IDA

INSTITUTE FOR DEFENSE ANALYSES

Storytelling with Receiver Operating Characteristic (ROC) Curves for Environmental Remediation of Unexploded Ordnance (UXO)

(Presentation)

Shelley M. Cazares
Jacob B. Bartel

July 2021

IDA Document NS D-22766

Log: H 21-000287

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About This Publication

This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-19-D-0001, Project AM-2-1528, “Assessment of Traditional and Emerging Approaches to the Detection and Classification of Surface and Buried Unexploded Ordnance (UXO),” for the Director, Environmental Security Technology Certification Program (ESTCP) and Strategic Environmental Research and Development Program (SERDP), under the Deputy Assistant Secretary of Defense (Environment). The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

For More Information

Shelley M. Cazares, Project Leader
scazares@ida.org, 703-845-6792

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

Copyright Notice

Executive Summary

Receiver operating characteristic (ROC) curves are often used to assess the performance of binary classification systems, allowing stakeholders to understand tradeoffs between Type I (false positive) and Type II (false negative) errors. This works well in textbook cases. In real-world experiments, however, ROC curves can have unexpected subtleties that make them difficult to construct and interpret. For example, the Department of Defense is sponsoring the development of novel sensors and software to identify unexploded ordnance (UXO) in the midst of clutter. UXO are duds—munitions that were previously armed and fired but failed to explode. UXO can still pose a risk of detonation even decades later, threatening the safety of nearby humans, animals, vegetation, and structures. Conducting blind tests to demonstrate finding UXO is fraught with safety, logistical, and cost constraints that make it difficult to construct the textbook ROC curves. Yet, with careful planning and a few key assumptions, ROC-like curves can still be crafted to quickly tell the story of how well novel sensors and software can detect, classify, and locate UXO versus clutter in terrestrial and underwater experiments.

Storytelling with Receiver Operating Characteristic (ROC) Curves for Environmental Remediation of Unexploded Ordnance (UXO)

Shelley Cazares, Jacob Bartel
Science and Technology Division
Institute for Defense Analyses
August 2021

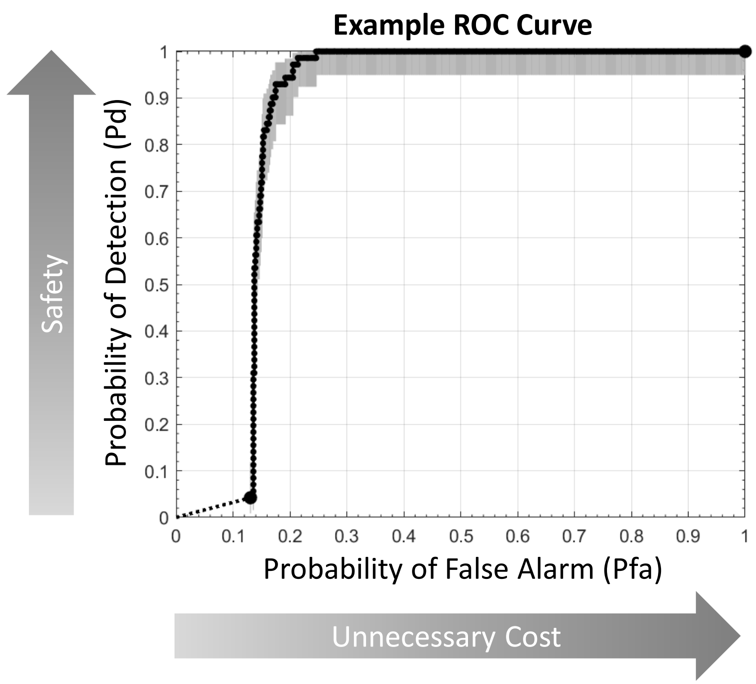
The Institute for Defense Analyses (IDA) is a non-profit Federally Funded Research and Development Center. IDA's mission is to provide objective analyses of national security issues, particularly those requiring scientific, technical, and analytic expertise—including the use and communication of statistics.

Over the past several years, IDA has supported the U.S. Department of Defense (DoD) with the design and evaluation of novel technology for environmental remediation of unexploded ordnance (UXO). This presentation describes how IDA researchers use Receiver-Operating Characteristic (ROC) curves to tell the story of how well this novel technology can meet the DoD's needs.

This presentation can be cited as follows:

Cazares, Shelley and Jacob, Bartel. 2021. Storytelling with Receiver Operating Characteristic (ROC) Curves for Environmental Remediation of Unexploded Ordnance (UXO). Presented at the *ASA Joint Statistical Meeting 2021*, virtual, 7–12 August 2021. (DD1910 cleared for open publication 21 July 2021)

Receiver-Operating Characteristic (ROC) Curves



- ROC curves convey tradeoffs between Type I and II errors
  - False Positives vs Negatives
  - Unnecessary Cost vs Safety
- Constructing ROC curves requires full ground truth
- Full ground truth is hard to obtain for some use cases: **Underwater UXO remediation**

Contact: scazares@ida.org

ROC curves are often used to assess the performance of *binary* classification systems. In the DoD, this is often a matter of detecting and classifying *threats* among *clutter*. ROC curves provide a quick and intuitive way for stakeholders to understand the tradeoffs between Type I and II errors, otherwise known as false positives and negatives. In many situations, this can be recast into a tradeoff of:

- *Safety*, on the vertical axis—using a metric like the Probability of Detection (Pd) to summarize how many *threats* are *found*—versus
- *Unnecessary costs*, on the horizontal axis—using a metric like the Probability of False Alarms (Pfa) to summarize how many *alarms* are *false*.

This approach works well in textbook cases.

In real world experiments, however, ROC curves can have unexpected subtleties that make them difficult to construct and interpret. In particular, constructing ROC curves requires full ground truth—knowledge about which threats and clutter were *truly* present.

Unfortunately, full ground truth can be difficult to obtain in some use cases, such as underwater UXO remediation.

This presentation will discuss:

- What *is* underwater UXO remediation?
- Why is it so *difficult* to construct ROC curves for this use case?
- What can we *do* about it?

