SOUTH-EAST EUROPE
INTERIM REPORT FIELD TRIAL CROATIA

(Continuation of the ITEP-Project Systematic Test and Evaluation of Metal Detectors - STEMD)

25 September – 18 October 2006

By Dieter Guelle, Mate Gaal, Marija Bertovic, Christina Mueller, Martina Scharmach, Mato Pavlovic

Berlin, March 2007

Content & Annexes

List of Tables

List of Figures

Table 1: *Abbreviations*

BT	Technical Board (body of the CEN)
CCMAT	Canadian Centre for Mine Action Technologies
CEN	Comité Européen de Normalisation, European Committee for Standardization
COTS	commercial off the shelf
CWA	CEN Workshop Agreement
CW	continuous wave
EC	European Commission
FAR	false alarm rate
FFE	free from explosive
FRY	Federal Republic Yugoslavia
GC	Ground Compensation
HD	Humanitarian Demining
IMAS	International Mine Action Standards
IPPTC	International Pilot Project for Technology Co-operation
JRC	Joint Research Centre
MA	Mine Action
MAC	Mine Action Centre, (national) organ responsible for MA in a country
MDD	maximum detection distance
NDT	Non destructive testing
NGO	Non-governmental organisation
POD	Probability of detection
UN MAC	United Nations' MAC
UN MAPA	UN Mine Action Programme for Afghanistan
ROC	Receiver Operating Characteristic
UXO	Unexploded Ordnance Explosive ordnance that has been primed, fused, armed, or otherwise prepared for use and used in an armed conflict. It may have been fired, dropped, launched, or projected and shoul have exploded but failed to do so.
WG 126	Working Group 126 of CEN (dealing with HD)

1. Introduction

The report describes the field trial carried out by BAM (Federal Institute for Materials Research and Testing, Berlin, Germany), in Benkovac, Croatia. The Centre for Testing Development and Training of the Croatian Mine Action Centre (HCR-CTRO) supported the trial preparation and execution with personnel, the test facility, and logistics. The trial comprised a reliability test, maximum detection distance measurements, pinpointing and the establishment of the footprint. This trial belongs to a series of lab and field tests of the ITEP (**I**nternational **T**est and **E**valuation **P**rogram) project 2.1.2.3 Systematic Test and Evaluation of Metal Detectors (STEMD), a campaign to assess the capabilities of the available commercial-off-the-shelf (COTS) metal detectors. In the STEMD campaign, a two-stage approach had been followed. The first stage was testing and evaluation under laboratory conditions, performed by the EC Joint Research Centre at their test facilities in Ispra. The second stage were in-field tests in several mine-affected regions (South East Europe, Southern Africa, South East Asia), under realistic conditions using the CEN Workshop Agreement CWA 14747:2003 protocol. Field trials had been carried out in Laos in autumn 2004. They focused on the assessment of metal detectors with large search heads for UXO detection in comparison with normal metal detectors. In spring 2005 the Mozambican trial focused on the influence of soil on metal detectors. The results of those trials are laid out in reports available at the ITEP and Joint Research Centre websites (www.itep.ws and http://serac.jrc.it/ , the latter look Archive TETHUD).

2. Background

Metal detectors are the main tool for detecting landmines in humanitarian demining (HD) and a correct understanding of their capabilities and limitations is of great importance to people working in the field. Since the end of the eighties, the start of first humanitarian mine clearance operations in Afghanistan, the metal detector is still the only trusted sensor used in humanitarian demining. Since World War 2 the sensitivity and the construction of metal detectors has been changed and new features have been added, while the general physical principle, the electromagnetic induction, remained the same. When mechanically supported clearance operations, the use of dogs or other sensors took place – still the human being has to move to the suspected area with a metal detector to find and neutralise the danger. The assessment of this sensor system – ***the metal detector, the human being and other environmental factors*** – are to be investigated to give an answer, which detector is appropriate to be used under which circumstances. Clearance organisations (deminer) care that their detectors are affordable and long-lasting, that they can use them for six hours without straining their arms or changing batteries and above all, that they will find low-metal content mines reliably enough that they can hand over land confident that all mines have really been removed. But the detector is just a part of the system; the human being with its own capabilities and the environment are factors that may influence each other and in most cases reduce the intrinsic capabilities of the metal detector. In this report, the main factors influencing the performance of the system are investigated.

The United Nations Mine Action Centres began to organise metal detector trials from 1997. It was soon recognised that resources would be wasted if testing was duplicated by every interested organisation. In June 2000 representatives of six donor governments and the EC [1] signed the Memorandum of Understanding of the International Test and Evaluation Program (ITEP), with the remit of conducting joint test and evaluation projects and exchanging the results.

The results from earlier tests were not comparable due to different approaches and focus on specific requirements. Simultaneously with the aspiration to unite the testing efforts, steps were undertaken to establish internationally recognised rules that may be used for the test and evaluation of humanitarian demining equipment. The CEN (Comité Européen de Normalisation, European organisation for

---

[1] ITEP currently has seven partners: Belgium, Canada, Germany, The Netherlands, Sweden, UK, USA. Discussions are ongoing for enlargement and for the establishment of cooperation agreements with other governments and organisations.

standardisation) received the mandate from the EC (European Commission) and ITEP to establish standard-like rules for the main tool in HD, the metal detector. Methods developed in trials were standardised in CEN Workshop Agreement CWA 14747[2] in 2003 and incorporated as a normative reference in the International Mine Action Standards (IMAS). The CWA is comprehensive, defining tests for all factors relevant to the user, www.itep.ws search standards (see summary ANNEX 1). Of key importance are measurements of detector sensitivity to targets under various conditions and the evaluation of the overall detector performance. Taken together, the information collected in all the trials demonstrates the changes and development of techniques, capabilities, and design of detectors during this relatively short period. A summary of different trials can be found in– see ANNEX 2 (Table of the STEMD and other trials)

During the process of establishing and publishing the CWA for metal detectors, the preparation for the STEMD project started. The JRC took over the lead for this project and purchased the available metal detectors (COTS) to include them into the lab tests and field trials. This project was also used for the implementation of the CWA. One of the main aims of the STEMD project was to give the HD community an overview of the commercially available detector fleet and to keep that information updated. Unfortunately the foreseen reliability trial[3] in South East Europe in autumn 2005 could not be carried out by the JRC (SERAC Unit) because of administrative reasons. Due to similar reasons the detectors were not made available to BAM, which took over the organisation and the execution of the trial in Croatia in 2006. This report includes the missing STEMD reliability trial of the current detector fleet.

Current mine situation in South-East Europe
At a meeting of the South-Eastern Europe Mine Action Coordination Council (SEEMACC) on 13 October 2005, CROMAC's director suggested that Croatia, Bosnia and Herzegovina, and Serbia and Montenegro should jointly identify the mine situation on their common borders and send their demining priorities to donors [3]

At the end of 2005, Bosnia and Herzegovina claimed that more than 2,146 square kilometres (4.14 percent of the country territory) was suspected to be contaminated [2]. The 2005-2009 mine action strategy planned to clear 21 square kilometres of "priority 1" area in highly impacted communities, to release 53 square kilometres through technical survey, to conduct general survey on 510 square kilometres, and to carry out systematic survey on 716 square kilometres of land.

Two years after the end of the armed conflict, in 1997, 23% of the Croatian territory was considered mine suspected. By the year 2005 mine suspected areas were reduced to 2.1% of the country's area, which is 1,147 square kilometres [2] . About 135 square kilometres of that area are known to be mined. In 2006, Croatia planned to spend approximately $50 million on clearance and technical survey, releasing a total of 28 square kilometres of land [1].

In Serbia and in Montenegro, two areas of border territory remain contaminated by landmines and explosive remnants of war (ERW) [2]. Much larger contamination was caused by cluster bomblets and large aerial bombs, especially in Kosovo. Cluster bomblets remained in Serbia affect approximately 24 square kilometres, while the mined areas occupy 4.3 square kilometres. In Montenegro, an area of 1.5 square kilometres is affected by cluster bomblets, while a smaller area is still mined.

3. Purpose and objectives of the trial

The **purpose** of the trial was to:

- Assess recent commercial off-the-shelf detectors believed to be appropriate to South East Europe (SEE) and for humanitarian demining generally, and

---

[2] Today standard like CEN workshop agreements exist for mechanical mine clearance, personal protective equipment, other are on their way.

[3] Blind or reliability trials are complex trials for the assessment of the sensor in connection with the user/operator and other influencing environmental factors including the working rules and conditions that are to be followed.

- Make the data available for the humanitarian demining community.

The objectives of the trial:

- Compare the reliability test results and other detector performance data in different types of soils.
- Create a lane layout usable for metal detector tests and dual-sensor tests.
- Implement new knowledge concerning the reliability trials and measurement techniques
- Psychological approach to the investigation of the human factor and its influence on detector performance
  - Collect useful information on how to improve the effect of human factor on the testing process and the safety of the end users.
- Measure sensitivity and accuracy of detectors to a typical local target of interest
- Train local staff in selected issues of the CWA.

4. Trial preparation

The participating detectors were, in alphabetic order:

- AKA (Moscow, Russia) – Condor 7252*
- AKA (Moscow, Russia) –Vector 7260*
- CEIA S.p.A. – Mil D1
- Inst. Dr. Foerster GmbH and Co. KG - MINEX 2FD 4.530*
- Minelab Pty. Ltd – F3
- Minelab Pty. Ltd – F1A4
- Schiebel Elektronische Geräte GmbH – ATMID
- Vallon GmbH – VMC1*
- Vallon GmbH – VMH3CS*

The detector models marked with an asterisk (*) were new in comparison to the STEMD trial performed in Mozambique. Unfortunately, the JRC could not make its detectors available to BAM as it was expected from earlier contacts. The BAM trial management was forced to change the trial matrix according to new circumstances due to a very short notice. The number of participating metal detector types was reduced from 12 to 9 (two types of Ebinger detectors and one type of detector produced by Beijing Geological Instrument Factory, China, had to be omitted) and the overall number of detectors from 22 to 14. A quick reaction of the manufacturers made the trial possible: they sent the new detector models or they were brought in the hand luggage directly to the training.

In an optimized sensor system, such as detecting mines in humanitarian demining, there are three factors interacting among each other, together influencing on the results:1) intrinsic capability, in this case, of the metal detector; 2) application parameters (e.g. the influences of the environment, training, noise, time-pressure, SOP, etc.); and 3) the human factor. By human factor we consider all possible factors within the person and outside of the person which influence on his work performance.

Although human factor influences in humanitarian demining had been recognized, not enough has been done to quantify those factors and to find ways directly to influence on these factors. Most of the accidents occurring in demining are based upon human error or because of the violation of existing rules. Mine clearance operations and reliability trials are carried out by deminers and therewith highly affected by the human being. Most activities for improving the reliability of demining emphasized the improvement of safety, management and techniques.
It was for the first that a decision was made to assess the human factor in more detail. Questionnaires had been prepared for this purpose. The detector operators filled out a personality questionnaire (NEO PI-R), a concentration test (d2) and a specially designed questionnaire (see ANNEX 3) that consisted of questions about:

- personal data

- previous experience in demining and qualification
- experience with certain detectors and
- knowledge about the detectors.

The answers were used to establish groups of operators for an unbiased trial. The aim was to make all groups of detector operators similar in their age, demining experiences and experience with the specific detector they are going to use. We ensured that none of the operators had any experience with the desired detector. We tried to achieve that they were of different age and that they had different amount of experience in demining. In this way all groups of operators can be considered similar in their preconditions to be involved in the trial.
Due to the small number of operators it could not be completely ensured. The first criteria was that they had no previous experience with the detector they are supposed to use, then the amount of demining experience in general, and then their age. The assignment of operators in detector groups is presented in Table 2. The top line shows the 5 groups and the manufacturers belonging to one group. The left column lists the selection criteria and the next the subdivision of those criteria. The figures of the other columns give the number of operators belonging to one of the sub-criteria. With this approach it can be expected that the groups can deliver unbiased results.

Table 2. Placement of operators in detector groups based on experience, qualification and age

MANUFACTURER		**Group 1 Minelab**	**Group 2 Schiebel, Ceia**	**Group 3 Vallon**	**Group 4 AKA**	**Group 5 Foerster**	**Total**
EXPERIENCE	Less than a year	0	0	2	0	0	2
	1 to 6 years	2	3	2	1	3	11
	More than 6 years	2	1	0	3	1	7
QUALIFICATION[4]	Less than a month	1	0	1	0	1	3
	1 to 3 months	3	3	1	1	1	9
	More than 3 months	0	1	1	3	2	7
AGE	25-34	1	2	2	3	3	11
	35-44	3	1	0	1	1	6
	45-54	0	1	2	0	0	3

The training of the operators was carried out by the manufacturers. Four operators (one group) were trained to use two detectors, for each detector going through 2 days of training. An exception was the two Russian detectors, Kondor and Vektor, where the ITEP support from Belgium carried out the training, after the Russian manual had been translated into English. It has to be mentioned that the training to the Russian detectors was limited to the four preset factory programs of the detectors. The operators were additionally allowed to change/use the ground compensation, sensitivity settings, and power output of the search head. Further available features had not been included into the training and trial.

The basic training for the use of the Leica Total Station was carried out for the BAM personnel in Berlin and additionally together with the training for the supporting personnel from ITEP during the training of the operators.

The Friday in both training weeks was used for carrying out the pinpointing test with the taught detectors for better understanding of the factors influencing detector sensitivity; the sensitivity profile was established in different soil types to different targets. During that time the operators learned how to perform some tests described in the CWA. Their knowledge about metal detectors and their use was also tested. Surprisingly, the questions to the operators how to establish the safe sweeping advance to a target and factors influencing on detection ability where only partially known. The last part of the preparation was the instruction of the operators and supporting staff to the test procedures and rules to be followed.

---

[4] One operator in the Vallon group did not state his qualification

4.1. *Personnel and Resources*

- <u>Data Gathering Team:</u> D. Guelle (also trial team leader), M. Scharmach, M. Gaal, M. Pavlovic, – all BAM;
- <u>Human factor investigation</u>: M. Bertovic (BAM)
- <u>ITEP & GICHD support:</u> S. Dillien, M. Devroedt (Belgium Army), A. Schoolderman, F. de Wolf (both TNO NL)
- <u>Local personnel:</u>
  The main team of HCR-CTRO: including N. Pavkovic, I. Steker, 24 operators, supervisors,
- <u>Equipment:</u> 2 Leica Total Stations 1x lent, another station & WET-Sensor given by JRC

4.2. *Technical details of the known and foreseen detectors*

The data shown in Table 3 had been collected during the STEMD trial and are important to the user of the detectors.

Detectors	Manufacturers	**Principal Features**											
		Mode		Coil		Set-up						Software Access	Signal[4]
						Sensitivity Adjust-ment			Ground Compen-sation				
		Static	Dynamic	Single	Double-D	Fixed	Stepped	Continuous	Automatic	Manual	None	Yes/No	A/L/V
Condor 7252	AKA	x	x	x	-	-	x	-	x	-	-	-	A^5/V
Vector 7260	AKA	x	x	x	-	-	x	-	x	-	-	-	A^5/V
MIL-D1	CEIA	x	-	-	x	-	-	x	x	-	-	Y	A
EBEX® 421 GC	Ebinger	-	x	x	-	-	-	x	-	x	-	-	A
EBEX® 420HS	Ebinger	-	x	x	-	-	-	x	-	-	x	-	A
Minex 2FD 4.500	Foerster	x	-	-	x	-	3	-	x	-	-	-	A
Minex 2FD 4.510	Foerster	x	-	-	x	-	3	-	x	-	-	Y^3	A
Minex 2FD 4.530	Foerster	x	-	-	x	-	x	-	x	-	-	Y	A
MD8+	Guartel	-	x	-	x	-	3	-	-	-	x	-	A/L
F1A4	Minelab	-	x	x	-	x	-	-	x	-	-	-	A
F3	Minelab	x	-	x	-	x^1	-	-	x	-	-	-	A
ATMID™	Schiebel	-	x	x	-	-	-	x	x	-	-	-	A
M90	SHIRMT	-	x	x	-	-	-	x	-	-	x	-	A
VMC1	Vallon	x	x	x	-	-	-	x^2	x	-	-	Y	A/L/V
VMH3	Vallon	x^6	x	x	-	-	-	x^2	x	-	-	Y	A/L/V
VMH3CS	Vallon	x	x	x	-	-	-	x^2	x	-	-	Y	A/L/V
VMH3 (M)	Vallon	x^6	x	x	-	-	-	x^2	x	-	-	Y	A/L/V

[1] The sensitivity level is normally fixed but can be changed

[2] A large number of digitized levels are available, so the adjustment is effectively continuous.

[3] Will be made available for this model.

[4] The signal can be delivered to the operator as an audio signal (A), LED-display (L), or a vibration (V) of the handle.

[5] The audio signal can be changed for different targets. The visual indication allows a differentiation between magnetic and non-magnetic metals to a certain degree.

[6] The detectors may be updated to static mode since summer 2006 but had been tested without this mode.

Table 3: *Detector capabilities with immediate importance*

The "mode" may be either static, if the detector continues to emit a sound when it is held stationary over a metal target or dynamic if it must be moved over the target to signal.

Some detectors have the receive coil divided into two halves, the “double-D” design, which have a zero line in the middle where the signal stops or changes, to enhance pinpointing. The manner and capability of detection and pinpointing depend on both these factors. The deminer should be aware of them and they should be emphasised during training. A detector with a double-D coil behaves very differently from one with a simple circular coil and it is dangerous to confuse the two, because the shapes of the sensitive areas are different. Similarly, it is important to understand that a dynamic mode detector can be silent, even over a metal object, when it is not moved.

Sensitivity adjustment in some detectors is made with a switch having a limited number of positions, such as low, medium and high, with others it is made with a continuously variable knob and others have fixed sensitivity. Setting of the soil compensation, where the detector has it, is usually made by invoking an automatic procedure which allows the detector to “learn” the soil properties. The Ebinger 421GC is the only detector tested during the STEMD trials, which has a completely manual adjustment. The CEIA Mil D-1 makes its soil compensation adjustment automatically, but the manual-adjusted sensitivity setting affects it. The detailed procedures are different for each detector and it is important to follow precisely the instructions of the manufacturer for the model in question. Some of the most recent detectors allow the user access to the software via a communications port, for example it may be possible to download updates from the manufacturer, or make special changes to adapt the detector to particular conditions on the operational site. All detectors have an audio indication when metal is detected and this is generally considered superior to visual indication to avoid distracting the operator from looking at the ground. The VMH3 and MD8+ do provide also visual indication by LEDs on the handle. Vallon have also recently introduced a vibrator in the handle as a tertiary indication.

5. Methodology and procedures of the trial

5.1. Selection of CWA tests

In this trial the focus was on the reliability of detection as described in the CWA test 8.5. That test is called reliability test and it is a blind test, meaning that the operator of the detector does not know where the targets are and how many targets are in the search area. The data for analysing and establishing ROC (receiver operating characteristics) and POD (probability of detection) curves were collected during blind trials in six lanes, in 3 types of soil. Other in-soil tests had been carried out during the preparation time of the deminer for the next run. Only the reliability trial includes over 8300 data sets including more 96 000 single information. The main advantage of this approach was that it permitted the testing of a greater number of detectors but at the same time a quite large number of operators.

The trial site conditions allowed the simultaneous use of six detectors in three different soil types against three target groups at continuous depths. This amount of data gives an overview of the different factors influencing the detector performance. These include the human factor, the technical solutions of the manufacturer, the targets and their position and finally the ground properties, in particular magnetic susceptibility[5].

The final selection of tests had been as follows:

CWA Test 8.5 Reliability trial
CWA Test 8.4 Fixed depth detection tests in soil
CWA Test 6.7.2 Footprint – sensitivity area of the detectors to different targets (variation) during the training
CWA Test 9.2 Target location accuracy during the operator assessment

5.2. Human factor methodology

For the assessment of the human factor, the following tests have been used[6]:

- NEO PI-R (Costa & McCrae, 1989) - a standardized psychological instrument used for an overall assessment of personality

---

[5] Magnetic susceptibility is the degree to which a material can be magnetized in an external magnetic field. If the ratio of the magnetization is expressed per unit volume, volume susceptibility is defined as $\acute{K}= M / H$, where M is the volume magnetization induced in a material of susceptibility by the applied external field H.

[6] A more detailed description of these tests would be given in the chapter about human factor investigation.

- Test of Attention (d2) (Brickenkamp, 1962) – this test is used to examine attention and the ability to concentrate

As earlier mentioned, an additional questionnaire was also used to get basic information on the deminers and to set up the groups of the deminers.

The purpose of these tests was to investigate which factors have influence on the fact that deminers' results differ. The use of these specific tests was chosen based on assumptions that a) individual differences are, in fact, personality differences; and b) attention ability is definitively important in demining.

5.3. Selection of targets

The selection of targets was based on the regional needs and at the same time with the orientation that a similar set of targets and the lay out will be used for a further trial with dual sensors in 2007. The metal content of the clutter was chosen with the aim that the metal detector will find it at all used depths. The selected targets included:

Figure1: PMA-2

Figure 2: PMA-3

The original PMA-2 and PMA-3 mines rendered safe by replacing the percussion cap to an aluminium cap with similar measures

- The clutter for the lanes has been selected by HCR-CTRO and BAM for the evaluation of discrimination, small clutter has the size of 7.62mm bullets or mortar grenade fragments in similar size.

Experiences from earlier tests have shown that there are practical limitations to the accuracy of targets used in tests. It is difficult to find simulants for minimum metal mines if the original metal part is not available; the less metal is present the more difficult it is to find one faithful simulant, common to all detectors. It should also be borne in mind that mines left in the ground will change over time, generally becoming more difficult to detect e.g. as steel parts rust away, so the reality that is being simulated itself changes. In our case we have mainly aluminium parts in the mine simulants that will reduce the chance for changes in their detectability by the used detectors.

5.4. Test matrix

The test included:

- 20 operators, (they were split into two shifts; one @ 8 operators – the other @ 12operators – the 1st working with 4 types of detectors, the 2nd shift with five types; 4 operators belongs to one group using 2 detector types
- 9 detector types
- 6 lanes on the site
- 2 copies of each type, excluding the Russian and Minelab models where only 1 copy of each type had been available.
- Each operator works in 6 lanes with two types of detectors
- Two mine simulants (PMA-2, PMA-3 and a group of typical clutter for the region had been planted in different depth
- The lanes have three different soil types (neutral, uncooperative homogeneous and uncooperative heterogeneous)

Table 4: Test Matrix

Matrix cut out		**start 1**			**start 2**			**start 3**			**start 4**			**start 5**			**start 6**		
		Deminer	Detector	Specimen	Deminer	Detector	Scimen	Deminer	Detector	Specin	Deminer	Detector	Specimen	Deminer	Dector	Specimen	Deiner	Detector	Specimen
Round 1	**Lane 1**	**A**	**1**	1	**B**	**2**	2	**C**	**2**	2	**D**	**1**	2	**E**	**3**	1	**F**	**4**	2
	Lane 2	**D**	**2**	2	**E**	**4**	1	**F**	**3**	1	**A**	**2**	2	**B**	**1**	2	**C**	**1**	2
	Lane 3	**C**	**2**	1	**D**	**1**	1	**E**	**3**	2	**F**	**4**	1	**A**	**1**	1	**B**	**2**	2
	Lane 4	**F**	**3**	2	**A**	**2**	1	**B**	**1**	1	**C**	**1**	1	**D**	**2**	2	**E**	**4**	1
	Lane 5	**E**	**3**	1	**F**	**4**	2	**A**	**1**	2	**B**	**2**	1	**C**	**2**	1	**D**	**1**	1
	Lane 6	**B**	**1**	2	**C**	**1**	2	**D**	**2**	1	**E**	**4**	2	**F**	**3**	2	**A**	**2**	1

The matrix demonstrates the exploitation of the operators and metal detectors in the different lanes. It shows the manner in which the detectors are cycled through all the lanes, and used by different operators.

For example, operator A starts with detector 1 in lane L1, after he has finished operator B works with detector 2 in that lane. At the same time with operator A, L2 to L6 will be occupied by the operators D, C, F, E, B with their detectors and so on. After the operator has finished the marking of the detected targets, the coordinates will be measured with two total stations. Then the detector will move to other lanes.
The process will be continued in accordance with the matrix so that every operator has been with the two types of detectors in each lane 3 times (at all 18 runs), with 1speciman of detectors 2 times and with the second one 1 time. It is the aim that every deminer will carry out this amount of starts during five days (the full matrix, demonstrates just four days and may be interrupted at every place and then continued).

5.4. *Target layout in lanes*

Each lane contained 29 targets:

- 10 PMA-2 antipersonnel mines,
- 9 PMA-3 antipersonnel mines,
- 10 pieces of metal clutter.

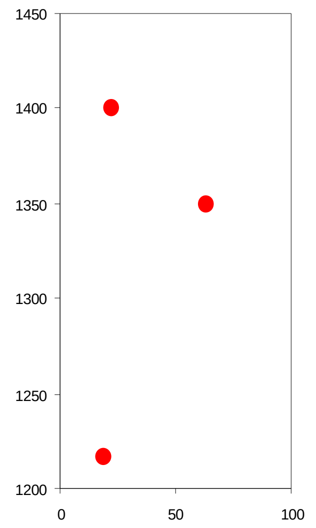


Figure 3: Fragment of a target Lane

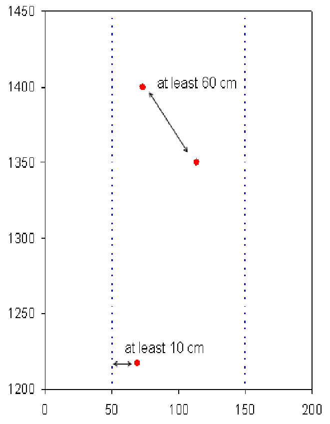


Figure 4: Fragment of a target Lane

The targets were buried to random positions according to the prescriptions of the CWA 14747:2003. The mutual distance between the targets was at least 60 cm and they will lie within the 1-m wide stripe in the lane with their entire halo areas. The mines were buried to depths between 1 and 14.5 cm, in steps of 1.5 cm. The depths of the clutter pieces were determined just before the burial of the targets, in a way that all pieces are easily detectable by most metal detectors in the test. The content of this section will be held confidential to the operators. Figure 3 presents a fraction of a standard lane. The red dots indicate the target positions. This diagram is only an example; the actual positions will be different.

5.5. *Lane preparation*

The targets were placed in accordance with the randomly generated distribution. The original coordinates of the targets' placement were established with the total station during the process of burying the targets.

Due to difficult weather conditions during the target placement no lane check for removing unwanted metal fragments was possible. This check was carried out after the trial by comparing the placement of the markers with real targets. It was very obviously by the placement of markers where still metal could be found. During the check carried out after the trial more than 20 metal pieces had been recovered from the lanes. Due to the used computer program those "true alarms" could be eliminated and had not been counted as false alarms.

Figure 5: Different stages of lane preparation – a) depth, b) buried, c) used for position & depth measurements

The accuracy of the measurements achieved by the operators using the total stations was in average ± 5.86 mm, the most accurate ± 4.83 and the "less" accurate achieved "only" ± 7.65 mm standard deviation. The time for the measurements of one lane (30m x 1m) depended from the amount of markers laid by the operator, the weather (heavy wind), and the operator. The measuring time was between 10 to 15 minutes when 25 to 40 markers had been placed. The use of two stations allowed keeping up with the speed of 6 operators in six lanes. Experiences from the STEMD Laos trial had been confirmed. Manual measurement would have taken much more time and personnel (3 to 4 times). For the use of two stations two different data transmission frequencies are to establish in advance to avoid interference and data confusion. If possible the stations should be placed in a way that the reflectors are not crossing each other during measurements. If this is impossible they should use different height levels of the reflectors for measurements.

The settlement of ground was more than a month. A precision in planting depth of ±5mm was achieved by compacting the ground below the target, using rulers vertically on top of the target, and carefully filling the hole with the removed ground. The stick in Figure 5 c) stayed at centre of the targets and was used for the correct depth and position measurement after the ground had been made even. In this way correction up to 25mm for the greater depths had been made.

5.6. *Detection Depths in Soils*

Figure 6: Target placement for fixed depth

Figure 6 demonstrates the approach for establishing the maximum detection depth for the PMA-2 surrogates (PMA-2S). The depths at which the targets were buried were chosen based on previous trial results (reliability trial in Croatia 2005 [Müller et al, ITEP 2.1.1.8 Final Rep.] and STEMD Mozambique trial 2005 [cite]). The targets for the test of maximum detection depth were placed at the required depth with fixed increments and wooden boards placed on surface to avoid that the search head distance can change to the target depth. The boards will support the height level and hinder the change of depth.

Four deminers with both detectors established the maximum detection depth in soil for reducing the individual influence on the result. Only detectors with ground compensation (GC) were used during the trial. The detector operator carried out the ground compensation before entering the test lane and after he established the maximum detection depth to the PMA-2S. Where the supervisor was not sure about the detection signal, he instructs the operator to make a comparison with the signal from the previous depth. Only when both operator and supervisor agreed the signal, it was accepted as detection.

