SERDP

FINAL REPORT – PART B

Development of a Decision Tool: Underwater Munitions Expert System

Sarah Rennie
Johns Hopkins University Applied Physics Laboratory

March 2022

SERDP Project MR19-1126

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 31-03-2022	**2. REPORT TYPE** SERDP Final Report	**3. DATES COVERED** *(From - To)* *9/30/2021 - 9/30/2022*

4. TITLE AND SUBTITLE

Development of a Decision Tool: Underwater Munitions Expert System

5a. CONTRACT NUMBER
W912HQ-21-P-0047

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Sarah E Rennie

5d. PROJECT NUMBER
MR19-1126

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

Johns Hopkins Applied Physics
Laboratory
11100 Johns Hopkins Road
Laurel, Maryland 20723

8. PERFORMING ORGANIZATION REPORT NUMBER

FPS-R-22-1131

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

Strategic Environmental and
Development Program
4800 Mark Center Dr,Suite 17
Alexandra, VA 20003

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

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
MR19-1126

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The development of a computer software tool to predict the mobility and burial of underwater munitions is summarized. The Underwater Munitions Expert System (UnMES) is based on a Bayesian Network framework and acts to synthesize results from multiple research projects supported by the SERDP Munitions Response program investigating the behavior of underwater munitions on the seabed. The underlying concepts and philosophy influencing software design decisions used to construct the expert system prototypes as a guide for developers enacting further improvements to UnMES is documented. In addition to laying out how UnMES is assembled, discussion of the rationale behind certain choices is provided to inform future expert system advances, along with the statistical analysis of underwater mobility and burial measurements required to demonstrate predictive utility, with an emphasis on the specific need for further data collection and testing

15. SUBJECT TERMS
Munitions, Remediation, Bayesian Network, Probabilistic modeling, Burial, Migration

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Sarah E Rennie
a. REPORT unclassified	**b. ABSTRACT** unclassified	**c. THIS PAGE** unclassified	unclassified	34	**19b. TELEPHONE NUMBER** *(include area code)* 443-538-3089

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

Development of a Decision Tool: Underwater Munitions Expert System
MR19-1126 Final Report

S.E. Rennie
Johns Hopkins Applied Physics Laboratory

March 2022

Table of Contents

Acronyms and Symbols

API – Application Programmer Interface
B – burial depth of the UXO
BN – Bayesian Network
BS – Brier Score
C. I. - Confidence Interval
CPT – Conditional Probability Table(s)
CSM – Conceptual Site Model
D – Diameter of the UXO
ESTCP – Environmental Security Technology Certification Program
FRF – Field Research Facility
GIS – Geographic Information System
GoM – Gulf of Mexico
GUI – Graphical User Interface
KML – Keyhole Markup Language for geographic visualization and annotation
IPCC – Intergovernmental Panel on Climate Change
MBP – Mine Burial Program
MC – Monte Carlo
MEC – Munitions of Concern
MR – Munitions Response
N – Sample size
NRL-SSC – Naval Research Laboratory at Stennis Space Center
ONR – Office of Naval Research
PDT – probability distribution table
RPM – Remedial Project Manager
RPS – Ranked Probability Score
RPSS – Relative Ranked Probability Score
SERDP – Strategic Environmental Research and Development Program
UnMES – Underwater Munitions Expert System
USACE – United States Army Corps of Engineers
UXO – Unexploded Ordnance

α – confidence level, (1-Confidence Interval)
$\boldsymbol{k}$ – bins or states in the discreet probability distribution (PDT)
ε – tolerance (uncertainty) in estimates of probability
σ – standard deviation

Development of a Decision Tool: Underwater Munitions Expert System

Abstract

This report summarizes the development of a computer software tool to predict the mobility and burial of underwater munitions. The Underwater Munitions Expert System (UnMES) is based on a Bayesian Network framework and acts to synthesize results from multiple research projects supported by the SERDP Munitions Response program investigating the behavior of underwater munitions on the seabed. This document describes the underlying concepts and philosophy influencing software design decisions used to construct the expert system prototypes as a guide for developers enacting further improvements to UnMES. An accompanying report, the *Programmer's Guide to UnMES Construction*, provides technical coding details. In addition to laying out how UnMES is assembled, discussion of the rationale behind certain choices is provided to inform future expert system advances. Also included is a discussion of the statistical analysis of underwater mobility and burial measurements required to demonstrate predictive utility, with an emphasis on the specific need for further data collection and testing.

1. Introduction to the Underwater Munitions Expert System

The SERDP Munitions Response (MR) program is focused on developing innovative methods to characterize, remediate and sustainably administer sites contaminated with unexploded ordnance (UXO) or other munitions of concern [SERDP, 2010]. The possibility of buried UXO presents a significant challenge to site management, as they are difficult to detect, monitor, or remediate. Of particular concern is the potential for UXO to migrate from their current location into an area with high likelihood of contact with human receptors. Due to the complexity of the environment and UXO response, decision making under uncertainty is required. The objective of this effort is the construction of a predictive Underwater Munitions Expert System (UnMES) providing computer-based decision support for management of aquatic sites.

The Prototype Version of UnMES, was documented in Rennie, Brandt and Ligo [2019, hereafter RBL2019]. A streamlined version (V2) is discussed in the *Programmer's Guide to UnMES Construction* [Rennie, 2022, hereafter referred to as the *Programmer's Guide*], and is diagrammed in Figure 1.1. UnMES is based upon a Bayesian Network (BN), a graphical probability model that reflects current knowledge of the probable UXO migration and burial in an underwater environment. The difficulty of obtaining environmental data, unknown details of the physical processes, and the exact disposition of the munitions result in a high level of uncertainty for UXO location and subsequent behavior. A Bayesian Network is a useful method of modeling systems with complex relationships in a probabilistic manner. A BN captures the uncertainty in relationships between the variables, quantified as the spread in the probability distributions [Plant and Holland, 2011]. Predictions can be made with uncertainty in the inputs, and the uncertainty is propagated by probabilistic inference using Bayes theorem.

Each variable (node in the network) is represented by a probability distribution, connected by arrows (directed links) that symbolize dependent relationships, as illustrated in Figure 1.1. Each relationship is quantified by a conditional probability table (CPT) associated with the dependent node. In Figure 1.1, nodes in blue indicate environmental inputs, while nodes in red describe the UXO characteristics and conditions. The green nodes show output distribution predictions.

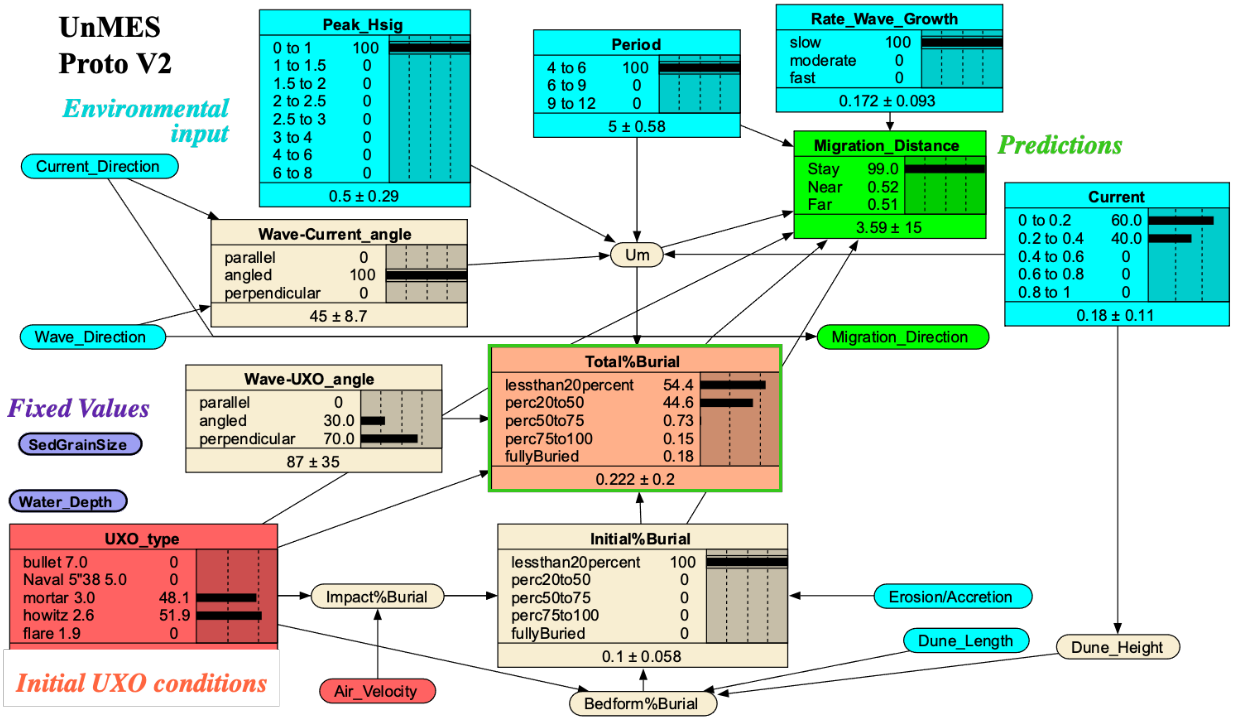


Figure 1.1 Bayesian Network for UnMES Prototype V2.

1.1 Software Design

When implemented in a practical BN, each variable (node) in the network is discretized into a set of bins, or states, so that the probability distribution is no longer a continuous function, but consists of a finite set of values. As such, a variable, or node, is represented by Probability Distribution Table (PDT). Within the BN, a relationship between nodes is characterized by a conditional probability table (CPT). A common approach to building a Bayesian Network is to train the CPT using a dataset of example cases. This requires a large number of data cases. However, due to the limited amount of field and laboratory observations applicable to UXO burial and mobility, an alternative approach was taken to construct UnMES. In this approach, deterministic models were developed to capture the first-order physics of the processes of interest, and then utilized to produce datasets of cases linking input conditions to output UXO behavior. Monte Carlo simulations employing these models are then run over the relevant combinations of input variable domains in order to output a large set of cases to train the CPTs in the expert system BN. Steps for this procedure are detailed in Section 2.1 of the *Programmer's Guide* [Rennie, 2022].

Storms are generally the primary drivers of burial and migration events. UnMES, as currently implemented, is designed to predict UXO response on the time scale of a single storm event. Bottom sitting objects during high-energy wave conditions will be subjected simultaneously to local burial processes and the potential for mobilization. Over an extended time period, under seasonal weather patterns with repeated storms, far-field morphologic changes to the seabed occur which can also influence UXO burial [Rennie *et al.,* 2019]. The version of UnMES reported on here focuses on short timescale processes and localized response. Several approaches to extending mobility burial and migration prediction were investigated during UnMES development, including time-stepped, or dynamic Bayesian Networks. However, it was found that the recorded uncertainty rapidly overwhelmed the predictions, so that the results had little operational utility. Future development of long-term predictions, simulating evolution of UXO migration and burial in response to multiple storms, will likely need to rely on improvements in regional numerical modeling, such as Xbeach [Palmsten *et al.,* 2021].

Community feedback indicates that scenario-based forecasting may be the preferred application for a management decision aid that predicts UXO mobility and burial [Rennie and Brandt, 2020]. Given the existing information, based on previous surveys, of UXO disposition at a remediation site, one management concern might be to predict the magnitude of storm that could significantly modify the current status. The spatial variation of UXO response across the site would be of particular interest (see Section 4).

