the PD was between 0.25 and 0.35.

Performance can be separated by target type and depth, and even by particular instance of a target at a fixed location in one of the lanes. Recall that the system traversed each of the lanes several times such that the variance in sensor/algorithm responses for each target instance/type/depth/class can be studied.

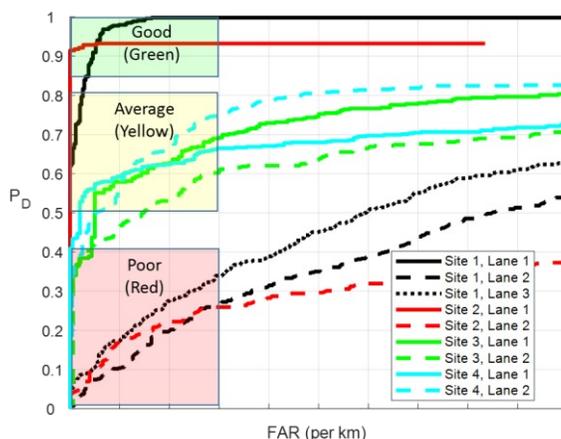


Figure 4. Detection performance ROC curves by site and lane for Australian data collection.

The maximum PD is highest in lane 1 at site 1 and lowest in lane 2 at site 2. We can plot the GPR detection performance as a function of target type for each of these lanes to investigate if a particular target class or type is the driving factor in performance. Figure 5 shows ROC curves for each target class in lane 1 at site 1 (left plot) and in lane 2 at site 2 (right plot). As in Figure 4, the labels on the x-axis have been removed since the performance comparison between the two lanes is of more interest than the raw false alarm rate. However, the limits of the x-axis are the same in both plots so that a direct comparison can be made.

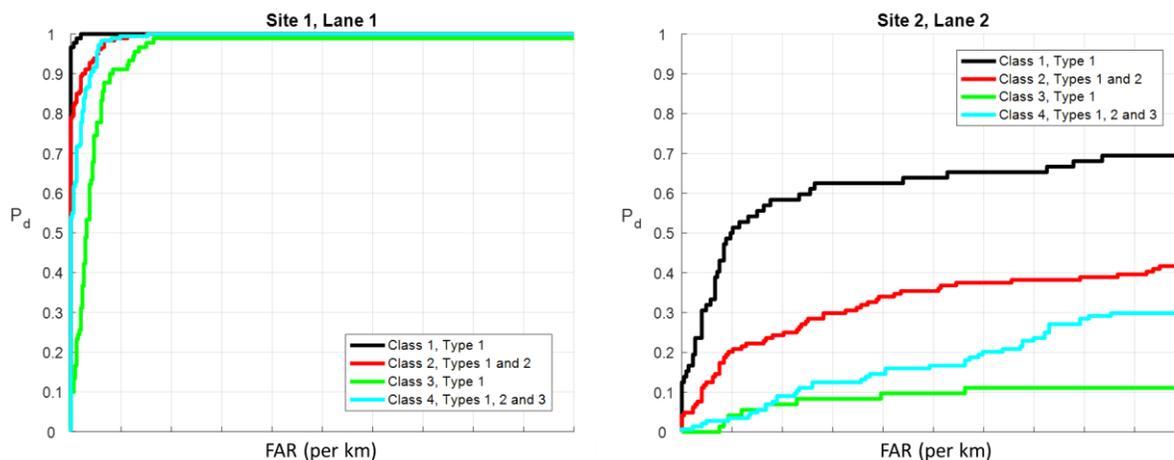


Figure 5. ROC curves by target class for the lanes with best (left) and worst (right) PD in Australia.

All targets in each class are detected in lane 1 at site 1, though at varying FAR values. The FAR at which the maximum PD is reached is lowest for Class 1 targets, which are metallic, and highest for Class 3 targets, which are non-metallic. Class 2 and Class 4 targets of all types are low-metallic and all are detected at a FAR between that of Class 1 and Class 3

targets. The performance against Class 1 metallic targets is best in lane 2 at site 2, though the PD plateaus at about 0.7. The PD falls off as a function of metal content in lane 2 at site 2, with the worst PD against non-metal Class 3 targets and the PD for low-metal Class 2 and Class 4 targets falling in-between the PD for Class 1 (metal) and Class 3 (non-metal) targets.

It is worth noting that the ROC curves shown in Figure 5 are not a balanced comparison since lane 1 at site 1 contains targets from all four quadrants, as shown in Figure 1a, while only quadrants 1, 2 and 4 are populated in lane 2 at site 2 due to time restrictions. The target layout is identical in quadrants 1, 2, and 3 and thus there are fewer statistics against the target set found in those quadrants in lane 2 site 2, since quadrant 3 was left empty. Though the target set is the same for quadrants 1, 2, and 3, the performance can vary from quadrant to quadrant due to different soil conditions, which can vary along the length of a lane. We can examine the PD as a function of each quadrant in a lane to attain a measure of performance variability in a lane. Figure 6 shows the maximum PD for each quadrant for three lanes, each located at a different site in Australia.

