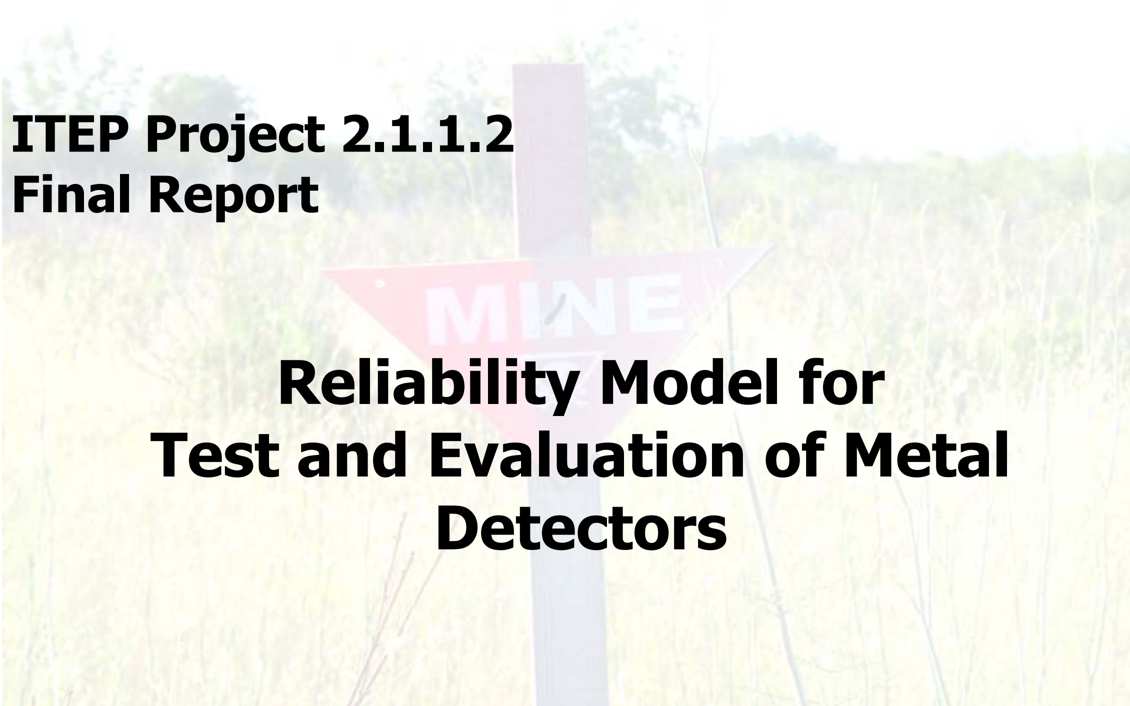

ITEP Project 2.1.1.2
Final Report

Reliability Model for Test and Evaluation of Metal Detectors

Project leader: Christina Mueller, Federal Institute for Materials Research and Testing (BAM), Berlin

Mate Gaal, BAM
Martina Scharmach, BAM
Uwe Ewert, BAM
Adam Lewis, Joint Research Centre of the European Commission (JRC), Ispra
Thomas Bloodworth, JRC
Peter-T. Wilrich, Freie Universität, Berlin
Dieter Guelle, ITEP Secretariat

Berlin, September 2004

Table of Content

1. Background

Metal detectors used in demining have been tested within the frame of the International Test and Evaluation Programme project ITEP 2.1.1.2. Detection reliability tests were performed in accordance with the CEN workshop agreement describing specifications for test and evaluation (CWA 14747:2003) (Ref. 1). Results of these tests are aimed to verify and should help to optimise the proposed testing procedures.

The CEN Workshop Agreement CWA 14747:2003 standardises methods for test and evaluation of metal detectors for humanitarian demining. It covers in-air and in-soil laboratory sensitivity measurements, immunity to operational conditions and electromagnetic interference as well as ergonomics, shock and bump tests etc. Reliability trials are covered by Section 8.5 of the CWA.

In that document detection reliability is defined as "the degree to which the metal detector is capable of achieving its purpose, which is to have maximum capability for giving true alarm indications without producing false alarm indications." An overall reliability of a mine detection system (R) can be understood as a result of three factors: an intrinsic capability (IC) describing the physics and basic technical capability of the devices and representing an upper limit of R, factors of application such as special environmental conditions in the field (AP) generally diminishing R and finally the human factor (HF), which lowers R. All three factors are described in a concept called the Reliability Model (Ref. 2, 3, 4):

$$\mathrm{R} = \mathcal{F}\big(f(\mathrm{IC}), g(\mathrm{AP}), h(\mathrm{HF})\big) \qquad (1)$$

It is important to bear in mind that there may be interactions between the factors. In humanitarian demining the influence of the last two factors has already been recognised as very important, since the conditions in the field and the behaviour of the operators have proven significant impact on the overall performance. Only the Intrinsic Capability and a part of the Application Factors are determined in laboratory measurements. The overall reliability, including the Human Factor, can be evaluated only in blind trials.

In a reliability test, targets are placed in metal free lanes at positions not known to detector operators. While scanning, the operators mark the places of indications and, later, supervisors measure and record the spatial co-ordinates of the markers. A target is considered to have been detected when a marker is dropped within a prescribed radius ("halo") around the true target location.

CWA 14747:2003 makes recommendations about lane widths and soil depths, soil types, target types, numbers, depths, orientation, separation and halo size and gives some practical instructions about lane preparation. Reliability tests described in this report were performed to determine the optimum choice of all parameters for a reproducible and repeatable trial, which would provide knowledge about true performance of metal detectors under field conditions.

In December 2003 participants of the workshop on Reliability Tests for Demining discussed the method and the results of these trials (Ref. 5). The conclusions of the workshop will form a basis for a proposal of an addendum to CWA 14747:2003, which will thus include practical experiences of the trials described in this report.

2. Project Overview

Three sets of reliability trials were conducted at test sites of the German Federal Armed Forces at Military Engineering Department 52 (WTD 52) in Oberjettenberg, Germany and at test sites of Croatian Mine Action Centre – Centre for Testing, Development and Training (CROMAC-CTDT) in

Benkovacko Selo near Benkovac, Croatia. Laboratory measurements were performed at Humanitarian Security Unit of Joint Research Centre of the European Commission (JRC) and on the test sites. This section gives a summary of all activities and all factors involved in the trials. A comprehensive overview is given in Table 1. More details and results of the trials are presented in the next sections.

Table 1: Overview of three series of trials.

Place	**Oberjettenberg**	**Benkovac**	**Oberjettenberg**
Date	May 2003	July 2003	November 2003
Lanes	4 lanes of 20m^2	8 lanes of 30m^2	7 lanes of 20m^2 (4 lanes same as May, 3 new ones)
Soils	1 artificially uncooperative 3 neutral (not completely metal-free)	3 types: Neutral Uncooperative homogeneous Uncooperative heterogeneous	New lanes: 1 strongly uncooperative 2 neutral (all metal-free)
Detector Models	4	4	5
Detector Specimens	2	2	2, except one manufacturer, which provided 1 specimen of 2 different models
Sensitivities	2	2	1
Operators	8 soldiers (inexperienced)	8 deminers (3 currently active)	8 soldiers (inexperienced)
Training	Brief	Brief	Extended
Starts per day	8	8	6

Four detector models tested in the trials were European-manufactured models designed for demining and currently in use in many mine affected countries. Manufacturers participating in the trials are listed alphabetically: CEIA, Ebinger, Foerster and Vallon. It was agreed that all detector models were kept anonymous, since the purpose of these trials was not a comparison between metal detectors, but optimisation of the testing conditions. A new model from one of the manufacturers was introduced in the last set of trials, i.e. in November. In that set of trials only one specimen of this model and one of the older model of that manufacturer were tested.

Detector models used in these trials operated on different physical principles. Some of them were time domain detectors (also called pulse induction detectors), some frequency domain ones (also known as continuous wave detectors). The shapes of their coils were also different: some of them used a single coil, some a "double-D" configuration. Some of the detectors were static mode detectors, some dynamic mode ones. They all had some data processing to compensate the soil background effects.

Laboratory measurements were performed at JRC (Joint Research Centre) to determine maximum detection distances in air and in soil for different targets, as well as soil electromagnetic properties (Ref. 6). These measurements were continued in the field, where blind tests were already in progress.

The first set of trials in Oberjettenberg (May 2003) was performed on four lanes, with each one containing a different type of soil. One of the lanes was covered with a 2-cm layer of blast furnace slag to emulate uncooperative soil. Eight lanes in Benkovac (July 2003) consisted of three different types of soil found in mined regions of Croatia. Two of these types were highly uncooperative. In the last set of trials in Oberjettenberg (November 2003), seven lanes were used. Four of them were the lanes from May trials, and three additional lanes contained three new types of soil. Test lanes were cleared of metal debris with the aid of detectors, to the extent achievable in the limited time of test preparations.

The targets used in the trials were real mines modified to be safe and ITOP standard targets (defined in the International Test Operations Procedure, Ref. 7). Mines used in Oberjettenberg trials are encountered in minefields all over the world and those used in Benkovac are typical for minefields of south-eastern Europe.

In Oberjettenberg the operators were soldiers without previous demining experience. In the Benkovac trial the operators were experienced deminers, three of them currently active as deminers. A new training scheme was applied in the last trials, with fewer encumbrances for the operators.

The Oberjettenberg May trials and the Benkovac trials were performed with two sensitivities. The higher sensitivity was the highest achievable sensitivity of devices and the lower sensitivity was calibrated with a standard target buried to a specified depth. Trials in November were performed only with high sensitivity.

A concept for statistical evaluation of results and design of the test is proposed. A non-linear regression model was used to describe the performance of metal detectors in dependence on depth (Ref. 5). Orthogonal design of the test enables influences of different factors to be distinguished with significantly fewer repetitions than a full factorial design with all combinations of all factors would require (see Section 4, Orthogonal Design). In each of the trials a total number of 256 passes was performed, which means 64 per each detector model. A typical number of targets in a lane was 25 or more, which gives at least 1600 opportunities to detect a target with each detector model.

After reliability trials an additional trial was performed, when operators investigated all signals and excavated all targets and sources of alarms. This test was carried out to get closer to the clearance process in a real minefield.

All operators answered a questionnaire about the detector models they have used. The questions were based on the questionnaire of the Afghan trial in 2002 organised by the UN (Ref. 8), and they varied from ergonomic aspects to the overall confidence in the device. The questions and the results are in the Annex 9 of this report.

Tables with complete schedules of all three trials, including lists of all targets with their positions, are also available in the Annexes.

3. Summary of the Project Strategy

CEN Workshop 07 began the process of standardizing test and evaluation methods for Metal Detectors in Humanitarian Demining, including both laboratory measurements of detection capability and blind field trials (reliability tests) to measure Probability of Detection versus False Alarm Rate. The conclusions were published in June 2003 as CEN Workshop Agreement 14747. One of the authors of this report (T. Bloodworth) served as the Secretary of the Workshop and several of the other authors contributed to the discussions. The project described in this report was begun before the final publication of CWA 14747 with the aim of better understanding how to specify the blind trial set-up and the statistical rules necessary to achieve true, repeatable and reproducible results under representative field conditions. It was conducted under the umbrella of ITEP as Project No. 2.1.1.2. The method adopted was to perform a series of trials in which a relatively small number of detectors would be tested very thoroughly, so that any sources of statistical or systematic error would be revealed without ambiguity. The trial scenarios ranged from straightforward detection of a large, metallic anti-tank mine, buried near the surface in a soil that does not give metal-detector signals, to the most difficult challenge of detecting low-metal antipersonnel mines, deeply buried in magnetic soil that affects detectors strongly. Individual human factors, such as training and currency of skills were assessed.

In order to ensure that the requirements of practical demining were met and that the analysis was done on a sound scientific basis, the authors organized an international workshop to discuss the problems of reliability test trials in December 2003.

3.1. Workshop on Reliability Tests for Demining

About 100 international experts in demining met for the "Workshop on Reliability Tests for Demining". The proceedings (Ref. 5) contain presentations of the oral sessions where the general national, European and international concepts in demining are described as well as the main activities and results of the ITEP-trials. An up-to-date series of "Lessons learnt and problems to be solved" was presented from international Mine Action Centres. In four focused sessions the authors and a number of competent international experts discussed the following specific topics: Configuration of test lanes and test target selection, Soil influence and ground compensation, Human Factors, and Rules for test planning and statistical evaluation.

A highlight was the second session, which addressed the problem of soils that influence metal detectors - such soils are described variously as "noisy" (CWA 14747:2003), "uncooperative" or "difficult". These effects were recognized to be due principally to magnetic properties of the soil; both the magnitude of the magnetic susceptibility and also its frequency dependence (see especially the presentation by S Billings et al, Ref. 5). The fundamental magnetic properties were related to the empirical "ground reference height" measurement, developed by D. Gülle: the maximum distance above the ground at which a calibrated, static-mode detector gives an alarm due to that ground.

Further presentations dealt with conclusions for future practical activities, such as the GICHD Manual Demining study (T. Lardner) or a world-wide accident data base (A. L. Smith). One of the conclusions for future research requirements was that there was still a need to get a more comprehensive understanding of soil influences (S Billings et al). Finally, the Workshop assembly expressed "Findings and Recommendations" with recommendations for how to deal further with the topic of reliability and with modelling for the improvement of demining techniques.

3.2. POD and ROC – Summary of Detection Rates and False Alarms

The ROC (Receiver Operating Characteristic) of a mine detection system (Ref. 2) shows the detection rate or probability of detection versus the false alarm rate or number of false alarms per unit area (Figure 1). The ROC shows how well the system discriminates between signal and noise. It shows how successful the system is in distinguishing between a signal from a mine and a noise signal arising from any other possible perturbation (from the soil, from other buried artefacts, from the electronics). The closer to the upper left corner the position of a ROC point is the better is the system.

When land is cleared of mines where minimum-metal mines are the main threat, the "metal free" procedure is sometimes used. This means that detectors are used on the maximum sensitivity possible and all metallic pieces found are removed from the ground. In trials for metal detectors to be used in this way, any metal piece found should be considered a true detection, not a false alarm.

In some mine/UXO clearance operations, relatively large metal objects are sought. In this scenario, it is often possible to reduce metal detector sensitivity to avoid detecting all of the possible metallic clutter that may be present, while still having the detection capability to find the targets. In trials designed for this type of operating procedure, it is possible to consider detection of extraneous small pieces of metal as a false call. However, the validity of this approach depends on the sizes of metal pieces in the test lanes. If metal pieces are present that have an equivalent response to the targets, then the test becomes rather meaningless because reporting these detections as false calls does not indicate that the detector is not performing as required.

A simple way of obtaining the detection rate curves is by plotting the mean values of the experimentally measured detection rates for each step of burial depth (Section 6).

For a fixed amount of false alarms the ROC point or operating point of the system for a fixed sensitivity can be taken and further analysed for its dependence on the main influencing factors like the mine depth or the metal content of the mine (Figure 2). All these points and curves need to be interpreted in connection with the corresponding confidence limits to consider the scatter of results. The latter scatter depends on the underlying statistical basis (the number of opportunities to detect the mine) and the natural variability of the factors. A simple way of obtaining the POD curves (also called detection rate curves) is by plotting the mean values of the experimentally measured detection rates for each step of burial depth (Ref. 9) (see Figure 16 and Figure 17 in Section 6). The smooth POD or detection rate curves, schematically presented in Figure 2, were determined by an advanced logistic regression model (see Section 4.2 and Ref. 5).

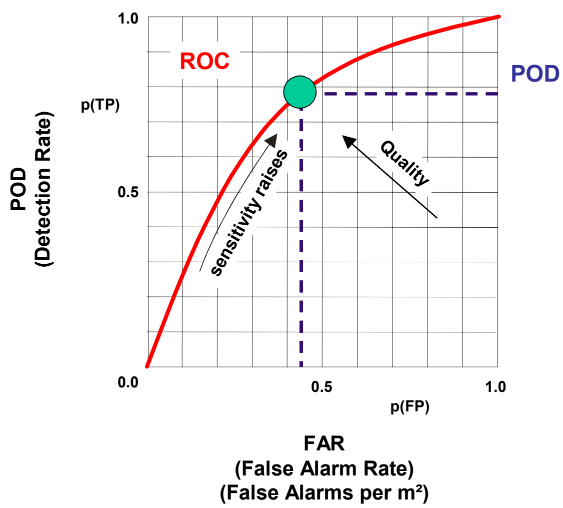

Figure 1: Explanation of ROC and POD diagrams.

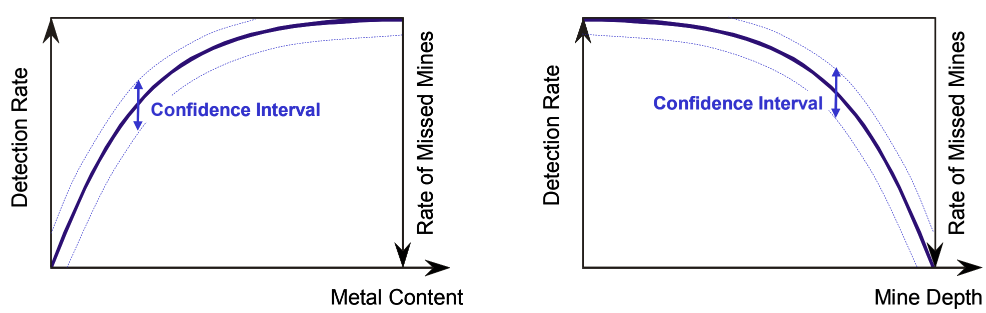

Figure 2: Schematic representation of typical POD curves.

3.3. Overview of the Parameter Matrix of the Trials

The main aim of the trials was to investigate how the device performance manifests itself in different application circumstances. The authors organized three sets of trials for which the main parameter set up can be seen in Table 1 and Table 2. The first and third took place in Oberjettenberg WTD 52 on the testing ground of the German Army.

The conditions for the first trial in May 2003 were representative of poor circumstances, likely to yield low performance: inexperienced operators with a short training period and test lanes with significant metal contamination. Three neutral soils were used and a fourth lane was artificially made "uncooperative" by adding a layer of magnetic blast-furnace slag. (With the benefit of hindsight, the authors would not recommend this technique because the slag was found to contain metallic particles, creating additional metal contamination). The buried mines were characterized by a large to medium metal content. Some generic "ITOP" targets were also used, irregularly distributed over a predefined depth range.

The second trial set was organized in Benkovac, Croatia with eight experienced Croatian operators, three of whom were active as deminers at the time of the trials. A brief training period (half a day for each detector) was given. There were three types of soil on eight lanes: neutral soil, homogeneous uncooperative soil and heterogeneous uncooperative soil. Both of the latter had frequency-dependent susceptibility. The mines had large, medium or very small metal content and were systematically distributed over a depth ranging between 0 and 20cm to allow statistical analysis. For testing metal detectors the normal target depth should be to the limits of the physical detection capability in the soil. The depth of 20cm was chosen because it is the required depth for mine clearance under Croatian law. The lanes were “almost” clean of metal pieces.

The lessons learnt from first two trials were applied to the third trial set in Oberjettenberg November 2003, with the intention of creating conditions likely to yield better performance. Three new lanes were set up, in addition to the ones available from the previous trial in May, and carefully cleaned of any metal fragments. Mines with large to medium and small metal content were selected and distributed systematically at a depth ranging from 0 to 20cm. The operators, who were inexperienced, were trained carefully in open and blind exercises until they were confident about the reaction of each detector to each mine in each soil and at different depths. To avoid confusion between the different detector operating procedures the operators were assigned during the training, as well as during the first week of the trial, detectors belonging to one class only (double-D coil, static mode or single coil, dynamic mode). In the second week they changed to the other class of detectors.

