Proposals for Performance Demonstration and Modular Reliability Assessment for Humanitarian Demining

Christina Müller
Martina Scharmach
BAM Berlin
Germany
christina.muller@bam.de
martina.scharmach@bam.de

Mate Gaal
BAM Berlin/
University of Zagreb
Croatia
mate.gaal@bam.de

Dieter Guelle
ITEP Secretariat Ispra
Italy
Dieter.guelle@jrc.de

Adam Lewis,
Alois Sieber
JRC Ispra
Italy
Adam.lewis@jrc.de
alois.sieber@jrc.de

Abstract

For safe and reliable demining it is necessary to determine the actual true performance of mine searching equipment in detecting mines. The subject to detect a hidden subject by penetrating physical interaction with the target is similar to that of non-destructive testing where it is looked for hidden cracks etc. in material via waves and rays. The non destructive testing profession is now about 100 years old and developed some procedures to check reliability of testing. Those principles like the performance demonstration where the successful detections are statistically evaluated against false calls rates and their implementation in an industrial standard (ASME section XI appendix VIII) are analysed. A first adoption to demining was accomplished in the prescription for blind trials in the CEN workshop agreement CEN BT 126 CWA07 for test and evaluation of metal detectors. A number of blind trials were accomplished within an ITEP project to learn about the necessary statistical layout of those trials to achieve true, reproducible and repeatable results to give guidance to selection and improvement of metal detectors. The special focus in these investigations was on the influence of the human factor due to the degree of experience of the operators and the infulence of uncooperative soil. The correlation of the results of the physical parameter measurement and the statistical results is analysed in a first attempt.

1. Introduction

The aim of the paper is to show attempts to determine the reliability of mine searching processes quantitatively. Apart from the motivation of performance measurement the capability to handle the real existing remaining risk is of imortance. At the second European American workshop about NDE reliability, 1999 in Boulder, USA, the following definition of NDE reliability was elaborated /1/: NDE reliability is the degree that an NDT system is capable of achieving its purpose regarding detection, characterization and false calls. Where the NDE system consists of the procedures, equipment and personnel that are used in performing NDE inspection.

In the CEN BT 126 CW07 group a transformation to mine detection was proposed: Detection reliability is 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. For the success of the detection process the mine detection system has to be considered as composed of the detector, the procedure and the human being that are used in performing mine searching under specified operational and environmental conditions. A compact view yields the reliabilty formula (Fig. 1a, b). The expression defines a total reliability R, which consists of: an intrinsic capability IC describing the physics and basic capability of the devices, factors of industrial application such as special evironmental conditions in the field, AP, and finally the human factors HF. It is of high importance to emphasize that the reliability in thefield is always composed of all the three factors.

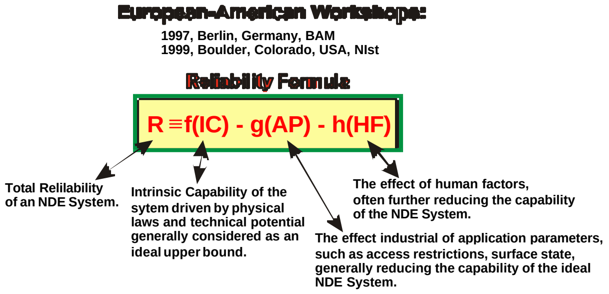

Figure 1: a) The Reliabilty Formula

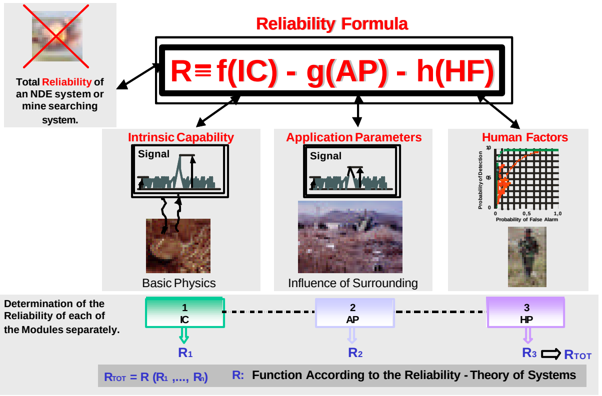

Figure 1: b): Modular Reliability Model for Demining

To measure the total R the so called integral approach has to be applied:

Integral approach: ROC – Receiver operating characteristic : Probability of Detection (POD) versus False Call Rate

The Receiver Operating Characteristic (ROC) /7, 8/ is deviated from the general theory of signal detection and widely used since the second world war in fields of evaluation of diagnostic systems like radar techniques, test of human perception and in medical diagnosis and since the eighties also in NDE. The general four possible situations in NDT (Nondestructive Testing) diagnosis are presented in figure 2. For both "true situations", defect/mine present or no defect/mine present, we have the possibility to recognize the truth (TP, TN) or to miss the truth with a false indication (FN, FP).

