Journal of Conventional Weapons Destruction

Volume 8
Issue 2 *The Journal of Mine Action*

Article 43

November 2004

ITEP Test Trials for Detection Reliability Assessment of Metal Detectors

Christina Mueller
BAM Berlin

Mate Gaal
Federal Institute for Materials Research and Testing

Martina Scharmach
BAM Berlin

Sylke Bär
BAM Berlin

A.M. Lewis
European Commission

See next page for additional authors

ITEP Test Trials for

Detection Reliability Assessment of Metal Detectors

Abstract

The total detection reliability of a mine-searching system is governed by the following three elements:

- Intrinsic capability, which describes the basic physical-technical capability of the method.
- Application factors, which include those due to environment.
- Human factor, which is the effect of human operators on the detection reliability.

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

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

A series of three field trials was performed in the International Test and Evaluation Program for Humanitarian Demining (ITEP) project 2.1.1.2, "Reliability Model for Test and Evaluation of Metal Detectors," in order to specify the optimal conditions to obtain reliable trial results with affordable effort. Each set of specific working conditions is characterized in terms of a combination of one mine type in one soil with one detector handled by local personnel. For each set of conditions, the searching system will deliver a working performance, expressed as mine-detection rates as a function of mine depth, and a certain overall false alarm rate. During the ITEP trials in Benkovac and Oberjettenberg, the authors learned to determine this function separately for each mine type in each soil. This is especially important for low-metal mines in soil that can influence metal detectors, as will be illustrated for the case of the PMA-2. Two discussion points still remain: how representative the trials are of field conditions and what statistical setup is required if we are to distinguish between the capabilities of individual detectors.

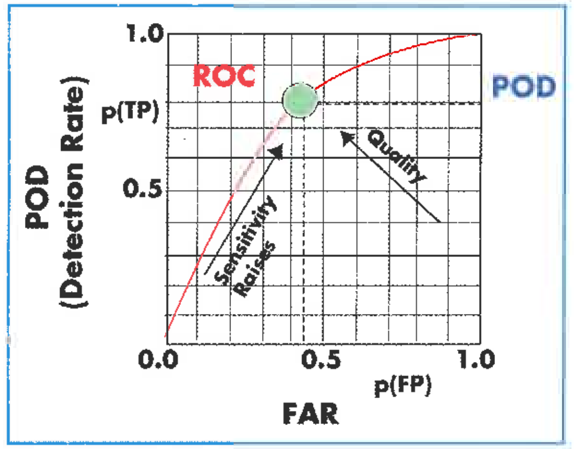


Figure 1: *Explanation for ROC and POD diagrams (FAR= false alarms per m²)*

Introduction and Background

The CEN Working Group 07 began the process of standardizing test and evaluation methods for metal detectors in humanitarian demining, including both laboratory measurements of detection capability and blind field trials (reliability tests). In reliability tests, the probability of detection (POD) and receiver operating characteristics (ROC) curves help to summarize the performance results. Under the umbrella of ITEP, a number of test trials with metal detectors have been conducted. The aim was to specify the trial setup and the statistical rules necessary to achieve true, repeatable and reproducible results under representative field conditions. The trial scenarios ranged from straightforward detection of a large, metallic anti-tank mine buried near the surface in a soil that does not give metal-detector signals to the most difficult challenge of detecting low-metal anti-personnel mines deeply buried in magnetic soil that affects detectors strongly. Individual human factors, such as training and currency of skills, were assessed. A full report about the trial conditions and results, including rules for minimum number of targets, operators, and test repetitions necessary to achieve **true** and reproducible results, was published on the ITEP website in the summer of 2004 (see http://www.itep.ws/reports/last_reports.php?projectid=293).

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

Workshop on Reliability Tests for Demining

About 100 international experts in demining met for the "Workshop on Reliability Tests for Demining." The proceedings (published at http://www.kb.bam.de/ITEP-workshop-03/) contain presentations of the oral sessions in which the general national, European and international concepts in demining are described as well as the main activities and results of the ITEP trials. An up-to-date series of lessons learned and problems to be solved was presented by international mine action centres. In four focused sessions, the authors and a number of competent international experts discussed the following specific topics: configuration of test lanes and test target (mine) selection, soil influence and ground compensation, human factors, and rules for test planning and statistical evaluation.

