

T.C
YEDİTEPE UNIVERSITY
GRADUATE INSTITUTE OF SCIENCE AND ENGINEERING

85286

A FUZZY APPROACH TO LAND CONTAMINATION DETECTION PROBLEM

By

Mustafa Alper ÖZPINAR

Submitted to the Institute for Graduate Institute of Science and Engineering
in partial fulfilment of the requirements for the degree
of Master of Science
in Systems Engineering

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

İSTANBUL, 1999

85286

A FUZZY APPROACH TO LAND CONTAMINATION DETECTION PROBLEM

Approved by:

Asst. Prof. Dr. Melek DEMİRHAN
(Supervisor)



Prof. Dr. Linet ÖZDAMAR



Asst. Prof. Dr. Ender ÖZCAN



Date of Approval : 23 – 09 -1999

TABLE OF CONTENTS

	page
LIST OF SYMBOLS	V
LIST OF ABBREVIATIONS.....	VII
LIST OF FIGURES	IX
LIST OF TABLES	XI
ACKNOWLEDGMENTS	XIII
ABSTRACT.....	XIV
ÖZET	XV
 1. INTRODUCTION	 1
 2 AREAL CONTAMINATION AND REMEDIATION.....	 4
2.1 BROWNFIELDS	4
2.2 SITE ASSESSMENT	5
2.3 SITE INVESTIGATION (IDENTIFY THE SOURCE, NATURE, AND EXTENT OF CONTAMINATION):	6
2.4 BROWNFIELDS REMEDIATION.....	7
2.5 BROWNFIELDS REMEDIATION METHODS :	8
2.5.1 Bioventing	8
2.5.2 Enhanced bioremediation	8
2.5.3 Land treatment	8
2.5.4 Natural attenuation in soils.....	9
2.5.5 Phytoremediation	9
2.5.6 Electrokinetic separation	9
2.5.7 Fracturing.....	10
2.5.8 Soil flushing.....	11
2.5.9 Soil vapor extraction	11
2.5.10 In situ solidification/ stabilization.....	12
2.5.11 Thermally enhanced soil vapor extraction	12
2.5.12 Biopiles.....	12
2.5.13 Composting.....	13
2.5.14 Fungal biodegradation.....	13
2.5.15 Landfarming.....	14
2.5.16 Slurry Phase Biological Treatment	14
2.5.17 Chemical extraction	14

6	PERFORMANCE EVALUATION OF SPATIAL INTERPOLATION METHODS IN THE PRESENCE OF NOISE	91
6.1	THE TEST FUNCTIONS.....	91
6.2	THE SAMPLING PATTERNS.....	92
6.3	AN EXAMPLE DEMONSTRATING THE DISTORTION CAUSED BY ADDITIVE WHITE NOISE.....	94
6.4	NUMERICAL EXPERIMENTS WITH SIM.....	94
	CONCLUSION	100
	REFERENCES.....	103
	APPENDIX A : TEST FUNCTIONS	107
	APPENDIX B: INTERPOLATION RESULTS FOR CHAPTER 6	111
	APPENDIX C: SOURCE CODES FOR PERFORMANCE EVALUATIONS.....	116
	C.1 KEY MAP GENERATION AND EVALUATION.....	117
	C.2 HERRINGBONE SAMPLE PATTERN CODE	119
	C.3 REGULAR GRID SAMPLE PATTERN CODE.....	121
	C.4 LINEAR SAMPLE PATTERN CODE.....	122
	C.5 CIRCULAR SAMPLE PATTERN CODE	123
	C.6 GAUSSIAN ERROR GENERATOR CODE.....	125

LIST OF SYMBOLS

Δ	Triangulation of the convex hull of V
$\rho(h)$	Correlation function
$\gamma(h)$	Variogram function
σ_{+h}	The standard deviation of all the data values whose locations are +h away from some other data location
σ_{-h}	The standard deviation of all the data values whose locations are -h away from some other data location
$\alpha_i(t)$	Sub region
$\lambda_j^{OK}(x)$	OK weights
$\alpha_S(t)$	Non-overlapping smaller sub-regions
σ_x	Standard deviation for $x_1, \dots, x_n$
σ_y	Standard deviation for $y_1, \dots, y_n$
$\{v_i, v_j\}$	Vertices
μ	Mean
$\mu(x)$	The mean of $Z(x)$
$C(h)$	Covariance function
$C(x_i, x_j)$	Covariance matrix
C_o	Nugget effect
d_i	Denotes the Euclidean distance between x and x_i
d_q	Radius of influence about the node x_i
d_w	Radius of influence around the point x
g	Nugget variable
gc	The number equivalent square grids
m_{+h}	The mean of all the data values whose locations are +h away from some other data
m_{-h}	The mean of all the data values whose locations are -h away from some other data
m_x	Mean for $x_1, \dots, x_n$

m_y	Mean for $y_1, \dots, y_n$
n	Number of data
n'_i	Number of new samples
$n_i(t)$	The number of samples falling within the area
p^*	User-specified constant between 0 and 1
q_i	Bivariate quadratic functions
r	Range variable
$r_i(t)$	The fuzzy potential of a sub-region
$r_{\max}(t)$	The maximum $r_i(t)$ over all existing sub-regions
s	Sill variable
S_i	A Voronoi tessellation is a collection of "tiles"
u	The interpolating function.
w	Power variable
w	The weight assigned to $Z(x_i)$
$x_1, \dots, x_n$	The data values for the first variable
$y_1, \dots, y_n$	The data values for the first variable
Z_β	The average value of $Z(x)$ over a particular block B
$Z(x)$	Regionalized random variable
$Z'(x_0)$	Linear predictor for $Z(x_0)$
$Z_{SK}(x)$	The linear regression estimator or "simple kriging" (SK) estimator

LIST OF ABBREVIATIONS

Avg	Average
B.L.U.E.	Best linear unbiased predictor
BTEX	Benzene, toluene, ethylbenzene, xylene
COVER	Percentage of contaminated area recognized correctly
DEM	Digital elevation modelling
DoE	Department of the Environment
EPA	Environmental Protection Agency
ER	Electrokinetic remediation
E-W	East-west
FASA	Fuzzy areal site assessment
GIS	Geographical information systems
GMF	Gaussian membership function
HerrB	Herringbone
HTRW	Hazardous, toxic, and radioactive waste
MNC	Minimum curvature method
moi	Moment of inertia
NE-SW	North-east South-west
NHS	The number of hot-spots per Site
N-S	North-south
NW-SE	North-west South-east
OK	Ordinary kriging
PAH	Polynuclear aromatic hydrocarbons
PCA	The percentage of contaminated Area
PCB	Polychlorinated biphenyl
PCP	Pentachlorophenol
PF	Pneumatic fracturing
RBF	Radial basis functions
RF	Random function
S/S	Solidification/stabilization

SAMP	The sampling pattern
SCRAP	Clean area mis-identified as contaminated (false positives)
SIM	Spatial interpolation methods
SK	Simple kriging
Std. Dev.	Standard Deviation
SVE	Soil vapor extraction
TIO	Technology Innovation Office
VCP	Voluntary cleanup program
VOC	Volatile organic compounds



LIST OF FIGURES

Figure 3.1 Examples of how data can be paired to obtain an h-scatterplot	30
Figure 3.2 h-scatterplots for four separation distances.	31
Figure 3.3 Theoretical Variograms	38
Figure 3.4 Linear $\phi(r)=r$ RBF Function	46
Figure 3.5 Thin-plate spline $\phi(r)=r^2 \log r$ RBF Function.....	46
Figure 3.6 Gaussian $\phi(r)=\exp(-ar^2)$ RBF Function.....	46
Figure 3.7 Multiquadratic $\phi(r)=(r^2+c^2)^{0.5}$ RBF Function	46
Figure 3.8 An example of a triangulation of points in R^2	54
Figure 3.9 Two possible triangulations.....	55
Figure 3.10: The two possible triangulations.....	56
Figure 3.11 Two min-max angle locally optimal triangulators	58
Figure 3.12 Delaunay triangulation and Voronoi tessellation of a point set V.....	59
Figure 3.13 A Delaunay triangulation vs. a non-Delaunay triangulation	60
Figure 3.14 A typical representation of triangulation algorithm	60
Figure 4.1. Pseudocode of FASA.....	64
Figure 4.2 Representation GMF for a given set of data.....	65
Figure 4.3 An exemplary contaminated site	67
Figure 5.1 Three sampling patterns.....	70
Figure 5.2 A hypothetical site with Herringbone sampling pattern.....	77
Figure 5.3 FASA.....	78
Figure 5.4 Triangulation	78
Figure 5.5 Spherical Kriging.....	79
Figure 5.6 Linear Kriging	79
Figure 5.7 Quadratic	80
Figure 5.8 Gaussian Kriging.....	80
Figure 5.9 Minimum Curvature	81
Figure 5.10 RBF- Thin Plate	81
Figure 5.11 RBF- Multiquadratic.....	82
Figure 5.12 Shepards Method	82
Figure 6.1 Sampling patterns used for noisy situation.....	93

Figure A.1 Function #1	107
Figure A.2 Function #2	107
Figure A.3 Function #3	108
Figure A.4 Function #4	108
Figure A.5 Function #5	109
Figure A.6 Function #6	109
Figure A.7 Function #7	110
Figure A.8 Function #8	110
Figure B.1 Original Function	111
Figure B.2 (Kriging) Herring Bone sampling pattern /Noise-Free	111
Figure B.3 (Kriging) Grid sampling pattern /Noise-Free	112
Figure B.4 (Kriging) Linear sampling pattern /Noise-Free	112
Figure B.5 (Kriging) Circular sampling pattern /Noise-Free.....	113
Figure B.6 (Kriging) Herring Bone sampling pattern /With noise	113
Figure B.7 (Kriging) Grid sampling pattern /With noise.....	114
Figure B.8 (Kriging) Linear sampling pattern /With noise.....	114
Figure B.9 (Kriging) Circular sampling pattern /With noise	115

LIST OF TABLES

Table 2.1 Various contaminants found at brownfields	20
Table 3.1. Conditions imposed on data for various radial basis interpolants.	47
Table 4.1. The iterations of FASA.....	68
Table 5.1 Performance of sampling patterns on 1000 hypothetical sites for PCA=1%.....	72
Table 5.2 Performance of sampling patterns on 1000 hypothetical sites for PCA=5%.....	72
Table 5.3 Results of various SIM and FASA on a synthetic site.....	76
Table 5.4 Linear kriging results.	85
Table 5.5 RBF- Multiquadratic results.	85
Table 5.6 Minimum curvature results.	85
Table 5.7 Shepard's results.....	86
Table 5.8 Triangulation results.	86
Table 5.9 FASA results.....	86
Table 5.10 Linear kriging results.	87
Table 5.11 RBF- Multiquadratic results	87
Table 5.12 Minimum curvature results	87
Table 5.13 Shepard's results.....	88
Table 5.14 Triangulation results.	88
Table 5.15 FASA results.....	88
Table 5.16 Overall results of methods for NHS=1	89
Table 5.17 Overall results of methods for NHS=3	89
Table 5.18 Overall results of sampling patterns for NHS=1.....	89
Table 5.18 Overall results of sampling patterns for NHS=1.....	90
Table 5.19 Overall results of sampling patterns for NHS=3.....	90
Table 5.20 Paired t-test statistics for 360 sites (*: significant at 5% level)	90
Table 6.1. Performance of four sampling patterns using Kriging interpolation method.	94
Table 6.2 Performance evaluations for Kriging without Noise	96
Table 6.3 Performance evaluations for Kriging with Noise	96
Table 6.4 Performance evaluations for Minimum Curvature without Noise.....	96
Table 6.5 Performance evaluations for Minimum Curvature with Noise.....	97
Table 6.6 Performance evaluations for Radial Basis without Noise.....	97
Table 6.7 Performance evaluations for Radial Basis with Noise.....	97

Table 6.8 Performance evaluations for Shepard's without Noise.....	97
Table 6.8 Performance evaluations for Shepard's without Noise.....	98
Table 6.9 Performance evaluations for Shepard's with Noise.....	98
Table 6.10 Performance evaluations for sampling methods without Noise	98
Table 6.11 Performance evaluations for sampling methods with Noise.....	99
Table D.1 Performance evaluation items for computer program.....	116



ACKNOWLEDGMENTS

I would like to thank to my thesis advisor Asst. Prof. Melek Demirhan and her colleague Prof. Dr. Linet Özdamar for their great aid, understanding and their advices throughout my study.

I would like to express my best regards to my dean Prof. Dr. Eralp Özil for his invaluable comments and motivation for my study and life.

Next I am grateful to my colleagues and friendly staff in Yeditepe University namely Iffet Doğan, Zeynep Temel, Ilker Birbil, Rana Belen, Zeynep Yazıcıoğlu, Ayşegül Ergin, Erhan Küçükgülzel and Ünal Ergenç for their supports.

Finally I wish to acknowledge the constant encouragement and support that I received from my family.

ABSTRACT

One of the most major concern, in the last decades is to minimize the deteriorating effect of the ecological system resulting from uncontrolled industrialization process and urbanization. This issue is essential in preserving and improving the quality of life for the next generations.

Scientists have developed mathematical and statistical tools in order to cope with the detection of pollution and remediation of polluted areas.

This thesis involves the performance evaluation of the conventional techniques, which are utilized in locating the contaminated zones by testing them on sites with different characteristics. These techniques are as follows;

- Kriging
- Minimum Curvature
- Radial Basis Function
- Shepard's Method
- Triangulation

A further contribution of this work is development of a novel site assessment technique, which is based on fuzzy evaluation zones. Numerical experiments are conducted to demonstrate and compare efficiency of these techniques.

In the last phase of the experiments Kriging, Minimum Curvature, Radial Basis and Shepard's method are tested according to their robustness to the experimental and evaluation errors.

ÖZET

Son on yıl içerisinde önemi gün geçtikçe artan konulardan bir tanesi de plansız ve kontrolsüz şehirleşme ve sanayileşmenin neticesinde bozulmakta olan ekolojik sistemin zararlarını en aza indirmek olmuştur. Bu konu; gelecek nesillerin yaşam standartlarının korunması ve geliştirilmesinin ayrılmaz bir parçasıdır.

Bilim adamları kirlenmiş bölgelerin tesbiti ve tekrar kullanıma açılması için çeşitli matematiksel ve istatistiksel araçlar ve methodlar geliştirmişlerdir.

Bu tez çalışması kirlenmiş bölgelerin tesbitinde kullanılan geleneksel methodların, değişik özelliklerdeki test bölgeleri üzerindeki performans analizlerini içermektedir. Tez çalışmasında karşılaştırılan methodlar aşağıdaki gibidir;

- Kriging
- Minimum Curvature
- Radial Basis Fonksiyonu
- Shepard's Metodu
- Triangulation

Bu çalışmada ayrıca, bulanık mantık kullanan yeni bir bölgesel değerlendirme metodu anlatılmış ve diğer methodlarla birlikte test edilmiştir.

Son etapta Kriging, Minimum Curvature, Radial Basis and Shepard's metodlarının hatalara karşı duyarlılıkları test edilmiş ve sonuçlar değerlendirilmiştir.

1. INTRODUCTION

After the years of rapid and unplanned urbanization and industrialization within the cities nowadays the regional economy has been making the transition from heavy manufacturing to more service-oriented sectors as a result of changing markets, international competition and advances in production technologies. Many communities have experienced plant downsizing and shutdowns as a result; leaving behind thousands of underused or abandoned, and often contaminated, industrial sites, commonly known as "brownfields."

Land contamination is more likely to take place in metal mines, iron and steel works, foundries, electroplating, anodising, and galvanizing works, engineering works such as ship building and breaking sites, scrap yards; gas works, power stations, chemical works, refineries, tar distilleries, waste disposal sites (contamination by asbestos). Obviously, areas, which host diverse industries, are susceptible for land contamination. Hazardous land contaminants such as arsenic, cadmium, chromium, lead, mercury and selenium (which may also appear in sewage sludge) jeopardize agricultural land, domestic gardens, allotments, parks, playing fields and open space. Sulphates, chloride, tarry substances, phenols and mineral oils contaminate housing developments, commercial and industrial buildings (Alloway, 1995).

The decades of hazardous wastes of 1980's are considered to start with the incidents in Lekkerkirk near Rotterdam, Netherlands and in Love Canal, USA in 1979 and 1980, respectively (Alloway, 1995). The environmental problems in both of these incidents were principally originated from the dirty legacy of past industries once situated on these lands. The huge volume of efforts needed and very high costs incurred for the remediation of these sites drew the attention to adverse environmental impacts of past anthropological activities on the environment. Since 1986 (the year that the International Standardization Organization established a technical committee to consider the standardization of soil quality) the issue of contaminated land has been drawing more and more attention from all concerned parties: from land owners to environmental professionals, from authorities to lawmakers. As of today, contaminated land problems and related investigation and

remediation projects constitute a large portion of consulting services provided by environmental professionals in Europe and in the U.S.

The conventional tools to detect the location of contaminated zones are usually selected from geostatistical techniques (Zirchky and Harris, 1986) originally devised for detecting natural resources (Isaaks and Srivastava, 1989) or geographical topology (Oliver and Webster, 1990, Rosenbaum and Soderstrom, 1996). In general terms, geostatistical techniques deduce attribute levels of unsampled locations from sample information according to the relative spatial topology of samples and the interpolated locations. This deduction is carried out by utilizing statistical relationships based on distance. A common nomenclature for these approaches is “special interpolation methods (SIM)”.

However, in the context of land contamination, the assumptions which SIM are based on, are difficult to validate in practice. Hence, there is a need for establishing the performance of SIM in this field. The aim of this thesis is to conduct a numerical survey to evaluate and compare the functionality of SIM in brownfields investigations. A new approach, which is independent of the statistical assumptions (Ripley, 1981, Cressie, 1993), underlying some SIM, is also developed here. This approach partitions the site under investigation systematically into smaller sections leading to a dynamic group assessment of sample information. This approach is denoted as Fuzzy Areal Site Assessment (FASA) technique.

As suggested by its name this approach involves a fuzzy measure in evaluating the potential of a given zone to contain contaminated spots.

In order to evaluate the performance of SIM and FASA, a test bed of hypothetical contaminated sites are generated randomly according to predefined characteristics. Performance is measured by two quantifiable criteria. These criteria are developed here and indicate the precision of areal identification in terms of true and false positives namely, correctly identified contaminated areas and misidentified areas. Furthermore, since all data collection and analysis processes inevitably contain errors, the performance of SIM is also assessed in the presence of noise.

However the numerical experiments concerning noise are conducted by using mathematical functions rather than randomly generated hypothetical sites. This is because the former site evaluation method does not provide the precise comparison of the contaminated zone. However in the case of noisy data the affect of noise on the interpolated values at every location is of primary importance. Mathematical functions naturally provide such a continuous test bed.

Finally, apart from investigating several site characterization techniques, some sampling patterns acquired from the literature as well as the ones developed here are assessed during numerical experiments.

The site characterization phase is the most crucial procedure prior to the remediation of an area due to very high remediation costs involving the removal of contaminated land and its replacement with clean re-fills. The experiments conducted in this thesis thus present an important contribution to this area of research.

2 AREAL CONTAMINATION AND REMEDIATION

2.1 Brownfields

After the years of rapid, unplanned urbanization and industrialization, the regional economy of nowadays has been making the transition from heavy manufacturing to more service-oriented sectors due to changing markets, international competition and advances in production technologies. The consequences of this urbanization trend include traffic congestion; air, water and noise pollution; contaminated soils; higher housing costs; disinvestment in older communities and related social disruption; loss of wildlife habitat, agricultural lands and other open spaces; and heightened infrastructure investment requirements. Therefore many communities have experienced plant downsizing and shutdowns leaving behind thousands of underused or abandoned, and often contaminated, industrial sites or commonly known as "brownfields."

The U.S. Environmental Protection Agency (EPA) has defined brownfields as abandoned, idled, or under-used industrial and commercial facilities where expansion or redevelopment is complicated by real or perceived environmental contamination. Preparing brownfields for productive reuse requires the integration of many elements; financial issues, community involvement, liability considerations, environmental assessment and cleanup. The assessment and cleanup of a site must be carried out in a way that integrates all those factors into the overall redevelopment process. In addition, the cleanup strategy will vary from site to site. At some sites, cleanup will be completed before the property is transferred to new owners. At other sites, cleanup may take place simultaneously with construction and redevelopment activities. Regardless of when and how cleanup is accomplished, the challenge to any Brownfields program is to cleanup sites quickly and redevelop the land in ways that benefit communities and local economies.

In developed countries, innovative technologies are being used in many cleanup programs to assess contamination and treat a variety of hazardous substances and petroleum products that have been released into the environment.

Land contamination is more likely to take place in metal mines, iron and steel works, foundries, electroplating, anodising, and galvanizing works, engineering works such as ship building and breaking sites, scrap yards; gas works, power stations, chemical works, refineries, tar distilleries, waste disposal sites (contamination by asbestos). Obviously, areas which host diverse industries are susceptible for land contamination. Hazardous land contaminants such as arsenic, cadmium, chromium, lead, mercury and selenium (which may also appear in sewage sludge) jeopardize agricultural land, domestic gardens, allotments, parks, playing fields and open space. Sulphates, chloride, tarry substances, phenols and mineral oils contaminate housing developments, commercial and industrial buildings. (Alloway, 1995)

In brownfields, particularly ones that are located in central urban areas, redevelopment is sometimes not possible for a variety of reasons, including high cleanup costs, unclear cleanup standards, which encourage development to migrate to outlying areas or undeveloped "greenfields."

2.2 Site Assessment

The purpose of this step is to determine the likelihood of contamination at a particular site by collecting and reviewing applicable information about a site. This "environmental audit" is an initial investigation that usually is limited to a search of historical records. The data to be collected also include information about past and current environmental conditions and historical uses of the site. During the site assessment phase, it is important to consider the activities and requirements described in the subsequent chapters and determine how they can be combined with or initiated during the site assessment. The collection of data during this initial step of the cleanup process is extremely important for use in identifying and evaluating the applicability of site assessment and cleanup technologies, as well as in determining whether the property can be cleaned to the level necessary for its intended reuse. It also is essential to assess and address the needs and concerns of the community (for example, the development of social and economic profiles and the identification of acceptable environmental risk).

To ensure that sufficient data are collected, the potential applicability of innovative technologies to the site also should be considered. Since much of the work at this stage involves a search of hard and soft copies of records, applicable technology options may be somewhat limited.

Factors that should be considered during this phase include:

1. What is known about the site? What records exist that indicate potential contamination and past use of the property? Have other environmental actions occurred (such as notices of violation)? Has an environmental audit been conducted? What level of site assessment is needed to identify the types and extent or the absence of contamination?
2. Is the site located in an area targeted for redevelopment?
3. Are there governmental or private requirements for site assessment? Is there a voluntary cleanup program (VCP)? (U.S. Environmental Protection Agency, Road Map to Understanding Innovative Technology Options for brownfields Investigation and Cleanup, 1997)

2.3 Site Investigation (Identify the Source, Nature, and Extent of Contamination):

This phase focuses on identifying, locating, and characterizing the nature and extent of contamination at a site. It is essential that an appropriately detailed study of the site be performed to identify the cause, nature, and extent of contamination and the possible threats to the environment or to any people living or working nearby. For brownfields, the results of such a study can be used in determining goals for cleanup, quantifying risks, determining acceptable and unacceptable risks, and developing cleanup plans that do not cause unnecessary delays in the remediation and reuse of property.

A site investigation is based on the results of the site assessment, which is discussed in the preceding section. The site investigation phase may include the analysis of samples of soil and soil gas, groundwater, surface water, and sediment. According to the risk factors identified in the historical survey, the number of samples collected for various contaminants are determined considering the possible interactive effects of contaminants. Next, the pattern of the sampling strategy is decided on. The criterion for the selection of the sampling pattern is particularly obtaining a uniform cover of the site with minimum

risk of missing contaminated zones. Once the samples are collected, laboratory analysis follows. Considering possible sampling and analysis errors, the distribution of contaminants is attempted to be identified so that the site can be remediated. The main concern of this thesis is to determine the effects of different spatial interpolation methods on the identification of the contaminant dispersion map considering different sampling strategies.

The migration pathways of contaminants also are examined during the later phase, and a baseline risk assessment may be needed to calculate risk to human health and the environment. (U.S. Environmental Protection Agency, Road Map to Understanding Innovative Technology Options for brownfields Investigation and Cleanup, 1997)

2.4 Brownfields Remediation

New life is what land recycling offers these debilitated and seemingly undesirable places. With infrastructure already in place, or access to a nearby workforce, brownfields provide an opportunity for more efficient and more sustainable development.

However because of the position and location of these brownfields they cannot be easily extracted from the future planning of the city. In the near future these areas should be considered as the potential areas for the new city regions or in other words “the city of tomorrow”. (U.S. Environmental Protection Agency, Road Map to Understanding Innovative Technology Options for brownfields Investigation and Cleanup, 1997)

The purpose of screening various technologies is to evaluate those technologies for their capability to meet specific cleanup and redevelopment objectives. For brownfields, it is also important to consider budget requirements and to maintain a work schedule so that the project remains profitable.

2.5 Brownfields Remediation Methods :

2.5.1 Bioventing

Bioventing is a promising new technology that stimulates the natural *in situ* biodegradation of any aerobically degradable compounds in soil by providing oxygen to existing soil microorganisms. The technology is as follows oxygen is delivered to contaminated unsaturated soils by forced air movement (either extraction or injection of air) to increase oxygen concentrations and stimulate biodegradation. In contrast to soil vapor vacuum extraction, bioventing uses low airflow rates to provide only enough oxygen to sustain microbial activity. Oxygen is most commonly supplied through direct air injection into residual contamination in soil. In addition to degradation of adsorbed fuel residuals, volatile compounds are biodegraded as vapor move slowly through biologically active soil.

2.5.2 Enhanced bioremediation

In this remediation technique the activity of naturally occurring microbes is stimulated by circulating water-based solutions through contaminated soils to enhance *in situ* biological degradation of organic contaminants or immobilization of inorganic contaminants. Nutrients, oxygen, or other amendments may be used to enhance bioremediation and contaminants desorption from subsurface materials. There are two types of bioremediation according to the oxygen dependency of the method.

- Aerobic: In the presence of sufficient oxygen (aerobic conditions), and other nutrient elements, microorganisms will ultimately convert many organic contaminants to carbon dioxide, water, and microbial cell mass.
- Anaerobic: In the absence of oxygen (anaerobic conditions), the organic contaminants will be ultimately metabolised to methane, limited amounts of carbon dioxide, and trace amounts of hydrogen gas.

2.5.3 Land treatment

In land treatment contaminated surface soil is treated in place by tilling to achieve aeration, and if necessary, by addition of amendments. Periodically tilling, to aerate the waste, enhances the biological activity. Land Treatment is a full-scale bioremediation technology

in which contaminated soils, sediments, or sludges are turned over (i.e., tilled) and allowed to interact with the soil and climate at the site. The waste, soil, climate, and biological activity interact dynamically as a system to degrade, transform, and immobilise waste constituents. Wastes are periodically tilled to aerate the waste.

2.5.4 Natural attenuation in soils

Natural processes - such as dilution, dispersion, volatilization, biodegradation, adsorption, and chemical reactions with soil materials - are allowed to reduce contaminant concentrations to acceptable levels. Surface and subsurface soils have different characteristics in natural attenuation. Mobile contaminants in subsurface soils diffuse into soil vapor and aqueous phase and thus are relatively easily subject to natural subsurface processes that can attenuate these contaminants. Most high molecular weight (persistent) organic and many inorganic contaminants will be immobilised in the subsurface soil matrix. These persistent organic contaminants often are difficult to degrade and the inorganic metals are conserved. Without probable exposure routes, however, they do not represent significant risk unless unlikely events such as fresh solvent releases, chemical or biochemical transformation or physical disturbances that increase their mobility or open exposure routes. Thus natural attenuation in subsurface soils may be considered on a site-specific-basis as a possible remedy.

2.5.5 Phytoremediation

Phytoremediation is a process that uses plants to remove, transfer, stabilise, and destroy contaminants in soil and sediment. Contaminants may be either organic or inorganic. The mechanisms of phytoremediation include enhanced rhizosphere biodegradation, phyto-extraction (also called phyto-accumulation), phyto-degradation, and phyto-stabilization.

2.5.6 Electrokinetic separation

The Electrokinetic Remediation (ER) process removes metals and organic contaminants from low permeability soil, mud, sludge, and marine dredging. ER uses electrochemical and electrokinetic processes to desorb, and then remove, metals and polar organics. This in

situ soil processing technology is primarily a separation and removal technique for extracting contaminants from soils.

The principle of electrokinetic remediation relies upon application of a low-intensity direct current through the soil between ceramic electrodes that are divided into a cathode array and an anode array. This mobilizes charged species, causing ions and water to move toward the electrodes. Metal ions, ammonium ions, and positively charged organic compounds move toward the cathode. Anions such as chloride, cyanide, fluoride, nitrate, and negatively charged organic compounds move toward the anode. The current creates an acid front at the anode and a base front at the cathode. This generation of acidic condition in situ may help to mobilize sorbed metal contaminants for transport to the collection system at the cathode.

2.5.7 Fracturing

Cracks are developed by fracturing beneath the surface in low permeability and over-consolidated sediments to open new passageways that increase the effectiveness of many in situ processes and enhance extraction efficiencies.

Fracturing is an enhancement technology designed to increase the efficiency of other in situ technologies in difficult soil conditions. The fracturing extends and enlarges existing fissures and introduces new fractures, primarily in the horizontal direction. When fracturing has been completed, the formation is then subjected to vapor extraction, either by applying a vacuum to all wells or by extracting from selected wells, while other wells are capped or used for passive air inlet or forced air injection. Technologies commonly used in soil fracturing include pneumatic fracturing (PF), blast-enhanced fracturing and Lasagna process.

Blast-enhanced Fracturing: Blast-enhanced fracturing is a process used at sites with fractured bedrock formations. The increased well yields, hydraulic conductivity values, and capture zones occur as a result of the highly fractured area created by detonation of explosives in boreholes.

Lasagna Process: Lasagna is an integrated, in situ remedial technology, which combines electroosmosis with treatment zones that are installed directly in the contaminated soil. In Lasagna process, hydraulic fracturing is used to create sorption/degradation zones horizontally in the subsurface soil.

Pneumatic Fracturing (PF): In the PF process, fracture wells are drilled in the contaminated vadose zone and left open (uncased) for most of their depth. A packer system is used to isolate small (0.6-meter or 2-foot) intervals so that short bursts (~20 seconds) of compressed air (less than 10,300 mmHg or 200 pounds per square inch) can be injected into the interval to fracture the formation. The process is repeated for each interval within the contaminated depth.

2.5.8 Soil flushing

Water or water containing an additive to enhance contaminant solubility, is applied to the soil or injected into the ground water to raise the water table into the contaminated soil zone. Contaminants are leached into the ground water, which is then extracted and treated. In situ soil flushing is the extraction of contaminants from the soil with water or other suitable aqueous solutions. Soil flushing is accomplished by passing the extraction fluid through in-place soils using an injection or infiltration process. Extraction fluids must be recovered from the underlying aquifer and, when possible, they are recycled.

2.5.9 Soil vapor extraction

Vacuum is applied through extraction wells to create a pressure/concentration gradient that induces gas-phase volatiles to be removed from soil through extraction wells. This technology also is known as in situ soil venting, in situ volatilization, enhanced volatilization, or soil vacuum extraction.

Soil vapor extraction (SVE) is an in situ unsaturated (vadose) zone soil remediation technology in which a vacuum is applied to the soil to induce the controlled flow of air and remove volatile and some semivolatile contaminants from the soil. The gas leaving the soil may be treated to recover or destroy the contaminants, depending on local and state air discharge regulations. Vertical extraction vents are typically used at depths of 1.5 meters (5

feet) or greater and have been successfully applied as deep as 91 meters (300 feet). Horizontal extraction vents (installed in trenches or horizontal borings) can be used as warranted by contaminant zone geometry, drill rig access, or other site-specific factors.