1.2 Bayesian Network Concept

The hypothesis underlying the expert system is that underwater UXO behavior can be predicted from observable attributes of the marine environment at the site, along with knowledge of the munitions. For example, based on analysis of coastal research, a proposed relationship between UXO burial and the local environmental conditions, combined with the UXO characteristics, can be in the form

$$\mathrm{TB}(\vec{x}) = f(\ \mathrm{UXOtype}(\vec{x}),\ \mathrm{Alpha}(\vec{x}),\ \mathrm{IB}(\vec{x}),\ \mathrm{UM}(\vec{x})\ ) \qquad \text{Eq (1.1)}$$

where TB is the percentage Total Burial, Alpha is the wave-UXO angle, IB is the percentage initial burial, and UM is the combined wave-current nearbed flow velocity. The vector $\vec{x}$ represents the set of relevant findings (observations) applied to the input nodes that define that forecast scenario (shown as blue and red in Figure 1.1). The variables within the function f are called parent nodes, while the node that they influence (TB) is called a child node.

When we model this hypothesis using a Bayesian Network, Eq (1.1) is conveyed as a probabilistic relationship for $\boldsymbol{p}$, the estimated probability,

$$\boldsymbol{p}(\text{ TB} \mid \text{ UXOtype, Alpha, IB, UM}) = f[\boldsymbol{p}(\text{ TB, UXOtype, Alpha, IB, UM}] \qquad \text{Eq (1.2)}$$

noting that Total Burial also appears as input, because BNs include prior knowledge of each variable. Bayesian Networks are directed acyclic graphs that apply Bayes' theorem [Jensen and Nielsen, 2007] to relate the probability of some event H given the occurrence of evidence E representing a scenario. Bayesian relationship are written as

$$\boldsymbol{p}(H_i \mid E_j) = \frac{\boldsymbol{p}(E_j \mid H_i)\, \boldsymbol{p}(H_i)}{\boldsymbol{p}(E_j)} \qquad \text{Eq (1.3)}$$

In Eq (1.3), the left side, known as the posterior probability, expresses the conditional probability of a forecast variable H, given the set of input observations E_j. In the numerator, the first term is called the likelihood and is the probability of the observing the evidence E if the event H were known to occur. The likelihood term is a measure of the correlation between the observed scenario and the response. The second term in the numerator, $\boldsymbol{p}(H)$, is the prior probability of event H, estimated before the evidence (findings) is observed. The term in the denominator provides a normalization factor to account for the likelihood of the observations. In the UnMES BN, the posterior probability is interpreted as predicted UXO behavior. Posterior PDTs are updated by imposing constraints corresponding to the input values, indicated by subscript j. For spatial applications, as discussed in Section 4, j can represent observations at different locations; or j can denote an index into the test scenarios used compute validation statistics (Section 5). See Jensen and Nielsen [2007] for further details on BN formulation and implementation.

2.0 Design Considerations for UnMES Bayesian Network

UnMES was designed and initially implemented using the commercial software package 'Netica' [Norsys, 2022], that provids a convenient user interface for graphically designing, building, testing, and modifying Bayesian networks. Further operational manipulation of the UnMES BN then requires use of the Netica Application Programmer Interface (API). The UnMES wrapper code developed for the Netica API, discussed in the *Programmer's Guide* [Rennie, 2022], was initially written in Matlab [Mathworks, 2002], and has been transitioned to Python for improved performance. Future modifications to enhance end-user operability are being undertaken under the ESTCP project MR21-5207 [Penko and Ioup, 2021].

The initial design phase of the UnMES BN focused on modeling UXO in sandy, wave-dominated coastal settings [RBL2019]. Input nodes for environmental observables represent standard wave measurements such as wave height and period. UnMES results are interpreted as UXO action over a single storm, therefore the wave height node is envisioned as representing the peak storm height, rather than the distribution of waves that occurs during the event. The primary variable driving underwater UXO behavior is the near-bed flow velocity, that is represented by the intermediate node `Um` in UnMES. Additional required input is the water depth, *h*, where the UXO is sited, and the local sediment characteristic, i.e., sand grain size. It was assumed that the original UnMES formulation represented near-field processes, and, in order to reduce the size of the BN and speed up Bayesian inference updating, the expert system design specified that a single implementation of UnMES had fixed values of these two variables (shown as purple nodes in Figure 1.1).

In practical applications, wave measurements for a specific remediation site will generally be available at some offshore location. However, the relevant wave height is that at the munition's location. Due to wave breaking dynamics across shoaling topography, the local wave height changes substantially throughout a management area with varying bathymetry. In addition, sediment characteristics can often change over fairly short spatial scales. Rough values for the spatially-varying wave heights can be estimated at each near-field location of interest from simple engineering algorithms. However, in order to more accurately understand the far-field pattern of wave variation over the site, a better tool is a calibrated numerical hydrodynamic model such as Delft3D [Deltares, 2022]. This open source modeling suite computes the combined circulation and wave transformation, and is being used by SERDP Project MR2733 to simulate far-field conditions at several test sites [Palmsten and Penko, 2020]. The connection of `Um` values computed by Delft3D into a spatial application of UnMES is discussed in Section 4.

2.1 Choice of Nodes and Links

A Bayesian Network node with no links coming in is considered a parent node, and will contain an unconditional (prior) probability function. In UnMES, these parent nodes act as input variables, providing the boundary conditions for an environmental scenario (blue nodes in Figure 1.1). Using Netica, probabilistic inference (belief updating) requires that every node must be discretized into a finite number of states. Some nodes, like `UXO_Type` are naturally discrete. Most environmental variables describe physical processes and are continuous in nature, so that the total range of the variable must be broken into a number of states, or bins. Therefore, the parent's unconditional probability function is quantified by a probability distribution table (PDT) over the states of the variable.

Adding or rearranging nodes and links in a BN is easily performed in the Netica Graphical User Interface (GUI), which also invokes pop-up boxes to specify the bin intervals. Netica saves the BN in two output formats: `.neta` is the space-efficient binary form commonly used, while `DNET` (`.dne` or `.dnet`) is a human-readable ascii text version that can be manipulated by any text editor (see Section 3.3). While Netica can read files created by Hugin, another widely-used software package for Bayesian network modeling [HUGINEXPERT, 2022], there is not yet a format to exchange Netica BN with another expert system development technology. For operational use, currently the UnMES BN would be manipulated by the Netica API.

A child node has one or more links pointing into it, and has an associated conditional probability table (CPT). When the BN is fully trained, each entry in the multi-dimensional CPT array contains a probability for a specific state of that child node, conditional on a given configuration of the states of its parents. The number of entries in the CPT will be the number of child states multiplied by the number of states in each parent. Clearly if variables are discretized into numerous states, the size of the CPT will grow rapidly. The choice of discrete intervals by which to represent each node is a balance between the desired resolution, and the sensitivity of the response, with an objective to cover the range of interest in as few states as possible, so as to form a compact CPT [Plant and Holland, 2011]. For example, the discretization intervals initially chosen for Total%Burial reflected the target detection performance sensitivity of available underwater survey sensors. As improvements and new technologies arise in underwater sensors, those bin intervals will need to be reassessed. Generally, most nodes are discretized to between 3 to 6 bins, based on a subjective assessment of data accuracy and sensitivity.

It is important to guard against overcomplexity in an effective predictive Bayesian Network. The design for the structure of UnMES, i.e., which nodes and links to include in the BN, was based on background knowledge of the near-field physical processes, with the desire to focus on first-order dominant forces, given the systemic uncertainty. Early choices in the structure of UnMES were influenced by the design of SERDP field and laboratory tests. For example, the initial Prototype version V1 [RBL2019] included an input node representing the angle of the long UXO axis to the shoreline, UXO Angle. This was inspired by field tests [e.g., Calantoni *et al.,* 2014] where this information was recorded, in concert with laboratory experiments showing that flow-UXO angle of attack is an important factor in onset of UXO motion [Garcia and Landry, 2018], and also affects scour burial [Friedrichs *et al.,* 2016].

However, in the operational world it is not reasonable to expect detailed information regarding the munitions' placement to be available, so that the UXO Angle variable was removed for UnMES V2. However, the node for the angle between the wave direction and the UXO main axis remains (Wave-UXO_angle), based on expert opinion that informative prior distributions can be provided in some situations. For example, at former firing ranges, a reasonable initial distribution of UXO angle might be strongly biased in the direction of firing, likely close to shore normal, and therefore mostly parallel to incoming waves in shallow water. For the general case of discarded munitions on sand where little impact burial is anticipated, observations indicate that a cylinder re-orients quickly to lie perpendicular to the dominant flow [Garcia and Landry, 2018], while tapered cylinders take up an angled position. Based on this knowledge, an appropriate prior distribution can be input for Wave-UXO_angle [RBL2019].

Several recent papers have discussed the pitfalls of excessive complexity in BN design and in bin resolution when applied to networks designed to resolve physical processes. Fienen and Plant

[2015] examine how increased complexity may improve the BN fit to a calibration data set (hindcast error), but lead to decreased predictive skill against future data scenarios. They then created a tool box of performance metrics, `CVNetica`, using the Netica API which relies on k-fold cross validation techniques, where different groups of the observations are repeatedly held back from training, and used only for skill evaluation. As of now, no underwater UXO data set has had a sufficient number of entries to allow use of this evaluation technique with the UnMES BN. Future development of the underwater munitions expert system should consider utilization of this analysis methodology when data sample numbers increase.

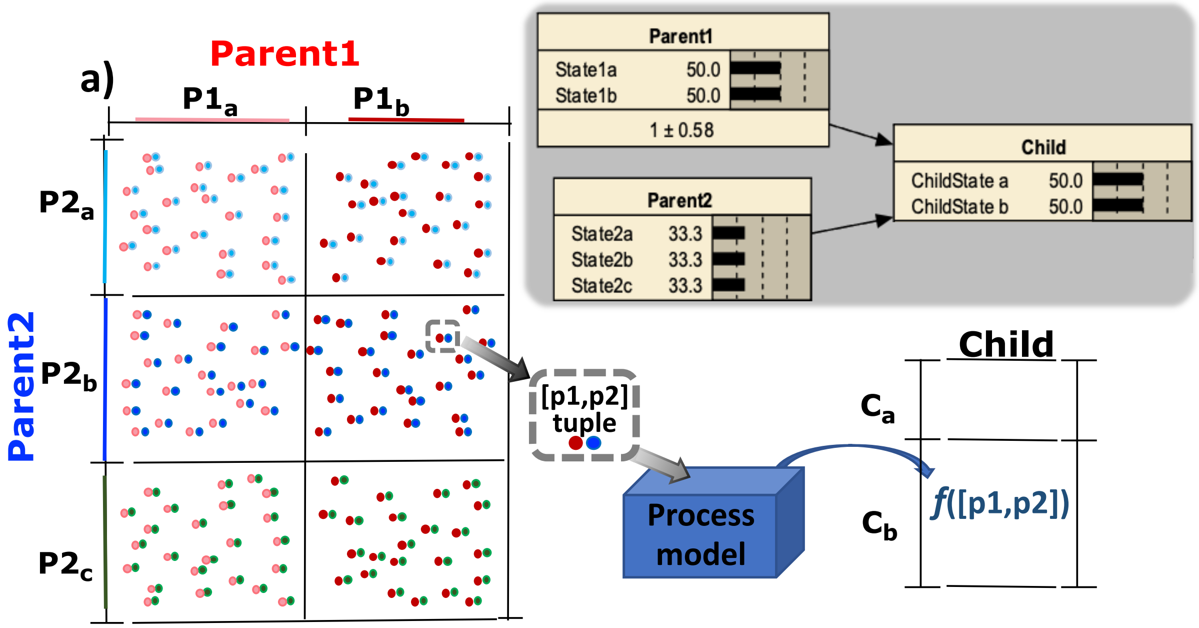
A relevant example showing successful application of `CVNetica` is found in Gutierrez *et al.,* [2015] for a Bayesian Network to model geomorphic storm response of barrier islands. A set of BNs with varying complexity were evaluated for hindcast and validation skill using simple intuitive metrics, which required combined data sets with O(1000) observations. UnMES lacks an adequate sample size for this approach, however, guided by an awareness of the issues of network complexity versus robustness, a re-examination of UnMES input nodes can be made. In particular, the node `Rate_Wave_Growth` had been added to improve hindcast skill for the 2014 surf zone data set from Martha's Vineyard [Traykovski and Austin, 2017]. The initiation of motion recorded for the UXO during the 2014 field test exhibited a range of behavior that, it was suggested, could be explained by variations in the onset speed of storms; information captured in `Rate_Wave_Growth`. However, no further data sets have shown this node to provide any useful predictive value, so that the expansion of the BN has not yet been justified. In fact, a 2018 surf zone study at the same Martha's Vineyard location showed UXO migration may be dominated by varying locations of high order moments of the shoaling wave transformation [Traykovski and Jaffre, 2020]. Recalling that the underwater munitions expert system concept strives to focus on the most straightforward, first order physics driving the observed behavior, the use of additional nodes representing proposed forcings is not warranted until future data sets are available to confirm the necessity for increased complexity.