5.7. General limitations of detector sensitivity measurements in the field

This paragraph content concern not only the establishment of the maximum detection depth to the PMA-2 placed at fixed depth, Figure 6. Some aspects of the establishment of a signal had to be taken into account by the operator that is different during the reliability trial. As the whole detection system was part of the evaluation during the blind trial no consultation or similar advice to the operators had been allowed during the blind trial.
Any measurement made in the field is likely to be less controlled than a laboratory measurement, as is recognised in the CWA. We describe here some factors, which could have influence to signals being recorded incorrectly as detections, even after all the precautions described above. Depending on the soil conditions and the efficiency of the ground compensation, it may be possible to eliminate completely the soil noise, so the only reaction of the detector is to metal. Some of the detectors still have background soil noise either continuously or in reaction to inhomogeneities in the ground, i.e. parts with different electromagnetic properties to the surrounding area or “hot stones” in a neutral environment or the other way around. Some detectors may give background noise due to drift of the electronics or the presence of electromagnetic fields from external sources. A noise cancel function is provided in most of the detectors, typically activated by holding the detector in the air and pushing a button. The detectors vary in the sophistication of their noise cancellation: from simple zeroing to complex intelligent processing. Particularly for the detectors with less effective noise cancellation, there is always a risk of electromagnetic noise being falsely declared as detection. It should also be remembered that ground compensation circuits might be subject to electronic drift.

To a certain extent, the deminer is able to recognise background noise and distinguish it from a true detection. But he is normally not able to distinguish between a signal from a test target and one from other sources. If a specific source of false alarms is located near to a test target, its signals could influence the result. We cannot be sure that every false alarm source will be removed as such and investigated during the trial. Some obvious signal sources had been eliminated other may come up and will be explained as such. In particular, we will highlight in the results cases where there are major discrepancies between the same types of detectors.

Not all operators were aware that some of the detectors could not be used on highest sensitivity because the background noise may overwhelm the signal strength of any target. Where this occurred, it is included into the individual assessment of the detectors.

An obvious possible source of error can be incorrect adjustment or handling of the detector. Although precautions of training and supervision had been to avoid these errors, we cannot be absolutely certain that none occurred.

5.8. Estimate of uncertainty

When trial data have to be used to judge whether or not a detector is able to achieve the sensitivity required for a particular task, or to compare the merits of different detectors, it is important to assess the experimental uncertainties, which are inevitably present. If two results differ by an amount less than the calculated uncertainty they should be regarded as essentially indistinguishable. We attempt here to quantify the known contributions of uncertainty in our measurements and to explain how we combined the estimates to arrive at overall figures. The measurements of depth in field conditions allow a ±10mm error in accordance with the CWA; other errors may be calculated by the achieved results and the amount of measurements.

The maximum detection depth measurements (the CWA test 8.4, “fixed depth detection test”) were performed with several operators. The measurements with each detector model repeated at least twice with four operators. It was expected that repeated measurements will produce different results. The uncertainty of the maximum detection depth was thus estimated.

The results of the reliability test (the CWA test 8.5) are reported in form of ROC diagrams (probability of detection versus false alarm rate) and POD curves (probability of detection versus target depth). The uncertainty of the probability of detection will be estimated based on the assumption that the number of detections is binomially distributed. The uncertainty of the false alarm rates will be estimated based on the Poisson distribution of the number of false alarms. The POD curves are estimated with a generalised linear model and a logistic regression. ROC diagrams and POD curves are generally described and the specific results explained in the individual assessment of the detectors.

5.9. CWA test 6.7.2 Sensitivity profile (footprint))

The establishment of the detectors’ sensitivity profile to different targets was carried out by the operators after the training with both detectors they would use during the trial. The sand boxes with neutral soil and homogeneous uncooperative soil were used for this purpose. Almost all of the operators got for the first time a visual impression about the sensitivity area (footprint, sensitivity cone/profile – other names used) of their detectors[7].

---

[7] One of the operators had his normally used detector with him and repeated the test again after the official working time was finished.

6. Soil properties Benkovac trial site

This chapter contains the pedological description and magnetic susceptibility of the natural soil adjacent to the test site of the Croatian Training and Testing Centre (HCR-CTRO) in Benkovac, Croatia. The results in this chapter are from two investigations of the Leibniz Institute for Applied Geosciences (Preetz & Igel 2005, 2006)[22], additional remarks concerning humanitarian demining are by the authors.

The test site includes 6 test lanes for multiple purposes. There are six test lanes with 3 types of soil, which are later described with their properties.

6.1. Pedological description with estimated specifications for the adjacent area

Soil type (WRB 2006[8]):	Skeletic Chromic Cambisol
Soil depth:	35 cm (on average)
Texture:	Clay to silty clay
Humus content:	1 - 2 %
Soil colour:	reddish brown and brownish grey
Lime content:	5 %
Stone content:	60 - 80 % (limestone with slightly rounded angles)
Rock outcrop on surface:	80 - 90 %
Parent material:	Tertiary limestone

Figure 7: View of the test area

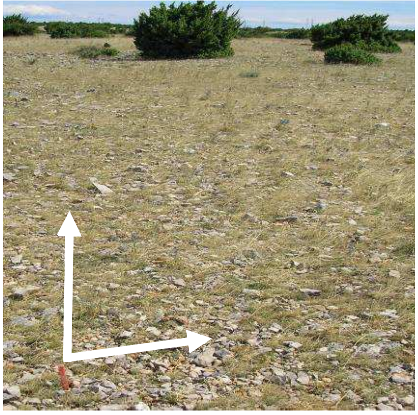


Figure 8: View of the soil surface

Figure 9: Soil profile in a pit

The red marker (stick) in the foreground represents the coordinate 0/0. The line of the sight is along y-axis (see. Figure 7 above).

6.2. Geophysical field measurements

Among the amount of the magnetic susceptibility, the spatial distribution of this parameter is one important factor which can adversely affect a metal detector.

As shown in the description and the pictures before, the soil has a very high content of limestone. We know from several laboratory measurements that the susceptibility of the fine grained material of this soil type in this region is pretty high whereas the one of the limestone is very low (see end of the report). With this combination a high variability of the parameter has to be expected.

---

[8] World reference base for soil resources 2006 - a framework for international classification, correlation and communication.- World soil resources reports: http://www.fao.org/ag/agl/agll/wrb/doc/wrb2006final.pdf

On the test area shown in Figure 8, measurements of the magnetic susceptibility have been carried out due to the determination of the spatial variability of this parameter. Therefore an area of 100 m² has been levelled and measurements were conducted with the MS2D search loop sensor from Bartington Instruments[9]. The distance of the profiles is 1 m with measuring point spacing of 10 cm. The measuring grid is depicted in the following Figure 10 and a plot of the spatial distribution is illustrated in Figure 11.

Figure 10: Grid of the susceptibility measurement

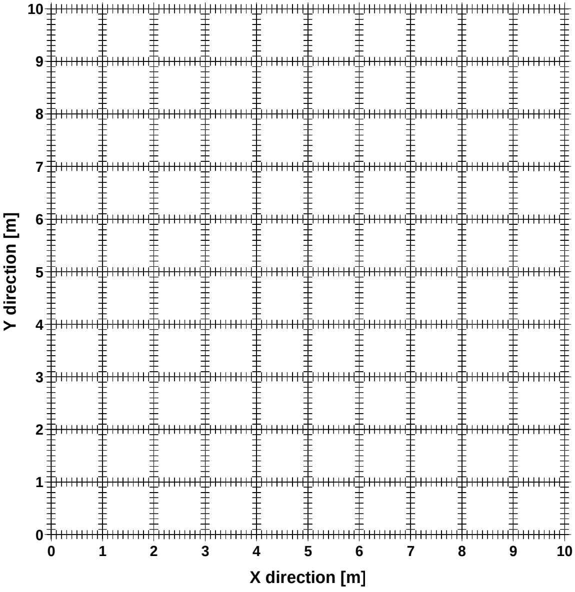
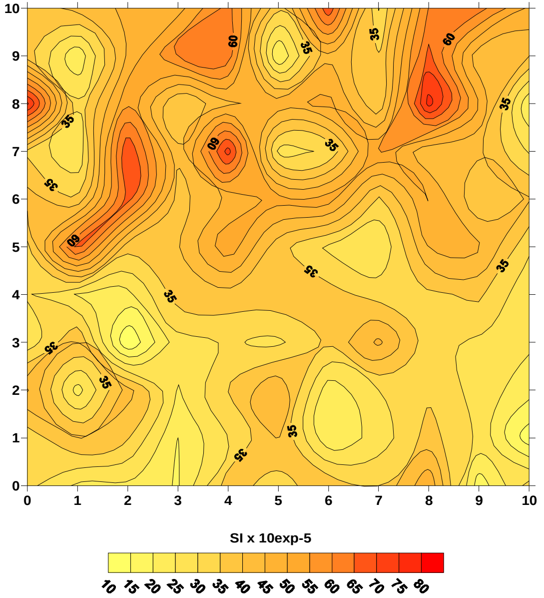


Figure 11: Spatial distribution of susceptibility (κ)



It consists of 2222 measuring points. Figure 11 is the plot of the magnetic susceptibility showing the spatial distribution of the parameter. There are surprisingly high changes of the magnetic susceptibility within the 10m². The deminers have to be aware about those possible changes and carefully check the set up of their detectors; especially the detectors without ground compensation and those that are sensitive to magnetic susceptibility. This fact also underlines to have a field capability for a rough soil assessment. One very simple way is the the measurement of the ground reference height (GRH) that can be done by most of the metal detectors. The detector should be set up to maximum sensitivity and will signal when brought close to the ground if there is a soil problem connected with the magnetic susceptibility. The bigger the distance to the soil the more the sensitivity may be reduced. This is for some detectors up to 60%. More details the reader can find in the STEMD report Mozambique and a detailed description of the GRH measurement in the the

Figure 12: Susceptibility values in Y-direction (κ)

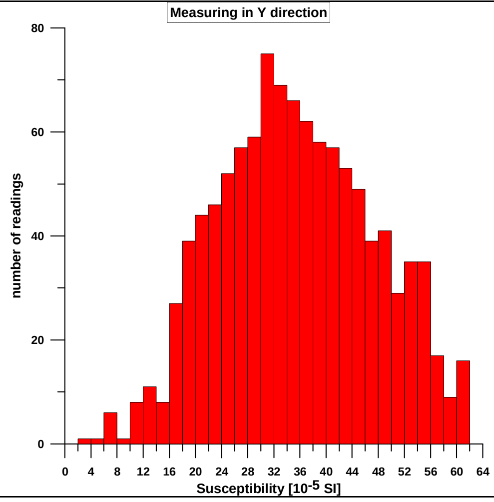
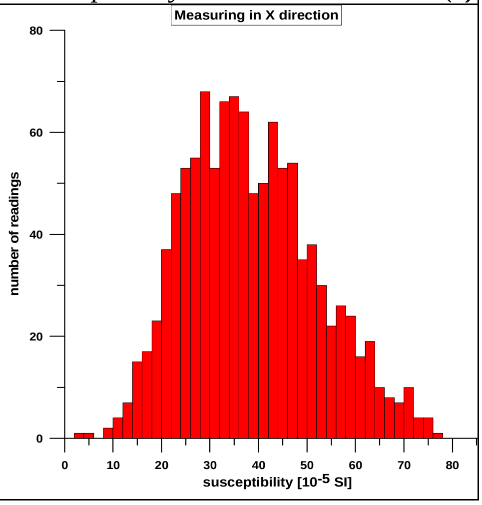


Figure 13: Susceptibility values in X-direction (κ)



---

[9] DEARING, J. (1999): Environmental magnetic susceptibility - Using the Bartington MS2 system; Kenilworth, England.

The histogram Figure 12 shows the measured values in Y direction. The mean average is 37.7 x 10^{-5} SI units for the magnetic susceptibility, the median is 36 and the coefficient of variation is 53 %.
Figure 13 show the same measured in X direction. The mean average value is 39.4 x 10^{-5} SI units for the magnetic susceptibility, the median is 38 and the coefficient of variation is 36 %.
Comparing the two measuring directions, the first statistical approach shows that the mean values as much as the median is nearly similar in both directions. Whereas the variance in Y direction is much higher than along X.

Further information about the spatial variability is illustrated by the variograms below:

Figure 14: Variogram in Y direction (κ)

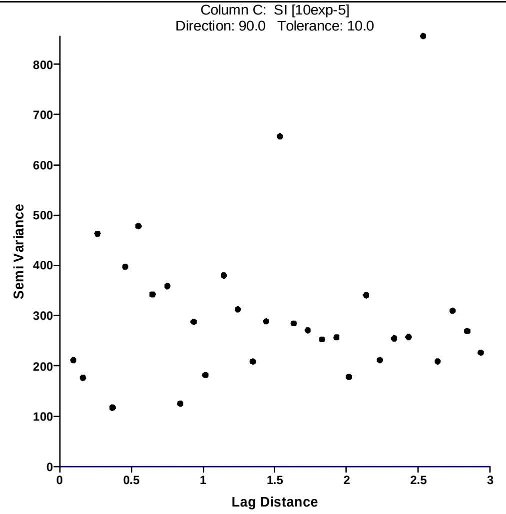
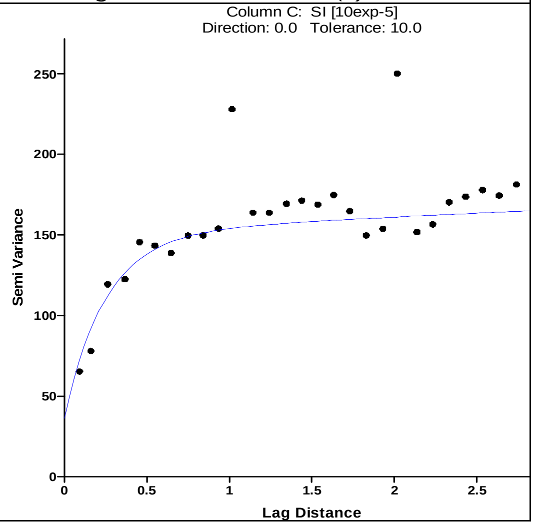


Figure 15: Variogram in X direction (κ)



The variograms in Figure 14 and 15 display different spatial distributions. The readings in Y direction do not have a spatial correlation. The distribution is just randomly in the range of the variance and no model can be fitted to these values. A spatial correlation of the measurements in X direction is clearly visible. The correlation length is approx. 0.5 m; i.e. the distances within the readings are similar to each other.

The reasons for the distinct anisotropy may be in the pattern of the fracture system of the limestone close to the soil surface. The weathering of the limestone is the highest in the fracture zone and the joint filling of these spaces consist of the fine grained soil material which is protected against the erosion at this position. The susceptibility is much higher in the soil material than the adjacent limestone.

Looking at this data from the mine clearance requirements it would make sense to approach the site from the Y-axis and have the clearance direction similar to the X-axis if possible from the terrain configuration. This means also that the deminer will have a similar pattern with the soil properties concerning the magnetic susceptibility along the X-axis. The detector's direct contact with the soil is reduced to few places due to the cover with limestone. There the deminer easily will experience difficulties in detection (signal from the soil) if his detector does not have ground compensation and is not set up to the soil.

6.3. Geophysical laboratory measurements

To give a review about the magnetic properties of the soil in Benkovac concerning the functionality of metal detectors the results of the analysis of the frequency dependent complex magnetic susceptibility of a soil sample are appended. The measurement had been carried out in 2005 and the object matter was a sample from a test lane in Benkovac used for metal detector tests. This soil sample is from the same area and has the same properties as those soils on our measuring field.

The real and imaginary part of the susceptibility at 12 different frequencies (50 Hz - 10 kHz) was determined with a Magnon VFSM susceptibility bridge[10]. The magnetic field strength was 161 A/m.

---

[10] www.magnon.de

Figure 16: Benkovac Soil / Lane 7 / B

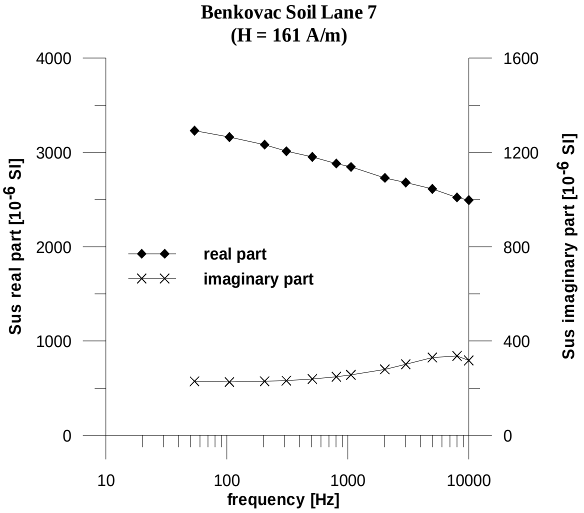
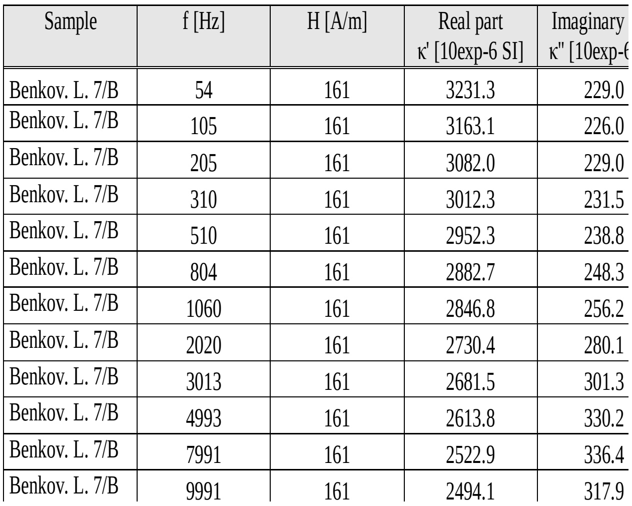


Figure 17: Frequency dependent complex κ
Benkovac soil



Sample	f [Hz]	H [A/m]	Real part κ' [10exp-6 SI]	Imaginary κ'' [10exp-
Benkov. L. 7/B	54	161	3231.3	229.0
Benkov. L. 7/B	105	161	3163.1	226.0
Benkov. L. 7/B	205	161	3082.0	229.0
Benkov. L. 7/B	310	161	3012.3	231.5
Benkov. L. 7/B	510	161	2952.3	238.8
Benkov. L. 7/B	804	161	2882.7	248.3
Benkov. L. 7/B	1060	161	2846.8	256.2
Benkov. L. 7/B	2020	161	2730.4	280.1
Benkov. L. 7/B	3013	161	2681.5	301.3
Benkov. L. 7/B	4993	161	2613.8	330.2
Benkov. L. 7/B	7991	161	2522.9	336.4
Benkov. L. 7/B	9991	161	2494.1	317.9

As shown in Figure 16 and 17 the soil has a frequency dependence of 11 % over 10 decades which is a strong evidence for the presence of a super paramagnetic compound of the magnetic minerals. The absolute values of the frequency dependence can be read in table 1 as well. Moreover, the susceptibility of the soil sample measured in the laboratory is nearly ten times higher than the mean of the field measurements. This is because the sample consisted only of pure soil without any limestone, which is reducing the susceptibility of the field measurement significantly. As mentioned above the limestone content in the topsoil is up to 90 % (see pedological description). Therefore the susceptibility of the field measurement with the Bartington loop MS2D is at about 10 % of the laboratory measurement of the pure soil. A further reason for the low level of measurements is that the placement of the coil could rarely be done cases directly on soil. In most cases the loop could only be placed at the neutral stones sticking out of the ground.

For comparison with the other available soil types and the evaluation of the influence of the soil properties on detector performance the susceptibility measurements of both other available soils on the Benkovac test site are added below, Figure 18 to 21.

Figure 18: Obrovac soil Lane 1 & 2

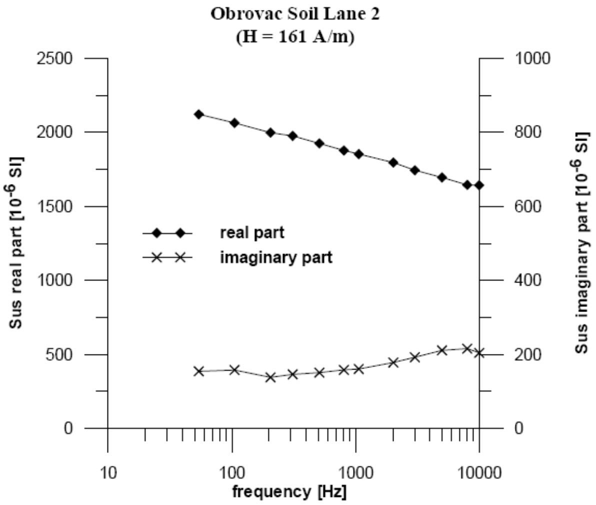
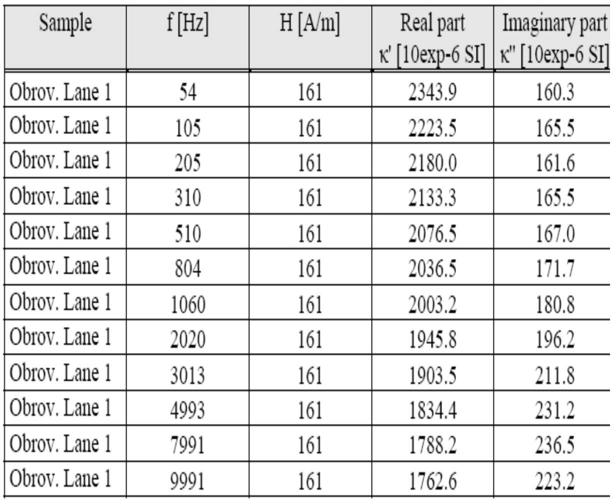


Figure 19: Frequency dependent complex κ of Obrova soil



Sample	f [Hz]	H [A/m]	Real part κ' [10exp-6 SI]	Imaginary part κ'' [10exp-6 SI]
Obrov. Lane 1	54	161	2343.9	160.3
Obrov. Lane 1	105	161	2223.5	165.5
Obrov. Lane 1	205	161	2180.0	161.6
Obrov. Lane 1	310	161	2133.3	165.5
Obrov. Lane 1	510	161	2076.5	167.0
Obrov. Lane 1	804	161	2036.5	171.7
Obrov. Lane 1	1060	161	2003.2	180.8
Obrov. Lane 1	2020	161	1945.8	196.2
Obrov. Lane 1	3013	161	1903.5	211.8
Obrov. Lane 1	4993	161	1834.4	231.2
Obrov. Lane 1	7991	161	1788.2	236.5
Obrov. Lane 1	9991	161	1762.6	223.2

When assessing the influencing parameter one should pay attention to the frequency dependency of the magnetic susceptibility and to the soil structure. Limestone is a magnetic neutral material and will create for some detectors additional complications that are sensitive to in-homogeneities in a soil with high magnetic susceptibility.

Figure 20: Sisak soil Lane 3

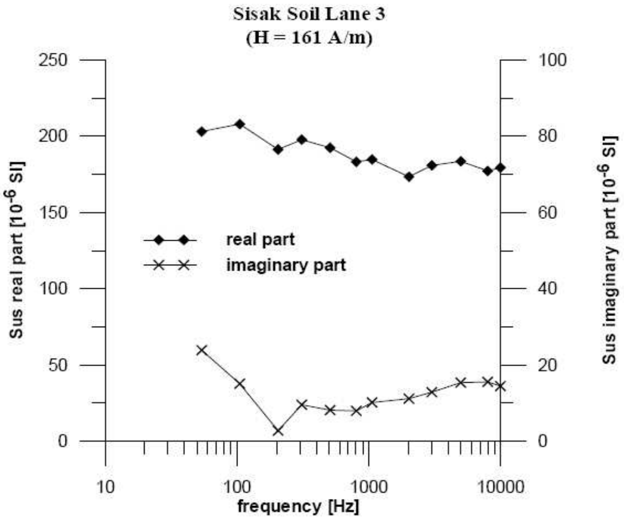
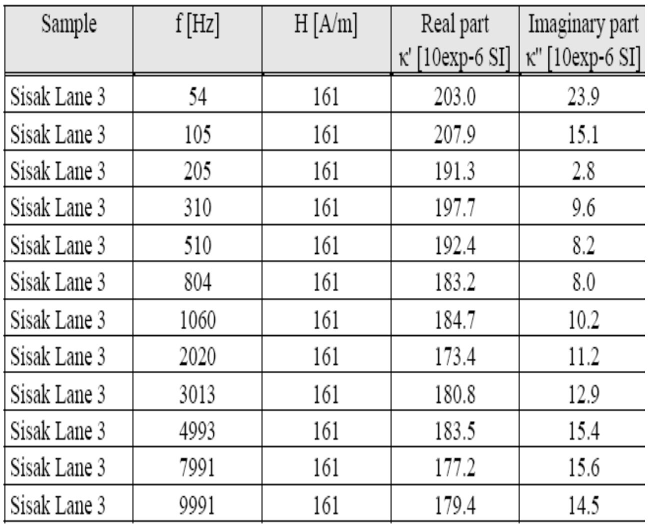


Figure 21: Frequency dependent complex κ of Sisak soil



Sample	f [Hz]	H [A/m]	Real part κ' [10exp-6 SI]	Imaginary part κ'' [10exp-6 SI]
Sisak Lane 3	54	161	203.0	23.9
Sisak Lane 3	105	161	207.9	15.1
Sisak Lane 3	205	161	191.3	2.8
Sisak Lane 3	310	161	197.7	9.6
Sisak Lane 3	510	161	192.4	8.2
Sisak Lane 3	804	161	183.2	8.0
Sisak Lane 3	1060	161	184.7	10.2
Sisak Lane 3	2020	161	173.4	11.2
Sisak Lane 3	3013	161	180.8	12.9
Sisak Lane 3	4993	161	183.5	15.4
Sisak Lane 3	7991	161	177.2	15.6
Sisak Lane 3	9991	161	179.4	14.5

The soil properties presented in Figure 18 to 21 demonstrates a decreasing susceptibility. In a similar way the influence of the soil on detection ability of the detectors decreases from uncooperative to neutral soil depending on the frequency dependency of the soil.

The next chapter contains the trial results and their discussion.

7. Results: comparison of the detector and operator performance

7.1. Introduction,

This chapter compiles the result of the trial concerning the metal detector tests. First an overview about the results not belonging to the blind trial is given and what was included to collect information about the detector sensitivity to the main targets used in the blind trial. That will allow better to analyse the results achieved during the main test.
During the preparation and training we could establish that most of the demining organisations do not include simple tests of metal detectors into the training of deminers that will give them a clear picture and better understanding to the sensitivity of their main tool – the metal detector. Here a variation of the CWA sensitivity profile test was used to explain factors influencing on the sensitivity of the metal detector.
The pinpointing test gives information with which accuracy the operator can define the position of a target not visible to him. At the same time this information is necessary to understand marking errors or false alarms occurring during the blind trial.

Tests of maximum detection depth were performed as described in Section 5.6. Surrogates of PMA-2 mines (marked PMA-2S) were used in that test. Measurements of maximum detection distance in air were performed with the goal to investigate the difference between the PMA-2 and the surrogate PMA-2S and to investigate the variability among the specimens of the same target type.

The main results of the trial are included into the blind trial and its evaluation. The reader will get an overview and a description how to understand the used graphs and curves. The follow on text will focus on the explanations that are important for the comparison of the data. We want also to demonstrate that our approach is not comparable with the real results achieved in mine clearance operations.

7.2. Results from other tests, not included into the blind trial

7.2.1. Sensitivity profile CWA test 6.7.2 (variation)

The CWA Test 6.7.2 Method 2 for establishing the sensitivity profile[11] (footprint) – the CWA describes the establishment of the footprint by a scanner and a second way manually to balls with different diameters. This is foreseen to be done in the lab and will give the detailed picture of the footprint to those targets. For international comparison this way was recommended.
For field use and more important for regional interest is the easily repeatable version used during this trial described in the "METAL DETECTOR HANDBOOK FOR HUMANITARIAN DEMINING"[21]: The detector is set up to the soil conditions, half buried and will not be moved. The operator used the target that is the object of his task and sees at which depth and within which radius he can find it during the search in the test lanes. This approach also allows comparing the detectors directly and with different targets in field conditions. This was of special interest to the deminers.

Figure 22: Sensitivity target 1

Figure 23: sensitivity 3 targets

Figure 24: Sensitivity 2 operators

---

[11] Sensitivity profile – defines the area below the search head within a target can be detected; other used names are footprint, sensitivity cone or area

At the same time this profile was used to explain the search advance dependent of the target and target depth. A turn of the detector by 90° will demonstrate the profile along the other axis. In this way the operator/deminer receives a three-dimensional picture about the sensitivity area of his detector and may easily confirm or check his clearance depth.