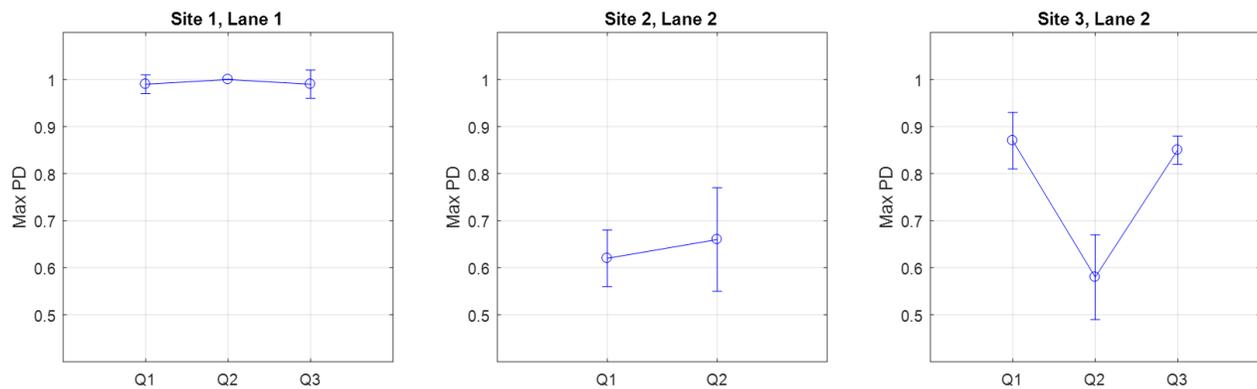


Figure 6. Maximum PD as a function of quadrant for select lanes in Australia.

It is clear from Figure 6 that not all quadrants within the same lane resulted in similar detection performance. The maximum PD varies little in lane 1 at site 1, which is the lane with the highest PD in the Australia collection. There is more variability in maximum PD in lane 2 at site 2 in Australia, which is missing quadrant 3 and has the lowest PD in the Australia collection. The variability in detection performance as a function of quadrant is largest in lane 2 at site 3, which falls in the yellow or average performance range in Figure 4. It is interesting to note that quadrants 1 and 3 have similar maximum PDs in lane 2 at site 3, while quadrant 2 has a maximum PD similar to those found in lane 2 at site 2 where detection performance was the worst. The poor detection performance in quadrant 2 in lane 2 at site 3 could be due to variations in grass cover, which can increase subsurface clutter, or due to changing soil conditions along the lane.

3.2 Scatter Boxplots

To understand drivers of detection performance in the lanes of interest it is worthwhile to examine the confidence value of alarms that are detections of targets and false alarms. An alarm is a system declaration and the confidence value of the alarm is the magnitude of the primary decision statistic. Thus, the larger the confidence value of the alarm, the more likely the alarm is due to a target. The ideal algorithm or decision statistic would assign to all targets alarms of higher confidence value than all false alarms, so that a threshold could be set so that all targets would be detected with no false alarms. Poor detection performance could be due to target detections being assigned a low confidence value, false alarms being assigned a high confidence value, or both. The larger the separation of confidence value distributions for targets and false alarms, the better the decision statistic can identify targets versus false alarms and the better the detection performance.

In Figure 7, the ROC curve for each target class in lane 1 at site 1 and lane 2 at site 2 is on the left and the corresponding scatter boxplots of the confidence values for all targets classes and false alarms are on the right. The scatter boxplot shows the 50 percent quartile in yellow and a green line denotes the median value. The blue boxes indicate the upper and lower 25 percent quartile and any points outside the boxes are outliers. Each individual target detection is a black dot in the scatter box plot and each individual false alarm is a red dot in the scatter boxplot.

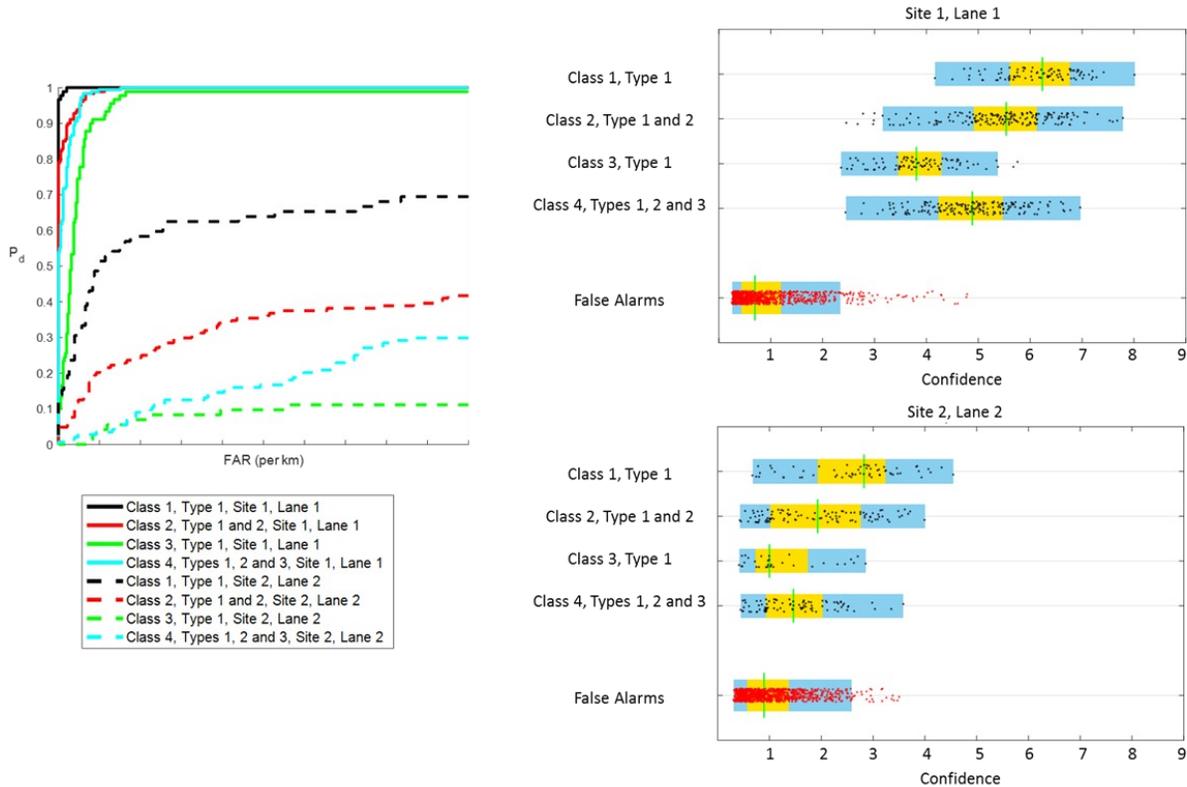


Figure 7. ROC curve (left) and scatter boxplot of confidence values (right) for all target classes and false alarms in lane 1 at site 1 and lane 2 at site 2 in Australia.