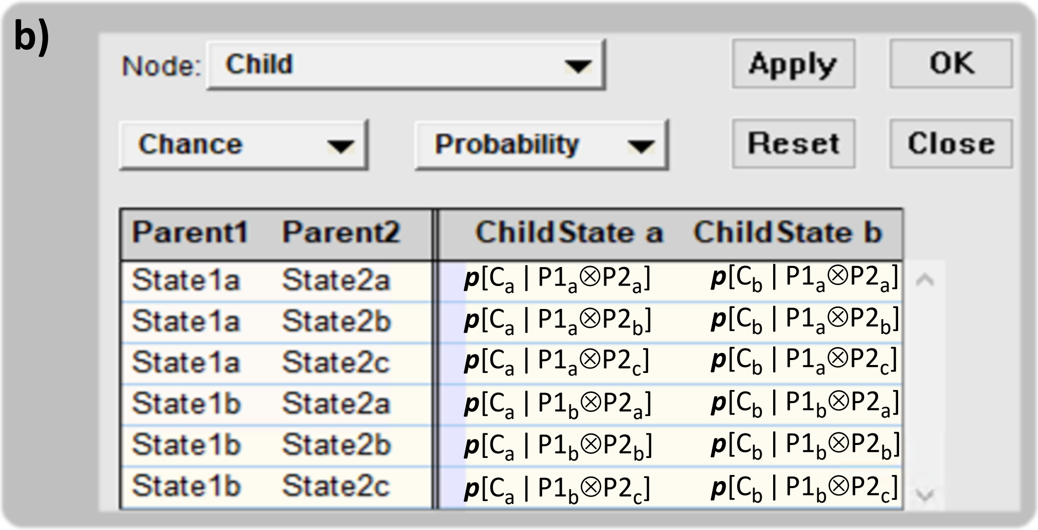
3.0 Training the Conditional Probability Tables

Wrapper code written in Matlab implements the Monte Carlo (MC) generation of training cases. Each case is output as a line in an ascii text file in the columnar format required by Netica, where the header line supplies the node name for each column. Further details are found in Section 2 of the *Programmer's Guide to UnMES Construction* [Rennie, 2022].

3.1 Learning from Cases

In constructing the network for UnMES, deterministic models compute example cases that are used to estimate the conditional probabilities in the tables of the non-input (child) variables of the Bayesian network. The MC procedure generates random, uniformly distributed samples over the joint domains of each child's parent variables, and calculates, for each sampled tuple of parent values, the corresponding value of the child variable, as predicted by the deterministic process model. An example is detailed in Figure 3.1 where there are two parent nodes: P1 discretized into two state partitions (bins), $P1_a$ and $P1_b$; and P2 with three states, $P2_a$, $P2_b$ and $P2_c$. A $[p1_i, p2_i]$ pair (2-tuple) is drawn from the joint parents' domains for case i.





Parent1	Parent2	ChildState a	ChildState b
State1a	State2a	$p[C_a \mid P1_a \otimes P2_a]$	$p[C_b \mid P1_a \otimes P2_a]$
State1a	State2b	$p[C_a \mid P1_a \otimes P2_b]$	$p[C_b \mid P1_a \otimes P2_b]$
State1a	State2c	$p[C_a \mid P1_a \otimes P2_c]$	$p[C_b \mid P1_a \otimes P2_c]$
State1b	State2a	$p[C_a \mid P1_b \otimes P2_a]$	$p[C_b \mid P1_b \otimes P2_a]$
State1b	State2b	$p[C_a \mid P1_b \otimes P2_b]$	$p[C_b \mid P1_b \otimes P2_b]$
State1b	State2c	$p[C_a \mid P1_b \otimes P2_c]$	$p[C_b \mid P1_b \otimes P2_c]$

Figure 3.1 a) schematic of Monte Carlo sampling from joint domains of two parent nodes, with assignment to states within the child node depending on the computed model value. Inset shows example Netica BN. b) Conditional Probability Table displayed by Netica GUI for the example BN.

The example child node value computed by the model is categorized into two states, C_a and C_b. Each sampled set of parent values along with a resulting value for the child node, forms a case. The MC output essentially produces a synthetic database for use in training the BN.

When Netica trains the nodes using its "Learning from Cases" function, the fastest and simplest Counting-Learning algorithm is used to compute the conditional probability table. This is possible because the model output cases do not have missing data, or uncertain findings, and there are no latent variables. These cases are then categorized and counted according to the states defined for the applicable nodes, and the conditional probabilities of the states defined for a child variable (with respect to values in its parent variables) are estimated by calculating the corresponding proportions among these counts. For the example shown in Figure 3.1a, the CPT for child node C requires 12 conditional probability entries, listed in Figure 3.1b. For example, $\boldsymbol{p}[C_a \mid P1_a \otimes P2_a]$ is calculated from relative frequencies of samples in the relevant event categories of P1, P2, and C by:

$$\#(\{ f([p1,p2]) \in C_a \mid [p1,p2] \in [P1_a \otimes P2_a])\}) / \#(\{[p1,p2] \in [P1_a \otimes P2_a]\})$$

where:

f is the deterministic function implemented by the process model
#(**s**) is the number of elements in set **s**
and ⊗ stands for set cross product.

In other words, $\boldsymbol{p}[C_a \mid P1a \otimes P2a]$ is estimated by the proportion of the f(p1,p2) that fall in the event category C_a for all [p1,p2] such that p1 falls in $P1_a$ and p2 falls in $P2_a$. It is of interest to evaluate how good this estimate is, as discussed in the following section.

3.2 Evaluation of Confidence in Conditional Probability Table

Based on available training, the BN supplies a probability with which probabilistic inference, (belief updating) can be used to calculate new beliefs for a set of variables (output predictions) given some findings (input values). But a probability by itself does not represent the full confidence, or lack thereof, one can have in these beliefs. Netica uses the concept of "experience" as a measure of the confidence in its computed probabilities. As the node is trained, an experience table is maintained with an experience number for each row of the CPT. For UnMES, the experience number is essentially the number of training cases that have been seen, because all cases are treated as having a weight, or multiplicity, of 1.

Before learning begins, the net starts off in a state of ignorance, enforced by removing the node's existing CPT and experience tables prior to training (see Section 2 of the *Programmer's Guide*). At each node, all CPT probabilities begin as uniform, and each experience starts at its lowest value, representing ignorance. Each case increments the experience by one. The experience tables are not readily visible through the Netica GUI, but can be accessed in the text version of the BN (`.dne` file), under the node's `numcases` section.

Following Davenport [1970] for insight into the number of cases required for adequate confidence, IC_a is defined to be the indicator function for the event C_a, a binomial distribution where:

$IC_a(x) = 1$ when $x \in C_a$
$IC_a (x) = 0$ when $x \notin C_a$

If $\boldsymbol{N}$ cases (samples) are taken, $\{(p1_i, p2_i) \mid 1 \le i \le \boldsymbol{N}\}$ from $P1_a \otimes P2_a$, each sampled uniformly and identically, then the set of random variables $\{I_i = IC_a(f(p1_i, p2_i)) \mid 1 \le i \le \boldsymbol{N}\}$is a set of identically distributed Bernoulli random variables, with:

$$\boldsymbol{p}[I_i = 1] = \boldsymbol{p}[f([p1_i, p2_i]) \in C_a \mid (p1_i, p2_i) \in P1_a \otimes P2_a] \text{ (for all } 1 \le i \le \boldsymbol{N})$$

By the weak law of large numbers (or Bernoulli's theorem), the sample mean of $\{I_i \mid 1 \le i \le \boldsymbol{N}\}$ approaches the shared actual mean of the I_i arbitrarily closely as $\boldsymbol{N} \to \infty$. The sample mean is:

$$[\textstyle\sum_{i=1\ldots N} I_i] / N$$

and is just the proportion of the $f([p1_i, p2_i])$ that fall in C_a, while the actual mean is:

$$\begin{aligned}(1 * \boldsymbol{p}[I_i = 1]) &+ (0 * \boldsymbol{p}[I_i = 0]) \\ &= \boldsymbol{p}[I_i = 1] \\ &= \boldsymbol{p}[f([p1_i, p2_i]) \in C_a \mid [p1_i, p2_i]) \in P1_a \otimes P2_a]\end{aligned}$$

from above. Thus, as the number of samples [p1, p2] drawn from $P1_a \otimes P2_a$ goes to infinity, the proportion of those samples for which $f([p1, p2]) \in C_a$ approaches the conditional probability $\mathbf{p}[C_a \mid P1_a \otimes P2_a]$ arbitrarily closely. To judge how many samples from each joint event over the parent variables (e.g,. $P1_a \otimes P2_a$) need to be drawn in order to get good estimates of the conditional probabilities for the child events, the rate of convergence is needed. This is also given by Bernoulli's theorem. For the example above:

$$\text{Prob}[\,|(\boldsymbol{N_{Ca}}/\boldsymbol{N}) - \boldsymbol{p}| < \boldsymbol{\varepsilon}] \ge (1 - 1/(4\boldsymbol{N}\boldsymbol{\varepsilon}^2)) \qquad \text{Eq.(3.1)}$$

where:

$\boldsymbol{\varepsilon}$ = desired tolerance for the probability estimates
$\boldsymbol{p} = \boldsymbol{p}[C_a \mid P1_a \otimes P2_a]$ (actual probability of interest)
$\boldsymbol{N} = \#(\{[p1,p2] \in [P1_a \otimes P2_a]\})$ (number of (p1,p2) in $[P_1a \otimes P_2a]$)
$\boldsymbol{N_{Ca}} = \#(\{f([p1_i, p2_i]) \in C_a\})$ (number of $f([p1, p2])$ in C_a)

with the estimated, or "learned", probability $\hat{\boldsymbol{p}} = \boldsymbol{N_{Ca}}/\boldsymbol{N}$. Note that $\boldsymbol{\varepsilon} \le 1.0$, and that we must have $\boldsymbol{N} > 1/(4\boldsymbol{\varepsilon}^2)$ in order to get a useful limit on the probability. Figure 3.2 shows the resulting probability that the estimated conditional probability will be within the tolerance $\boldsymbol{\varepsilon}$ of the true probability over a range of sample sizes, for several values of **ε**. For **ε** = 0.10, a sample size of ***N*** ≥ 500 will result in confidence over 0.95 ($\boldsymbol{\alpha}$ = 0.05), but for smaller **ε** = 0.05, it takes ***N*** ≥ 2000 for the same confidence. Smaller tolerances, e.g., **ε** = 0.01, require an order of magnitude more samples (see inset).

