DEVELOPMENT
TECHNOLOGY
UNIT

D
T
U

Working Paper No. 54

The Problem of Evaluating the Performance of Landmine Detection Systems

February 2000

R Gasser & T H Thomas

Development Technology Unit
School of Engineering, University of Warwick
Coventry CV4 7AL UK
Tel: +44 (0) 2476 523122 Fax: (0) 2476 418922
email: dtu@eng.warwick.ac.uk
http://www.eng.warwick.ac.uk/DTU/

DTU Working Paper No. 54 **February 2000**

The Problem of Evaluating the Performance of Landmine Detection Systems

by Russell Gasser and Terry Thomas

Contents

1 Introduction

Even in densely mined areas mines tend to be widely separated, and very few mines will remain undetected by a good mine detector system. Objective analysis of the performance of mine detectors based solely on the percentage of mines not detected is thus difficult at more than an anecdotal level. Field conditions, the types of target and operating procedures vary so widely that testing in a laboratory or test area may not reveal the limits of the performance of a mine detector in a specific mined area with a particular type of mine/UXO contamination. Generalisations about humanitarian demining equipment performance can have only a very limited validity.

Reports of promising technologies of necessity quote results such as a "95% success rate" or "16 out of 16 targets found" [Ano00] but usually do not attempt to present an analysis of the statistical confidence of such data. This working paper examines the difficulties in assessing the performance of mine detecting equipment both quantitatively, by developing a statistical analysis and presenting the results, and qualitatively, by analysing the practical difficulties of evaluating equipment performance specific to demining. Use of the concept of "margin of detection" is proposed as a possible way forward.

2 The three main problems

The three principal difficulties in assessing prototype demining equipment are:

1. Testing equipment that is still under development, and hence not yet capable of finding every mine, in live areas is not possible because of the risk to the operator. Testing under simulated conditions does not yield the same results; the major impediments to finding mines, such as vegetation, have to be removed or altered to place surrogate mines. Deminers, no matter how carefully they seek to re-create their working practices, are likely to act differently in live areas from ones they know to be safe. The ethical justification for asking deminers to perform a trial in a safe area while leading them to think that it is mined in order to simulate live conditions more accurately, is debatable.
2. Finding enough mines or surrogate mines to provide a meaningful statistical analysis of the detection rate at a useful level of confidence is quite impractical as meaningful trials require hundreds of targets.
3. Sensitivity to factors beyond the control of the test protocol may be greater than sensitivity to the parameters being measured. For example, the exact depth of a small buried target may strongly influence the probability of detection. Placing a target and then re-filling above with soil makes precise determination of the depth difficult, moreover it may well not give the same results for some methods of detection as long-buried mines, and the precise depth of soil may vary slightly after heavy rain or vegetation growth.

The rest of this working paper addresses problems two and three above.

3 The statistics of missed mines

The only published paper of note on detection probability in humanitarian demining and associated confidence levels appears to be that of Voles [Vol98] who presented his argument somewhat cryptically. His method permits the calculation of results for only a limited number of levels of confidence, namely the values of the cumulative Poisson function $F(x;\lambda)$ where $x=$ 2, 3, 4 or 5 and $\lambda = 1$. The analysis outlined below is both simpler and more general, and is believed to illuminate an inconsistency in Voles' assumptions.

Mine detection satisfies the conditions for a Bernoulli trial [MF85, page 57]:
(i) there are only two outcomes (mine detected or mine not detected),
(ii) the probability of success is the same for each trial,
(iii) there is a constant number of trials (the total number of mines), and
(iv) the trials are independent (locating a mine does not affect the performance of the mine detector when attempting to locate the next mine).

Thus the use of the Binomial distribution for probability calculations is justified. The conditions for using the Poisson approximation to the Binomial are also satisfied if the number of mines (n) is large and the probability of failing to detect each mine (p_{fail}) is small. In general the Poisson approximation can be considered valid if $n \geq 20$ and $p \leq 0,05$ or if $n \geq 100$ and $n.p \leq 10$ [MF85] [Bar94].

3.1 A conceptual model

A useful conceptual aid to understanding the statistics of mine detection can be to consider the model of a population of mines of which a fraction u are undetectable, and a detector capable of always finding all the rest. This gives the same results as a population entirely made up of theoretically detectable mines and a detection system with a probability of only $(1-u)$ of locating each mine. A practical example of this conceptual model is the type of minimum-metal mines that are supplied with steel discs which can be optionally fitted as the mines are emplaced in order to make them easy to locate later with a standard metal detector (e.g. the TMA-4 anti-tank mine formerly made in Yugoslavia). If the detection performance required is the commonly accepted 99,6% success rate then it can be assumed that on average four mines per thousand have not been fitted with discs, and the metal detector will find all of the mines with discs and none that lack discs. This is a reasonable assumption in practice if the mines have been laid at the correct depth. Defining the exact reason for indetectability makes no difference to the statistical analysis provided that it is a random process. In practice this condition may not be completely satisfied as failure to detect a mine may be due, for example, to specific soil conditions that prevail over some mines and not others in a non-random manner, but the consequences of this are considered to be negligible.

