ORNL/TM-13588

Computer Science and Mathematics Division

EVALUATION OF STATISTICAL METHODOLOGIES USED IN U.S. ARMY ORDNANCE AND EXPLOSIVE WORK

George Ostrouchov
Gregory P. Zimmerman [a]
John J. Beauchamp [b]
Valerii V. Fedorov [c]
Darryl J. Downing [c]

[a] Center for Energy and Environmental Analysis
Energy Division
Oak Ridge National Laboratory
[b] Department of Mathematics
Lipscomb University
Nashville, Tennessee
[c] SmithKline Beecham
King of Prussia, Pennsylvania

Date Published: September 1999

Prepared for the U.S. Army Engineering and Support Center, Huntsville under Interagency Agreement 2231-K085-A1 with the U.S. Department of Energy

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831-6285
managed by
Lockheed Martin Energy Research Corp.
for the
U.S. DEPARTMENT OF ENERGY
under contract DE-AC05-96OR22464

This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information. P. O. Box 62, Oak Ridge, TN 37831; prices available from (615) 576-8401.

Available to the public from the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Contents

List of Tables v
List of Figures vii
Acknowledgment ix
Abstract x
1 Introduction 1
2 Background and Overview 2
2.1 Background 2
2.2 Overview of the *SiteStats/GridStats* Approach and Methodology . 2
2.2.1 *SiteStats* Specifics 6
2.2.2 *GridStats* Specifics 7
2.3 Summary of Overall Findings 8
3 Detailed Comments on Each Component 9
3.1 *SiteStats* 10
3.1.1 Sampling of Anomalies 10
3.1.2 Hopkins-Like Statistic and Its Use 10
3.1.3 Spatial Interpolation Algorithm 16
3.1.4 The Migrating Means Algorithm 18
3.2 *GridStats* 18
3.2.1 Discriminator and Stopping Rules 18
3.3 Point Estimators and Variability 19
4 Recommendations for New Methodology 21
4.1 Density Estimation 21
4.2 Predicting UXO from Anomalies and Other Information 22
5 References 23
Appendix
A Minor Errors in *SiteStats/GridStats* Final Report 25
B Responses to Technical Comments on the Draft Version of This Document 27

List of Tables

1 *SiteStats* hand calculations for sample in Fig. 4. 15
2 50 point estimates from an unbiased estimator of the mean UXO per grid in sectors A and B 20

List of Figures

1 Sector and grid approach as used in *SiteStats/GridStats*. 3

2 Flowchart of the conceptual process incorporated into *SiteStats*. *Source:* [12] 4

3 Flowchart of the conceptual process incorporated into *GridStats*. *Source:* [12] 5

4 A 10×10 sector with 16 grids sampled. Grids marked a have a low number of UXO and grids marked b have a high number of UXO; subscripts show the sampling order for all 16 sampled grids. 13

5 A 10×30 sector with 16 grids sampled. All grids have the same number of projected UXO; subscripts show the sampling order for all the sampled grids. 14

6 Interpolation with linear decay weights (a) and exponential decay weights (b). Boldface values represent results from actual sampling; values in italics are the result of interpolation. 17

7 Two 10×10 sectors with different density variation but with same average UXO per grid. 20

Acknowledgements

The authors of this report gratefully acknowledge the programmatic guidance provided by Mr. Arkie Fanning of the U.S. Army Engineering and Support Center, Huntsville, Ala. The authors also acknowledge the assistance provided by Dr. Robert Mog in providing insight into the mathematics behind *SiteStats/GridStats* and in conducting numerous computer runs on sample data sets.

The authors also appreciate the telephone discussions conducted with Dr. Jeffery Riggs of QuantiTech, Inc., (QuantiTech is the developer of the software, Professor Bruce Barrett of the University of Alabama, Tuscaloosa, and Professor George Milliken of Kansas State University. These three individuals, in conjunction with Dr. Mog, provided valuable insights into the principles behind the statistical methods in *SiteStats/GridStats* and/or offered viewpoints on alternative statistical approaches.

This work was sponsored by the U.S. Army Engineering and Support Center, Huntsville (USAESCH) under Interagency Agreement 2231-K085-A1 with the U.S. Department of Energy. The work was conducted primarily from October 1997 to March 1998. A draft report was delivered in March 1998. After receiving comments from USAESCH, a final report that included the comments and our replies was produced in August 1998. Formal publication as ORNL/TM-13588 occurred in September 1999.

Abstract

Oak Ridge National Laboratory was tasked by the U.S. Army Engineering and Support Center (Huntsville, AL) to evaluate the mathematical basis of existing software tools used to assist the Army with the characterization of sites potentially contaminated with unexploded ordnance (UXO). These software tools are collectively known as *SiteStats/GridStats*. The first purpose of the software is to guide sampling of underground anomalies to estimate a site's UXO density. The second purpose is to delineate areas of homogeneous UXO density that can be used in the formulation of response actions.

It was found that *SiteStats/GridStats* does adequately guide the sampling so that the UXO density estimator for a sector is unbiased. However, the software's techniques for delineation of homogeneous areas perform less well than visual inspection, which is frequently used to override the software in the overall sectorization methodology. The main problems with the software lie in the criteria used to detect nonhomogeneity and those used to recommend the number of homogeneous subareas.

SiteStats/GridStats is not a decision-making tool in the classical sense. Although it does provide information to decision makers, it does not require a decision based on that information. *SiteStats/GridStats* provides information that is supplemented by visual inspections, land-use plans, and risk estimates prior to making any decisions.

Although the sector UXO density estimator is unbiased regardless of UXO density variation within a sector, its variability increases with increased sector density variation. For this reason, the current practice of visual inspection of individual sampled grid densities (as provided by *SiteStats/GridStats*) is necessary to ensure approximate homogeneity, particularly at sites with medium to high UXO density. Together with *SiteStats/GridStats* override capabilities, this provides a sufficient mechanism for homogeneous sectorization and thus yields representative UXO density estimates.

Objections raised by various parties to the use of a numerical "discriminator" in *SiteStats/GridStats* were likely because of the fact that the concerned statistical technique is customarily applied for a different purpose and because of poor documentation. The "discriminator," in *SiteStats/GridStats* is a "tuning parameter" for the sampling process, and it affects the precision of the grid density estimates through changes in required sample size.

It is recommended that sector characterization in terms of a map showing contour lines of constant UXO density with an expressed uncertainty or confidence level is a better basis for remediation decisions than a sector UXO density point estimate. A number of spatial density estimation techniques could be adapted to the UXO density estimation problem.

1. Introduction

One of the principal challenges facing the U.S. Army Corps of Engineers (COE) is the characterization and evaluation of sites (such as former bombing ranges, artillery ranges, or munition burial grounds) potentially containing unexploded ordnance (UXO). Within the U.S. Department of Defense, the U.S. Army Engineering and Support Center, Huntsville (USAESCH), has been designated as a center of expertise through which all UXO work is coordinated. The responsibilities of the USAESCH include identification of potential UXO sites, site surveys, site characterization, risk assessment, site prioritization, and any necessary UXO removal actions.

To accomplish its goals, the USAESCH has developed a suite of mathematical tools and software packages that assist in the characterization of UXO sites and the assessment of risks from UXO. This report documents the effort of staff members at the Oak Ridge National Laboratory (ORNL) in evaluating the mathematical tools employed by the U.S. Army to characterize UXO sites. ORNL was specifically tasked by the USAESCH to evaluate the mathematics behind a set of software codes collectively called *SiteStats/GridStats*.

For this effort, ORNL relied heavily on the existing documentation [12, 11, 10, 13] for *SiteStats/GridStats*. ORNL was unable to obtain direct assistance from QuantiTech, Inc., the developer of the software, because of the unavailability of key QuantiTech personnel during the performance period for ORNL's effort. However, ORNL was able to interact significantly with Dr. Robert Mog, formerly of QuantiTech, one of the developers of *SiteStats/GridStats*. Dr. Mog provided additional insight beyond what was contained in the existing documentation for *SiteStats/GridStats* and also conducted software runs on sample data sets created by ORNL.

The purpose of ORNL's effort was to offer an assessment and commentary on the applicability and appropriateness of the mathematics and statistics currently employed in *SiteStats/GridStats*. ORNL was not tasked to conduct investigations into the computer codes themselves but rather was asked to evaluate the mathematical equations as currently documented. Evaluation of the USAESCH's risk assessment tools (specifically, a software tool called OECert) was beyond the scope of ORNL's effort.

Section 2 of this report provides some necessary background information for the reader to understand the focus of ORNL's effort. This section also highlights and summarizes several of the findings of the evaluation. Section 3 presents the details of the investigation into the statistical methodology behind *Site-*

Stats/GridStats. Section 4 provides some recommendations on better approaches.

2. Background and Overview

2.1. Background

SiteStats/GridStats is not a decision-making tool in the classical sense. Although it does provide information to decision makers, it does not require a decision based on that information. *SiteStats/GridStats* provides information that is supplemented by visual inspections, land-use plans, and risk estimates prior to making any decisions.

The basic approach by *SiteStats/GridStats* is illustrated in Fig. 1. In this figure, a sector has been identified as potentially contaminated with UXO. The initial identification and definition of a sector and its boundaries is accomplished by methods outside the scope of *SiteStats/GridStats*. These methods include archival searches on the previous uses of the area, interviews with personnel familiar with the general prior usage of the area, and/or physical inspection of the area.