ROC Curve figure adapted from:

Cazares, Shelley, Elizabeth Ayers, and Michael Tuley. 2017. UXO Live Site Classification Demonstrations: A Retrospective Study. Presented at *SERDP & ESTCP Symposium 2017*, Washington DC, 28–30 November 2017. (DD1910 cleared for open publication 3 Nov 2017)

Unexploded Ordnance (UXO)

- UXO are duds: Munitions that were previously armed and fired but did not explode
- UXO can still pose a risk of detonation, even decades later
- Millions of acres of land in the continental U.S. are contaminated with UXO, due to their previous uses as military training camps and test ranges

UXO

vs

Clutter

Contact: scazares@ida.org

UXO stands for Unexploded Ordnance.

UXO are duds. They are munitions (mortars, artillery, and so forth) that were previously armed and fired but did not explode.

They can still pose a risk of detonation, though, even decades later.

____

UXO and Clutter photos reproduced from:

Cazares, Shelley, Elizabeth Ayers, and Michael Tuley. 2017. UXO Live Site Classification Demonstrations: A Retrospective Study. Presented at *SERDP & ESTCP Symposium 2017*, Washington DC, 28–30 November 2017. (DD1910 cleared for open publication 3 Nov 2017)

Tuesday 31 January 2012

The Telegraph

HOME NEWS SPORT FINANCE COMMENT BLOGS CULTURE TRAVEL LIFEST

UK World Politics Obituaries Education Earth Science Defence Health N

HOME » NEWS » NEWS TOPICS » WORLD WAR 2

Second World War bomb leads to mass evacuation German town

An unexploded Second World War bomb is leading to the evacuation of nea the population of the German town of Koblenz.

Some 45,000 of the 106,000-strong population will be cleared from an evacuation zone 1.8 kilometres in radius i post-war evacuation in Koblenz's history Photo: ALAMY

The New York Times

World War II-Era 'Earthquake Bomb' Explodes in Polish Waters

Polish Navy divers tried to remotely neutralize a six-ton "Tallboy" dropped by the British, but the bomb had no intention of going quietly.

A British World War II bomb exploded during an effo specialists in northwestern Poland on Tuesday.

By John Ismay

Oct. 14, 2020

A 12,000-pound World War II-era canal off the Polish port city of S plume of water high into the air a working to neutralize it remotely

USNews NEWS » News Best Countries Best States Healthiest Communities Cities Sign in

Home / News / Best States / Hawaii News

Hawaii Plans to Leave Maui Underwater War Ordnance in Place

Hawaii plans to leave undetonated World War II ordnance in place off Maui until a nonexplosive removal option is developed

By **Associated Press**, Wire Service Content Aug. 21, 2020, at 10:06 a.m.

WAILUKU, HAWAII (AP) — Hawaii plans to leave undetonated World War II ordnance in place off Maui until a nonexplosive removal option is developed.

The state Department of Land and Natural Resources announced Wednesday that the decades-old explosives will remain untouched for now in the Molokini Marine Life Conservation District, The Maui News reported.

The Department of Defense Explosives Safety Board recommended against detonating the munitions in place following a state request for an assessment.

Community members, environmentalists and lawmakers expressed fears of irreversible damage to coral reefs and ocean life in the Molokini Crater.

The military used the crater for bombing practice when the U.S. entered WWII. Detonations in place have previously been conducted to remove dangerous munitions

This risk is often reported on in the popular news.

Many sites around the world are contaminated with UXO from WWI, WWII, and more recent wars.

Many sites in the *United States* are also contaminated with UXO, due to their prior uses as military training camps and test ranges.

The DoD is committed to remediating the threat of UXO in order to keep ourselves and future generations safe.

The Telegraph figure reproduced from:

https://www.telegraph.co.uk/news/newstopics/world-war-2/8920347/Second-World-War-bomb-leads-to-mass-evacuation-of-German-town.html

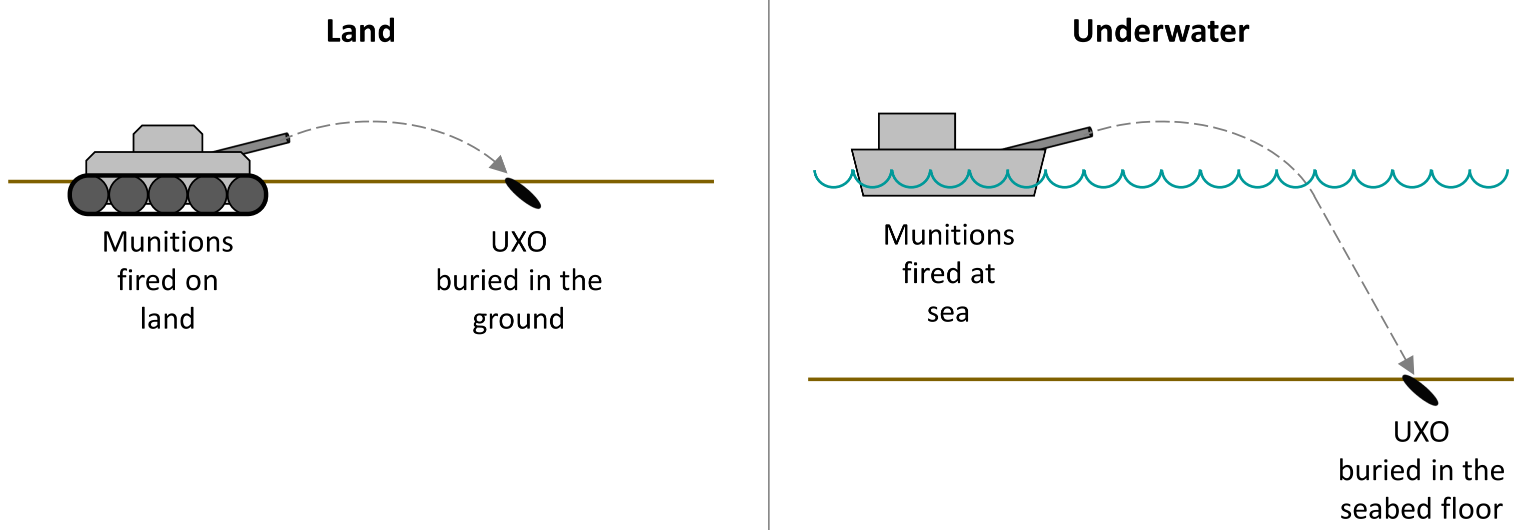
The New York Times figure reproduced from:

https://www.nytimes.com/2020/10/14/world/europe/poland-bomb.html

The US News figure reproduced from:

https://www.usnews.com/news/best-states/hawaii/articles/2020-08-21/hawaii-plans-to-leave-maui-underwater-war-ordnance-in-place

