To the Graduate Council:

I am submitting herewith a thesis written by Robert N. Stewart entitled “A Geospatial Based Decision Framework for Extending MARSSIM Regulatory Principles into the Subsurface.” I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Geography.

Dr. Shih-Lung Shaw, Major Professor

We have read this dissertation or thesis
and recommend its acceptance:

Dr. Bruce Ralston

Dr. Liem Tran

Dr. Louis Gross

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

A Geospatial Based Decision Framework for Extending MARSSIM Regulatory Principles into the Subsurface

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Robert Nathan Stewart
August 2011

Copyright © 2011 by Robert N. Stewart
All rights reserved.

DEDICATION

This dissertation is dedicated to my family; wife Debra and children Chelsea, Jacob, and Nate who walked on this journey with me the entire way.

ACKNOWLEDGEMENTS

There is an old saying that it takes a village to raise a child. I think it also takes a village to produce a Ph.D. I simply could not have completed this dissertation without the kindness and support of my family, friends, coworkers, and committee members. The biggest appreciation goes to my lovely wife Debra who ran my life for me while I muttered, fussed, and stared into a laptop for months on end at the corner of our dining room table. In an ideal world, spouses would receive an honorary doctoral degree for their efforts. I must also extend a generous thanks to our three children Chelsea (17), Jacob (15), and Nate (14) who recognized when dad needed a break and would pull me away to play outdoors, talk, or play various video games (that I clearly had no ability to win).

I am very fortunate to have so many friends and coworkers that understood the journey I was on and routinely checked in with me, listening to the many small success and failures, and encouraging me to keep moving forward. Thanks especially to Tom Purucker who convinced me to pursue the degree and then gladly served as a fantastic sounding board for ideas and problems. Thanks to Wilson McGinn, a die-hard fan, who listened to every complaint and every success, never dozing off once. Thanks to Dr. George Powers for encouraging me to pursue this degree, for the kind words he wrote in a graduate school recommendation letter, and for his support throughout the process. Thanks to Doug Freels whom I play a great deal of music with for putting band-life on hold until I finished the program and for cheering me on all the way from the sidelines. Thanks to my sister Linda Ward and parents Forrest and Lucy Stewart (all three UT alumni) for their encouragement and unwavering faith that I could succeed. Thanks to my new coworkers and friends Amy Rose and Marie Urban for cheerfully checking in on my daily progress and holding me accountable to the numerous weekly goals I set for myself. A sincere thanks to Budhu Bhaduri, group leader at the Oak Ridge National Laboratory (ORNL), for his very real

support and encouragement to quickly finish the program. I appreciate ORNL permitting use of their computing resources to run many of the models developed herein. I would be remiss not to mention an entire host of other friends and colleagues that made a difference including Brad Lyon, Pastor Steve Damos, Richard Medina, Fred Dolislager, Leslie Galloway, and Jamison Conley, and many other FaceBook friends who kept tabs on me as the journey unfolded.

I would like to acknowledge particular contributions to this work. Thanks to Dr. Keith Eckerman for helping with the radiological dose calculations. Thanks to Robert Coleman for helping with the emulation of the High Performance Germanium Detector portion of the example chapter. Thanks to Dr. Louis Gross, a member of my committee, but also a long time supporter and co-principal investigator for the SADA project who also encouraged me to pursue this degree. Thanks to Dr. Bruce Ralston for his support and encouragement particularly in development of the Check and Cover algorithm. Thanks to Dr. Liem Tran for his sound advice on how to narrow the original broad dissertation proposal down into a more specific and practical work. Last, but by no means least, thanks to my major professor, Dr. Shih-lung Shaw who helped me stay the course with a sharp focus on the specific details and research contributions that would make the dissertation stronger.

ABSTRACT

The Multi-Agency Radiological Site Survey Investigation Manual (MARSSIM) is a regulatory guidance document regarding compliance evaluation of radiologically contaminated soils and buildings (USNRC, 2000). Compliance is determined by comparing radiological measurements to established limits using a combination of hypothesis testing and scanning measurements. Scanning allows investigators to identify localized pockets of contamination missed during sampling and allows investigators to assess radiological exposure at different spatial scales. Scale is important in radiological dose assessment as regulatory limits can vary with the size of the contaminated area and sites are often evaluated at more than one scale (USNRC, 2000). Unfortunately, scanning is not possible in the subsurface and direct application of MARSSIM breaks down.

This dissertation develops a subsurface decision framework called the Geospatial Extension to MARSSIM (GEM) to provide multi-scale subsurface decision support in the absence of scanning technologies. Based on geostatistical simulations of radiological activity, the GEM recasts the decision rule as a multi-scale, geospatial decision rule called the *regulatory limit rule* (RLR). The RLR requires simultaneous compliance with all scales and depths of interest at every location throughout the site. The RLR is accompanied by a compliance test called the *stochastic conceptual site model* (SCSM). For those sites that fail compliance, a remedial design strategy is developed called the Multi-scale Remedial Design Model (MrDM) that spatially indicates volumes requiring remedial action. The MrDM is accompanied by a sample design strategy known as the Multi-scale Remedial Sample Design Model (MrsDM) that refines this remedial action volume through careful placement of new sample locations. Finally, a new sample design called “check and cover” is presented that can support early sampling efforts by directly using prior knowledge about where contamination may exist.

This dissertation demonstrates how these tools are used within an environmental investigation and situates the GEM within existing regulatory methods with an emphasis on the Environmental Protection Agency's *Triad* method which recognizes and encourages the use of advanced decision methods. The GEM is implemented within the Spatial Analysis and Decision Assistance (SADA) software and applied to a hypothetical radiologically contaminated site.

TABLE OF CONTENTS

Chapter Page

List of Acronyms xvi
Chapter 1: Introduction 1
Objectives 1
Background 1
Research Need 3
Research Objective 4
Research Method 6
Organization of the Dissertation 8
CHAPTER 2: KEY CONCEPTS 10
Introduction 10
Phases of Environmental Investigation 11
MARSSIM 12
Geostatistical Simulation 20
Triad 31
Closely Related Work 34
Summary 42
Chapter 3: Theoretical Derivation of the GEM Framework 44
Introduction 44
The Regulatory Limit Rule 46
The Stochastic Conceptual Site Model Test 53
The Multi-scale Remedial Design Model 58
Multi-scale Remedial Sample Design Model 75
Summary 81
Chapter 4: Implementing the GEM in SADA 83
Introduction 83
The GEM Prototype Within the SADA Modeling Environment 83

Remedial and Exposure Units in SADA 89
SADA Workflow 94
GEM Class Structures 94
The GEM Model Workflow 104
Summary 107
Chapter 5 Check and Cover Sampling Strategy 108
Introduction 108
Method 109
Implementation in SADA 117
Creating a Prior CSM 117
Check and Cover Design in the SADA GUI 118
Check and Cover Workflow 121
Conclusion 123
CHAPTER 6: Example Application 124
Establishing the Synthetic Data 124
Simulating the Sampling Process 125
Defining the Exposure Scenario 126
Historical Site Assessment 126
Scoping Phase 127
Characterization Phase 131
Remedial Phase 142
Compliance Phase 146
Performance Issues 147
Summary 151
Chapter 7: Conclusions 152
Summary of the Study 152
Future Research 155
Summary 160
References 162
Appendix 173

A.1 Class SADA.clsExposureUnitSpecifications 174
A.2 Class SADA.colExposureUnitSpecifications 176
A.3 SADA.clsExposureUnit 180
A.4 Class SADA.clsGEMPhysicalStructure 188
A.5 Class SADA.clsGEMStructure 189
Table A.6 Class SADA.clsPreviouslyConstructedGEMStructures: Properties 194
Table A.6 Class SADA.clsPreviouslyConstructedGEMStructures: Methods 195
Vita 196

LIST OF TABLES

Table Page

Table 1.1 Summary of the GEM development activities. 9
Table 2.1 Survey Unit Classifications.....16
Table 3.1 Relationship between Decision Components.....45
Table 3.2 Input requirements for RLR/SCSM 58
Table 4.1 Draw and Area of Concern Map Steps.....88
Table 6.1 External Exposure Limits for 9 Exposure Unit Geometry/Volumes ...126

LIST OF FIGURES

Figure Page

Figure 2.1 Concentration values are propagated through models to produce dose or risk estimates. 13
Figure 2.2 Dose or risk values are reverse propagated through the model to produce corresponding concentration estimates. 13
Figure 2.3 Local elevation (emphasized in red) within a larger exposure unit. 15
Figure 2.4 Two random variable distributions from the set of 168. 22
Figure 2.5. Three realizations of the random function. 25
Figure 2.6 A simple CSM drawn in graphical software (USEPA, 2008, p. 1). 33
Figure 2.7 A variety of information is provided in this CSM, including quantitatively derived subsurface conditions (USEPA, 2008, p. 2). 33
Figure 2.8 Relationship of the GEM to MARSSIM, Triad, and Geostatistical Simulation 43
Figure 3.1 The GEM workflow. 46
Figure 3.2 Two example external exposure units positioned in the subsurface. 47
Figure 3.3. Exposure unit E′ compliance as a function of spatial location. E′ in green passes compliance while E′ in red fails. 48
Figure 3.4 The GEM spatial resolution grid. 49
Figure 3.5 A moving window view of exposure unit compliance. 51
Figure 3.6 Four of the eighty nine exposure unit instantiations are shown. 51
Figure 3.7. Selected instances of two exposure unit classes are shown. 52
Figure 3.8. Site specific calculation of the number of failing exposure units within each class. 57
Figure 3.9 Volume of grid points locations where the corresponding exposure unit instance demonstrated a probability greater than alpha = 0.1. 57
Figure 3.10 Scenario where no single RU improves compliance probability. 62
Figure 3.11 Two overlapping exposure unit instances with compliance failures. 65

Figure 3.12 Demonstration of Algorithm 3.2. Remedial units slated for remedial action are highlighted in red. 66
Figure 3.13 Global remedial design is reduced when concurrently considering the remedial designs of local exposure unit instances. 67
Figure 3.14 The iterative construction of MrDM. 70
Figure 3.15 MrsDM Evaluation of proposed locations. 78
Figure 4.1 Conceptual view of SADA's ISPR style interface, used with permission from Stewart and Purucker (2011). 84
Figure 4.2 SADA's ISPR GUI. 85
Figure 4.3 SCSM and MrDM implemented within the ISPR GUI. 86
Figure 4.4. MrsDM Sample Design Parameters. 89
Figure 4.5 Three dimensional grid systems in SADA. 90
Figure 4.6 Exposure unit specification as a function of remedial unit specification. 91
Figure 4.7 Exposure unit specification within the GEM grid system. 93
Figure 4.8 EU Specification File format (a) and example (b). 93
Figure 4.9 SADA's Macro-Level Workflow 94
Figure 4.10 Class diagram for the clsGEMStructure. 96
Figure 4.11 Class diagram for the *clsGEMPhysicalStructure* 97
Figure 4.12 Class diagram for *clsExposureUnitSpecifications*. GEM operations refer to methods enclosed within the red box in Figure 4.10. 99
Figure 4.13 Class diagram for *clsExposureUnitSpecifications*. 100
Figure 4.14 Class diagram for *clsExposureUnitSpecification*. 102
Figure 4.15. Class diagram for *clsPreviouslyDevelopedGEMs* 104
Figure 4.16 The SCSM workflow 105
Figure 4.17 The MrDM workflow. 106
Figure 4.18 The MrsDM workflow. 106
Figure 5.1 Scale of concern for contamination. 110
Figure 5.2 A three dimensional conceptual site model indicating contamination concerns. 111

Figure 5.3 Prior CSM confidence scale. 113
Figure 5.4 Map reliance factor effect on scale of concern values. 114
Figure 5.5 Map Reliance factor effect on initial guess. 115
Figure 5.6. Check and Cover: sample size versus design metric. 116
Figure 5.7 Users define the 3D raster model by specifying a grid and layer resolution. 118
Figure 5.8. Users “paint” numerical values into the model. 119
Figure 5.9 Two access points for check and cover. 120
Figure 5.10 Check and Cover parameters within SADA GUI. 120
Figure 5.11 Example check and cover design. 121
Figure 5.12 Check and Cover Workflow 122
Figure 6.1 A 3D rendering of the “true” Cs-137 values at Cesium Site. 124
Figure 6.2 Map of Cesium Site. 128
Figure 6.4 shows vertical profiles for two of the most contaminated cores (using SADA’s vertical profile tool). 128
Figure 6.3. Scoping survey results. 129
Figure 6.4 Vertical profiles for scoping survey results. 130
Figure 6.5 Three dimensional view of scoping results. 130
Figure 6.6 Prior contamination concern model for Cs-137 contamination shown in the SADA 3D viewer (a) for levels 2 and higher and (b) layer by layer in the SADA 2D viewer. 132
Figure 6.7 Grid spacing versus sample size and p-median values. 132
Figure 6.8 Number of samples (for grid design) and p-median metric. 133
Figure 6.9 Check and Cover places 9 core locations (red triangles) based on the prior concern model. 134
Figure 6.10 Technology selection by corehole. 135
Figure 6.11 Lab and HPGe measurements in 3D viewer. 136
Figure 6.12 Unbounded areas (red dashed boundaries) are supplemented by additional judgmental locations (blue triangles). 137
Figure 6.13 Additional HPGe characterization samples. 138

Figure 6.14 4/50 SIS simulations. 139
Figure 6.15. Simulation average values above 32pCi/g. 139
Figure 6.16 Cesium site exposure unit specification file. 140
Figure 6.17 SCSM Test parameters. 141
Figure 6.18 Number of exposure unit failures by exposure unit specification and by depth interval (S1 = surface layer). 141
Figure 6.19 Location of EU instance compliance failures. 142
Figure 6.20 a) Baseline design b) MrDM design c) MrDM improvement regions. 143
Figure 6.21 Ten Adaptive Fill locations from SADA. 144
Figure 6.22 MrsDM parameters 144
Figure 6.23 MrsDM produces winning corehole locations and illuminates those volumes forecast to be removed from MrDM design (top layer shown only). 145
Figure 6.24 Candidate corehole performance in reducing MrDM design. 145
Figure 6.25 MrDM before (a) and after (b) MrsDM Sample Design. 146
Figure 6.27 Algorithm duration by stage. 148
Figure 6.28 Parallelization in MrDM Algorithm 150

List of Acronyms

DCGL	Derived Concentration Guideline Level
$DCGL_{EMC}$	DCGL for local exposure scenarios
$DCGL_{w}$	DCGL for site wide exposure scenarios
EPA	Environmental Protection Agency
NRC	Nuclear Regulatory Commission
GEM	Geospatial Extension to MARSSIM
GUI	Graphical User Interface
MARSSIM	Multi-Agency Radiological Site Survey Investigation Manual
RLR	Regulator Limit Rule
PRV	Post-remedial Replacement Value
MrsDM	Multi-scale Remedial Sample Design Model
MrDM	Multi-scale Remedial Design Model
SADA	Spatial Analysis and Decision Assistance
SIS	Sequential Indicator Simulation

Chapter 1: Introduction

Objectives

The Multi-Agency Radiological Site Survey Investigation Manual (MARSSIM) is a regulatory guidance document regarding compliance evaluation of radiologically contaminated soils and buildings (USNRC, 2000). MARSSIM is a comprehensive decision framework for surface contamination but stops short of formalizing a process for the subsurface. In this dissertation, a decision framework called the Geospatial Extension to MARSSIM (GEM) is developed to address this need. The goal of the GEM is not to establish full regulatory policy on the matter, but rather to provide the technical foundation upon which future subsurface guidance may be built. To meet this goal, this dissertation develops the GEM as a numerically explicit decision framework that draws upon, extends, and situates advances in geostatistical decision support within the context of radiological regulatory compliance.

Background

Federal guidance documents provide and interpret environmental regulation for federal agencies and the public (USOMB, 2007, pp. 1,19). These documents often translate policy within a scientific context, promoting responsible and consistent methods in responding to environmental pollution. The *Data Quality Objectives* (DQO) process is a cornerstone of regulatory guidance for investigating contaminated lands. Guidance is provided on setting project objectives, specifying decision errors, and identifying information needs, including type, quantity, and quality of data (USEPA, 2006a). First appearing in the 1980s (USEPA, 1980), the DQO process has motivated a number of follow up guidance documents (e.g., USEPA, 1987a, 1987b, 1994b, 1989a, 1989b, 1992a, 1992b, 1994a, 1997, 2000a, 2001c, 2002a, 2002b, 2002c, 2003b, 2006b, 2006c, 2006d; and USNRC, 2000) and has shaped the landscape of environmental investigations for the last 30 years.

During this time, the environmental community has seen the emergence of advanced sampling and remote sensing technologies, statistical and mathematical models, and decision support systems that deal with various aspects of site investigation. Members of the regulatory community, particularly at the Environmental Protection Agency (EPA), have called for a substantial update of the DQO process that integrates these new and powerful approaches into a *second generation* DQO process (Crumbling, 2002). Unfortunately, the response to such calls for revision has been slow, primarily because the implications of change are difficult to ascertain (Crumbling, 2004).

While no such sweeping update has occurred, the EPA has articulated the *Triad* model (Crumbling, 2001a). Triad represents a concerted effort by experts from the public and private sector to create a modern approach that lays the groundwork for a second generation DQO process (Crumbling, 2002). Triad methodology spans the project life cycle, providing continuity between management practices, scientific methods, and technological advances that emphasizes the quality of the decision. At the center of Triad is the *conceptual site model* (CSM). A CSM is a representation of site knowledge that evolves over the course of investigation. CSMs communicate knowledge about a variety of issues, including geology, exposure pathways, spatial distribution of contamination, and transport mechanisms (Crumbling, 2001a; USEPA, 1992b). Under Triad, the CSM drives data collection by identifying knowledge gaps. The CSM is reciprocally informed and evolved by the outcome of those data (Crumbling, 2001a). Triad recognizes the value of accurate laboratory analysis but also calls for the inclusion of screening and field detection methods that are typically faster and less expensive to collect (Crumbling, 2004). The combination of speed and reduced costs can result in a greater sampling density and better support for CSM evolution.

CSMs that describe spatial processes, such as the pattern of contamination (USEPA, 1992b), may rely on some form of geospatial modeling (Isaaks and Srivastava, 1989; Goovaerts, 1997). Regulatory guidance has positively commented on the use of geostatistics, in particular, to support decision processes (e.g., EPA 1987b, 1989a, 1992b; and NRC, 2000). Three EPA guidance documents (USEPA 1987b, 1989a, and1992b) substantially discuss the use of the geostatistical estimator *kriging* (Isaaks and Srivastava, 1989). Now over two decades old, these documents represent evolution of the DQO process and the computational resources available to investigators at that time. Since those issuances, the literature has evolved by developing advanced and often computationally demanding geostatistical decision support tools (e.g., Ahmed et al., 2008; Brus et al., 1997; D'Or, 2005; Demougeot-Renard et al., 2004; England et al., 1992; Goovaerts, 1997, 1999, 2001; Pilger et al., 2001; Savelieva et al., 2005; and Saito et al., 2003). Tools such as geostatistical *simulation* provide a more rigorous assessment of uncertainty than kriging and greater capabilities in characterizing spatial processes. Key advances include uncertainty assessment across different spatial scales and methods for integrating various kinds of information (e.g., field and laboratory data) under a single model (Goovaerts, 1997). These abilities represent a substantial opportunity for investigators to develop, evolve, and use the CSM as envisioned under Triad.

Research Need

Given these recent advances, it may be time to identify opportunities within regulatory guidance where Triad principles and geostatistical advances can be drawn together into a regulatory process. This dissertation engages with this idea by re-examining how MARSSIM principles may be extended into the subsurface. MARSSIM focuses on radiological contamination of surface soils and building surfaces and provides a uniform approach for evaluating contamination at those

locations. The specific objective is to determine whether a particular section of land or building is safe for certain uses (USNRC, 2000).

In MARSSIM, compliance is required at two spatial scales: the entire survey unit[1] and smaller *local* areas identified within the unit. In order to be in compliance, radiological contamination may not exceed limits established at either scale. For the entire survey unit case, a classical hypothesis test is typically applied. At the local scale, investigators must demonstrate that no smaller areas of contamination within the unit, referred to as *hotspots*, are present that could impact public health. Scanning technologies (e.g., radiological detectors) that can exhaustively detect radiological activity at the surface are used to assess this local scale requirement.

Unfortunately, radiological contamination can also migrate to the subsurface, where such scanning is impossible and the approach for local compliance breaks down. In the absence of exhaustive measurement devices, geostatistical modeling and Triad methodology present an opportunity to reformulate core MARSSIM principles within a fully spatial context.

Research Objective

This dissertation develops the GEM framework that extends MARSSIM principles into the subsurface, by integrating geostatistical decision support tools with elements of Triad. The challenge is to develop the GEM as a rigorous and cohesive workflow within a fully geospatial context that resonates with the MARSSIM community, establishes decision processes analogous to those at the surface, and implements methods that are operable within the standard phases of environmental investigation (USNRC, 2000).

[1] A survey unit is a defined section of land for which a decision will be made (USNRC, 2000).

There are three major stages in developing the GEM. First, create a geospatial decision rule focused on protecting public health. Such a rule, like MARSSIM, would require compliance evaluation at different spatial scales (e.g., survey unit and local). Second, develop a path for demonstrating compliance with this decision rule. Third, demonstrate how the GEM framework operates across established phases of environmental investigation and provides direct support to key activities, including sampling and remedial design strategies. These stages are organized as follows:

Part 1: Develop a decision rule

How can a geospatial decision rule be developed that

a. accounts for limits at different scales,
b. places a premium on uncertainty about exceeding those limits, and
c. is analogous to compliance in MARSSIM?

Part II: Demonstrate compliance with the decision rule

How can geostatistical modeling and decision support tools be developed and organized under Triad principles to develop a compliance test with the geospatial decision rule in Part 1?

Part III: Strategies for moving failing sites into compliance.

For those sites that fail compliance, what decisions or actions (such as remediation) might be taken to efficiently move those sites into compliance?

Part IV: Integration with existing environmental investigation phases

How can the approaches in Parts I-III be accomplished during the course of an environmental investigation? Can these methods provide support to key activities in the investigation, such as sample design and remedial action plans?

Research Method

This dissertation responds to these questions by developing the GEM, which establishes a decision rule, a path for evaluating compliance with that rule, and key tools based on both that create viable options for investigators.

In Part I, a *Regulatory Limit Rule* (RLR) is developed to establish how investigators evaluate compliance for the entire survey unit (global scale) as well as hotspots (local scale) of any size and shape. Compliance is therefore evaluated at multiple scales of interest. Compliance is accomplished by demonstrating, for each scale, that the chance of contaminant levels exceeding safe limits is less than a maximum probability value. The RLR requires a geospatial model that is capable of estimating probability values at any spatial scale.

In Part II, the dissertation turns to geostatistical simulation as the basis for testing the RLR in Part I. Geostatistical simulation permits calculation of the probability that contaminant levels are exceeding a regulatory limit over any spatial scale. Simulation also permits the inclusion of various kinds of data into the estimation (Goovaerts, 1997), including both laboratory and field sampling methods emphasized by Triad (Crumbling, 2001a). The *Stochastic Conceptual Site Model* (SCSM) test post-processes geostatistical simulations to determine the probability of exceeding the RLR for any scale.

In Part III, a remedial design algorithm called the *Multi-scale Remedial Design Model* (MrDM) is developed. MrDM estimates the minimum location and size of the contaminated soil volume to remediate that would bring the site into compliance. A companion to MrDM is the *Multi-Scale Remedial Sample Design Model* (MrsDM), which seeks to place additional samples in locations that may further refine the MrDM outcome.

In Part IV, there are five phases of an investigation under MARSSIM: historical site assessment, scoping, characterization, remediation, and compliance (USNRC, 2000). For the scoping phase, an adaptation of the P-median algorithm (Miller and Shaw, 2001; Ostresh, 1977) called *Check and Cover* (C&C) is used to create an initial sampling design that takes advantage of expert knowledge regarding the location of contamination in early characterization efforts. The characterization phase is also where the SCSM test is applied to determine if remedial actions are necessary. If so, then in the remedial phase, both the MrDM and MrsDM are used to inform remedial action decisions. Finally, the compliance phase sees a reapplication of the SCSM test following completion of remedial actions.

To demonstrate the GEM and Check and Cover design, the Spatial Analysis and Decision Assistance (SADA) freeware package developed at the University of Tennessee's Institute for Environmental Modeling will be extended to include these new approaches. SADA was developed with funding from three federal regulatory agencies (Nuclear Regulatory Commission, Environmental Protection Agency, and the Department of Energy) in collaboration with other universities, national laboratories, private sector companies, and individual consultants. The software provides a tool that makes a direct, practical connection between data analysis, modeling, and decision-making within a spatial context (Stewart et al., 2009) and is commonly recognized as a Triad tool (www.triadcentral.org). Multi-disciplinary tools include geographic information systems, sample design, statistics, data management, two- and three-dimensional visualization, spatial modeling, uncertainty analysis, human health and ecological risk assessment, remedial design, and cost/benefit analysis.

SADA capabilities, as an environmental management computational toolkit (Holland et al. 2003), have led to its mention within regulatory frameworks, and appearances in the literature have continued to grow. Applications include

underground storage tanks, landfill disposal sites (Butt et al., 2008) and Brownfield sites (USEPA, 2005a), Triad applications (USEPA, 2003, 2005), US Army Corp of Engineers sites (Puckett et al., 2004), ecological risk (Carlon et al., 2008; Purucker et al., 2007), human health risk (Butt et al., 2008), and others (USEPA/state of Pennsylvania, 2003). Others include investigations of microbial community structure (Franklin and Mills, 2003), multi-criteria decision analyses (Linkov et al., 2004), delineating the boundaries of soil polygons for terrain analysis (Sunila et al., 2004), examination of interactions between habitat and contamination on ecological dose (Purucker et al., 2007), frameworks for soil remediation (Norman et al.; 2008; Rügner et al., 2006), hotspot delineation (Sinha et al., 2007), and level of laboratory analytical support necessary to support field-level data collection (Puckett and Shaw, 2005). SADA is well poised to serve as a computational platform for adding new geospatial decision methods. In particular, aspects of both Triad and MARSSIM are already present within the code and are well positioned for implementation of the methods developed here. Table 1.1 summarizes the major activities in developing the GEM.

Organization of the Dissertation

Chapter 2 presents key concepts important to this work. These include a brief overview of risk assessment and necessary details concerning MARSSIM, Triad, and geostatistical simulation. An overview of relevant developments in the literature is also provided.

The research objectives discussed above (Part I-IV) provide a natural flow to the remainder of this work. Chapter 3 provides a theoretical derivation of the GEM framework, including the RLR, SCSM, MrDM, and MrsDM. Chapter 4 discusses implementation of the GEM in SADA. Chapter 5 presents the Check and Cover design including both derivation and implementation in SADA. Chapter 6

Table 1.1 Summary of the GEM development activities.

Developments	Acronym	Description
Regulatory Limit Rule	RLR	Specifies the subsurface decision rule across multiple scales
Stochastic Conceptual Site Model Test	SCSM	Test for compliance with RLR
Multi-scale Remedial Design Model	MrDM	Considers multiple decision scales at once in designing remedial plans
Multi-scale Remedial Sample Design Model	MrsDM	Samples locations that might reduce the size of the MrDM remedial design
Check and Cover	C&C	Locates samples using prior knowledge
GEM Implementation	NA	Extends SADA capabilities to include RLR, SCSM, MrDM, and MrsDM

introduces a hypothetical contamination scenario called “Cesium Site” where these methods are applied under the phases of investigation. Finally, Chapter 7 evaluates the subsurface framework, discusses its limitations, and identifies where specific research and development activities are needed.

CHAPTER 2: KEY CONCEPTS

Introduction

There are several key concepts that set both context and methodology in formally deriving the GEM and the additional Check and Cover design. Those concepts are presented here and include key details about environmental investigation, MARSSIM, geostatistical simulation, decision support, and Triad. In addition, an enumeration of closely related work is provided.

The discussion begins with the major phases of radiological investigation under which MARSSIM operates. Relevant aspects of MARSSIM are then presented, including why direct application to the subsurface breaks down. This point of failure motivates the derivation of the GEM and the move toward geostatistical simulation as a mechanism for assessing subsurface compliance. A brief overview of geostatistical simulation and associated decision making is presented with a focus on those concepts critical to the GEM methodology. Additionally, it is common for guidance authors to show the connection between proposed and existing guidance. As a technical approach to potentially new regulatory guidance, the discussion here will follow suit and situate the GEM with respect to both Triad and MARSSIM. Finally, as evidenced in the following discussion, while the literature is abundant with application of advanced geostatistical methods in environmental investigation, regulatory guidance has been slow to respond. In this larger picture, the GEM represents an opportunity for visiting the issue of guidance revision and mainstreaming within the regulatory life of those advanced and formal geospatial decision frameworks already apparent in the literature.

Phases of Environmental Investigation

The investigation life cycle for a radiologically contaminated site is divided into five separate phases: historical site assessment, scoping, characterization, remediation, and compliance (USNRC, 2000, p. 2-15). Each phase has particular objectives and builds on previous phases in characterizing and responding to public health risks. A brief overview drawn from MARSSIM is provided here.

Historical Site Assessment (HSA)

In this first stage, investigators collect all relevant information regarding the potential study area. This is usually a desk study paired with site visits to characterize operating history, identify potential sources of contamination, and estimate the likelihood of contaminant migration.

Scoping Phase

This phase provides site-specific information based on a limited number of sample measurements. Often, the number and location of samples is based on expert judgment. These results, along with knowledge from the HSA, are used to determine if characterization will be necessary.

Characterization Phase

In the characterization phase, investigators estimate the nature and extent of contamination. This can be a highly spatial exercise with multiple objectives in play. Characterization may begin initially as an exploratory refinement on the scoping survey results but should mature into a result usable in evaluation of remedial alternatives and technologies.

Remedial Phase

In this phase, a remedial action plan is developed and executed. Additional measurements can be collected as the remedial process is underway to inform the remedial process as it unfolds. At the end of this phase, a site should be well prepared to meet compliance.

Compliance Phase

In this phase, regulators evaluate whether the site is safe for release under its intended use. In MARSSIM, an independent final status survey and associated decision rule are applied to support this judgment.

As with MARSSIM, information gained in these phases is directly used in the GEM framework. Specifically, Chapter 6 demonstrates the GEM within these phases using a hypothetical site. The discussion now continues with the MARSSIM decision rule and why direct application fails in the subsurface.

MARSSIM

Human exposure to radioactively polluted soil creates the potential for harmful, ionizing radiation to enter the body by various pathways, such as ingestion, inhalation, dermal contact, and external radiation (Cember and Johnson, 2009; Byrd and Cothern, 2000; USEPA, 1997; Eckerman and Ryman, 1993). Dose refers to the amount absorbed by the body during this exposure (Byrd and Cothern, 2000). Excessive dose may lead to cancer, and regulatory agencies have established limits that are protective of public health. For example, the USNRC imposes a 25 mrem/year[2] limit in Title 10 of the USNRC code of Regulations (USNRC, 2009).

[2] A milli-rem is one millionth of a *Roentgen Man Equivalent* (rem) which is a measurement unit for dose. (Cember and Johnson, 2009)

Risk assessors are therefore concerned with determining the dose associated with exposure to contaminated soils. The amount of dose received by the body cannot be measured directly, but must be modeled (USNRC, 2000, p. 2-2). This is a broad and a highly complex field of study requiring scientific methods that consider numerous factors, including type of radionuclide, duration of exposure, target pathways, and even specific organs that may be vulnerable (Eckerman and Ryman, 1993). In the simplified view of the process illustrated in Figure 2.1, concentration levels are processed by an exposure/dose model, producing an estimate of dose to the body[3] (Eckerman and Ryman, 1993).

The opposite is also possible. Given a dose value, assessors can invert the exposure model to produce a corresponding concentration limit for the soil (NRC, 2000, p. 2-2), as in Figure 2.2. In MARSSIM, the Derived Concentration

Figure 2.1 Concentration values are propagated through models to produce dose or risk estimates.

Figure 2.2 Dose or risk values are reverse propagated through the model to produce corresponding concentration estimates.

[3] This discussion provides only a very broad view of exposure and dose assessment, both of which are large and complex areas of scientific activity. For interested readers, the United States Environmental Protection Agency has produced accessible introductions, including USEPA (1989a), USEPA (1992a), and USEPA (1997). Other valuable introductions include Byrd and Cothern (2000), which presents environmental risk analysis within a larger risk context and Cember and Johnson (2009), which provides an introduction to health physics and radiological risk assessment.

Key Concept: MARSSIM Decision Criteria

Soil concentration limits (DCGLs) limit the amount of dose that exposed individuals can receive and form the foundation of the MARSSIM compliance.

Guideline Level (DCGL) corresponds to a concentration value associated with a maximum dose limit.

Exposure usually occurs over a particular spatial domain[4] called the exposure unit (USEPA, 2002a, p. 1). Exposure units can vary in size and shape and depend on how the property will be used. For example, residential properties are often associated with small spatial areas, around 1/8th acre (USEPA, 1989a, p. 6-28). In contrast, an agricultural scenario might consider exposures over a much larger area.

Investigators must demonstrate that the exposure unit *average concentration* does not exceed the DCGL (NRC, 2000, p. 8-6). In the interest of public health, investigators will conservatively estimate the average concentration. For example, an upper confidence limit on the average may be compared to the concentration limit (USEPA, 1989a, p. 6-19).

By using the average concentration, an assumption is made that contamination is relatively uniform throughout the site (USNRC, 2000, p. 2-3) and that receptors will not preferentially engage with any portion of the exposure unit. This may not be the case. Contamination is often heterogeneously distributed and human

[4] It is worthwhile to note that not all exposures occur over a spatial domain. For example, consider the scenario where contamination filters from the soil into groundwater and is then ingested at downstream wells or public intake locations. In this situation, investigators can reverse calculate acceptable soil concentration limits or total contaminant mass that are protective of ground water. This results in a set of soil-based decision criteria that the contaminated site can be assessed against. The methods presented under GEM may be well suited for this downstream scenario; however this is outside the scope of this work.

0.1	0.1	0.2	0.1
0.1	0.2	0.1	0.1
0.2	600	0.1	0.1
0.1	0.2	0.2	0.1
0.1	0.1	0.1	0.1

Figure 2.3 Local elevation (emphasized in red) within a larger exposure unit.

behavior is difficult to predict. Regulatory agencies are therefore interested in addressing potential "hotspots" where locally elevated radiation levels are too high for even small exposures (USEPA, 1989a, p. 5-22) or where humans might preferentially engage (USEPA, 1989a, p. 6-28). Figure 2.3 illustrates how such a scenario might appear, using a simplified exposure unit and a hypothetical DCGL of 50pCi/g.[5] The average concentration for this exposure unit is 30.1pCi/g and is well below the limit of 50pCi/g; however, an area near the center exceeds 600pCi/g and may pose a health hazard, particularly if human activity is preferentially located in that area.

Hence, the concern over local hotspots really represents a concern about smaller exposure scenarios that could happen within the larger exposure unit. MARSSIM responds to this concern by defining two DCGL values at two different exposure scales.

The $DCGL_W$ [6] is based on exposure to the entire area. For exposure to small areas of elevated activity, a separate limit known as the $DCGL_{EMC}$[7] is derived, potentially under different exposure assumptions (USNRC, 2000, p. 2-3). The

[5] The Curie is a unit of radioactive decay defined as 3.7×10^{10} decays per second. A pico-Curie (pCi) is one trillionth of a Curie (pCi).
[6] Originally the "W" indicated that a statistical test called the Wilcoxon Rank Sum test would be used to test exceedance of this DCGL. MARSSIM, however, permits other tests, such as Sign test, but continues to use the "w" notation (USNRC, 2000 p.2-3).
[7] EMC stands for "Elevated Measurement Comparison", referring to the method for evaluating compliance with this DCGL (USNRC, 2000 p.2-3).

size of the local area is determined by the regulatory agency, and no explicit direction is provided on how this is done (USNRC, 2000, p. 5-38). For the purposes of this dissertation, these facts point to the presence of multi-scale (only two in this case) decision making in MARSSIM. As the reader will see in later sections, this core principle of MARSSIM is preserved and indeed expanded in the GEM framework. In addition, MARSSIM guidance does not include direction on how DCGLs are calculated. Rather DCGLs are input to that process, and the same will hold true for the GEM.

Under MARSSIM, once DCGLs are available, the site is divided into a series of *survey units* (USNRC, 2000, p. 5-22). A survey unit is a section of land with specified size and shape for which a decision will be made regarding compliance with DCGLs (USNRC, 2000, p. 2-4). Survey units are chosen and classified by how likely it is that unacceptable contamination exists within them. There are three classes to choose from (USNRC, 2000, p. 4-12), as seen in Table 2.1.

Classification decisions and survey unit selections are based on expert evaluation of the site's operating history and previous survey results. Particular

Table 2.1 Survey Unit Classifications

Class	Description
1	Areas with potential (or known) contaminated levels higher than the $DCGL_W$
2	Areas with potential (or known) contamination present but unlikely to exceed the $DCGL_W$
3	Areas with potential (or known) contamination levels expected to be no more than a small fraction of the $DCGL_W$

boundaries are therefore judgmental, potentially subjective, and likely vulnerable to the modifiable area unit problem (MAUP)[8] (O'Sullivan and Unwin, 2003). For this reason, the GEM does not require a division of the site into survey units. Rather, the decision is applied to the entire survey area, where the geospatial model delineates the likelihood of contamination in a more explicit manner.[9] These survey units can then serve as exposure units for which a decision rule involving both the $DCGL_W$ and $DCGL_{EMC}$ is used to assess compliance.

Key Concept: MARSSIM Decision Criteria Specified at Two Spatial Scales

- Survey unit average concentration is limited by the $DCGL_W$.
- Local concentrations within the survey unit are limited by the $DCGL_{EMC}$.

The first step is to determine whether the average concentration exceeds the $DCGL_W$ using hypothesis testing. This is accomplished with a final status survey and corresponding statistical test. In particular, investigators establish a null hypothesis that the $DCGL_W$ is exceeded (USNRC, 2000, p. 2-26). Two traditional tests that assume data are independent and identically distributed are emphasized by MARSSIM (USNRC, 2000, p. 2-27): Wilcoxon Rank Sum test (if the radionuclide naturally occurs in background) or Sign test (if the radionuclide is not present in background).[10] If the data fail to reject the null hypothesis, the site is out of compliance (USNRC, 2000, 8-11, 8-17). Before any sampling occurs, the test is selected, permitting investigators to develop a sample design

---

[8] The MAUP refers to the fact that if spatial units were arranged in a different way, a different result might arise.

[9] If so desired, one could continue the practice of survey units and then apply GEM within separate units.

[10] Under the sign test, measured values are subtracted from the DCGL. A large number of positive differences (when compared to sign test critical values) indicates failure to comply (USNRC, 2000, p. 8-12). For the Wilcoxon Rank Sum test, reference area measurements are first increased by the DCGL value, combined with site values, and then ranked. The ranks from the adjusted reference area values are summed and compared to critical test values. If the sum exceeds the critical value, the site fails compliance (USNRC, 2000, p. 8-18). Both the Sign and Wilcoxon Rank Sum tests are actually tests for the median and not the average (Miller et al., 1990). MARSSIM looks past this by arguing that used in this fashion, a test for the median is a good approximation for a test to the mean (USNRC, 2000, p. 2-28).

that supports that test. Given an assumption about what the variance is likely to be, it is possible to estimate the number of samples required to conduct the statistical test at desired Type I and Type II error rates (USNRC, 2000, pp. 5-28, 5-33).

For Class 3 survey units, where hotspots presumably do not exist, these samples are distributed randomly across the site (USNRC, 2000, p. 2-31). For Class 1 and Class 2 sites, where hotspots may exist, the data are distributed as a grid to maximize the probability of encountering an elevated area (Gilbert, 1987). In practice, this probability is usually unsatisfactorily low for small hotspots. In order to provide further assurance that no hotspots exist, field detection devices such as scanning technologies are used to identify local hotspots (USNRC, 2000, p. 5-47).[11]

Soil samples for comparison with the $DCGL_W$ are collected, assumptions about the variance and independence[12] are checked, and the hypothesis test is conducted. If this test passes, then for Class 2 and Class 3 units, scanning is implemented, and both scanning and lab measurements are compared to the $DCGL_{EMC}$. If values exceed the $DCGL_{EMC}$, the results are flagged for further investigation and potentially greater characterization (USNRC, 2000, p. 8-9). These steps form the *decision rule* for MARSSIM compliance.

KEY Concept: MARSSIM's Two Part Decision Rule

- Compliance with $DCGL_W$ is assessed with a hypothesis test.
- Compliance with $DCGL_{EMC}$ is assessed with field detection devices.

---

[11] Some detection devices are not sensitive enough to detect small local hotspots, and the number of actual laboratory samples must be increased, creating a denser grid to accommodate this shortcoming (USNRC, 2000 p.5-36).

[12] Although MARSSIM emphasizes the need for spatial independence, it says nothing about what should be done if the data are found to be spatially auto-correlated. In that sense, GEM may eventually play a role for surface applications as well. This is discussed in the final chapter.

It is within this decision rule that application to the subsurface faces a major obstacle. In the subsurface, the shielding effects of soil (Eckerman and Ryman, 1993) and the inaccessibility of the subsurface exclude the possibility of thorough sensing for local hotspots, and the MARSSIM process breaks down. It is this breakdown that motivates development of the GEM.

KEY Concept: MARSSIM Decision Rule and Subsurface Failure

- Field detection devices cannot provide evidence of compliance with a $DCGL_{EMC}$ in the subsurface and the decision rule breaks down.
- This breakdown motivates development of the GEM.

Since investigators cannot scan for concentration values between known locations, there are currently only three options. First, one could require that the entire survey unit be remediated. For small areas, this indeed might be economically viable and lead to lower risk of future litigation. For large survey areas, this option may not be practical. Second, one could apply MARSSIM one layer at a time, beginning with the surface. If MARSSIM fails, the soil layer of some specified depth is removed, exposing a new surface. MARSSIM is reapplied to the new surface and the process is repeated until a layer passes the compliance. This approach represents a kind of exploratory remedial process that can also be costly and may miss deeper contamination underlying compliant layers.

Another option (used by the GEM) is to model values between samples to assess compliance with $DCGL_{EMC}$ prior to any remedial action.[13] Using geospatial models, though, generally assumes that some form of spatial continuity exists between points (spatial auto-correlation). This violates the assumption of independence central to the WRS and Sign test and $DCGL_W$

[13] In fact, for non-radiologically contaminated sites, investigators often have no means for exhaustive scanning, even at the surface. This is where GEM may benefit non-radiological guidance as well. This is discussed in the final chapter.

evaluation. Griffith (2005) shows how auto-correlation can impact statistical confidence.

Rather than build a decision rule based on different approaches (e.g., hypothesis test for $DCGL_W$ and scan for $DCGL_{EMC}$), the GEM shifts entirely to a geospatial modeling paradigm where a single multi-scale decision rule may be applied. The core principle[14] of the GEM is that a site will fail compliance if the probability of exceeding a DCGL for any exposure unit of any size and shape, situated anywhere within the survey area (including the survey unit itself), exceeds an established probability limit.

Development of the GEM will require a geospatial model that can:

1) model the uncertainty (probability) about exceeding a DCGL for any exposure unit situated anywhere within the study area, and
2) integrate different forms of data in the model (field methods, laboratory methods, etc.) consistent with Triad methodology.

Geostatistical simulation meets both of these requirements, and the discussion now turns to key concepts from this field that the GEM draws upon.

Key Concept: Geostatistical Simulation and Decision Support

- Fills the knowledge gap formed by the absence of exhaustive scanning.
- Characterizes the probability that a DCGL is exceeded at any spatial scale.
- Is an input to the GEM, where the decision rule is based on a probability of exceedance limit applied uniformly over all scales and locations.

Geostatistical Simulation

Geostatistics is concerned with assessing and modeling attributes that vary in space (or time). Only geostatistical concepts specific to the GEM are presented

[14] While GEM is formally developed in Chapter 3, the key principle of GEM will be stated here in order to enable discussions regarding additional key concepts required for development.

here, and interested readers are encouraged to begin with Isaaks and Srivastava (1989), which provides an accessible introduction. Goovaerts (1997) is an excellent continuation of the subject, providing mathematically thorough yet accessible explanations illustrated with numerous examples and frank discussions about the limits and misuse of geostatistics in environmental characterization. Deutsch and Journel (1992) add to this discussion and provide users with a computational library known as GSLIB to facilitate the use of geostatistical methods.[15]

Deutsch and Journel (1992, pp. 9-18) provide a concise introduction to the fundamentals of geostatistics. The discussion here draws heavily on that work. The primary goal of geostatistics is to characterize the attribute of interest at unsampled locations, in this case, radiological concentration levels. It is common in the literature to denote the spatial coordinates as a vector $\mathbf{u} = (x, y, z)$, where x and y represent horizontal position and z represents depth below the surface. For any location **u**, geostatistical models treat the unknown concentration c(**u**) as a random variable C(**u**). The probability distribution function (pdf) and the cumulative distribution function (cdf) for C(**u**) characterize uncertainty about c(**u**). These distributions are determined or *conditioned* by existing samples. In this case, the cdf is referred to as the *conditional cumulative distribution function* (ccdf). Using DCGL notation, the expression for the ccdf is:

$$F(\mathbf{u}; DCGL \mid (n)) = pr\{C(\mathbf{u}) \leq DCGL \mid (n)\} \qquad \text{Eq 2.1}$$

In this equation, (n) represents the conditioning sample size and *pr* refers to probability. These distributions permit investigators to characterize c(**u**) in a variety of ways (Deutsch and Journel, 1992):

- What is the probability that c(**u**) < DCGL?
- What is the probability that c(**u**) lies in [a,b]?

[15] SADA's geostatistical algorithms are based largely on GSLIB routines (Stewart et al., 2009).

- What is an estimate for c(**u**)?

Typically, investigators are interested in creating a “continuous” characterization of concentration using a raster model. The raster is created in 3d space by dividing the spatial domain with a grid system. At the center of each cubic grid cell lies a random variable, C(**u**). In Figure 2.4, a 21x8 raster grid (two dimensional only) is presented along with seven sample locations symbolized by colored circles. This grid contains 168 cells, and therefore 168 random variables are present, one at the center of each cell. Example distributions are illustrated for two of the 168 random variables.

This set of random variables forms a *random function* characterized also by a conditional probability distribution function and corresponding cumulative distribution function written as (Deutsch and Journel, 1992; Goovaerts, 1997):

$$F(\mathbf{u}_1,...,\mathbf{u}_K; DCGL \mid (n)) = P\{C(\mathbf{u}_1) \le DCGL,...,C(\mathbf{u}_K) \le DCGL \mid (n)\} \qquad \text{Eq 2.2}$$

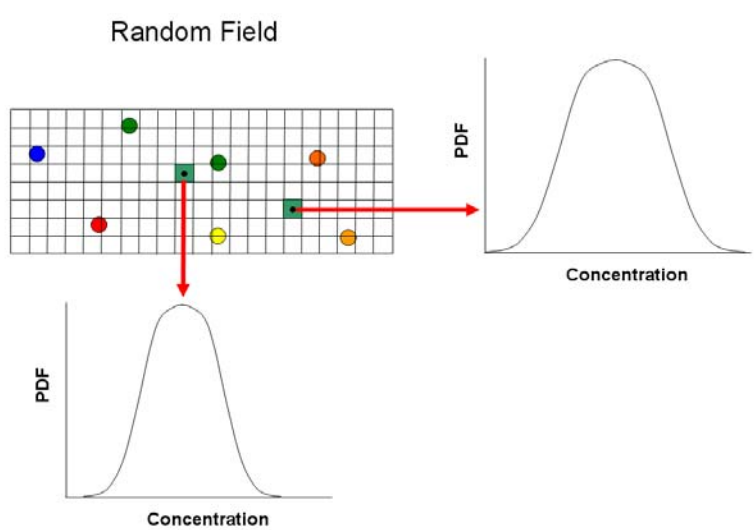


Figure 2.4 Two random variable distributions from the set of 168.

