Proc. Vth International Workshop, Advances in Signal Processing for Non Destructive Evaluation of Materials
Québec City (Canada),2-4 Aug. 2005. © X. Maldague ed., É. du CAO (2006), ISBN 2-9809199-0-X

Reliability Tests for Demining

Christina Mueller, Mate Gaal, Mato Pavlovic, Martina Scharmach,
BAM Berlin, Germany
Unter den Eichen 87,
D12205 Berlin, Germany
Christina.Mueller@bam.de
Peter Wilrich FU Berlin, Germany

Abstract

The total detection reliability of a mine searching system is analogous to NDE-systems governed by three elements;

- intrinsic capability - which describes the basic physical-technical capability of the method
- application factors - including those due to environment
- human factor - the effect of human operators on the detection reliability.

Some of these can be determined in simple laboratory measurements in which the effect on detection capability of individual parameters is measured. However, the human factor and some aspects of the effects of environmental conditions on the system need to be treated statistically.

By far the most common "mine searching system" in use today is the metal detector. The test and evaluation procedures for metal detectors described in CEN CWA 14747: 2003 include the above ideas. This is why, in addition to parameter tests, they include detection reliability or blind field tests under local conditions with local personnel.

A series of three field trials was performed in the ITEP-project 2.1.1.2 in 2003 “Reliability Model for Test and Evaluation of Metal Detectors” and another trial named ITEP-project 2.1.1.8 in 2005, in order to specify the optimum conditions to obtain reliable trial results with affordable effort. Each set of specific working conditions is characterized in terms of a combination of one mine type in one soil with one detector handled by local personnel. For each set of conditions, the searching system will deliver a working performance, expressed as mine detection rates as a function of mine depth, and a certain overall false alarm rate. During the ITEP-trials in Benkovac and Oberjettenberg, the authors learned to determine this function separately for each mine type in each soil. This is especially important for low-metal mines in soil that can influence metal detectors, as will be illustrated for the case of the PMA2. Two discussion points remained after 2003; how representative the trials are of field conditions and what is the statistical set-up required if we are to distinguish between the capabilities of individual detectors. In the trials of the ITEP-project 2.1.1.8 a more suitable design of experiment was applied (only two mine types in two soil types on fewer depths). The human factor was considered more carefully in applying some elements of the local SOP(Standard Operating Procedure), thus coming closer to the local practise and field conditions. Until then two detector models had been improved as a result of the previous trials.

Introduction and Background

The CEN Working Group 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). In reliability tests, the probability of Detection (POD) and Receiver Operating Characteristics (ROC) curves help to summarize the performance results. Under the umbrella of ITEP (International Test and Evaluation Program for Humanitarian Demining) a number of test trials with metal detectors have been conducted. The aim was to specify the trial set-up and the statistical rules necessary to achieve true, repeatable and reproducible results under representative field conditions. 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. A full report about the trial conditions and results, including rules for minimum number of targets, operators and numbers of test repetitions necessary to achieve true and reproducible results, have been published on the ITEP website in September 2004 and December 2005.

POD and ROC – Summary of Detection Rates and False Alarms

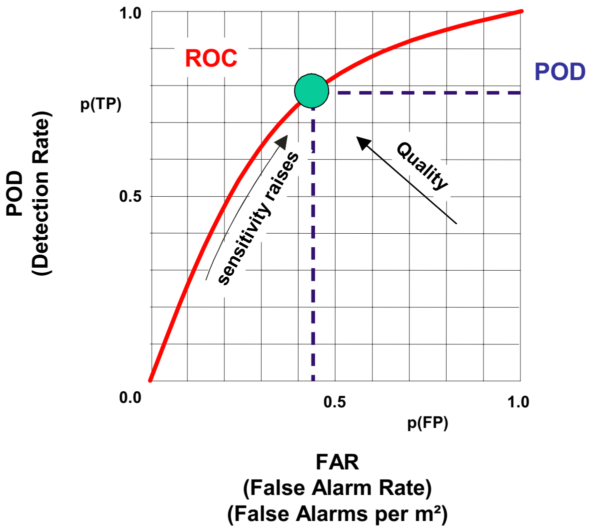


Figure 1: Explanation for ROC and POD diagrams

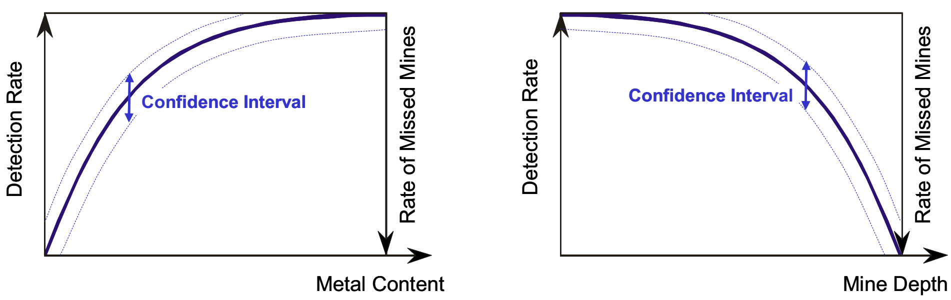


Figure 2: Schematic Representation of POD Curves

The ROC (Receiver Operating Characteristic) of a mine detection system /1/ 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. The ROC shows how successful the system is in distinguishing between a real 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.