Anomalies are investigated in grids within the sector (grids are shown in Fig. 1). Some common grid sizes are 50×50 ft, 100×100 ft, 100×200 ft, and 200×400 ft. Before running the *SiteStats/GridStats* software (i.e., before implementing the embedded statistical methodologies), the USAESCH would survey the entire grid with magnetometers or gravitometers to identify buried anomalies. Within the sector, the *SiteStats/GridStats* software assists with the statistical sampling of these anomalies.

According to the *SiteStats* documentation, "the concept of SiteStats is simple: accept a small amount of uncertainty in characterizing individual grids in exchange for a much greater understanding of the contamination of the overall site using sequential sampling techniques to minimize costs" [12]. *SiteStats/GridStats* applies statistical methods that provide estimates of UXO densities without the need for sampling or investigating 100% of the anomalies.

2.2. Overview of the *SiteStats/GridStats* Approach and Methodology

The primary purposes of *SiteStats/GridStats* are to provide guidance for sampling anomalies within a sector and to use this information to delineate homogeneous areas within the sector. The statistical methods built into *SiteStats* direct this grid-by-grid sampling process as illustrated in Fig. 2. (Note that in Fig. 2, activities within the box labeled "Sample Grid" are conducted within *GridStats* as

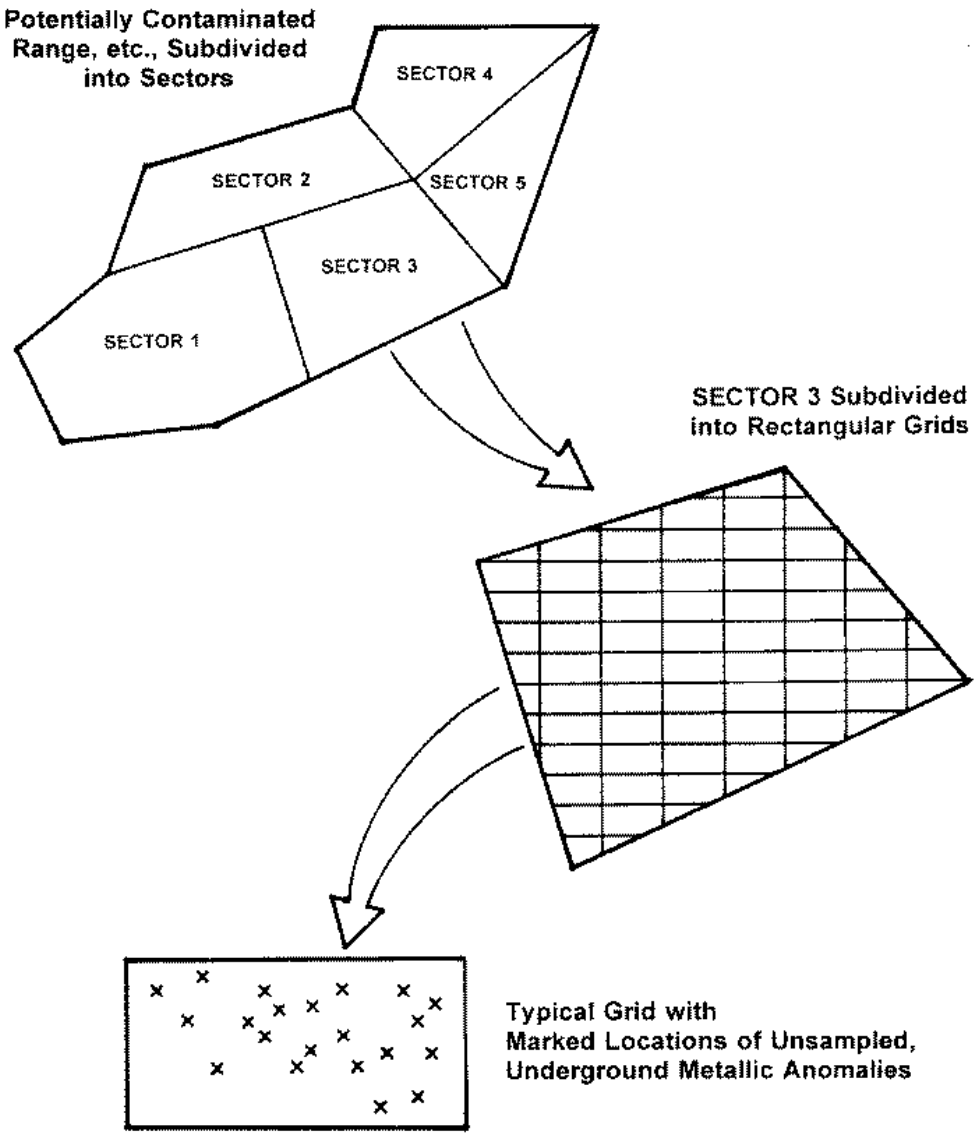

Figure 1: Sector and grid approach as used in *SiteStats/GridStats*.

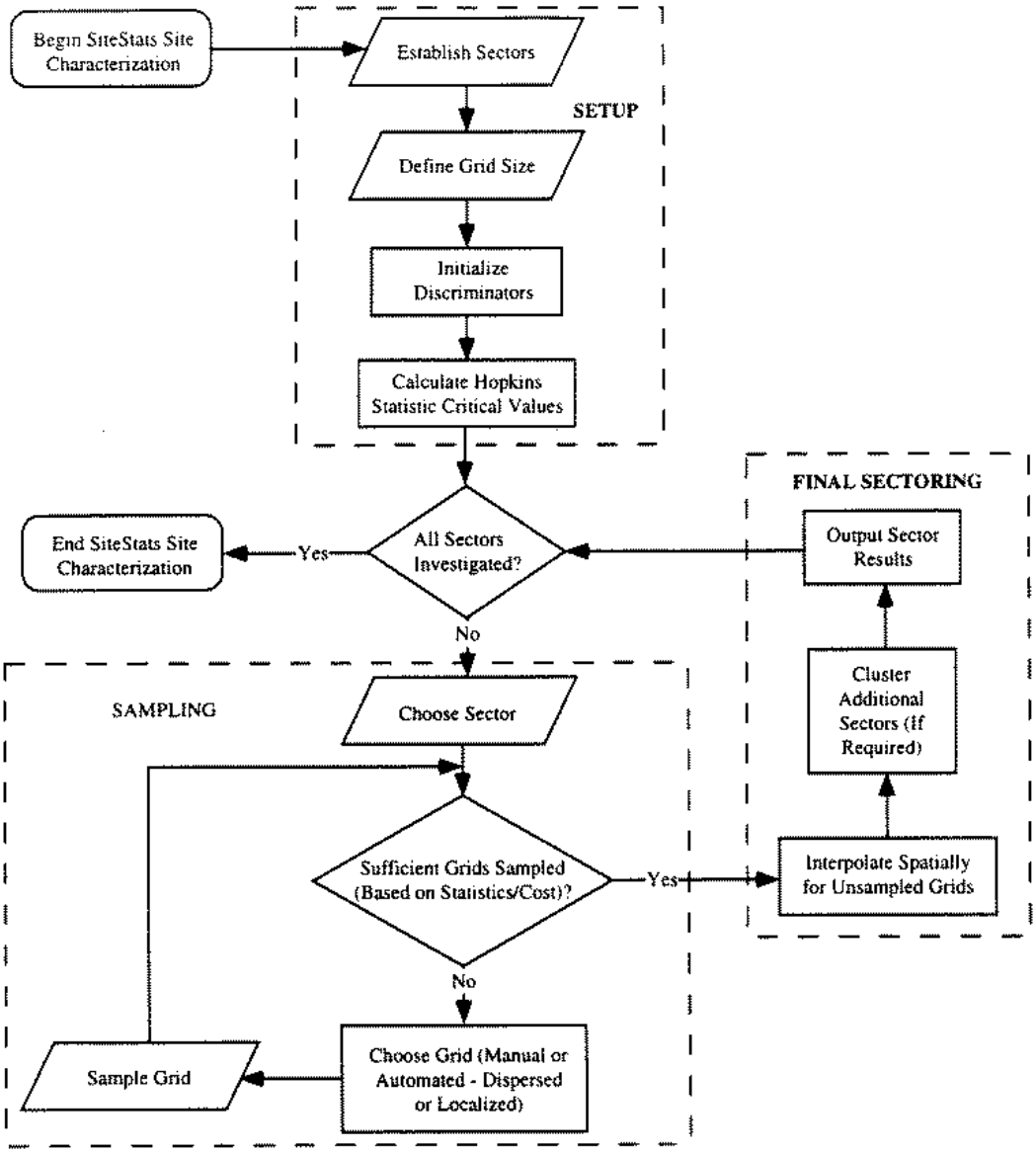

Figure 2: Flowchart of the conceptual process incorporated into *SiteStats*. *Source:* [12]