This ccdf is called the *joint ccdf* and permits investigators to characterize spatial uncertainty across multiple locations at once rather than only the local uncertainty at a single **u** (Eq.2.1).

Geostatistical characterization relies on the presence of spatial correlation as the basis for how sample values actually condition the ccdf. Geostatistical methods can be organized into two major groups distinguished by how the data and associated spatial correlation are used to condition the ccdf: kriging and simulation.

Kriging estimates a value for c(**u**) as a weighted combination of nearby samples. The simplest form of kriging, known conveniently as simple kriging (SK), is written as (Deutsch and Journel, 1992, p. 14):

$$C^*_{SK}(\mathbf{u}) = \sum_{i=1}^{n} \lambda_i(\mathbf{u}) C(\mathbf{u}_i) + \left[1 - \sum_{i=1}^{n} \lambda_i(\mathbf{u}) \right] m \qquad \text{Eq 2.3}$$

C*(**u**) is the estimated value of c(**u**), C($\mathbf{u}_i$) is in practice the measured value at the ith location,[16] n is the number of existing samples, λ_i is the weight assigned to C($\mathbf{u}_i$) and *m* is the mean of the random function. The weights (λ_i) are selected such that they minimize a quantity known as the *kriging variance,* formally written in Eq. 2.4 where G is a covariance function describing spatial correlation over distance.

$$\sigma^2_{SK}(\mathbf{u}) = G(0) - \sum_{i=1}^{n} \lambda_i(\mathbf{u}) G(\mathbf{u} - \mathbf{u}_i) \qquad \text{Eq 2.4}$$

[16] In the random function approach, every location (sampled or not) is represented by a random variable C(**u**). In cases where an actual measurement is taken, the value is a particular realization c(**u**) of C(**u**). Indeed this realization (or reality) is the most important to investigators; however, the most generalized formulation C(u) is used in the equation for kriging.

There is a history of using the kriging variance as a model of uncertainty about C*. For example, under the decision to assume normality, the kriging estimate becomes the mean and the kriging variance becomes the variance of a normal distribution assigned to C(**u**). There is a fundamental problem with using the kriging variance this way. Notice that in Eq 2.4, there is no term that involves the actual value of any sample point[17] but only the distance between values. The result is that variance is only a function of the spatial distribution of points and not their values (Goovaerts, 1997; Deutsch and Journel, 1992). Hence, in areas where samples collected close together demonstrate widely different concentration values, the estimates in that area will present low variances and overestimate the confidence about the true value.

On the other hand, geostatistical simulation permits the empirical development of the ccdf at C(**u**) by creating equiprobable realizations of the random function (Goovaerts, 1997) based on actual values. Figure 2.5 shows 3 such realizations for a two dimensional exposure unit.

The ccdf at any random variable C(**u**) is numerically constructed by the realizations $c^{(q)}(\mathbf{u})$ at location **u**, where (q) refers to the qth realization. Realizations are generated using the sequential simulation algorithm described in detail in Goovaerts (1997, p. 377) and briefly summarized here. The algorithm begins by randomly selecting a starting node and modeling its cdf based on the data. A simulated value drawn from that ccdf becomes a conditioning datum for all subsequent drawings. Each remaining node is randomly selected one at a time, with ccdfs developed using both the original data and any previously simulated node values. This process is repeated until all nodes have received a simulated value. The resulting set of simulated nodes represents one spatial

[17] Even the weight λ is not based on actual values but rather only the difference between values (Deutsch and Journel, 1992, pp. 14-15)

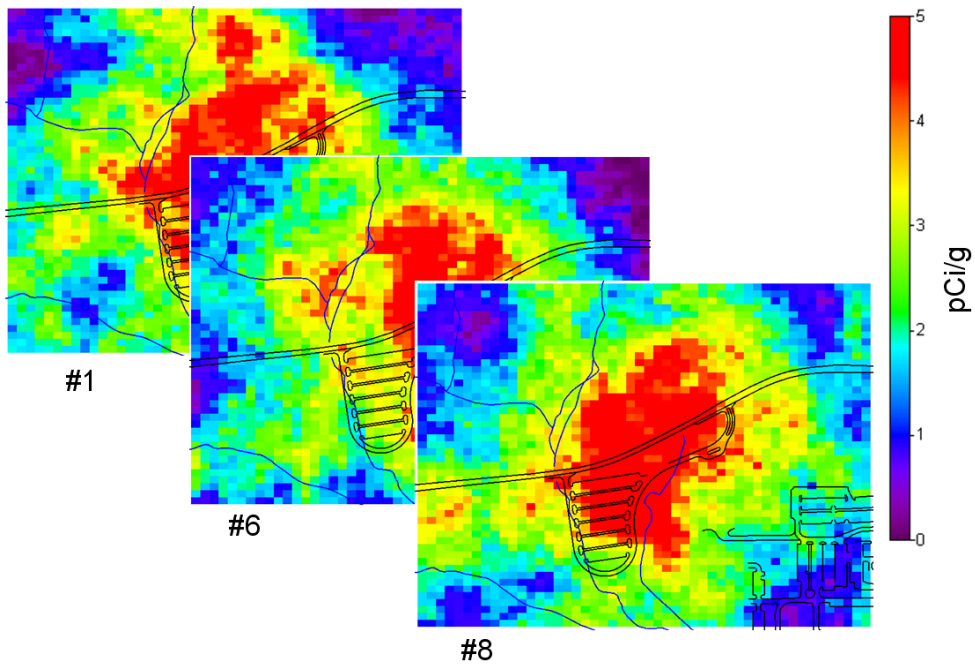


Figure 2.5. Three realizations of the random function.

realization. Many such realizations can be developed to create a ccdf at each location based on both the geometrical arrangement of samples *and* their value (Goovaerts, 1997).

In addition, post processing these realizations permits estimation maps (Eq.2.5), probability maps (Eq. 2.6), and variance maps (Eq. 2.7), all based on spatial correlation, geometric arrangement of samples, and sample values.

$$c^*(\mathbf{u}) = \frac{\sum_{q=1}^{Q} c^{(q)}(\mathbf{u})}{Q} \qquad \text{(Eq. 2.5)}$$

$$prob\{C(\mathbf{u}) \succ DCGL\} = \frac{\sum_{q=1}^{Q} \begin{Bmatrix} 1 & if\ c^{(q)}(\mathbf{u}) \succ DCGL \\ 0 & if\ c^{(q)}(\mathbf{u}) \leq DCGL \end{Bmatrix}}{Q} \qquad \text{(Eq. 2.6)}$$

$$\mathrm{var}(C(\mathbf{u})) = \frac{\sum_{q=1}^{Q} \left(c^{(q)}(\mathbf{u}) - c^{*}(\mathbf{u})\right)^2}{Q-1} \qquad \text{(Eq. 2.7)}$$

Simulation also poses another powerful property vital to the construction of the the GEM. Under geostatistical simulation, it is possible to characterize uncertainty about the *average concentration* over any area of interest (i.e., an exposure unit).[18] This is accomplished through a *change of support*. The support refers to the spatial scale at which information is presented. When samples are collected, the volume or support of the soil sample is quite small. When characterizing unknown concentrations at that same support level (e.g. point estimation), it is appropriate to use the ccdf C(**u**). Suppose that interest exists in determining whether an average concentration over a larger support, such as an exposure unit, exceed a DCGL value. In this case, the average concentration is also treated as a random variable over a spatial volume **E,** C(**E**), characterized by averaging the point realizations $c^{(q)}(\mathbf{u})\ \ \forall \mathbf{u} \in E$ within each simulation (Goovaerts, 1997). In other words:

$$C^{(q)}(\mathbf{E}) = \frac{\sum_{i=1}^{N} c^{(q)}(\mathbf{u}_i)}{N}\ where\ \{\mathbf{u}_i, ..., \mathbf{u}_N\} \in \mathbf{E} \qquad \text{Eq 2.8}$$

[18] Another approach is to use *block kriging* where average concentrations are estimated over a larger block. Problems with using the kriging variance as a model of uncertainty, however, still persist.

The set of Q realizations $\left\{C^{(1)}(\mathbf{E}), ..., C^{(Q)}(\mathbf{E})\right\}$ develop the ccdf for the average concentration within the exposure unit E for any exposure unit threshold (e.g., $DCGL_E$):

$$F(E; DCGL_E \mid (n)) \;=\; pr\left\{C(E) \le DCGL_E \mid (n)\right\} \;=\; \frac{\sum_{q=1}^{Q}\begin{Bmatrix} 1 & if\ C^{(q)}(\mathbf{u}) \succ DCGL_E \\ 0 & if\ C^{(q)}(\mathbf{u}) \le DCGL_E \end{Bmatrix}}{Q} \qquad \text{Eq 2.9}$$

For the GEM, this permits assessment of the probability that an exposure unit will exceed a corresponding DCGL value. This process can be conducted for any domain E (i.e., exposure unit) of any shape or size (including the entire survey area),[19] a key requirement in the development of the GEM decision rule in Chapter 3.

There are many forms of geostatistical simulation from which to choose, each with their own strengths and weaknesses (Goovaerts, 1997). Geostatistical simulation is an input to the GEM which is agnostic to the type of simulation used. The simulation method should be selected based on the kind of site-specific circumstances that investigators may face. Indeed, one cannot defensibly claim that one simulation algorithm is best for all cases (Goovaerts, 2001). At the same time, one popular simulation approach, known as sequential indicator simulation (SIS) (Goovaerts, 1997, p. 393), may be an accessible choice for investigators to consider for three particular reasons. Indicator-based approaches:

---

[19] Geostatistical simulations are parameterized by the data set at hand and seek to preserve various statistical properties of the data (e.g. mean, histogram, correlation structure). This is only done *on average* over numerous simulations. Hence even at the global scale (entire survey area), it is possible to model uncertainty about the global average due to *ergodic fluctuations* in the realizations (Deutsch and Journel 1992, p. 127). This is a particularly attractive trait about simulation that permits a single probability calculation about the average concentration and the DCGL at all scales and locations of interest.

- are non-parametric approaches to modeling (Goovaerts, 1997, p. 284). Non parametric methods are preferred in MARSSIM, evidenced by the emphasis on Sign and WRS tests (USNRC, 2000),.
- can capture *non-linear* patterns in the distribution of radiological contamination (Deutsch and Journel, 1992, p. 71), and
- provide an accessible way to encode different kinds of information into the model (Goovaerts, 1997, p. 395). This includes both hard (laboratory) and soft (field detection measurements) data alike.

A very brief overview of SIS will support understanding of key concepts in the upcoming discussion of Triad. In addition, SIS is used to demonstrate the GEM in the Cesium Site example presented in the next chapter. The discussion begins with an overview of indicator formalism.

Indicator approaches (e.g., SIS) make no assumption about the shape of the ccdf[20] at any point **u**. Rather, the ccdf is empirically derived as follows. First, the range of values is divided into a series of K threshold values c_k.[21] For a given threshold value, the N sampled values undergo the indicator transform $I(\mathbf{u}; c_k) = 1$ if $c(\mathbf{u}) \leq c_k$ and zero otherwise. For each location **u**, kriging is applied[22] using the N transformed data values and the associated model of spatial correlation for those transforms. The kriging estimate represents the probability that an indicator transform of the true but unknown value at **u** would be zero. The probability that the true value is less than c_k is by definition the ccdf value at threshold c_k (Eq. 2.1). Repeating this process for all monotonic increasing values c_k, the empirical ccdf $F(\mathbf{u})$ is constructed at each **u** (Eq. 2.10).

---

[20] Another popular form of simulation, known as sequential Gaussian simulation, assumes normality.
[21] Guidance on selecting those thresholds is provided in Goovaerts (1997, p. 285).
[22] In this case, kriging is being used strictly as a method of interpolation and the kriging variance is not considered as a form of uncertainty.

$$F(\mathbf{u}; c_k \mid (n)) = pr\{C(\mathbf{u}) \le c_k \mid (n)\} \qquad \text{Eq 2.10}$$

Indeed this process is applied to every **u** of interest in the site to provide a raster of ccdfs (Goovaerts, 1997, p. 284).

For any threshold value c_k, the transformation of hard values, c(**u**), into 0s and 1s is an *encoding* of conditioning information into the model. The indicator formalism permits investigators to encode values other than 0 and 1 as well. This opens the door for *soft* data, such as less accurate field measurement methods to be used in the model (Goovaerts, 1997, p. 292). While complete scanning of the subsurface is indeed impossible, there are field detection methods for producing quick soil measurements (not exhaustive scans) that may not have the accuracy or precision of laboratory methods but may be sufficient for the decision at hand (MARSSIM, 2000, p.6-1, Appendix H). Measurements typically include both a hard constraint interval [a,b] and a probability distribution describing variability within this interval (USNRC, 2000, p. 6-54). This probability distribution permits the calculation of the probability that the true value is less than c_k. This encoded value at location **u** is found in Eq. 2.11 (Goovaerts, 1997, p. 292).

$$i(\mathbf{u}; c_k) = \begin{Bmatrix} 1 & if\ c_k \succ b \\ F^*(c_k) & if\ c_k \in [a,b] \\ 0 & if\ c_k \prec a \end{Bmatrix} \qquad \text{Eq 2.11}$$

Here, $F^*(c_k)$ is the cumulative distribution function representing the field measurement uncertainty with respect to c_k in the interval [a,b].

This is an important concept in making the GEM a viable option. First, integration of field measurement results is an important principle of Triad (discussion following). Second, a concern exists within the environmental community about the number of samples required to support a geostatistical evaluation (USEPA,

1992b, p. 2-1). The concern arises over the cost of sampling, and cheaper field detection methods can mitigate this financial burden. This is where geostatistical methods and Triad come together under the GEM to provide a way for inexpensive field methods to be integrated directly into a decision process, permitting greater sampling density at potentially viable cost. Other methods are also available for integration of field measurement data, including sequential Gaussian simulation, where more complex methods of integration are required (Goovaerts, 1997, pp. 385-392).

Under this indicator formalism, SIS produces joint realizations (see Goovaerts 1997, p.395) of the random function by sequentially drawing realizations from each of the local ccdfs. These realizations represent the joint behavior of multiple C(**u**)s and permit the uncertainty about exceedances at different spatial scales (Eq. 2.9). The algorithm for sequential simulation is given by Goovaerts (1997, p. 377).

KEY Concepts: GEM and Geostatistical simulation

- Geostatistical simulation permits calculation that a decision criterion is exceeded at any scale.
- Geostatistical simulation permits inclusion of cheaper, faster field measurements, an activity emphasized by Triad.
- GEM is agnostic to the particular form of simulation.

It is important to emphasize that simulation is considered an input into the GEM process and not itself the focus of this dissertation. Methods for establishing and assessing the quality of a simulation are outside the scope of this work. Readers are encouraged to review Deutsch and Journel (1992) and Goovaerts (1997) for details on building a simulation model.

A close connection exists between the GEM and the emerging methods in Triad. The discussion now continues with a brief overview of Triad and how the GEM is connected to that effort.

Triad

The Triad model is an EPA initiative to foster modernization of technical practices in characterization and remediation of contaminated sites. Triad is a result of the combined efforts and expertise of experienced practitioners from the public and private sector to formulate a framework for managing decision uncertainty and increasing confidence that decisions are made as efficiently and accurately as possible (Crumbling, 2003, 2004).

Focus on decision quality is a hallmark of the Triad method and a departure from narrow notions of “data quality” that focus primarily on measurement accuracy. This has been driven to some degree by regulatory pressure, evidenced in the rejection of screening and field detection methods in many final decisions. Unfortunately, higher analytic accuracy comes at greater cost. As a result, project managers may limit the number of samples collected (Crumbling, 2002). This is particularly problematic for geostatistical models, which typically require more data for proper calibration (USEPA, 1992b).

Triad approaches expand the concept of data quality from an analytic quality to decision quality. In a perfect world, “decision quality” would be equivalent to “decision correctness”; however, decision correctness is often unknown at the time a decision must be made. In many cases, correctness may never be known, due to the situational complexity and conditions that have evolved over time. The term ‘‘decision quality’’ therefore means that decisions are defensible against reasonable scientific or legal challenges (Crumbling, 2002), given the best

available information and knowledge afforded by financial and professional resources at the time of investigation.

In the interest of decision quality, emphasis is placed on the use of alternative field and real-time measurements that may have reduced accuracy but impart valuable information relative to the decision. As a trivial example, suppose a decision limit of 100pCi/g is established. Method A, an expensive sampling approach, is able to detect radiation levels as low as 0.1 pCi/g and measure it to several significant digits. Method A, however, does no better in supporting the decision than less expensive method B, which can detect activities as low as 20 pCi/g and measure it within +/- 10pCi/g. Both are well below the criteria of 100pCi/g. Therefore, overly accurate sampling wastes valuable resources. As previously discussed, geostatistical simulation is well suited to integrate these results directly into the model.

The foundation of Triad is the conceptual site model. A CSM is a representation of site knowledge that evolves over the course of investigation. CSMs communicate knowledge about a variety of issues, including geology, exposure pathways, spatial distribution of contamination, and transport mechanisms (Crumbling, 2001a; USEPA, 1992b). CSMs can take on a variety of forms. Some CSMs are simple graphical depictions, as in Figure 2.6, or complex and quantitatively derived models, as in Figure 2.7. Both figures are taken from USEPA (2008).

Under Triad, the CSM drives data collection by identifying knowledge gaps. The CSM is reciprocally informed and evolved by the outcome of those data (Crumbling, 2001a). The CSM ultimately informs the decision making process, and a focus on increasing the content and information value of the CSM should direct activities throughout the investigation life cycle. The GEM is based on this

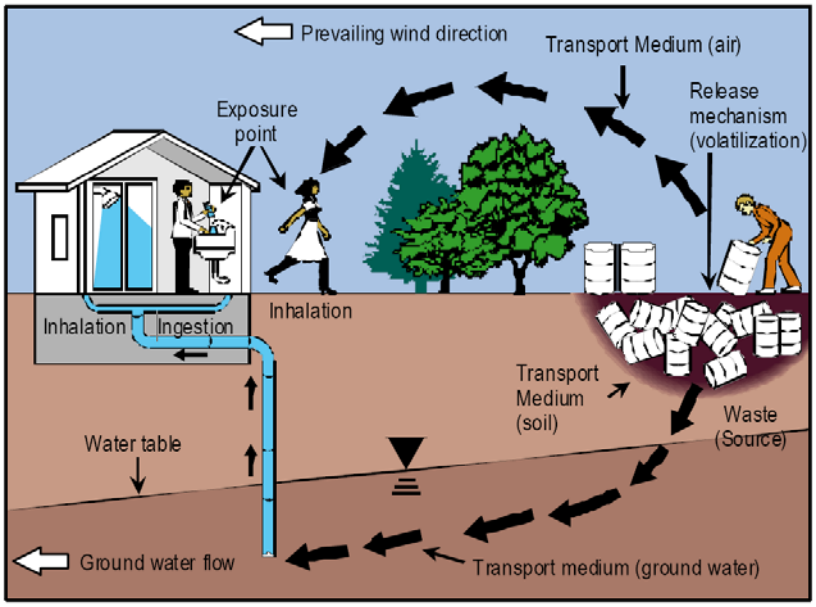


Figure 2.6 A simple CSM drawn in graphical software (USEPA, 2008, p. 1).

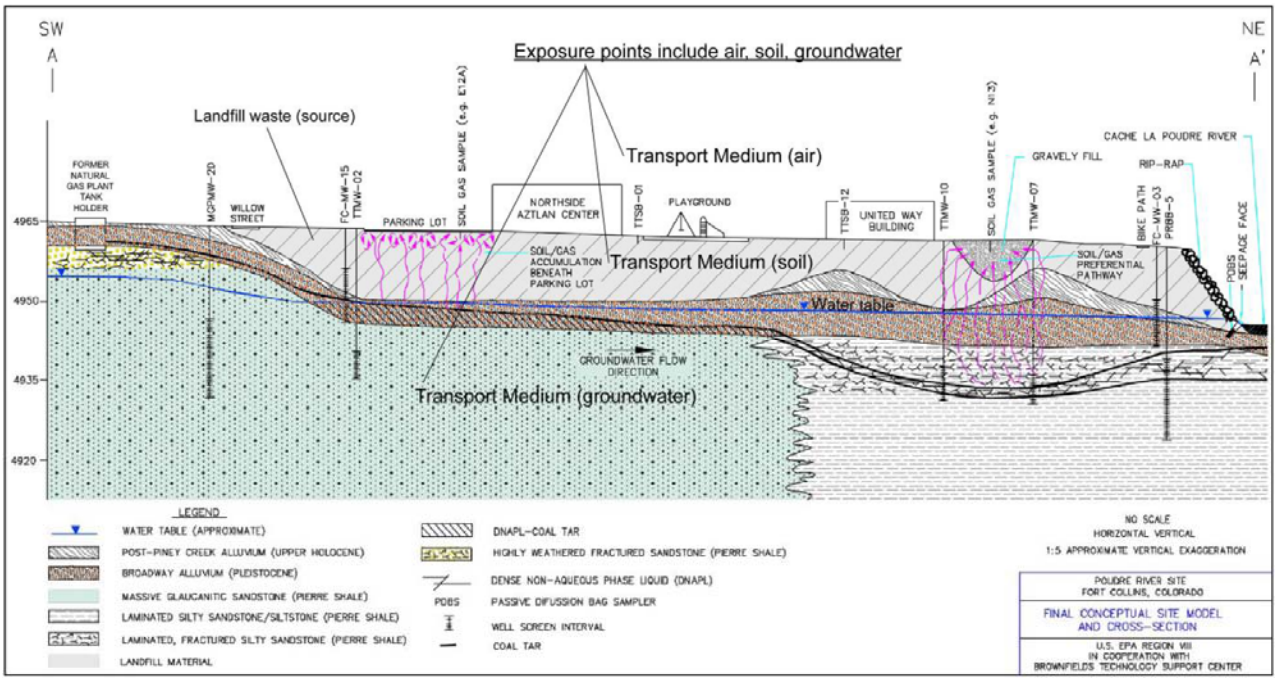


Figure 2.7 A variety of information is provided in this CSM, including quantitatively derived subsurface conditions (USEPA, 2008, p. 2).

same approach. The CSM in the GEM is a *stochastic conceptual site model* (SCSM) which provides the test against the RLR. The SCSM is a multi-scale model of compliance that reflects the current state of knowledge regarding compliance.

KEY Concept: GEM and Triad

- The foundation of both Triad and GEM is a CSM, which supports the decision process.
- Triad and GEM emphasize the use of both laboratory and field measurements to offset the cost of sampling and to improve the decision.

With these concepts in hand, it is possible to now derive the GEM. Prior to this, however, a review of the literature is presented that canvases relevant regulation and the vibrant activity that has occurred in the literature over the past two decades in geostatistical decision support, sample design, and regulatory revision. The publications presented herein are designed to provide a sense of the regulatory and modeling communities from which the GEM arises.

Closely Related Work

The GEM draws on a growing body of work in spatially-based decision making, sample design, remedial design and calls for regulatory guidance such as Triad. Many of these publications have already been mentioned in discussing key concepts. Others are presented here and organized by what element of the GEM they are closely related to. As various aspects of the GEM are developed in the upcoming chapters, several of these will be discussed in greater detail.

Regulatory Guidance and Geostatistical Decision Making

The USEPA has produced a number of regulatory documents that direct environmental characterization and cleanup activities under the DQO process (USEPA, 2006a). From the vantage of decision support, these documents are

largely centered on issues of estimating the mean concentration level, geometric designs for hotspot searches,[23] and statistical hypothesis testing under the assumption of spatial independence. Sample designs usually include standard design methods (Delmelle and Goovaerts, 2009), including random, stratified, systematic and grid, ranked set, adaptive cluster, and composite sampling (USEPA, 2002c). Regulatory documents commonly follow suit, including USEPA (1989, 1994a, 2000, 2002c, 2006a, 2006b, 2006d). The USEPA also produced software to help investigators implement many of the sample designs that arise from these statistical tests (USEPA, 2001b). The GEM framework presented here assumes that spatial dependence does in fact exist and that decision needs often require model-based designs (Delmelle and Goovaerts, 2009). These designs may be biased with respect to such things as estimating the mean, but they are powerful in delineating contamination and supporting the remedial cleanup.

The presence of geostatistical methods in guidance is minimal (Verstraete and Meirvenne, 2008). In the U.S., it is older regulatory guidance that addresses the possible role of geostatistics in environmental investigation but only considers kriging (USEPA, 1989b, 1992b, 2006c). Neither of these develops a compliance framework for geostatistics to operate under. Perhaps the best indication that a multi-scale geostatistical framework such as the GEM can resonate with the regulatory community arises in USEPA (1989b). This guidance document dedicates a chapter to instructing readers about what geostatistics does and shows how kriging can be used to estimate probability of exceeding a single decision criterion. The document stops short of developing a formal framework and states that more work effort in understanding these methods is first required (p. 10-9). The greatest regulatory discussion of geostatistics was found in USEPA (1992b), where soil sampling and handling protocols were extensively

[23] These hotspot search methods amount to applying triangular or rectangular grids on the site with node spacing sufficient to encounter a hotspot of a given size. For small hotspots, the number of required samples to meet a target probability may not be economically viable. For details, see Gilbert (1987).

described with respect to kriging. Finally, USEPA (2006c) acknowledges that geostatistical methods (kriging) are valuable tools in Chapter 2 (B.2). These publications indicate that the regulatory community is willing to consider the use of geostatistical approaches. The goal of the GEM is to bring advanced geostatistical methods to the forefront of debates over guidance revision by demonstrating their viability within a formal framework.

Geostatistics in the Characterization and Remedial Design Literature

Geostatistical-based decision making has been a vibrant publication area for the last 15 years or more. This section provides an overview of the more relevant publications to this work and gives a sense of the large amount of activity in this research area.

An important area of research regards model-based sampling designs, which intend to refine or improve geostatistical decision support. In this dissertation, two sample designs (MrsDM and Check and Cover) are presented. A common theme among many of these approaches is the concept of *simulated sampling*. In a simulated sampling, a location(s) for the next sample(s) is identified using some decision rule. A modeled value(s) at that location is then added to the set of real data, and the geostatistical model is reapplied. This process is repeated to generate each new sample location(s). Typically, the decision on how many samples to collect is based on a cost-benefit analysis, where the cost of additional samples is compared to the potential benefit of collecting the sample (Freeze et al., 1992). A number of these are based on kriging rather than simulation and focus on minimizing uncertainty at specific nodes rather than across larger and possibly multiple spatial units (exposure units).

Methods based on kriging are briefly enumerated here. These include Groenigen et al. (1999), Vasat et al. (2010), Delmelle and Goovaerts (2009), Watson et al.

(1995), Simbahan and Doberman (2006), Juang et al. (2008), and Stewart et al. (2009). One method presented by Johnson (1996) for including expert judgment has made appearances in the literature over the last decade and has had a small but collaborative relationship with the SADA project, where some methods were implemented. In this approach, each grid node is assigned a Bayesian prior in the form of a beta distribution function. Additionally, a prior covariance structure is assumed as well. The Bayesian posterior is accomplished heuristically by combining the kriging estimate and the prior beta distribution (Johnson et al., 1996, Eq. 7).

In Johnson's approach to sample design (Johnson et al., 2005), sample locations are optimized in one of two ways: "Outside-in" and "Inside-out". In "Outside-in", samples are collected based on their expected minimization of the contaminated area in the posterior update. In "Inside-out", the samples are selected based on their expected maximization of the contaminated area in the posterior update. With respect to multi-scale compliance, simulation was not the basis of the approach; therefore, it is not possible to rigorously aggregate compliance evaluation at higher spatial scales. Still, the concept of single scale "Outside-in" informs the multi-scale GEM sample design developed in Chapter 3 (MrsDM), and detailed discussion is provided in that section. In application of this method to the subsurface (Johnson et al., 1999), this two-dimensional approach is repeatedly applied to each subsurface layer as it is removed and remediated.

Numerous examples exist in the literature where geostatistical simulation is used during characterization and remediation activities to support decision making. For example, Pilger et al. (2001) use sequential Gaussian simulation to model the uncertainty associated at each grid node. In this approach, nodes exhibiting the greatest variability in the ccdf are selected as new sample locations. The benefit of sampling is measured to be the reduction in local variability of nearby nodes and the globally averaged reduction in local node variability. Unlike the GEM, no

decision criterion drives the measure of uncertainty, and only local uncertainty at the nodes is considered (rather than multiple spatial scales).

Verstraete and Van Mervenne (2008) suggest a sample design based on minimizing local (node) uncertainty about exceeding a single decision criterion. Geostatistical simulation rather than kriging was used to build the local ccdf. Goovaerts (1999) work "Geostatistics in soil science: state-of-the-art and perspectives" provides a concise synopsis of his 1997 book and discusses briefly the use of simulation to produce area-based probabilities (e.g., Eq 2.8, 2.9). Saito and Goovaerts (2003) use geostatistical simulation as the basis for planning a remedial design for a single decision criterion, for geographically fixed exposure units. The GEM-based remedial design (MrDM), developed in Chapter 3, extends and modifies this approach to include a continuum of exposure unit sizes and shapes potentially occurring anywhere on the site. This paper is more closely examined in that chapter.

Goovaerts (2001) provides an excellent discussion of uncertainty assessment in soil science, compares kriging to simulation, makes recommendations for when one choice is better than the other, and provides useful tools in assessing the quality of a geostatistical model. Emery (2008) adds to these methods by providing statistical tests for validating geostatistical simulation algorithms. Brakewood (2000) use a moving window approach to scan a contaminated area (data or modeled) for violations of a single concentration limit for a single exposure unit size. Brakewood does not consider multiple criteria, scales, or uncertainty in the process (Saito and Goovaerts, 2003). Similarly Van Tooren and Mosselman (1996) rely on a moving window approach based on kriging.

During the mid 1990s, a great deal of activity surrounded data worth in sample design (Freeze et al., 1992). England et al. (1992) propose a method of sample optimization using geostatistical simulation that would minimize a given cost

function at the remedial unit scale[24] for a single decision criterion. James et al. (1994) discuss data worth regarding aquifer remediation design. Lyon et al. (1994) discuss estimating the value of perfect information in sample design. Van Groeningen et al (1997 and 2000) discuss methods for optimizing soil sampling locations against a single decision criteria using kriging based probability maps. Dakins et al. (1996) discuss the expected value of sample information in risk-based environmental remediation for fate and transport models. McNulty et al. (1997) discuss value of information analysis within the context of groundwater modeling activities at the Nevada Test Site.

More recently, Demougeot-Renard et al. (2004) also demonstrate a sample design that attempts to minimize the uncertainty about cost associated with volume of removal for a single decision criterion. This paper is revisited when the GEM-based sample design (MrsDM) is developed in Chapter 3. Norberg et al. (2006) proposes a Bayesian method for computing the number of samples based on a data worth analysis. Back (2006, 2007) provides an excellent laundry list of publications in this area, adds additional content regarding measurement accuracy in hotspot delineation, and concludes by connecting the value of information analysis with the DQO process. In the GEM, the MrsDM sampling design (developed in Chapter 3) considers the value of data only within the context of how much uncertainty in the remedial design is reduced while remaining protective of public health. Investigators may apply a cost function to the GEM output to translate failure risk, if so desired. This is outside the scope of this work.

Finally, Meyer et al. (1988) use the P-median algorithm to locate groundwater wells at places most likely to encounter contamination in flow. In this paper, the method is extended to a new GEM-based sample design called “Check and

[24] A remedial unit is the smallest soil volume that can be removed or remediated (e.g. backhoe scoop)

Cover" that may support sample designs in the scoping phase or early characterization. Check and Cover is offered here as an additional sampling design strategy (Chapter 5) that may support the process but is not explicitly connected to the GEM.

MARSSIM, Triad, and SADA

Publications for the recently developed Triad process include Crumbling (2001a, 2001b, 2002, 2003, and 2004). The EPA produced a technology primer that focused on management strategies (USEPA, 2003) under Triad. More recently, a Triad issue paper was published by USEPA (2008) that discussed the role of geophysics in Triad. SADA is listed as a Triad resource on its website (see www.Triadcentral.org, last accessed 3/18/2011) and was included as a training course in the Triad National Conference and Training in 2008 (www.umass.edu/tei/conferences/courses_description.html). SADA is identified as a Triad code again in USEPA (2005) and (2005a). Applications of Triad include Byrn (2003) and Puckett and Shaw (2004), which document the use of SADA in Triad.

The Interstate Technology and Regulatory Council (ITRC) is a highly active panel of state environmental agency members who produce publications that supplement and interpret federal policies and new technology innovations within a state regulatory context. An excellent discussion of Triad can be found in ITRC (2003), where a paragraph is devoted to the promising connection between Triad and MARSSIM, although no specifics are given. ITRC also produced a regulatory supplement on decontaminating and decommissioning radiologically contaminated facilities (ITRC, 2008). Additionally, the ITRC provided support for SADA in a 2008 technology transfer workshop (see www.itrcweb.org/conf_aram.asp, last accessed 2/9/2011).

Johnson et al. (2004) makes an early connection between Triad and MARSSIM. In this paper, Johnson demonstrates that X-Ray fluorescent measurements serve well as surrogates for total uranium in a stream bed characterization for the purpose of detecting exceedances of local activity limits (e.g., $DCGL_{EMC}$). This use of secondary measurements in characterization is a Triad principle.

Many of the publications that discuss SADA were already enumerated in the first chapter of this paper and are therefore not discussed in detail here (USEPA, 2003, 2005, 2005a; USEPA/state of Pennsylvania, 2003; Franklin and Mills, 2003; Linkov et al., 2004; Sunila et al., 2004; Puckett Puckett and Shaw, 2004; Rügner et al., 2006; Purucker et al., 2007; Sinha et al., 2007; Butt et al., 2008; Carlon et al., 2008; Norman et al., 2008).

Publications key to the GEM

From this body of literature, four publications figure prominently in the dissertation work conducted here. First, the development of the GEM decision rule extends the single decision criteria normally found in many geostatistical publications (such as those listed here) to multiple, scale-dependent criteria required for the evaluation of a continuum of exposure unit sizes situated anywhere on the site. Second, the SCSM test is essentially a model of compliance based on geostatistical (stochastic) simulation. Using such a stochastic CSM or a geostatistical simulation model directly in the compliance decision, however, is believed to be a new approach for regulatory guidance and may lay the groundwork for a geospatial paradigm in regulatory decision making.

From this GEM decision rule, the MrDM is developed by extending and modifying the methods published by Saito and Goovaerts (2003) from single to multi-scale decision criteria and from a fixed set of exposure units to a continuum of exposure unit sizes and shapes that can be placed anywhere across the site.

Other modifications were made as well and are discussed in Chapter 3. The sample design strategy (MrsDM) accompanies the MrDM approach and is informed by both Demougeot-Renard et al. (2004) and the Johnson principle of “outside in” (1996). Unlike either of these methods, MrsDM is a multi-scale sampling strategy specifically designed to reduce uncertainty in MrDM designs. Finally, the Check and Cover design extends the method proposed by Meyers et al. (1988) and supports the early characterization efforts.

Summary

The GEM represents a technical extension to MARSSIM that permits systematic and probabilistic evaluation of the subsurface. The GEM does two other things as well. First, it represents a technical basis for expanding the role of geospatial modeling within the body of guidance work. Second, it re-enforces the principles embodied in Triad and adds to the growing motivation for broader guidance revision. Figure 2.8 summarizes the core principles of the GEM and how those principles are situated within MARSSIM, Triad, and geostatistics.

The discussion is now prepared to move forward in developing the new GEM framework. The next chapter formally derives the GEM decision rule (RLR) and the stochastic conceptual site model (SCSM) and defines how together they form a basis for demonstrating dose-based compliance limits.

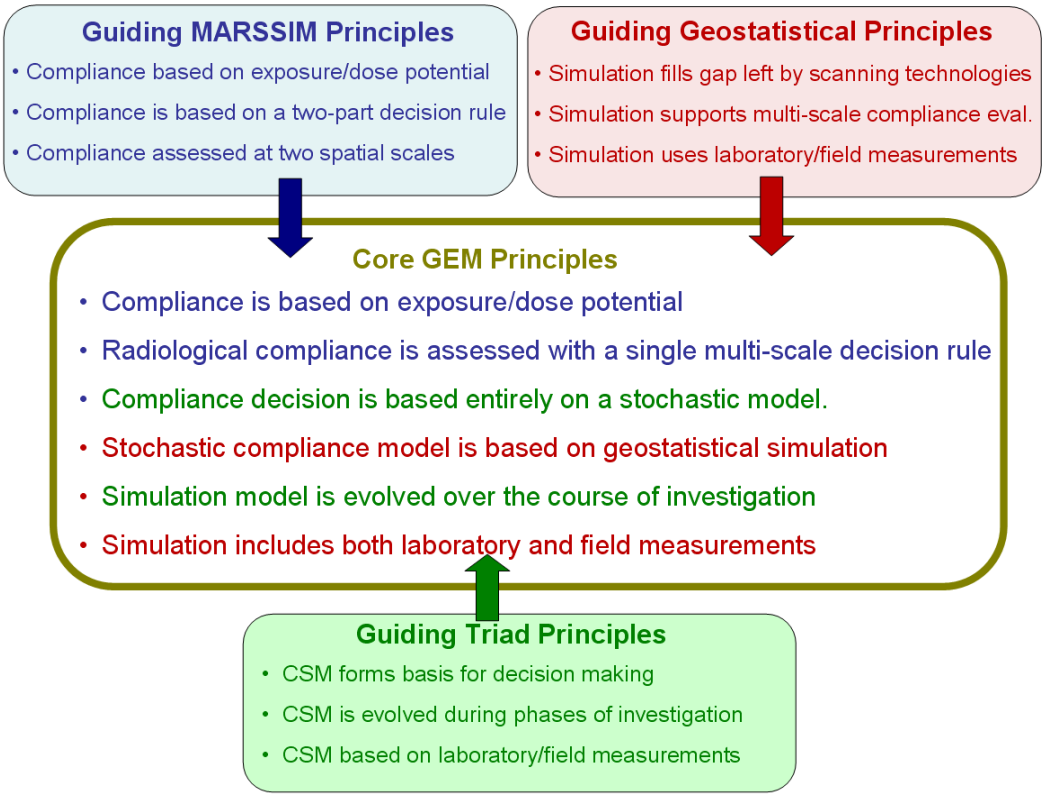


Figure 2.8 Relationship of the GEM to MARSSIM, Triad, and Geostatistical Simulation

Chapter 3: Theoretical Derivation of the GEM Framework

Introduction

As previously discussed in Chapter 2, MARSIMM presents a two part decision rule applied at two distinct spatial scales: A) The average concentration may not exceed the $DCGL_W$ and B) no local area of a specified size may have an average concentration exceeding the $DCGL_{EMC}$. Adherence to each part of the decision rule is demonstrated separately. Part A is demonstrated by applying a statistical hypothesis test. Part B is demonstrated using a radiological scan (USNRC, 2000). In the subsurface, exhaustive radiological scans are not possible and the method of demonstrating adherence breaks down. In order to avoid this breakdown, a new approach is needed.

The regulatory limit rule (RLR) provides a *model based*[25] decision rule that requires that for any volume of interest, the probability that the mean concentration exceeds the associated DCGL is less than a specified limit α. The decision rule test is provided by the stochastic conceptual site model (SCSM). The SCSM test is based on geostatistical simulations of radiological activity across the site and provides the probability of exceedance for any decision scale required by the RLR. Additionally, neither the RLR nor the SCSM test require the investigator to decide on a particular local scale. Rather, a range of possible sizes and shapes for the volume can be simultaneously considered relieving the investigator of this decision requirement. Table 3.1 provides a comparison of the decision components in the MARSSIM and the GEM.

When a site is found to be out of compliance according to the SCSM test, a remedial design is required to bring the site into compliance. A remedial design

[25] Model based means that an underlying model informs the basis of investigation rather than the data alone (National Research Council, 2007).

Table 3.1 Relationship between Decision Components.

Component	MARSSIM	GEM
Rule	$Mean_{Site} \leq DCGL_W$ $Mean_{Local} \leq DCGL_{EMC}$	$P(Mean_E \geq DCGL_S) \leq \alpha$
Test	Hypothesis Test Radiological Scan	SCSM

here means a spatial delineation of soil volumes that will be cleaned or replaced with clean soils. The goal then is to define the boundaries of such a soil volume. Within the GEM framework, the MrDM serves this role with respect to the RLR/SCSM approach and produces an estimate of the minimal remedial design that would result in compliance. In some situations where a degree of confidence (α) is sought, the remedial design will necessarily include areas that are more likely uncontaminated than not. This is due to the uncertainty in the spatial distribution of contaminants and the requested high degree of confidence about successful remediation. If the MrDM remedial design is deemed too costly, investigators can use MrsDM to estimate strategic positions where additional samples may decrease uncertainty in the spatial pattern and lead to a reduction in the MrDM remedial design volume. There is a strong connection between these different components that order compliance activities as shown in Figure 3.1.

There are two kinds of input into the GEM process. Investigators must supply the exposure scenarios and associated DCGL values to establish the RLR. Environmental measurements inform geostatistical simulations which join the RLR requirements as inputs into the SCSM test. If the test passes, the site meets compliance under the GEM. If not, then investigators may use the MrDM to

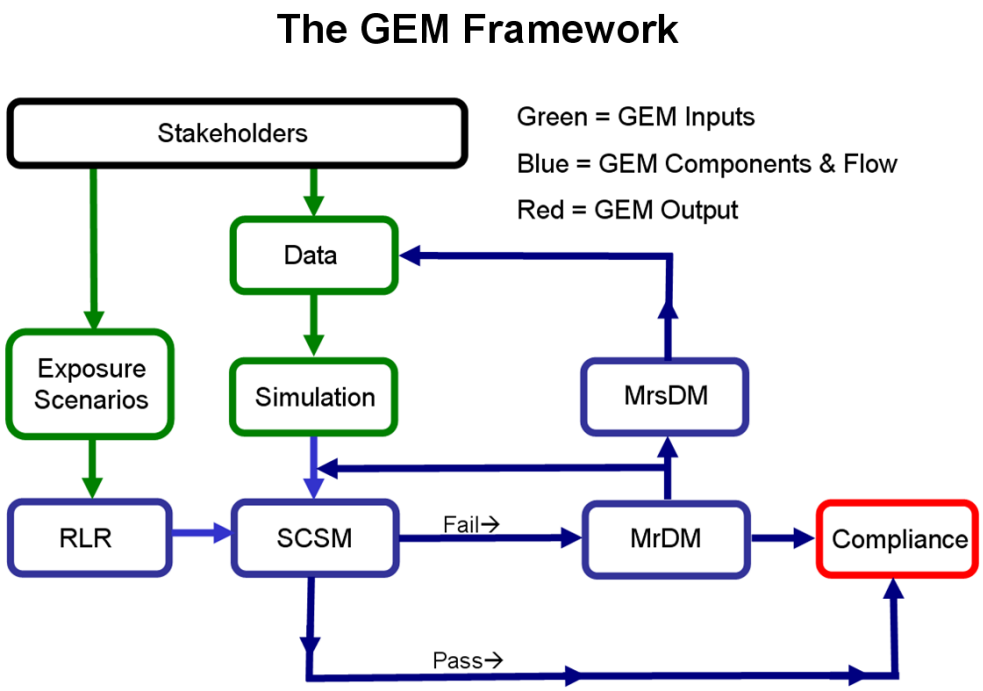


Figure 3.1 The GEM workflow.

develop a remedial design and further refine the design using the MrsDM sample design. The SCSM test is reapplied given the new samples and/or the remedial actions that were taken. The discussion now continues with a derivation of each of these components in the order they appear in the workflow (RLR→SCSM→MrDM→MrsDM).

The Regulatory Limit Rule

Let E(v,g,d,DCGL) represent a three dimensional soil exposure unit characterized by volume (v), geometry (g), exposure unit depth (d), and exposure unit concentration limit (DCGL). In Figure 3.2, two example cubic geometries are presented, one 2ft below the receptor and the other just at the surface with concentration limits of 30pCi/g and 50pCi/g respectively.

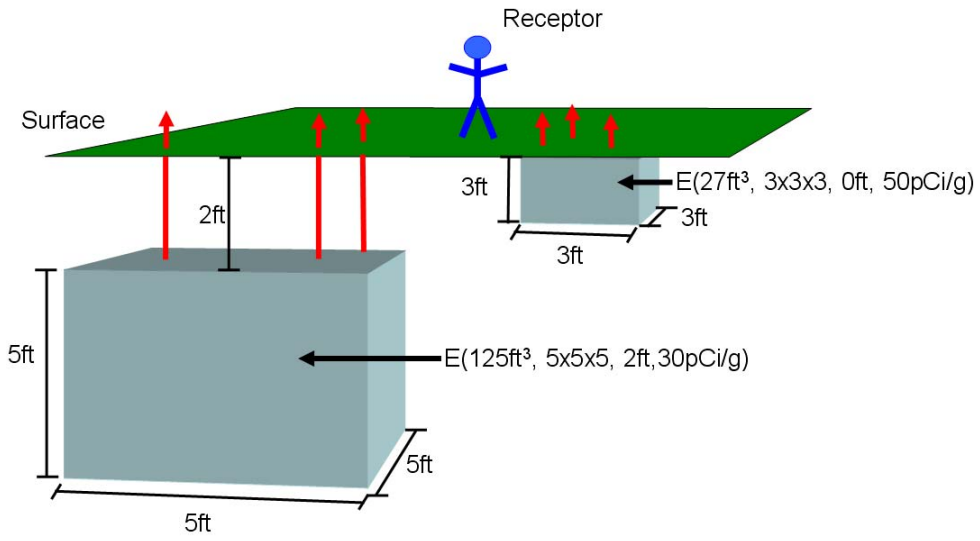


Figure 3.2 Two example external exposure units positioned in the subsurface.

In application, whether or not an exposure unit specification fails compliance may depend on where it is located. Positioning an exposure unit in a high concentration area results in a compliance failure whereas a low area will pass. Consider the situation in Figure 3.3 where a single exposure unit E(400ft^3,20x20x1ft,0ft,55pCi/g) represented by a square box is positioned in two different locations. In the lower left, the exposure unit (green) has a much better chance of passing the DCGL of 55pCi/g than the unit in a highly contaminated area near the center of the site (red).

Ideally, compliance would be checked at every coordinate (x, y, z) but this amounts to an infinite number of locations and poses an intractable computational problem for the SCSM (discussed below). Under RLR, decision makers will evaluate a finite set of positions defined by a three dimensional grid system. The *GEM spatial resolution grid system* is formed by overlaying the site in 3D space with a 3D grid specified with origin (x_0, y_0, z_0) and grid cell size

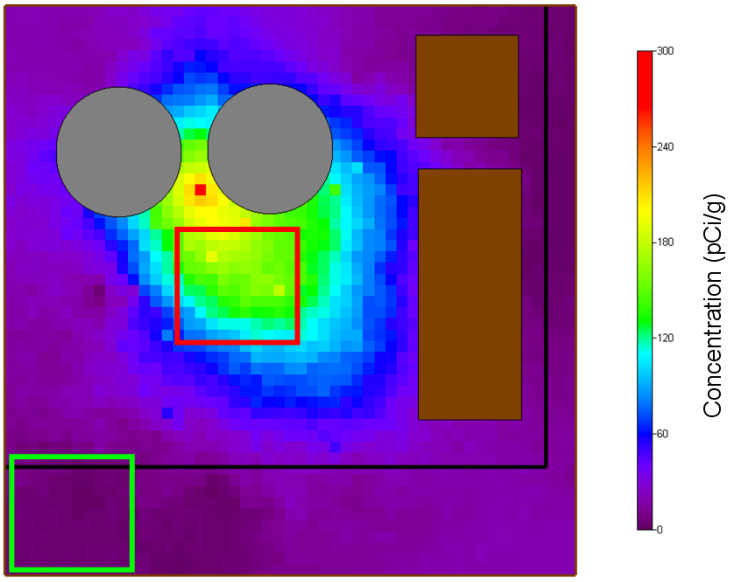


Figure 3.3. Exposure unit E′ compliance as a function of spatial location. E′ in green passes compliance while E′ in red fails.