7.2.2. Pinpointing of targets (*CWA Test 9.2 target location accuracy)*

The pinpointing abilities of the detectors in lab conditions are from 2mm to 14mm. The accuracy depends on the detector type and on the ability of the deminer. The conditions in the field do not allow this accuracy. The marking on the ground is often only done visually by surface characteristics. Additionally every organisation has its own experiences in marking and signal investigation so that different requirements to accuracy are possible and acceptable. In general the field accuracy should not be above the diameter of the smallest used mine in the region. The achieved results in Table 3 are the average achieved by all deminers using the same detector.
In our trial only one operator was in the area of critical accuracy with 70mm when pinpointing a PMA-2. To the end of the trial he achieved a much better accuracy.
The results in Table 5 demonstrate that the achieved average accuracy will provide that the deminer when investigating the signal will not miss the mine.

Table 5*: Pinpointing*

Pinpointing Accuracy (cm)								
CEIAM1	Condor 7252	Minelab F3	Minelab F1A4	MINEX 2FD 4.530	Schieb AN19	Vektor 7260	VallonVMC1	VallonVMH3CS
2.3	5.0	2.2	2.0	1.4	3.0	3.4	2.6	3.5

7.2.3. Maximum detection depth in soil

The maximum detection depth (MDD) measurements were performed on PMA-2S, a surrogate of PMA-2, as described in Section 5.6. The number of repeated measurements with each detector model was between 4 and 12. The original intention of the trial organisers was to consider the largest depth at which the target can be detected as the MDD and to find the variability of the MDD. However, the results of many measurements did not allow such an interpretation. It often happened in a single MDD measurement that a target was detected, for example, at 4, 5, 6, ... 10 cm, not detected at 11 cm, detected at 12 cm and not detected at 13, 14, 15 cm. It is not possible to know whether the signal at 12 cm depth is a false alarm or a true detection (see Table 4). Such results occurred in all soil types, but the results were especially unclear in the uncooperative soils (Benkovac soil and Obrovac soil). *Table 6* presents two examples of an ambiguous result produced by detectors CEIA MIL-D1 and AKA Condor both in Benkovac soil. Most other detectors had similar results.

Table 6*: Two examples of an ambiguous result*

depth [cm]	CEIA	Condor
	detected [1=yes, 0=no]	
4	1	1
5	1	0
6	1	1
7	1	1
8	1	0
9	1	1
10	1	0
11	0	0
12	1	1
13	0	0
14	0	0
15	0	0

The main reason for such behaviour of metal detectors is the local electromagnetic properties of the soil where the targets are buried, including the possible presence of small metal clutter. Other reasons are the uncertainty of the acoustic signal, subjectivity of the operator, sensitivity changing in time, and other unknown influences. For example, the decision not to perform 5 sweeps with a clear signal over each target, but to leave the decision whether the target is detected to the personal judgment of the operator, certainly increased the variability of the results. However, this alone cannot explain irregularities such as in the last column of *Table 6*. There the knowledge in using the detector, the signal interpretation, and ground compensation capability may have influenced on the results.

The results point to the conclusion that the definition of the maximum detection distance as the largest depth at which a target is detected can not be used in the experiment described in this report. Instead, the term "maximum detection depth" is used in this report to refer to the measurements and the analysis described in this section. Since the MDD could not be established as a result of a single measurement, different data analysis was applied. The detections of each target were treated as results of Bernoulli experiments, with 1 (detected) or 0 (not detected) as an outcome and with the probability of detection (POD) as a parameter of the Bernoulli distribution. A generalised linear model with a logistic regression was used to estimate the POD in dependence on depth for each detector-soil combination. The curves of POD versus depth are called POD curves. The logistic transformation means that the logistic function

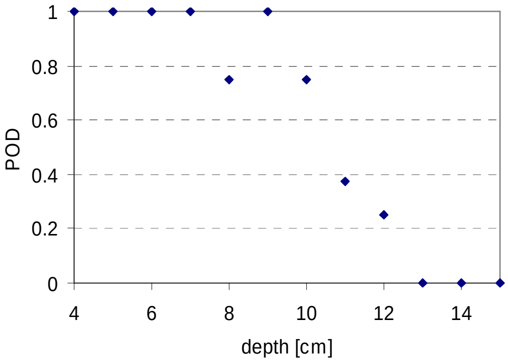
$$POD = \frac{1}{1 + e^{a+bx}}$$

defines the shape of the POD curve, where x is the depth and a and b parameters estimated by maximum likelihood estimation. 95% confidence bounds are produced to these curves. The depth at which POD equals 0.5 is marked $d_{0.5}$ and it is equivalent to the average maximum detection depth. That can be demonstrated on the example of Minelab F1A4. In the results of this detector, there were no irregularities such as those described at the beginning of this section. Thus it was possible to establish the MDD in a classical way, as the largest depth at which the target is detected. The average MDD for this detector was 138 mm. The same results were analysed with the generalised linear model. The value of $d_{0.5}$ was 139 mm, which is almost identical to the “classical” MDD. For a more exact explanation, see [5]. The depths at which the confidence bounds of the POD curve cross the straight POD=0.5 are the confidence limits of $d_{0.5}$. The whole procedure is illustrated in Table 7, Figure 25 and Figure 26 on the example of Foerster MINEX in Obrovac soil. For example, let us examine the depth 10 cm. From Table 7 we read that the target buried to 10 cm depth was detected 6 times out of 8 attempts. The estimated POD at that depth is therefore 6/8=0.75. This estimated POD is represented by a point on Figure 25. The curve on Figure 26 is a result of a generalised linear regression described above (and in more detail in [5]). The generalised linear model was used also in the analysis of the reliability test results, see Section 7.3.

Table 7: MDD measurements, an example of Foerster MINEX in Obrovac soil.

depth [cm]	detected [1=yes, 0=no]								total number of detections
4	1	1	1	1	1	1	1	1	8
5	1	1	1	1	1	1	1	1	8
6	1	1	1	1	1	1	1	1	8
7	1	1	1	1	1	1	1	1	8
8	1	1	0	1	1	1	0	1	6
9	1	1	1	1	1	1	1	1	8
10	1	1	1	1	1	0	0	1	6
11	0	1	0	1	1	0	0	0	3
12	0	0	0	1	1	0	0	0	2
13	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0

Figure 25: MDD measurements, Foerster MINEX in Obrovac soil, estimated POD for each depth present in the test.



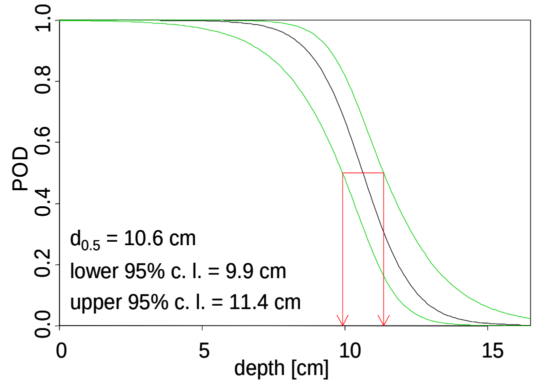


Figure 26: MDD measurements, Foerster MINEX in Obrovac soil, estimated POD curve with 95% confidence bounds.

An overview of the MDD results is in *Figure 27*. It presents the depth at which the POD falls to 0.5. In Sisak soil, detectors Minelab F3 and Vallon VMC1 detected almost all targets, while Vallon VMH3CS detected all of them. For these detectors in Sisak soil, it was not possible to use the generalised linear model and to produce POD curves with 95% confidence bounds. Nevertheless, it is clear that the result of Minelab F3 is $d_{0.5}$ = 16 cm, while the results of the two Vallon models are higher than 17 cm, which is why these two columns are marked with a darker nuance of yellow.

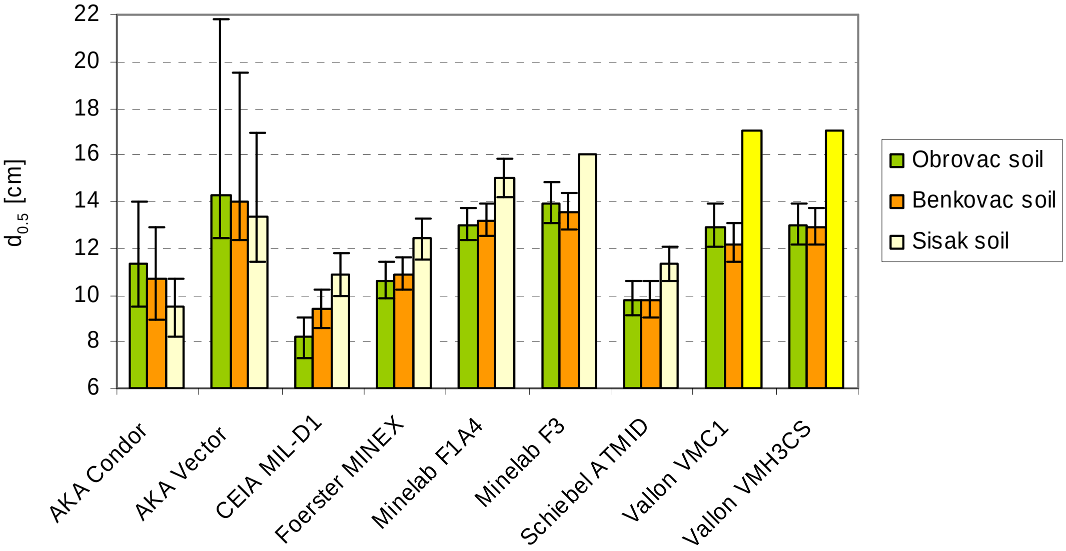


Figure 27: MDD measurements, the depth $d_{0.5}$ at which POD=0.5, comparison of soils and detectors.

The diagram on *Figure 27* contains less information than the POD curves. The POD curves are given in Annex 7. By looking only at $d_{0.5}$, we lose the information about how much sensitivity a metal detector loses at smaller depths. To get that information, we need also the slope of the POD curve at POD=0.5. This value carries the information about the stability of the metal detector: the larger the slope, the more stable is the detector. As an alternative with a more intuitive appeal, one can also look for $d_{0.9}$, i.e. the depth at which POD reaches 0.9. If $d_{0.9}$ is closer to $d_{0.5}$, than the POD curve is steeper and the detector is more stable. A comparison of metal detectors should be based on both $d_{0.5}$ and $d_{0.9}$ values. The $d_{0.9}$ values are not presented in this report, but the reader can read them approximately from the POD curves of the MDD measurements given in Annex 7. Detector AKA Vector is a good example: its POD falls to 0.5 at the highest depth compared with other detectors (in other words, its $d_{0.5}$ is the highest), but a comparison of $d_{0.9}$ values (which can be read from the diagrams in Annex 7) reveals that the Minelab models and the Vallon models are more stable: targets at smaller depths are more often missed by Vector.

The results of this trial (STEMD Croatia 2006) were surprisingly high, compared with the previous trials. During the MDD test three detectors detected all or almost all targets in Sisak soil. These detectors were: Minelab F3, Vallon VMC1 and Vallon VMH3CS.

There is no statistically significant difference between the performance in Obrovac and Benkovac soil (see *Figure 27*). The ranking in both of these soils is as follows: the largest depths were achieved by AKA Vector, both Minelab models and both Vallon models, with no statistically significant difference between these five models. Detector AKA Condor achieved lower results, but the difference is not statistically significant. All other detectors (CEIA, Foerster and Schiebel) achieved lower results, with a significant difference. The ranking in Sisak soil is almost identical.

This ranking can be compared with the results of the reliability test, which was performed with the same detectors, operators, soils and targets and is described later in more detail in chapter 7.4. There are large differences between the MDD measurement results and reliability test results. This supports the conclusion that the MDD measurements cannot predict the outcome of a reliability test. Reliability tests give a more reliable estimate of the performance of metal detectors in real minefields, since these tests come closest to representing the real field conditions in demining. An analysis of $d_{0.9}$ values leads to the same conclusion.

The main cause of the difference between the ranking in MDD results and in the reliability tests is the following: In a reliability test, the operators do not know the positions of the targets. During their search for the targets, they may receive many false alarms. They sometimes reduce the sensitivity of their devices, or they decide to ignore smaller signals to reduce the false alarm rate. Thus they also reduce the probability to find a target. During the MDD measurements, some alarms actually caused by the soil or by small metal clutter are interpreted as a detection of a buried target.

7.2.4. Maximum detection distance in air

As a part of this trial, some measurements of maximum detection distance (MDD) in air were performed. The two goals of these measurements were: to investigate the difference between the PMA-2 and the surrogate PMA-2S and to investigate the variability among the specimens of the same target type. Four specimens of PMA-2 and four specimens of PMA-2S were used in the measurements, so that the variability could be estimated. The MDD for all eight targets was measured with 12 combinations operator-detector. For each of these 12 combinations there were four pairs of targets (PMA-2 and PMA-2S). For each target the MDD was measured and for each pair the MDD difference was calculated. The average difference in detection between the PMA-2 and PMA-2S was (1.8±0.3) cm, where ±0.3 marks the 95% confidence limits. In average, the real mine had a 1.8cm larger maximum detection distance than the surrogate. In other words, it is easier to detect the real PMA-2 mine than its surrogate PMA-2S.

7.3. Reliability of detection

7.3.1. Introduction

This section gives an overview about the reliability trial in Croatia, and the results achieved by the detectors in all lanes. The comparison of the results will demonstrate different factors influencing on the detector performance and some limits in evaluation.

A target which can be found in the lab conditions may nevertheless be missed in the field if the operator loses concentration, or does not sweep over the point where it lies, or sweeps too fast or too slowly for the detector electronics, or misinterprets a weak mine signal as a soil signal. An operator may also incorrectly signal the presence of a mine when none is there if there is a signal from a small area of magnetic soil minerals or a small piece of metal clutter or electronic noise. Such errors are unpredictable but one may measure the probability of their happening in statistically-based blind reliability trials, in which a team of operators attempts to find rendered safe mines or other targets buried at locations unknown to them. The Probability of Detection (POD) for a given target in given conditions and the False Alarm Rate (FAR) depend on the detector design, the operator behaviour as well as environmental factors mutually influencing each other. This complex of influencing factors is part of test and evaluation. Details of the method are now standardised in CWA 14747 Section 8.5 under the name “reliability tests”.

At first glance, the results of all reliability tests performed up to the present are surprisingly low. However, reliability tests are purposely designed to be difficult. The choice of depths in a trial does not represent the actual situation in a minefield, but rather a difficult scenario. In reality, most mines are found near the ground surface. In a trial, some mines have to be buried deeper to bring the detectors to their limits. Such a choice of depths makes the differences between detectors more apparent. For example, the targets used in our trial were buried between 1cm and 14.5cm. The POD for lower depth can easily be taken from the ROC curves where the POD is expressed in dependence of target depth.

The estimated POD for a particular choice of detectors, soils and targets is the ratio of the number of detected targets and the total number of opportunities to detect a target. For example, if two runs were made on a lane with 30 targets, and if 22 out of 30 targets were detected in the first run and 26 targets in the second run, than the estimated POD is (22+26)/(30+30)=0.8. The estimated FAR is defined as the number of false alarms counted on an area divided by the size of that area, or the average number of false alarms per square metre. The area is calculated as the area of the test lane minus the area of all detection halos. For example, let us consider 30 antipersonnel mines PMA-2 buried to a 30m long and 1 m wide lane. The halo radius of this target is r=10cm, the total area of all halos is $30 \times r^2\pi = 0.9m^2$, so that the lane area needed for the calculation of the FAR is $29.1m^2$.

POD and FAR are related in the sense that they both decrease if the operator reduces the sensitivity of the instrument, or implicitly does so by requiring a clearer sound before calling an indication. It is therefore the usual practice to quote POD and FAR together. The POD and FAR are combined in a diagram called an ROC diagram, where ROC stands for “receiver operating characteristic”. An ideal detector would have POD=1 and FAR=0 and it would be represented by a point in the upper left corner of an ROC diagram. To each point on an ROC diagram, 95% confidence limits are attributed. They describe the uncertainty of the estimate of the POD and of the FAR. The difference between two detectors can be roughly estimated as statistically significant if the confidence intervals of the two points representing those detectors do not overlap.

The used clutter had been distributed in depth as the mine targets. The small clutter not deeper than 60mm. Minelab F3 detected the Clutter with POD 0.98; Vallon MVMH3CS 0.9, and CEIA MIL-D1with a POD 0.86

7.3.2. Discussion of results

Overall detector performance

Figure 28 and *Figure 29* demonstrate the overall results of the reliability trial for all detectors in all lanes. All curves presented in this section are described with the same legend given in *Figure 29*. In *Figure 28* the POD in dependence of depth is demonstrated. There is a clear difference between the detectors and their behaviour to the change of the target depth. The probability of detection drops to 0.5 at the depth $d_{0.5}$ =4.5cm for the AKA Vector , for a group of detectors (Condor, CEIA, Schiebel, Vallon) at the depth about 8 to 9.5 cm and for the Minelab and Foerster detectors about 13 to 15cm target depth.

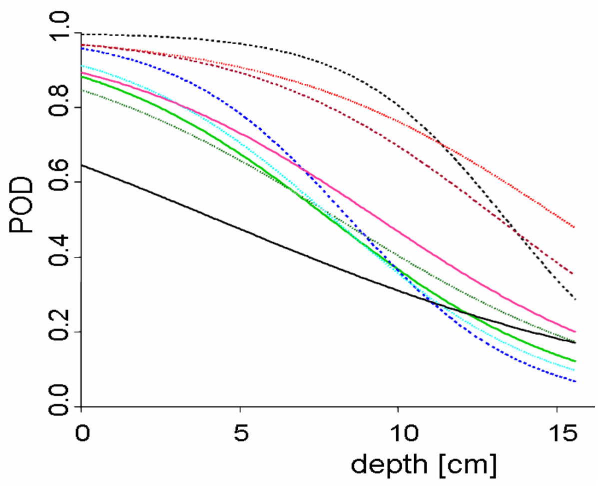


Figure 28: POD in dependence of depth, all detectors in all lanes

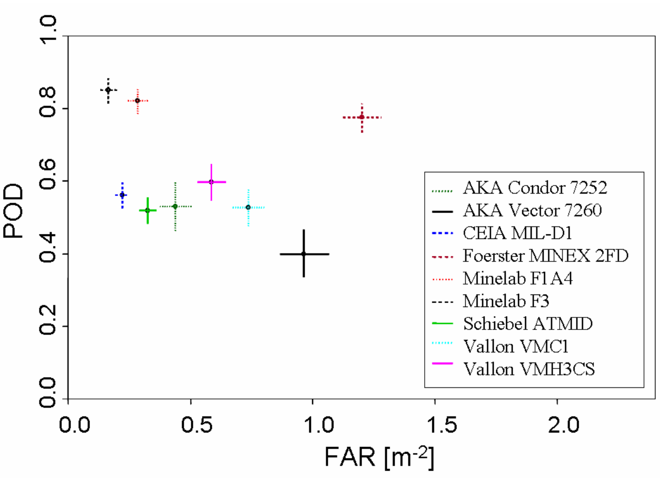


Figure 29: ROC diagrams summarising detector results in all lanes

Figure 29 demonstrates the overall performance of the detectors in a ROC diagram. Regarding POD, two groups of detectors can be distinguished. The two Minelab models and the Foerster MINEX 4.530 achieved clearly better results concerning the POD. Their POD was about 0.8. The difference between these detectors and other detectors is statistically highly significant. All other detectors had a POD performance between 0.4 and 0.6. If one looks at the false alarm rate the results are much more different. There is nearly no grouping but more a split along the x-axis with significant differences between the detectors. The three detectors which achieved the highest POD are very different with regard to the FAR: the two Minelab models have a much lower FAR than the Foerster MINEX.

Figure 30, Figure 32 and Figure 35 are ROC diagrams for each soil type separately. The results are somewhat higher in Sisak soil because the frequency dependent magnetic susceptibility is in that soil lower than in the other soil types present in the test. Figure 31, Figure 33 and Figure 34 are POD curves (POD in dependence on depth) with the results of all detectors in three soil types separately. The 95% confidence bounds of these curves can be found in Annex 7.

Sisak soil results

Figure 30 is an ROC diagram of the Sisak soil results. The two Minelab detectors had the highest POD and the lowest FAR in the trial. Minelab F3 had a smaller FAR and the difference to Minelab F1A4 was at the limit of statistical significance, while there was no difference between their PODs. Regarding POD, the difference to AKA Vector 7260 was not statistically significant, and the difference to Foerster MINEX 2FD was at the limit of statistical significance. Regarding FAR, the difference between AKA Vector 7260 and Minelab F3 was at the limit of statistical significance, in favour of the Minelab detector. There are other detectors with a FAR smaller than the FAR of AKA Vector, but they all had low PODs. Figure 31 shows POD curves for the same soil. It can be seen that they all dropped to 0.5 at larger depths than in the Obrovac soil. That was expected, since the electromagnetic properties of Sisak soil were easier for metal detectors. In the depth range down to 5 cm depth, there were no differences between most of the detectors: they detected more than 90% of the targets. There were two exceptions: Schiebel ATMID and Vallon VMHCS, with a POD about 80%. The depth at which the POD dropped to 0.5 depended on the detector: the weakest was Schiebel ATMID with 7 cm, and the best were Minelab F1A4 and Foerster MINEX 2FD with about 18 cm depth.

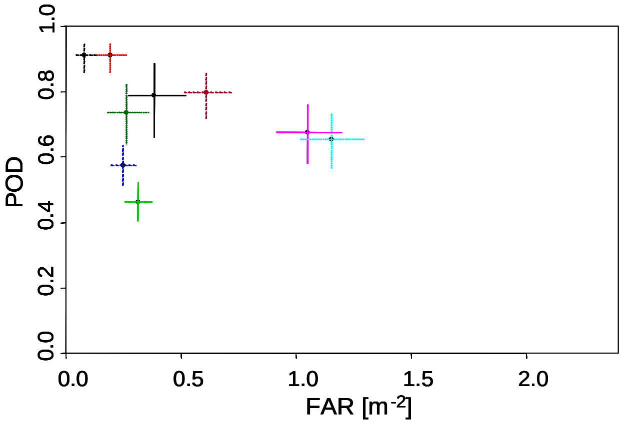


Figure 30: ROC diagram, Sisak soil.

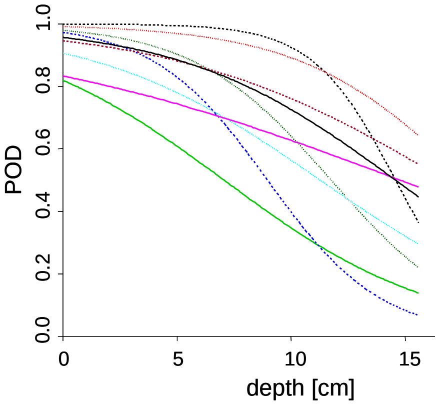


Figure 31: POD curves, Sisak soil.

Obrovac soil results

Figure 32 is an ROC diagram of the Obrovac soil results. The two Minelab detectors and the Foerster detector had a clearly higher POD than other detectors. Regarding the POD, the difference between the Minelabs and the Foerster was not significant, while the difference to other detectors was highly significant. However, the two Minelab models had a much lower FAR than the Foerster model. An obvious ideal choice for demining operations in this soil type would be Minelab F3 or Minelab F1A4. *Figure 33* shows POD curves for the same soil type. The depth range between 0cm and 5cm is especially interesting, because mines are mostly laid near the surface. At those depths Minelab F3 and Foerster MINEX 2FD had the highest POD. They were closely followed by Minelab F1A4, but it is not possible to see from this diagram whether the difference between them is statistically significant. An interested reader should study ANNEX 7, where all POD curves are published with their corresponding 95% confidence bounds. The lowest POD in that depth range was achieved by AKA Vector 7260. At the depth of about 10cm, the two Minelab models and the Foerster were clearly above the other detectors: their POD was between 0.7 and 0.9, while all other detectors were below POD=0.5. The depth at which the POD drops to 0.5 was about 12cm for Minelab F1A4, Minelab F3 and Foerster, while other detectors achieved weaker results equal to 0.5, and for Minelab F1A4 it wasis 15 cm.

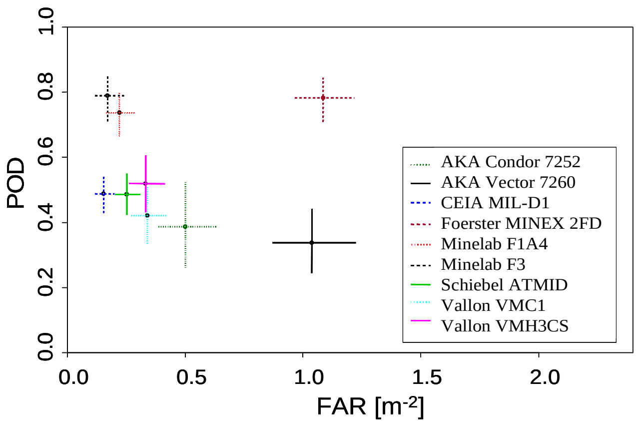


Figure 32: ROC diagram, Obrovac soil.

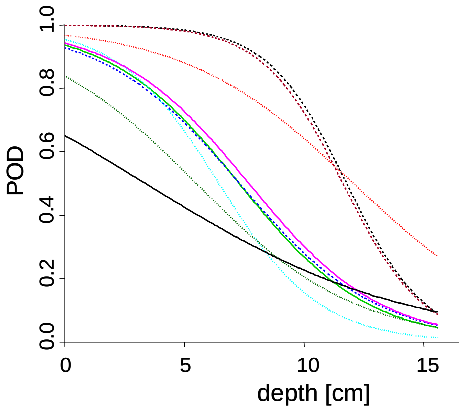


Figure 33: POD curves, Obrovac soil.

Benkovac soil results

Figure 34 is an ROC diagram from Benkovac soil. The differences between detectors were very similar to the differences appearing in Obrovac soil, since the two soils were quite similar. Again the two Minelab models would be the best choice for this type of soil. *Figure 35* shows POD curves for the same soil type. In the range between 0cm and 5cm depth, the best detectors were Minelab F3 and CEIA MIL-D1, but very closely followed by Minelab F1A4, Foerster MINEX 2FD and then by other detectors. The depths at which the POD reaches 0.5 were very different: from 0cm for AKA Vector to about 20cm for Minelab F1A4.

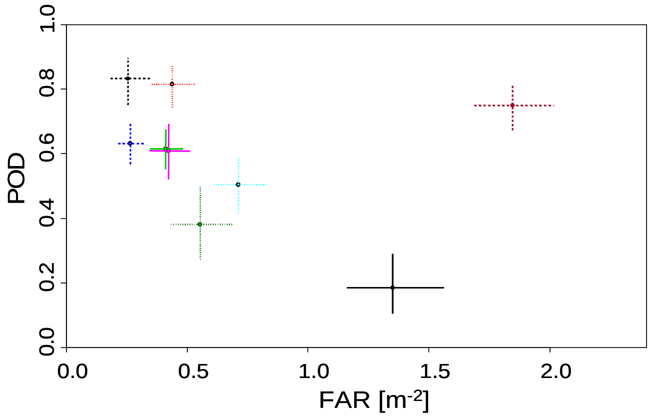


Figure 34: ROC diagram, Benkovac soil.

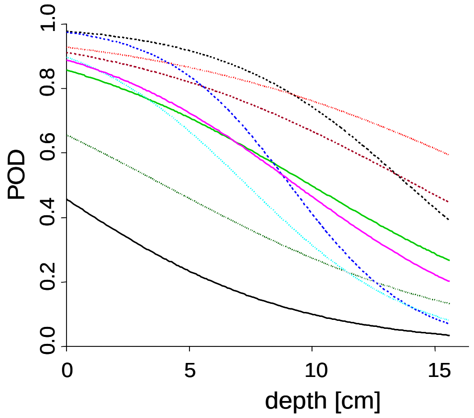


Figure 35: POD curves, Benkovac soil.

A comparison of MDD measurements with the reliability test results

Annex 7 provides a comparison between the MDD results and the reliability test results for each detector in each soil separately. It should be born in mind that the MDD measurements were performed on PMA-2S, which is slightly more difficult to detect than the PMA-2 used in the reliability test (see Section 7.2.4). Here we see once more that the two results differ. It should be emphasised again that the reliability tests provide a more reliable comparison of metal detectors, since they are performed in conditions closer to those in a real minefield. In a reliability test, the false alarm rate is also evaluated. The operators often reduce the sensitivity of their detectors to reduce the false alarm rate, what also reduces the probability of detection. In a MMD test, the influence of false alarms is not evaluated. Some indications might as well be caused by the soil – an event that would be counted as a false alarm in a reliability test.

7.4. Operator groups evaluation in connection with detector use and time

The chapter includes the assessment of the operator groups concerning their results under the view of their time needed in the lanes for the different detectors which allows to a certain degree to evaluate the ease of use of the detector and also the interaction of the deminer with the detector.

Figure 36 and 37 demonstrates the overall time the detector had been in a lane in average used by all deminers and the time a single deminer needed in average in the lanes. The general time shows already serious differences and indicates that there should be reasons for this. The operator groups have been chosen to avoid biased data collection so we can assume that the detectors and their interaction with the operators are the reason for time differences. The first shift (Group 1 and 2), 8 operators were divided into two groups using two detectors each. The first used the two Minelab, the second the MIL-D1 and ATMID detectors. The second shift, 12 operators, with Group 3 to 5 used the both Vallon's, the AKA detectors (Condor, Vektor) and the MINEX 4.530 only the last group. The second shift needed in general much more time than the first one. The time variables within the first two groups are acceptable and indicate an easy understanding and use of the detectors. The time for the other groups increased essentially and indicates that there are problems in the use and the interface between the detector and the operator. Outstanding is the time needed by the MINEX 4.530, 1:20h as average for all deminers in all lanes. It should be kept in mind that this time was in average needed to search an area of about 30m^2, to pinpoint the readings (detector signals) and to mark them[12]. The time difference between the fastest and slowest operators is above one hour in a lane.