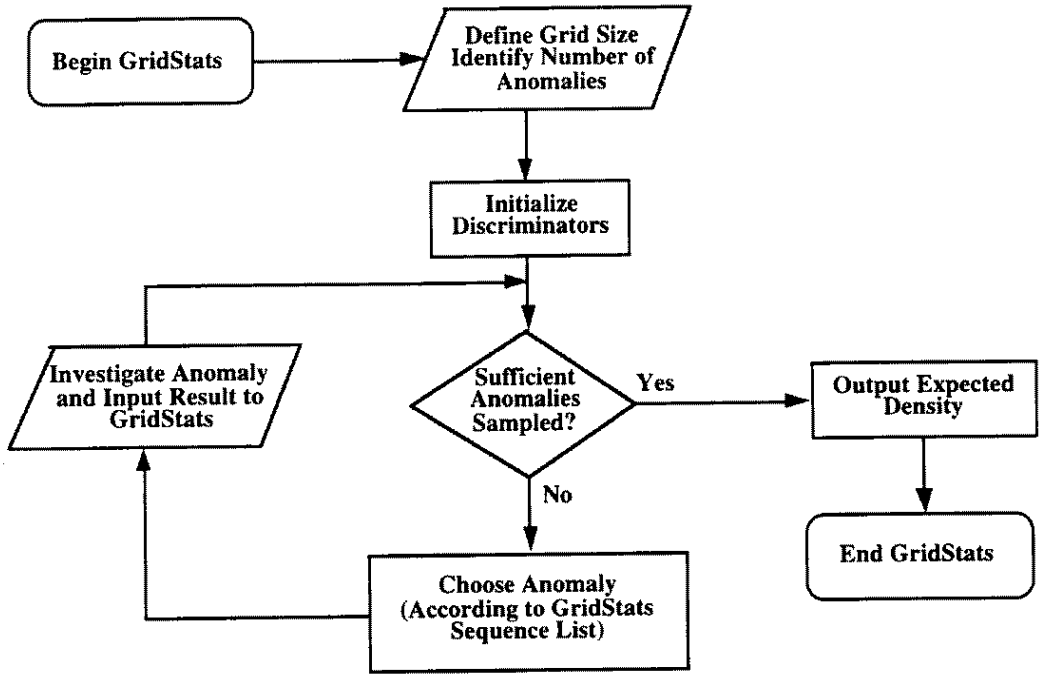

Figure 3: Flowchart of the conceptual process incorporated into *GridStats*. *Source:* [12]

described subsequently.) The objective of *SiteStats* is to identify a region (defined as a set of grids, or even the entire sector) where a common response action can be undertaken. Such response actions include removal actions and cleanup to various levels as well as "no further action required." The principal step in this identification and determination process is the evaluation of the "homogeneous" nature of a sector in regard to the spatial distribution of UXO within that sector.

SiteStats incorporates a collection of statistical methods and ad hoc rules to declare a sector homogeneous or to subdivide it into smaller sectors that are themselves homogeneous. A desired outcome of this process is an estimate of UXO density in each homogeneous sector. This is done in *GridStats* by sampling anomalies within the grids. Each time the sampling of a new grid is completed, a statistical test is applied by *SiteStats* to determine whether the sector is homogeneous. If no determination can yet be made, another grid is sampled.

Anomaly sampling within a grid is controlled by *GridStats*, as illustrated in Fig. 3. Anomalies are sampled until the statistical tests in *GridStats* categorize

the grid as either having (a) more than a certain number of UXO per grid or (b) less than this same number of UXO per grid. This value is called the "UXO discriminator" and, as currently used in *GridStats*, is numerically equivalent to 5 UXO per grid. Note that the amount of UXO found does not in and of itself require a removal action. Finding 5 UXO per grid across a sector does not necessarily infer that the sector will be cleared (although it might be), more importantly, finding less than 5 UXO per grid across a sector definitely does not infer that the sector will not be cleared. Decisions regarding removal actions are based on more information than the numerical UXO density estimates obtained from *SiteStats/GridStats*.

When a statistical conclusion regarding the degree of contamination has been reached, the grid's *projected* UXO count is presented to *SiteStats*, which continues its sector homogeneity testing with the information from this additional sampled grid. The results of the sampling that led to a conclusion regarding homogeneity are then used to estimate UXO density.

2.2.1. *SiteStats* Specifics

As discussed previously, *SiteStats* evaluates the grid sampling within a sector by using a statistical measure of UXO homogeneity. This measure is called the Hopkins Statistic and is described in greater detail in Sect. 3. Within *SiteStats*, the Hopkins Statistic is embedded into a sequential probability ratio test (SPRT) that attempts to answer the question about homogeneity. According to the *SiteStats* documentation [12], the question addressed by the SPRT can be stated as "Is the sector of interest homogeneous with respect to the UXO spatial distribution, or does it appear that a significant spatial variation in UXO indicates that more than one random process is appropriate for modeling the presence of UXO in this sector?"

SiteStats employs "stopping rules" to determine when to halt the sequential grid sampling that occurs within a sector. The stopping rules are associated with three parameters built into *SiteStats*:

- a Type I error value,
- a Type II error value, and
- Hopkins Statistic critical value 0.62.

The Type I error is defined as the probability associated with the statistical conclusion that an area has nonhomogeneous UXO density when actually the

density is homogeneous. The Type II error is defined as the probability associated with the statistical conclusion that an area has homogeneous UXO density when actually the density is nonhomogeneous across this area.

The Hopkins Statistic critical values are used to determine whether clustering is required after the sampling is complete. A value of the Hopkins Statistic above 0.62 indicates that the sector is nonhomogeneous. If a sector is nonhomogeneous, the Hopkins Statistic for the last sampled grid is used to indicate a preferred number of clusters (i.e., new sectors to be subdivided from the original sector) that can encompass the nonhomogeneous portions of the sector. To aid in the identification of similar grids that can be "clustered" into a new sector, *SiteStats* uses a migrating means algorithm with inverse Manhattan distances for interpolation.

2.2.2. *GridStats* Specifics

GridStats provides the user with a sampling sequence to be followed for the actual field investigation and characterization of anomalies within a grid. The sampling can result in identification of the anomaly as (a) UXO, (b) UXO-related scrap, or (c) other ferrous items. This anomaly-by-anomaly sampling process continues until the statistical "stopping rules" in *GridStats* indicate that sampling can be halted. The statistical methods used in *GridStats* to halt sampling include a SPRT.

Three specific parameters are used to determine when to halt sampling of anomalies within a grid. These three parameters are

- the cost error α,
- the risk error β, and
- the UXO discriminator, D.

The cost error is defined as the probability associated with the statistical conclusion that an area has UXO present at levels higher than the value of the UXO discriminator, D, when the area actually does not. This conclusion could result in more expansive cleanup actions than necessary. Within *GridStats*, the value of the cost error is 0.20 (i.e., 20% probability).

The risk error is defined as the probability associated with the statistical conclusion that an area has UXO present at levels less than the value of the UXO discriminator, D, when the area actually has higher levels of UXO. This

conclusion could result in inadequate cleanup actions for the existing level of UXO present. Within *GridStats*, the value of the cost error is 0.10 (i.e., 10%).

According to the *SiteStats/GridStats* documentation [12], the value of the UXO discriminator, D, depends on the total number of anomalies within a grid. The documentation states that D equals 5 if the total number of anomalies is 213 or less and that D equals 0.0235 times the total number of anomalies in the grid if the total number of anomalies is more than 213. However, Dr. Robert Mog, one of the developers of the *SiteStats/GridStats* software, has told ORNL[1] that the latter computation was never implemented in the *SiteStats/GridStats* software and that the value of the discriminator, D–as currently used–always equals 5.

2.3. Summary of Overall Findings

ORNL's overall evaluation is that *SiteStats/GridStats* does provide useful information for site characterization. However, several problems were uncovered as discussed in the following summary paragraphs. Additional details are presented in Sect. 3.

ORNL's first observation and finding is that there are several instances where the mathematics and/or modules presented and described in the existing *SiteStats/GridStats* documentation differ from what is actually implemented in the software. One such example is the constant value of the UXO discriminator value, D, discussed previously. An additional instance is discussed in the next paragraph.

ORNL staff found that the Hopkins [7] statistic is not the best available statistical tool for determining homogeneity. This issue is discussed further in Sect. 3.1.2. Furthermore, the mathematical equations implementing the statistic, as documented in the *SiteStats/GridStats* manuals [12], are only *motivated* by the Hopkins statistic but are different in form. The Hopkins-like statistic does not perform adequately. After consultation with Dr. Robert Mog, it was further discovered that the documented equations were modified before implementation into the software. These modifications are not covered in the existing *SiteStats/GridStats* documentation. This issue is discussed in detail in Sect. 3.1.2. Based on our theoretical investigations described in Section 3.1.2 and a few grid-based examples, the homogeneity analysis portion of *SiteStats/GridStats* software performs better than the documented version but still significantly less well than visual inspection. It should be noted that in practice, visual inspection is fre-

[1] Telephone conversations with G. Zimmerman, January 2, 1998

quently used to override the software in the overall sectorization methodology.

For sectors determined to be nonhomogeneous, the ORNL staff found that the migrating means algorithm is an appropriate method for forming clusters that define new sectors. However, better methods than the inverse Manhattan distance interpolation algorithm (as presently used in *SiteStats/GridStats*) do exist. This issue is discussed in detail in Sect. 3.1.3, below.