(Δx,Δy,Δz). Because some sites have an irregular shape only those grid cells whose center lies within the site are considered. The surface layer of the 3D grid is shown in Figure 3.4. Valid cells are identified with green blocks (89 total). Hashed blocks represent those nodes failing this condition and are not part of the system.

For any given exposure unit E(v,g,d,DCGL), one centers[26] the unit at each grid node center ($\mathbf{u_i}$) and evaluates whether compliance has been met. This permits investigators to systematically determine whether the exposure unit meets compliance throughout the site by iteratively positioning and evaluating the unit at each node. For the purpose of discussion, grid nodes are enumerated by first

[26] For irregularly shaped exposure units, the center may be ambiguously defined. Methods such as the mean center (Sullivan and Unwin, 2003) are available in this regard. It is up to the investigators to define the center for their exposure units Additionally, it is possible for parts of an exposure unit to fall outside the study area. There are reasons why this may be desired (e.g. contaminant occurs naturally in background) and the GEM permits investigators to choose.

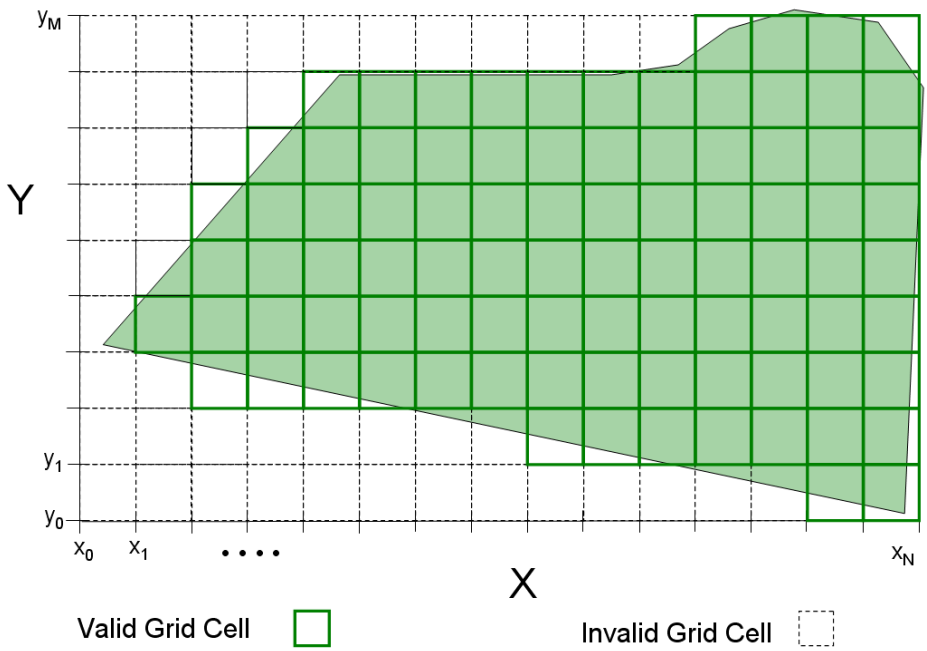


Figure 3.4 The GEM spatial resolution grid.

Each node within the GEM spatial grid is identified by the center of the grid cell with a spatial coordinate **u=** (x,y,z).

assigning an index value of 1 to the westernmost cell located in the southernmost row in the surface layer. Indices are increased by one cycling first on x, then y, and then z (Deutsch and Journel, 1992). One way to conceptualize this is as a moving decision window (Brakewood, 2000).While this approach may support compliance assessment, another approach is taken that provides an object-oriented architecture that equivalently supports compliance and facilitates the MrDM and MrsDM modeling that follows.

Object-Oriented Approach

Consider the positioning of a *base* exposure unit E(v,g,d,DCGL) at any grid node $\mathbf{u_i}$ as an *instantiation* E(v,g,d,DCGL,$\mathbf{u_i}$) of the base at that grid node. Instantiation is a coding term found in object-oriented programming (Burke, 2003) that will serve well in this case. In object-oriented programming, one can define a base object with certain characteristics. Programmers can then create multiple separate copies of that base object and use them as needed in the program flow. These copies are called instances or instantiations of the base object.

In the case of the RLR and Figure 3.5, an exposure unit instance would be placed at all 89 grid cells. Figure 3.6 shows 4 of these exposure instances (#3, #18, #20, and #40).

The focus of the RLR is therefore on these groups of *instantiated* exposure units referred to here collectively as exposure units *collections*. Let $\xi^{v,g}_{d,DCGL}$ represent a collection of K exposure units $\{E(v,g,d,DCGL,u_1),...,E(v,g,d,DCGL,u_K)\}$ where each is an instantiation of the base E(v,g,d,DCGL) at K grid nodes $\mathbf{u}_i = (x_i, y_i, z_i)\ \ i = 1,...,K$. Let the vector $\{c_{true}(\mathbf{u}_1),\ ...,c_{true}(\mathbf{u}_K)\}$ represent the true (but unknown) average concentrations within each exposure unit instance at $\mathbf{u_i}$ in the class $\xi^{v,g}_{d,DCGL}$. Under RLR, the instance E(v,g,d,DCGL,$\mathbf{u}_i$) is in compliance if Eq. 3.1 is satisfied.

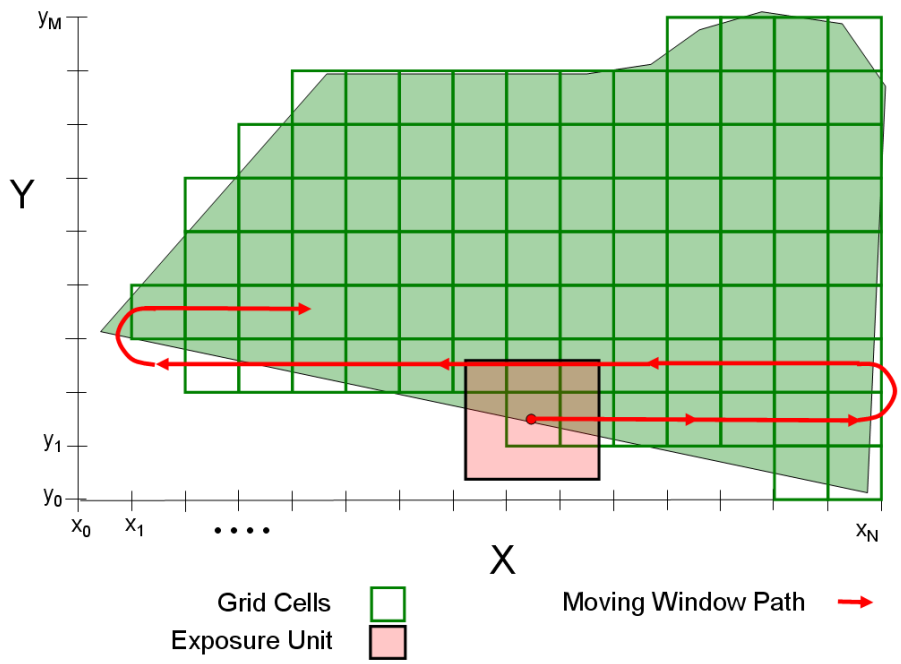


Figure 3.5 A moving window view of exposure unit compliance.

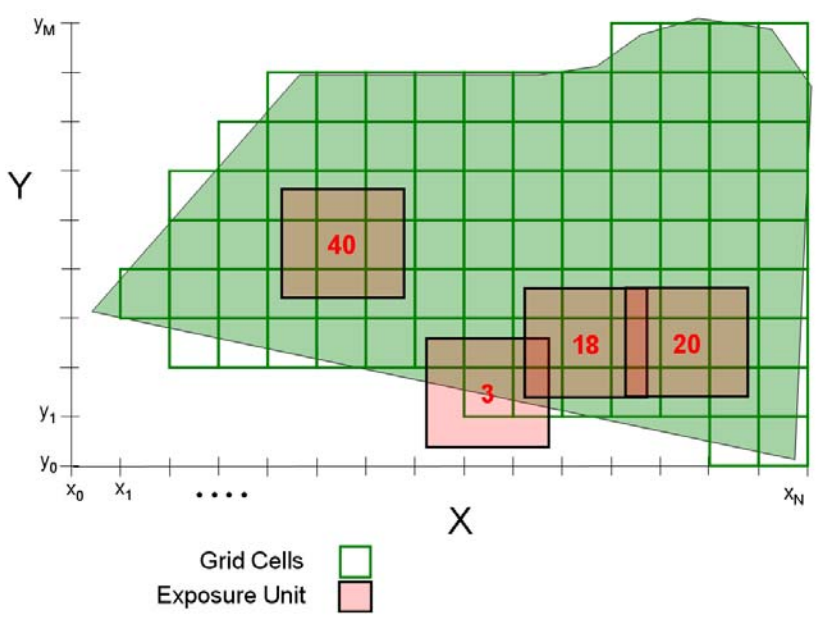


Figure 3.6 Four of the eighty nine exposure unit instantiations are shown.

$$pr\{c_{true}(\mathbf{u}_i) > DCGL\} \leq \alpha \qquad \text{(Eq. 3.1)}$$

Under RLR, the entire collection $\xi^{v,g}_{d,DCGL}$ is in compliance if Eq. 3.2 is satisfied.

$$Compliance[\xi^{v,g}_{d,DCGL}] = \begin{cases} True & if \;\; pr\{c_{true}(\mathbf{u}_i) > DCGL\} \leq \alpha \;\; for \;\; i = 1,\ldots, K \\ False & otherwise \end{cases} \qquad \text{(Eq. 3.2)}$$

Recall that a range of base exposure unit sizes and shapes is possible under RLR. Therefore, for a site to meet regulatory compliance under the RLR Eq. 3.2 must be met for every exposure unit collection. In Figure 3.7 there are two exposure unit collections of interest. Instances from the larger collection, previously shown in Figure 3.6, are now joined by selected instances from the smaller exposure unit collection (#8, #28, #33, #56) in Figure 3.7.

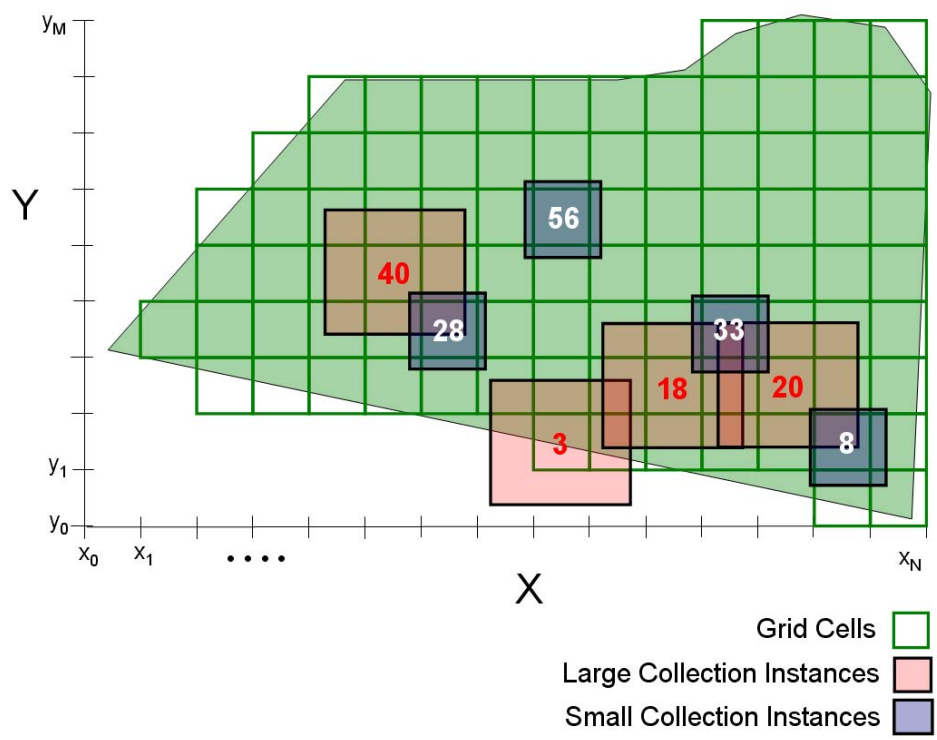


Figure 3.7. Selected instances of two exposure unit classes are shown.

Figure 3.7 provides an illustration of how these classes overlap each other providing a spatial topological integration that leads to important methods under MrDM.

Let Ω represent the *set* of N exposure unit collections $\left\{\xi^{v,g}_{d,DCGL}(\mathbf{u})^1,\ ...,\ \xi^{v,g}_{d,DCGL}(\mathbf{u})^N\right\}$ derived from the set of exposure unit bases $\{E(v_1, g_1, d_1, DCGL_1),\ ...,\ E(v_N, g_N, d_N, DCGL_N)\}$.[27] The RLR formally defines subsurface compliance for site **S** as follows in Eq. 3.3:

$$Compliance[\mathbf{S}] = \begin{cases} True & if\ \ Compliance[\xi^{v,g}_{d,DCGL}] = True\ \forall\ \ \xi^{v,g}_{d,DCGL} \in \Omega \\ False & otherwise \end{cases} \qquad \text{(Eq. 3.3)}$$

While Eq. 3.3 explicitly defines compliance, it does not indicate how the probabilities in Eq. 3.1 and 3.2 may be calculated. The SCSM defines how these probabilities will be calculated based on geostatistical simulations and serves as the test for compliance with Equation 3.3.

The Stochastic Conceptual Site Model Test

As mentioned above, a conceptual site model is the foundation of the Triad model and strongly emphasized by guidance such as MARSSIM. A conceptual site model captures various important aspects about the circumstances regarding the site and contamination, including exposure, buildings, etc. More than one conceptual site model is possible as not all types of information can be conveyed in a single model. Under the GEM, the test for compliance with the RLR is conducted using a stochastic conceptual site model which spatially delineates the probability of complying with the RLR. The term stochastic is used to refer to

[27] Note that subscripts for base exposure unit arguments move into the superscript position for the entire exposure unit classes rather than appearing as subscripts there as well. This is done to reduce the growing notational complexity.

the fact that the SCSM is based on geostatistical simulations of concentrations across the site. Indeed as discussed in Chapter 2, geostatistical simulations provide the means to estimate the probabilities indicated in Eq. 3.1 and 3.2. Simulations are themselves based on an underlying grid system that must be at least as fine as the GEM spatial resolution grid. To simplify the discussion, the simulation grid is assumed to be the same as the GEM grid.

A remedial unit is the decision scale for the actual removal or remediation of soil. For example, one might consider a remedial unit to be as small as a backhoe scoop. As with exposure units, investigators will discretize the site into a three dimensional grid system where each cell represents a single remedial unit[28]. In the interest of clarity, this discussion will assume that the remedial unit grid and the GEM spatial resolution grid are the same although this is not required.[29]

Let RU_i represent the ith remedial unit in the spatial resolution grid. A remedial scenario occurs over a spatial domain (S) when a set of remedial units $\Theta(S) \equiv \{RU_1, \ldots RU_k\}$ is selected from the full set of remedial units $\{RU_1, \ldots RU_N\}$ for remediation. The SCSM can be updated to consider either actual remediation or simulated remediation (e.g. by MrDM) by setting the remedial unit concentration value equal to a specified post-remediation value[30] (prv) for every realization, specifically:

$$c^{(q)}(RU_i) = prv \quad for\, q = 1, \ldots, Q \;\; if \;\; RU_i \in \Theta(S) \qquad \text{(Eq. 3.4)}$$

Suppose there are Q geostatistical simulations of remedial unit concentrations over the study area S and a remedial design $\Theta(S)$.[31] Algorithm 3.1 specifies how the simulations are processed to develop the SCSM.

---

[28] The remedial unit grid must be no finer than the simulation grid with cells small enough to be contained within the smallest exposure unit under consideration.
[29] Depending on the size of the remedial unit, investigators may wish to use a finer spatial grid to adequately model the change of support from sample size to remedial unit size.
[30] For example, one may choose 0pCi/g for a post-remediation value.
[31] If no remedial design is available then $\Theta(S)$ is empty.

Algorithm 3.1: The SCSM Test Algorithm.

1. Select the ith exposure unit class $\xi_{d,t}^{v,g}(i) \in \Omega$, and set k = 1.
2. Select the kth exposure unit instance in that class.

$$E(v,g,d,DCGL,\mathbf{u}_k) \in \xi_{d,DCGL}^{v,g}(i)$$

3. Post process the set of Q geostatistical simulations within the exposure unit space defined by E(v,g,d,DCGL,$\mathbf{u}_k$) to generate each realization average at u_k (Goovaerts, 1997):

$$c_{ave}^{(q)}(\mathbf{u}_k) = \frac{\sum_{i=1}^{N} c^*(\mathbf{u}_i)}{N}$$

where

$$c^*(\mathbf{u}_i) = \begin{cases} c^{(q)}(u_i) \ if \ RU(u_i) \notin \Theta(S) \\ prv \quad if \ RU(u_i) \in \Theta(S) \end{cases} \qquad \text{(Eq. 3.5)}$$

and

$$\{\mathbf{u}_1,...,\mathbf{u}_N\} \in \mathbf{E}(v,g,d,DCGL,\mathbf{u}_k)$$

4. Given the set of Q average concentration values $\{c_{ave}^{(1)}(\mathbf{u}_k),...,c_{ave}^{(Q)}(\mathbf{u}_k)\}$, compute the probability that the true (but unknown) average concentration $\mathbf{c}_{true}(\mathbf{u}_k)$ is less than the DCGL and assign to grid node u_k. From Eq. 3.1 (Goovaerts, 1997, 1999, 2001; Demougeot-Renard et al. 2004; Saito et al. 2004.):

$$prob\{c_{true}(\mathbf{u}_k) \succ DCGL\} = \frac{\sum_{q=1}^{Q} \begin{Bmatrix} 1 & if \ c_{ave}^{(q)} \succ DCGL \\ 0 & if \ c_{ave}^{(q)} \leq DCGL \end{Bmatrix}}{Q} \qquad \text{(Eq. 3.6)}$$

5. Add 1 to k and repeat Steps 2-4 for each exposure unit instance in $\xi_{d,t}^{v,g}(i)$ to form a complete 3D raster model of probability values, $RM[\xi_{d,t}^{v,g}(i)]$, for the ith collection.

6. Add 1 to i, and return to Step 1 repeating the algorithm until all exposure unit classes have been visited. The set of 3D raster models forms the stochastic conceptual site model given by Eq. 3.7.

$$SCSM[\Theta] \equiv \left\{ RM[\xi_{d,t}^{v,g}(1)], ..., RM[\xi_{d,t}^{v,g}(N)] \right\} \qquad \text{(Eq. 3.7)}$$

Simply stated, the SCSM is a set of 3D raster models reporting the probability that an exposure unit instance centered at every grid node will fail compliance for every exposure unit class.

The SCMS model then demonstrates compliance against Eq. 3.2 if the set of all nodes **u** in SCSM have a probability value less than α. Specifically:

$$P(\mathbf{u}) \leq \alpha \quad \forall\, \mathbf{u} \in RM[\xi_{d,t}^{v,g}]\ \forall\, RM[\xi_{d,t}^{v,g}] \in SCSM[\Theta] \qquad \text{(Eq. 3.8)}$$

Compliance evaluation could be accomplished by the SCSM process returning a simple yes/no answer. A great deal more can be reported using this simple answer. For example, it will be useful to determine the severity of compliance failure. Severity can be expressed graphically by plotting for each class the number of exposure unit instances that fail compliance (Figure 3.8).

In addition, one can produce a 3d volume indicating where exposure unit instances are failing compliance for a particular exposure unit class. For those exposure unit positioning grid nodes who have a probability of failing compliance greater than alpha = 0.1, the following image in Figure 3.9 is produced.

Roadmap

Regulatory guidance such as MARSSIM often includes a *roadmap* section that quickly summarizes the methods of interest. While derivation was involved, the input requirements by the regulator are reasonably simple. Table 3.2 summarizes the five inputs to the RLR.

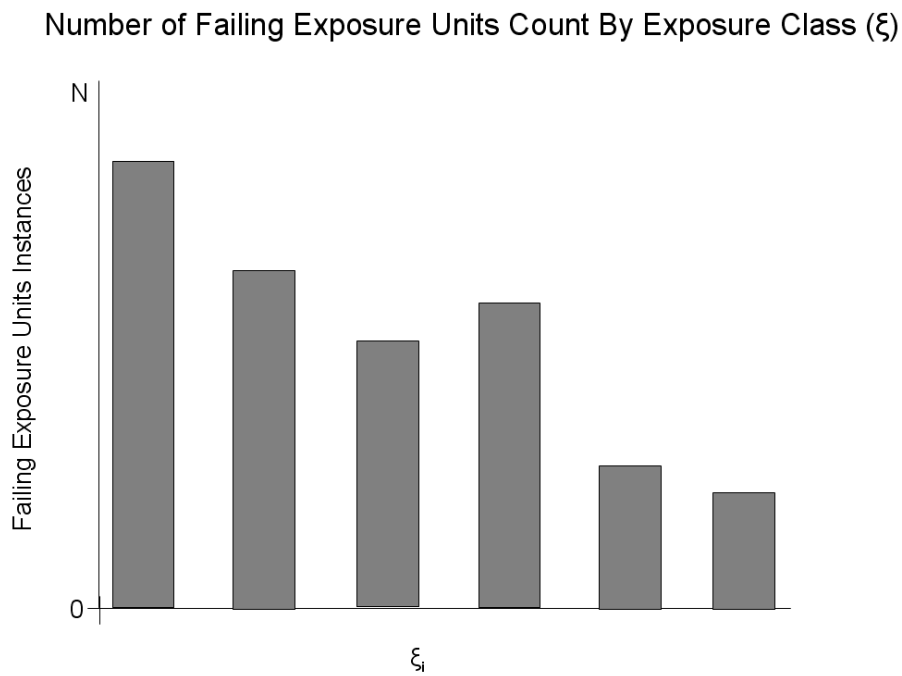


Figure 3.8. Site specific calculation of the number of failing exposure units within each class.

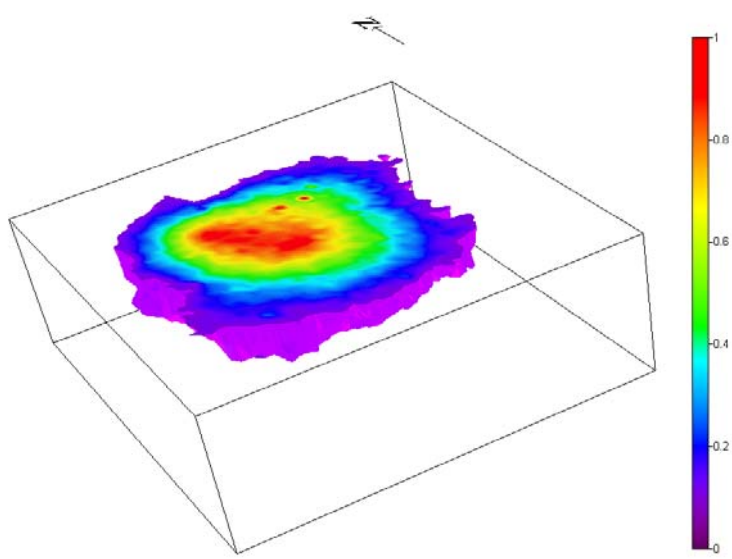


Figure 3.9 Volume of grid points locations where the corresponding exposure unit instance demonstrated a probability greater than alpha = 0.1.

Table 3.2 Input requirements for RLR/SCSM

Description	Symbol
Base Exposure Units	E(v,g,d,DCGL)
Spatial Boundaries of the Site	S
GEM Spatial Grid System	G
Probability limit	α
Geostatistical Simulation Set	Q

For those sites that fail compliance, there is interest in developing a remedial design through replacement of realizations with post-remediation values. Furthermore, the goal is to identify the smallest remedial volume possible to achieve this compliance. The MrDM approach estimates this minimum design by simultaneously considering all exposure unit instances and the spatial overlaps between them.

The Multi-scale Remedial Design Model

When a site fails compliance, investigators may wish to know what soils to remediate to bring the site into compliance: in particular, the minimum volume of contaminated soil that must be removed, replaced, or cleaned to move the site into compliance with respect to the RLR. In addition, investigators will want to know where this volume is positioned on the site.

The goal is to identify a set of remedial units $\Theta_{MIN}(S)$, such that the cardinality[32] $|\Theta_{MIN}(S)|$ is minimized subject to Eq. 3.8. In other words, $SCSM[\Theta_{MIN}(S)]$ must be in compliance with respect to the RLR. Eq. 3.9 formally states this goal.

[32] Cardinality means the number of elements in the set.

$$\text{Identify a remedial design } \Theta_{MIN}(S) \text{ with minimum } |\Theta_{MIN}(S)| \ \forall\ \Theta(S) \text{ such that } P(\mathbf{u}) \le \alpha \quad \forall\, \mathbf{u} \in RM[\xi_{d,t}^{v,g}]\ \forall\, RM[\xi_{d,t}^{v,g}] \in SCSM[\Theta(S)] \qquad \text{(Eq. 3.9)}$$

Notice that there is no requirement or expectation by the GEM that the solution to Eq. 3.9 is unique. For example, there may be two remedial designs $\Theta_1(S)$ and $\Theta_2(S)$, each having the minimum number of remedial units. From a compliance perspective, no rule for deciding between them is offered since both lead to compliance, the primary goal. Other additional considerations may offer means to choose between them. For example, from an economic standpoint, one may be more cost effective to implement than another. For example, $\Theta_2(S)$ may include deeper subsurface remedial units that require more effort to access than $\Theta_1(S)$. Assigning a cost as a function of depth is not a straightforward solution. Suppose that two units are tied with respect to the minimization objective. One is at the surface and one is at depth. It is not necessarily true that the unit at depth incurs greater cost to extract. It depends on whether remedial units situated above it might be removed (in the final solution) that expose it for easy extraction. This would require knowing the final solution (at least in part) before it is developed. Therefore, assigning a cost a priori would be problematic at best. This complex issue is not taken up in this dissertation.

Minimizing the number of remedial units required for the SCSM to demonstrate compliance can be a computationally challenging objective if approached in a brute force manner. Consider a situation where the set of failing exposure units instances includes only 20 remedial units. Investigators wish to know the minimum number of remedial units to choose from in order to bring these failing instances (and the site) into compliance. In a brute force approach, one would determine if any 1 of the 20 remedial units would lead to compliance. If not, then pairs of units would be considered and so forth. At any given stage, this amounts to a combinatorics problem where one has N possibilities from which to select k

combinations, the value of which is provided by the following (Brockett and Levine, 1984):

$$\binom{N}{k} = \frac{\mathrm{N!}}{(\mathrm{N-k})!} \qquad \text{(Eq. 3.10)}$$

In the small case of examining 10 units out of 20, the number of permutations under which k= 10 could occur is 6.7 x 10^{11}. The task is considerably more daunting when one recalls that for each of the 6.7 x 10^{11} possibilities, Algorithm 3.1 would need to be calculated.[33]

Two alternative approaches for selecting remedial units within a *single exposure unit instance* have been identified in the literature. Saito and Goovaerts (2003) select remedial units (within a single exposure unit) by first identifying the RU, that when remediated, produces the greatest reduction in the risk of compliance failure. Once identified, a second remedial unit is identified that maximizes further risk reduction in compliance failure and then a third, and so forth until the risk falls below a risk limit. Three problems exist with applying this approach to multiple and topologically integrated exposure units.

First, there is no clear method for extending the approach from a single exposure unit to multiple, topologically integrated units. This complication is best explained by example. For a single remedial unit involved in N failing exposure unit instances, the unit provides N probability reductions. Indeed, many remedial units will be in this situation. One could choose the remedial unit whose vector of probability reductions is greatest. However, this selection only provides the greatest remedial benefit to the one exposure unit to which it applies. The goal of minimizing the global remedial design over all exposure units would not be part of this selection process.

[33] In the upcoming example, there are over thousands of remedial units to consider.

One could modify the algorithm to better suit the current situation by selecting that remedial unit which produces the greatest reduction in the sum of exposure unit probabilities that fail compliance. Two major obstacles stand in the way of this approach. For a modestly sized problem in three dimensions, the number of remedial units may be quite large. For example, in the upcoming example (Chapter 6), over 3500 remedial units are found within non-compliant exposure unit instances. This means that for the first remedial unit to be identified, Algorithm 3.1 would need to be executed for each remedial unit in this set (with simulated values replaced by post-remediation values for each unit) 3500 times. Identification of the second remedial unit would require processing Algorithm 3.1 3500 times. Keeping in mind that each execution of Algorithm 3.1 requires the processing of a potentially large set of simulations, the computational demand makes this an unattractive property of the approach.

Finally, it is possible to reach a point where no single RU when remediated would reduce the probability of compliance failure for any exposure unit instance. A simple demonstration of this is presented in Figure 3.10 for a single 1m x 4m rectangular exposure unit. Here only two geostatistical realizations of node values are used in order to simplify the discussion. Note that no remediation of an *individual* RU (noted in blue) leads to a reduction in the probability that the exposure unit instance exceeds 3pCi/g and the algorithm breaks down.

Another approach found in the literature (Brakewood et al. 2008; Stewart et al. 2009; Stewart and Purucker 2011) first estimates remedial unit concentration values based on averaging data within a remedial cell or by some geospatial interpolator (e.g. kriging). The remedial units are then sorted by concentration value in descending order. The remedial unit with the highest concentration estimate is added to the remedial design and the average is compared to the DCGL. If the average is greater than the DCGL, then the next highest remedial unit is added and so forth until the average is sufficiently low. This approach as

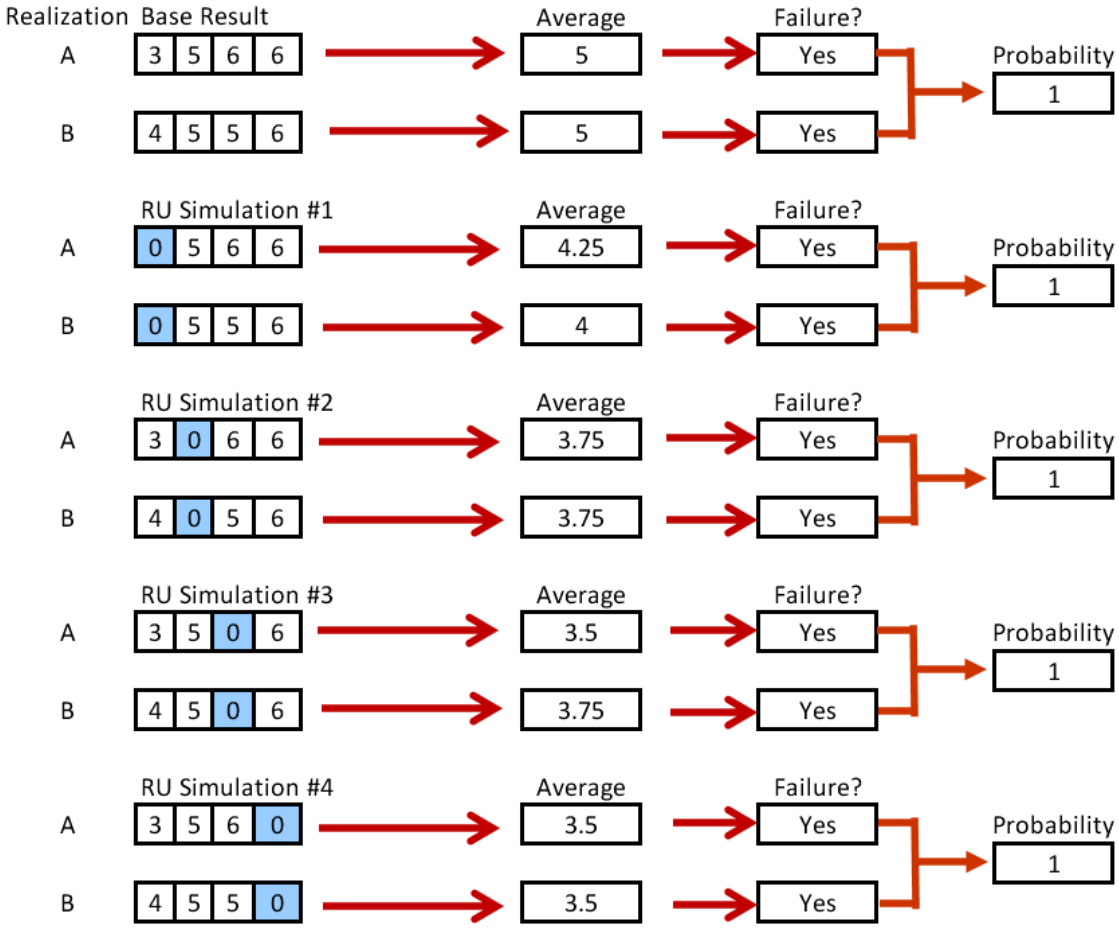


Figure 3.10 Scenario where no single RU improves compliance probability.

implemented in those works does not consider the uncertainty in the exposure unit average (Saito and Goovaerts, 2003).[34]

Consider the following approach developed as a building block for the MrDM algorithm called the *local remedial design*. This design is a *hybrid* of the two former methods but remains applicable only to an individual exposure unit instance. The local remedial design is written as $\Theta_{Local}(E)$. The first step in this approach is to estimate the remedial unit concentration values within an exposure unit by averaging the geostatistical realizations within each remedial unit cell. Like Brakewood (2000) and Stewart and Purucker (2011) these values are then sorted in descending order. But unlike these former works, remedial units are sequentially included in the remedial design until the *probability* that the exposure unit instance's true (but unknown) average concentration is less than α according to Eq. 3.5 and 3.6. Hence the local remedial design is a hybrid of these two approaches permitting consideration of the uncertainty in the exposure unit concentration but guaranteed not to break down, since the method under which remedial units are added is clear even when no immediate reduction in the probability may be observed in a given iteration. Furthermore, the targeted sort design also considers any existing remedial activities or plans that have or will occur; it is expressed as $\Theta_{local}(E_i \mid \Theta)$. In these situations, average values for those remedial units included in the remedial design will first be replaced by the post-remediation value prior to sorting as in Eq. 3.4. The local remedial design is formalized in Algorithm 3.2.

[34] In the implementation by Stewart and Purucker (2011), uncertainty at the remedial unit level is incorporated by permitting the user to choose an upper percentile from the ccdf rather than the mean. This is different than considering the uncertainty in the exposure unit average concentration and relies on local uncertainty (e.g. kriging variance) rather than joint uncertainty, a problem which was previously discussed in Chapter 2.

Algorithm 3.2: Local Remedial Design

1. For any exposure unit instance failing compliance, E_i, estimate remedial unit concentration values by averaging the set of geostatistical realizations for each remedial unit according to Eq. 2.5.
2. If an existing remedial design, Θ, is in place, replace the average of those remedial units within E_i that are part of Θ with the post-remediation value as in Eq. 3.4.
3. Sort these remedial unit concentration estimations in descending order. Set k = 1.
4. Add the kth remedial unit to the local remedial design $\Theta_{local}(E_i \mid \Theta)$.
5. Calculate the probability that E_i fails compliance according to Eq. 3.5 and 3.6.
6. If compliance is failed, set k = k + 1 and repeat 4-5 until compliance is met. The resulting set $\Theta_{local}(E_i \mid \Theta)$ is the local remedial design.

Like the previous methods, the local design works well for a single exposure unit but is not yet suited for direct application to multiple exposure units. In the case of multiple and overlapping exposure units, direct application of Algorithm 3.2 produces a *naïve design.* The design $\Theta_{local}(E_i \mid \Theta)$ is considered naïve when it does not recognize the fact that other remedial units within E_i might be included in a separately executed remedial design $\Theta_{local}(E_k \mid \Theta)$ due to the sharing of remedial units between E_i and E_k. In this situation, the notation $\Theta_{naive}(E_i \mid \Theta)$ is used rather than $\Theta_{local}(E_i \mid \Theta)$ to emphasize this lack of information.

The following scenario considers two exposure unit instances (E_1 and E_2) and a single geostatistical realization[35] shown in Figure 3.11.[36]

---

[35] Throughout this dissertation, when examples are presented, it will be common to show only a single geostatistical realization to demonstrate a point. Showing many would take up an unacceptable amount of space and would not add any value to the explanation.
[36] To simplify the discussion, no existing remedial design is in play.

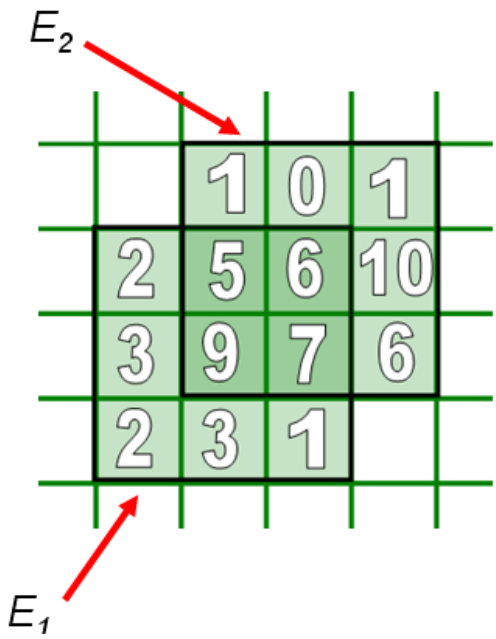


Figure 3.11 Two overlapping exposure unit instances with compliance failures.

With a DCGL value of 4.1pCi/g, each instance is experiencing a compliance failure. The lower left presents an average concentration realization of 4.2 pCi/g and the upper right a realization greater than 5.0pCi/g. With only a single realization, the probability of exceeding a DCGL of 4.1pCi/g is 1 for each unit and a remedial design is required. Figure 3.12 graphically illustrates the application of Algorithm 3.2 for each exposure unit instance, and the global remedial design, $\Theta(S)$, formed by a union of local naïve designs $\Theta_{naive}(E_1)$ and $\Theta_{naive}(E_2)$ together.

The resulting global design result includes two remedial units. Had the remedial design for E_1 been taken into consideration when developing E_2 under Algorithm 3.2, only one remedial unit would have sufficed in bringing both units into compliance as demonstrated in Figure 3.13.

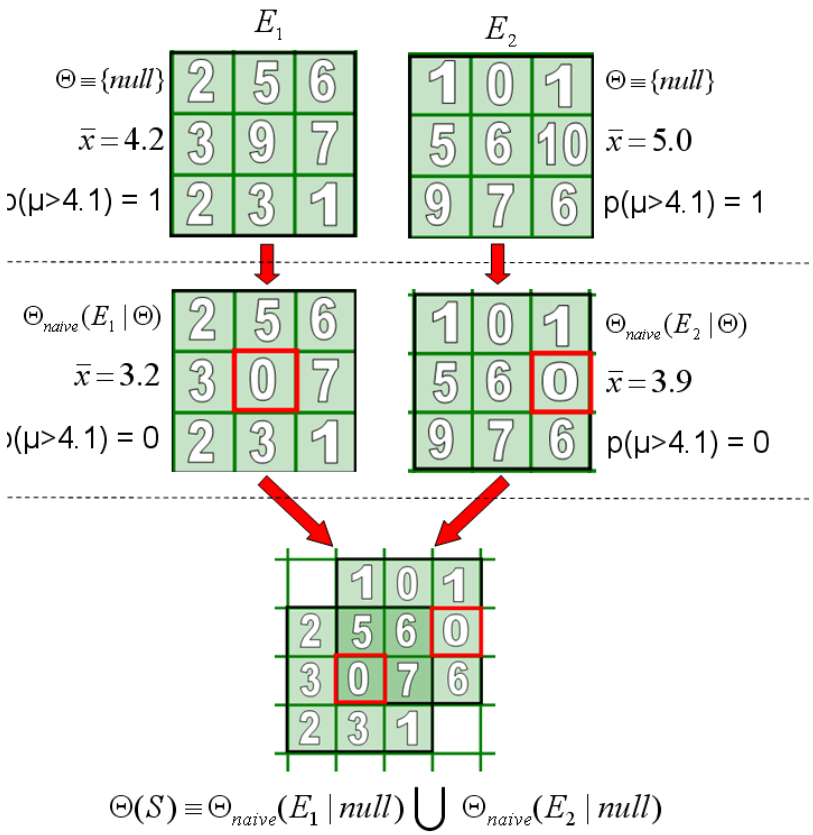


Figure 3.12 Demonstration of Algorithm 3.2. Remedial units slated for remedial action are highlighted in red.

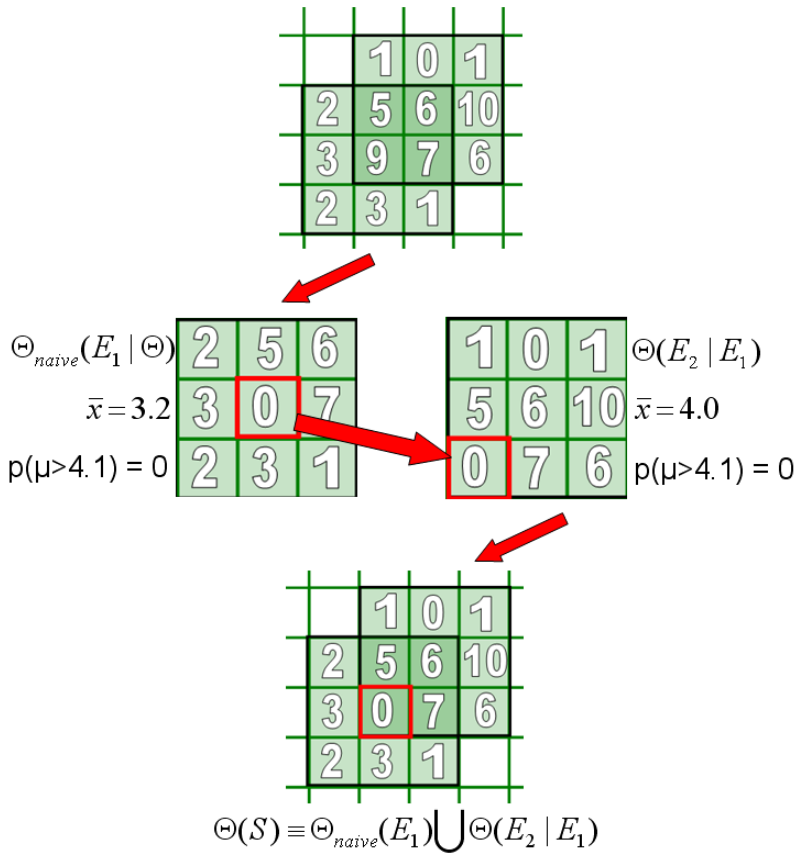


Figure 3.13 Global remedial design is reduced when concurrently considering the remedial designs of local exposure unit instances.

In this situation, the remedial design for E_2 is further *conditioned* by naïve remedial activities imposed by E_1:

$$\Theta(E_2 \mid E_1) \equiv \Theta_{local}[E_2 \mid \Theta, \Theta_{naive}(E_1 \mid \Theta)] .$$

Note however, that if the roles were reversed $\Theta_{naive}(E_2) \bigcup \Theta(E_1 \mid E_2)$, calculation reveals no improvement in the union of naïve designs shown in Figure 3.12 . The challenge then is to strategically choose conditioning exposure units that lead to fewer remedial units in the final design. For a large number of exposure unit instances, a brute force approach leads to the same combinatorics problem encountered previously in Eq. 3.10.

Instead, MrDM approaches this minimization problem by first identifying a feasible solution to the minimization problem and then seeking to minimize that first solution.[37] Let the *baseline remedial design,* $\Theta_{baseline}(S)$, be formed by union of the set of naïve exposure unit remedial designs, $\Theta_{naive}(E_i)$ conditioned only by an existing remedial design, Θ. In other words, exposure unit instances operate unaware of each other's remedial design plans given Θ.

$$\Theta_{baseline}(S) \equiv \bigcup_{i=1}^{N} \Theta_{naive}(E_i \mid \Theta) \qquad \text{(Eq. 3.11)}$$

Any additional remedial units added to this baseline design would be superfluous as every single exposure unit already passes compliance by definition (Algorithm 3.2). Hence the cardinality $|\Theta_{baseline}(S)|$ provides a reasonable *upper bound* for the minimize value $|\Theta_{MIN}(S)|$ and the initial estimate for MrDM, $\Theta_{MrDM}(S)$. The goal is then to maximize the difference between the baseline and final solution for MrDM stated by Eq. 3.12

$$\Delta_{MAX} = Maximum \;\; |\Theta_{baseline}(S)| - |\Theta_{MrDM}(S)| \qquad \text{(Eq. 3.12)}$$

[37] This is a common numerical approach used in minimization problems. For example, see Gass (1985, p.239).

Maximizing this difference amounts to identifying the right set of conditioning exposure units such that the baseline remedial design is reduced. Unfortunately, approaching the problem in a brute force manner once again results in the same combinatorics problem as previously discussed with global selection of remedial units (Eq. 3.10).

Instead, MrDM approaches the selection process by strategically adding local remedial designs to the global design over a series of stages. At each stage, the remedial design of the exposure unit instance that serves best as a conditioning design (in reducing the baseline) is added to the global remedial design. Remedial units added in this manner are considered *"optimal"* for the current stage. The optimal remedial units at the jth stage, Θ_j, serves as the existing remedial design in the search for the next best local conditioning design in the j+1 stage. At any new stage j+1, the local design whose addition to the optimal remedial units results in the greatest decrease in the baseline becomes part of the optimal set. The full estimate for the remedial design Θ_{MrDM} at any stage is comprised of the optimal remedial units plus those remedial units contributed by the naïve application of Algorithm 3.2 to any remaining, non-compliant exposure units. Hence at any stage j, the jth estimate of MrDM is comprised of two parts: 1) optimal and 2) naïve. As the stages progresses, the portion that is optimal increases and the naïve part decreases until the entire design is optimal. Figure 3.14 demonstrates this for 3 stages of development.

This is formally stated as follows. Let Θ_j represent the set of optimal remedial units at stage j. At the j+1st stage, let E_k represent the next local remedial design whose addition to the optimal set, produces the greatest reduction between theta MrDM and theta baseline. The Θ_{MrDM}^{j+1} is then given by Eq 3.13.

$$\Theta^{0}_{MrDM} \equiv \{\underbrace{RU_1, \ldots, RU_k}_{\Theta_{naive}[E_1]}, \underbrace{RU_{k+1}, \ldots, RU_m}_{\Theta_{naive}[E_2]}, \underbrace{RU_{m+1}, \ldots, RU_h}_{\Theta_{naive}[E_3]}\} \equiv \Theta_{baseline}$$

$$\Theta^{1}_{MrDM} \equiv \{\underbrace{RU_1, \ldots, RU_k}_{\Theta_1}, \underbrace{RU_{k+1}, \ldots, RU_m}_{\Theta_{naive}[E_1 \mid \Theta_1]}, \underbrace{RU_{m+1}, \ldots, RU_n}_{\Theta_{naive}[E_2 \mid \Theta_1]}\} \qquad n \le h$$

$$\Theta^{2}_{MrDM} \equiv \{\underbrace{RU_1, \ldots, RU_k, RU_{k+1}, \ldots, RU_t}_{\Theta_2}, \underbrace{RU_{t+1}, \ldots, RU_v}_{\Theta_{naive}[E_1 \mid \Theta_2]}\} \qquad v \le n$$

$$\Theta^{3}_{MrDM} \equiv \{\underbrace{RU_1, \ldots, RU_k, RU_{k+1}, \ldots, RU_l, RU_{l+1}, \ldots, RU_i}_{\Theta_3}\} \qquad i \le v$$

Figure 3.14 The iterative construction of MrDM.

$$\Theta^{j+1}_{MrDM} = \Theta_j \cup \Theta_{\text{naive}}(E_k \mid \Theta_j) \bigcup_{k \neq kk} \Theta_{naive}[E_{kk} \mid \Theta_j \cup \Theta_{\text{naive}}(E_k \mid \Theta_{j-1})]$$

where $\Theta_{\text{local}}(E_k)$ produces the greatest remedial reduction at stage $j+1$ (Eq. 3.13)

$$\Delta^{j+1}_{MAX} = \left|\Theta_{\text{baseline}}\right| - \left|\Theta^{j+1}_{MrDM}\right|$$

MrDM is a *heuristic estimate* to the optimal solution since there is no guarantee that a series of stage-specific optimal solutions produces the true optimal solution to the problem.[38]

[38] This method is similar to the greedy *add solution* to the P-median location problem (Miller and Shaw, 2001).