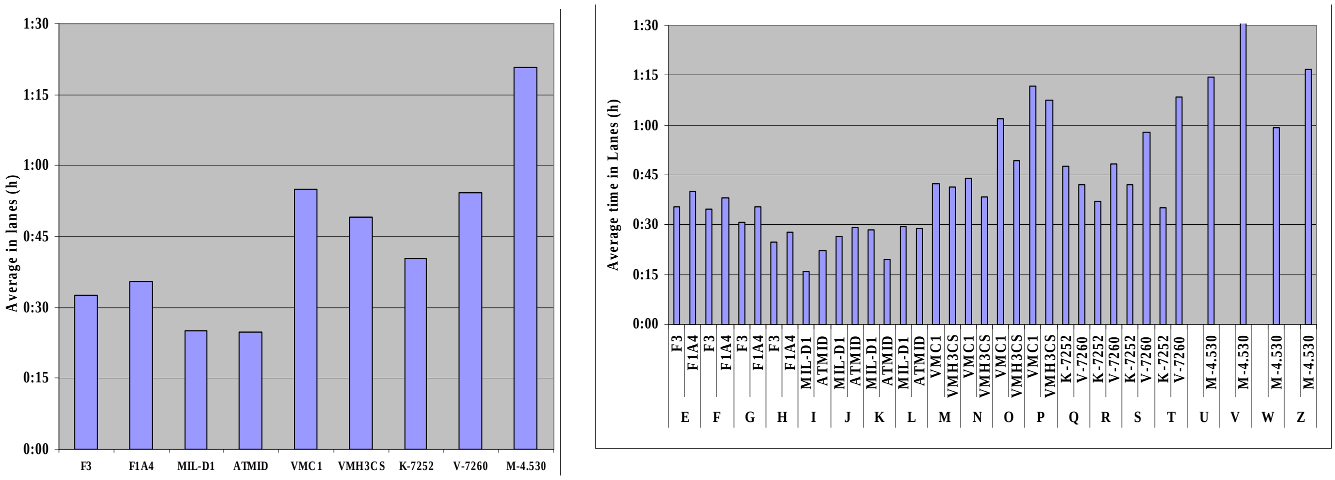


Figure 36 & 37: Average time all detectors, deminers, and all lanes

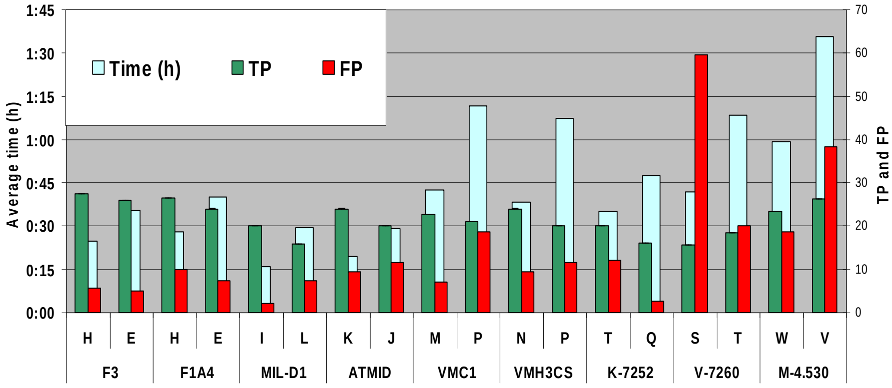


Figure 38: Time needed in connection with operator and detector

False alarms are a serious problem in demining because they considerably slow down clearance operations and also the work in the test lanes. In this trial, the operators achieved very different FARs, mainly because they used different detectors but also within the groups using the same detector. In most cases, deminers with a higher FAR needed more time to complete their work in a lane. In clearance operations deminers spend a significant portion of their time investigating false alarms and excavating metal clutter, so that false alarms slow down their work even more than they do in a test.

The results are based on the individual time of the operators, Figure 37, and in more detail, they are directly connected with the TP and FP signals demonstrated in Figure 38. The left y-axis shows the time (turquoise

---

[12] A field evaluation in Mozambique showed that a deminer can achieve in a minefield a clearance rate of about 25 to 30 square metres following the established rules of the local SOP within 30 minutes if there is no vegetation to cut and no false alarm or mine signal to investigate.

column) and the right y-axis the TP=detected mines (green column) & FP=false alarms (red column). For better overview we included only two operators out of four in each group, the fastest and the slowest. If we take a simple example about the practical effect of the time needed to “clear” the area of 30m^2 then the importance is very obviously. The operator E using the F3 and V using the 4.530 have the same number of TP = 26, E has 5 FP and V 38. The investigation time of a signal will differ from place to place depending on vegetation, ground condition, time of the year etc. The assumption is that the average is about 10 to 15 minutes for one signal investigation independent if a mine or a fragment is later found, the vegetation is already cut down. E will need

26+5x10&15’= 310 to 465 minutes,

when V needs 26+38x10&15’= 640 to 960 minutes.

At the end it would be about one day (or 1.5 days) work in a minefield for E when V would need about two days (or 3 days). In short, the operator V needs twice the time due to the FAR.

Concerning the used time in lanes it has also to be noted that individual abilities of the operators have influence on this factor. Some operators are slowly working some faster and there is no pattern that the slowly working has the better results when using the same detector. On the other hand there is no rule that the faster operator has the better results. Figure 39 shows the ROC diagrams of the same operators as above in Figure 38. There are no significant differences in the POD between the operators using the same detector. This is important because the POD is much more essential than the FAR because it “costs just time” but not life due to a missed mine. There are significant differences in the FAR for two detectors, the Vector and MINEX 4.530, bigger ROC points. The opposite results had been achieved with the detectors. The operators kept statistically the same level of the POD but their FAR was significantly different. The slower Condor operator had a much lower FAR with a slightly higher POD than the faster operator. The slower operator of the MINEX has a very significant higher FAR and a slightly lower POD. Here obviously the human being plays its role.

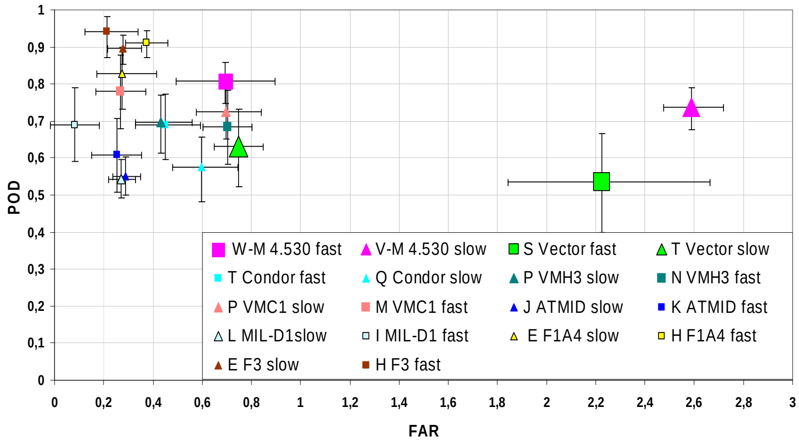


Figure 39: ROC diagrams of slowest and fastest operators

7.5 Operators and their performance with the detectors

This chapter will give the main results of the operators achieved with their detectors as well as limits found in the evaluation of the results. The test results are compared only within the groups and there only within the operators using the same detector. The detector is shown with the same sign and the operator is shown with the same colour for both detectors.

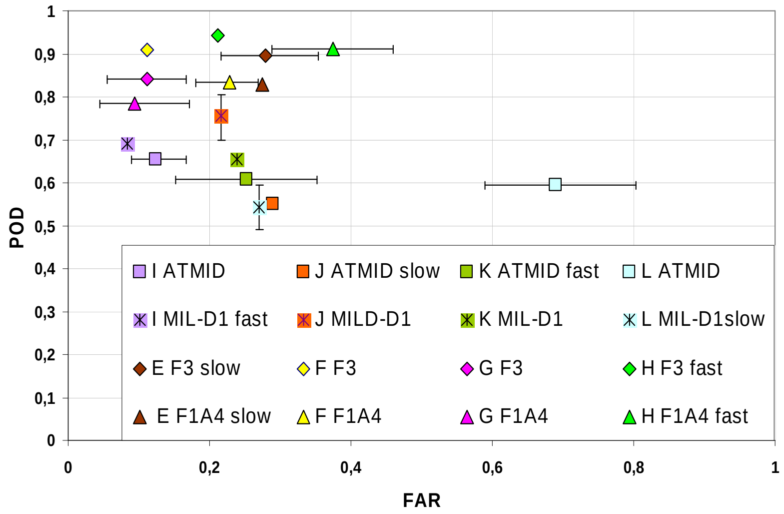


Figure 40:Operator/detector results group 1 and 2

Figure 40 demonstrates the result of the first two groups using the F3 & F1A4 Minelab, Gr.1 and Gr. 2 with the ATMID and MIL-D1. The points with error bars demonstrate cases where the results within the group using the same detector differ statistically significantly. The error bars are placed in accordance with axis they belong to. Within the groups we find both cases of significant differences POD and FAR. Group 1 using the both Minelab have with the increase of POD significant differences in the FAR. They confirm the general rule that an increase of POD will be connected with a higher false alarm rate. The Group 2 used detectors of two manufacturers. The ATMID kept the level of POD but the FAR shows essential differences when the MIL-D1 had significant differences in POD (operator J & L). The opposite happened with the operators of the ATMID, operator I has significant lower FAR in comparison to L. For this behaviour and result it can be assumed that the human factor has its influence. In general the results confirm the above mentioned assessment concerning the time in connection with TP and FP.

Figure 41 shows the results of the second shift. The approach of the demonstration is similar to the description for Figure 38 above. The detector has the same sign, the operator has the same colour when using two detectors and the 5% confidence bars are taken away where there are no significant differences in the results. Please note the different scale of the x-axis in comparison to Figure 40. The picture has changed significantly there are more differences in POD and FAR.

There are two detectors without statistically significant changes in POD the Vector, and the MINEX 4.530. But all groups have within their performance data significant changes in the false alarm rate. The VMC1, VMH3 and Condor have in addition significant changes in POD too.

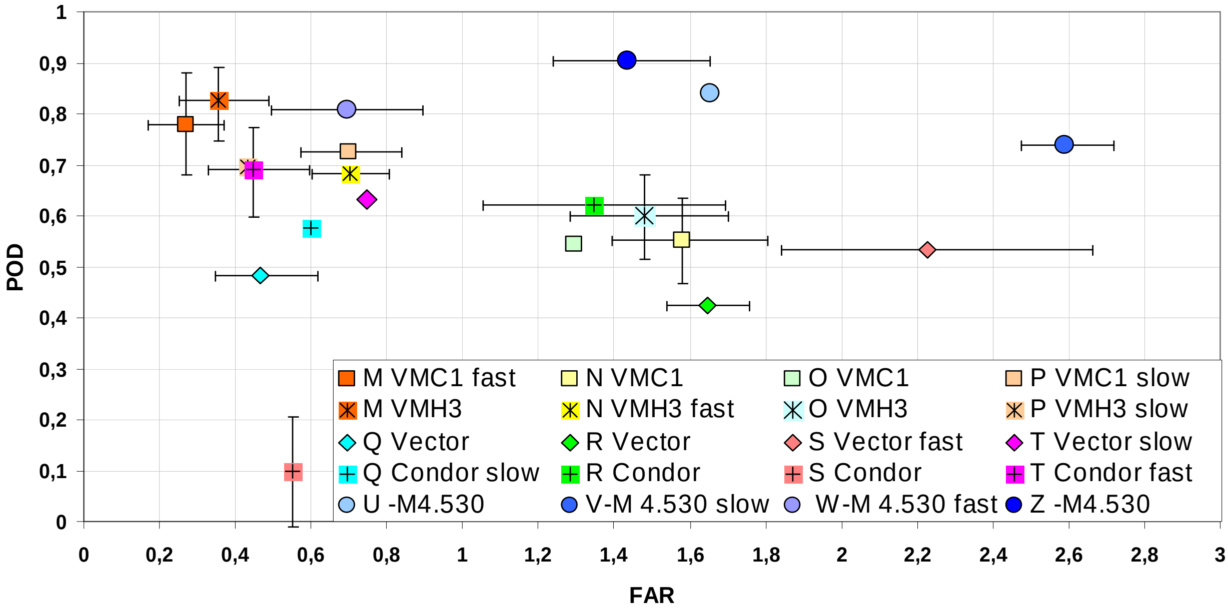


Figure 41: Operator/detector results group 3 to 5

It seems that the operators have a larger influence on the detector performance when the detector has a larger number of possibilities for the setup. The differences between the individual operators' results are in such a situation larger. The technical capabilities of the detector, the targets and environment did not change for the operators. The only change had been the operator himself so he created the very different results from the POD to the FAR. This question will be dealt with in the next chapter.

8. Human factor investigation

Overview

The following chapter is giving the description of the human factor investigation performed during the trial. This investigation is an attempt to see which criteria of the personality characterisation and other factors should be considered for establishing the human influence on detection performance.

The *introduction* gives a short theoretical background about the idea for this research and what it is based on. The *methodology* describes the included tests for investigating of the human factor, and in the *results and discussion* section we elaborate why we got these results, discuss about problems we encountered, and give suggestions for future research.

8.1. Introduction

The idea for this research is based on the following assumption: the performance of the detector in the same conditions is normally always the same. When different persons use the detector in those conditions, the results will differ. The determination of these differences is the goal of the human factor investigation.

The results demonstrated above (Chapter 7) show that the operators can influence on the results of the test in a way that the detector seems, under certain conditions, to be unusable in a minefield. The lack of proper training and the knowledge about its use and other environmental influences might also be the reasons. The aim of this particular investigation was to establish the most important influencing factors of the operators on detector performance.

Among many possible influencing factors, the choice was to investigate the individual differences between operators and the influence of these differences on the results of the trial. The main focus was directed on the personality differences. A personality is defined as a stable set of individual characteristics that make us unique (http://allpsych.com/dictionary/p.html). There are many personality theories, and the one that is currently most popular and very often used in personality assessment was used here, and that is the Five-Factor Model (FFM), more known as the Big Five theory (Digman, 1990). The test which we used to investigate the personality, NEO PI-R, is based on assumption that personality traits can be organized in 30 facets, which then group around five main domains of personality – the five factors. The descriptions of the five main factors and the 30 facets which give us closer and more thorough descriptions can be seen in *Table* 8.

Table 8. *Description of the five personality factors (NEO PI-R)*

FACTOR	DEFINITION	People with high results:	People with low results:	FACETS
Neuroticism	Neuroticism (N) represents a level of adjustment or emotional stability of the person on one side of the scale, and maladjustment or neuroticism, on the other.	-prone to show negative emotions such as anger, guilt, fear, sorrow etc. - have irrational ideas, poor impulse control, poor stress coping abilities and poor tolerance to frustration.	-usually calm, well-balanced, relaxed and cope with stress well.	1. anxiety, 2. angry hostility, 3. depression, 4. self-consciousness, 5. impulsiveness 6. vulnerability
Extraversion	Extraversion (E) is a measure of social adjustment. Primarily, it represents a desirable quantity and intensity of interpersonal relationships.	- warm, sociable and assertive. They seek for excitement, activity and positive emotions; they are full of energy, optimistic, speak fast and are prone to be leaders.	-reserved, serious, distanced, quiet and independent; prefer to be alone rather then surrounded by many people	1. warmth, 2. gregariousness, 3. assertiveness, 4. activity, 5. excitement-seeking 6. positive emotions
Openness	Openness (O) is defined as a need for gaining and questioning experience.	-have a need for diversity, are independent in their judgments and are sensitive to their inner feelings -have vivid imagination, intellectual curiosity, non-conventional attitudes and are very flexible.	-more traditional, conservative and prone to have habits.	1. fantasy, 2. aesthetics, 3. feelings, 4. actions, 5. ideas, 6. values
Agreeableness	Agreeableness (A) represents quality of social interactions from compassion on one side, and antagonism, on the other side of the scale.	-humble, considerate, honest, well-intentioned, willing to help and to cooperate with others having positive expectations from them.	-antagonistic, egocentric, mistrustful and not willing to cooperate	1. trust, 2. straightforwardness, 3. altruism, 4. compliance, 5. modesty, 6. tender-mindedness
Conscientiousness	Conscientiousness (C) relates to a degree of being organized, persistent, in control and orientated to goal achievement.	-well organized, determined, punctual and reliable, ambitious, finishes his/her affairs and fulfils all duties.	-without goals, unreliable, lazy, disorganized, not careful, careless and hedonistic	1. competence, 2. order, 3. dutifulness, 4. achievement striving, 5. self-discipline, 6. deliberation

Previous investigations have shown that a personality has an influence on work performance:

- High conscientiousness and low neuroticism (or high emotional stability) for instance, is considered to be significantly related to high work efficiency (Barrick, Mount & Judge, 2001).
- Investigations on US pilots have shown that high emotional stability is a very important predictor in dangerous occupations (Salgado, 1998).
- Barrick and Mount (1993) discovered that conscientiousness and extraversion are significantly related to job performance.

Therefore, we can assume that personality might have an influence on deminers as well. Personality tests are also widely used in organizational psychology – in selection of workers, follow up of the career, adapting to new working conditions, creating better managing skills, etc. and this justifies the idea to include it into humanitarian demining area as well.

The attention (concentration) had been chosen for investigation too as a very important predictor of good performance. It is defined as efficient, constant and directed selection of stimuli; the ability of the individual to dedicate himself to certain internal and external stimuli and to analyze them quickly and correctly without interruptions and selectively (excluding irrelevant stimuli) (Brickenkamp & Karl, 1986; according to Brickenkamp, 1999). It is logical to assume that deminers need to have high attention abilities to do their job quickly, completely and efficiently.

Previous metal detector trials have shown that the deminers who are currently working and who have more experience in demining show better results (Mueller, Gaal, Scharmach, Ewert, Lewis, Bloodworth, Wilrich & Guelle, 2004). These hypotheses will also be tested in this trial, together with age and qualification (training in demining).

The hypotheses were:
1. the results in trials should increase to a certain age and then start decreasing;
2. operators who are currently working in demining should have better results;
3. more training in demining (qualification) should have a positive influence on the results;
4. more experienced operators should have better results;
5. some personality traits increase and some decrease the POD and FAR (the task is to discover which and what is the nature of their influence); and
6. operators with better concentration skills have higher POD and lower FAR

8.2. Methodology

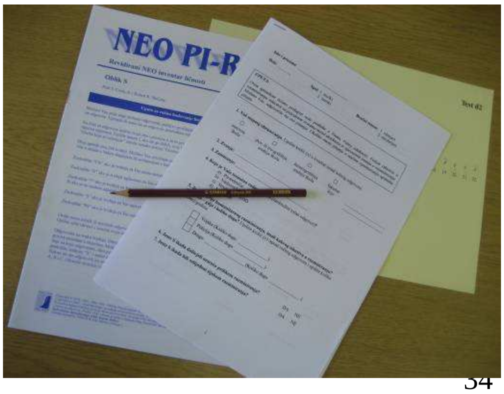
8.2.1. Sample

20 operators agreed to participate in the human factor investigation by filling out the questionnaires. Two of them, due to a lack of some detectors, did not participate in the actual trials so their data could not be analysed. So the sample counts 18 operators (all male, age 25 to 54; mean – 35; standard deviation – 8.36).

8.2.2. Instruments

SELECTION QUESTIONNAIRE

At the beginning of the training, all of the operators filled out a so-called “selection questionnaire” (see ANNEX 3). This questionnaire gives us information about their age, marital status, education, current occupation and current activity in demining, other experience in working with metal detectors beside humanitarian demining, involvement in accidents (or having injuries in accidents), training to become a deminer / EOD specialist / supervisor or to be promoted (qualification), experience in demining / supervising / EOD / managing and, finally, some data about their knowledge about the detectors.



The data on knowledge about the detectors were used to get an overview about the general knowledge deminers have on detectors and on which questions should the attention, during the check after training, be concentrated.

Other data derived from this questionnaire were used for the analysis of the human factor. The aim was to see whether age, current activity in demining, qualification and experience influence on the operators` ability to detect a mine or, in other words, their probability of detection (POD) and false alarm rate (FAR).

This questionnaire also had an important role in the selection of the operators for handling certain detectors, as mentioned in the introduction of this report. The selection we did can be considered as reducing the human factor influence to some extent (each detector being tested by people of various characteristics). If we had more operators, their influence would have been decreased even more.

Reducing the human factor influence can be viewed from two directions – one is reducing this influence in the reliability test, and the other is reducing it in an actual mine field. Doing the selection of the operators as we have done it had been a step in the direction of reducing the human influence in testing of the demining equipment which was indeed the aim of this trial. Further analyses of the age, experience, personality, attention etc. can be considered as investigations in both directions. The difference between the actual field work and metal detector trials held in controlled conditions will be discussed in the following chapters.

NEO PI-R

NEO PI-R by Costa & McCrae (1989; adaptation into Croatian by Naklada Slap, 2005) is a standardized psychological instrument used for the overall assessment of a human's personality. It is consisted of 240 items which measure five main factors or dimensions of personality (neuroticism (N), extraversion (E), openness (O), agreeableness (A) and conscientiousness (C)). Each of these scales has 6 facets which more thoroughly and diversely describe each factor (see Table 8). Items are formed as sentences about oneself and the task of the operators was, for each of the 240 items, to state how much they agree or disagree with the given statements on a five point Likert scale (1 means *strongly disagree;* 2 – *disagree;* 3 – *neutral(neither agree nor disagree);* 4 – *agree* and 5 – s*trongly agree*). It is important to note that having high or low results on any of these scales *do not* represent any disorder. A personality test only aims to *describe* a person in terms of behaviour, thoughts, and feelings (Petz, 1992).

ATTENTION TEST

D2 is a short name for Brickenkamp`s **Test of Attention** (1962). It is mainly focused on visual attention, which is, besides auditory (listening to the sound of the detector), most important considering that a large amount of mines can be actually found just by sight. The task of all participants in this test was to find certain letters among other letters represent distracters in a determined period of time. The operators were instructed to do the task as fast and as punctual as possible. This test gives us data on level of concentration (below average, average, above average for Croatian sample) and the type of attention (for example – highly concentrated, precise, impulsive, etc.). In this investigation only the level of concentration was used. The test is also sensitive to different ways of executing this task. If a person did not behave according to instruction, it might seem that his concentration is very low. But this test recognizes this and therefore makes his result not valid.

8.3. Results and discussion

The main objective of this investigation is to discover whether the proposed influencing factors can predict operators' performance in this trial. Regression analysis was used for this purpose

Influencing factors:

- Age;
- Current activity in demining (currently working as a deminer/ not currently working as a deminer);
- Qualification (low – less than 1 month; middle – 1-3 months; high – more than 3 months);
- Experience in demining (low – less than a year; middle – 1 – 6 years, high – more than 6 years);
- Personality types (N, E, A, O, C); and
- Concentration (below average, average, above average).

Performance measures:

- Probability of detection (POD);
- False alarm rate (FAR).

8.3.1. Statistical analysis of data

To discover which factors have important influences on the operators’ performance and what the nature of their influence is, we need to *compare the operators’ performance* and discover which factors might increase or decrease it.

The problem we encountered was that only four operators used the same detector. In this way a direct comparison was impossible due to different detectors used. These results are shown in Figure 42. To protect the operators´ anonymity in giving their personal data, instead of alphabetical letters which were given to each operator during the trial, in the human factor investigation we randomly called them O1, O2, O3…etc.

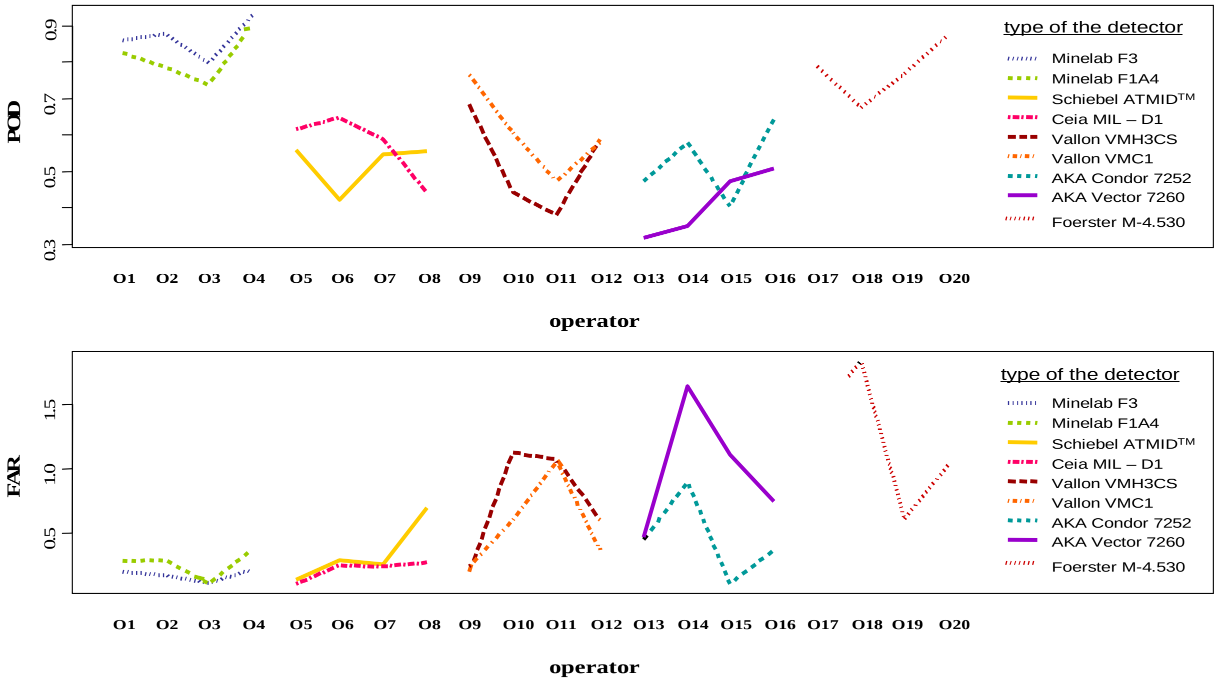


Figure 42: POD and FAR for all operators and all detectors

Figure 42 demonstrates the POD and the FAR, at y-axis, in relationship to the operators, x-axis. The performance of the four operators using two detectors is expressed in comparable lines showing the achieved POD and FAR.

The figure shows several outcomes:

- individual performance of the operators within their group
- the difference in results of one operator using two detectors
- the direct comparison of the detectors and identification of possible interactions between the detector and operator

To investigate which factors cause these differences, the operators should be directly comparable excluding the detector performance. The POD and FAR values had been converted to coefficients according to these formulas:

$$\frac{\overline{X}\mathrm{PODoperator}_{2detectors} - \overline{X}\mathrm{POD}_{2detectors}}{\overline{X}\mathrm{POD}_{2detectors}}$$

$$\frac{\overline{X}\mathrm{FARoperator}_{2detectors} - \overline{X}\mathrm{FAR}_{2detectors}}{\overline{X}\mathrm{FAR}_{2detectors}}$$

The coefficients for each operator represent relative deviation of the POD and FAR of that operator (mean of two detectors he used) from the average POD and FAR of the two detectors he used. These coefficients are shown in the Figure 43 (see ANNEX 5 the complete list of coefficients)**.** Operators who have an average performance will have coefficients close to zero. If a POD coefficient is above zero, than the result of this operator is better than the results of other operators in his group and vice versa. FAR coefficient below zero represents a lower FAR than what is average for his group and vice versa. It is desirable to have a high POD and a low FAR coefficient.