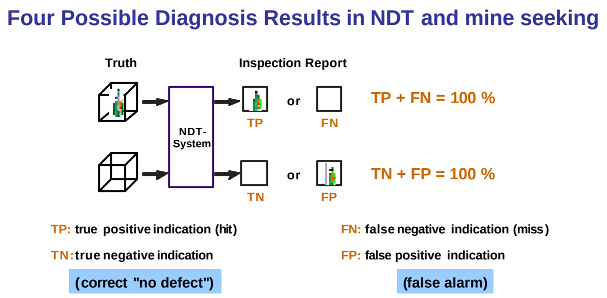

Figure 2: The Principles of ROC (Reciver Operating Characteristic) The Possible Diagnosis Results in NDT

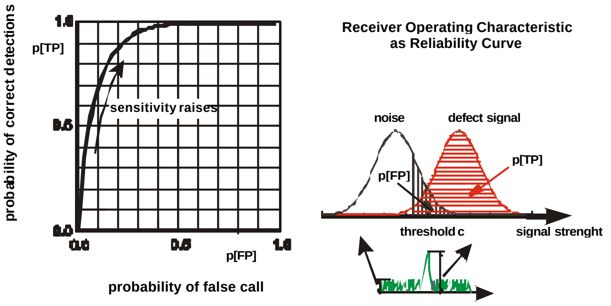

Figure 3: Characteristic of one NDT-System by an ROC curve

The idea of the ROC method is to characterise the accuracy of an inspection system by evaluating the true positive detection rate versus the false positive detection rate for a set of possible decision criteria or recording levels in the language of NDT which represents a varying sensitivity. Following theROC-curve in figure 3 from the lower left corner to the upper right - the sensitivity of the system raises. So- in the lower part of the curve the highest signals (correct indications) are included and only a small amount of noise (false calls). In the higher part more and more all of the defects are taken into account but also a greater amount of false calls has to be paid as price. The underlying mathematical model in terms of the two Gaussian signal distribution curves for the defect signals and the noise respectively are shown on the right hand side. Fig. 4 shows the comparison of systems with different reliability. For the fictive systems the performance of the system increases from curve 1 to curve 7.

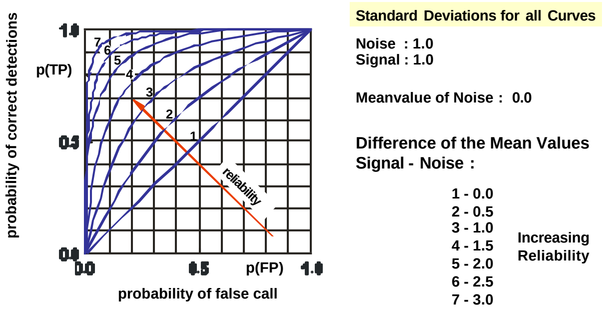

Figure 4: Comparison of Different NDT-Systems

We emphasize here the importance to look to the dependence of performance or reliability from the system sensitivity especially when comparing different systems.

When we consider just the actual applied sensitivity recommended by the manufacturer we receive just one operational ROC point for one system. An illustration is given for the results of the different devices from the IPPTC /11/ test trials for cooperative soil (green points) and uncooperative soil (red) points in an ROC diagram in figure 5. It is easilly seen how the detection performance is dropped down by uncooperative soil in both lowering the detection probability and raising the false alarms.