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.

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).

To fit a curve for the dependence of POD on depth, non-linear logistic regression was applied. The POD is transformed according to the following equation and a linear dependence on depth is assumed:

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

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

All these points and curves need to be interpreted in connection with the corresponding confidence bounds 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. The smooth POD or detection rate curves, presented in Figure 2 were determined by the advanced logistic regression model mentioned above/2/. 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 /3, 4, 5, 7/.

Overview about the Parameter Matrix of the Trials

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**
	<u>Trial 2005:</u> **→ Cooperative (Soil from Sisak)** **→ Homogenous uncooperative, frequency dependent susceptibility (Soil**	**PMA1A and PMA2 only**	**→ Local SOP and System** **→ Quality Assurance** **→ long training** **→ 5 working**

from Obrovac)		**hours**

Figure 3: Test Parameters

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 in 2003 and an other one in 2005 for which the main parameter set up can be seen in Figure 3. The first and third took place in Oberjettenberg WTD 52 on the testing ground of the German Army and the second and the fourth in Benovac, Croatia (CROMAC testing center CTRO).

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, we 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.

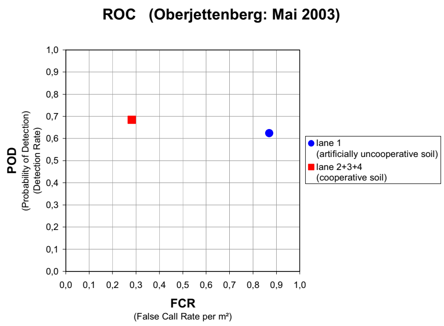
The new trials in May 2005 (ITEP-project 2.1.1.8) were conducted in Benkovac, Croatia. The same training scheme as in Oberjettenberg November trials was applied, but the training was twice longer. Other improvements were: a reduction to two soil types, two mine types distributed to depths only between 0 cm and 15 cm and the implementation of the working conditions as close as possible to the conditions in minefields (protective equipment, working hours, section leader and quality assurance). Two metal detector models were new in these trials; the manufacturers improved the models used in the last trials based on their results.

Results of the trials

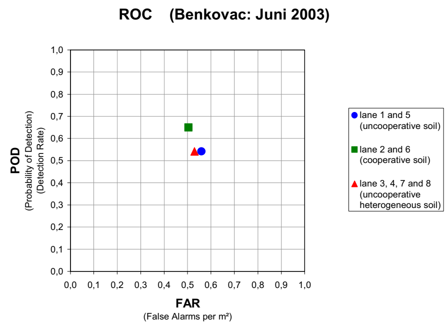
Figure 4 shows the overall results of each trial set, in ROC diagrams. These diagrams illustrate the influence of the factors (Application factor and Human factor) degrading the performance of all the detectors, without distinguishing between individual detectors. The result of inexperienced operators 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:

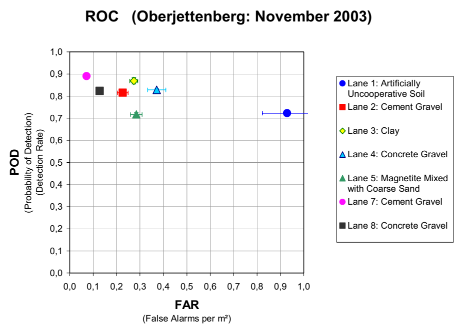
1) Many of the targets were very deeply buried and in some cases beyond the physical capability of some of the detectors. Minimum metal mines, which are inherently difficult to detect, were buried according to a systematic depth distribution, ranging from 0 to 20cm in order to evaluate the detection 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.
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.



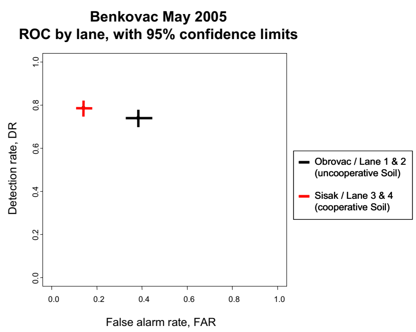
a



b



c



d

Figure 4: ROC diagrams for different soil and human factor conditions

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 fig. 4c 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. We tried to find the answers to these questions in the trials accomplished in May 2005 again in Benkovac Croatia. This time all 4 deminers were actually working, the training was long and careful like in the November 2003 trial and the local SOP(Standard Operating Procedure) was implemented. The improvements from a) to c) and from b) to d) show how important the proper treatment of the human factor is.