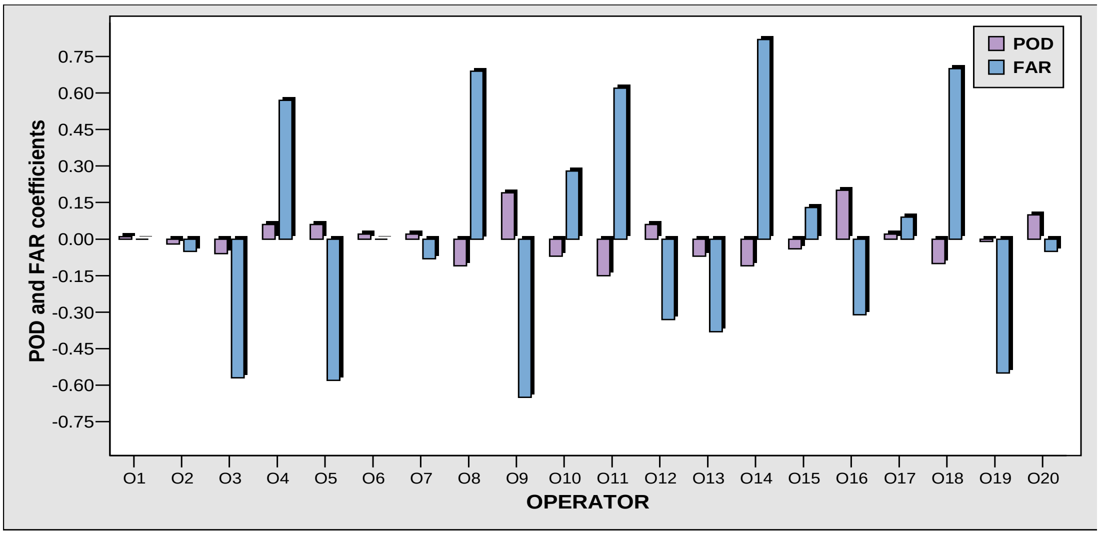


Figure 43. POD and FAR coefficients for all operators excluding the influence of the detecto [13]

The POD and FAR conversion into coefficients gives us the opportunity to compare the operators' results among each other regardless of the detector they used. Additionally, it gives us information which operators performed the best (O16 – the highest POD coefficient, 0.20; O9 – the lowest FAR coefficient, -0.65; O9 - combined best qualities - high POD and low FAR –the best operator), or the worst (O11 – the lowest POD, the highest FAR)). Of course, this information still includes the combined results of two different metal detectors. The problem with these coefficients is that they are based on means of POD and FAR values for two detectors which might have completely different performance therefore making the results average, which might not give us exact picture of how the deminers performed. But since we needed only one data for each operator, this approximation of two detectors was used.

To examine whether *age*, *current activity in demining*, *qualification*, *experience*, *personality types* and *concentration* have an influence on the operators´ POD and FAR a **linear regression analysis** had been performed. This model was used to examine whether these factors can predict the POD and FAR, regardless of the detector type and soil properties. This analysis (see ANNEX 6 for full set of results) has shown that the proposed factors do not have a statistically significant effect on the performance of deminers (POD, FAR).

In the following chapter, possible reasons for not getting the expected effects will be discussed together with some other problems we encountered during this investigation, followed by suggestions for future research.

---

[13] Although the sample for the human factor investigation (the psychological analyses) contains the sample of 18 operators, the POD and FAR coefficients are also added for two operators which participated in the trial but not in the human factor investigation.

8.3.2. Problems and suggestions

a) Design

The trial was not specially designed for the human factor research. It was designed to give us the most useful and reliable information about the detectors. The human factor investigation was added later and performed as an attempt to see which factors might have influence on the results as a function of people who performed it. During the execution and analysis of the data we have faced some problems (e.g. use of two detectors by each operator). Suggestion is to prepare the next trial so that the needs of the human factor investigation can be met in an efficient way too, for example, including the operator as a separate factor. The best approach would be to design a separate trial only for investigating this problem. It would enable the researchers to make different manipulations of the conditions, control undesired influences and make various approaches to the problem.

b) Sample size

Small samples are a problem for all analyses concerning the statistics. When having only 18 people representing the population of about 650 deminers (data obtained from CROMAC) that are actively working in Croatia[14], we can make two types of errors in making conclusions. One is concluding that there are statistical differences when in population they actually don't exist (in statistics known as error type 1, *alpha error*), or saying that there are no differences (for example, between different levels of experience), when in population there are (error type 2, *beta error*) (Petz, 2004). Both mistakes are dangerous. This is why we cannot say with certainty that not getting the expected differences in this investigation means that the differences do not actually exist. Beta error highly depends on the sample size. Since we have only a sample of only 18 people representing the population of almost 650 deminers (data obtained from CROMAC) that are actively working in Croatia the beta error will be large. This is why we cannot say that non-significant test results in this investigation means that the effects do not actually exist. If the sample could be increased in the next trials, using the same methodology (target placement, types of soil, experimental design) it might be possible to draw some conclusions and have statistically and theoretically valid data.

Another problem we encounter is the problem with how the sample is obtained. When we talk about representatives of the sample to the population, it does not only apply on the size of the sample, but also on the way how the sample was chosen. To represent the population of the deminers well, the sampling should be made randomly out of the population of, for example, Croatian deminers (if a trial is made in Croatia). Therewith, the sample would consist of people with different knowledge, experience, age, social status, from different areas of the country, etc. and represent the population in all its characteristics respectively.

c) Choice of measurement

The personality assessment measures are never completely reliable. The results of these kinds of tests are often influenced by the motivation of the participants, and their desire to give socially desirable answers. According to Petz (1992), socially desirable answers are those answers which do not reflect real attributes of the person but instead his desire to show his attributes as socially desirable, which means acceptable, acknowledged and appreciated because of his need to be approved and acknowledged by his social surroundings. Some tests have a scale which can recognize this tendency in answering, but NEO PI-R doesn't. The authors of this test believe that the tendency to give socially desirable answers also reflects the personality of a person which makes it unreasonable then to exclude (Costa McCrae, 2005). The question here is not if the test should be sensitive to this or not, but did we choose a proper test for this investigation? Although, NEO PI-R gives a quite thorough description of a person's personality, maybe a different test should have been chosen – one which would recognize socially desirable answers or one which would be shorter. The time needed to fill out this 240-item test is about 45 minutes. This could decrease operators´ motivation and then resulting in random answering, tendency to give neutral answers (which causes getting average profiles) etc. On the other hand, this test is commonly used and has proven to be very reliable in personality measuring. Its big quality is having 6 facets describing more thoroughly each of the five main factors. If we had a bigger sample, and, with that, actually some variance in these results, it would be interesting to find out more detailed qualities that make us different and which might be interesting in this investigation.

---

[14] Not all operators participating in this trial were from Croatia. A few were from Serbia and Montenegro which makes the size of the population even bigger and our sample even smaller in comparison to it.

There is a strong possibility that this measurement of personality is too wide and not all personality traits which this test measures are equally important for this investigation. Trials like this can give us a deeper insight at what else should be measured, or what, of the already measured, is of greater importance. Therefore it is suggested that in future research a more specific measures are chosen, perhaps using only a few scales of a bigger test, such as this. This way we can get more useful and specific information on the problem we are investigating, excluding the ones which are irrelevant.

The attention test we used is primarily designed to measure visual attention. As already said, it is important considering that in a real minefield a lot of mines can be spotted visually. But in trials like this, auditory attention is of bigger importance; therefore the choice of this measurement is also questionable.

d) Pre-selection

Another possible reason why we didn't get the expected results is the pre-selection of deminers for the demining occupation. Although in most of the mine-inflicted countries there are no special demands if someone wants to work in demining, in Croatia, there are stringent criteria for all organizations which say that all applicants must have (Lardner, 2005):

- A high school education
- Completed military service
- No criminal record
- Good physical and mental health; and
- Attended a special Police Academy training course (6 months) and successfully passed the final examination.

The "mental health" examination is usually performed by psychiatrists and psychologists, and it includes tests, one of which is often a personality test. Not only are they pre-selected with the personality test, they are also pre-selected by their desire and motivation to do this job. It is assumed that a person who wants to work in this rather dangerous and demanding occupation is already motivated, reliable, calm, deals with stress well, and more of those qualities we think a good deminer should have.

e) Standard operating procedures (SOP)

There are very strict rules which must be obeyed in a controlled situation, such as an operation of mine clearance. When behaving according to the SOPs, there is not much space for individual differences, and not much is given to the operator to do outside the limits proposed by the SOPs. Therefore, it is also reasonable to assume that in such a controlled situation, personality, for instance, would not play an important role. It is suggested, therefore, to focus on management, for instance, which has an influence on how the SOPs are performed and how well the operators follow them.

f) Detector influence

Although using the POD and FAR coefficients can solve some of our problems and make the comparison between operators possible, we must take in consideration two things. First, by creating these coefficients, statistics can give us a way to ignore the type of the detector used to some extent; but it still has an influence on these coefficients. If the detector is not functioning well or is not understandable to the operator, it can cause frustration to which not all people are equally resilient, and then cause poorer results in comparison to others. Russian detectors (AKA Condor and AKA Vector) were, for instance, quite difficult for the operators to understand - mostly because of lack of training by the manufacturer, but also because of a totally new way of presenting signals to the operators. Some operators are not open to new things, or this not-understanding could really be frustrating, therefore causing poorer results than they would be if a situation was that they have had a proper training and better understanding of the equipment.

Next, some operators handled two different detectors, but the coefficients shown in Figure 43 are the approximation of the both which still decreases or increases the operator's performance with one of two detectors in order to give us data to compare to others. Because it was not possible to do the analyses with more than one data for POD and FAR for each operator, the statistical analyses were performed on this data. Different results for each operator and each detector are shown in Figure 44 and Figure 45.

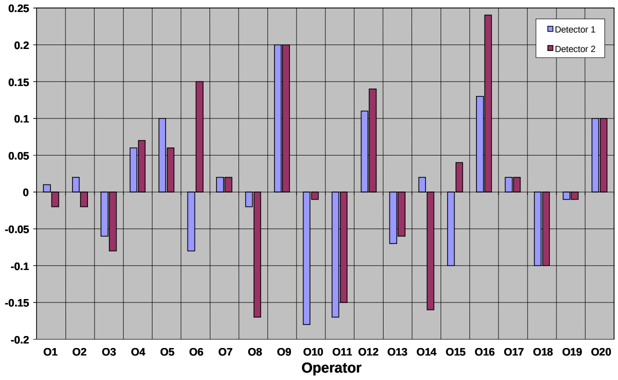


Figure 44. POD coefficients for all operators and bo detectors

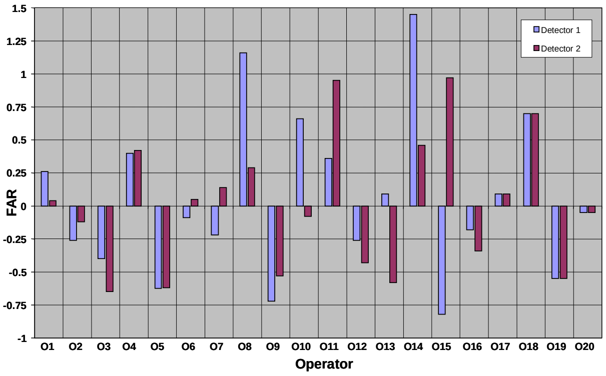


Figure 45. FAR coefficients for all operators and bot detectors

This is why it is suggested that, that for the purposes of the human factor investigation a separate trial is made where, ideally, all operators go through all soil types with the same detector. This would give us a good opportunity to reliably compare the individual performance and to see which factors might have influence on performance. Another suggestion is to take only the results with one detector to avoid problems like with operator O15 who has opposite results with two detectors which is then equalized and identified as average if we take the mean of the results with these both detectors. The problem is the researcher’s bias – which detector result to use?

In praxis, having all operators going through lanes with one detector, just for the human factor investigation often creates a problem, in money and time. There is often not enough time to train all operators to work with the same detector. Maybe, one detector that is already in use with the majority of operators can be used for this investigation, for the desired quality in human factor investigation is not the detectors but the people who use them. This option is good if the human factor is not the main aim of the trial. If it was human factor that is of most interest, including the test of new technologies and the adaptation of people to this technology, a comprehensive trial should be made, with the operators as an additional factor, together with the soil, detector type, target type etc.

8.3.3. Other possible human factor influences

During the trial, there was a chance to observe the deminers, and in this section we will discuss some other possible influences on the results of the tested metal detectors. It is not only important to recognize what the human factor influences are. It is also of great value to make some changes in the planning of the trials in order to give the operators a chance to do their best.

a) Training

From the deminers´ point of view, training to learn how to work with the detectors (two days for one detector) was sometimes too long. Although it is unlikely that this can cause bad results, it can contribute to the loss of motivation, and boredom. For some detectors the training was not done appropriately, by the manufacturers, which might have decreased the quality of training and therefore their performance, as well.

The quality of training is also of big importance. It should be performed by properly trained people by the manufacturers. In this trial we had problems with obtaining the proper training for some detectors. Although, the training for these detectors had been made by people with experience in demining, not having a proper training might have affected the results. From the human factor perspective – the emphasis should not only be put on the operators – the performers of test, but also on the performers of the training. They might differ in the quantity of knowledge, but also in their ability to transfer their knowledge well. These differences could have a big influence on the results as well, and that should be considered in the selection of people who will perform the training. Therefore, it should be ensured that the training is performed by specially educated people with good communication skills with the adjustment of time needed for the training according to the needs of the deminers.

b) Motivation

Motivation should also be considered as an important influence in trials like this. The atmosphere in the metal detector trials significantly differs from the one in the actual mine field – lack of danger, stressful surroundings, seriousness, etc. which may result in poorer performance in the test field. The goal should be to simulate the atmosphere similar to the one in the real minefield, where operators are highly motivated to perform well for saving their life, and the life of others.

In trials, it is not the goal to create danger but to maintain the motivation which is similar to one in the real minefield. For example, it could be done in the following way: Huge motivators in all occupations are achievement, recognition and money, or certain privileges. If we create a situation which involves public recognition of achievement or a certain reward for it, we can increase the motivation of the operators to behave according to the instructions of the trial. For example, at the beginning of the trial a reward for the best operator can be announced. This reward doesn't have to be financial; it can also be in a form of privileges – a vacation, or similar. But in order for this to work, the competition[15] must be public; all the operators must see the results of everyone; and the prize must be attractive to all the participants.

Feedback is often mentioned as a good motivator. If the operator would be given instant feedback about his performance, after inspecting each lane, it might increase his motivation next time to do better. Although this idea is feasible and could easily improve the operators' results, it cannot be performed in a blind trial like this. The rules of a blind trial strictly prohibit giving any suggestions to the operator how he should behave in a lane or give information on how many targets are buried. By giving him information about his performance (for example, "you missed 10% of the targets", or "you made 5 false alarms") we could risk that the operators might calculate how many targets are buried therefore compromising the whole idea of a blind trial.

The question on how to increase the motivation in a reliability trial still stays open, and an important issue to be further discussed.

c) Selection of operators

Although the selection of operators for each detector type, in this trial, was carefully planned, with such a small group of deminers, not all desired solutions could be achieved. It turned out that in all detector groups, except for Vallon detectors, we placed people with experience, varying from 1 to more than 10 years, and the only two operators with a very few or no experience were placed in the Vallon detector group. The reason for this was that most of the deminers in Croatia have already worked with different Vallon types.

Figure 46 shows the results of operators in the Vallon group of detectors. There is an obvious difference between the results of operators O9 (3 to 6 years of experience) and O11(less than one year/no previous experience) which might be a result of difference in experience. O12 shows good results in spite of his lack of experience (less than 1 year/no experience) which may be the result of different factors which might have greater influence on the results than experience does.

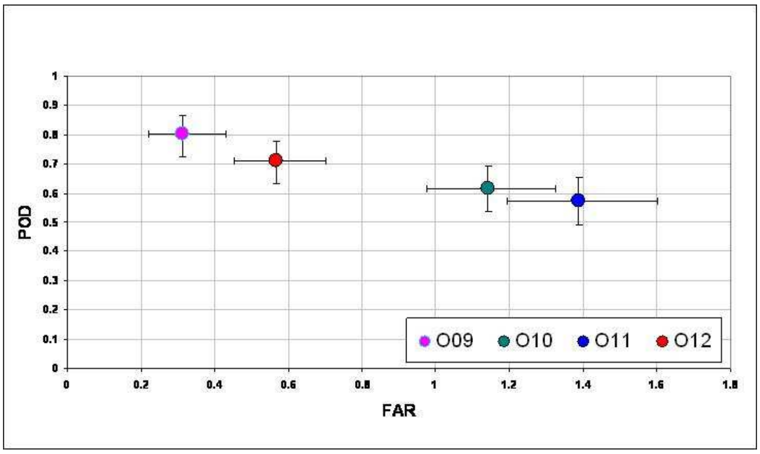


Figure 46. Comparison of operators´ performance with Vallon detectors

d) Management

If the assumption that the deminers are already pre-selected for this occupation is true, or that personality has no or just a small influence on the results on this trials, we need to look for other ideas for possible influencing factors.

---

[15] Note that competition in the real minefield is strongly discouraged and not desirable.

A lot of work, in demining and other occupations, has been done in investigating the influence of management on the performance (e.g. Lardner, 2005) which might indicate that this influence might be stronger than the individual differences between the operators. The performance of operators under different conditions caused by the management can be viewed from the psychological aspect, as well. So far, it has been noticed that operators in trials behave differently if their supervisor is present. Therefore, it is suggested for further research to focus on management and motivational influences.

e) Cognitive processes

Personality is not the only measurement of individual differences. There are many factors in which we differ, and maybe the personality, as a measurement, is a too general approach to investigate these differences. Most of the work in psychology in this direction has been done in investigating the differences in cognitive processes – human information processing, decision making, signal detection, etc. Making a research in this direction would give us useful and more reliable information on the influencing factors we are searching for.

Many human factor investigations in non-destructive testing (e.g. the PISC III study[16]) have shown that that there are no clear correlations between *single* human factors and inspection performance. This is why the approach to human factors should be more extensive (involving cognitive, perceptual, social, organizational and technical influences).

8.3.4. Application of experimental results in “real life”

One of the biggest problems is the application of the knowledge gained from research (human factors; the results about the performance of the metal detectors) into the “real life” situation. Different behaviour of the operators in reliability trial was already mentioned (e.g. poorer motivation). Sometimes, people tend to do their best when knowing that they are being tested. They also know that there are targets buried which might make them try harder to find all the targets. NDT operators (manual ultrasonic inspection in nuclear power plants) indicated that the way of inspection they use in experiments was unrelated to the way of inspecting that they would use during an actual inspection (Wheeler et. al, 1986; according to Enkvist, 2003). Further, experiments are usually performed in a context of relatively low complexity. This leads to the conclusion that generalizing from experiment to reality should be taken with great caution.

8.4 Conclusions

The investigation of the human factor, performed during the metal detector trials held in Benkovac, in October 2006, did not give us the answers we needed - to discover which factors cause different performance of different operators. The analysis showed that age, current activity in demining, qualification, experience, concentration and personality have no significant influence on the results of this trial. This could be a result of numerous factors: design of the trial, sample size, choice of measurement, problems with handling two different detectors, pre-selection of the operators, SOPs etc.

It is suggested to continue with this investigation (same methodology, POD and FAR of metal detectors; but with an increased sample) to be able to make more reliable conclusions, for the influences we have measured.

This investigation has also given us ideas about other possible factors which should be considered in future investigations – influence of training, management, motivation and cognitive processes; and the way we should perform the investigations (design of the trial, sample size, different instruments, etc). To get a clearer idea on the human factors, a separate trial should be designed, with a model describing the behaviour of operators from individual, social, technical and organizational perspective.

---

[16] PISC III. Human Reliability in Inspection, Final Report on Action 7 in the PISC III Program, PISC Report 31. Nuclear Energy Agency. 1994.

9. Conclusions and recommendations

- The used methodology is based on the CWA 14747 : 2003 and approved itself by flexible reaction for changes in the participation of detectors and personnel without losing value concerning the statistics. The CWA approach was improved with statistical design and the results presented with ROC diagrams and curves. The used method of data analysis allowed the placement of targets in different depth without grouping them in two or more fixed depth. The stepped depth gives an accurate understanding about the POD relationship in dependence of depth.

- The use of the two total stations made it possible to keep up with the operators speed in marking in six lanes. The accuracy of coordinate measurements was significantly improved in comparison to the manual measurements. The standard error of the coordinate measurements was about ± 6mm and therewith more accurate as required by the standard. Together with the time and personnel reduction for this purpose we strongly recommend to use this approach in similar trials. Over 8300 data sets with more than 96 000 single data had been collected and could be transferred digitalised to the computer for further evaluation.

- Concerning the soil investigation the demining organisations should be aware about the spatial distribution of the magnetic susceptibility. This trial again confirmed the essential reduction of the sensitivity with increasing magnetic susceptibility and its frequency dependence. Simple measuring methods are known and described and must be explained to the deminers that they will use it in their daily work. The Ground Reference Height is a measurement connected with the frequency dependency of the magnetic susceptibility of the soil and that is accurate enough for field use. Additionally the in-homogeneities in the local Benkovac soil created an essential increase of false alarms -for the metal detectors.

- The results of the questionnaires and the tests not belonging to the blind trial confirmed that the operators have in general a reasonable knowledge about the taught detectors. But only few operators were able to define the “safe search head advance”[17] for their detectors to the mines involved in the trial. This might have influenced on the detection results and the relatively high differences in the false alarm rate.

- The results of the test for establishing the maximum detection depth explains in a simple way the connection of this test to a full reliability test (blind test) and the advantage of a blind test. The reliability test is more close to the reality and gives much more possibilities for detection than the establishment of the maximum detection depth. The result is therefore more reliable.

- The results of the blind trial demonstrated that there are three detectors (Foerster MINEX 2FD 4.530; Minelab F1A4 and F3) with the same POD level of about 0.8. The other detectors are grouped about a POD level of 0.5. Five detectors have the FAR between 0.2 and 0.4 false alarms per m^2 (both Minelabs, CEIA, Schiebel, and AKA Vector). Although having a good POD the Foerster detector has with 1.2 the highest FAR.

    - The results of the detectors differ clearly in the three available soil types where the local Benkovac soil created the greatest problems with its in-homogeneities.

    - The importance of the false alarm reduction is clearly visible by the time the operators needed to pinpoint and mark the detected targets. Two detectors with the same level of mine detection – one would need about 1 day to search an area of 30m^2 and to investigate the marked signals, while the other would need two days.

---

[17] Term used for defining the advance after the deminer has made a sweep from the left to the right in his lane. It is connected with the sensitivity cone to the target. At the edge of the cone Figure 22 to 24 the advance may be less than the half search head for full cover of the lane area with the sensitivity area of the detector at full depth.

  - Looking at the performance of the operators there is clearly the individual influence of the operators on the results of the detectors visible. The speed of “clearance” does not have a clear influence (good/bad) in one or the other direction on the detector performance. The individual performance of the operators is more clearly visible when the detector gives more choices to change sensitivity and other parameters of the detector influencing on it.

- In general the reliability trial is designed to bring the tested equipment to its limits and therewith the results demonstrate more clearly the advantage in detecting one or the other target. The approach of analysing allows also the evaluation of the POD for real situation placement of targets. This can be in detail analysed by using the ANNEX 7 Test Results, Dependency of POD on Depth where each detector result is shown.

- The first attempt for a more detailed and professional analysis of the human influence on the results of the metal detector trial did not give the expected results. This attempt will be continued by extending the sample size but the idea is to make new trials using more directed approaches describing the behaviour of operators from individual, social, technical and organizational perspective.

- With the reliability trial in Benkovac Croatia the STEMD project is formally finished. But there is no follow on project that will keep the demining community updated about new detectors and their performance under different circumstances as laboratory, soil and other environmental conditions. Already now the authors had been contacted to include detectors from Germany, China, Japan, Russia, into future evaluations. It is recommended that ITEP should find an approach to make such information further available. Some of the detectors are with interesting characteristic worth to be tested in detail in the field and in lab conditions. This includes target learning and discrimination of different metals.

10. References

[1] Vlada Republike Hrvatske. Plan razminiranja za 2006. godinu, 5 April 2006. Available at http://www.hcr.hr.

[2] International Campaign to Ban Landmines. Landmine Monitor Report 2005. Available at http://www.icbl.org/lm/.

[3] SEEMACC, "Minutes of the 15th Meeting of the South-Eastern Europe Mine Action Council held in Bijela, Montenegro on 13th October 2005," www.see-demining.org, accessed 20 March 2006.

[4] D. Guelle et. al "**METAL DETECTOR TRIALS;** DETECTOR TEST RESULTS AND THEIR INTERPRETATION" (2006 JRC ISPRA)

[5] Gaal, M. Trial Design for Testing and Evaluation in Humanitarian Mine Clearance. PhD thesis submitted to BTU Cottbus, Germany, January 2007. Soon available at www.itep.ws.

[6] Barrick, M. R., Mount, M. K., Judge, T. A. (2001). Personality and job performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, **9**, 9-30.

[7] Barrick, M.R. & Mount, M.K. (1993). Autonomy as a moderator of the relationships between the big five personality dimensions and job performance. *Journal of Applied Psychology, 78,* 111-118.

[8] Brickenkamp, R. (1999). *Test opterecenja paznje: prirucnik za Test d2.* Jastrebarsko: Naklada Salp (*engl.* Test of Attention: Handbook)

[9] Costa, P.T. & McCrae, R.R. (2005). *Revidirani NEO inventar licnosti (NEO PI-R); NEO petfaktorski inventar (NEO-FFI): prirucnik*. Jastrebarsko: Naklada Slap (*Engl.* NEO personality inventory – revised (NEO PI-R), NEO five factor inventory (NEO-FFI): handbook)

[10] Croatian Mine Action Centre (CROMAC). A list of licensed companies in humanitarian demining in Croatia. Available at http://www.hcr.hr/razminiranje_download/20061124100008_1.pdf

[11] Dalton, M. & Wilson, M. (2000). The relationship of the five-factor model of personality to job performance for a group of Middle Eastern expatriate managers. *Journal of Cross-Cultural Psychology, 31 (2),* 250-258.

[12] Digman, J.M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of Psychology, 41*, 417-440.

[13] Enkvist, J. (2003). *A human factors perspective on non-destructive testing (NDT). Detection and identification of cracks.* Doctoral dissertation submitted at Stockholm University, Department od Psychology.

[14] Jackson, C.J. & Corr, P.J. (1998) Personality-performance correlations at work: individual and aggregate levels of analyses. *Personality individual differences, 24 (6),* 815-820.

[15] Lardner, T. (editor) (2005). *A study of manual mine clearance: 2. the management of manual mine clearance programmes.* Geneva: GICHD.

[16] Mueller, C.; Gaal, M.; Scharmach, M; Ewert, U.; Lewis, A.; Bloodworth, T.; Wilrich, P-T. & Guelle, D. (2004). *Reliability model for test and evaluation of metal detectors – ITEP project 2.1.1.2.*Available at www.itep.ws.

[17] Petz, B. (2004). *Osnovne statisticke metode za nematematicare.* Jastrebarsko: Naklada Slap. (*Engl.* Basic statistical methods for non-mathematicians).

[18] Petz, B. (1992). *Psihologijski rjecnik.* Zagreb: Prosvjeta (*Engl.* Psychological dictionary)

[19] Psychological dictionary, online version. Available at http://allpsych.com/dictionary/p.html

[20] Salgado, J. F. (1998). Big Five personality dimensions and job performance in army and civil occupations: a European perspective. *Human Performance, 11,* 271–288.