2.5.10 In situ solidification/ stabilization

Contaminants are physically bound or enclosed within a stabilized mass (solidification), or chemical reactions are induced between the stabilizing agent and contaminants to reduce their mobility (stabilization). Solidification/stabilization (S/S) reduces the mobility of hazardous substances and contaminants in the environment through both physical and chemical means. Unlike other remedial technologies, S/S seeks to trap or immobilize contaminants within their "host" medium (i.e., the soil, sand, and/or building materials that contain them), instead of removing them through chemical or physical treatment. Leachability testing is typically performed to measure the immobilization of contaminants. S/S techniques can be used alone or combined with other treatment and disposal methods to yield a product or material suitable for land disposal or, in other cases, that can be applied to beneficial use. These techniques have been used as both final and interim remedial measures.

2.5.11 Thermally enhanced soil vapor extraction

Steam/hot air injection or electrical resistance/electromagnetic/fiber optic/radio frequency heating is used to increase the volatilization rate of semi-volatiles and facilitate extraction. Thermally enhanced SVE is a full-scale technology that uses electrical resistance/electromagnetic/fiber optic/radio frequency heating or hot-air/steam injection to increase the volatilization rate of semi-volatiles and facilitate extraction.

2.5.12 Biopiles

Excavated soils are mixed with soil amendments and placed in aboveground enclosures. It is an aerated static pile composting process in which compost is formed into piles and aerated with blowers or vacuum pumps.

Biopile treatment is a full-scale technology in which excavated soils are mixed with soil amendments and placed on a treatment area that includes leachate collection systems and some form of aeration. It is used to reduce concentrations of petroleum constituents in excavated soils through the use of biodegradation. Moisture, heat, nutrients, oxygen, and pH can be controlled to enhance biodegradation.

2.5.13 Composting

Contaminated soil is excavated and mixed with bulking agents and organic amendments such as wood chips, hay, manure, and vegetative (e.g., potato) wastes. Proper amendment selection ensure adequate porosity and provides a balance of carbon and nitrogen to promote thermophilic, microbial activity.

Composting is a controlled biological process by which organic contaminants are converted by microorganisms (under aerobic and anaerobic conditions) to innocuous, stabilized by products. Typically, thermophilic conditions (54 to 65°C) must be maintained to properly compost soil contaminated with hazardous organic contaminants. The increased temperatures result from heat produced by microorganisms during the degradation of the organic material in the waste. In most cases, this is achieved by the use of indigenous microorganisms. Soils are excavated and mixed with bulking agents and organic amendments, such as wood chips, animal, and vegetative wastes, to enhance the porosity of the mixture to be decomposed. Maximum degradation efficiency is achieved through maintaining oxygenation (e.g., daily windrow turning), irrigation as necessary, and closely monitoring moisture content, and temperature.

2.5.14 Fungal biodegradation

Fungal biodegradation refers to the degradation of a wide variety of organopollutants by using their lignin-degrading or wood-rotting enzyme system. White rot fungus has been tested under two different treatment configurations: in situ and bioreactor.

2.5.15 Landfarming

Contaminated soil, sediment, or sludge is excavated, applied into lined beds, and periodically turned over or tilled to aerate the waste. Landfarming is a full-scale bioremediation technology, which usually incorporates liners and other methods to control leaching of contaminants, which requires excavation and placement of contaminated soils, sediments, or sludges. Contaminated media is applied into lined beds and periodically turned over or tilled to aerate the waste.

2.5.16 Slurry Phase Biological Treatment

An aqueous slurry is created by combining soil, sediment, or sludge with water and other additives. The slurry is mixed to keep solids suspended and microorganisms in contact with the soil contaminants. Upon completion of the process, the slurry is dewatered and the treated soil is disposed of.

Slurry phase biological treatment involves the controlled treatment of excavated soil in a bioreactor. The excavated soil is first processed to physically separate stones and rubble. The soil is then mixed with water to a predetermined concentration dependent upon the concentration of the contaminants, the rate of biodegradation, and the physical nature of the soils. Some processes pre-wash the soil to concentrate the contaminants. Clean sand may then be discharged, leaving only contaminated fines and washwater to biotreat. Typically, a slurry contains from 10 to 30% solids by weight.

The solids are maintained in suspension in a reactor vessel and mixed with nutrients and oxygen. If necessary, an acid or alkali may be added to control pH. Microorganisms also may be added if a suitable population is not present. When biodegradation is complete, the soil slurry is dewatered. Dewatering devices that may be used include clarifiers, pressure filters, vacuum filters, sand drying beds, or centrifuges.

2.5.17 Chemical extraction

Waste contaminated soil and extractant are mixed in an extractor, dissolving the contaminants. The extracted solution is then placed in a separator, where the contaminants

and extractant are separated for treatment and further use. Chemical extraction does not destroy wastes but is a means of separating hazardous contaminants from soils, sludges, and sediments, thereby reducing the volume of the hazardous waste that must be treated. The technology uses an Extracting chemical and differs from soil washing, which generally uses water or water with wash-improving additives. Commercial-scale units are in operation. They vary in regard to the Chemical employed, type of equipment used, and mode of operation.

Physical separation steps are often used before chemical extraction to grade the soil into coarse and fine fractions, with the assumption that the fines contain most of the contamination. Physical separation can also enhance the kinetics of extraction by separating out particulate heavy metals, if these are present in the soil.

2.5.18 Chemical Reduction/Oxidation

Reduction/oxidation chemically converts hazardous contaminants to non-hazardous or less toxic compounds that are more stable, less mobile, and/or inert. The oxidizing agents most commonly used are ozone, hydrogen peroxide, hypochlorites, chlorine, and chlorine dioxide.

Reduction/oxidation (Redox) reactions chemically convert hazardous contaminants to nonhazardous or less toxic compounds that are more stable, less mobile, and/or inert. Redox reactions involve the transfer of electrons from one compound to another. Specifically, one reactant is oxidized (loses electrons) and one is reduced (gains electrons). The oxidizing agents most commonly used for treatment of hazardous contaminants are ozone, hydrogen peroxide, hypochlorites, chlorine, and chlorine dioxide. Chemical redox is a full-scale, well-established technology used for disinfection of drinking water and wastewater, and it is a common treatment for cyanide wastes. Enhanced systems are now being used more frequently to treat contaminants in soils. Chemical reduction/oxidation is a short- to medium-term technology.

2.5.19 Dehalogenation

Reagents are added to soils contaminated with halogenated organics. The dehalogenation process is achieved by either the replacement of the halogen molecules or the decomposition and partial volatilization of the contaminants.

Contaminated soil is screened, processed with a crusher and pug mill, and mixed with reagents. The mixture is heated in a reactor. The dehalogenation process is achieved by either the replacement of the halogen molecules or the decomposition and partial volatilization of the contaminants.

2.5.20 Separation

Separation techniques concentrate contaminated solids through physical and chemical means. These processes seek to detach contaminants from their medium (i.e., the soil, sand, and/or binding material that contains them).

Ex situ separation can be performed by many processes. Gravity separation and sieving/physical separation are two well-developed processes that have long been primary methods for treating municipal wastewaters. Magnetic separation, on the other hand, is a much newer separation process that is still being tested.

2.5.21 Soil Washing

Contaminants sorbed onto fine soil particles are separated from bulk soil in an aqueous-based system on the basis of particle size. The wash water may be augmented with a basic leaching agent, surfactant, pH adjustment, or chelating agent to help remove organics and heavy metals.

Soil washing is a water-based process for scrubbing soils ex situ to remove contaminants. The process removes contaminants from soils in one of two ways:

2.5.22 *Soil Vapor Extraction (Ex Situ)*

A vacuum is applied to a network of aboveground piping to encourage volatilization of organics from the excavated media. The process includes a system for handling off-gases.

Ex situ soil vapor extraction (SVE) is a full-scale technology in which soil is excavated and placed over a network of aboveground piping to which a vacuum is applied to encourage volatilization of organics. Soil piles are generally covered with a geomembrane to prevent volatile emissions and to prevent the soil from becoming saturated by precipitation. The process includes a system for handling off-gases. Advantages over its in situ counterpart include that the excavation process forms an increased number of passageways, shallow ground water no longer limits the process, leachate collection is possible, and treatment is more uniform and easily monitored. The major disadvantage over in situ SVE is the increased excavation costs.

2.5.23 *Solar Detoxification*

Solar detoxification is a process that destroys contaminants by photochemical and thermal reactions using the ultraviolet energy in sunlight. In this process, vacuum extraction is used to remove contaminants from soils. After condensation, contaminants are mixed with a semiconductor catalyst such (e.g., titanium dioxide), and fed through a reactor which is illuminated by sunlight. Ultraviolet light activates the catalyst, which results in the formation of reactive chemicals known as "radicals". These radicals are powerful oxidizers that break down the contaminants into non-toxic by-products such as carbon dioxide and water.

A big advantage of solar detoxification over conventional treatment processes such as those using granular activated carbon or air stripping is that it completely destroys the toxic compounds in the water instead of simply removing or displacing them. The solar process also has no atmospheric emissions.

2.5.24 Solidification/ Stabilization (Ex Situ)

Contaminants are physically bound or enclosed within a stabilized mass (solidification), or chemical reactions are induced between the stabilizing agent and contaminants to reduce their mobility (stabilization).

There are many innovations in the stabilization and solidification technology. Most of the innovations are modifications of proven processes and are directed to encapsulation or immobilizing the harmful constituents and involve processing of the waste or contaminated soil. Nine distinct innovative processes or groups of processes include: (1) bituminization, (2) emulsified asphalt, (3) modified sulfur cement, (4) polyethylene extrusion, (5) pozzolan/Portland cement, (6) radioactive waste solidification, (7) sludge stabilization, (8) soluble phosphates, and (9) vitrification/molten glass.

2.5.25 Hot Gas Decontamination

The process involves raising the temperature of the contaminated equipment or material for a specified period of time. The gas effluent from the material is treated in an afterburner system to destroy all volatilized contaminants.

The process involves raising the temperature of the contaminated equipment or material to 260 °C (500 °F) for a specified period of time. The gas effluent from the material is treated in an afterburner system to destroy all volatilized contaminants. The method eliminates a waste that currently is stockpiled and requires disposal as a hazardous material. This method will permit reuse or disposal of scrap as nonhazardous material.

Hot gas decontamination can also be used for decontamination of explosives-contaminated masonry or metallic structures. The method involves sealing and insulating the structures, heating with hot gas stream to 260 °C (500 °F) for a prescribed period of time, volatilizing the explosive contaminants, and destroying them in an afterburner. Operating conditions are site-specific. Contaminants are completely destroyed.

2.5.26 *Thermal Desorption*

Wastes are heated to volatilize water and organic contaminants. A carrier gas or vacuum system transports volatilized water and organics to the gas treatment system. Thermal desorption is a physical separation process and is not designed to destroy organics. Wastes are heated to volatilize water and organic contaminants. A carrier gas or vacuum system transports volatilized water and organics to the gas treatment system. The bed temperatures and residence times designed into these systems will volatilize selected contaminants but will typically not oxidize them.

2.5.27 *Landfill Cap*

Landfill caps are used for contaminant source control. Landfill caps can be used to:

Caps can minimize exposure on the surface of the waste facility.

Prevent vertical infiltration of water into wastes that would create contaminated leachate.

Contain waste while treatment is being applied.

Control gas emissions from underlying waste.

Create a land surface that can support vegetation and/or be used for other purposes.

Landfill Capping is the most common form of remediation because it is generally less expensive than other technologies and effectively manages the human and ecological risks associated with a remediation site.

2.6 Contaminants Found At Typical Brownfields

The following table identifies several activities that may have caused contamination at brownfields. The table summarizes contaminants that are related to such activities and identifies sources for the contaminants; however, it is not an exhaustive list of contaminants that can be found at a Brownfields. Identifying contaminants that may be present should be determined on a site-by-site basis. Such a determination should be conducted thoroughly and carefully. (U.S. Environmental Protection Agency, Road Map to Understanding Innovative Technology Options for brownfields Investigation and Cleanup, 1997)

Table 2.1 Various contaminants found at brownfields

Agriculture	Volatile organic compounds (VOC); arsenic, copper, carbon tetrachloride, ethylene dibromide, and methylene chloride; pesticides; insecticides; herbicides; grain fumigants
Automotive refinishing and repair	Some metals and metal dust; various organic compounds; solvents; paint and paint sludges; scrap metal; waste oils
Battery recycling and disposal	Lead; cadmium; acids
Chloro-alkali manufacturing	Chlorine compounds; mercury
Coal gasification	Polynuclear aromatic hydrocarbons (PAH)
Cosmetics manufacturing	Heavy metals; dusts; solvents; acids
Dry cleaning activities	VOCs such as chloroform and tetrachloroethane; various solvents; spot removers; fluorocarbon 113
Electroplating operations	Various metals such as cadmium, chromium, cyanide, copper, and nickel
Glass manufacturing	Arsenic; lead
Herbicide manufacturing and use	Dioxin; metals; herbicides
Hospitals	Formaldehyde; radionuclides; photographic chemicals; solvents; mercury; ethylene oxide; chemotherapy chemicals
Incinerators	Dioxin; various municipal and industrial waste
Landfills :municipal and industrial	Metals; VOCs; polychlorinated biphenyl (PCB); ammonia; methane; household products and cleaners; pesticides; various wastes
Leather manufacturing	Toluene; benzene
Machine shops/metal fabrication	Metals; VOCs; dioxin; beryllium; degreasing agents; solvents; waste oils
Marine maintenance industry	Solvents; paints; cyanide; acids; VOC emissions; heavy metal sludges; Degreasers
Munitions manufacturing	Lead; explosives; copper; antimony
Paint/ink manufacturing	Metals (such as chromium, cadmium, lead, and zinc);

	VOCs;chloroform; ethyl benzene; solvents; paints; inks
Pesticide manufacturing	VOCs; arsenic; copper; pesticides; insecticides; herbicides; fungicides;xylene; chlorinated organic compounds; solvents
Petroleum refining and reuse	Petroleum hydrocarbons; benzene, toluene, ethylbenzene, xylene (BTEX); fuels; oil and grease
Pharmaceutical manufacturing	Lead; various organic chemicals; organic solvents
Photographic manufacturing and uses	Silver bromide; methylene chloride; solvents; photographic products
Plastics manufacturing	Polymers; phthalates; cadmium; solvents; resins; chemical additives; VOCs
Printing industry	Silver; solvents; acids; waste oils; inks and dyes; photographic chemicals
Railroad yards	Petroleum hydrocarbons; VOCs; BTEX; solvents; fuels; oil and grease;lead; PCBs
Research and educational institutions	Inorganic acids; organic solvents; metals and metal dust; photographic waste; waste oil; paint; heavy metals; pesticides
Scrap metal operations	Various metals (such as lead and nickel); PCBs; dioxin; transformers
Smelter operations	Metals (such as lead, copper, and arsenic)
Semiconductor manufacturing	Metals; VOCs; carbon tetrachloride; degreasing agents; solvents
Wood pulp and paper manufacturing	Chlorinated organic compounds; dioxin; furans; chloroform; resin acids
Wood preserving	Creosote; pentachlorophenol (PCP); arsenic; chromium; copper; PCB; PAHs; beryllium; dioxin; wood preservatives

3 SPATIAL INTERPOLATION METHODS (SIM) USED IN IDENTIFYING CONTAMINATED ZONES

The environmental remediation arena is inherently an interdisciplinary, multi-faceted field with many scientific disciplines cooperation to develop solutions to hazardous waste remediation problems. While solving these problems physical and financial constraints limits the methodology.

It can be clearly seen from the previous chapter that the remediation techniques mostly depend on the type and the level of contamination. Without knowing the level and point by point estimation of the contaminated area, it is very hard to define the process conditions and methodology for the remediation techniques. Such as required time for remediation, concentration of the solutions, type of the organics and chemicals used for remediation and their dosages. These points are very important for the success and the efficiency of the remediation process. In order to obtain the contamination map of the area some interpolation and estimation techniques used.

Interpolation techniques especially geostatistics are used for mapping hazardous, toxic, and radioactive waste (HTRW) sites. (U.S. Army Corps of Engineers, HTRW Report, 1997). One very fundamental aspect of perhaps all HTRW site investigations that deal with environmental contamination is the need to characterise the extent and spatial distribution of contamination. Such a characterization typically would include the determination of spatial trends and variability using a variety of statistical or analytical tools. A principal difficulty in doing this is the fact that measurements may be few, or may be sparsely scattered over large regions. Thus it is required to make predictions (or estimates) at locations where measurements of contaminant concentration are not available.

HTRW site investigations involve complex administrative, scientific, and engineering functions and are truly interdisciplinary. Scientists and engineers, for instance, may be confronted with administrative findings or directives, associated with fiscal, managerial, or regulatory input, that may either guide or constrain their work. Frequently, a HTRW site

investigation will benefit from input from earth-science disciplines such as geology, hydrogeology, and chemistry, among others. Some HTRW site investigations are large enough to use several individuals from each field associated with implementing or operating of these disciplines, as well as many others disciplines.

Interpolation techniques are utilised in identifying contaminated zones in potentially contaminated sites (Zirschky and Harris, 1986), in image re-construction or visualization (eg. medical imaging, Carr, 1996, Rohling et al. 1998), remotely sensed satellite imagery (eg. Fogel, 1997), identification of mining resources (eg. Isaaks and Srivastava, 1989) and topographical mapping (Oliver and Webster, 1990, Rosenbaum and Soderstrom, 1996) or Digital Elevation Modelling (DEM) (USGS, 1987, Hutchinson, 1997). Based on the specific area in which they are used, these methods are called "geostatistical tools", "GIS methods", or "DEM". The most general term for these methods is "Spatial Interpolation Methods (SIM)" and their major assumption is that spatially closer locations are more likely to have similar observation values than those, which are far apart.

SIM are the procedures for estimating the value of characteristics at unsampled locations and are used to provide contours. SIM are used to provide contours for displaying data graphically (DEM); to estimate some property of the surface at a given location (mining); to visualise surfaces of human organs in clinical investigation; and, to aid the spatial decision making process in detecting hot-spots (environmental assessment) (U.S. Army Corps of Engineers, HTRW Report, 1997, U.K. DoE Contaminated Land Research Report No. 4, 1994).

SIM are conventional techniques used in the above-mentioned contexts, and they usually construct surfaces from point observations. Yet, there are uncertainties about these re-constructed surfaces due to the fact that they assume that value of a data point is valid for the whole area covered by the grid in which it lies. Among other problems implied by SIM are:

1. Different interpolation algorithms yield different contours from the same data
2. Depending on the algorithm used to contour data, the raw data are not necessarily honoured

3. Small areas with high contamination levels are difficult to model using these algorithms due to the smoothing effect.

A further handicap in representing surfaces is the existence of noise in the observed values. For example, sampling errors resulting from the sampling methodology and laboratory analysis errors may exist (Ramsey et al, 1995). Consequently, the performance of SIM should not only be investigated on data without noise.

3.1 A Methodology for SIM

Spatial interpolation methods are used to estimate the contamination data at unsampled site locations. Spatial interpolation can also be used when preparing irregularly scattered data to construct a contour map or a contour surface, which is a two-dimensional representation of a three dimensional surface. The spatial interpolation methods differ in their assumptions. They may have a local or global perspective, or be of a deterministic or stochastic nature. SIM may be classified under five different categories (Waters, 1998).

3.1.1 *Point versus areal interpolation methods*

In a point interpolation method, unobserved locations are interpolated using observed locations of the area whereas in an areal interpolation method, values of data for a target zone are determined using another set of data on a given source zone. In point interpolation unobserved points are located as within a grid. The interpolated values on the grid are used for drawing contours by connecting adjacent grid points with an interpolated straight line. Hence, the width of the interpolated grid affects the performance of the interpolation algorithm.

3.1.2 *Global versus local interpolation methods*

Estimation over a large area within which we have many samples is global estimation. In local estimation, we are considering a smaller area, one in which there are few samples; in such situations we often use nearby samples located outside the area being estimated.

A global interpolator uses all data in the whole site and hence, the data collected from the whole region affects a single interpolated point. On the other hand, local interpolators consider a neighbourhood around the interpolated point while calculating the interpolated value. Consequently, a global interpolator can be localised by adjusting some parameters such as the number of data included in the calculations, the radius, etc.

In practical situations there is usually a large target area over which samples are collected. Hence a single global estimate rarely satisfies the goals of a study and we usually also require a complete set of local estimates. For example in a pollution study, an estimate of the overall concentration does not give us the information we need to decide which specific locations have unacceptably high concentration.

Global estimation is fairly straightforward if the samples are located on irregular grid or are located randomly. Global estimation methods should therefore account for the possibility of clustering.

Local estimates need to account for the distance from the point(s) we are estimating to the individual sample locations. Some samples will be much closer than others will and the values at these closer locations should be given more weight than the values that are further away. Local estimates like the global ones are often affected by clustered sampling similar to the global ones. A group of closely spaced samples with similar values contains redundant information. Our local estimate would likely improve if we could locate samples uniformly over the small area we are interested in. The weights assigned by local estimation methods should account both for their from distance to the samples and also for the possible redundancy between samples.

To summarize;

Local estimation

- estimates a value using the values of cells in the neighbourhood
- needs to know the extent of the neighbourhood
- usually uses the distance weighted mean value of surrounding values

Global approximation

- uses all values in the grid

- mathematically intensive but very accurate
- fits a surface passing through all points

3.1.3 *Exact versus approximate interpolation methods*

An exact interpolator generates a surface, which passes through all observations (with known values), whereas approximate interpolators do not necessarily do so. The implication is that an exact interpolator assumes the value of the observation value even when an unobserved location on the generated grid falls upon the observed location. Thus, the unobserved location is not interpolated, but assumes the value of the real data. Approximate interpolators should be used when the data involves uncertainty (sampling errors including wrong method of sample collection, erroneous recording and/or analysis errors) and it is desired that the interpolator smooths out the sampling errors since interpolated values override observed values.

To summarize;

Exact interpolators

- honours the data points upon which the interpolation is based.
- the surface passes through all points whose values are known
- honouring data points is seen as an important feature in many applications eg. the oil industry

Approximate interpolators

- used when there is some uncertainty about the given surface values this utilises the belief that in many data sets there are global trends, which vary slowly, overlain by local fluctuations, which vary rapidly and produce uncertainty (error) in the recorded values
- the effect of smoothing will therefore be to reduce the effects of error on the resulting surface

3.1.4 *Stochastic versus deterministic interpolation methods.*

In a probabilistic model, the available sample data are viewed as the result of random process. The processes that actually do create an ore deposit, a petroleum reservoir or a hazardous waste site are certainly complicated, and our understanding of them may be so

poor that their complexity appears as random behaviour to us, but it does not mean that they are random; it simply means that we are ignorant.

Unfortunately, our ignorance does not excuse us from the difficult task of making predictions about how apparently random phenomena behave where we have not sampled them. Though earth science data are not, in fact, the result of random processes, this conceptualization does not turn out to be a useful one for the problem of estimation. Though the word random often connotes “unpredictable”, it turns out that viewing our data as the outcome of random process does not help us with the problem of predicting unknown values. Not only does it give us estimation procedures that, in practice, have sometimes proved to be very good; it also gives us ability to gauge the accuracy of our estimates and to assign confidence intervals to them.

Stochastic interpolators assume that there is an underlying random process explaining the distribution of concentration values from which the collected data result. For instance, Kriging is a stochastic interpolator, which assumes stationarity of mean and variance. Often, stochastic assumptions do not hold for the underlying process. Deterministic interpolators do not use probability theory. An example is the inverse distance method, which, while interpolating an unobserved location, calculates the weight of each observation simply using its distance from the unobserved location. However deterministic modelling is possible only if the context of the data values is well understood. The data values, by themselves, do not reveal what should be the appropriate model. Stochastic methods incorporate the concept of randomness and stochastic interpolators include trend surface analysis procedures such as trend surface analysis allow the statistical significance of the surface and uncertainty of the predicted values to be calculated

3.1.5 Gradual versus abrupt interpolators

Gradual interpolators generate a smooth surface on a continuum with slow changes. However, one may convert a gradual interpolator into an abrupt one by reducing the number of neighbouring observations while interpolating a grid point. Then, the interpolated values reflect sudden changes in nearby observations.

Typical examples of gradual interpolators are the distance weighted and moving average and gradual interpolators. Gradual interpolators usually produces an interpolated surface with gradual changes however, if the number of points used in the moving average is reduced to a small number, or even one, there would be abrupt changes in the surface. Then it may be necessary to include barriers in the interpolation process such as

- Semipermeable barriers (e.g. weather fronts will produce quickly changing but continuous values)
- Impermeable barriers (e.g. geologic faults will produce abrupt changes)

3.2 Conventional SIM

In this thesis the effectiveness of five popular SIM (Radial Basis Functions (RBF), Kriging, Minimum Curvature (MNC), Shepard's Inverse Distance Method , Triangulation) are tested for their performance on mapping the contaminant dispersion over the site.

3.2.1 Kriging

Kriging is an interpolation method commonly used in various fields, such as mining, geographical mapping, and environmental assessment of sites which is originated by Krige (1951) and developed by Matheron (1971). The idea is to consider all observations as a realization of a random spatial process. Following paragraph explains the rationality behind the well known geostatistical method, Kriging.

For introductory purposes kriging can be defined as a technique for determining the optimal weighting of measurements at sampled locations for obtaining predictions, or estimates, at unsampled locations. Kriging is well suited for making point and block estimates. However, much of the advantage of using geostatistical procedures, such as kriging, lies not just in the point and block estimates they provide, but in the information they provide concerning uncertainty associated with these estimates. The uncertainty information is usually quantified as either the standard deviation (or variance) associated with kriging estimates and is referred to as kriging standard deviation (or kriging variance).

Kriging is a technique for determining in an optimal manner the weighting of measurements at sampled locations for obtaining predictions at unsampled locations. These optimal weights depend on spatial trends and correlations that may be present. And kriging is a stochastic interpolation method. (U.S. Army Corps of Engineers, HTRW Report, 1997).

A spatial process can be viewed as having a large-scale or regional component and a smaller scale or local component; both of these components need to be accounted for when modelling spatial process. The large-scale component is referred to as the mean field and is most often modelled by a spatial trend, which may or may not be constant over the region. The smaller scale component is a random fluctuation, which is mathematically combined with the trend to make up the sample at a point. The random component is usually assumed to be zero on the average but can be either positive or negative in individual samples. The separation of the trend from the random components is a problem which is scale-dependent and requires some judgment to determine. There can be several solutions to the problem of “separating the trend” and random components that may be useful for various geostatistical purposes when using a single set of data.

In stochastic methods, it is assumed that the measurements, actual or not, constitute a single realization of a random (or stochastic) process. One advantage of assuming the existence of such a random process is that measures of uncertainty, such as the variance used in kriging, can be defined. These measures of uncertainty permit objective assessment of the performance of a spatial prediction technique on the basis of how small such measures are. Once a measure of uncertainty has been selected, the weights to be used in spatial prediction may be determined so as to explicitly minimize the measure of uncertainty. In short, the use of stochastic techniques provides the investigator with a way of objectively quantifying errors and determining weights. In practice, spatial predictions obtained using kriging are almost always accompanied by a measure of the associated error.

3.2.1.1 Variograms

A central idea in geostatistics is the use of spatial correlation to improve spatial

predictions, or interpolations. The variogram is the principal tool used to characterise the degree of spatial correlation present in the data and is fundamental to kriging. (U.S. Army Corps of Engineers, HTRW Report, 1997).

A measure of the similarity between data (the co-variance between the two data) for distance h apart is obtained. This is repeated for all samples that are h distance apart and the average squared difference obtained. This similarity measure is called $\gamma(h)$, $g(h)$ or h -scatterplots. These are plotted on a x - y plot with the x -axis being the distance h , and $g(h)$ on the y -axis.

An h -scatterplots shows all possible pairs of data values whose locations are separated by a certain distance in a particular direction. On most of the scatterplots, the x -axis is labelled $v(t)$ and the y -axis is labelled $v(t+h)$. The x -coordinate of a point corresponds to the v value at a particular location and the y -coordinate to the v value a distance and direction h away. (Isaaks and Srivastava, 1989)

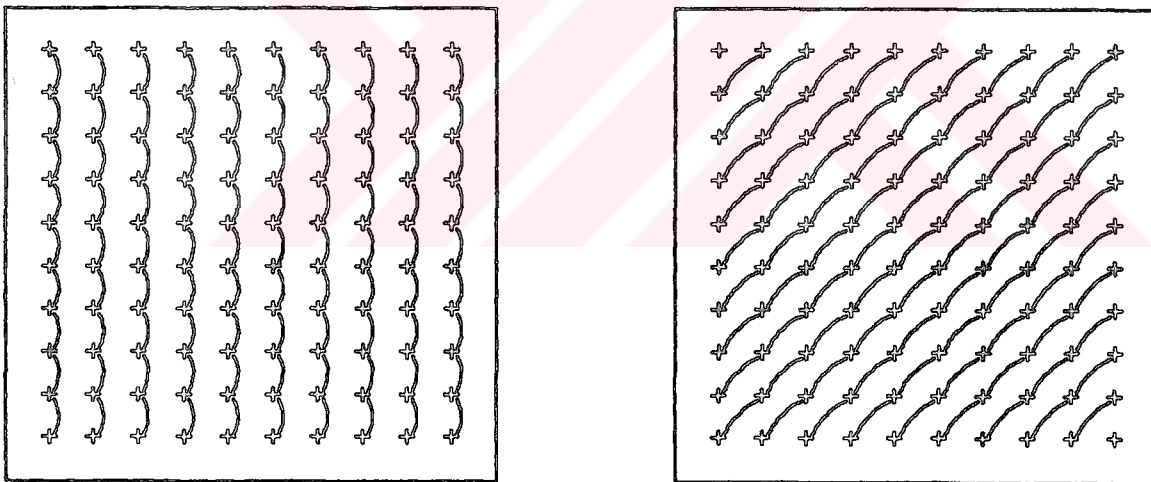


Figure 3.1 Examples of how data can be paired to obtain an h -scatterplot

The shape of the cloud of points on an h -scatterplots tells us how continuous the data values are over a certain distance in a particular direction. If the data values at locations separated by h are very similar then the pairs will plot close to the line $x=y$, a 45-degree line passing through the origin. As the data values become less similar, the cloud of points on the h -scatterplots becomes fatter and more diffuse. (Figure 3.2)

The correlation coefficient, ρ , is the statistic that is most commonly used to summarise the linear relationship between two variables. It is calculated from;

$$\rho = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - m_x)(y_i - m_y)}{\sigma_x \sigma_y} \quad (3.1)$$

where n is the number of data; $x_1, \dots, x_n$ are the data values, m_x is their mean, and σ_x is their standard deviation for the first variable, $y_1, \dots, y_n$ are the data values, m_y is their mean, and σ_y is their standard deviation for the second variable.