3.2 Definition of the problem of sampling

The problems of obtaining meaningful results from limited-size trials can be expressed thus:

Test results show that in a sample of n mines (the trial size) x $(= 0, 1, 2, 3, \ldots, n)$ mines were ***not*** detected. What is the probability that an arbitrary proportion π of similar mines from the same population would also escape detection? What is the limit of confidence in this result?

Clearly if the number of mines used in the test (n) is very large there is a high degree of confidence that $\pi = \frac{x}{n}$. If the sample size is small it will less accurately represent the stock of mines as a whole and there is a certain likelihood that the value of π is larger than $\frac{x}{n}$ (or smaller than $\frac{x}{n}$ if $x > 0$). The trial result therefore depends on (i) the probability that any individual mine is detectable and (ii) the confidence that the sample accurately reflects the stock as a whole.

Taking the standard promoted by the United Nations of 99,6% mine clearance, an example of the conceptual model outlined above might be a crate of 1 000 mines of which 996 are detectable (they have steel discs that

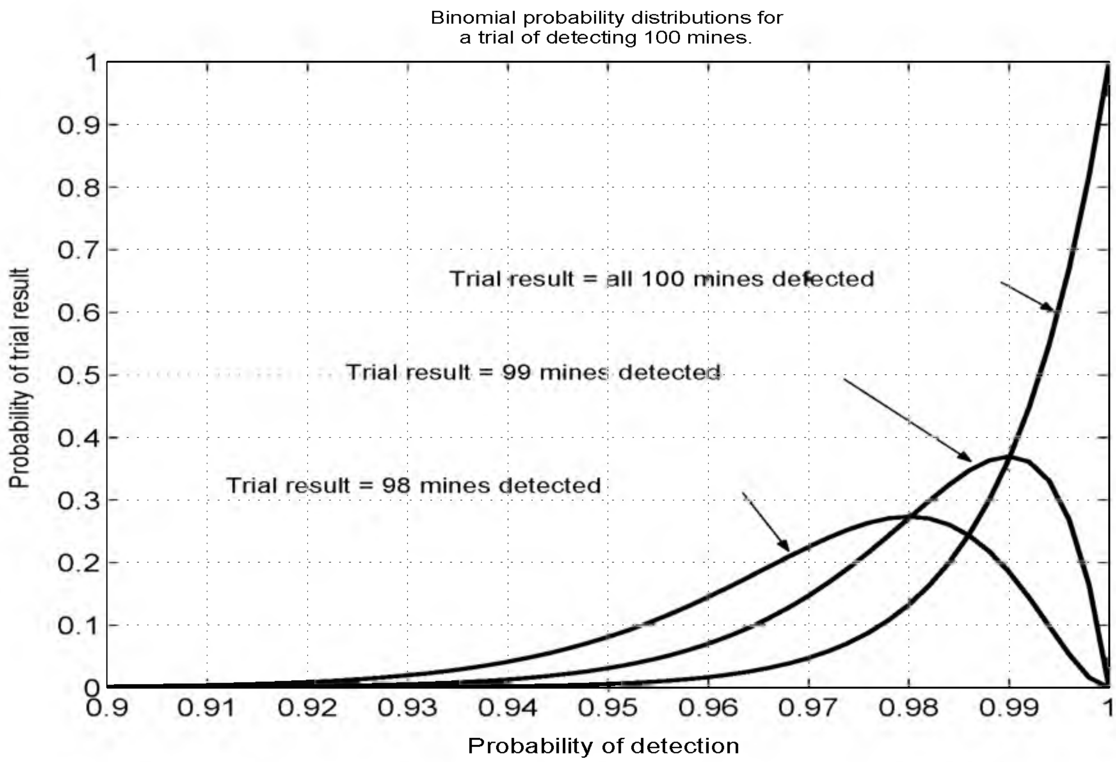

Figure 1: **Probability of result versus assumed probability of detection (fraction of mines that are detectable) for trials of 100 mines.**

make detection possible correctly installed) and four are undetectable (without the discs). It is obvious that a small sample, such as 20 mines, taken at random from the crate might well not include any undetectable mines. Clearly, it is not correct to conclude that because a small sample contains no undetectable mines that there are no undetectable mines in the crate. The question is, how many mines must be detected in a test to be sure that there are indeed very few undetectable mines in the crate, and what are the precise percentage probabilities?

3.3 Analysis using the Binomial probability distribution

The likelihood that a sample of size n, taken at random from a stock of mines having a proportion π of undetectable mines, will contain exactly x undetectable mines is given by the Binomial probability distribution

$$f(x; n, \pi) = {}_nC_x \pi^x (1-\pi)^{n-x}$$

where ${}_nC_x$ is the number of combinations of x events in n events.

Thus the likelihood that the sample will contain from zero to k (inclusive) undetectable mines is given by the cumulative Binomial function

$$F(k; n, \pi) = \sum_{x=0}^{x=k} {}_nC_x \pi^x (1-\pi)^{n-x}$$

Figure 1 shows the results of calculating the probability of the outcome of trials of attempting to find 100 mines, plotted against the probability of detection; the Binomial distribution relates the two probabilities as

$p_{result} = binomial(x, n, \pi)$ where x mines escape detection in a trial of n mines by a detector with a p_D of $(1 - \pi)$. Clearly if the detector is perfect (p_D=100%) the probability of finding all the mines is one, and the probability of failing to detect any number of mines is zero. When the p_D is 98% the most likely outcomes are one or two mines not detected.

