INSTITUTE FOR DEFENSE ANALYSES

2022 SERDP/ESTCP/OE-Innovation Symposium Side Meeting on Scoring in the Underwater Environment

Daniel W. Kolodrubetz
Jacob B. Bartel

January 2023

Approved for public release;
distribution is unlimited.

IDA Document D-33361

Log: H 23-000007

INSTITUTE FOR DEFENSE ANALYSES
730 East Glebe Road
Alexandria, Virginia 22305-3086

About This Publication

This work was conducted by the IDA Systems and Analyses Center under contract HQ0034-19-D-0001, project AM-2-1528, "Assessment of Traditional and Emerging Approaches to the Remediation of Unexploded Ordnance (UXO) and Other Related Analyses," for the Executive Director, Strategic Environmental Research and Development Program (SERDP) and Environmental Security Technology Certification Program (ESTCP), under the Deputy Assistant Secretary of Defense, Environment and Energy Resilience. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments

The authors would like to thank Mike Tuley, Frank Rotondo, and Steven Rabinowitz of the IDA Systems and Analyses Center for their technical review of this report.

For More Information

Daniel W. Kolodrubetz, Project Leader
dkolodru@ida.org, 703-845-2579

Leonard J. Buckley, Director, Science and Technology Division
lbuckley@ida.org, 703-578-2800

Copyright Notice

INSTITUTE FOR DEFENSE ANALYSES

IDA Document D-33361

2022 SERDP/ESTCP/OE-Innovation Symposium Side Meeting on Scoring in the Underwater Environment

Daniel W. Kolodrubetz
Jacob B. Bartel

Executive Summary

On December 2, 2022, the Strategic Environmental Research and Development Program and Environmental Security Technology Certification Program (SERDP/ESTCP) Munitions Response (MR) Program Office hosted a side meeting during the SERDP/ESTCP/Operational Energy (OE)-Innovation Symposium. The side meeting focused on Scoring in the Underwater Environment, and was intended as the beginning of a conversation between different stakeholders working to solve the problem of underwater remediation of unexploded ordnance (UXO). The meeting included technology developers, test bed managers, test designers and scorers, and site managers. The key points of the presentations and discussion were:

- The current scoring for underwater blind tests uses the halo radius method to count the number of False Alarms and False Negatives. Based on these, the scoring package includes summary statistics, a visual map of performance, an estimate of the "best" detection halo radius, and a receiver-operating characteristic (ROC) curve (which summarizes classification performance).
- In the terrestrial realm, the accreditation is based on probability of correct classification of targets of interest (TOI) and the False Alarm Rate. Underwater, these metrics may not capture all variables that need to be considered.
- Variation between sites and situations underwater will necessitate variation in scoring and accreditation.
- Currently, there is no single body to address this regulation. There are many interested parties on a site-by-site basis.
- Three key metrics emphasized by site managers were:
  - Minimum size of TOI detectable
  - Depth to which system can detect (both water depth and burial depth)
  - Geolocation accuracy of the system
- Cost is an important consideration, and it could be worth convening a working group of stakeholders to consider the interaction between stakeholder needs, technology capabilities, and cost.

Contents

1. Introduction

As part of the 2022 Strategic Environmental Research and Development Program and Environmental Security Technology Certification Program (SERDP/ESTCP) Operational Energy (OE)-Innovation Symposium, the SERDP/ESTCP Munitions Response (MR) Program Office hosted a side meeting focused on Scoring in the Underwater Environment. This meeting took place from 9 a.m. to 12 p.m. on December 2, 2022. The session was chaired and planned by Kevin Williams from Applied Physics Laboratory, University of Washington (APL-UW), Stephen Billings from Black Tusk Geophysics, Jacob Bartel from the Institute for Defense Analyses (IDA), and Dan Kolodrubetz from IDA. Sarah Barlow from Noblis provided planning and administrative support. This meeting summary was compiled by Dan Kolodrubetz and Jacob Bartel.

The goal of the side meeting was to have a general discussion regarding the correct approaches to scoring underwater unexploded ordnance (UXO) detection tests. In particular, the list of attendees (see Appendix A) included stakeholders from various parts of the UXO detection and remediation process. This included technology developers, testbed managers, test designers and scorers, and site managers.