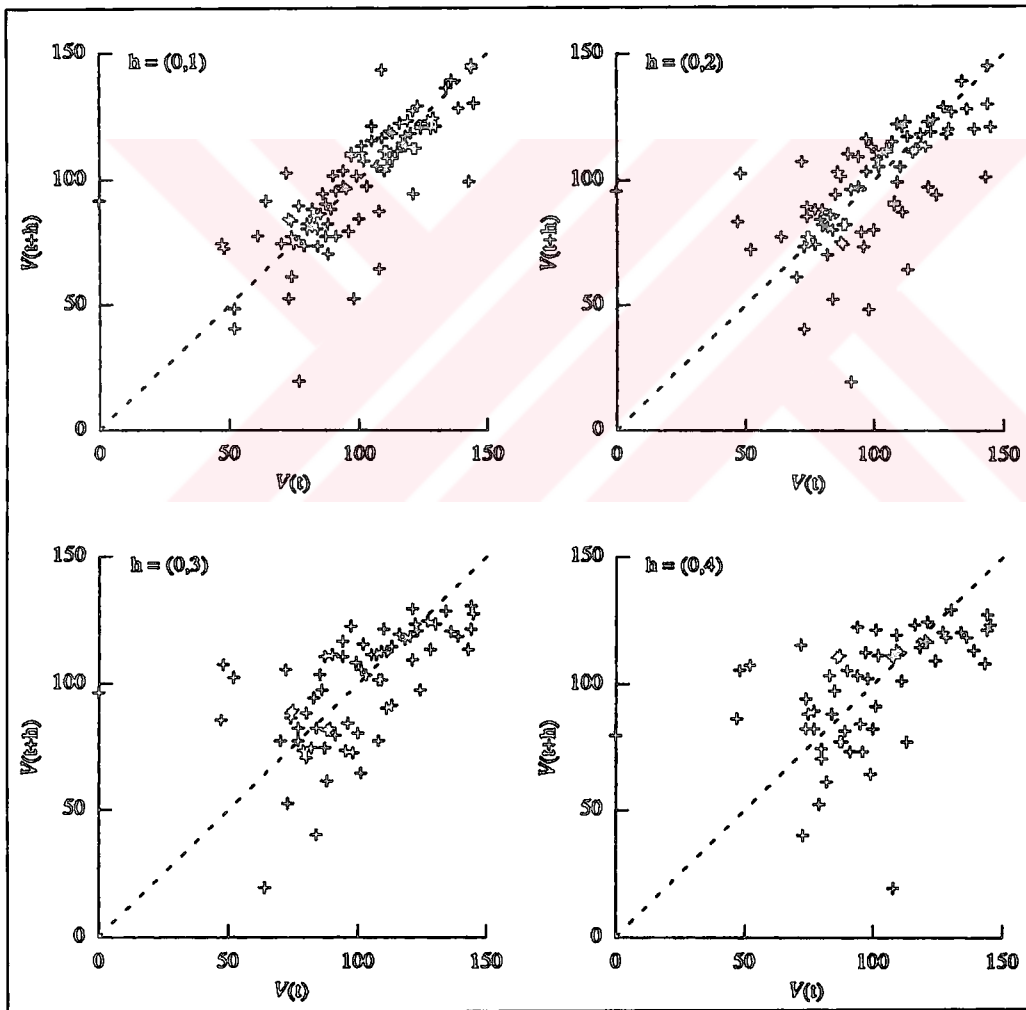


Figure 3.2 h-scatterplots for four separation distances.

As the cloud of points in the h-scatterplots spread out, we expect the correlation coefficient, ρ , to be close to zero. The relationship between the correlation coefficient of an h-scatterplots and h is called the correlation function or correlogram. The correlation coefficient depends on which being a vector, has both magnitude and direction.

An alternative index for spatial continuity is the covariance. Covariance is the numerator of the correlation. The relationship between the covariance of an h-scatterplots and h is called the covariance function.

Another plausible index for the spread out of the cloud is the moment of inertia (moi) about the line $x=y$, which can be calculated from the following;

$$moi = \frac{1}{2n} \sum_{i=1}^n (x_i - y_i)^2 \quad (3.2)$$

Unlike the other two indices of spatial continuity, the moment of inertia increases as the cloud spread outs. The relationship between the moment of inertia of an h-scatterplots and h is traditionally called the semivariogram or simply the variogram.

Although the h-scatterplots contain much more information than any of the three summary statistics, it is quite common to bypass the actual h-scatterplots and go directly to either $\rho(h)$, $C(h)$, $\gamma(h)$ to describe spatial continuity. It is convenient, therefore, to have the formulas for these functions expressed directly in terms of the data values rather than in terms of x and y coordinates of an h-scatterplots.

Then the covariance function, $C(h)$, can be calculated from the following equation;

$$C(h) = \frac{1}{N(h)} \sum_{(i,j)_{h_{i,j}=h}} v_i v_j - m_{-h} m_{+h} \quad (3.3)$$

The data values are $v_1, \dots, v_n$; the summation is over only the $N(h)$ pairs of data whose locations are separated by h.

m_{-h} is the mean of all the data values whose locations are $-h$ away from some other data

location:

$$m_{-h} = \frac{1}{N(h)} \sum_{i, h_{i,j} = -h} v_i \quad (3.4)$$

m_{+h} is the mean of all the data values whose locations are $+h$ away from some other data location:

$$m_{+h} = \frac{1}{N(h)} \sum_{i, h_{i,j} = +h} v_i \quad (3.5)$$

The values of m_{-h} and m_{+h} are generally not equal in practice. The correlation function, $\rho(h)$, is the covariance function standardised by the appropriate standard deviations:

$$\rho(h) = \frac{C(h)}{\sigma_{-h} \sigma_{+h}} \quad (3.6)$$

σ_{-h} is the standard deviation of all the data values whose locations are $-h$ away from some other data location:

$$\sigma_{-h}^2 = \frac{1}{N(h)} \sum_{i, h_{ij} = -h} v_i^2 - m_{-h}^2 \quad (3.7)$$

σ_{+h} is the standard deviation of all the data values whose locations are $+h$ away from some other data location:

$$\sigma_{+h}^2 = \frac{1}{N(h)} \sum_{i, h_{ij} = +h} v_i^2 - m_{+h}^2 \quad (3.8)$$

Like the means, the standard deviations are usually not equal in practice. The variogram, $\gamma(h)$, is the half average squared difference between the paired data values:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(i,j), h_{ij}=h} (v_i - v_j)^2 \quad (3.9)$$

The values $\rho(h)$, $C(h)$ and $\gamma(h)$ are unaffected if we switch all of the i and j subscripts in the preceding equations. Then the above equation becomes

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(j,i), h_{ji}=h} (v_j - v_i)^2 \quad (3.10)$$

So over all (j,i) pairs that are separated by h , we could sum over all (i,j) pairs that are separated by $-h$. As a result we can obtain

$$\gamma(h) = \gamma(-h) \quad (3.11)$$

This result entails that the variogram calculated for any particular direction will be identical to the variogram calculated in the opposite direction. The correlation and the covariance function share this property. (U.S. Army Corps of Engineers, HTRW Report, 1997)

Variograms are obtained from the data for different directions. Typically, these are N-S, E-W, NE-SW and NW-SE. Windows, or allowable deviations from these directions are often used if the data is not on a regular grid. Note that no matter what direction, a window of $\pm 90^\circ$ will incorporate all the data. The directional convention used in geostatistics is E-W is 0° , N-S is 90° , NE-SW is 45° and NW-SE is -45° . The window is set at the direction plus or minus the deviation angle. (Isaaks and Srivastava, 1989)

There are two cases for the distribution of the data points. The first is they are located on a regular grid. In this case $g(h)$ may be calculated using an arithmetic average of the distances. If the data is not located on a regular grid, then a modified version of the equation needs to be used to use a weighted average rather than the arithmetic average. This is called the *moment centre*.

Two other measures are obtainable. The first is the *average distance* of all the pairs of data points whose distance fell within a given class interval. This is calculated from the cumulative distance divided by the number of samples. The second is a measure of *drift*, or the general increase or decrease in grade with distance. This is calculated from the cumulative assay difference divided by the number of samples.

If the variogram rises, and then levels off or stabilises around some value, or in other words if the variogram reaches either reaches or becomes asymptotic to a constant value as h increases to a value and that is said to have reached a *sill*. This is theoretically the sample variance.

The distance at which the rising variogram reaches the sill or the variogram remains at or close to the sill is called the *range*. The range is the distance at which the covariance becomes zero, so it marks the limit of the zone of influence of a single sample. Beyond the range samples are no longer correlated and are independent.

If the sample interval is greater than the range, then the semi-variance is equal to the variance for all h . There will be no rising part to the variogram so the plot will show a scatter about the variance. If this is the case then it is safe to use conventional statistics for estimation.

If the variogram rises without levelling off, then no sill is present. This may indicate drift if the variance is crossed, or one of the models which do not possess a sill (de Wijsian, power, linear). The models without a sill result from a RF with unlimited capacity for dispersion.

The intercept of the variogram on the gamma axis is the semi-variance when $h=0$. If the intercept is greater than zero then a random or unstructured component of variation at $h=0$ is present. This is called the *nugget effect* and is represented by C_0 . Ideally, the nugget effect should be zero as two samples from the same point should have the same value. A non zero value could indicate errors in sampling or it represents the sill of a very small scale structured component whose range is much less than the sample interval.

Typically, four variograms with different orientations are constructed. These orientations are usually N/S, E/W, NW/SE and NE/SW. If the variograms are the same, then the distribution of values is said to be *isotropic*. If they are not the same then distribution is said to be *anisotropic*. If anisotropy is present, and is reflected by the same sill but different ranges of the same model, then the anisotropy is called *geometric anisotropy*. This anisotropy can be removed by a distance transformation of the coordinates. If the variograms show different models or different sills then *zonal anisotropy* does exist. Removal of zonal anisotropy is complex and not discussed here.

Relative variograms are sometimes useful in correcting for zonal anisotropy. They normalise the variogram by dividing $g(h)$ by either the mean square or the variance of the sequence used to compute the variogram. This makes the variograms more directly comparable.

Logarithmic variograms computes a variogram of the logs of the assays. This transformation reduces the effect of erratic highs in the data and smooths the variogram. In many cases the overall shape is little different from untransformed variograms.

Indicator variograms are a special case of variograms. The indicator is a binomial variable, which takes a value of 0 or 1. The 0's may represent barren holes or values below cut off, and the 1's represent holes above cut off. The resulting variogram of these 0's and 1's is the indicator variogram. These may show features that are not apparent on the assay variogram such as clustering of high-grade zones. (U.S. Army Corps of Engineers, HTRW Report, 1997)

A simple monotonic function is usually selected to approximate the variogram. Four such functions are often used in practice are:

3.2.1.1.1 The exponential variogram

The exponential model is also linear at the origin, but reaches the sill asymptotically, well beyond the value of the true range, ie it approaches the sill gradually without ever reaching

it.

Parameters: sill, $s > 0$; nugget, $0 < g < s$; range, $r > 0$)

$$\gamma(h) = \begin{cases} g + (s - g) \left[1 - \exp\left(-3\frac{h}{r}\right) \right], & h > 0 \\ 0, & h = 0 \end{cases} \quad (3.12)$$

3.2.1.1.2 The spherical variogram

The spherical model is linear at the origin indicating good continuity.

Parameters: sill, $s > 0$; nugget, $0 < g < s$; range, $r > 0$)

$$\gamma(h) = \begin{cases} s, & h > r \\ g + (s - g) \left[1.5 \frac{h}{r} - 0.5 \left(\frac{h}{r} \right)^3 \right], & 0 < h \leq r \\ 0, & h = 0 \end{cases} \quad (3.13)$$

3.2.1.1.3 The gaussian variogram

The Gaussian model is parabolic at the origin (indicating perfect continuity) and reaching the sill asymptotically

Parameters: sill, $s > 0$; nugget, $0 < g < s$; range, $r > 0$)

$$\gamma(h) = \begin{cases} g + (s - g) \left[1 - \exp\left(-3\frac{h}{r}\right)^2 \right], & h > 0 \\ 0, & h = 0 \end{cases} \quad (3.14)$$

3.2.1.1.4 Power Model variogram

Power $0 < w < 2$, positive slope b , and the distance h may be anisotropic in 2D or 3D

$$\gamma(h) = bh^w \quad (3.15)$$

3.2.1.1.5 The linear variogram

Parameters: nugget, $g > 0$; slope, $b > 0$)

$$\gamma(h) = \begin{cases} g + bh, & h > 0 \\ 0, & h = 0 \end{cases} \quad (3.16)$$

3.2.1.1.6 Hole Effect Variogram

The model defined by a length r to the first peak (size of the underlying cyclic features) and positive variance contribution value c

$$\gamma(h) = c \left[1 - \cos\left(\frac{h}{r} \pi\right) \right] \quad (3.17)$$

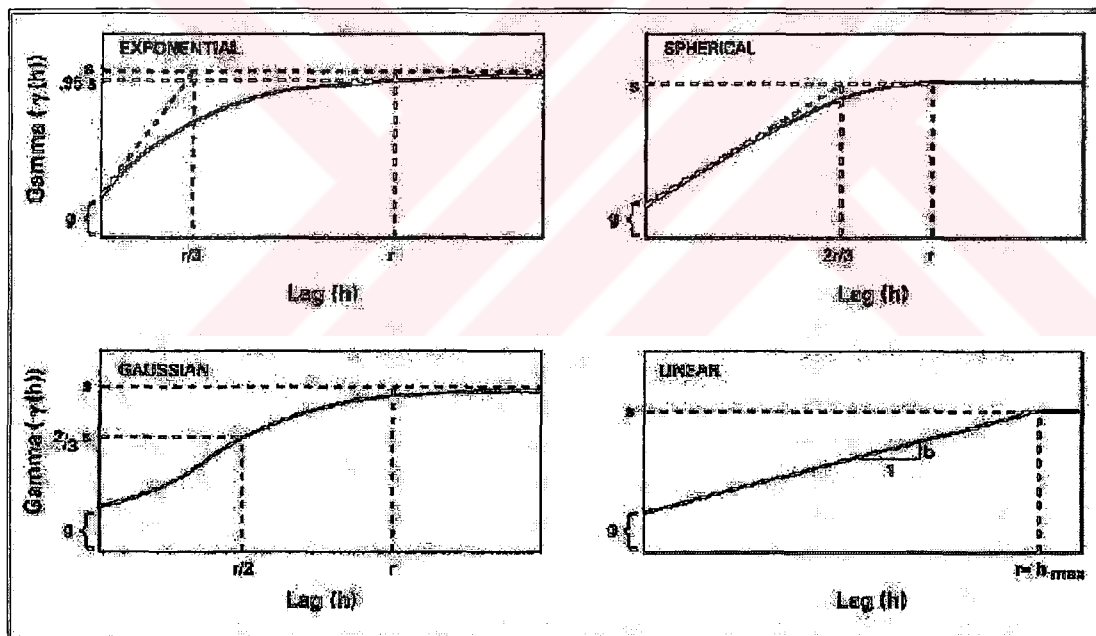


Figure 3.3 Theoretical Variograms

3.2.1.1.7 Regionalized Random Variable

Given a regionalized random variable $Z(x)$ with a known theoretical variogram, the question is: how can the value of $Z(x)$ be predicted at an arbitrary location, based on

measurements taken at other locations. Suppose that Z is measured at n specified locations: $Z(x_1), \dots, Z(x_n)$. For example, Z could correspond to lead contamination level and the locations might correspond to n points over an old refinery location. Let a new location be given by $x_0 = (u_0, v_0)$ and denote the i^{th} measurement be denoted by $x_i = (u_i, v_i)$. Suppose that, based on prior knowledge of the geology, there are no prevailing trends in lead contamination, so the mean of $Z(x)$ is assumed to be constant over the entire region:

$$\mu(x) = \mu(\text{constant})$$

Suppose the investigator wants to predict the value of $Z(x_0)$ by using a linear predictor, $Z'(x_0)$, which is defined as a weighted linear combination of the measured data

$$Z'(x_0) = \sum_{i=1}^n w_i Z(x_i) \quad (3.18)$$

where w is the weight assigned to $Z(x_i)$. To determine specific values for the weights, some criteria need to be specified for $Z(x_0)$ to be a good predictor of $Z'(x_0)$. The first criterion is that $Z'(x_0)$ be an unbiased predictor of $Z(x_0)$, which is expressed as

$$E[Z'(x_0) - Z(x_0)] = 0 \quad (3.19)$$

An unbiased predictor will neither consistently overpredict nor underpredict $Z(x_0)$ because the statistical expectation of the prediction errors is zero. The second criterion for a good predictor is that it has small prediction variance, defined by

$$\text{Var}[Z'(x_0) - Z(x_0)] = E[(Z'(x_0) - Z(x_0))^2] \quad (3.20)$$

The smaller the prediction variance, the closer $Z'(x_0)$ will be (on average) to the true value $Z(x_0)$. The geostatistical method of kriging deals with the Best Linear Unbiased Predictor or Estimator of $Z(x_0)$ with the smallest possible prediction variance. (B.L.U.E.)

The form of the best linear unbiased predictor will depend on the mean of $Z(x)$. For example, if $Z(x)$ has a constant mean and a pure nugget variogram [$\gamma(h) = s$ for all $h > 0$], the best linear unbiased predictor of $Z(x_0)$ will simply be the average of measured data

$$Z'(x_0) = \frac{1}{n} \sum_{i=1}^n Z(x_i) \quad (3.21)$$

Because the variogram is the same for all $h > 0$ and there is no trend in the data, there is no trend in the data, there is no reason to favour any of the measurements over any of the other measurements. Therefore the weights are all the same.

3.2.1.2 Simple Kriging

Simple Kriging (SK), deals with best linear unbiased prediction in the case when the mean of $Z(x)$ is fixed and known. Simple kriging is not widely used because in most application, the mean is not known and has to be estimated. The basic generalised linear regression algorithm and corresponding estimator for Simple kriging is;

$$[Z_{SK}(x) - m(x)] = \sum_{i=1}^n \lambda_i(x) [Z(x_i) - m(x_i)] \quad (3.22)$$

where $Z(x)$ is the RV model at the location x , the x_i 's are the n data locations, $m(x) = E[Z(x)]$ is the location dependent expected value of RV $Z(x)$, and $Z_{SK}(x)$ is the linear regression estimator. Also called the "simple kriging" (SK) estimator.

The SK algorithm requires prior knowledge of the $(n+1)$ means $m(x_i)$, $m(x_i), i=1, \dots, n$, and the $(n+1)$ by $(n+1)$ covariance matrix $[C(x_i, x_j), i, j=0, 1, \dots, n]$ with $x_0 = x$. In most practical situations, inference of these means and covariance values requires a prior hypothesis (rather a decision) of stationarity of random function $Z(x)$.

If the random function $Z(x)$ is stationary with constant mean m , and covariance function $C(h) = C(x, x+h)$, the SK estimator reduces to its stationary version:

$$Z_{SK}(x) = \sum_{i=1}^n \lambda_i(x) Z(x_i) + \left(1 - \sum_{i=1}^n \lambda_i(x)\right) \cdot m \quad (3.23)$$

With the traditional stationary SK system:

$$\sum_{j=1}^n \lambda_j(x) C(x_j - x_i) = C(x - x_i), i = 1, \dots, n \quad (3.24)$$

Stationary SK does not adapt to local trends since it assumes the mean value m is constant and known through the area. In some situations secondary information allows a prior determination of the locally varying means $m(x)$, the SK estimator can be used with the system assuming a stationary residual covariance. In the above formula, the covariance values $C(h)$ cannot be replaced by semivariogram values $\gamma(h) = C(0) - C(h)$ unless $\sum \lambda_j(x) = 1$, which is the ordinary kriging constraint.

3.2.1.3 Ordinary Kriging

Ordinary Kriging (OK) filters the mean from the SK estimator by requiring that the kriging weights sum to one. This results in the following ordinary kriging estimator:

$$Z_{OK}(x) = \sum_{i=1}^n \lambda_i^{OK}(x) Z(x_i) \quad (3.25)$$

And the stationary OK system:

$$\sum_{j=1}^n \lambda_j^{OK}(x) \cdot C(x_j - x_i) + \mu(x) = C(x - x_i), i = 1, \dots, n \quad (3.26)$$

Where the $\lambda_j^{OK}(x)$ are the OK weights and $\mu(x)$ is the Lagrange parameter associated with the constraint

$$\sum_{j=1}^n \lambda_j^{OK}(x) = 1 \quad (3.27)$$

It can be clearly seen that SK weights are different from the OK weights. Moreover it can be shown that ordinary kriging amounts to re-estimating at each new location x , the mean m as used in the SK expression. Since ordinary kriging is most often applied within moving search neighbourhoods. Thus the OK estimator is, in fact, a simple kriging type where the constant mean value is replaced by the location-dependent estimate $m(x)$:

$$Z_{OK}(x) = \sum_{i=1}^n \lambda_i^{OK}(x) Z(x_i) \quad (3.28)$$

By substituting

$$Z_{OK}(x) = \sum_{i=1}^n \lambda_i^{SK}(x) Z(x_i) + \left(1 - \sum_{i=1}^n \lambda_i^{SK}(x)\right) \cdot m(x) \quad (3.29)$$

Where λ_i^{OK} are the OK weights given by the system

$$\sum_{j=1}^n \lambda_j^{OK}(x) \cdot C(x_j - x_i) + \mu(x) = C(x - x_i), i = 1, \dots, n \quad (3.30)$$

and λ_i^{SK} is the SK weights given by the system

$$\sum_{j=1}^n \lambda_j(x) C(x_j - x_i) = C(x - x_i), i = 1, \dots, n \quad (3.31)$$

Hence ordinary kriging as applied within moving data neighbourhoods is already a nonstationary algorithm, in the sense that it corresponds to a nonstationary RF model with varying mean but stationary covariance. This ability to rescale locally the RF model $Z(x)$ to a different mean value $m(x)$ explains the extreme robustness of the OK algorithm.

3.2.1.4 Universal Kriging

Universal Kriging is an extension of ordinary kriging that, due to the fact that environmental data often contain drift, can be important in HTRW site investigations. Universal kriging addresses the case of a non-constant mean $\mu(x)$. Generally, the mean is assumed to have a functional dependence on spatial location of the form

Where the $f_i(u,v)$'s are known deterministic functions of $x=(u,v)$ (that is, these functions serve as independent variables) and the β_i 's are regression coefficients to be estimated from the data.

The form assumed for the mean is also generally used in standard linear regression analysis. In regression, ordinary least squares is generally used to solve for the coefficients;

when this done, it is assumed that the residuals are independent and identically distributed. Universal kriging is an extension of an ordinary least square regression that allows for spatially correlated results. Assuming that $Z(x)$ is a regionalized random variable with a mean and residual correlation function gives the best linear unbiased predictor by solving $n+p$ equations called the universal kriging equations :

$$\sum_{j=1}^n w_j \rho_{ij} + \frac{1}{s} \sum_{k=1}^p \lambda_k f_k(x_i) = \rho_{i0}, i = 1, \dots, n \quad (3.32)$$

$$\sum_{j=1}^n w_j f_k(x_j) = f_k(x_0), k = 1, \dots, p \quad (3.33)$$

Where in contrast to the ordinary kriging equations, there are now p coefficients $\lambda_1, \dots, \lambda_p$ resulting from unbiased condition on the predictor. The first term in the mean will usually be a constant, or intercept, for which $f_1(x)=1$. Therefore, the universal kriging model includes ordinary kriging as a special case. The universal kriging variance is given by

$$\sigma_k^2(x_0) = s \left[1 - \sum_{i=1}^n w_i \rho_{i0} \right] - \sum_{k=1}^p \lambda_k f_k(x_0) \quad (3.34)$$

These equations can be easily solved to obtain universal kriging predictors and kriging variances for any desired location. The estimated trend surface does not actually need to be computed to obtain the universal kriging predictor. If a particular application needs an estimate of the trend surface, then generalised least squares regression can be used to estimate the coefficients (β_j 's) in the regression equation.

3.2.1.5 Block Kriging

Up to this point, the problem of predicting the value of a regionalized random variable at a given location in the region over which the variable is defined has been considered. Implicit in this analysis is the assumption that the support of the variable being predicted is defined in exactly the same way as the variables that make up the measurements. However,

there may be applications where it is necessary to estimate the average of Z over an estimation block of much larger area than is represented by an individual sample. For example, an estimate of the average concentration of a contaminant over entire aquifer based point measurements at various locations might be needed. In other applications, an estimate of average concentration of soil contaminant in daily excavation volumes that are much larger than volume of an individual sample may be needed. Let Z_β be the average value of $Z(x)$ over a particular block B ,

$$Z_B = \frac{1}{m} \sum_{i=1}^n Z(x_{0i}) \quad (3.35)$$

where x_{0i} , $i=1, \dots, m$, denotes m predictions locations in block B . The object is to predict this average rather than the regionalized variable at a single location. In many applications, the location x_{0i} might correspond to nodes of a regular grid or finite element nodes in a groundwater model. Results of the block kriging are dependent on m and on the placement of the prediction locations. Selecting a large number of locations in block B , where each location has approximately the same representative area, is the best approach.

The objective of block kriging is to obtain the best linear unbiased predictor of Z and a Z_β and an estimate of the block kriging variance based on the measurements. The model for $Z(x)$ can be the constant-mean model assumed for ordinary kriging or the more general linear regression model assumed for universal kriging. In either case, the predicted value of Z_β coincides with the average of predicted values of the individual measurements in the block; that is;

$$Z'_B = \frac{1}{m} \sum_{i=1}^n Z'(x_{0i}) \quad (3.36)$$

In this equation, the individual predicted values are obtained from either the ordinary or universal kriging equations.

3.2.2 Radial Basis Functions

Radial functions constitute the basis for a nonlinear transformation of a given input vector,

$\mathbf{x}_i \in \mathbb{R}^d$ Here, $\mathbf{x}_i$ are the observed locations within the site. Their observed values are denoted by $f(\mathbf{x}_i)$. These functions are characterised by the fact that the relation between the value to be interpolated decreases monotonically with the distance from a central point (an observed data point).

The principle of a radial basis function derives from the theory of functional approximation. We consider a real valued function f , such that $f: \mathbb{R}^d \rightarrow \mathbb{R}$. Given the distinct set of points $\{\mathbf{x}_i \in \mathbb{R}^d: i=1,2,\dots,n\}$ and their functional values $\{f(\mathbf{x}_i): i=1,2,\dots,n\}$, the unknown function f is to be approximated by another real valued function $s: \mathbb{R}^d \rightarrow \mathbb{R}$. The radial basis function approximations are of the form

$$s(x) = \sum_{i=1}^n \lambda_i \phi(\|x - x_i\|) + p_m(x) \quad x \in \mathbb{R}^d, \lambda_i \in \mathbb{R} \quad (3.37)$$

where p_m is a polynomial of a low degree m , $\phi(\cdot): \mathbb{R}^+ \rightarrow \mathbb{R}$ is a function of the Euclidean distance between each sample data $\mathbf{x}_i$, and the given location $\mathbf{x}$ to be interpolated. Thus, the radial basis function s is a linear combination of translates of radially symmetric functions and a low degree polynomial. The coefficients λ_i are also unknown and have to be computed. If the space of all the polynomials of degree at most m in d variables is denoted by π_m^d , then the coefficients λ_i of the approximation $s(\cdot)$ and the coefficients of the polynomial $p_m(\mathbf{x})$ are determined by requiring that $s(\cdot)$ satisfy the following interpolation conditions,

$s(\mathbf{x}_i) = f(\mathbf{x}_i)$, for all $i=1,2, \dots, n$, and the side conditions,

$$\sum_{i=1}^n \lambda_i q(\mathbf{x}_i) = 0 \text{ for all } q \in \pi_m^d \quad (3.38)$$

where $q(\mathbf{x}_i)$ denotes a multivariate polynomial of degree m . These side conditions enforce that the data span the space $\mathbb{R}^d$, in order to constrain the polynomial $p_m(\mathbf{x})$.

Usually a function $\phi(\cdot)$ which has its maximum at a distance of zero is used. The popular choices of $\phi(\cdot)$ employed are linear, Gaussian, multiquadratic, and thin-plate splines functions (Powell 1992). (See Figure 3.4-3.7)

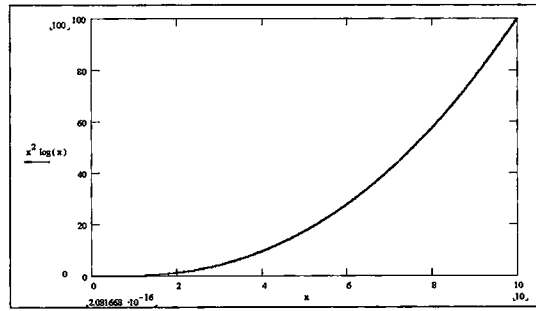


Figure 3.4 Linear $\phi(r)=r$ RBF Function

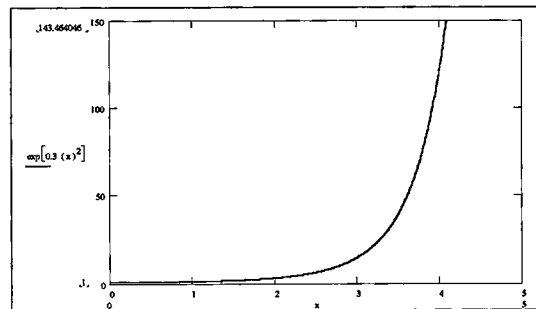


Figure 3.5 Thin-plate spline $\phi(r)=r^2 \log r$ RBF Function

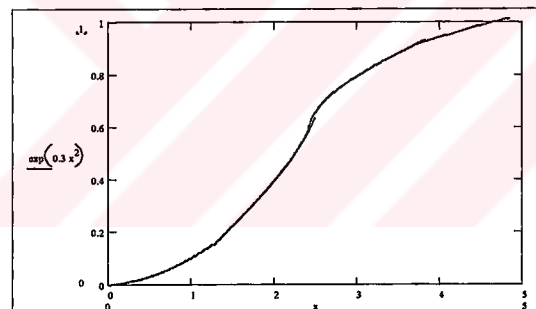


Figure 3.6 Gaussian $\phi(r)=\exp(-ar^2)$ RBF Function

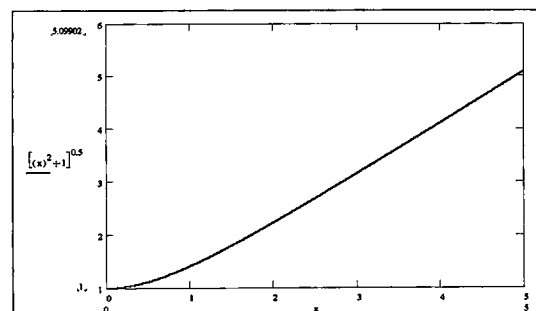


Figure 3.7 Multiquadratic $\phi(r)=(r^2+c^2)^{0.5}$ RBF Function

The choice of $\phi(\cdot)$ is determined by the dimension of the problem d , the interpolation conditions and the desired properties of the interpolant. Table 1 specifies conditions for various radial basis interpolants. In our context, we have $d=2$, and the non-collinearity condition of the samples is not satisfied and we therefore select the multiquadratic function in the assessment of hypothetical sites.

Table 3.1. Conditions imposed on data for various radial basis interpolants.

ϕ	Spatial Dimension d	Polynomial degree m	Restriction on data
Linear	Any	1	Data not coplanar
Thin-plate	2	1	Data not colinear
Gaussian	Any	-	None
Multiquadratic	Any	-	None

3.2.3 Minimum Curvature

Briggs (1974) has first proposed the minimum curvature method for automatic contouring of geophysical data. Wherever values of a function are available at discrete points only, additional information is required for determining the entire function. In case of a time series, eg. limitation of bandwidth constitutes the additional information. In case of a surface sampled at randomly scattered points the additional information “total curvature of surface to be minimum” has proven successful.

Grid interpolation algorithms derived from the minimum total curvature assumption work well where the average spacing of sample points is not too large as compared to the grid constant. However, within large voids between sample points they suffer from poor convergence. In a given grid point, local curvature is calculated using grid values “remote” grid points. In some cases additional information is available on the surface to be interpolated, most notably the main direction of geologic features. Such information may be taken into account in the curvature minimization process.

Briggs (1974) pointed out the optimum properties of the spline fit can be obtained by solving the differential equation equivalent to third order spline:

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) = 0 \quad (3.39)$$

Where u is the interpolating function.

Here, a two dimensional cubic spline is employed to fit the data by solving the corresponding difference equations with the objective of keeping the total squared curvature minimum. The interpolated surface generated by minimum curvature is analogous to a thin linearly elastic plate passing through each of the data values with a minimum amount of bending. Minimum curvature generates the smoothest possible surface while attempting to honour the data as close as possible, but it is not an exact interpolator.