In practice it is usual to discuss the case of a ***lower limit of probability of detection*** on the basis that any improvement in performance is welcome. When a detector is referred to as "having a p_D of 99%" the more complete statement is that the p_D is 99% *or greater* under certain conditions.

3.4 Using trial outcome to predict p_D

Predicting the outcome of a trial from a knowledge of the detector is less useful than using the results of a trial to find values for the probability of detection and the confidence in that probability. This is a standard problem in statistical quality control and is covered in text-books, for example Yamane [Yam73] who bases his text on the earlier work in the 1930s of Clopper and Pearson [CP34] and Barnes [Bar94] who presents a nomogram adapted from Johnson and Kotz [JK69].

Two hypotheses are tested. These are:
H_0 : The mine detector performs to the required standard.
H_A : The mine detector is defective.

The associated errors are α = $P(H_0$ rejected when true) and β = $P(H_0$ accepted when false). In quality assurance α is known as the producer's risk and β as the consumer's risk.

In testing mine detectors (1-α) is the confidence that the detector is accepted correctly from the test results (confidence in the result), and (1-β), known as the power of the test, is the confidence that an unsatisfactory detector will be rejected. In the case of a trial where all the mines are detected β is clearly meaningless as the detector has been shown to function with a theoretical maximum p_D of 100%. In trials where one or more mines are missed β can be used to define the confidence with which an unsatisfactory detector would be rejected, though this is likely to be so low that it is not useful.

In any distribution the parameter α can be visualised as the fraction of the area under the tail of the probability curve lying below the value p_{limit}. This is illustrated in figure 2.

In most statistical quality assurance the areas under both "tails" of the distribution curve contribute to the probability α; manufactured items that are oversize as well as those that are undersize should be rejected. In mine detection there is no meaningful interpretation of a detector that is "too good" so α is exclusively the probability that the detector is not good enough, the area under the curve below the required minimum value of the probability of detection. This is the lower limit of the possible range of values of p_D for that value of α.

An arbitrary decision has to be made as to the relative magnitude of the p_D and confidence (1-α) in order to analyse test results. This decision can also be visualised as moving the vertical "decision" line to the left or to the right on the graph of the probability function. As the line moves leftwards the area under the curve to the right of the line increases illustrating that as the minimum probability of detection required is reduced the confidence that this can be achieved increases.

Direct calculation of the Binomial distribution is straightforward and can be performed rapidly using a digital computer. There is no longer any need to employ analytical methods including further assumptions to reduce the problem to a form that is more readily calculable; this was necessary until computer power

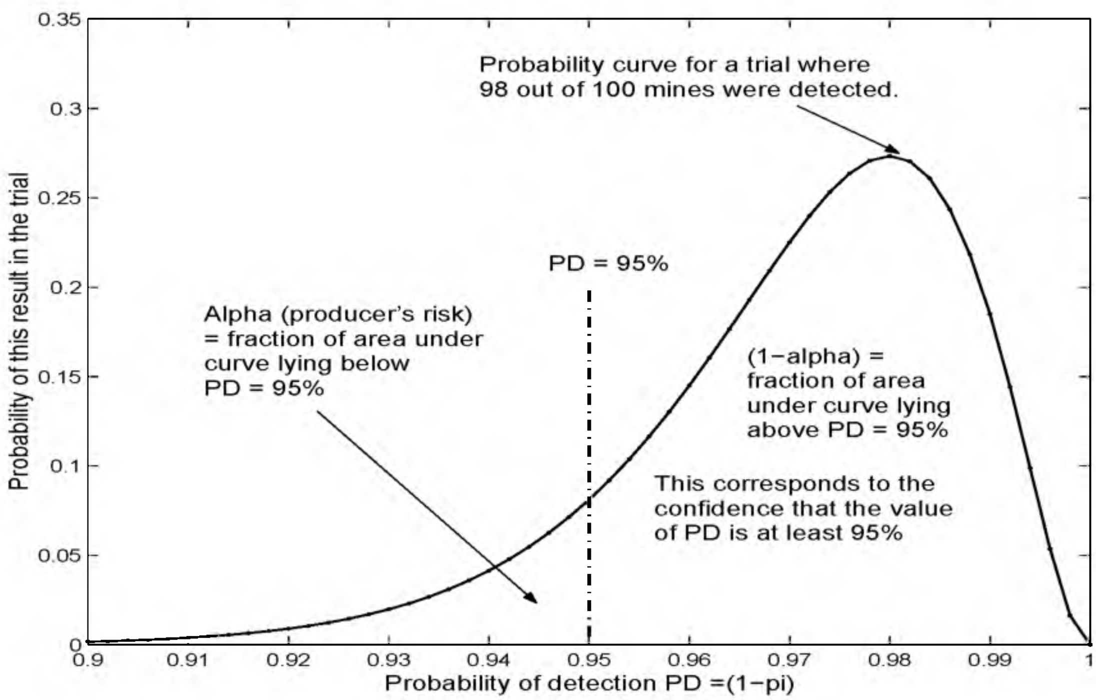

Figure 2: **Probability of result versus probability of detection for a trial of 100 mines, to illustrate α for a p_D of 95% .**

became cheaply available in recent years and was therefore widely covered in textbooks. The solution to the calculation for evaluating p_D from the results of a trial of a detector can be performed by a direct numerical approach of seeking solutions that fit. Although this may appear clumsy and inelegant it produces unequivocal results without the need to make further assumptions about the data, and once the computer programming has been done it is a very quick method. To avoid computational difficulties that arise when the sample size is large, the result can be calculated using the identity $b(x; n, \pi) = b(n - x; n, 1 - \pi)$.

