

ORDNANCE AND EXPLOSIVES SITE STATISTICAL

SAMPLING METHODOLOGY (SITESTATS)

Dr. Jeffrey L. Riggs

Quantitech Inc.

Huntsville, Alabama

Arkie Fanning PE

U.S. Army Engineering and Support Center

Huntsville, Alabama

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE AUG 1996	2. REPORT TYPE	3. DATES COVERED 00-00-1996 to 00-00-1996			
4. TITLE AND SUBTITLE Ordnance and Explosives Site Statistical Sampling Methodology (Sitestats)		5a. CONTRACT NUMBER			
		5b. GRANT NUMBER			
		5c. PROGRAM ELEMENT NUMBER			
6. AUTHOR(S)		5d. PROJECT NUMBER			
		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Corps of Engineers, Army Engineering and Support Center, Huntsville, PO Box 1600, Huntsville, AL, 35807-4301		8. PERFORMING ORGANIZATION REPORT NUMBER			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)			
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)			
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES See also ADM000767. Proceedings of the Twenty-Seventh DoD Explosives Safety Seminar Held in Las Vegas, NV on 22-26 August 1996.					
14. ABSTRACT see report					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 8	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

ABSTRACT

SiteStats/GridStats is a statistical sampling decision support tool that statistically characterizes the density of ordnance at FUD sites. SiteStats/GridStats was jointly developed by the Huntsville Division Corps of Engineers and QuantiTech Inc. This tool is designed to minimize the sampling requirements necessary to characterize Ordnance and Explosives (OE) contamination for a specified level of statistical confidence. The characterization includes a confidence interval and point estimate of contamination density using the hypergeometric probability distribution and the sequential probability ratio test. SiteStats is used to determine the ordnance density for the FUD site. GridStats (a submodule of SiteStats and a stand alone program) is used to determine the ordnance density for an individual grid. GridStats has significantly reduced the number of anomalies that must be investigated in a grid to determine the density of that grid. Under the current GridStats configuration, no more than 40% of anomalies have to be investigated to achieve statistical confidence of the density of the grid. Normally much less than 40% of the anomalies have to be investigated (on average 20-25%). SiteStats/GridStats uses sophisticated statistical techniques such as a sequential probability ratio test and clustering algorithms to both increase the statistical confidence and to decrease the data required to be investigated. SiteStats/GridStats is designed to be resident on a laptop computer, is used in the field by non-technical people and to be extremely user friendly. SiteStats/GridStats has been used at numerous FUD sites and has saved the Government significant amounts of money in the investigation stage of the Engineering Evaluation and Cost Analysis phase and has led to the development of much more accurate ordnance density estimates for the sites (the ordnance density is required to determine the public risk inherent at the site). Future anticipated enhancements to SiteStats/GridStats include developing a Bayesian version based on historical usage, use of a proportion test in GridStats rather than a discrete test, and the development of a binomial module that would eliminate the need for flagging individual anomalies.

1.0 INTRODUCTION

The U.S. Army has over 1,300 Formerly Used Defense Sites (FUDs) that have possible ordnance contamination. Many of these sites are significantly contaminated but many others are either not contaminated or the contamination is not extensive. A major problem has been to decide which sites should be cleaned and to what extent. One of the major inputs into this decision is what is the density of unexploded ordnance at the site. To arrive at this decision point given a minimum of data has been a major goal of the U.S. Army Engineering and Support Center, Huntsville, (USAEDCH). One of the engineering tools developed to achieve this goal is SiteStats/GridStats.

SiteStats/GridStats is a statistical methodology that predicts the amount of ordnance at a given site based on the results of a statistical survey. This survey utilizes about 20-40% of the information in a grid to determine the most probable ordnance density for that grid. Further, the methodology utilizes representative grids to determine the most probable ordnance density for a sector (a homogenous portion of a site) and the sector data is utilized to determine the most probable ordnance density for a site. GridStats is a subset of SiteStats but has proved so useful that a GridStats stand alone model has been developed. GridStats determines the ordnance density within a grid. SiteStats determines the ordnance density within a sector and within a site.

The density estimate developed from SiteStats is then utilized by another USAEDCH methodology (Ordnance and Explosive Cost-Effectiveness Risk Tool-OECert) to determine the public and individual risk of encountering ordnance at a site. This model provides the decision maker with statistically determined information about the relative safety of the site. Given this information, the decision maker can more readily determine which sites to concentrate on and what degree of clean up is necessary at the site.

2.0 SEQUENTIAL PROBABILITY RATIOS

SiteStats and GridStats utilize Sequential Probability Ratio Tests (SPRT) in order to decrease the amount of data required to make a decision. In a nutshell, SPRTs utilize not only the content of the data found but also the order of the data found to make predictions about the density of the grid. For instance, if you investigate ten anomalies and all have been scrap, the chances of the eleventh anomaly being scrap are higher than if all ten anomalies had been ordnance.

The use of SPRTs provide the same decision with about 50% less data investigation. This increase in predictability comes at a price. Not only must the number of anomalies be known (as would be required in a fixed sampling plan) but also the order of investigation of the anomalies must be known. Therefore when investigating a grid, the Government cannot know with any certainty when the analysis will be complete. That was the reason for the incorporation of stopping rules within GridStats. When a grid requires more investigation than the stopping rules allow, the grid is truncated and another grid is investigated. The information from both these grid feed SiteStats so the ultimate decision on clean up is not affected.