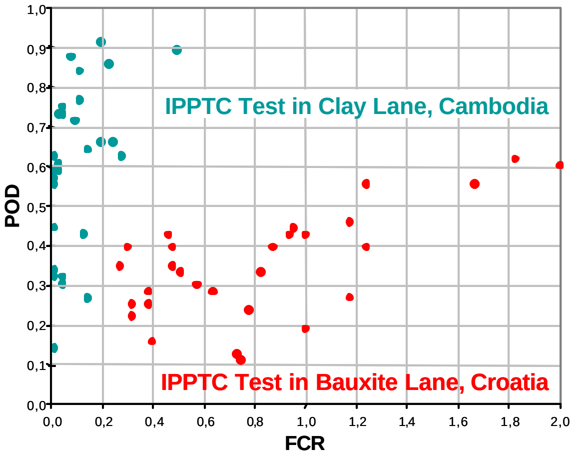

Figure 5: Results of the IPPTC Trials in a ROC diagram

A first 'hands on' adoption of the ASME code to demining was accomplished in the prescription for blind trials in the CEN workshop agreement CEN BT 126 CWA07 for test and evaluation of metal detectors to make the results of field trials comparable. A number of blind trials were accomplished within an ITEP project – ITEP 2.1.1.2 'Reliability Model for Test and Evaluation of Metal Detectors' to learn about the necessary statistical layout of those trials to achieve true, reproducible and repeatable results to give guidance to selection and improvement of metal detectors. To set up a first relation to the modular model the correlation of the results of the physical parameter measurement and the statistical results is analysed in a first attempt.

The parameter measurements were carried out to a great extent at the JRC Ispra a long the lines of /10, 12/ and will be published in detail within a JRC report. As described in /10, 12/ maximum detection distances were measured for different steel balls and ITOP targets in air and in soil. From the in air values we can learn a rough forecast of detection performance in cooperative soil while the specific condtions in uncooperative soil might even reverse those results. The maximum detection depths in specific soils of different mines and the applied detectors were carried out also at the testfield location. As well as the measuremt of the magnetic susseptibility using a Bartington device as described in /9/ and the degree of uncooperativeness by setting a detector without soil-compensation to a fixed sensitivity and measuring how close it can be brought to the soil surface before the alarm sounds: called the soil reference height. This measurement was made on all the soils using a Schiebel AN19 Mod 7 detector, adjusted so that it could just detect its calibration pin at 10cm distance from the baseline mark.

Two sets of blind field trials using realistic mines were carried out in accordance with the description in CWA07 draft 10, section 8.5 detection reliability tests.

The first set of trials took place on the test fields of the German army in Oberjettenberg at WTD52 with unexperienced young soldiers as opeartors during two weeks in May 2003. Each week started with two days training on the 4 detectors for 4 operators folllowed by four days test:

Design of the test trials in Oberjettenberg:

- 4 fields - 4 types of soil
- 8 devices - 4 types, 2 specimens
- 2 sensitivities
- 8 unexperienced deminers
- 2 repetitions (same fields)

The second set of trials took place near to actually mine afected regoins in Croatia with experienced deminers in realistic soils.

Design of the test trials in Benkovac:

- 8 fields - 3 types of soil
- 8 devices - 4 types, 2 specimens
- 2 sensitivities
- 8 experienced deminers
- 2 repetitions (different fields)

In spite the importance of different sensitivity setting was emphasized and also aplied we will present in this first evaluation only the high sensitivity setting – so we are comparable to the IPPTC results. The focus of the evaluation is here how the human factor influences the performance of detectors for diffeent mines in different soils.

Evaluation parameters for In-field tests:

• Probability of Detection:

POD = Number of Mines detected devided by the number of present mines, were a detection is defined as an indication which falls in the area of the halo radius of a mine.

• False Call Rate: FCR

This is simply the number of indications per square meter which fall outside of the halo circles.

Since the aim of the trials was to define the statistical layout of test design it was decided to treat the company names of detector manufacturers anonymously and name them U, X, Y, Z. The table of general device types applied is given in figure 6.

Detector	Search Head Coil	Mode	EM Wave
U	single	static and dynamic	time domain
X	single	dynamic	time domain
Y	double-D	static	frequency domain
Z	double-D	static	frequency domain

Figure 6: Detectors

2. Test Trials in Oberjettenberg

All test lanes were 20 m long and 1m in width containing about 30 tragets which are listed in detail in the table of figure 7. Test lane 1, the uncooperative soil) was simulated using garden soil plus a 2 cm deep layer of steel owen slag

- Magnetic susceptibility measurement with Bartington yields 200 ... 330 * 10-5 in SI units
- The empirical distance measurement with the Schiebel detector according to Dieter Guelle yields distances of 3 cm ...7,5 cm

Test lane 2-4, the essentially cooperative soils, contained two types of gravel and clay with Bartington values 0.3...7,6 * 10-5 SI units.