Table 2: Trial parameters

Devices	Soil	Mines	Human Factor
→2 pulse time domain U, X, W →2 continous wave Y, Z	Types of soil: →Cooperative (neutral) →Uncooperative (Frequency dependent; Constant susceptibility) →Metal contamination of the soil →Homogeneous / heterogeneous	Types of mines: →(metal content): biggest TM  smallest PMA 2  → Depths of mines	→working time →Training mode: →Brief and extended →Status of experience, pre-experience with one device type, age →Current activity →Personal capability

3.4. Results of the Trials

Figure 3 shows the overall results of each trial set, in ROC diagrams. These diagrams illustrate the influence of the factors (Application factor and Human factor) degrading the performance of all the detectors, without distinguishing between individual detectors. The result of inexperienced operators with a short training on metal contaminated ground shows a mean detection rate of 70% and 0.3 false alarms per m^2. The artificial uncooperativeness reduces the performance to 60% detection rate and almost one false alarm per m^2, which is surprisingly poor.

Even more surprising are the total overall results for Benkovac in June 2003, where the operators consisted of eight experienced Croatian deminers. The detection rate of about 65% in neutral soil decreases to almost 50% in a real, local, uncooperative soil with frequency dependent susceptibility.

The false alarm rate grows from 0.5 false alarms per m^2 to almost 0.6. Possible reasons for this extremely poor result are:

1) Many of the targets were very deeply buried and in some cases beyond the physical capability of some of the detectors. Minimum metal mines, which are inherently difficult to detect, were buried according to a systematic depth distribution, ranging from 0 to 20cm in order to evaluate the detection rate as a function of depth. The maximum depth of 20cm was chosen because it is the requirement of the Croatian clearance law. A more realistic mean value of detection rate for the region could be determined, if the real depth distribution of mines is known, by using the POD as a function of depth measured in the trial. Usually, AP mines are mainly buried at a depth ranging from 0 to 5cm, which is much shallower than the range used in the trial and would be detected with higher average POD than measured in the trial.

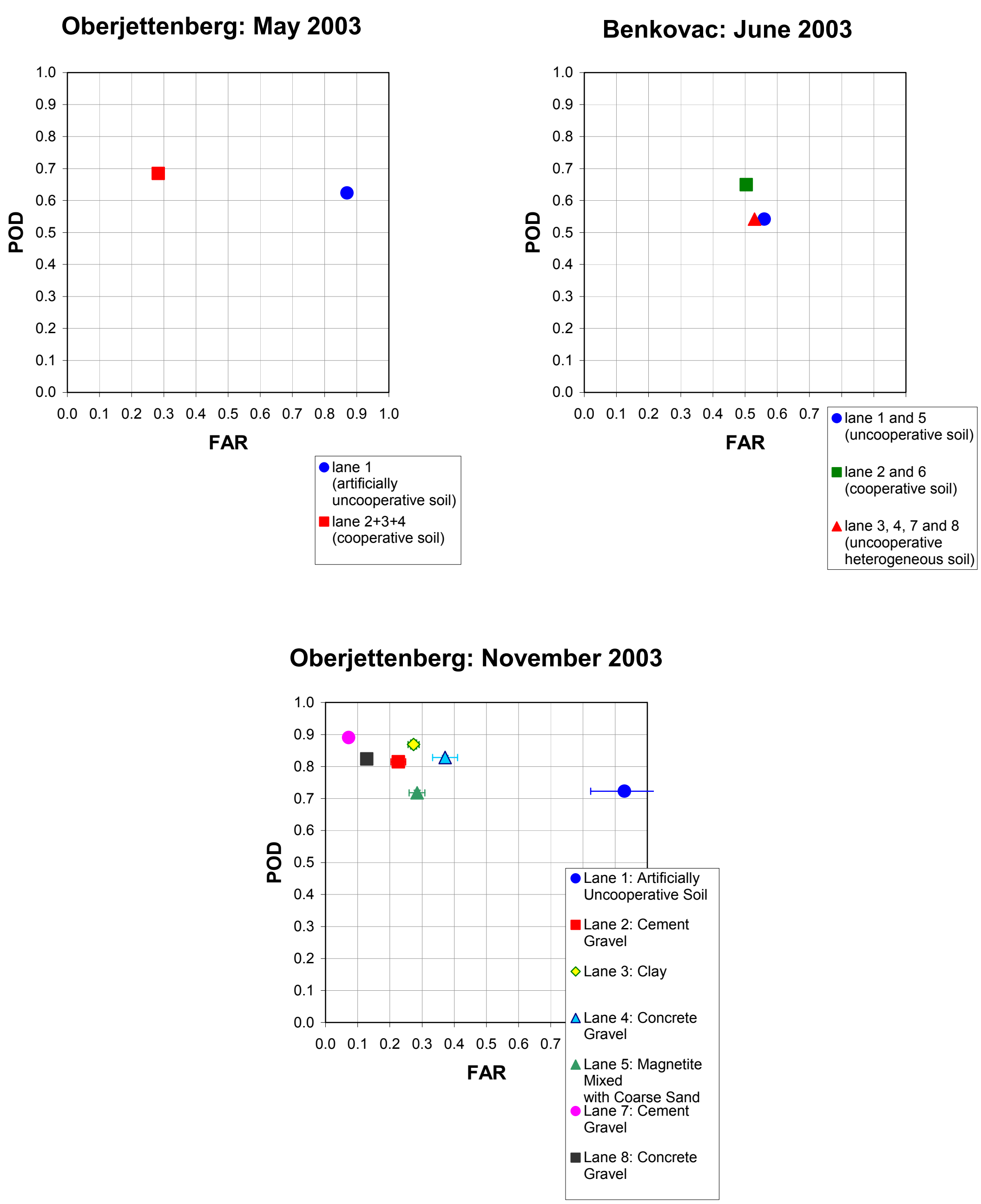

Figure 3: ROC diagrams for different soil, target and human factor conditions

2) Only three of the deminers are currently active.
3) It has been suggested that experienced deminers may need a longer training phase because they are generally accustomed to using a particular detector model and cannot handle too many different device types at the same time.
4) In the trial, the deminers are not in danger and are less motivated to be careful than they would be in a real minefield.
5) The test schedule required the deminers to work more quickly and for longer hours than they would normally do.
6) The test lanes were contaminated with metal.
7) Heterogeneous soil with strong frequency-dependent magnetic susceptibility is a challenge for all detectors, especially in combination with minimum metal mines, since the soil signals often mask the mine signal.
8) Investigation of the source of the signal was not executed as it is requested in manual clearance (see Section 3.5.).

The performance in the third trial is much better than in the first two, as expected from the conditions of the test with respect to the human factors and application factors. In Figure 3 Oberjettenberg November upper left corner the ROC point is 90% detection rate and false alarms below 0.1 per m^2. The "secret" is in carefully-conducted and longer training, reduced workload, neutral and very clean soil and targets that are easier to detect. If we want to estimate a realistic POD it is therefore necessary to ask what is the appropriate scenario of application and human factors for the situation we want to investigate.

3.5. Full Process Simulation

In Oberjettenberg in November one additional test was conducted, on the advice of Dieter Guelle (Ref. 5), which simulated the full manual demining process, including prodding and excavation. Since the statistical basis was too small to be representative, results of this test must be considered indicative only and any conclusions provisional. The detection rate of the manual clearance process appeared to be higher than of the detection process without excavation, probably due to instances where a minimum metal mine was hidden by a larger false-alarm item. Indications which could be assigned to identifiable metal fragments were excluded (according to a "metal free" approach), so the false alarm rate is lower. The latter is, of course, a matter of definition rather than performance. A more detailed investigation is planned within the GICHD study of manual demining methods mentioned above.

3.6. Example of a set of Resulting Curves: Detection Rates as Function of Depth and False Alarms for the PMA-2 in Different Soils

In the following figures the individual detector results are illustrated for the PMA-2 minimum metal mine under ideal conditions, i.e. neutral soil without metal contamination, well trained operators and optimized working hours. Figure 4a-d shows the detection rates as function of the burial depth for each device separately and Figure 4e shows the ROC points of all devices together.

Figure 5 presents the same results for the most difficult soil. A more detailed analysis is given in Section 6 Reliability Tests, including a correlation between detector parameters and POD and ROC.

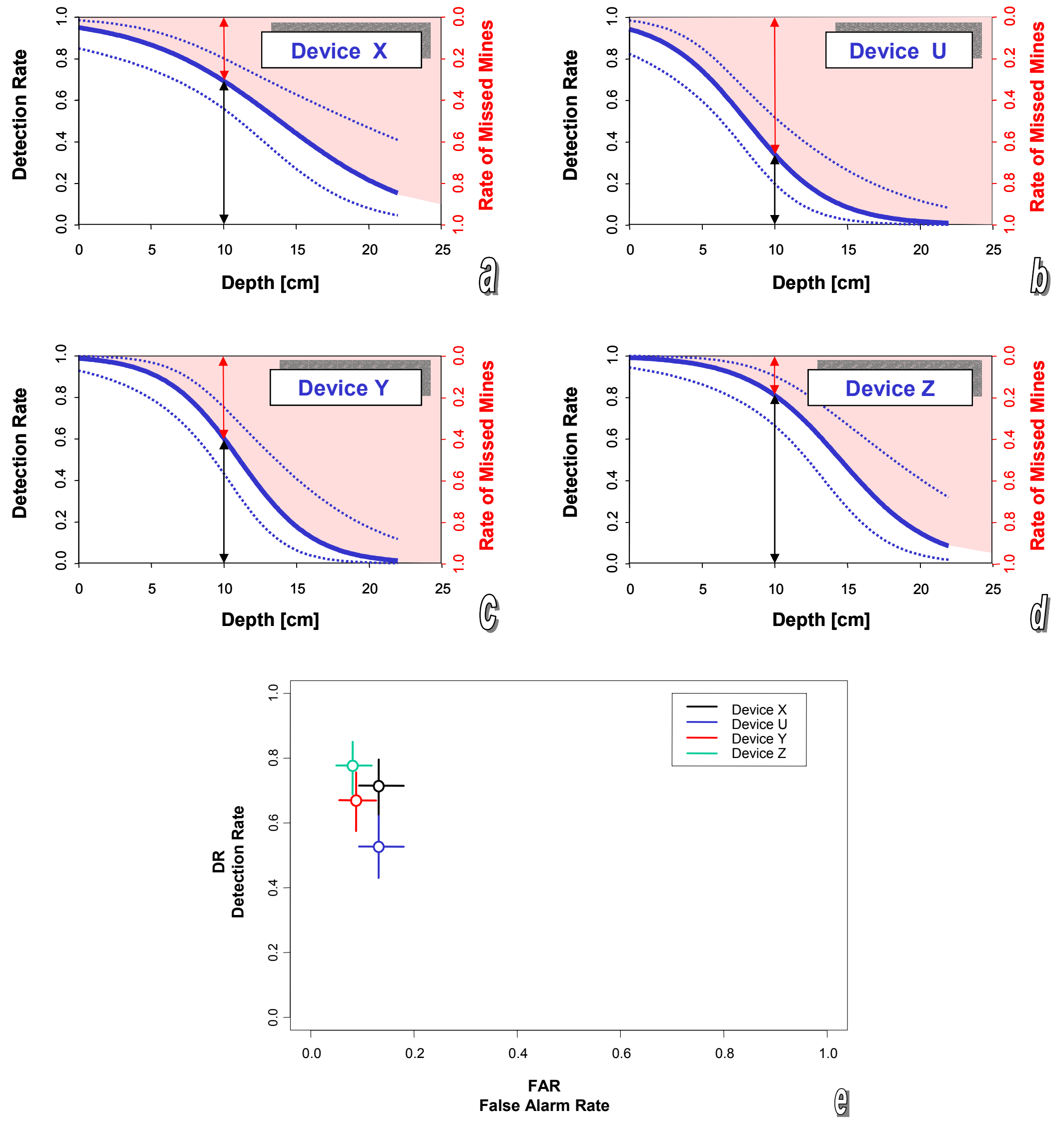

Figure 4: (a)-(d) POD versus depth, (e) ROC diagram. Neutral cooperative soil, very clean, only surrogate of PMA-2, with 95% confidence limits.

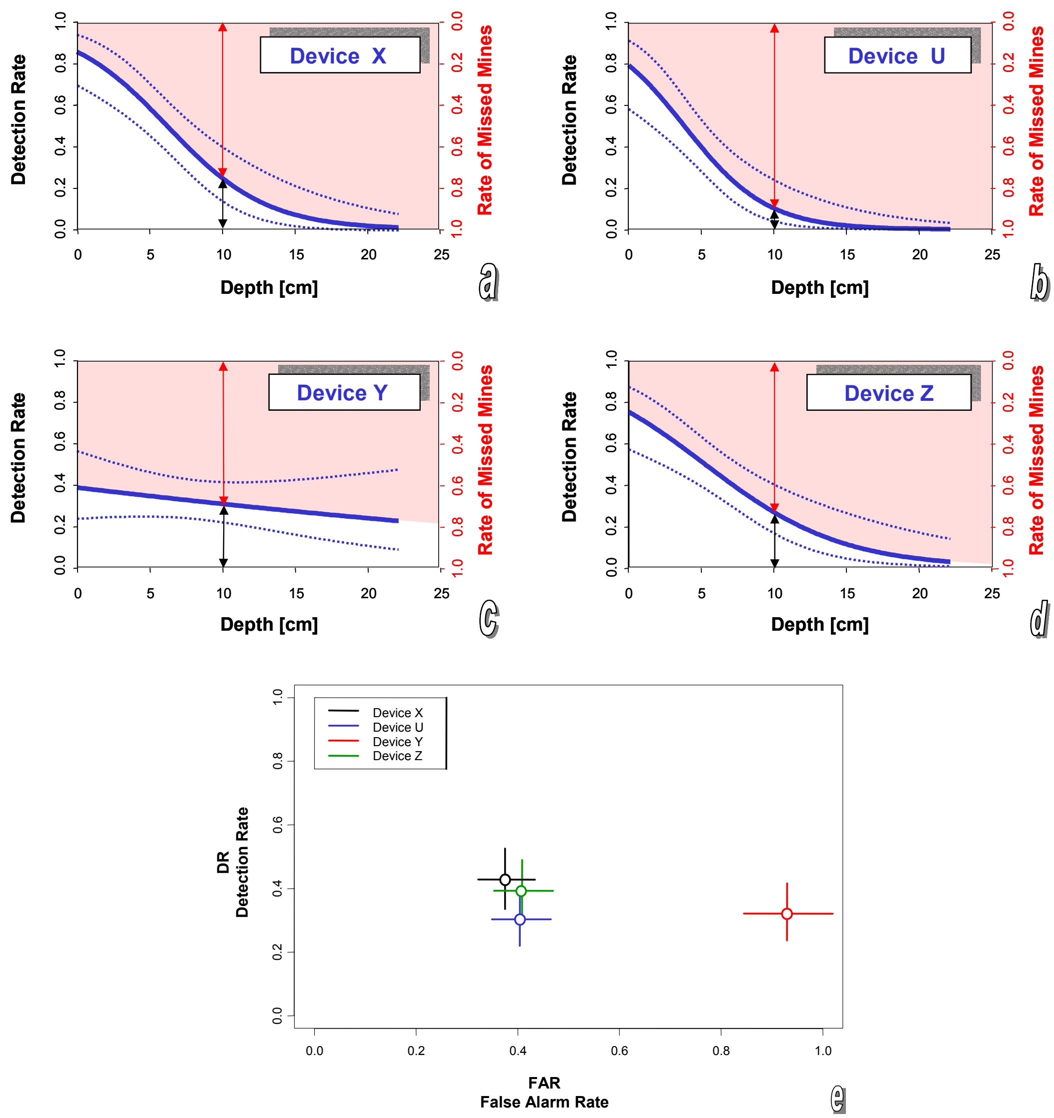

Figure 5: (a)-(d) POD versus depth. (e) ROC diagram. Uncooperative soil, heterogeneous, with frequency dependent susceptibility, red bauxite with neutral stones, target PMA-2, 95% confidence limits.

In the opinion of the authors this combination of receiver operating characteristic curves (Figure 4 and Figure 5) provide the information that the end user ought to know about the device that he/she is going to operate in the field. It is therefore recommended that receiver operating characteristic curves, with appropriate explanation and interpretation, be included in device catalogues for the main categories of soils encountered in mine affected areas.

3.7. Conclusions and Outlook

For detection reliability field tests the combined scenario of soil type, soil metal contamination and human factor has to be set up with care and must be appropriate for the local field situation. The characteristics of one detector should be determined in terms of the detection rate as function of depth in each soil for each mine type and completed with the information about the corresponding false alarm rate. An expected mean value of the performance of a detector in a certain region can then be determined from these basic curves, knowing the local mine distribution. The full demining process should be simulated to assess true clearance performance and might be introduced as a correction factor within a modular reliability model.

4. Statistical Considerations

This section presents two basic diagrams used in the report and a statistical model behind calculations of fitted curves and confidence limits. The orthogonal experimental design applied in the tests is also discussed.

Reliability of diagnostic systems can be described with ROC diagrams (receiver operating characteristics). It is usual in the scientific community of non-destructive testing to plot estimated probability of detection against estimated probability of false alarm (Ref. 2). Results of mine detection reliability tests are presented in an adapted kind of ROC diagrams: estimated probability of detection, PÔD (in this report also called Detection Rate), is plotted against estimated false alarm rate, FÂR (Ref. 3, 4). POD for a single pass through a test lane is estimated as the number of mines found divided by the number of mines in the lane. FAR is defined as the number of false alarms per square meter. A false alarm is each indication of an operator which falls outside a prescribed radius around a target, so called halo radius, which is defined in CWA 14747:2003. Result of each pass is a pair of values, PÔD and FÂR, and that pair corresponds to a single point on a ROC diagram. More reliable detection systems give results closer to the upper left corner of the ROC diagram, where POD is high and FAR low.

Probability of detection can be plotted as a function of depth, for each of the soils and for each type of target separately. This kind of diagram is called a POD curve and it gives important information about the performance of a detector in a soil of interest, for a specific mine type, in dependence on depth. POD curves are used in non-destructive testing to describe the influence of some physical characteristic of the defect on the probability of detection (Ref. 2).

4.1. Pointwise Confidence Limits

The estimated POD and FAR are subject to uncertainty caused by random influences and by the extent of the statistical basis. Sufficient numbers of targets and passes must be used to reduce this uncertainty to an acceptable level. No conclusions should be drawn about detector or operator performance beyond what the statistics allow. Therefore it is essential to estimate the standard errors and confidence limits associated with the results. In fact, these estimates for different possible outcomes should be made prior to conducting the test, when the test matrix is being designed.

If the true value of POD would be the same for all targets, the number of detections would follow a binomial distribution. The estimated POD is PÔD=*y/n*, where *y* is the number of detections and *n* is the number of opportunities to detect a target. Upper and lower confidence limits for PÔD can be found from the following equations:

$$POD_{lower} = \frac{1}{1+\dfrac{n-y+1}{yF_{INV}\,[1-\alpha/2, 2y, 2(n\text{-}y+1)]}} \qquad (1)$$

$$POD_{upper} = \frac{1}{1+\dfrac{n-y}{yF_{INV}\,[\alpha/2, 2(y+1), 2(n-y)]}} \qquad (2)$$

where $1-\alpha$ is the level of confidence and F_{INV} is F-quantile (the inverse F-function).