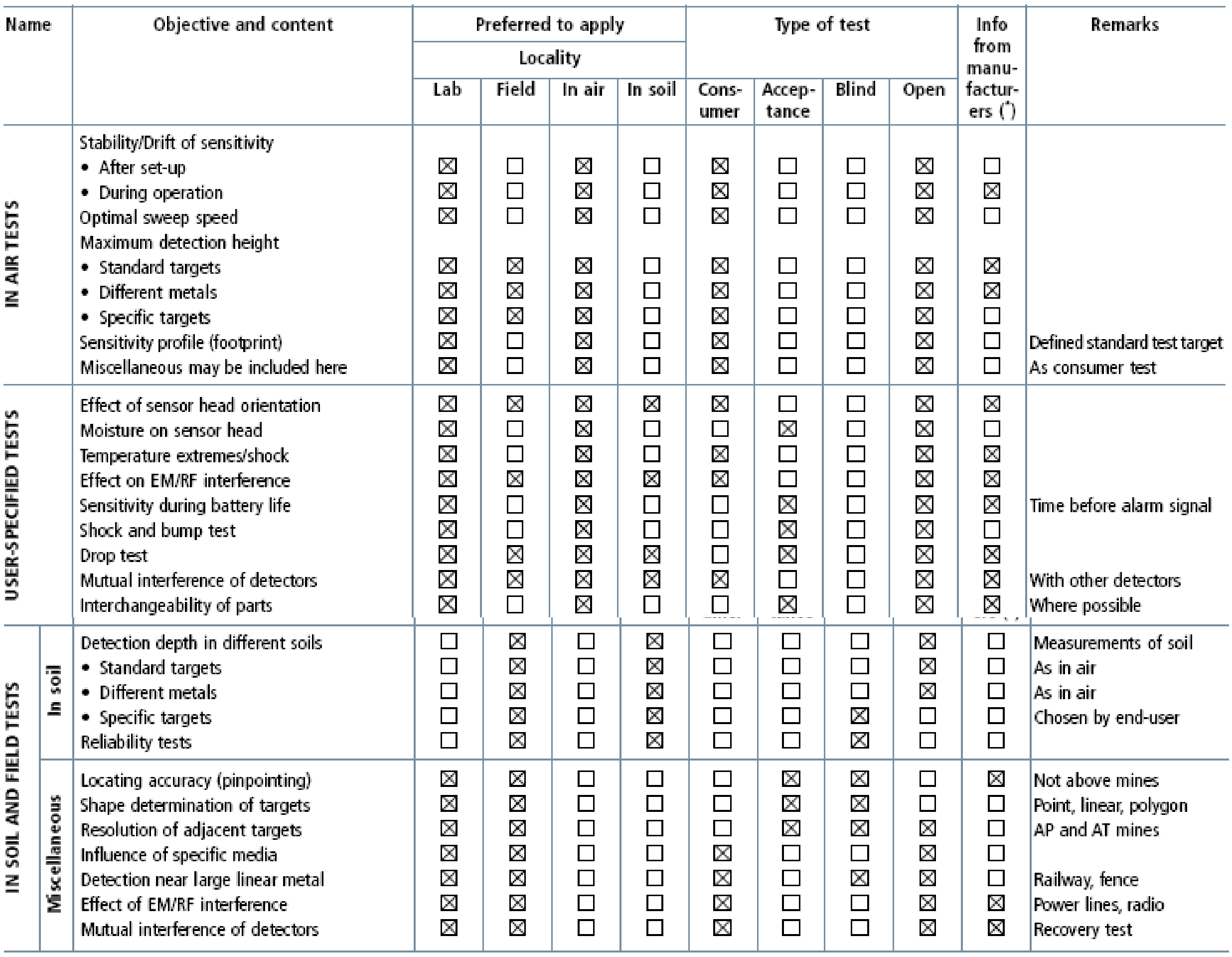
[21][4] D. Guelle, A. Smith, A. Lewis , T. Bloodworth "METAL DETECTOR HANDBOOK FOR HUMANITARIAN DEMINING" HTTP://WWW.ITEP.WS/PDF/METAL_DETECTOR_HANDBOOK.PDF

[22] H. Preetz & J. Igel (Leibniz Institute for Applied Geosciences, Hannover) 2006: Pedological description and magnetic susceptibility of the natural soil nearby the test site of the Croatian mine action centre (CROMAC) in Benkovac, Croatia. Available at http://www.gga-hannover.de/app/gga_appl/produkte_docs/20070116105453.pdf

[23]

ANNEX 1 CWA14747 Test Content

CWA 14747:2003 Test Content



Name		Objective and content	Preferred to apply				Type of test				Info from manufacturers (*)	Remarks
			Locality									
			Lab	Field	In air	In soil	Consumer	Acceptance	Blind	Open		
IN AIR TESTS		Stability/Drift of sensitivity										
		• After set-up	☒	☐	☒	☐	☒	☐	☐	☒	☐	
		• During operation	☒	☐	☒	☐	☒	☐	☐	☒	☒	
		Optimal sweep speed	☒	☐	☒	☐	☒	☐	☐	☒	☐	
		Maximum detection height										
		• Standard targets	☒	☒	☒	☐	☒	☐	☐	☒	☒	
		• Different metals	☒	☒	☒	☐	☒	☐	☐	☒	☒	
		• Specific targets	☒	☒	☒	☐	☒	☐	☐	☒	☐	
		Sensitivity profile (footprint)	☒	☐	☒	☐	☒	☐	☐	☒	☐	Defined standard test target
		Miscellaneous may be included here	☒	☐	☒	☐	☒	☐	☐	☒	☐	As consumer test
USER-SPECIFIED TESTS		Effect of sensor head orientation	☒	☒	☒	☒	☒	☐	☐	☒	☒	
		Moisture on sensor head	☒	☐	☒	☐	☐	☒	☐	☒	☐	
		Temperature extremes/shock	☒	☐	☒	☐	☒	☐	☐	☒	☒	
		Effect on EM/RF interference	☒	☒	☒	☒	☒	☐	☐	☒	☒	
		Sensitivity during battery life	☒	☐	☒	☐	☐	☒	☐	☒	☒	Time before alarm signal
		Shock and bump test	☒	☐	☒	☐	☐	☒	☐	☒	☐	
		Drop test	☒	☒	☒	☒	☐	☒	☐	☒	☒	
		Mutual interference of detectors	☒	☒	☒	☒	☒	☐	☐	☒	☒	With other detectors
		Interchangeability of parts	☒	☐	☒	☐	☐	☒	☐	☒	☒	Where possible
IN SOIL AND FIELD TESTS	In soil	Detection depth in different soils	☐	☒	☐	☒	☐	☐	☐	☒	☐	Measurements of soil
		• Standard targets	☐	☒	☐	☒	☐	☐	☐	☒	☐	As in air
		• Different metals	☐	☒	☐	☒	☐	☐	☐	☒	☐	As in air
		• Specific targets	☐	☒	☐	☒	☐	☐	☒	☐	☐	Chosen by end-user
		Reliability tests	☐	☒	☐	☒	☐	☐	☒	☐	☐	
	Miscellaneous	Locating accuracy (pinpointing)	☒	☒	☐	☐	☐	☒	☒	☐	☒	Not above mines
		Shape determination of targets	☒	☒	☐	☐	☐	☒	☒	☐	☐	Point, linear, polygon
		Resolution of adjacent targets	☒	☒	☐	☐	☐	☒	☒	☒	☐	AP and AT mines
		Influence of specific media	☒	☒	☐	☐	☒	☐	☐	☒	☐	
		Detection near large linear metal	☒	☒	☐	☐	☒	☐	☒	☒	☐	Railway, fence
		Effect of EM/RF interference	☒	☒	☐	☐	☒	☐	☐	☒	☒	Power lines, radio
		Mutual interference of detectors	☒	☒	☐	☐	☒	☐	☐	☒	☒	Recovery test

CWA 14747:2003 Test Content

CRT – consumer report trial – aims to test equipment against standard general standard tests, so that the results are of general interest to metal detector users;
AT – acceptance test – aims to test equipment against specific costumer requirements for purchase decisions

Clause	Test	Testing category							
		CRT	AT	open	blind	lab	field	air	soil
6	*Detection capability testing in-air*								
6.3.3 & 6.4.1	**General test –** **Measuring the maximum detection height**	●	●	●		●	●	●	
6.4.2	**Sweep speed – mechanized movement**	●	○	●		●		●	
6.4.3	**Sweep speed – manual movement**	●	○	●			●	●	
6.4.4	**Repeatability of sensitivity on set-up**	●	○	●		●	○	●	
6.4.5	**Sensitivity drift**	●	○	●		●	○	●	
6.5.2	**Minimum target detection curves for steel balls**	●	●	●		●	●	●	
6.5.3	Minimum target detection curves for other metals	●	○	●		●	○	●	
6.6	**Detection capability for specific targets**	●	●	●		●	●	●	
6.7.1	Sensitivity profile (footprint) measurement - Method 1	●	○	●		●		●	
6.7.2	Sensitivity profile (footprint) measurement – Method 2	●	○	●		●	○	●	
7	*Immunity to environment and operational conditions*								
7.2	Sensor head orientation and shaft extension	●	○	●		●	○	●	
7.3	Moisture on sensor head	●	○	●		●	○	●	
7.4	Temperature extremes	●	○	●		●	○	●	
7.5	Temperature shock	●	○	●		●	○	●	
7.6	Sensitivity during battery life	●	●	●		●	○	●	
7.7	Effect of EM/RF interference	●	○	●		●		●	
8	*Detection capability for targets buried in soil*								
8.2	**Minimum detectable target as a function of depth**	●	●	●		●	○		●
8.3	**Detection capability for specific targets in soil**	●	●	●		●	○		●

Clause	Test	Testing category							
		CRT	AT	open	blind	lab	field	air	soil
8.4	**Fixed-depth detection test**	●	●	●		○	●		●
8.5	**Detection reliability tests**	●	●		●		●		●
8.6	Additional detection reliability testing	●	●		●		●		●
9	*Operational performance characteristics*								
9.2	**Target location accuracy**	○	●		●	○	●	●	○
9.3	Shape determination of targets	○	●		●	○	●	○	●
9.4	**Resolution of adjacent targets**	○	●		●	○	●		●
9.5	The influence of specific media on detection		●	●		○	●		●
9.6	Detection near large linear metal objects	○	●		●	○	●		●
9.7	Effect of specific electromagnetic interference sources		●	●			●	●	
9.8	Mutual interference between detectors	○	●	●		○	●	●	
10	*Evaluation of ergonomic and operational aspects*								
10.1.1	Shock and bump tests	●	○	●		●		●	
10.1.2	Drop tests	●	●	●			●	●	
10.3	Interchangeability of parts	○	●	●		○	●	●	

ANNEX 2 Table of Trials

Order of international test campaigns

Table 9: Trials under view

Date	Location	Organisation	Comments
January 1997	Sarajevo, Mostar	UN Mine Action Servi (UNMAS)	16 detectors, 11 manufacturers; to provide a *list of detectors acceptable for FRY*, support decision for purchase
1998-2000	Cambodia, Croatia, Canada, Netherlands	IPPTC, US, UK, Netherlands, Canada, E – later ITEP members	28 detectors, 13 manufacturers; to provide a *COTS overview of different capabilities of metal detectors tested under lab and field conditions, soil properties measured (conductivity, susceptibility)*
September 1999 to march 2000	Peshawar, Jalalab and Kabul	MAPA	13 detectors, 8 manufacturers; *soil properties measured - Bartington D*, support decision for purchase
Autumn 1999 & 20 for 5 months time in summary	Maputo, Gaza, Inhambane provinces	UNADP Mozambique	9 detectors, 6 manufacturers; *field trial in minefields and focus on soil influence metal detectors (GRH)*, support decision for purchase
2001	Nicaragua	US-Army	7 detectors, 5 manufacturers; support decision for purchase small-scale trial
February 2002	Jalalabad, Kabul	MAPA, UNOPS, ITEP (inv.)	7 detectors, 7 manufacturers; support decision for purchase
July 2003	Colombia	Defence R&D Canada	5 detectors, 5manufacturers; *First use of a Total Station* support for purchase armed forces
May – Nov 2003	Germany, Croatia	BAM Germany ITEP (inv.)	4 detectors, 4 manufacturers; *Reliability trials based on non-destructive testing and evaluation*
August 2004	Cambodia	CMAC, ITEP (inv.)	5 detectors, 4 manufacturers; support decision for purchase
October 2004	Laos	STEMD JRC, ITEP (in	8 detectors (4 of them UXO), 6 manufacturers; *comparison of UXO and normal metal detectors*, support decision for purchase
April 2005	Mozambique	STEMD JRC, ITEP (in	12 detectors, 8 manufacturers; *overview about current COTS and the influence of soil*
Sep-Oct 2006	Croatia	STEMD BAM, ITEP	9 detectors, 6manufacturers; *newest knowledge about reliability trials*

ANNEX 3 The selection questionnaire

Name and surname: ___________________________

Age: _______ **Gender**: 1. male 2. female **Marrital status:** 1. married 2. not married

INSTRUCTION:

This questionnaire is designed to collect some information about you, your education, your demining experience, experience with certain metal detector types and general knowledge in demining. Please answe to all questions. If you have any questions, please ask.

1. What is your level of education? Put a cross in the square above the correct answer.

☐	☐	☐	☐	
elementary school	2 or 3 years secondary school	4 years of secondary school	university which:	____________ ____________ ____________

2. What is your profession? ____________________________________

3. What is your current occupation? ____________________________________

4. What is your current position?

a) Deminer
b) Supervisor
c) EOD specialist
d) Manager

5. Before humanitarian demining have you had any other experience in demining?
If yes-where and how long? Put a cross in the square in front of the correct answer and write how many months / years.

☐ Military (How long: ______________)
☐ Police (How long: ______________)
☐ Other: ___________________ (How long: _______________)

6. Have you ever been involved in an accident? YES NO

7. Have you ever been injured during mine clearence? YES NO

DEMINING QUALIFICATION

In the next section, we would like to know about your education in demining. If you have gone through education in more than one field, please answer for all of them. Please, put a cross in the square above the correct answer.

8. How long did your education last?

a) to become a deminer

☐	☐	☐	☐
less than a month	1-2 months	2-3 months	3 months or more

b) to become a supervisor

☐	☐	☐	☐
less than a month	1-2 months	2-3 months	3 months or more

c) to become an EOD specialist

☐	☐	☐	☐
1 month	3-6 months	6-12 months	over 12 months

DEMINING EXPERIENCE

In this section we want to know about your demining experience. If you have had experience in more than one position, please answer for all of them.Please, put a cross in the square above the correct answer.

9. How much experience do you have?

a) as a deminer

☐	☐	☐	☐	☐
less than a year	1-3 years	3-6 years	6-10 years	more than 10 years

b) as a supervisor

☐	☐	☐
less than a year	1-5 years	more than 5 years

c) as an EOD specialist

☐	☐	☐	☐	☐
less than a year	1-3 years	3-6 years	6-10 years	more than 10 years

EXPERIENCE WITH THE DETECTORS

In front of you there is a table with a list of detectors. Please, put a cross in a field next to the detectors which you have used, stating how much experience you have had in handeling that detector. For example, you have had experience with CEIA's MIL-D1 of more than two years, and Shiebel's ATMID™ of 3 months put crosses in both of those squares.

			EXPERIENCE IN MONTHS/YEARS				
No.	Manufacturer	Detector	Less than 6 months	6-12 months	1–2 years	2-3 years	More than 3 years
1	**CEIA**	**MIL – D1**					
2	**Ebinger**	**EBEX®421 GC**					
3		**EBEX®420 HS**					
4	**Foerster**	**Minex 2FD 4.500**					
5		**Minex 2FD 4.510**					
6	**Minelab**	**F1A4**					
7		**F3**					
8	**Schiebel**	**ATMID™**					
9	**SHRIMT**	**M90**					
10	**Vallon**	**VMH3**					
11		**VMH3 (M)**					
12	**AKA**	**Condor**					
13		**Medusa**					
14	**Chinese Institute**	**GT 115-2**					

KNOWLEDGE ABOUT DETECTORS

Finally, we have prepared 10 questions to examine your knowledge about handling the detectors. There is always only one correct answer. Please, answer on all questions.

1. What is <u>not included</u> in the daily routine (control of the detector before entering the mine field)?

a) Checking the battery contact
b) Checking the maximum sensitivity of the detector to a mine
c) Checking the connection between the search head and the electronic unit
d) Using the test piece for the control of the set up

2. What does ground compensation mean?

a) Reducing the influence of the electromagnetic properties of the soil on detector performance
b) Reducing the sensitivity of the metal detector to the ground
c) Reducing the loudness of the signal

3. Which are <u>not</u> the working principles of the detector?

a) Electromagnetic induction
b) Pulse induction
c) Continuous wave induction
d) Movement induction

4. What does the test piece of the manufacturer tell you after you have switched on the detector?

a) The detector is at maximum sensitivity
b) The detector is usable in uncooperative ground
c) The detector is functioning as designed

5. What signal the detector <u>does not</u> provide?

a) Baterry low
b) Confidence click (control that the detector is still functioning)
c) A permanent signal
d) Temperature warning

6. How do you establish the safe advance for the detector sweep?

a) By the size of the search head
b) By the sensitivity area of the detector to the target
c) By measuring the size of the mines
d) By assesing the surface conditions

7. What does not influence the detector performance?

a) Eletromagnetic properties of the ground
b) The salt content of wet soil
c) Use of different batteries
d) Human factor

8. How do you pinpoint a target? Choose the type of the search head you have worked with and answer to that question.

- *If you have worked with a double D search head?*

a) by searching crosswise above the signal
b) by approaching from at least 4 directions

- *if you have worked with a single coil search head?*

a) by searching crosswise above the signal
b) by approaching from at least 4 directions

9. How do you establish the depth you can reliably clear to?

a) Using a rendered safe original target and place it on different depths on the ground and establish maximum detection height
b) By mesuring maximum detection height in air
c) Setting up the detector to the ground and measure maximum detection height with the detector set up in the air

10. What is the minimum distance that should be between two metal detectors in a way so that they do not interfere with each other? Write the name of the detector with which you are currently working and write the distance.

Name of the detector: __________________________________

Minimum distance (in cm): ______________________________

Please check that you have answered to all questions.
Thank you very much for participating.

ANNEX 4. Questionnaire about the detectors

INSTRUCTION:
The reason we are giving you this questionnaire is to collect valuable information about the detectors from the people who used it – you! *Please answer to these questions by circling out if you agree (YES) or disagree (NO) with the statement for each of the two detectors you handled.*

Write down the NAMES OF THE DETECTORS here

detector 1 detector 2

1	Is the user manual easy to use and to understand?		YES	NO		YES	NO
2	Is the field card easy to use and understand?		YES	NO		YES	NO
3	Is the detector easy to assemble and disassemble?		YES	NO		YES	NO
4	Is there a risk that you can assemble the detector wrongly?		YES	NO		YES	NO
5	Are the controls easy to understand?		YES	NO		YES	NO
6	Is the start up procedure simple?		YES	NO		YES	NO
7	Is the detector easy to operate?		YES	NO		YES	NO
8	Is the detector easy to adjust for comfort?		YES	NO		YES	NO
9	Is the confidence tone easy to understand?		YES	NO		YES	NO
10	Are the alarm tones easy to distinguish and understand?		YES	NO		YES	NO
11	Is it possible to set the sound level?		YES	NO		YES	NO
12	Are you comfortable with the weight of the detector?		YES	NO		YES	NO
13	Do external cables get in the way?		YES	NO		YES	NO
14	Do you often need to adjust the search head and telescope arm?		YES	NO		YES	NO
15	Is the ground compensating procedure easy to understand?		YES	NO		YES	NO
16	Is it easy to pinpoint a target?		YES	NO		YES	NO
17	Is the detector robust?		YES	NO		YES	NO
18	After complete training would you feel confident with this detector in a live minefield?		YES	NO		YES	NO

19. Do you think it is important to be able to set the sound level? YES NO

20. Do you want to be able to choose between different sensitivity levels? YES NO

21. What is your overall impression of this detector? (descriptive answer):

DETECTOR 1: _________________ (write the name)

____

____

DETECTOR 2: _________________ (write the name)

____

____

ANNEX 5 Comparable deminer coefficients

POD and FAR coefficients

operator	POD coeff.	FAR coeff.
O1	.01	.00
O2	-.02	-.05
O3	-.06	-.57
O4	.06	.57
O5	.06	-.58
O6	.02	.00
O7	.02	-.08
O8	-.11	.69
O9	.19	-.65
O10	-.07	.28
O11	-.15	.62
O12	.06	-.33
O13	-.07	-.38
O14	-.11	.82
O15	-.04	.13
O16	.20	-.31
O17	.02	.09
O18	-.10	.70
O19	-.01	-.55
O20	.10	-.05

ANNEX 6 Full set of results of regression analysis for the human factor investigation

Regression Summary for Dependent Variable: POD
R= .70951763 R²= .50341527 Adjusted R²= -.32422596
F(10,6)=.60825 p<.76780 Std.Error of estimate: .10832

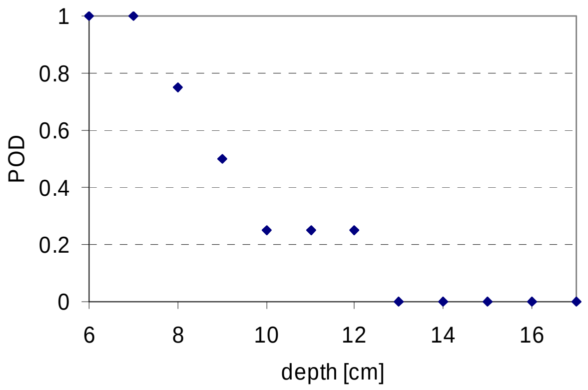
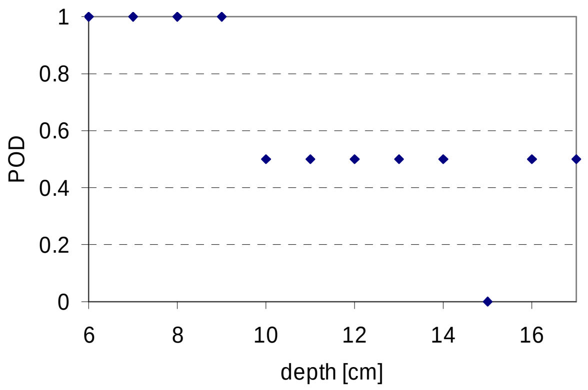
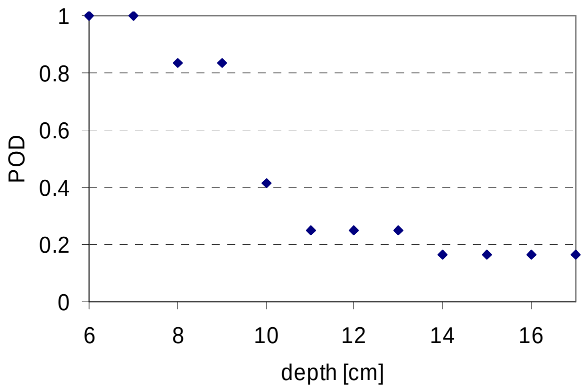
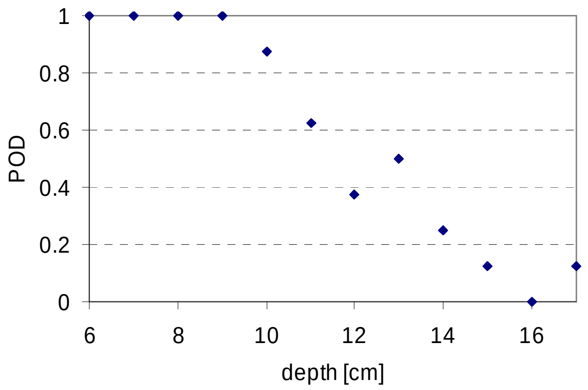
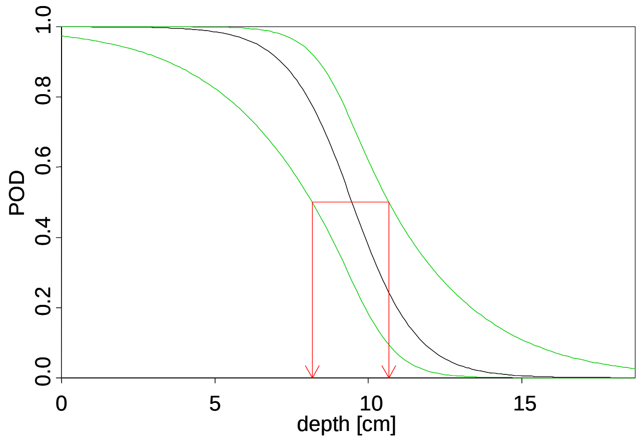
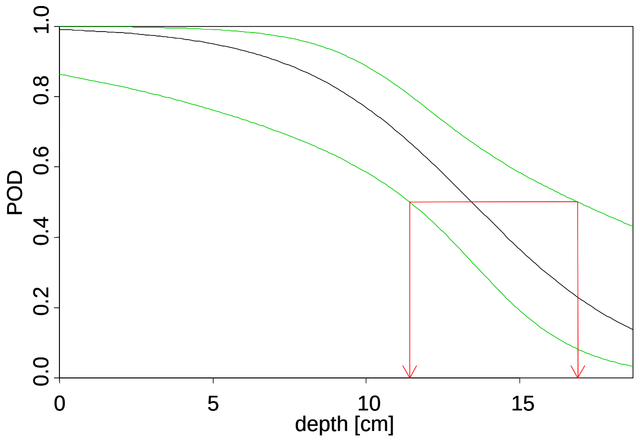
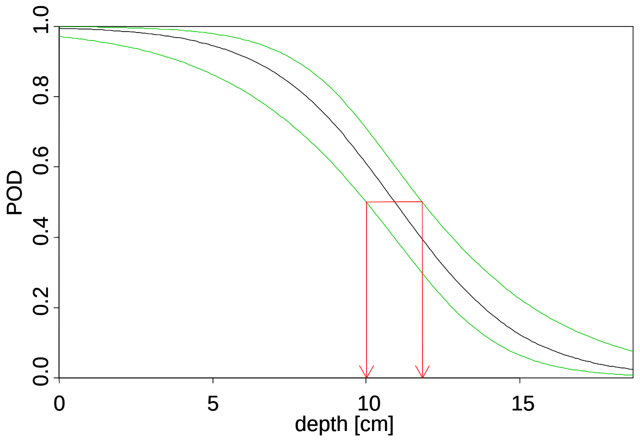
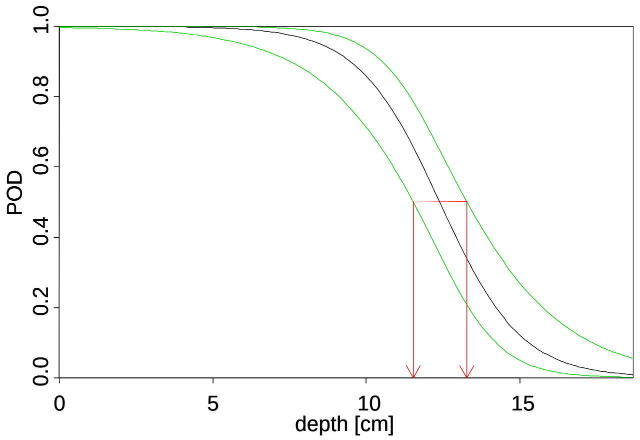
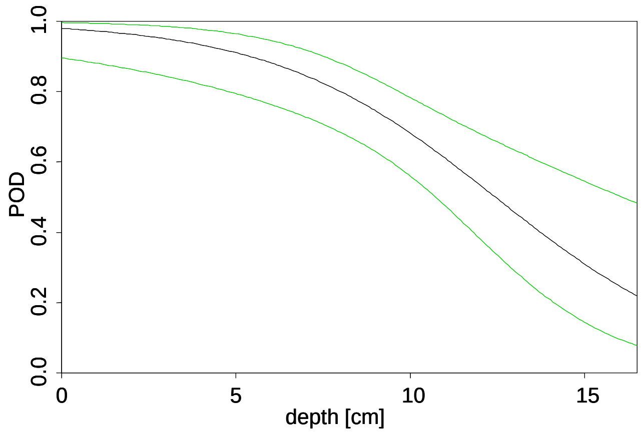
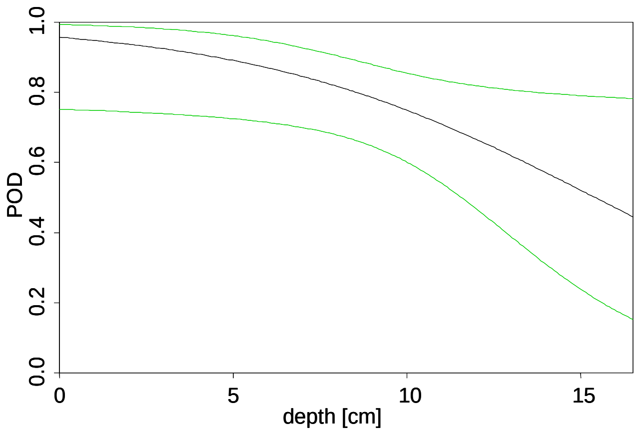
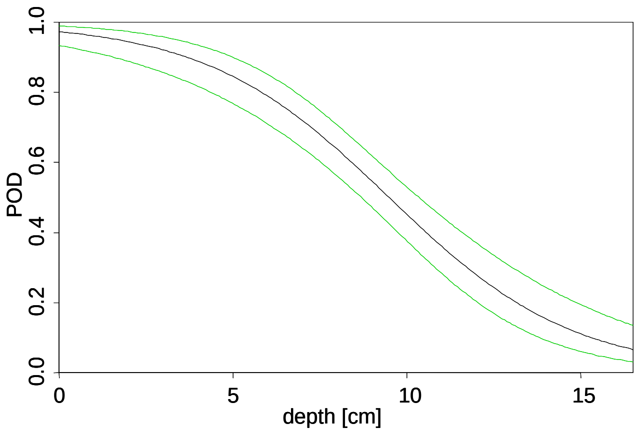
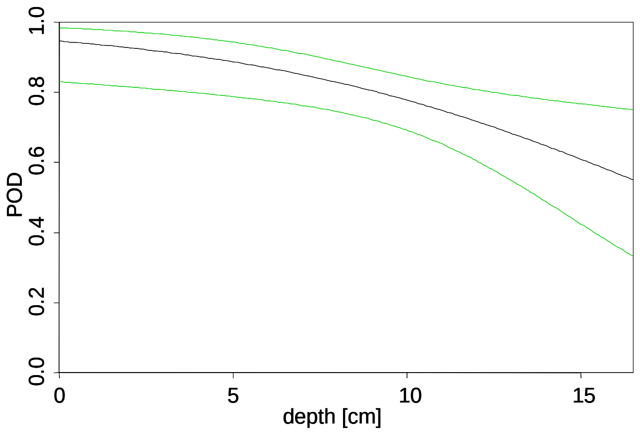
N=17	Beta	Std.Err. of Beta	B	Std.Err. of B	t(6)	p-level
Intercept			-0.281753	0.685746	-0.41087	0.695438
current activity	-0.075771	0.460113	-0.014060	0.085377	-0.16468	0.874604
age	-0.704086	0.457544	-0.009340	0.006069	-1.53884	0.174766
concentration	-0.110902	0.363125	-0.012055	0.039470	-0.30541	0.770372
qualification	-0.241215	0.386943	-0.019468	0.031230	-0.62339	0.555972
experience	0.197926	0.419073	0.016472	0.034876	0.47230	0.653404
N	0.858480	0.537079	0.008363	0.005232	1.59843	0.161064
E	0.199755	0.686103	0.001978	0.006794	0.29114	0.780744
O	0.329190	0.539429	0.003844	0.006298	0.61026	0.564076
A	0.555877	0.720753	0.005344	0.006929	0.77124	0.469835
C	-0.332531	0.609658	-0.004985	0.009139	-0.54544	0.605120