One of the statistical parameters, called the "UXO discriminator," D, used in *SiteStats/GridStats* is a "tuning parameter" for the sampling process within a grid. In *SiteStats/GridStats*, this discriminator has units of "number of UXO per grid" and has a numerical value of 5. It affects the precision of the grid density estimates through changes in required sample size and does not itself generate remediation decisions. Objections raised to the "discriminator" in *SiteStats/GridStats* by various parties were likely because of the fact that the concerned statistical technique is customarily applied for a different purpose (making decisions between two states) and because of poor documentation. This issue is discussed in more detail in Sect. 3.2.1, below.

While more sophisticated and computationally simple statistical methods are available, *SiteStats/GridStats* appears to work acceptably; however, this conclusion must be qualified by the observation that it is based not so much on ORNL's evaluation of the mathematics and statistical methods as documented, but more so on the results of sample data sets used with the actual *SiteStats/GridStats* software. The performance of the software is moderated by the common sense of its users, who are able to see a map of grid results and override some of the software's homogeneity recommendations. Our reasons for looking at this map are discussed in Sect. 3.3. There does not appear to be a compelling reason to question the applicability of *SiteStats/GridStats* within anticipated "real world" values of UXO contamination for grid sampling, sector identification, grid clustering, and UXO density estimation. A more detailed discussion of ORNL's findings is in Sect. 3 and some recommendations for better approaches are in Sect. 4 of this report.

3. Detailed Comments on Each Component

Most of the comments in this section are based on existing *SiteStats/GridStats* documentation [12] and on a few examples that were constructed and executed with the software to check that our theoretical findings agree with software performance. The document [12] is often difficult to understand because it is mathematically inaccurate and frequently vague. In some cases it describes features

that are not present in the software. Apparently the authors of [12] had difficulty in communicating and documenting many of the statistical concepts that are presented. In addition to the discussion in this section, Appendix A lists some detailed errors and inconsistencies in [12].

3.1. *SiteStats*

3.1.1. Sampling of Anomalies

SiteStats/GridStats first randomly selects a grid and then does a random sample of anomalies within each grid. This is called two-stage sampling (see, for example, [2]). Without making any assumptions about spatial uniformity of UXO density (homogeneity), the spatial uniformity of the sampling scheme guarantees that a simple projection (multiplication by total-area/area-sampled) of the UXO found is an unbiased estimator of the total UXO within the sector. Thus, even if the homogeneity analysis fails, the UXO density estimate is unbiased for the sector because of the spatial uniformity of the sample.

3.1.2. Hopkins-Like Statistic and Its Use

Although a central purpose of *SiteStats/GridStats* is to delineate homogeneous sectors with respect to UXO density, homogeneity is not defined explicitly. *SiteStats* implicitly defines homogeneity by the use of a Hopkins [7] motivated statistic that we discuss in this section.

SiteStats computes a Hopkins-like statistic for each grid that is sampled. After each grid is sampled, a hypergeometric SPRT attempts to conclude that either less than 50% or more than 50% of the grids in the sector have a Hopkins statistic over 0.62. In the former case ($< 50\%$), the sector is declared homogeneous, and in the latter case ($> 50\%$) it is declared not homogeneous. Each inconclusive test will result in sampling an additional grid. If the SPRT is inconclusive after $n_s(\max)$ (see p. A-3 in [12]) grids are sampled, the Hopkins value for the last grid sampled determines the conclusion[2]. If this value is 0.62 or less, the sector is declared homogeneous, and otherwise it is declared not homogeneous. When a sector is declared not homogeneous, the Hopkins value of the last grid sampled determines the number of clusters.

First, we discuss the Hopkins statistics as reported in statistical literature, and then we turn to its adaptation in *SiteStats*. We also make some comments on the

[2]Telephone conversations between R. Mog and G. Ostrouchov, December 1997, and January 1998

SPRT and its implementation, and finally we comment on the recommendation for the number of clusters.

The Hopkins statistic [7] considers two kinds of Euclidean distances. In terms of the present application, one distance is from a random point in space to the nearest UXO, denoted by U_i. The other kind of distance is from a random UXO to its nearest neighbor UXO, denoted by W_i. The statistic, for a sample of n random points and n random UXO, is then defined as

$$\frac{\sum_{i=1}^{n} U_i^2}{\sum_{i=1}^{n} U_i^2 + \sum_{i=1}^{n} W_i^2}. \tag{1}$$

When the UXO are spatially random with uniform density, this has a beta(n,n) distribution [7, 3]. Often, the Hopkins statistic is also presented on page 604 of [3] as

$$\frac{\sum_{i=1}^{n} U_i^2}{\sum_{i=1}^{n} W_i^2}, \tag{2}$$

in which case it has an $F_{2n,2n}$ distribution under the same spatial randomness assumption.

Although sensitive to density variation, the Hopkins statistic measures the tendency to cluster. Because there is a difference between clustering and density variation, and for this reason there are better methods for assessing density variation. The Hopkins statistic is a ratio of distances from random points in space to nearest UXO and distances from random UXO to nearest UXO. The intuition is that if the UXO are clustered (there is empty space between clusters), distances from random points in space will tend to be greater than distances from random UXO. If the UXO are regularly spaced the reverse is true. Hopkins statistic measures the continuum between regularity and clustering. Spatial randomness (usually defined as a homogeneous Poisson process) falls somewhere between regularity and clustering. Although clustering implies density variation, it is not true that density variation always implies clustering.

The difficulty in implementing the Hopkins statistic is that it requires the selection of a random sample of UXO to compute the W_i (see pp.609 and pp.611 of [3]). A random sample requires either knowledge of all UXO or knowledge of UXO density. Apparently, the developers of *SiteStats* realized this and developed

a different statistic as an adaptation of the original Hopkins statistic.

SiteStats does not record the positions of individual UXO, but rather reports a projected total for each sampled grid. Consequently it does not use Euclidean distances between individual UXO. A *distance* function between two grids is defined as the sum of the Manhattan distance of grid coordinates and absolute difference in UXO projections for the two grids concerned (see p.H-4 in [12]). This *distance* function may have poor properties because a sum implies that a UXO is exchangeable for a Manhattan block. A better way to combine a UXO difference and distance would be a mathematical product.

Because it is not possible to implement the Hopkins statistic in the present application, *SiteStats* [12] defines a different Hopkins-like statistic for each sampled grid as

$$H = \frac{U}{U + W}, \tag{3}$$

where U is the *distance* from the sampled grid to its nearest neighbor sampled grid and W the average of the Us for all other sampled grids (see Appendix C in [12]). An H is computed for each grid sampled and then an SPRT is used to test whether more than 50% of the sector grids have $H > 0.62$ (see Appendix E in [12]).

This Hopkins-like statistic (3) fails to detect density differences in simple situations. Consider a sector with 10×10 grids, as shown in Fig. 4, where the left half has low UXO density a and the right half has high UXO density b. We will refer to this as the $a - b$ sector. The 16 labeled grids are sampled. The subscripts give the order of grid selection but this order is not relevant at this point. Regardless of what value of UXO density we assign grids a_i and what other value is assigned to grids b_i, U is 3 for every grid. Since $U = 3$ for every grid, we have $H = 0.5$ for every grid. That is, the Hopkins-like statistic (3) and the associated SPRT indicate homogeneity regardless of the UXO assigned to each half of the sector! In this example, all UXO information in the sample cancels out.

The regularity of the above sample was chosen only for simplicity of the exposition. The same failure mode persists if grids are sampled randomly and when the UXO found has a more random distribution. The problem with (3) is that it is based on local (nearest-neighbor) *distance*s. As soon as each density region has two grids sampled, the density cancels out in the U calculation because the nearest neighbor is in the same density region. In fact, large density differences discourage nearest neighbors in different density regions, thus minimizing density differences detected. This is the opposite of what common sense dictates: to dis-

	a_1			a_4			b_3		
		a_{12}			b_9			b_7	
a_{14}			a_6			b_{15}			b_{11}
	a_{16}			a_{13}			b_{10}		
		a_8			b_5			b_2	

Figure 4: A 10×10 sector with 16 grids sampled. Grids marked a have a low number of UXO and grids marked b have a high number of UXO; subscripts show the sampling order for all 16 sampled grids.

cover whether different density regions exist, the focus should be on the largest density differences.

When the example in Fig. 4 is presented to the *SiteStats/GridStats* software, the results are different from what is expected in the preceding calculations. For $a = 1$ and $b = 10$, the conclusion is that the sector is homogeneous. But for $a = 1$ and $b \geq 20$, the conclusion is that the sector is not homogeneous. Apparently the *SiteStats/GridStats* development team discovered that the Hopkins-like statistic was not working and made some changes that do not appear in the documentation[3]. Based on a few conversations with Dr. Robert Mog, we describe what appear to be the Hopkins calculations implemented in the software. A few hand calculated examples agree with the results obtained by the software.