The function $u(x,y)$ minimizes the total squared curvature;

$$C(u) = \iint \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right)^2 dx dy \quad (3.40)$$

And this equation obeys Equation (3.39), and conversely, if a function u obeys Equation (3.39) it minimizes $C(u)$. By expressing the discrete total squared curvature as,

$$C(u) = \sum_{i=1}^I \sum_{j=1}^J \frac{(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1})^2}{h^2} \quad (3.41)$$

Where h is the grid spacing in both the x and y directions and $u_{i,j}$ is the interpolated value at grid point (i,j) , we obtain the following system of linear equations from the necessary conditions to minimize C :

$$\begin{aligned} &u_{i+2,j} + u_{i,j+2} + u_{i-2,j} + u_{i,j-2} + 2(u_{i+1,j+1} + u_{i-1,j+1} + u_{i+1,j-1} + u_{i-1,j-1}) \\ &+ 8(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}) + 20u_{i,j} = 0, i = 1, \dots, I, j = 1, \dots, J \end{aligned} \quad (3.42)$$

For observation points coinciding with grid point. For observation points not on a grid point, an approximate expression for $\partial^2 u / \partial x^2 + \partial^2 u / \partial y^2$ at given point (x_0, y_0) is obtained by expanding u in Taylor series around (x_0, y_0) and using the Equation (3.41)

Another system of equation is then obtained from the necessary conditions to minimize C . Solving the above mentioned linear systems of equations leads to the interpolated values $u_{i,j}$, $i=1, \dots, I$, $j=1, \dots, J$

The resulting equations determine a set of relations between neighbouring locations and they are solved to identify the coefficients of the cubic polynomial. If one of the

$$g(x) = \frac{\sum_{i=1}^n w_i(x) q_i(x)}{\sum_{i=1}^n w_i(x)} \quad (3.43)$$

vertices does not coincide with an observation, then additional difference equations are used for the missing vertex.

3.2.4 Shepard's Inverse Distance Method

Inverse distance weighting is the simplest interpolation method. A neighbourhood about the interpolated point is identified and a weighted average is taken of the observation values within this neighbourhood. The weights are a decreasing function of distance. The user has control over the mathematical form of the weighting function, and the size of the neighbourhood (expressed as a radius or a number of points).

3.2.4.1 Weighting function

The simplest weighting function is inverse power:

$$w(d) = \frac{1}{d^p} \quad (3.44)$$

ELİF ÖKSEKİ
DOKÜMANTASYON MERKEZİ

With $p > 0$. The value of p is specified by the user. The most common choice is $p = 2$. For $p = 1$, the interpolated function is "cone-like" in the vicinity of the data points, where it is not differentiable.

The original "Shepard's method" presented in Shepard, (1968) is a variation on inverse power, with two different weighting functions using two separate neighbourhoods. The default weighting function for Shepard's method is an exponent of 2 in the inner neighbourhood and an exponent of 4 in the outer neighbourhood. The form of the outer function is modified to preserve continuity at the boundary of the two neighbourhoods. This method has been modified by Franke and Nielson (1980) and Renka (1988). The modified version also uses an inverse distance weighted least squares approach. Given potential function f with values f_i at the sampling points x_i , where the q_i are bivariate quadratic functions which locally approximate f at the sampling points

$$q_i(x) = f_i \quad (3.45)$$

in a weighted least squares sense. The relative weights are defined by the inverse distance functions

$$w_i = \left(\frac{(d_w - d_i)^2}{d_w d_i} \right) \text{ for } (d_w - d_i) = \begin{cases} d_w - d_i & \text{if } d_i < d_w \\ 0 & \text{if } d_w \geq d_i \end{cases} \quad (3.46)$$

where d_i denotes the Euclidean distance between x and x_i and d_w is a radius of influence around the point x . The data point f_i only influences interpolated values at points within this radius. It can be shown that, for all i and j , the weights

$$w'_i = \frac{w_i}{\sum_{j=1}^n w_j} \quad (3.47)$$

Satisfy,

$$w'(x_j) = \delta_{ij}$$

$$(3.48)$$

$$\sum_{i=1}^n w'_i(x) = 1 \quad (3.49)$$

And

$$\frac{\partial}{\partial x_1} w'_i(x_j) = \frac{\partial}{\partial x_2} w'_i(x_j) = 0 \quad (3.50)$$

for all i and j . Also these weights, and hence also interpolant g , have continuous first partial derivatives. The interpolatory of g follows directly from equations 3.45-3.48

$$g(x_j) = \sum_{i=1}^n w'_i(x_j) q_i(x_j) = q_j(x_j) = f_j \quad (3.51)$$

From equations 3.47 to 3.50

$$\frac{\partial}{\partial x_{1,2}} g_j(x_j) = \frac{\partial}{\partial x_{1,2}} q_j(x_j) \quad (3.52)$$

Thus, g maintains the local shape properties if the nodal functions q . If f is quadratic then $q_i(x)$ is $f(x)$ for all i and from equation 3.48 $g(x)=f(x)$. Thus g is also quadratic. We define the nodal function q_i as;

$$q_j(x) = c'_{j,1} (x_1 - x_{1,j})^2 + c'_{j,2} (x_1 - x_{1,j})(x_2 - x_{2,j}) + c'_{j,3} (x_2 - x_{2,j})^2 + c'_{j,4} (x_1 - x_{1,j}) + c'_{j,5} (x_2 - x_{2,j}) \quad (3.53)$$

Where the coefficients minimize

$$\sum_{\substack{i=1 \\ i \neq j}}^n w_i(x_j) (c'_{j,1} (x_{1,i} - x_{1,j})^2 + \dots + c'_{j,5} (x_2 - x_{2,j}) + f_j + f_i)^2 \quad (3.54)$$

For

$$w_i(x) = \left(\frac{(d_q - d_i)}{d_q d_i} \right)^2$$

where d_q is again a radius of influence about the node x_i . An implementation of method is given Renka(1988).

3.2.4.2 Neighbourhood size

The neighbourhood size determines how many points are included in the inverse distance weighting. The neighbourhood size can be specified in terms of its radius (in km), the number of points, or a combination of the two. If a radius is specified, the user also can specify an override in terms of a minimum and/or maximum number of points. Invoking the override option will expand or contract the circle as needed. If the user specifies the number of points, an override of a minimum and/or maximum radius can be included. It also is possible to specify an average radius based upon a specified number of points. Again, there is an override to expand or contract the neighbourhood to include a minimum and/or maximum number of points. In Shepard's method, there are two neighbourhoods; the inner neighbourhood is taken to be one-third the radius of the outer radius.

3.2.4.3 Anisotropy correction

In many instances, the observation points are not uniformly spaced about the interpolation points, with several in a particular direction and fewer in others. This situation produces a spatial bias of the estimate, as the clustered points carry an artificially large weight. The anisotropy corrector permits the weighted average to downweight clustered points that are providing redundant information. The user selects this option by setting the anisotropy factor to a positive value. A value of 1 produces its full effects, while a value of 0 produces no correction.

This correction factor is defined by computing the angle between every pair of observation points in the neighbourhood, relative to the observation point.

3.2.5 Triangulation

Triangulation play an important role in computer aided geometric design, finite element methods, numerical analysis, geology, meteorology, approximation theory and elsewhere. Triangulation is often used to obtain a partioned domain on which various functions are built, to be used to approximate or model surfaces.

Triangulation is a way of obtaining 3 dimensional interpolation of a surface, given the scattered data points $V=\{v_i\}$ where $v_i=(x_i, y_i)$, and heights $\{z_i\}$ for $i=1,\dots,n$. To find such an interpolant we partion the surface into triangles and by some method construct a piecewise function $f_j(v)$ on each triangle T_j which interpolates the surface values on the vertices $v_i=(x_i, y_i)$ for $i=1,\dots,n$. Then we set $f(v)=f_j(v)$ for $v\in T_j$. On each intersection of the triangles we want a certain C^r continuity to obtain the smooth surface f , i.e. f might be a polynomial or a spline.

In order to understand the theory of triangulation it is necessary to give the following definitions.

- *Convex Set*: A set is convex if, given two points in the set, the straight-line segment joining the two points is also contained in the set.
- *Convex Hull*: The convex hull of a set is the smallest convex set containing it, namely the intersection of all convex sets containing it.
- *A Triangle T*: A triangle T is the convex hull of any three noncollinear points from the set V .
- *A Triangulation Δ of V* : A set $\Delta=\{T_1,\dots,T_N\}$ of triangles T_j is called a triangulation of V provided that
 1. The points in v are vertices in Δ .
 2. Any pair of triangles T_i and T_j intersects at most at one common vertex or along a common edge.
 3. The union of the triangles is a connected set in R^2 , which means there is always a T^h of edges connecting any two vertices in the triangulation. The union of the triangles is equal to Ω , and it is the convex hull of V .

Δ is called the triangulation of the domain $\Omega = \bigcup_{j=1}^N T_j$, which implies that Ω contains no holes. The convex hull Ω of the vertices in V , is formally described by the convex combinations

$$\Omega = \left\{ w_i = \sum_{i=1}^n \lambda_i v_i \quad \text{such that} \quad \lambda_i \geq 0, \quad \sum_{i=1}^n \lambda_i = 1 \right\} \quad (3.56)$$

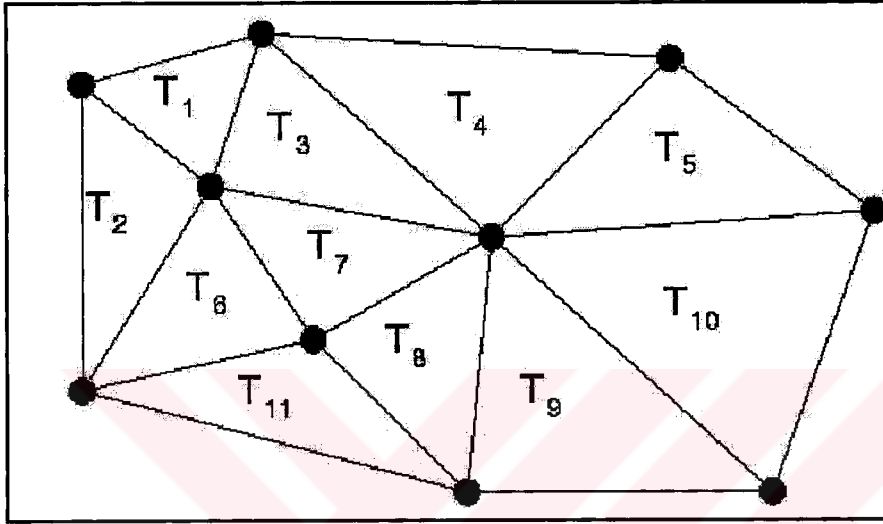


Figure 3.8 An example of a triangulation of points in $\mathbb{R}^2$.

There is no standard way of determining a triangulation. Given the set of points V , different approaches can be used depending on their application and how restrictive we want the definition to be.

3.2.5.1 Construction of a Triangulation

The construction of a triangulation depends on the given data and their application. For example, the data can be given as

1. Only a set of the vertices V
2. The set of vertices V and some edges to be included in the triangulation.
3. The boundary edges and an estimate on the number of inner vertices required.

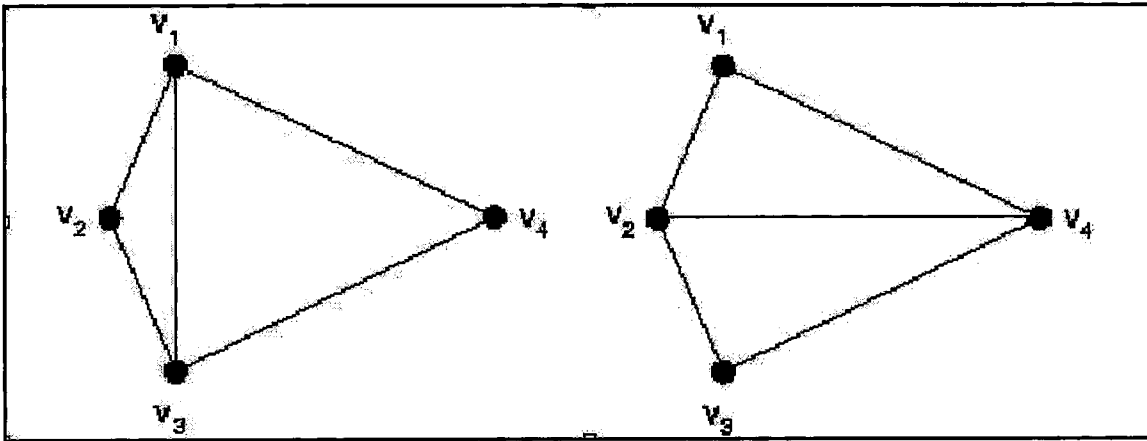


Figure 3.9 Two possible triangulations

In first case there is no information about the configuration of Δ . The natural thing to do will be to let Δ be the convex hull of V . In the second case we will typically be given all the boundary edges and a selection of inner edges. This often happens in the approximation of scattered data. The last case often appears in finite element methods where the problem of constructing Δ is referred to as grid generation because we are free to select the number and position of the vertices. This means we do not necessarily include all data points. We also often want to use different sizes of triangles in different parts of the domain, which will result in a gridded mesh.

3.2.5.2 Criteria for optimality

An important problem in triangulation is to generate the optimal triangles by maximizing the smallest angle in the triangulation.

The max-min angle criterion: Associated with a triangulation, Δ there is a number $\alpha(\Delta)$ which denotes the smallest angle between the edges in Δ . A triangulation Δ^* is said to be optimal with respect to the max-min angle criterion provided that $\alpha(\Delta^*) \geq \alpha(\Delta)$ for all possible triangulations Δ of Ω using the same vertices.

For this criterion, there may be more than one optimal triangulation. For example, for the two quadrilaterals shown in Figure 3.10, the two competing triangulations are equally good, yielding the neutral case. If the four vertices of a quadrilateral do not lie on a circle, then one of the triangulations is better than the other triangulations.

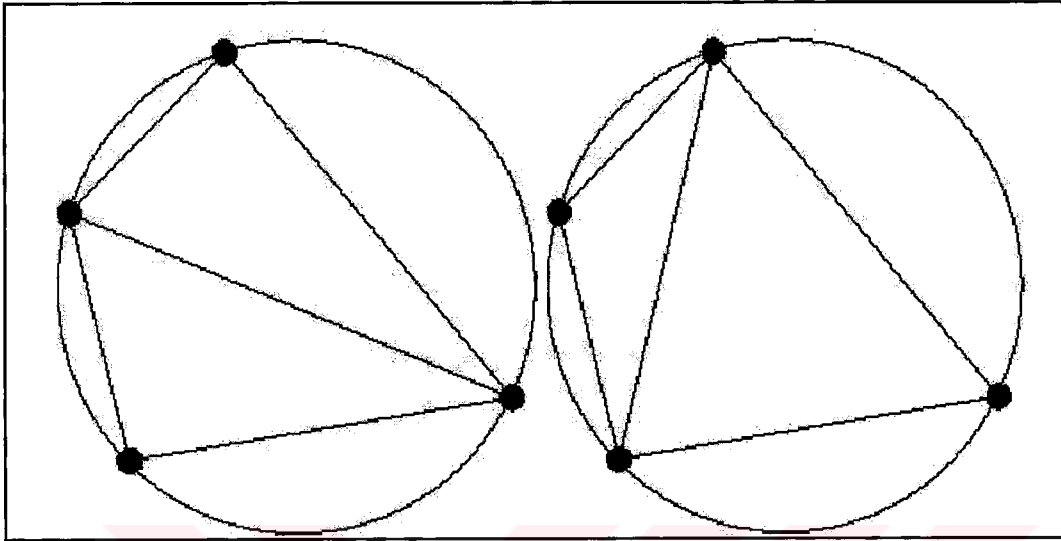


Figure 3.10: The two possible triangulations

Triangulations of four points where the common edge have been swapped. As the four points are on the same circumcircle, the triangulations are equally good and the optimal triangulation is not unique.

The classical error bounds for the error in approximating a smooth function over a triangle by a polynomial depend explicitly on the size of the smallest angle in the triangle.

The min-max angle criterion: Associated with a triangulation Δ there is number $\beta(\Delta)$ which denotes the largest angle between two edges in Δ . A triangulation Δ^* is said to be optimal with respect to the min-max angle criterion provided that $\beta(\Delta^*) \leq \beta(\Delta)$ for all possible triangulations Δ of Ω using the same vertices.

That is, we want to minimize the largest angle in the triangulation.

The max-min and the min-max criteria will often result in the same triangulation of a quadrilateral, but not always. Some other possibilities of triangulating a quadrilateral include:

1. The max-min radius criterion: Choose the triangulation, which maximizes the minimum radius of the circumcircles in the two triangles.
2. The min-max radius criterion: Choose the triangulation, which minimizes the maximum radius of the circumcircles in the two triangles.
3. The max-min area criterion: Choose the triangulation, which maximizes the minimum area of the two triangles, that is, the smallest of the areas of the two triangles.
4. The max-min height criterion: Choose the triangulation, which maximizes the minimum height of the two triangles.

All of the criteria discussed above are different from each other, but may lead to the same triangulation of four points in special cases.

3.2.5.3 Optimal triangulations of general point sets

The fact that for a given polygonal domain has many triangulation possibilities suggests that we might look for a best triangulation. In this section optimal triangulation criteria will be defined for a general point set in the plane. First we need the concept of local optimality:

Locally optimal triangulations: A triangulation is called locally optimal with respect to some criterion provided that every convex quadrilateral (consisting of two triangles sharing an interior edge) is optimal with respect to that criterion.

Theorem 1: Given any set of vertices V , there always exists at least one (globally optimal triangulation. Moreover, every optimal triangulation is locally optimal.

The converse of the second half of the theorem, namely that locally optimal triangulations are globally optimal, is only proved for the max-min angle criterion. It is impossible to design an algorithm for finding a best triangulation for the other criteria. Indeed, for these criteria, we can't even decide whether a given locally optimal triangulation is globally optimal or not without checking all possible triangulations.

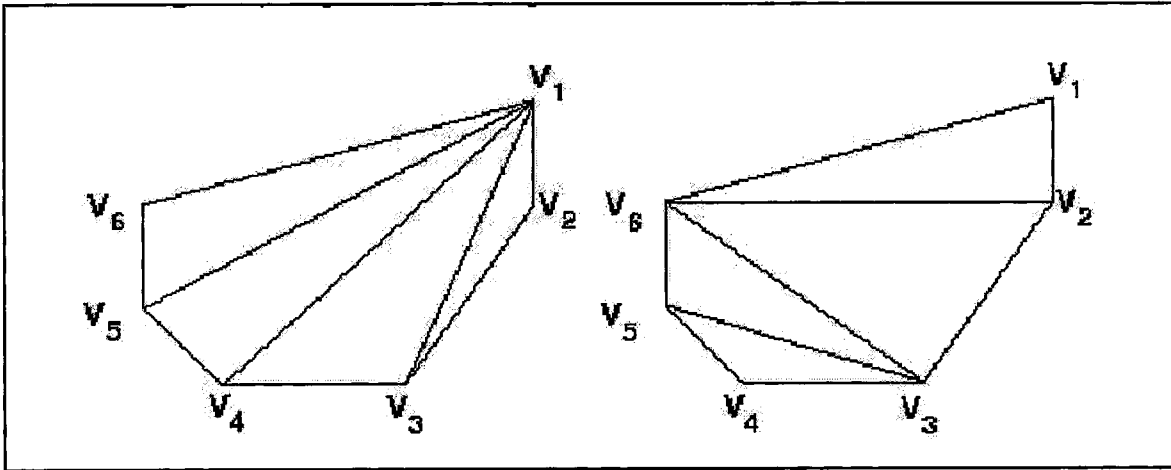


Figure 3.11 Two min-max angle locally optimal triangulators

3.2.5.4 Delaunay triangulation

Given a set of points $V = \{v_i, i=1,2,\dots,n\}$; voronoi tessellation is a collection of "tiles" S_i that cover the plane. The i th tile consists of the set of all points that are closer to the point v_i than to any other point v_j . We see that every vertex v_i has a counterpart S_i for $i = 1, \dots, n$. The interiors of the S_i 's are pairwise disjoint, and their union covers all of $\mathbb{R}^2$. In this sense the set $\Sigma = \{S_i\}_{i=1}^n$ forms a tiling or tessellation of the plane. The Voronoi tessellation is also referred to as a Dirichlet or Thiessen tessellation. Formally, we can describe a Voronoi diagram as in this definition

$$S_i = \{x \in \mathbb{R}^2 : \|x - v_i\| \leq \|x - v_j\|, \quad j = 1, \dots, n, \quad j \neq i\} \quad (3.57)$$

Voronoi diagram: Given a set V of vertices, and voronoi tiles S_i are given by for every $i=1, \dots, n$. The union of all tiles is called a Voronoi diagram.

Delaunay triangulations arise as the duals of Voronoi tessellations of point sets in $\mathbb{R}^2$, and they are, by definition, triangulations of convex domains in the plane.

Delaunay triangulation : If the set of tiles $\Sigma = \bigcup_{i=1}^n \{S_1, \dots, S_n\}$ denotes a Voronoi diagram of the set V , then $[v_i, v_j]$ is an edge in the Delaunay triangulation if and only if $S_i \cap S_j \neq \emptyset$,

that is, if the corresponding tiles $\{S_i, S_j\}$ of the vertices $\{v_i, v_j\}$ share a common vertex or a common edge. See Figure 3.12 for a geometrical interpretation.

There are several ways of constructing a Delaunay triangulation. For example, a max-min angle optimal triangulation is a Delaunay Triangulation. However, for a geometrical interpretation, the following theorem might be more useful;

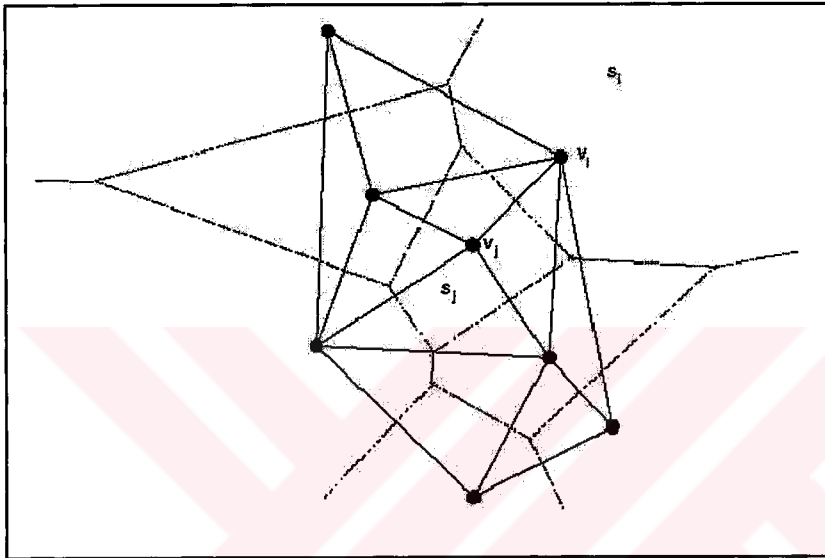


Figure 3.12 Delaunay triangulation and Voronoi tessellation of a point set V

Considerable effort has gone into the design of algorithms for the construction of a Delaunay triangulation. There are essentially three approaches:

1. Swapping edges: Construct some initial triangulation. Then iteratively go through the list of convex quadrilaterals and make swaps where necessary to assure local optimality.
2. Adding points: Start with one triangle and add one point at a time, making sure that at each step the current triangulation is optimal.
3. Divide and conquer: Recursively divide the data up into pieces, find locally optimal triangulations for each piece, then merge the resulting triangulations.

Consider Figure 3.13 for an example;

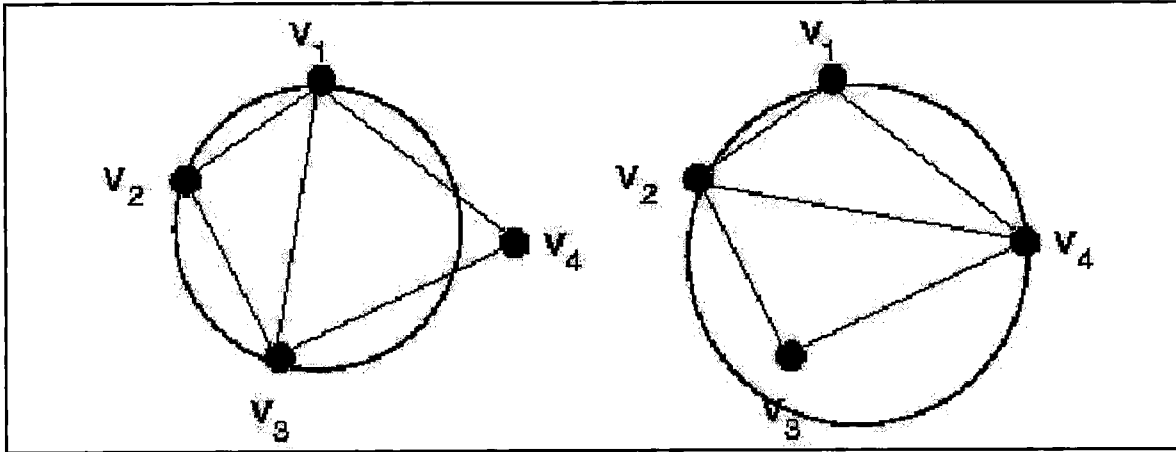
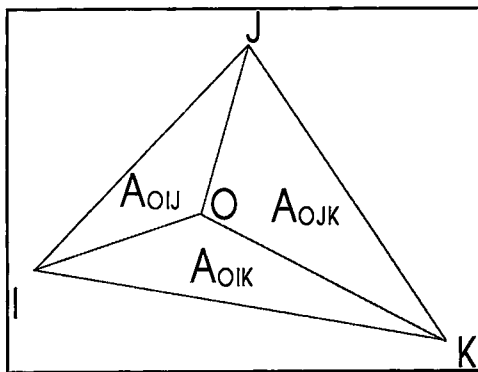


Figure 3.13 A Delaunay triangulation vs. a non-Delaunay triangulation

As a conclusion we can say that a good algorithm will always end in a globally optimal triangulation, that is, a Delaunay triangulation. For local changes we should always apply the max-min angle criterion which is the only proved local method that ensures a globally optimal triangulation.

The following is an example of showing how three nearest data can be weighted by triangular areas to point estimate. The data are located at the corners of the triangle. The data value at I is weighted by the triangle area A_{OJK} , at J by area A_{OIK} , and K by A_{OIJ} . And the corresponding values for these locations are z_i, z_j, z_k and z_o .



$$z_o = \frac{A_{OJK}z_I + A_{OIK}z_J + A_{OIJ}z_K}{A_{IJK}} \quad (3.58)$$

Figure 3.14 A typical representation of triangulation algorithm

4 A FUZZY AREAL SITE ASSESSMENT APPROACH

The fuzzy areal site assessment approach (FASA) is based on the Fuzzy Adaptive Partitioning Algorithm originally developed for identifying the global optimum solution in highly multi-modal mathematical functions (Demirhan et al. 1998, Özdamar and Demirhan 1998). FASA iteratively partitions the site into non-overlapping regular sub-regions ($\alpha_i(t)$), and evaluate their potentials, ($r_i(t)$), to contain hot-spots via the set of samples ($A_i(t)$) of size n collected from each sub-region. Here, t indicates the partitioning iteration. The sub-region with the highest contamination potential ($\alpha_s(t)$), is selected to be re-partitioned into k subregions ($\alpha_i(t+1)$, $i=1\dots k$) from which new samples are collected (the sampling is now more intensive in these regions because their areas are reduced). New samples are also collected from the existing unpartitioned subregions. The unpartitioned subregions and the re-partitioned daughter subregions make up the set of existing subregions at iteration t , $C(t)$.

4.1 The Partitioning Algorithm

After a sample of size n is collected within the site according to any given sampling pattern, FASA is applied to indicate boundaries of the sub-areas in the site, which are most likely to contain contaminated spots. This is achieved by partitioning the site into smaller and smaller sub-regions where the set of observations falling within these small areas indicate "as a group" the possibility of areal contamination. Obviously, the re-partitioning should be stopped whenever a sub-region is under a given tolerance area size or when very few data remain in the area. The stopping criteria are tied to user-defined parameters. At the extreme case, the user may end up with single observation areas, but the latter is not desirable due to the fact that using a potential estimator, who is an averaging fuzzy measure, carries out the assessment of each sub-region.

The eventually reached smaller areas whose boundaries are indicated by FASA are either reclaimed or re-investigated with new samples. Since FASA does not generate interpolated

isolines, it does not claim any concentration levels for these indicated areas, but calls for re-investigation.

Initially, FASA divides the site to be investigated (denoted as $\alpha_0(0)$) into k non-overlapping sub-regions. In the notation $\alpha_i(t)$, i is the partition index and t is the iteration index. The index t is initially set to zero. In the next iteration ($t=1$), and, in each subsequent iteration t , FASA identifies the sub-region, $\alpha_s(t)$, among existing sub-regions, which has the highest potential of containing contaminated areas and re-partitions $\alpha_s(t)$ into k non-overlapping smaller sub-regions, $\alpha_i(t+1)$, $i=1..k$, where $\cup_i$

$$\alpha_i(t+1) = \alpha_s(t), \text{ and } \alpha_i(t+1) \cap \alpha_j(t+1) = \emptyset. \quad (4.1)$$

The potential $r_i(t)$ of a sub-region $\alpha_i(t)$ to contain hot-spots is calculated by using the observation values of the sample set $A_i(t)$ falling within $\alpha_i(t)$. While selecting $\alpha_s(t)$, all existing sub-regions including the ones which have not been partitioned in the previous iteration, $t-1$, are considered. Thus, all the accumulated sub-regions up to and including iteration t make up the set of existing sub-regions denoted by $C(t)$.

This areal partitioning scheme leads to a natural dynamic grouping of observations into specific sets (lying within given corresponding sub-regions) which reflect contamination levels as a group. The finally indicated sub-regions are prone to further investigation where a more intensified re-sampling should be carried out in order to refine areal boundaries for remediation.

In order to reduce SCRAP, sub-regions whose potential values do not hold any implications of contamination are discarded. At each iteration, there are three ways to dispose of sub-regions: i) the sub-region may not contain the threshold number of samples, tresamp , defined by the user, i.e., the number of samples $n_i(t)$ falling within the area, is below tresamp , and therefore, its boundaries are stored for further investigation later on, i.e. $\alpha_i(t)$ is added to the set of stored areas, ST ; ii) the area of the sub-region may be smaller

than a tolerance value, areatol , specified by the user; i.e., $\text{size}(\alpha_i(t)) < \text{areatol}$; in this case the sub-region is also stored in the set ST , because the area is so small that either the site investigator is ready to remediate the area even if it is not really contaminated or the investigator will investigate it further in the second phase; iii) the potential to contain hot-spots is so low, $(r_i(t) < p^* r_{\max}(t))$, that it will no longer be considered as a potentially contaminated area, so this area is chopped off. Here, $r_{\max}(t)$ is the maximum $r_i(t)$ over all existing sub-regions, and p^* is a user-specified constant between 0 and 1.

Thus, if FASA is used in a multi-phase site investigation, once the first analysis ends with no sub-regions to assess (either they have been put aside for further investigation, or, they have been discarded), the investigator may use FASA once more after collecting new samples from each of the sub-regions stored in the set ST in the previous phase. The number of new samples n'_i collected from each stored sub-region depends on its size as well as the pre-specified number of samples to be collected in this phase.

The pseudocode of FASA is given briefly in Figure 4.1. Finally, we re-emphasize that unlike the previously mentioned interpolation methods, FASA has no claim on the concentration value surface within the sub-regions indicated. FASA simply cautions the site investigators about the areas indicated by the boundaries and reports their potential to contain hot-spots on a $[0-1]$ scale.

4.2 The Fuzzy Potential Estimator

In order to discuss the nature of the fuzzy potential estimator, $r_i(t)$, first we should analyze how we normalize the contamination level indicated by each observation through a membership function.

Membership function. A membership function, μ_{ij} , maps the concentration value $f(x_{ij})$ to the unit interval. Here, x_{ij} is the j th observation in $A_i(t)$. Membership values represent the scaled position of $f(x_{ij})$ against the maximum concentration f^* obtained within the initial set of samples $A_0(0)$. μ_{ij} indicates the grade that observation x_{ij} belongs to the set of points

lying in the hot-spot areas. Therefore, μ_{ij} is zero if $f(x_{ij})$ is below the remediation trigger level of the contaminant.

