Mine Neutralization Case Study: Applying Experimental Design to Operational Testing

Kelly McGinnity

Institute for Defense Analyses

August 4, 2014

Motivation

- **Operational testing (OT):**
  - Purpose to provide objective assessments of how systems perform under realistic combat conditions
  - Accomplished by evaluating systems in operational scenarios when employed by units against realistic threats

Realism of OT events presents unique challenges to the process of applying sound statistical design and analysis techniques

- **Mine neutralization system case study highlights the challenges of testing in restricted conditions and outlines proposed solutions using statistical methods**

Outline

- **Mine Neutralization System Description**

- **Designing the Test**
  - Response variables / Factors and Levels of Interest
  - Survival Model Considered
  - Challenges and Limitations
  - Sequential Testing Solution

- **Calculating Power**
  - Monte Carlo Approach
  - Results

Mine Neutralization System Description

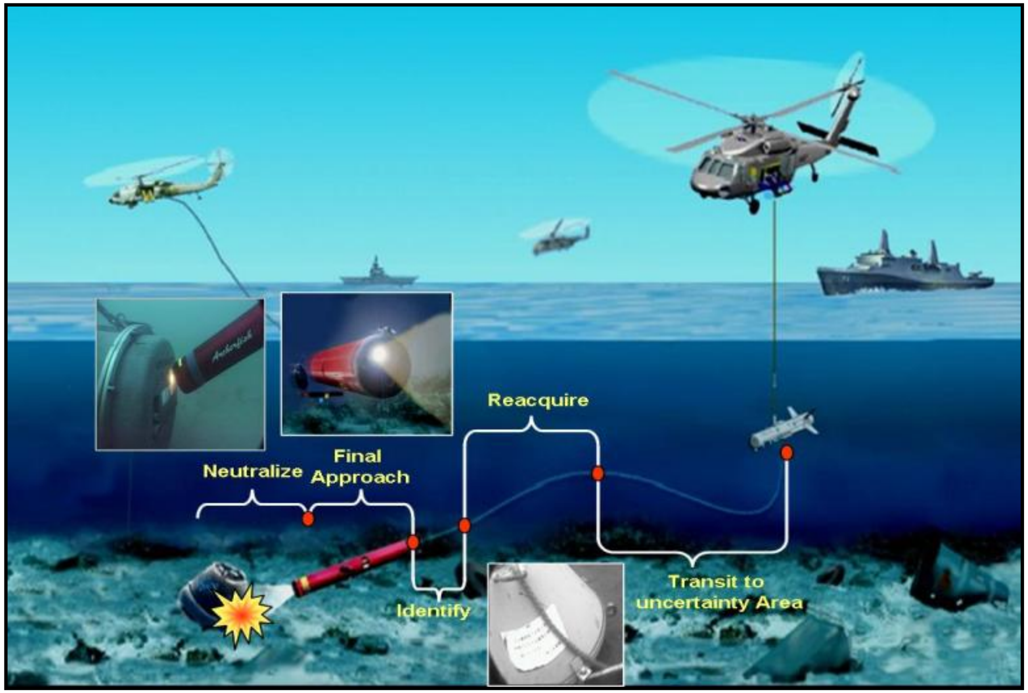
- **Description:** Three major subsystems: (1) Helicopter, (2) Expendable Neutralizer, deployed from a (3) Launch and Handling System (LHS) and controlled from the helicopter through a fiber optic cable.

- **Platform:** Deployed from Littoral Combat Ship (LCS)

- **Employment:** Identify and explosively neutralize moored mines previously detected by sonar and explosively neutralize bottom mines identified by sonar. Each of up to four neutralizers is launched, one at a time, without having to recover the LHS. The neutralizer provides sonar and video data a sensor operator who performs positive identification before commanding warhead detonation.