After grid i is sampled, U_i is computed as the *distance* (Manhattan + UXO) to its current nearest neighbor, and W_i is computed as the current average *distance* of all other grids to their nearest neighbors [that is, W_i is the same as W in (3) when grid i is sampled]. The new Hopkins-like statistic for grid i is

$$H_i = \frac{\overline{U}_i}{\overline{U}_i + W_i}, \tag{4}$$

[3]Telephone conversations between R. Mog and G. Ostrouchov, December 1997, and January 1998

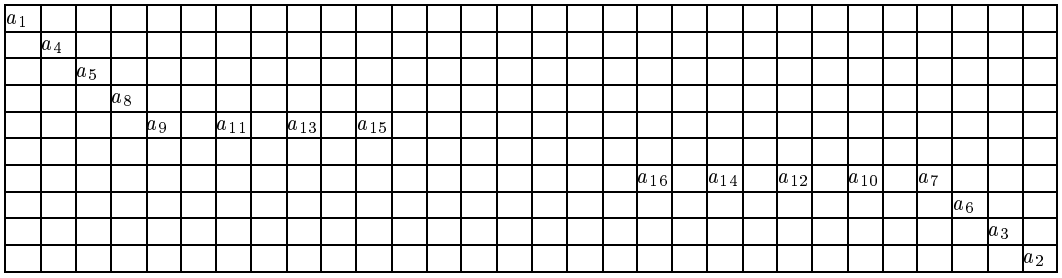

Figure 5: A 10×30 sector with 16 grids sampled. All grids have the same number of projected UXO; subscripts show the sampling order for all the sampled grids.

where $\overline{U}_i = \frac{1}{i}\sum_{j=1}^{i} U_j$. In contrast with the documented calculations (3) where U, W, and H are recomputed for all grids after a new grid is sampled, U_i, W_i, and H_i are computed only once when grid i is sampled.

Consider the sample in Fig. 4, drawn in the order indicated by the subscripts. All sampled grids a_i have a UXO, and all sampled grids b_i have b UXO. Table 1 shows the quantities U_i, W_i, and $\overline{U}_i$ involved in the new Hopkins-like statistic calculations for this sample. We show the individual components involved in the calculations. Note that all nearest neighbor *distance*s U_i, with the exception of U_2, do not involve the UXO density. This means that only one UXO difference out of the 16 sampled points contributes to the analysis of this $a - b$ sector; all other UXO information cancels out. This is only slightly better than the originally documented version where all UXO information cancels out when four points are sampled in the $a - b$ sector.

The preceding calculations indicate that *SiteStats* uses only a small fraction of the UXO information that it collects in making its homogeneity determination. As a result, its Hopkins statistics can be strongly influenced by sample placement.

Table 1: *SiteStats* hand calculations for sample in Fig. 4.

grid$_i$	nearest neighbor	U_i	W_i	$\overline{U}_i$
a_1	–	0	0	0
b_2	a_1	$15 + \lvert a - b \rvert$	$15 + \lvert a - b \rvert$	$\frac{15+\lvert a-b\rvert}{2}$
b_3	b_2	9	$\frac{(6+\lvert a-b\rvert)+9}{2}$	$\frac{24+\lvert a-b\rvert}{3}$
a_4	a_1	3	$\frac{3+9+9}{3}$	$\frac{27+\lvert a-b\rvert}{4}$
b_5	b_2	3	$\frac{3+3+9+3}{4}$	$\frac{30+\lvert a-b\rvert}{5}$
a_6	a_4	5	$\frac{3+3+9+3+3}{5}$	$\frac{35+\lvert a-b\rvert}{6}$
b_7	b_3	3	$\frac{3+3+3+3+3+5}{6}$	$\frac{38+\lvert a-b\rvert}{7}$
a_8	a_6	5	$\frac{3+3+3+3+3+5+3}{7}$	$\frac{43+\lvert a-b\rvert}{8}$
b_9	b_7	3	$\frac{3+3+3+3+3+5+3+5}{8}$	$\frac{46+\lvert a-b\rvert}{9}$
b_{10}	b_2	3	$\frac{3+3+3+3+3+5+3+5+3}{9}$	$\frac{49+\lvert a-b\rvert}{10}$
b_{11}	b_7	3	$\frac{3+3+3+3+3+5+3+5+3+3}{10}$	$\frac{52+\lvert a-b\rvert}{11}$
a_{12}	a_1	3	$\frac{3+3+3+3+3+3+3+5+3+3+3}{11}$	$\frac{55+\lvert a-b\rvert}{12}$
a_{13}	a_6	3	$\frac{3+3+3+3+3+3+3+4+3+3+3+3}{12}$	$\frac{58+\lvert a-b\rvert}{13}$
a_{14}	a_6	3	$\frac{3+3+3+3+3+3+3+4+3+3+3+3+3}{13}$	$\frac{61+\lvert a-b\rvert}{14}$
b_{15}	b_{11}	3	$\frac{3+3+3+3+3+3+3+3+4+3+3+3+3+3+3}{14}$	$\frac{64+\lvert a-b\rvert}{15}$
a_{16}	a_8	3	$\frac{3+3+3+3+3+3+3+3+3+3+3+3+3+3+3+3}{15}$	$\frac{67+\lvert a-b\rvert}{16}$

tween two values of a parameter is required (a "simple hypothesis" in the language of statistics) and the underlying process operates at or near one of the two values. In other cases, the average sample size may be greater than the sample size of an equivalent nonsequential test (see, for example, p.23 in [19]). The hypothesis tested in *SiteStats* is whether the proportion of Hopkins statistics over 0.62 is below or above 0.5 (a "compound hypothesis" in the language of statistics). This is a case where sequential tests are not optimal.

What large expected sample sizes mean for *SiteStats* is that $n_s(\max)$ (see p. A-3 in [12]) is often reached while the SPRT is still inconclusive. In this case, only the single Hopkins value of the last sampled grid determines the entire homogeneity analysis outcome. Because decisions are forced when the SPRT is inconclusive, the nominal Type I and Type II error rates are rendered meaningless.

Given a nonhomogeneity conclusion, the size of the last Hopkins value determines the number of clusters. Smaller values indicate more clusters. In terms of the original Hopkins [7] statistic, large values indicate a few strong multi-point clusters and small values indicate regularity. One could argue that regularity is a very large number of single-point clusters and thus that smaller values indicate more clusters. However, we know of no published results that investigate this. We further caution that the *SiteStats* Hopkins-like statistic is quite different from the original Hopkins statistic and we expect its properties to be rather different.

In summary, the statistical implications of the poor use of UXO information by the *SiteStats* Hopkins analysis have an analogy in terms of signal-to-noise ratio. The *SiteStats* Hopkins analysis has a very low signal to noise ratio; that is, to detect a signal that is not very loud, the volume has to be high. However, the accompanying noise causes incorrect conclusions. This is why *SiteStats* sometimes recommends too many subsectors to the clustering algorithm, while at other times it may overlook a need to subdivide.

The implicit definition of homogeneity that results from the use of *SiteStats* is not consistent between applications. In our estimation, visual inspection of a sector map of the grid UXO projections will provide better and more consistent guidance for sectorization than the *SiteStats* Hopkins analysis.

3.1.3. Spatial Interpolation Algorithm

Interpolation is required by the migrating means algorithm, which expects a UXO value in each grid. Spatial interpolation is the process of assigning a UXO value to each grid that has not been sampled. This value is a weighted average of the grids that have been sampled, and determination of these weights can have a great

5	*23*	*24*	**5**	*33*	*50*	*49*	*53*	*60*	**85**
24	*28*	*29*	*27*	*44*	**85**	*56*	*55*	*56*	*62*
27	*27*	*24*	**5**	*34*	*50*	*49*	*53*	*53*	*56*
28	*28*	*25*	*26*	*37*	*47*	*51*	*58*	*55*	*55*
28	*25*	**5**	*26*	*38*	*48*	*56*	**85**	*60*	*57*
29	*29*	*25*	*27*	*38*	*49*	*52*	*59*	*55*	*55*
29	*29*	*26*	**5**	*36*	*52*	*51*	*54*	*54*	*55*
30	*32*	*32*	*32*	*48*	**85**	*58*	*56*	*56*	*58*
25	*31*	*32*	*31*	*45*	*60*	*55*	*56*	*57*	*63*
5	*26*	*28*	**5**	*44*	**85**	*58*	*57*	*62*	**85**

(a)

5	*6*	*8*	**5**	*25*	*62*	*66*	*77*	*83*	**85**
6	*10*	*15*	*17*	*50*	**85**	*78*	*80*	*83*	*85*
6	*7*	*8*	**5**	*25*	*62*	*66*	*77*	*80*	*84*
6	*6*	*6*	*9*	*25*	*62*	*74*	*83*	*83*	*84*
6	*6*	**5**	*9*	*26*	*63*	*79*	**85**	*84*	*84*
6	*7*	*7*	*9*	*27*	*64*	*75*	*83*	*83*	*84*
7	*8*	*8*	**5**	*28*	*65*	*68*	*78*	*79*	*82*
9	*16*	*20*	*23*	*60*	**85**	*80*	*82*	*83*	*84*
6	*10*	*15*	*18*	*52*	*78*	*78*	*81*	*83*	*85*
5	*8*	*13*	**5**	*46*	**85**	*77*	*80*	*84*	**85**

(b)

Figure 6: Interpolation with linear decay weights (a) and exponential decay weights (b). Boldface values represent results from actual sampling; values in italics are the result of interpolation.

influence on the outcome. Consider two examples in Fig. 6. Inverse Manhattan distance is used in *SiteStats* and in (a) of Fig. 6. Part (b) of this figure illustrates inverse exponential Manhattan distance. This example illustrates the strikingly different results of two different weight functions. Considering that the bold grids are sampled, (b) seems to be the more appealing interpolation here. Euclidean distance would further enhance the result.