The membership function used here is the gaussian membership function (GMF) (Ross, 1995) which results in an exponential bias for observation values near f^* . GMF is given below.

$$\mu_{ij}(t) = \exp(-[f^* - f(x_{ij})]^2 / \sigma^2) \quad (4.2)$$

where σ^2 is the variance of all observation values. In Figure 4.2, we provide membership values for a data set consisting of (21.5, 29.1, 33.9, 34.8, 42.3) where $f^* = 42.3$ and the variance of the data values in the set is 58.92.

Pseudocode of FASA

- Step 0. Initialization: Identify the initial site α_0 . Set $\alpha_s(0) = \alpha_0$, $C(0) = \{\alpha_0\}$, and $ST = \emptyset$, $t = 1$.
- Step 1. Partitioning: Partition the subregion $\alpha_s(t-1)$ into k subregions to form the set C .
Let $C(t) = [C(t-1) \setminus \alpha_s(t-1)] \cup C$.
- Step 2. Evidence collection: Identify the samples falling into $\alpha_i(t)$ to form $A_i(t) \in A_0(0)$ from each sub-region
 $\alpha_i(t) \in C(t)$. ($A_0(0)$ consists of the n samples collected initially.)
- Step 3. Evaluation: Compute from each $A_i(t)$, the potential $r_i(t)$ of sub-region $\alpha_i(t)$ to contain hot-spots. Calculate $r_{\max}(t)$.
- Step 4. Storage & Disposal: For all $\alpha_i(t) \in C(t)$, check:
- If $size(\alpha_i(t)) < areatol$, then $ST = ST + \alpha_i(t)$, and $C(t) = C(t) - \alpha_i(t)$.
 - If $n_i(t) < tresamp$, then $ST = ST + \alpha_i(t)$, and $C(t) = C(t) - \alpha_i(t)$.
 - If $r_i(t) < p^* r_{\max}(t)$, $C(t) = C(t) - \alpha_i(t)$.
- Step 5. Selection: If $C(t) \neq \emptyset$, $\alpha_s(t) = \alpha_i(t) : r_i(t) = r_{\max}(t)$. Set $t = t + 1$ and go to step 1.
- Else, if single-phase investigation, then stop.
Else, if multi-phase investigation, then go to Step 6.
- Step 6. Set $t = 1$. $C(1) \leftarrow ST$. $ST = \emptyset$. Collect a set $A_i(1)$ of n'_i new observations from each $\alpha_i(1) \in C(1)$.
 $A_0(1) = \cup A_i(1)$ Go to Step 3.
- i

Figure 4.1. Pseudocode of FASA

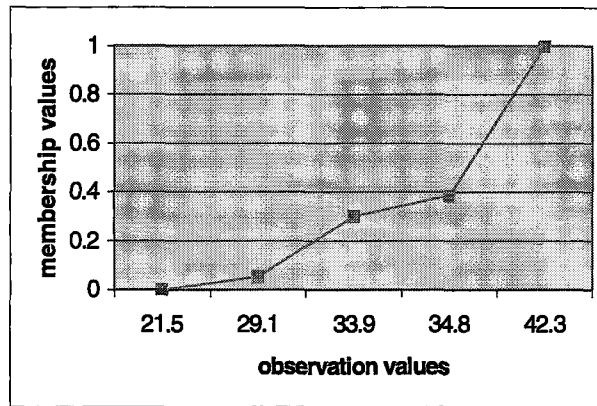


Figure 4.2 Representation GMF for a given set of data

The assessment of each sub-region $\alpha_i(t)$ is achieved through a fuzzy potential estimator $r_i(t)$ which is based on the membership values of samples lying in that particular region. The concentration values (observed and unobserved) within a given area constitute a fuzzy set, since it is not possible to know exactly the concentration level of each and every areal unit within the sub-region. Similarly, the boundaries of the hot-spot areas are also fuzzy, because not all concentration values can be identified within a hot-spot and one may not know where the area ceases to be a hot-spot. Hence, the hot-spot areas represent fuzzy sets.

The fuzzy potential estimator used in this study is derived from an entropy measure described in Pal and Bezdek (1994). An entropy measure is a function of membership values and it reflects the average ambiguity in a given fuzzy set. In the entropy measure, the uncertainties related to both the set itself and to its complement are included. Thus, in our context, an entropy measure is factored into two parts: The part reflecting the presence of hot spots in an area (the boundaries of the hot spots are fuzzy and therefore, ambiguity is involved on the issue of whether the location of an observed value within a sub-region is a member of the hot-spot area) and the part reflecting its absence. A fuzzy set's uncertainty level is minimum when the membership values (the components of the entropy measure) approach either to zero or to one. The uncertainty level reaches the maximum when the membership values approach the value of 0.5. Here, we omit the part of the entropy measure which reflects the uncertainty related to the complement of the set, because we are only interested in the presence of high contaminant concentrations rather than low levels.

The multiplicative potential estimator which ranges between [0-1] is given below.

$$r_i(t) = 1 - \left[\sum_{j=1}^{n_i(t)} \mu_{ij}(t) (1 - \mu_{ij}(t)) \right] / n_i(t) \quad (4.3)$$

As observed in the mathematical expression for $r_i(t)$, the effect of an observation on $r_i(t)$ is weighted by $1/n_i(t)$. Hence, since the number of observations lying in sub-regions of different sizes vary, the weight of the same observation becomes adaptive and dependent on the specific group of observations of which it becomes a member. For instance, an observation's weight is usually lower in a parent sub-region as compared to its child within which it currently lies. The reason is that, in general (if uniform sampling is used), there are more samples in the parent sub-region than in the child. Hence, samples in areas which are potentially more contaminated than others (re-partitioned sub-regions) are assigned higher weights while the samples in the remaining unpartitioned sub-regions preserve their previous weights. Thus, the adaptive area-partitioning scheme of FASA results in an adaptive observation-weighting scheme.

4.3 An Example Illustrating the Mechanics of FASA

The following example demonstrates how FASA identifies hot spots. Figure 4.3 illustrates a hypothetical contaminant distribution over a site whose boundaries are given as $(1,5) \times (2,6)$. Three hot spots exist in the site and they are indicated by lighter tones. The threshold isocline for intolerable contamination level is the second inner isocline. In Table 2, the partitioning iterations of FASA and the resulting partitioned zones are indicated with their respective potential values. Note that if a tie between the width and length of a sub-area exists, then the sub-area is partitioned by randomly selecting the x- or y- dimension. Here, p^* is equal to 0.75 and many partitions are discarded according to p^* . The parameter areatol is assumed to be 10% of the site area. The partitions which are stored due to the areatol criterion are also indicated in Table 2.

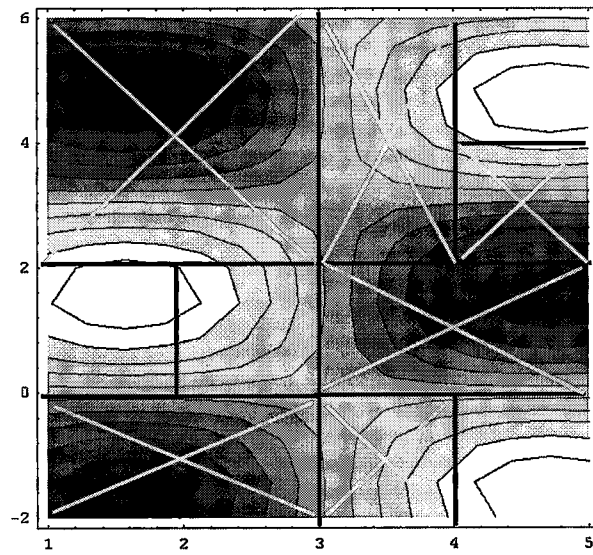


Figure 4.3 An exemplary contaminated site

Table 4.1. The iterations of FASA.

Partitioning Iteration (t)	Partition Index (i)	Partition Bounds (x) x (y)	$r_i(t)$	Decision on Partitions
0	1	(1,3) x (2,6)	0.91	Preserved
	2	(3,5) x (2,6)	0.96	Selected
1	1	(1,3) x (2,6)	0.91	Preserved
	2	(3,5) x (4,6)	0.97	Selected
	3	(3,5) x (2,4)	0.93	Preserved
2	1	(1,3) x (2,6)	0.91	Preserved
	2	(3,4) x (4,6)	0.45	Discarded
	3	(4,5) x (4,6)	0.98	Selected
	4	(3,5) x (2,4)	0.93	Preserved
3	1	(1,3) x (2,6)	0.91	Preserved
	2	(4,5) x (5,6)	0.99	Stored (areatol criterion)
	3	(4,5) x (4,5)	0.33	Discarded
	4	(3,5) x (2,4)	0.93	Selected
4	1	(1,3) x (2,6)	0.91	Preserved
	2	(3,5) x (3,4)	0.20	Discarded
	3	(3,5) x (2,3)	0.95	Selected
5	1	(1,3) x (2,6)	0.91	Selected
	2	(3,4) x (2,3)	0.72	Discarded
	3	(4,5) x (2,3)	0.98	Stored (areatol criterion)
6	1	(1,3) x (4,6)	0.64	Discarded
	2	(1,3) x (2,4)	0.87	Selected
7	1	(1,3) x (3,4)	0.94	Selected
	2	(1,3) x (2,3)	0.47	Discarded
8	1	(1,2) x (3,4)	0.99	Stored (areatol criterion)
	2	(2,3) x (3,4)	0.97	Stored (areatol criterion)

5 PERFORMANCE EVALUATION OF SIM AND FASA

5.1 Sampling Patterns Used in the Experiments

In the first part of the experiments three different sampling patterns are used. The first one which was stated by Ferguson (1992) is the Herringbone pattern, the other two patterns are more basic patterns and widely used in geostatistics and known as regular grid, and stratified random sampling. These three patterns are selected due to the suggestions made on various sampling strategies mentioned in the report by the Department of the Environment (U.K. DoE Contaminated Land Research Report No. 4, 1994) which is a brief guide for contaminated land research basics, methodologies and concept. It is shown that in order to have a certain probability for at least one sample point to hit the hot-spot (contaminated part of the area), the investigators should collect a given number of samples from the potential contaminated area. It is claimed that environmental sampling strategies should satisfy the following conditions.

1. It should be stratified: (that is, the area to be sampled should be partitioned into regular sub-areas;
2. Each sub-area should carry only one sampling point;
3. It should be systematic;
4. Sampling points should not be aligned.

The Herringbone pattern, which is generated by offsetting points by one quarter of a unit grid by starting from the left-bottom of the area, satisfies all four conditions stated above. A family of Herringbone designs can be generated by offsetting points by other fractions. However, according to U.K. DoE Contaminated Land Research Report No. 4, 1994, computer experiments showed that the one-quarter unit offset provides the most efficient search pattern. This type of herringbone pattern can be generated by four square grids each of the same orientation and spacing. Thus it is almost as easy to set out on site as a single square grid.

The regular grid pattern, which is generated by dividing the area into equal sub regions and put the sample locations to the corners of these areas, satisfies all conditions except the

fourth one. The disadvantage of the regular grid is its much reduced ability to detect elongate hot-spots whose long axis is parallel to the grid axis.

Stratified random pattern, which is generated by using a pure random number generator for both x and y coordinates in each strata of the sample location pattern satisfying all conditions except the third one. The weakness of the stratified random sampling pattern is tendency for irregular sampling.

In Figure 5.1 the three sampling patterns are depicted. The source codes for these sampling patterns are in Appendix D.

The performance of sampling strategies are evaluated by their frequency of hitting a point in the contaminated area. Namely, if 100 partially contaminated sites are investigated, a pattern, which most frequently results in at least one sample hit a hot spot, is considered to be the most efficient. The results provided by Ferguson (1992) indicate that the Herringbone pattern is the most efficient one among the strategies investigated within an experiment where various geometric shapes (square, elliptical, circular, rectangular with several aspect ratios) representing hot-spots are considered.

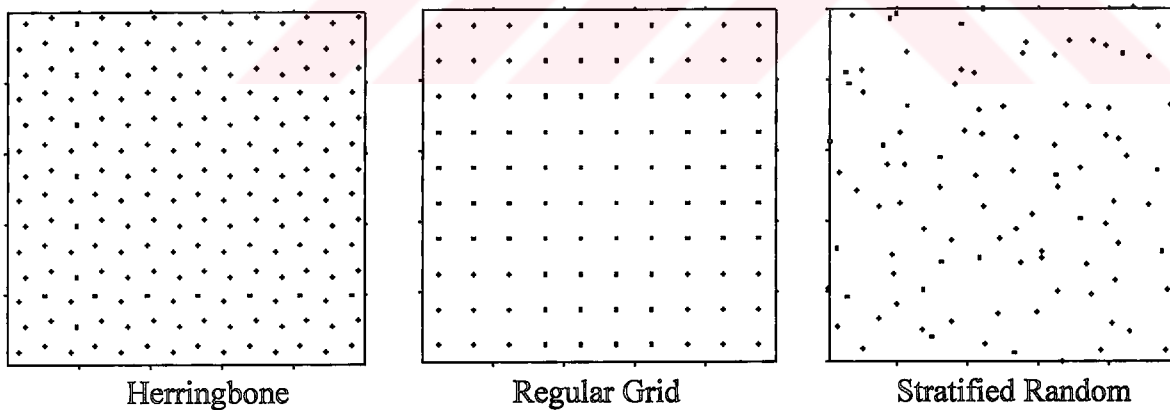


Figure 5.1 Three sampling patterns

5.2 Design of Hypothetical Test Sites

In order to test the SIM and FASA, 360 hypothetical sites have been generated by the computer according to an experimental design considering the following factors:

- The Percentage of Contaminated Area (area covered by hot-spots) over the area of the site (factor PCA). Factor PCA has two levels: the contaminated area constitutes 5% and 1% of total site area,
- The Number of Hot-Spots per Site (NHS). Factor NHS also assumes two levels of contamination: a single hot-spot and three non-overlapping hot-spots exist.
- The Sampling Pattern (SAMP). As described in the previous section SAMP has three types: Herringbone (Ferguson, 1992), regular grid, and stratified random sampling.

These three major factors are the most significant part of the evaluation procedure. By using the above mentioned factors and the formula given below 360 site is synthetically generated;

$$3 \text{ (SAMP)} \times 2 \text{ (PCA)} \times 2 \text{ (NHS)} \times 30 \text{ (Replicates)} = 360 \text{ Hypothetical Sites} \quad (5.1)$$

In Table 5.1 and 5.2 the number of times that at least one sample falls on a hot-spot among 1000 hypothetical sites (indicated by # of Success) and the average number of samples falling within a hot-spot (indicated by Ave. # of Samples) are reported for different sample sizes, for two PCA and three NHS levels, and three different sampling patterns. The area of each site is 10,000 squared meters. In Table 5.3, we observe that the regular grid and the Herringbone patterns provide the best results. When PCA equals 5%, a sample size of 100 provides satisfactory results regardless of NHS. Similarly, a sample size of 200 seems to work well when PCA is 1%.

The synthetic hot-spot generation scheme within a site generated as follows. First, it is randomly decided on how many (between 5 and 10) equivalent square grids, gc, each hot-spot area will lie. Then, each edge of the site is divided into G equivalent intervals where,

$$G = \text{edge length} / [\text{site area} * \text{PCA} / (\text{gc} * \text{NHS})]^{0.5} \quad (5.2)$$

Table 5.1 Performance of sampling patterns on 1000 hypothetical sites for PCA=1%

PCA=1%						
NHS=1						
Sample Size=50			Sample Size=100		Sample Size=200	
	#Success	Ave. # of Samples	#Success	Ave. # of Samples	#Success	Ave. # of Samples
Herring Bone	488	1.06	798	1.28	974	2.01
Stratified Sam.	420	1.12	738	1.36	939	2.11
Grid	504	1.10	813	1.27	982	2.08
NHS=2						
Sample Size=50			Sample Size=100		Sample Size=200	
	#Success	Ave. # of Samples	#Success	Ave. # of Samples	#Success	Ave. # of Samples
Herring Bone	441	1.16	796	2.06	961	2.17
Stratified Sam.	418	1.19	671	1.46	916	2.15
Grid	472	1.17	779	2.09	965	2.20
NHS=4						
Sample Size=50			Sample Size=100		Sample Size=200	
	#Success	Ave. # of Samples	#Success	Ave. # of Samples	#Success	Ave. # of Samples
Herring Bone	443	1.24	719	1.71	914	2.19
Stratified Sam.	399	1.21	676	1.53	893	2.17
Grid	484	1.27	779	2.20	933	2.25

Table 5.2 Performance of sampling patterns on 1000 hypothetical sites for PCA=5%

PCA=5%				
NHS=1				
Sample Size=25			Sample Size=100	
	#Success	Ave. # of Samples	#Success	Ave. # of Samples
Herring Bone	902	1.75	1000	4.75
Stratified Sam.	827	1.55	1000	5.08
Grid	918	2.36	1000	4.95
NHS=2				
Sample Size=50			Sample Size=100	
	#Success	Ave. # of Samples	#Success	Ave. # of Samples
Herring Bone	989	2.52	1000	4.83
Stratified Sam.	955	2.60	998	4.95
Grid	984	2.53	1000	4.88
NHS=4				
Sample Size=50			Sample Size=100	
	#Success	Ave. # of Samples	#Success	Ave. # of Samples
Herring Bone	973	2.66	1000	6.37
Stratified Sam.	948	2.62	998	5.05
Grid	975	2.79	1000	8.06

Next, one starting grid is randomly selected from G^2 grids for the constructing of the hot-spot area. One can continue to construct the hot spot with eight possible neighbour grids, 4 on each side of the starting grid and 4 on its vertices. The neighbour is randomly selected.

This neighbour selection process goes on until the number of selected grids becomes g_c and one hot-spot area is completely constructed. The procedure goes on to construct the various levels of hot-spots with the condition that none of the grids belonging to a given hot-spot has a common edge or vertex within the existing hot-spots.

As an example, by using the following numbers $PCA=1\%$, $NHS=2$ and g_c as 8, in Equation 5.2; G is found as; $G=100/[(10000*0.01/16)]^{0.5}=40$ and a total of $40 \times 40=1600$ grids will cover the site, each with an area 6.25 square meters and edge of 2.5 meters. When 200 samples are taken according to the regular grid pattern, the distance between a pair of observations in the x- and in the y-directions becomes 7.07 meters which is much wider than an edge of a grid. In order to create a realistic pollution distribution, each grid among the g_c grids of a hot-spot is randomly assigned one of the two following concentration ranges: moderately contaminated range [150-600), and highly contaminated range [600-1000]. The reason behind this is the desire to construct a hot-spot area consisting of a non-homogeneous contamination pattern. Clean areas are assigned a concentration range of [0-150). These ranges are selected from the critical concentration levels of a heavy metal (Pb) reported to be found in contaminated and non-contaminated areas (Alloway, 1995).

The observations whose locations are determined by a sampling pattern are assigned random concentration levels in the ranges specified by their locations, e.g., a sample falling into a clean area is assigned a random value uniformly distributed in the range [0-150).

5.3 Performance Criteria in Environmental Site Characterization

The relevant performance criteria in environmental site characterization is to determine the contaminated regions in the site as precisely as possible (e.g., identify the areas which contain high contamination levels completely) and to carry out this task so that it results in minimum possible remediation costs. The latter implies that contaminated areas need to be identified without leading to over-estimated zones. Obviously, these two objectives are conflicting. The best way to make sure all contaminated zones are remediated, is to remediate the whole site. However, such a policy results in the highest remediation costs of transporting tons of soil (some of which are already clean and this operation is partially

futile) and replacing them by clean re-fills. Yet, the inability to cover the contaminated zones completely leads to high health hazards. Consequently, the ultimate goal of site characterization is expressed in the following statement. "Maximize the percentage of correctly recognized contaminated areas with minimum areal cover". In other words, we wish to identify 100% of the contaminated zones while having 0% of clean area mis-identified as contaminated (false positives). From now on, we call false positives "SCRAP" and calculate it by dividing the sum of mis-identified areas by the area of the whole site. The percentage of areas (with respect to all contaminated area) recognized correctly as contaminated is called "COVER".

5.4 FASA and SIM Implemented on a Hypothetical Site

A hypothetical site is depicted in Figure 5.2 with Herringbone sampling pattern where slightly lighter colour tone indicates moderate contamination and darker tone indicates high contamination. PCA is 5% in this site. It is observed that 5 samples fall within the hot-spot. In Figure 5.3 to 5.12, the areas indicated as contaminated by various SIM and FASA are shown. The COVER and SCRAP percentages of different methods are provided in Table 5.3 Furthermore, since SIM provide isolines with given concentrations, we also provide the percentages of moderately and highly contaminated areas identified correctly by SIM. (In Figure 5.2 two isolines with values of "greater than 150" and "greater than 600" are indicated. The smaller circled isoline represents high contamination.) No such information is made available by FASA, since the method has no claim on the specific concentrations within the interior of the areas it has identified and it simply claims that these areas are prone to contain contamination levels above the threshold value. Note that the COVER percentage is calculated without regard to isolines, i.e., the intersection of the areas identified as contaminated (independent of moderate or high) by a SIM with the actual hot-spot is reported as COVER. In fact, SIM often incorrectly label a highly contaminated area as moderately contaminated, but rarely vice versa.

Various interpolation basis functions and different variogram models which are given in Chapter 3.2, are tested and their performance is indicated in Table 5.3 the purpose here is to provide a visual insight for each method's performance. For instance, kriging with gaussian variogram model (Figure 5.8) claims small contaminated areas all over the site

whereas other variogram models result in very small SCRAP percentages. The performance of kriging as compared with FASA is partly due to the well-known effect of hot-spots situated in the corners of a given site. Also kriging tends to smooth sharp data points and therefore, the outer boundary of the hot-spot is indicated as relatively clean.

Prior to discussing the performance of the remaining SIM in Table 5.3, let us recall that all gridding/interpolation methods have some common problems such as sensitivity to grid size, partial ignorance of observations, and sensitivity to minor changes in the data which usually result in major changes in the outcome (Wingle 1992). Furthermore, using gridded data, it is not possible to represent sharp singularities due to the smoothing effect. This aspect, common to the interpolation methods discussed here, is intuitively contradictory in the context of detecting hot-spots where we need to identify singularities. Due to this reason, in Table 5.3, the percentages of correctly identified moderately contaminated areas are much higher as compared to the percentages of correctly identified highly contaminated areas. Thus, when the site investigator uses interpolation methods, he/she faces a high risk of overlooking the most essentially polluted zones.

The minimum curvature method (Figure 5.9) usually generates SCRAP near the corners of the site, because the interpolated locations are not completely surrounded by observations. (This is the usual trend of this method in other hypothetical sites too.)

On the other hand, Shepard's method (Figure 5.12) splits the main body of the hot-spot into more than one part and indicates clean areas in-between. This is due to the inward thrust of the hot-spot near the upper part which results in a singular square grid and more clean observations around that grid affect the interpolated values. The problem with misidentified contaminated areas dispersed over the corners of the site remains to be valid for this method too. In other hypothetical sites, we observe that even though peaks are away from the corners, due to the lack of surrounding data, they affect the interpolation near the corners. Again, these visual characteristics belong not only to this example, but to all other hypothetical sites used in the experiment.

In triangulation (Figure 5.4), the border of the indicated contaminated area resembles a mosaic consisting of triangles of various sizes and therefore, the borders are quite sharp (piecewise linear).

Table 5.3 Results of various SIM and FASA on a synthetic site.

METHOD	Identified Area % (moderately contaminated)	Identified Area % (highly contaminated)	COVER %	SCRAP %
FASA	-	-	95.38	7.70
Triangulation	87.47	10.61	79.38	2.62
Kriging-Linear	88.84	6.08	79.69	2.52
Kriging Spherical	85.91	6.76	77.99	2.44
Kriging-Quadratic	85.16	4.38	77.16	2.28
Kriging-Gaussian	80.91	18.07	74.92	14.77
Minimum curvature	87.29	8.68	78.76	4.91
RBF-Thin Plate	82.55	14.60	77.76	6.29
RBF-Multiquadratic	83.55	12.20	77.02	2.99
Shepard's Method	70.83	14.22	57.95	5.66

As observed in Figure 5.3, in FASA, the site area is divided into four parts in the first two iterations and three of the partitions are immediately discarded by the p^* criterion (0.85 of the maximum potential value.) (In Figure 5.3, partitions without samples represent discarded ones.) Next, the remaining partition in the upper right corner is divided into four parts in the next two iterations, one of which is totally discarded. The remaining undiscarded areas almost totally cover the hot-spot. The superfluous large SCRAP to the left of the cover is due to a single high concentration observation value falling on the boundary of the scrapped part. Notice that the high SCRAP area resulting from FASA is also due to the rectangular block-partitioning scheme implemented on a hot-spot with irregular shape.