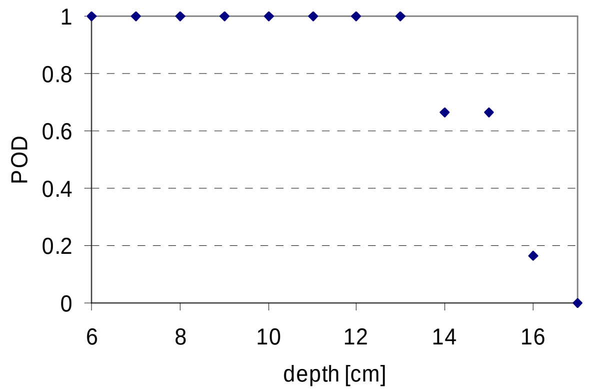
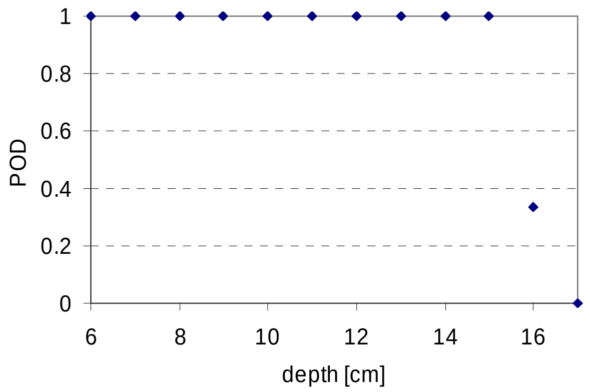
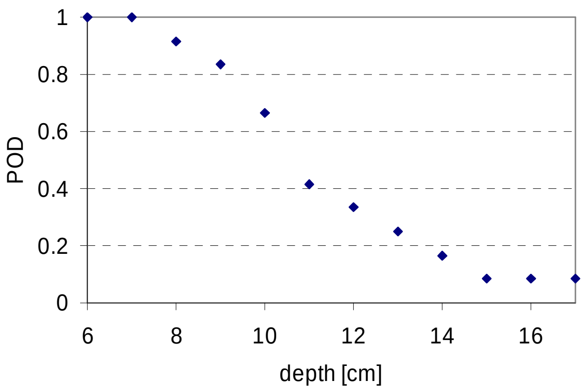
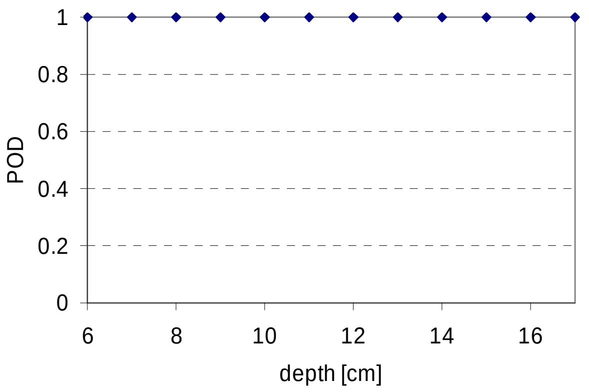
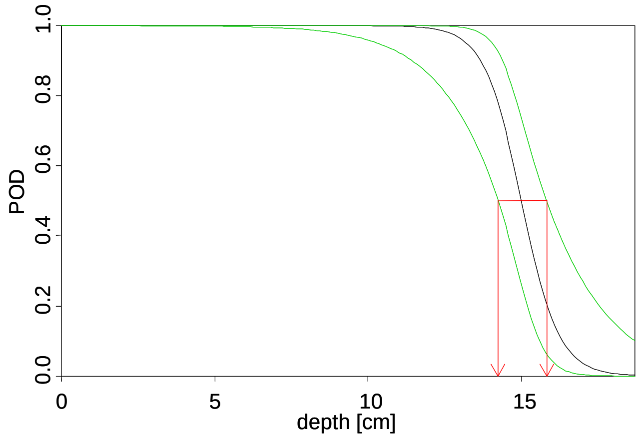
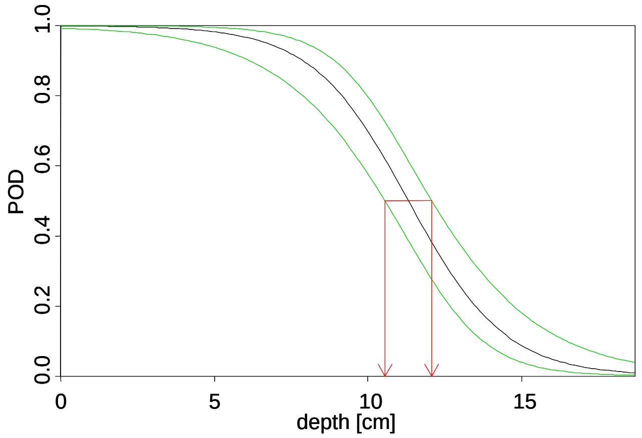
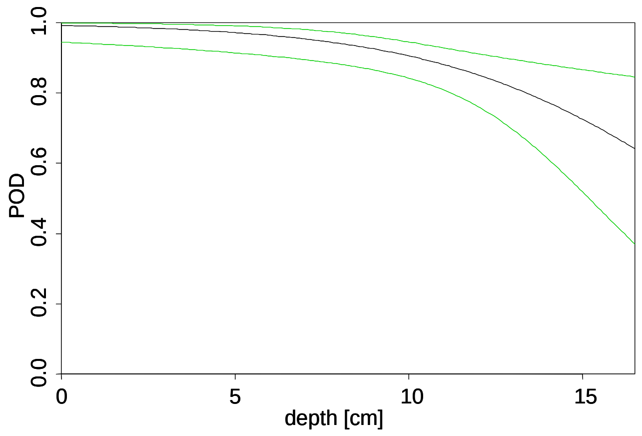
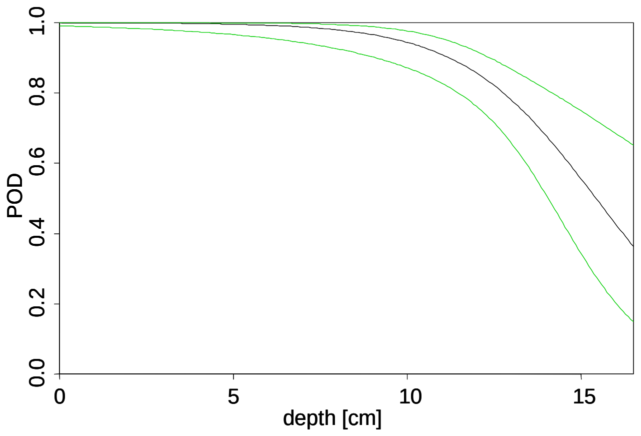
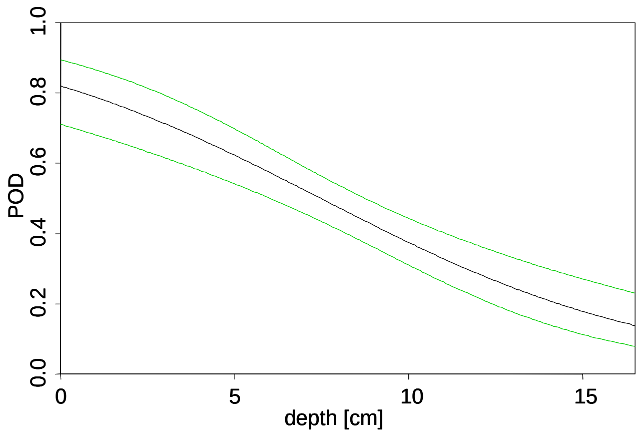
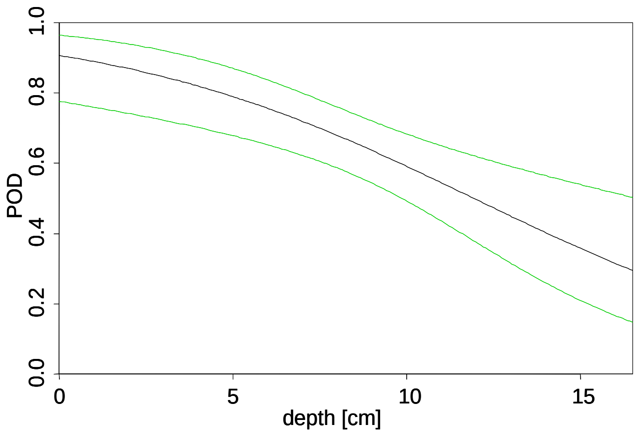
Regression Summary for Dependent Variable: FAR
R= .81491886 R²= .66409275 Adjusted R²= .10424734
F(10,6)=1.1862 p<.43529 Std.Error of estimate: .44102

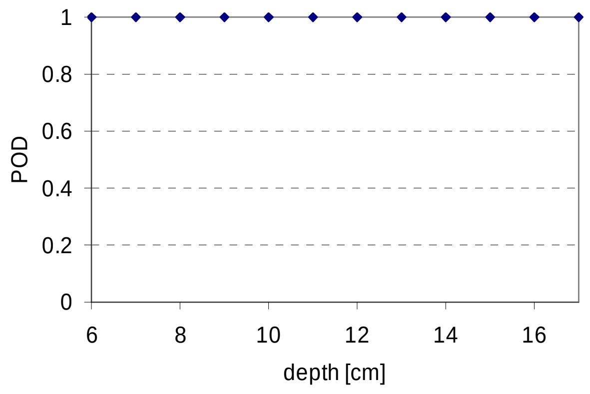
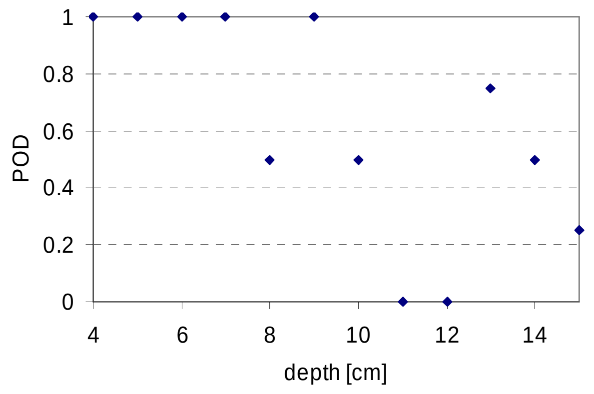
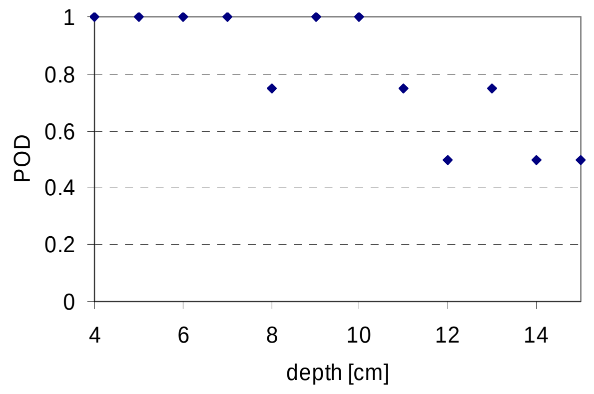
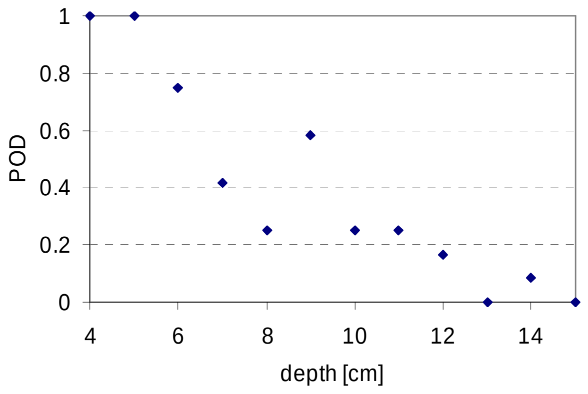
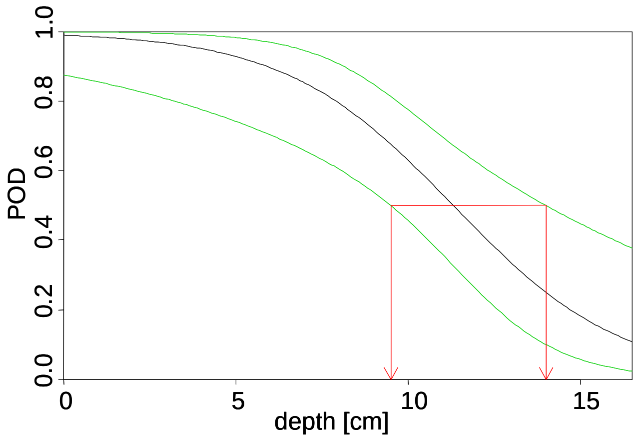
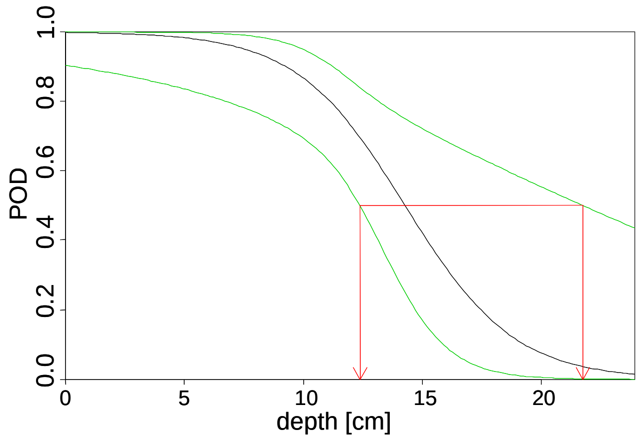
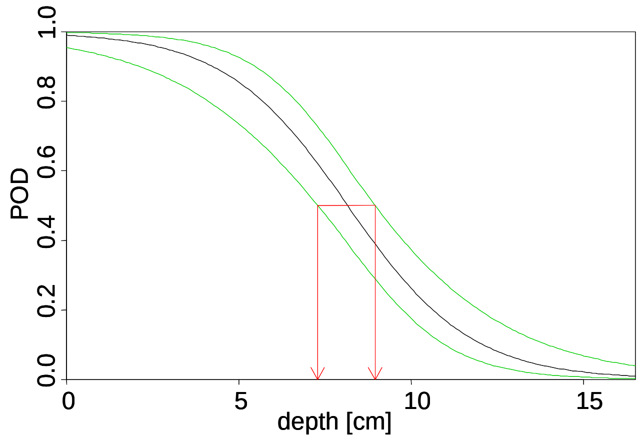
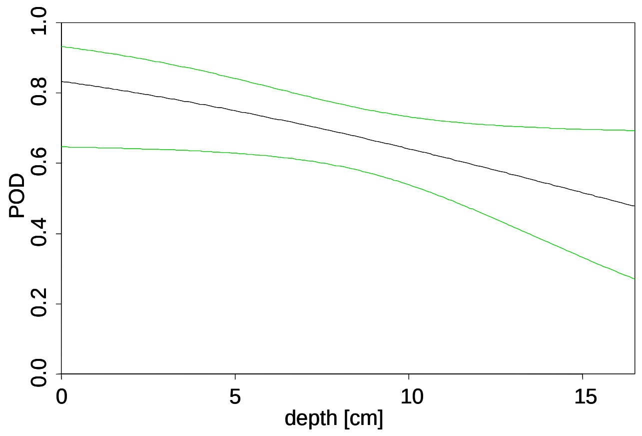
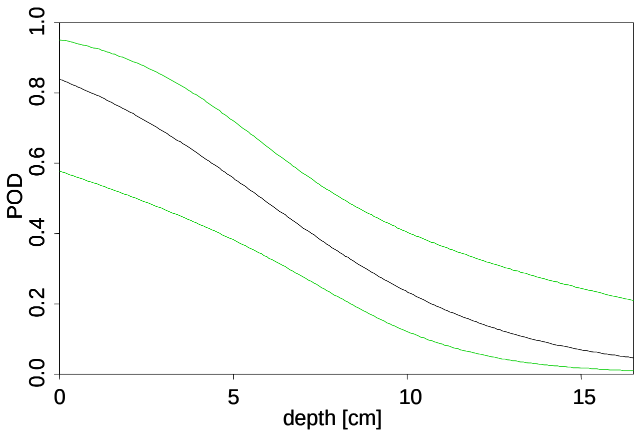
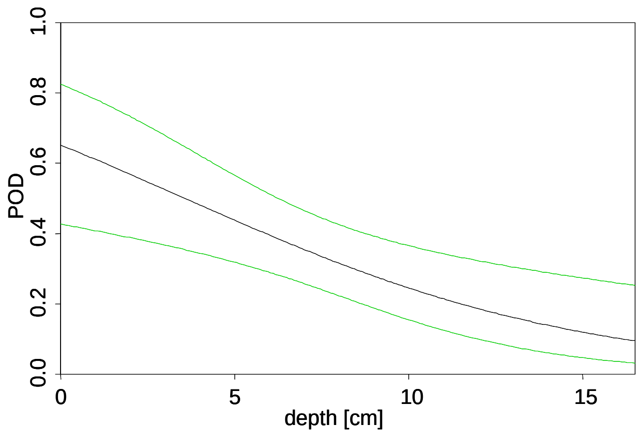
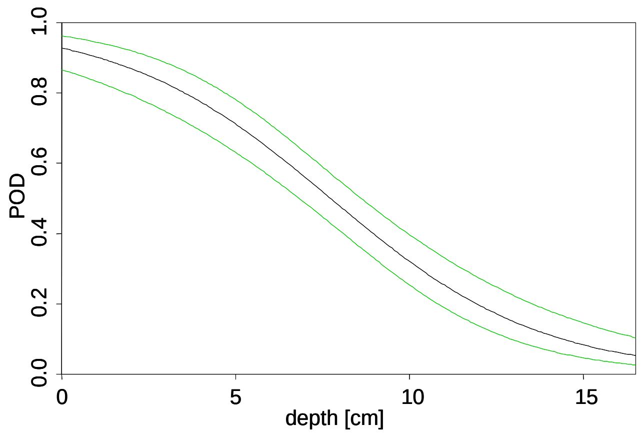
N=17	Beta	Std.Err. of Beta	B	Std.Err. of B	t(6)	p-level
Intercept			-0.255097	2.791875	-0.09137	0.930172
curent activity	0.58781	0.378423	0.539924	0.347596	1.55331	0.171340
age	0.57268	0.376310	0.037603	0.024709	1.52182	0.178878
concentration	0.13960	0.298654	0.075114	0.160694	0.46744	0.656682
qualification	0.10913	0.318244	0.043600	0.127147	0.34291	0.743361
experience	-0.29698	0.344669	-0.122343	0.141990	-0.86163	0.421985
N	-0.74109	0.441724	-0.035737	0.021301	-1.67772	0.144411
E	-0.99408	0.564290	-0.048730	0.027661	-1.76164	0.128603
O	0.61774	0.443657	0.035703	0.025642	1.39238	0.213221
A	-1.01565	0.592788	-0.048330	0.028208	-1.71335	0.137482
C	0.93931	0.501417	0.069701	0.037207	1.87332	0.110171

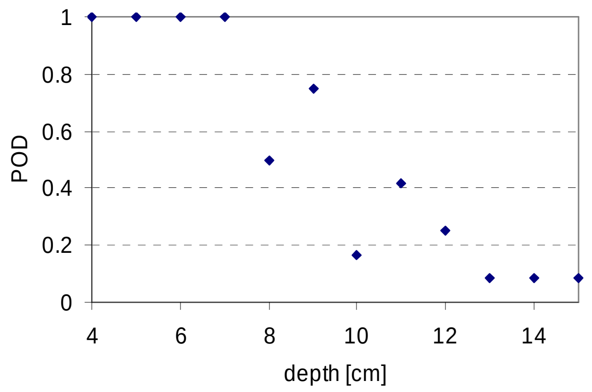
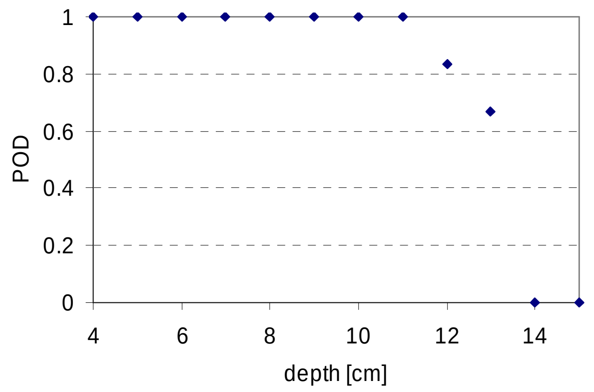
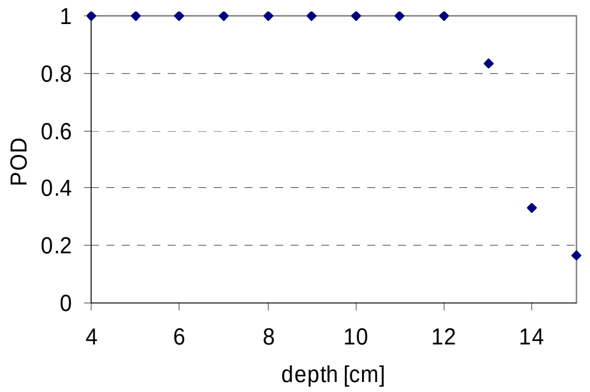
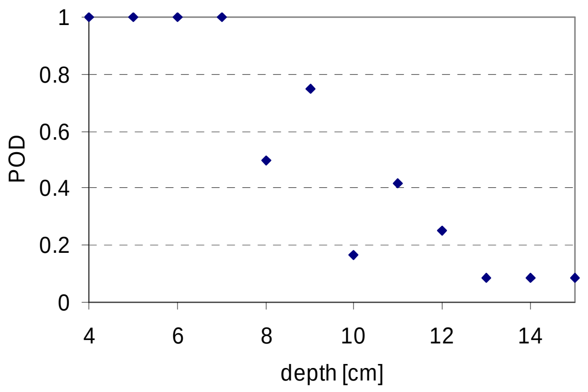
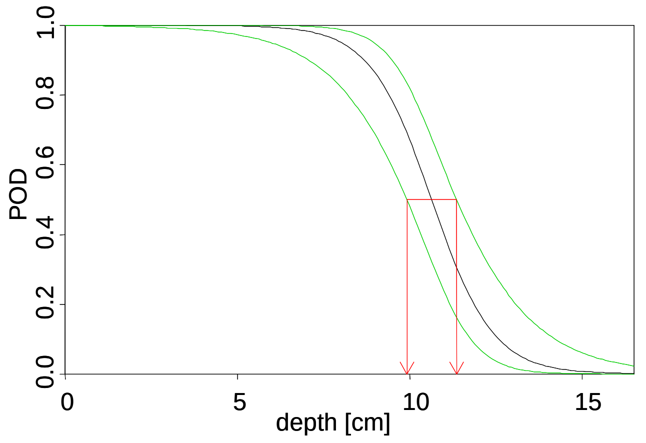
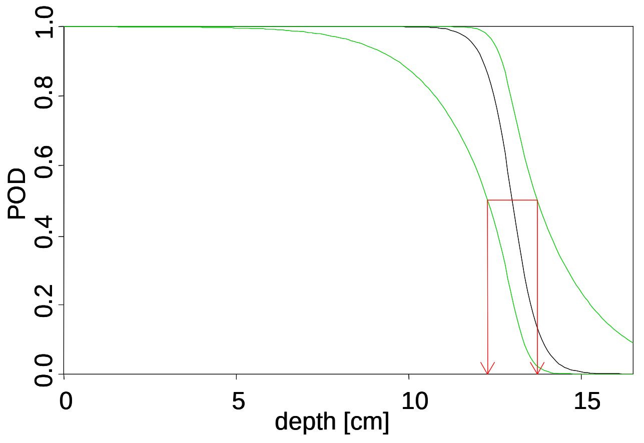
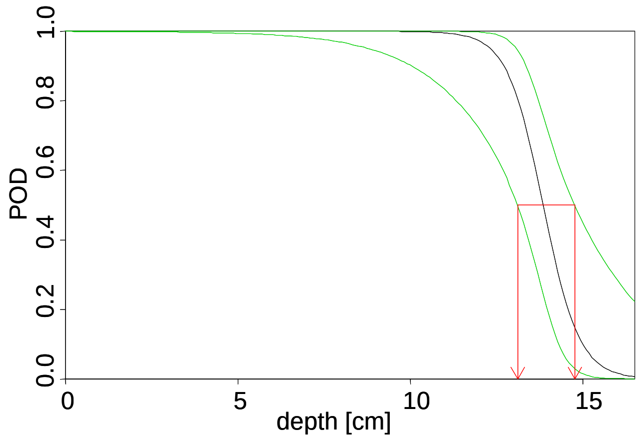
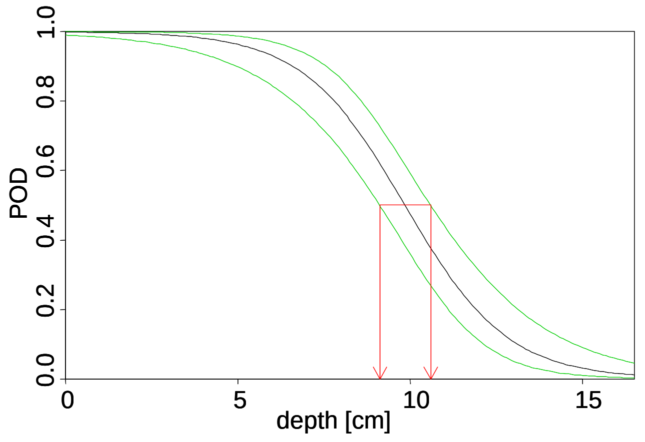
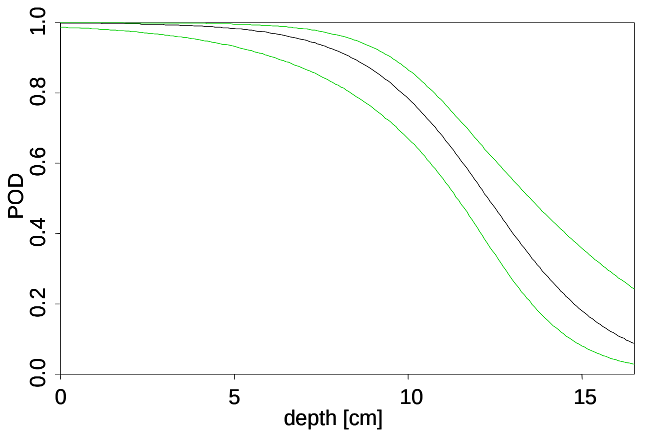
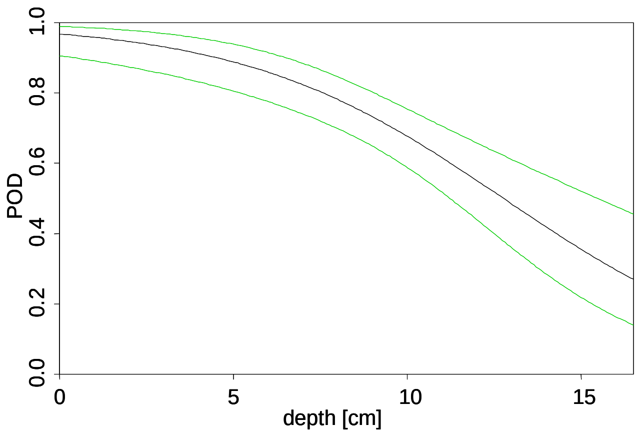
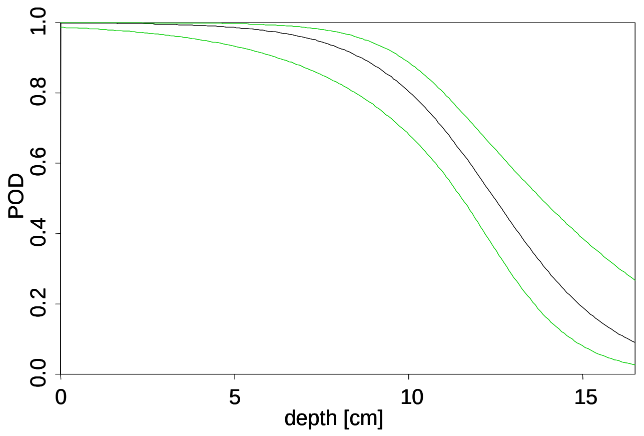
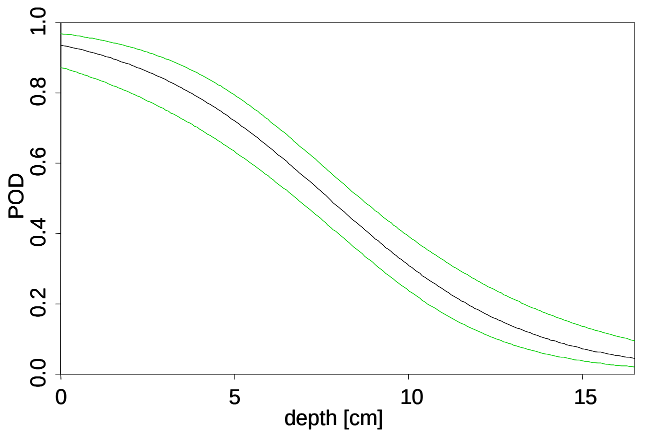
ANNEX 7 Test Results, Dependency of POD on Depth