Example of a set of Resulting Curves from Year 2003: Detection Rates as Function of Depth and False Alarms for the PMA2 in different Soils

The following figure gives an overview of soil parameters of all trials 2003.

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	5 ± 2	244 ± 64	6,1
Lane 2 cement gravel	no signal	0 ± 1	- 0,2
Lane 3 clay	no signal	2 ± 1	- 0,5
Lane 4 concrete gravel	no signal	6 ± 1	- 0,5
Lane 5 magnetite mixed with coarse sand	4,5 ± 0,7	3000 ± 500	6 ± 7
Lane 7 cement gravel	no signal	-1,0 ± 0,2	-0,1 ± 0,2
Lane 8 concrete gravel	no signal	7 ± 1	-0,1 ± 0,1

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 from Sisac	no signal	13 ± 2	0,6
Lanes 1, 5 (uncooperative) Laterite soil from Obrovac	18,8 ± 0,9	154 ± 13	25,5
Lanes 3, 4, 7, 8 (uncooperative heterogeneous) local red Bauxite from Bencovac	19,7 ± 2,5	190 ± 36	35,4

Figure 5: Overview of magnetic properties of the soils

In the following Figures the individual detector results are illustrated for the PMA2 minimum metal mine under ideal conditions, i.e. neutral soil without metal contamination, well trained operators and optimized working hours. Figures 6a-d show the detection rates as function of the burial depth for each device separately and Figure 6e shows the ROC points of all devices together.

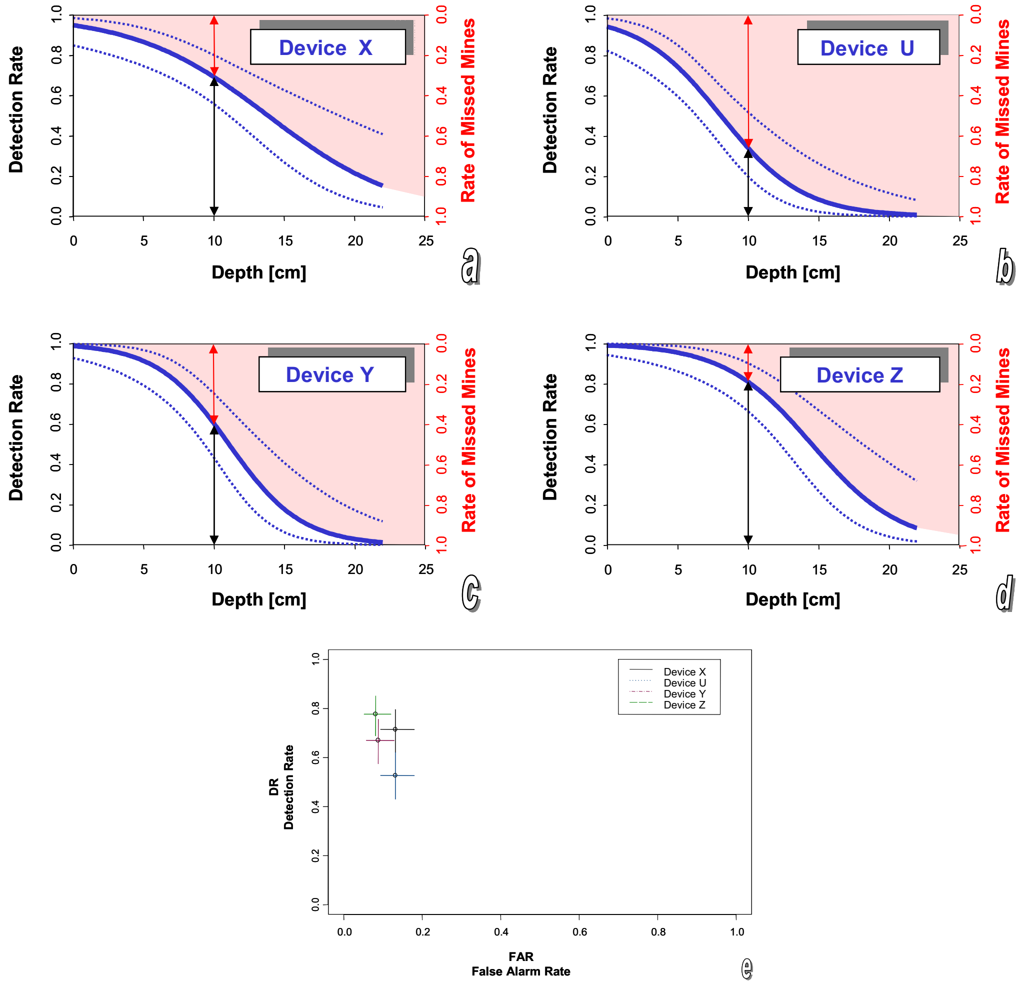


Figure 6: Neutral cooperative soil; only mine PMA-2. Mean value of ROC (detection rate versus false alarm rate) with 95% confidence limits for the different devices