Eq. 3.13 is well suited to an algorithmic approach. Let $\mathbf{E}_{fail}(j)$ represent the set of exposure units failing compliance at the jth iteration. Let Θ_j represent the set of optimal remedial units at the jth stage. The MrDM algorithm is formally given by Algorithm 3.3.

Algorithm 3.3 The MrDM Algorithm

Step 1: Determine initial conditions.

Set the initial remedial design iteration $\Theta_0 = \{null\}$. The next step is to determine the initial set of exposure unit instance failures $\mathbf{E}_{fail}(0)$ according to Algorithm 3.1. If there are no failures, then no remedial design is required and this algorithm terminates with $\Theta_{MrDM} = \Theta_0 = \{null\}$. Otherwise, create the base line remedial design, $\Theta_{baseline}$, using $\mathbf{E}_{fail}(0)$ according to Eq. 3.11 and continue to Step 2 with j = 0.

Step 2: Remediate special case exposure unit instances

If the investigator specifies a base exposure unit structurally equivalent to the remedial unit specification, then any corresponding instance that fails compliance results in mandatory remediation of the remedial unit on which it is situated regardless of what other remedial activities may occur. All remedial units associated with such an exposure unit instance are automatically added to the remedial design producing the next iteration, Θ_1. Using Θ_1, determine $\mathbf{E}_{fail}(1)$ according to Algorithm 3.1 (modified). Continue to Step 3 with j = 1.

Step 3: Begin the j+1st iteration.

If there are no more exposure unit instance compliance failures ($\mathbf{E}_{fail}(j) \equiv \{null\}$), then set $\Theta_{MrDM} = \Theta_j$ and exit Algorithm 3.3. Otherwise, at least one exposure unit is still out of compliance. In this case, set j = j + 1, k = 1, and continue to Step 4.

Step 4: Remediate any isolated exposure unit instances.

As Θ_j is iteratively constructed and more optimal remedial units are added, exposure units will begin moving into compliance. In many cases, this may lead to non-compliant exposure unit instances that no longer share an overlap with any other non-compliant exposure unit instances. In other words, they are spatially isolated cases of compliance failure. For these cases, remediation has no effect on any other exposure unit instance and vice versa. Hence their local remedial designs as defined by Algorithm 3.2 and expressed as $\Theta_{local}(E^{f}_{fail} \mid \Theta_j)$ may be added directly to the remedial design creating the next iteration. Specifically, for F equal to the total number of isolated exposure unit instances set

$$\Theta_{j+1} \equiv \Theta_{naive}(E^{f}_{fail} \mid \Theta_j) \bigcup \Theta_j \quad for\ f = 1,...,F$$

and then return to step 3.

Step 5: Establish the next candidate for addition to the optimal set.

Select the kth exposure instance $\mathbf{E}^{k}_{fail}$ from the set $\mathbf{E}_{fail}(j)$ and compute the conditional remedial design $\Theta_{naive}(E^{k}_{fail} \mid \Theta_j)$. Tentatively add this design to the optimal set of remedial units

$$\widetilde{\Theta}_{j+1} \equiv \Theta_{local}(E^{k}_{fail} \mid \Theta_j) \bigcup \Theta_j$$

Where $\widetilde{\Theta}_{j+1}$ tentatively represents the candidate set of optimal units.

Step 6: Assess remaining exposure instances response to candidate design.

Given this adjustment remedial design, $\widetilde{\Theta}_{j+1}$, assess the adjustment of the remaining exposure unit instances as they respond to this design change

(Algorithm 3.2). Specifically, for all failing instances that topologically overlap[39] E_{fail}^{k} with indices $kk \neq k$ compute $\Theta_{naive}(E_{fail}^{kk} \mid \widetilde{\Theta}_{j})$ and add each one to the tentative candidate for the next MrDM remedial design.

$$\widetilde{\Theta}_{MrDM}^{j+1} \equiv \widetilde{\Theta}_{j+1} \bigcup_{k \neq kk} \Theta_{local}(E_{fail}^{kk} \mid \widetilde{\Theta}_{j+1})$$

For each instance with kk<>k, the local remedial design is still conducted without the knowledge of the other kk<>k exposure units. The only conditioning at this point is the previous iteration and the kth local remedial design.

Step 7: Determine the reduction in the number of remedial units.

$$\Delta^{k} = \left|\Theta_{baseline}\right| - \left|\widetilde{\Theta}_{MrDM}^{j+1}\right| .$$

If $k < \left|\mathbf{E}_{fail}\right|$ then set k = k + 1 and return to step 5.

Step 8: Determine the exposure unit instance whose conditioning results in the greatest reduction of remedial units from the baseline.

$$\Delta_{Max}^{j+1} = Maximum\ \Delta^{k}\ k = 1, \ldots, K$$

The winning unit, E_{winner}, is the E_{Fail}^{k} where $\Delta^{k} = \Delta_{Max}^{j+1}$. If there are ties, then the following tie breaker rules are applied in order until a winner emerges. The winner is determined by:

a. which adds the minimum number of *additional* remedial units,
b. which moves the greatest number of failing units into compliance,
c. which minimizes probability of failure summed over all failing units,
d. or wins in a random draw.

Step 9: Update the remedial designs.

[39] Only those exposure units that overlap the current candidate exposure unit will be affected by any remediation. Hence only those units need to be updated saving a considerable amount of time in the algorithm.

The j+1st optimal remedial unit set, is therefore the union of the winning local design and the jth optimal set design:

$$\Theta_{j+1} \equiv \Theta_j \bigcup \Theta_{local}(E_{winner} \mid \Theta_j)$$

The j+1st MrDM remedial design unions this with the remaining naïve designs.

$$\Theta_{MrDM}^{j+1} \equiv \Theta_{j+1} \bigcup_{k=1,k\neq kk}^{K} \Theta_{naive}(E_{Fail}^{k} \mid \Theta_{j+1})$$

Step 10: Check the stopping rule.

As j increases and more remedial units are added to the optimal remedial design Θ_{j+1}, the pool of remaining remedial units that might be removed from the baseline becomes smaller. Indeed, as the remedial unit set becomes increasingly optimal, at some point, additional iterations might not produce any additional reduction in the baseline. For this reason, MrDM is equipped with a stopping rule. Namely, when the number of consecutive iterations exhibiting a zero reduction in the baseline ($\Delta_{MAX}^{j} = 0$) exceeds a specified limit, the algorithm terminates with $\Theta_{MrDM} \equiv \Theta_j$. If one wished to continue iterations regardless of any consecutive stretch of zero reductions, until all exposure unit have been added, this can be accomplished by setting the limit very high. If the criteria for stopping the algorithm has not yet been made, then continue with Step 9.

Step 11: Update the set of failing exposure units $\mathbf{E}_{fail}$(j+1)*.*

Using Θ_{j+1}, update $\mathbf{E}_{fail}(j+1)$ according to Algorithm 3.1 (modified), set j = j + 1 and return to step 3.

If there are K non-compliant exposure unit instances at any stage j, this will require K implementations of Algorithm 3.2 (Step 6) at that stage. For very large values of K, this results in a computationally intense approach, one that is well suited for parallel computational methods (discussed in the final chapter). On the other hand, the number of iterations j is difficult to determine since the benefit of any remedial design also depends on the number of exposure units that are

moved to compliance. In a worst case scenario, only the E_{winner} exposure unit is moved into compliance in every round. This is unlikely for every iteration however as the example (Chapter 6) demonstrates. Indeed, one can plot for each round the number of remedial units reduced and also the number of exposure units moved into compliance as result of the most recent expansion in the optimal remedial unit set. In addition, as exposure units are moved into compliance, some topological integrations may break down leading to spatially isolated compliance failures discussed in Step 2. These exposure unit instance designs are immediately added to the optimal set and moved into a state of compliance further reducing the size of K. At some point either the stopping rule will engage (Step 8) or K = 0, and the algorithm is complete. The number of iterations is therefore a function of the severity of compliance failures, topological integration, and the limit imposed for the stopping rule.

In some situations, the MrDM may necessarily produce expansive remedial designs that are too costly for the investigation. The remedial design size can be large due to 1) uncertainty in the spatial distribution of the contaminant and 2) a requirement for a high level of certainty (α) or both. One way to reduce the uncertainty in the spatial distribution is to take additional samples. Strategically locating those samples may indeed reduce the remedial design imposed by MrDM. This is the purpose of the MrsDM which is derived in the following discussion.

Multi-scale Remedial Sample Design Model

One way to maintain a high compliance standard, while potentially reducing the volume of unnecessarily remediated soil, is to strategically collect new samples. Within the GEM framework, new samples collected at the proper locations can further refine the spatial behavior of the contaminant and potentially increase the confidence about compliance failure. The value of taking additional samples and

the associated sampling cost must be weighed against the forecast reduction in the global remedial design and associated cost savings. Examples for a single decision criterion over a single exposure area are available in the literature (Pilger et al., 2001; Verstraete and Van Mervenne, 2008; Freeze et al., 1992; England et al., 1992; Demougeot-Renard et al., 2004; Norberg et al, 2006; and Back 2006, 2007). These approaches do not consider multiple-topologically integrated exposure units with a range of different decision criteria. In this dissertation, a new remedial design is developed (MrsDM) to supplement the MrDM algorithm by identifying sample locations that may increase the decision confidence and reduce the remedial design requirements.

MrsDM selects an optimal subset of corehole locations from a larger set of proposed locations, by forecasting the benefit these locations may provide in reducing the remedial design imposed by MrDM. A corehole here is defined as a set of sample locations taken at the same x/y coordinate at different depths. Corehole notation is given by $w_i \equiv \{(x', y', z_1),...,(x', y', z_j)\}$ where j is the number of vertical samples taken at (x', y'). Specifically, the MrsDM objective is to select a subset of coreholes $\mathbf{K} \equiv \{w_1,...,w_k\}$ from set $\mathbf{N} \equiv \{w_1,...,w_n\}$ that maximizes

$$\left|\Theta_{\mathrm{MrDM}}(S)\right| - \left|\Theta_{MrDM}(S \mid \mathbf{K})\right| \qquad \text{(Eq. 3.14)}$$

As with the brute force selection of remedial units, the brute force solution to Eq. 3.14 could require examination of a very large set of combinations (Eq. 3.10). For each combination, the MrDM algorithm would be run (Algorithm 3.3), itself a computationally intense operation.

MrsDM begins with the set of proposed locations **N** and the requested subset of size k and determines the optimal set k by selective removal of locations from N such that the impact on the total reduction in remedial design is minimal.[40]

---

[40] This is similar in nature to the drop algorithm approach to the P-Median solution (Miller and Shaw, 2001).

MrsDM takes advantage of the property that simulation outcomes for any remedial unit cell is more influenced by nearby data points than by those further away due to the presence of spatial auto-correlation. Based on this guiding principle, MrsDM selectively removes proposed sample locations from the design by observing their *local performance* in reducing the design. Remedial cells removed from the remedial design following the addition of new locations are changed to reflect the nearest newly proposed location. The assumption is that this change in status is due *largely* to the presence of the nearest new location. It is important to note that no claim is made that any remedial cell changes its status *exclusively* due to addition of the nearest proposed location. Exclusivity is known not to be true as more than one new sample location may be involved in the re-estimation. The assumption that the nearest new location is primarily responsible is based on the fact that greater weight is given to this new proposed location during simulation than any other new location. Remedial areas that are closer to a proposed sample location than others are in that new location's *Voronoi/Thiessen region* (Sullivan and Unwin, 2003). Note that in the presence of spatial anisotropy (correlation is stronger in one direction than in another) the Voronoi/Thiessen regions should be adjusted to reflect this. This is accomplished by transforming geographic space into an isotropic space through the use of rotational matrices indicated by the spatial auto-correlation model (Deutsch and Journel, 1992). Once the Voronoi areas are in place, the task then is to identify those remedial units that change their remedial status within each new sample location's Voronoi region.

An example may prove useful at this point. Suppose 13 existing coreholes result in the MrDM remedial design shown as a gray set of remedial cells in Figure 3.15(a). Suppose that investigators wish to choose the best 2 out of 3 proposed locations shown in Figure 3.15(b) along with their volumes of influence (VOI). Values for these three proposed locations are estimated and added to the full data set resulting in a new set of geostatistical realizations and a new MrDM

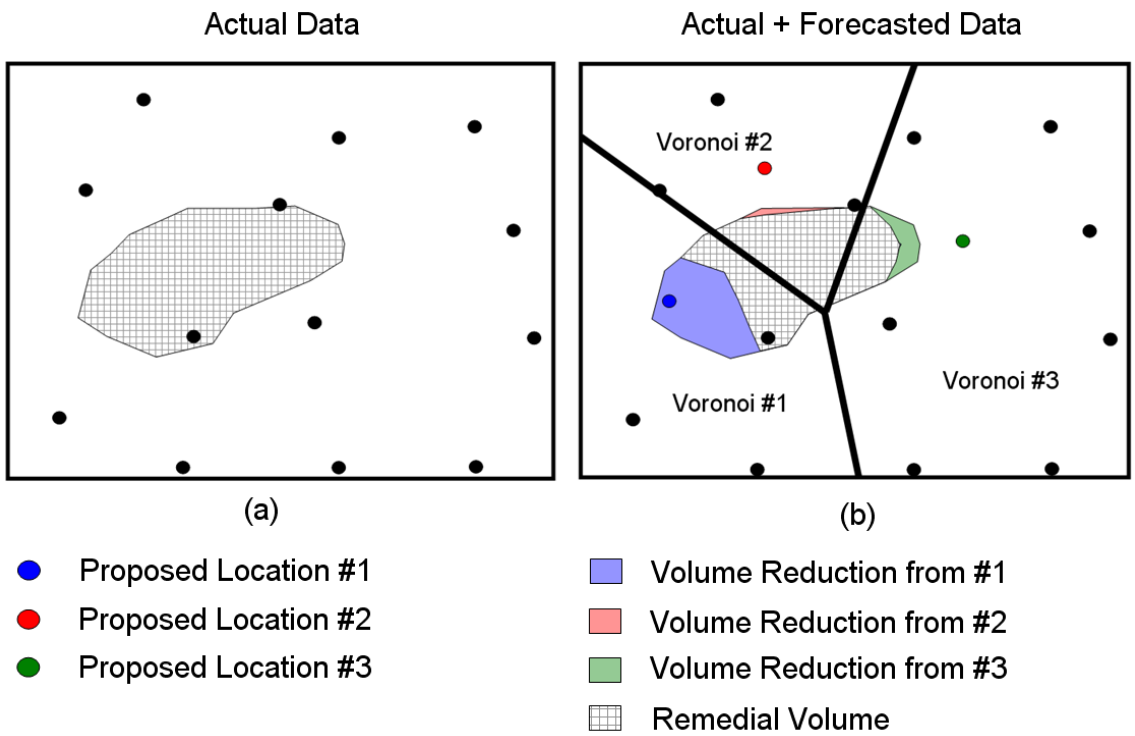


Figure 3.15 MrsDM Evaluation of proposed locations.

shown as the set of gray remedial cells in Figure 3.15(b). In addition to the new remedial area, the area reduced by the addition of the new proposed locations is shown in colors corresponding to the proposed location color.

In this example, proposed location #2 is assigned the least remedial reduction. Under MrsDM the global reduction afforded by all three new sample locations is predicted to be the least diminished by the removal of proposed sample location #2. Hence, the recommended new locations are #1 and #3. If only one sample was requested from the three, then the process would be repeated using only #1 and #3. The one assigned the least reduction benefit is removed.

There is a connection then between the cost associated with taking an additional sample and the cost savings associated with the remedial reduction assigned to

a proposed node. Such a relationship assists in determining the number of samples to collect. Indeed, when the cost of each additional sample exceeds the remedial benefit that sample is adding to the total reduction then the sample should be removed. This can continue until the savings incurred by adding the proposed design outweigh the costs associated with taking the sample. Estimation of these costs is beyond the scope of this dissertation. However, one could request that the best 1 out of N samples be selected and observe the remedial reduction associated with each removed sample design. Knowing both the cost of a sample and the savings imposed by the remedial design suggests when sampling is no longer economically viable. Under MrsDM here one may specify a stopping rule by cost consideration or by simply stating the number of desired samples.

Algorithm 3.4: MrsDM

Step 1: Compute the baseline MrDM.

The first step is to compute the baseline MrDM $\Theta_{\mathrm{MrDM}}(S)$ according to Algorithm 3.3 using simulations from the existing data.

*Step 2: Develop the set of proposed corehole locations **N**.*

Investigators can develop **N** in a variety of ways. One way is to presume nothing about the location of potentially valuable sample locations. In this case, investigators could use the GEM spatial resolution grid, specifying that new sample locations at every grid node be considered. Alternatively, investigators may recognize certain features of the site that lends itself to a well informed selection of N and reduce the computational requirements. This step is entirely the decision of the investigator.

Step 3: Forecast the values for each corehole sample.

In order to forecast the effect of additional corehole sampling on the baseline design, one must forecast the sample value at each location. In this case, a

corehole sample is emulated by assuming that one sample per vertical layer within the corehole will be collected. A forecast sample is obviously different than actually collecting the sample. How optimal the resulting design is depends on how well the sample values are forecast. Hence, MrsDM is most appropriately applied during the remedial design phase when the geostatistical simulation model is in a mature state and the SCSM is likewise stable. The sample values are forecast by selecting the median simulation value at each sample location within each candidate corehole. At the end of this step, a forecast value will be available for every sample location within every proposed corehole location.

Step 4: Compute MrDM given full set of forecast samples ***N.***

In this step, the corehole locations are added to the original conditioning data and a new set of simulations is created. This new set of simulations is used to develop a new MrDM $\Theta_{\mathrm{MrDM}}(S \mid N)$ and the baseline value $\left|\Theta_{\mathrm{MrDM}}(S)\right| - \left|\Theta_{MrDM}(S \mid \mathbf{N})\right|$. Set K = N and proceed to Step 5.

Step 5: Assign remedial units to sample locations.

For each core location w_i in **K**, assign remedial units in the remedial design to the nearest proposed sample location.

Step 6: Select the next corehole to remove from the design.

Compute the location w_{min} presenting the lowest reduction in remedial units. If there are ties then they are broken using the following rule hierarchy:

a. Location farthest from any other core,[41] or
b. Closest to center of the site,[42] or
c. Chosen randomly.

---

[41] Preference is first given to locations filling the larger spatial gaps in the data.
[42] In many cases, study areas are more or less centered over the contaminated area. Hence the one closer to the center is where the decision tends to be more important.

Step 7: Recompute MrDM.

If K-1 > K then continue to Step 8. Otherwise continue to Step 9.

Step 8. Redevelop the MrDM

Calculate the newly forecast MrDM design, $\Theta_{\mathrm{MrDM}}(S \mid K_{new})$, and return to Step 5.

Step 9: Set the final Design

The MrsDM sample design is the most recent K_{new}.

Investigators may then use the MrsDM sample design to collect new samples from the site for the purpose of the final compliance decision. The resulting benefit in reducing the remedial design will vary depending on how accurately the model is forecasting the sampling results at the proposed locations. Indeed if samples sufficiently differ from forecast values enough to drastically alter the forecast remedial design values, there may be evidence that the model is not yet stable and more data collection is warranted. In this case, the investigation returns to the characterization phase where additional samples are collected to improve the simulation model.

Summary

This chapter provides the theoretical derivation of the GEM framework. Under this framework four interrelated methods were developed: the RLR, SCSM, MrDM, and MrsDM. Each one is intended to build on the method preceding it. Beginning with the RLR a formal definition for geospatial compliance is established. Using the RLR, the SCSM test determines if compliance has been met. MrDM estimates the optimal remedial design that brings site into compliance. Using MrDM, MrsDM seeks to further reduce the remedial design through the careful positioning and acquisition of new samples. The following

chapter discusses how these components were implemented with the SADA modeling environment and the challenges that face such an implementation.

Chapter 4: Implementing the GEM in SADA

Introduction

The GEM framework is implemented as a prototype (McConnell, 1993) extension within SADA Version 5.0. The purpose of this prototype is to demonstrate how the GEM may be accessed and used. Specifically, this chapter presents how the GEM algorithms are integrated into the SADA's modeling environment, including integration into the Graphical User Interface (GUI) and within SADA logical workflows. An example application of this implementation is reserved for Chapter 6 where a hypothetical, radiological contaminated site is assessed for compliance and remediation within the five stages of radiological investigation (USNRC, 2000). Presentation of the prototype proceeds by discussing how users encounter the GEM within the SADA GUI, the architecture of new GEM class structures that implement the Chapter 3 algorithms, and finally how these classes are mainstreamed into SADA work flows.

The GEM Prototype Within the SADA Modeling Environment

The SADA freeware package provides a rich modeling and problem solving environment that well supports development of a GEM prototype. Written largely in .NET 2003, an earlier version of Visual Basic, SADA provides a rich set of modules and objects that provide ready access to geostatistical simulation models, data management tools, graphical algorithms, and a *scalable* GUI called the *Interviews-Steps-Parameters-Results* interface or ISPR (Stewart and Purucker, 2011).

Over the course of approximately 15 years of development, SADA's ISPR design was created to deal with a continually expanding set of models and an increasingly complex parameter set. The ISPR divides the problem solving environment into a series of work flows called *Interviews*. Each *Interview* in SADA represents a specific process to perform, usually a workflow of integrated

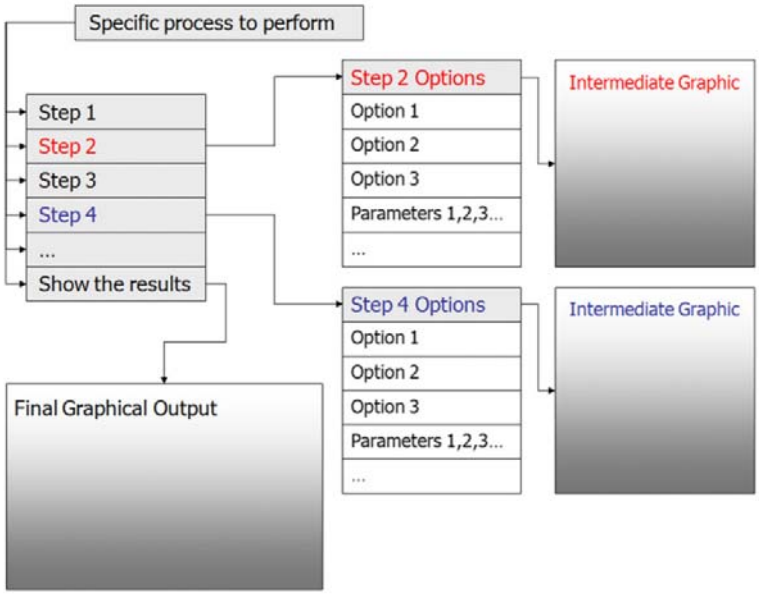


Figure 4.1 Conceptual view of SADA's ISPR style interface, used with permission from Stewart and Purucker (2011).

models. Each *Interview* is associated with a custom set of *Steps* that present the user with a logical ordering of options and steps related only to that *Interview*. Some *Steps* also allow users to view intermediary results that may occur within a larger work flow. Figure 4.1 (taken with permission from Stewart and Purucker (2011)) shows a conceptual view of the ISPR.

Figure 4.2 shows the SADA GUI and the associated ISPR.

Users begin by selecting the interview, visit each step choosing options and setting values in the associated parameter window, and selecting the *Show The Results* step (or button) to generate the results in either the 2D or 3D viewer (Stewart et al., 2009).

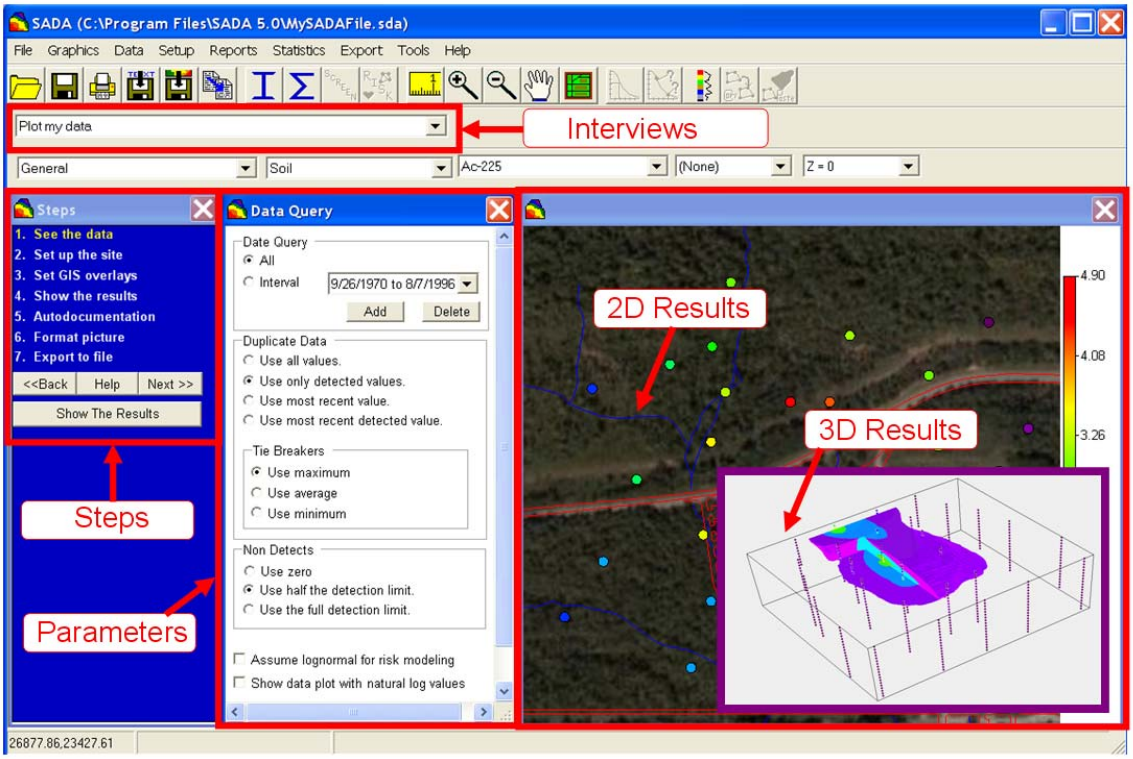


Figure 4.2 SADA's ISPR GUI.

The GEM framework is implemented within two existing interviews: *Draw an Area of Concern Map* and *Develop a New Sample Design* (Stewart et al, 2009). In *Draw an Area of Concern Map*, users develop areas of concern that are slated for remediation in order to meet a decision criteria. Prior to GEM implementation only two scales were possible (block and site). Block scale applies a single decision criteria to each remedial unit and site scale applies a single decision criteria to the entire site by cleaning remedial units from most to least contaminated until the site average is less than the criteria.[43] Figure 4.3 shows where SCSM and MrDM calculation are implemented simply by extending the *Specify Decision Criteria* step under the *Draw an Area of Concern Map*.

[43] Neither was based on geostatistical simulation.

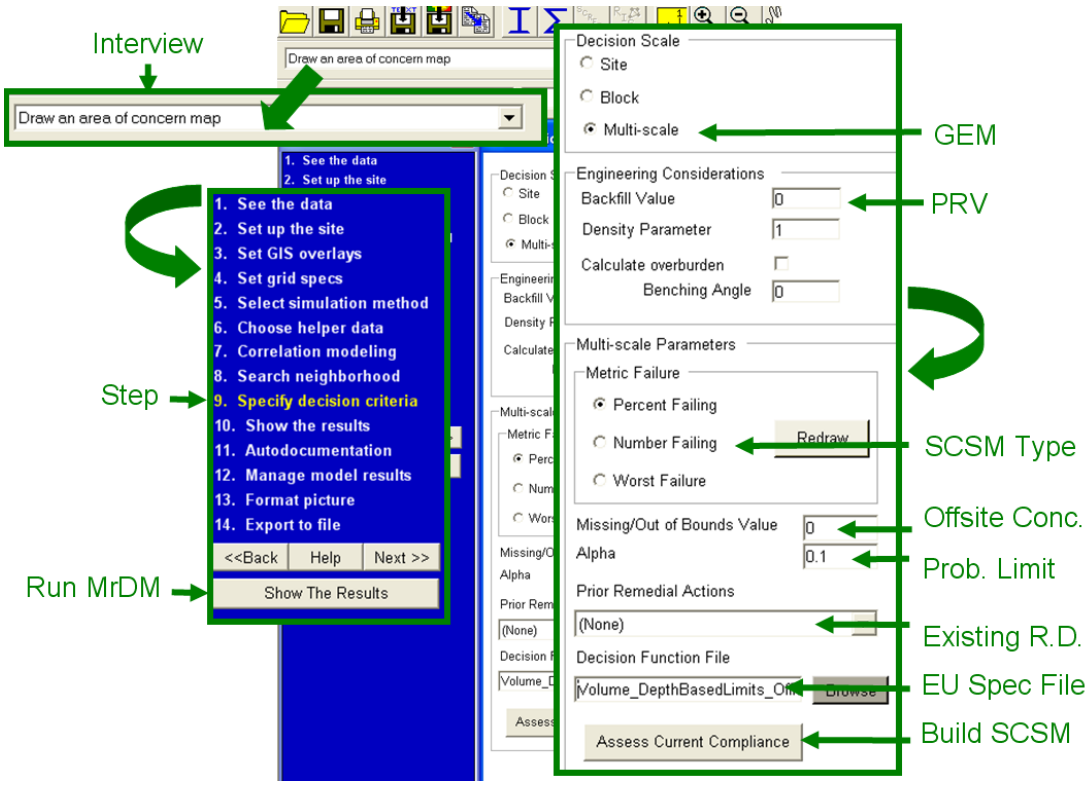


Figure 4.3 SCSM and MrDM implemented within the ISPR GUI.

Under *Draw An Area Of Concern*, MrDM is the primary goal with SCSM calculation treated as an intermediary step accessible to the user. In the parameters window, selection of the option *Multi-scale* indicates the GEM framework should be used. The *Backfill* value is where the post-remedial concentration values are entered (prv). The *Density* parameter is used for calculating the mass of contaminant. *Calculate Overburden* estimates the amount of clean soil that is physically located over contaminated soil. This clean soil will have to be removed incurring costs. *Benching Angle* is an engineering parameter (USDOL, 2008) specifying the allowable steepness of the slopes within the resulting pit. Each of these parameters is not GEM-specific and existed before this prototype. The remaining parameters are specific to the GEM implementation. The parameter set *Metric-failure* indicates the kind of

compliance graph users wish to see following a SCSM calculation. The *Missing/Out-of-Bounds* specifies what concentration value should be used in the event users wish to consider exposure unit instances that lay partially off site. Prior remedial actions drop-down box enumerates any existing remedial design strategies to consider in the work flow. The *Decision Function File* is where the user specifies an exposure unit specification file (discussed below) and finally *Assess Compliance* implements a SCSM calculation. A brief summary of the other steps is provided in Table 4.1. For a comprehensive discussion see Stewart et al. (2009)

The MrsDM model is integrated under the Develop sample design interview and affects two existing steps within that model flow: *Set Sampling Parameters* and *Specify Decision Criteria*. Sampling parameters are shown in Figure 4.4.

The step *Specify Decision Criteria* appears exactly as it does in 4.3 and is not repeated here. When the *Multi-scale option* is selected in that step, the *Area of Concern Boundary* option under *Set Sampling Parameters* provides access to the MrsDM algorithm. Users can specify the set of candidate locations (Selection Set) from which to choose an optimal MrsDM set. Users also specify the desired sample size from this selection set. In addition to the location of the samples, users have three different options for showing the behavior of MrDM: Metric, Base, and Improvement. Under Metric, a simple XY graph is plotted that shows the reduction in benefit with each additional sample removed from the selection set. The Base shows overlays of the new sample locations with the MrDM obtained without any new samples (same result as users would see under *Draw an Area of Concern map*). The Improvement shows this same map but with areas removed from this baseline highlighted.

Table 4.1 Draw and Area of Concern Map Steps

Step	Description
See the Data	Select the data set.
Setup the Site	Sets horizontal boundaries/vertical layers
Set GIS Overlays	Permits addition of GIS files in results window.
Set Grid Specs	Sets horizontal grid specifications.
Select Simulation Method	Permits selection of simulation approach
Choose Helper Data	Allows users to include field detection data.
Correlation Modeling	Facilitates modeling of spatial auto-correlation.
Search Neighborhood	Search neighborhood geometry for simulation
Specify Decision Criteria	Where decision criteria such as GEM are entered.
Show the Results	Executes MrDM (if multi-scale is selected)
Autodocumentation	SADA's report writing feature[44]
Manage Model Results	Store remedial designs developed here.
Format Picture	Access to graphical formatting controls
Export to File	Exports results to SADA standard format.

[44] Autodocumentation is not connected to the GEM prototype implementation.

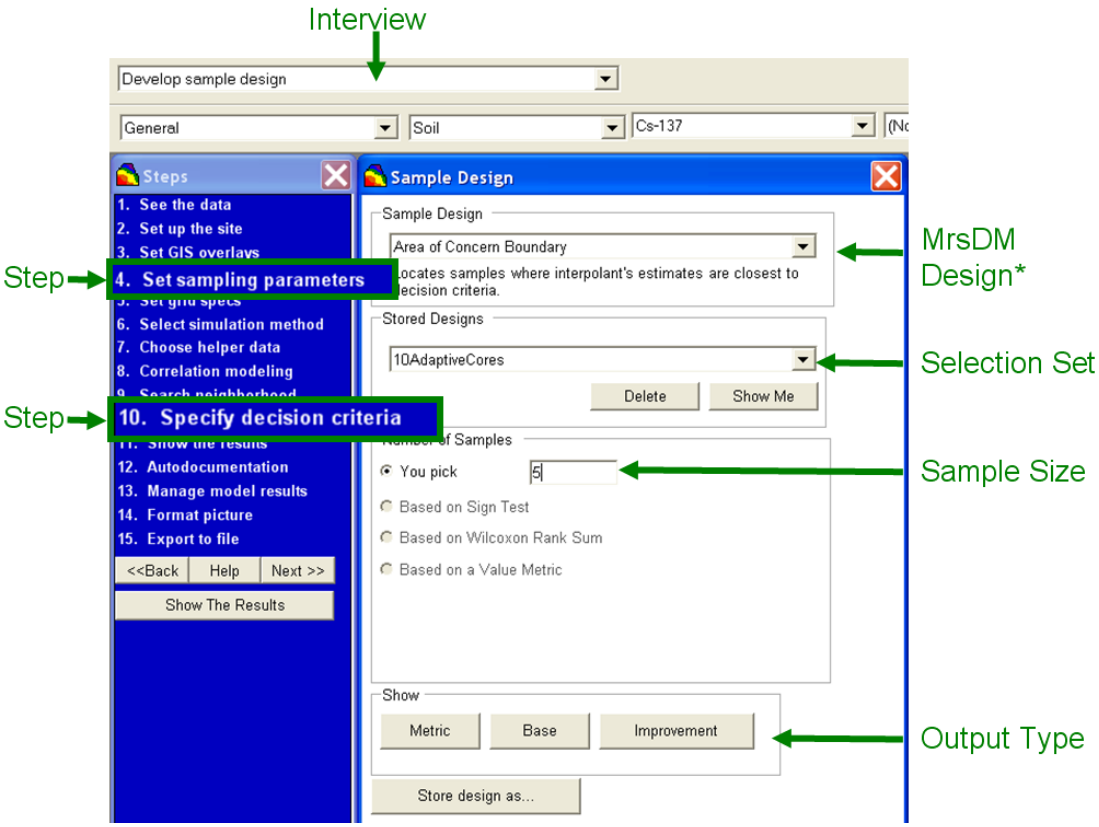


Figure 4.4. MrsDM Sample Design Parameters.

Within these steps and parameters, there are choices regarding how remedial units and exposure units are selected, both of which are tightly connected to SADA's existing grid and layering architectures. The discussion now turns to grid and layer specifications and how these relate to both remedial and exposure units specifications.

Remedial and Exposure Units in SADA

In this implementation, both remedial and exposure units are based on SADA's underlying three dimensional grid system. In Chapter 3 this is referred to as the GEM spatial grid system. Understanding how SADA builds a 3D grid system begins with SADA's vertical layering scheme. Figure 4.5 illustrates how SADA

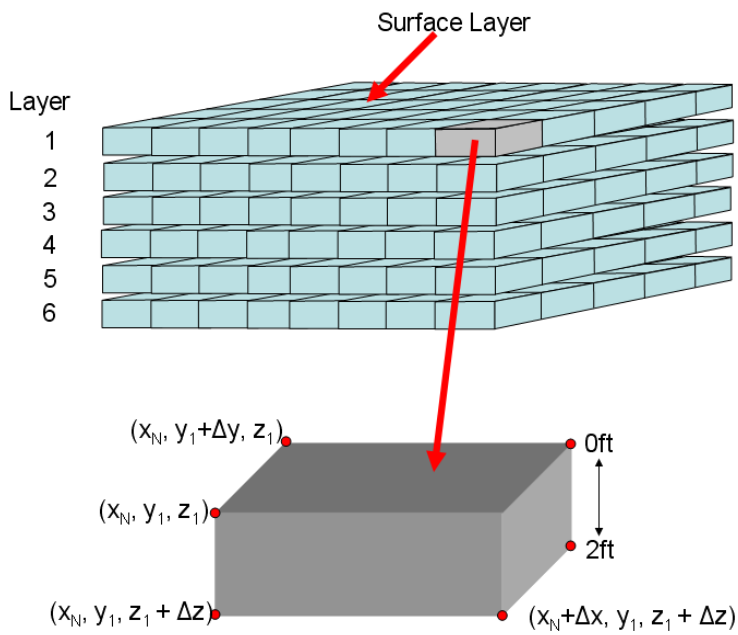


Figure 4.5 Three dimensional grid systems in SADA.

deals with the vertical component of a grid system.[45] Users specify a number of layers (e.g. 6) and an associated depth.[46] This is then combined with the horizontal grid system (Δx, Δy) to form the three dimensional grid system. In this implementation, this system becomes the GEM spatial resolution grid which is the basis for both the simulation resolution and remedial unit size in this prototype.

The last cell in the first row of the first layer is highlighted to illustrate how a vertical layering system and a horizontal grid system come together for a particular layer.

[45] Layers do not actually have gaps between them. Gaps are added here for visual clarity.
[46] In SADA applications, users choose to use variable depths. However, in this prototype implementation of GEM variable depths is not yet supported.

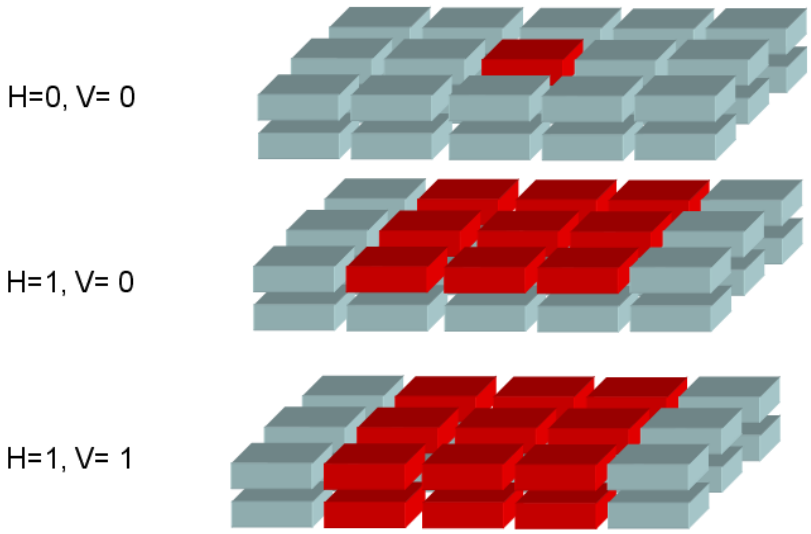


Figure 4.6 Exposure unit specification as a function of remedial unit specification.

For this prototype, a rectangular (cubic) geometry shape for exposure units will be used in order to take advantage of a number of existing data management and process handling procedures in SADA dedicated to this type of geometry. It is important to note that the GEM framework does not limit users to only a cubic geometry. Here however, for the purposes of prototyping, such a selection facilitates rapid development and demonstration of the framework. Exposure unit geometries are then defined based on the number of remedial units the span in both the horizontal and vertical direction. This span is called the horizontal neighborhood and vertical neighborhood respectively. Figure 4.6 illustrates this principle.

In Figure 4.6 three scenarios are presented; the smallest exposure unit permitted is the remedial unit, and it is specified by a horizontal and vertical neighborhood of zero (H=0,V=0). An exposure unit with a horizontal neighborhood of one and a vertical neighborhood of zero (H=1,V=0) would include a total of nine remedial

units. An exposure unit with a horizontal neighborhood of one and a vertical neighborhood of 1 (H=1,V=1) would include a total of 18 remedial units.
In addition to the geometry specification, the depth at which the exposure unit becomes relevant is required.

In addition, not all exposure units may be relevant at all depths (Figure 3.2). For depth specification, users indicate the vertical layer where the *top* of the geometry is located by specifying that layer's top and bottom depth. For example, in Figure 4.7, the exposure unit indicated in red would be accomplished by a horizontal neighborhood of one, a vertical neighborhood of one, a layer top of two and a layer bottom of three.

For any given depth and neighborhood specification, the user must also provide the DCGL. Taken together these form the base exposure unit E(v,g,d,DCGL).

Users indicate these specifications by using an exposure unit specification file. The specification file is a comma delimited custom format file developed specifically for the prototype GEM implementation. The format of the file is given in Figure 4.8 along with an example.

It is important for investigators using this prototype to remain cognizant that DCGL values are a function of neighborhood sizes which are expressed as remedial unit increments and not distance. For example, suppose that a remedial unit specification is given by a 5ft x 5ft x 5ft. Users interested in an exposure unit of the same size (125ft^3) would enter the associated DCGL under the neighborhood 0/0 at the depth of interest in the specification file.

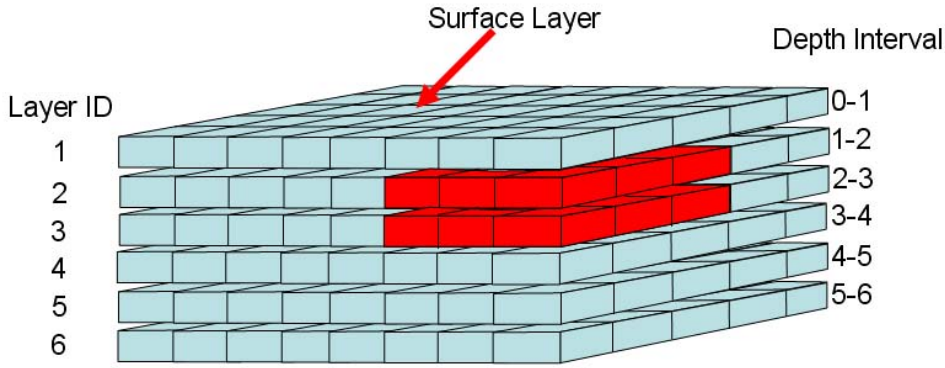
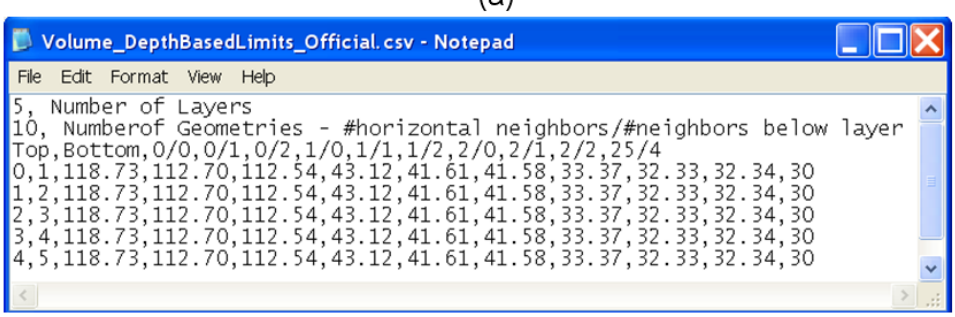


Figure 4.7 Exposure unit specification within the GEM grid system.

K	Number of Depth Specific Exposure Units			
G	Number of Geometries			
Top	Bottom	H_1V_1	……	H_G/V_G
T1	B1	$DCGL_{T1,B1}^{H1/V1}$		$DCGL_{T1,B1}^{HG/VG}$
.	.	.	.	.
.	.	.	.	.
TK	BK	$DCGL_{TK,BK}^{H1/V1}$		$DCGL_{TK,BK}^{HG/VG}$

(a)



(b)

Figure 4.8 EU Specification File format (a) and example (b).

SADA Workflow

When the user selects *Show the Results* or *Assess Compliance*, this event initiates a workflow that begins by gathering the relevant parameters and selected options and ends with a graphical product in the results viewer. A simplified view of a SADA workflow is presented in Figure 4.9.

Model-specific workflows are encapsulated algorithms that operate independently of the GUI and the larger workflow. Within this model-specific workflow is where new components SCSM, MrDM, or MrsDM will be situated. These new components exist as specific class structures that handle data and parameters passed by the calling workflow, execute the required algorithm, and produce the outcomes in a format suitable for the graphing routines. The architecture for these class structures is briefly discussed.

GEM Class Structures

Five new VB.Net classes were created to handle the GEM workflow. Each class is fairly complex and only the most important features are discussed in this

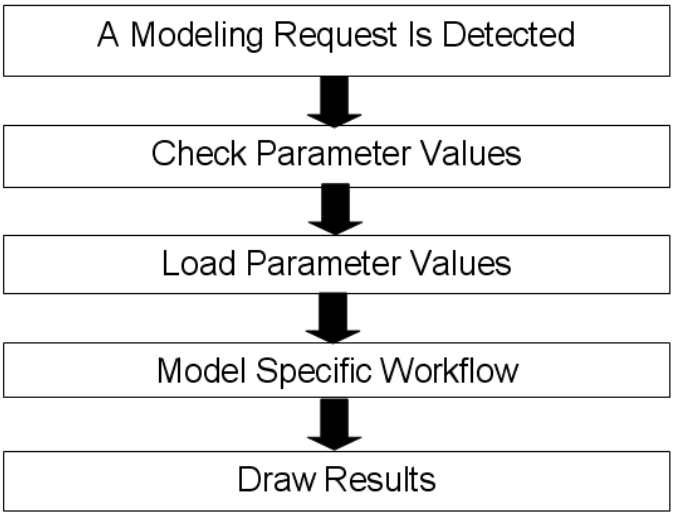


Figure 4.9 SADA's Macro-Level Workflow

chapter. A complete disclosure of all the public properties and methods can be found in Appendix A.

SADA.clsGEMStructure

The foundation of GEM implementation is the class *clsGEMStructure* which supplies the methods for computing the GEM framework, specifically, SCSM calculation, MrDM, and MrsDM. Figure 4.10 illustrates the behavior of the structure as it receives, processes, and outputs results.

The class can fully instantiate and populate itself using one or more of the methods included in the blue box. These methods can accept data from either a flat exposure unit specifications file or a stored clsGEMStructure parameters set from the SADA file.

Execution of these methods leads to a full specification of the class properties (green) which in turn supports the calculation of the primary GEM components (red). The method *CalculateSCSM* implements Algorithm 3.1, *CalculateMrDM* implements Algorithm 3.3, and *CalculateMrsDM* implements Algorithm 3.4. Which algorithm to execute is included in the flow of parameters and data.

Methods exist for creating and populating the class (blue) based on spatial grid system specifications and exposure unit specifications either directly or retrieved from previous implementations of this class (from the SADA File). These methods for establishing the class (blue) fully populate the properties (fields) *Name*, *ExposureUnitArray*, and *PhysicalStructure*. The *Name* is a unique assigned string identifier. The other two properties are themselves rich class structures discussed in the next two sections.