	type	halo r.	depth	metal part diameter (mm)	**halo radius (cm)**	halo surface (m²)	ease of detection with a MD
APM	PPM-2	16	2	120	**16**	0,08	easy
	PPM-2	16	3				
	PMN	16	2	112	**16**	0,08	easy
	PMN	16	3				
	PMN	16	15				
	Maus	14	1	89	**14**	0,06	easy
	Maus	14	2				
	Maus	14	3				
	SchAMi DM 31	15	5	102	**15**	0,07	easy
	SchAMi DM 31	15	6				
AVM	TM-46	25	5	305	**25**	0,20	easy
	TM-46		6				
	PT-Mi-Ba-III	11	5	10	**11**	0,04	very difficul
	PT-Mi-Ba-III	11	6				
	PT-Mi-Ba-III	11	7				
	TM-62 P2		6	125	**16**	0,08	difficult
	TM-62 P2	16	5				
	TM-62 P3	16	5	125	**16**	0,08	difficult
	TM-62 P3	16	6				
	TM-62 M	26	5	320	**26**	0,21	easy
	TM-62 M	26	6				
	TM-62 M	26	15				
	TM-62 M	26	16				
ITOP	small C0	10	5	< 10	**10**	0,04	difficult
	small E0	10	10				
	small G0	10	5				
	small I0	10	10				
	small K0	10	5				
	small K0	10	10				
	large G0	10	10				
	large G0	10	5				
	large I0	10	5				
	large K0	10	10				
balls	100Cr6 - 16mm	11	20	16	**11**	0,04	difficult
	O0	10	20				

Figure 7: Mines in Oberjettenberg

Operator	Preferred detector	Years of experience in mine action
A	Y	0
B	Y	0
C	Y	0
D	Y	0
E	Y and Z	0
F	Y and Z	0
G	Y and Z	0
H	Y	0

Figure 8: Operator properties, Oberjettenberg Trials, May 2003

The table in figure 8 shows the list of the unexpierenced operators with their preferred detecor (result of questionaire). Fig. 9 shows the young soldiers during the training in Oberjettenberg on the little calibration fields which were prepared for detector set up for each sol type.

Figure 9: Unexperienced Operators during Training

For an overall impression the meanvalues over all devices, operators and repetions has been taken for cooperative and uncooperative soils respectively and plotted in figure 10 together with the corresponding standard deviations. We see a performance just below 70% of detection probability with a lower value for uncooperative soil and an increasing false call rate. Looking to figure 11, the POD distribution in correspondence to the target depths, we see that this values are mainly due to deeper targets which are more difficult to detect – as well known – with exception to the deeper big anti tank mines.

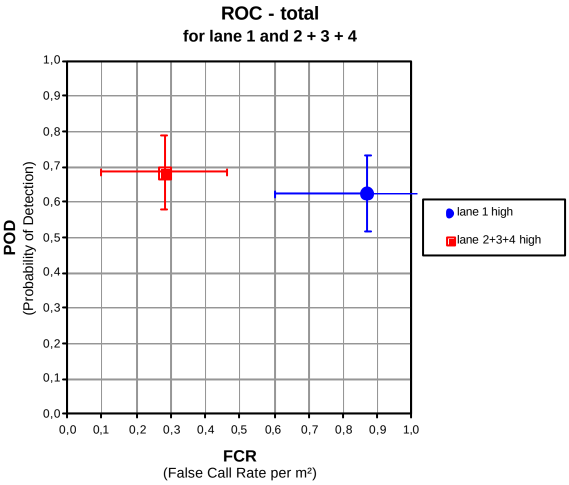

Figure 10: ROC-total

In figure 12 upper and lower part we see the mean value rsults for each of the detector types for the first and second Repetition seperately. In spite ths is not exactly the same result we see the pattern of their relative postions (Z and X in upper left close together with reasonable detection probabilities and low false call rates, the Y in the upper right corner paying for the higher detection with high false calls and U below them in the middle) repeated. So we can conclude the repeatability of the test result. Even more the differences of detection rates between the devices are statistically highly significant/13/.