UXO on Land and Underwater



Contact: scazares@ida.org

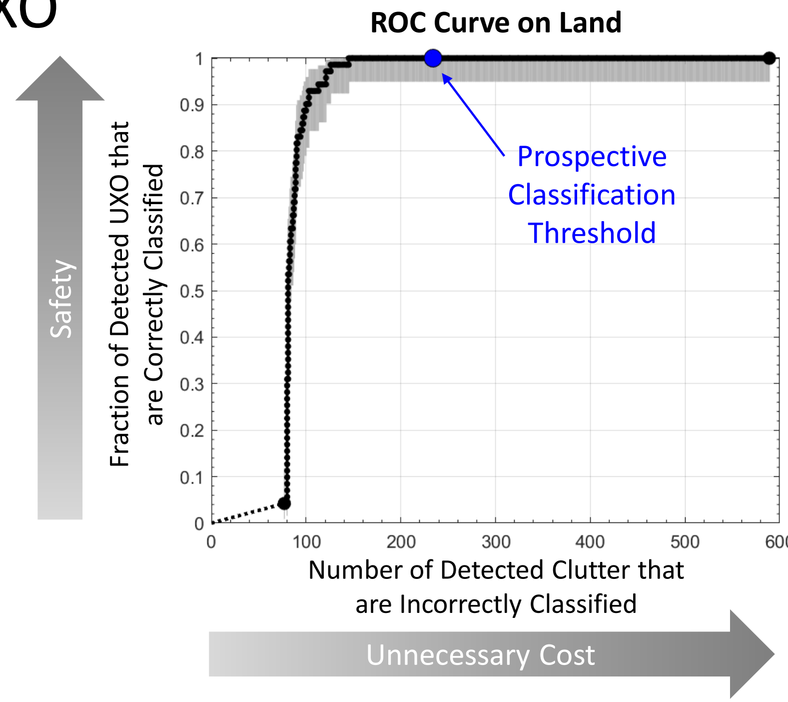
UXO can contaminate both land and underwater sites.

The cartoon on the *left* shows that munitions can be fired on *land*, sometimes resulting in duds that burrow into the ground and remain undetonated. These UXO can later unearth themselves during construction or farming efforts.

The cartoon on the *right* shows that shots can also be fired into oceans, rivers, and lakes, sometimes resulting in duds that sink through the water and bury into the seabed floor. These *underwater* UXO can later roll up onto the beach.

Land Remediation of UXO

Time-domain Electromagnetic Multi-sensor Towed Array Detection System (TEMTADS), developed by the Naval Research Laboratory and Nova Research



Land systems have demonstrated that they can now successfully detect, classify, and geolocate UXO among clutter

Contact: scazares@ida.org

Over the past several decades, the DoD has invested in the research and development of novel systems to find UXO buried in the ground. For example, the TEMTADS, shown here, collects data over the ground as an operator pushes it in a lawnmower-like pattern. The TEMTADS then uses straightforward machine learning methods to process that data in order to:

- Detect objects buried in the ground and
- Classify them as “likely UXO” versus “likely Clutter”.

IDA researchers have designed experiments for the DoD to demonstrate how well different systems can detect and classify UXO on *land*. This is largely done by measuring every single object in the test site, in order to collect ground truth. Ground truth is information about the *true* locations of all objects buried in the ground, along with their *true* labels—UXO versus clutter. The ground truth is then compared back to the system alarms in order to construct ROC curves, like this:

- Each point on this ROC curve corresponds to a different classification threshold that the system *could have* used to differentiate between “likely UXO” versus “likely clutter” objects.
- A vertical grey line is drawn through each point, to indicate the 95% confidence internal around that point’s Pd value. Each point’s confidence interval was calculated with the beta distribution to the binomial distribution, with no adjustments for multiple comparisons.
- The blue dot is the *prospective* classification threshold—the threshold selected by the system demonstrators *during* the experiment, without *any* knowledge of ground truth.

The blue dot on this ROC curve can quickly tell a very good story about the system performance: This system was able to correctly classify 100% of the UXO (at the top of the vertical axis) while generating only about 230 false alarms, less than half (less than halfway to the right of the horizontal axis) of the almost 600 false alarms that would have otherwise been produced if no classification algorithms were used at all.

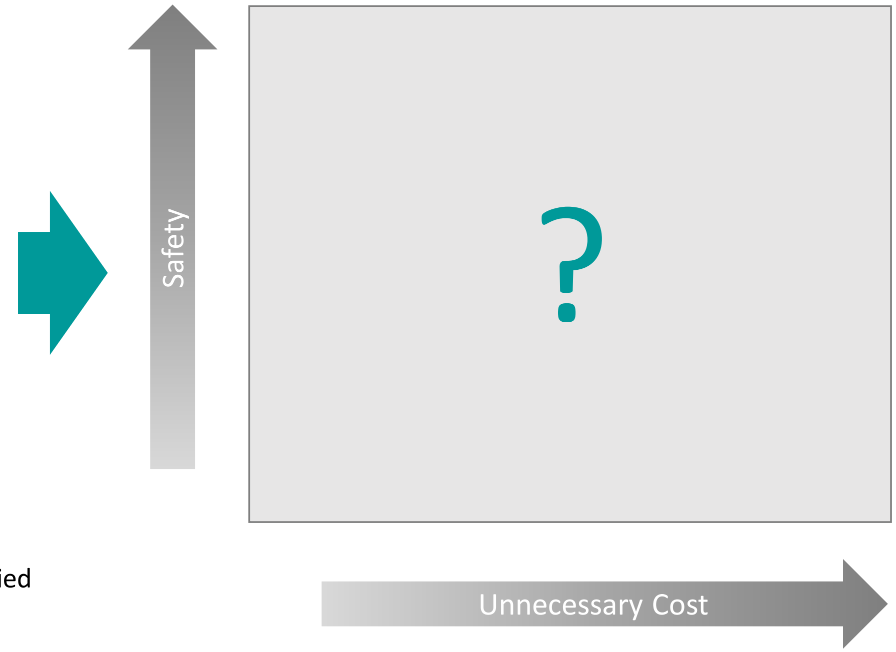
Similar results have also been demonstrated in a wide variety of challenging terrains and conditions, on land.