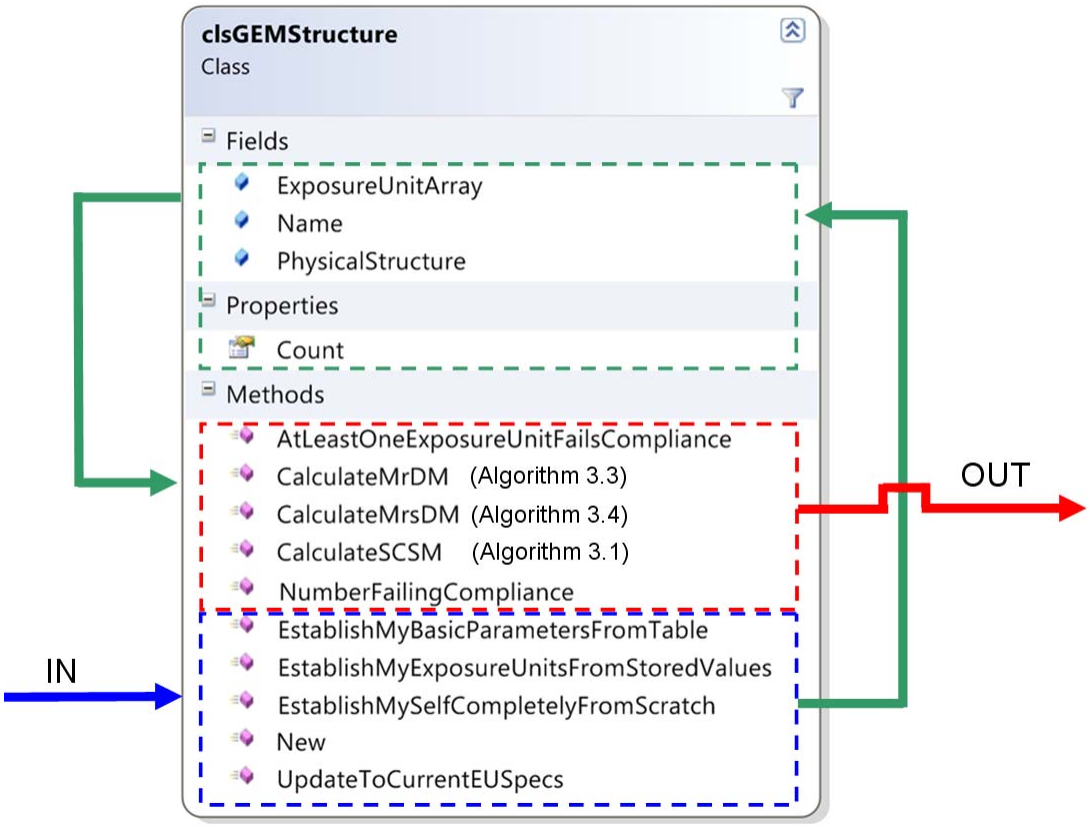


Figure 4.10 Class diagram for the clsGEMStructure.

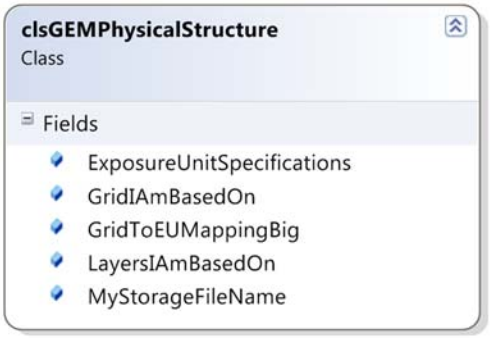


Figure 4.11 Class diagram for the *clsGEMPhysicalStructure*

SADA.clsGEMPhysicalStructure

The property *PhysicalStructure* is the class *clsGEMPhysicalStructure* diagramed in Figure 4.11.

This class holds the specifications for the GEM spatial resolution grid comprised of the horizontal grid (*GridIAmBasedOn*), and vertical layers (*LayersIAmBasedOn)*. Property *GridIAmBasedOn* is an original SADA class structure called *clsGridDefinition* that provides a rich set of properties and methods for using a spatial grid system. The property *LayersIAmBasedOn* is an original SADA class structure called *colLayers* that provides a rich set of properties and methods for creating and managing subsurface layers. The property *GridToEUMappingBig* is an array mapping every node in the spatial resolution grid to each exposure unit assigned to it. This permits code to efficiently move between remedial grid units and exposure units. The property *MyStorageFileName* is a unique name assigned to this class for the purpose of data management. Finally, the property *ExposureUnitSpecifications* is a new GEM class structure named *colExposureUnitSpecifications* that holds the contents of the exposure unit specification file.

SADA. colExposureUnitSpecifications

The class *colExposureUnitSpecifications* holds the contents of the exposure unit specification file and is used to create exposure unit instances as well as a collection point for gathering compliance results. A class diagram is provided in Figure 4.12.

In Figure 4.12, properties and methods are organized into five types indicated by different colors. Methods for instantiation and parameterization of the class are indicated in blue. The method *PopulateWithValuesFromFile* parameterizes the class directly from the exposure unit specifications file. The other method *ParameterizeWithParameterString* parameterizes the class based on a string of stored parameter values developed during a previous use of this class and passed in by SADA's data management tools. Both methods completely parameterize the properties indicated in green. These are then consumed by GEM operations executed within the *clsGEMStructure* class. Following these operations, the methods enclosed in purple consolidate these results by unique volume and populate the remaining fields indicated in red. These fields are then ready for the drawing routines to use. The areas indicated in black refer to methods that are tools that support calculations throughout the workflow (e.g. within GEM operations). Refer to Appendix A for details. This class is actually a collection[47] of GEM structures known as *clsExposureUnitSpecifications*. Each particular structure is assigned a unique key and is accessible through the *Item* property.

SADA.clsExposureUnitSpecifications

The class, *clsExposureUnitSpecifications*, stores, manages, and utilizes the specifications (e.g. each $DCGL_{T/B}^{H/T}$) for a single cell in the exposure unit specification file. Hence if there are N geometry and M layer specifications in the

[47] Collections are an alternative method to an array of storing multiple items.

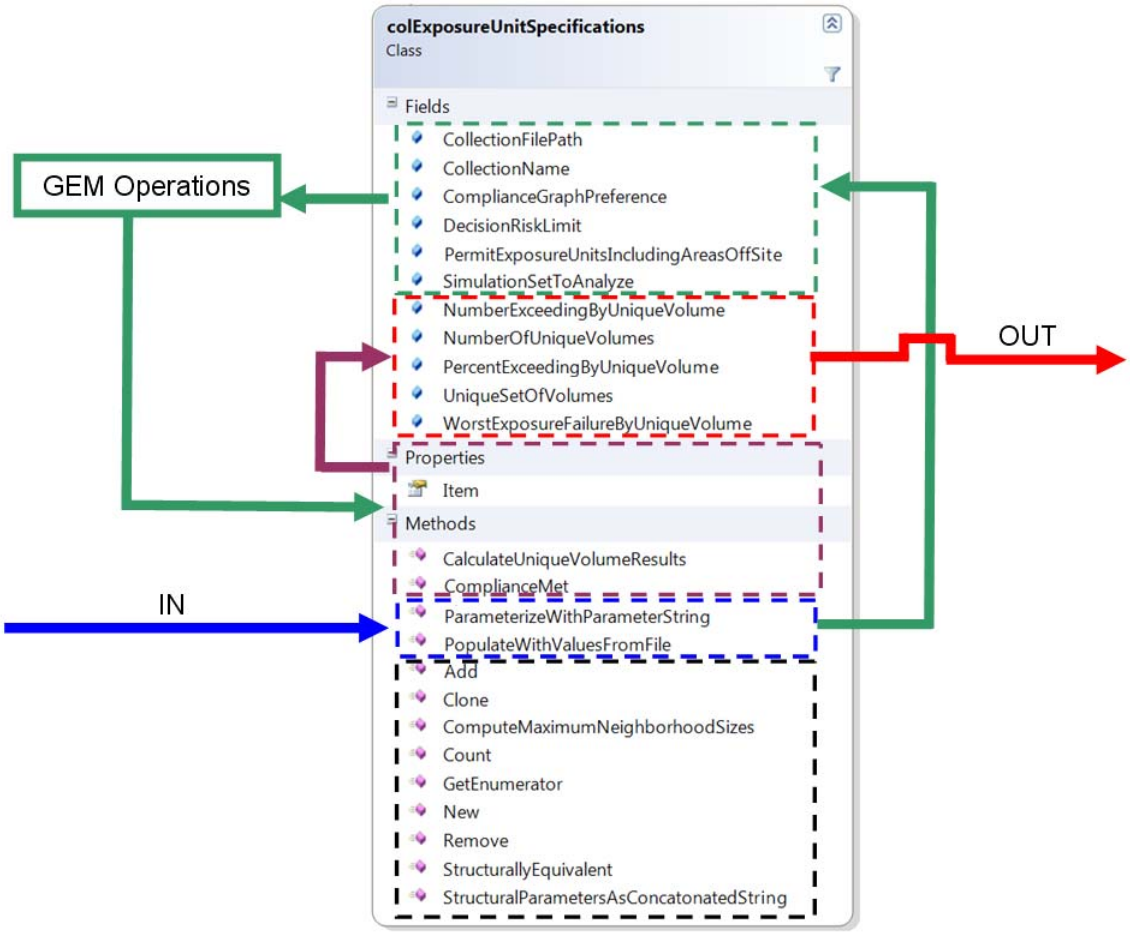


Figure 4.12 Class diagram for *clsExposureUnitSpecifications*. GEM operations refer to methods enclosed within the red box in Figure 4.10.

exposure unit specification file there will be N x M instances of this class. This class records the geometry of the exposure unit and the vertical placement but also provides a number of important features that support GEM operations including reporting the number of exposure unit instances arising from this specification such as the number, the worst case compliance failure, and other methods that support various data management operations. Figure 4.13 shows a class diagram for this structure and its relationship to the parameterization methods in *colExposureUnitSpecifications*.

As with the other diagrams, the methods associated with parameterizing the class, outlined in blue, parameterize the properties, outlined in green. These are in turn used by the GEM operations (via *colExposureUnitSpecifications*) which

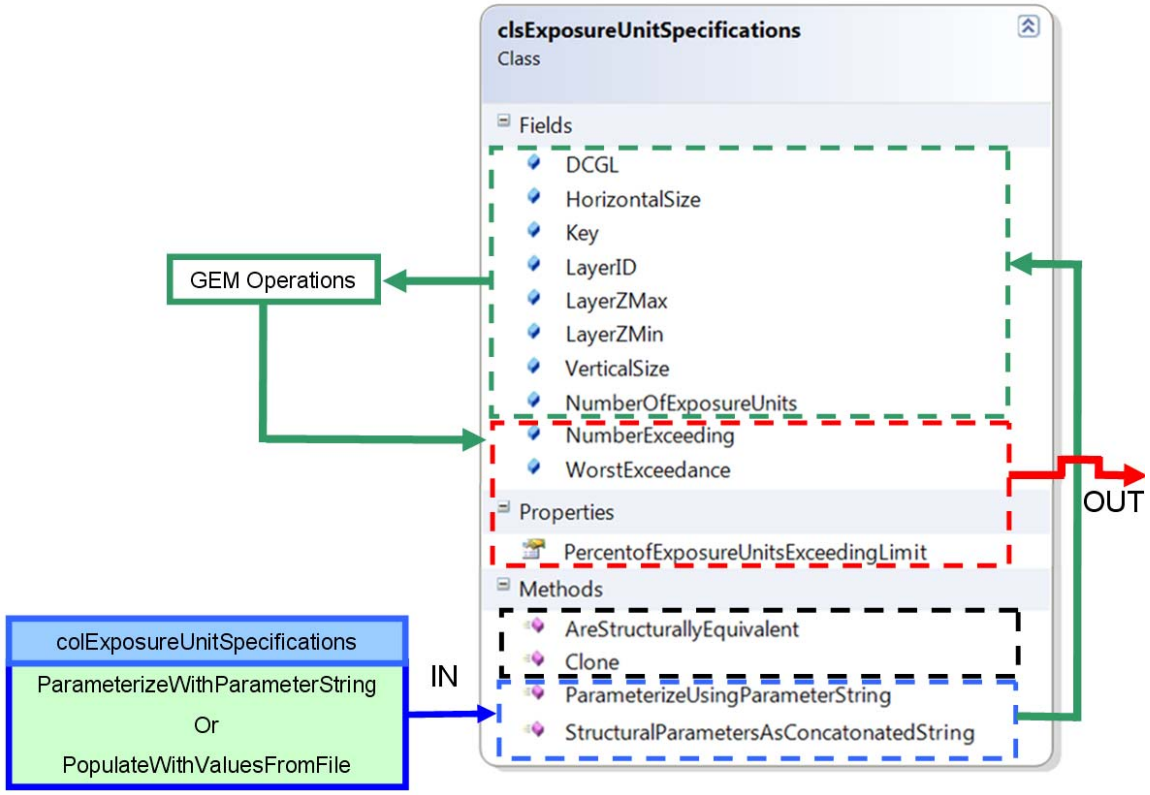


Figure 4.13 Class diagram for *clsExposureUnitSpecifications*.

sets the failure compliance values outlined in red. Care was taken in naming properties and methods in such that their meaning is clear. For greater detail about each, please see Appendix A.

SADA.clsExposureUnit

The other major property in the *clsGEMStructure* (Figure 4.10) is the *ExposureUnitArray*. This property is an array of *clsExposureUnit* class structures which represent the instantiated exposure units at each grid node for each *clsExposureUnitSpecification*. Each member of this array is instantiated by mapping each *clsExposureUnitSpecifications* within the *colExposureUnitSpecifications* of the *clsGEMPhysicalStructure* to every appropriate grid node in the spatial resolution grid. This class supplies numerous properties and methods that are directly accessed during the execution of *CalculateMrDM, CalculateMrsDM, and CalculateSCSM* methods within the *clsGEMStructure*. A full disclosure of these is found in Appendix A. In this discussion a select few are emphasized as they are recognizable components of the GEM algorithms. Figure 4.14 shows the full set of public properties (fields) and methods and places colored circles next to those which are specifically discussed here.

As with the other class diagrams, blue indicates methods by which this class is instantiated by passing parameters from the calling routine. These lead to setting or calculation of property values. Among these are parameters mentioned here and in Chapter 3 including the depth at which the unit should be positioned (*ApplicableDepth*), the DCGL (*DCGL)*, horizontal neighborhood size (*HorizontalNeighorhoodSize)*, probability limit (*ProbabilityLimit*), vertical neighborhood size (*VerticalNeighorhoodSize*), the post-remedial concentration value (*RemedialReplacementValue*), and the identification key for the *clsExposureUnitSpecification* on which this unit is based (*ExposureUnitSpecificationKey*). The property *ExposureUnitsIShareRUsWith*

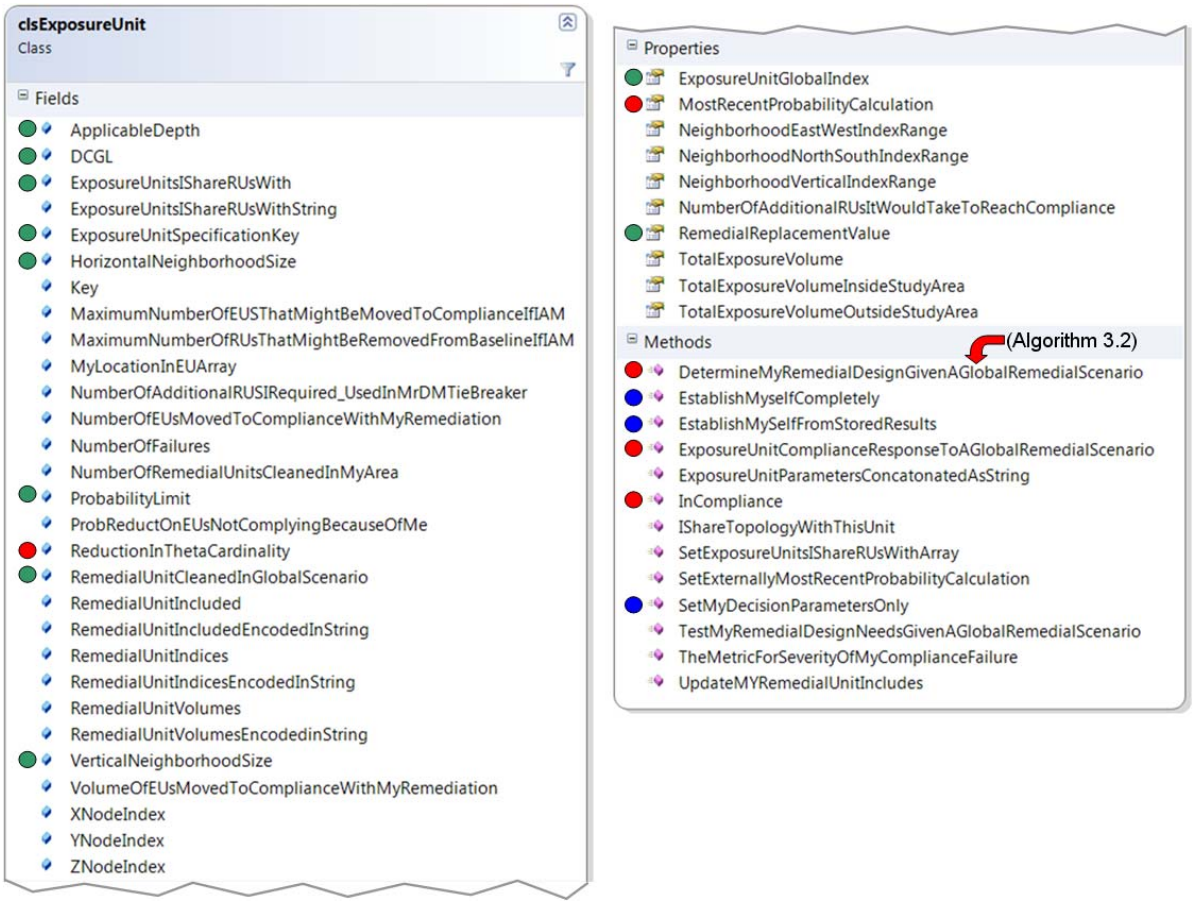


Figure 4.14 Class diagram for *clsExposureUnitSpecification*.

which records the *key* property for all the other *clsExposureUnitInstances* sharing space with this unit. The property *ExposureUnitGlobalIndex* records the spatial resolution grid node ID where this unit is found. Finally, *RemedialUnitCleanedInAGlobalScenario* keeps track of which remedial units within the spatial domain of this class are (or should be) included in the global design to meet compliance. Three central methods are:

1) *ExposureUnitComplianceGivenAGlobalRemedialScenario* which calculates the probability of compliance failure in the SCSM,
2) *MostRecentProbabilityCalculation* where this probability is recorded,
3) *DetermineMyRemedialDesignGivenAGlobalRemedialScenario* which produces the local naïve remedial design in Algorithm 3.2 and is used by the MrDM workflow.
4) *ReductionInThetaCardinality* where the benefit of using this instance as a conditioning design is recorded (calculated by *clsGEMStructure.CalculateMrsDM*).

One final class of importance is the *clsPreviouslyGEMStructures*. When a GEM model-specific workflow is first entered, this class determines which, if any, previously developed *clsGEMStructures* may be used given the exposure unit specification file and the grid/layering system provided by the user. Using a previously created *clsGEMStructure* and updating it with current decision criteria avoids time-consuming events associated with calculating topography between exposure unit instances. If no previous development can be used, a new *clsGEMStructure* is created and added to the collection. A class diagram of *clsPreviouslyDevelopedGEMs* can be found in Figure 4.15.

The property (field) *DevelopedGEMs* is an array of previously developed *clsGEMStructures* and the *CurrentGEM* is the *clsGEMStructure*,the appropriate class to use during the workflow.

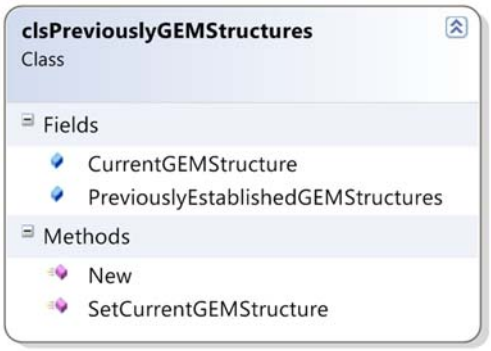


Figure 4.15. Class diagram for *clsPreviouslyDevelopedGEMs*

Summary

While the details of each structure are necessarily involved, from the SADA workflow perspective, executing a GEM algorithm amounts to creating a *clsGEMStructure*, passing it a simulation set, a grid/layer system (given within the simulation set class), a set of exposure unit specifications, and various decision criteria previously mentioned. The following discussion illustrates how these classes are arranged into a model-specific workflow.

The GEM Model Workflow

The following diagrams illustrate how these five classes form the model workflow and are situated within the larger SADA macro workflow for calculating the SCSM, MrDM, and MrsDM algorithms.

The SCSM work flow begins when the user selects *Assess Compliance* (Figure 4.3). This initiates the workflow presented in Figure 4.16.

As indicated in Figure 4.16, the task of the SADA workflow is to access the exposure unit specifications, the simulation set (which sets the GEM spatial resolution grid), and identify the proper *clsGEMStructure* to call. This *clsGEMStructure* executes the remaining tasks and produces the SCSM model.

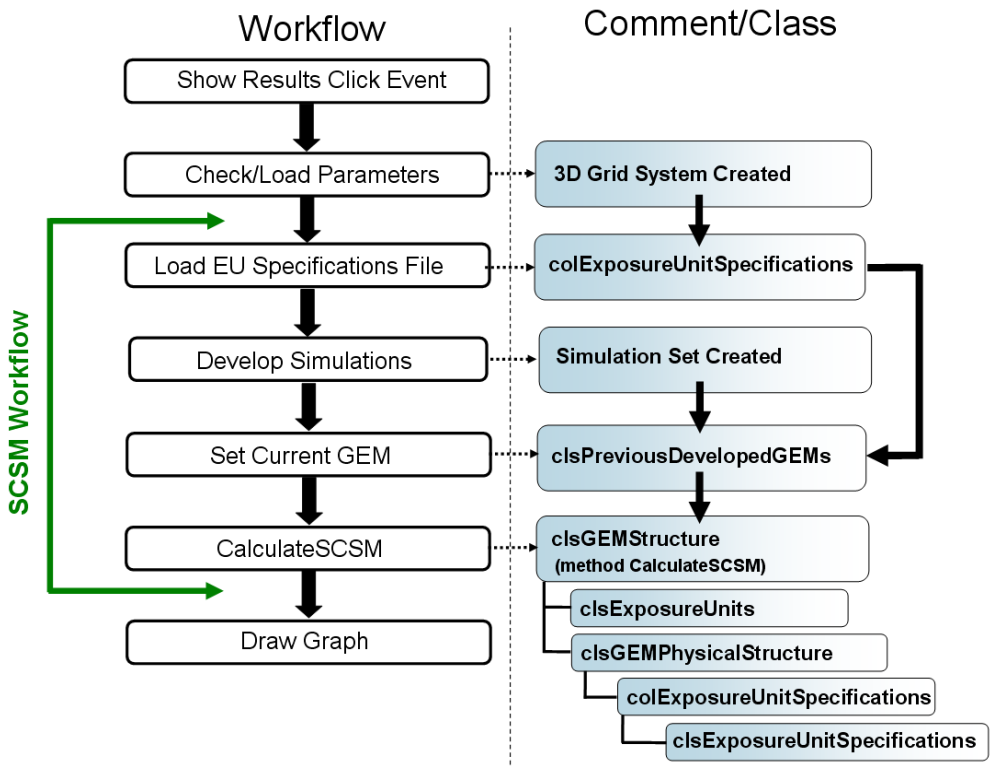


Figure 4.16 The SCSM workflow.

The workflow for MrDM is virtually identical. Rather than call the method *CalculateSCSM*, the routine *CalculateMrDM* is called instead. Figure 4.17 shows the MrDM workflow.

The MrsDM workflow begins by first using the MrDM workflow to create the baseline design. Then the same *clsGEMStructure* is used to access the method *CalcualteMrsDM*. Figure 4.18 shows the MrsDM workflow.

MrsDM is the most computationally demanding workflow as the method *CalculateMrsDM* repeatedly calls the MrDM workflow each time it must decide which new sample location to remove from the list (see Algorithm 3.4). A discussion of the computational demands of the MrsDM and the GEM framework is provided in the example chapter and the conclusions.

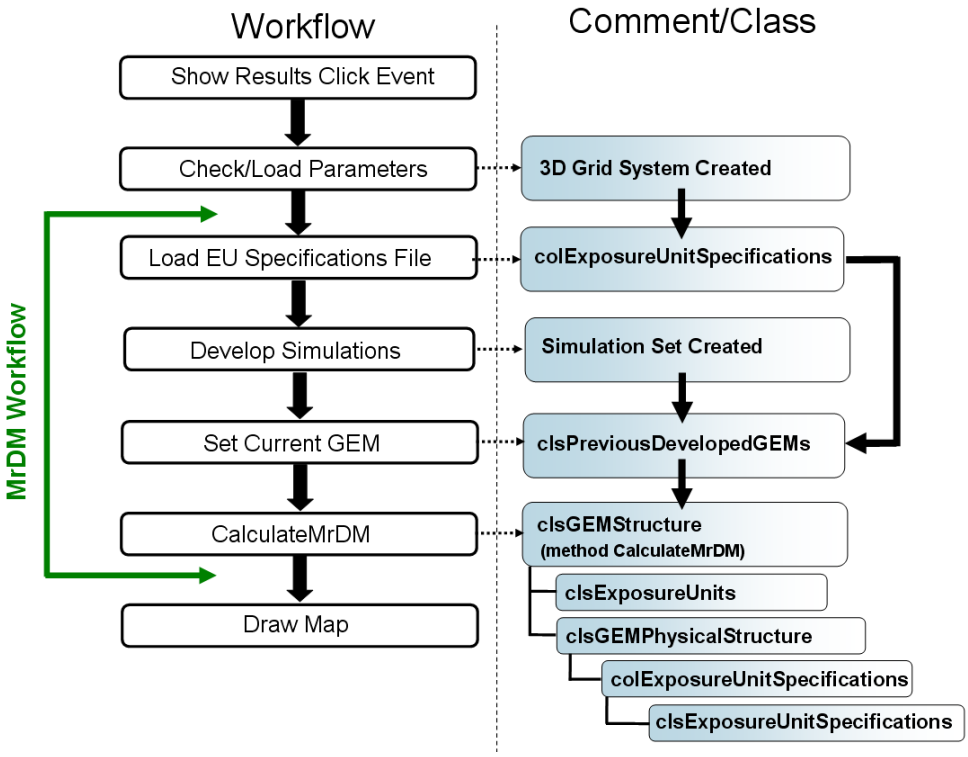


Figure 4.17 The MrDM workflow.

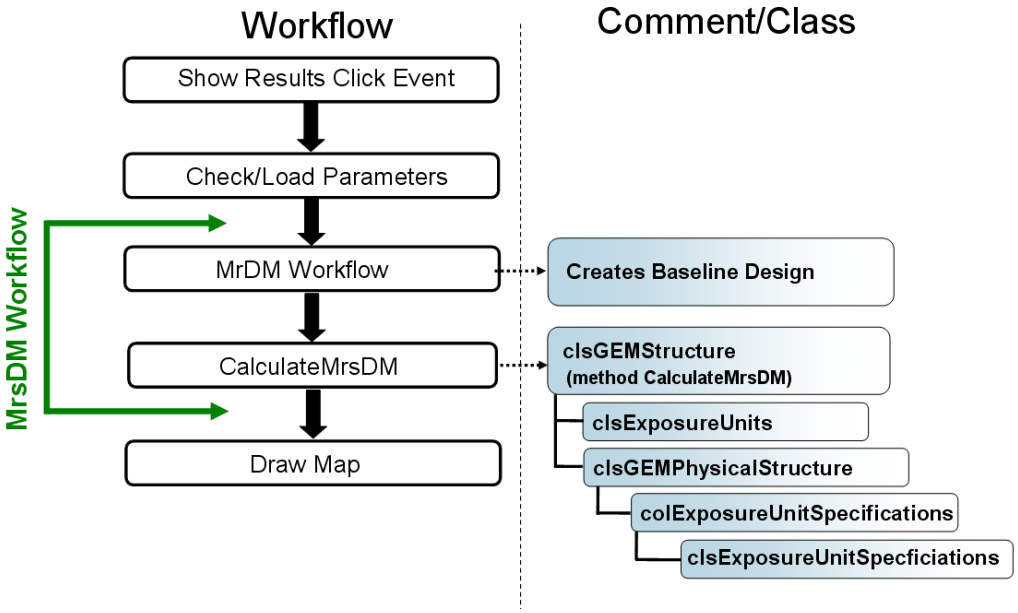


Figure 4.18 The MrsDM workflow.

Summary

This chapter provided an overview of how the GEM components are implemented within the SADA modeling environment. The major points of this chapter can be summarized as follows:

- GEM is implemented as prototype within SADA Version 5.0.
- GEM is positioned within the GUI under two existing interviews.
- Interviews require that GEM modeling reside within the SADA workflow.
- GEM modeling is handled by *clsGEMStructure*
- The *clsGEMStructure* encapsulates SCSM, MrDM, and MrsDM algorithms
- Under this prototype, the spatial resolution grid forms the set of remedial units and exposure units are derived from subsets of remedial units.

An example application of this implementation is reserved for Chapter 6 where the GEM framework will be demonstrated within the phases of environmental investigation. Before the example is presented, the study will discuss a separate geospatial tool which was developed that may support the geospatial methods in the subsurface in general, and the GEM framework in particular. This method is a sample design referred to as “Check and Cover” and is intended for early characterization efforts. The following chapter presents both the theoretical derivation and implementation within SADA. As with the GEM framework, this design is also demonstrated within Chapter 6.

Chapter 5 Check and Cover Sampling Strategy

Introduction

During the scoping phase investigators are focused on determining whether contamination exists at the site. A handful of samples are collected, usually in areas considered most likely to be contaminated based on the HSA findings. If samples indicate that a problem may exist, the investigation enters the characterization phase. A central goal of characterization is to find and delineate the spatial distribution of contamination across the study area. Early in characterization the objective is to encounter the body of contamination through sampling in order to determine the spatial magnitude of contamination. As the investigation unfolds, the goal shifts to defining the boundary of this contaminant body in order to identify the areas of concern that will be slated for remedial action.

In practice it is common to begin by applying a uniform sampling grid across the site. Indeed for two dimensional applications hot spot search algorithms have been previously developed that estimate sampling density required to encounter a hotspot with a given probability. These approaches assume that nothing is known about the location of contamination, assume the contamination is elliptical, and in some cases require the user to specify the size of the hotspot they wish to find (Gilbert, 1987). Application of a uniform grid like these provides equal sampling coverage across the entire site and implies (from a decision perspective) that each region of the site is equally important to measure. Suppose this was known not the case (with respect to encountering contamination). Investigators may indeed have some initial knowledge about where contamination exists. While this knowledge tends to be more qualitative than quantitative in the early stages, it remains a valuable piece of the characterization puzzle.

Given this knowledge, investigators may wish to focus sampling in this area in order to encounter and determine the magnitude of the contaminated zone. Targeted designs such as SADA's high value design (Stewart et al., 2009) accomplish this, but the risk in committing substantial resources to such prior knowledge is of course that the prior knowledge is wrong and contamination is indeed elsewhere on the site. Even under the scenario that the prior knowledge is largely correct and contamination is encountered, investigators may not wish to expend so much of the sampling budget that later sampling efforts aimed at delineating the boundary lack sufficient funds. Indeed, sampling in areas that are not contaminated can also be desirable for a variety of reasons including delineating the boundary between contaminated and uncontaminated zones.

Methods have been developed for casting this risk of sampling into a value of information formulation where the cost of gaining the information is traded off against the value it provides the decision maker. These methods are usually highly quantitative in nature and require a decision endpoint be defined.

The challenge is therefore twofold. First, how to use qualitative expert knowledge to strike a balance between taking samples in areas that are highly suspected of contamination and those areas that might not be contaminated but nonetheless require some sort of quantitative evidence of this fact early in the characterization phase. Second, how does one account for the level of confidence in the prior knowledge? These are the objectives of the Check and Cover sample design strategy.

Method

The Check and Cover design begins by establishing a Likert-like (Trochim, 2006) *scale of concern* where investigators rate their concern for the presence of contamination on a scale from 1-10. A value of 1 indicates a very low concern, 10 indicates a very high level of concern, and 5 indicates a complete lack of

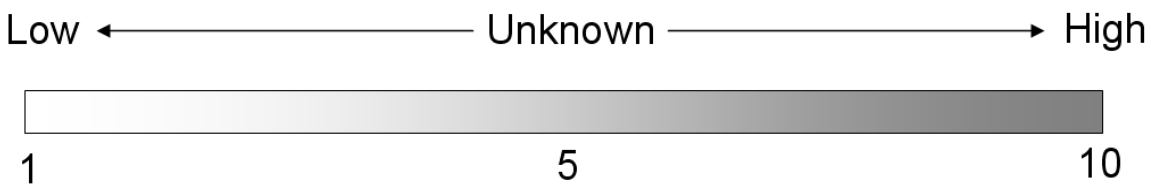


Figure 5.1 Scale of concern for contamination.

knowledge about whether contamination is present or not. Figure 5.1 shows this scale of concern.

Using this scale, investigators develop a three dimensional conceptual site model called the *concern model* that spatially delineates where contamination is thought to exist. The model is a three dimensional raster model (much like the GEM spatial resolution grid) where cell values are assigned values from the scale of concern. SADA version 5.0 provides access to a *User Defined Model* tool where a finite number of values (e.g. 1, 2, …10) can be easily assigned to a three dimensional raster model using various graphical tools (Stewart et al., 2009). Figure 5.2 shows an example.

When samples are collected in an area, they provide a service to the investigator by meeting a demand for knowledge about whether contamination exists. This demand is measured by the level of concern provided in the model. Considering sampling and contamination concerns in this light reformulates the process into a service/demand problem. Indeed such a question is at the center of a problem known as the *location* problem (Ostresh, 1978) or P-median problem (Miller and Shaw, 2001; Dai and Cheung, 1997).

In the p-median problem one wishes to optimally locate a number of supply or service locations among a set of demand locations. In particular, the problem can

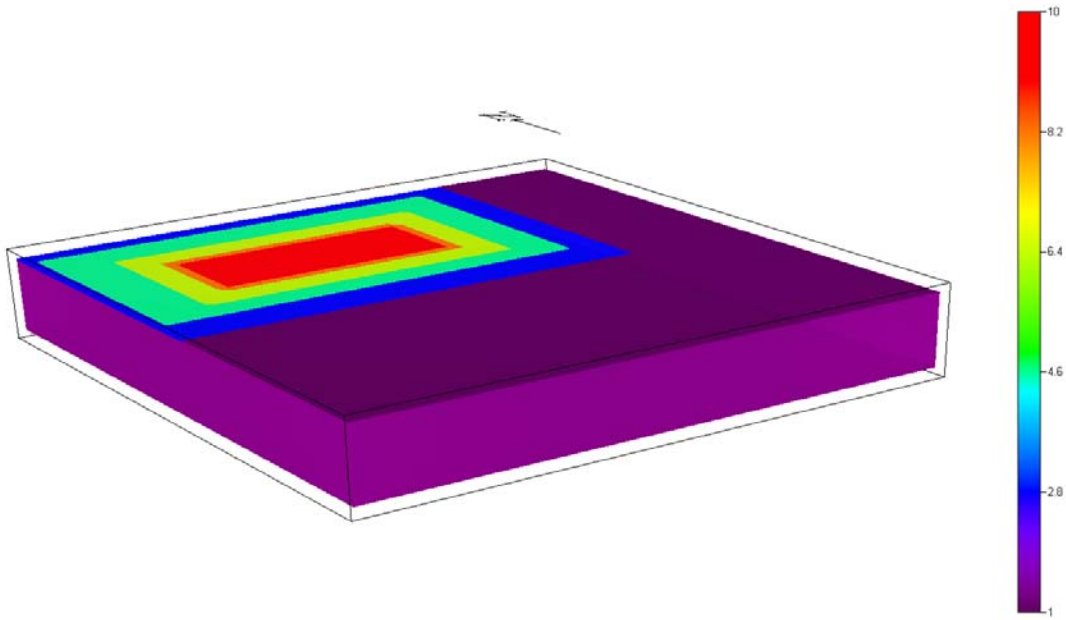


Figure 5.2 A three dimensional conceptual site model indicating contamination concerns.

be formulated as follows (Ostresh, 1978) for two dimensional space. Let $U = (\mathbf{u}_1, ..., \mathbf{u}_N)$ be a set of fixed distinct (x_i,y_i) points in a two-dimensional space. Let $W = (w_1, ..., w_N)$ be a set of weights associated with U. The goal is to find a set of P new points $S = (\mathbf{s}_1, ..., \mathbf{s}_P)$ such that minimize

$$\sum_{i=1}^{n}\sum_{j=1}^{p} w_i d_{ij}$$

Where dij is the Euclidean distance from $\mathbf{u}_i$ to $\mathbf{s}_j$.

Similar applications of p-median to the study of environmental applications are already available in the literature for other applications. For example, Meyer and Brill (1988) apply a variation of p-median called the maximum covering location problem to the optimal placement of groundwater wells.

In the situation here, **U** is the set of grid nodes for the model of concern, **W** is the level of concern about contamination at those nodes, and **S** is the location of the new sample locations. The problem is actually quite challenging and a number of methods for solving it have been provided. See Miller and Shaw 2001 for an enumeration of methods. Here the existing code[48] based on Ostresh (1978) was used to solve the p-median for Check and Cover.

Using the scale of concern, the prior concern model, and P-median provides a means to answer the first question posed (how to strike a balance). P-median in this context has some potentially appealing properties to an environmental decision maker. If the prior concern model is correct, then samples are indeed placed within the contaminant body as desired. Furthermore, some samples are placed in clean areas allowing a good start to delineating the contaminant boundary (a later objective in characterization). If the prior is incorrect, resources may have been unnecessarily expended in a clean area. Nonetheless, those samples outside this domain may have encountered the true contaminant body or at a minimum have further reduced the area in which contamination may exist.

An important factor to consider therefore is the level of confidence about the prior knowledge. At such an early stage, it is not likely that investigator confidence can be stated quantitatively. Therefore, for stating confidence qualitatively, the method turns again to the use of a Likert-like scale. In this instance, the decision maker selects a confidence level about the concern model from a 5-point scale (Figure 5.3).

Each level on the scale is associated with a parameter called the map reliance parameter which lies in the interval [0,1]. Given this parameter, the prior concern

[48] Code provided during personal correspondence with Dr. Bruce Ralston in 2009 in support of the SADA project.

○ Complete (1.00)

○ High (0.75)

◉ Moderate (0.50)

○ Low (0.25)

○ None (0.00)

Figure 5.3 Prior CSM confidence scale.

model is adjusted according to Equation 5.1.

$$CSM_{adjusted}(u_i) = CSM(u_i) + (5 - CSM(u_i))(1 - R) \;\; for \;\; i = 1, \ldots, N_{nodes} \qquad \text{(Eq. 5.1)}$$
$$R \in [0,1]$$

Notice that for a reliance factor of 1 (Complete) the $CSM_{adjusted}$ and the CSM are equal. As the reliance factor decreases from 1 to 0, the $CSM_{adjusted}$ converges to the unknown scale of concern value (5) everywhere. Figure 5.4 shows this convergence for the top layer of a three dimensional raster concern model.

Note that the “None” reliance factor means that the level of concern everywhere is the same. Hence there is no greater value in preferentially core sampling in one region over another. Under Check and Cover, an initial triangular grid is used as the initial guess. A triangular grid is created by offsetting every other row of a regular grid by half the grid spacing. The effect is to create a triangular pattern in the sampling design. The literature suggests (Gilbert, 1987) that a triangular grid is the optimal search approach when no prior knowledge is known. If there are no

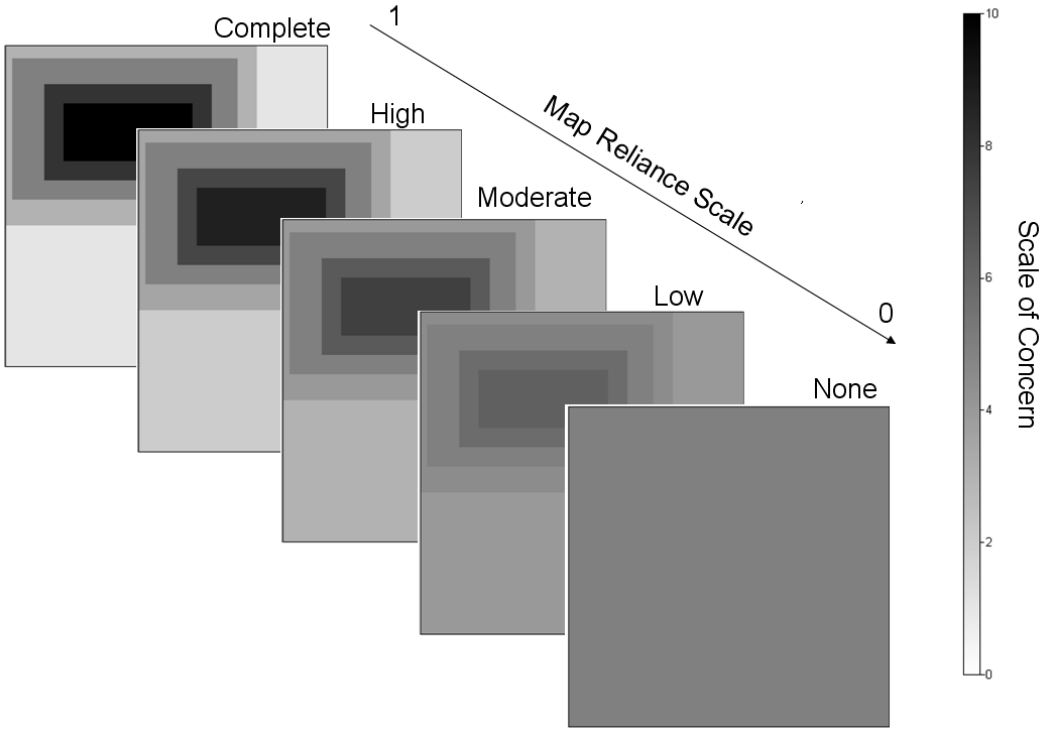


Figure 5.4 Map reliance factor effect on scale of concern values.

variations in concern, p-median should adjust this initial guess only to spatially balance samples within the site based on site boundaries. If concern levels do vary, p-median moves away from a triangular grid into a more clustered or biased design. Figure 5.5 demonstrates the effect of the map reliance factor on an initial guess.

Notice how the P-median values tend to provide a regular triangular distribution under the zero reliance (know-nothing) state. On the other end of the spectrum (reliance factor = 1) the design tends to provide preferential sampling balanced by the spatial distribution of concern. This movement by the sample locations reflects the relative change in the concern when adjusted by the reliance factor. For the *None* case, the adjusted level of concern is the same everywhere and P-

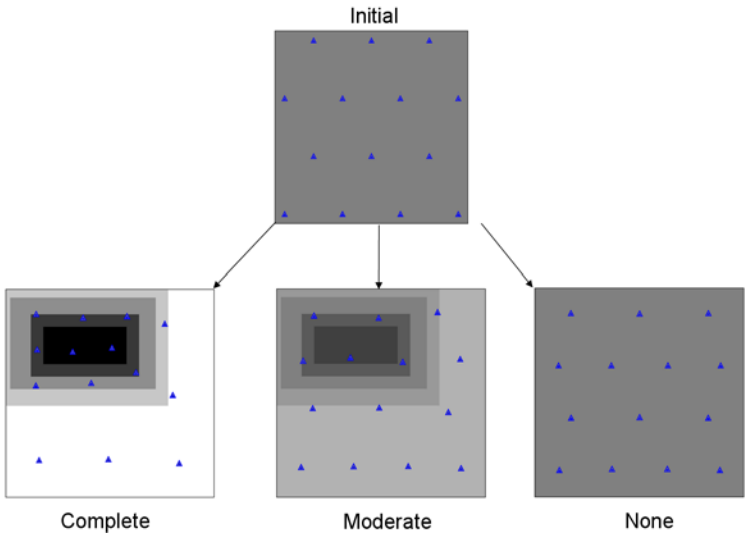


Figure 5.5 Map Reliance factor effect on initial guess.

median finds no real improvement in the minimization other than adjustments related to the location of site boundaries. The complete scenario places a premium on the level of concern in the northern portion of the site. P-median responds by moving more samples into that area to minimize the sum.

In subsurface sampling, samples are collected by corehole. A corehole location located at (x,y) will result in multiple samples collected at different depth intervals. Interest in placing a corehole at this phase depends on the likelihood of encountering elevated concentration levels somewhere in the vertical profile. Hence it is appropriate to think of corehole locations rather than individual sample locations. In Check and Cover, one can consider projecting the three dimensional concern model onto a two dimensional model by either a) taking the maximum value or b) taking the average value. Given a three dimensional raster concern model with grid nodes (x_i, y_j, z_k) map concern scale values (v) to a two dimensional grid with nodes (x_i, y_j) either by average or by maximum. Specifically:

$$\tilde{v}(x_i, y_j) = \frac{\sum_{k=1}^{N} v(x_i, y_j, z_k)}{N}$$

$$v_{\max}(x_i, y_j) = \max(v(x_i, y_j, z_k)) \; for \; k = 1, ..., N$$

Here N is the number of vertical layers in the raster model. In this regard, one may apply a traditional two dimensional p-median algorithm to an aggregated 3d concern. This is precisely the approach taken under Check and Cover.

Investigators can select the number of samples based on external factors such as cost. Alternatively, Check and Cover can indicate the relative change in the p-median minimized sum (or *p-median metric*) as the number of samples increase. As sample size increases, the effect of each additional sample becomes less pronounced on the minimized sum of concern weighted distances. From a cost perspective, one could choose the number of samples where this asymptotic effect becomes most apparent (Figure 5.6).

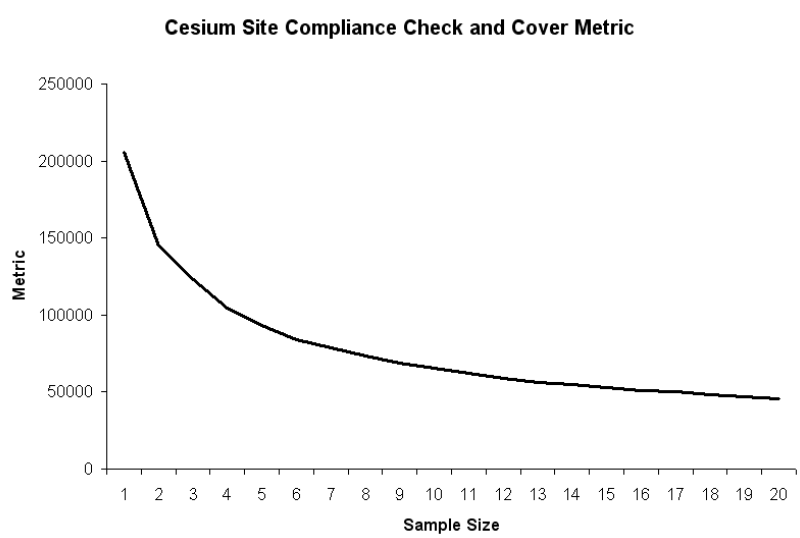


Figure 5.6. Check and Cover: sample size versus design metric.

Summary

Check and Cover provides investigators with a formal means to utilize expert judgment without requiring undue statements of probabilities or quantitative statements that investigators may feel uncomfortable providing at the time. The method is intended to provide a start to the characterization process that future sample designs may add to as new objectives emerge. The discussion continues with an explanation of how Check and Cover is implemented in SADA.