The information plotted in the scatter boxplots in Figure 7 informs the shape of the ROC curves of detection results for each lane. The lower left hand corner of the ROC curve corresponds to the highest confidence value alarms, which are on the right of the scatter boxplots. Alarms are plotted on the ROC curve by descending confidence value, or moving from right to left in the confidence value distributions in the box plots. If alarms are a detection, the curve moves up the y-axis (PD) and if alarms are false alarms, the curve moves to the right on the x-axis (FAR).

Lane 1 at site 1 had the maximum PD at a low FAR, as shown in the ROC curves on the left of Figure 7. From the scatter boxplot, we can see that the confidence value distributions for all targets are higher than the false alarms. There is little overlap between the confidence value distributions of target detections and false alarms, though there are some outlier false alarms with higher confidence values. The higher confidence value false alarms affect the shape of the ROC curve for all target classes in lane 1 at site 1: all targets in each class are detected, but the high confidence false alarms cause the ROC curve to go to the right along the x-axis (FAR) prior to reaching the maximum PD.

The confidence value distributions for the target detections in lane 2 at site 2 are much lower than those in lane 1 at site 1. The median confidence value for the false alarm distribution is slightly higher in lane 2 at site 2 than in lane 1 at site 1. However, the poor detection performance in lane 2 at site 2 is not due to high confidence false alarms but due to low confidence value target detections. For example, the performance in lane 2 at site 2 was worst against Class 3 targets in the ROC curve on the left of Figure 7. When we examine the scatter boxplots for Class 3 targets, we see that the confidence value distribution overlaps nearly entirely with the confidence value distributions of the false alarms and that the median target detection confidence value is only slightly higher than the median false alarm confidence value. This indicates that the algorithm or decision statistic cannot tell the difference between a false alarm and a Class 3 target. The overall poor performance in lane 2 at site 2 is due to nearly all target detection confidence value distributions overlapping with the distribution of false alarm confidence values.

3.3 Probability of Detection and False Alarm Rate as a Function of Confidence Value

The scatter boxplots in Figure 7 show that PD and FAR are a function of confidence value. Figure 8 is a plot that shows the relationship between confidence value and PD and FAR. This plot has three axes: the left y-axis is PD, the right y-axis

is FAR, and the x-axis is confidence value. The thick lines correspond to plots of PD versus confidence value, and the thin line corresponds to FAR versus confidence value.

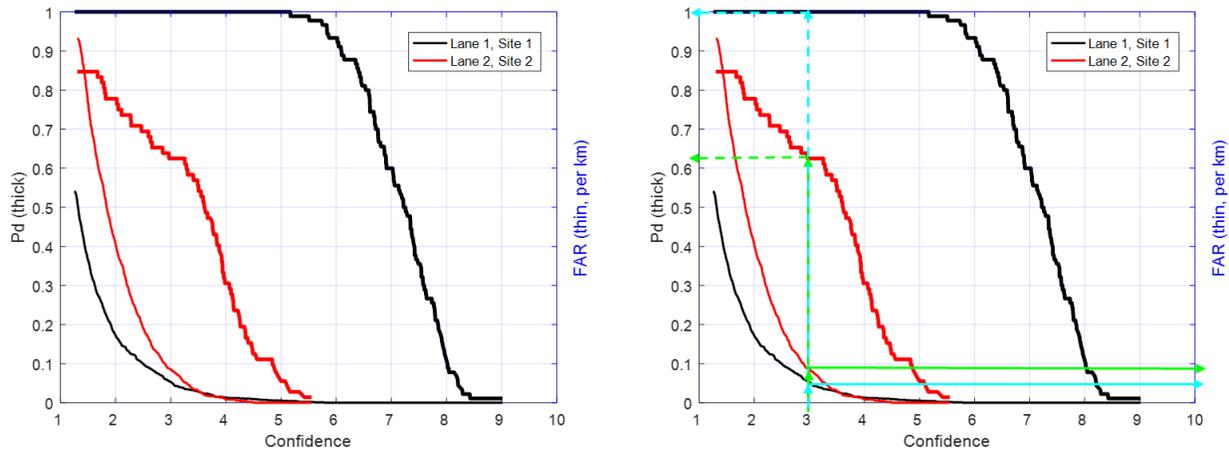


Figure 8. PD and FAR as a function of confidence value for the same target in lane 1 at site 1 and lane 2 at site 2 in Australia.

The plot on the right in Figure 8 is a guide for how to read the plot shown in the left. The cyan lines are used to read the plot for lane 1 at site 1 and the green lines are used to read the plot for lane 2 at site 2. In this example, we are interested in comparing performance between the two lanes at a confidence value of 3. To determine the FAR in lane 1 at site 1 at the confidence value of 3, we begin at 3 on the x-axis and follow the cyan line until it hits the thin black curve. We then map this intersection to the right y-axis, shown by the solid cyan line, to find the FAR per km for lane 1, site 1 at the confidence value of 3. To determine the PD in lane 1, site 1, we begin at 3 on the x-axis and follow the dashed cyan line until it intersects the solid black curve. We then map this intersection to the left y-axis to determine the PD in lane 1, site 1 at a confidence value of 3. One could follow this identical process with the green lines to find the FAR (solid green line) and PD (dashed green line) at a confidence value of 3 in lane 2 at site 2. In the example shown in Figure 8 the PD is 1 in lane 1, site 1 and just over 0.6 for lane 2, site 2 at a confidence value of 3. At this same confidence value, the FAR is higher in lane 2 at site 2 than in lane 1 at site 1.