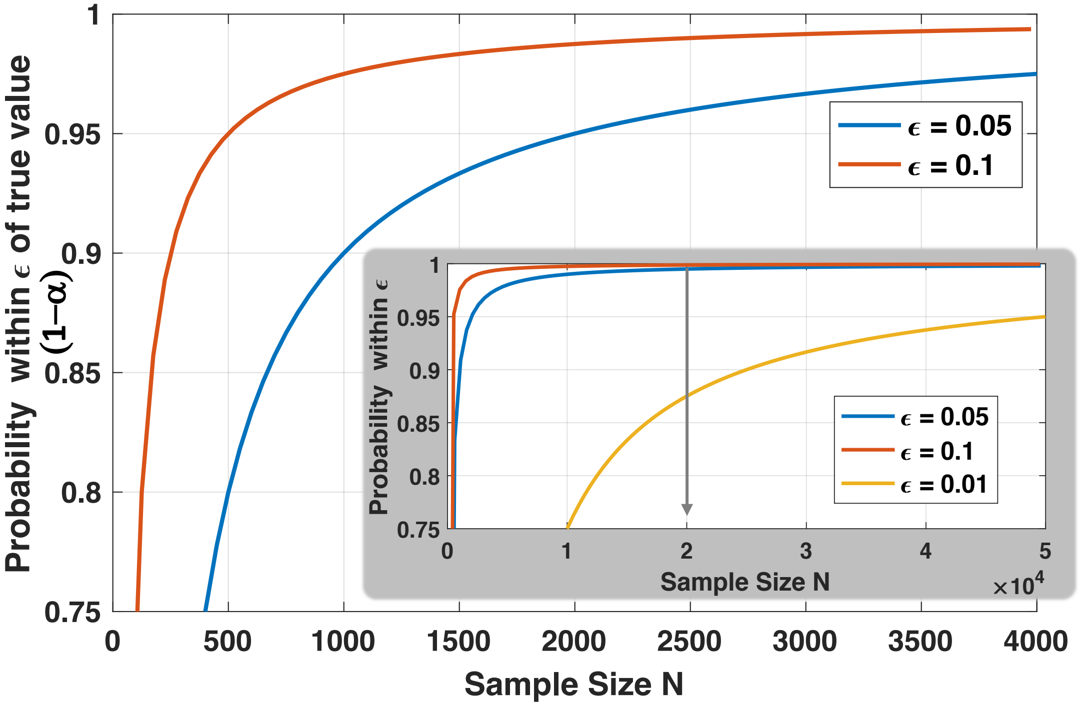


Figure 3.2 Probability that estimated conditional probability will be within ε of the actual probability plotted versus sample size. The y-axis plots the confidence level as (1 - α).

This sample size ***N*** ≥ 2000 is required for each entry in the child CPT, so that a full MC exploration of the example BN in Figure 3.1 would necessitate ***N*** ≥ 24000 at **ε** = 0.05 for all 12 entries. For the Prototype UnMES, the largest CPT is built for the `Total%Burial` node, which has 5 parents, all with multiple states, for a total of 2250 CPT entries. Note that Eq. (3.1) is an upper bound on the probability of error, derived under the assumption that the true probability, ***p***, is equal to 0.5, the worst case for which the probability of error is maximal for any given ***N***. If ***p*** differs significantly from 0.5, then fewer samples are sufficient to achieve a specified degree of precision. In other words, near-uniform probability distributions are the most difficult to estimate accurately, while $\hat{\boldsymbol{p}}$ for strongly modal distributions converge more rapidly towards ***p***. Since generally ***p*** is unknown, the more stringent requirement is used.

An extensive MC exploration of the full input domain was run for a total of several million cases when building UnMES, with the mean ***N*** per CPT entry of over 20,000 (marked on Figure 3.2 inset with grey arrow). Even these large samples numbers do not provide a 90% confidence ($\boldsymbol{\alpha}$ = 0.1) at a very strict tolerance (**ε** = 0.01). Given the substantial uncertainties still acknowledged in the deterministic process models (see *Programmer's Guide*), an acceptance of larger **ε** is appropriate for the expert system, and the larger values shown for **ε** will generally be applied. Display of the relevant tolerance and confidence limits is challenging, however it is important that UnMES users be educated as to how to interpret the wide tolerance and moderate confidence resulting from training sampling limitations so to avoid overemphasis of insignificant details in predicted probability distributions.

3.3 Direct Specification of a Conditional Probability Table in Netica

In certain situations, it can be preferable to specify the entries in the CPT directly, rather than computing the values from a training data set. One example of this occurred when building the Bedform%Burial node where a geometrical argument based on ratios between the UXO diameter and the dune height and length resulted in an algorithm directly defining the CPT values [RBL2019]. For this node, the Matlab wrapper code does not generate training cases, but computes the probability values and outputs them in a text file formatted to be compatible with the Netica's .dnet file layout. This text is inserted directly into the Netica definition file using a standard text editor as described in Appendix D of the see *Programmer's Guide*.

Another example of the direct specification of CPT values is seen in the present formulation of the Migration_Direction node. There has been minimal evidence about the UXO direction of migration, although a few observations indicate that the direction of the current (node cDir) plays a dominant role. For the coastal UnMES V2, Migration_Direction is coarsely binned into 4 states: Alongshore Left, Onshore, Alongshore Right and Offshore. Probabilities conditional on wave and current directions have been proposed based on a rule of thumb informed by the available observations. These relationships act as a placeholder until further research clarifies UXO migration behavior. In this case, the small numbers of CPT values are easily entered using the Netica GUI.

4. Spatial Application and Visualization

The UnMES V2 BN as shown in Figure 1.1 expresses the present understanding of underwater UXO burial and mobility behavior at a single location, where the water depth and sediment type are represented by fixed values. Most munitions response sites cover conditions that vary spatially across the area of interest. Often a Geographic Information System (GIS) to organize map-based information within the Conceptual Site Model (CSM) is created early in the operational remediation process [Rennie and Brandt, 2020]. The project manager needs to distinguish between different regions of the site, and prioritize remedial actions among those regions. UnMES is envisioned as one tool within a larger decision support system that, given the inherent spatial structure of the problem, would best be implemented within a GIS framework.

The first steps towards such an approach are demonstrated by the connection of the UnMES BN with gridded model output for waves and currents computed by the regional hydrodynamic model Delft3D by SERDP Project MR2733 [Palmsten and Penko, 2020]. The bathymetric data from the remediation CSM is the first important input required to initialize the Delft3D process. Delft3D has been implemented and calibrated at several relevant SERDP field locations [Palmsten and Penko, 2020], including the US Army Corps of Engineers (USACE) Field Research Facility (FRF) at Duck, North Carolina, where the DUCK15 field experiment made observations of UXO burial during winter storm conditions [Calantoni, 2016].

4.1 Delft3D Connection to UnMES

The transfer of spatially gridded results from Delft3D numerical simulations to the UnMES BN is currently performed by a combination of Matlab and Python software working with the Netica API. Future revisions and upgrades will transition fully into Python for a more uniform and stable environment. The Matlab script `make_network_input_D3D.m` reformats Delft3D output into a columnar text file with one line for each grid cell and a column for each input node, with a format matching the Netica case file requirements. Different components of Delft3D are computed on different grids, waves being on a lower resolution latitude-longitude grid than velocity, and the bathymetry-morphology grids slightly offset. Particular care must be taken to align and resolve all the input variables to a common grid, and to transform wave and current directions to the shore normal coordinate orientation employed by UnMES. For the simulation of DUCK15 conditions, the FRF Northing, Easting coordinate system was chosen as the standard grid.

A Delft3D simulation is run over an extended time period to ensure accurate calibration, whereas the present version of UnMES is designed to represent UXO response over the duration of one storm. To connect to the BN, the Delft3D time step centered on the peak of the largest storm waves in the model run is usually chosen, designating the most energetic conditions to which the UXO were subjected. Input for UnMES V2 describing the wave conditions consists of three variables: Peak_Hsig, Period, and Wave_Direction, (Figure 1.1). Values from the PDTs of these nodes are used to calculate the near-bed wave orbital velocity given the fixed value Water_Depth. This is combined with values from the Current and Current_Direction nodes to compute the combined near-bed flow velocity (the intermediate variable Um). When using Delft3D input, the numerical simulation provides the bottom orbital velocity components directly, so that the wave component input nodes are not needed.

A version of the expert system called UnMES-D3D was built to facilitate interconnection with Delft3D, in which Um is now treated as an input node. As presently implemented for SERPD locations, Delft3D models a depth-averaged current, therefore the `make_network_input_D3D` code estimates the near-bed current components assuming a logarithmic velocity profile (law of the wall) with a bed roughness length appropriate to hydrodynamically rough flow [Soulsby, 1997]. A new input node for `Um_Direction` is introduced in UnMES-D3D, replacing the `Wave_Direction` and `Current_Direction` nodes. For input to UnMES-D3D, `make_network_input_D3D` computes `Um_Direction` combining the near-bed wave and current components, rotated into UnMES shoreline oriented coordinate system. Use of UnMES-D3D is being investigated by USGS and NRL-SSC [Palmsten *et al.*, 2021], with further details given in the *Programmer's Guide*.

The next step uses the Python wrapper, `UnMES_D3Dconnect.py`, that inputs this text case file from the Delft3D results and loops through the grid locations. At each grid cell, the local Delft3D values are entered as findings into their corresponding UnMES-D3D nodes, and the designated output nodes are queried for the posterior probabilities. The Python interconnections to the UnMES BN are based on the `pynetica` package developed by Sam Bateman at NRL-SSC that uses Netica API functions. The libraries and modules required to form the necessary Python environment are described in Section 3 of the *Programmer's Guide*. Custom versions of `UnMES_D3Dconnect.py`, are easily created where the specification of values for nodes such as `UXO_Type` or `Initial%Burial` can be manipulated for exploration of different remedial scenarios.

4.2 Visualization of UnMES-D3D Results

Code for plotting spatially-varying results from UnMES have been included in the `UnMES_D3Dconnect.py` wrapper with output in both a simple bitmap format (e.g., PNG), as well as the Keyhole Markup Language (KML) file format used to display geographic data in a browser such as Google Earth. Visualization of geospatial information in a Google Earth overlay, or as a Google Hybrid map, is a common standard, with the Google Earth web browser providing a familiar interface, and increasing interconnectivity with web-GIS that is readily available.

In order to visualize the scenario, it is can be helpful to display the dominant driving forces such as the `Um` vector field, which is plotted in Figure 4.1a as directional arrows scaled by magnitude, with the magnitude is repeated by the intensity of the color overlay. The time step selected from the Delft3D time series represents the peak waves of the major storm on February 11, 2015 during the DUCK15 experiment. It is clear that at this time the wave orbital velocities (nearly shore normal) dominate over the alongshore current contributions to the `Um` bottom flow.

There are multiple choices for display of predicted probabilities. Optimal methods for visually representing and interacting with probabilistic results are an area of active research [Rennie and Brandt, 2020]. Uncertainty can be difficult to successfully communicate to the operational community, and spatially-varying displays present an increased challenge. UnMES predictions result in a multi-state PDT, while color scaled presentations in map format can only summarize the PDT into one scalar value. The simplest choice is to display the probabilities from one state of particular interest. In Figure 4.1b the probabilities from the state of most concern, that of complete burial, are plotted.