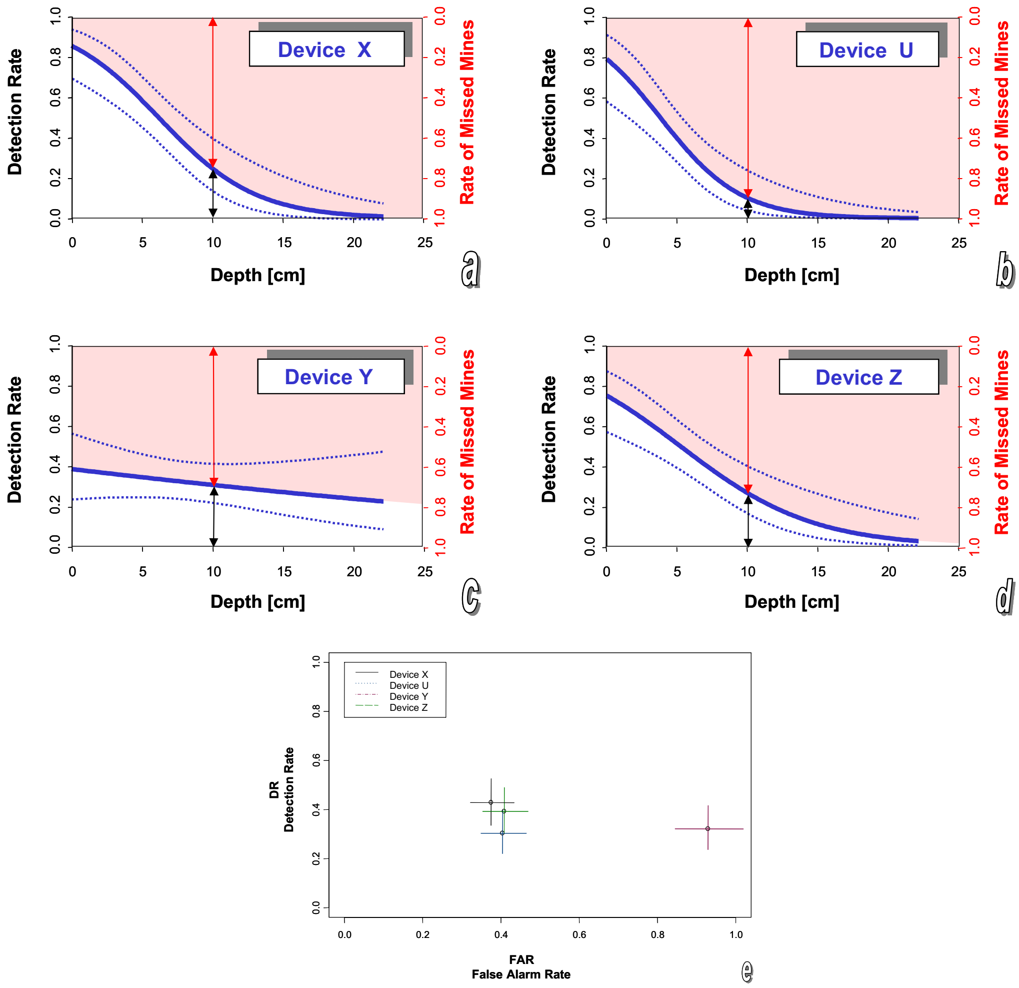
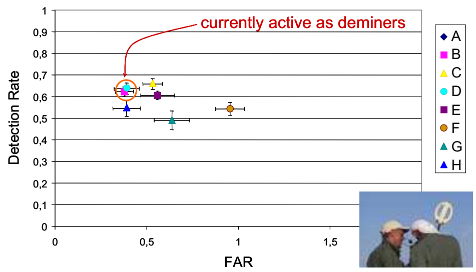
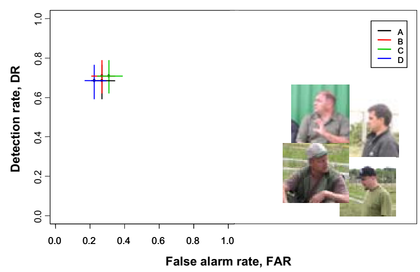