TEMTADS photo and ROC Curve figure reproduced from:

Cazares, Shelley, Elizabeth Ayers, and Michael Tuley. 2017. UXO Live Site Classification Demonstrations: A Retrospective Study. Presented at *SERDP & ESTCP Symposium 2017*, Washington DC, 28–30 November 2017. (DD1910 cleared for open publication 3 Nov 2017)

IDA **Underwater** Remediation of UXO

Multi-Sensor Towbody (MuST), developed by the Applied Physics Laboratory, University of Washington



Underwater systems for detection/classification/geolocation of UXO among clutter are difficult to demonstrate and assess with ROC curves

Contact: scazares@ida.org

The DoD has now turned its eye to the sea.

Systems are now being developed to deploy sensors in the *water*, such as those used by the MuST system circled in yellow on the left. The MuST is towed by a boat in a lawnmower-like pattern through the water. IDA researchers are now designing experiments to demonstrate how well these underwater systems can detect and classify UXO that are lying on or buried in the seabed floor.

Here's where things get tricky: Designing and executing these experiments is much more difficult *underwater* than on *land*, due to the additional safety, logistical, and engineering constraints of the underwater environment. In short, ROC curves are much more difficult to construct.

Underwater remediation system photos and ROC curve reproduced from:

Williams, Kevin, Tim Marston, Dana Woodruff, Shelley Cazares. 2021. Results from An Informal Demonstration of a Buried UXO Detection, Classification, and Geo-location System. Presented at the 2021 Underwater Acoustics Conference and Exhibition, virtual, 21–24 June 2021. (DD1910 cleared for open publication 24 May 2021)

Challenges in Underwater UXO Demonstrations

Challenges	Mitigations/Assumptions
• Greater regulatory hurdles (permit restrictions due to shipping lanes and flora/fauna protections)	• Carefully select test site and season • Reduce time objects are emplaced in seabed floor
• Potential movement of emplaced objects between ground truthing and demonstration (due to waves and currents, dragging by ship anchors, etc)	• Carefully select test site and season • Reduce time b/t ground truthing and demonstration • Assume objects do not move in time between
• Less precise geolocation systems (for ground truth and alarms)	• Larger detection halo radius R must be used • All munitions must be emplaced ≥2R apart
• More limited resources (time, funds) to excavate detected objects that were not emplaced	• Assume the only UXO and clutter in test area were those that were emplaced • Plot FP (not Pfa) on horizontal axis of ROC curve
• Diver safety concerns during emplacement and ground truthing	

Contact: scazares@ida.org

The following slides will explain some of the real-world, practical challenges to underwater demonstrations that make it difficult to construct and interpret the textbook ROC curves. We'll also explain what we *do* about it—what mitigations and assumptions we make to address those challenges.

IDA **Challenges** in Underwater UXO Demonstrations

Challenges	Mitigations/Assumptions
• Greater regulatory hurdles (permit restrictions due to shipping lanes and flora/fauna protections)	• Carefully select test site and season • Reduce time objects are emplaced in seabed floor
• Potential movement of emplaced objects between ground truthing and demonstration (due to waves and currents, dragging by ship anchors, etc)	• Carefully select test site and season • Reduce time b/t ground truthing and demonstration • Assume objects do not move in time between
• Less precise geolocation systems (for ground truth and alarms)	• Larger detection halo radius R must be used • All munitions must be emplaced ≥2R apart
• More limited resources (time, funds) to excavate detected objects that were not emplaced	• Assume the only UXO and clutter in test area were those that were emplaced • Plot FP (not Pfa) on horizontal axis of ROC curve
• Diver safety concerns during emplacement and ground truthing	

Contact: scazares@ida.org

First, there are often much greater regulatory hurdles when operating underwater. This is due to permit restrictions to avoid shipping lanes, as well as to protect underwater flora and fauna, including coral, kelp beds, and other marine life.

To overcome these hurdles, we must carefully select the test site as well the month or season of the year in which we conduct our demonstration.

Once the test site and season are chosen, the first step of any demonstration is to have a scuba diving team emplace objects in the test site, in or on the seabed floor, so that the novel systems can have the opportunity to detect and classify them during the demonstration. These emplaced objects include both UXO (inert or surrogate UXO) and clutter (old scuba tanks, ship anchors, crab pots, and so forth). Permit restrictions may limit the amount of time we can leave those objects in or on the seabed floor. Therefore, the more we can compress our demonstration schedule, the better.

These two mitigations are not for free, though. Being picky about our test site and demonstration season, as well as compressing our demonstration schedule, can lead to increased demonstration *costs*. Much more funding is needed to conduct *underwater* demonstrations than on land, and these mitigations are just two reasons why.

IDA **Challenges** in Underwater UXO Demonstrations

Challenges	Mitigations/Assumptions
• Greater regulatory hurdles (permit restrictions due to shipping lanes and flora/fauna protections)	• Carefully select test site and season • Reduce time objects are emplaced in seabed floor
• Potential movement of emplaced objects between ground truthing and demonstration (due to waves and currents, dragging by ship anchors, etc)	• Carefully select test site and season • Reduce time b/t ground truthing and demonstration • Assume objects do not move in time between
• Less precise geolocation systems (for ground truth and alarms)	• Larger detection halo radius R must be used • All munitions must be emplaced ≥2R apart
• More limited resources (time, funds) to excavate detected objects that were not emplaced	• Assume the only UXO and clutter in test area were those that were emplaced • Plot FP (not Pfa) on horizontal axis of ROC curve
• Diver safety concerns during emplacement and ground truthing	

The objects emplaced in the test site for the demonstration can lead to other challenges, as well.

On *land*, the emplaced objects stay put. *Underwater*, though, they can move. This movement could occur anytime between emplacement and removal. Of particular concern, though, is when the objects move *in between* collecting ground truth and the demonstration itself. If the objects move in between, then the ground truth is no longer true, and it becomes very difficult, if not impossible, to objectively score the demonstration.