Numerical solution of the Binomial approximation leads to the conclusion that Voles [Vol98] may have misinterpreted one of his own assumptions. Not only is a numerical solution to the case of no undetected mines possible (which is not the case for his method) but the results obtained for the case of zero undetected mines are the same as for the results he offers for the case of zero or one undetected mines. It appears that his assumption that the number of mines undetected is *less than or equal to one* should in fact be that the number undetected is *strictly less than one.* The interpretation of the limiting case is that the mine detector just managed to find all the mines and with any reduction in p_D would have missed a mine, instead of the case that the detector just failed to find one mine and any increase in p_D would have led to finding all the mines.

3.5 Selection of p_D and confidence

The selection from the results of an appropriate p_D and its associated confidence depends on the circumstances of the testing. The result of a trial is the set of points that form the curve of p_D versus (1-α) for the number of mines detected and undetected, as shown in figure 3. While investigating a new detection

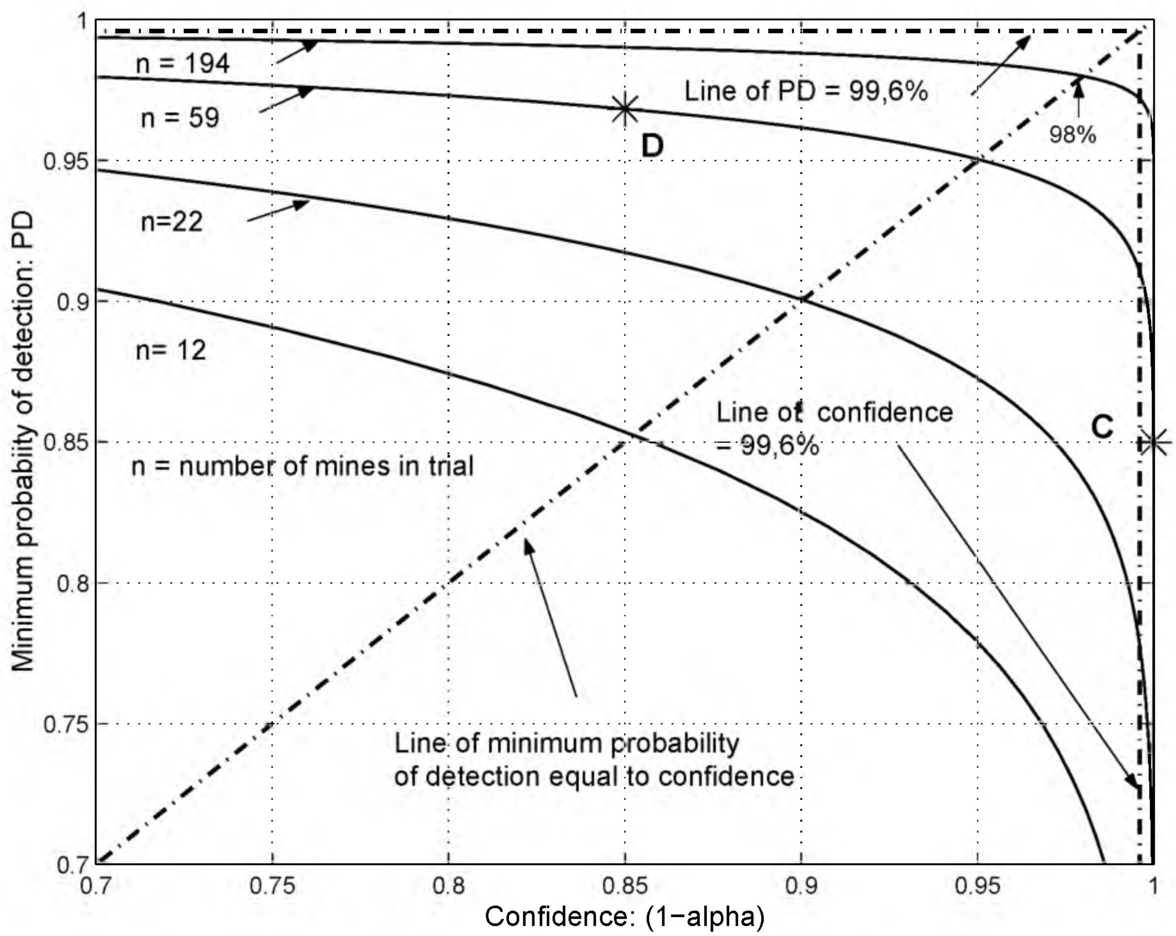

Figure 3: **Minimum probability of detection vs confidence in the result, for trials in which 100% of the mines are detected.**