Neutralizer Loaded on Helicopter

Suspended By Tow Cable



Variants of the Weapon

Inert, recoverable

Live

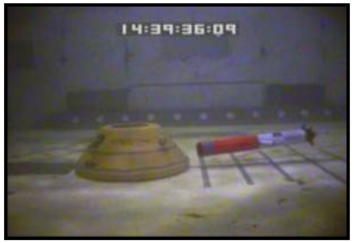
Outline

- Mine Neutralization System Description
- **Designing the Test**
  - Response variables / Factors and Levels of Interest
  - Survival Model Considered
  - Challenges and Limitations
  - Sequential Testing Solution
- Calculating Power
  - Monte Carlo Approach
  - Results

Response Variables

- **Main goal of testing is to determine the effectiveness of the system in reacquiring and neutralizing mines**
- **Probability of detection and reacquisition**
- **Probability of neutralization**
- **Time to Neutralize (censored)**
  - 15 min = requirement
  - 45 min = search cutoff – maximum neutralizer run time
  - Continuous variable → more informative than simple binary measures
  - Censored and right skewed → presents challenges

Mine Neutralized



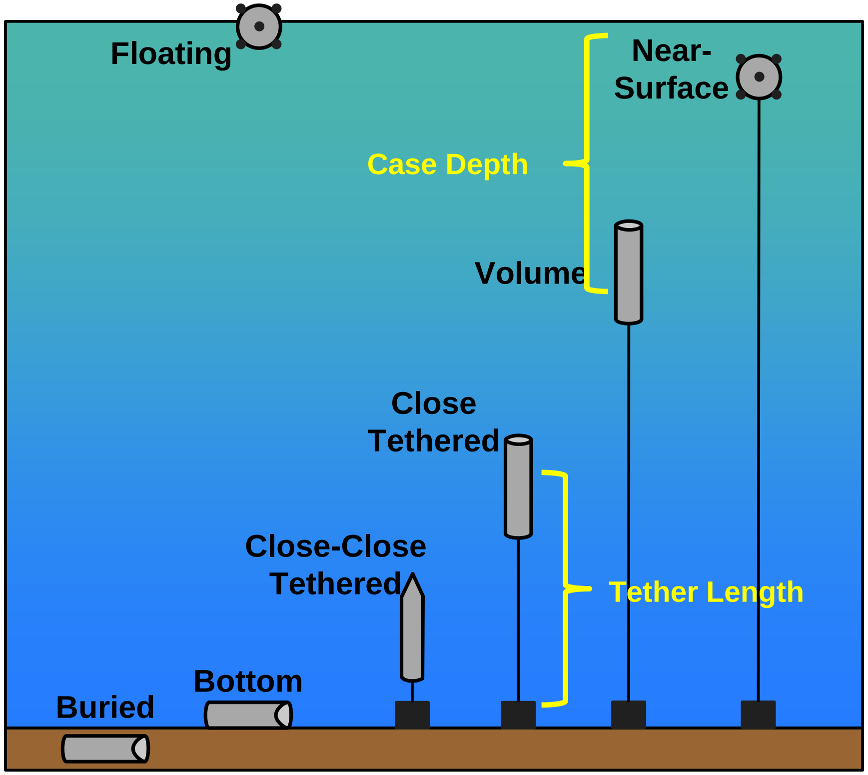
Maneuver to Neutralize ← Final Approach ← Reacquire ← Transit ← Weapon Launched

Time is recorded if steps completed within 45 minutes

Run is censored if steps NOT completed within 45 minutes

Factors and Levels

- **Mine Type**
  - Moored / In-Volume
  - Bottom
  - Stealth
- **Tether Length**
- **Case Depth**
- **Sonar Localization Accuracy**
- **Environmental Conditions**
  - Ocean Depth
  - Current
  - Sea State
  - Bottom State
  - Water Clarity
  - Wind Speed



Statistical Model

- **Time to Neutralize can be fit using lognormal survival model with censoring**
  - $L = \prod_{i=1}^{n}[f(t_i)]^{\delta_i}\,[S(t_i)]^{1-\delta_i}$, where

    $f(t)$ is the lognormal probability density function,
    $S(t)$ is the lognormal survival function,
    δ is a censoring indicator (1 if the time is observed, 0 if not), and
    n is the number of runs in the experiment
  - Assumes that the mine would *eventually* be detected given infinite time

- **Considerations**
  - Many environmental factors may be correlated → check correlations / VIFs to ensure model is valid