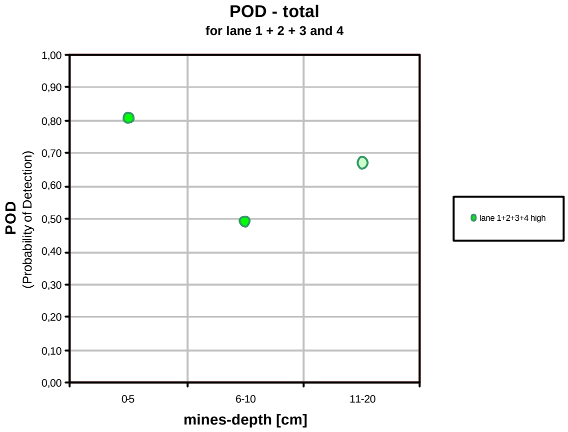

Figure 11: POD-total

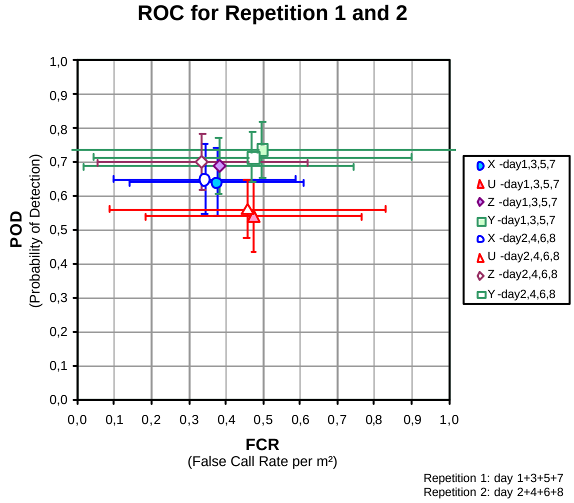

Figure 12: ROC for Repetition 1 and 2

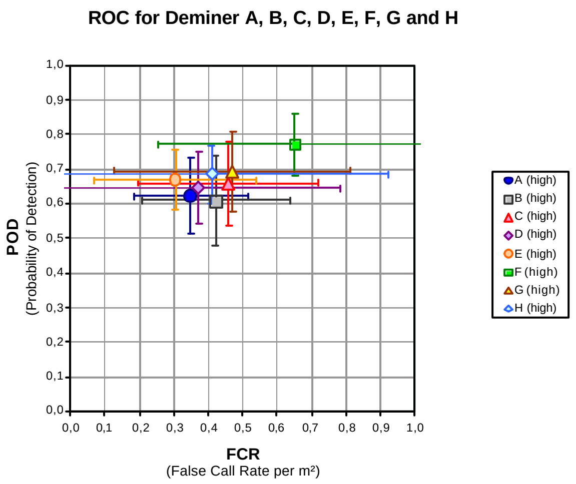

Figure 13: ROC for Deminers

Fig. 13 shows the mean ROC-points for each of the operators. The spread is between 80% and 60% POD and a false call scatter from almost zero to almost 1 false call per square meter. This high scatter in test results of unexperienced operators is well known from NDT /4, 8/.

3. Test Trials in Benkovac

	Susceptibility in 10^5 SI	Schiebel distance in cm
Lane 2 and 6 Neutral soil from Sisak (grey)	13 ± 2	0 ± 0
Lane 1 and 5 Uncooperative soil from Orbovac (Bauxite – **red**)	154 ± 13	18,8 ± 0,9
Lane 3, 4 and 7 Local soil from Benkovac (Calcocambisol) heavy uncooperative and heterogeneous, change after digging (**red**)	190 ± 36	19,7 ± 2,5

Figure 14: Properties of the 8 Testlanes, length 30 m, width 1 m

	type	halo r.	depth	metal part diameter (mm)	halo radius (cm)	halo surface (m²)	ease of detection with a MD
Testsamples	ball	11	10				
	F0	10	5				
	G0	10	5				
	K0	10	5				
APM:	PMA-1A	13	0	65	**13**	0,05	possible
	PMA-1A	13	5				
	PMA-1A	13	5				
	PMA-1A	13	5				
	PMA-1A	13	10				
	PMA-1A	13	13				
	PMA-1A	13	20				
	PMA-2	11	0	20	**11**	0,04	difficult
	PMA-2	11	5				
	PMA-2	11	5				
	PMA-2	11	5				
	PMA-2	11	10				
	PMA-2	11	13				
	PMA-2	11	20				
	PMA-3	10	0	5	**10**	0,03	difficult
	PMA-3	10	5				
	PMA-3	10	5				
	PMA-3	10	5				
	PMA-3	10	10				
	PMA-3	10	13				
	PMA-3	10	20				
	PROM-1	14	0	75	**14**	0,06	possible
	PROM-1	14	0				
	PROM-1	14	5				
	PROM-1	14	5				
	PROM-1	14	5				
AVM:	TMA-3	22	10	230	**22**	0,15	possible
	TMA-4	23	10	250	**23**	0,17	possible or easy
	TMRP-6	19	10	170	**19**	0,11	easy
	TMM-1	26	10	325	**26**	0,21	very easy