Implementation in SADA

As with the GEM implementation, Check and Cover is implemented as a prototype (McConnel, 1993) extension within SADA Version 5.0. The purpose of this prototype is to demonstrate how Check and Cover may be accessed and used. Specifically, this chapter presents how the Check and Cover design is integrated into the SADA's modeling environment, including integration into the SADA GUI and within SADA logical workflows. An example application of this implementation is reserved for Chapter 6 where a hypothetical, radiologically contaminated site is assessed for remediation and compliance within the five stages of radiological investigation. Presentation of the prototype proceeds by discussing how users encounter Check and Cover within the SADA GUI, the new Check and Cover work flow, and finally how this workflow is mainstreamed into the larger SADA workflow.

Creating a Prior CSM

Check and Cover requires the construction of a prior conceptual site model regarding investigator concerns about the location of contamination. SADA Version 5.0 provides users with a means of defining a three dimensional grid then manually assigning numerical quantities such as the scale of concern to individual grid cells. Such a model is referred to within SADA as a *User Defined Model*. Figure 5.7 and Figure 5.8 show how users create a user defined model

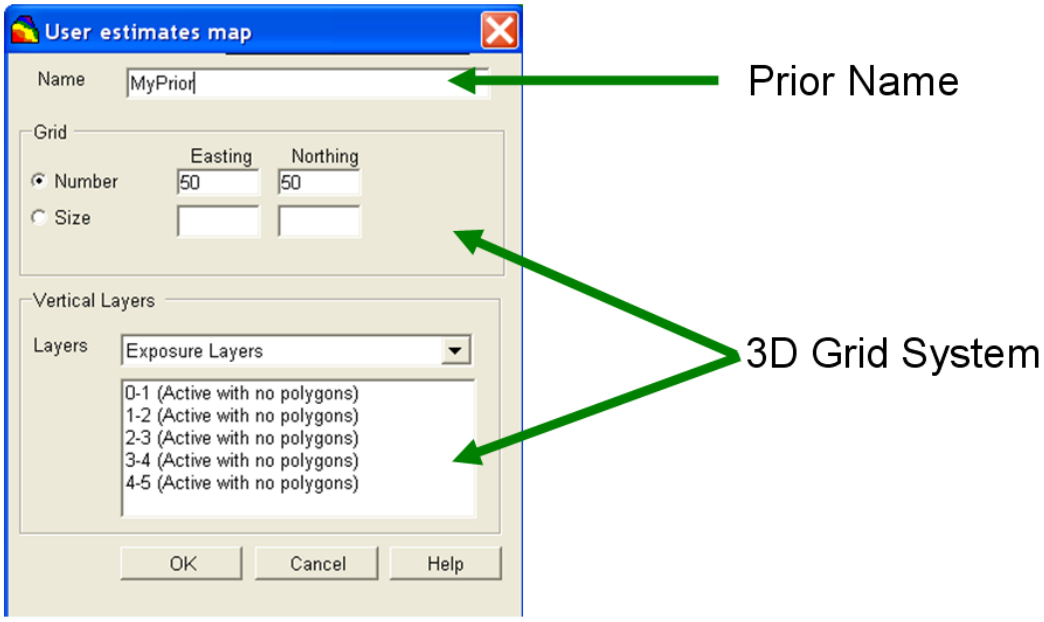


Figure 5.7 Users define the 3D raster model by specifying a grid and layer resolution.

within SADA interface. Details may be found in Stewart et al. (2009). In Figure 5.7 users begin by specifying a three dimensional grid system.

In Figure 5.8, Users are presented with an opportunity to create numerical values (i.e. scale of concern) and use paint tools to assign these to the 3D model manually.

Check and Cover Design in the SADA GUI

Check and Cover sample design is implemented within the *Develop a New Sample Design Interview* (Stewart et al, 2009). The steps that appear for Check and Cover under this interview depend on the data or model that has been selected. If the user has some actual data they would like p-median to consider in determining the arrangement of new locations, then users are met with a *See*

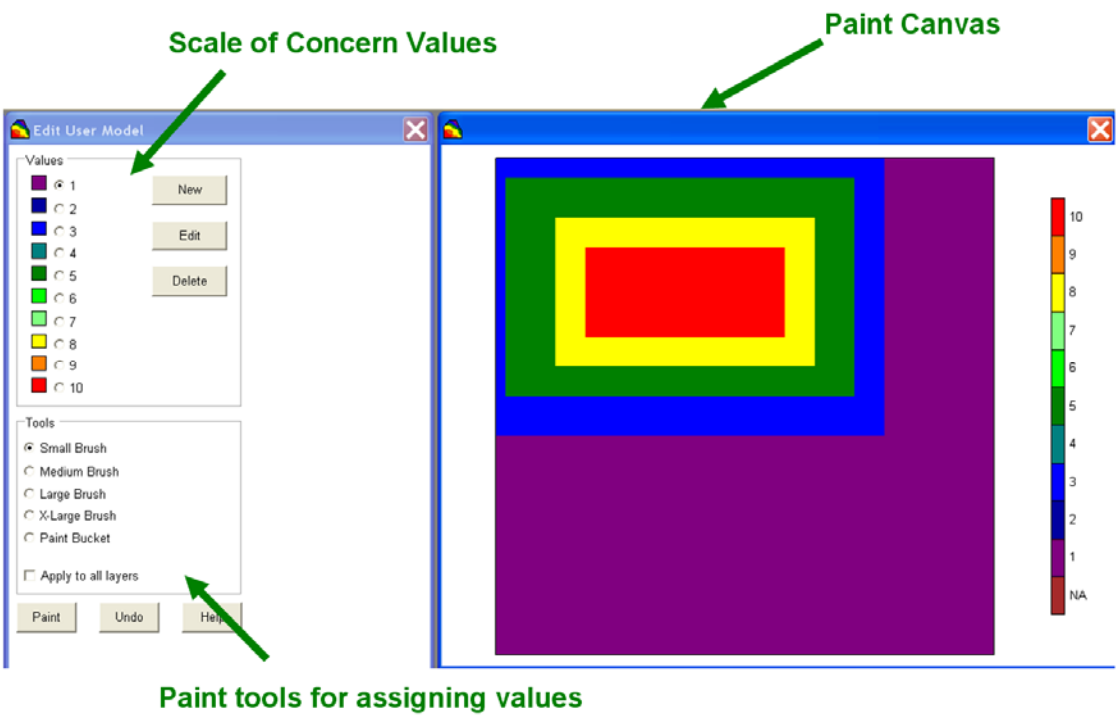


Figure 5.8. Users "paint" numerical values into the model.

the Data step. This step allows users to choose the data they wish to use. The user may then select the prior conceptual site model (Figure 5.2) as an "interpolation method" in step 5.[49] If the user has no data to consider, then the prior concern model previously developed (Figure 5.2) is selected from the list of data/models (Stewart, et al., 2009) and the first step simply becomes *See the model*. Figure 5.9 illustrates these two scenarios.

In either scenario, a method for selecting the prior concern model is provided. Under Step 4, *Set Sampling Parameters*, users will select *Check and Cover*. Figure 5.10 shows the parameter window and the associated Check and Cover options.

[49] The prior conceptual site model is clearly not an interpolation of data; however, this was a convenient location to provide the model selection. In a future public release, an additional step may be added specifically labeled "select the prior concern model."

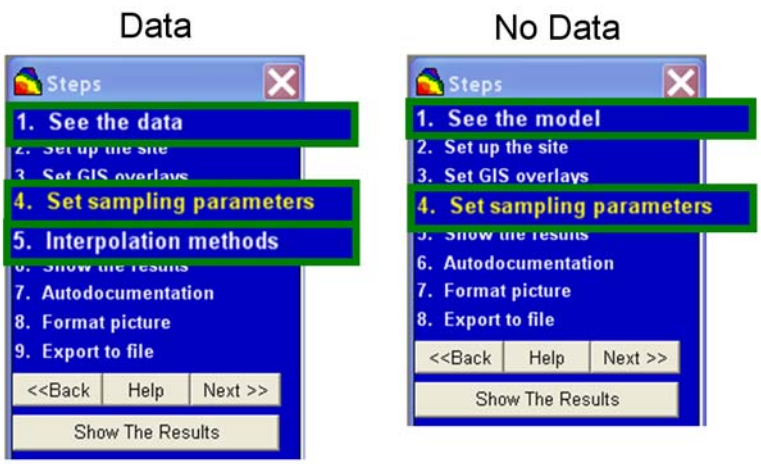


Figure 5.9 Two access points for check and cover.

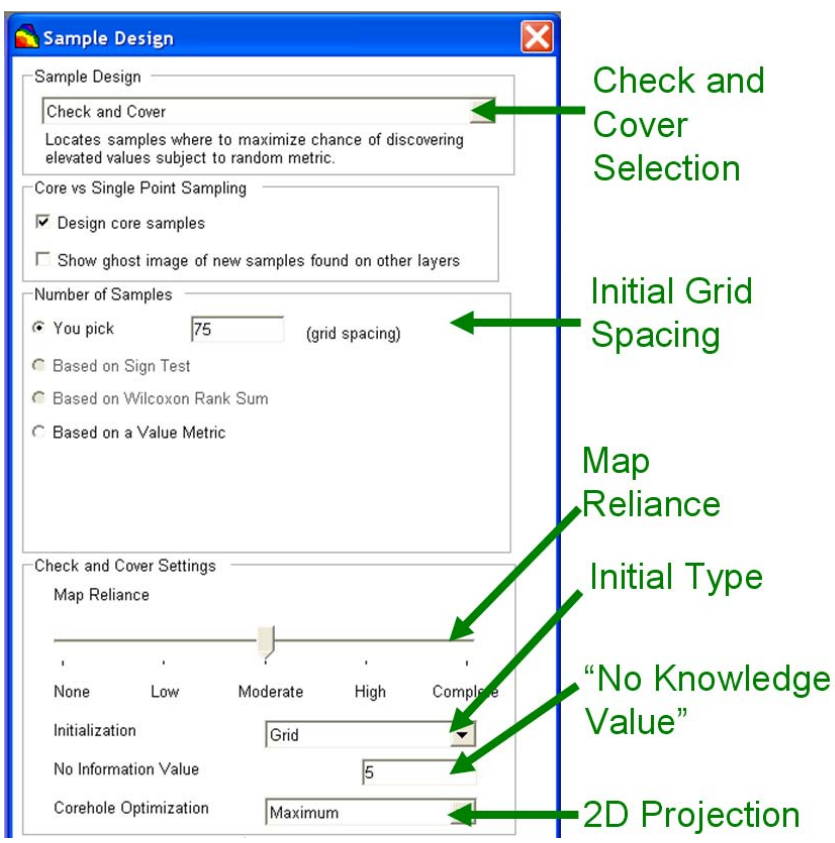


Figure 5.10 Check and Cover parameters within SADA GUI.

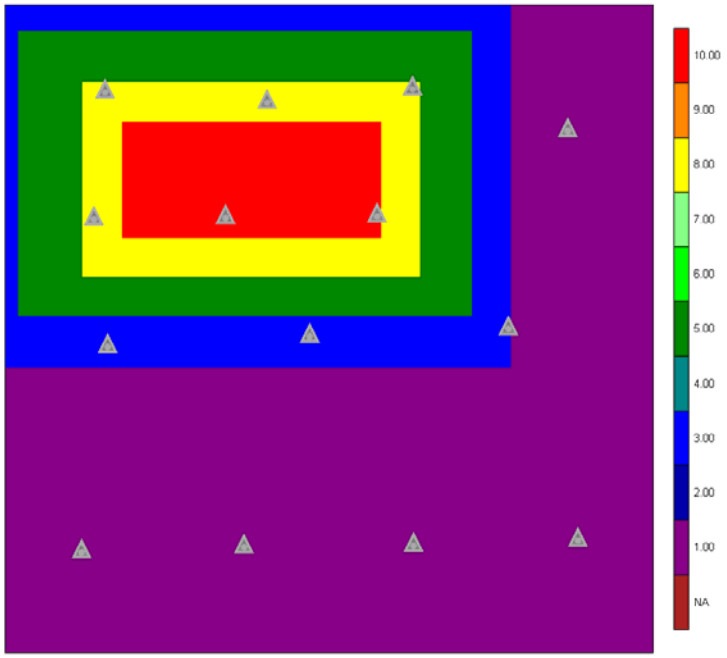


Figure 5.11 Example check and cover design.

Pressing *Show the Results* executes the Check and Cover, producing the sample design in the results window (Figure 5.11).

When the user selects *Show the Results* step or button this initiates a SADA workflow (Figure 4.9) that includes the Check and Cover workflow. The following discussion presents the resulting workflow.

Check and Cover Workflow

Unlike the GEM framework, the check and cover was written using a simple public subroutine called *CalculateCheckAndCoverSampleDesign*. The work flow for this routine is illustrated in Figure 5.12

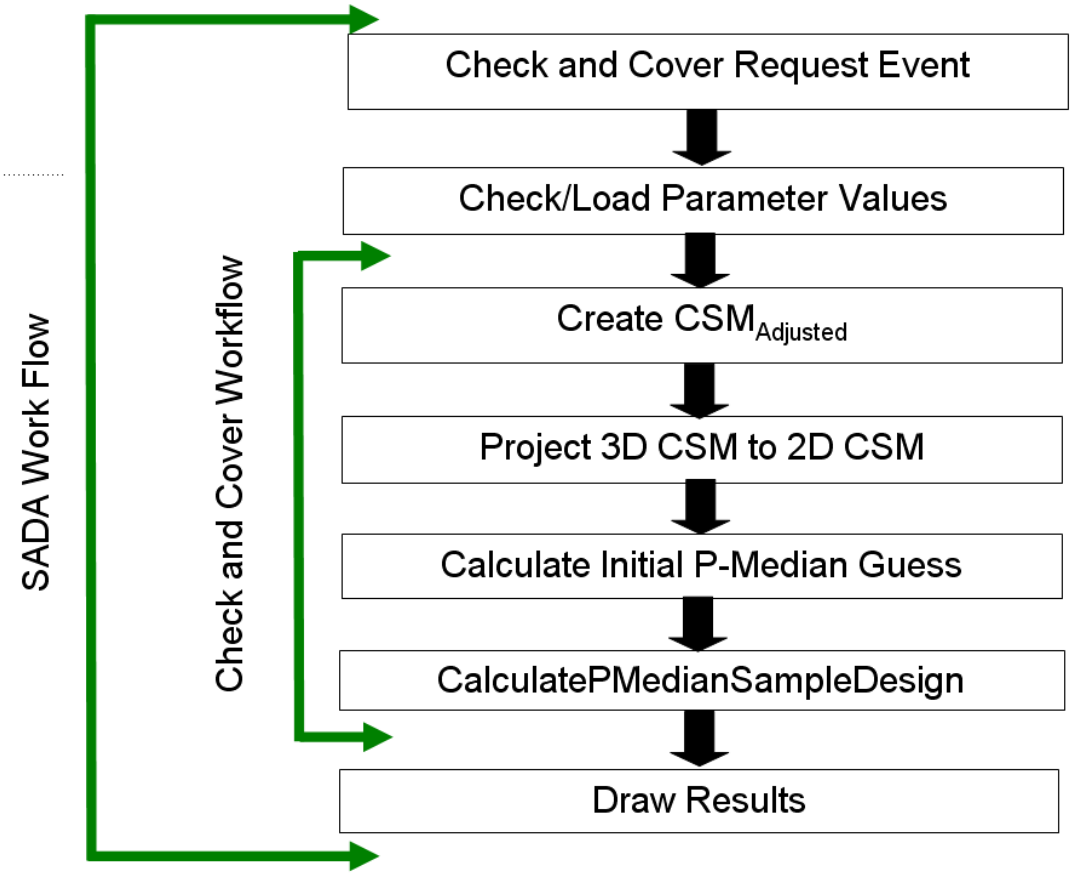


Figure 5.12 Check and Cover Workflow

The routine *CalculatePMedianSampleDesign* is based on existing p-median code[50] derived from the method by Ostresh (1978) for solving the planar p-median problem.

Conclusion

This chapter presents a new sampling strategy for incorporating expert judgment in sample designs early in characterization. This is accomplished by applying the p-median algorithm to a prior raster concern model adjusted for investigator confidence. This approach is included here as a means to initiate characterization efforts with a focus on finding and delineating contamination boundaries, a goal central to supporting the GEM framework. The next chapter demonstrates this approach and the GEM framework; it illustrates how both may be situated within the normal phases of environmental investigation.

[50] The original code was provided during personal correspondence with Dr. Bruce Ralston as part of the SADA project in 2009.

CHAPTER 6: Example Application

A hypothetical, radiologically-contaminated site is used to demonstrate the prototypes for the GEM and for Check and Cover implemented in SADA 5.0. The site, referred to as “Cesium Site”, engaged in production activities that led to Cs-137 contamination in the subsurface; investigators are interested in determining what (if any) remedial activities might be necessary to bring the site into compliance under the GEM framework’s RLR. The site is hypothetical and any similarity of Cesium Site to any real site is completely coincidental. Furthermore no insistence is made that similar scenarios must be approached in exactly the same way. Cesium Site is simply a demonstration tool.

Establishing the Synthetic Data

A complete, 3D, synthetic model of Cs-137 concentrations was created and presented in Figure 6.1. This synthetic model represents the “true” but unknown state of Cesium-137 contamination.

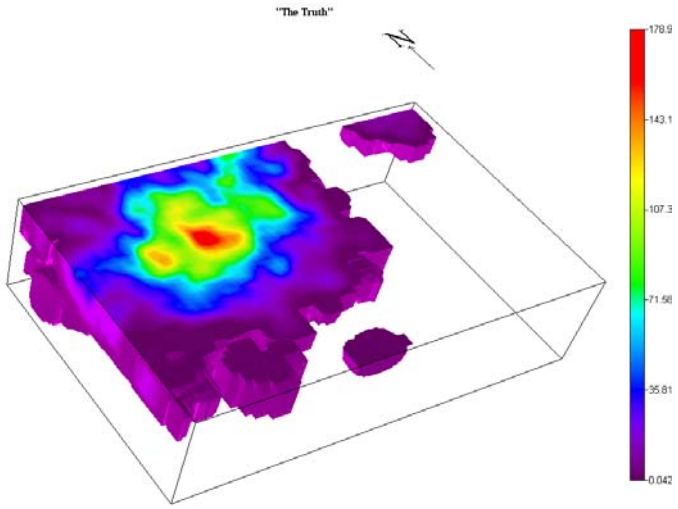


Figure 6.1 A 3D rendering of the “true” Cs-137 values at Cesium Site.

The investigation will know about this true state only in so far as it is sampled using the following method to simulate the sampling process.

Simulating the Sampling Process

A utility program (SIMSAMPLE) was created to emulate data collection from the synthetic model in two ways: laboratory and field measurements. For laboratory measurements, SIMSAMPLE returns the exact value from the true volume (Figure 6.1). There are no simulated measurement errors in this process.

For field measurements, the behavior of particular field sampling technology called a *high purity germanium* (HPGe) spectrometer is simulated. As a part of the SADA project (Stewart et al., 2009) a report was prepared characterizing the uncertainty regarding on-site measurements of Cs-137 using a 50% relative efficiency (RE50%) high purity germanium (HPGe) spectrometer (Coleman, 2009). Several factors that are normally controlled in a laboratory will vary under field conditions. These include moisture content, homogeneity of the soil sample, and count uncertainty. Coleman (2009) estimated that given these various factors one can expect the uncertainty characterized by a standard nominal deviation of approximately .22X where X is the “true value”. In this case one may simulate the information provided by an HPGe by assuming a normal probability distribution (USNRC, 2000, p6-54), centered about X with standard deviation of .22X. Suppose that the concentration at point **u** from the “true” volume (Figure 6.1) is 30pCi/g. The SIMSAMPLE would assume a normal distribution characterized by (μ,σ) of (30,6.6). In the case of sequential indicator simulation, for each indicator threshold k, the SIMSAMPLE will compute the probability that the true value is less than k by inverting the normal distribution. For example, for a threshold value of 30pCi/g, SIMSAMPLE returns a probability of 0.5. For a threshold value of 35pCi/g SIMSAMPLE returns a probability of 0.75. Hence the

SIMSAMPLE HPGe simulator returns both estimated values (average) and probability values alike. The use of field detection methods emphasizes the presence of TRIAD methodology within the GEM processes where rapid detection methods are encouraged within the decision process (Crumbling, 2004).

Defining the Exposure Scenario

For this hypothetical example, investigators are concerned about an external exposure scenario. Based on the methods in Eckerman and Ryman (1993) and DCGL calculations for a set of three dimensional, subsurface exposure units where calculated[51] under an external exposure scenario. These are reported in Table 6.1

The example will proceed by carrying out a simplified mock investigation under the five phases of investigation.

.

Historical Site Assessment

In this first stage, investigators collect all relevant information regarding the

Table 6.1 External Exposure Limits for 9 Exposure Unit Geometry/Volumes

Exposure Unit Geometry	Exposure Unit volume	DCGL
5 x 5 x 1	25	118.7
5 x 5 x 2	50	112.7
5 x 5 x 3	75	112.5
15 x 15 x 1	225	43.1
15 x 15 x 2	450	41.6
15 x 15 x 3	675	41.6
25 x 25 x 1	625	33.4

[51] Values were provided through personal correspondence with Dr. Keith Eckerman in 2010.

potential study area. This includes site history, potential sources of contamination, the identification of impacted areas, and estimates of the likelihood of contaminant migration (USNRC, 2000).

Investigation reveals that Cesium Site is a 250ft x 250ft span of property originally occupied by two buildings and two storage tanks on the northern half of the property. The facility has ceased operations and both the buildings and tanks have been decommissioned (removed). During decommissioning, it became apparent that structural damage to the tank system existed and that Cs-137 may have leaked into the soil. GIS layers for the operating facility were found (or created) and imported into SADA. Figure 6.2 shows the results of the resulting map.

The site is covered with gravel in the upper left hand corner and grass covers the remainder. A road leads into the facility from the west and turns north at the far side of the site. In the gravel area are two tanks and an underground pipe suspected of leaking Cs-137. Finally the subsurface is sandy and could permit migration of Cs-137 into the subsurface. As a result, subsurface contamination is a concern.

Scoping Phase

During this phase it is not uncommon to conduct a scoping survey which is intended to provide site-specific information based on a limited number of measurements. Often, the samples are located based on expert judgment. The results collected in this phase along with the knowledge from the HSA can be used to determine if a site has been impacted.

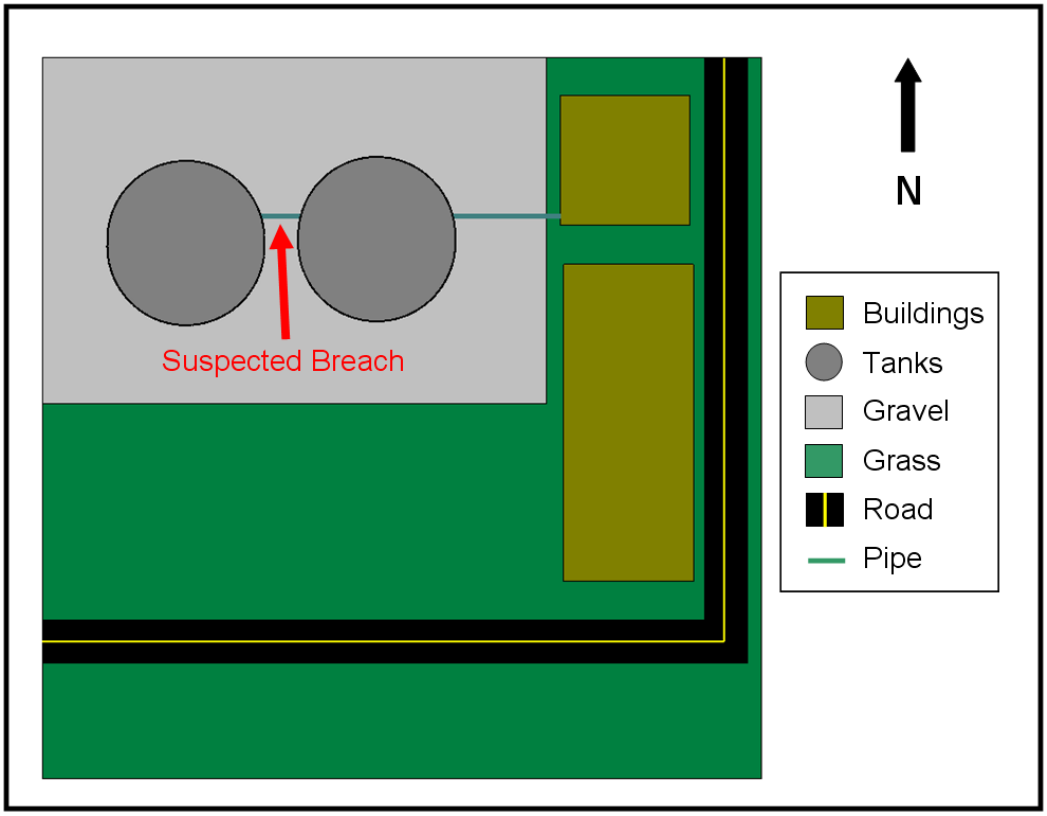


Figure 6.2 Map of Cesium Site.

For Cesium Site, investigators used the Judgmental Sampling Design in SADA (Stewart et al, 2009) to locate six coreholes near the tanks and near the boundaries of the site (Figure 6.3).

Figure 6.4 shows vertical profiles for two of the most contaminated cores (using SADA's vertical profile tool).

Finally, Figure 6.5 shows the scoping results in three dimensions.

Scoping results indicate that a reasonable depth for the site investigation is 5 feet since even the most contaminated cores reach near zero values at that depth. A

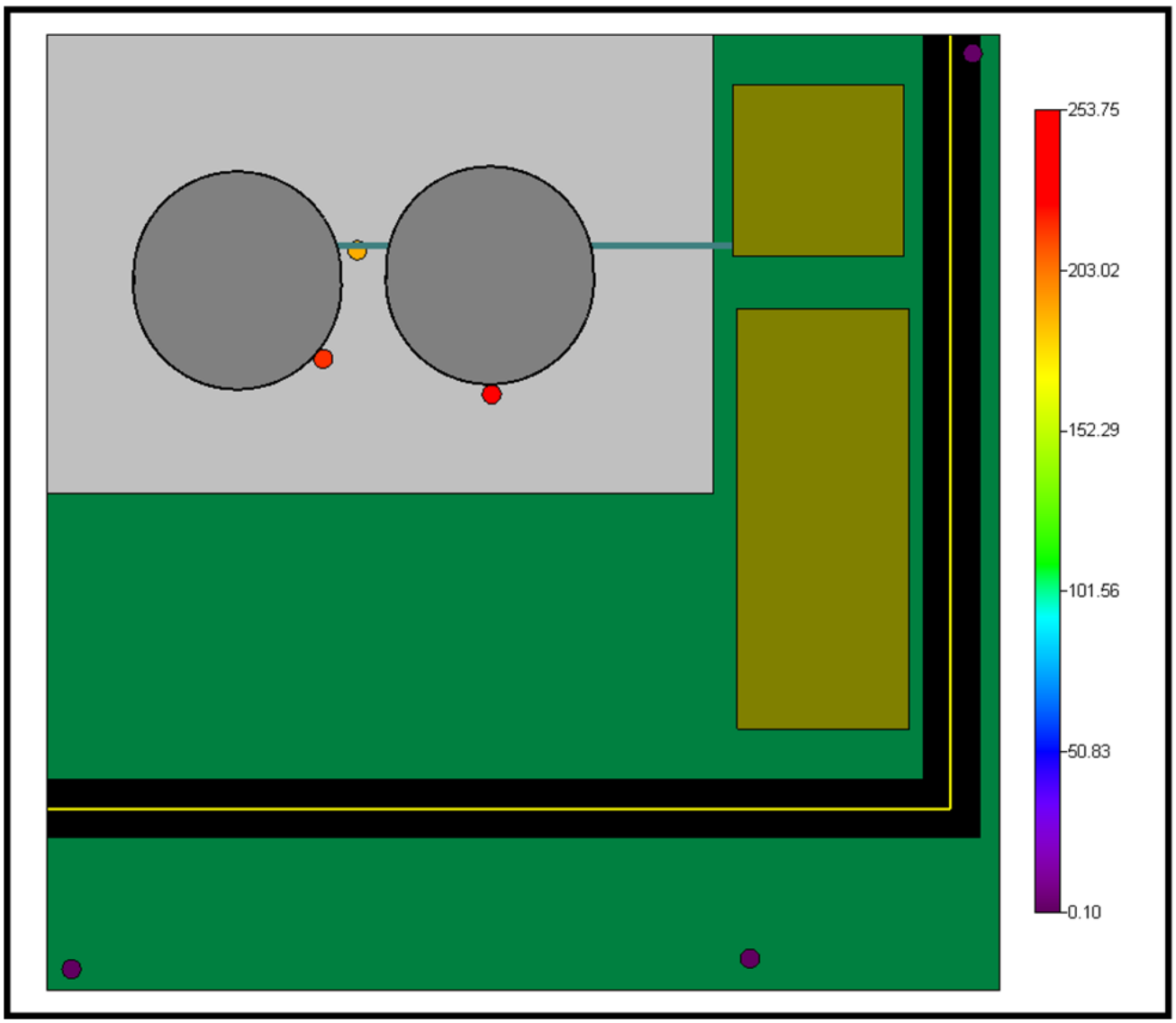


Figure 6.3. Scoping survey results.

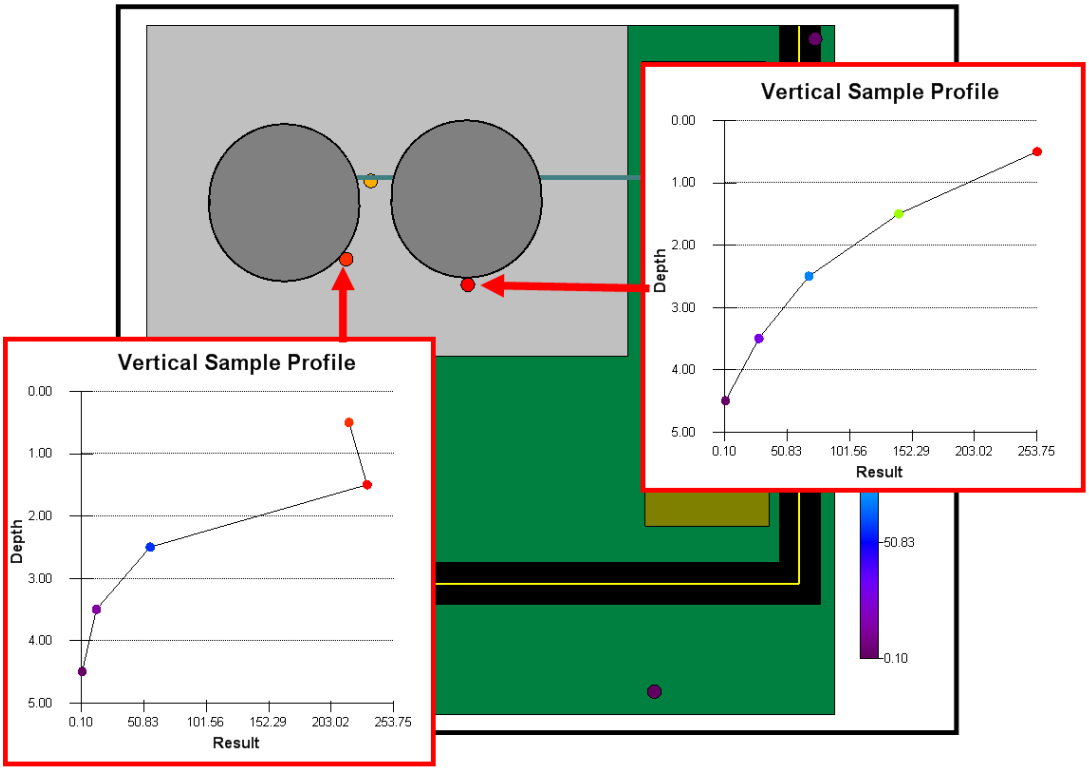


Figure 6.4 Vertical profiles for scoping survey results.

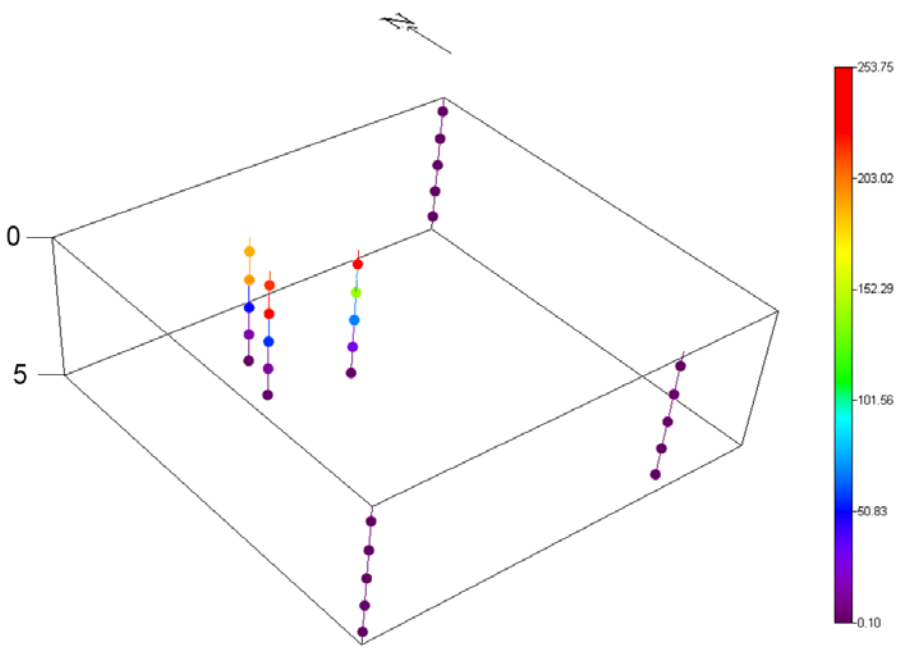


Figure 6.5 Three dimensional view of scoping results.

number of the core results indicate high levels of Cesium-137 at depth (greater than even the largest DCGL). Hence a characterization is required to determine the extent and exposure risk of the contamination.

Characterization Phase

In this phase, investigators attempt to estimate the nature and extent of contamination. This can be a highly spatial exercise with multiple objectives in play. Characterization may begin initially as an exploratory refinement on the scoping survey results but should mature into a result useable in evaluation of remedial alternatives and technologies. As discussed in chapter 5, Check and Cover can play a role particularly in early stages of characterization by using a prior conceptual site model called the “raster of concern” model to position samples. For Cesium Site, investigators created a prior concern model for where contamination may exist based on the information gained in the decommissioning phase regarding potential tank leakage and supported by findings in the scoping survey results. Figure 6.6 shows this prior CSM.

This contamination concern model is based on a grid system of 5x5x1ft cubes. This grid system will later serve as the GEM spatial grid system as well, although this is not a requirement.

Investigators used the Check and Cover sample design to locate the first round of cores. Investigators agreed to a complete level of confidence in the prior knowledge and decided to project from 3d to 2d using the vertical average. To determine the sample size, investigators relied on SADA’s *Based on A Value Metric* option to calculate the minimized p-median values for a range of grid spacings. The results are shown in Figure 6.7.

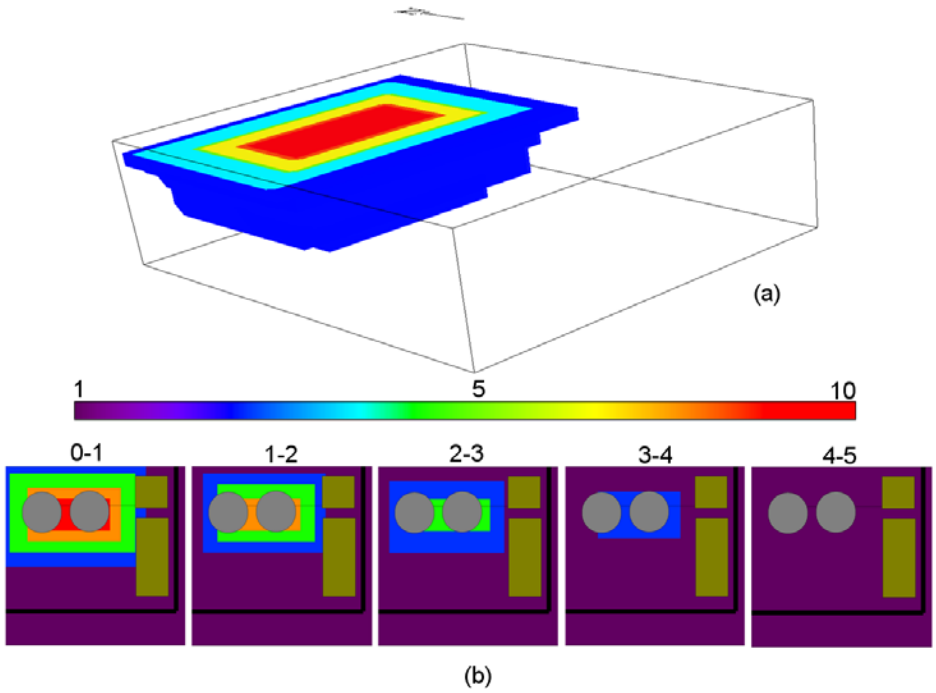


Figure 6.6 Prior contamination concern model for Cs-137 contamination shown in the SADA 3D viewer (a) for levels 2 and higher and (b) layer by layer in the SADA 2D viewer.

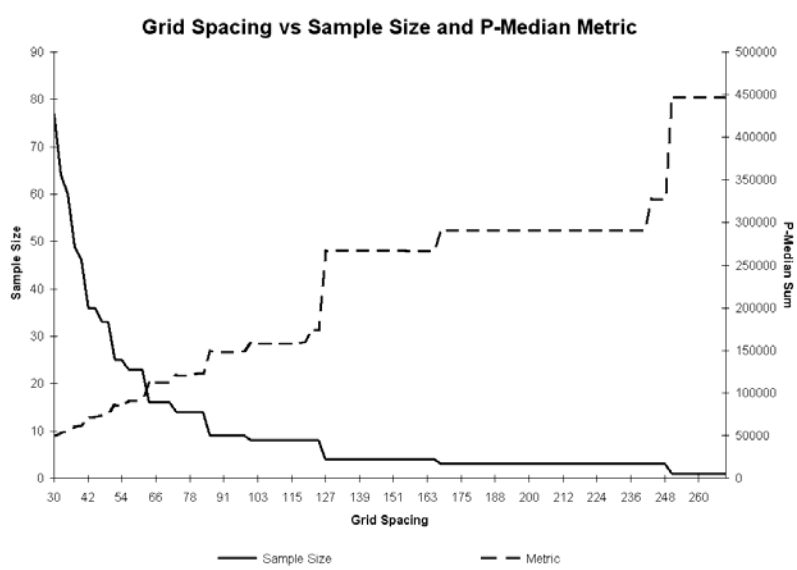


Figure 6.7 Grid spacing versus sample size and p-median values.

Notice that there are several areas along both the sample size line and the metric line where no change in value occurs. For certain spacing size changes there is not a corresponding change in the number of samples due. Consider the one dimensional case with an extent from 1 to 10 ft. A grid spacing of 2 ft allows five samples. A grid spacing of 2.1 ft also yields five samples. Plotting the sample size against the p-median metric yields Figure 6.8.

Observing that the p-median metric results behave asymptotically and that a sample size of only 9 samples produces 75% of the p-median metric reduction that a very high sample size of 77 produces, investigators select a spacing of 100ft (9 samples) to begin characterization. Execution of Check and Cover produces the sample design in Figure 6.9.

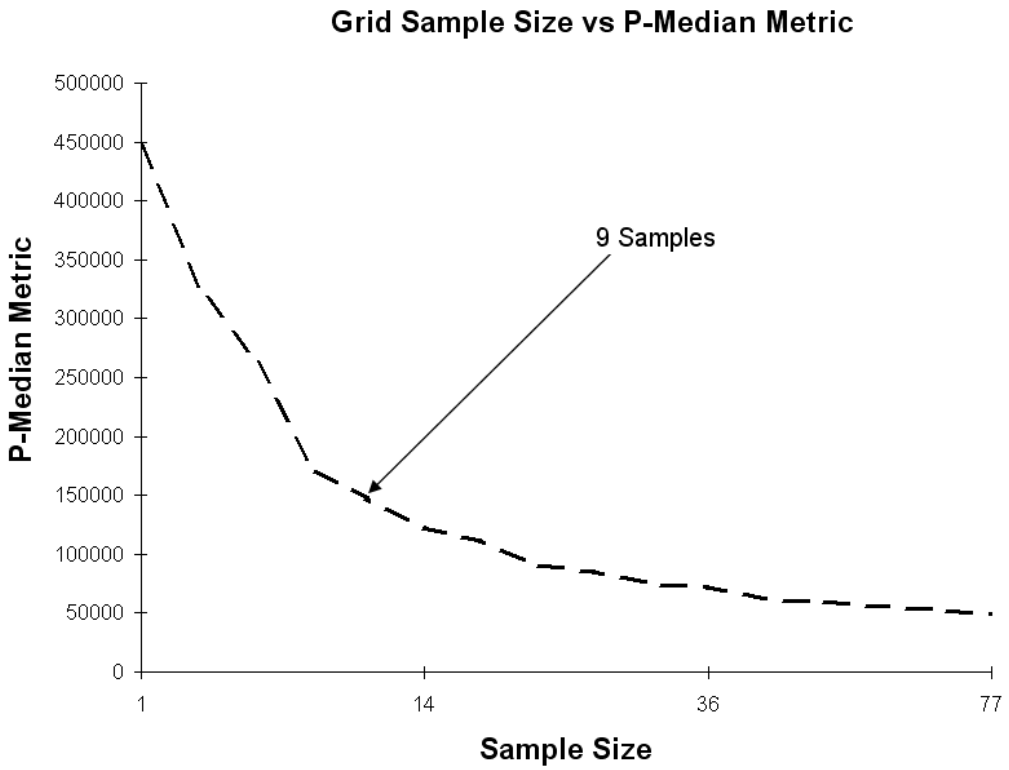


Figure 6.8 Number of samples (for grid design) and p-median metric.

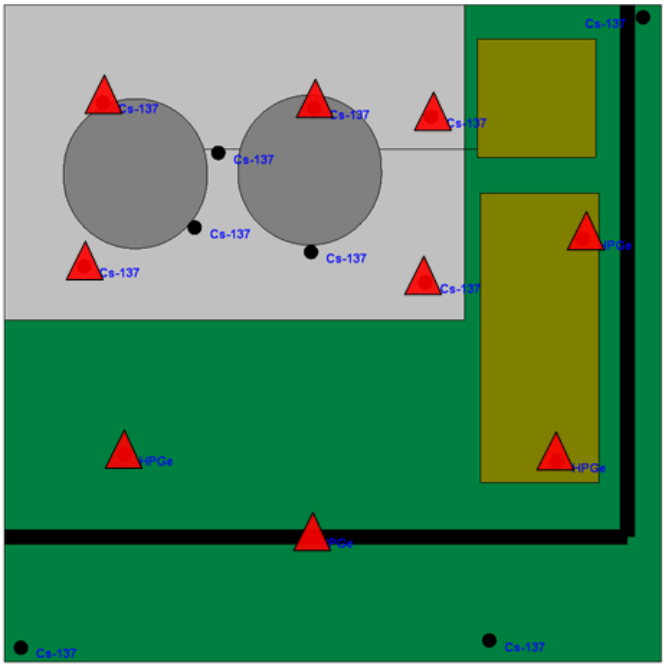


Figure 6.9 Check and Cover places 9 core locations (red triangles) based on the prior concern model.

In keeping with Triad principles, the investigators decided to use the high performance germanium detector on those areas that are likely not contaminated according to the prior concern model. Figure 6.10 shows the method used for each corehole. Points labeled as Cs-137 are lab measurement locations while points labeled HPGe are slated for field analysis.

The samples were “collected” using SIMSAMPLE. Figure 6.11 presents the results for “lab measured” Cs-137 samples.

High values were encountered near the northern border and along the edge of a fairly open area near the center of the site. This situation is illustrated in Figure 6.12 along with four new sample core locations added using SADA’s judgmental design.

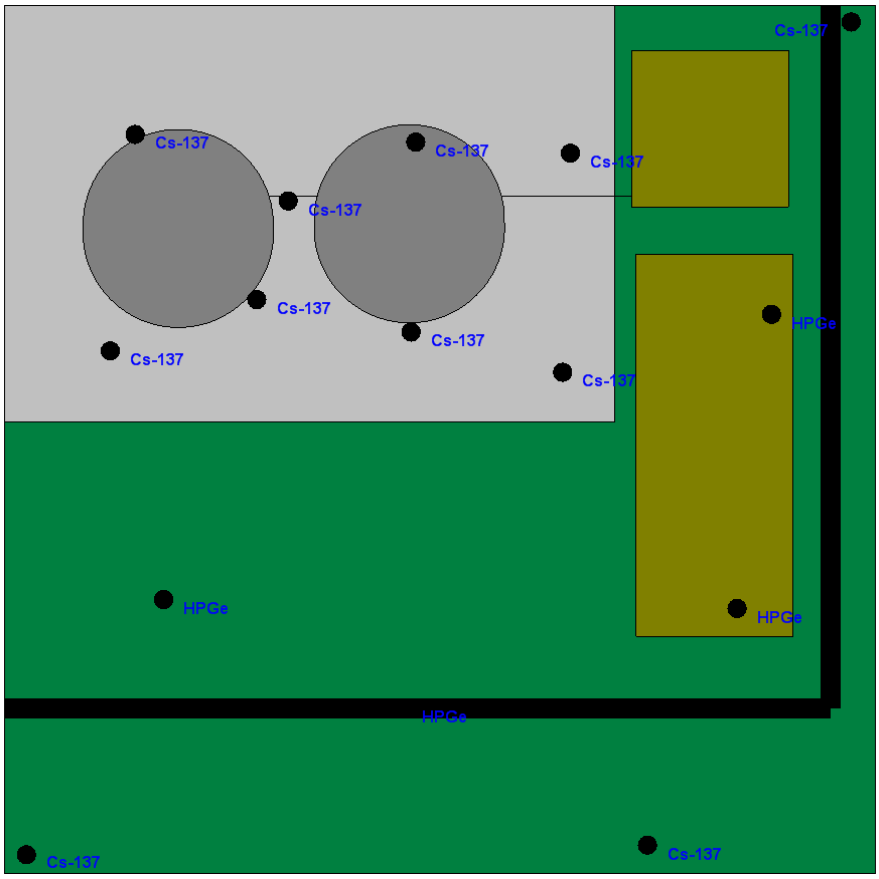


Figure 6.10 Technology selection by corehole.

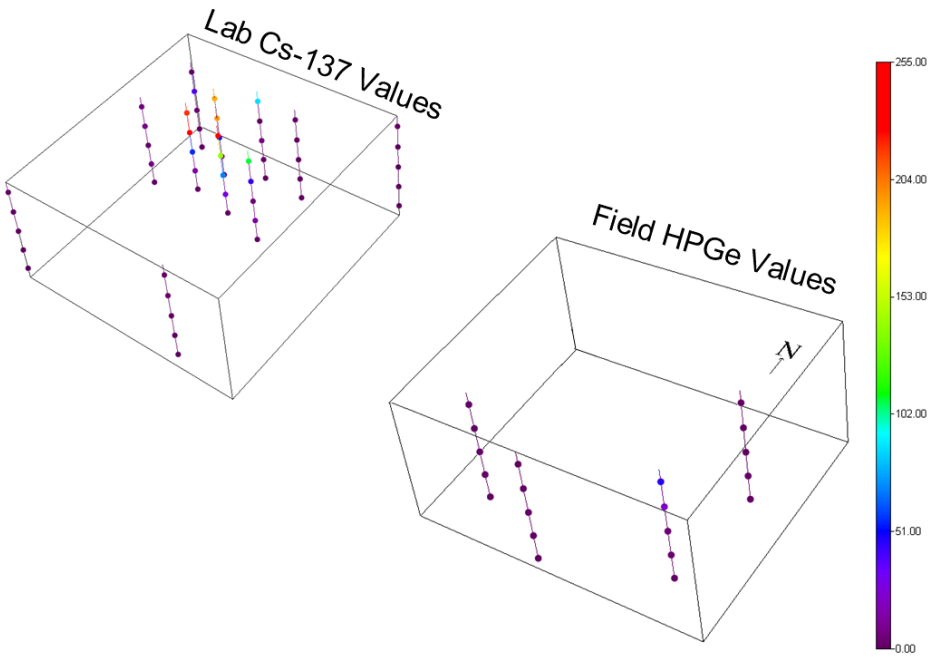


Figure 6.11 Lab and HPGe measurements in 3D viewer.[52]

[52] SADA does not permit them to be shown in the same view.