One way to reduce the risk of objects moving is to carefully choose the location of a test site so that the seabed floor sediments make it more difficult for objects to move (i.e., mud, not sand). We can also carefully choose the month or season of the demonstration to make it less likely that the water currents will be strong enough to move the objects through the sediment.

Another way to mitigate this risk is to compress the demonstration schedule, leaving little opportunity for the emplaced objects to move in between the ground truthing and the demonstration.

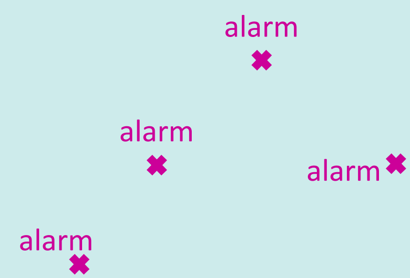
These two mitigations are similar to those we already employ to address the additional regulatory hurdles of operating in the underwater environment—as discussed on the previous slide.

Even with these mitigations, there is still a residual risk that the emplaced objects can move. However, the show must go on. Therefore, to construct our ROC curves, we must simply *assume* that the emplaced objects do *not* move in the time between ground truthing and the demonstration.

IDA **Challenges** in Underwater UXO Demonstrations

Challenges	Mitigations/Assumptions
• Greater regulatory hurdles (permit restrictions due to shipping lanes and flora/fauna protections)	
• Potential movement of emplaced objects between ground truthing and demonstration (due to waves and currents, dragging by ship anchors, etc)	
• Less precise geolocation systems (for ground truth and alarms)	
• More limited resources (time, funds) to excavate detected objects that were not emplaced	
• Diver safety concerns during emplacement and ground truthing	

Bird's Eye View of Test Site



Contact: scazares@ida.org

The next challenge is technical in nature. Less precise geolocation instruments are available in underwater environments. On *land*, we can use a Real Time Kinematic (RTK) Global Positioning System (GPS) to collect extremely precise information about the locations the objects were emplaced (the ground truth) and the locations detected by the system during the demonstration (the alarms). However, GPS signals do not travel well through water. Therefore, different types of geolocation instruments must be used *underwater*, and they usually have precision levels that are one to two orders of magnitude worse than on land.

Therefore, our scoring scheme must be flexible enough to handle this increased geolocation error.

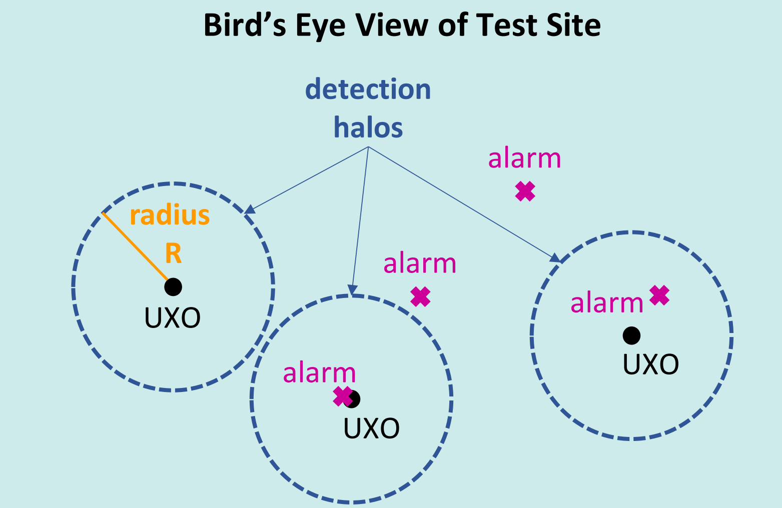
For example, here we show a bird's eye view of a test site. Imagine looking down from the sky, through the water, and to the seabed floor. And imagine that a system collected data over this test area during the demonstration, processed that data, and generated four alarms—four detected objects classified as "likely UXO".

Challenges in Underwater UXO Demonstrations

Challenges	Mitigations/Assumptions
• Greater regulatory hurdles (permit restrictions due to shipping lanes and flora/fauna protections)	**Bird's Eye View of Test Site**
• Potential movement of emplaced objects between ground truthing and demonstration (due to waves and currents, dragging by ship anchors, etc)	
• Less precise geolocation systems (for ground truth and alarms)	
• More limited resources (time, funds) to excavate detected objects that were not emplaced	
• Diver safety concerns during emplacement and ground truthing	

We can score the performance of the system by comparing its alarms to ground truth—the locations where inert UXO were *truly* emplaced. In this case, it's easy to see that the left UXO was missed, and the middle UXO was found by an alarm. But what about the right UXO—was it found? It's a little far from the right-most alarm—but should we consider this "close enough", bearing in mind the large geolocation error of both the alarms and the emplaced UXO? And what about the two alarms in the top half of figure, both of which are nowhere near a true UXO—are these *false* alarms?

Challenges in Underwater UXO Demonstrations

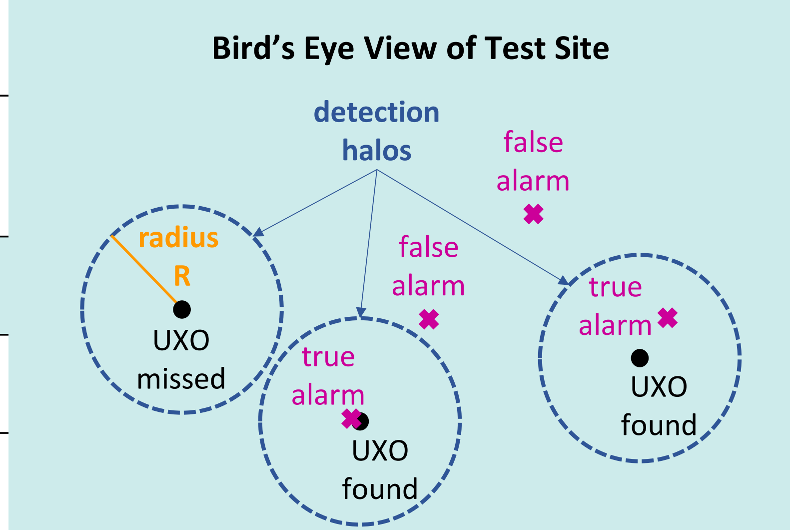
Challenges	Mitigations/Assumptions
• Greater regulatory hurdles (permit restrictions due to shipping lanes and flora/fauna protections)	
• Potential movement of emplaced objects between ground truthing and demonstration (due to waves and currents, dragging by ship anchors, etc)	
• Less precise geolocation systems (for ground truth and alarms)	
• More limited resources (time, funds) to excavate detected objects that were not emplaced	
• Diver safety concerns during emplacement and ground truthing	

Contact: scazares@ida.org

This is where the concept of *detection halos* come in. Detection halos are imaginary circles centered around each emplaced UXO. They all have the same radius of R.