In many cases a binomial distribution can be approximated with a normal one. It has been proposed (Ref. 10) that an approximation with a normal distribution can be used if the following two conditions are fulfilled:

$$n \cdot P\hat{O}D > 5$$
$$n \cdot \left(1 - P\hat{O}D\right) > 5 \tag{3}$$

Confidence limits of that normal distribution would be (Ref. 11):

$$POD_{upper/lower} = P\hat{O}D \pm \frac{\hat{\sigma}\, t_{INV}\left(\alpha, n-1\right)}{\sqrt{n}}, \tag{4}$$

where n is the number of opportunities to detect a target, $1-\alpha$ is the level of confidence, t_{INV} is the inverse t-function and $\hat{\sigma}$ is the estimated standard deviation in PÔD, which equals

$$\hat{\sigma} = \sqrt{\frac{P\hat{O}D\left(1 - P\hat{O}D\right)}{1 - \frac{1}{n}}}. \tag{5}$$

For level of confidence $1-\alpha$ = 95% and sufficiently large n, t_{INV} is approximately 2, so that the expression for the 95% confidence limits can be further simplified:

$$POD_{upper/lower} = P\hat{O}D \pm 2\sqrt{\frac{P\hat{O}D\left(1 - P\hat{O}D\right)}{n-1}}, \tag{6}$$

where n is the total number of opportunities to detect a target (e.g. number of passes through a lane times number of targets in a lane). It should be kept in mind that this simple relation holds only if PÔD is sufficiently far from 0 and 1 and when the number of opportunities to detect a target is sufficiently large.

For the very first information about the size of the error bars, and that could be helpful during the planning of the trials, the equation (6) can be even more simplified. For PÔD=0.5 it becomes

$$POD_{upper/lower} = P\hat{O}D \pm \frac{1}{\sqrt{n}}, \tag{7}$$

where n is the number of opportunities to detect a target. In the region 0.29<POD<0.71 this approximation gives less than 10% larger confidence interval than the equation (6).

If POD is the same for each target, it behaves according to a binomial distribution. However, the assumption of equal probability would not be appropriate for these trials, since different targets on different depths are used. The resulting distribution is not binomial, but it is unknown and it depends on the combination of targets, depths and possibly other factors. This unknown distribution can be in many cases approximated with a normal distribution if POD is sufficiently far from 1 and 0 and if the number of opportunities to detect a target is sufficiently large.

The estimated standard deviation is obtained only after separating the data of all possible situations, which all have different standard deviations. However, the resulting confidence limit would not be much different from the one obtained using the assumption of equal targets and equal POD's discussed above, that is the equation (6).

An example with PÔD=0.8 is illustrated on Figure 6. For this PÔD, the conditions of equation (3) require that n>25. It can be seen that even for a PÔD as high as 0.8, and for n>25, 95% confidence limits of both normal distributions are similar to those of the binomial distribution. For PÔD closer to 0.5 they would be even more similar. This is why the approximation with normal distribution, i.e. the equation (6), was used throughout this report. In some cases confidence limits were calculated assuming binomial distribution, equations (1) and (2), which is more convenient for results close to PÔD=1 or 0. For simplicity, the estimated values PÔD and FÂR are denoted as POD and FAR (without the "hats") in all diagrams of this report.

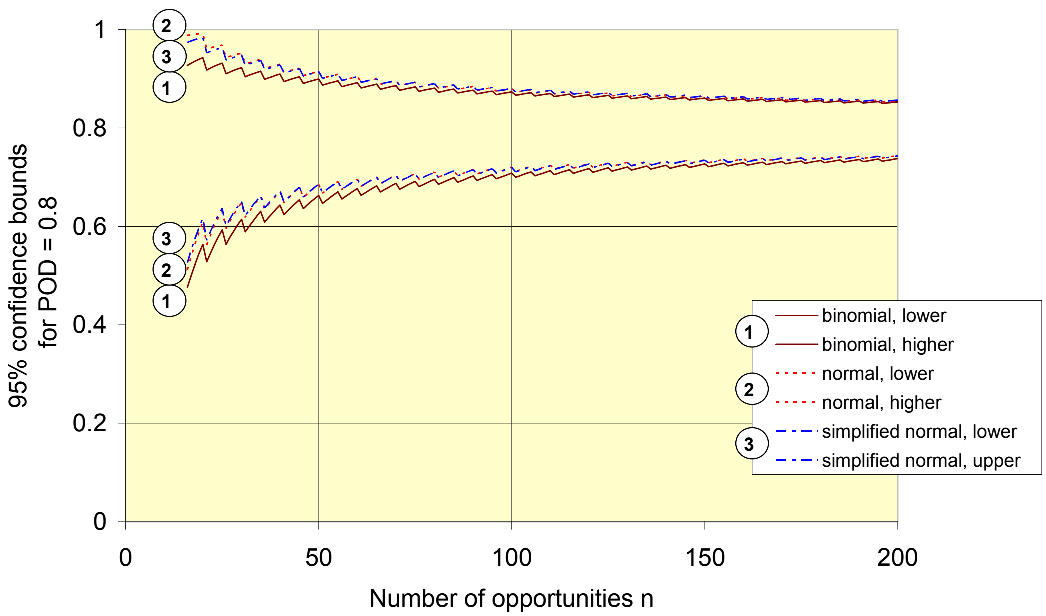

Figure 6: 95% pointwise confidence limits for PÔD = 0.8 depending on the number of opportunities to detect a mine. The full brown line belongs to the binomial distribution (equations (1) and (2)), the red dotted line to the normal approximation (equations (4) and (5)), and the blue dashed line to the simplified normal approximation $t_{INV}(\alpha, n\text{-}1) = 2$ (equation (6)).

Similarly simple formula can be found to describe the 95% confidence limits of the false alarm rate, FAR (Ref. 12). The number of false alarms follows a Poisson distribution, and the variance and the mean of the Poisson distribution are equal. That distribution can be approximated as normal. Setting the variance of that normal distribution to be equal to the variance of the Poisson distribution gives the following result for FÂR (which is the number of false alarms per area):

$$FAR_{upper/lower} = F\hat{A}R \pm 2\sqrt{\frac{F\hat{A}R}{N \cdot A}}, \qquad (8)$$

where *N* is the number of passes and *A* area of a lane. Throughout this report two methods of calculating pointwise confidence limits are used: the one described above in equation (8) and the one assuming Poisson distribution (see Ref. 12). Both methods give very similar results.

It is also important to note a simple rule: doubling the number of repetitions decreases the pointwise confidence intervals approximately by factor $\sqrt{2}$. This rule is valid both for PÔD and for FÂR, if they are not too close to PÔD=0 or 1 and FÂR=0.

4.2. POD Curves

To fit a curve for the dependence of POD on depth, non-linear logistic regression was applied (Ref. 5, presentation P. T. Wilrich). The POD is transformed according to the following equation and a linear dependence on depth is assumed:

$$\ln\left(\frac{POD}{1-POD}\right)=ax+b, \tag{9}$$

where x is the depth and a and b are parameters of the fit. The parameters a and b are found by Maximum Likelihood.

The POD does not fall abruptly to zero at certain depth but falls gradually with depth, as shown in Figure 7. The confidence intervals show that this is not just an artifact of the model. This behaviour may be understood semi-quantitatively from a simple picture of the sweep pattern and the spatial distribution of the intrinsic sensitivity of the detector.

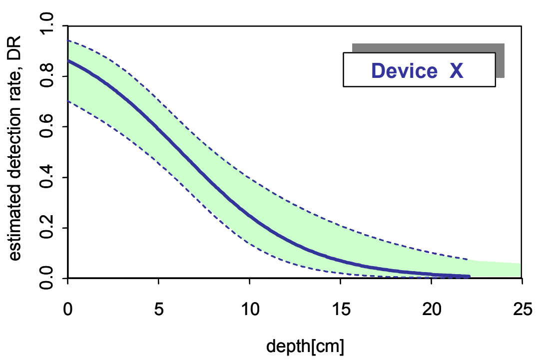

Figure 7: Typical POD-versus-depth curve for a target with small metal amount, with 95% confidence limits.

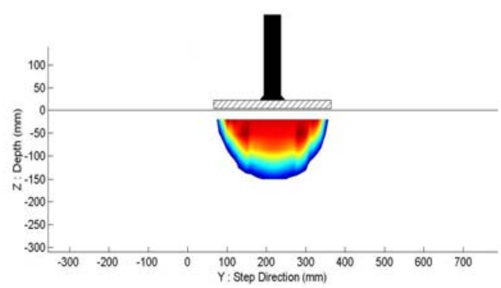

Figure 8: Sensitivity cone of a metal detector.

All detectors give an alarm indication when the target is within a finite volume under the head, termed the "sensitivity cone". A horizontal cross-section of this sensitivity cone is termed the "footprint". The plots on Figure 8 show a false-colour representation of the audio alarm as the detector is moved over a target on a mechanical scanner. The effective area swept out by a deminer in one sweep (i.e. in one left-right movement) is proportional to the front-back width of the footprint at the depth at which the target lies. The POD is therefore proportional to the fraction of the total area covered by this narrow strip. If the sensitivity cone is approximated as literally a geometric cone, the POD is then predicted to be flat to a certain depth and then to fall linearly to zero (Figure 9). Although this is an extremely simple model which does not take into account the details of the sweeping pattern, it does predict quite well the observed dependence of POD on depth. A more comprehensive model would need to take into account influences of many factors, for example: noise coming from the electronics, electromagnetic noise of the surroundings, heterogeneity of soil, uneven sweeping speed, height and step of the operator, and last but not least interpretation of the signal by the operator, following the general reliability model presented in Section 1.

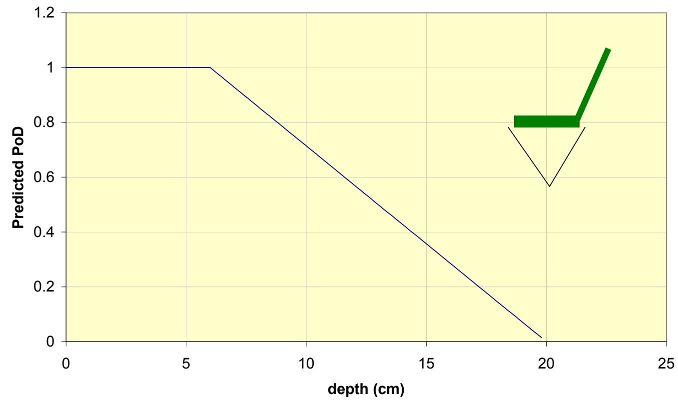

Figure 9: Simplified model of a sensitivity cone and prediction of POD. The sensitivity cone is approximated as literally a geometric cone.

4.3. Orthogonal Design

The factors which are used to identify each pass in the test are: Test Lane, Operator, Detector Model and Detector Specimen. Each factor has many levels, e.g. Detector Model has four levels if four detector models were used in a trial. A test matrix which includes combination of all of the levels of all of the factors (full factorial design) is the most obvious way to obtain unbiased estimates of factor effects, but it would also be a large matrix. As discussed in the previous section, there are diminishing returns as *n* is increased, so very large trials entail much cost to little benefit. It is possible, instead, to use a matrix in which each detector is tested with each level of each factor, but not with all the possible combinations. This design based on orthogonal design will give an unbiased test with considerably fewer passes than the full factorial design (Ref. 13).

If the object of the test is to determine how each detector model performs in each type of soil, one may consider the Test Lane and Detector Model to be systematic variables and the Operator and Specimen random variables, termed "nuisance factors". A suitable design based on two "Graeco-Latin squares" (see Ref. 13) is shown in Table 3. Taking the May Oberjettenberg test as an example: While the full factorial design would require a total of 512 passes, the Graeco-Latin square design requires only 128.

Table 3: Double Graeco-Latin square test matrix devised by P. Th. Wilrich, Day 1 of trials May Oberjettenberg. A, B, C, D are operators, α, β, γ, δ are detectors, two specimens of each model. The operators changed directions for starts 5-8, to minimise the possibility to remember the positions of their indications.

		start							
		1	2	3	4	5	6	7	8
test lane	1	A α	C γ′	B β	D δ′	C γ	A α′	D δ	B β′
	2	B γ	D α′	A δ	C β′	D α	B γ′	C β	A δ′
	3	C δ	A β′	D γ	B α′	A β	C δ′	B α	D γ′
	4	D β	B δ′	C α	A γ′	B δ	D β′	A γ	C α′

4.4. Optimal Choice of Some Values

The appropriate value of *n* to be used for achieving desired confidence limits will depend on what aspect is being compared e.g. ability of the detector for all mines, ability of the detector for one type of mine etc. It may be better to reduce the number of levels of a factor, e.g. a number of soil types, in the test matrix, whilst keeping the total size the same, in order to be more confident about fewer things.

The estimated standard deviation can be broken down into contributions from operator σ_o, the specimen of the detector σ_s and residual random variation on repetition σ_ε. If abbreviations n_O, n_S and n_ε are used to denote the number of operators, the number of specimens of each detector model and the number of opportunities to detect a target respectively, than the estimated standard deviation is

$$\hat{\sigma} = \sqrt{\frac{\hat{\sigma}_O^2}{n_O} + \frac{\hat{\sigma}_S^2}{n_S} + \frac{\hat{\sigma}_\varepsilon^2}{n_O n_S n_\varepsilon}} \tag{10}$$

Wilrich (Ref. 5) remarks that since σ_o was found to be the largest, it would be most beneficial to increase the number of operators. Similarly, it is pointless to increase the number of repetitions with one specimen beyond the level where the standard deviation is dominated by σ_S. It is, though, usually easier to organise tests with many repetitions than with many specimens of the detectors.

Target depths should be chosen to avoid POD = 0 or 1. For example, it would not be useful to bury the PMA-2 mines much deeper than 20cm in a trial, since no known detector could find them, so these targets would not provide any information about the relative performance of devices. It may be important to focus on certain depths which are of particular interest, such as the standard clearance depth of an organisation or the regulation mine-laying depth employed by a combatant army in the theatre of interest. However, for statistically optimal trials (optimal use of resources) it is recommendable to aim for a POD around 0.5. This can be achieved only with an appropriate choice of difficult targets, in which case the selection of targets and depths would not represent a realistic situation. It is important to see that this is not a disadvantage of the method here proposed, since the purpose of reliability trials is a comparison of metal detectors and not an estimate of their real performance in a minefield.

5. Laboratory Measurements

In the demining community the impact of the soil is regarded as the severest factor that reduces the reliability of the mine detection process. The influences of magnetic susceptibility and, to a lesser degree, electric conductivity are considered as key factors that determine the performance of metal detectors. It has been proposed elsewhere that magnetic viscosity of the soil, the dependence of susceptibility on frequency, is the major reason for the soil effects encountered in metal detectors. Measurements of soil susceptibility were made using a Bartington MS2 Magnetometer, in two ways. The first one was measuring in situ with an 18.5cm diameter circular loop probe operating at 958Hz (κ_{958}). The second way was obtained from 10cm^3 samples, in a sample chamber operating at 465Hz and 4650Hz (κ_{LF} and κ_{HF} respectively). The difference $\kappa_{FD} = \kappa_{HF} - \kappa_{LF}$ was also investigated, as a measure of magnetic viscosity.

A simple empirical measurement of the effect of a soil on detectors can be made by setting a detector without soil-compensation to a definite sensitivity and measuring the minimum distance to the soil surface at which the detector starts giving signals. This distance is called the ground reference height. This measurement was made on all the soils using a Schiebel AN19 Mod 7 detector, adjusted in a way that it could just detect a calibration pin at 10cm distance from the baseline mark. Equivalent calibration procedure is using a 10mm diameter chromium steel ball, material 100Cr6, at 14cm distance (Ref. 14). Results of soil measurements are given in Table 4 and Table 5, presenting the mean values and their standard deviations.

Table 4: Soils in Benkovac trials. Ground reference height and susceptibility measurements.

Soil Types in Benkovac Trials	Ground Reference Height (cm)	Susceptibility at 958 Hz (10^{-5} SI)	Susceptibility difference at 465 and 4650 Hz (10^{-5} SI)
Lanes 2, 6 (neutral) clay	no signal	13 ± 2	0,6
Lanes 1, 5 (uncooperative) bauxite	18,8 ± 0,9	154 ± 13	25,5
Lanes 3, 4, 7, 8 (uncooperative heterogeneous) bauxite with neutral stones	19,7 ± 2,5	190 ± 36	35,4

Table 5: Soils in Oberjettenberg trials. Ground reference height and susceptibility measurements.

Soil Types in Oberjettenberg Trials	Ground Reference Height (cm)	Susceptibility at 958 Hz (10^{-5} SI)	Susceptibility difference at 465 and 4650 Hz (10^{-5} SI)
Lane 1 (artificially uncooperative soil) humus + layer of blast furnace slag	5 ± 2	244 ± 64	6,1
Lane 2 (cooperative) cement gravel	no signal	0 ± 1	- 0,2
Lane 3 (cooperative) clay	no signal	2 ± 1	- 0,5
Lane 4 (cooperative) concrete gravel	no signal	6 ± 1	- 0,5
Lane 5 (uncooperative) magnetite mixed with coarse sand	4,5 ± 0,7	3000 ± 500	6 ± 7
Lane 7 (cooperative) cement gravel	no signal	-1,0 ± 0,2	-0,1 ± 0,2
Lane 8 (cooperative) concrete gravel	no signal	7 ± 1	-0,1 ± 0,1

As seen from the table, two of the three soils used in Benkovac were highly uncooperative, with ground reference height close to 20 cm, while one soil type was neutral. In Oberjettenberg five lanes contained neutral soil and two were uncooperative. The frequency dependence of uncooperative soils was very different. Lane 5 of Oberjettenberg contained magnetite with extremely high susceptibility, but with almost no frequency dependence. (Standard deviations are given only when the number of measurements was at least 8.) In the first trials in Oberjettenberg only Lanes 1-4 were used, other lanes were built for the trials in November, when seven lanes were used in the test.

Lane 1 in Oberjettenberg was artificially made uncooperative by adding a layer of blast furnace slag. The authors would not recommend this technique because the slag was found to contain metallic particles, creating additional metal contamination (see Section 7).

During the trials it has been observed that the ground reference height measured with the Schiebel detector is strongly correlated with κ_{FD} and that it follows the approximate relationship

$$S = \sqrt{10^6 \cdot \kappa_{FD}}\ [cm] \tag{11}$$

where S is the ground reference height measured in cm. This observation was based on measurements from test lanes in Mozambique and Benkovac (Figure 10) (Ref. 15). Correlation between ground reference height and susceptibility was found to be much less pronounced. This indicates that magnetic viscosity has a dominant role in the performance of Schiebel AN19 Mod 7 detector, which is a pulse induction (time domain) detector. Theoretically, susceptibility falling with frequency, i.e. having large κ_{FD} values, results in a time-decaying magnetisation in response to a pulse. Such a soil would be expected to affect both time-domain and frequency-domain metal detectors. To demonstrate this, κ_{FD} should ideally be measured over a bandwidth representative of metal detectors (a few 10's of kHz), so the two-frequency chamber used here has rather too low a bandwidth. Nevertheless, these results still show a clear correlation.