Figure 15: The properties of the soils of the 8 testlanes in Benkovac and Mines in the Test Lanes in Benkovac

Operator	Currently active as deminer	Years of experience in mine action	Detector already used [1]	Preferred detector
A	yes	7	X, Y, Schiebel	X and Z
B	yes	7,5	X, Schiebel	X and Z
C	no	8	X, Y, Schiebel	Z
D	yes	2,5	X, Minelab	X and Z
E	no	10	X, Y, Schiebel	Z
F	no	12	X, Schiebel, Guartel	X and Z
G	no	7	U, X, Y, Schiebel, Guartel	Y and Z
H	no	10	Schiebel, (one-month training: U, X, Y, Fisher, GIAT, Guartel, LG Precision, Reutech, Schiebel, Whites Spectrum, P roscan, Minelab, Adams)	X

Figure 16: Operator Properties, Benkovac Trials, July 2003

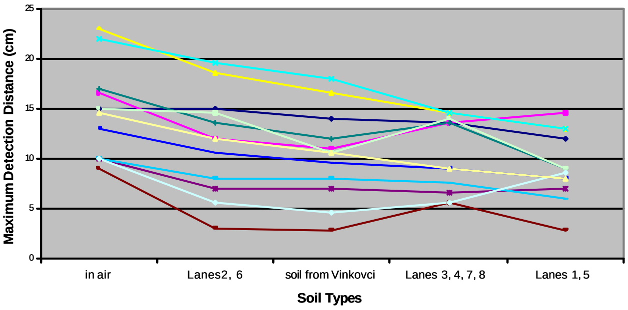

Figure 17: Detection Distances of PMA-1A, -2, -3 for Different Detectors

Are listed in the table in figure 14 in terms of the susceptibilty value measured with the Bartington and the Schiebel soil reference. We see, that the degree of uncooperativeness is much higher here than in Oberjettenberg. There is also an inverse proportionality seen between the soil reference and the susceptibility on the one side and the maximum depths of different smaller mines for all detectors for the corresponding soils as shown in figure 17 on the other side.

The figures 15 and 16 show the mine type and depth distribution and the properties of the experienced operators, respectively. Also concerning mine types (lower metal content of PMA -1, 2, 3 in deeper depths) distribution the Benkovac situation was a challenge for the detection performance. This “being at the limit” concerning the physics, the application parameters and the human capability is reflected in the overall results for the coopperative (lane 2, 6) and uncooperative (lane 1, 5 and 3, 4, 7, 8) test fields which is below 70% of POD for cooperative soil and even below 60% for uncooperative soil shown in figure 18.

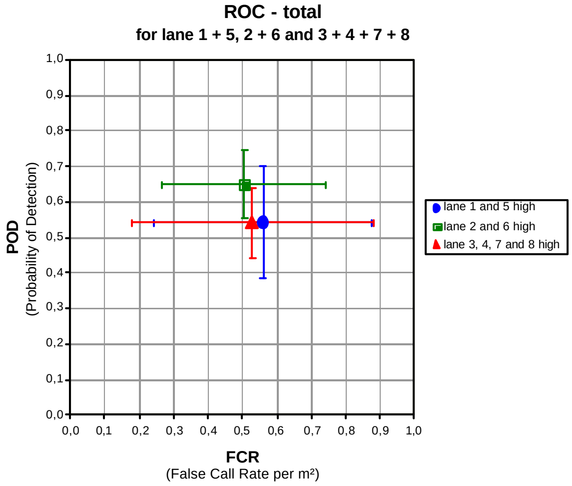

Figure 18: ROC-total

The depths distribution of the POD for the different soils is shown in figure 19. The statistical results reflect of course in the average the properties of the physical parameters of maximum detection depths in figure 17. For all the soils but especially for the uncooperatives the detection depths greater than 10cm are critical. The POD for the individual detectors is shown in figure 20 as an illustration. For the purpose of practical application those plots should be craeted for each mine type in each soil to select the appropriate detector for the region.