IDA Challenges in Underwater UXO Demonstrations

Challenges	Mitigations/Assumptions
• Greater regulatory hurdles (permit restrictions due to shipping lanes and flora/fauna protections)	
• Potential movement of emplaced objects between ground truthing and demonstration (due to waves and currents, dragging by ship anchors, etc)	
• Less precise geolocation systems (for ground truth and alarms)	
• More limited resources (time, funds) to excavate detected objects that were not emplaced	
• Diver safety concerns during emplacement and ground truthing	



Contact: scazares@ida.org

We score a UXO as *found* (a True Positive) if an alarm is within R of it, such as the two true alarms matching up with the two found UXO in the middle and right of this figure.

In this simple cartoon, we also have two False Positives—the two false alarms in the top half of the figure.

We also have one False Negative. One out of three UXO were missed, resulting in a Pd of 2/3, or 67%, in this cartoon.

Challenges in Underwater UXO Demonstrations

Challenges	Mitigations/Assumptions
• Greater regulatory hurdles (permit restrictions due to shipping lanes and flora/fauna protections)	• Carefully select test site and season • Reduce time objects are emplaced in seabed floor
• Potential movement of emplaced objects between ground truthing and demonstration (due to waves and currents, dragging by ship anchors, etc)	• Carefully select test site and season • Reduce time b/t ground truthing and demonstration • Assume objects do not move in time between
• Less precise geolocation systems (for ground truth and alarms)	• Larger detection halo radius R must be used • All munitions must be emplaced ≥2R apart
• More limited resources (time, funds) to excavate detected objects that were not emplaced	• Assume the only UXO and clutter in test area were those that were emplaced • Plot FP (not Pfa) on horizontal axis of ROC curve
• Diver safety concerns during emplacement and ground truthing	

Contact: scazares@ida.org

In *land*-based demonstrations, our detection halo radius R could theoretically be as small as a few *centimeters*. In practice, though, we set R to be around 20–30 centimeters, approximately the width the bomb squad's shovel when they carefully dig up and neutralize a buried UXO.

However, in *underwater* demonstrations, our R must be much larger, for all of the reasons discussed in the previous slide—sometimes 2 or 3 *meters*.

Furthermore, to avoid any strange corner cases in scoring, we must also ensure that all inert UXO are emplaced farther than 2R away from each other. As technology matures, and system become able to detect and classify more closely-spaced UXO, we'll need to relax this constraint. Therefore, future demonstrations will require additional rules to handle these corner cases.

IDA Challenges in Underwater UXO Demonstrations

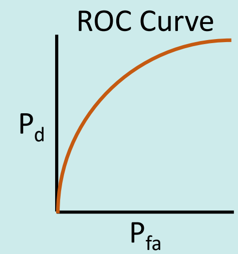
Challenges

- Greater regulatory hurdles (permit restrictions due to shipping lanes and flora/fauna protections)
- Potential movement of emplaced objects between ground truthing and demonstration (due to waves and currents, dragging by ship anchors, etc)
- Less precise geolocation systems (for ground truth and alarms)
- More limited resources (time, funds) to excavate detected objects that were not emplaced
- Diver safety concerns during emplacement and ground truthing

Mitigations/Assumptions

Vertical axis of ROC curve:

$$P_d = \frac{TP}{TP + FN} = \frac{\#\ \text{True Alarms}}{\#\ \text{True UXO}} \approx \frac{\#\ \text{True Alarms}}{\mathbf{\#\ Emplaced\ UXO}}$$



Horizontal axis of ROC curve:

$$P_{fa} = \frac{FP}{FP + TN} = \frac{\#\ \text{False Alarms}}{\#\ \text{True Clutter}} \approx \frac{\#\ \text{False Alarms}}{\mathbf{\#\ Detections} - \mathbf{\#\ Emplaced\ UXO}}$$

Contact: scazares@ida.org

Our last two sets of challenges go together. In an underwater demonstration, there is never enough time or funding to dig up every single object in the test area, in order to construct full ground truth. This *is* possible to do on land, with strong sponsorship to provide the necessary time and funding. But in an underwater environment, it's largely impossible. And there are safety concerns about how much time the scuba divers can spend underwater to do this job, anyway.

Therefore, we must reconsider our definition of the metrics we plot on the vertical and horizontal axes of our ROC curves.

Probability of Detection (Pd) is usually plotted on the *vertical* axis of the ROC curve. Textbooks define Pd as the number of true alarms divided by the number of true UXO. However, it is impossible to know how many UXO are *truly* present at a test site. We know how many UXO were *emplaced* at the site for the purpose of the demonstration. But we don't know if there were any *other* UXO present at the site *before we even arrived*—remnants of previous military training camps or test sites. By carefully selecting the location of our test sites, though, we can make a pretty good assumption about the denominator of our Pd metric—that the only UXO present at the test site were those that were emplaced for the demonstration.

A similar assumption can be made for Probability of False Alarm (Pfa) on the *horizontal* axis of the ROC curve. Textbooks define Pfa as the number of false alarms divided by the number of true clutter. There are many reasons why it's difficult to estimate the denominator of this metric, especially for demonstrations like these, in which we are using a ROC curve to assess both *detection* and classification performance. However, we can address this challenge by making some assumptions about the denominator of our Pfa metric– that the only clutter present at the test site were detected objects that were *not* emplaced UXO.