Confidence (1-α) = Minimum probability of detection, %	**Number of mines in trial**		
	All mines detected	1 mine not detected	2 mines not detected
50	1	3	7
75	5	10	18
80	8	14	18
85	12	22	35
90	22	38	59
95	59	94	139
98	194	290	411
99,0	459	662	913
99,2	602	861	1177
99,4	851	1204	
99,6	1378		

Table 1: **Number of mines required in trial for a given minimum probability of detection, when minimum probability of detection is numerically equal to confidence in the result.**

method it may be initially useful to build prototype equipment that is known to have an unacceptably low p_D in order to investigate how, under field conditions, this p_D varies with soil type, temperature, moisture or other factors. In this case a higher confidence and lower p_D would be appropriate. On the other hand, during detailed testing of a detector that is known to perform extremely well setting the p_D to the desired performance level and calculating the resultant confidence level might appear a better strategy.

Figure 3 shows that the curve is not symmetrical about the line of equality (p_D = (1-α)) and whereas in a 100% successful trial of 59 mines a probability of detection of 85% can be stated with practically 100% confidence — point C on graph, a confidence level of 85% corresponds to a p_D of only 97% — point D. This serves to re-emphasise the difficulty of obtaining any meaningful results from measurement of the rate of detection in reasonably sized trials. The artificial separation of the two parameters of a test result, and the difficulty of the concept of confidence compared to the relative simplicity of probability of detection, lead to correct, but misleading, claims of such figures as 95% or even 100% success in trials. A p_D of 100% can only occur when the confidence is zero but the asymmetry allows confidence levels of 100% to be closely approached at values of p_D of well over 50%.

A useful approach to reduce the difficulty of presenting the results of testing mine detection equipment is to calculate a single value for both p_D and confidence by using the point where the two are numerically equal. The locus of these points is shown on figure 3 as a diagonal line. Solving numerically for the limiting case of $(1-\alpha) \geq \pi$ (i.e. confidence equal to or very slightly greater than probability of detection) gives the results shown in figure 4. Table 1 cites examples from the data. A trial of reasonable size can be seen to have p_D well below 100% by this method, finding 22 mines and missing none in a trial has a p_D (= confidence) value of just over 90%. In practice this is probably a more useful measure than expressing the same result as, for example, "99,6% p_D at a confidence of just over 8%." A small improvement to the equipment is unlikely to change the 99,6% p_D figure by a significant amount but will be clearly indicated by an increase in the value of joint value for p_D = confidence.

Table 1 shows the size of a trial needed to demonstrate a probability of detection and confidence level both equal to 99,6% to be 1 378 mines successfully detected and none missed. This is completely unrealistic.

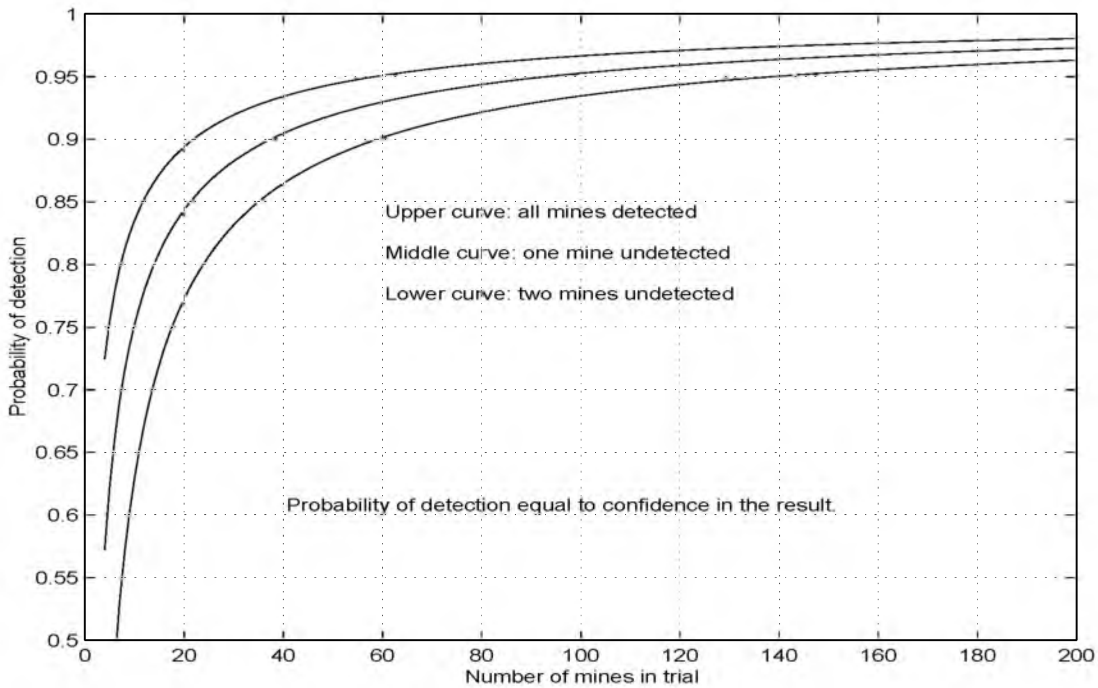

Figure 4: **Probability of detection vs size of trial when the minimum probability of detection is numerically equal to the confidence in the result.**

4 Qualitative factors

Testing mine detectors "in the field" under realistic circumstances introduces so many variables that trials of a size that are feasible cannot be expected to yield results within an order of magnitude of the desired failure-to-detect rate. More information than just the crude detection rate must be included, either in a comparative manner or analytically, if the results are to be useful.