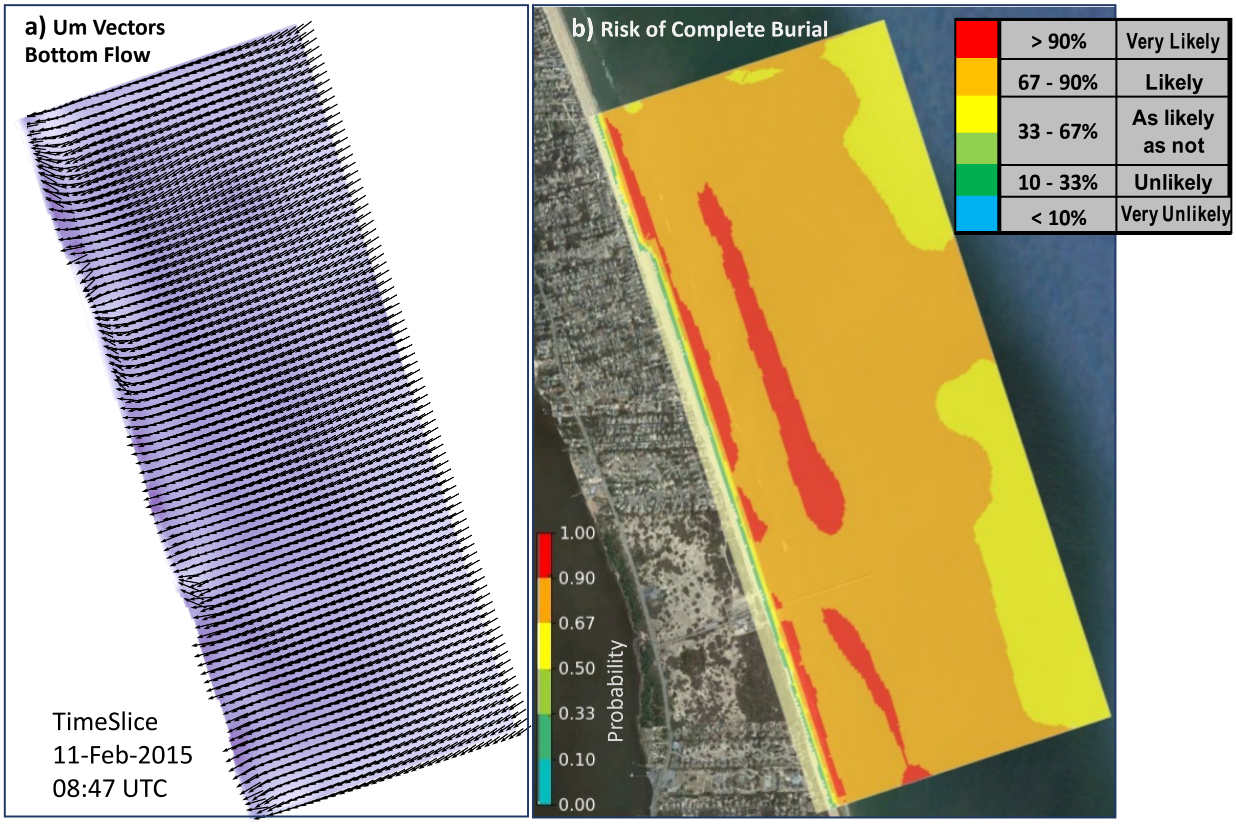
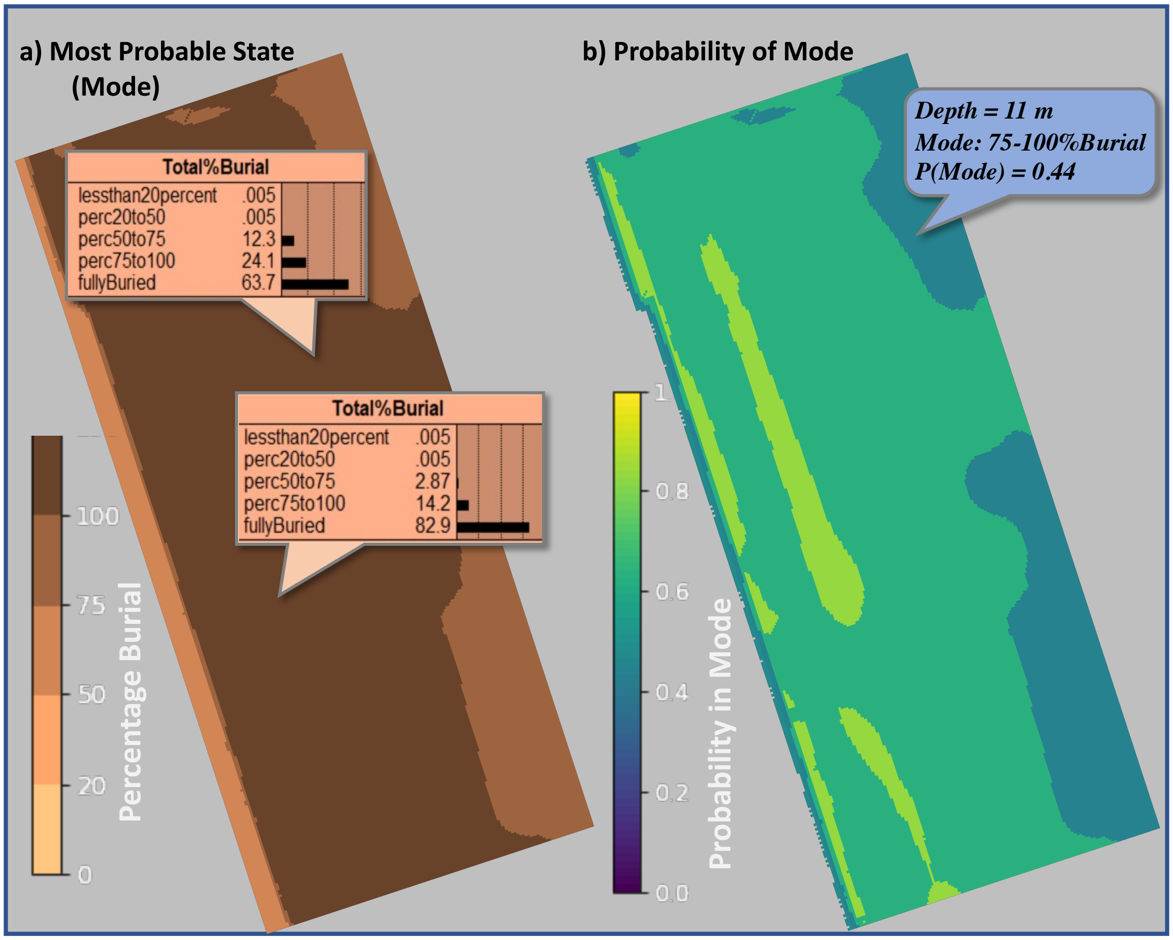


Figure 4.1 a) display of the combined wave and current near-bed flow velocities computed by `make_network_input_D3D.m` from Delft3D model output of DUCK15 storm conditions. b) Probability of complete burial from UnMES-D3D color scaled into IPCC risk categories.

From the map in Figure 4.1b a manager can discern what is the risk level of a UXO being completely undetectable under the sediment surface; an alternate interpretation would be that the map represents the percentage of UXO expected to be totally buried. However, no information is communicated about the spread, or uncertainty, of the burial PDT. In addition, no knowledge about the possibility of partial burial, or probabilities in the lower burial bins, is imparted.

The burial probabilities are displayed using a color scale that is mapped to a logarithmic division of risk levels for decision making proposed by the Intergovernmental Panel on Climate Change (IPCC). The accompanying natural language terminology (inset in Figure 4.1b) expresses perceptive acknowledgement of the uncertainty, while each level is specifically related to a numeric probability level, rather than being merely a relative level. This set builds on the familiar stoplight approach with High (red), Medium (yellow) and Low (green) risk levels, but with 5 grades providing increased discrimination among scenarios. Future implementation of UnMES-D3D with sufficient validation to support further discretization of the probability ranges would allow risk terminology to be extended to include two end categories useful for management concerns: "Exceptionally Unlikely" for probability < 1% and "Virtually Certain" for probability > 99%, as presented in Gutierrez *et al*., [2015].

Two maps are needed to more fully represent probability: the first displays a univariate quantity of interest summarized from the probability distribution, while the second map represents the uncertainty associated with that quantity. Usual metrics are Expected Value (mean of the PDT) for the quantity, and Standard Deviation for the uncertainty. As can be seen in the belief bars displayed in Figure 1.1, Burial and Migration PDTs often have strongly skewed, rather than normal, distributions, confounding the usual intuitive interpretation of mean as center, and standard deviation as width, of a distribution. Instead, a map of the Most Probable State (the mode of the PDT), along with the probability of that the mode, as shown in Figure 4.2, may be a more informative presentation for management purposes.



a) Most Probable State (Mode)

b) Probability of Mode

Total%Burial	
lessthan20percent	.005
perc20to50	.005
perc50to75	12.3
perc75to100	24.1
fullyBuried	63.7

Total%Burial	
lessthan20percent	.005
perc20to50	.005
perc50to75	2.87
perc75to100	14.2
fullyBuried	82.9

Depth = 11 m
Mode: 75-100%Burial
P(Mode) = 0.44

Figure 4.2. Prototype example of geospatial presentation options for UnMES-Delft3D implementation displaying Total%Burial predictions for UXO Type of Mortar: a) %Burial quantity expressed as mode, b) %Burial uncertainty expressed as probability of modal state. Example pop-ups illustrate possible user interaction methods.

Discretization intervals for many nodes in UnMES are irregular, with the final state encompassing a much larger range than the other states, which distorts the calculation of standard deviation. For a BN with discretized nodes, rather than using a continuous color mapping, it is helpful to divide the color map into color bands that align with the discrete states as in Figure 4.2.

Techniques for visualization of uncertainty sometimes use additional visual cues including texture, scaling, transparency, and glyphs to convey multidimensional information on a map. Depending on the target audience, too much information in a crowded display can be difficult to assimilate. An possible effective presentation allows the user to interact with alternate images through a self-driven investigation, such as the pop-up PDT displays illustrated in Figure 4.2. Future visualization software design for the expert system needs to support exploration in a manner that helps the user make sense of the probabilistic nature of UnMES results in order to answer management questions.

5. Statistical Validation of UnMES

The goal for validation of the Underwater Munitions Expert System is to document the utility of UnMES predictions for use by the remediation community. Satisfactory validation with relevant field data is required to establish model credibility with end-users (site managers) and promote confidence in the operational value of the expert system results. As there does not exist a strict requirement for the exact degree of statistical confidence that a decision tool must achieve in MR management, the working principle is to validate the model "to the regulator's satisfaction". In general, the operational viability for complex expert systems has often been assessed simply using a visual comparison showing similarity between prediction and observations.

With probabilistic predictions, performance must be determined by more than one metric. For example, statistical bias, or measure of the central tendency in the predictions compared to that of the observations is a useful statistic; along with accuracy, a measure of the average difference between the predicted PDT and the PDT formed from the field data, that depends on the variance or shape of the distribution. Another helpful metric is the skill, a non-dimensional measure of performance relative to some baseline prediction [Sutherland *et al.,* 2004].

In general, the Remedial Project Manager (RPM) recognizes predictive usefulness in UnMES results if they differ sufficiently from the current background state of knowledge at a particular site, and if they can be shown to be accurate enough so that these differences are significant. Often, the current state of RPM knowledge for a given scenario is close to complete uncertainty, i.e., the prior PDT is a uniform distribution [Rennie and Brandt, 2020]. If an UnMES prediction for a specified scenario is highly skewed or strongly modal (majority of prediction in a single bin), this can provide actionable information for the site manager, if it has been shown that UnMES predictions lie within some acceptable tolerance of the true (but unknown) underlying distribution.

It is difficult to establish absolute accuracy for probabilistic models, so a common approach is to compare the predictions relative to a baseline or reference distribution which represents a "No Skill" model. In meteorological forecasting, the established regional climatology usually provides a meaningful reference. Figure 5.1 illustrates two possible reference PDTs that might be meaningful to site managers in assessing UnMES burial predictions: the first is a uniform PDT representing a state of no knowledge. The second refence PDT has probability = 1 in the >100% burial bin, or an assumption of complete burial of all UXO. Several validation metrics are discussed in in Section 5.2 using this relative skill approach.