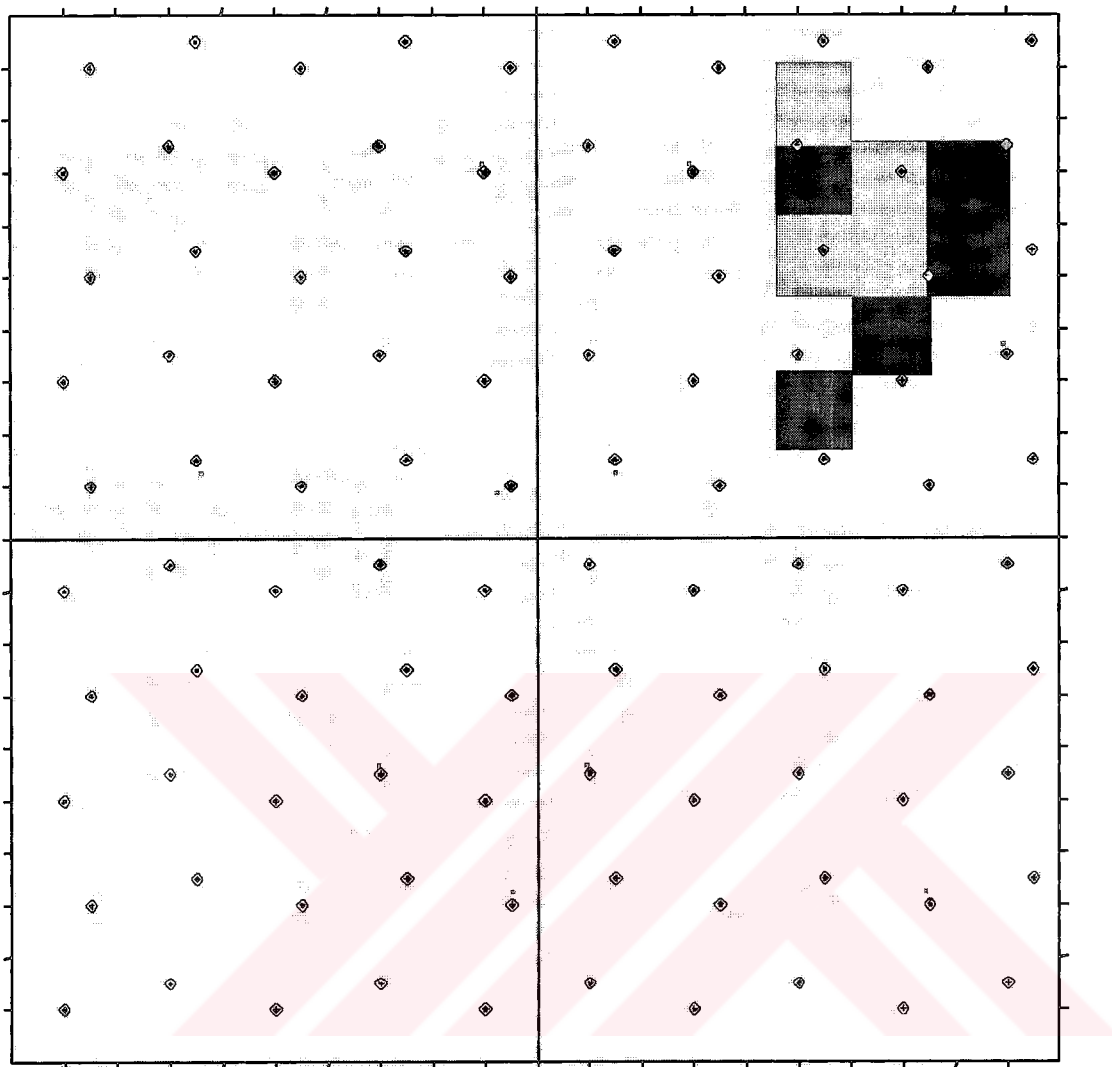


Figure 5.2 A hypothetical site with Herringbone sampling pattern

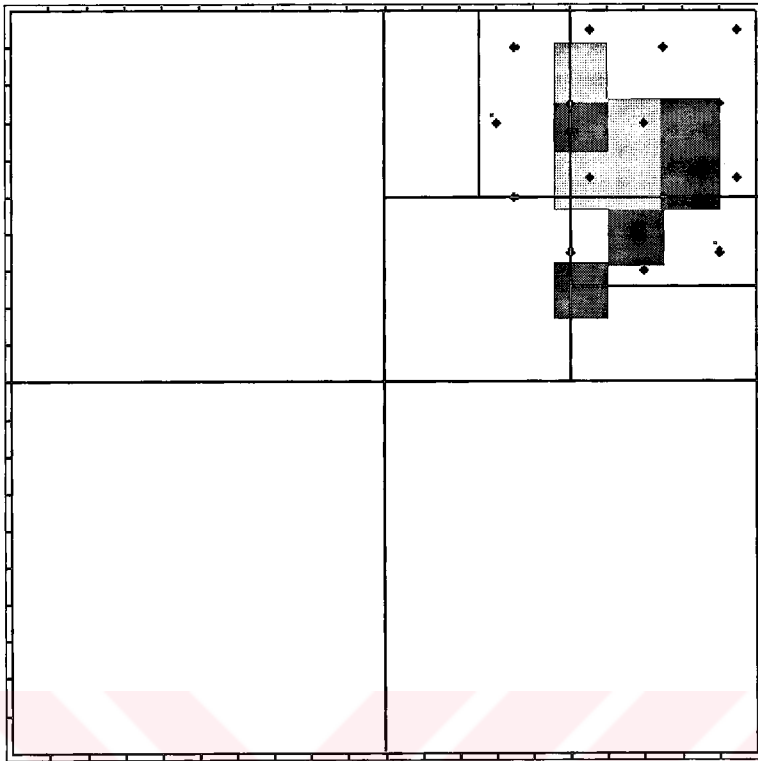


Figure 5.3 FASA

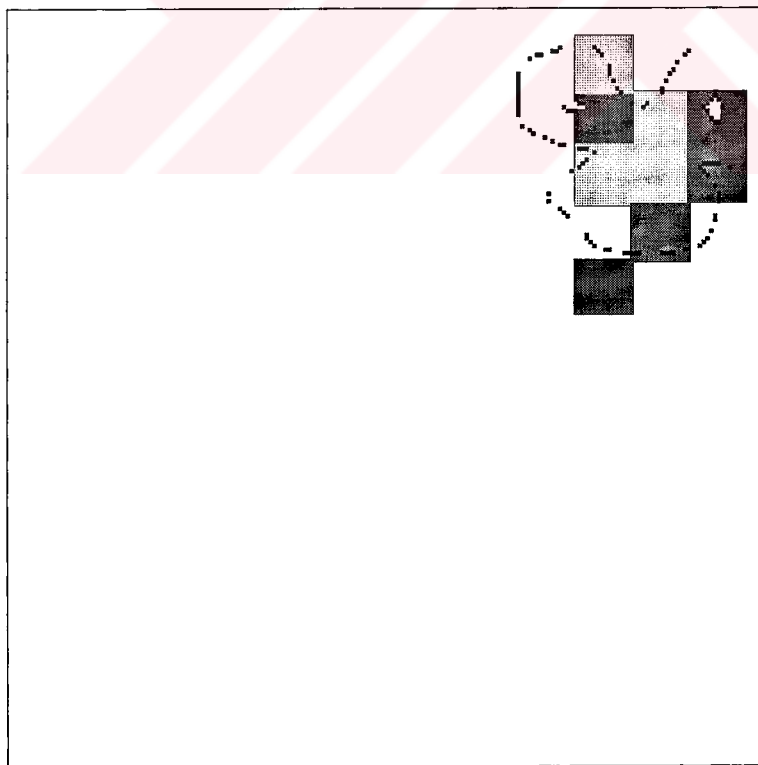


Figure 5.4 Triangulation

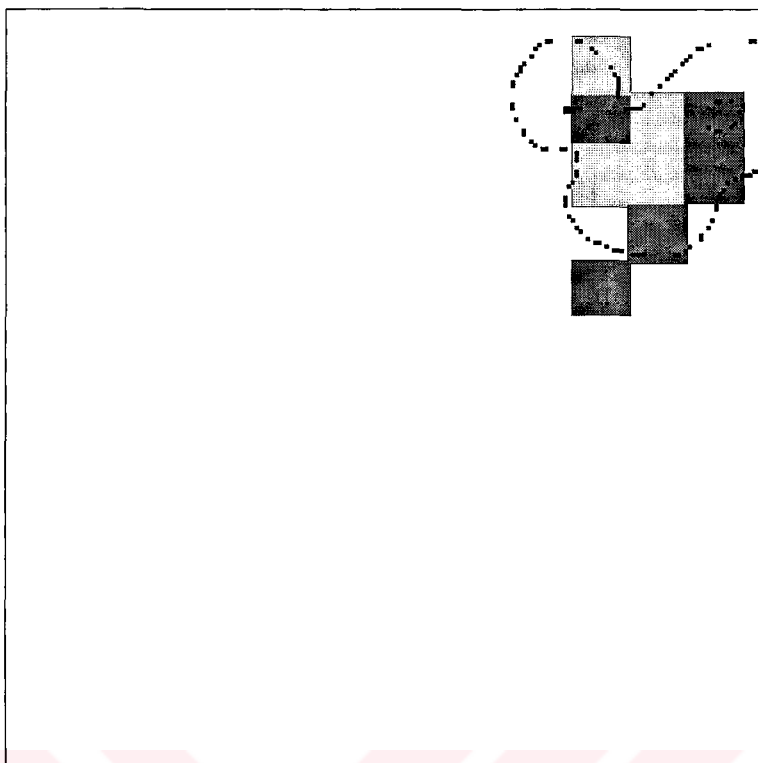


Figure 5.5 Spherical Kriging

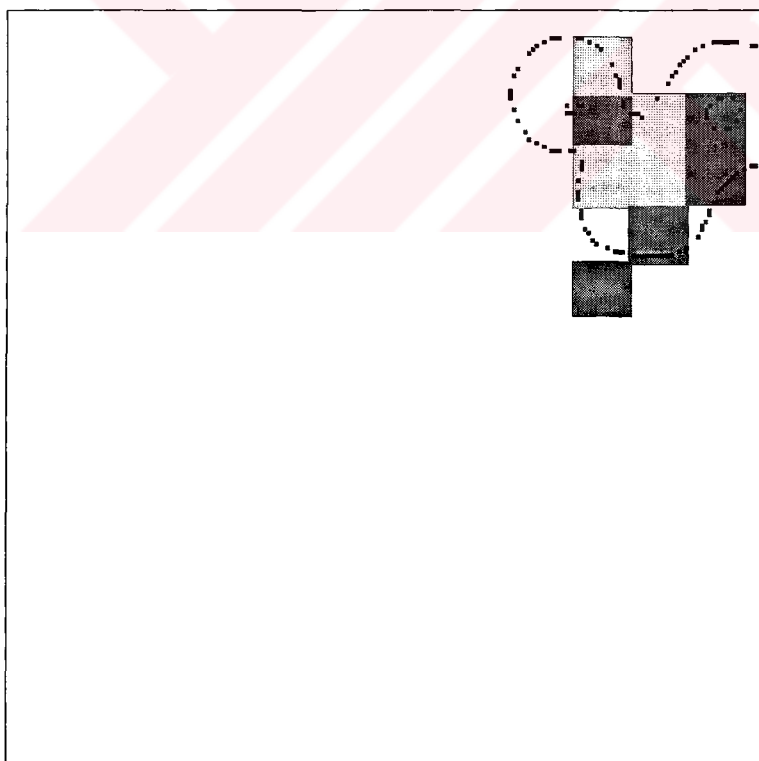


Figure 5.6 Linear Kriging

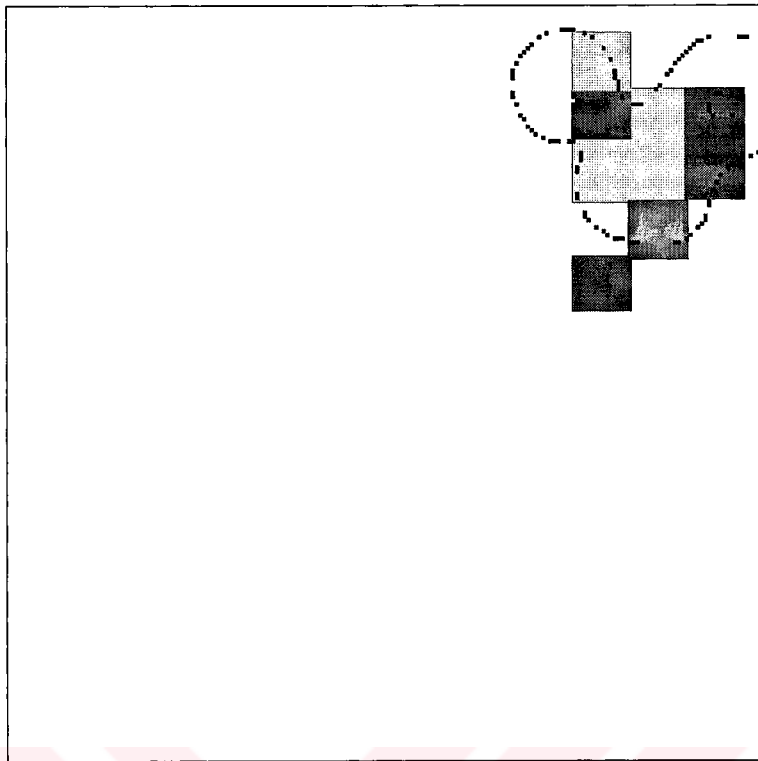


Figure 5.7 Quadratic

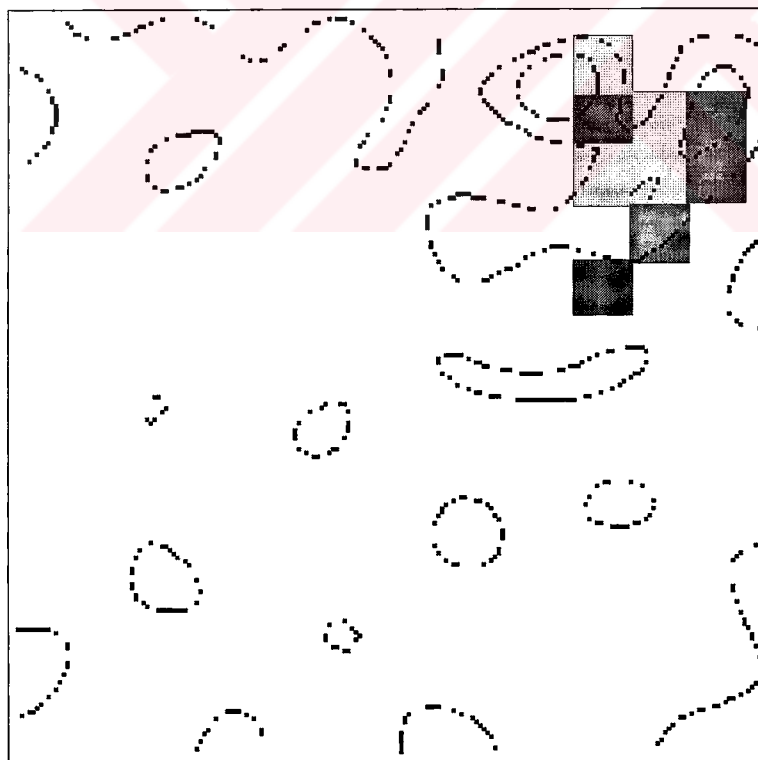


Figure 5.8 Gaussian Kriging



Figure 5.9 Minimum Curvature



Figure 5.10 RBF- Thin Plate

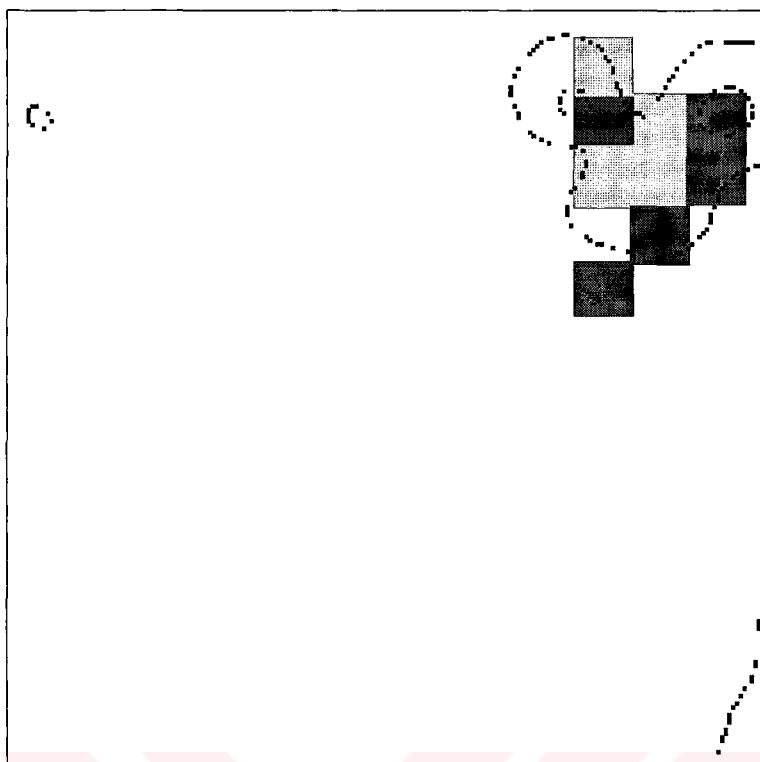


Figure 5.11 RBF- Multiquadratic

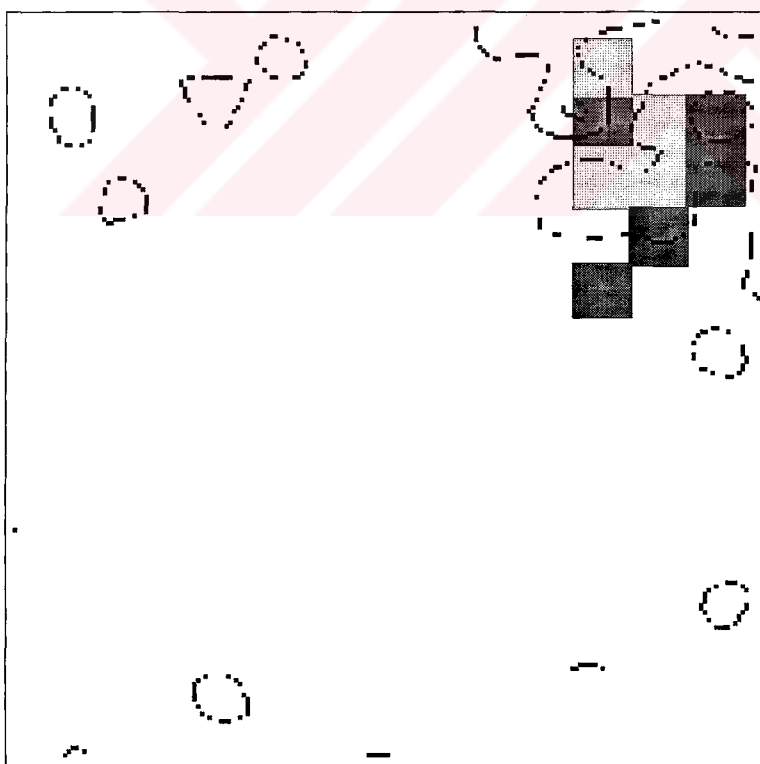


Figure 5.12 Shepards Method

5.5 Results

The results obtained on 360 synthetic sites by five SIM and FASA are provided in Tables 5.4-5.15. The sample size is 100 when PCA is 5% and 200 when PCA is 1% independent of the number of hot-spots, NHS. The results for each method are indicated for two percentages of hot-spot areas (with respect to whole site area) (PCA=1% and 5%), two NHS values and three sampling patterns. Performance is indicated by COVER (ratio of correctly identified hot-spot area to total hot-spot area) and SCRAP (ratio of clean area indicated as contaminated by the method to whole site area). Average, standard deviation and minimum levels of COVER and SCRAP are provided in Tables 5.4-5.15. Tables 5.4-5.9 provide the results for sites with a single hot-spot and Tables 5.10-5.15 provide the ones for the sites with three hot-spots. It is observed that the performance of all methods significantly deteriorate when NHS=3 irrespective of PCA level. On the other hand, when NHS=1, there is no statistically significant difference between performances at PCA levels of 1% and 5%, except for FASA. However, when NHS=3, performances of all methods deteriorate significantly as PCA decreases to 1%.

SIM are applied using the corresponding modules in the commercially available software Surfer (Golden Software). In all methods, no anisotropy is incorporated. RBF results are obtained using multi-quadratic basis function and all data are used in interpolation. Kriging is carried out with linear, quadratic gaussian and spherical variogram models, but only linear results are reported due to the fact that the three models do not lead to statistically significant results. There is no nugget effect and no drift. All data are honoured. Thus, kriging is used as an exact interpolator. On the other hand, in minimum curvature, all data may not be honoured. The maximum iteration number for grid node correction is set to 10000. In Shepard's methods the exponent p is set to 2 and the smoothing parameter is set to zero so that high observation values are honoured appropriately. All data are used in the interpolation.

FASA is coded by L. Özdamar and M. Demirhan using Borland C++ 5.01. Program implemented with the following parameters: the number of child areas, k , emanating from a parent is 2; the partitioning is carried out by dividing the longest edge of the parent region; the tolerance level, $areatol$, for discarding and storing a region for further

investigation is set to the given PCA level. In actual applications areatol may be as large as the expected percentage of contaminated area; tresamp is set to 3; the parameter p^* for discarding regions without further investigation because of low potential values is 0.85.

As observed in Tables 5.4-5.15, the Herringbone and the Grid sampling patterns seem to work best with regard to COVER whereas Stratified Random Sampling is relatively more inferior specially when NHS is 3.

Two-tailed paired t-tests are applied to detect differences among pairs of means (with respect to COVER criterion, and for the total number of 360 hypothetical sites, see Table 5.16. The values in Table 5.9 show that FASA is superior to all SIM and there exist no statistically significant differences in the performances of kriging, Minimum Curvature, RBF and Triangulation. Shepard's method is significantly inferior to all other methods.

With respect to the criterion of SCRAP, (detailed tables are omitted here), the performance of RBF, kriging and Shepard's method are not statistically different. However, minimum curvature has statistically significantly higher SCRAP. Furthermore, triangulation's performance is totally disappointing with regard to SCRAP possibly due to the reason mentioned previously, it provides a very coarse cover of the site with too large triangles. As mentioned previously, FASA's performance is compatible with that of minimum curvature when PCA=1%, but significantly higher when PCA is 5%. The parameter areatol, which is automatically set to 5%, seems to be too high and leaves superfluous SCRAP. Naturally, in all methods SCRAP increases significantly when PCA=5%.

Table 5.4 Linear kriging results.

PCA	%1		%5	
Herring Bone	COVER %	SCRAP %	COVER %	SCRAP %
AVG	75.42	1.36	86.08	3.61
STD	26.51	0.75	12.84	1.60
MIN	0.24	0.00	60.57	0.84
Grid				
AVG	75.36	1.28	87.03	3.80
STD	24.80	0.54	13.27	1.59
MIN	0.26	0.00	52.76	0.79
Stratified Random				
AVG	71.55	1.45	89.01	4.25
STD	29.83	0.93	7.14	1.73
MIN	0.24	0.06	71.46	1.33

Table 5.5 RBF- Multiquadratic results.

PCA	%1		%5	
Herring Bone	COVER %	SCRAP %	COVER %	SCRAP %
AVG	75.03	1.57	83.74	3.62
STD	25.03	0.63	13.29	1.48
MIN	0.24	0.27	57.38	1.15
Grid				
AVG	74.45	1.36	85.95	3.68
STD	24.38	0.52	13.45	1.54
MIN	1.36	0.02	53.27	1.18
Stratified Random				
AVG	69.50	2.22	86.45	4.61
STD	29.00	0.83	7.83	1.75
MIN	0.24	1.04	68.70	2.15

Table 5.6 Minimum curvature results.

PCA	%1		%5	
Herring Bone	COVER %	SCRAP %	COVER %	SCRAP %
AVG	75.22	2.94	83.36	4.95
STD	24.36	0.71	13.45	1.64
MIN	6.05	1.61	57.74	2.11
Grid				
AVG	74.17	2.84	84.69	5.21
STD	24.06	0.72	13.59	1.72
MIN	1.17	1.46	51.61	2.43
Stratified Random				
AVG	70.07	3.10	86.19	6.11
STD	27.39	1.03	7.53	1.94
MIN	0.24	1.43	69.89	2.88

Table 5.7 Shepard's results.

PCA	%1		%5	
Herring Bone	COVER %	SCRAP %	COVER %	SCRAP %
AVG	62.68	0.75	73.87	2.25
STD	26.87	0.51	17.62	1.53
MIN	0.24	0.00	40.12	0.12
Grid				
AVG	63.88	0.77	75.13	3.15
STD	28.30	0.51	19.54	1.77
MIN	0.26	0.00	25.85	0.42
Stratified Random				
AVG	60.88	0.79	78.93	4.28
STD	30.34	0.69	16.76	2.25
MIN	0.24	0.00	42.23	0.35

Table 5.8 Triangulation results.

PCA	%1		%5	
Herring Bone	COVER %	SCRAP %	COVER %	SCRAP %
AVG	74.29	12.62	84.30	21.52
STD	25.85	0.75	11.92	1.63
MIN	0.24	11.43	54.71	18.06
Grid				
AVG	72.57	15.67	86.14	24.42
STD	24.80	0.50	12.87	1.93
MIN	0.26	14.72	50.76	20.77
Stratified Random				
AVG	68.84	10.44	86.06	15.15
STD	28.44	0.92	9.21	2.48
MIN	0.24	8.88	66.74	10.93

Table 5.9 FASA results.

PCA	%1		%5	
Herring Bone	COVER %	SCRAP %	COVER %	SCRAP %
AVG	84.83	2.93	94.76	14.63
STD	14.44	1.22	8.97	6.00
MIN	48.97	0.77	62.11	7.50
Grid				
AVG	81.77	2.46	97.74	16.36
STD	17.77	1.46	6.57	5.90
MIN	29.60	0.59	70.82	2.70
Stratified Random				
AVG	71.10	1.69	89.91	9.25
STD	25.62	0.76	12.30	4.50
MIN	13.74	0.67	50.37	1.87

Table 5.10 Linear kriging results.

PCA	%1		%5	
Herring Bone	COVER %	SCRAP %	COVER %	SCRAP %
AVG	49.34	1.72	74.33	6.41
STD	28.72	1.11	15.50	2.44
MIN	0.23	0.00	44.47	2.13
Grid				
AVG	51.61	1.75	73.09	5.66
STD	26.44	1.14	14.98	2.00
MIN	0.23	0.00	32.47	2.27
Stratified Random				
AVG	41.69	1.81	65.49	6.35
STD	25.62	1.21	18.84	1.72
MIN	0.24	0.00	17.71	3.46

Table 5.11 RBF- Multiquadratic results

PCA	%1		%5	
Herring Bone	COVER %	SCRAP %	COVER %	SCRAP %
AVG	50.36	1.90	74.25	6.15
STD	28.82	1.01	14.77	2.19
MIN	0.23	0.04	43.88	2.04
Grid				
AVG	52.60	1.96	72.87	5.45
STD	25.92	1.12	14.63	1.78
MIN	0.23	0.29	31.98	2.68
Stratified Random				
AVG	43.81	2.59	64.71	6.75
STD	25.21	1.13	18.07	1.89
MIN	1.47	0.33	16.91	3.51

Table 5.12 Minimum curvature results

PCA	%1		%5	
Herring Bone	COVER %	SCRAP %	COVER %	SCRAP %
AVG	51.22	3.23	73.76	7.35
STD	29.35	1.13	14.19	2.21
MIN	0.23	1.52	43.50	3.29
Grid				
AVG	52.87	3.51	72.43	6.88
STD	26.05	1.11	14.63	1.95
MIN	0.23	1.67	31.09	4.08
Stratified Random				
AVG	44.50	3.47	65.40	7.86
STD	24.69	1.36	18.88	1.96
MIN	0.24	1.09	17.64	4.15

Table 5.13 Shepard's results.

PCA	%1		%5	
Herring Bone	COVER %	SCRAP %	COVER %	SCRAP %
AVG	39.30	0.76	61.69	4.60
STD	23.92	0.59	16.83	3.45
MIN	0.23	0.00	26.21	1.14
Grid				
AVG	41.15	0.81	58.70	3.71
STD	23.18	0.89	15.58	1.97
MIN	0.23	0.00	29.60	0.84
Stratified Random				
AVG	34.82	0.98	53.76	4.48
STD	24.24	0.87	15.22	1.91
MIN	0.24	0.00	16.34	1.87

Table 5.14 Triangulation results.

PCA	%1		%5	
Herring Bone	COVER %	SCRAP %	COVER %	SCRAP %
AVG	49.94	12.75	75.87	23.00
STD	27.57	0.86	13.71	2.21
MIN	0.24	11.49	43.82	18.23
Grid				
AVG	54.28	16.18	72.70	25.81
STD	23.46	0.86	12.78	2.26
MIN	0.23	14.85	40.96	21.69
Stratified Random				
AVG	43.45	10.73	64.58	17.60
STD	22.45	1.24	17.58	2.40
MIN	6.17	8.24	21.39	11.24

Table 5.15 FASA results.

Herring Bone	COVER %	SCRAP %	COVER %	SCRAP %
AVG	55.09	3.73	86.20	21.93
STD	24.49	2.24	14.54	7.66
MIN	3.77	1.05	41.27	9.16
Grid				
AVG	50.99	3.63	88.90	2.59
STD	23.06	1.74	14.15	12.91
MIN	8.20	1.22	43.16	9.16
Stratified Random				
AVG	41.60	2.65	67.14	15.80
STD	25.10	1.45	19.33	5.16
MIN	5.13	1.23	29.08	4.61

Table 5.16 Overall results of methods for NHS=1

PCA	1%		5%	
Minimum Curv.	COVER %	SCRAP %	COVER %	SCRAP %
avg	73,15	2,96	84,75	5,42
std	25,27	0,82	11,52	1,77
Linear Kriging				
avg	74,11	1,37	87,37	3,89
std	27,04	0,74	11,08	1,64
Triangulation				
avg	71,90	12,91	85,50	20,36
std	26,36	0,72	11,34	2,01
RBF-Multiquadric				
avg	72,99	1,72	85,38	3,97
std	26,14	0,66	11,52	1,59
Shepard's				
avg	62,48	0,77	75,98	3,23
std	28,50	0,57	17,97	1,85
FASA				
avg	79,23	2,36	94,14	13,41
std	19,28	1,15	9,28	5,47

Table 5.17 Overall results of methods for NHS=3

PCA	1%		5%	
Minimum Curv.	COVER %	SCRAP %	COVER %	SCRAP %
avg	49,53	3,40	70,53	7,36
std	26,69	1,20	15,90	2,04
Linear Kriging				
avg	47,55	1,76	70,97	6,14
std	26,93	1,15	16,44	2,05
Triangulation				
avg	49,22	13,22	71,05	22,14
std	24,49	0,99	14,69	2,29
RBF-Multiquadric				
avg	48,92	2,15	70,61	6,12
std	26,65	1,08	15,82	1,95
Shepard's				
avg	38,42	0,85	58,05	4,26
std	23,78	0,78	15,87	2,44
FASA				
avg	49,23	3,34	80,75	13,44
std	24,22	1,81	16,01	8,58

Table 5.18 Overall results of sampling patterns for NHS=1

PCA	1%		5%	
HerringBone	COVER %	SCRAP %	COVER %	SCRAP %
avg	74,69	3,69	84,73	8,43
std	23,84	0,76	13,01	2,31
Grid				
avg	73,39	4,06	86,53	9,44
std	24,02	0,71	13,21	2,41
Stratified Random				
avg	68,66	3,28	86,09	7,28
std	28,44	0,86	10,13	2,44

Table 5.19 Overall results of sampling patterns for NHS=3

PCA	1%		5%	
HerringBone	COVER %	SCRAP %	COVER %	SCRAP %
avg	49,55	4,01	74,17	11,57
std	27,14	1,16	14,92	3,36
Grid				
avg	49,31	4,64	72,03	8,35
std	24,69	1,14	14,46	3,81
Stratified Random				
avg	41,65	3,70	63,51	9,81
std	24,55	1,21	17,99	2,51

Table 5.20 Paired t-test statistics for 360 sites (*: significant at 5% level)

	FASA	Kriging	MINC	RBF	Shepard's	Triangulation
FASA		9.057*	4.083*	5.201*	15.201*	5.283*
Kriging			1.5261	1.5647	33.026*	1.7580
MINC				1.837	18.279*	1.614
RBF					20.258*	0.180
Shepard's						-19.418*

Two-tailed t-test is applied. Positive t-test values indicate that the method in row i is superior to the one in column j in cell i,j and negative values indicate the converse.

6 *PERFORMANCE EVALUATION of SPATIAL INTERPOLATION METHODS in the PRESENCE OF NOISE*

A handicap in site characterization is the existence of noise in the observed values. For example, sampling errors (often, duplicate samples taken from same location result in different observation values) resulting from the sampling methodology dominate laboratory analysis errors. Consequently, a realistic performance investigation method should involve data with noise. Therefore, we carry out experiments where data are distorted by additive white noise as is the case in many applications. We demonstrate the effectiveness of SIM considered here, that is Multi-quadratic Radial Basis Function, Kriging, Minimum Curvature, and Shepard's Inverse Distance method, by using noisy data. We omit Surfer's Triangulation method due to extremely bad results obtained in preliminary testing.

Here, in order to conduct more refined experiments, we utilize mathematical functions to represent sites. The reason is that in the previous test sites, we only measured the COVER and SCRAP in areal terms, we did not measure how imprecise were the contours of the interpolated surfaces. In the previous design group this precision could not be measured, since we did not know the contaminant concentration at every point within the site. In the case of noisy data, we are interested in the amount of errors generated by the reconstruction of the site surface. Consequently, we evaluate performance according to the absolute error between the true value of a given mathematical function and its corresponding interpolated value. Again, due to the fact that FASA is not an interpolator, it is omitted from the second set of experiments.

6.1 The Test Functions

8 test functions are devised to illustrate various possible characteristics of hypothetical surface patterns on which the performance of SIM are to be evaluated. Initially, SIM are utilized to approximate the behaviours of the original forms of the functions based a given number of sample data. (Here, 196 samples are collected.) Then, the surface of each function is distorted with additive Gaussian noise (noise is generated in proportion to the

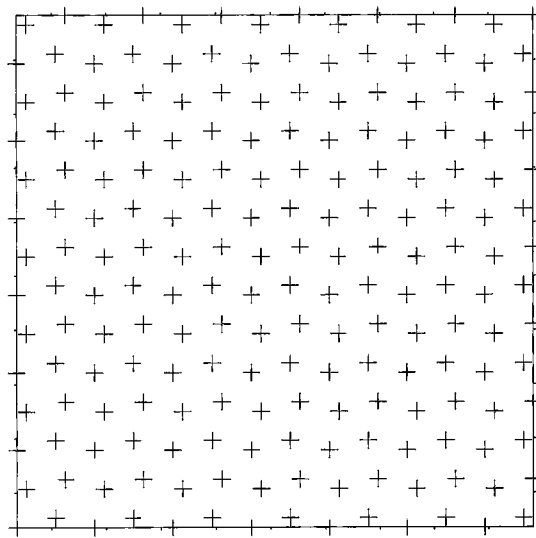
standard deviation of function values pertaining to 196 data) and the performances of SIM are re-evaluated.

The three dimensional images of the test functions are provided in Appendix A. The first test function represents a plateau type of surface where sharp rises occur near the corners. The second, third and the fourth test functions are generated according to the scheme described in Stuckmann (1988). The third function represents very sharp stepwise changes of attribute level whereas the fourth one involves stepwise changes of smaller magnitudes, thus presenting a smoother surface. The second one (No. 2a) is again a discontinuous function where changes are relatively smoother except at one corner of the site. Furthermore, when the steps of the second function are removed, a very slightly undulating function (No. 2b) results. The fifth test function is a symmetric function with very frequent changes in attribute levels achieved over small distances. The sixth function has a trend in attribute level and a relatively lower number of extremities. On the opposite hand, the seventh one has many extremities in between which there lie plateaus and valleys.

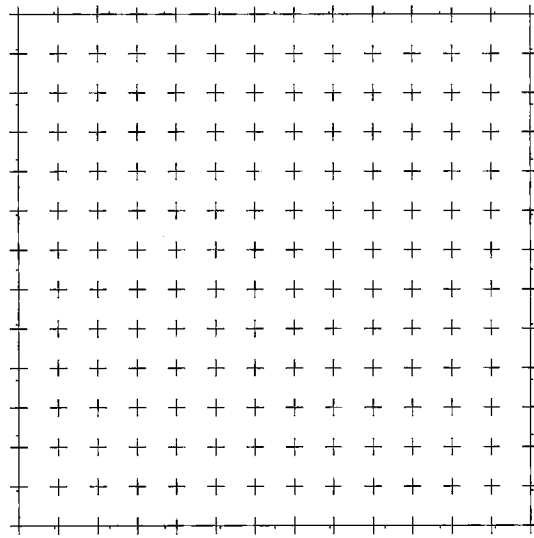
6.2 The Sampling Patterns

Since in the previous experiments, we observe that the Herringbone and grid patterns are most successful, we implement these two sampling patterns in the case of noisy data. Furthermore, we develop and introduce two additional sampling patterns: namely the linear and circular sampling patterns. In Figure 6.1, the four sampling patterns are depicted.

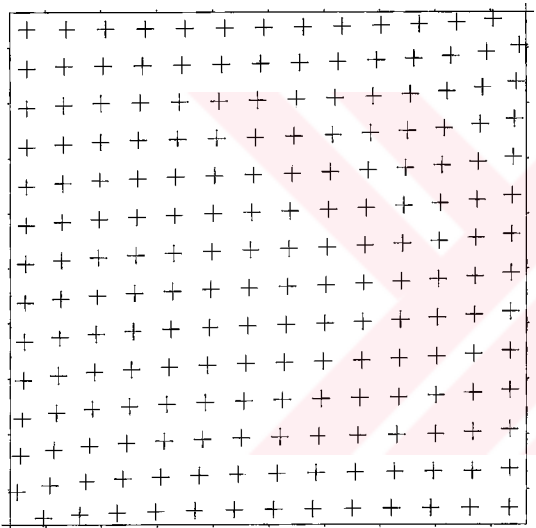
The *linear sampling pattern* is generated by placing observations on k parallel equidistant lines drawn across the surface. These lines divide the surface into equidistant cross-sections and each makes an angle of 45° with the x -axis. The first line contains a single observation while the second one contains two observations. In this manner, the number of observations on each line becomes equivalent to its index until $k/2$ lines are drawn. Then, the number of observations in each succeeding line decreases by one. Hence, given a fixed number of observations, n , the number of equidistant lines to be drawn in parallel can be calculated easily.