4.1 Limitations of "crude detection rate" data

The approach of defining a single overall detection figure determined by limited testing is not only unworkable in practice but conceptually flawed. Manufacturers of safety-critical systems such as passenger aircraft or nuclear power plants do not define the probability of failure of their products by waiting for occasional failures and then projecting a probability of failure rate from the results. Whilst the development of an exact method for testing humanitarian demining equipment is beyond the scope of this working paper, it is clear that mine detectors are not the only safety-critical equipment that requires testing, and that sophisticated methods have been developed to deal with this situation in other industries. The most distinctive characteristic of humanitarian demining is the wide variability of the environment. This requires focusing on the development of test methods and procedures that are as insensitive to the environment as possible. In this regard a crude detection rate of 99,6% is a very poor measure as it depends heavily on the environment, and on the training, skills and supervision of the operator and not just on the performance of the mine detector.

4.2 Human factors

The limiting sensitivity of a mine detector can depend in part on the skills and hereditary characteristics of the operator. The ability of different people to detect a small change in the pitch of a tone is known to vary by more than an order of magnitude [VLFM99], yet many metal detectors use a change of pitch to indicate a target.

The operator interface of mine detecting equipment has a large impact on how the equipment is used and ultimately the probability of detection. There are many non-technical aspects to designing equipment with an intuitive feel; this topic is beyond the scope of this working paper.

4.3 False alarms and unwanted alarms

Analysing the results of testing humanitarian demining equipment requires a workable method for dealing with false alarms. Taking once again the example of metal detectors, increasing the sensitivity increases the probability of detecting a minimum metal mine, but also increases the number of small pieces of scrap metal detected. In areas where it is considered that there are no minimum metal mines, deminers are known to reduce the sensitivity of the metal detectors to decrease the false alarm rate. Yet in doing so they are making a decision to reduce p_D.

At present no distinction is made in defining the false alarm rate between mine-like objects that are identified by the detector as possible mines, and false alarms due to interference or sensitivity to non-mine-like objects. The statistical analysis presented above does not include any analysis of false alarm data.

For example, a small piece of scrap metal found by a metal detector is considered a false alarm just as an area of highly mineralised soil that triggers the detector is considered a false alarm. For the purpose of analysing the detector's performance these are two separate cases. The small piece of metal might be the firing pin of a minimum metal mine that had already detonated; clearly it falls within the range of items that should be detected by this method and as such is an unwanted alarm but not a false alarm. To distinguish an unwanted alarm from a mine, a separate detection method must be used which seeks to identify a different characteristic of the mine such as its dielectric constant in the case of radar systems, the presence of explosive in the case of olefactory and quadrupole resonance systems, or the physical presence of a mine case by prodding and excavation.

To a deminer both types of alarm are a nuisance. However, for the purposes of analysis one is an unwanted but correctly identified alarm and the other is truly a false alarm. If false alarms can be clearly distinguished from unwanted alarms then the high unwanted alarm rate of metal detecting can be used to provide an improvement in the statistical basis of quality assurance. By considering all unwanted alarms as valid targets in a statistical analysis, the number of targets is increased to a level at which a measure of the crude rate of detection can begin to offer some meaningful results. However, this is not enough of a benefit to outweigh the nuisance of large numbers of unwanted alarms.

Currently, it is common to have a detector with the highest possible probability of detecting both mines and unwanted alarms in order to reduce the number of mines not found (i.e. p_D is at a maximum). A subsequent discrimination method is then used to distinguish between mines and unwanted alarms. The common secondary method in humanitarian demining is to excavate manually and look for a mine, which is costly and tedious. The maximum usable p_D may in practice depend not only on the equipment under test but also on the associated secondary discrimination method and the operating procedures used. This makes a meaningful statistical measure of p_D more difficult to define.

Area cleared, m^2	Mines found	UXO found	Metal fragments found
19 489	5	7	20 015
76 264	6	9	35 931
35 290	6	93 (in 20 groups)	72 220

Table 2: **Number of mines/UXO and scrap metal items found in three cleared areas in Cambodia.**

5 Quality Assurance

Sampling methods of quality assurance (QA) are not appropriate to demining, yet they have been used. Some humanitarian demining organisations have, for example, carefully rechecked a part of a cleared area to be sure that they are confident that the demining has been done thoroughly. The commonly adopted current practice of removing all metal fragments from an area and then performing a check that it is entirely free from metal is a non-analytical form of QA.

The scarcity of mines in a mined area can be seen from the data from three mined areas visited in Cambodia in 1999, presented in table 2. It is clear that with any reasonable probability of detection there will be few, if any, mines/UXO overlooked. If one mine has been overlooked then the chance of it being found in a sampled QA test area is minimal — if 1% of the entire site is re-sampled then the chance of finding the missed mine is 1 in 100. If there were five mines found in the entire area and one not detected this amounts to a probability of detection, p_D, of only 83%. QA methods must demonstrate a high probability of being able to detect such a low p_D, certainly far more than just 1%.