A highlight of the workshop was the second session, which addressed the problem of soils that influence metal detectors—such soils are described variously as "noisy" (CWA 14747:2003), "uncooperative" or "difficult." These effects were due principally to magnetic properties of the soil—both the magnitude of the magnetic susceptibility and its frequency dependence (see especially the presentation by S. Billings, et al.[1]). The fundamental magnetic properties were related to the empirical "ground reference height" measurement, developed by D. Güllc: the maximum distance above the ground at which a calibrated, static-mode detector gives an alarm due to that ground.

Further presentations dealt with conclusions for future practical activities, such as the Geneva International Centre for Humanitarian Demining (GICHD) Manual Demining Study (T. Lardner) or a worldwide accident database (A. L. Smith). One of the conclusions for future research requirements was that there was still a need to get a more comprehensive understanding of soil influences (S. Billings, et al.). Finally, the workshop assembly expressed "findings and recommendations" with recommendations for how to deal further with the topic of reliability and with modelling for the improvement of demining techniques.

by Christina Mueller, Mate Gaal, Martina Scharmach and Sylke Bär, *BAM Berlin*; Adam Lewis and Tom Bloodworth, *JRC Ispra*; Dieter Guelle, *Secretariat, ITEP*; and Peter-Th. Wilrich, *FU Berlin*

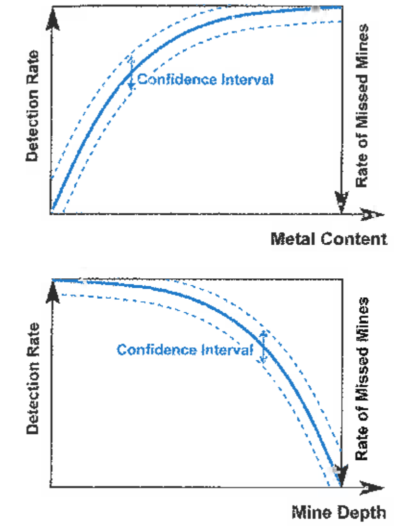


Figure 2: *Typical POD curves.*

Devices	Soil	Mines	Human Factor
2 pulse time domain U, X, W 2 continous wave Y, Z	Types of soil: • Cooperative (neutral) • Uncooperative (Frequency dependent susceptibility) Metal contamination of the soil Homogeneous/ heterogeneous	Types of mines: (metal content): • Biggest TM  • Smallest PMA-2  Depths of mines	Working time Training mode: • Brief • Extended Status of experience Pre-experience with one device type Age Current activity Personal capability

Figure 3: *Test parameters.*

POD and ROC—Summary of Rates for Detection and False Alarms

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

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

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

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

For a fixed amount of false alarms, the ROC point or operating point of the system for a fixed sensitivity can be taken and further analysed for its dependence on the main influencing factors such as the mine depth or the metal content of the mine