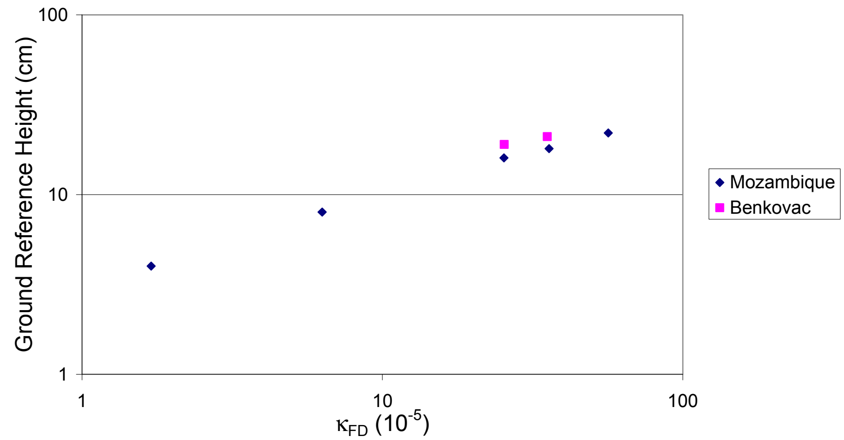

Figure 10: Ground reference height as a function of susceptibility difference $\kappa_{FD} = \kappa_{HF} - \kappa_{LF}$. Each point corresponds to a soil type.

Tests of in-air maximum detection height using parametric target sets of metal balls allow detection capability to be defined in terms of a minimum detectable ball diameter at a given height. The relative detection capability of the detectors for different metals is also studied. Evidence is given that supports the use of chrome steel balls as a standard parametric target set.

Extending the idea of using sets of metal balls, the detection capability in soil can be readily measured and compared to the capability in-air. Comparison of in-soil and in-air capability has been made in tests using an uncooperative magnetic soil. This gives a test for giving quantitative measurements of the influence of soil on the capability to detect buried metal targets.

To illustrate some of the laboratory measurements conducted at JRC, maximum detection distances for chromium steel balls measured with different detectors are given in Figure 11.

Maximum detection distance measurements were made on some of the targets used in Benkovac trials, in air as well as in soils present in Benkovac test lanes, and also on one target used in Oberjettenber trials in November. The results for all four detectors, three targets and four soils of Benkovac trials are plotted on a diagram on Figure 12, and results of Oberjettenberg measurements are on Figure 13. Measurements on PMA-S (Oberjettenberg) can be compared to those on PMA-2, since PMA-S is a surrogate of that mine. The first column contains detection distances in-air. It is obvious that detection capabilities of all detectors for all targets decrease with the increasing ground reference height. The only exception is detector U in Benkovac trials, which seems to achieve better results in soils with the highest ground reference height – so called uncooperative soils – than in soils with more moderate electromagnetic properties.

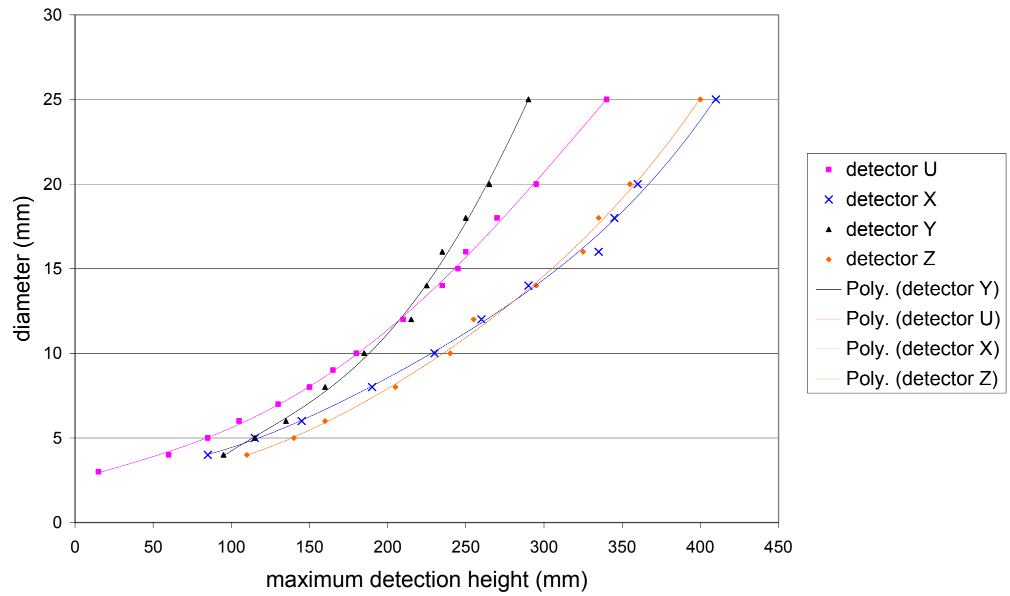

Figure 11: Maximum detection distance in air for 100Cr6 balls and all four detectors.

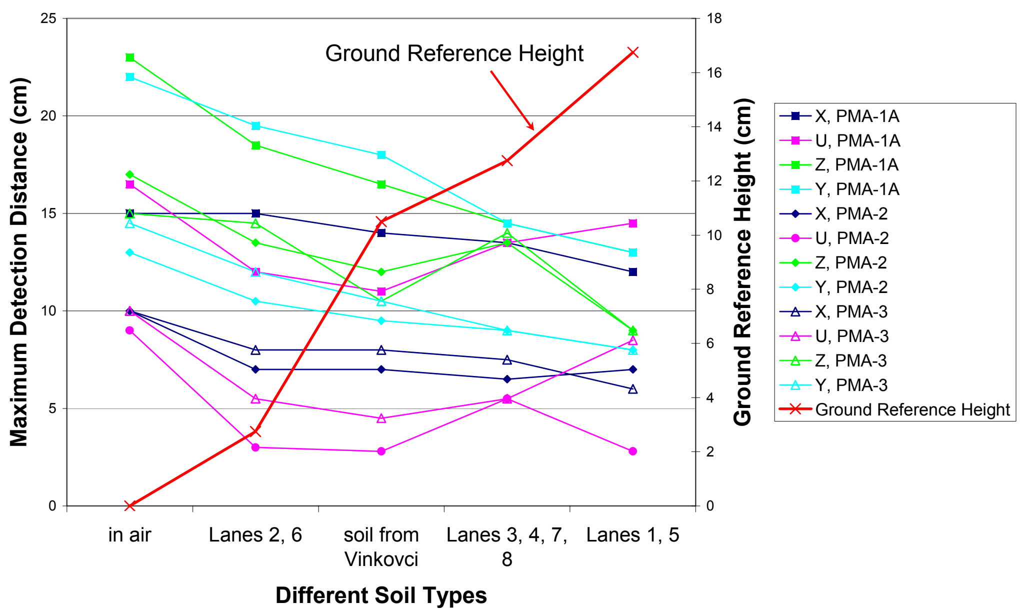

Figure 12: Maximum detection distance measurements, Benkovac. Detectors U, X, Y, Z, targets PMA-1A, PMA-2 and PMA-3. Measurements of ground reference height are given for comparison.

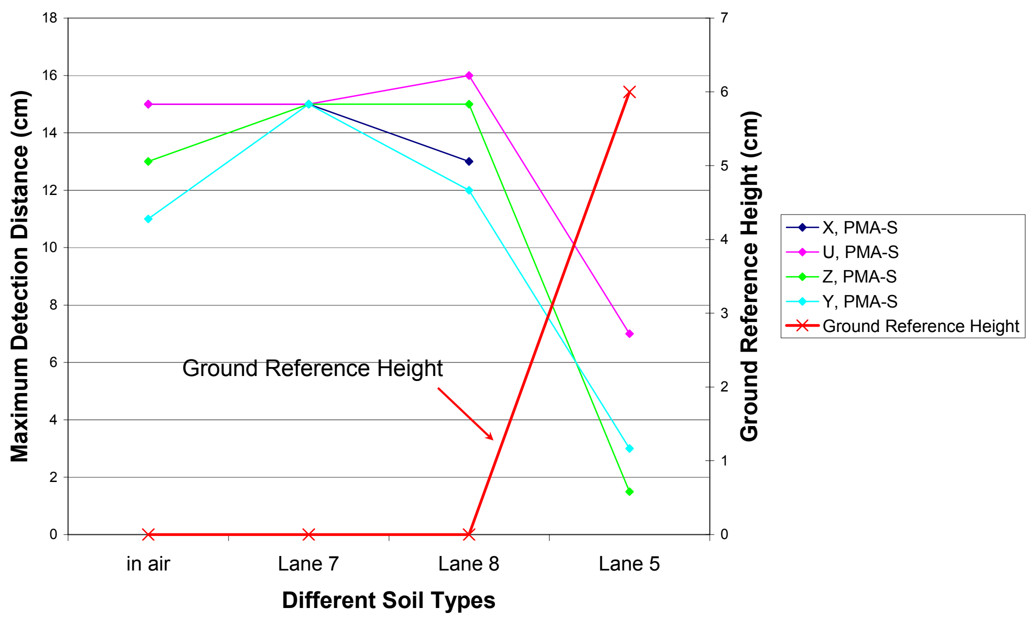

Figure 13: Maximum detection distance measurements, Oberjettenberg. Detectors U, X, Y, Z, target PMA-S (PMA-2 surrogate). Measurements of ground reference height are given for comparison.

These results are compared with results of blind reliability trials, namely with corresponding PÔD's. These PÔD's are calculated counting all targets of the same type, regardless of depth. The correlation between maximum detection distances (MDD) and the corresponding PÔD's obtained in Benkovac trials is very low (Figure 14), but in Oberjettenberg trials it is much higher (Figure 15). The authors suggest that this difference might have been caused by different methods of measuring MDD in soil. In Benkovac measurements with all detector models were performed separately and each target was repeatedly buried until the MDD was established. Thus the soil was disturbed many times before the measurement result was obtained, what might have influenced local electromagnetic properties of the soil. In Oberjettenberg November trials measurements were organised differently. Several targets of the same type were buried to certain depths and they were checked with metal detectors. The largest depth detected was recorded as MDD. This way the soil was disturbed only once or twice. Therefore, for MDD measurements on larger targets the authors would recommend to use the method described above, which was applied in Oberjettenberg.

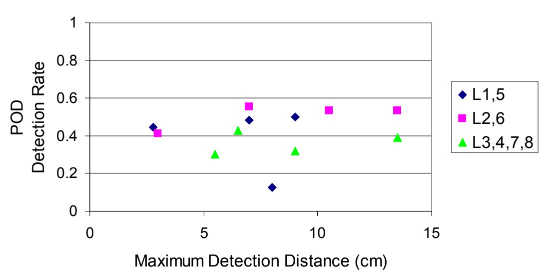

Figure 14: Correlation of PÔD and maximum detection distance for PMA-2, Benkovac trials. Each point corresponds to a certain detector in a certain soil. Each soil type is represented with a different symbol and colour.

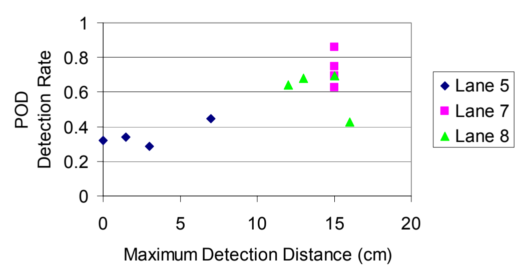

Figure 15: Correlation of PÔD and maximum detection distance for PMA-S, Oberjettenberg trials. Each point corresponds to a certain detector in a certain soil. As expected, lower result are attained in the uncooperative soil.

The mines Maus, MS3 and PMN were detected by all detectors in air at distances larger than 40cm. Measurements in soils were not performed.

These results point to the conclusion that performance of each detector is highly influenced by soil. None of the minimum metal targets could be detected at 20cm depth in any of the soils. In more difficult soils only a few could be detected at 13cm depth with only the most sensitive detectors. The clearance standard established by Croatian laws is 20cm depth, and the UN standards (IMAS 09.10, "Clearance standards") recommend that at least top 13cm of soil be cleared of all mines. Some of the targets used in these measurements, PMA-1A, PMA-2, PMA-3 and PROM-1, are the most frequently found antipersonnel landmines in Croatia and Bosnia and Herzegovina, therefore this choice of targets can be considered representative for that region. The PMA mines are usually put to 1-5 cm depth, but sometimes they sink lower with time, or are covered with vegetation. Also after mechanical clearance they can be found on larger depths. Databases containing the typical depths on which mines are found do still not exist.

6. Reliability Tests

This section presents the method and the results of three reliability trials. The layout of the lanes is described, also the choice of targets and their depths, as well as training of the operators. After that most important results of reliability tests are presented. An overview of all parameters of all three trials is given in Table 1 in Section 2 and in Table 2 Section 3.

Each set of trials was performed in a two-week period, with additional few days of preparations in the field. During the preparation test lanes were cleared of metal debris with the aid of detectors and targets were placed (except those already in the ground).

The test lanes used in Oberjettenberg were 20m long and 1m wide. In Benkovac eight test lanes with 30m length and 1m width were prepared. Electromagnetic properties of soils are discussed in Section 5. All lanes were planned to be cleaned of all metal debris with the aid of metal detectors. Within the time limitations that was impossible to achieve: Lane 8 of Benkovac trials was so contaminated with metal pieces, that it was decided not to use it in the trials. Instead, all passes that were planned to be performed in Lane 8 were performed in Lane 4, which contained the same soil type. (This expedient was, of course, not ideal. Bias could have occurred if the deminers had remembered where they had previously located targets in Lane 4. Fortunately, analysis of the results does not allow a conclusion that this happened). The same problem occurred in Oberjettenberg with Lane 6, which contained huge amounts of metal debris. For that reason Lane 6 in Oberjettenberg was abandoned and no measurements were performed. All other lanes were cleaned with the aid of metal detectors, to the extent that is possible to achieve. (It is discussed in Section 7 how effective this way of clearing is.)

The number of targets in May trials varied between 24 and 28. Besides real mine bodies without explosives, standard targets that simulate metal components of mines (ITOPs, see Ref. 7) were also used. Most of the targets were buried to random positions and to depths between 2cm and 8cm as measured from the surface to the top of the mine. The depths were slightly different in different lanes. All explosives were removed from the mines for safety reasons, leaving the metal constituents unchanged.

All lanes in Benkovac contained the same 32 targets, buried with the same depth distribution, varying from 0cm (just below the soil surface) to 20cm, but positioned randomly in each lane. All targets used in Benkovac tests, mines as well as ITOPs, are given in Table 6, together with the corresponding depths. Targets TMA-3 and TMA-4 were treated as the same target, since their metal content is very similar. Targets TMRP-6 and TMM-1 are both anti-vehicle mines with a metal case, so they were also treated as the same target. Since each of PMA landmine types was buried in each lane to five different depths, an analysis of detection with respect to depth was feasible, which was not the case in May trials, when the positions and depths of targets were given and could not be changed. The initial explosive was removed from all targets by an authorised company and the metal constituents remained unchanged. The main body of explosive was not removed, since the targets are used by CROMAC-CTDT for testing explosive detection dogs.

Table 6: Mines buried in each lane, Benkovac trials

	targets	*pieces*	*depth (cm)*
antipersonnel mines	PMA-1A	1	0
		3	5
		1	10
		1	13
		1	20
	PMA-2	1	0
		3	5
		1	10
		1	13
		1	20
	PMA-3	1	0
		3	5
		1	10
		1	13
		1	20
	PROM-1	2	0
		3	5
antitank mines	TMA-3 or TMA-4 (similar metal content)	1	10
	TMRP-6 or TMM-1 (similar metal content)	1	10
ITOP targets	E0	1	5
	G0	1	5
	K0	1	5
metal ball	100Cr6 (16mm Ø chromium steel ball)	1	10

In the last set of trials in November all mines in Lanes 1-4 were kept on their original locations, to test reproducibility of the trials. Only positions of ITOP test targets were changed and their number was reduced. In the new lanes 5, 7 and 8 a similar scheme to the one in Benkovac was applied: all targets are listed in Table 7. PMN and MS3 are here treated as the same target, since their metal content is exactly the same. Target PMA-S is a surrogate of PMA-2, containing the same metal piece as the real mine and having approximately the same shape, without any explosive. As in the May trials, real mine bodies without explosives were used.

Table 7: Mines buried in lanes 5, 7 and 8, Oberjettenberg trials

	targets	*pieces*	*depth (cm)*
antipersonnel mines	Maus	1	0
		3	5
		1	10
		1	13
		1	20
	PMN or MS3 (same metal content)	1	0
		3	5
		1	10
		1	13
		1	20
	PMA-S	1	0
		3	5
		1	10
		1	13
		1	20
antitank mines	TM-62 M	1	10
ITOP targets	C0	1	5
	E0	1	5
	G0	1	5
	I0	1	5
	K0	1	5
metal ball	100Cr6 (16mm Ø chromium steel ball)	1	10

In three series of trials two different training schemes were applied. In the first and the second series (Oberjettenberg May and Benkovac July) all operators underwent two-day training, during which they were introduced to four detector models. In the last series of trials (Oberjettenberg November) the training was twice longer, the number of detector models was 5, and the training for two groups of detectors (pulse induction and continuous wave) was separated, as well as the tests. This way, operators had more time to practise on hidden targets. An overview is given in Table 8.

Table 8: Training scheme, Oberjettenberg November

Date	Operators	Detectors
17th, 18th Nov	A, B, C, D	U, W, X
	E, F, G, H	Y, Z
24th, 25th Nov	A, B, C, D	Y, Z
	E, F, G, H	U, W, X

Another novelty was introduced to November trials: the number of passes per deminer per day was not larger than 6, while in the earlier trials deminers made an average of 8 passes per day. This change, as well as the new training scheme, was applied to lower the stress for operators and to increase their performance, thus coming closer to the real conditions in a minefield. Results of these changes are presented later in this section.

A definite sensitivity of a diagnostic system, in our case a metal detector, yields one operating point on a ROC diagram. Changing the sensitivity causes a shift along the ROC curve: the increase in sensitivity causes the increase of POD, but also raises probability of false positive indications. Each ROC curve corresponds to different reliability: curves of systems with higher reliability will lay higher. To study the shift along a ROC curve, the first two series of trials were performed with two sensitivities. The higher sensitivity was the highest achievable sensitivity of devices and the detectors were set up according to the manufacturers' instructions. The lower sensitivity was calibrated so that a steel ball was buried to some specified depth, different in each soil type, and the sensitivity of each device was adjusted so that the ball was just detected. Trials in November were performed only with high sensitivity.

All passes in all trials were performed according to the Graeco-Latin square test matrix discussed in Section 4. In Oberjettenberg May trials the same matrix was used on the second day. The scheme was repeated at reduced sensitivity on days three and four and the entire scheme by Operators E-H in the second week. Each lane contained about 25 objects, so there were typically n=800 opportunities for each detector using high sensitivity. The 95% confidence interval for n=800 at p=0.7 is 0.66 to 0.74, calculated by any of the methods of Section 4. In the Benkovac trial an improvement was introduced: the operators were permuted for the repetitions. The same was done in Oberjettenberg November trial.

The results presented in this report are predominantly based on high sensitivity measurements from Benkovac and Oberjettenberg November trials. The low sensitivity results and some results from Oberjettenberg May are presented only when necessary allowing a conclusion.

6.1. Tests Benkovac

The performance of each detector was assessed in each of the soils and for each type of target. An example is given in Figure 16, illustrating the performance of detector X in Lanes 2 and 6, neutral soil from Sisak. Each curve corresponds to one of the PMA antipersonnel mines. The estimated probability of detection (PÔD) for each point on this diagram is calculated from eight different scans: four operators made two passes each. Although this is a small data sample, two aspects become obvious: the falling trend of all curves and in general higher PÔD's for PMA-1A.