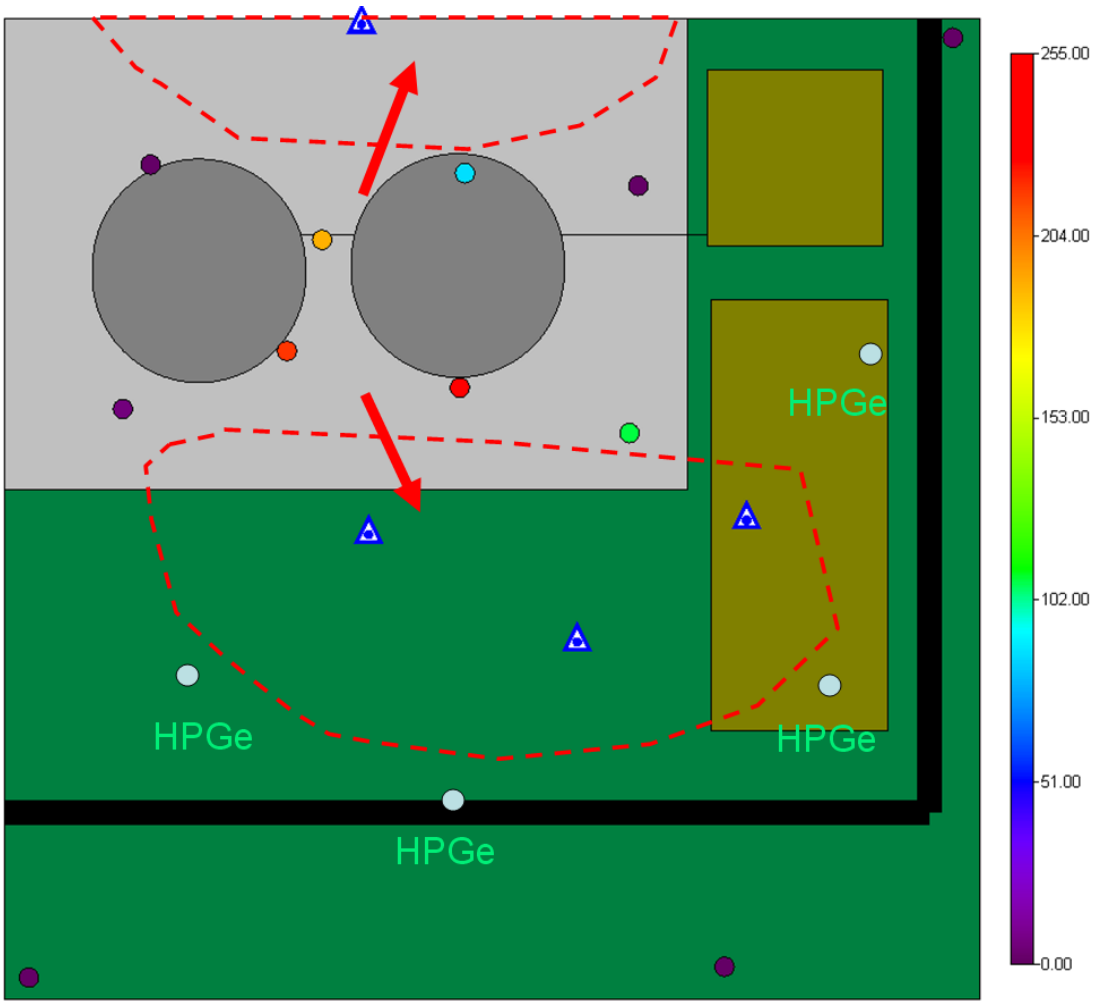


Figure 6.12 Unbounded areas (red dashed boundaries) are supplemented by additional judgmental locations (blue triangles).

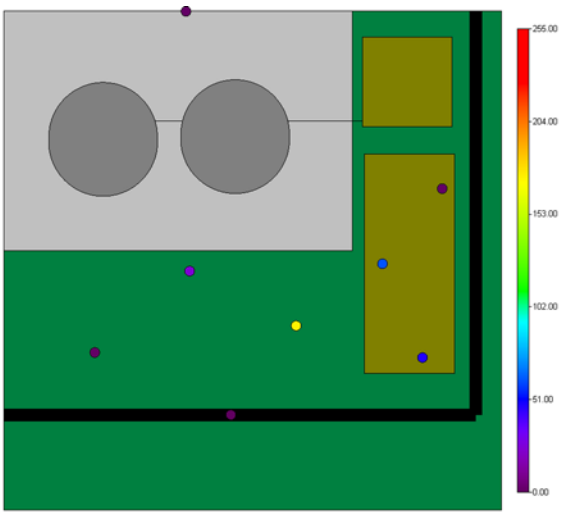


Figure 6.13 Additional HPGe characterization samples.

These cores were also collected and measured using the less expensive high performance germanium detector. The results are shown for the top layers in Figure 6.13.

The geostatistical simulation model sequential indicator simulation (Chapter 2) was developed using the data (lab and HPGe) in preparation for the RLR test against the exposure scenarios in Table 6.1. A total of fifty simulations were calculated, four of which are presented in Figure 6.14. Only values of 32pCi/g are shown to permit a view of those areas above the smallest DCGL value (32pCi/g).

Post processing these simulations to produce a contour map (averaging simulated values) yields the model in Figure 6.15 where values above 32pCi/g are shown.

Investigators now decide to apply the SCSM test to the RLR. Using a GEM spatial resolution grid cell size of 5x5x1 ft, the DCGL values for each associated geometry are encoded in the exposure unit specifications file (Figure 6.16).

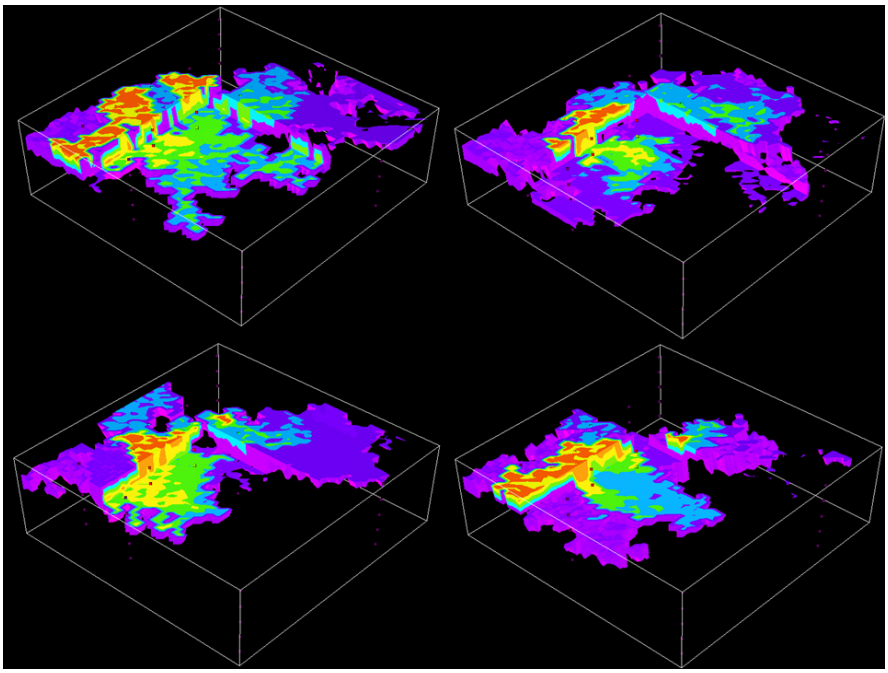


Figure 6.14 4/50 SIS simulations.

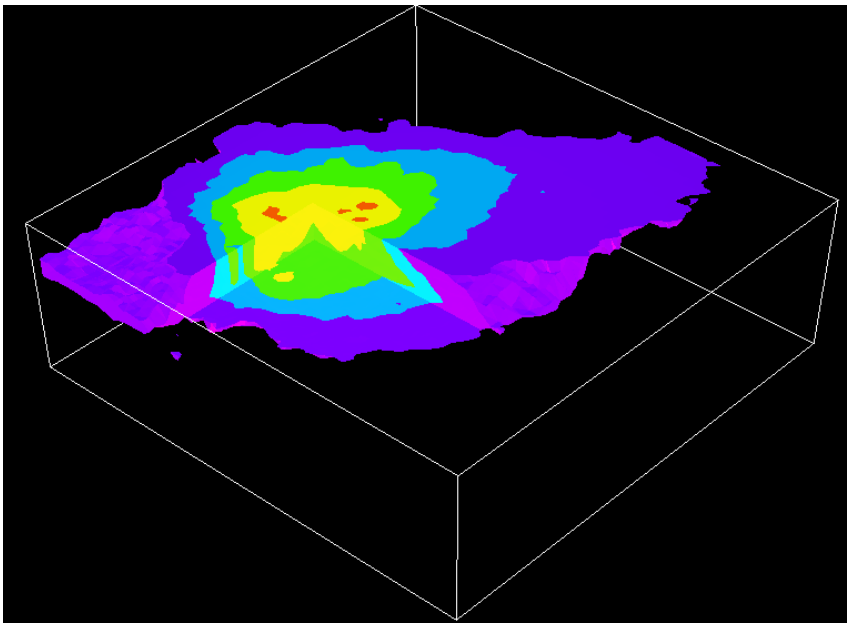


Figure 6.15. Simulation average values above 32pCi/g.

```
5, Number of Layers
7, Numberof Geometries - #horizontal neighbors/#neighbors below layer
Top,Bottom,0/0,0/1,0/2,1/0,1/1,1/2,2/0
0,1,118.73,112.70,112.54,43.12,41.61,41.58,33.37
1,2,118.73,112.70,112.54,43.12,41.61,41.58,33.37
2,3,118.73,112.70,112.54,43.12,41.61,41.58,33.37
3,4,118.73,112.70,112.54,43.12,41.61,41.58,33.37
4,5,118.73,112.70,112.54,43.12,41.61,41.58,33.37
```

Figure 6.16 Cesium site exposure unit specification file.

Investigators wish to maintain a high degree of certainty in their decision and choose a decision limit (α) of 0.1. Since Cs-137 does not occur naturally in background, the remedial design of replacing contaminated soil with clean soil should result in a back fill concentration value of 0 pCi/g. These parameters are entered in the Specify Decision Criteria step under the *Draw an area of concern map interview* (Figure 6.17).

Choosing the simulation model in Figure 6.14, under the *Select Simulation Method* (Figure 4.3), is the final step prior to pressing *Assess Current Compliance*. SADA produces the following results indicating that Cesium site fails the SCSM test for the RLR rule in Figure 6.18.

The SCSM model also produces maps of failure by exposure unit class. Figure 6.19 shows surface layer failures. Each GEM spatial resolution grid cell is colored blue if the exposure unit instance positioned on there has failed compliance, green otherwise.

These results of the SCSM test indicate that remediation will be required in order to comply with the RLR, and thus the investigation enters the remedial phase.

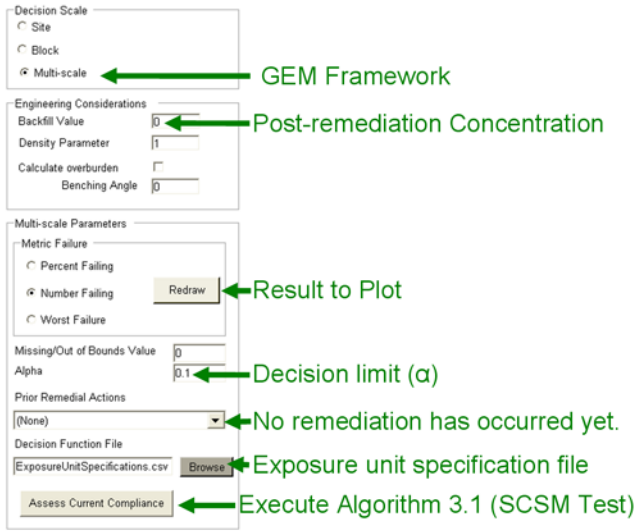


Figure 6.17 SCSM Test parameters.

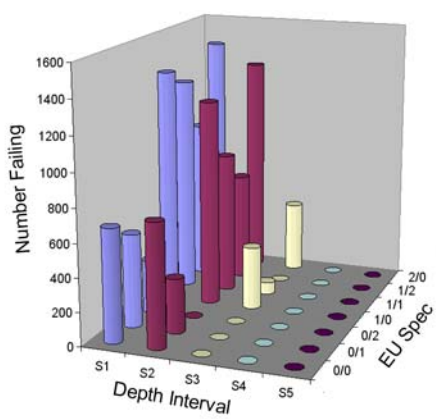


Figure 6.18 Number of exposure unit failures by exposure unit specification and by depth interval (S1 = surface layer).

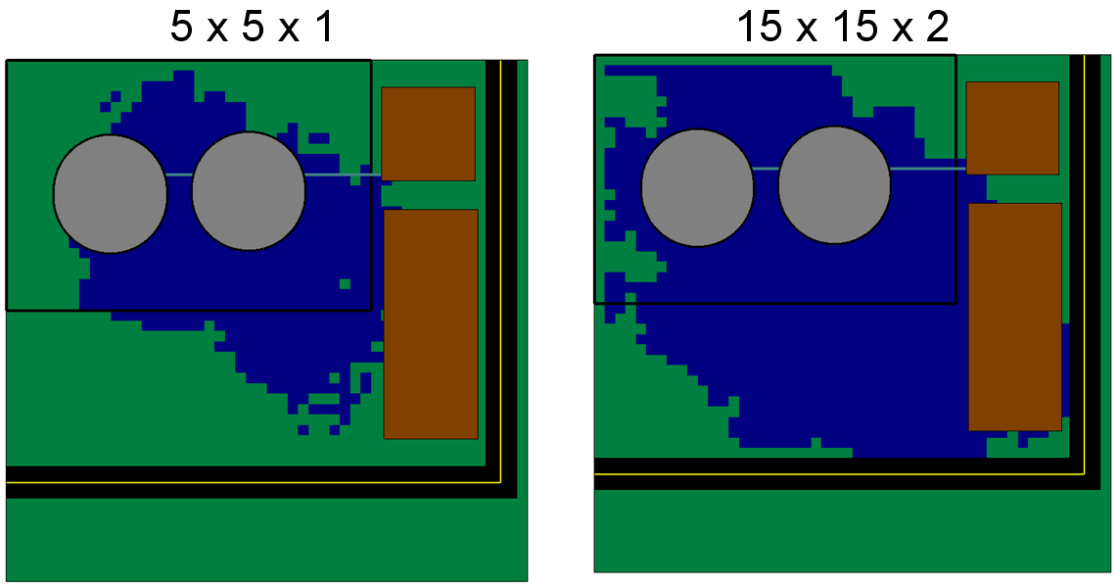
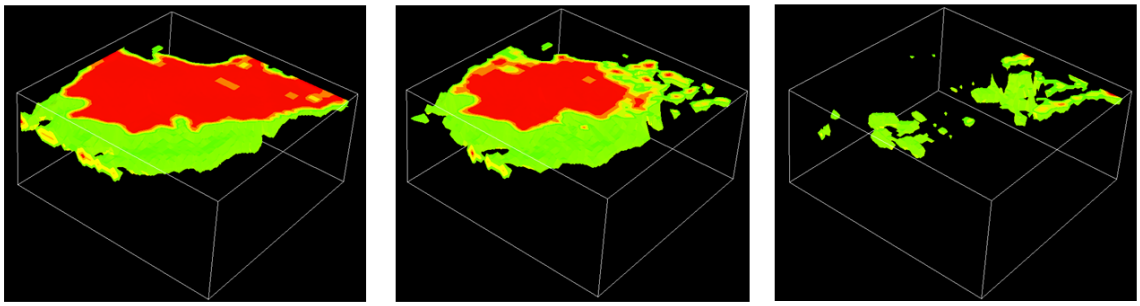


Figure 6.19 Location of EU instance compliance failures.

Remedial Phase

During this phase, investigators turn their attention to what remedial actions will allow Cesium Site to pass the SCSM test and be in compliance with the RLR. In order to build a remedial design base, investigators turn to the MrDM. Using precisely the same calibrations as seen in Figure 6.17, modelers choose *Show the Results* instead of *Assess Compliance* to execute MrDM (Algorithm 3.3). Figure 6.20 shows three results: the baseline remedial design (Step 1 of Algorithm 3.3), the final MrDM remedial design, and the volume removed from the baseline by the MrDM algorithm.

Given the cost of remediating this volume, investigators wonder if careful selection of a few more cores might decrease the remedial volume required at this high confidence level. MrsDM was applied to determine what cores (if any), if correctly estimated, might lead to a smaller remedial volume. Investigators create a set of 10 candidate locations using the Adaptive Fill design from which



(a) (b) (c)

Figure 6.20 a) Baseline design b) MrDM design c) MrDM improvement regions.

to choose the best 5. Adaptive fill places new candidate samples in the largest spatial data gaps (Stewart et al., 2009).These candidate locations are seen in Figure 6.21.

Selecting the interview *Develop a Sample Design*, specifying “Multi-scale” under *Set Decision Criteria*, and selecting *Area of Concern Boundary Design* gives investigators access to MrsDM (Chapter 4). Figure 6.22 shows the parameter set for MrsDM.

Both the simulation model and the exposure unit specifications are exactly the same as in MrDM. Execution of MrsDM identifies the best 5 based on Algorithm 3.4 and provides additional information as follows. Figure 6.23 shows the five selected locations along with the portion of the original remedial volume forecast to be removed from the remedial design by collecting data from the candidate coreholes.

Volume reductions associated with each corehole (and cumulative totals) are shown in Figure 6.24.

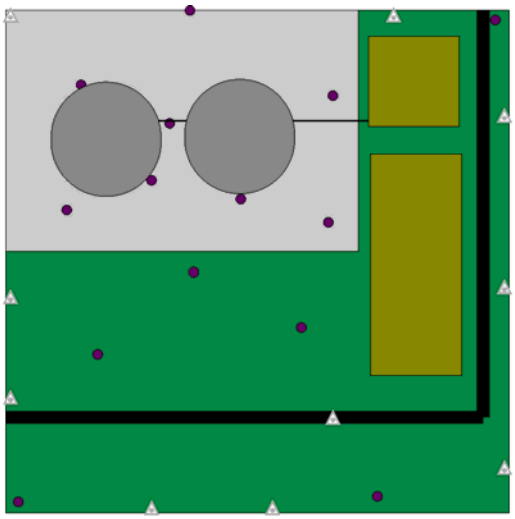


Figure 6.21 Ten Adaptive Fill locations from SADA.

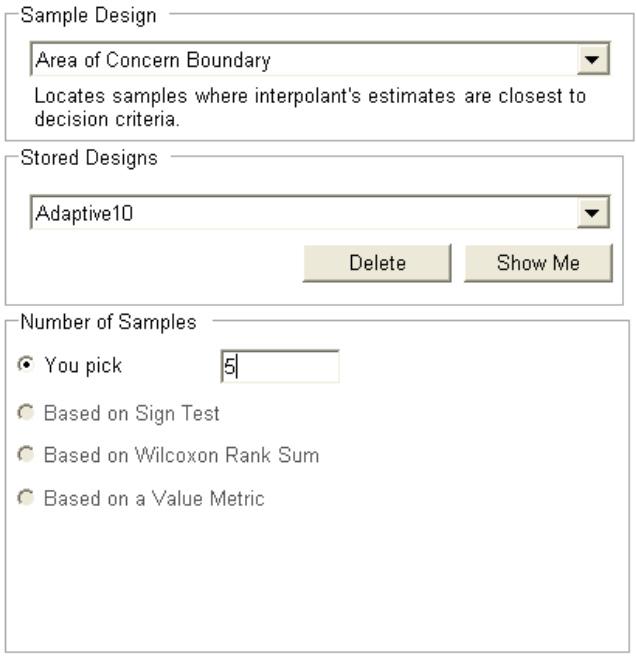


Figure 6.22 MrsDM parameters

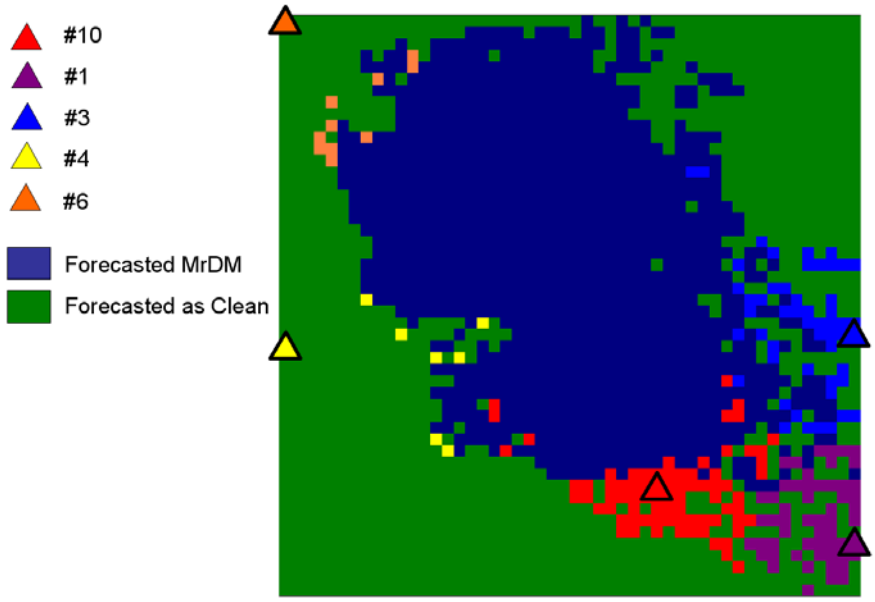


Figure 6.23 MrsDM produces winning corehole locations and illuminates those volumes forecast to be removed from MrDM design (top layer shown only).

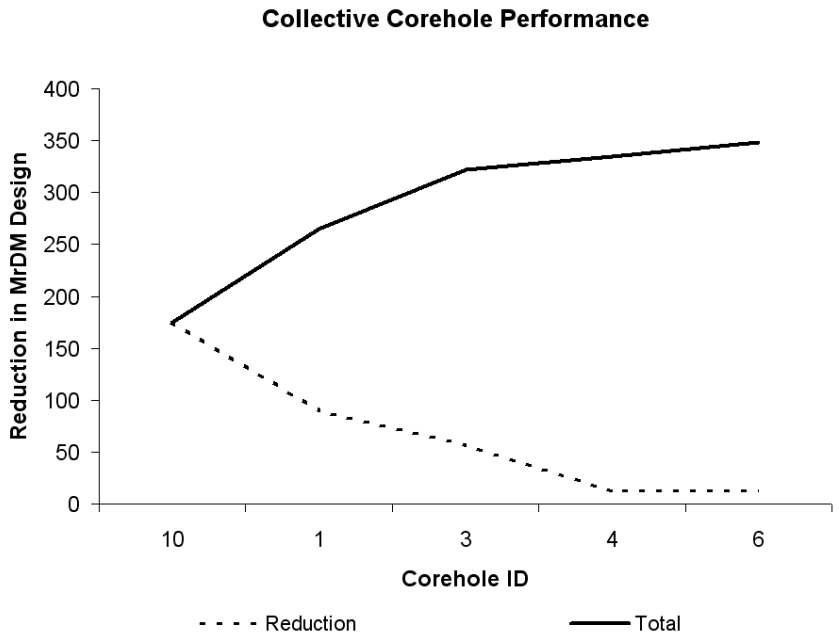


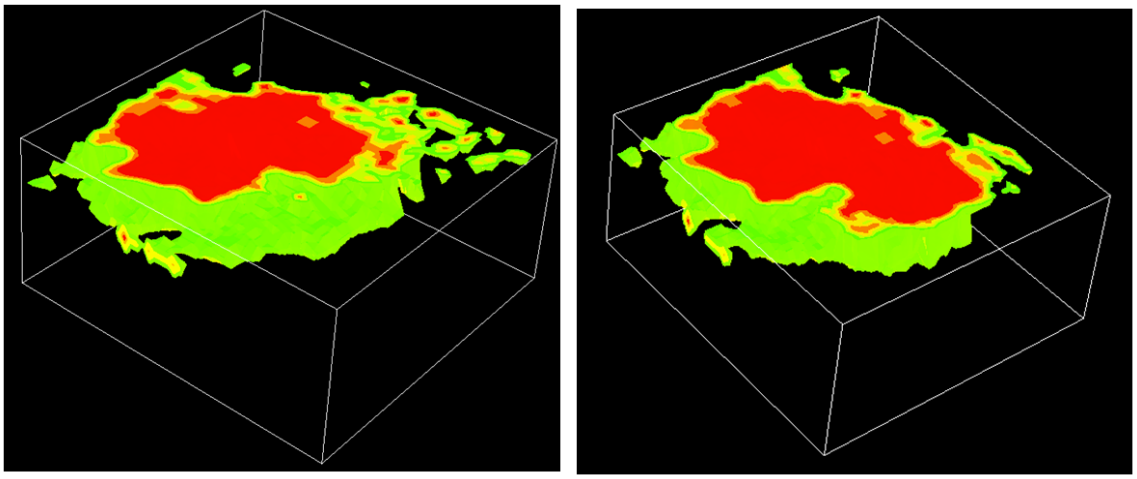
Figure 6.24 Candidate corehole performance in reducing MrDM design.

Investigators decide to select only the three highest performing coreholes (#10, #1, and #3). Lab data at these locations is collected and the simulation model is updated again. Reapplication of MrDM yields the following final remedial design in Figure 6.25 (b). The pre-MrsDM sampling round MrDM is shown again in (a) for comparison.

Implementation of this remedial design is conducted at the site. In practice, variations in the remedial design may occur due to unforeseen obstacles, unexpected contamination, and the like. If these are encountered, more data would be collected and used to update the simulation model. Application of the actual remedial design would then be used in the SCSM test. The investigation is prepared to move into the compliance phase.

Compliance Phase

Investigators enter this phase with regulators to determine if the final remedial action permits Cesium Site to pass the SCSM test for the RLR. By now, this phase should be little more than a formality. If regulators have been involved in



(a) (b)

Figure 6.25 MrDM before (a) and after (b) MrsDM Sample Design.

the development of the simulation model, the exposure scenario, and the SCSM parameters, there should be no unexpected surprises. If regulators are unable to appraise the SCSM test elements, an independent and qualified 3rd party could be tasked with an independent review. Such a review may illuminate weakness that need to be corrected in the process. If so, then investigators may need to return to the characterization phase.

For Cesium Site, the SCSM model was rerun with the actual remedial design in place. All exposure unit instances at all locations now pass the RLR using a transparent, repeatable process.

Performance Issues

Execution of the MrDM and MrsDM algorithms can be computationally demanding for a laptop or desktop computer. Within the MrDM algorithm, the algorithm calls for the systematic evaluation of every failing exposure unit's remedial design as a conditioning design for the remaining, failing exposure units in each round. Algorithm 3.3 was designed with a number of time saving measures including automatically adding all exposure units that are remedial units or are topologically isolated to the MrDM design. A significant time savings comes from recognizing that only those exposure units overlapping the remediated unit require an update. This moves the remedial design calculation for any single stage from N x N-1 calculations to only N x k where k is the number of failing exposure units sharing the same remedial units. Despite these efforts the computational times can be demanding. Figure 6.27 shows computing times for Cesium Site as a function of the number of failing exposure units at each stage for a Dell Mobile Workstation 6400M (laptop) with 4 GB of RAM and a processor speed of 2.8 GHz.

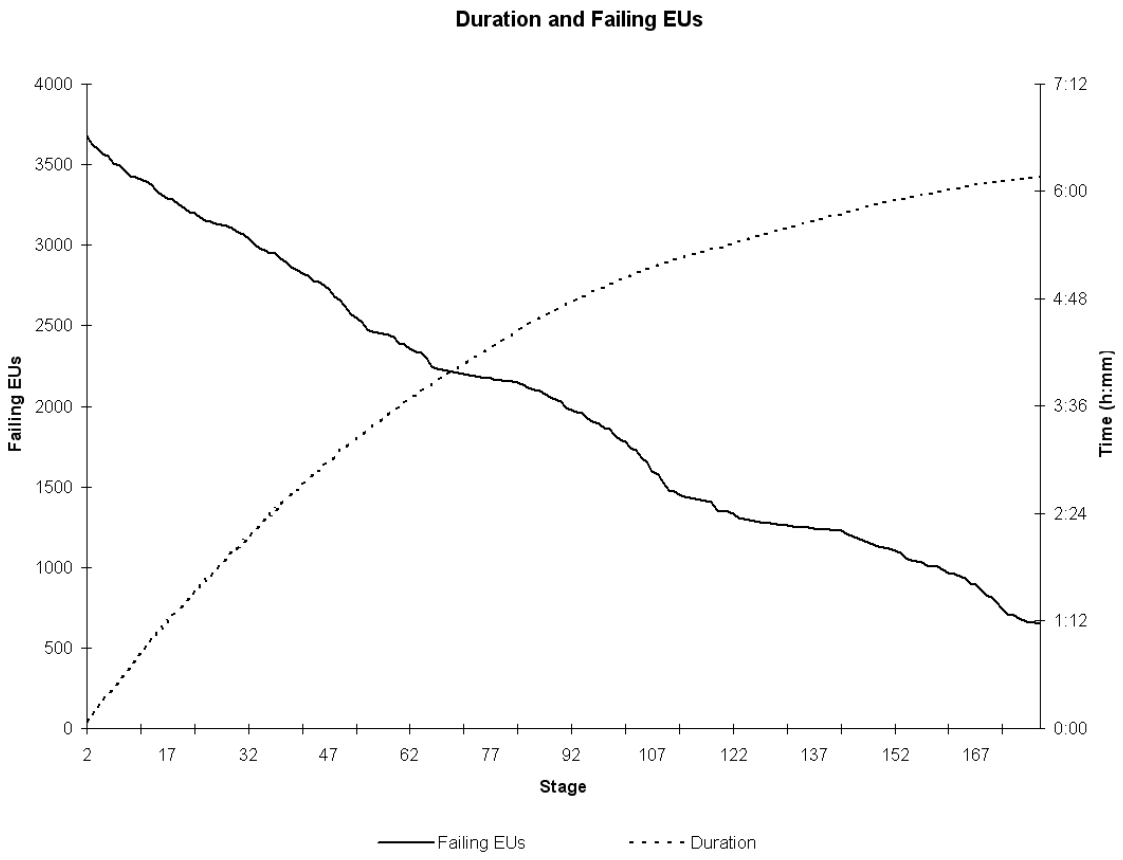


Figure 6.27 Algorithm duration by stage.

Figure 6.27 shows the relationship between calculation time and number of failing exposure units. As the number of failing exposure units decreases, the calculation time required for each stage drops off in a non-linear fashion. This is evidenced in the exponential behavior of the cumulative time line (dashed line). The total time for computing MrDM on the Dell 6400M was about 6 hours for about 3700[53] failing exposure units. This has very negative implications for MrsDM where MrDM is calculated k-m times where k is the candidate set size and m is the number of requested sample locations.

One approach to alleviate these computational burdens is to *parallelize* Algorithm 3.3. Parallelization refers to the fact that certain aspects of an algorithm are independent and can be conducted concurrently or in *parallel*. Within a multi-processor environment, code can be modified to task individual processors to work these independent tasks at the same time. Steps 5-7 of Algorithm 3.3 are certainly candidates for parallelization. While one processor could execute 5-7 for the 1st unit, a second processor could execute 5-7 for the 2nd unit right away because it does not depend on the outcome of the 1st unit. The details of parallelization are quite interesting and form an entire area of expertise within computer science. Figure 6.28 demonstrates one way in which Steps 5-7 could be parallelized for handling three failing exposure units.

In Figure 6.28 (a), a single processor handles each stage sequentially. The n+1st exposure unit cannot be addressed until the nth unit is complete. Figure 6.28 (b) shows how a quad processor could handle parallelization. The first processor acts as the master to three slave processors (#2, #3, and #4). The first processor assigns a separate exposure unit to each process. All three processors simultaneously calculate the reduction benefit of each unit and report findings back to the primary processor.

---

[53] This was approximately the number of failing exposure units after every failing EU structurally equivalent to an RU was included in the design.

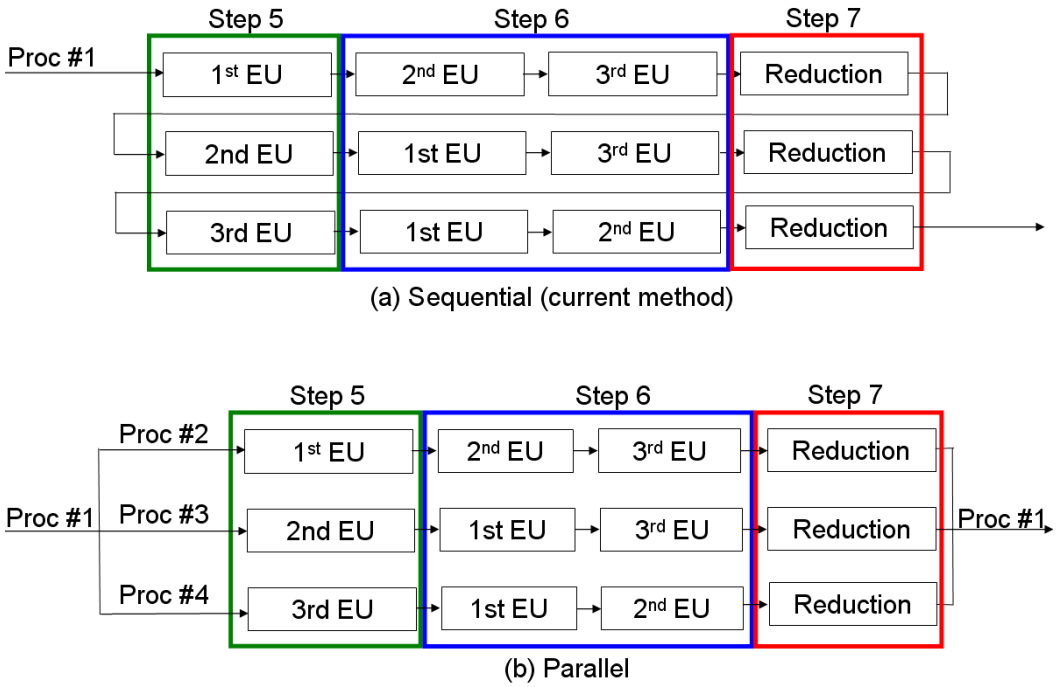


Figure 6.28 Parallelization in MrDM Algorithm

While the current version of the .NET framework permits parallelization, SADA was written in an early version where such a process was not possible. Hence parallelization was not considered in this work. According to Amdahl's law (Sun and Chen, 2010) the maximum speed up is less than or equal to linear for this section of the algorithm. For example a quad processor could theoretically reduce the calculation time to just over an hour, which is a reasonable amount of time, and MrsDM could be expected to be complete in a just a few hours.
This is a recommended next step for MrDM and MrsDM if SADA is fully upgraded to the current version of .NET. Additionally, implementations outside of SADA by others should consider parallelization of this particular process in their implementations.

Summary

Applying the Check and Cover and the GEM framework to Cesium Site demonstrated how the methods can be used to place geospatial decision support at the center of the compliance process. In this example, the phases of investigation were used to build a geostatistical simulation model, assess compliance using the RLR and SCSM, determine where to remediate (MrDM) and take additional samples (MrsDM). Triad principles within the GEM were emphasized including evolving the conceptual models (simulations, SCSM, and MrDM) and the use of field detection models in the decision process. Opportunities for improvement in computational speeds were clear and some recommendations regarding the use of multi-processor environments were discussed. Further discussions on the strengths and weaknesses are continued in the following chapter.

Chapter 7: Conclusions

Summary of the Study

The Multi-Agency Radiological Site Survey Investigation Manual provides a comprehensive decision framework for assessing compliance of radiologically contaminated surface soils and buildings with safe dose and exposure limits (USNRC, 2000). Compliance is determined by comparing radiological measurements to established limits using a combination of hypothesis testing and scanning measurements. Scanning plays a critical role in MARSSIM by allowing investigators to identify localized pockets of contamination missed during sampling as well as assess radiological exposure at different spatial scales. In the subsurface, exhaustive scanning is not possible and the process breaks down.

This dissertation presents a decision framework called the Geospatial Extension to MARSSIM (GEM) which addresses this problem by moving the problem into a geospatial modeling paradigm. The approach is based on geostatistical simulations which provide a model of uncertainty regarding the true but unknown radiological levels between sampled locations. Furthermore, geostatistical simulations permit the evaluation of uncertainty at different spatial scales (Goovaerts, 1997) and provide a surrogate for the absent subsurface scans. The goal of the GEM is to recast the MARSSIM principles of scale-dependent compliance within the context of geostatistical modeling and perch upon these models a decision system that both defines a compliance rule over 3D space and provides a test for demonstrating compliance for specific sites.

The GEM RLR is a rule set that requires that exposure scenarios (units) of any size, thickness, and depth, situated anywhere on the site, will not exceed the corresponding DCGL with a specified probability. This work recognizes that

scenarios for direct radiological exposure to the subsurface remain an unresolved issue within regulatory agencies and therefore a highly flexible method for specifying multiple scenario scales defined by exposure units that can vary in soil thickness, depth, and allowable limits is provided.

The GEM SCSM test is a method for demonstrating compliance with the regulatory limit rule for a particular site. The SCSM accepts as inputs the parameters of the RLR and a set of geostatistical simulations and outputs for each scenario a three dimensional model indicating the probability of exceeding allowable limits across the entire site. Hence if the investigators specify N scenarios, there will be N associated probability of exceedance models, one for each exposure scale. If any raster cell in any of these probability models exceeds the specified decision risk limit the site fails compliance.

The GEM MrDM provides investigators a method for determining what minimal amount of soil remediation or replacement would move the site into compliance. This amounts to a computationally demanding minimization problem that must consider uncertainty about unsampled concentrations as well as multiple and topologically integrated exposure units of varying sizes, limits, and positions over depth. The MrDM model provides a heuristic solution to this problem by first identifying a feasible solution for the soil volume and location and then sequentially improving (reducing) the design by recognizing that remediation within exposure units can have a benefit to other exposure units which have remedial units in common. The result is a three dimensional geospatial map indicating what soils to remove or remediate.

The GEM MrsDM indicates where additional core hole sampling might improve understanding of the spatial distribution of a contaminant and result in a smaller MrDM remedial design. In this approach, the investigator provides a set of candidate corehole locations. The MrsDM then simulates the collection of data at

those coreholes by assigning the median simulated value from nearest node in the simulation set. These simulated values are then added to the geostatistical model as if they were actual data and the MrDM is rerun. The corehole location that represents the worst reduction in the design in their local area is eliminated and the process repeats until the specified number of requested coreholes is reached. The MrsDM provides data on which coreholes were removed and their local performance. Additionally a three dimensional raster model is produced indicating which areas would still require remediation and identifying any reductions in the original design with the nearest corehole.

Finally, this dissertation also presents an additional sample design called Check and Cover. This corehole design strategy applies the location-allocation approach to a subjective model of concern indicating early on in the investigation where investigators are most concerned about finding contamination. Check and Cover seeks to check those locations that are of greatest concern while providing some coverage of areas considered not contaminated. This approach can mitigate the risk of a incorrect concern model by providing some coverage throughout the site. More importantly for correct prior concern models it strikes a balance between the initial interest in finding the contamination (sampling where it is likely found) with later interests in spatially bounding the contamination (sampling where it is likely not). While there is no explicit connection between Check and Cover and the GEM, both share a common goal of distinguishing between impacted and non-impacted areas. Therefore, it is anticipated that Check and Cover is supportive of the GEM framework at the early stages where expert judgment can play a valuable role in characterization.

Both the GEM and Check and Cover were implemented within the SADA 5.0 freeware package as prototypes and applied to a hypothetical, radiologically contaminated site called “Cesium Site.” Application to this site demonstrated the viability of both methods in supporting the investigation and compliance process

in two ways. First the example was carried out using the standard phases of a MARSSIM investigation (USNRC, 2000) demonstrating that the methods presented here are well situated within that regulatory method and culture. In addition, a particular connection with the emerging EPA Triad method exists here as well. Second, the SCSM test and MrDM represent the kind of conceptual model that Triad insists should be developed and evolved across the course of an investigation and ultimately used in the decision process. Secondly, the example application demonstrated that the GEM is in step with Triad emphasis on using field detection methods in the decision process. Indeed for Cesium Site, field detection results and the evolution of their associated measurement uncertainty were folded into the geostatistical simulation set with direct implications for the GEM components. Finally, the prototype application demonstrates that the method can be implemented into a publically available GIS/decision system and take advantage of geostatistical modeling algorithms already available. From a research and development perspective, implementation and application of these methods also indicates future research directions and opportunities for improvement.

Future Research

Research opportunities exist in the three areas of methodology, implementation, and other kinds of applications. Many of these research needs have already been mentioned during discussions in earlier chapters.

Methodology

The following discusses potential research and development direction for the algorithms and methods themselves.

Multiple Contaminants

The dissertation has considered only one contaminant. At some sites, there are multiple contaminants. MARSSIM is faced with a similar problem and addresses this situation with application of the *unity rule* (USNRC, 2000, p. 4-8). Suppose that N radionuclides are present in the subsurface each with a DCGL value and an average concentration for the exposure unit of interest. The unity rules says that taken together, these N radionuclides are in compliance if the sum of their ratios of their average concentration to DCGL value is less than or equal to one.

$$\frac{C_1}{DCGL_1} + \frac{C_2}{DCGL_2} + ... + \frac{C_N}{DCGL_N} \leq 1$$

In the case of the RLR/SCSM, the decision threshold is no longer a DCGL but the value 1. For the SCSM, a set of simulations would be calculated for each radionuclide. Each set would be transformed by dividing by the corresponding DCGL and the transformed sets would be added together. In other words the realization of the unity value U(**u)** at node **u** would be:

$$U^{(q)}(\mathbf{u}) = \frac{c_1^{(q)}(\mathbf{u})}{DCGL_1} + \frac{c_2^{(q)}(\mathbf{u})}{DCGL_2} + ... + \frac{c_N^{(q)}}{DCGL_N}$$

The compliance checks would be identical to those in Eq. 3.1, 3.2, and 3.3 and in the SCSM algorithm by replacing $c_{true}(\mathbf{u}_i)$ with $u_{true}(\mathbf{u}_i)$ and DCGL with “1”.

MARSSIM also suggests the use of surrogate measurements to reduce the number of radionuclides that must be sampled at each location. Under this approach, only one radionuclide is measured at every location. At some locations, the other radionuclides are also measured and ratios with the surrogate are estimated. The number of measurements to use in estimating the ratio is selected using the Data Quality Objectives (DQO) Process and based on the chemical, physical, and radiological characteristics of the nuclides and the site (USNRC, 2000, p. 4-4). In the case of the RLR/SCSM, the simulation set would first be produced for the surrogate measurement. For each of the remaining radionuclides, simulation values would be multiplied by the appropriate

ratio providing essentially a linear transform of the simulation set. Each simulation set would be processed according to the unity rule previously described.

Spatial Connectivity in MrDM

As seen in Chapter 6, the remedial designs produced by MrDM can be topologically disconnected (small islands). This occurs because the minimization problem does not contain a directive for maintaining connectivity among remedial units slated for action. A future research question could be how to constrain the MrDM by connectivity requirements. This is not necessarily a serious problem but can create practical engineering problems if investigators must burrow for small remedial locations at depth. Until this problem is solved engineers may likely disregard very small remedial areas or may sweep them together in a single removal (defeating some of the benefit of minimizing the remedial design).

Assessing the Quality of the Geostatistical Simulations

In the GEM, geostatistical simulation results are an input to the process and guidance for proper selection and evaluation of the quality of the simulations is outside the scope of this current dissertation where the framework itself was derived. Development and assessment of the simulations should be conducted by a qualified geostatistician in collaboration with environmental investigators. Geostatisticians have access to a wealth of published methods in the literature to draw from including Goovaerts (1997), Emery (2008), Simbahan et al. (2006), Lark (2002), Isaaks and Srivastava (1989), and Baraba's et al. (2001). Evaluation of the simulation set should be conducted during the phases of investigation creating a mature SCSM upon which a compliance assessment can be made. From a regulatory standpoint, should regulators consider the adoption of a GEM-like model for compliance, additional guidance will be required to assess the quality of the simulation.

Connecting Check and Cover to the GEM

In this dissertation Check and Cover is not explicitly connected to the GEM framework although each represents similar goals at different stages of investigation. The question becomes: is it possible to explicitly connect these two models? Is it possible to carry expert judgment through geostatistical simulations and through the GEM framework within the context of compliance? Certainly, incorporation of secondary forms of information into the simulation (including expert judgment) is nothing new (Deutsch and Journel, 1992; Goovaerts, 1997). The problem under the context of compliance occurs when the prior concern model under Check and Cover is “too wrong”. Suppose that under a future GEM framework, one begins simulations at the earliest stages. In other words, as soon as Check and Cover samples are collected, they are simulated with support from the prior concern model under a multi-covariant simulation model such as Markov-Bayes. Suppose however, that collected samples indicate that at least some portions of the prior concern model are incorrect. From a decision maker’s viewpoint, what is the next step? Should the prior concern be completely discounted? This is not necessarily an automatic solution as the prior could be based on years of experience with the site or historical sampling efforts. Should the prior be made to match the data? This is not necessarily rigorous in a strict Bayesian sense since the prior update would consist of a manual update of the prior followed by a second update of the prior in simulation. If you keep the prior “as is” and use it together in a co-simulation model they will essentially “compete” with each other with the prior winning out in open unsampled areas and the data winning out in regions close to the corehole. How should this be interpreted? These pose interesting and challenging research questions that would be valuable to address within this GEM context.

Comparing the GEM and MARSSIM at the Surface

Comparing the GEM outcomes with MARSSIM outcomes at the surface for a particular set of case studies may prove interesting. MARSSIM does not explicitly

indicate remedial design or associated sampling designs directly. However, a case study might indicate differences in the compliance test outcomes and when those differences could occur. For those sites compliance such a study may brightline efficiencies incurred under GEM through the use of MrDM and MrsDM.

Implementation

The following discusses potential research and development opportunities in the implementation of the GEM as software.

Speed

Application of the model to Cesium Site in Chapter 6 demonstrated the computational complexity anticipated for the MrDM and the MrsDM during derivation in Chapter 3. One future research question regards how parallelization of the MrDM algorithm may lead to substantially better computation times. As multi-core, multi-processor desk and laptop computers have become the norm, the use of parallel computing techniques is a very real possibility. Indeed, recent versions of Microsoft's .NET have recognized this opportunity and examples of parallel codes are available, such as at the Microsoft Developers Network, "Parallel Programming in the .NET Framework" (msdn.microsoft.com/en-us/library/dd460693.aspx). Chapter 6 provided a basic discussion of how MrDM could be parallelized; however, more research would be needed as different ways to parallelize the method are possible[54].

Non-cubic Exposure Units

The prototype implemented in SADA is limited only to square (cubic) exposure units. This is not a requirement of the GEM but constraining the implementation in this way facilitated quick prototyping within SADA where existing infrastructure was in place to support this geometry.

---

[54] Note that constraints on connectivity in the MrDM model could impact the parallelization algorithm presented in Figure 6.28.