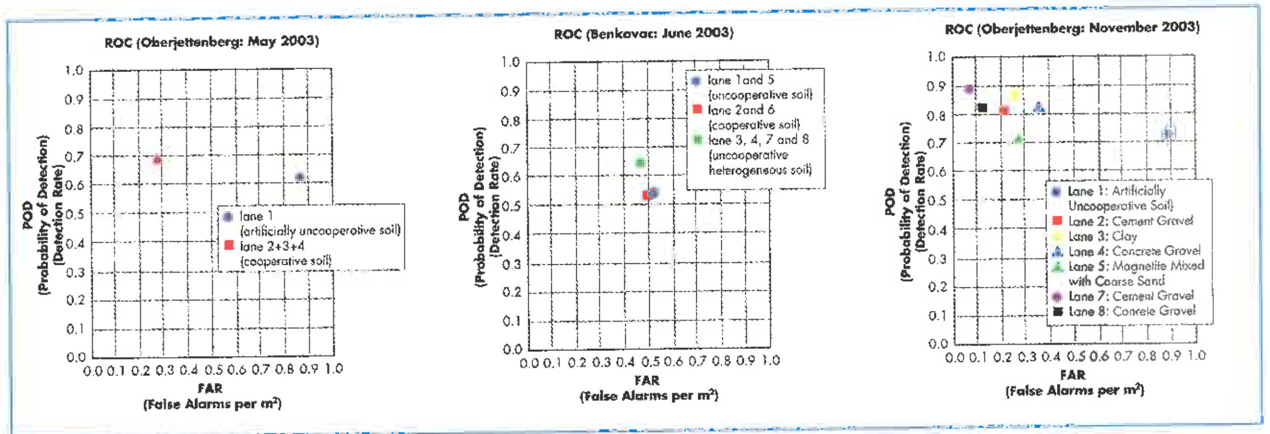


Figure 4: *ROC diagrams for different soil and human factor conditions.*

(Figure 2). All these points and curves need to be interpreted in connection with the corresponding confidence bounds to consider the scatter of results. The latter scatter depends on the underlying statistical basis (the number of opportunities to detect the mine) and the natural variability of the factors. The smooth POD or detection rate curves, presented in Figure 2, were determined by an advanced logistic regression model.[3] A simple way of obtaining the detection rate curves is by plotting the mean values of the experimentally measured detection rates for each step of burial depth.[4, 5, 6, 8]

Overview of the Parameter Matrix of the Trials

The main aim of the trials was to investigate how the device performance manifests itself in different application circumstances. The authors organized three sets of trials for which the main parameter setup can be seen in Figure 3. The first and third took place in Oberjettenberg WTD 52 on the testing ground of the German army.

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

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

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

Results of the Trials

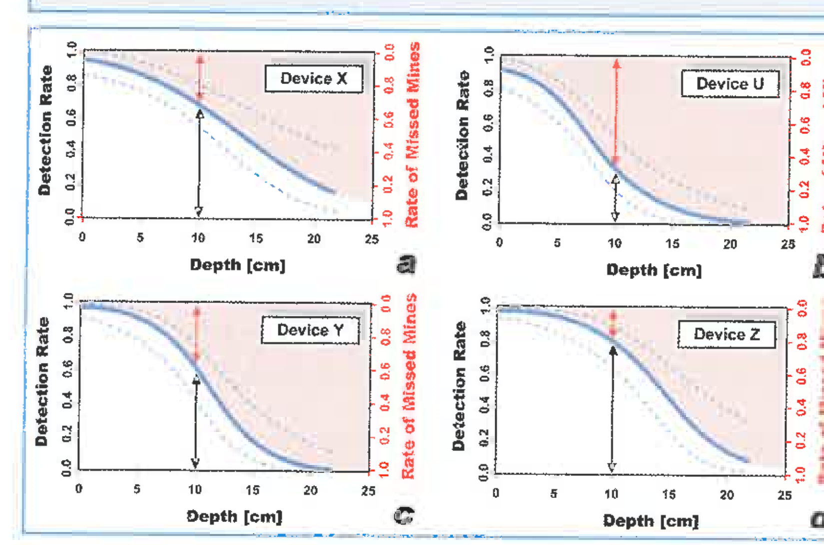
Figure 4 shows the overall results of each trial set, in ROC diagrams. These diagrams illustrate the influence of the factors (application factor and human factor) degrading the performance of all the detectors, without distinguishing among individual detectors. The result of inexperienced operators with a short training on metal-contaminated ground shows a mean detection rate of 70 percent and 0.3 false alarms per square meter. The artificial uncooperativeness reduces the performance to a 60-percent detection rate and almost one false alarm per square meter, which is surprisingly poor.