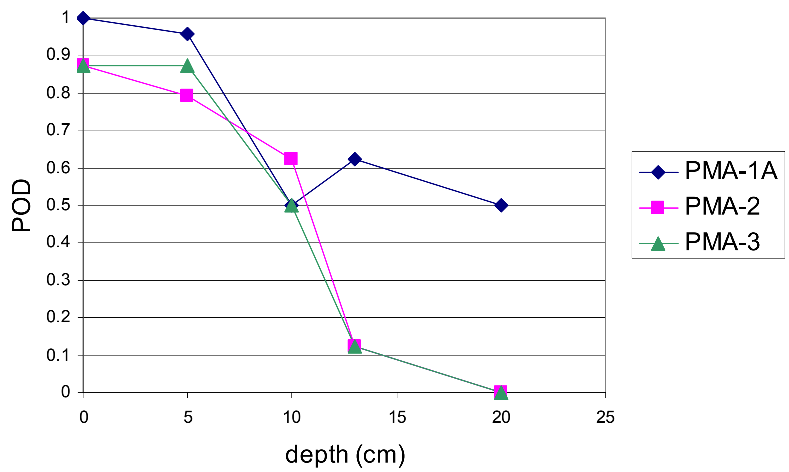

Figure 16: Performance of detector X in lanes 2 and 6, neutral soil from the surroundings of Sisak. Each set of points corresponds to a different target: PMA-1A, PMA-2 and PMA-3.

The following diagrams present results of tests on Lanes 3, 4, 7 and 8, which contained original soil from Benkovac region, with a target PMA-2, one of the most dangerous threats in Southeast Europe. Four detector models are shown separately, thus presenting differences between the detectors.

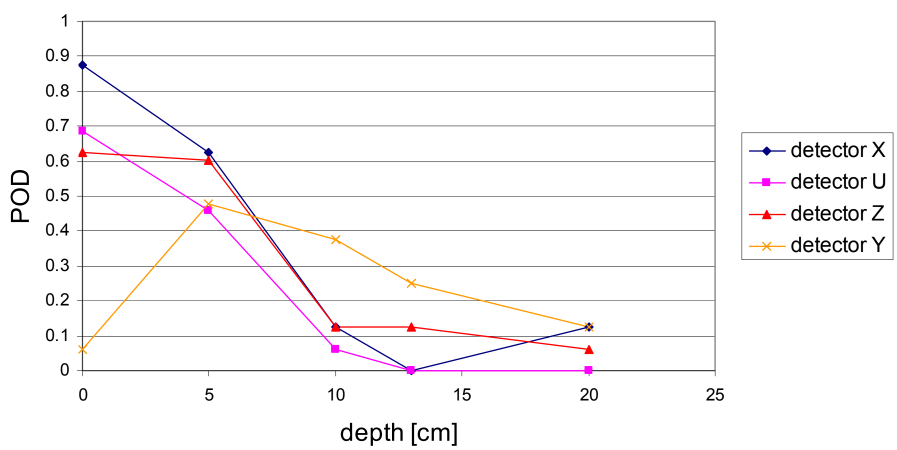

Figure 17: Benkovac trials, results in Lanes 3, 4, 7 and 8, target PMA-2, comparison of detectors.

Non-linear logistic regression was applied to these results. The four curves corresponding to four detectors together with their 95% confidence limits are presented in separate diagrams in Figure 18. Four curves are plotted together on the next diagram, Figure 19, without the corresponding confidence limits. This diagram allows comparisons between detectors. All confidence limits of POD curves in this report are based on a normal distribution of a transformed variable on the left hand side of the equation (7), Section 4.

The width of the confidence interval in case of device Y (Figure 18) and also a comparison with the positions of the points on Figure 17 indicate that the assumption about the shape of the POD curve (logistic regression) is not the most appropriate. In this very specific case of detector Y with PMA-2 in Benkovac soil, it can be seen (Figure 17) that the POD was unexpectedly low for PMA-2 mines buried to 0 cm depth (just below the surface). Results are based on four passes performed in each of the four lanes with the same soil type, each of the lanes containing one target of interest. Separated results for Lanes 3, 4 and Lanes 7, 8 (Figure 21 and Figure 22) show the same unexpected behaviour. It is assumed that in this specific case a signal coming from the target causes an error in pinpointing.

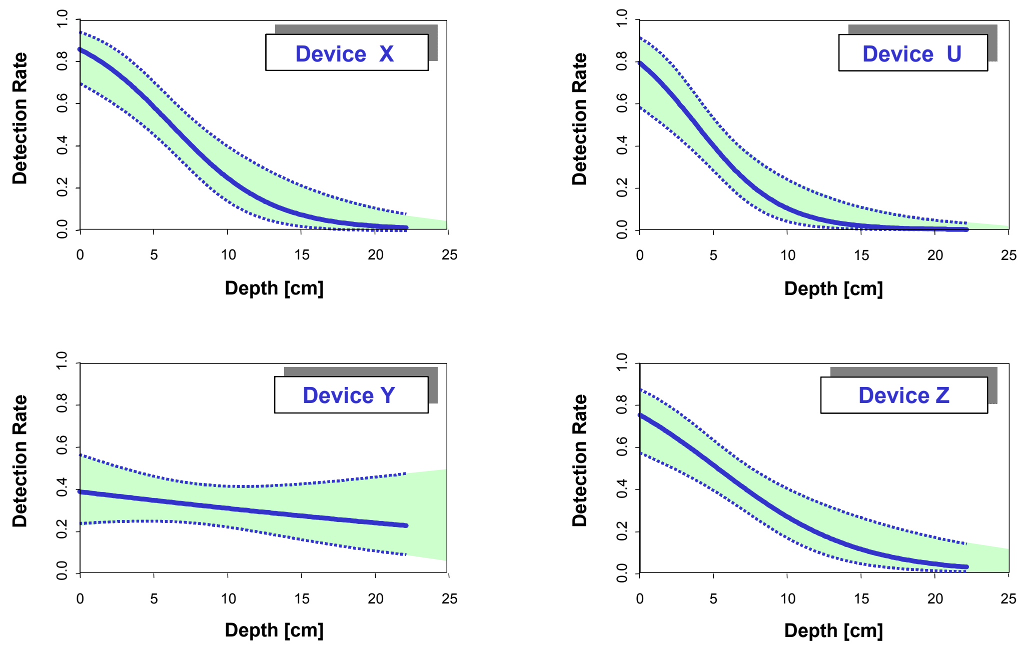

Figure 18: Benkovac trials, results in Lanes 3, 4, 7 and 8, target PMA-2, with non-linear regression and 95% confidence limits. The estimated probability of detection is here called Detection Rate.

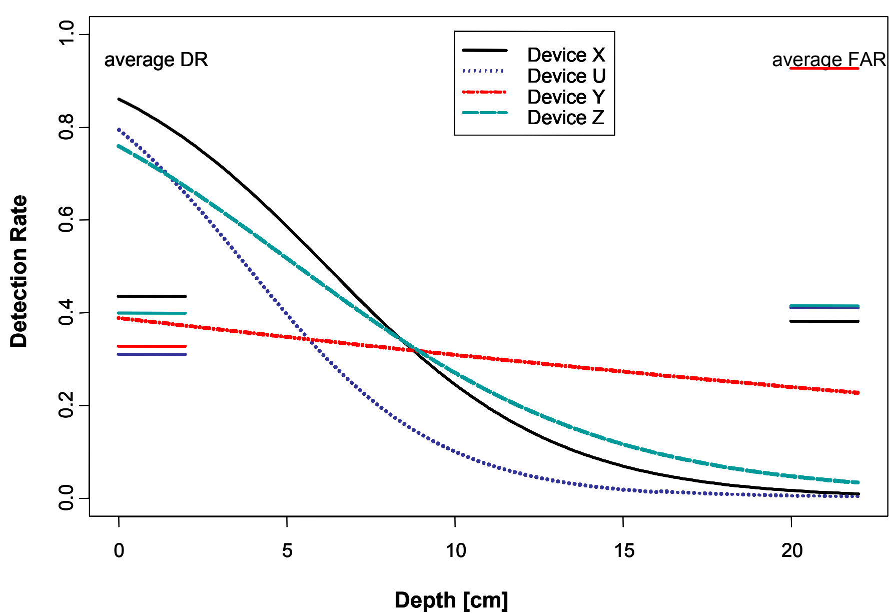

Figure 19: Benkovac trials, results in lanes 3, 4, 7 and 8, target PMA-2, non-linear regression. Average estimated probability of detection (detection rate) and false alarm rate are indicated with short horizontal lines to the left and right on the diagram.

ROC diagrams present probabilities of detection and also give information about false alarms. On the next diagram (Figure 20) PMA-2 mines at different depths are counted together. A more sophisticated approach would be using the curves of POD versus depth and attributing some statistical weights to different depths. For example, if the targets are expected closer to the surface (which is the case in most real minefields), then a higher statistical weight would be given to smaller depths, thus increasing the total POD for this target.

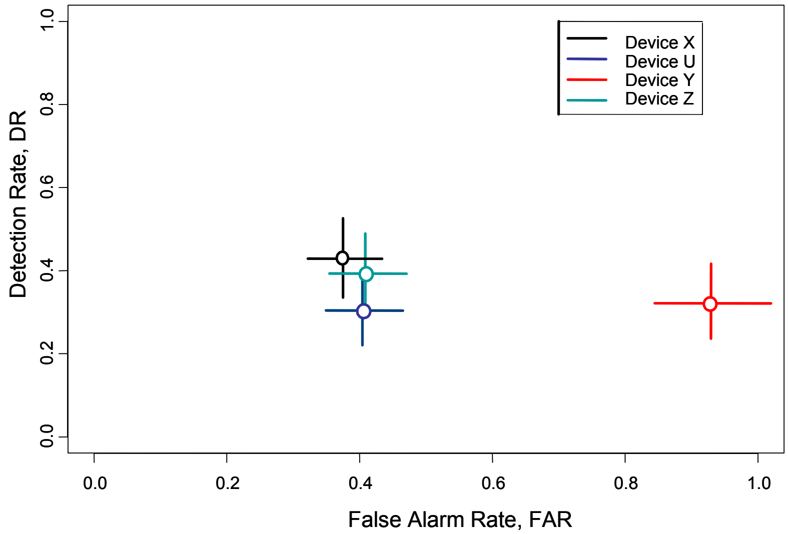

Figure 20: ROC diagram of Benkovac trials, target PMA-2, Lanes 3, 4, 7 and 8, 95% confidence limits of POD based on binomial distribution, FAR 95% confidence limits based on Poisson distribution.

A rough check of reproducibility of the trials could be a comparison between results in Lanes 3, 4 with those of Lanes 7, 8. These results are presented on Figure 21 and Figure 22. It can be seen that the results from Lanes 3, 4 are very similar to those from Lanes 7, 8. The confidence intervals are wider than those on the diagram with all four lanes containing that soil, Lanes 3, 4, 7 and 8 (Figure 18), because the total number of opportunities to detect a target is twice smaller.

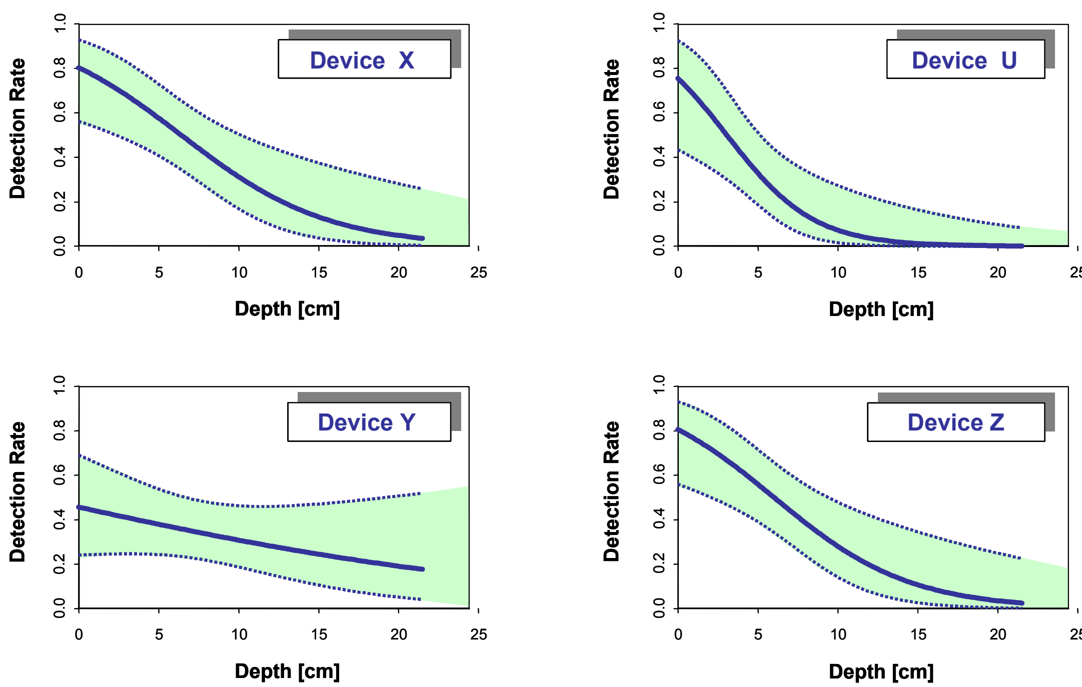

Figure 21: Benkovac trials, results in Lanes 3 and 4, target PMA-2, non-linear regression, 95% confidence limits.

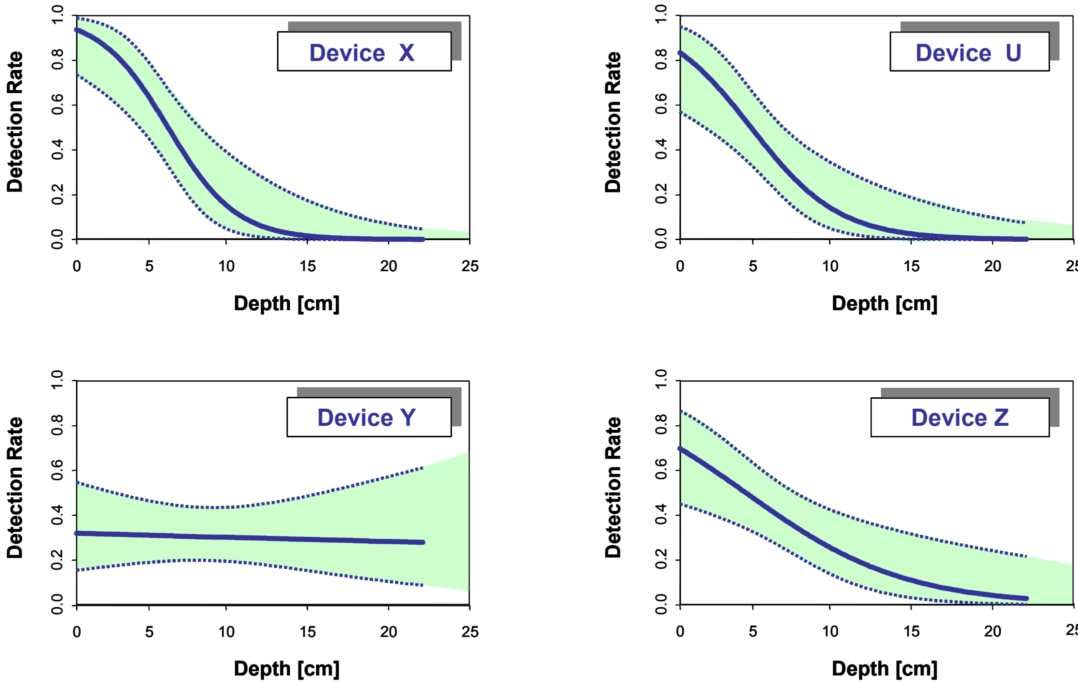

Figure 22: Benkovac trials, results in Lanes 7 and 8, target PMA-2, non-linear regression, 95% confidence limits.

ROC curves and POD curves can also be used to indicate differences between operators. Results from Benkovac indicate that deminers who currently work with detectors day-to-day achieve higher results than their colleagues former deminers who work on higher level positions such as demining supervisors. Performance of deminers is illustrated on Figure 23 and Figure 24. The second of these figures clearly shows that also the relative performance of detectors, that is their ranking, is influenced by the operators' skills. The group A-B-C-D, which achieved higher results (see Figure 23), has the best results with detector Z, while the other group achieved the best results with detector X. This confirms the opinion that tests should be performed by experienced and well trained deminers, i.e. those who will actually do the clearance in a minefield.

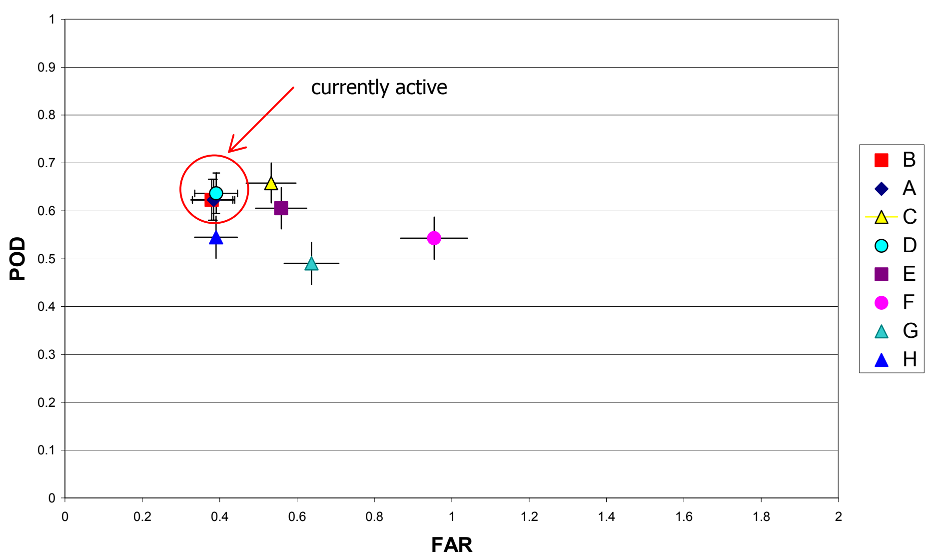

Figure 23: Benkovac trials, comparison of the performance of deminers. Persons A, B and D are currently active as deminers. 95% confidence limits are based on normal distribution.

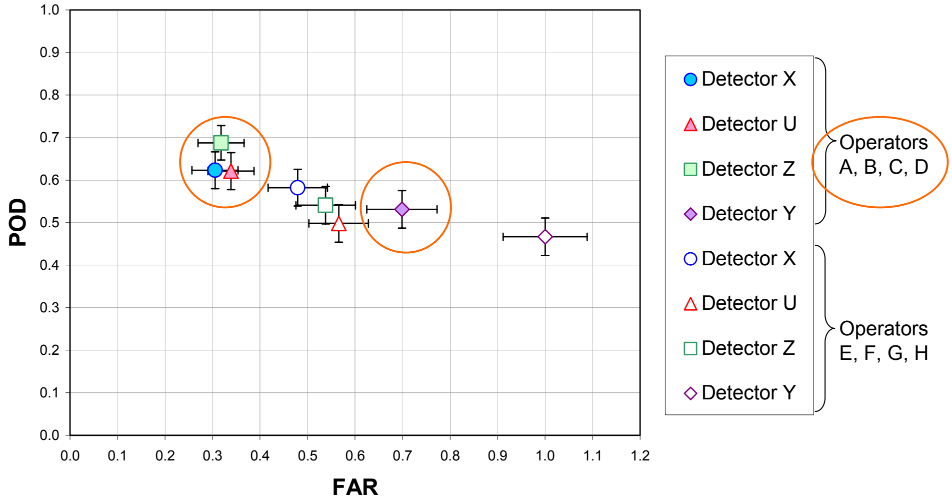

Figure 24: Benkovac trials, comparison of the two groups of deminers. The ranking of detectors is influenced by the skills of the operators. 95% confidence limits are based on normal distribution.