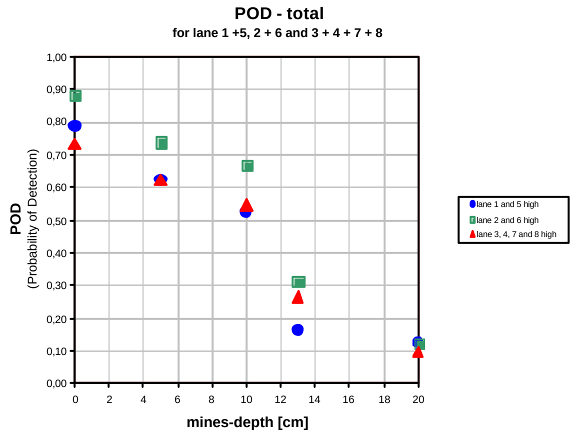

Figure 19: POD-total

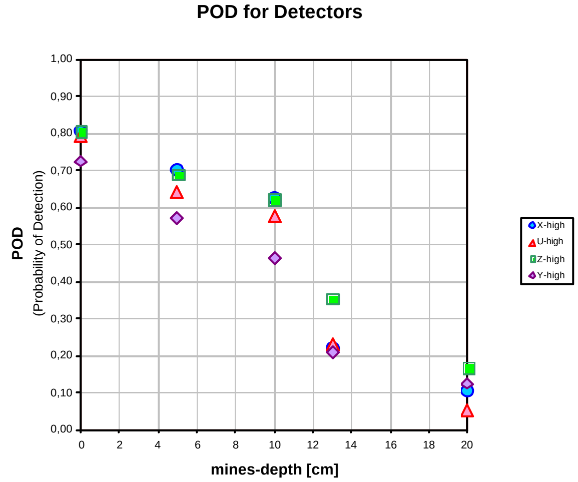

Figure 20: POD for Detectors

Since the actual digging depth of almost all mines in Croatia is beow or equal 5cm we selected the results up to this depth to show a typical every day situation: In figure 21 the corresponding meanvalue for the devices are shown. We see very roughly the pattern from figure 12 repeated but the Y decreased in performance of detection and raised the false call rate due to the difficulty in soil compensation while the U gained in performance the more difficult the soil situation is. In Figure 22 are the results up to 5 cm for all mines and all soils for each of the operators shown. The operators A, B, D – the ones which are currently actually active deminers – sho up the highest performace with ROC points in the upper left corner. Revealing this way our testing method being suited also for performance demonstration tests for operators.

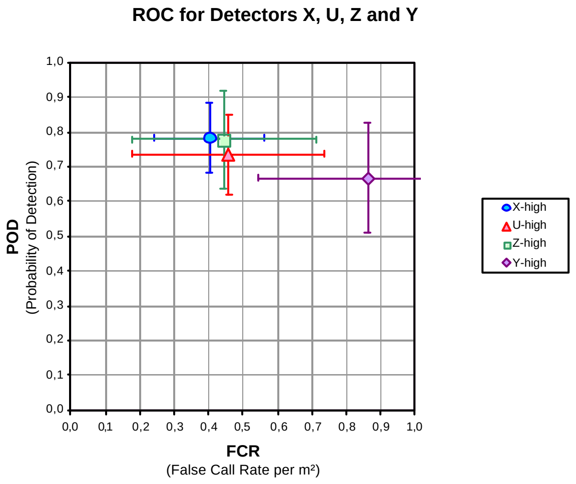

Figure 21: ROC for Detectors for 0-5 cm mines-depth

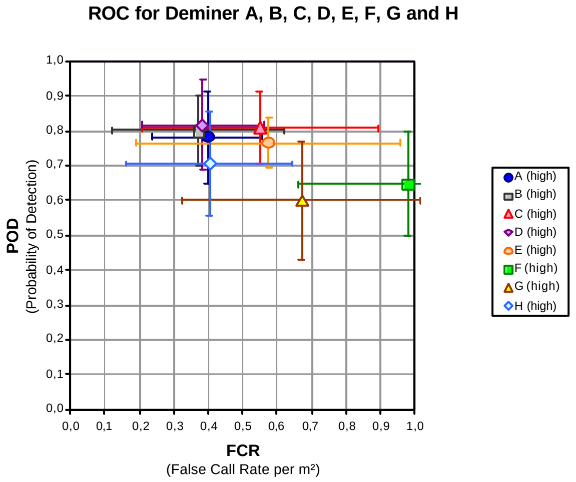

Figure 22: ROC for the Deminers for 0-5 cm mines-depth

Figure 23: Experienced Deminer during the test in Croatia

4. Summary and Conclusions

The test results of the blind field tests are highly influenced not only by the intrinsic physical capability of the detectors but also by the specific mine soil combination and especially the human factor.