Statements about the required clearance rate of 99,6% lack value if only five mines have been found; the confidence level with p_D equal to 99,6% in a trial of five mines, even when they are all found, is tiny — less than 2%.

There are also important issues concerning the need for an independent mine locating technology to be used for QA (i.e. one that is unrelated to the primary detection method) to avoid missing the same mine for the same reason. One of the more promising methods is the use of dogs or artificial noses which can tell if explosive vapours are present in an area without necessarily being able to locate any individual mine. Any suspect area can be rechecked by hand.

In practice, the only way to guarantee mine clearance to a very high standard is to introduce not post-clearance sampling methods but methods that evaluate the performance of the demining operation as it is taking place. One way of doing this by using the concept of the "margin of detection" is proposed in section 6 below.

The strict operating procedures and supervision of manual demining are a method for maintaining an adequate quality of clearance. It is generally not the quality of current clearance methods that is a problem, but the time and cost of achieving the required standard and the lack of a way to guarantee that it has been achieved.

6 "Margin of detection"

In a typical trial two metal detectors may both find a set of targets without failure. However, connecting suitable measuring equipment, such as an oscilloscope, may show that one is at the limit of its ability to distinguish the targets but the other has a substantial reserve of performance and could still detect the targets under substantially more demanding conditions. Clearly their performance is not identical but the crude detection rate does not distinguish between them. To do so requires the introduction of a measure of how close the "signal" from each mine is to the limit of detection of the equipment being used. An estimate of how easily the detector identified all the targets, or almost all the targets, adds considerable useful information to its evaluation.

Similarly, in an area being cleared of mines/UXO, if all the mines already located have been found easily, and the geographical conditions are similar throughout the area to be cleared, then it is possible to be confident that further mines of a similar type and depth could be found readily. However, if the mine detector had been functioning at the limit of its performance while one or more mines were detected then it is possible that a mine buried slightly deeper, or one encountered by a tired deminer at the end of a working shift, might be missed.

Thus the concept of the ease of detecting of a target, or the "margin of detection" is one way to resolve the problems of statistically meaningful testing of mine detectors as well as QA.

A further weakness of testing methods that rely solely on the crude maximum detection rate is that a crucial evaluation is made when the desired signal from the mine/UXO is only just distinguishable from the "background" which is noise, clutter, interference and other undesired signals, depending on the detector type, soil, vegetation and other factors. The measurement of a signal barely different from the background is unlikely to give reliable or repeatable results. To avoid this it is common throughout engineering to use methods of extrapolation; the signal is measured under less critical conditions and a curve fitted to the results which is then extrapolated to define a point at which detection is just possible. The goodness of fit of the curve can be analysed statistically to provide measures of confidence and probable error limits. Individual manufacturers of mine detectors may well be using this method to enhance their products, what is required is a more general technique which can be used to compare different detectors.

A suitable measure of the margin of detection might be the ratio of signal to background noise at some point in the detector circuit. For complex detection methods a measure of the effective "signal to noise ratio" can be made from probabilistic considerations. However, in defining the conditions of the measurement of the margin of detection the number of variables is large. The sensitivity of the detection process to some of these variables is also large; in the case of the limiting distance to a small metal target in air, the metal detector's received signal depends on the inverse fourth power of the distance from the detection coil to the metal fragment. Moving from 100 mm to 110 mm causes a reduction in the received signal of 32%. A small pebble on the surface of a test area can clearly cause significant variation in the results depending on whether operators touch the pebble and push it aside or raise the detector over it.

The margin-of-detection parameter attempts to give a readily understandable result for a wide variety of targets. It can be evaluated by presenting a range of known targets to the detector under controlled conditions. Extrapolation of these results should give an acceptable estimate of the limits of the performance of the detector under ideal conditions. The performance of a known detector in the field can then be used to define the effect of the field conditions using the ratio of margin of detection. For example, a soil with a high metal ore content (e.g. laterite) might be considered three times more difficult for detection using a metal detector than a sandy soil, or dry sand might be considered three times more difficult for a radar

Ratio of the power of the signal to the power of the background noise	Ratio expressed as decibels (dB)
100 000	+50
10 000	+40
1 000	+30
100	+20
50	+17
20	+13
10	+10
2	+3
1	0
0,5	-3
0,1	-10
0,01	-20

Table 3: **Power ratios expressed as decibels (dB).**

system than moist soil.

These ratios may be conveniently expressed in decibels (dB) which are defined as $10 \times log_{10}(ratio)$. Table 3 gives decibel values for some ratios of the strength of the signal to that of the background, measured from the power of each.

In practice, an easy-to-detect target might be defined as 45 dB above background, and a difficult to detect target perhaps 5 dB above background using a standard mine detection system. A different detector that is being evaluated could then be measured and might perhaps give figures of 40 dB for the first and 6 dB for the second; this detector is more capable of finding the smaller target. Such ratiometric measurements allow direct comparison with existing "reference" equipment with which deminers are familiar, thus producing numbers with an immediate practical application.