For optimal detection performance the FAR and PD lines on the plot in Figure 8 would have as much separation as possible in the x-axis, which would indicate that the confidence values assigned to false alarms are lower than the confidence values assigned to targets. This behavior is shown in Figure 8 for lane 1 at site 1. If a user wanted to choose a confidence value to set as a threshold for system operation, 4.5 would be an ideal choice for lane 1 at site 1, since all targets would be detected with no false alarms.

3.4 B-scans

To evaluate various approaches to target detection, we study the DLGPR sensor data in its fundamental form, adopting the terminology used in Daniels's book on GPR³. The fundamental response of any DLGPR system is the A-scan, which is the radar response of a single element, and is associated with a particular down-track location and a particular across-track channel. Figure 9 shows an A-scan response collected over a buried metallic target. The largest response is from the ground, followed by two distinct high-magnitude reflections from the target itself. Nearly all other parts of the response are zero, including the response before the ground bounce, the response between the ground and the buried target, and the response after (or deeper) than the buried target.

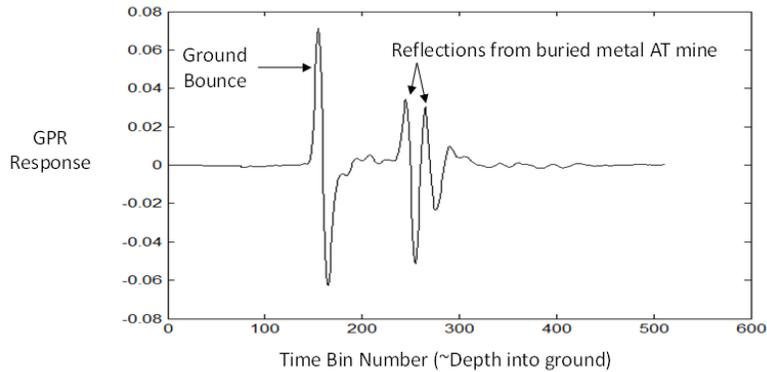


Figure 9. DLGPR A-scan for a buried target.

If we rotate the A-scan of Figure 9 by 90 degrees so that the ground bounce is at the top and then plot neighboring A-scans to the left and right of it, we get the resultant waterfall plot shown in Figure 10a. All the A-scans of Figure 10a come from channel 40, which happens to be the channel in which the target is centered. Thus, the adjacent A-scans are from successive down-track scans. In Figure 10a, the x -axis is the local scan number and the y -axis is the time bin number, which can be thought of as depth into the ground. To create a down-track B-scan, we create a grayscale image of the GPR responses found in Figure 10a, which is shown in Figure 10b. The peaks and valleys of the GPR responses are mapped to a color on the grayscale color bar, with white being a high-magnitude peak and black being a high-magnitude valley. A target manifests itself as a parabola in the B-scan view of the GPR data. The down-track B-scan is essentially a slice of the 3D volumetric DLGPR data taken from one channel of the GPR in the down-track direction. The automatic target recognition algorithms also analyze companion across-track B-scans.

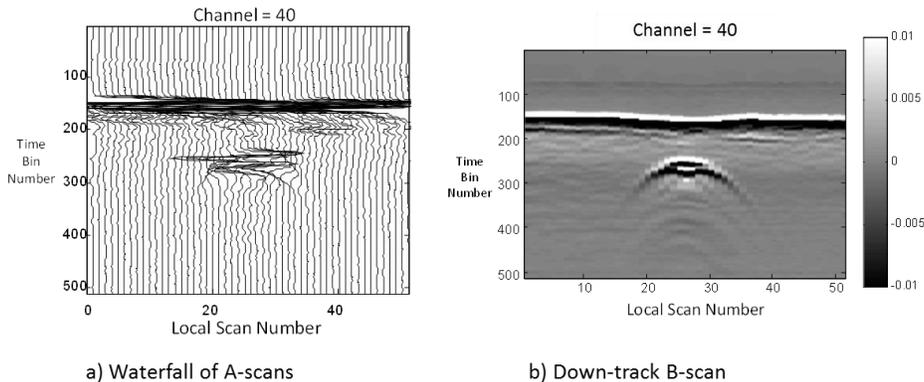


Figure 10. DLGPR waterfall plot of A-scans (a) and DLGPR Down-Track B-scan (b) for a buried metallic target.

Examining the DLGPR sensor data in the form of B-scans is insightful for understanding detection performance. Figure 11 shows down-track B-scan examples of the same target at various depths for lane 1 at site 1 and lane 2 at site 2.

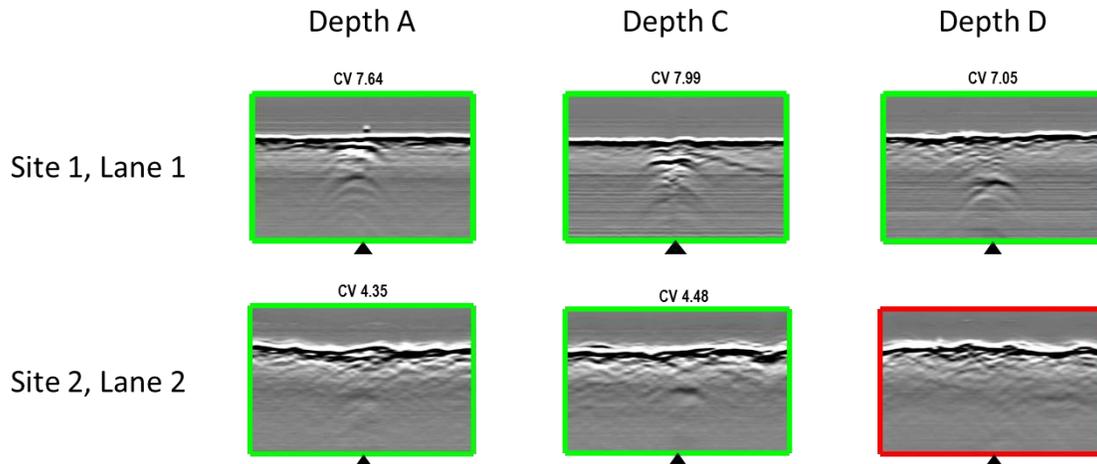


Figure 11. B-scan examples of the same target in lane 1 at site 1 and lane 2 at site 2 in Australia.