There are operational problems associated with SPRTs. The contractor must feed the information from each dig into a computer program and the program will tell him which anomaly to dig next. Similarly, the contractor must feed complete the current grid before the SPRT is certain whether enough information has been gathered about that sector and that site. How these operational problems where overcome will be detailed later.

3.0 GRIDSTATS METHODOLOGY

The current GridStats model incorporates an SPRT based on the hypergeometric distribution and some stopping rules to determine when enough sampling has been completed. The Hypergeometric distribution is the mathematical distribution that most resembles the way a

grid is investigated. For instance the hypergeometric distribution is used extensively when sampling without replacement from a known sample size. The sample size is known because the investigation crew investigates the number of anomalies in a grid (uses a magnetometer and flags all of the suspect anomalies). This number is used to determine when enough data has been gathered.

Other data necessary to run GridStats are values for cost errors, risk errors and a discriminator value. The cost error (alpha) used is .20. This is the probability that GridStats has overestimated the number of anomalies in a grid. The risk error (beta) used is .10. This is the probability that GridStats has underestimated the number of anomalies in a grid. The discriminator value used is 5. This value is used in the hypothesis test that GridStats performs to determine if enough data has been collected. The hypothesis test is:

H(O): Ordnance items in grid \geq Discriminator (5)

H(1): Ordnance items in grid $<$ Discriminator (5)

Stopping rules have also been incorporated into GridStats to ensure efficient grid investigation. These rules are:

a) At least 5% of the anomalies in a grid must be sampled unless 20 ordnance items in a row are discovered

b) No more than 40% of anomalies in a grid will be sampled

The GridStats computer program is menu driven, easy to use and terminates automatically when the statistical goals have been reached.

4.0 SITESTATS METHODOLOGY

Sitestats utilizes a poisson sequential ratio test to determine when the sector has been sufficiently sampled. The SiteStats hypothesis test is:

H(0): The sector has Poisson homogenous density

H(1): The sector does not have Poisson homogenous density

The underlying technique that determines if the hypothesis is accepted is a Hopkins Statistic. This statistic determines the probability that the sector has homogenous density. It is an iterative process. At a minimum two grids must be evaluated in a sector. The statistic then calculates the probability the sector is homogenous. If two grids are insufficient then another grid is investigated and the procedure iterates again. This process continues until the calculated statistic satisfies the alpha (.20) and beta (.10) constraints.

Embedded in SiteStats is a clustering algorithm. This algorithm determines if the sectoring that has been postulated is accurate. The clustering algorithm uses the ordnance densities for the grids and the spatial differences between the grids to calculate the probability that the grids belong to an homogenous sector. This provides an excellent check on the initial sectoring activity. Often very little information is available about the density of a sector and the OE team assumes a great deal. The clustering algorithm helps ensure that we have properly identified the appropriate densities for the appropriate sectors at the site.

The SiteStats computer program is menu-driven, easy to use and will terminate automatically when the sectors have been statistically proven.

5.0 GRIDSTATS OPERATIONAL ISSUES

A significant problem was the sequential and random nature of the anomaly investigation. When GridStats was first used the magnetometer team spent a large amount of time finding out which sub-grid to go to next to investigate. Since the computer uses a set random number sheet to determine this sequence, the magnetometer team was provided a hard copy of the sequence. Now the team investigates the sub-grids and just calls in the result of every five digs. The computer program will automatically terminate when enough data has been gathered and the GridStats operator calls the team (via two way radios normally) to end the grid search.

6.0 SITESTATS OPERATIONAL ISSUES

The biggest SiteStats operational problem was the fact that the Government did not know how many grids were necessary for site characterization. To alleviate this problem a non-linear regression model was developed from site data to determine the most probable grid requirements. The upper bound for this value is used to establish the most probable needs for the Government. An option for more grids is also included in the contract in case more grids are required to characterize the site.

Another operational problem was the sequential nature of the program. The program is set up to provide a location for another grid only after the grid currently being investigated is completed. This problem was solved by having the OE team decide ahead of time which grids would be investigated. This will not affect the results as long as the team selects the grids within the sector at random. This allows the clearance team to stay well ahead of the magnetometer team.

7.0 FUTURE ENHANCEMENTS

Another version of SiteStats is being developed that is based on a binomial sequential probability ratio test. The use of the binomial will allow the Government to investigate a grid without having to first determine the number of anomalies in a grid (mag and flag). A value engineering study found that this procedure would save significant funds in future applications. The binomial is less expensive in areas where there is a large number of anomalies per grid. The hypergeometric model will continue to be used for those grids with relatively fewer anomalies. The cross over point is about 700 anomalies. The determination of which models to use will be made at the site and will be another operational issue that must be dealt with.

The future version of SiteStats will also have an audit capacity. This version will save all the data and decisions that have been made in both GridStats and SiteStats. The evaluation of the data and decisions will assist USAESCH in determining the effectiveness of the programs and in deciding what improvements could be made in the future.

8.0 COMPUTER REQUIREMENTS

SiteStats is implemented in Visual Basic and requires the following:

- a) IBM compatible with 90286 processor or higher
- b) Minimum hard drive of 1 Megabyte
- c) Floppy disk to load (5 1/4" or 3 1/2")
- d) EGA, VGA, 8514, Hercules, or compatible monitor
- e) Minimum memory of 1 Megabyte
- f) Mouse
- g) Microsoft MS-DOS version 3.2 or later
- h) Windows version 3.0 or later in standard or enhanced mode