Test Design Challenges

- **Ideal Design**
  – Same exact design with identical conditions replicated in all test event locations
- **Reality**
  – Environmental conditions not controllable and change with season
  – Limited number of test ranges (mine fields)
  – Restraints on operations (live weapon allowance, availability of surrogate targets)
  – Most events are shore-based (large distance to deep water conditions makes these runs time-consuming and expensive)
- **Can address some if not all of these with smart statistical design**

Solution: Sequential Testing

- **Sequential Testing:**
  - Use outcomes/runs gathered from previous test to inform/help plan for the future events
  - Combine the results of multiple test events and locations to make final evaluation

Location (Season)	# of Test Events	Total # of Runs	Conditions Likely to be Captured
Panama City, FL (Spring / Summer)	4 (2 Completed, 2 Planned)	84	Clear Seas, Low Current, Smooth Sandy Bottom, Low Sea State
Norfolk, VA (VACAPES) (Spring / Summer)	1 (Completed)	31	Less Clear, Higher Current, Various Bottom Conditions
San Diego, CA (Fall)	1 (Planned)	4	Higher Sea State
Riviera Beach, FL (TBD)	1 (Planned)	24	Medium Current

Notional Design Matrices

Run Number	Test Location	Mine Type	Case Depth (ft)	Tether Length (ft)	Water Depth (ft)	Error Percentile	Shot type	Current	Sea State	Bottom State	Water Clarity	Wind Speed
1	VACAPES	Stealth	75	0	75	10	Live	High	2	Rough	Low	Medium
2	VACAPES	Stealth	75	0	75	50	Live	High	1	Rough	Low	High
3	VACAPES	Bottom	100	0	100	90	Live	High	4	Rough	Low	High
4	VACAPES	Bottom	200	0	200	10	[illegible]	High	2	Rough	Low	Low
5	VACAPES	Stealth	200	0	200	10	[illegible]	Medium	2	Rough	Low	Low
6	VACAPES	Moored	190	10	200	[illegible]	[illegible]	[illegible]	2	Rough	Medium	Medium
7	VACAPES	Moored	295	20	[illegible]	[illegible]	[illegible]	[illegible]	2	Rough	Medium	High
8	VACAPES	Moored	270	25	[illegible]	[illegible]	[illegible]	[illegible]	3	Rough	Low	High
9	VACAPES	Moored	245	30	[illegible]	[illegible]	[illegible]	High	2	Rough	Medium	Low
10	VACAPES	Moored	50	150	[illegible]	[illegible]	Live	High	1	Rough	Low	Medium
11	VACAPES	Moored	85	285	[illegible]	50	Inert	High	3	Rough	Medium	Low

⋮

Run Number	Test Location	Mine Type	Case Depth (ft)	Tether Length (ft)	Water Depth (ft)	Error Percentile	Shot type	Current	Sea State	Bottom State	Water Clarity	Wind Speed
32	Panama City	Stealth	103	0	103	10	Inert	Low	[illegible]	[illegible]	[illegible]	Low
33	Panama City	Bottom	195	0	195	90	Inert	Low	[illegible]	[illegible]	[illegible]	Medium
34	Panama City	Bottom	100	0	100	10	Inert	Low	[illegible]	[illegible]	[illegible]	[illegible]
35	Panama City	Stealth	195	0	195	50	Inert	Low	1	[illegible]	[illegible]	[illegible]
36	Panama City	Moored	180	15	195	10	Inert	Low	1	Smooth	[illegible]	[illegible]
37	Panama City	Moored	180	15	195	50	Inert	Low	2	Smooth	[illegible]	[illegible]
38	Panama City	Moored	180	15	195	90	Inert	Low	2	Smooth	High	[illegible]
39	Panama City	Moored	60	190	250	10	Inert	Low	1	Smooth	High	[illegible]
40	Panama City	Moored	60	190	250	50	Inert	Low	2	Smooth	High	Low
41	Panama City	Moored	60	190	250	90	Inert	Low	1	Smooth	High	Low
42	Panama City	Moored	60	290	350	10	Inert	Low	2	Smooth	High	Medium
43	Panama City	Moored	75	525	600	10	Inert	Low	1	Smooth	High	Medium