The GPR response to the ground is the nearly horizontal white/black line in the B-scan. The target responses in the B-scans are the parabolas of varying intensity shown under the ground. The depths correspond to the burial depth to the top of the target. The B-scans shown in Figure 11 are Down-track B-scan examples with the target centered in the B-scan plot.

The colored boxes in Figure 11 indicate whether the target has been detected. A green box indicates the target was detected and the confidence value is listed above the box. A red box indicates that the target was not detected. In lane 1 at site 1, all examples of the target are detected and the GPR response to the target is visible at all depths. As the burial depth increases, the GPR response to the target appears deeper in the B-scan. It is interesting to note that at Depth A the top of the target response blends into the GPR response to the ground.

The target is detected at depth A and C in lane 2 at site 2, but not at depth D. The target response is faint in the B-scan for depth A, more pronounced in the B-scan for depth C and not visible in the B-scan for depth D. It is observed in the B-scans that there is more clutter at the ground location and just below the ground in lane 2 at site 2. This subsurface clutter is potentially why the GPR response to the target is fainter in the B-scan at depth A than at depth C, and why the confidence value of the target detection is slightly higher at depth C than at depth A. The target responses in the B-scans for lane 1 at site 1 are more crisp and sharp than those for the same target in lane 2 at site 2, which is another observation from the B-scans that could affect detection performance.

4. PERMITTIVITY CALCULATION FROM DUAL-PLATE CAL TARGETS

To understand the detection performance of a DLGPR system in different soil conditions, it is necessary to measure characteristics of the soil at target locations in a timeframe close to that during which the data is collected. Soil moisture, permittivity, conductivity, and magnetic susceptibility are all quantities that affect the strength of the radar reflection off a target. It is difficult to measure many of these soil attributes in situ with accuracy and precision, but the U.S. data collection team and NVESD are working on expanding our knowledge, expertise and data collection of soil properties. The Australian data collection team at certain sites measured many of the soil properties mentioned above. The U.S. and Canadian team collected soil moisture measurements during the Canada collect when time allowed, though it was not a comprehensive sampling of all lanes in all conditions.

The dual-plate calibration target was emplaced in all data collection lanes in both the Australia and Canada data collection and serves as an opportunity to calculate the real part of the permittivity using the response of the radar, without requiring collection or maneuvering of soil samples. The calibration target consists of two non-overlapping metal plates that are 6 inches in diameter and separated in height by exactly 4 inches. All calibration targets were buried such that they were oriented across-track. By observing the magnitude of the radar response to each of the metal plates, we can determine the time it takes the radar wave to travel from the top plate to the bottom plate. Since the distance between the top and bottom plate is fixed and known, the velocity of the radar wave as it travels through the soil can be calculated, which can in turn be used to estimate the real part of the permittivity of the soil at the location of the dual-plate calibration target. The setup for calculating the real permittivity of the soil is shown in Figure 12.

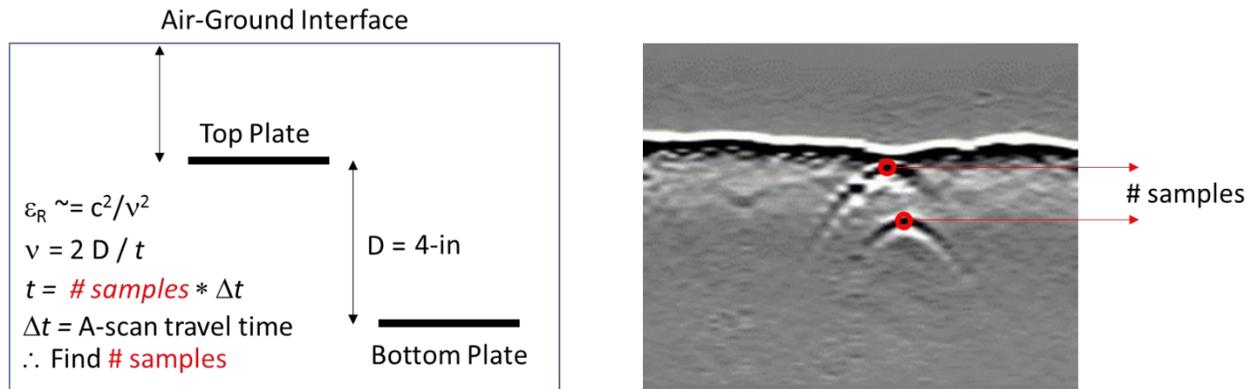


Figure 12. Setup for calculating the real permittivity of the soil from the dual-plate calibration target.

Equations required for calculating the real permittivity are listed on the left of Figure 12. The real part of the permittivity is denoted by ϵ_r and is the result of dividing the speed of light (c) squared by the velocity of the radar wave in the soil squared. The velocity of the radar wave in the soil is calculated by dividing two times the distance (D) by the time it takes to travel the distance D . To find out the time of travel, we must calculate the number of samples between the magnitude of the radar response to the top plate and the bottom plate and multiply this number of samples by the A-scan point spacing or the time for all samples of one A-scan to travel into the ground.