Figure 7: Uncooperative soil, heterogeneous, with frequency dependent susceptibility, red bauxite with neutral stones. Detection rate as function of mine (PMA-2 only) depth for the different four devices with 95% confidence limits



Results of the year 2003



Results of the year 2005, with 95% confidence limits, only PMA-2

Figure 8: ROC-Results for different deminers of the years 2003 and 2005 in Benkovac

Figures 7a-d and Figure 7e present the same results for the most difficult soil. The anomalous result for detector Y is due to a high FAR in the uncooperative soil, up to one false alarm per m^2 and the spuriously higher detection rate at large depth. The latter phenomenon can be explained by the fact that some of the “true” positive indications appear to be signals from the soil that happened to fall within the halo of a target, so that the apparent POD does not approach zero at large depth. To avoid this type of anomaly, the soil compensation and sensitivity of the detector should be adjusted to produce an acceptable low FAR prior to starting the blind trial. CWA 14747: 2003 section 8.1.5 specifies a procedure for checking the adjustment of a metal detector to the soil under test. The test is only to be considered valid if the detector can be adjusted in a representative 1m × 1m set-up area so that no false alarms are given when placed on the soil surface and then raised 30mm above it. It seems likely that detector Y was not adjusted (or not adjustable) according to this procedure.

Figure 8 compares the results of each individual deminer in the two trials in Benkovac 2003 and 2005 in ROC diagrams. The lower scatter between their results in 2005 clearly indicate that the human factor influence improved significantly by selecting currently active deminers and providing them adequate training and working conditions according to the local SOP.

Example of a set of Resulting Curves from Year 2005: Detection Rates as Function of Depth and False Alarms for the PMA2 in different Soils

Soil Types in Benkovac Trials	Susceptibility at 958 Hz (10-5 SI)	Susceptibility difference at 465 and 4650 Hz (10-5 SI)
Lanes 1 and 2 (uncooperative)	154 ± 13	25,5
Lanes 3 and 4 (cooperative)	13 ± 2	0,6

Figure 9: Overview of magnetic properties of the soils

In the following figures the individual detector results are illustrated for the PMA2 mine under cooperative conditions for Sisak soil (fig 10a-d): detection rates as function of the PMA2 depth for each device separately and e) the ROC points of all devices together.