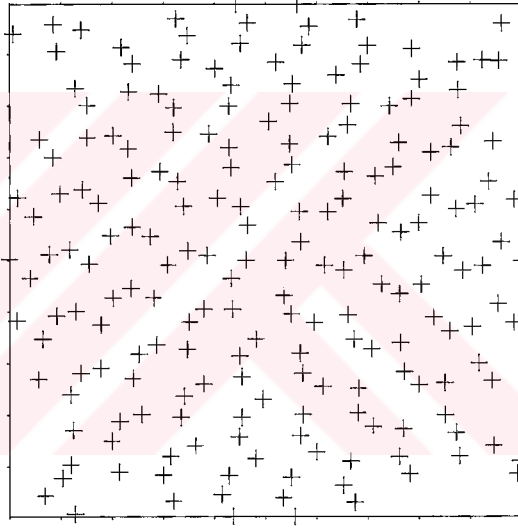
Herringbone



Regular Grid



Linear



Circular

Figure 6.1 Sampling patterns used for noisy situation

In the *circular sampling pattern*, the first observation is placed at the geometric center of the surface. Similar to the linear pattern, one more observation is added to each additional circle drawn around the center of the surface. However, in order to eliminate aligned positions, the first observation on each circle is offset by a given degree from the last observation placed on the previous circle. The degree of offsetting is computed by dividing 360° by the number of circles to be drawn.

6.3 An Example Demonstrating the Distortion Caused by Additive White Noise

Test function #1 is used to demonstrate the distortion of the surface's shape caused by white noise. In Figure B.1 the contours of the function are depicted. In Figures B.2 – B.5, the surfaces interpolated by kriging using the four sampling methods are given. The last four contours (Figures B.6 – B.9) belong to the corresponding four sampling methods' interpolated surfaces when noise is added to the data. Table 6.1 provides the average and standard deviation of the absolute errors (absolute value of the functional value minus the interpolated value) obtained by the four sampling patterns using the Kriging interpolation method (50x50 grid). Although the regular grid pattern provides the minimum average absolute value, its performance in the presence of noise is intolerable (see Figure B.3 and Figure B.7). When the data are noise-free, 81% of the interpolation errors are within 1% of the function's range over the defined domain. On the other hand, when noise is added, only 20% of the errors fall within this range.

Table 6.1. Performance of four sampling patterns using Kriging interpolation method.

Kriging	Noise-Free				Additive Gaussian Noise			
Pattern	HerrB	Grid	Linear	Circular	HerrB	Grid	Linear	Circular
Avg. Error	7.63	6.61	11.52	19.38	63.00	47.40	55.69	53.80
Std. Dev.	11.23	9.10	23.64	47.73	41.52	35.83	37.58	65.36

6.4 Numerical Experiments with SIM

The test functions given in Appendix A, provide our medium of experimentation. Observations are made on the surface domains given for each function according to each of the four sampling patterns described above. Interpolation is then applied to each function in both cases where the original and the distorted observation values (196 data for each function domain) are utilized as functional information. The range of the noise added to each observation value is between ± 2 standard deviations around the mean. In Tables 6.2-6.10 each function's interpolation error results are provided under four interpolation methods and four sampling patterns. In tables 6.11-6.12 the average of all methods is given

in order to make a the method base performance evaluation. The performance of each SIM is measured in terms of both absolute error and percentage error. However, percentage error details are not given here due to space restrictions. The first two rows in each Table indicate average and standard deviation of the absolute difference of interpolated values from the original function values. (The number of interpolated points is equal to 2500 resulting from a 50x50 grid.) The numbers indicated under the column "method average" provide the average results obtained by each method over the four sampling patterns. The rows starting with "Sample average" indicate the average results of each sampling pattern over four SIM. The third row under the "Sample average" provides the range of the function values (over the 2500 interpolated locations) so that the reader can compare average absolute errors against this range.

Inspecting the performance of the four SIM without including Gaussian noise in the data, Shepard's method is the worst performing one except for the fifth function.

The performance of Minimum Curvature is considerably superior to that of Shepard's. However, it is the third SIM in terms of performance. The performance of RBF is consistently the best or second best. At times Kriging is slightly superior to RBF. However, when RBF is superior to Kriging, there is considerably difference between the performance of the two methods. RBF performs specifically well in smooth functions where abrupt changes do not take place (e.g., functions #1, #3, #6). On the other hand, when noise is included in the data Kriging's performance is superior to that of RBF. Thus, Kriging turns out to be more robust against noise.

Considering the differences of sampling patterns (again excluding noise from data), the circular sampling pattern is the worst performing one due to the fact that it does not cover the site uniformly. The grid and linear patterns behave in a similar manner. The Herring Bone pattern follows closely the latter two patterns. Consequently, in the noise-free case, patterns generating a uniform cover of the domain work better.

In the case where noise is included, the regular grid is again the best performing sampling pattern and the linear pattern is a close follower of the regular grid. The Herring Bone and circular patterns share the worst performing position. Consequently, the regular grid seems to be the most robust sampling pattern against white noise.

Table 6.2 Performance evaluations for Kriging without Noise

Without Gaussian Noise					
	Herringbone	Regular Grid	Linear	Circular	
Kriging					Method Avg
avg	7,63	6,61	11,52	19,38	11,29
std	11,23	9,10	23,64	47,73	22,93
range	1839,24	1532,81	1532,81	2149,45	1763,58
min	-1839,35	-1532,87	-1532,87	-2149,47	-1763,64
max	-0,11	-0,06	-0,06	-0,02	-0,06
%1	84,00	81,24	80,76	83,24	82,31
%5	99,92	100,00	96,32	95,04	97,82
%10	100,00	100,00	99,72	98,72	99,61
%20	100,00	100,00	100,00	99,80	99,95

Table 6.3 Performance evaluations for Kriging with Noise

With Gaussian Noise					
	Herringbone	Regular Grid	Linear	Circular	
Kriging					Method Avg
avg	63,00	47,40	55,69	53,80	54,97
std	41,52	35,83	37,58	65,36	45,07
range	1839,24	1532,81	1532,81	2149,45	1763,58
min	-1839,35	-1532,87	-1532,87	-2149,47	-1763,64
max	-0,11	-0,06	-0,06	-0,02	-0,06
%1	15,24	20,80	14,76	27,44	19,56
%5	76,04	80,64	73,32	91,88	80,47
%10	99,24	99,28	98,84	97,80	98,79
%20	100,00	100,00	100,00	99,48	99,87

Table 6.4 Performance evaluations for Minimum Curvature without Noise

Without Gaussian Noise					
	Herringbone	Regular Grid	Linear	Circular	
Minimum curvature					Method Avg
avg	12,66	10,51	11,68	24,91	14,94
std	13,53	10,67	16,87	56,25	24,33
range	1839,24	1532,81	1532,81	2149,45	1763,58
min	-1839,35	-1532,87	-1532,87	-2149,47	-1763,64
max	-0,11	-0,06	-0,06	-0,02	-0,06
%1	72,12	73,80	77,00	79,56	75,62
%5	100,00	100,00	98,36	93,32	97,92
%10	100,00	100,00	100,00	97,60	99,40
%20	100,00	100,00	100,00	99,84	99,96

Table 6.5 Performance evaluations for Minimum Curvature with Noise

With Gaussian Noise					
	Herringbone	Regular Grid	Linear	Circular	
Minimum curvature					Method Avg
avg	71,30	52,56	63,08	59,19	61,53
std	49,23	38,97	48,49	61,05	49,44
range	1839,24	1532,81	1532,81	2149,45	1763,58
min	-1839,35	-1532,87	-1532,87	-2149,47	-1763,64
max	-0,11	-0,06	-0,06	-0,02	-0,06
%1	14,80	18,76	14,32	24,64	18,13
%5	69,36	75,28	67,28	87,64	74,89
%10	97,64	98,56	95,64	97,44	97,32
%20	100,00	100,00	99,80	99,68	99,87

Table 6.6 Performance evaluations for Radial Basis without Noise

Without Gaussian Noise					
	Herringbone	Regular Grid	Linear	Circular	
Radial Basis					Method Avg
avg	5,28	4,92	9,11	14,85	8,54
std	9,36	8,56	20,26	41,02	19,80
range	1839,24	1532,81	1532,81	2149,45	1763,58
min	-1839,35	-1532,87	-1532,87	-2149,47	-1763,64
max	-0,11	-0,06	-0,06	-0,02	-0,06
%1	90,32	86,00	85,04	85,28	86,66
%5	99,92	100,00	97,44	96,56	98,48
%10	100,00	100,00	99,80	99,12	99,73
%20	100,00	100,00	100,00	99,92	99,98

Table 6.7 Performance evaluations for Radial Basis with Noise

With Gaussian Noise					
	Herringbone	Regular Grid	Linear	Circular	
Radial Basis					Method Avg
avg	68,23	52,06	58,48	55,20	58,49
std	47,18	38,85	41,71	61,02	47,19
range	1839,24	1532,81	1532,81	2149,45	1763,58
min	-1839,35	-1532,87	-1532,87	-2149,47	-1763,64
max	-0,11	-0,06	-0,06	-0,02	-0,06
%1	15,52	18,36	14,88	26,04	18,70
%5	70,68	76,48	71,00	90,60	77,19
%10	98,20	98,44	97,44	98,16	98,06
%20	100,00	100,00	100,00	99,48	99,87

Table 6.8 Performance evaluations for Shepard's without Noise

Without Gaussian Noise					
	Herringbon	egular Gri	Linear	Circular	
Shepards					Method Avg
avg	116,03	110,56	100,99	143,51	117,77
std	68,80	63,92	71,54	134,16	84,61
range	1839,24	1532,81	1532,81	2149,45	1763,58
min	-1839,35	-1532,87	-1532,87	-2149,47	-1763,64
max	-0,11	-0,06	-0,06	-0,02	-0,06
%1	8,08	6,56	6,84	7,92	7,35
%5	38,56	32,92	40,40	45,04	39,23
%10	82,24	72,92	83,00	85,40	80,89
%20	99,88	100,00	98,40	95,76	98,51

Table 6.9 Performance evaluations for Shepard's with Noise

With Gaussian Noise					
	Herringbon	egular Gri	Linear	Circular	
Shepards					Method Avg
avg	161,39	141,12	135,07	166,12	150,93
std	76,70	74,34	69,75	155,40	94,05
range	1839,24	1532,81	1532,81	2149,45	1763,58
min	-1839,35	-1532,87	-1532,87	-2149,47	-1763,64
max	-0,11	-0,06	-0,06	-0,02	-0,06
%1	3,12	3,80	4,28	5,84	4,26
%5	20,68	21,08	20,76	33,96	24,12
%10	59,64	56,00	61,16	79,32	64,03
%20	99,88	99,08	98,52	95,48	98,24

Table 6.10 Performance evaluations for sampling methods without Noise

Without Gaussian Noise					
	Herringbone	Regular Grid	Linear	Circular	Method Avg
Sample Avg	35,40	33,15	33,33	50,66	38,14
Sample Std	25,73	23,06	33,08	69,79	37,92
range	1839,24	1532,81	1532,81	2149,45	1763,58
min	-1839,35	-1532,87	-1532,87	-2149,47	-1763,64
max	-0,11	-0,06	-0,06	-0,02	-0,06
%1	63,63	61,90	62,41	64,00	62,99
%5	84,60	83,23	83,13	82,49	83,36
%10	95,56	93,23	95,63	95,21	94,91
%20	99,97	100,00	99,60	98,83	99,60

Table 6.11 Performance evaluations for sampling methods with Noise

With Gaussian Noise					
	Herringbone	Regular Grid	Linear	Circular	Method Avg
Sample Avg	90,98	73,28	78,08	83,58	81,48
Sample Std	53,66	47,00	49,38	85,71	58,94
range	1839,24	1532,81	1532,81	2149,45	1763,58
min	-1839,35	-1532,87	-1532,87	-2149,47	-1763,64
max	-0,11	-0,06	-0,06	-0,02	-0,06
%1	12,17	15,43	12,06	20,99	15,16
%5	59,19	63,37	58,09	76,02	64,17
%10	88,68	88,07	88,27	93,18	89,55
%20	99,97	99,77	99,58	98,53	99,46

Table 6.12 Average performance comparisons of all methods

	without error			
	Kriging	MNC	RBF	Shepards
	Method Avg	Method Avg	Method Avg	Method Avg
avg	11,29	14,94	8,54	117,77
std	22,93	24,33	19,80	84,61
%1	82,31	75,62	86,66	7,35
%5	97,82	97,92	98,48	39,23
%10	99,61	99,40	99,73	80,89
%20	99,95	99,96	99,98	98,51
	with error			
	Kriging	MNC	RBF	Shepards
	Method Avg	Method Avg	Method Avg	Method Avg
avg	54,97	61,53	58,49	150,93
std	45,07	49,44	47,19	94,05
%1	19,56	18,13	18,70	4,26
%5	80,47	74,89	77,19	24,12
%10	98,79	97,32	98,06	64,03
%20	99,87	99,87	99,87	98,24

CONCLUSION

The issue of preserving the environment in its natural form becomes more and more crucial in developing countries as well as the developed ones since the industrialization process gains speed. Joint projects are being taken up in developed countries involving related units in the public administration, scientists, and industrial partners so as to support the environmental protection legislation with scientific standards. These projects concentrated on the following topics:

- development of decision support systems for the characterization of potentially contaminated brownfields and the reduction of their remediation costs (e.g., use of Geographical Information Systems (GIS), for spatial data management and planning,
- use of geostatistical techniques for the determination of contaminant distribution and behaviour in soil and groundwater),
- state risk assessment procedures and models for analysing the relative contribution of multi-source contaminants to human exposures,
- determination of the health risks from contaminated soil in urban environments over time and space,
- rehabilitation of contaminated sites, sediments and groundwater bodies to prevent water pollution, monitoring and assessment of water/land quality, early warning and communication systems in support of pollution prevention.

Brownfields remediation is also important in the preservation of groundwater resources and air quality as well as the development and re-use of underground construction works and other earthworks. Furthermore, the inevitable requirement that housing settlements, green parks and agricultural land need to be situated on clean soil whose contaminants are well below tolerance levels, re-emphasize the crucial issue of land protection and remediation.

The investigation of a potentially contaminated site involves the following phases: Selection of the sampling strategy (spatial distribution of samples); Collection of samples;

Laboratory analysis, Evaluation of data and Assessment of the site's contamination levels. The phases where the major decisions are made are the first and last phases.

In the last phase of site investigation, the usual techniques to identify the distribution of the contaminant over the site are Spatial Interpolation Methods (SIM). These methods are usually interpolation- or regression-based approaches which use sampled observations to predict the contamination levels at unsampled locations (spatial interpolation). Originally, most of these approaches come from mineral resource estimation in mining, and are classified as geostatistical methods. The rationale behind spatial interpolation is the observation that points close together in space are more likely to have similar values than points far apart.

In this thesis, the most widely used SIM and an alternative areal site assessment method (FASA) compared on hypothetical contaminated sites. FASA is based on a fuzzy evaluation of partitioned zones within a given site. The numerical survey is conducted on 360 hypothetical sites and linear ordinary kriging, radial basis functions, minimum curvature method, Shepard's method, triangulation and FASA are tested. In the experiments, the effects of factors, such as the number of hot-spots in a site and the percentage of the contaminated zones within the site, are demonstrated. A visual insight on all of the above mentioned methods by demonstrating their performance on an exemplary site is also provided.

Furthermore, as most of the investigations are conducted using data with uncontrollable noise, we extend our experiments to the case of assessing the performance of SIM in the presence of noise data. It is important to investigate the robustness of these techniques against noise. The experiments conducted here also evaluate two novel sampling patterns introduced here. These experiments are realized on a test bed of 8 mathematical functions and it is observed that the simplest sampling pattern, namely, the regular grid, is the most robust sampling pattern. Furthermore, among the four spatial interpolation techniques tested here, kriging is the most robust one against noise.

A further project as a consequence of this thesis may be the evaluation of the SIM and FASA on real world instead of the computer based test functions and environments. After the field experiments and evaluation of the results with the ones in this thesis, a roadmap for the land contamination detection problem may be prepared.



REFERENCES

- Alloway, B.J., *Understanding our Environment and Introduction to Environmental Chemistry and Pollution*, Chapter 5, Ed. R.M. Harrison, Second Edition, The Royal Society of Chemistry, London, 1995.
- Bardossy, A., 1987, Notes on the robustness of the kriging system. *Mathematical Geology*, 20, 189-203.
- Briggs, I.C., 1974, Machine contouring using minimum curvature. *Geophysics*, 39, 39-48.
- Carr, J., 1996. Surface reconstruction in 3D medical imaging, unpublished PhD Thesis, Electrical and Electronic Eng., University of Canterbury, New Zealand.
- Cressie, N.A., 1993, *Statistics for Spatial Data*, John Wiley & Sons, New York.
- Davis, J.C., 1986, *Statistics and Data Analysis in Geology*, (2nd edition), John Wiley and Sons, New York.
- Deutsch, C.V., Journel, A.G., *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, 1998.
- Ferguson, C.C., 1992, The statistical basis for spatial sampling of contaminated land. *Ground Engineering*, 25, 34-38.
- Fogel D.N., 1997, Image rectification with radial basis functions: Application to RS/GIS data integration. Available on site <http://polux.geog.ucsb.edu/~ifogel/sf/santafe.html>.
- Fogel, D.N., Tinney, L., 1996, Image registration using multi-quadric functions. Technical Report No. 96-01, National Center for Geographic Information and Analysis, University of California, Santa Barbara.
- Henley, S. and Watson, D.F. 1997, Geostatistics: A critical review. Working Paper, CSIRO Exploration and Mining, Nedlands, Western Australia.
- Hutchinson, M.F., 1997, A locally adaptive approach to the interpolation of digital elevation models. Working Paper, Australian National University, Center for Resource and Environmental Studies, Canberra, Australia.
- Isaaks, E.H., and Srivastava, R.M., *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 1989.

Journel, A.G. and Huijbregts, C., *Mining Geostatistics*, Academic Press, 1981.

Kerry, K.E., and Hawick, K.A., 1996, Interpolation on high performance computers, *ASCE '96*, Available on site <http://www.cs.adelaide.edu.au/users/katrina/hpcn/index.html>.

Krige, D.G., A statistical approach to some mine valuations and allied problems at witwatersrand. MSC. Thesis, University of Witwatersrand, South Africa, 1951.

Luenberger, D., *Optimization by Vector Space Methods*. John Wiley and Sons, New York, 1969.

M. Demirhan, M., Özdamar, L., Helvacioğlu, L., Birbil, Ş. I. 1998, FRACTOP: A Geometric partitioning metaheuristic for global optimization. To appear in the *Journal of Global Optimization*.

Matheron, G., 1971, The theory of regionalized variables and its applications. *Les Cahier du Centre de Morphologie Mathematique*, 5: pp211.

Olea, R. (Ed.), *Geostatistical Glossary and Multilingual Dictionary*, Oxford University Press, New York, 1991.

Oliver, M.A., Webster, R., 1990, Kriging: A method of interpolation for geographical information systems. *Int. J. Geographic Information Systems*, 4, 313-332.

Özdamar, L. and Demirhan, M., 1998, Comparison of partition evaluation measures in an adaptive partitioning algorithm for global optimization. To appear in *Fuzzy Sets and Systems*.

Pal, N.R., Bezdek, J.C., 1994. Measuring fuzzy uncertainty, *IEEE Transactions on Fuzzy Systems*, 2, 107-118.

Powell, M.J.D. 1992, The theory of radial basis function approximation in 1990. *Advances in Numerical Analysis*, W. Light, ed., Oxford. Oxford Science Publications, pp 105-210.

Practical Aspects of Applying Geostatistics at Hazardous, Toxic, and Radioactive Waste Sites, U.S. Army Corps of Engineers, CEMP-RT, Technical Report ETL 1110-1-175, 1997.

Rauth, M, 1998. Gridding of Geophysical Potential Fields from Noisy Scattered Data, Institut für Meteorologie und Geophysik der Universität Wien

Ripley, B.D., *Spatial Statistics*, John Wiley and Sons, New York, 1981.

Rohling, R.N., Gee, A.H., Berman, L., 1998. Radial Basis Function Interpolation for 3-D Ultrasound, Working Paper, Cambridge University Engineering Dept., UK.

Rosenbaum, M. and Soderstrom, M., 1996, Geostatistics as an aid to mapping. <http://www.esri.com/base/common/userconf/europroc96/PAPERS/PN11/PN11F.HTML>

RTD Fifth Framework Programme (EC), 1998. *Environment and Climate Programme*, Related key actions: "Sustainable Management and Quality of Water", "City of Tomorrow and Cultural Heritage". <http://europa.eu.int/comm/dg12/fp5.html>.

Sampling Strategies for Contaminated Land. U.K. DoE Contaminated Land Research Report No. 4, 1994.

Shurtz, 1997, Some kriging results. Private publication.

Shurtz, R.F., 1985, Letter to the Editor. *Journal of the International Association for Mathematical Geology*, 17, 861-868.

U.K. DoE Contaminated Land Research Report No. 4, 1994. *Sampling Strategies for Contaminated Land*.

U.S. Army Corps of Engineers, HTRW Report, 1997. Practical Aspects of applying geostatistics at Hazardous, Toxic, and radioactive Waste Sites, CEMP-RT, Technical Report ETL 1110-1-175.

U.S. Environmental Protection Agency, Road Map to Understanding Innovative Technology Options for brownfields Investigation and Cleanup, 1997. Washington.

USGS, 1987. Digital Elevation Models, Data Users Guide 5. US Dept. of the Interior, USGS, Reston, VA.

Waters, N.M., 1998, Spatial interpolation lecture notes, (University of Calgary), available at the web site <http://www.gisca.adelaide.edu.au/kea/gisrs/gisrsrc/courses/gis/ngia/u40.html>.

Watson F.D., *Contouring: A Guide to the Analysis and Display of Spatial Data*, Pergamon Press, 1992.

Wingle, W.L., and Poeter, E.P., Evaluating subsurface uncertainty using zonal kriging. *Uncertainty '96 (ASCE)*, Wisconsin, 1996.

Wingle, W.L., Examining common problems associated with various contouring methods particularly inverse-distance methods using shaded relief surfaces. *Geotech '92*, Lakewood, Colorado, 1992.

Zirschky, J.H. and Harris, D.J., 1986, Geostatistical analysis of hazardous waste site data. *Journal of Environmental Engineering*, 112, 770-784.



APPENDIX A : TEST FUNCTIONS

Figure A.1 Function #1

$$f(x, y) = (x^2 + y - 11) \cdot (x^2 + y - 11) + (x + y^2 - 7) \cdot (x + y^2 - 7)$$

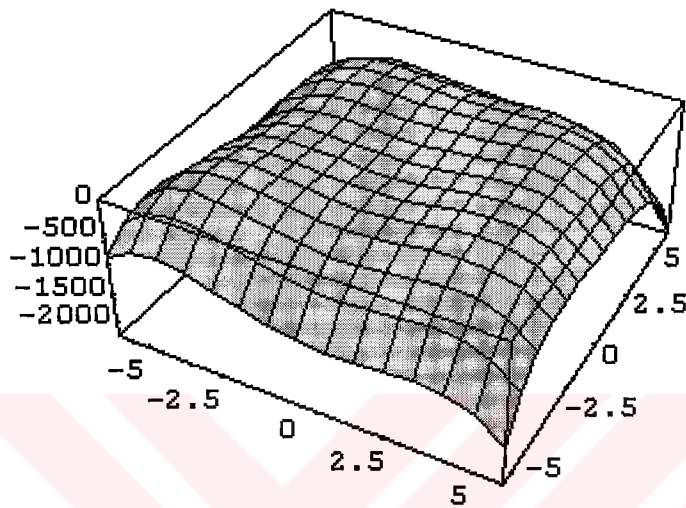


Figure A.2 Function #2

$$a_1 = \text{int}(x - 1.87) + \text{int}(y - 9.3) \quad f(x, y) = \text{int}(75.95 + 0.5) \cdot \frac{\sin a_1}{a_1}$$

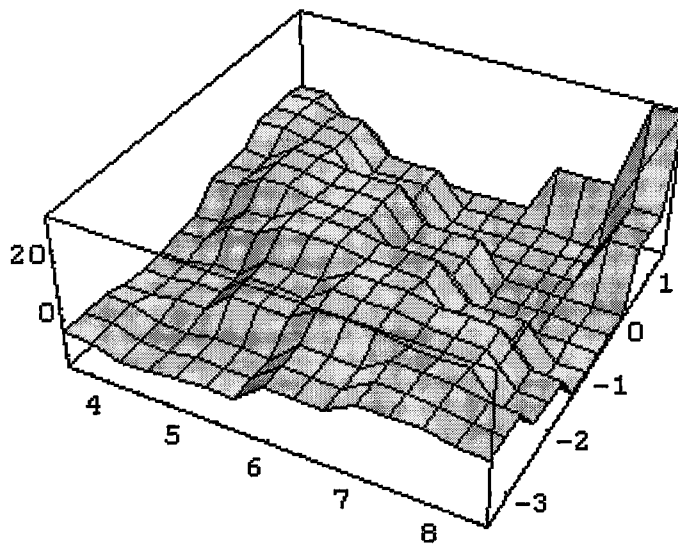


Figure A.3 Function #3

$$a_1 = (x - 1.87) + (y - 9.3) \quad f(x, y) = (75.95 + 0.5) \cdot \frac{\sin a_1}{a_1}$$

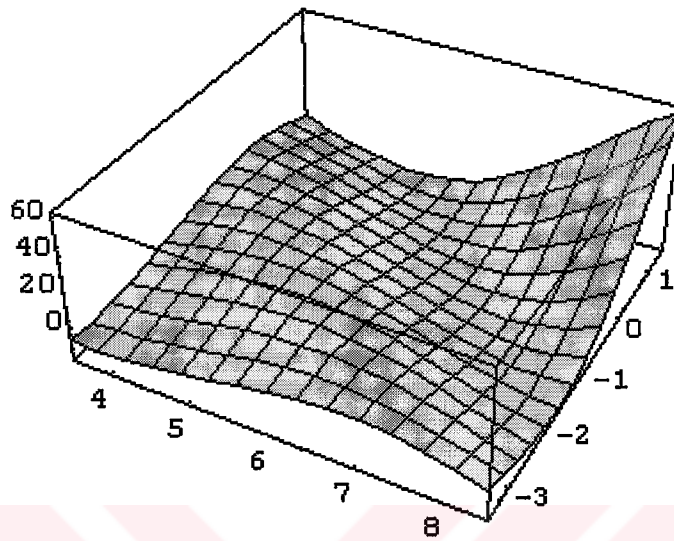


Figure A.4 Function #4

$$a_1 = \text{int}(x - 6.5) + \text{int}(y - 9.8) \quad f(x, y) = \text{int}(13.02 + 0.5) \cdot \frac{\sin a_1}{a_1}$$

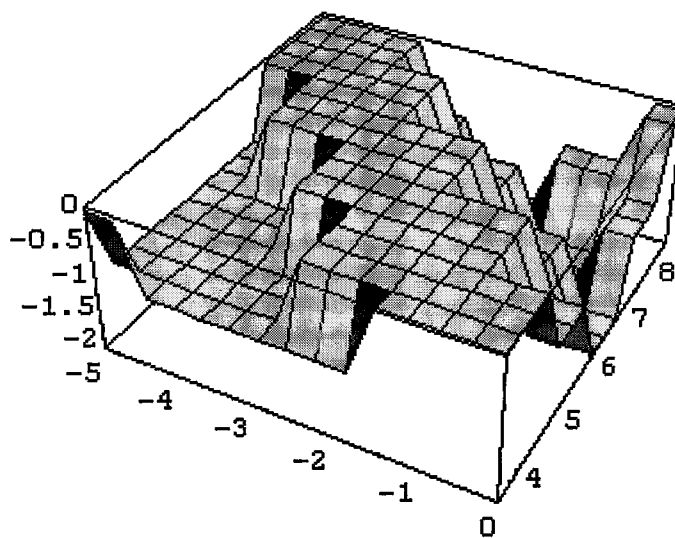


Figure A.5 Function #5

$$a_1 = \text{int}(x - 2.4) + \text{int}(y - 4.92) \quad f(x, y) = \text{int}(45.43 + 0.5) \cdot \frac{\sin a_1}{a_1}$$

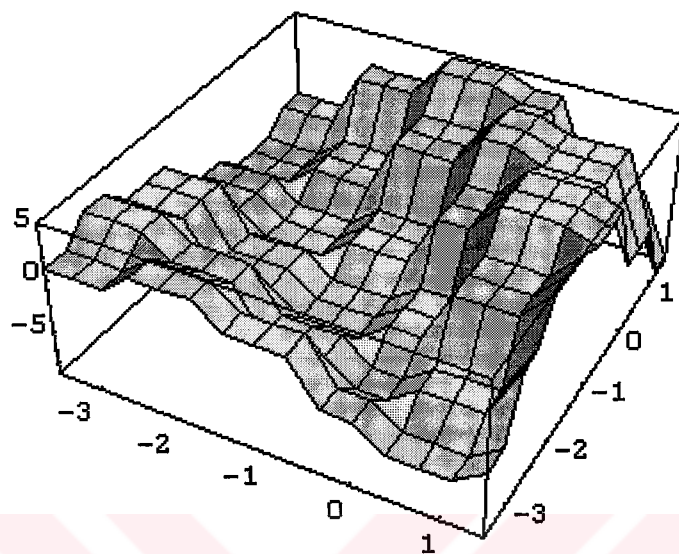


Figure A.6 Function #6

$$f(x, y) = \left[(x^2 + y^2)^{0.25} \right] \cdot \left[1 + \sin(50 \cdot (x^2 + y^2)^{0.1}) \right] \cdot \left[\sin(50 \cdot (x^2 + y^2)^{0.1}) \right]$$

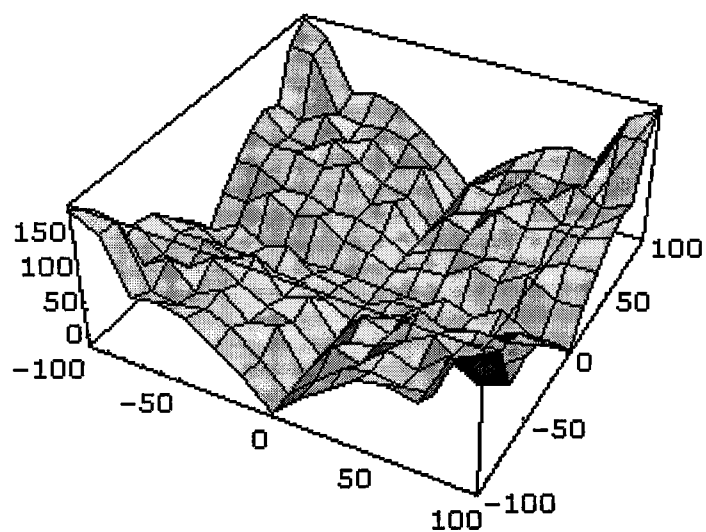


Figure A.7 Function #7

$$f(x, y) = \frac{(x+2y-7)^2}{x+1} + \sin(3 \cdot (2x+y-5) \cdot (x^2-y))$$

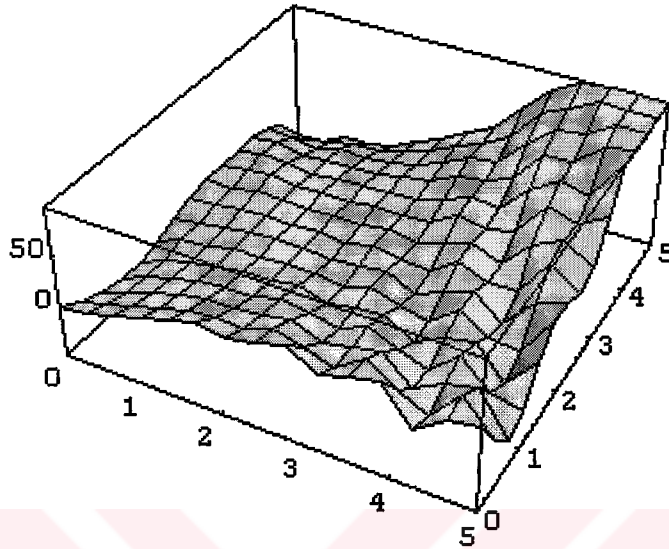
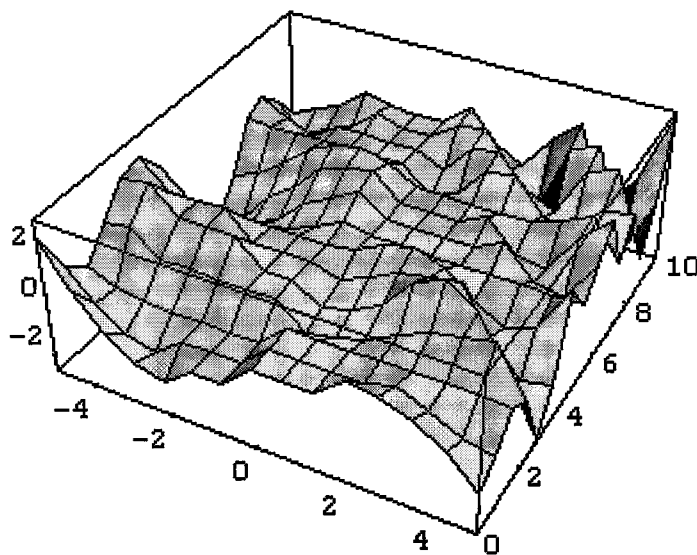