All quantities in the equations on the left of Figure 12 are known, except for the number of samples. The B-scan on the right of Figure 12 identifies how the number of samples is found: the sample number corresponding to the largest magnitude radar response to both the top and bottom plate are identified and the number of samples is the sample number of the top plate subtracted from the sample number of the bottom plate. This calculation can be done for every GPR pass on a given lane to attain a measure of permittivity as weather or other conditions may change during the data collection. This calculation was completed for all lanes at both the Australia and Canada data collection. The real part of the permittivity varied from lane to lane and site to site in Australia and Canada. The permittivity ranged from 3.0 to 12.0 in Australia and from 4.0 to 6.1 in Canada.

5. CORRELATION OF DETECTION PERFORMANCE WITH DLGPR DATA CHARACTERISTICS

We have seen that detection performance over the same set of targets varied significantly from site to site and lane to lane. But the objective of the data collection effort was not just to quantify performance, but to understand the reasons for good/average/poor performance and to investigate algorithmic approaches for improving performance. There are several ways in which the collected data could be leveraged to improve performance. One fairly straightforward approach would be to add the new data to sets of old data collected at other test sites, retrain features and classifiers using all the data, and arrive at an algorithm that works best using all the diverse data combined. This robust algorithm might be the best of all algorithms for all the data available, but it may underperform other algorithms when data is restricted to a particular site or lane or soil type. Thus, there is a choice in either developing a one-size-fits-all algorithm, or instead adopting a several-algorithm solution in which a particular algorithm is essentially tuned to specific soil/terrain types and conditions.

In order to determine if the several-algorithm approach has the potential for success, we will need to study the GPR data itself to see if there are differences that distinguish one soil/terrain type from another, or one soil condition from another. We begin by examining the data from one of the good lanes and one of the poor lanes. We would like to determine if there are characteristics of the GPR data that give rise to good performance or poor performance. Finding correlations between GPR data metrics and performance is the first step.

In Figure 13, we show down-track B-scans from excursions on site 1, lane 1 and site 2, lane 2, as well as a photo of each lane and the ROC curve comparison. The down-track B-scans correspond to 10 meters of data, where the ground bounce response of each A-scan has been aligned to time bin 50. The red box is drawn around the portion of the GPR response prior to the ground bounce peak (the in-air response), while the green box is drawn around a portion of the GPR response after the ground bounce (the shallow subsurface response). The presence of greater clutter signals within the red and green

boxes of site 2, lane 2 manifests as lighter shades of white and darker shades of black, and in general more heterogeneity than the more uniform or homogeneous responses in the site 1, lane 1 boxes (i.e., more unchanging gray).

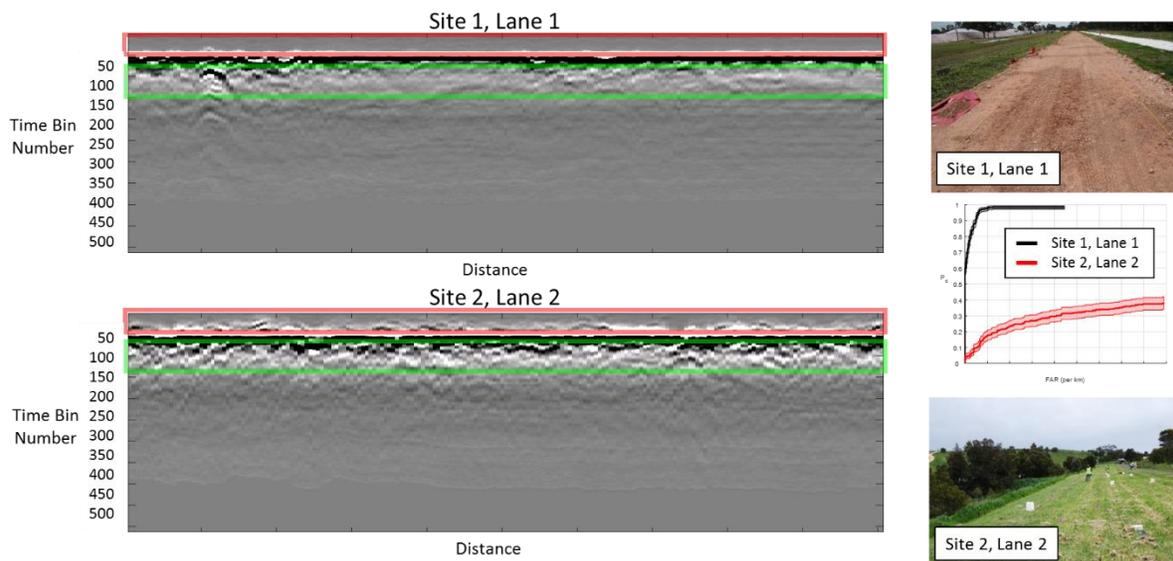


Figure 13. Down-track B-scans for two lanes resulting in the extremes of detection performance.

If for every A-scan we compute standard deviations over different time/depth regimes, we create multiple types of C-scans where for every scan and channel we have a positive scalar value. For the red box in-air metric, we choose our start and stop time bins as 25 and 40. For the green box shallow subsurface metric, we choose our start and stop time bins as 85-150. The resultant in-air C-scans for 50-m samples of data taken from each site/lane are shown as grayscale images in figure 14, where the elevated levels of clutter from site 2, lane 2 are obvious. It is likely that scattering from the grass is the cause of the elevated clutter in the in-air response on the grass lane at site 2. And it is possible that the elevated shallow subsurface clutter in the grass lane is caused by root structures that are not present in the dirt lane at site 1. We do not suggest that correlation is necessarily causation here, but we next determine if there is a relation between in-air noise and PD.