To reduce the FAR professional deminers sometimes use sensitivities lower than maximum. This can be done when it is known with certainty what mine types can be found in a minefield and if the expected depths are known. Sensitivity is set so that the mine type most difficult to detect can be detected at the expected depth. Results of these trials indicate that changing the sensitivity of a metal detector causes a shift along the ROC curve (explained in Section 4). This is illustrated with Figure 25, where overall results for all detectors, high and low sensitivity scans separately, are plotted together.

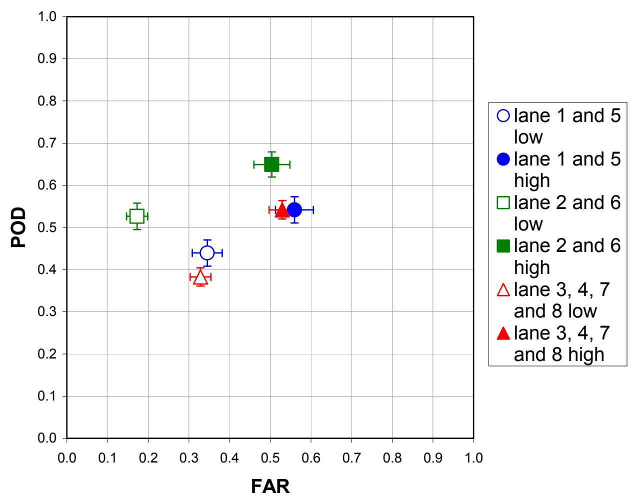

Figure 25: Benkovac trials, results of high and low sensitivity tests. A change in sensitivity causes a shift along a ROC curve. 95% confidence limits are based on normal distribution.

The same diagram illustrates the influence of soil on detection results averaged over detectors. Results in the neutral soil of Lanes 2 and 6 are the best, both in terms of POD and FAR. Once more it has been confirmed that the ground reference height and the magnetic susceptibility indicate the difficulties that metal detectors have with a particular soil.

Even experienced operators from Benkovac had difficulties to detect large metal content antitank mines (TMM-1 and TMRP-6), though all detectors gave very strong signals. It is assumed that the reason of surprisingly low performance on large metal content mines is poor pinpointing; in case the mine was not detected, operators' indications were found about 40cm from the centre of the target, outside the halo, and were counted as false alarms. A better training would probably result in a better performance.

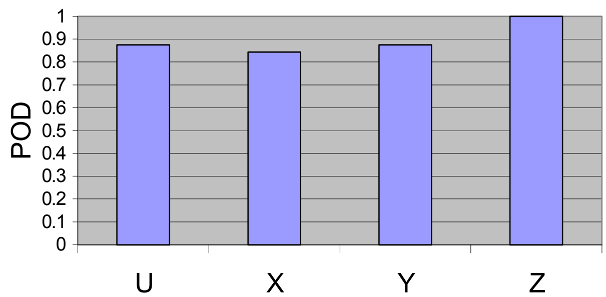

Figure 26: Benkovac trials, detections of metal antitank mines TMM-1 and TMRP-6.

6.2. Tests Oberjettenberg November

Since the target PMA-S used in Oberjettenberg November trials is a surrogate of PMA-2, curves of POD depending on depth are very similar for these two targets. An example is given to illustrate the performance of metal detectors in cooperative soil of lanes 7 and 8 of Oberjettenberg test fields, Figure 28.

Figure 27: PMA-S, a surrogate of PMA-2. The metal part is the same as in the original mine.

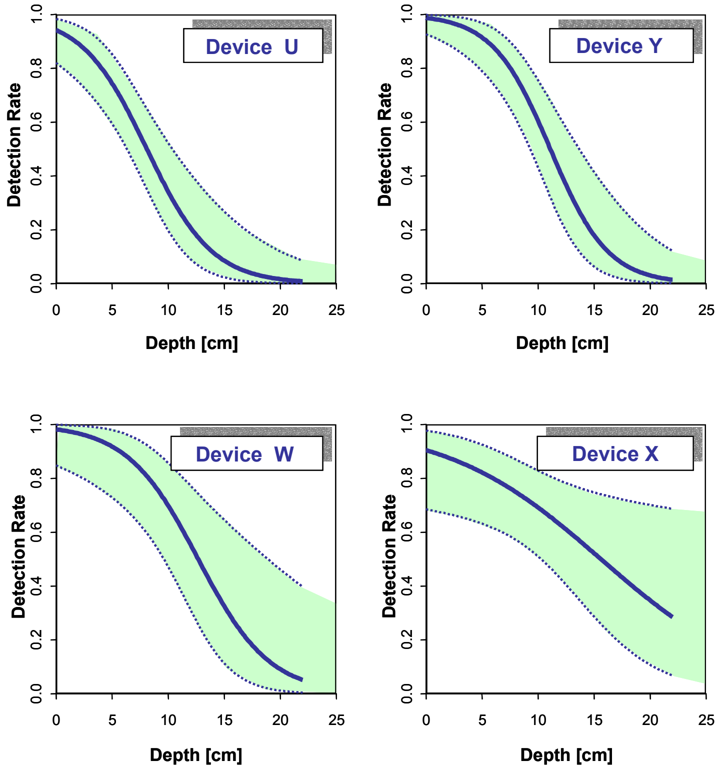

Figure 28: Oberjettenberg November trials, results in Lanes 7 and 8, target PMA-S, non-linear regression, 95% confidence limits.

Besides PMA-S, two other targets were buried to five different depths: Maus and PMN+MS3. Two targets, PMN and MS3, were treated as the same, since their metal content is the same. These targets were more easily detected by all metal detectors (Figure 29, also Figure 30). Even in uncooperative magnetite, PÔD was well above 80% regardless of depth, which is not surprising, having in mind the results of maximum detection distance measurements (all higher than 40cm). In this case logistic regression would not be the most appropriate choice. The width of the pointwise confidence intervals is larger than the difference between the detectors, which makes every comparison of devices very unreliable. This is generally the case whenever PÔD is close to 1 or 0.

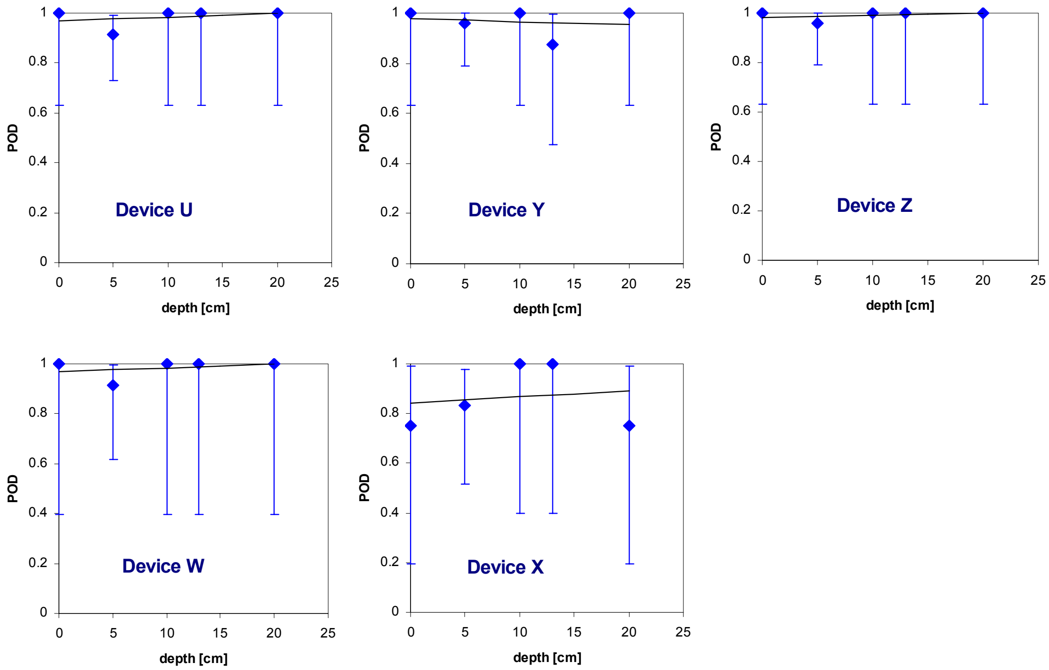

Figure 29: Oberjettenberg November trials, results in Lane 5, target PMN and MS3, linear fit, 95% pointwise confidence limits based on a binomial distribution (equations (1) and (2)).

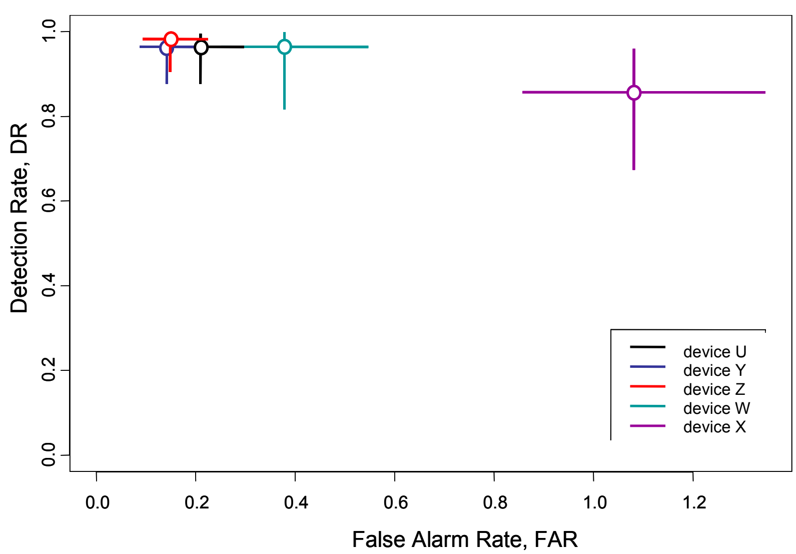

Figure 30: Oberjettenberg November trials, results in Lane 5, target PMN and MS3, ROC diagram. 95% confidence limits of POD based on binomial distribution, FAR confidence limits based on Poisson distribution.

A diagram on Figure 31 illustrates differences between the lanes. Results of the last part of the trials, when targets were excavated, point to the conclusion that many of the FA come from metal pieces overlooked during the preparation of the trials. This problem is discussed in Section 7. The diagram also shows that PÔD's were lower in soils where ground compensation was used, namely in Lanes 5 and 1. This implies that ground compensation reduces sensitivity if all detectors are counted together. This result is in agreement with the maximum detection distance measurements, which showed that each detector lost some sensitivity in uncooperative soils (Section 5).

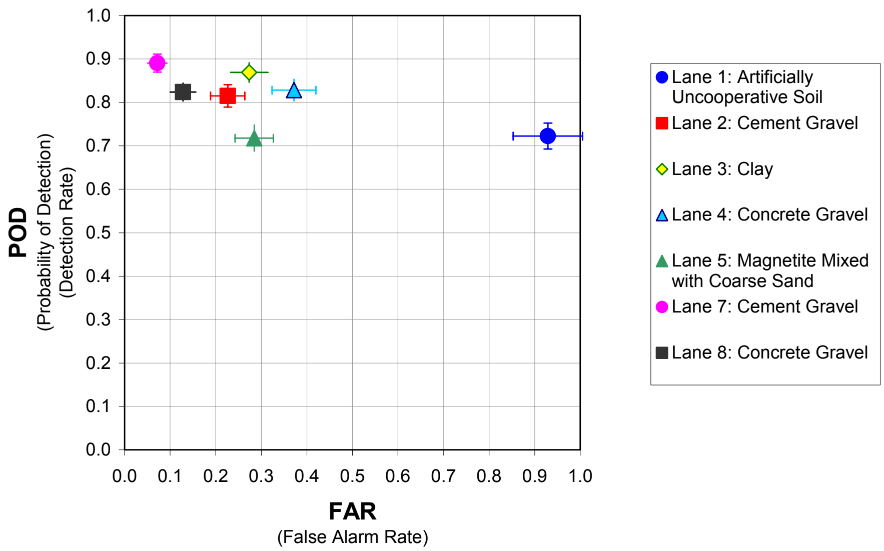

Figure 31 Oberjettenberg November trials, comparison of results in different lanes. 95% confidence limits based on normal distribution.

The effect of the improvement of the human factor is presented on Figure 32. Results of Oberjettenberg May and November trials are compared, counting only the targets and the soils that were common to both trials. In the November trial only 10% of these targets were missed, while in May 16% were not found by the operators. The authors believe that a longer training and less starts per day (discussed at the beginning of this section) are the main reasons for improved performance.

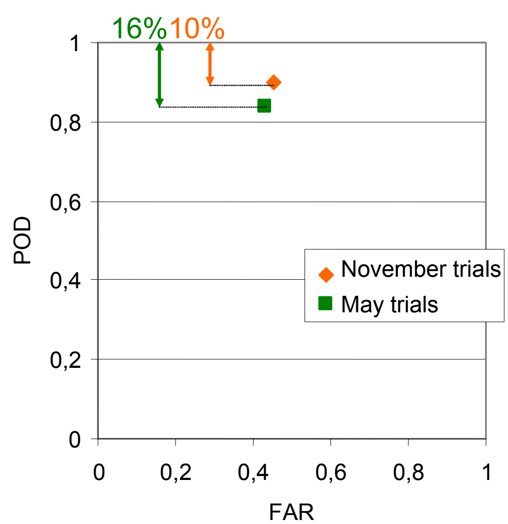

Figure 32: Compared results of Oberjettenberg May and November. Only the targets and the soils common to both trials are counted, so that the results are comparable. 95% confidence intervals are smaller than the size of the symbols representing the results, hence not visible on this diagram.

7. Test with Excavation of Targets

After all passes from the experimental design were finished, an additional trial was performed. The operators indicated positions of all alarms, they were recorded by supervisors, and the operators prodded the ground and excavated all sources of alarms. In this test seven operators completed only one pass each, using one detector in one lane, thus excavating targets from all seven lanes, so that the test could have not been repeated. Their indications were measured and it was recorded whether they come from a known target, from a metal piece overseen during the preparation of the trials, or from the soil. The positions and the depths of all targets were once more measured and compared with the data.

Results of this test were compared with average results achieved in the reliability trials on the same lane with the same detector model. Four results of the excavation test were higher, two were lower, and one was the same as the corresponding results of reliability trials. However, the error bars attributed to these results are so large, that only very uncertain conclusions can be made about a higher performance in average. Repeated excavation trials on the same lanes are not feasible, therefore longer lanes would be needed with more targets to reduce the error bars, which would increase the expenses significantly. An orthogonal design with repeated measurements on the same lanes could not be used in that case.

In all cooperative lanes (with zero ground reference height) all alarms came from the buried targets or from metal pieces found and excavated during this last pass, there were no signals coming from the soil. The only exception was Lane 2, where three "hot stones" were found – stones that cause a detector to alarm. These detections were counted as false alarms, since metal detectors are designed to detect metal and to compensate the influence of soil and hot stones. Detections coming from metal debris were ignored and not counted as false alarms. These findings indicate that the vast majority of false alarms of the reliability trials in the cooperative lanes actually came from very small metal pieces overlooked during the preparation of the trials and that they are not a consequence of the electromagnetic properties of the soils. In Lane 5 with ground reference height different from zero there were some false alarms coming actually from soil. These were counted as false alarms, whilst those coming from metal debris were ignored. All alarms in Lane 1 were coming from metal pieces and not from the soil.

In the case discussed here, the mine detection systems being tested are metal detectors. Whether detection alarms caused by metal pieces in the ground are considered "true" or "false" detections depends on the aims of the detection reliability trial. An ideal mine detection system would, in principle, be able to distinguish between a mine and a piece of scrap metal. Metal detectors currently used in demining do not have this capability.

During this test depths of buried targets were measured once more. Results of these measurements are in most cases a few centimetres different from the measurements performed in the preparation phase. The differences could be caused by measurement errors during burying and excavation, but also by moving the surface of the soil during the trials (the operators and the supervisors were allowed to step on the lanes).

8. Practical Conclusions

The authors would recommend that at least one week is dedicated to the preparation at the test site. Presence of metal debris in lanes is the most difficult practical problem the authors have encountered. Some simple technical improvements, apparently trivial, can save resources: for example, a simple measurement device for measuring positions of the markers was very helpful, since measurement of the marker positions was the most time consuming process in the trials.

All of the data entry and some data evaluation were performed on site. This enabled the authors a quick reaction to errors. For example, anomalies noticed in the resulting diagrams indicated errors in data entry, which could be corrected on site.

The authors would recommend to measure maximum detection distances using the method used in Oberjettenberg November trials, rather than the one used in Benkovac (discussed in Section 5). An adequate care should be taken about pinpointing, during the training and during the test, since errors in pinpointing can influence the results. Proper training in general is one of the most important influencing factors.

It is very recommendable that all participants of the trials understand and accept the importance of all rules and procedures. Even then, constant surveillance of the trials is recommended, from the preparation to the end of the trials, due to complexity of the whole process.

Many repeated measurements are needed to achieve reliable and repeatable result. For example, for PÔD=0.5 with pointwise confidence limits ±0.1 a total number of n=100 opportunities to detect a target is needed (see equation (6), Section 4). This can be achieved, for example, with 4 repetitions (preferably with different operators) and 25 targets in a lane.

It is recommended that targets of the same type are buried to a set of depths, the same in each lane, as specified in the CWA 14747:2003 (see Ref. 1), that is to some of these depths: 0, 5, 10, 13, 15 and 20 cm. Location of the targets should be randomly chosen and different in each lane. The lanes should be 1 m wide and between 10 and 30 m long and the overall number of targets should be smaller than 1.2 targets per square metre. If the trials include a larger number of different targets, than 10 m long lanes could be insufficient, and in the case of 30 m long lanes fatigue and time pressure could influence results. The design of the test should be an orthogonal design. Each detector model should be tested with two specimens. The training should last at least a day per detector model. A training scheme similar to the one applied in Oberjettenberg November trials is recommended. The number of operators should be as large as possible, at least four. Test operators should be chosen randomly from a group of trained deminers who will perform clearance operations.

Additional practical recommendations can be found in the Summaries of Breakout Sessions of the ITEP Workshop Reliability Tests for Demining (Ref. 5). The Summaries and the Recommendations of the Workshop can be found in the Annexes of this report.

9. General Conclusions

Series of tests in Benkovac and Oberjettenberg were executed following the standard CWA 14747:2003 for testing metal detectors. Some novelties were introduced and are recommended by the authors for future use, most important being the orthogonal design of the test in combination with the use of ROC diagrams and POD curves.

No systematic difference between time domain and frequency domain detectors was noticed. Significant differences between specimens of the same detector model were also not noticed. It is confirmed that experienced, active and skilled deminers should operate the detectors during the trials. Simple approximate formulas for the error bars (equations (6) and (8)) are suitable for preliminary results in the field, since they do not require complicated calculations. In many cases they can also be used in final reports.