The interpolation algorithm in *SiteStats* is adequate, but better results can be obtained. The weight function that is appropriate in a given situation depends on the amount of spatial correlation present. This can be estimated from the data and is probably the central subject of most spatial statistics books (see, for example, [3, 14, 4, 8]).

3.1.4. The Migrating Means Algorithm

The algorithm described as "migrating means" (see p. 18 of [12]) is more widely known as the *Isodata Procedure* (see [5], for example). It is a simple procedure for unsupervised clustering, commonly used in image analysis. It is considered a computationally fast approximation to maximum likelihood estimation of cluster means. It needs to be told how many clusters to produce, which is why the Hopkins analysis gives a recommendation.

The migrating means algorithm, as implemented in *SiteStats*, appears to be appropriate for the task of clustering similar grids into homogeneous sectors. However, its performance could be enhanced by the use of better weight functions in the interpolation algorithm discussed in Sect. 3.1.3.

3.2. *GridStats*

GridStats samples anomalies within a grid according to a spatially random pattern. A hypergeometric SPRT is used to stop sampling when a conclusion can be reached regarding whether the UXO content of the grid is above or below the discriminator $D = 5$. In low UXO situations (< 5 UXO per grid), sampling is stopped by the *Final Stopping Rule*.

3.2.1. Discriminator and Stopping Rules

Much concern has arisen that the *GridStats* discriminator value somehow determines the amount of remediation performed at a given site. This issue has been addressed by several experts (including Professor Bruce Barrett, Professor George Milliken, and Dr. Robert Mog). These experts have all concluded, and we concur, that the discriminator is a tuning parameter that affects the sample size within grids and it has no significant impact on eventual remediation.

However, we would like to add that perhaps the reason for the original objections to the discriminator arise from an unusual application of the SPRT. The SPRT is a statistical tool designed to make decisions between two possible separated states. It is not easily seen in the documentation [12] that its application in *GridStats* throws the decision away and merely uses the fact that a decision was made to stop the sampling. Much attention in [12] is spent on discussing the Type I and Type II errors of a decision the outcome of which is not used, and this misleads many readers. We suggest that it is better to address the variance of the UXO estimate for a grid directly (and that is very simple), thus avoiding an SPRT entirely.

The hypothesis tested by the *GridStats* SPRT is "compound." As we discussed near the end of Sect. 3.1.2, an SPRT may require very large average samples in such situations. In particular, if the true UXO content is close to the discriminator value, the required samples will be large. This is because the SPRT must determine that the UXO content is statistically different from the discriminator value in order to terminate. The shorter this distance, the more precise estimate of the UXO content is needed, resulting in larger samples.

Because the SPRT required large sample sizes when less than five UXO were found, QuantiTech [12] introduced the *Final Stopping Rule*. We suspect that the SPRT stops the sampling only in highly contaminated grids. In applications where the UXO contamination is low, the SPRT is inconclusive; hence, the *Final Stopping Rule* applies. Similarly, as in the *SiteStats* application of an SPRT, Type I and Type II error rates are rendered meaningless. They also become tuning parameters that do not have simple interpretations.

3.3. Point Estimators and Variability

In the case of the Hopkins statistic [7], the null hypothesis tested is that the data are generated by a homogeneous Poisson process. If the null hypothesis is accepted, then the spatial variability is approximately that of a homogeneous Poisson process with the estimated mean. Although the statistic in *SiteStats/GridStats* is motivated by the Hopkins statistic [7], its form is rather different and its sampling properties have not been investigated. As a result, a sector declared homogeneous by *SiteStats/GridStats* does not necessarily have the variability of a homogeneous Poisson process. We include this section to illustrate a potential problem that can result from the use of a point estimator without an estimate of variability.

Suppose we have two sectors with 10×10 grids, as illustrated in Fig. 7. Sector A has half of the grids with 1 UXO and the other half with 10 UXO. Sector B has half of the grids with 5 UXO and the other half with 6 UXO.

The average UXO per grid is 5.5 for both sectors A and B. If our estimation process randomly selects 10 grids and computes the average UXO per grid, we have an unbiased point estimator of the average UXO per grid in the sector. Table 2 shows 100 point estimates obtained by repeating this estimation process 50 times in each of the sectors; this was done with the aid of Splus [15]. Note that for sector B all of the results characterize the sector well. On the other hand, the results for sector A are mixed.

If 2.8 would result in the same remediation decision as 9.1, then no problem ex-

1	10	1	1	1	1	1	1	1	1
1	10	1	1	1	1	1	10	10	1
1	10	10	1	1	1	10	10	1	1
1	10	10	1	1	1	1	1	1	1
1	10	10	1	1	1	1	1	1	1
10	10	10	10	10	10	1	10	10	10
10	10	10	10	10	1	1	10	10	10
10	10	1	1	10	1	10	10	10	10
10	1	1	10	10	1	10	1	10	10
10	1	1	10	10	10	10	10	10	10

Sector A

5	5	5	5	5	5	5	5	5	5
5	6	5	6	6	6	5	5	5	5
5	6	6	5	6	5	5	6	6	5
5	6	5	5	6	6	6	5	5	5
5	5	5	6	6	5	5	5	5	5
6	6	6	6	5	6	6	6	6	6
5	6	5	6	5	5	6	5	6	6
6	6	6	6	6	6	6	6	5	6
5	5	5	6	6	6	6	5	5	6
6	5	5	6	6	6	6	6	5	6

Sector B

Figure 7: Two 10×10 sectors with different density variation but with same average UXO per grid.

ists. If it would not, we need to look at the individual grid results or some measure of variability in the sample. The 2.8 came from the sample (10,10,1,1,1,1,1,1,1,1), and the 9.1 came from the sample (10,10,10,10,10,10,10,10,10,1). Seeing the individual values gives a much better idea of what can be found in the sector than the point estimates. This is why we recommend continuing the current practice of visual inspection of the individual sampled grid densities.

Table 2: 50 point estimates from an unbiased estimator of the mean UXO per grid in sectors A and B

Sector A					Sector B				
4.6	6.4	5.5	4.6	5.5	5.6	5.7	5.7	5.3	5.4
3.7	3.7	7.3	7.3	6.4	5.6	5.7	5.5	5.5	5.6
7.3	3.7	4.6	7.3	5.5	5.4	5.5	5.8	5.2	5.6
5.5	6.4	7.3	5.5	5.5	5.4	5.7	5.4	5.6	5.6
5.5	3.7	7.3	4.6	7.3	5.4	5.3	5.5	5.6	5.7
7.3	6.4	6.4	6.4	3.7	5.3	5.1	5.5	5.3	5.3
5.5	7.3	7.3	4.6	6.4	5.5	5.6	5.5	5.6	5.4
6.4	6.4	4.6	2.8	5.5	5.6	5.3	5.4	5.6	5.7
5.5	5.5	2.8	8.2	9.1	5.4	5.5	5.5	5.4	5.6
3.7	6.4	5.5	5.5	6.4	5.0	5.4	5.6	5.5	5.4

4. Recommendations for New Methodology

Sector definition and remediation decisions are linked and are dependent on many factors. One of these factors is the UXO content. Rather than condensing the UXO information into a single number (mean UXO density in the sector), as is currently done, we think that the sampling results should be presented graphically as a map showing contour lines of constant UXO density with an expressed uncertainty or confidence level. As a result, this should be approached as a spatial density estimation problem.

The UXO are clearly a spatial point process (see Chapt. 8 of [3]). Most available statistical techniques for point processes (including the Hopkins statistic, incidentally) require a random sample of the points. We can randomly sample space that contains the UXO, but we cannot randomly sample the individual UXO.

Because we can randomly sample space, we can sample small areas Δx and record their UXO content. This formulation of the problem opens the door to the application of statistical methods developed for geostatistical data (see, for example, [8, 18] or Chapt. 2 through 5 of [3]). Many of these methods were originally developed for the mining industry to help with ore reserve calculations but have since spread into other areas of earth science.

4.1. Density Estimation

Let us assume that the distribution of "precipitation" of UXO in a sector can be described by some density function. If we select any small area Δx and consider the number of UXO in this area as a random variable U, then the distribution of U can be approximated by the Poisson distribution with the parameter $\mu(\Delta x)N$, where

$$\mu(\Delta x) = \int_{\Delta x} p(x)dx$$

and N is the total number of UXO precipitated on the sector. The realization $u(\Delta x)$ of $U(\Delta x)$ is an estimator of $\mu(\Delta x)N$ with the estimated variance $u(x)$.

Suppose now that we have $u(\Delta x_1), \ldots, u(\Delta x_k)$, where $\Delta x_1, \ldots, \Delta x_k$ are given and can have different shape and size for every location. Then the following problems can be posted:

1. Estimation of $p(x)$: If $p(x)$ is parameterized, then the problem can be solved using the maximum likelihood method (see, for example, [9]). If $p(x)$ is not parameterized but some assumptions on its smoothness can be made, then

we can use moving regression to interpolate/extrapolate the function $p(x)$ (see, for example, [18]).