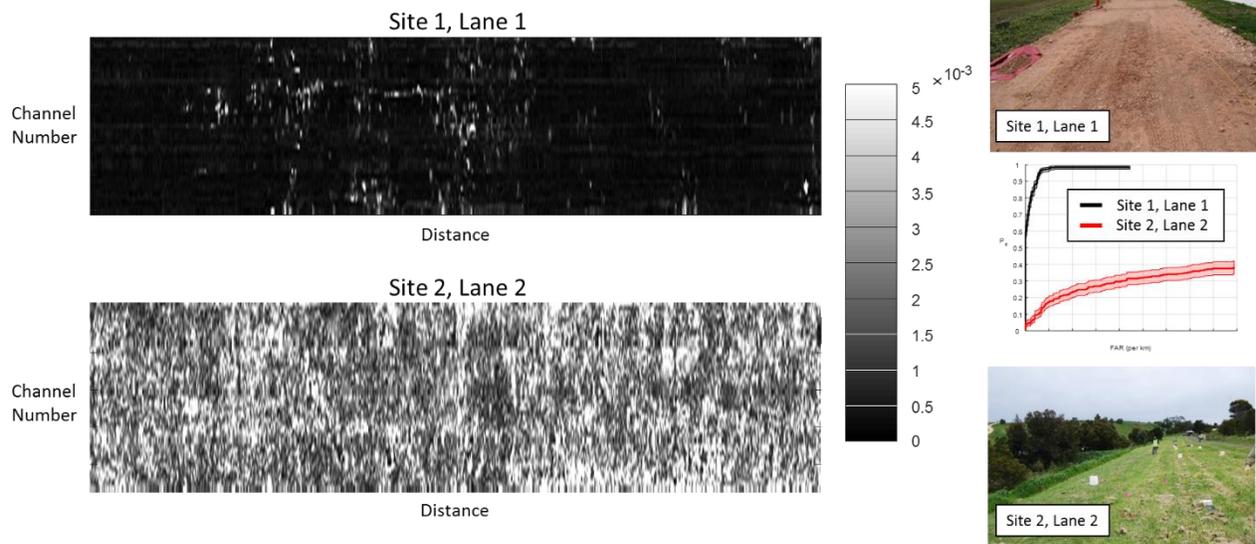


Figure 14. C-scans of the in-air standard deviations for two lanes resulting in the extremes of detection performance.

In figure 15a, we plot the mean value of an entire in-air C-scan corresponding to a 200 meter long excursion on a given lane against the maximum PD for that lane—taken from the ROC curves of figure 4, though the curves are truncated in that figure. We then use a linear least squares fit to the data to predict what the maximum PD would be in a given lane based on the in-air metric for the entire lane. In figure 13b, we compare the actual maximum PD to the predicted maximum PD. The results indicate that by merely computing the standard deviation of the A-scan response prior to the ground bounce, one can predict the PD within 10 percent or less. Expanding the analysis to include results by quadrant instead of lane, as well as including data collected in Canada over the same target set is necessary to ensure the correlations are rooted in physical causes—that elevated in-air noise/clutter is a cause of degraded detection performance.

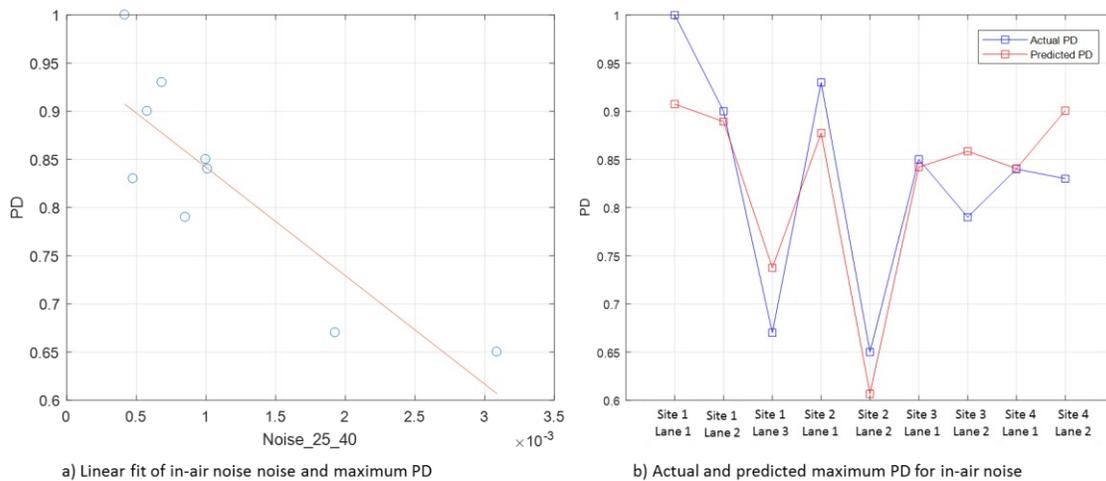


Figure 15. Linear fit of in-air noise versus maximum PD for each of the nine lanes, and predicted-actual PD.

We explored the concept of a discrete meter that might be used by an operator to determine if detection conditions are degraded, or are favorable. We chose thresholds below and above in which we color each C-scan pixel green for low in-air noise, red for high in-air noise, and yellow for in-air noise level between. It is not clear how well these thresholds will translate to the remainder of the data, but we simply demonstrate the possibilities here. Figure 16 shows what an operator might see scrolling by as the system is driven down a roadway, where green suggests favorable detection conditions and red warns of degraded conditions.