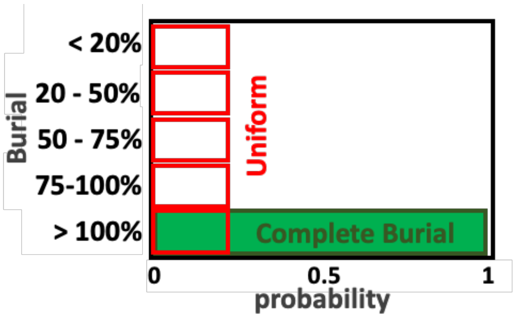


Figure 5.1 Example reference (baseline) distributions for evaluating relative skill scores.

To establish user acceptance in the expert system, evaluation at high confidence levels, such as commonly cited for unidimensional statistics, e.g., 95% Confidence Interval (C.I.), would be desirable. However, the sampling demands for statistics of multidimensional probabilities make those confidence levels difficult to attain. Also, considering the remaining acknowledged uncertainties in the underlying physical models, selection of more modest confidence criteria, such as 90% or even 80% C.I., are suitable for UnMES assessment. Particular effort may be needed to build the user community comfort with the operational value furnished by these more modest validation scores and the use of relative skill metrics.

5.1 Considerations of Sample Size for Validation

Design of field tests to provide data for validation of UnMES need to take into consideration the sampling demands for accurate comparison with multidimensional probabilistic predictions. To obtain independent and identically distributed observations, the measured UXO must be from the same background distribution, i.e., with the same munitions characteristics. Observations obtained for UXO burial can be formed into a histogram with the same state (bin) divisions as used in the UnMES node Total%Burial, and the differences between the discreet probability distributions (PDTs) evaluated to determine the skill of the prediction. Assuming that the PDTs have $\boldsymbol{k}$ bins, the question arises: how well does the PDT of the observations (the estimated probability $\widehat{\boldsymbol{p}}$ **,** now a vector of length $\boldsymbol{k}$) represent the true distribution $\vec{\boldsymbol{p}}$ of the population of UXO burial at this site?

The analysis for sample size presented in Section 3.2 was extended to multinomial distributions by Thompson [1987] that formed the basis for recommendations of validation test sampling requirements in RBL2019, and is further considered here. Thompson [1987] presented a method to establish the smallest number of observations needed achieve a desired statistical significance (specified $\boldsymbol{\alpha}$) for simultaneous estimation of multinomial population. If there is no prior knowledge of $\vec{\boldsymbol{p}}$, the analysis is based on the "worst case" distribution, where several bins have the same value $\boldsymbol{p_i} = \mathbf{1/}\boldsymbol{m}$ for $\boldsymbol{m}$ bins, and the remaining $\boldsymbol{k\text{-}m}$ bins have value $\boldsymbol{p_i} = \mathbf{0}$. In addition, it is assumed that the proposed precision, or tolerance **ε** (the half-width of the uncertainty interval) is the same for all bins.

In the procedure for validation of the expert system, the UnMES PDT can be considered to represent $\vec{\boldsymbol{p}}$. With this prior knowledge of the distribution, Thompson describes an iterative algorithm from Tortora [1978] for determining the minimum sample size $\boldsymbol{N}$ required to satisfy a desired confidence level $\boldsymbol{\alpha}$, given an acceptable tolerance **ε**, by computing the components $\boldsymbol{\alpha_i}$ for each bin $\boldsymbol{i}$ as

$$\alpha_i \ = 2(1 - \Phi(\varepsilon \frac{\sqrt{N}}{\sqrt{\vec{p}_i(1-\vec{p}_i)}})) \qquad \text{Eq. (5.1)}$$

for trial values of $\boldsymbol{N}$, and select the minimum sample size for which $\sum \alpha_i \ \leq \alpha$. The symbol Φ represents the cumulative standard normal distribution. Again, the most stressing situation is when non-zero PDT values are the same for all bins. Two example PDTs, both with $\boldsymbol{m} = \boldsymbol{k} = 4$, are shown in Figure 5.2a: a uniform distribution (plotted in red) versus a strongly model $\vec{\boldsymbol{p}}$ (blue) which has $\boldsymbol{p_1} = 0.8$ in the first bin and probabilities less than 0.1 in the other bins. A tolerance of **ε** = 0.1 about the value in each bin is illustrated by the dotted lines.

The sample sizes needed to achieve $\boldsymbol{\alpha}$ = 0.1 (90% confidence) are compared in Figure 5.2b. At **ε** = 0.1, required $\boldsymbol{N}$ = 95 for $\vec{\boldsymbol{p}}$ uniform, or $\boldsymbol{N}$ = 50 for the modal PDT. If an 80% confidence level is acceptable ($\boldsymbol{\alpha}$ = 0.2), then the sampling demands are less onerous, with $\boldsymbol{N}$ ~ 35 for the modal PDT. This **ε** = 0.1 tolerance is fairly wide, almost half the range of the uniform bins, and one cannot distinguish between the values in the smaller bins of the modal distribution. However, to reduce the tolerance, e.g., **ε** = 0.05, would require over two hundred observations at 90% C.I. (Figure 5.2b). For validation tests performed with smaller samples size, it will be important to convey the resulting confidence interval widths, to discourage interpretation of small PDT variations as actionable information. It should be emphasized that UnMES results must be interpreted as guidance in a manner appropriate to the probabilistic expert system approach.

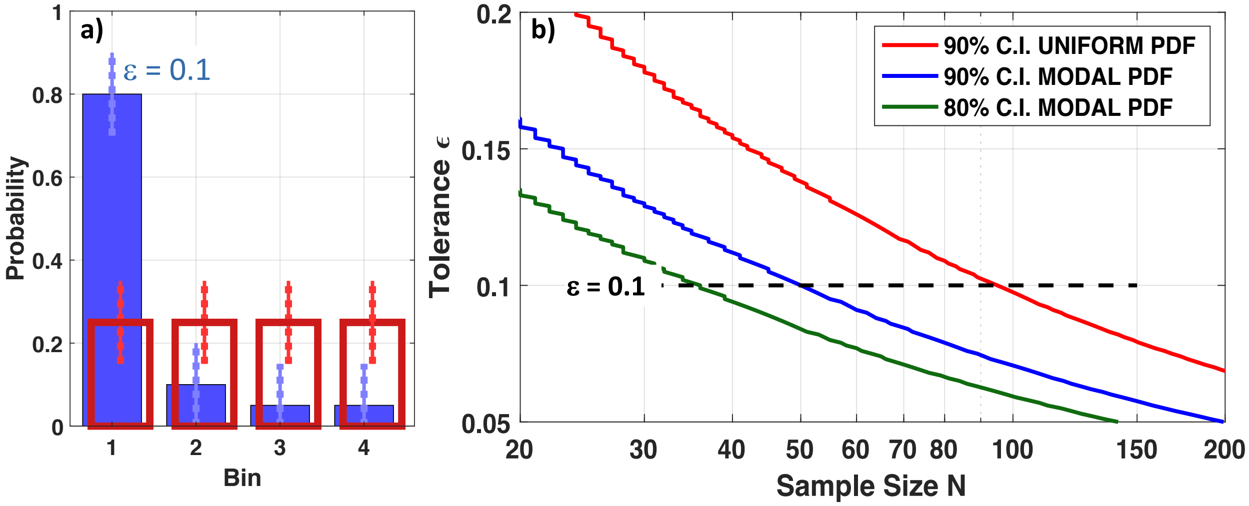


Figure 5.2 a) example 4-bin distributions: uniform (red) and strongly modal (blue). The dashed lines indicate ± ε about the bin probability. b) Required sample size plotted against the resulting tolerance (ε) for 90% C.I. (red and blue), and 80% C.I. (green) for the modal distribution.

When designing a validation field test, a version of UnMES should be built that is populated with the `UXOtype` to be deployed in the proposed experiment. Then, with an estimation of the environmental forcing, the BN can be queried to learn the predicted $\vec{\boldsymbol{p}}$, and Eq (5.1) used to justify the planned sample size, given designated confidence levels. The conclusion that a highly modal distribution requires fewer samples to be verified agrees with our intuitive sense that extreme conditions will produce more conclusive results. However, observations of UXO behavior under environmental end conditions has varying operational value. A field test conducted during very calm conditions will generally result in distributions strongly skewed to the smallest amount of burial or migration, while extreme storm conditions could be expected to produce PDTs with large modes in the deepest (farthest) states. Either situation will likely be sufficient to validate a UnMES prediction. However, the more interesting investigation from a management perspective will target the threshold conditions at which significant burial or mobility occurs, conditions which tend to result in a broad spread in $\vec{\boldsymbol{p}}$.

Even the smallest $\boldsymbol{N}$ considered above that attains a reasonable degree of statistical confidence represents a sample size many times larger than has yet been observed during SERDP field tests exploring UXO burial and migration. The most extensive field data set on surf zone migration [Traykovski, 2017] contains about two dozen samples, but these are divided among several UXO sizes and densities. Extensive observations from swash zone experiments number over 200 samples [Puleo and Cristaudo, 2020], however eight different UXO types were recorded. The ability to obtain accurate munitions behavior data has been increase by many improvements in sensor technology over recent years, e.g., "smart" munitions such as described in Bruder, Cristaudo, and Puleo [2018]. However, obtaining large sample sizes, especially in storm conditions, is still expensive and challenging. To realistically meet the requirement for both sample size and identically distributed observations, validation test design will need to focus on a careful choice using just a few UXO types whose characteristics span the domain of interest in a way that will answer both physical and management questions, as discussed in Calantoni [2019].

5.2 Skill Assessment for Probabilistic Predictions

In the previous section we considered validation from the point of view of accuracy within a single estimate of probability in an individual state. The most familiar approach for statistical assessment considers one model parameter at a time. For probabilistic models, suitable performance is assessed by multiple metrics in concert, so that the ability of the predicted PDT to recreate the observed distribution are evaluated for several aspects of the distribution. Acceptance of UnMES as a applied decision tool depends on buy-in across the range of stakeholders involved in the remediation process, therefore the validation metrics should be simple to understand, robust, and intuitively effective for conveying model success. A number of statistics have been used evaluate probabilistic predictive skill for discretized continuous variables, similar to UnMES. Several of the statistical analyses discussed in this section can be computed using the Netica software, either within the GUI, or when using the API under Netica's `NetTester` tool.

The initial validation metrics computed are based on the first moments of the probability distribution: an estimate of the bias, or difference between the expected value of the prediction, and that of the observations. The second moment, or variance, relates the width or spread of the predicted compared to the observed distribution. The expected value as mean of the PDT is always displayed by the Netica GUI along the bottom of nodes shown in the "Belief Bar" style, as illustrated in Figure 1.1, followed by ± sqrt(variance), or standard deviation, σ. An alternate measure of the expected value is the Mode, or Most Likely bin, i.e., the state with the highest probability.

A simple skill metric that has been used frequently is the Error Rate, which records the percent of observations that do not match the Most Likely predicted bin [e.g. Gutierrez *et al.,* 2015]. The Error Rate is not truly meaningful in an absolute sense, but provides an intuitive comparative measure between probability models, where the modeled distribution with the lower Error Rate clearly provides a better prediction. The Error Rate is a summary of the percentage of off-diagonal values in the corresponding Confusion Matrix. A Confusion, or Error, Matrix is a helpful tool to investigate the performance of an probabilistic system, summarizing which data classifications fall into the Most Likely predicted state [Fielding and Bell, 1997]. Because the Error Rate metric focuses on the mode of the PDT, it is important that the node's chosen bin intervals define states that are meaningful in a remediation management scenario (Section 2.1). Note that while a Confusion Matrix good for providing detailed feedback to the BN developer about where over-and under-predictions occur, it is not generally suitable for conveying validation results to the end-user, being complicated to explain.