2. As soon as $p(x)$ is estimated, we can look for an area of a given size V with the greatest UXO density

$$\max_X \int_X \hat{p}(x)dx,$$

such that $\int_X dx = V$. Or we can look for an area of the smallest size that contains a given proportion P of the total UXO

$$\min \int_X dx,$$

such that $\int_X \hat{p}(x)dx = P$. These are problems of estimating a density contour or a boundary and specific methods exist for this problem [6, 16, 17].

3. Optimal choice of Δx_i, $i = 1, \ldots, n$ can be formulated as a design problem for nonlinear regression. This can be used to select sampling locations that minimize a variance criterion of the density estimate.

4. Let X be a given sector of interest,and let Δx_i be selected at random. For simplicity, the Δx_i can be of the same size. Then,

$$K \sum u(x_i) = \frac{\text{area}(X)}{k\text{area}(\Delta x)} \sum_i u(x_i) = \hat{\mu} \tag{5}$$

is an unbiased estimator of $E[U(X)] = N \int_X p(x)dx$ (see Chapt. 10 of [18]).

We can improve $\hat{\mu}$ through

- Selecting weights w_i and introducing $\hat{\mu}^* = K \sum_i w_i u(x_i)$ to reduce the variance of $\hat{\mu}$. Kriging [3, 18] presents one possible approach to selecting the weights.
- Instead of a uniform distribution for selecting x_i, some sequential "optimization" procedure can be considered. For example, new sample areas can be selected in the place of greatest uncertainty about $p(x)$.

4.2. Predicting UXO from Anomalies and Other Information

Because it is much easier to discover anomalies than UXO, it is worth

locations. This relationship may not be simple and will require other information. Today's Global Positioning System (GPS) and Geographical Information System (GIS) technologies enable easy recording of individual anomaly and UXO positions along with some "signature" information from data collection equipment, as well as other attributes of the location such as target type, likely ammunition types used at that target, soil type, topography, vegetation, etc. that all provide information about the anomaly and the probability that it is a UXO. Soft information such as expert opinion on ordnance failure can be incorporated with the use of Bayesian methods (see, for example, [1]). The availability of a good model of UXO probability in terms of anomalies and other information may provide substantial improvements and cost savings in both UXO density estimation and remediation methods.

5. References

[1] George E. P. Box and George C. Tiao. *Bayesian Inference in Statistical Analysis.* Addison-Wesley Pub. Co., 1973. Wiley Classics Library edition published 1992.

[2] William G. Cochran. *Sampling Techniques.* John Wiley & Sons, Inc., second edition, 1963.

[3] Noel A. C. Cressie. *Statistics for Spatial Data.* Wiley Interscience, 1991.

[4] Peter J. Diggle. *Statistic*

[10] QuantiTech. *Grid Statistical Sampling Based Methodology (GridStats), Version 1.2, User's Manual.* Huntsville, Ala., September 30 1995. prepared for U.S. Army Engineer Division, Huntsville, Alabama.

[11] QuantiTech. Oew site statistical sampling based methodology (SiteStats), version 1.0, user's manual. Technical Report 95-R-017, QuantiTech, Huntsville, Ala., May 16 1995. prepared for U.S. Army Engineer Division.

[12] QuantiTech. Ordnance and explosives site statistical sampling based methodology (SiteStats) final report. Technical Report 95-R-011, QuantiTech, Huntsville, Ala., 30 September 1995.

[13] QuantiTech. Development of an audit capability in SiteStats (SiteStats/GridStats, version 2.0, user's manual). Technical Report 96-R-005, QuantiTech, Huntsville, ALa., February 19 1996. prepared for the U.S. Army Engineering and Support Center, Hontsville, Alabama.

[14] B. D. Ripley. *Statistical Inference for Spatial Processes.* Cambridge University Press, 1988.

[15] Statistical Sciences. *S-PLUS Guide to statistical and mathematical analysis, Version 3.4.* StatSci, a division of MathSoft Inc., Seattle, 1996.

[16] A. B. Tsybakov. Multidimensional change-point problems and boundary estimation. In Edward Carlstein, Hans-Georg Muller, and David Siegmund, editors, *Change-point Problems*, volume 23. Insitute of Mathematical Statistics, 1994.

[17] A. B. Tsybakov. On nonparametric estimation of density level sets. *The Annals of Statistics*,

Appendix

A. Minor Errors in *SiteStats/GridStats* Final Report

1. Pages 10ff and Appendix A: The definitions of N_s on page 10 and page A-2 do not appear to be the same.

2. Page 12, E(n) equation: This is actually the average sample number for the binomial case when

$$p = \frac{ln((1-p_0)/(1-p_1))}{ln(p_1/p_0) - ln((1-p_1)/(1-p_0))}.$$

   See page 101 of Wald (1947).

3. Page 12, second paragraph, first sentence: If the given values of α, β, p_0, and p_1 are substituted into E(n), the answer is 53, not 63.

4. Pages 21ff, Sect. 3.1: When the variance of the hypergeometric distribution given near the top of page 22 is modified to have N/A (=0.01065) and D/A (=0.00122), the expression given at the top of page 23 does not follow. There is a factor of N missing on the right-hand side. Therefore, the vertical axis in Figure 3.1-1 is actually the log of s^2/N for this special case.

5. Page 26: There is no need to define Type III and IV errors because these are still simply Type I and II errors in another setting. In statistics, Type I and II errors have a specific meaning.

6. Page 30, Fig. 3.4-1: In other parts of the report, we had α=0.2 and $\beta = 0.1$. Has there been a change and why?

7. Page 35, Table 3.4.2-1: The results for the second column of this table

10. Fig. C-1 describes U_j as "density variation between randomly chosen grids and their nearest sampled neighbor." This is inconsistent with other definitions of U in the report. This is also not possible, because a randomly chosen grid does not have a density estimate unless it is sampled!

11. The next to last equation on p. C-5 is missing some exponents.

12. In the last sentence on p. C-5, should three actually be four?

13. Appendix G: The expression for d(i,j) may be incorrect. It is difficult to tell since nothing has been defined. Why not define each point by (x_i, y_i) and (x_j, y_j), then $d(i,j) =\mid x_i - x_j \mid + \mid y_i - y_j \mid$? Why introduce weights when they are never used?

14. Appendix H: At the end of Sect. H.1, "error" is not defined?

15. Appendix J: For the Poisson distribution, the mean and variance are equal, not the mean and standard deviation.

Appendix

B. Responses to Technical Comments on the Draft Version of This Document

Comments from Mr. Arkie Fanning:

1. (Executive Summary, p. ii, 2nd paragraph): I am concerned that we are leaving the impression that *SiteStats* sectorization was a black box. This is not the case. HNC and the contractor would look at the distribution of UXO by grid to determine post sampling sectorization (as happened at Buckley) regardless of what SiteStats stated.

   Response: The second paragraph has been changed as follows. See also the responses to Comments No. 5 and No. 10.

   > It was found that *SiteStats/GridStats* does adequately guide the sampling so that the UXO density estimator for a sector is unbiased. However, the software's techniques for delineation of homogeneous areas perform less well than visual inspection, which is frequently used to override the software in the overall sectorization methodology. The main problems with the software lie in the criteria used to detect nonhomogeneity and those used to recommend the number of homogeneous subareas.

2. (Executive Summary, p. ii, 3rd paragraph): The first sentence might be interpreted as saying we had bad UXO estimates. Please make it plain that the estimates coming out of the software package were good but the sectorization is in question.

   Response: The third paragraph has been changed as follows.

   > Although the sector UXO density estimator is unbiased regardless of UXO density variation within a sector, its variability increases with increased sector density variation. For this reason, the current practice of visual inspection of individual sampled grid densities (as provided by *SiteStats/GridStats*) is necessary to ensure approximate homogeneity, particularly at sites with medium to high UXO density. Together with *SiteStats/GridStats* override capabilities, this provides a sufficient mechanism for ho-

mogeneous sectorization and thus yields representative UXO density estimates

3. (Acknowledgments, p. v): Please make it clear that QuantiTech developed the software as a company (not Jeff Riggs). Actually, I don't believe I had more than a couple of conversations with Dr. Riggs concerning SiteStats during the entire contract period.

   Response: The change has been made as suggested in the comment.

4. (Page 7, Sect. 2.2.1): The first paragraph is referring to *GridStats* not SiteStats. A complete survey of buried items in a sector is seldom done. There is a complete survey within grids sometimes.

   Response: The entire paragraph has been deleted from this final version of the report.

5. (Page 10, Sect. 2.3): Visual inspection is required as well as a SiteStats conclusion. This is not a black box. It is an aid to the decision makers.

   Response: A new sentence has been added to the end of the third paragraph in Sect. 2.3 to read: "It should be noted that in practice, visual inspection is frequently used to override the software in the overall sectorization methodology."

6. (Page 10, next to last sentence): In the phrase "to of" remove "of."

   Response: The change has been made as suggested in the comment.

7. (General): You did not really address why the SiteStats idea of sectorization fails. If only certain scenarios cause it to give nonsensical answers that still does not mean it is invalid. Is the Hopkins statistic calculated incorrectly? Is the spatial difference stated by the Hopkins incorrect?