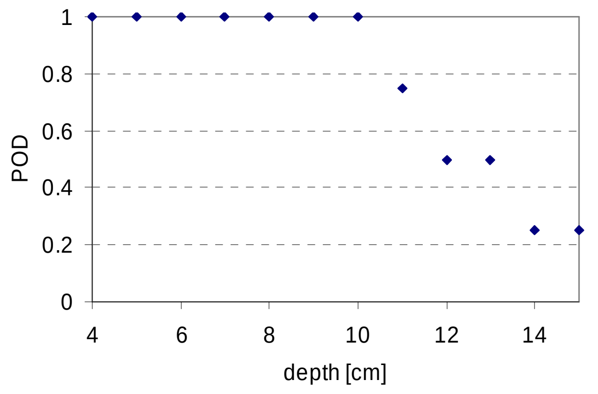
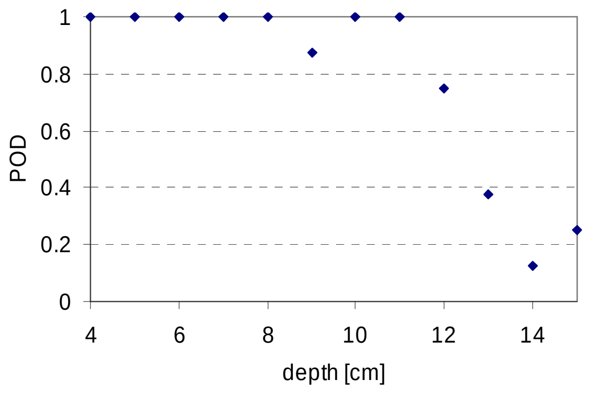
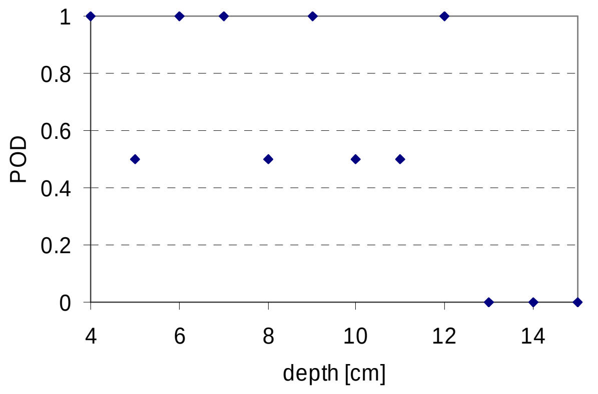
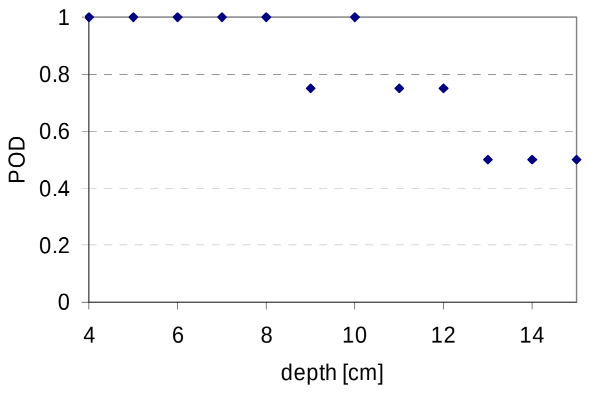
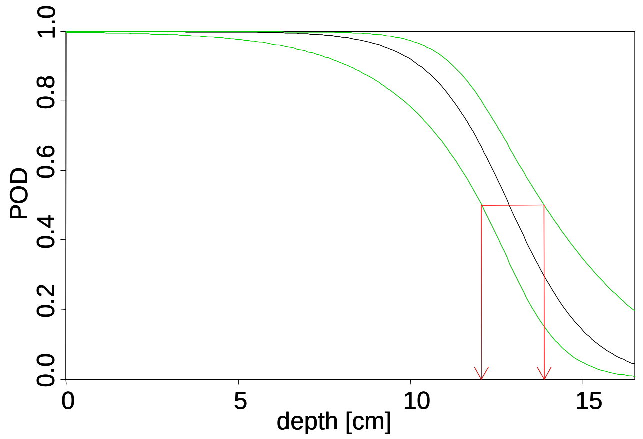
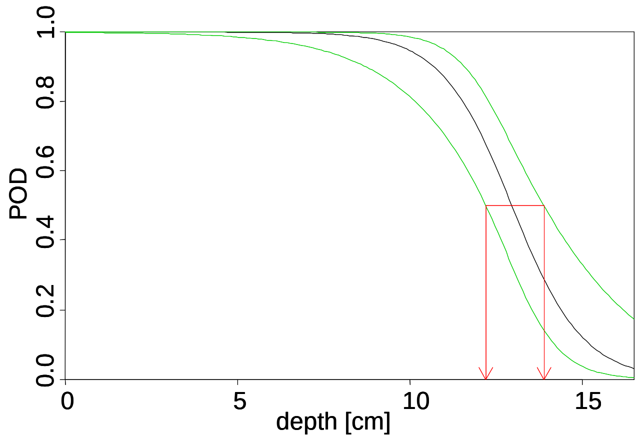
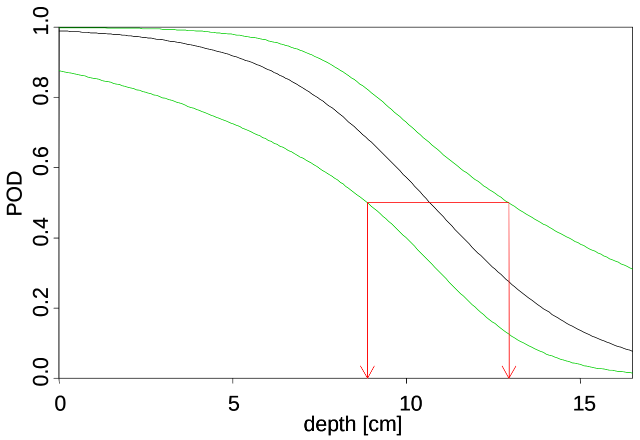
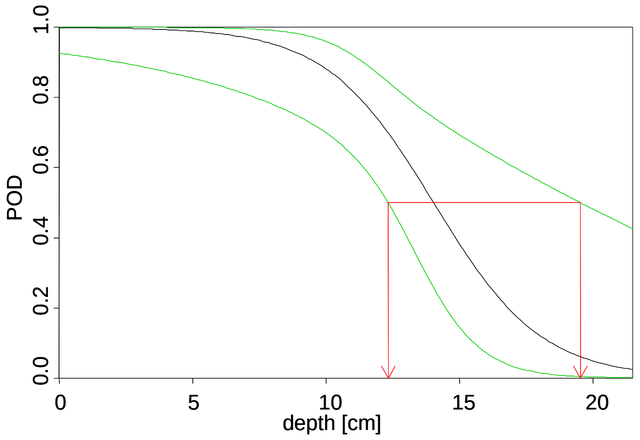
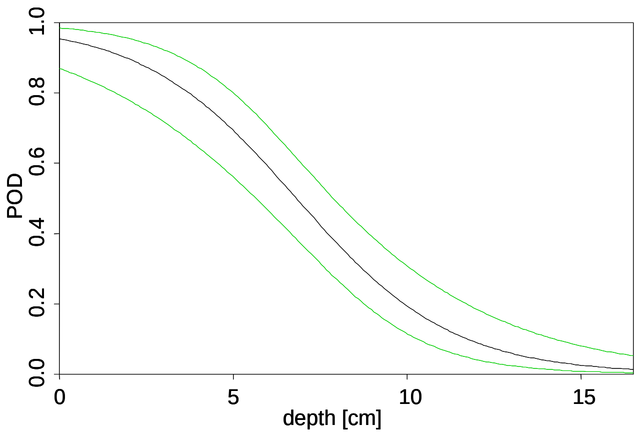
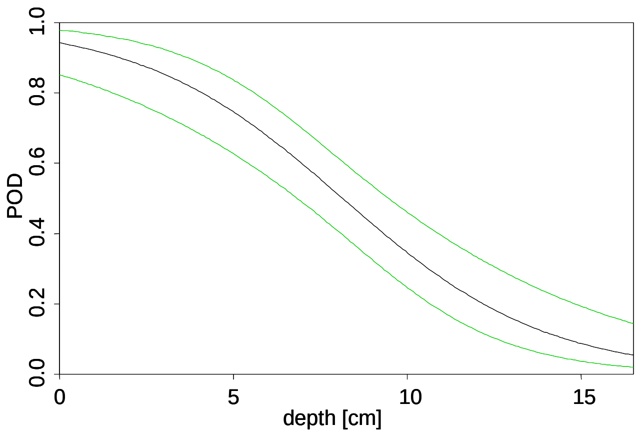
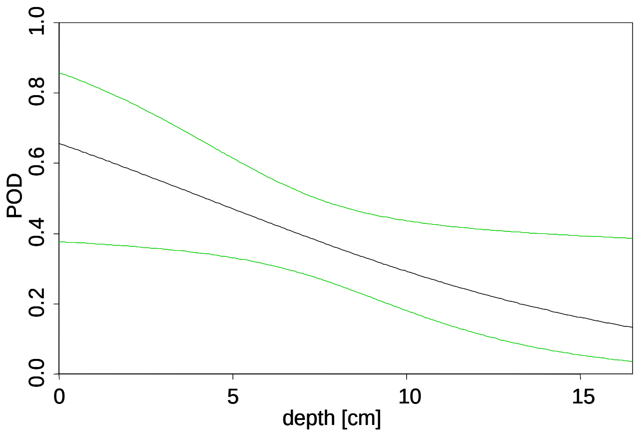
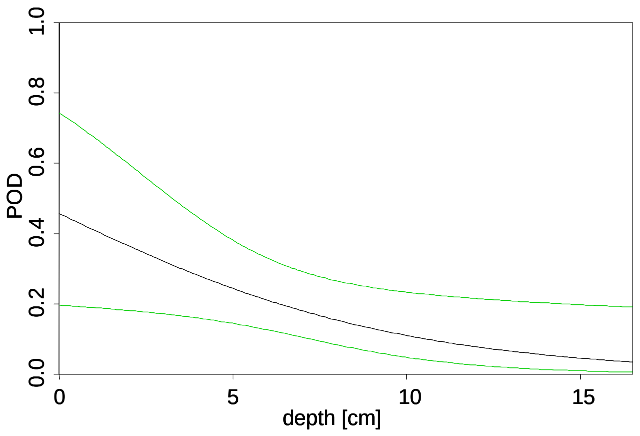
The diagrams present the results of the maximum detection depth (MDD) measurements and the reliability test for PMA-2 and PMA3 mines. The MDD measurements were performed on PMA-2S, surrogates of PMA-2, while the reliability test was performed with real PMA-2 mines rendered safe. The PMA-2S surrogates are slightly more difficult to detect than the real mines (see Section 7.2.4).

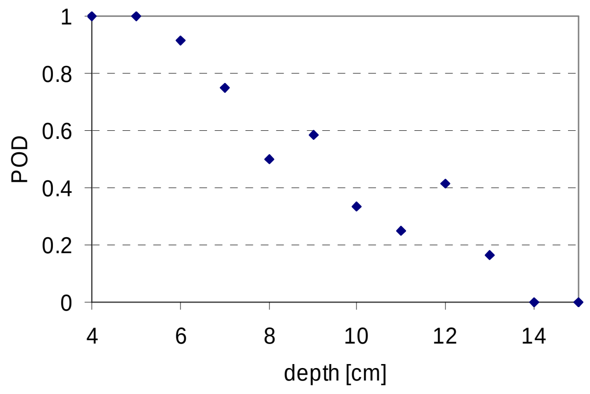
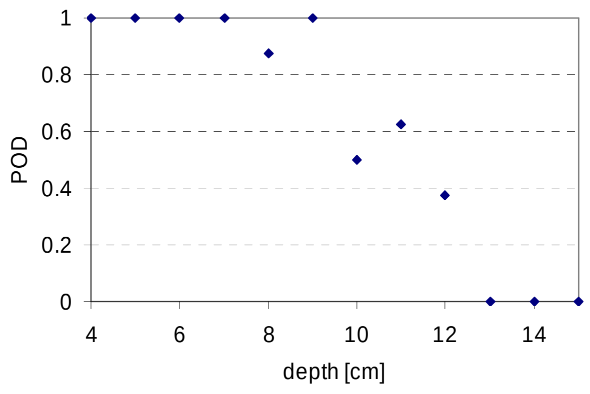
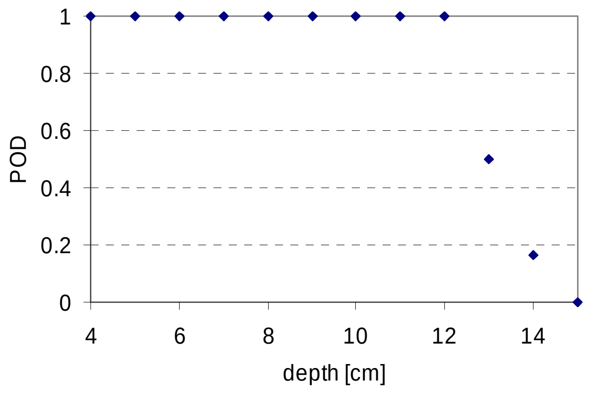
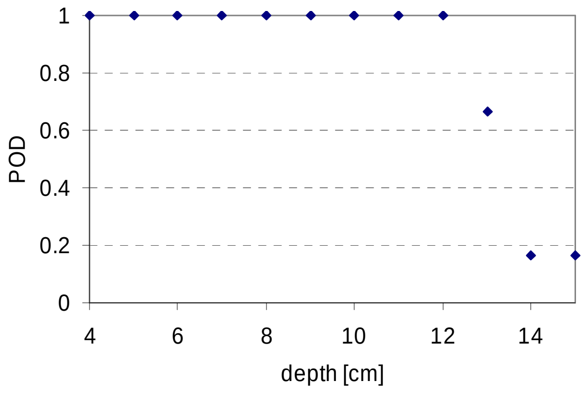
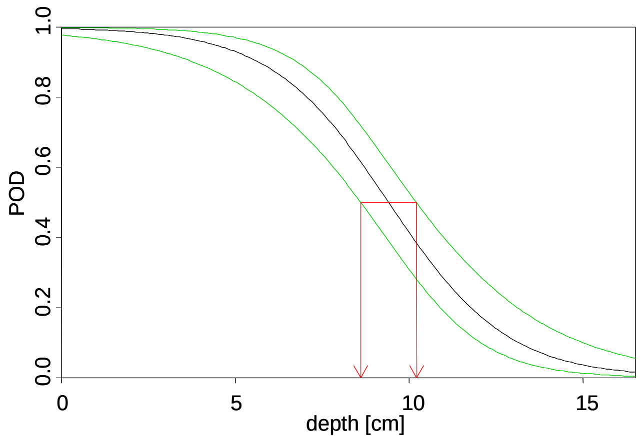
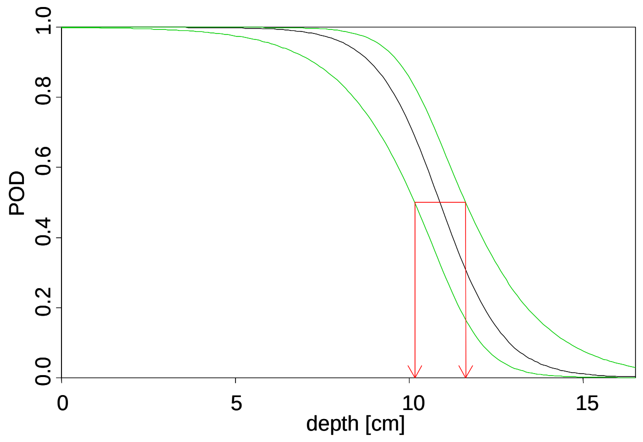
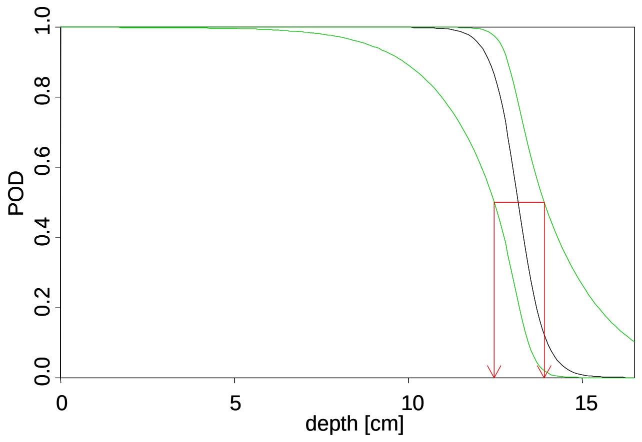
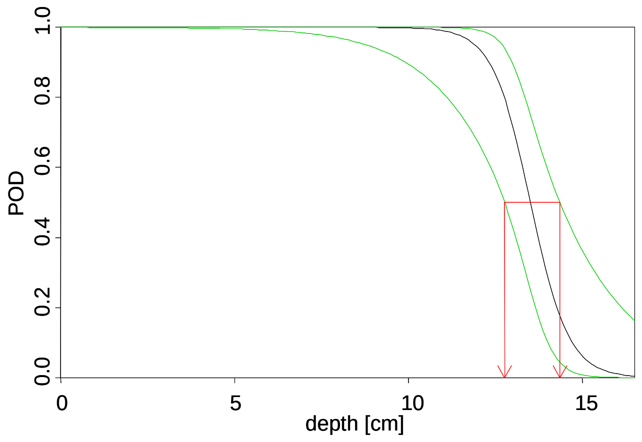
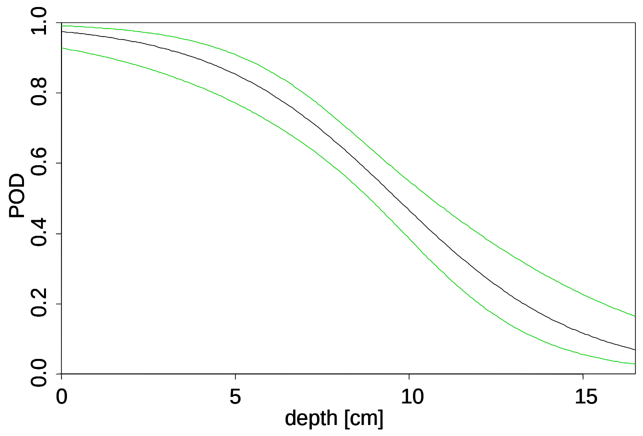
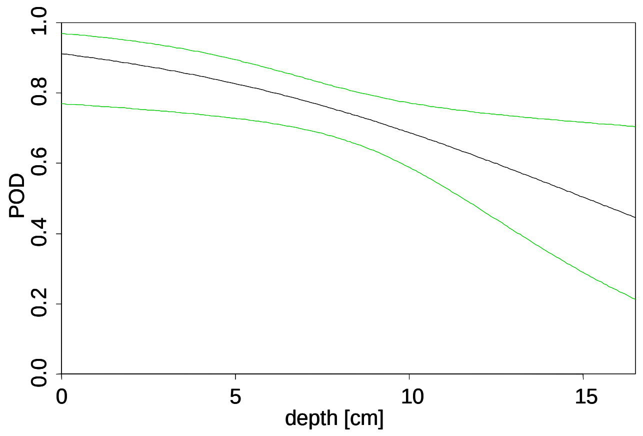
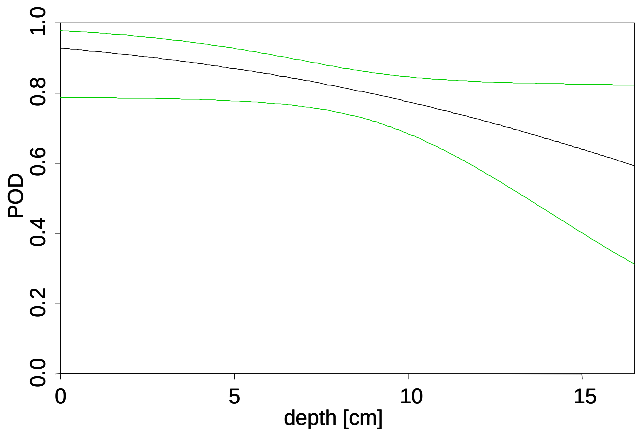
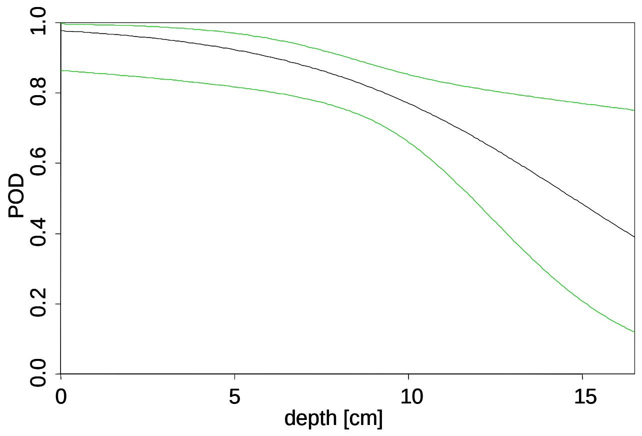
	AKA Condor 7252 in Sisak soil (lanes 3, 4)	AKA Vector 7260 in Sisak soil (lanes 3, 4)	CEIA MIL-D1 in Sisak soil (lanes 3, 4)	Foerster MINEX 2FD 4.530 in Sisak soil (lanes 3, 4)
MDD measurements				
MDD measurements, POD curves				
reliability test, POD curves				

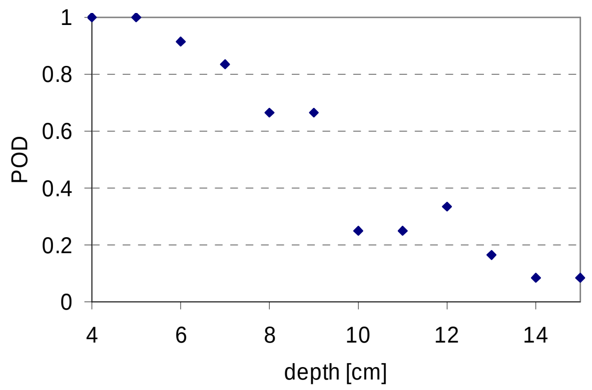
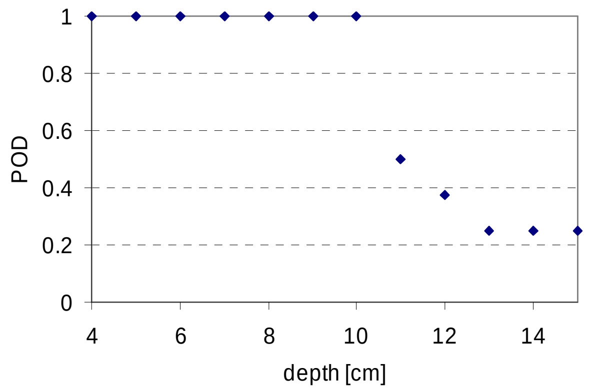
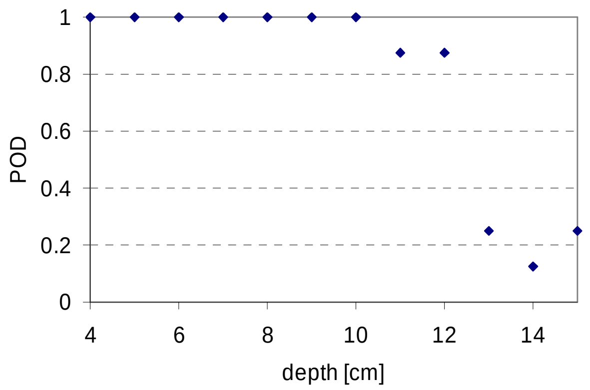
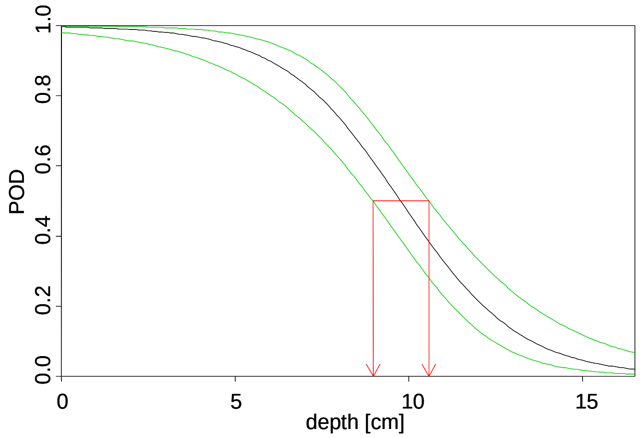
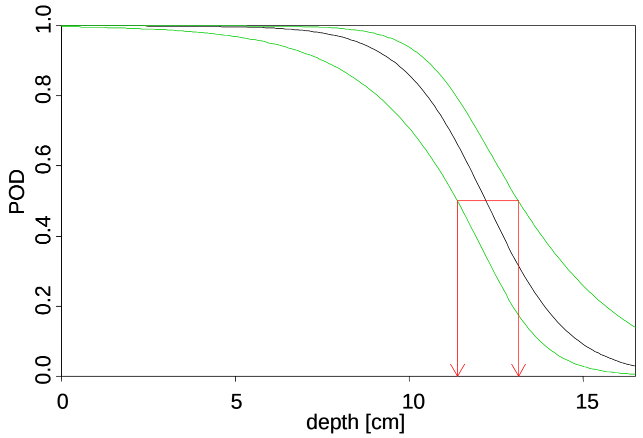
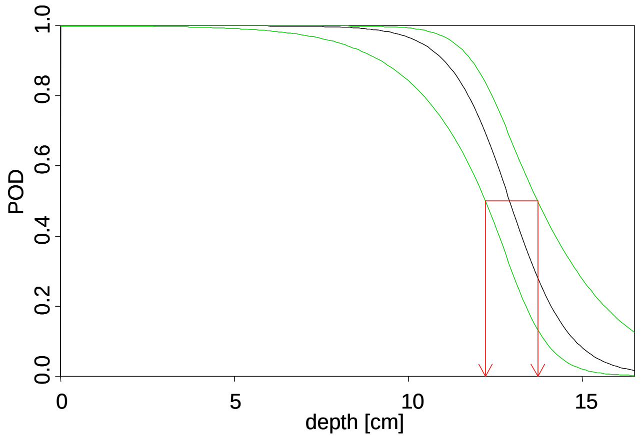
	Minelab F1A4 in Sisak soil (lanes 3, 4)	Minelab F3 in Sisak soil (lanes 3, 4)	Schiebel ATMID in Sisak soil (lanes 3, 4)	Vallon VMC1 in Sisak soil (lanes 3, 4)
MDD measurements				
MDD measurements, POD curves		not available (modelling with the generalised linear model not possible)		not available (modelling with the generalised linear model not possible)
reliability test, POD curves				

	Vallon VMH3CS in Sisak soil (lanes 3, 4)	AKA Condor 7252 in Obrovac soil (lanes 1, 2)	AKA Vector 7260 in Obrovac soil (lanes 1, 2)	CEIA MIL-D1 in Obrovac soil (lanes 1, 2)
MDD measurements				
MDD measurements, POD curves	not available (modelling with the generalised linear model not possible)			
reliability test, POD curves				

	Foerster MINEX 2FD 4.530 in Obrovac soil (lanes 1, 2)	Minelab F1A4 in Obrovac soil (lanes 1, 2)	Minelab F3 in Obrovac soil (lanes 1, 2)	Schiebel ATMID in Obrovac soil (lanes 1, 2)
MDD measurements				
MDD measurements, POD curves				
reliability test, POD curves				

	Vallon VMC1 in Obrovac soil (lanes 1, 2)	Vallon VMH3CS in Obrovac soil (lanes 1, 2)	AKA Condor 7252 in Benkovac soil (lanes 5, 6)	AKA Vector 7260 in Benkovac soil (lanes 5, 6)
MDD measurements				
MDD measurements, POD curves				
reliability test, POD curves				

	CEIA MIL-D1 in Benkovac soil (lanes 5, 6)	Foerster MINEX 2FD 4.530 in Benkovac soil (lanes 5, 6)	Minelab F1A4 in Benkovac soil (lanes 5, 6)	Minelab F3 in Benkovac soil (lanes 5, 6)
MDD measurements				
MDD measurements, POD curves				
reliability test, POD curves				

	Schiebel ATMID in Benkovac soil (lanes 5, 6)	Vallon VMC1 in Benkovac soil (lanes 5, 6)	Vallon VMH3CS in Benkovac soil (lanes 5, 6)
MDD measurements			
MDD measurements, POD curves			
reliability test, POD curves