UnMES predictions for burial and migration distance represent continous variables, that have been discretized into a fixed number of probability bins. The Error Rate, simply counting the observed hits not in the predicted modal bin (i.e., unsuccessful classification), does not provide information about how far the data lies from the prediction. Another score based on classification success which is sensitive to this distance is the ranked probability score (RPS), introduced by Murphy [1969], and Epstein [1969], and used extensively in the evaluation of weather forecasting. The RPS is a squared quantity that compares the predicted cumulative density function with the cumulative density of the observational distribution, computed as

$$\mathrm{RPS} = \sum\left(\sum_1^k \boldsymbol{p}_i - \sum_1^k \boldsymbol{o}_i\right)^2 \qquad \text{Eq.(5.2)}$$

with $\boldsymbol{o}_i = 1$ when the observation falls in bin $\boldsymbol{i}$, and $\boldsymbol{o}_i = 0$ elsewhere; and where $\boldsymbol{p}_i$ is the posterior probability in the $\boldsymbol{i}$th bin, as in Section 5.1. In this form, a perfect forecast yields RPS = 0; more intuitively, ranked probability score is sometimes reported as 1- Eq.(5.2), so that a good score approaches 1. The RPS is also referred to as the Brier score (BS) [Brier, 1950], although BS is strictly for a binary classification (PDT consisting of 2 bins). The version of BS computed by Netica is called the "quadratic loss" score. RPS takes into account both the shape and the overall tendency of the PDT. This statistic can be recast as a relative ranked probability skill score (RPSS) by taking the ratio of the RPS for the UnMES predictions to the score of a reference distribution, such as shown in Figure 5.1:

$$\mathrm{RPSS} = 1 - \frac{\langle RPS \rangle}{\langle RPS_{ref} \rangle}; \qquad \text{Eq.(5.3)}$$

where $\langle \cdot \rangle$ indicates the mean over set of {prediction, observation} pairs. For this formulation, RPSS approaches 1 when the observations match the predicted probability better than they match the reference. If the UnMES posterior PDT is no better than the reference compared to the observations, then its RPSS will be zero; if worse, the RPSS will take on a negative value. Examples of the use of this skill score for evaluation of complex coastal system models can be found in Gutierrez *et al*., [2015], and Poelhekke *et al*., [2016]. For UnMES assessment, there are very few observations to use for the validation set. The effects of small $\boldsymbol{N}$ on RPSS estimation are considered in Weigel *et al*., [2007], where calculation of the negative bias due to inadequate sample size is proposed, and should be considered when computing scores for UnMES.

To illustrate these metrics, an evaluation of burial prediction has been performed for the coastal version UnMES V2 trained to represent an underwater location with fine sand at depth h = 8 m. Observations of surrogate UXO burial are available from two SERDP field experiments where both had surrogate munitions laid in water depths of about 8 m. The first, called TREX13, was in the Gulf of Mexico (GoM) [Calantoni *et al*., 2014], and the second one was at the FRF site for

DUCK15 [Calantoni, 2016]. One of the GoM TREX13 deployments experienced storm conditions, while there was mild weather during the other deployment. Details of the burial measurements derived from TREX13 field observations are presented in Rennie and Brandt [2017a, 2017b]. For this analysis, only data for larger UXO (diameter > 25 mm) were used, due to the difficulty in estimating percentage burial for the small bullet-size munitions.

Figure 5.3 displays the observed (blue) and predicted (red) PDTs for UXO burial, overlaid with mean and σ for the compared distributions. In calm conditions (Figure 5.3a), the UnMES predictions generally agree with the observations reporting only moderate burial: both PDTs have a strong mode in the bin for 20 to 50% burial. The UnMES results do, however, have more spread, and a larger mean burial predicted. However, due to the very small sample size ($N = 7$), with an 80% C.I. the resulting tolerance (uncertainty) for estimating the observed PDT is ±0.3, or ±30%, as illustrated by the orange line in Figure 5.3a. With the PDT so poorly resolved, even that strong mode is only marginally significant.

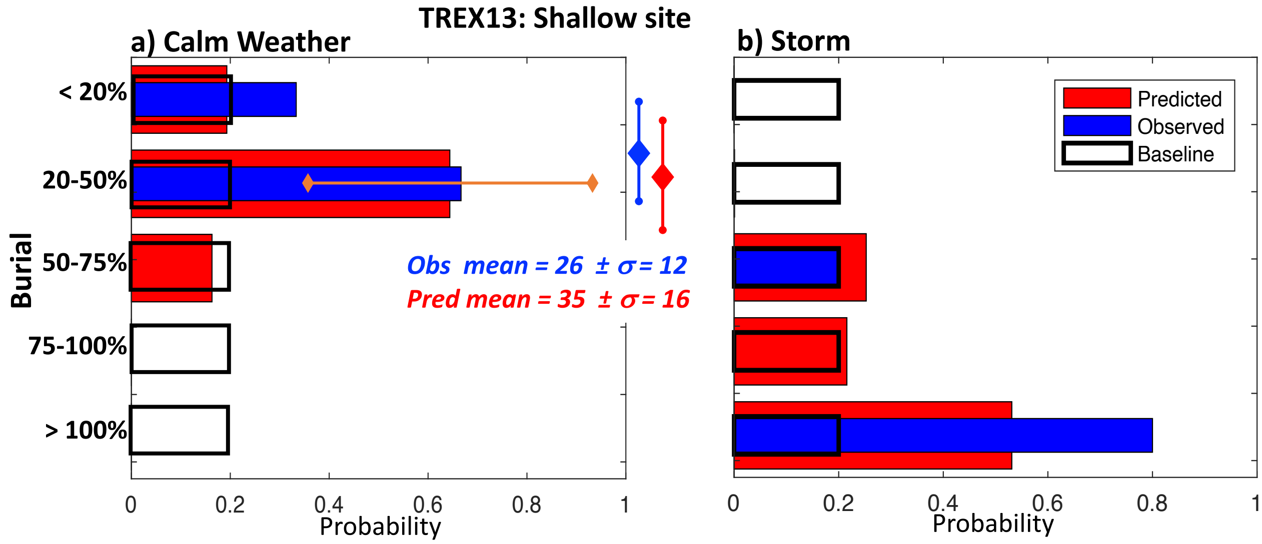


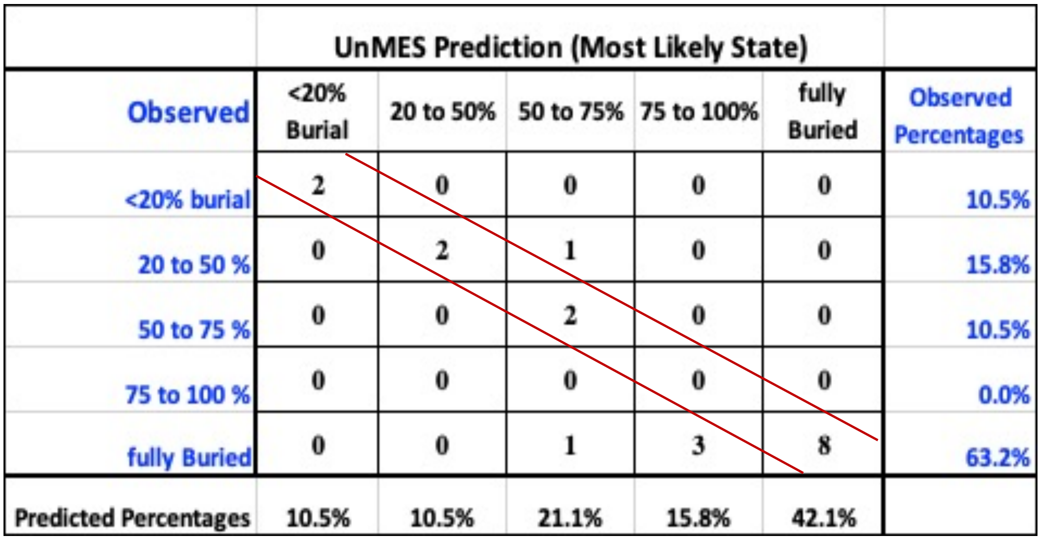
Figure 5.3. Comparison of observed (blue) and UnMES predicted (red) probability distributions for UXO burial. Meteorological forcing and burial measurements from the TREX13 shallow site for two deployments are shown: a) low wind and wave conditions and b) storm conditions. The black boxes indicate a uniform probability distribution. The orange bar shows the large uncertainty in the observed PDT (ε = 0.3 at 80% C.I.) due to the small sample size ($N = 7$).

For storm conditions observed during the other TREX13 deployment, Figure 5.3b compares the observed PDT overlaid on the UnMES predictions. Both modes fall in the deepest burial bin, resulting a low Error Rate (< 20%). A problem arises in how to define the bin describing full burial, which strongly affects metrics that treat the bins as quantitative. For initial considerations of scour burial alone, fully buried was defined at 115%. But the addition of burial processes such as impact penetration, liquefaction, and far-field accretion can result in extreme percentage burial.

In fact, observations from the DUCK15 field experiment showed burial to many times the UXO diameters. In that situation, the deepest burial state is labeled "fully Buried", as in Figure 1.1, and the bins should be treated as categories, rather than numerical intervals.

If the reference distribution is chosen to be uniform (all categories of burial equally likely, shown as black boxes in Figure 5.3), then an RPSS assessment of the TREX13 data reveals that the UnMES predictions show a ~ 90% improvement over a state of "no knowledge". If the "no knowledge" assumption is one of complete burial, i.e., the reference PDT is 100% in the deepest burial state (illustrated by green bar in Figure 5.1), then the RPSS score is very high (RPSS = 0.98) for the calm weather (low burial) prediction, but shows only a moderate improvement (RPSS = 0.18) for the storm-driven burial. It can require multiple examples to compile a sense of how well the expert system is performing using these situational measures, with the choice of relevant reference distributions meeting the end-user needs. Clearly, further field tests for UXO burial are needed with larger sample sizes, and for similar surrogate munitions and environmental conditions, as discussed above.

Table 5.1 Error Matrix for UnMES Burial Predictions



	UnMES Prediction (Most Likely State)					
Observed	<20% Burial	20 to 50%	50 to 75%	75 to 100%	fully Buried	Observed Percentages
<20% burial	2	0	0	0	0	10.5%
20 to 50 %	0	2	1	0	0	15.8%
50 to 75 %	0	0	2	0	0	10.5%
75 to 100 %	0	0	0	0	0	0.0%
fully Buried	0	0	1	3	8	63.2%
Predicted Percentages	10.5%	10.5%	21.1%	15.8%	42.1%	

Adding the DUCK15 burial data to the GoM observations increases the sample size to ***N*** = 19. A combined Confusion Matrix is presented in Table 5.1, revealing only a few non-zero off-diagonal entries. The resulting Error Rate of 26% (5 off-diagonal values out of 19), is an error level sufficiently low that the expert system is considered to demonstrate skillful prediction capability [Gutierrez *et al.,* 2015] for the most likely value of `Total%Burial`. A more thorough statistical presentation of UnMES utility will await future data, especially addressing the predictability of migration distances.

6. Summary and the Way Forward

The development of an expert system (UnMES) for the prediction of underwater munitions burial and mobility has provided a focal point for the collection and synthesis of information obtained by the on-going research projects within the SERDP Munitions Response program. Our efforts at JHU/APL, summarized in Table 6.1, have ranged from laboratory experiments, refinement of physics-based process models, and analysis of multiple field tests in collaboration with colleagues among the SERDP researchers. Extensive background knowledge is clearly essential to building a meaningful Bayesian Network, and a substantial portion of recent SERDP research has been incorporated into the foundational version of UnMES. The BN framework allows easy extension of UnMES as our understanding is augmented by forthcoming results.