   Response: The *SiteStats* idea of sectorization is based on a statistic that QuantiTech calls "Hopkins," which it is not. Rather, it is a statistic that QuantiTech designed and which has not had an independent review until this time. Although its mathematical form somewhat resembles the Hopkins statistic, the "distances" used in its computation are completely different (see the discussion immediately preceding Equation 3 in Sect. 3.1.2). The UXO density difference is a component of the "distance." Because the statistic is based on nearest neighbors, it picks neighbors that minimize UXO difference (see the second full paragraph below Equation 3 in that

same section), thus ignoring the larger differences present in the sample. The statistic tends to downplay UXO density differences. The examples given are for easy illustration, but the same low sensitivity to differences in UXO density persists in all real applications.

Because of its low sensitivity to density variation, it is used in a very "loud" mode inside the *SiteStats* sequential probability ratio tests (see the signal-to-noise discussion in the next to last paragraph in Sect. 3.1.2). This sometimes results in spurious resectorization conclusions by *SiteStats*. Of course, the display of grid results and the availability of overrides provides an opportunity to correct the *SiteStats* conclusion if necessary.

8. (Page 11, Sect. 3): The last sentence is very strong. Are you sure you wish to state it this bluntly?

   Response: The sentence has been reworded, as follows: "It appears that the authors of [14] had difficulty in communicating and documenting many of the statistical concepts that are presented."

9. (General): The fact that the amount of sampling required and the cost of sampling required (accuracy of data-gathering versus data-gathering cost) were the basis for GridStats/SiteStats decision and not just statistical accuracy were not addressed. Is the method unreasonable in this light?

   Response: The *SiteStats/GridStats* function to allocate sample between within-grid items and the number of grids is reasonable, although not optimal. The QuantiTech report cites sequential probability ratio test (SPRT) error rates as criteria, but the SPRT is often stopped when it is inconclusive, thus rendering the nominal error rates meaningless. Nevertheless, there are other criteria (such as the "final stopping rule") that ensure an economical sampling result.

10. (General): *SiteStats/GridStats* is not a decision tool in the classical sense. It provides information to the decision makers but does not require a decision based on that information. The report should state in the executive summary that the amount of UXO found does not in and of itself require a removal action. Particularly, the fact that finding five in a grid on average across a sector does not mean that sector will necessarily be cleaned (although it might). Even more importantly is the fact that finding four in a grid on average across a sector does not mean that the sector will not be cleaned.

Response: In regard to decision making, the following paragraph has been added as the new third paragraph in the Abstract and also as the new first paragraph under Sect. 2.1:

> *SiteStats/GridStats* is not a decision making tool in the classical sense. While it does provide information to decision makers, it does not require a decision based on that information. *SiteStats/GridStats* provides information that is supplemented by visual inspections, land use plans, and risk estimates prior to making any decisions.

In regard to the UXO density and decision making, the following new sentences have been added to the end of the next to last paragraph in Sect. 2.2:

> It should be noted that the amount of UXO found does not in and of itself require a removal action. Finding 5 UXO per grid across a sector does not necessarily infer that the sector will be cleared (although it might be). But even more importantly, finding less than 5 UXO per grid across a sector definitely does not infer that the sector will not be cleared. Decisions regarding removal actions are based on more information than simply the numerical UXO density estimates obtained from *SiteStats/GridStats*.

11. (General): Courtesy comments of Dr. Mog are attached (do not need to address but you may if you wish).

    Response: Dr. Mog's comments are addressed in the following.

Comments from Dr. Robert Mog:

1. (Page iii): Agree that "the techniques for delineation of homogeneous areas perform less well than visual inspection." This was recognized by the original developers and was the underlying reason for including an override option in the software. The ORNL authors should address the mechanism for handling the well-known drawbacks of visual inspection, which include the tendency for the human eye to see patterns in random data and the varying degrees of judgment used in pattern choice by different observers, or even the same observer under different conditions.

Response: It appears that ORNL and Dr. Mog agree that, despite the drawbacks noted in the comment, visual inspection performs better. Because visual inspection is not a specific part of the *SiteStats/GridStats* package, it was not included in the scope of ORNL's evaluation of these software tools. Hence, the drawbacks discussed in the comment cannot be specifically addressed in this report.

2. (Page 4, Sect. 2.2, 2nd paragraph): "GridStats" should be replaced by "SiteStats" in ". . . a statistical test is applied by *GridStats* to determine if the sector is homogeneous or not." Only *SiteStats* uses such a test.

   Response: The change has been made as suggested in the comment.

3. (Page 7, first paragraph): "Homogeneity" should be replaced by "degree of contamination" or some other suitable phrase in "When a statistical conclusion regarding homogeneity has been reached . . . ."

   Response: The change has been made as suggested in the comment.

4. (Page 7, Sect. 2.2.1, first paragraph): Do not understand the first sentence. Thought that the software only requires a complete survey of anomalies in each investigated grid, but not within the entire sector.

   Response: The entire paragraph has been deleted in this final version of the report.

5. (General): Agree with the overall analysis and recommendations of the report. Believe that the inadequacies of the homogeneity criteria used in *SiteStats* are probably overstated from a practical standpoint, but not from a theoretical standpoint. Also, geostatistical techniques, including kriging, spatial covariance, etc., were investigated by the original development team. The reason for their rejection was that they required computational complexities inconsistent with the software development cost and simpler approach. Welcome a revisit of that judgment however, particularly with regard to the homogeneity criteria.

   Response: The agreement stated in the first sentence of the comment is noted. ORNL would also like to add that currently there are commercially available software packages that implement the types of standard geostatistical techniques mentioned in the comment.

Comments from Mr. Bruce Railey:

1. (General): I have reviewed the subject document. For a technical report that is the result of a review of a very technical document and process, it was well written and follows a logical format. I have the following comments for consideration.

   Response: The comment is noted.

2a. (Abstract, 1st paragraph, 1st sentence): It would probably be more appropriate to state that ORNL was "tasked", or some more formal term, since this is not some informal review that was done on the back of an envelope.

   Response: The change has been made as suggested in the comment.

2b. (3rd sentence, and throughout the report): Confirm that the use of the term "metallic" is correct, since I'm sure non-metallic items, although rare, are located and reported during sampling activities. See also "ferrous items" in Sect. 2.2.2, 1st paragraph, 2nd sentence.

   Response: The word "metallic" has been deleted; however, in regard to "ferrous items", it is ORNL's understanding that the logged results of anomaly sampling within a grid, as directed by the software, specifically include such a category.

3. (Abstract, 3rd paragraph). Please clarify the phrase "its usefulness decreases with increased density variation." I don't believe the intent is to say that usefulness for SS/GS decreases for low density sites, but that usefulness decreases for sites where the density of UXO doesn't change drastically across the sector. Also, for low density sites, is the percentage of change the way variation is measured (where a distribution of 0 to 2 UXO access across the site has a much greater variation, while distribution from 5 to 7 is less so), or another measure? Please simplify this confusing discussion.

   Response: The phrase refers to absolute (rather than relative) variation (in which case 0 to 2 is the same as 5 to 7). ORNL would like to clarify that we consider an estimator with high variability less useful than one with low variability (Sect. 3.3 deals with this concept). See also the response to Comment No. 2 from Mr. Arkie Fanning.

4. (Abstract, last paragraph). Explain what is meant by "a UXO density map with known uncertainty." This implies, and is probably true, that the human eye can look at a distribution on a map and make a guess that is about as good as SS/GS.

Response: The phrase "a UXO density map with known uncertainty" has been replaced with "a map showing contour lines of constant UXO density with an expressed uncertainty or confidence level" to add clarification.

5. (Sect. 2.2, 2nd paragraph, 4th sentence): Confirm that "GridStats" shouldn't be "SiteStats," to read "a statistical test is applied by *SiteStats* to determine ..."

   Response: The change has been made as suggested in the comment.

6. (Sect. 2.2.2, 1st paragraph, 2nd sentence): "Ferrous," see comment 2b.

   Response: See the response to Comment 2b.

7. (Sect. 2.2.2, 4th paragraph, 2nd sentence): Remove one of the "that"s, to read "... states that D equals 5 ..."

   Response: The change has been made as suggested in the comment.

8. (Sect. 4.2, 3rd sentence): Change to read "Today's Global Positioning System (GPS) and ... (GIS) technologies enable easy recording ..."

   Response: The change has been made as suggested in the comment.

9. (Sect. 4.2, 3rd sentence): Change to "data collection equipment," or other appropriate term, since we are not limited to magnetometers, although they are mainly used with SS/GS.

   Response: The change to the word "magnetometers" has been made as suggested in the comment.

Comments from Mr. John Loyd:

1. (Executive Summary, p. ii, 2nd paragraph): We should state in the first sentence that the primary use of the tool was to guide sampling and to provide an unbiased sector estimator. HNC rarely (if ever) used the sectorization methodology but rather depended on visual inspection.

   Response: See the responses to Comments No. 1 and No. 10 from Mr. Arkie Fanning.

2. (Executive Summary, p. ii, 3rd paragraph). ORNL needs to define what is meant by the first statement of this paragraph. Does this mean in determining sector density (is the density estimate not as good at high density sites) or in determining whether the sector is homogeneous. Again, HNC

rarely (if ever) depended