Even more surprising are the total overall results for Benkovac in June 2003, where the operators consisted of eight

Soil Types in Oberjettenberg Trials	Ground Reference Height (cm)	Susceptibility at 958 Hz (10^{-5}SI)	Susceptibility difference at 465 and 4650 Hz (10^{-5}SI)
Lane 1 artificially uncooperative soil	5 ± 2	244 ± 64	6.1
Lane 2 cement gravel	no signal	0 ± 1	-0.2
Lane 3 clay	no signal	2 ± 1	-0.5
Lane 4 concrete gravel	no signal	6 ± 1	-0.5
Lane 5 mangnetite mixed with coarse sand	4.5 ± 0.7	3000 ± 500	6 ± 7
Lane 7 cement gravel	no signal	-1.0 ± 0.2	-0.1 ± 0.2
Lane 8 concrete gravel	no signal	7 ± 1	-0.1 ± 0.1

Soil Types in Benkovac Trials	Ground Reference Height (cm)	Susceptibility at 958 Hz (10^{-5}SI)	Susceptibility difference at 465 and 4650 Hz (10^{-5}SI)
Lanes 2, 6 (neutral) Clay from Sisac	no signal	13 ± 2	0.6
Lanes 1, 5 (uncooperative) Laterite soil from Obrovac	18.8 ± 0.9	154 ± 13	25.5
Lanes 3, 4, 7, 8 (uncooperative heterogeneous) local red Bauxite from Benkovac	19.7 ± 2.5	190 ± 36	35.4

Figure 5 (above): *Overview of magnetic properties of the soils.*



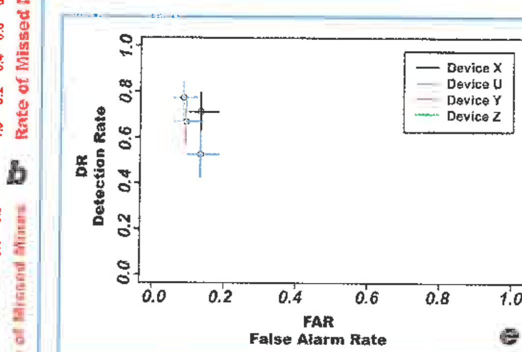


Figure 6: *This soil sample is neutral and very clean. It is the only mine PMA-2 with a mean value of ROC (detection rate versus FAR) that exceeds 95-percent confidence limits for the different devices.*

experienced Croatian deminers. The detection rate of about 65 percent in neutral soil decreases to almost 50 percent in a real, local, uncooperative soil with frequency-dependent susceptibility. The false alarm rate grows from 0.5 false alarms per square meter to almost 0.6. Possible reasons for this extremely poor result are as follows:

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

2. Only three of the deminers are currently active.

3. It has been suggested that experienced deminers may need a longer training phase because they are generally accustomed to using a particular detector model and cannot handle too many different device types at the same time.

4. In the trial, the deminers are not in

danger and are less motivated to be careful than they would be in a real minefield.

5. The test schedule required the deminers to work more quickly and for longer hours than they would normally.

6. The test lanes were contaminated with metal.

Heterogeneous soil with strong frequency-dependent magnetic susceptibility is a challenge for all detectors, especially in combination with minimum metal mines, since the soil signals often mask the mine signal.

The performance in the third trial is much better than in the first two, as expected from the conditions of the test with respect to the human factors and application factors. In Figure 4c, the upper left corner of the ROC point is 90-percent detection rate and false alarms below 0.1 per square meter. The "secret" is in carefully conducted and longer training, reduced workload, neutral and very clean soil, and targets that are easier to detect. If we want to estimate a realistic POD, it is therefore necessary to ask "What is the appropriate scenario of application and human factors for the situation we want to investigate?"

Full Process Simulation

In Oberjettenberg in November, one additional test was conducted, on the advice of Dieter Guelle.[7] The test simulated the full manual demining process, including prodding and excavation. Since the statistical basis was too small to be representative, results of this test must be considered indicative only and any conclusions provisional. The detection rate of the manual clearance process appeared to be higher than that of the detection process without excavation, probably due to instances where a minimum-metal mine was hidden by a larger false-alarm item. Indications that could be assigned to identifiable metal fragments were excluded (according to a "metal-free" approach), so the false alarm rate is lower. The latter is, of course, a matter of definition rather than performance. A more detailed investigation is planned within the GICHD program for improvement of the manual demining methods mentioned above.