IDA **Challenges** in Underwater UXO Demonstrations

Challenges

- Greater regulatory hurdles (permit restrictions due to shipping lanes and flora/fauna protections)
- Potential movement of emplaced objects between ground truthing and demonstration (due to waves and currents, dragging by ship anchors, etc)
- Less precise geolocation systems (for ground truth and alarms)
- More limited resources (time, funds) to excavate detected objects that were not emplaced
- Diver safety concerns during emplacement and ground truthing

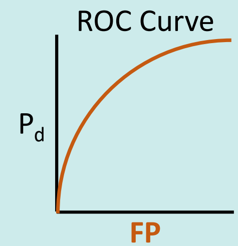
Mitigations/Assumptions

Vertical axis of ROC curve:

$$P_d = \frac{TP}{TP + FN} = \frac{\#\ \text{True Alarms}}{\#\ \text{True UXO}} \approx \frac{\#\ \text{True Alarms}}{\#\ \textbf{Emplaced UXO}}$$

Horizontal axis of ROC curve:

FP = # False Alarms



Contact: scazares@ida.org

Alternatively, we can skip using Pfa completely, and plot FP on the horizontal axis of our ROC curve—the number of false positives, also known as false alarms, with no denominator. There are many advantages to doing this—namely, the FP metric can be more easily translated into the unnecessary dollar costs spent investigating alarms that turn out to be false.

IDA **Challenges** in Underwater UXO Demonstrations

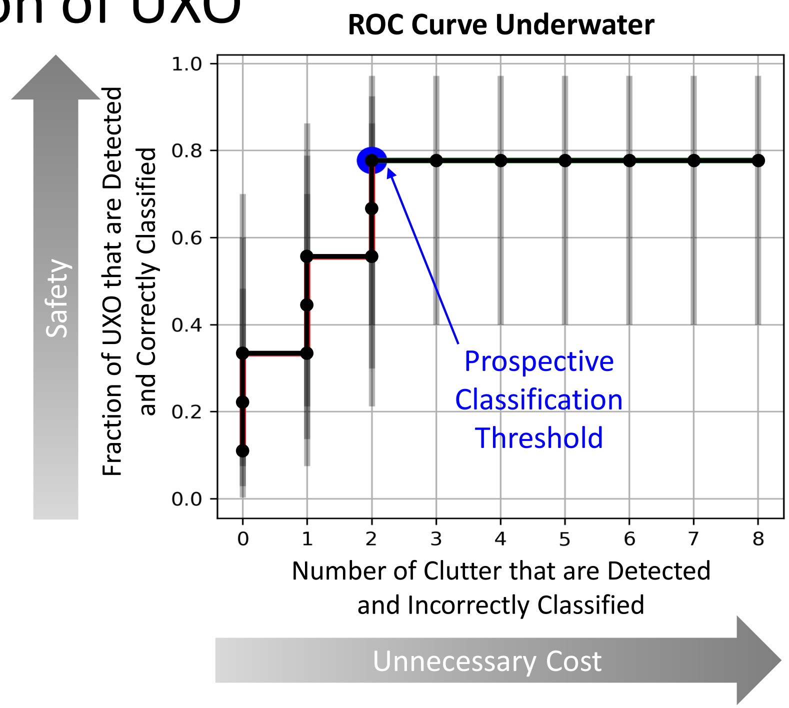
Challenges	Mitigations/Assumptions
• Greater regulatory hurdles (permit restrictions due to shipping lanes and flora/fauna protections)	• Carefully select test site and season • Reduce time objects are emplaced in seabed floor
• Potential movement of emplaced objects between ground truthing and demonstration (due to waves and currents, dragging by ship anchors, etc)	• Carefully select test site and season • Reduce time b/t ground truthing and demonstration • Assume objects do not move in time between
• Less precise geolocation systems (for ground truth and alarms)	• Larger detection halo radius R must be used • All munitions must be emplaced ≥2R apart
• More limited resources (time, funds) to excavate detected objects that were not emplaced	• Assume the only UXO and clutter in test area were those that were emplaced • Plot FP (not Pfa) on horizontal axis of ROC curve
• Diver safety concerns during emplacement and ground truthing	

Contact: scazares@ida.org

All of these mitigations and assumptions are needed to overcome the challenges of designing and running experiments in an underwater environment.

IDA Underwater Remediation of UXO

Multi-Sensor Towbody (MuST), developed by the Applied Physics Laboratory, University of Washington



With careful planning and some key assumptions, ROC curves can still be crafted for demonstrations of underwater UXO remediation systems

Contact: scazares@ida.org

With these mitigations and assumptions in place, ROC curves can still be crafted to quickly tell the story of how well novel systems can detect and classify UXO versus clutter in *underwater* environments. The blue dot on this ROC curve tells us that this system can correctly detect and classify almost 80% of the emplaced UXO while generating only two false alarms, an imperfect but promising balance between safety and unnecessary costs. Further research and development is on-going.

Underwater remediation system photos and ROC curve reproduced from:

Williams, Kevin, Tim Marston, Dana Woodruff, Shelley Cazares. 2021. Results from An Informal Demonstration of a Buried UXO Detection, Classification, and Geo-location System. Presented at the 2021 Underwater Acoustics Conference and Exhibition, virtual, 21–24 June 2021. (DD1910 cleared for open publication 24 May 2021)

IDA **Testbeds** for Underwater UXO Demonstrations

The DoD is preparing four different testbeds to demonstrate underwater UXO remediation systems

Contact: scazares@ida.org

For now, the DoD is currently preparing four different testbeds to demonstrate underwater UXO remediation systems. The first official blind test is scheduled for September 2021 in Sequim Bay, WA. Stay tuned for results…

World map obtained from:

https://www.1001freedownloads.com/free-clipart/world-map-7

IDA

Further Information

Recap of Land Demonstrations

https://fasttimesonline.co/the-live-site-demonstrations-for-advanced-geophysical-classification-of-terrestrial-unexploded-ordenance/

Planning for Underwater Demonstrations

https://www.ida.org/research-and-publications/publications/all/s/sc/scoring-underwater-demonstrations-for-detection-and-classification-of-unexploded-ordnance-uxo

Further Discussion
Shelley Cazares
703 845 6792
scazares@ida.org

Any Questions?

Please look online for additional information, including a recap of previous land demonstrations and plans for upcoming underwater demonstrations. Please also reach out to us directly with any questions.