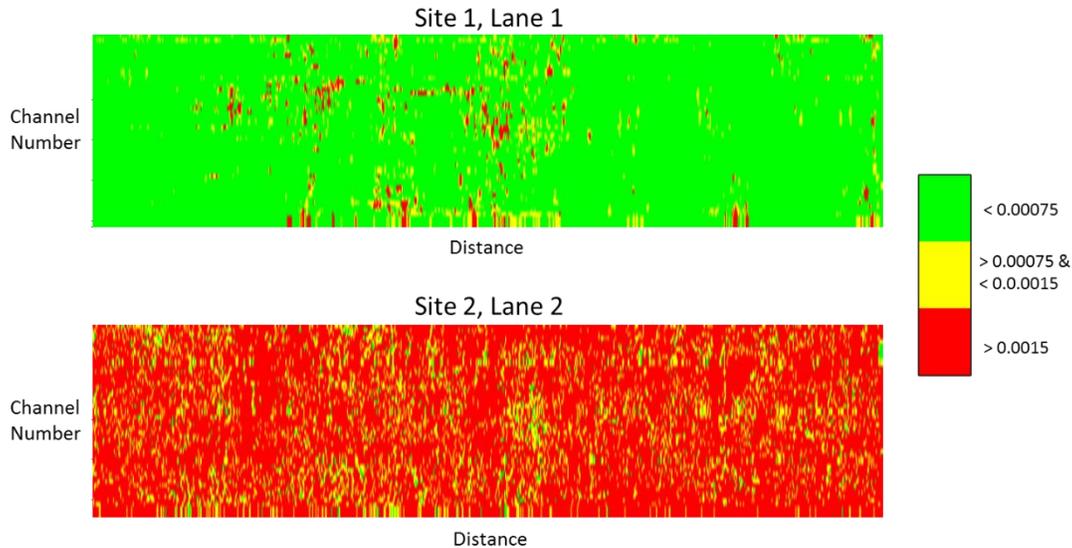


Figure 16. Mapping of in-air noise to discrete G/Y/R meter for operator notification of potential degraded detection.

6. FUTURE WORK

6.1 Toward Adaptive Context-Dependent Algorithms

The ultimate objective of a DLGPR detection system/algorithm is to be so highly adaptable that it essentially senses its environment in real-time and adjusts thresholds and parameters in such a way that detection performance is optimized. There have been efforts in context dependent algorithm development using DLGPR data⁵, but no algorithm was ever adopted due in part to the limitations of the available datasets. The more diverse data collected for this effort may renew context dependent algorithm approaches.

6.2 Soil Heterogeneity and Variogram Analysis

In section 5.1, we demonstrated the potential of standard deviations of A-scans as predictors of detection performance. It is likely there are other physical factors that affect DL GPR detection performance but that which cannot be captured in the variance of A-scan responses. The C-scans of figure 9 suggest that textural differences in the DLGPR data may be of importance to detection performance. Some authors⁶⁻⁷ have found that detectability of targets in DLGPR data is correlated to soil heterogeneity, and that variogram analysis can be used to identify and quantify these soil differences. The diverse data collected for this effort is suitable for this type of analysis.

6.3 Analysis of Canada Data

Most work shown in this paper is the result of applying techniques and methodologies to the Australian data set. We would like to apply the same methodologies to the Canada data set from September 2017, collected in temperate conditions, and from January 2018, which was collected in frozen and snowy conditions.

7. ACKNOWLEDGMENTS

We would like to thank the individuals who participated in the planning and execution of this significant data collection effort. From the U.S Army Night Vision Electronics Sensor Directorate (NVESD), Lee Offen was the team leader most responsible for the successful execution of the effort. Other NVESD team leads, system operators, and surveyors included Tom Becker, Mike Donnelly, Blaine Froeschl, Kerry Hart, Bryan Hucker, Michael Scholl and Brett Smith. The Australian team from DSTG led by Canicious Abeynayake and John Wilson were gracious hosts and provided support for the data collection. The Canadian team from DRDC—Anthony Faust, Kevin Russell, Cristian Mosquera, Scott Irvine, Brian VanderGaast, Wayne Sirovyak and Haley Dittmann—was crucial to the success of the Canadian data collection and were great partners to work with on the test site. Field representative Jeremy Poe contributed to the success of the Australian

data collection. And finally, the correlation and predictive analysis was inspired by creative energies of Martin Milner and Michael Younger of AARD.

REFERENCES

- [1] Rhebergen, J. B., Lensen, H. A., Wijk, R., Hendrickx, J. M. H., van Dam, R., Borchers, B., "Prediction of soil effects on GPR signatures," Proc. SPIE 5415, 705-715, 2004.
- [2] Wilson, J. N., Gader, P., Lee, W., Frigui, H. and Ho, K. C. "A large-scale systematic evaluation of algorithms using ground-penetrating radar for landmine detection and discrimination," IEEE Transactions on Geoscience and Remote Sensing 45(8), 2560-2572 (2007).
- [3] Daniels, D. J. [Ground Penetrating Radar, 2nd Edition], The Institution of Engineering and Technology (2004).
- [4] van Kempen, L. and Sahli, H., "Ground penetrating radar data processing: A selective survey of the state of the art literature," Technical Report, IRIS-TR-0060, Department of Electronics and Information Processing, Vrije Universiteit Brussel (1999).
- [5] C. R Ratto, P. A Torrione, and L. M Collins, "Context-dependent feature selection for landmine detection with ground-penetrating radar," Proc. SPIE 7303, p. 730327, (2009).
- [6] H. Frigui, P. D. Gader, and A. C. B. Abdallah, "A generic framework for context-dependent fusion with application to landmine detection," Proc. SPIE 6953, p. 69531F, (2008).
- [7] Takahashi, K., Igel, J., Preetz, H. and Sato, M., "Influence of Heterogeneous Soils and Clutter on the Performance of Ground-Penetrating Radar for Landmine Detection," IEE Transactions on Geoscience and Remote Sensing, 52 (6), 3464-3472, 2014.