⋮

Completed Test Events

Run Number	Test Location	Mine Type	Case Depth (ft)	Tether Length (ft)	Water Depth (ft)	Error Percentile	Shot type	Current	Sea State	Bottom State	Water Clarity	Wind Speed
71	San Diego							High	4	Smooth		
72	San Diego							Medium	4	Smooth		
73	San Diego							High	4	Smooth		
74	San Diego							High	4	Smooth		
75	Riviera Beach							Medium	1	Smooth		
76	Riviera Beach							Medium	1	Smooth		
77	Riviera Beach							Medium	1	Smooth		
78	Riviera Beach							Medium	1	Smooth		

Planned Future Events

⋮

Outline

- **Mine Neutralization System Description**
- **Designing the Test**
  - Response variables / Factors and Levels of Interest
  - Survival Model Considered
  - Challenges and Limitations
  - Sequential Testing Solution
- **Calculating Power**
  - Monte Carlo Approach
  - Results

Power Calculation Challenges

- **Power is an important part of determining test adequacy**
  - Tells you the probability of detecting differences in performance between levels of a factor, given that differences occurred
  - Depends on many inputs including number/type of factors, the model of interest, confidence level, effect size, and variability of the response
  - Low power for certain factor(s) can drive the need for testing in those conditions in future phases of test
- **Standard DOE software assumes the response variable is normally distributed and every observation is observed**
  - Time to neutralize is right-skewed (lognormal)
  - Possible unobserved (censored) data points
- **Monte Carlo approach developed**
  - Comes with its own unique challenges

Simulation Inputs and Considerations

- **Assume time to neutralize (T_N) is lognormally distributed**
  - Fix the scale parameter σ at 1*
- **Assume the median T_N under the null hypothesis is at the requirement**
  - Equates to setting the mean of the lognormal distribution under the null hypothesis to a value, μ_0, such that $e^{\mu_0} \approx 15$
  - Can also base the null hypothesis on a set censoring rate
- **Specify an operationally important effect size, Δ, to detect, e.g. an increase in median T_N of Δ minutes**
  - For a 2 level factor, equates to setting the mean under the alternative hypothesis to a value, μ_1, such that $e^{\mu_1} \approx 15 + \Delta$
  - Carefully consider effects for 3+ level factors; power very sensitive to direction of change and number of “active” levels
- **Ensure the censoring rate is reasonable under both hypotheses**
  - If P(*lognormal*(μ,1) > 45) is more than about ½, power analyses are dramatically affected
  - *Can consider adjusting σ to control censoring rate under the alternative

Subset of Power Results for Mine Neutralization System

- **Δ set to 10 minutes; α set to 0.2**
- **Sum to zero contrasts with two "active" levels (conservative power)**

Completed Tests Only

Effect	Power
Mine Type	0.62
Localization Error	0.75
Current	0.29
Sea State	0.44

With Future Events

Effect	Power
Mine Type	0.80
Localization Error	0.87
Current	0.60
Sea State	0.50

- **Power analyses document the risk associated with not collecting enough data under certain conditions**
  - Can drive the need for a new test event / location
- **Check correlations of design matrix prior to reporting power**
  - Two highly confounded effects should not both be included in the model!

Conclusions

- **Make the most of multiple-phased tests**
  - Many factors cannot be controlled
  - Employ smart sequential design to adequately cover all operational conditions
  - Combine results to make final evaluation

- **Perform power analyses to show the tradeoffs between risks and operational feasibility**
  - Continuous metrics better than binary, but make sure to analyze right-skewed, censored data appropriately
  - Proposed power simulation violates some statistical assumptions (i.e. the design is not completely randomized), but it still provides a framework we find useful

- **Future Work**
  - Further investigate censored data power analysis and how power is affected by the size / direction of the effect size, standard deviation of the lognormal distribution, censoring rate, etc.
  - Modify the simulation to allow for specification of Δ in terms of P(Neutralization)

- **BACKUP**

IDA

Simulation Details

1. Input the design matrix
2. Generate data under the alternative hypothesis (that changing levels of a certain factor causes a change of Δ in the response)
3. Censor any data points > 45
4. Fit a lognormal survival model using *all* factors AND a model *without* the factor causing the change
5. Perform a likelihood ratio test to determine if there is a significant difference between the fits of the two models
6. If the p-value < α, count as "correct"

Iterate many times

Power for that factor is simply the proportion correct!