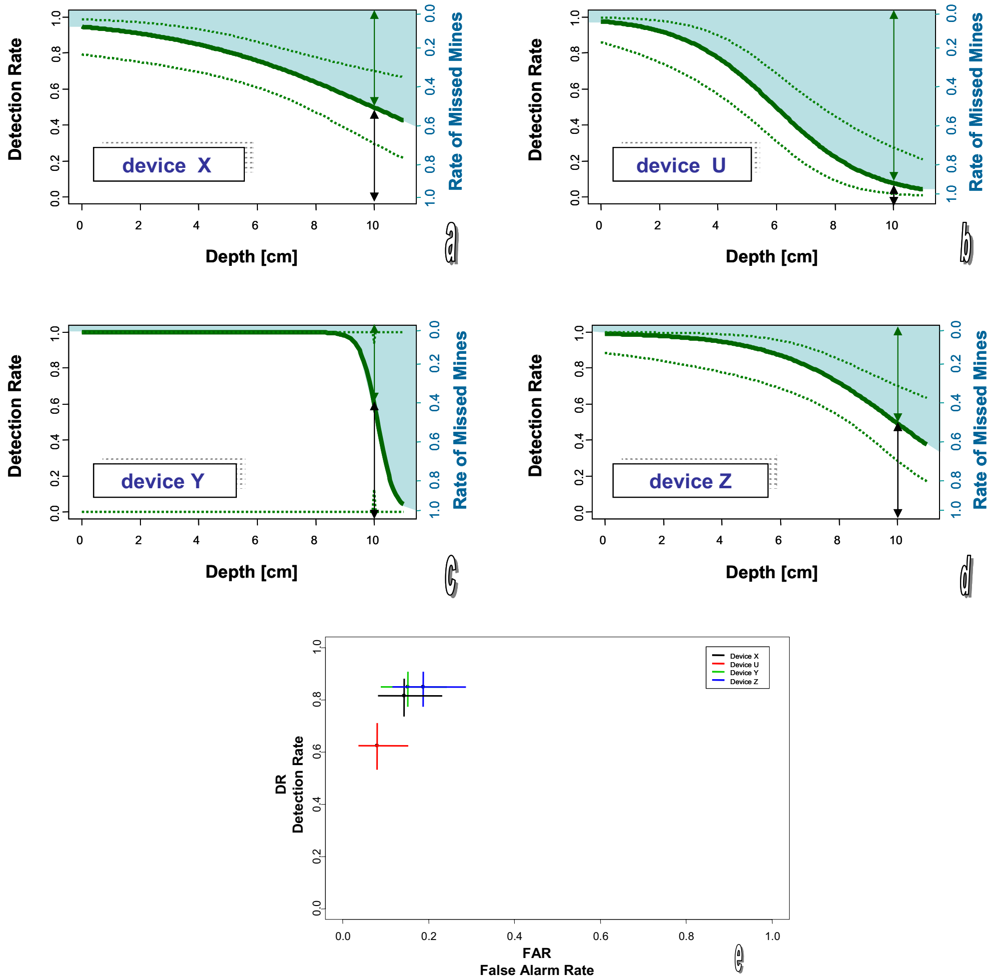


Figure 10: Lanes 3 an 4 (cooperative soil), estimated detection rate with 95% confidence limits for PMA2 – This diagram is comparable with Figure 7

The same curves for each detector but for uncooperative conditions (Obrovac soil) are shown in the figures 11a-e. Especially the comparison of the results for uncooperative soil (fig. 7 – 2003 and fig. 11 – 2005) reveals clearly the improvement by i) better trial conditions and ii) the improvement of the hardware/software of detectors. To make the point i) clearer we put the results of 2003 and 2005 of the identical detectors for the same mines and soils in one diagram in figure 12. This diagram is closer to the realistic detection rates and also indicates more clearly the differences in device performance.
It is the opinion of the authors that the trial procedure with a clear experimental design, careful training and implementation of local working conditions should be added to the existing standard of testing metal detectors CWA 14747:2003, as well as the new insight about maximum detection distance measurements.

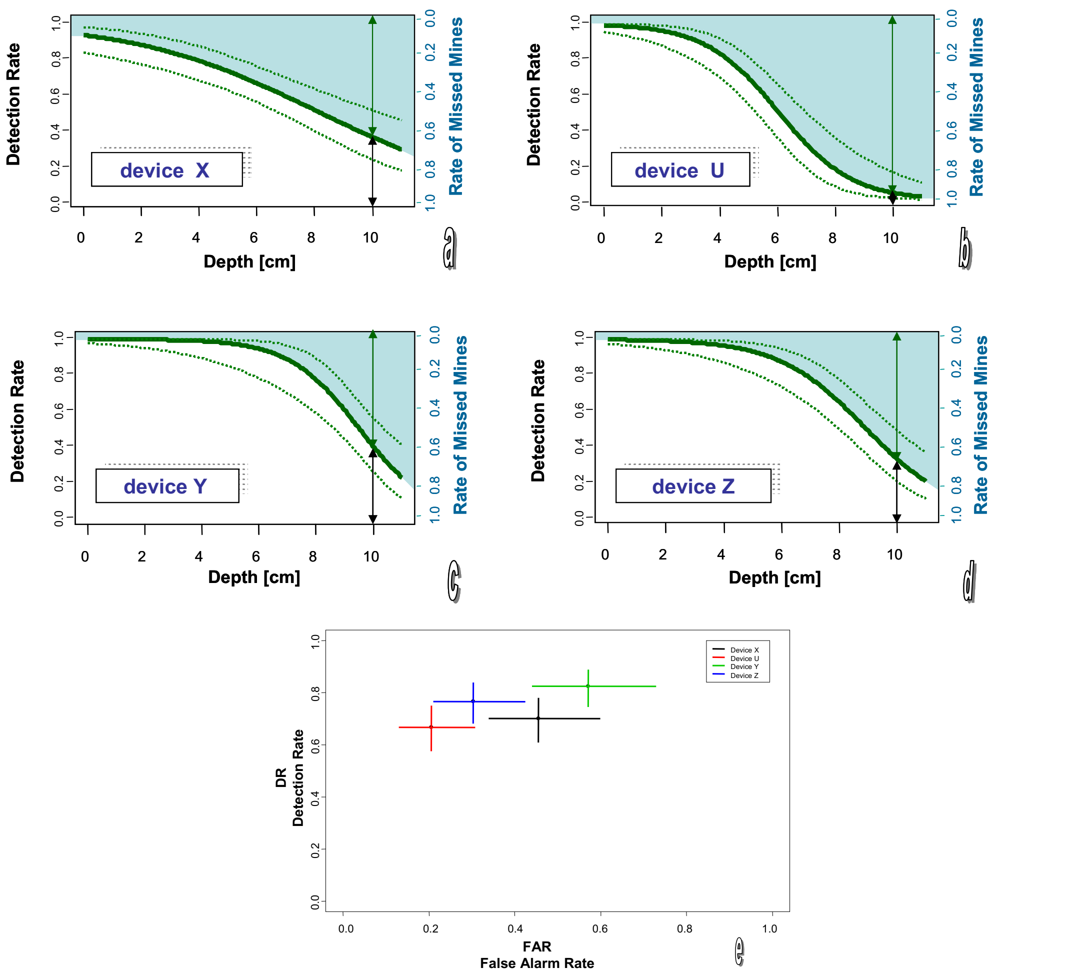


Figure 11: uncooperative soil from Obrovac with a high frequency-dependent magnetic susceptibility; estimated detection rate with 95% confidence limits for PMA2. This soil is very similar to that on Figure 8, though less heterogeneous.