Similarly, by inserting a standard target to a known depth, field conditions can be measured. A standard target might be described as producing, for example, a level 3dB above background in a particular area, whereas it is known to produce 5 dB above background under "standard conditions". This area thus has a background level 2 dB worse than standard conditions; the impact of this on the maximum depth of detection of small targets can be directly calculated. As most demining organisations check the operation of equipment at regular intervals against known targets this operation adds little extra effort to the work of demining.

6.1 Application of "margin of detection" to QA

When applied to quality assurance the concept of margin of detection has considerable appeal as it provides a working check on both the equipment and its operation.

If the depth of a target, either a mine/UXO or an unwanted alarm is known then the margin of detection can be used to provide a measure of the level of confidence of finding another similar target at the required depth of clearance in the same soil. If, in the case of a metal detector, the target is a small piece of scrap metal that is excavated, it can be compared to an objective detection performance in air by measuring the margin of detection for the object at a known distance. This can be readily measured by dropping the scrap metal into a plastic container of known depth and placing the metal detector on top of the container. For example, if a particular target gives a margin of detection of 10 dB under the test conditions and was found at a particular depth then predicting the probability of detection of known minimum metal mines at a similar depth should be possible. Many of the factors that affect detector performance at the particular site can be combined into a single measure by following such a procedure.

By maintaining a record of all items found and the corresponding margin of detection, the performance of the equipment and operator can be continuously evaluated and a statistically meaningful quality assurance may be possible.

6.2 Use of unwanted alarms and margin of detection to enhance QA

Quality assurance could be immediately enhanced by introducing a simplified form of measurement of the margin of detection. This can be directly implemented with existing detectors. In terms of metal detection, small metal fragments can be characterised after excavation by the distance at which they can be detected in air; this gives a simple measure of their "detectability," and depends on their size, shape and composition. If the depth at which they were detected in the soil is noted, the limits of detection for the particular combination of soil, mine detector and operator can be approximately categorised. From this simple analysis it is possible to verify that a target similar to a minimum-metal mine would be found at a depth that would give adequate safe clearance of all mines. Research into this method should be able to determine if a rule of thumb that is easily memorised can be deduced, or a simple tool based on a nomogram developed. Simple categories that can be readily coded as "acceptable'," "marginal," and "unacceptable" could be used, with appropriate colours or symbols.

Overall, measurement of the margin of detection is a simple and powerful tool that could be applied to both detector evaluation and quality assurance methods in demining.

7 Summary and conclusions

Producing useful results from the evaluation of mine detection systems is surprisingly difficult; testing prototype equipment of unproven reliability in live areas may be hazardous, and testing in safe areas offers limited realism.

The provision of tests of sufficient size to give a statistically meaningful outcome is difficult and the impact of the test conditions and operator training may dominate the results.

Measurement of the crude detection rate of mine detectors in realistically sized trials fails to offer meaningful information on their performance. This paper presents a rigorous statistical analysis that demonstrates this, using a clear conceptual model.

An alternative to quoting *probability of detection*, p_D, and *confidence in the result*, (1-α), data separately is suggested and examined. Combining these two aspects of the result by using the value of p_D that is numerically equal to (1-α) provides a simple way of expressing the two crucial measures of a test result in a single figure, and permits a simple and direct comparison between different tests.

The large impact of qualitative factors such as the effects of the environment and the operator are discussed, and the need for test methods that are not highly dependent on these factors is explained. This is particularly important when using crude detection rate data.

The problem of achieving a statistically significant method of quality assurance after demining is examined and found to be so severe that such post-clearance QA is infeasible.

The concept of "margin of detection" is offered as a potential way to resolve some of these problems. Its advantages and implementation are discussed in relation to both testing mine detectors, and QA during clearance. Simple ways of implementing a crude form of margin of detection that are compatible with existing equipment and operating procedures are outlined, and the need for further research into possible ways of implementing the concept identified.

References

[Ano00] Anonymous. Croatia Sets Feb. 11 "ELF" Lab Tests to Validate Loch/ChemTech Landmine Detector. Internet Website http://biz.yahoo.com/bw/000201/tx_loch_ha_1.html, Feb 2000. Valid as of 12 Feb 2000.

[Bar94] J Wesley Barnes. *Statistical Analysis for Engineers and Scientists.* McGraw Hill, 1994. ISBN 0 07 839608 5.

[CP34] C J Clopper and E S Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, (26):404–413, 1934.

[JK69] Johnson and Kotz. *Discrete Distributions.* Wiley Interscience, 1969.

[MF85] Irwin Miller and John E Freund. *Probability and Statistics for Engineers.* Prentice-Hall, third edition, 1985. ISBN 0 13 711938 0.

[VLFM99] N L Vause, T R Letowski, L G Ferguson, and T J Mermagen. Auditory issues in handheld land mine detectors. In Abinash C Dubey, James F Harvey, J Thomas Broach, and Regina E Dugan, editors, *Detection and Remediation Technologies for Mines and Minelike Targets IV*, volume 3710, pages 778–786. SPIE, Apr 1999.

[Vol98] R Voles. Confidence in the assessment and use of mine detection systems. In IEE, editor, *The Detection of Abandoned Landmines*, volume 458 of *Conference Publications*, pages 28–30. IEE, Oct 1998.

[Yam73] Taro Yamane. *Statistics, An introductory analysis.* Harper and Row, third edition, 1973. ISBN 06 047313 4.