MrsDM and Spatial Anisotropy

The creation of Voronoi volumes (used to determine a corehole's local neighborhood) in the MrsDM prototype does not consider anisotropic conditions in spatial correlation. Prior to moving from a prototype stage to a beta stage, modification to this code section would be required to match any anistropic behaviors modeled by the geostatistical simulations.

Other Applications

GEM and Surface Contamination

Radiological contamination at the surface is clearly the regulatory purview of MARSSIM. However, it would be interesting to consider a GEM surface application (current algorithms would apply as they are) when spatial auto-correlation is present and assumptions for statistical hypothesis testing are violated.

Non-radiological Applications

Radiological contamination is not the only kind of environmental pollution. In many cases, non-radiological contaminants such as metals, volatile organic compounds and the like also pose a threat that may vary over different spatial scales. There are no exhaustive "scanning" devices available for every kind of contaminant or scanning might be impeded due to obstacles at the surface. In these situations, the GEM may also play a role in supporting decision making. Check and Cover can also play a role similar to the one presented here.

Summary

The work presented here accomplishes three major goals. First it demonstrates how MARSSIM principles can be extended into the subsurface by shifting to a

geospatial paradigm. Secondly, it emphasizes the important role that spatial statistics can play in regulatory guidance and adds to the growing body of literature tying decision support and GIS systems. Finally, it is believed that this work provides a starting point upon which a future subsurface technical guidance may be built in collaboration with regulators, environmental scientists, and the public.

References

Ahmed, Shakeel, Dewashish Kumar, et al. (2008). Application of Geostatistics in Optimal Groundwater Monitoring Network Design. *Groundwater Dynamics in Hard Rock Aquifers*: 179-190.

Back, Par-Erik (2006). *Value of Information Analysis for Site Investigations in Remediation Projects*, Dissertation. Chalmers University of Technology, Sweden.

---. (2007). A model for estimating the value of sampling programs and the optimal number of samples for contaminated soil. *Environmental Geology*. 52:573–585.

Baraba`s, N., P. Goovaerts, and P. Adriaens. (2001). Geostatistical assessment and validation of uncertainty for three-dimensional dioxin data from sediments in an estuarine river. *Environmental Science and Technology*. 35:3294–3301.

Brakewood, Liv H. and Domenico Grasso. (2000). Floating Spatial Domain Averaging in Surface Soil Remediation. *Environmental Science and Technology* 34(18); 3837–3842

Brockett, P. and A. Levine. (1984). *Statistics and Probability*. New York, NY, CBS College Publishing.

Burke, R. (2003). *Getting to Know ArcObjects*. Redlands, CA, ESRI Press.

Butt, T. E., E. Lockley, et al. (2007). Risk Assessment of landfill disposal sites – State of the art. *Waste Management* 28(6): 952-964.

Byrd, D. and R. Cothern. (2000). *Introduction to Risk Analysis*. Rockville, MD, Government Institutes.

Carlon, C., L. Pizzol, et al. (2008). A spatial risk assessment methodology to support the remediation of contaminated land. *Environment International* 34(3): 397-411.

Cember, Herman and Thomas Johnson. (2009). *Introduction to Health Physics*, McGraw-Hill.

Coleman, R. (2009). Uncertainty Associated with On-Site Analysis of Unprepared Soil Samples Containing Cs-137 Using High Resolution Gamma Spectrometry. Knoxville, TN, *A Report Prepared for the University of Tennessee in Support of Nuclear Regulatory Commission Contract Number NRC-06-066*, "Sparse Radiological Subsurface Survey Data Analysis." 2006-2009.

Crumbling, Deana (2001a). Current perspectives in site remediation and monitoring: Using the Triad Approach to improve the cost-effectiveness of hazardous waste site cleanups. EPA 542-R-01-016. Office of Solid Waste and Emergency Response. Washington, D.C.

Crumbling, D. (2001b). Using the TRIAD approach to improve the cost-effectiveness of hazardous waste site cleanups. EPA 542-R-01-016, Washington, DC.

----. (2002). In Search of Representativeness: Evolving the Environmental Data Quality Model. *Quality Assurance: Good Practice, Regulation, and Law* 9(3): 179 - 190.

-----. (2004). Summary of the TRIAD Approach. Washington, D.C., Office of Superfund Remediation and Technology Innovation, United States Environmental Protection Agency.

-----. .M. Griffith, J., Powell, D. (2003). Improving decision quality: Making the case for adopting next-generation site characterization practices. *Remediation Journal* 13(2): 91-111.

D'Or, D. (2005). Towards a real-time multi-phase sampling strategy optimization. In *Geostatistics for Environmental Applications*. P. Renard. (ed). Berlin, Heidelberg, Springer-Verlag: 355-366.

Dakins, Maxine E., John E. Toll, et al. (1996). Risk-Based Environmental Remediation: Bayesian Monte Carlo Analysis and the Expected Value of Sample Information. *Risk Analysis* 16(1): 67-79.

Deutsch, C.V. and A.G. Journel (1992). *GSLIB: Geostatistical Software Library and User's Guide.* New York, Oxford University Press.

Delmelle, E.M.. and P. Goovaerts (2009). Second-phase sampling designs for non-stationary spatial variables. *Geoderma* 153(1-2): 205-216.

Demougeot-Renard, H., Chantal de Fouquet, et al. (2004). "Forecasting the Number of Soil Samples Required to Reduce Remediation Cost Uncertainty." *J Environmental Qurterly* 33(5): 1694-1702.

Eckerman, K. and Jeffrey R. (1993). External Exposure to Radionuclides in Air, Water, and Soil (FGR12). Washington, D.C., Office of Radiation and Indoor Air, United States Environmental Protection Agency. EPA 402-R-93-081.

Emery, Xavier (2008). Statistical tests for validating geostatistical simulation algorithms. *Computers & Geosciences* 34(11): 1610-1620.

England, Evan and Naser Heravi (1992). "Conditional Simulation: Practical Application for Sampling Design Optimization." In *Geostatistics Troia*, ed A.Soares. Kluwer Academic Publ., Dordrecht, the Netherlands

Englund, E., and N. Heravi. (1993). Conditional simulation: Practical application for sampling design optimisation. p. 613–624. In *Geostatistics Troi*a, ed. A.Soares.. Kluwer Academic Publ., Dordrecht, the Netherlands.

Englund, E., and N. Heravi. (1994). Phased sampling for soil remediation. *Enviromental . Ecology. Stat.* 1:247–263.

Franklin, R.B., Mills, A.L., (2003). Multi-scale spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. *FEMS Microbiology Ecology* 44:335-346.

Freeze, R. A., J. Bruce, J. Massman, T. Sperling, and L. Smith. (1992). Hydrogeological decision analysis. 4. The concept of data worth and its use in the development of site investigation strategies. *Ground Water* 30:574–588.

Gass, Saul. (1985). *Linear Programming Methods and Applications.* New York, NY, McGraw-Hill Book Company.

Gilbert, Richard (1987). *Statistical Methods for Environmental Pollution Monitoring*. New York, NY, Van Nostrand Reinhold.

Goovaerts, P. (1997). *Geostatistics for Natural Resource Evaluation*. New York, Oxford University Press.

---. (1999). Geostatistics in soil science: state-of-the-art and perspectives.*Geoderma* 89(1-2): 1-45.

---. (2001). Geostatistical modelling of uncertainty in soil science. *Geoderma* 103(1-2): 3-26.

Groenigen, J. W., W. Siderius, et al. (1999). Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma* 87(3-4): 239-259.

Griffith, Daniel A. (2005). Effective Geographic Sample Size in the Presence of Spatial Autocorrelation. *Annals of the Association of American Geographers* 95(4): 740-760.

Holland, J., Bartell, S.M., Hallam, T.G., Purucker, S.T., Welsh, C.J.E., (2003). Role of Computational Toolkits in *Environmental Management. Ecological modeling for resource management.* Ed. V. H. Dale. New York, Springer-Verlag.

Hortal, J., M. B. Araújo, et al. (2009). Testing the effectiveness of discrete and continuous environmental diversity as a surrogate for species diversity."*Ecological Indicators* 9(1): 138-149.

Isaaks, E., Srivastava, R.M. (1989). *An Introduction to Applied Geostatistics*. New York, Oxford University Press.

ITRC (2003). *Technical and Regulatory Guidance for the Triad Approach: A New Paradigm for Environmental Project Management*. Washington, D.C., Interstate Technology and Regulatory Council.

---. (2008). *Decontamination and Decommissioning of Radiologically Contaminated Facilities*. Washington, DC, Interstate Technology & Regulatory Council.

James, B. R. and S. M. Gorelick. (1994). “When enough is enough: the worth of monitoring data in aquifer remediation design." *Water Resource Res*. 30:3499–3513.

Johnson, R. (1996). A Bayesian/Geostatistical Approach to the Design of Adaptive Sampling Programs. In *Geostatistics for Environmental and Geotechnical Applications*. R. Srivastava . et. al. (eds). Philadelphia, PA, American Society for Testing and Materials.

Johnson, Robert, John Quinn, et al. (1997). Adaptive sampling and analysis programs for contaminated soils. *Remediation Journal* 7(3): 81-96.

Johnson, Robert, Karen Smith, et al. (1999). *The Application of Adaptive Sampling and Analysis Program (ASAP) Techniques to Norm Sites*.. Argonne National Laboratory Environmental Assessment Division Argonne, Illinois. DOE/BC/W-31-109-ENG-38-9.

Johnson, Robert, David LePoire, et al. (2005). Bayesian Approaches for Adaptive Spatial Sampling: An Example Application ANL/EAD/TM-05-01. Argonne, Illinois, prepared by the Environmental Assessment Division, Argonne National Laboratory.

Johnson, Robert and Lisa Durham (2004). Triad Case Study: Rattlesnake Creek. *Remediation: The Journal of Environmental Cleanup Costs, Technologies and Techniques* 15(1).

Juang, Kai-Wei, Wan-Jiun Liao, et al. (2008). Additional sampling based on regulation threshold and kriging variance to reduce the probability of false delineation in a contaminated site. *Science of The Total Environment* 389(1): 20-28.

Lark, R. M. (2002). Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. *Geoderma* 105(1-2): 49-80.

Linkov, I., A. Varghese, et al. (2004). *Comparative Risk Assessment and Environmental Decision Making.* Springer Netherlands: 15-54.

Lyon, B.F., S.T. Purucker, et al. (1994). The Value of Perfect Information: How Much is a Crystal Ball Worth? International Specialty Conference, Pittsburgh, PA, Air & Waste Management Association.

McConnell, Steve. (1993). *Code Complete: A practical handbook of software construction*. Redmond, WA, Microsoft Press.

McNulty, G., B. Deshler, and H. Dove. IT Corporation. (1997). Value of Information Analysis; Nevada Test Site. CONF-970335–37, US Department of Energy, Washington, DC,. 12. pp.

Meyer, P. D. and E. D. Brill, Jr. (1988). A Method for Locating Wells in a Groundwater Monitoring Network Under Conditions of Uncertainty. *Water Resource Res*. 24.

Microsoft Developers Network. (2011). Parallel Programming in the .NET Framework. Microsoft.com. (http://msdn.microsoft.com/en-us/library/dd460693.aspx) Accessed on 15 April 2011.

Miller, Harvey and Shih-Lung Shaw (2001). *Geographic Information Systems for Transportation.* London, Oxford University Press.

Miller, Irwin, John Freund and Richard Johnson. (1990). *Probability and Statistics for Engineers*, 4th Edition. Englewood Cliffs, NJ, Prentice Hall.

National Research Council (2007). *Models in Environmental Regulatory Decision Making*. Washington, D.C., The National Academies Press.

Norberg, T. and L. Rosén (2006). Calculating the optimal number of contaminant samples by means of data worth analysis. *Environmetrics* 17:705–719.

Norrman, J., Purucker, S.T., Stewart, R.N., Back, P.-E., Englelke, F., (2008). Framework for optimizing the evaluation of data from contaminated soil in

Sweden. Proceedings of ConSoil 2008, 10th International Conference on Soil-Water Systems; Milan, Italy.

Ostresh, Lawrence M. (1977). On the Convergence of a Class of Iterative Methods for Solving the Weber Location Model. *Opertations Research* 26(4):597-609.

O'Sullivan, David and David Unwin. (2003). *Geographic Information Analysi*s. Hoboken, NJ, John Wiley and Sons, Inc.

Puckett, G. and T. Shaw (2004). Triad case study: Former small arms training range. *Remediation Journal* 15(1): 79-87.

Purucker, S. T., C. J. E. Welsh, et al. (2007). Use of habitat-contamination spatial correlation to determine when to perform a spatially explicit ecological risk assessment. *Ecological Modelling* 204(1-2): 180-192.

Rügner, H., M. Finkel, et al. (2006). Application of monitored natural attenuation in contaminated land management--A review and recommended approach for Europe. *Environmental Science & Policy* 9(6): 568-576.

Saito, Hirotaka and Pierre Goovaerts (2003). Selective Remediation Of Contaminated Sites Using a Two-Level Multiphase Strategy and Geostatistics. *Environmental Science and Technology* 37.

Savelieva, E., Demyanov, V., Kanevski, M., Serre, M.L., Christakos, G., (2005). BME-based uncertainty assessment of the Chernobyl fallout. *Geoderma* 128(3-4): 312-324.

Simbahan, Gregorio C. and Achim Dobermann (2006). Sampling optimization based on secondary information and its utilization in soil carbon mapping. *Geoderma* 133(3-4): 345-362.

Sinha, P., Lambert, M.B., Schew, W.A. (2007). Evaluation of a risk-based environmental hotspot delineation algorithm. *Journal of Hazardous Materials* 149.

Stewart, R.N., Purucker, S.T., (2006). SADA: A Freeware Decision Support Tool Integrating GIS, Sample design, Spatial Modeling, and Risk Assessment. Proceedings of the Third Biennial Meeting of the International Environmental Modelling and Software Society, Burlington, Vermont.

Stewart, R.N., Purucker, S.T. Welsh, C. (2009). *An Introduction To Spatial Analysis and Decision Assistance Environmental Applications for Version*

5 User Guide. Washington, D.C., United States Nuclear Regulatory Commission.

Stewart, R. N., Purucker, S.T. (2011). An Environmental Decision Support System for Spatial Assessment and Selective Remediation. *Environmental Modelling & Software.*

Sun, Xian-He and Yong Chen. (2010) Reevaluating Amdahl's law in the multicore era. *Journal of Parallel and Distributed Computing* 70(2).

Sunila, R., E. Laine, et al. (2004). Fuzzy model and kriging for imprecise soil polygon boundaries. Proceedings of the 12th International Conference on Geoinformatics, Sweden.

Trochim, William M. (2006) The Research Methods Knowledge Base, 2nd Edition, at URL: http//www.socialresearchmethods.net/kb/.

USDOL (2008). Occupational Health and Safety Administration (OSHA), Sloping and Benching, Safety and Health Regulations for Construction (Excavations), 1926 Subpart P App B http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=10932

USEPA (1980). Interim Guidelines and Specifications for Preparing Quality Assurance Project Plans. QAMS-005/80. Washington, D.C

---. (1987a). Data Quality Objectives for Remedial Response Activities: Example Scenario. EPA/540/G-87/003. Office of Emergency and Remedial Response. Washington, D.C.

---. (1987b). Data Quality Objectives for Remedial Response Activities: Example Scenario. EPA/540/G-87/004. Office of Emergency and Remedial Response. Washington, D.C.

---. (1989a). Risk Assessment Guidance for Superfund: Volume I - Human Health Evaluation Manual. (Part A). EPA/540/1-89/002. Office of Emergency and Remedial Response, U.S. Environmental Protection Agency. Washington, D.C.

---. (1989b). Methods for Evaluating the Attainment of Cleanup Standards, Volume 1: Soils and Solid Media. Washington, D.C., Office of Policy, Planning, and Evaluation. EPA-230/02-89-042.

---. (1992a). Guidelines for Exposure Assessment. 600Z-92/001 U.S. Environmental Protection Agency, Risk Assessment Forum Washington, D.C.

---. (1992b). Preparation of Soil Sampling Protocols: Sampling Techniques and Strategies. EPA/600/R92/128. Office of Research and Development. Washington, DC.

---. (1994a). Statistical Methods for Evaluating the Attainment of Cleanup Standards, Volume 3: Reference Based Standards for Soils and Solid Media, Office of Policy, Planning, and Evaluation. EPA 230-R-94-004. Washington, D.C.

---.. (1994b). Guidance for the Data Quality Objectives Process. EPA/600/R-96/055. Washington, DC. EPA QA/G-4 (EPA/600/R-96/055). Washington, D.C.

---. (1994b). An SAB Report: Review of EPA's Approach to Screening for Radioactive Waste Materials at a Superfund Site in Uniontown, Ohio. Washington, D.C., Prepared by the ad hoc Industrial Excess Landfill Panel of the Science Advisory Board (SAB). EPA-SAB-EC-94-010.

---. (1997). Exposure Factors Handbook (Final Report). EPA/600/P-95/002F a-c. U.S. Environmental Protection Agency. Washington, DC

---. (2000a). Data Quality Objectives Process for Hazardous Waste Site Investigations. Washington, D.C., Office of Environmental Information. EPA/600/R-00/007.

---. (2000b). Guidance for the Data Quality Objectives Process. EPA QA/G-4 (EPA/600/R-96/055). Washington, D.C.

---. (2001a). Resources for Strategic Investigation and Monitoring, Office of Solid Waste and Emergency Response. EPA 542-F-01-030b.

---. (2001b). Data Quality Objectives Decision Error Feasibility Trials Software (DEFT) - USER'S GUIDE. Washington, D.C., Office of Environmental Information. EPA QA/G-4D.

---. (2001c). Requirements for Quality Assurance Project Plans. EPA QA/R-5. Office of Environmental Information. Washington, D.C.

---.. (2002a). Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites, 9285.6-10. Office of

Emergency and Remedial Response U.S. Environmental Protection Agency Washington, D.C.

---. (2002b). Guidance for Quality Assurance Project Plans for Modeling. Washington, D.C., Office of Environmental Information. EPA/240/R-02/007.

---. (2002c). Guidance on Choosing a Sampling Design For Environmental Data Collection. Washington, D.C., US Environmental Protection Agency. EPA QA/G-5S.

---. (2003). Using the TRIAD Approach to Streamline Brownsfield Site Assessment and Cleanup. Washington, D.C., Office of Solid Waste and Emergency Response.

---. (2003b). Guidance for Geospatial Data Quality Assurance Project Plans. Washington, D.C., USEPA. EPA QA/G-5G.

---. and Pennsylvania Department of Environmental Protection (2003). STATEMENT OF WORK: APPLYING THE TRIAD APPROACH TO ADVANCE LAND REUSE MARINO BROTHERS SCRAP YARD SITE.

---. (2005). Use of Dynamic Work Strategies Under a Triad Approach for Site Assessment and Cleanup – Technology Bulletin September 2005, Office of Solid Waste and Emergency Response. EPA 542-F-05-008.

---. (2005a). Case Study: Use of a Decision Support Tool: Using FIELDS and SADA to Develop Contour Maps of Contaminant Concentrations and Estimate Removal Volumes for Cleanup of Soil at the Marino Brothers Scrapyard Site, Rochester Borough, Pennsylvania. Washington D.C., Office of Superfund Remediation and Technology Innovation; Brownfields and Land Revitalization Technology Support Center.

---. (2006a). Guidance on Systematic Planning Using the Data Quality Objectives Process. EPA QA/G-4. Office of Environmental Information. Washington, D.C.

---. (2006b). Data Quality Assessment: A Reviewer's guide. Washington, D.C., US Environmental Protection Agency. EPA QA/G-9R.

---. (2006c). Data Quality Assessment: Statistical Methods for Practitioners. Washington, D.C., Office of Environmental Information. EPA QA/G-9S.

---. (2006d). Systematic Planning: A Case Study for Hazardous Waste Site Investigations. Washington, D.C. EPA QA/CS-1.

---. (2008). Triad Issue Paper: Using Geophysical Tools to Develop the Conceptual Site Model. OSWER. Washington, DC, USEPA. 542-F-08-007.

---. (2010) Triad Conceptual Site models http://www.Triadcentral.org/tech/dsp_sub.cfm?id=13 (last accessed 3/18/2010).

USOMB (2007). Final Bulletin for Agency Good Guidance Practices. M-07-07. U.S. Office of Management and Budget, Section I(3). Washington, D.C.

USNRC (2000). Multi-Agency Radiation Survey and Site Investigation Manual (NUREG-1575), Office of Regulatory Research. Washington, D.C.

USNRC (2009). Title 10 of the Code of Federal Regulations, Washington, D.C.

Van Groeningen, J.W., A. Stein, and R. Suurbier. (1997). Optimisation of environmental sampling using interactive GIS. *Soil Technology*10(2):83–97.

Van Groeningen, J.W., G. Pieters, and A. Stein. (2000). Optimizing sampling for multivariate contamination in urban areas. *Environmetrics* 11:227–244.

Vasát, R., G. B. M. Heuvelink, et al. (2010). Sampling design optimization for multivariate soil mapping. *Geoderma* 155(3-4): 147-153.

Verstraete, S. and M. Van Meirvenne (2008). A multi-stage sampling strategy for the delineation of soil pollution in a contaminated brownfield. *Environmental Pollution* 154(2): 184-191.

Viswanathan, Hari S., Bruce A. Robinson, et al. A geostatistical modeling study of the effect of heterogeneity on radionuclide transport in the unsaturated zone, Yucca Mountain. *Journal of Contaminant Hydrology* 62-63: 319-336.

Watson, Alan G. and Randal J. Barnes (1995). Infill sampling criteria to locate extremes. *Mathematical Geology* 27(5): 589-608.

Watson, Dave. (1999). The natural neighbor series manuals and source codes. *Computers & Geosciences* 25(4).

Appendix

A.1 Class SADA.clsExposureUnitSpecifications

This class manages activities and parameters associated with the base exposure unit specified by the user in the exposure unit specifications file.

Properties

Property	Type	Description
DCGL	Double	The DCGL for this EU.
HorizontalSize	Integer	The horizontal extent of the exposure unit geometry in number of remedial units.
Key	String	Unique string that identifies this unit.
LayerID	Integer	This is the collection key of the layer where EU instances should apply.
LayerZMax	Double	The bottom of the vertical layer where EU instances should apply.
LayerZMin	Double	The top of the layer where EU instances should apply.
NumberExceeding	Double	Number of EU instances exceeding the DCGL with probability greater than α.
NumberOfExposureUnits	Double	Number of EU instances associated with this specification.
VerticalSize	Integer	The vertical extent of the exposure unit geometry in number of base units.
WorstExceedance	Double	Highest probability of exceedance among EU instances.

Methods

Method	Type	Description
AreStructurallyEquivalent	Function	Returns true if the argument is structurally equivalent, false otherwise
acIsExposureUnitSpecifications	clsExposureUnitSpecifications	The clsExposureUnitSpecification to compare
Clone	Function	Creates an exact copy and returns it

Public Methods (continued)

Method	Type	Description
New	Subroutine	Creates a new instance of this class
ParameterizeUsingParameterString	Subroutine	Uses parameters string to recover previously created class
asParameterString	String	The complete set of parameters as a concatonated string
asSeparator	String	The concatonation separation character
PercentofExposureUnitsExceedingLimit	Function	Returns the percentage of exposure unit instances associated with this class failing compliance
StructuralParametersAsConcatonatedString	Function	Concatonates the entire set of parameters associated with this class
asSeparator	String	Concatonation separation character

A.2 Class SADA.colExposureUnitSpecifications

A collection of type clsExposureUnitSpecifications used to manage the set of clsExposureUnitSpecifications. This collection creates itself by reading base exposure unit specifications from the user geometry file or by recalling its previous instantiation state from the SADA file. This collection stores decision parameters such as the decision limit (alpha) and whether exposure unit instances must be entirely within the survey area. This manager also serves as a point for checking compliance over all exposure unit instances given a simulation set and provides that information back to the calling program in different formats such as number of exceedances, percent of exceedances, worst exceedance, and simply a boolean indicating pass or fail.

Properties

Property	Type	Description
CollectionName	String	The name of the collection.
CollectionFilePath	String	Full path for the text file containing base exposure unit geometries.
ComplianceGraphPreference	Integer	Encoded value defining what kind of compliance graph user prefers to see.
DecisionRiskLimit	Double	The RLR probability limit.

Properties (continued)

Property	Type	Description
PermitExposureUnitsIncludingAreasOff Site	Boolean	If True, exposure unit instances with part of their domain outside the study area are permitted.
UniqueSetOfVolumes	Double	Unique set of volumes over all clsBaseExposureUnitManagers. Used in plotting the Compliance Graph.
NumberOfUniqueVolumes	Integer	UniqueSetOfVolumes array size.
NumberExceedingByUniqueVolume	Double	Number of exposure unit instances sharing same volume and failing compliance
PercentExceedingByUniqueVolume	Double	Percent of exposure unit instances sharing same volume and failing compliance
WorstExposureFailureByUniqueVolume	Double	Exposure unit with highest probability of compliance failure among Exposure Unit Instances Sharing Same Volume

Methods

Method	Type	Description
Add	Function	Adds an already created clsExposureUnitSpecification into the collection with key string1. Returns this clsExposureUnitSpecification back.
clsExposureUnitSpecification	clsExposureUnitSpecification	The clsExposureUnitSpecification to add to the collection at location clsExposureUnitSpecification.Key

Methods (continued)

Method	Type	Description
CalculateUniqueVolumeResults	Subroutine	Creates arrays UniqueSetOfVolumes, NumberExceedingByUniqueVolume, PercentExceedingByUniqueVolume,WorstExposureFailureByUnique Volumeby canvasing compliance assessment results over all exposure unit specifications and organizing by EU volume.
adXSize	Double	Remedial unit width
adYSize	Double	Remedial unit height
adZSize	Double	Remedial unit depth
Clone	Function	Creates an exact copy of this collection and returns it.
ComplianceMet	Function	False if any clsExposureUnitSpecification contains a failing clsExposureUnitInstance, otherwise true.
Count	Function	Returns the number of clsExposureUnitSpecifications in this collection
GetEnumerator	Function	Permits “For each” logic in .NET collection browsing.
Item	Function	Returns clsExposureUnitSpecification from this collection with this key.
asKey	String	Key of clsExposureUnitSpecification to return.
ParameterizeWithParameterString	Subroutine	Used to rebuild collection using using parameters stored from previously created collections.
asEUSpecParameters	String	Concatonated string of exposure unit specifications
PopulateWithValuesFromFile	Subroutine	Accepts the user’s base exposure unit geometry file path and completely populates this collection using specifications in that file.
asFullPathandFileName	String	Specification file path
New	Subroutine	Creates an instance of this collection.

Methods (continued)

Method	Type	Description
Remove	Function	Removes the clsExposureUnitSpecification at location alIndex in the collection
asKey	String	Key of clsExposureUnitSpecification to remove.
StructurallyEquivalent	Function	Determines if the argument is structurally equivalent to this class.
acolExposureUnitSpecifications	acolExposureUnitSpecifications	The collection to compare with.
StructuralParametersAsConcatonatedString	Function: String	Concatonates parameters into a parameter string.

A.3 SADA.clsExposureUnit

This is the exposure unit instance (EU).

Properties

Property	Type	Description
DCGL	Double	The DCGL value for this exposure unit.
ApplicableDepth	Double	Depth below the current surface where this exposure unit instance is located.
ExposureUnitGlobalIndex	Integer	The grid node ID upon which this unit is centered.
ExposureUnitsIShareRUsWithString	String	A concatonated string keys from SADA.clsExposureUnit instances that share at least one remedial unit with this instance. Used for data storage purposes.
ExposureUnitsIShareRUsWith	Integer	An array containing all the keys of other SADA.clsExposureUnit instances that share at least one remedial unit with this instance.
HorizontalNeighborhoodSize	Integer	The horizontal size of the exposure unit defined by the number of RUs to the side of the RU unit where this EU.
Key	String	The unique key for this EU.
MaximumNumberOfEUSThatMightBe-MovedToComplianceIfIAM	Integer	The number of EU instances overlapping this EU that are out of compliance.
MaximumNumberOfRUsThatMightBe-RemovedFromBaselineIfIAM	Integer	The total number of remedial units that could be removed from the baseline if remediation of this unit were to completely remove the need for any other remedial unit remediation in any other overlapping exposure unit.

Properties (continued)

Property	Type	Description
MostRecentProbabilityCalculation	Double	The most recently calculated probability of exceeding the DCGL for this unit given any remedial designs currently imposed.
MyLocationInEUArray	Integer	The index of this unit within an array of SADA.clsExposureUnits
NeighborhoodEastWestIndexRange	Point	Given the property value HorizontalNeighorhoodSize, this is the index of the furthest remedial unit in this domain to the west (Point.x) and east (Point.y) (read only).
NeighborhoodNorthSouthIndexRange	Integer	Given the property value VerticalNeighorhoodSize, this is the index of the furthest remedial unit in this domain to the north (Point.x) and south (Point.y) (read only).
NeighborhoodVerticalIndexRange	Integer	Given the property value VerticalNeighorhoodSize, this is the index of the furthest remedial unit in this domain to above (Point.x) and below (Point.y) (read only).
NumberOfAdditionalRUsItWould-TakeToReachCompliance	Integer	This is the number of additional remedial units within the domain of this unit that would need to be added to the global remedial design to locally meet compliance (read only).
NumberOfEUsMovedToCompliance-WithMyRemediation	Integer	Given the set of remedials required to move this unit into compliance, this is the number of overlapping exposure units that would also be moved into compliance without further remediation (read only).
NumberOfFailures	Integer	This is the number of failures of this unit in the set of geostatistical realizations (read only).
NumberOfRemedialUnitsCleanedInMy Area	Integer	This is the total number of remedial units slated for remedial action within the spatial domain of this unit (read only).
ProbabilityLimit	Double	This is the probability decision limit.
ProbReductOnEUsNotComplying-BecauseOfMe	Double	This is the sum of the probability reduction experienced by all exposure units as a result of this unit moving into compliance through remediation. Used as a tie breaker (read only).

Properties (continued)

Property	Type	Description
ReductionInThetaCardinality	Integer	This is the reduction in the baseline design because of this unit's remedial design submitted as a conditioning design.
RemedialReplacementValue	Double	The value to use for a remediated cell.
RemedialUnitCleanedInGlobalScenario	Boolean	An array indicating whether which remedial units within this spatial domain are included in the proposed global design.
RemedialUnitIncluded	Boolean	An array indicating whether which remedial units within this spatial domain are included in the analysis at all.
RemedialUnitIncludedEncodedInString	String	A concatonated string of the values in RemedialUnitIncluded. Used for data storage purposes.
RemedialUnitIndices	Integer	The global indices of remedial units found in this exposure unit spatial domain.
RemedialUnitIndicesEncodedInString	String	A concatonated string of the values in RemedialUnitIndices. Used for data storage purposes.
RemedialUnitVolumes	Double	An array of remedial unit volumes for those remedial units found in this exposure units' spatial domain.
RemedialUnitVolumesEncodedinString	String	A concatonated string of the values in the RemedialUnitVolumes.
TotalExposureVolume	Double	This is the total volume within the spatial domain of this exposure unit (read only).
TotalExposureVolumeInsideStudyArea	Double	This is the total exposure volume within this exposure unit's spatial domain and also within the study area (read only).
TotalExposureVolumeOutsideStudyArea	Double	This is the total exposure volume within this exposure unit's spatial domain and but outside the study area (read only).
VerticalNeighborhoodSize	Integer	The horizontal size of the rectangular exposure unit defined by the number of units below the unit where this exposure unit situated.

Properties (continued)

Property	Type	Description
ExposureUnitSpecificationKey	String	This is the key of the clsExposureUnitSpecification on which this exposure unit instance was based.
VolumeOfEUsMovedToCompliance-WithMyRemediation	Double	The total volume of other exposure units moved to compliance automatically with the remediation of this unit.
XnodeIndex	Integer	The x column where this unit is centered in the GEM spatial resolution grid.
YnodeIndex	Integer	The y column where this unit is centered in the GEM spatial resolution grid.
ZnodeIndex	Integer	The layer where this unit begins in the GEM spatial resolution grid.

Methods

Method	Type	Description
DetermineMyRemedialDesignGiven_ AGlobalRemedialScenario	Subroutine	This routine returns the additional remedial units this unit would require to meet compliance (Algorithm 3.2?)
adbSADAFile	Dao.Database	SADA File
abGlobalRemedialScenario	Boolean (Array)	The global remedial design
adGlobalAverageRUValues	Double (Array)	RU Simulation Averages
aclsErrorReport	clsErrorReport	SADA Error Reporting System
aclsSimulationResult	clsSimulationResult	Class containing simulations
afrmAdvancedProgress	frmAdvancedProgress	SADA advanced progress form
alProgressBarToUse	Integer	Progress bar in form to use
aclsSADAConstants	clsSADAConstants	SADA's class of constants
abMakeUpdatedProbabilityPermanent	Boolean	T = make probability official
abMakeRemedialUnitsCleanedPermanent	Boolean	T= actually add F=simulate add
alAdditionalNumberOfRUSToRemediate	Integer(Array)	Additional RUs from local design

Methods (continued)

Method	Type	Description
EstablishMyselfCompletely	Subroutine	This routine completely specifies the exposure unit by receiving from the calling routine all its property values as well as a grid to EU mapping array that let's the GEM spatial grid know what Eus each node is assigned too.
aclsGridDefinition As clsGridDefinition	clsGridDefinition	GEM Grid System
acolLayers	colLayer	GEM Layers
alMyGlobalIndex	Integer	GEM grid node for this unit
adApplicableDepth	Double	Depth where this unit is positioned
alGlobalInclude	Boolean (Array)	Remedial unit include ids
adMyDCGL	Double	DCGL
adMyProbabilityLimit	Double	The probability limit for compliance
adRemedialReplacementValue	Double	The post remedial concentration
alMyHorizontalNeighborhoodSize	Integer	This EU's horizontal size
alMyVerticalNeighborhoodSize	Integer	This EU's vertical size
aclsErrorReport	clsErrorReport	SADA's error reporting system
adMissingOrOutOfBoundsValue	Double	Concentration value to use for out of bound regions.
asExposureUnitSpecificationKey	String	EU specification on which this EU is based.
asGridToEUMapping	Integer (Array)	A map of the GEM nodes to all EU mappings
alMyLocationInEUArray	Integer	EU's position in the clsGEMStructure.ExposureUnits array.
EstablishMySelfFromStoredResults	Subroutine	This routine completely specifies the exposure unit by receiving stored specifications previously calculated and kept in a file.
asExposureUnitStructuralString	String	Concatonated string of EU property values.
asSeparator	String	Concatonation character
aclsGrid	clsGridDefinition	GEM grid system
acolLayer	colLayer	GEM layers

Methods (continued)

Method	Type	Description
ExposureUnitComplianceResponse_ ToAGlobalRemedialScenario	Subroutine	Calculates the probability that the exposure unit will exceed the DCGL given the current global remedial design. This is returned to the calling routine as the argument adUpdatedProbability and also whether it passes in the argument abCompiles. The exposure unit property *MostRecentProbabilityCalculation* is not updated unless calling routine requests it. This is so that certain proposed designs can be tested without affecting the exposure units current probability of failure. Algorithm?
adbSADAFile	DAO.Database	The SADA file
abGlobal_RemedialUnitIsCleaned	Boolean	The global remedial design
adUpdatedProbability	Double	Returns probability of exceeding the DCGL
abComplies	Boolean	If Probability < ProbabilityLimit then true, else false
aclsSimulationResult	clsSimulationResult	The set of simulations
aclsErrorReport	clsErrorReport	SADA's error reporting system
aclsSADAConstants	clsSADAConstants	SADA's class of constants
abMakeResponsePermanent	Boolean	T = MostRecentProbabilityCalculation is updated.
ExposureUnitParametersConcatonated-AsString	Function	Concatonates the exposure unit parameters into a string using the separator character asSeparator.
asSeparator	String	Concatonation separator character
InCompliance	Function	Returns True if MostRecentProbabilityCalculation<= ProbabilityLimit, false otherwise.
IShareTopologyWithThisUnit	Function	Returns true if aclsExposureUnit shares remedial units with this unit, false otherwise.
aclsExposureUnit	clsExposureUnit	The exposure unit to make a comparison against.

Methods (continued)

Method	Type	Description
New	Subroutine	Creates an instance of this collection.
SetExposureUnitsIShareRUsWithArray	Subroutine	Splits the property *ExposureUnitsIShareRUsWithString* into the property array *ExposureUnitsIShareRUsWith*.
SetExternallyMostRecentProbability-Calculation	Subroutine	Permits an external routine to set the value for *MostRecentProbabilityCalculation***.**
adValue	Double	A probabability value
SetMyDecisionParametersOnly	Subroutine	In some situations, a user may request evaluation of an RLR that is identical structurally (grid, layers, unit sizes, etc) to one already established except for these parameters. This permits the code to quickly update just these parameters and begin evaluation.
adMyConcentrationLimit **adMyProbabilityLimit** **adRemedialReplacementValue** **adMissingOrOutOfBoundsValue**	Double Double Double Double	DCGL Probability limit Post remedial concentration value Concentration value to use for RUs falling off site.
TheMetricForSeverityOfMyCompliance-Failure	Function	Returns the metric indicating the severity of compliance failure (probability of failure).

Methods (continued)

Method	Type	Description
TestMyRemedialDesignNeedsGiven-AGlobalRemedialScenario	Subroutine	Allows calling routine to test a remedial design without affecting any current status of this exposure unit.
adbSADAFile	DAO.Database	SADA File
abGlobalRemedialScenario	Boolean(Array)	The global remedial design
adGlobalAverageRUValues	Double (Array)	RU simulation Averages
aclsErrorReport	clsErrorReport	SADA's error reporting system
aclsSimulationResult	clsSimulationResult	Class containing simulations
afrmAdvancedProgress	frmAdvancedProgress	SADA advanced progress form
alProgressBarToUse	Integer	Progress bar in form to use
aclsSADAConstants	clsSADAConstants	SADA's class of constants
abUpdateMyRemedialUnitsBeingCleaned	Boolean	If true, then update the property *RemedialUnitCleanedInGlobalScenario*
alAdditionalNumberOfRUSToRemediate	Integer	Number of additional RUs to be added
alRemedialUnitsThatAreOrShouldBeRemediat-ed	Integer(Array)	All remedial units within this unit that would be included in global scenario.
UpdateMyRemedialUnitIncludes	Subroutine	Updates the property *RemedialUnitIncluded* given a global set of include ids.
abGlobalInclude	Boolean	Remedial unit include IDs

A.4 Class SADA.clsGEMPhysicalStructure

This class contains the structural specifications for a GEM structure including grid system, layering system, exposure unit specifications, and a map from each node to every EU that is centered upon it.

Properties

Property	Type	Description
ExposureUnitSpecifications	SADA.colExposureUnitSpecifications	Contains the collection of user defined geometry and depth based DCGL values to use.
GridIAmBaseOn	SADA.clsGridDefinition	The GEM grid specifiction.
GridToEUMappingBIG	Integer	A two dimensional array mapping each GEM node to every EU that is centered on it.
LayersIAmBasedOn	SADA.colLayer	GEM layering design.
MyStorageFileName	String	The name to be identified with stored parameters sets.

A.5 Class SADA.clsGEMStructure

This class manages all the parameters needed to implement the GEM framework. This structure call can carry out the calculations for SCSM calculation, MrDM, and MrsDM.

Properties

Property	Type	Description
ExposureUnitArray	clsExposureUnit	Array of instantiated exposure units clsExposureUnit
Name	String	Name assigned to this GEM.
PhysicalStructure	clsExposureUnitCollectionStructure	Holds GEM Structural parameters

Methods

Method	Type	Description
AtLeastOneExposureUnitFailsCompliance	Function	Returns a value of true if at least one of the members of ExposureUnitArray fails compliance.
CalculateMrDM	Subroutine	Executes Algorithms 3.2 and 3.3 and places the MrDMRemedial design in the class clsResultDocumentation.
aclsDataQueryTools	clsDataQueryTools	SADA's data management class.
aclsErrorReport	clsErrorReport	SADA's error management class.
aclsSimulationResult	clsSimulationResult	Simulation results are held by this class.
aclsResultDocumentation	clsResultDocumentation	Manages all SADA modeling results.
aclsRemedialDesign	clsRemedialDesign	Contains a remedial design parameters.
afrmAdvancedProgress	frmAdvancedProgress	Indicates progress in calculating MrDM.
alTopProgressBarToUse	Integer	Bar on that form to use.
alChemicalID	Integer	Unique OID for the current contaminant.
abIncludeBlock()	Boolean	Indicates which RUs are included.
abExportBaseResult	Boolean	Indicates whether to export baseline.
asExportFileName	String	File name to use in exporting baseline .
asMRDMLogFile	String	Log file documenting MrDM calculation.
CalculateMrsDM	Subroutine	Executes Algorithm 3.4 (MrsDM).
afrmAdvancedProgress	frmAdvancedProgress	Indicating progress in calculating MrsDM.
aclsNewSampleDesignParameters	clsSampleDesignParameters	Manages new sample design parameters.
aclsDataQueryTools	clsDataQueryTools	SADA's data management class.
aclsBaseLineSimulationResult	clsSimulationResult	The baseline simulation result.
aclsBaseLineResultDocumentation	clsResultDocumentation	The MrDM result using existing data.
aclsErrorReport	clsErrorReport	SADA's error management class.

Methods (continued)

Method	Type	Description
alChemicalID	Integer	Unique OID for the current contaminant.
abUseOnlySelectedDataForInterpolation	Boolean	True means only data within polygon area are used.
acolVariographySets	colVariographySets	Spatial correlation modeling parameters
acolGeospatialParameters	colGeospatialParameters	Geostatistical modeling parameters.
aStatusBar	StatusBar	Status bar that shows MrsDM progress
aclsColorPreferencesForVariousItems	clsColorPreferencesFor-VariousItems	Contains color preferences.
acolLayerDesigns	colLayerDesigns	Collection of layer designs.
aclsInformationSet	clsInformationSet	Contains information regarding selected contaminant.
aclsBaseLineRemedialDesign	clsRemedialDesign	Baseline remedial design (no new data).
abLogTransformed	Boolean	Indicates whether data are transformed.
asMrsDMLogFile	String	Log file documenting the MrsDM process.
CalculateSCSM	Subroutine	Executes Algorithm 3.1 (SCSM) and places results in the property *ExposureUnitSpecifications* of the class property *PhysicalStructure*.
aclsDataQueryTools	clsDataQueryTools	This is SADA's data management class.
aclsSimulationResult	clsSimulationResult	Simulation results are held by this class.
afrmAdvancedProgress	frmAdvancedProgress	This is a form indicating progress in calculating SCSM
alBarIndexToUse	Integer	This is the particular bar on that form to use.
aclsErrorReport	clsErrorReport	This is SADA's error management class.
abIncludeBlock()	Boolean Array	This indicates which remedial units are included.
abBlockIsRemediated()	Boolean Array	This indicates which remedial units are already remediated.
EstablishMyBasicParameters-FromTable	Subroutine	Populates basic colExposureUnit parameters from aTable.
aTable	DAO.Recordset	SADA file recordset containing GEM parameters.
asMyName	String	Name of this instantiated clsGEMStructure.

Methods (continued)

Method	Type	Description
EstablishMyExposureUnits-FromStoredValues	Subroutine	This routine rebuilds the array of clsExposureUnits from a previously recorded in the SADA file.
adbSADAFile **abMaximumNumberOfEUSPer Node** **abSucceeded**	DAO.Database Integer Boolean	The SADA file The upper bound on the number of EUs that could be assigned to any one RU node. Stating this increases speed of method. True means the instance successfully parameterized itself.
EstablishMySelfCompletely-FromScratch	Subroutine	This routine combines the GEM grid and layer system together with the collection of *colExpsoureUnitSpecifciations* to create the array of instantiated exposure units **ExposureUnitArray.** This new instance is added to the array of previously created clsGEMStructures in the clsPreviouslyCreatedEU class.
afrmAdvancedProgress **alTopBarToUse** **acolLayers** **aclsGrid** **acolVolume_DepthBasedLimits** **abGlobalIncludeBlock()** **adProbabilityLimit** **adRemedialReplacementValue** **adMissingOrOutOfBoundsValue** **aclsErrorReport** **asName** **aclsPreviouslyCreatedEUs**	frmAdvancedProgress Integer colLayer clsGridDefinition colExposureUnitSpecifications Boolean Double Double Double clsErrorReport String clsPreviouslyCreatedEUCollections	SADA's advanced progress form. Progress bar on the form to use here. GEM layering. GEM grid definition. EU Specifications to build the instance with. Array indicating RU inclusion. The probability limit. The post remedial concentration value. The value to use for EUs with offsite domains. SADA's error reporting system. Name of this clsGEMStructure instance. Contains an array of previously created colExposureUnits.

Methods (continued)

Method	Type	Description
New	Subroutine	Creates an instance of this collection.
NumberFailingCompliance	Function	Returns the number of exposure unit instances within *ExposureUnitArray* failing compliance.
UpdateToCurrentEUSpecs		
acolExposureUnitSpecifications	colExposureUnitSpecifications	EU specifications to update this instance with.
adProbabilityLimit	Double	The probability limit.
adRemedialReplacementValue	Double	The post remedial concentration value.
adMissingOrOutOfBoundsValue	Double	The value to use for EUs with offsite domain.
afrmAdvancedProgress	frmAdvancedProgress	SADA's advanced progress form.
alBarToUse	Integer	Progress bar on the form to use here.
abGlobalIncludeBlock	Boolean	Array indicating RU inclusion.

Table A.6 Class SADA.clsPreviouslyConstructedGEMStructures: Properties

This set contains all previously created clsGEMStructures and determines which ones are most appropriate to use when requested. If an appropriate structure is not found, a new is created.

Property	Type	Description
CurrentEUCollection	clsGEMStructure	The current GEMStructure.
PreviouslyEstablishedGEMs	clsGEMStructure (Array)	This is the array of previously constructed clsGEMStructures.

Table A.6 Class SADA.clsPreviouslyConstructedGEMStructures: Methods

Method	Type	Description
SetCurrentGEMStructure	Subroutine	Determines if a clsGEMStructure is equivalent. If so, the clsGEMStructure can be used simply by updating it with the remaining parameters. If not a new one is created.
aclsSimulationResult	clsSimulationResult	Contains set of simulations.
acolExposureUnitSpecifications	acolExposureUnitSpecifications	EU specifications class.
adProbabilityLimit	Double	The probability limit.
adRemedialReplacementValue	Double	The post-remedial concentration value.
adMissingOrOutOfBoundsValue	Double	Concentration to use for off site areas.
afrmAdvancedProgress	frmAdvancedProgress	SADA's advanced progress form.
alTopBarToUse	Integer	Progress bar on the form to use here.
aclsErrorReport	clsErrorReport	SADA's error reporting system.
abGlobalIncludeBlock	Boolean	Array of included IDs for GEM remedial units.
adbSADAFile	DAO.Database	The SADA file.

Vita

Robert Stewart currently works in the Geographic Information Science and Technology Group at the Oak Ridge National Laboratory in Oak Ridge, TN (http://www.ornl.gov/sci/gist/index.shtml). There he engages in research and development activities broadly characterized as "geospatial risk assessment." These include areas of chemical and radiological hazards, spatial epidemiology, and population dynamics. Prior to ORNL he worked for nearly sixteen years at the University of Tennessee where he served as principal investigator, technical lead, and in most cases point of contact with sponsoring agencies such as the Environmental Protection Agency, the Nuclear Regulatory Commission, and the Department of Energy. Primary research and development activities at UT included environmental sample design, geospatial modeling, uncertainty analysis, and risk assessment. The emphasis is typically on integration of these methods within various decision analysis frameworks using GIS technologies. A central focus of his work has been the Spatial Analysis and Decision Assistance (SADA) project (http://www.tiem.utk.edu/~sada/index.shtml).

Robert earned a Bachelor of Science degree from the University of Tennessee in mathematics and statistics in 1992 and a Master's degree in Mathematics in 1995. Robert has a number of publications in his field, serves on two federal interagency working groups in environmental modeling, and has taught numerous short courses in the US and abroad, attended by academics, consultants, and federal employees.