STATISTICS INFORMS ENVIRONMENTAL REMEDIATION

STATISTICAL SIGNIFICANCE

Statisticians help the U.S. Department of Defense clean up after itself. They design experiments to show that new sensors and software can find munition duds from previous military activity, so that the duds can be safely removed from the land or sea. These experiments help remediation teams and environmental regulators decide which sensors and software they can trust to clean up our environment for future generations.

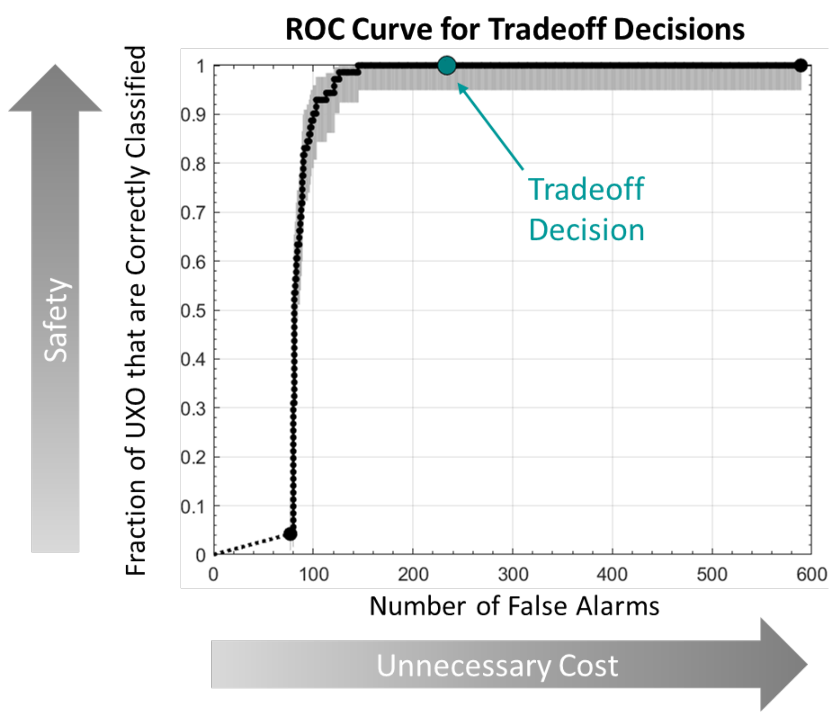
Environmental Remediation

UNEXPLODED ORDNANCE (UXO):
UXO are duds—munitions that were armed and fired but failed to explode. But they still *can* explode, even decades later. Cities, towns, and countrysides across the world are still rattled by WWI and WWII era UXO, unearthed from the ground or rolled up on the beach. And the problem persists—militaries make more and more UXO each day.

The U.S. Department of Defense (DoD) is committed to UXO remediation (cleanup). Over the past few decades, the DoD has funded new sensors and software to help find UXO in the ground. Bomb squads can then safely and quickly remove the UXO before anyone is harmed.

DESIGNING EXPERIMENTS, DRAWING CONCLUSIONS:
Statistics helped the DoD design and score a series of experiments to demonstrate how well different combinations of sensors and software could help clean up real pieces of land contaminated with real UXO. Researchers drew Receiver-Operating Characteristic (ROC) Curves to summarize the performance of each sensor/software combination. These curves helped UXO remediation teams and environmental regulators weigh the tradeoffs between finding all UXO (upwards) and declaring a false alarm (rightwards). The optimum tradeoff decision lies somewhere in between. A good sensor/software combination can deliver the best of both worlds (blue dot): full safety (at the very top) with low unnecessary costs (towards the left). ROC curves like these showed that the best systems for land-based UXO remediation use electromagnetic induction sensors and template matching software.

UNDERWATER, THE FINAL FRONTIER:
The DoD has now turned its eye to the sea. UXO also contaminate underwater sites. The DoD now sponsors research and development into additional sensors and software for the marine environment, including low-frequency acoustics and deep learning. Design of experiments is difficult underwater, due to additional safety, logistical, and engineering constraints. Using statistics, the DoD hopes to one day soon deliver robust environmental technology to provide safer, faster UXO remediation to keep us safe in our oceans, rivers, and beaches, free from the threat of UXO.



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

1. REPORT DATE *(DD-MM-YYYY)*: July 2021

2. REPORT TYPE: FINAL

3. DATES COVERED *(From - To)*:

4. TITLE AND SUBTITLE: Storytelling with Receiver Operating Characteristic (ROC) Curves for Environmental Remediation of Unexploded Ordnance (UXO) (Presentation)

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)

Cazares, Shelley M.
Bartel, Jacob B.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Institute for Defense Analyses
Systems and Analyses Center
4850 Mark Center Drive
Alexandra, VA 22311-1882

8. PERFORMING ORGANIZATION REPORT NUMBER: IDA Document NS D-22766

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

SERDP/ESTCP
Munitions Response
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 (22 July 2021).

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Receiver operating characteristic (ROC) curves are often used to assess the performance of binary classification systems, allowing stakeholders to understand tradeoffs between Type I (false positive) and Type II (false negative) errors. This works well in textbook cases. In real-world experiments, however, ROC curves can have unexpected subtleties that make them difficult to construct and interpret. For example, the Department of Defense is sponsoring the development of novel sensors and software to identify unexploded ordnance (UXO) in the midst of clutter. UXO are duds—munitions that were previously armed and fired but failed to explode. UXO can still pose a risk of detonation even decades later, threatening the safety of nearby humans, animals, vegetation, and structures. Conducting blind tests to demonstrate finding UXO is fraught with safety, logistical, and cost constraints that make it difficult to construct the textbook ROC curves. Yet, with careful planning and a few key assumptions, ROC-like curves can still be crafted to quickly tell the story of how well novel sensors and software can detect, classify, and locate UXO versus clutter in terrestrial and underwater experiments.

15. SUBJECT TERMS

environmental remediation; Probability of Detection (Pd); Probability of False Alarm (Pfa); Receiver Operating Characteristic (ROC) curve; unexploded ordnance (UXO)

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Bradley, David
Unclassified	Unclassified	Unclassified	Same as Report	47	19b. TELEPHONE NUMBER *(Include area code)* (571) 372-6388

Reset

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18