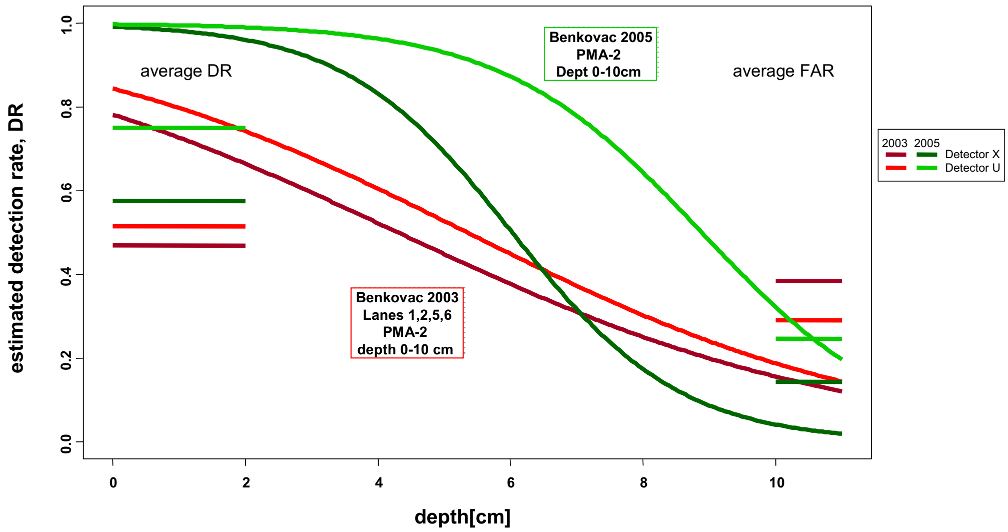


Figure 12: Comparison of individual results of the same detectors (U and X) in Benkovac in the same soils and on the same targets.

Maximum Detection Distance Measurements

For the first time ever (according to our knowledge) the maximum detection distance of a certain combination detector-target was measured several times with different operators (four deminers, twice). The maximum detection distance is the distance between the search head of the metal detector to the top of the target at which the detector starts to give clear signals. The results clearly show that there is a variance in the results that must be taken into account. It seems that the maximum detection distance measurements give results that change within minutes. These changes are not caused only by the operators (deminers), but also by the hardware of the devices.
Figure 13 shows the results for PMA2 in Obrovac soil (lane 1+2) and Sisak soil (Lane 3+4) and in air for PMA2 and PMA1A as mean value and standard deviation of 8 single measurements. The results point to the conclusion that the maximum detection distance measurements should be performed with several operators and repetitions and that the measurements in air can not be considered an indicator of the performance in soil.