Table 6.1 Summary of JHU/APL Major Documentation on UnMES Development

Title	Authors	SERDP Report	JHU/APL Report	Date	Significant Collaboration
Underwater Munitions Expert System to Predict Mobility and Burial: Domain Knowledge Extraction	Rennie and Brandt	Interim Report MR-2227	FPS-T-14-0179	December 2014	
Experimental Determination of Underwater Munitions Mobility and Burial	Rennie and Brandt	Interim Report MR-2227	FPS-T-14-0575	December 2014	
Underwater Munitions Expert System: Preliminary Design Report	Rennie and Brandt	Interim Report MR-2227	FPS-T-15-0333	August 2015	
Self-burial of objects on sandy beds by scour: A synthesis ofobservations	Friedrichs, Rennie and Brandt	*Scour and Erosion* CRC Press, pp. 179–189		2016	Carl Friedrichs, VIMS
Initiation of motion and scour burial of objects underwater	Rennise, Brandt and Friedrichs	*Ocean Engineering* , 131, pp. 282–294		2017	Carl Friedrichs, VIMS
Underwater Munitions Expert System: Demonstration and Evaluation Report	Rennie and Brandt	Final Report MR-2227	FPS-R-17-2321	November 2017	
Status of Underwater Impact Penetration Modeling for use in the Underwater Munitions Expert System	Rennie and Brandt	Interim Report MR-2645	FPS-T-17-0456	November 2017	Teichman and Cazares, IDA
Improved Models for Burial, Exposure, and Migration of Underwater Munitions	Rennie and Brandt	Interim Report MR-2645	FPS-T-19-0332	August 2019	Traykovski, WHOI
Prototype Underwater Munitions Expert System: Demonstration and User's Guide	Rennie, Brandt and Ligo	Final Report MR-2645	FPS-R-19-0695	November 2019	
Community Feedback and Design of Visualization Output for the Underwater Munitions Expert System	Rennie and Brandt	Interim Report, MR19-1126	FPS-R-20-0281	July 2020	Bryan Harre, NAVFAC Andrew Schwartz, USACE
Object Burial by High-Energy Forcing: Liquefaction and Granular Sorting	Rennie and Brandt	Interim Report, MR19-1126	FPS-R-20-0685	January 2021	Carl Friedrichs, VIMS Klammer and Calatoni, NRL-SSC
Extension of the Underwater Munitions Expert System to Beach Face and Estuarine Environments	Rennie and Brandt	Interim Report, MR19-1126	FPS-R-21-0476	August 2021	Jack Puleo, Art Trembanis and Carter Duval, UDEL

Outreach to the user community was undertaken as part of the JHUAPL efforts [Rennie and Brandt, 2020], with the help of SERDP's NAVFAC and USACE liaisons. However, UnMES was not yet at a stage of validation to elicit sustained interest from busy site managers, nor is JHU/APL in the position to provide the continued maintenance that an active operational tool requires.

Therefore, the Underwater Munitions Expert System is being transitioned to a software team at NRL-SSC under the direction J. Calantoni for support and further development. Additional documentation describing the construction and training of UnMES is detailed in the *Programmer's Guide to UnMES Construction* [Rennie, 2022].

In particular, an end-user interface appropriate for practical application by remedial site managers needs to be designed and implemented by the software professionals at NRL-SSC, as proposed by ESTCP project MR21-5207 [Penko and Ioup, 2021]. This ESTCP project, also overseen at NRL-SSC, will, in addition, focus on enhanced validation appropriate for regulatory acceptance as well as on-going support to end-users, with UnMES as one of a set of tools in a Munitions Response Library for site management.

Literature Cited

Brier, GW, 1950. Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78, 1–3.

Bruder, B, D Cristaudo, JA Puleo, 2018. Smart surrogate munitions for nearshore unexploded ordnance mobility/burial studies, *IEEE J. of Oceanic Eng.,* 99, 1-20.

Calantoni, J, T Staples, A Sheremet, 2014. "Long Time Series Measurements of Munitions Mobility in the Wave-Current Boundary Layer," SERDP Project MR-2320, Interim Report.

Calantoni, J, 2016. "Long Time Series Measurements of Munitions Mobility in the Wave-Current Boundary Layer," SERDP Project MR-2320, In-Progress Review.

Calantoni, J, 2019. "Towards Developing Demonstrations for Munitions Mobility and Burial in the Underwater Environment," SERDP Project MR19-1317 Proposal.

Deltares, 2022. https://oss.deltares.nl/

Davenport, WB, 1970. *Probability and Random Processes: An Introduction for Applied Scientists and Engineers.* New York, NY, McGraw-Hill Inc., pp 273-274.

Epstein, ES, 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.*, **8 ,** 985–987.

Fielding, AH, Bell, JF, 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24 (1), 38–49.

Fienen, MN, and NG Plant, 2015. "A cross-validation package driving Netica with python", Environ. Modell. Software, 63, 14–23, doi:10.1016/j.envsoft.2014.09.007.

Friedrichs, CT, Rennie, SE, Brandt, A, 2016, Self-burial of objects on sandy beds by scour: A synthesis of observations, *Scour and Erosion*. CRC Press, 179–189.

Garcia, MH, BJ Landry, 2018. "Large-Scale Laboratory Experiments of Incipient Motion, Transport, and Fate of Underwater Munitions under Waves, Currents, and Combined-Flows," SERDP Project MR-2410, Final Report, April 2018.

Gutierrez, BT, NG Plant, ER Thieler, A Turecek, 2015. "Using a Bayesian network to predict barrier island geomorphologic characteristics", J. Geophys. Res. Earth Surf., 120, pp. 2452–2475.

HUGINEXPERT, hugin.com, 2022.

Jensen, FV, Nielsen, TD, 2007. Bayesian Networks and Decision Graphs, Springer, New York.

Mathworks, 2022. https://matlab.mathworks.com/

Murphy, AH, 1969. On the ranked probability skill score. *J. Appl. Meteor.*, **8 ,** 988–989.

NATO Naval Armaments Group, Maritime Capability Group 3, Mine Countermeasures and Harbour Protection, 2007. Final Report by Specialist Team for Sea Mine Burial Expert System (SMBES), NATO Document AC/141(MCG/3)D(2007)0001, NATO Unclassified.

Norsys Software, 2022. **https://www.norsys.com/netica.html.**

Palmsten, M, A Penko, 2020. "Probabilistic Environmental Modeling System for Munitions Mobility," SERDP Project MR-2733, SERDP Interim Report, September 2020.

Palmsten, M, D Frank-Gilchrist, RC Mickey, A Penko, K Koetje, S Harrison, 2021. “Munition Mobility Hindcasts during Hurricane Sally,” SERDP Project MR2733, SERDP/ESTCP Symposium 2021.

Penko, A. Ioup, E, 2021. Munitions Response Library for Site Management, ESTCP Project MR21-5207, SERDP/ESTCP Symposium 2021

Plant, NG, Holland, KT, 2011. Prediction and assimilation of surf-zone processes using a Bayesian network. Part I: Forward models, *Coastal Eng., 58* (1), 119-130.

Poelhekke L, WS Jäger, A van Dongeren, TA Plomaritis, R McCall, Ó Ferreira, 2016. "Predicting coastal hazards for sandy coasts with a Bayesian Network,” Coastal Engineering 118, pp. 21-34.

Puleo, JA, Cristaudo, D, 2020. Quantification of Hydrodynamic Forcing and Burial, Exposure and Mobility of Munitions on the Beach Face, Final Report for SERDP Project MR-2503.

Rennie SE, A Brandt, N Plant, 2007. A Probabilistic Expert System Approach for Sea Mine Burial Prediction, *IEEE Journal of Oceanic Engineering*, Vol. 32, No.1, *Special Issue on Mine Burial*, 260–272, Jan. 2007.

Rennie SE, A Brandt, 2017a. “Underwater Munitions Expert System for Remediation Guidance”, SERDP MR-2645 Presentation at 3rd Workshop on Burial and Mobility Modeling of Munitions in the Underwater Environment, June 2017.

Rennie SE, A Brandt, 2017b. “Underwater Munitions Expert System: Demonstration and Evaluation Report”, Johns Hopkins University/Applied Physics Laboratory, FPS-R-17-0321, SERDP Final Report for MR-2227, November 2017.

Rennie, SE, Brandt, A., Friedrichs, CT, 2017. “Initiation of motion and scour burial of objects underwater”, *Ocean Engineering, 131*, 282–294. dx.doi.org/10.1016/j.oceaneng.2016.12.029.

Rennie SE, A Brandt, JG Ligo, 2019. “Prototype Underwater Munitions Expert System: Demonstration and User's Guide,” Johns Hopkins University/Applied Physics Laboratory, FPS-R-19-0695, SERDP Guidance Document, MR-2645, November 2019.

Rennie SE, A Brandt, 2020. “Community Feedback and Design of Visualization Output for the Underwater Munitions Expert System,” Johns Hopkins University/Applied Physics Laboratory, FPS-R-20-0281, SERDP Interim Report MR19-1126, July 2020.

Rennie SE, A Brandt, 2021. “Extension of the Underwater Munitions Expert System to Beach Face and Estuarine Environments,” Johns Hopkins University/Applied Physics Laboratory, FPS-R-21-0476, SERDP Interim Report MR19-1126, August 2021.

Rennie SE, 2022. “Programmer's Guide to UnMES Construction & Spatial Application,” Johns Hopkins University/Applied Physics Laboratory, FPS-R-22-**DRAFT,** SERDP Final Report Addendum MR19-1126, *in preparation.* June 2022.

SERPD, 2010. *Munitions in the Underwater Environment: State of the Science and Knowledge Gaps*, SERDP White Pater, June, 2010.

Soulsby, RL, 1997. Dynamics of Marine Sands, London: Thomas Telford.

Sutherland J, Peet, AH, Soulsby, RL. 2004. "Evaluating the performance of morphological models", Coastal Engineering, 51, 917-939. (DOI: 10.1016/j.coastaleng.2004.07.015)

Thompson, SK, 1987. “Sample Size for Estimating Multinomial Proportions,”, *The American Statistician*, Vol. 41, No. 1, 42-46.

Tortora, RD, 1978. "A Note on Sample Size Estimation for Multinomial Populations," *The American Statistician*, Vol. 32, 100-102.

Traykovski, P, T Austin, 2017. “Continuous Monitoring of Mobility, Burial and Re-exposure of Underwater Munitions in Energetic Near-Shore Environments”, SERDP Project MR-2319 Final Report, January 2017.

Traykovski, P, Jaffre, F, 2020. Rapid Response Surveys of Mobility, Burial and Re-exposure of Underwater Munitions in Energetic Surf-Zone Environments and Object Monitoring Technology Development, SERDP Project MR-2729 Final Report.

Traykovski, P, Palmsten, M, 2020. Wave Resolved and Averaged Numerical Modelling of UXO Migration, SERDP Proposal MR21-S1-1341.

Trembanis, A, DuVal, C, 2020. Unexploded Ordnance Characterization and Detection in Muddy Estuarine Environments, SERDP MR-2730 Final Report.

AP Weigel, MA Liniger, C Appenzeller, 2007. The Discrete Brier and Ranked Probability Skill Scores, Monthly Weather Review, Vol 135, pp 118-124.

Whitehouse, R, 1998. Scour at marine structures. London: Thomas Telford.

Whitehouse, R, Soulsby, R, Roberts, W, Mitchener, H, 2000. Dynamics of Estuarine Muds. A Manual for Practical Applications. HR Wallingford Report SR527, Wallingford, Oxon.