As each of these parties has potentially different concerns and priorities, this side meeting was an opportunity to start a conversation between the groups. The eventual goal of the SERDP/ESTCP MR program is to have the developed technologies transferred to real-world use. In order to do that, the tests of the technology must build enough trust from stakeholders and regulators. The scoring side meeting was a way for all involved parties to consider the best ways to build that trust through scoring.

This document summarizes what happened at the side meeting. It attempts to capture what was discussed as closely as possible.

2. Side Meeting Synopsis

The side meeting was split between several short presentations and ample time for discussion. In this section, we will summarize the presentations and the ensuing discussions.

A. Current State of Scoring

The session began with Jacob Bartel giving a presentation summarizing the current state of scoring. In his presentation, he explained that the current goals of scoring are to give quantitative performance measures and provide useful information for demonstrators. Scoring is well-developed in the land case, and underwater scoring uses the same basic procedure with some modifications due to underwater uncertainty.

In general, scoring uses a detection halo radius to identify which detections are true positives (TP, correctly detect an emplaced UXO), which are false positives (FP, incorrectly indicate a UXO is present when it is not), and when an emplaced UXO is not detected at all (FN or false negative). The scoring team reports summaries of these results, a visual map of performance, an estimate of the "best" halo detection radius, a list of difficult UXO, and a receiver-operating characteristic (ROC) curve (that summarizes classification performance). Further information about the scoring process can be found in the more in-depth reference: [1]. The main differences between land-based and marine scoring are:

- Use of the False Alarm Rate (FAR) metric underwater instead of Probability of False Alarm (Pfa)
- Larger halo detection radii

The use of FAR instead of Pfa is primarily because there are a wider range of technologies being tested in the marine environment than in the terrestrial case. The FAR metric allows for simpler comparisons between technologies that often use very different classification steps. A more detailed discussion of the different considerations that went into picking the ROC curve axes can be found in [2]. The larger halo detection radius is primarily a result of increased geolocation uncertainties in the marine environment. Discussion during and after the presentation included:

1. It would be helpful to incorporate typical types of clutter that would be found where UXO are found, including frag in an appropriate distribution of sizes and densities within the test site.

   a. This has been done at the Sequim Bay, Washington, test site to some extent already, but could be expanded in all test sites.

2. There were two important lessons learned from the land case:
   a. The center of mass of a target of interest (TOI) is not the same as the center from an electromagnetic interference (EMI) dipole. This can be relevant for larger TOI.
   b. It is important to keep track of mismatches between predicted and actual TOI size, even for TOI that are classified correctly.
3. It would be helpful to have standard and visually appealing ways to present scoring information to stakeholders and the public.
4. It would be helpful for the scoring package to have increased information about geolocation errors.

B. Accreditation

After the discussion of how scoring currently works, the meeting turned to thinking about the path toward accreditation in the marine environment. This began with a presentation from Dr. Stephen Billings. First, he laid out what system accreditation looks like in the terrestrial realm, where realistic detection and classification objectives were set. In particular, the test sites for system validation (that lead to the accreditation decision) are designed to be difficult, but not impossible. In the terrestrial case, they focus on two key metrics:

- Probability of correct classification of TOI (must be 1)
- False alarm probability at the “do not dig” point

In the marine environment there is more complexity and variability, and we do not have a solid base of what to expect. The two metrics from the terrestrial side might not be enough for underwater. Dr. Billings posed the following question to the group: What are the important metrics in the marine environment?

1. First Breakout Session

At this point, Dr. Billings opened the breakout sessions for people to discuss important underwater metrics and considerations. In order to guide the discussion, he posited two canonical case-studies:

- **Site A.** A marine site with future expected construction with high likelihood of potential interactions with UXO.
- **Site B.** A recreational marine site, with some light commercial activities such as fishing. A lower risk of UXO interaction.