But with the proposed statistical layout the difference of meanvalues between performance of devices and operators are statistical highly significant (F-test) and can be used for test and evaluation and performance assessment. One conclusion is that about 30 mines pers test lane are well suited. The final statistical layout will be defined after the third part of the trials in Oberjettenberg in October.

The unexperienced operators act more or less identical with each detector but have a high random scatter and miss the ”feeling” for the mine in the field.

The experienced operators have a ”feeling” for the mine in the field but are very specialised to the detector they used for longer time. The currently active deminers have a significant higher performance especially concerning the false call rates.

Both groups require a longer training period with each detector and a trial & error phase in practising blind trails with immediate feedback for response to specific mines in specific depth in specific soil.

The uncooperative soils were a challenge for the compensation capability of all detectors.

The properties of the local soil changed due to digging activities:

- the electromagnetic properties were not only uncooperative but also heterogeneous yielding additional numbers of false calls.

The maximum detection distances for the lower metal content mines:

- PMA-1A, PMA-2, PMA-3 were in all soils much lower than 20 cm. The legal clearance depth in Croatia is 20 cm. We included in the assembly a number of 20 cm deep mines of this type which were in almost all cases not found especially in uncooperative soil,
- the 20 cm clearance depth for these mines is a challenge for the physical capability of all the detectors. In spite the usual depth of these mines is below 20 cm they might occur in this depth due to aging, vegetation, wheather influences or as a result of uncomplete clearance by mechanical machines.

The capability of metal detectors should by analysed systematically within the modular approach concerning the influence of basic physical capability, application factors and human factors including simulations and experiments in the lab and in the field.

Realistic results will be obtained only in considering the whole PROCESS of demining including area reduction, searching, digging and neutralisation. With the current results we gave a first ‘taste’ only how the final reliability model of metal detectors should look like.

Acknowledgment

The authors like to thank the German Ministry for Foreign Affairs (Col. Detlef Schröder) and the German Ministry of Defense (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 Beknovac and the WTD52 in Oberjettenberg as well as the Humantitarian Security Unit of the JRC Ispra especially Tom Bloodworth.. We thank for the hands on support of all colleagues from WTD52, BAM and CROMAC and Ute Böttger (DLR) for scientific discussion directly in the field. Especially we thank all the soldiers of the German army and the deminers of Croatia for their brave work.

References

(1) C. Andrew, “Mathematical modelling of electro magnetic fields”, Transactions of Magnetism, vol. 37. No 5, 1998.

(2) G. S. Kino, Acoustic Waves: Devices, Imaging and Analog Signal Processing. Englewood Cliffs, Prentice Hall, 1987 [book]

(3) ASME Boiler and Pressure Vessel Code, Section XI “Rules for Inservice Inspection of Nuclear Power Plant Components.“ American Society of Mechanical Engineering. 1992 Edition

(4) Nockemann, C. 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)

(5) Non-Destructive Evaluation System Reliability Assessment, MIL STD 1823 (U.S. Department of Defense)

(6) Mueller, C; Fritz, T; Tillack, G.-R.; Bellon, C; Scharmach, M ,Theory and Application of the Modular Approach to NDT Reliability, Materials Evaluation, 2001, vol. 59, no. 7, pp. 871 874

(7) Barlow, R. E., Fussel, J. B., Singpurvalla, N. D., eds. (1975); ‘Reliability and Fault Tree Analysis, SIAM, Philadelphia‘.

(8) Proceedings of the “European-American Workshop Determination of Reliability and Validation Methods on NDE”, Berlin, Germany, June 18 20, 1997, ISBN 3931381 18 8

(9) F. Borry, D. Guelle, A. Lewis: “Soil Characterization for Evaluation of Metal Detector Performance”, submitted to EUDEM-SCOT-2003, Vreije Universeteit Brussel, Brussels, Belgium

(10) T.J. Bloodworth, A.M.Lewis: “Quantifying the Detection Capability of Metal Detectors Using Metal Spheres”, submitted to EUDEM-SCOT-2003, paper 144,

(11) Y. Das, J.T. Dean, D. Lewis, J.H.H. Roosenboom, G. Zahaczewsky, eds.: “International Pilot Project for Technology Co-operation IPPTC”, Final Report, 2001

(12) T.J. Bloodworth, A. Sieber: “Standardised Test and Evaluation of Metal Detectors”, submitted to EUDEM-SCOT-2003, paper 145,

(13) P.-Th. Wilrich, ‘Results of an F-test’, private communication, to be publishes in the ITEP report 2.1.1.2.