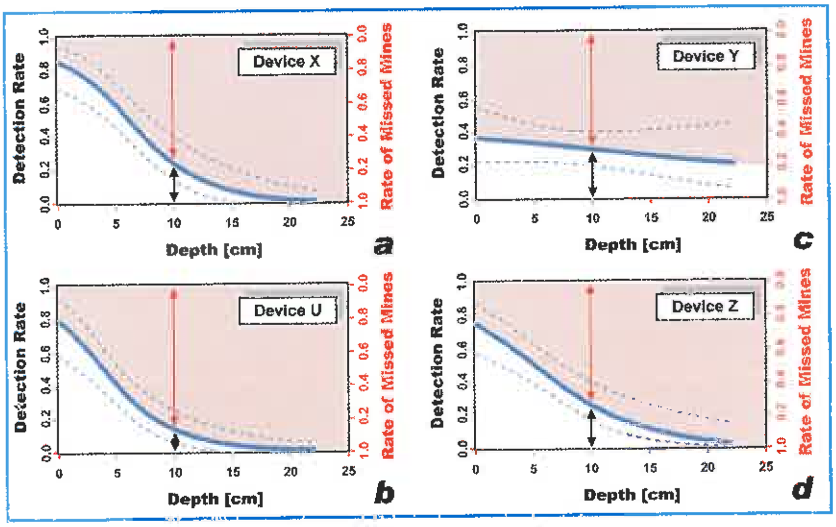


Figure 7: *This soil sample is heterogeneous and hence, uncooperative. Being that it has red bauxite with neutral stones it has frequency-dependent susceptibility. Its detection rate as function of mine depth (PMA-2 only) has four different devices with 95-percent confidence limits.*

Example of a Set of Resulting Curves:

Detection Rates as Function of Depth and False Alarms for the PMA-2 in Different Soils

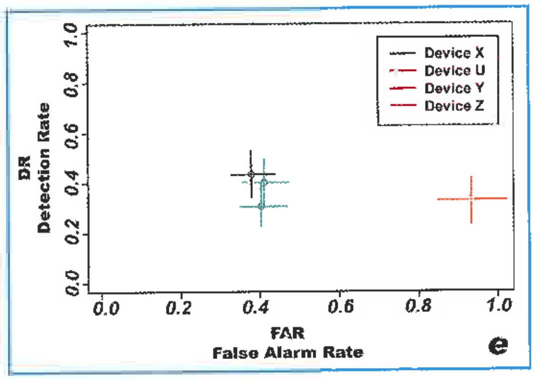


Figure 5 gives an overview of all the soils in the three trials.

In Figures 6 and 7, the individual detector results are illustrated for the PMA-2 minimum metal mine under ideal conditions (i.e., neutral soil without metal contamination, well-trained operators and optimized working hours.) Figures 6a–d show the detection rates as functions of the burial depth for each device separately and Figure 6e shows the ROC points of all devices together.

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

continued on page 109, ITEP

ITEP, continued from page 58

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

Conclusions and Outlook

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

* *Figures c/o author.*

References

1. S. Billings, L. R. Pasion and D. W. Oldenburg. "Characterising Magnetic Soils: State of the Art and Future Needs." *Reliability Tests for Demining*, 16–17. December, 2003, Berlin, Germany. http://www.kb.bam.de/ITEP-workshop-03/. File 0312171610_billings magnetic soils.pps on Workshop CD.

2. Ch. Nockemann, et al. "Performance Demonstration in NDT by Statistical Methods: ROC and POD for Ultrasonic and Radiographic Testing," Proceedings, 6th European Conference on Non-Destructive Testing, pp. 37–44 (1994).

3. P.-Th. Wilrich. "Statistical Design of Demining Experiments and Analysis by Logistic Regression," ITEP Workshop Reliability Tests for Demining, Federal Institute for Materials Research and Testing (BAM) Berlin, 16–17 December

continued on page 74, **ITEP**