The participants split into four different breakout groups to discuss these cases and the considerations for metrics of UXO detection and classification performance in the underwater environment. The breakout groups were chosen to give people of different backgrounds (e.g., test bed managers, scorers, developers, site managers) a chance to interact and discuss varying priorities. The key discussion points that came up during the breakout sessions are as follows:

- Differences between sites and situations are extremely important when looking at underwater testing and accreditation.
  - Site A would require the ability to detect and remove everything down to a certain size, while site B could rely on more of a wide-area assessment.
  - Testing could differ for systems designed to accomplish one goal vs. another. However, when testing the low-UXO-density scenario over a large area, it could be difficult to get good statistics when emplacing few UXO.
  - One idea was to have an initial acceptance gate: a blind test at an ESTCP demonstration site that all technologies must pass. Then have an additional site-specific test when the system is deployed to a new site. One suggested way to do the site-specific test is to use an Instrument Verification Strip (IVS) where items are emplaced and the contractor needs to demonstrate they can successfully detect (and classify) those items.
  - Additionally, there could be different types of accreditation. One option is one level for wide-area survey technologies, and another for precise detection and classification.
- Several key metrics are important, though the weight between different performance metrics will differ from situation to situation. Note that there may be relationships between these variables (for example, the minimum size a system can detect at a particular depth may be larger than the minimum size it can detect at a smaller depth).
  - **The minimum size of TOI a system can detect**
    - This is particularly true for the Site A case where all UXO above a certain size will likely need to be removed.
    - One suggestion from the discussions was to present a separate ROC curve for different size cutoffs to see how performance depends on TOI size.
  - **The depth to which a system can detect**
    - This refers to both the water depth that the system can operate in, and the burial depth under the seabed that the system can detect.

- **The geolocation accuracy of the system**
  - For the geolocation metric, it can be more important to get extremely precise measurements from a system classifying in a UXO-dense environment. Less precise measurements from a system doing a wide-area survey could be acceptable.

- For testing, it could be beneficial to create a standardized set of targets. This would minimize variability across tests and allow for easier comparison of other important variables, such as the test bed environment.
  - It will still be important to match the tests to real-life remediation environments as closely as possible.

C. Discussion of Compounding Factors

After the first breakout session, Dr. Kevin Williams presented about potentially compounding factors. In particular, he asked the group to think about:

- How site size plays a role in influencing the technological requirements. Two examples of where this differs is San Diego Bay, California, versus Minefield Fox in Key West, Florida.
- How cost comes into the equation of evaluating a technology.
- Could different technologies work together to create the best detection and classification solution.

Dr. Williams emphasized tradeoffs between different variables. The group discussed some of these issues as a whole. The main focus of this discussion was to try to understand the priorities and issues faced by site managers. The questions asked and corresponding answers were:

- What are the dynamics for underwater sites for regulators?
  - Generally, underwater sites have more stakeholders and regulators involved than terrestrial sites (see Culabera, Puerto Rico, as an example). Each has different end goals and expectations.
  - Site managers would appreciate an accounting of the limitations of each system to help with choosing the correct technology and communicating with regulators.
  - Stakeholders/regulators meet at the beginning of a remediation project to discuss end goals and the conceptual site model. Who is involved is decided case-by-case.

- Site managers confirmed three important metrics to ensure technology matches their needs: 1) minimum size of detectable target, 2) depth of detection (both water depth and burial depth), and 3) geolocation accuracy.
  - Detection curves would be nice to have.
- Is manage-in-place (i.e., leaving detected UXO where they are, but monitoring them to make sure that they do not move) coming up in underwater site discussions?
  - Different regulators respond to that philosophy in different ways. Some regulators want everything cleaned up.
  - It is likely that all sites will have long-term management.
- Is there a list of most important sites?
  - No, but politics play a role in determining what is a priority.
- Is there any effort to organize regulators similar to the Interstate Technology and Regulatory Council (ITRC)?
  - Not yet, but that would be useful.
- Is there a historical database to see what has been used at what sites?
  - There is not a single place to find that information.
- What percentage of the remediation cost is spent on detection?
  - Extremely rough estimate is 1/4 – 1/3 of the total remediation cost.
  - On land, the estimated cost is $100 per dig. Underwater, that estimate should be approximately 10x higher.

After this discussion, the breakout groups formed again to talk about the additional factors impacting scoring and accreditation.

1. Second Breakout Session

The following bullets summarize the discussion points that came out of the second breakout session.

- Considering cost is an important factor in determining the correct technology.
  - Simply lumping risk and cost together and trying to minimize both may not lead to the best solution.
  - Need to think about both survey costs and remediation costs. In other words, there should be a cost penalty for having more false alarms.