Low values of POD – not just in these tests, but in all other tests known to the authors – raised many concerns in the demining community. However, this does not reduce the value of the reliability test as described in the CWA 14747:2003. The purpose of such a test is a comparison of metal detectors, not an estimate of their true performance in a minefield. To achieve its purpose, the test *has* to be designed so that the average PÔD is about 0.5. Results of quality assurance and accident statistics indicate that the probability of detection in a real minefield is actually much higher. The main reasons for such low results in tests are:

- Targets are placed at depths that do not represent a realistic scenario, but a rather difficult one.
- Training of the operators is much shorter than in reality.
- Operators change many detector models and very often, sometimes with very little time to adjust to another device.
- The existence of a time schedule creates a pressure on the operators, which causes faster progression along the lanes than in a minefield.
- In absence of danger, lower alertness can be expected than in a minefield.

Four of these five points describe influences of human factors, what clearly underlines their importance. Every trial report that presents results in terms of POD should also give a detailed description of all conditions in which the trials took place. Those results should never be interpreted as realistic minefield results, but only as results of the test. The purpose of a reliability test is a reliable comparison of devices, with the final aim of choosing the most suitable device for certain conditions of application.

10. Acknowledgements

We would like to thank German Federal Foreign Office (Lieut-Col. Detlef Schroeder), Ministry of Defence (Col. Manfred Merk), the BAM Presidential Board and JRC (Dr. Alois J. Sieber) for sponsorship and support. We thank the staff of CROMAC-CTDT and WTD 52 for their help at the test sites, especially Col. Ivan Steker. We would also like to thank CEIA, Ebinger, Foerster and Vallon (listed alphabetically) for the devices and for the training. Last but not least, we wish to thank the deminers and soldiers who participated in the trials.

11. References

---

[1] CWA 14747:2003, CEN Workshop Agreement, Humanitarian Mine Action – Test and Evaluation – Metal Detectors, http://humanitarian-security.jrc.it/demining/cw07/
[2] C. Nockemann et al., "Performance Demonstration in NDT by Statistical Methods: ROC and POD for Ultrasonic and Radiographic Testing", Proceedings, 6th European Conference on Non Destructive Testing, pp. 37 44 (1994)
[3] C. Mueller, M. Scharmach, M. Gaal, D. Guelle, A. M. Lewis, A. Sieber, "Performance Demonstration for Humanitarian Demining", Materialpruefung Jahrg. 45, Vol. 11-12 pp. 504-512, 2003
[4] C. Mueller, M. Scharmach, M. Gaal, D. Guelle, A. M. Lewis, A. J. Sieber "Proposals for Performance Demonstration and Modular Reliability Assessment for Humanitarian Demining", International Conference on Requirements and Technologies for the Detection, Removal and Neutralization of Landmine and UXO, EUDEM2-SCOT, Brussels, 15-18 September 2003
[5] ITEP Workshop on Reliability Tests for Demining, http://www.kb.bam.de/ITEP-workshop-03/
[6] T. J. Bloodworth, A. Lewis, "Quantifying the Detection Capability of Metal Detectors Using Metal Spheres", EUDEM-SCOT-2003, International Conference on Requirements and Technologies for the Detection, Removal and Neutralization of Landmines and UXO, Vreije Universiteit Brussel, Belgium, 15-18 September 2003
[7] Mine Detection Equipment for Countermine and Demining (Hand-held or Vehicle Mounted), FR/GE/UK/US International Test Operations Procedure (ITOP) "non-paper" 4-2-523 AD No. 251795, 20 December 1999
[8] Summary of Metal Detector Trial Report, UN Mine Action Programme Afghanistan, February-March 2002, available at http://www.itep.ws/
[9] M. Gaal, C. Mueller, K. Osterloh, M. Scharmach, S. Baer, U. Ewert, A. Lewis, D. Guelle, "Metal Detector Test Trials in Germany and Croatia – ITEP Project 2.1.1.2", MATEST 2003 NDT Achievements and Challenges, Brijuni, Croatia, 28-30 September 2003

[10] K. M. Simonson, "Statistical Considerations in Designing Tests of Mine Detection Systems I – Measures Related to the Probability of Detection", http://www.uxocoe.brtrc.com/TechnicalReps/librarylistall.asp
[11] M. Gaal, S. Baer, T. J. Bloodworth, D. Guelle, A. M. Lewis, C. Mueller, M. Scharmach, "Optimising Detector Trials for Humanitarian Demining", SPIE Defense and Security Symposium, April 12-16, 2004, Orlando, Florida, USA
[12] K. M. Simonson, "Statistical Considerations in Designing Tests of Mine Detection Systems II – Measures Related to the False Alarm Rate", http://www.uxocoe.brtrc.com/TechnicalReps/librarylistall.asp
[13] NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, 28 June 2004
[14] D. Guelle, A. Smith, A. Lewis, T. Bloodworth, "Metal Detector Handbook for Humanitarian Demining", Luxembourg: Office for Official Publications of the European Communities, Italy 2003; available at http://www.itep.ws/
[15] D. Guelle: International Detector Test UNADP, Final Report, UNADP Mozambique, 2000, available at http://www.itep.ws/

Annex 1: ITEP Workshop on Reliability Tests for Demining, Summaries of Breakout Sessions

This document summarises the findings of four breakout sessions held on Wednesday 17 December, 2003 in Berlin, Germany as part of the workshop "Reliability Tests for Demining".

Summary of Breakout Session 1

Setup of Test Lanes; Mines Selection
Ivan Steker, Dieter Guelle, Adam Lewis, Mate Gaal

- Tests should be performed in the region in which the metal detector will be used in order to experience the local conditions. Where possible, soils used in the test should be the same as soils in which the mines/UXO are expected.
- Test soils shall be metal free or contain only controlled contamination. Metal free soil can usually be obtained at greater depths, or special machines and metal detectors should be used to clean the soil.
- Lanes shall be cleared of vegetation and have flat surface.
- Depth distribution shall be as specified in CWA 14747:2003. Maximum depth shall be chosen according to national requirements, and in accordance with local conditions. Several mines of the same type shall be buried to the specified depths. (See also a footnote in Breakout Session 4.)
- Mines relevant to the local region should be chosen as well as some reference targets.
- The minimum distance between targets should be 70 cm (not 50 cm, as specified in CWA 14747:2003). In addition, special care should be taken that a target with a small metal content is not hidden by a signal of a large metal target or contamination.

Open questions:

- Length of the test lanes
- Choice and number of targets

Summary of Breakout Session 2

Soil Influence and Ground Compensation

Stephen Billings, Dieter Guelle and Adam Lewis

- In agreement with Breakout Session 1, the soils should be sourced from the local area actually contaminated with mines. The soil conditions in the test lanes should mimic the local soils as closely as possible (e.g. degree of water-logging, sun exposure etc).
- Categories of neutral, moderate and heavily uncooperative soils defined in Annex A of CWA 14747:2003 should be extended by characterisation with frequency dependence of susceptibility, since it has an important influence on metal detector performance.
- Whilst pulse induction detectors are generally immune to constant susceptibility, some designs of continuous wave detectors may be affected by it. Therefore, a secondary classification based on constant susceptibility should be used as in the first version of the document.
- Categories of soils should be extended by characterisation with ground reference height[1]. This measurement may be applied as an alternative to measurements of frequency dependence of susceptibility.
- To define ground compensation capability of a detector the maximum detection distance of a standard target in soil shall be measured after the ground compensation is made and than compared to maximum detection distance measurements performed in air.

Open questions:

- The exact boundaries between different soil categories regarding frequency dependence of susceptibility and regarding ground reference height should be defined. There are few measurement results that can be used to define those boundaries. The table below gives a suggestion, which should be modified as more soils are measured.

Category	**Primary** (x 10^{-5} SI) Susceptibility difference (465 to 4650 Hz)	**Secondary** (x 10^{-5} SI) Susceptibility (958 Hz)	**Ground Reference Height** (cm) (Schiebel AN19 Mod 7)
Neutral	< 1	< 50	< 1
Moderate	1-10	50-500	1-10
Severe	10-40	500-2000	10-20
Very Severe	> 40	> 2000	> 20

- The values of 465 Hz, 4650 Hz and 958 Hz are frequencies used by Bartington MS2 magnetometer, but they may not be the best choice. The current knowledge about soils does not allow us to specify any set of values as standard. Besides, that would discourage manufacturers from developing instruments with other options and unfairly discriminate against other existing products.
- Reference soils would be soils with standard electromagnetic properties, belonging to different categories, which would be defined in a table like that above. Because knowledge about the influence of different factors on metal detectors is still evolving, this classification is not final and the reference soils still cannot be established.
- Characterisation with ground reference height raises the issue of how to relate susceptibility difference to ground reference height (to ensure both measurements give consistent categorisations). The following proposal, while not actually discussed at the breakout session, is raised as a possible solution. First, measurements of ground reference height and susceptibility difference could be taken over a wide range of soils and the relationship between the two could be determined. Secondly, the ground reference height could be used to form an initial, rapid, classification of soils. However, the susceptibility difference should always be measured once the test lanes are established, and a recategorisation made if required.

[1] A simple empirical measurement of the effect of a soil on detectors can be made by setting a static mode detector without soil-compensation to a definite sensitivity and measuring the minimum distance to the soil surface at which the detector starts giving signals. This distance is called the ground reference height. This measurement can be performed using a Schiebel AN19 Mod 7 detector, adjusted in a way that it could just detect a calibration pin at 10cm distance from the baseline mark.

Summary of Breakout Session 3

Human Factor

Christina Mueller, Josef Matulewicz, Dieter Guelle, Davor Laura, Maik Hamann

- Experienced and active deminers shall be selected for testing. They should be chosen from those who will perform the clearance operations in the region.
- For pre-testing purposes inexperienced deminers are acceptable, with appropriate training.
- Training shall last at least one day per detector type. It shall include exercising on all mine types present in the tests, buried in each soil to each depth. Special attention shall be given to pinpointing of AP and AT mines and also to the sweep advance, which will depend on the detection cone.
- If possible, separate training and test runs should be carried out for detectors working on different principles (continuous wave or time domain).
- Some deminers may already be familiar with one type of detector or the detectors of one manufacturer. In this case the training should be adjusted accordingly and previous experience recognised.
- Working hours shall be in accordance with national laws and regional conditions. In any case, a total of 5-6 working hours for operators (deminers) should not be exceeded.
- One run may be performed for reference. It may be conducted by a person representing the manufacturer.
- An additional pass or run at the end when all targets are removed when found and their positions recorded may be performed, to simulate the real demining process.

Open questions:

- none

Summary of Breakout Session 4

Rules for Plan of Experiments and Statistical Evaluation

Peter Wilrich, Mate Gaal, Christina Mueller

- The design of the test shall be an orthogonal design, in order to enable the estimate of the main effects of the factors not correlated with the main effects of other factors.
- An orthogonal design requires that the numbers of lanes, operators and devices are powers of two (4, 8, 16...). The number of operators should be as large as possible. Each type of device shall be tested with two specimens.
- The design shall use neutral abbreviations A, B, ..., α, β, ..., 1, 2, ... for names of devices, names of operators, designation of lanes etc. and these should be allocated randomly.
- Targets and depths shall be systematically[2] determined and the same for all lanes. Location of the targets shall be randomly chosen and different in each lane.
- If possible, only a small number of ITOP or other standard targets should be included; instead, more mines.
- If possible, test operators should be chosen randomly out of the group of available trained deminers. They should, however, not be untrained nor be selected out of a group of the most excellent deminers.
- Sensitivity level shall be set by the operators, according to manufacturer's instructions, to give the best detection capability for the given conditions. The sensitivity level and setting shall be recorded.

Open questions:

- Length of the test lanes
- Exact choice and number of targets
- Exact number of devices, lanes, operators, and starts per day.

[2] "Systematically" means that it shall be possible afterwards to create a POD (detection rate) curve as a function of depth for each mine type in each soil with reasonable confidence limits using logistic regression based on generalised linear model.

Annex 2: ITEP Workshop on Reliability Tests for Demining, Recommendations

1. The participants of the ITEP-Workshop on „Reliability Tests for Demining ", December , 16-17, 2003 made the following recommendations in the fields of reliability, testing and research:

RELIABILITY TESTING

Reliability Tests

As a consequence of the evaluation of the test trials within the ITEP-project 2.1.1.2 "Reliability Model for Test&Evaluation of Metal Detectors" and the workshop presentations and discussions it is recommended that:

- A reliability model, including a realistic system model, should be established based on the experience of mine action centers, scientists, deminers, manufacturers and sponsors and should be introduced into the international demining community as conceptual basis.

- Further exercises should be undertaken to develop and define the practical *reliability tests* for both device selection requirements and in-service reliability testing.

It should be noted that, following the above recommendations the results must be practically applicable and there may be:

- A need to live with the consequences and interpretation of the results (risk management technologies).
- Possible increased costs. However, the safety of demining programs should increase and the expenses can be optimized by selection of equipment for specific conditions.

Reliability of Metal Detection

The method of Reliability Tests, and the Modular Reliability Model, proposed by BAM, JRC and ITEP in the ITEP-project 2.1.1.2 and discussed in the break out sessions of the workshop will be subject of an ITEP report including further details of minimum/optimum test design. The issues of the report should be offered for inclusion in the CEN CWA 14747, and International Mine Action Standards (IMAS) as applicable and so far not already included.

When deciding on the selection of detectors for a specific local task it is most important to determine the detection rate (or POD) of each detector in each soil against each mine type, with representative local personnel, as a function of mine depth. The overall false alarm rate should also be recorded. From this "basic cell" the expected detection rates for the actual local scenario can be composed.

Performance reliability testing of other equipment

It is recommended that the experiences gained, through the metal detector reliability trials, are used to help form reliability tests and trials for other demining equipment and methods.

Support of the ITEP work program

The ITEP work program should promote the need for trials, to show where metal detectors are suitable and where they have to be used with care.

Training Standards

The workshop assembly recommends to promote the relevance of training/experience and the need for training standards

RESEARCH

2. Problems continue to exist for manual demining using metal detectors (less than 100% detection rates; high false alarm rates, no one single method can adequately clear mines, lack of any systematic evaluation of potential detection technologies) and these have been confirmed during the ITEP trials. Accordingly:

It is recommended that:

- A realistic system model should be developed and established in the demining R&D community. The basic idea of a systematic and realistic model conception like easyMine is to follow a comprehensive approach for a systematic scientific analysis of landmine detection processes in order to meet the interdisciplinary requirements for R&D in demining closing the gap between field people and scientists.

- Co-operation between the interested parties/institutions is established based on the idea of easyMine and, as part of the comprehensive approach, to increase the efforts in:
  - *Research of soils, mines, detection methods, sensor transfer functions, sensor signal processing, processing result assessment*
  - *socio-economic embedding*
  - *The implementation of International Mine Action Standards*
  - *Support of the toolbox concept*

- Cooperation on raising funds is encouraged.
  - The complexity of the systematic analysis of the holistic concept of landmine detection processes will require national and international level funding to ensure concerted action.

- A *committee/working group /international network of competence,* of mine action centers AND scientists AND deminers AND is manufacturers, is established to improve the communication between them and to integrate more the experience of deminers into further developments.

Annex 3: Oberjettenberg May, Target Positions

x, y - position in the lane (cm)
h - depth measured from the soil surface to the top of the target (cm)

	target	halo radius	Lane 1			Lane 2			Lane 3			Lane 4		
			x	y	h	x	y	h	x	y	h	x	y	h
1.	PPM-2	16	102	128	4	29	606	2	33	628	3	130	137	2
2.	PPM-2	16	90	627	5	31	1099	2	89	1215	3	130	928	2
3.	PPM-2	16	94	1539	4	102	1440	2	25	1658	3	129	1253	3
4.	PMN	16	35	272	5	54	1542	3	14	119	3	64	83	2
5.	PMN	16	42	1698	5	110	1877	2	86	931	2	134	390	3
6.	PMN	16	65	1442	17	98	101	15	98	745	15	45	623	15
7.	Maus	14	53	469	5	50	45	3	83	500	2	77	172	2
8.	Maus	14	88	1315	4	29	910	1	32	959	2	60	1376	2
9.	Maus	14	101	1651	5	55	1050	3	82	1422	1	80	1575	1
10.	Maus	14	29	1847	5	103	1976	2	31	1927	2	108	1901	1
11.	SchAMi DM 31	15	44	1279	7	28	1256	5	86	654	5	127	216	6
12.	TM-46	25	42	42	8	32	1800	6	80	397	5	64	289	5
13.	PT-Mi-Ba-III	11	43	178	8	96	1370	5	55	259	7			
14.	PT-Mi-Ba-III	11	75	1212	8	36	1947	6	55	1881	5			
15.	TM-62 P2	16				56	489	5	35	553	5	67	1003	6
16.	TM-62 P3	16	35	1057	7	63	814	5	73	55	6	110	729	5
17.	TM-62 P3	16	90	1919	7	32	1365	6	76	1830	5			
18.	TM-62 M	26	38	1512	7	74	678	15	33	1078	5	94	1171	15
19.	TM-62 M	26	36	555	17	35	1637	6	39	1513	16	104	1484	6
20.	C0	10	24	367	7	87	1178	5	90	162	5	121	33	5
21.	E0	10							15	211	10	78	443	10
22.	G0	10	34	1115	7	23	1885	5	56	1598	5			
23.	I0	10				24	210	10	19	678	10	43	224	10
24.	K0	10	55	899	7	46	306	5	23	1280	5	46	498	5
25.	K0	10				64	728	10	98	1665	10	45	1645	10
26.	G0	10	87	745	12	74	156	10	68	806	10	38	330	10
27.	I0	10				107	1003	5	19	1367	5	112	568	5
28.	K0	10	91	1800	12	87	1678	10				123	1657	10
29.	100Cr6 ball	11	73	85	22	87	423	20				46	812	20

Annex 4: Oberjettenberg May, Schedule

Days 1, 2, 3, 4

	Start 1	Start 2	Start 3	Start 4	Start 5	Start 6	Start 7	Start 8
Lane 1	A alpha-1	C gamma-2	B beta-1	D delta-2	C gamma-1	A alpha-2	D delta-1	B beta-2
Lane 2	B gamma-1	D alpha-2	A delta-1	C beta-2	D alpha-1	B gamma-2	C beta-1	A delta-2
Lane 3	C delta-1	A beta-2	D gamma-1	B alpha-2	A beta-1	C delta-2	B alpha-1	D gamma-2
Lane 4	D beta-1	B delta-2	C alpha-1	A gamma-2	B delta-1	D beta-2	A gamma-1	C alpha-2

Days 5, 6, 7, 8

	Start 1	Start 2	Start 3	Start 4	Start 5	Start 6	Start 7	Start 8
Lane 1	E alpha-1	G gamma-2	F beta-1	H delta-2	G gamma-1	E alpha-2	H delta-1	F beta-2
Lane 2	F gamma-1	H alpha-2	E delta-1	G beta-2	H alpha-1	F gamma-2	G beta-1	E delta-2
Lane 3	G delta-1	E beta-2	H gamma-1	F alpha-2	E beta-1	G delta-2	F alpha-1	H gamma-2
Lane 4	H beta-1	F delta-2	G alpha-1	E gamma-2	F delta-1	H beta-2	E gamma-1	G alpha-2

Days 1, 2, 5, 6: low sensitivity
Days 3, 4, 7, 8: high sensitivity

In starts 5-8 operators move in the opposite direction then in starts 1-4.