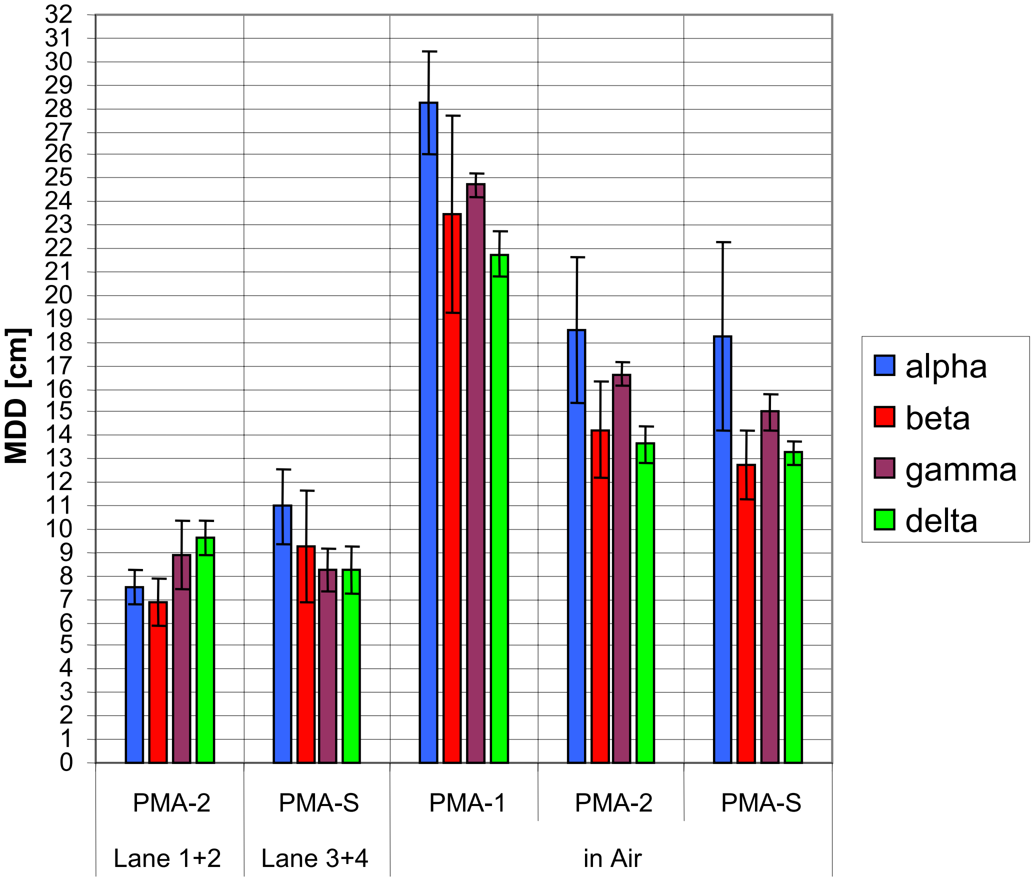


Figure 13: Maximum Detection Distance with the Corresponding Standard Deviations

In addition to the described experiments the trials in 2005 included the investigation of scanned voltage signals on the same test lanes as used for the blind trials. It is the aim of the authors to find a clear relationship between the maximum detection distance and the depth at 50% decrease of the POD from the signal response POD and an assessment of the human factor influence from the comparison of the signal response POD from the scanned results to the POD from blind trials. The results of these attempts will be published in a following paper.

Conclusions and Outlook

For the results of detection reliability field tests the embedding scenario in terms of soil type and cleanness and human factor treatment has to be set up with care and explicit consideration of the local field situation and working system. The characteristics of a 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 be determined from these basic curves by superposition, according to the local mine distribution.

Acknowledgment

The authors would like to thank the German Ministry for Foreign Affairs (Lt Col. Detlef Schröder) and the German Ministry of Defence (Lt Col. Manfred Merk) as well as the BAM's presidential Board for sponsorship and support of the project. We are indebted about the fruitful partnership with the CROMAC testing facilities in Benkovac and the WTD 52 in Oberjettenberg as well as We thank for the hands on support of all colleagues from WTD52, BAM and CROMAC. Especially we thank all the soldiers of the German army and the deminers of Croatia for their brave work.

References

1. Ch. 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)
2. P.-Th. Wilrich, “Statistical Design of Demining Experiments and Analysis by Logistic Regression*”, ITEP Workshop Reliability Tests for Demining,* December, 16-17, 2003 Federal Institute for Materials Research and Testing (BAM) Berlin, 16-17 Dec 2003, Workshop CD or ITEP-ws
3. Ch. Mueller, M. Scharmach, M. Gaal, D. Guelle A. M. Lewis and A. J. Sieber “Performance demonstration for humanitarian demining” *Materialpruefung* Jahrg. 45, Vol. 11-12 pp. 504-512, 2003
4. Ch. Mueller, M. Scharmach, M. Gaal, D. Guelle A. M. Lewis and 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 Sep 2003
5. M. Gaal, Ch. Mueller, K. Osterloh, M. Scharmach, S. Baer, U. Ewert, A. M. Lewis, D. Guelle “Metal detector test trials in Germany and Croatia – ITEP Project 2.1.1.2”, *MATEST 2003 NDT Achievements and Challenges*, Brijuni-Pula, Croatia 28-30 September 2003
6. Dieter Guelle and Martina Scharmach, ROC-Diagrams for the different lanes, 0312161445-Guelle-14-45.pps on Workshop CD, ITEP-workshop, “Reliability Tests for Demining”, 16.-17. December, 2003, Berlin, Germany
7. M. Gaal, S. Baer, T. J. Bloodworth, D. Guelle, A. M. Lewis, Ch. Mueller and M. Scharmach, Optimizing Detector trials for Humanitarian Demining, SPIE, Defense & Security Symposium, April 12-16, 2004, Conference Proccedingds

Contact:

Name:	Christina Müller
Phone:	+49 30 8104 1833
Fax:	+49 30 8104 1836
E-mail:	Christina.Mueller@bam.de
Address:	BAM – VIII.33 Unter den Eichen 87 12205 Berlin Germany