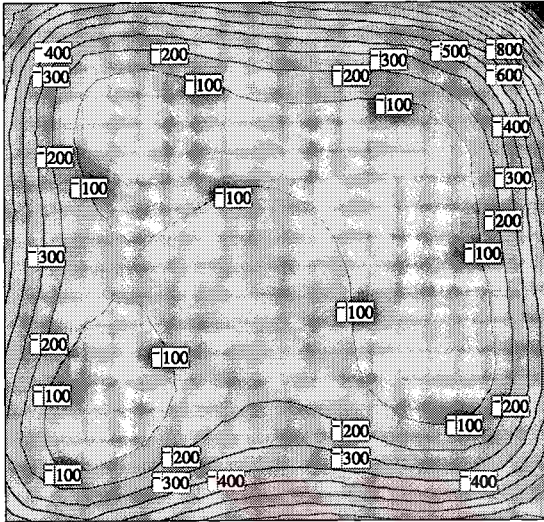
Figure A.8 Function #8

$$f(x, y) = \sin(0.1 \cdot (x+y)^2 + 1.5) \cdot \text{int}(0.5x \cdot \cos(y))$$

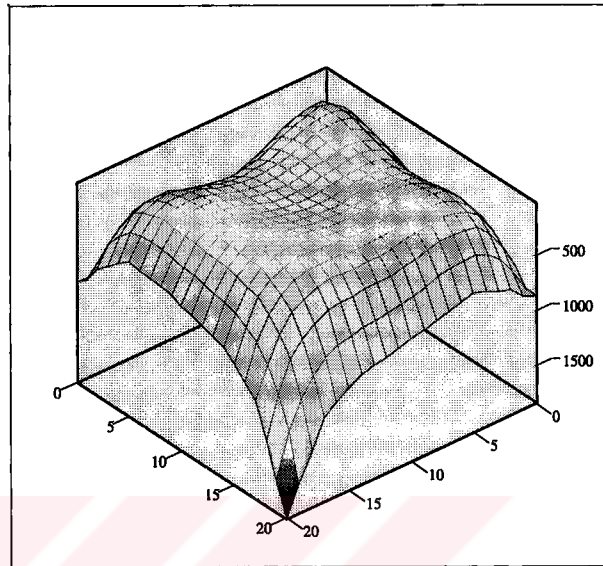


APPENDIX B: INTERPOLATION RESULTS FOR CHAPTER 6

Figure B.1 Original Function

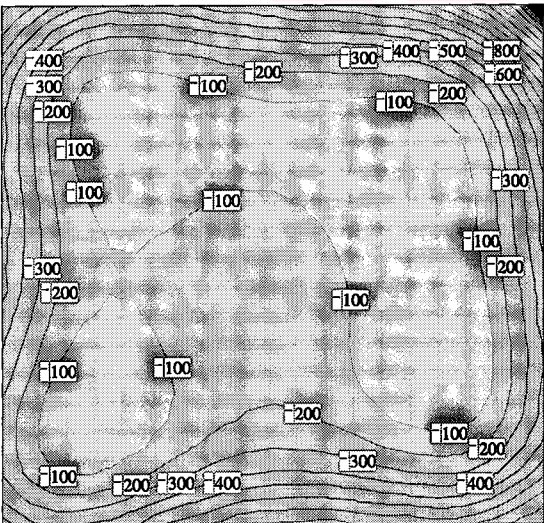


X, Y, Z

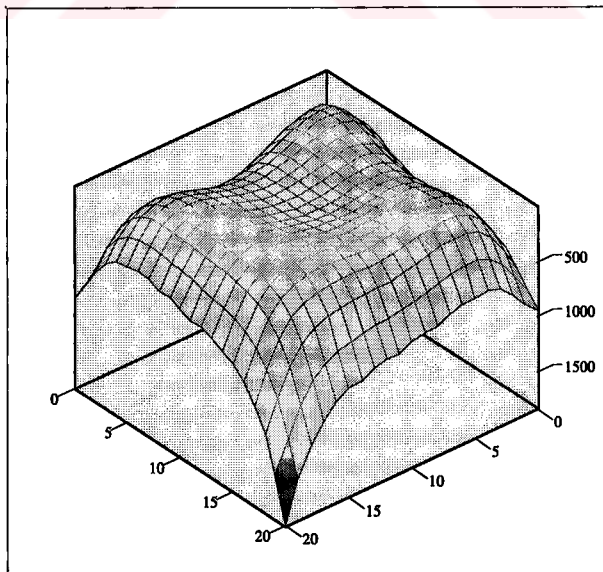


X, Y, Z

Figure B.2 (Kriging) Herring Bone sampling pattern /Noise-Free

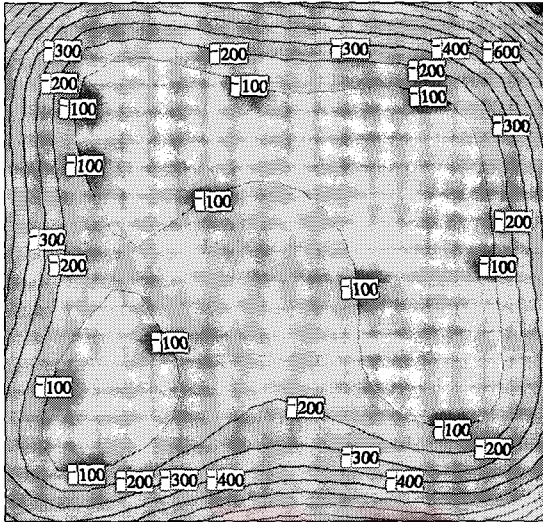


X, Y, Z

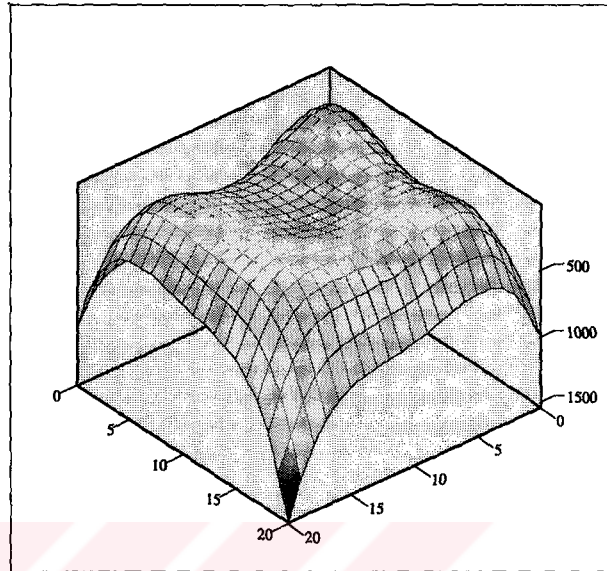


X, Y, Z

Figure B.3 (Kriging) Grid sampling pattern /Noise-Free

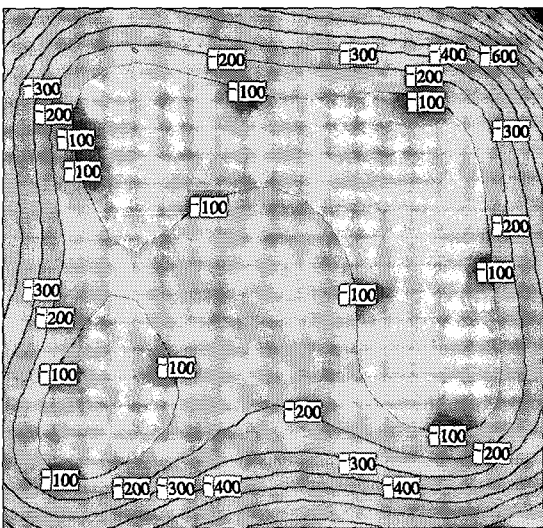


X, Y, Z

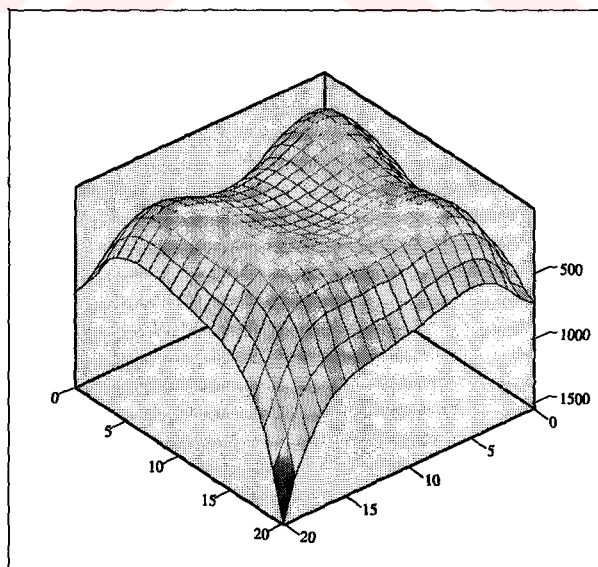


X, Y, Z

Figure B.4 (Kriging) Linear sampling pattern /Noise-Free

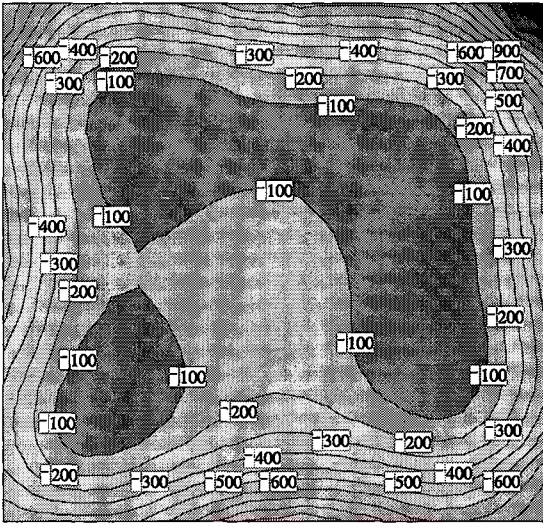


X, Y, Z

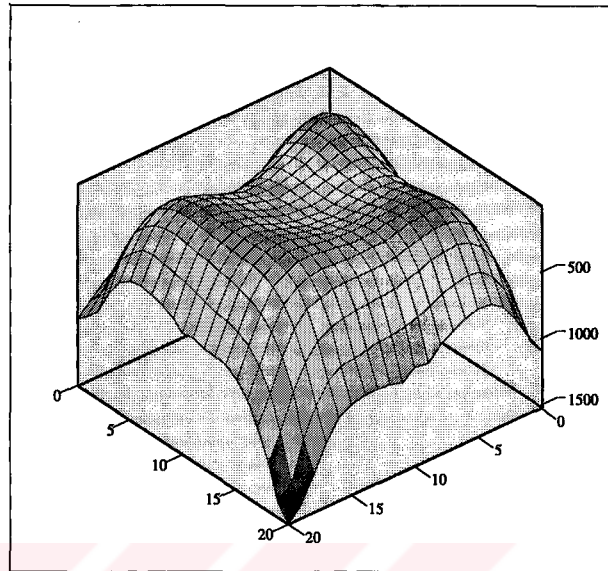


X, Y, Z

Figure B.5 (Kriging) Circular sampling pattern /Noise-Free

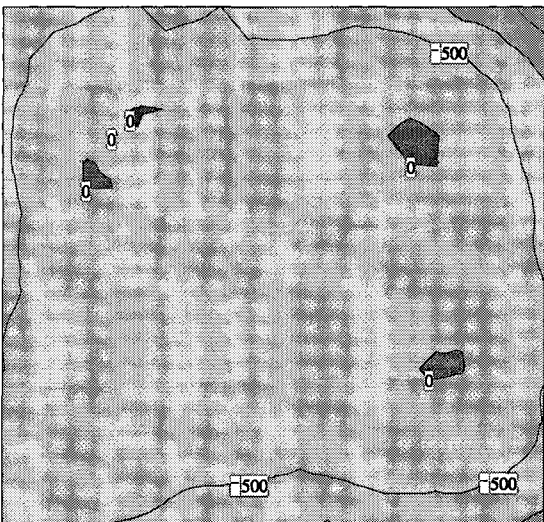


X, Y, Z

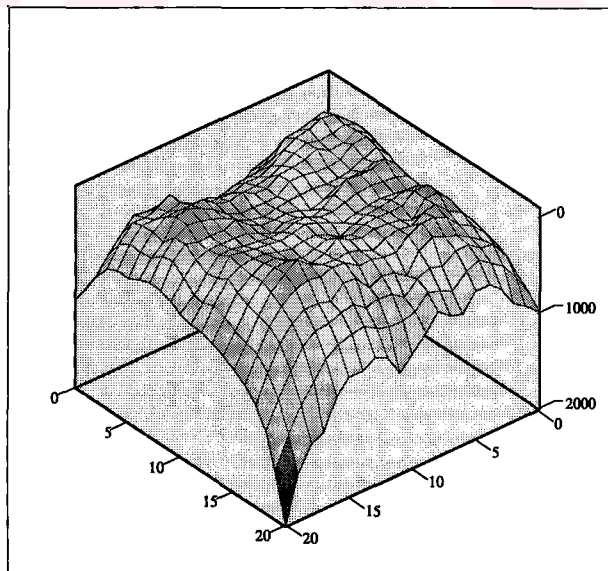


X, Y, Z

Figure B.6 (Kriging) Herring Bone sampling pattern /With noise

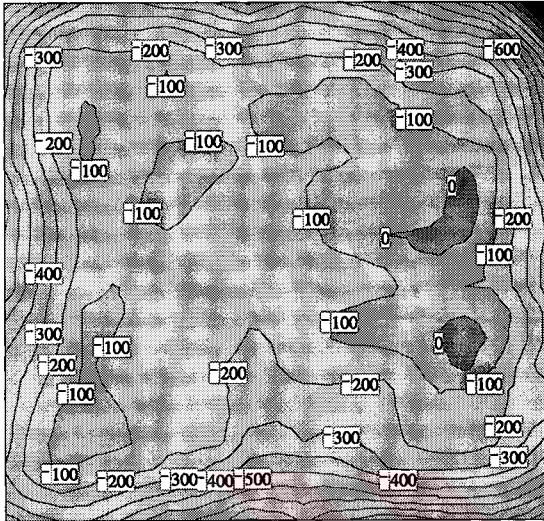


X, Y, Z

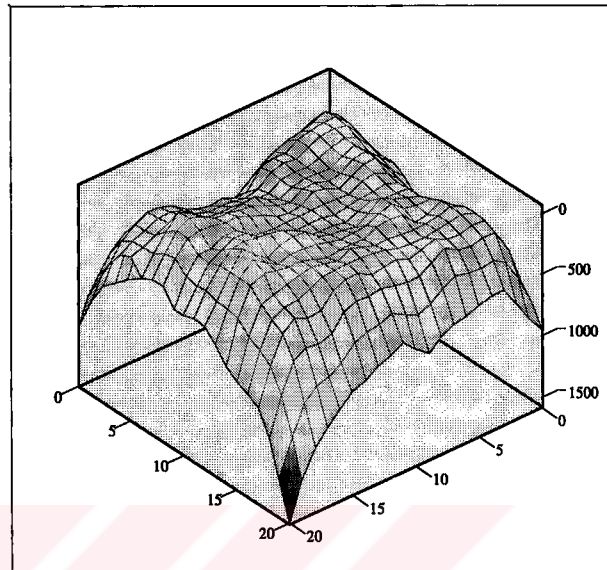


X, Y, Z

Figure B.7 (Kriging) Grid sampling pattern /With noise

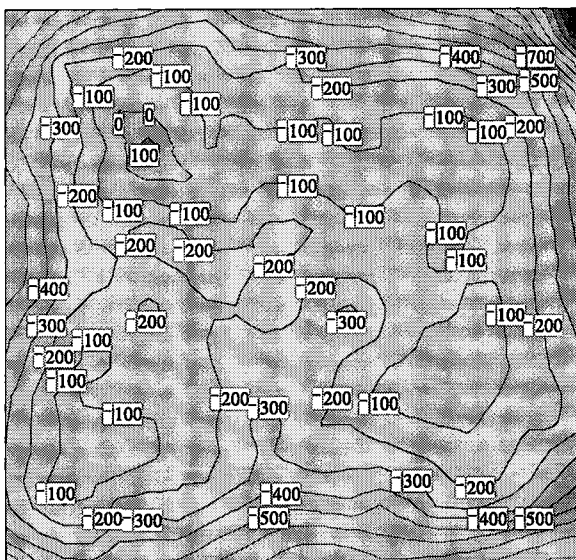


X, Y, Z

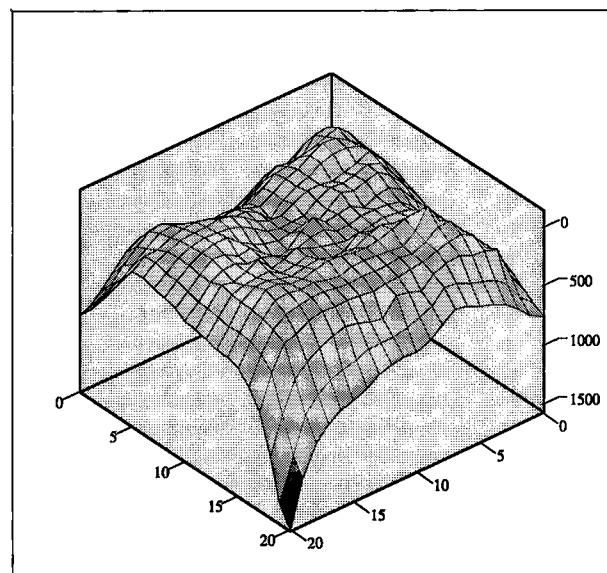


X, Y, Z

Figure B.8 (Kriging) Linear sampling pattern /With noise

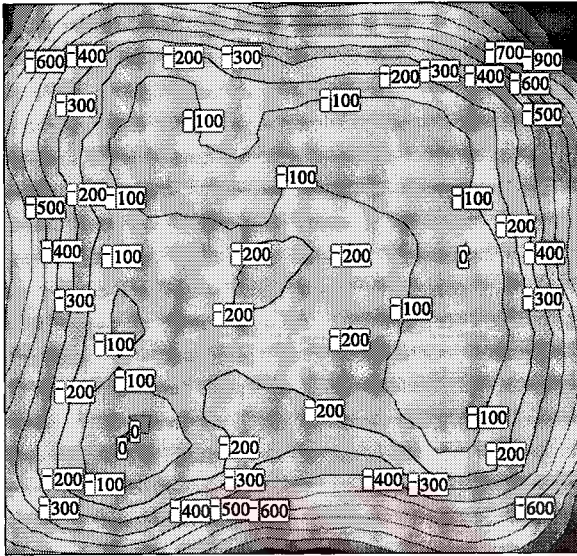


X, Y, Z

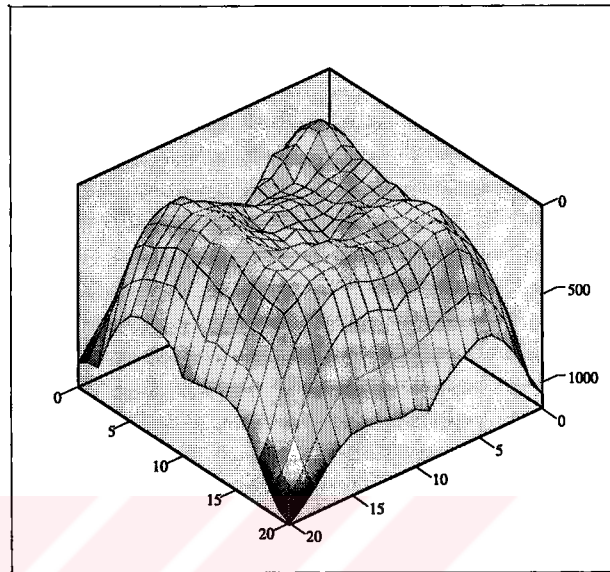


X, Y, Z

Figure B.9 (Kriging) Circular sampling pattern /With noise



X, Y, Z



X, Y, Z

APPENDIX C: SOURCE CODES FOR PERFORMANCE EVALUATIONS

During the performance evaluation steps Microsoft Visual Basic 5.0 Enterprise edition was used. The reason to choose Visual Basic among the other programming languages is its ability to perform image comparison operations. In the first part of the evaluation tests; the images of the original contamination area and the interpolated area are compared pixel by pixel in order to give the SCRAP and COVER ratios.

According to the concentration ranges: blue color is used for moderately contaminated range [150-600), and red color is used for highly contaminated range [600-1000]. By using these colors following items is obtained for performance evaluations

Table D.1 Performance evaluation items for computer program

Bb	Correctly interpolated
Br	Originally moderately contaminated but interpolated as highly contaminated
Bw	Originally moderately contaminated but interpolated as clean
Rb	Originally highly contaminated but interpolated as moderately contaminated
Rr	Correctly interpolated
Rw	Originally highly contaminated but interpolated as clean
Wb	Originally clean but interpolated as moderately contaminated
Wr	Originally clean but interpolated as highly contaminated
Ww	Correctly interpolated

For the second part of the experiments the above items are not used since in these experiments more precise evaluation method is used.

C.1 Key Map Generation And Evaluation

Rem key map generation

Dim xupper(50), xlower(50), yupper(50), ylower(50), clevel(50) as double

Dim xlowerv, xupperv, ylowerv, yupperv as integer 'define the variables

Keyfile = "c:\Thesis\1.dat" ' define target file name.

Open keyfile for input as 1

Input #1, nofarea

For i = 1 to nofarea

input #1, xlower(i), xupper(i), ylower(i), yupper(i), clevel(i)

xlowerv = int(3.83 * xlower(i))

xupperv = int(3.83 * xupper(i))

ylowerv = 384 - int(3.83 * ylower(i))

yupperv = 384 - int(3.83 * yupper(i))

for xi = xlowerv to xupperv

for yi = yupperv to ylowerv

if clevel(i) = 1 then frm360run.pickey.pset (xi, yi), 16711680

if clevel(i) = 2 then frm360run.pickey.pset (xi, yi), 255

next yi

next xi

Next i

Close 1

End sub

Rem performance evaluations

Dim xupper(50), xlower(50), yupper(50), ylower(50), clevel(50) as double

Dim xlowerv, xupperv, ylowerv, yupperv as integer 'define the variables

Open "d:\master\output.txt" for output as 2 'output file

Print #2, "d", "c", "bb", "br", "bw", "rb", "rr", "rw", "wb", "wr", "ww"

For d = 1 to 8

For c = 1 to 120

```
Mainfile = "d:\master\outwmf" + "\" + mid(str(d), 2, 1) + "\" + mid(str(c), 2, 3) + ".wmf"
```

```
Keyfile = "d:\master\location" + "\" + mid(str(c), 2, 3) + ".dat"
```

```
Input #1, nofarea
```

```
For i = 1 to nofarea
```

```
    input #1, xlower(i), xupper(i), ylower(i), yupper(i), clevel(i)
```

```
    xlowerv = int(3.83 * xlower(i))
```

```
    xupperv = int(3.83 * xupper(i))
```

```
    ylowerv = 384 - int(3.83 * ylower(i))
```

```
    yupperv = 384 - int(3.83 * yupper(i))
```

```
    for xi = xlowerv to xupperv
```

```
        for yi = yupperv to ylowerv
```

```
            if clevel(i) = 1 then frm360run.pickey.pset (xi, yi), 16711680
```

```
            if clevel(i) = 2 then frm360run.pickey.pset (xi, yi), 255
```

```
        next yi
```

```
    next xi
```

```
Next i
```

```
Close 1
```

```
Set frm360run.picmain = loadpicture(mainfile)
```

```
Bb = 1
```

```
Br = 1
```

```
Bw = 1
```

```
Rb = 1
```

```
Rr = 1
```

```
Rw = 1
```

```
Wb = 1
```

```
Wr = 1
```

```
Ww = 1
```

```
Rem comparison
```

```
For i = 1 to 383
```

```
    for j = 1 to 383
```

```
        if frm360run.picmain.point(i, j) = frm360run.pickey.point(i, j) then
```

```
            if frm360run.picmain.point(i, j) = 255 then rr = rr + 1
```

```

    if frm360run.picmain.point(i, j) = 16711680 then bb = bb + 1
    if frm360run.picmain.point(i, j) = 16777215 then ww = ww + 1
end if
if frm360run.picmain.point(i, j) = 16777215 then
    if frm360run.pickey.point(i, j) = 16711680 then wb = wb + 1
    if frm360run.pickey.point(i, j) = 255 then wr = wr + 1
end if
if frm360run.picmain.point(i, j) = 16711680 then
    if frm360run.pickey.point(i, j) = 16777215 then bw = bw + 1
    if frm360run.pickey.point(i, j) = 255 then br = br + 1
end if
if frm360run.picmain.point(i, j) = 255 then
    if frm360run.pickey.point(i, j) = 16777215 then rw = rw + 1
    if frm360run.pickey.point(i, j) = 16711680 then br = br + 1
end if
next j
Next i
Print #2, d, c, bb, br, bw, rb, rr, rw, wb, wr, ww
Frm360run.pickey.cls
Next c
Next d
Close 2
End
End sub

```

C.2 Herringbone Sample Pattern Code

```

Rem herringbone pattern
Dim xlower, xupper, ylower, yupper, j as double
Dim xint, yint as double
Dim nsmp as integer
Rem open output file

```

```

Open "c:\Thesis\1.txt" for output as 1
Rem open setup file for the function
Open "c:\Thesis\config.txt" for input as 2
Rem upper and lower bounds
Input #2, xlower, xupper, ylower, yupper
Nsmp = 200 'number of sample
J = int(sqr(nsmp))
Xint = (xupper - xlower) / j ' x interval
Yint = (yupper - ylower) / j ' y interval
Say = 0
For k = 0 to j - 1
  for g = 0 to j - 1
    Gd = g mod 2
    Kd = k mod 2
    If gd = 0 and kd = 0 then
      x = xint * 0.5 + g * xint + xlower
      y = yint * 0.5 + k * yint + ylower
    End if
    If gd = 0 and kd <> 0 then
      x = xint * 0.75 + g * xint + xlower
      y = yint * 0.5 + k * yint + ylower
    End if
    If gd <> 0 and kd = 0 then
      x = xint * 0.5 + g * xint + xlower
      y = yint * 0.75 + k * yint + ylower
    End if
    If gd <> 0 and kd <> 0 then
      x = xint * 0.75 + g * xint + xlower
      y = yint * 0.75 + k * yint + ylower
    End if
    F = fx(x, y) 'get the function value
    Print #1, x, y, f
  
```

```

Say = say + 1
Next g
Next k
Rem print #1, say
Close 1
Close 2
End sub

```

C.3 Regular Grid Sample Pattern Code

```

Rem regular grid
Dim xlower, xupper, ylower, yupper, j as single
Dim xint, yint, x, y as single
Dim nsmp as integer
Open "c:\Thesis\2.txt" for output as 1
Open "c:\Thesis\config.txt" for input as 2
Input #2, xlower, xupper, ylower, yupper
Nsmp = 200 'number of sample
J = int(sqr(nsmp))
Xint = (xupper - xlower) / j ' x interval
Yint = (yupper - ylower) / j ' y interval
Say = 0
For k = 0 to j - 1
    for g = 0 to j - 1
        x = xint * (0.5 + g) + xlower
        y = yint * (0.5 + k) + ylower
        f = fx(x, y)
    Print #1, x, y, f
    Say = say + 1
    Next g
Next k
Close 1

```

Close 2

End sub

C.4 Linear Sample Pattern Code

Rem linear

Dim xlower, xupper, ylower, yupper, j as double

Dim xint, yint as double

Dim x, y, deltaxy, intv as double

Dim nsmp as integer

Open "c:\Thesis\3.txt" for output as 1

Open "c:\Thesis\config.txt" for input as 2

Input #2, xlower, xupper, ylower, yupper

Say = 0

Nsmp = 100 'number of sample

J = int(sqr(nsmp))

Intv = (xupper - xlower) / j 'length of the interval

For k = 1 to j

for a = 1 to k

deltaxy = k * intv / (k + 1)

x = a * deltaxy + xlower

y = (k - a + 1) * deltaxy + ylower

f = fx(x, y)

print #1, x, y, f

say = say + 1

next a

Next k

For k = 1 to j - 1

for a = 1 to k

deltaxy = k * intv / (k + 1)

x = xupper - a * deltaxy

y = yupper - (k - a + 1) * deltaxy

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

```

    print #1, x, y
    say = say + 1
next a
Next k
Close 1
Close 2
End sub

```

C.5 Circular Sample Pattern Code

```

Rem curcular
Dim xlower, xupper, ylower, yupper, j as double
Dim xint, yint as double
Dim x, y, deltaxy, intv as double
Dim nsmp as integer
Open "c:\Thesis\4.txt" for output as 1
Open "c:\Thesis\config.txt" for input as 2
Input #2, xlower, xupper, ylower, yupper
Say = 0
Nsmp = 100 'number of sample
Ncyc = 9
Rlx = (xupper - xlower) / 2 ' radius of the mr
Rly = (yupper - ylower) / 2 ' radius of the mr
For cycle = 1 to ncyc
Rkx = rlx / ncyc * cycle
Rky = rly / ncyc * cycle
For k = 1 to cycle * 2
Startangle = (360 / ncyc) * (cycle - 1)
Q = (startangle + (k - 1) * 360 / (cycle * 2)) * 3.14 / 180
    x = xlower + rlx + rkx * cos(q)
    y = ylower + rly + rky * sin(q)
    say = say + 1

```

```

    f = fx(x, y)
Print #1, x, y, f
Next k
Next cycle
T = (rlx * sqr(2) - rlx) / sqr(2)
For k = 1 to 2
Rem bottom-left
X = rnd * t + xlower
Y = rnd * t + ylower
F = fx(x, y)
Print #1, x, y, f
Rem top-left
X = rnd * t + xlower
Y = yupper - rnd * t
F = fx(x, y)
Print #1, x, y, f
Rem top-rigth
X = xupper - rnd * t
Y = yupper - rnd * t
F = fx(x, y)
Print #1, x, y, f
Rem bottom-rigth
X = xupper - rnd * t
Y = rnd * t + ylower
F = fx(x, y)
Print #1, x, y, f
Next k
Rem print #1, say
Close 1
Close 2

```

C.6 Gaussian Error Generator Code

```

Dim x(1000), y(1000), fx(1000), fxe(1000) as double
For k = 1 to 4
    Erperc = 0.05
    Minfx = 100000
    Maxfx = 0
    I = 1
    A = "c:\Thesis\" & k & ".txt"
    B = "c:\Thesis\" & k + 4 & ".txt"
Open a for input as 1
Open b for output as 2
For i = 1 to 196
    Input #1, x(i), y(i), fx(i)
Next i
N = i - 1
For i = 1 to n ' n is the number of data points
    if fx(i) > maxfx then maxfx = fx(i)
    if fx(i) < minfx then minfx = fx(i)
Next i
Range = maxfx - minfx
For i = 1 to n
    sum = 0
    for m = 1 to 11
        sum = sum + rnd
    next m
    fxe(i) = fx(i) + (sum - 6) * range * erperc
    print #2, x(i), y(i), fxe(i)
next i
Close 1
Close 2
Next k

```