    - It could be worth convening a working group to discuss the intersection of stakeholder needs, technology capabilities, and cost. This group should include people with the problem (e.g., site manager), and those with the technical solutions (e.g., technology developers).
- There are possibilities for combining technologies to solve the problem.
    - Technologies can be combined additively. For example, one technology could be used for a wide-area survey, then another can be used for detailed surveys in identified problem areas.
    - Technologies could perhaps see even greater benefit by combining data for analysis (e.g., combining light detection and ranging (LiDAR) and acoustic data). However, this would require a significant investment to create the tools required for this analysis.
- For regulators and stakeholders, it could be helpful for SERDP/ESTCP to put together a database of technologies. This could include information about:
    - The expected specifications/limitations of the technology
    - The list of environments where the technology has been tested
    - The results of any of the tests of the technology

The side meeting then concluded with quick summaries from each breakout group.

Appendix A.
List of Attendees

Name	Organization
Jacob Bartel	IDA
Stefano Biagini (virtual)	CMRE
Stephen Billings	Black Tusk Geophysics
David Bradley	SERDP/ESTCP
Jill Brandenberger	PNNL
Kelly Enriquez	USACE
Richard Funk	Tetra Tech
Jeff Gamey	Tetra Tech
Joe Haxel	PNNL
John Jackson	USACE
Steve Kargl	APL-UW
Kevin Kingdon	Black Tusk Geophysics
Dan Kolodrubetz	IDA
Raymond Lim	NSWC PCD
Peter Menzel	Corvus Works UG
David Raudales	NRL
Mike Richardson	IDA
Greg Schultz	White River Technologies
Fridon Shubitidze	Dartmouth
Harry Simpson	NRL
Per Arne Sletner	CMRE
Dan Steinhurst	Nova Research, Inc. / NRL
Mike Tuley	IDA
Richard Whitehouse	HR Wallingford
David Williams	ARL-PSU

Name	Organization
Kevin Williams	APL-UW
Dana Woodruff	PNNL
Timothy Yoder	NRL

References

[1] Cazares, S. M. 2020. *Scoring Underwater Demonstrations for Detection and Classification of Unexploded Ordnance (UXO)*. IDA Document D-19436. Alexandria, VA: Institute for Defense Analyses. https://www.ida.org/research-and-publications/publications/all/s/sc/scoring-underwater-demonstrations-for-detection-and-classification-of-unexploded-ordnance-uxo.

[2] Cazares, S. M., and J. B. Bartel. 2021. *Practical Considerations in Constructing Receiver-Operating Characteristic (ROC) Curves to Assess the Performance of Advanced Detection and Classification Technology for Environmental Remediation of Unexploded Ordnance (UXO).* IDA Document D-23838. Alexandria, VA: Institute for Defense Analyses.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*
January 2023

2. REPORT TYPE
FINAL

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
2022 SERDP/ESTCP/OE-Innovation Symposium Side Meeting on Scoring in the Underwater Environment

5a. CONTRACT NUMBER
HQ0034-19-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER
AM-2-1528

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Kolodrubetz, Daniel W.
Bartel, Jacob B.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
Systems and Analyses Center
730 East Glebe Road
Alexandra, VA 22305-3086

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Document D-33361

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
SERDP and ESTCP
4800 Mark Center Drive
Suite 16F16
Alexandria, VA 22350-3605

10. SPONSOR/MONITOR'S ACRONYM(S)
SERDP/ESTCP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited (14 February 2023).

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The U.S. Department of Defense Strategic Environmental Research and Development Program (SERDP) and Environmental Security Technology Certification Program's (ESTCP) Munitions Response program hosted a side meeting focused on Scoring in the Underwater Environment as part of the 2022 SERDP/ESTCP/Operational Energy (OE)-Innovation Workshop. During this side meeting, participants discussed the problem of scoring unexploded ordnance (UXO) detection tests in the marine environment. While there are established metrics for this scoring in the terrestrial environment, the difficulties of the marine environment add complications. The goal of this side meeting was to begin a conversation amongst stakeholders about the appropriate approach to scoring, with an eye toward eventual accreditation of underwater UXO detection and classification technologies. This paper summarizes the discussions from the side meeting.

15. SUBJECT TERMS
acoustics; burial and migration; electromagnetic induction; munitions remediation; sonar; target classification; target detection; underwater; Unexploded Ordnance (UXO)

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	Same as Report	16	Bradley, David
					19b. TELEPHONE NUMBER *(Include area code)* 571-372-6388

Reset