Annex 5: Benkovac July, Target Positions

x, y - position in the lane (cm)
h - depth measured from the soil surface to the top of the target (cm)

	target	halo radius	h	Lane 1 x	Lane 1 y	Lane 2 x	Lane 2 y	Lane 3 x	Lane 3 y	Lane 4 x	Lane 4 y	Lane 5 x	Lane 5 y	Lane 6 x	Lane 6 y	Lane 7 x	Lane 7 y	Lane 8
1.	100Cr6 ball	11	10	90	326	86	2230	78	1038	56	139	85	1241	72	1852	14	757	All passes were performed in Lane 4.
2.	E0	10	5	27	60	58	491	36	699	85	2550	62	1461	20	1055	47	498	
3.	G0	10	5	87	683	33	1308	70	131	69	169	74	534	31	1146	19	2285	
4.	K0	10	5	62	1070	93	2750	79	2175	78	1238	12	2754	90	68	21	300	
5.	PMA-1A	13	0	76	120	68	1900	28	2522	50	68	80	1161	23	776	77	2167	
6.	PMA-1A	13	5	33	1585	80	1562	24	241	73	1099	15	1104	31	359	79	848	
7.	PMA-1A	13	5	40	2520	23	1590	20	535	29	2273	65	1697	78	705	31	1614	
8.	PMA-1A	13	5	24	2798	19	2905	88	2713	19	2843	19	1965	23	1702	23	2790	
9.	PMA-1A	13	10	75	1154	71	628	50	1640	21	576	52	2919	30	110	23	575	
10.	PMA-1A	13	13	16	1006	30	976	40	2770	60	961	20	66	77	1998	29	1984	
11.	PMA-1A	13	20	87	405	84	130	22	1941	65	1520	30	1531	76	1194	23	426	
12.	PMA-2	11	0	60	2698	22	1705	88	2821	44	2128	23	2350	53	1264	22	39	
13.	PMA-2	11	5	24	624	29	228	43	470	81	389	35	1877	87	394	31	1166	
14.	PMA-2	11	5	22	2120	30	709	52	628	30	1180	45	2050	17	1657	47	1868	
15.	PMA-2	11	5	88	2612	29	1199	20	1350	58	1624	80	2407	35	2788	78	2841	
16.	PMA-2	11	10	81	2788	25	2268	72	1154	51	2468	64	1304	59	913	43	1056	
17.	PMA-2	11	13	38	939	88	2078	79	282	30	332	59	140	32	1943	29	2980	
18.	PMA-2	11	20	86	863	51	2551	78	1230	57	2753	22	1401	38	254	82	70	
19.	PMA-3	10	0	56	2878	20	430	80	1390	57	519	30	616	17	2524	56	1303	
20.	PMA-3	10	5	18	1374	78	380	68	389	60	261	69	689	72	1395	69	1548	
21.	PMA-3	10	5	83	1695	66	2812	38	760	55	1355	86	1633	71	1691	78	2346	
22.	PMA-3	10	5	76	2222	85	2933	74	1550	59	2024	20	2532	15	2126	63	2637	
23.	PMA-3	10	10	38	510	50	2353	17	1490	18	1865	48	2851	61	1588	44	1465	
24.	PMA-3	10	13	23	1221	82	1029	20	344	58	1715	41	809	74	998	27	2579	
25.	PMA-3	10	20	76	2144	80	1130	64	2645	71	2219	49	961	55	2606	44	2441	
26.	PROM-1	14	0	15	1795	78	1463	79	1706	62	666	25	492	65	2275	57	947	
27.	PROM-1	14	0	15	2936	28	2428	66	1790	53	1437	77	1826	81	2732	20	2702	
28.	PROM-1	14	5	47	786	58	1796	21	65	58	1819	20	213	65	2193	61	369	
29.	PROM-1	14	5	61	1450	67	2152	50	2240	41	1934	76	452	45	2415	46	1743	
30.	PROM-1	14	5	81	2447	86	2470	58	2403	68	2915	24	2626	84	2928	85	2528	
31.	TMA-3	22	10	68	1843									56	545	36	165	
32.	TMA-4	23	10			84	273	40	2920	49	2653	81	2197					
33.	TMRP-6	19	10	40	1510	22	1968	57	961			54	290					
34.	TMM-1	26	10							60	851			30	1517	32	675	

Annex 6: Benkovac July, Schedule

Day 1 & Day 3

	Start 1	Start 2	Start 3	Start 4	Start 5	Start 6	Start 7	Start 8
Lane 1	A alpha-1	C gamma-2	B beta-1	D delta-2	C gamma-1	A alpha-2	D delta-1	B beta-2
Lane 2	B gamma-1	D alpha-2	A delta-1	C beta-2	D alpha-1	B gamma-2	C beta-1	A delta-2
Lane 3	C delta-1	A beta-2	D gamma-1	B alfa-2	A beta-1	C delta-2	B alpha-1	D gamma-2
Lane 4	D beta-1	B delta-2	C alpha-1	A gamma-2	B delta-1	D beta-2	A gamma-1	C alpha-2

Day 2 & Day 4

	Start 1	Start 2	Start 3	Start 4	Start 5	Start 6	Start 7	Start 8
Lane 5	C alpha-1	A gamma-2	D beta-1	B delta-2	A gamma-1	C alpha-2	B delta-1	D beta-2
Lane 6	D gamma-1	B alpha-2	C delta-1	A beta-2	B alpha-1	D gamma-2	A beta-1	C delta-2
Lane 7	A delta-1	C beta-2	B gamma-1	D alfa-2	C beta-1	A delta-2	D alpha-1	B gamma-2
Lane 8	B beta-1	D delta-2	A alpha-1	C gamma-2	D delta-1	B beta-2	C gamma-1	A alpha-2

Day 5 & Day 7

	Start 1	Start 2	Start 3	Start 4	Start 5	Start 6	Start 7	Start 8
Lane 1	E alpha-1	G gamma-2	F beta-1	H delta-2	G gamma-1	E alpha-2	H delta-1	F beta-2
Lane 2	F gamma-1	H alpha-2	E delta-1	G beta-2	H alpha-1	F gamma-2	G beta-1	E delta-2
Lane 3	G delta-1	E beta-2	H gamma-1	F alfa-2	E beta-1	G delta-2	F alpha-1	H gamma-2
Lane 4	H beta-1	F delta-2	G alpha-1	E gamma-2	F delta-1	H beta-2	E gamma-1	G alpha-2

Day 6 & Day 8

	Start 1	Start 2	Start 3	Start 4	Start 5	Start 6	Start 7	Start 8
Lane 5	G alpha-1	E gamma-2	H beta-1	F delta-2	E gamma-1	G alpha-2	F delta-1	H beta-2
Lane 6	H gamma-1	F alpha-2	G delta-1	E beta-2	F alpha-1	H gamma-2	E beta-1	G delta-2
Lane 7	E delta-1	G beta-2	F gamma-1	H alfa-2	G beta-1	E delta-2	H alpha-1	F gamma-2
Lane 8	F beta-1	H delta-2	E alpha-1	G gamma-2	H delta-1	F beta-2	G gamma-1	E alpha-2

Days 1, 2, 5, 6: low sensitivity
Days 3, 4, 7, 8: high sensitivity

Annex 7: Oberjettenberg November, Target Positions

x, y - position in the lane (cm)
h - depth measured from the soil surface to the top of the target (cm)

	target	halo radius	Lane 1			Lane 2			Lane 3			Lane 4			Lane 5			Lane 7			Lane 8		
			h	x	y	h	x	y	h	x	y	h	x	y	h	x	y	h	x	y	h	x	y
1.	100Cr6 ball	11	10	102	74	10	96	425	10	98	1660	10	89	815	10	38	960	10	82	1300	10	55	1905
2.	C0	10	5	94	240	5	96	1180	5	88	162	5	124	31	5	81	358	5	34	820	5	48	810
3.	E0	10	5	34	1608	5	96	1770	5	20	1370	5	119	1657	5	96	1391	5	96	1925	5	81	1376
4.	G0	10	5	98	1025	5	23	1878	5	64	1598	5	81	440	5	60	1311	5	47	745	5	93	285
5.	I0	10	5	37	652	5	77	182	5	68	808	5	41	345	5	64	1520	5	27	78	5	34	875
6.	K0	10	5	71	953	5	58	306	5	27	1285	5	46	498	5	31	298	5	95	1790	5	45	225
7.	Maus	14	5	56	470	3	50	49	2	83	500	2	72	177	20	42	210	20	93	283	20	86	1003
8.	Maus	14	4	89	1315	1	28	910	2	31	960	2	60	1376	0	106	260	0	37	495	0	81	1838
9.	Maus	14	5	107	1653	3	55	1060	1	87	1427	1	79	1585	5	82	625	5	91	872	5	35	730
10.	Maus	14	5	32	1847	2	104	1972	2	31	1927	1	105	1906	13	33	1140	13	86	1423	13	90	396
11.	Maus	14													5	38	1440	5	75	1530	5	67	1295
12.	Maus	14													10	95	1759	10	53	1857	10	29	1107
13.	Maus	14													5	93	1850	5	32	1732	5	31	1632
14.	PMN	16	5	35	270	15	96	105	3	14	119	2	60	86	5	38	533	5	104	646	5	64	660
15.	PMN	16	17	70	1440	3	56	1556	15	91	750	3	128	392	0	90	750	0	85	225	0	37	1186
16.	PMN	16	5	46	1698	2	110	1875	2	84	934	15	49	630	5	92	1244	5	37	678	5	50	1770
17.	MS3	16													10	39	671	10	45	610	10	71	133
18.	MS3	16													13	90	840	13	92	1599	13	96	1135
19.	MS3	16													20	94	995	20	91	450	20	89	1966
20.	MS3	16													5	33	1680	5	35	1135	5	45	1495
21.	PPM-2	16	4	103	123	2	29	607	3	37	633	2	121	140									
22.	PPM-2	16	5	88	624	2	31	1099	3	89	1215	2	130	928									
23.	PPM-2	16	4	98	1540	2	103	1438	3	28	1658	3	129	1253									
24.	PT-Mi-Ba-III	11	8	41	178	5	91	1368	7	55	259												
25.	PT-Mi-Ba-III	11	8	75	1211	6	38	1945	5	57	1876												
26.	SchAMi DM 31	15	7	47	1278	5	28	1256	5	86	654	6	120	216									
27.	TM-46	25	8	42	38	6	35	1811	5	78	397	5	62	301									
28.	TM-62 M	26	17	44	459	15	72	675	5	33	1079	15	94	1171	10	67	480	10	83	371	10	86	1703
29.	TM-62 M	26	7	42	1506	6	36	1639	16	36	1510	6	98	1490									
30.	PMA-S	11	5	58	491	5	36	558							5	37	37	5	91	525	5	27	63
31.	PMA-S	11													0	82	111	0	94	1084	0	84	1240
32.	PMA-S	11													13	96	170	13	70	1220	13	88	1572
33.	PMA-S	11													20	23	800	20	78	1017	20	45	560
34.	PMA-S	11	7	40	1054	5	61	816	6	75	56				5	52	885	5	91	1671	5	98	768
35.	PMA-S	11				6	36	1365	5	77	1828				10	54	1070	10	100	118	10	33	1423
36.	PMA-S	11													5	79	1620	5	58	1962	5	50	955

light blue	Positions of the mines kept from the May Trials (Lanes 1-4)
orange	ITOP standard targets
yellow	Positions of the mines analogous to the Benkovac trials (Lanes 5-8)

Annex 8: Oberjettenberg November, Schedule

Q1

	Day 1								Day 2							
Lane	Start 1		Start 2		Start 3		Start 4		Start 5		Start 6		Start 7		Start 8	
1	A	alpha 1	H	delta 2	C	beta 1	F	gamma 2	B	alpha 2	E	gamma 1	D	beta 2	G	delta 1
2	D	beta 1	E	gamma 2	B	alpha 1	G	delta 2	C	beta 2	H	delta 1	A	alpha 2	F	gamma 1
3	C	alpha 2	F	delta 1	A	beta 2	H	gamma 1	D	alpha 1	G	gamma 2	B	beta 1	E	delta 2
4	B	beta 2	G	gamma 1	D	alpha 2	E	delta 1	A	beta 1	F	delta 2	C	alpha 1	H	gamma 2
5	G	delta 1	B	alpha 2	E	gamma 1	D	beta 2	H	delta 2	C	beta 1	F	gamma 2	A	alpha 1
7	E	delta 2	D	alpha 1	G	gamma 2	B	beta 1	F	delta 1	A	beta 2	H	gamma 1	C	alpha 2
8	H	gamma 2	A	beta 1	F	delta 2	C	alpha 1	G	gamma 1	D	alpha 2	E	delta 1	B	beta 2

Q2

	Day 2				Day 3											
Lane	Start 1		Start 2		Start 3		Start 4		Start 5		Start 6		Start 7		Start 8	
1	C	alpha 1	F	delta 2	A	beta 1	H	gamma 2	D	alpha 2	G	gamma 1	B	beta 2	E	delta 1
2	B	beta 1	G	gamma 2	D	alpha 1	E	delta 2	A	beta 2	F	delta 1	C	alpha 2	H	gamma 1
3	A	alpha 2	H	delta 1	C	beta 2	F	gamma 1	B	alpha 1	E	gamma 2	D	beta 1	G	delta 2
4	D	beta 2	E	gamma 1	B	alpha 2	G	delta 1	C	beta 1	H	delta 2	A	alpha 1	F	gamma 2
5	E	delta 1	D	alpha 2	G	gamma 1	B	beta 2	F	delta 2	A	beta 1	H	gamma 2	C	alpha 1
~~6~~	~~H~~	~~gamma 1~~	~~A~~	~~beta 2~~	~~F~~	~~delta 1~~	~~C~~	~~alpha 2~~	~~G~~	~~gamma 2~~	~~D~~	~~alpha 1~~	~~E~~	~~delta 2~~	~~B~~	~~beta 1~~
7	G	delta 2	B	alpha 1	E	gamma 2	D	beta 1	H	delta 1	C	beta 2	F	gamma 1	A	alpha 2
8	F	gamma 2	C	beta 1	H	delta 2	A	alpha 1	E	gamma 1	B	alpha 2	G	delta 1	D	beta 2

Q3

	Day 4								Day 5							
Lane	Start 1		Start 2		Start 3		Start 4		Start 5		Start 6		Start 7		Start 8	
1	E	alpha 1	D	delta 2	G	beta 1	B	gamma 2	F	alpha 2	A	gamma 1	H	beta 2	C	delta 1
2	H	beta 1	A	gamma 2	F	alpha 1	C	delta 2	G	beta 2	D	delta 1	E	alpha 2	B	gamma 1
3	G	alpha 2	B	delta 1	E	beta 2	D	gamma 1	H	alpha 1	C	gamma 2	F	beta 1	A	delta 2
4	F	beta 2	C	gamma 1	H	alpha 2	A	delta 1	E	beta 1	B	delta 2	G	alpha 1	D	gamma 2
5	C	delta 1	F	alpha 2	A	gamma 1	H	beta 2	D	delta 2	G	beta 1	B	gamma 2	E	alpha 1
~~6~~	~~B~~	~~gamma 1~~	~~G~~	~~beta 2~~	~~D~~	~~delta 1~~	~~E~~	~~alpha 2~~	~~A~~	~~gamma 2~~	~~F~~	~~alpha 1~~	~~C~~	~~delta 2~~	~~H~~	~~beta 1~~
7	A	delta 2	H	alpha 1	C	gamma 2	F	beta 1	B	delta 1	E	beta 2	D	gamma 1	G	alpha 2
8	D	gamma 2	E	beta 1	B	delta 2	G	alpha 1	C	gamma 1	H	alpha 2	A	delta 1	F	beta 2

Q4

	Day 5				Day 6											
Lane	Start 1		Start 2		Start 3		Start 4		Start 5		Start 6		Start 7		Start 8	
1	G	alpha 1	B	delta 2	E	beta 1	D	gamma 2	H	alpha 2	C	gamma 1	F	beta 2	A	delta 1
2	F	beta 1	C	gamma 2	H	alpha 1	A	delta 2	E	beta 2	B	delta 1	G	alpha 2	D	gamma 1
3	E	alpha 2	D	delta 1	G	beta 2	B	gamma 1	F	alpha 1	A	gamma 2	H	beta 1	C	delta 2
4	H	beta 2	A	gamma 1	F	alpha 2	C	delta 1	G	beta 1	D	delta 2	E	alpha 1	B	gamma 2
5	A	delta 1	H	alpha 2	C	gamma 1	F	beta 2	B	delta 2	E	beta 1	D	gamma 2	G	alpha 1
~~6~~	~~D~~	~~gamma 1~~	~~E~~	~~beta 2~~	~~B~~	~~delta 1~~	~~G~~	~~alpha 2~~	~~C~~	~~gamma 2~~	~~H~~	~~alpha 1~~	~~A~~	~~delta 2~~	~~F~~	~~beta 1~~
7	C	delta 2	F	alpha 1	A	gamma 2	H	beta 1	D	delta 1	G	beta 2	B	gamma 1	E	alpha 2
8	B	gamma 2	G	beta 1	D	delta 2	E	alpha 1	A	gamma 1	F	alpha 2	C	delta 1	H	beta 2

Annex 9: Questionnaire

Please circle the most apropriate answer for each detector:
1 - no,
2 - I can not decide,
3 - yes.

1. Is the detector easy to assemble and disassemble?
2. Are the controls easy to understand?
3. Is the start up procedure simple?
4. Is the ground compensating procedure easy to understand?
5. Is it easy to pin point the target?
6. Are the alarm tones easy to distinguish and understand?
7. Is the confidence tone easy to understand?
8. Is the detector easy to adjust for comfort?
9. Are you comfortable with the weight and mechanical handling?
10. After complete training would you feel confident with this detector in a live minefield?
11. What is your overall impression of this detector? (descriptive answer)

Oberjettenberg May, average of all operators' answers

Question	detector U	detector X	detector Y	detector Z
1	1	3	3	2.625
2	1.375	2.375	3	2.75
3	2.25	2.5	3	3
4	2	2	3	2
5	1	2.5	3	2.625
6	1.875	2	2.875	2.75
7	2.125	1.25	3	2.875
8	1.125	1.875	3	2.125
9	1.5	1.5	3	2.125
10	1	1.5	3	2.25
average of 1-9	1.583333333	2.111111111	2.986111111	2.541666667

Benkovac July, average of all operators' answers

Question	detector U	detector X	detector Y	detector Z
1	2.875	3	3	3
2	2.25	3	2.75	3
3	1.625	3	2.75	2.75
4	2	2.25	1.875	2.625
5	2.375	3	1.875	2.75
6	2.625	3	1.875	3
7	2.75	3	2.25	3
8	2.75	3	2.75	2.75
9	2.875	2.125	3	1.375
10	2.5	3	2.25	3
average of 1-9	2.458333333	2.819444444	2.458333333	2.694444444

Oberjettenberg November, average of all operators' answers

Question	detector U	detector W	detector X	detector Y	detector Z
1	3	3	3	2.875	3
2	2.875	3	2.5	3	3
3	3	3	3	3	3
4	3	3	3	3	3
5	2.25	3	2.75	3	2.875
6	2.625	3	2.75	2.875	2.875
7	2.75	3	2.875	3	3
8	3	3	3	3	2.875
9	2.625	3	2.125	2.625	2.625
10	2.125	2.375	2.125	2.375	2.625
average of 1-9	2.791666667	3	2.777777778	2.930555556	2.916666667