KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KUMASI

COLLEGE OF SCIENCE

EFFECT OF SAMPLE SIZE AND SAMPLING SCHEME ON VARIOGRAM UNCERTAINTY

BY

BENJAMIN TOMMY BAVUG

(BSC. STATISTICS)

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF PHILOSOPHY (MATHEMATICAL STATISTICS)

June 23, 2024

Declaration

I hereby declare that this piece of submission is my own work towards the award of Master of Philosophy degree.

To the best of my knowledge, it contains no materials previously published by another person nor material of this kind, which has been presented for the award of any degree either in this University or elsewhere, except where due acknowledgment has been made in the text.

Benjamin Tommy Bavug	.....	.....
Student	Signature	Date

Certified by:

S. K. Appiah (PhD)	.....	.....
Supervisor	Signature	Date

Certified by:

Prof. A. O. Adebanji	.....	.....
Head of Department	Signature	Date

Dedication

I strongly dedicate this material to my family, friends and love ones for their kind support in many ways, making this work a reality. May the Almighty Father richly bless you in thousand folds.

Abstract

Uncertainty of variogram refers to variogram inaccuracy, error and imprecision. The impurity surrounding variogram estimates largely originates from imperfection in spatial sampling and its datasets. Stochastic fields come in and invariably mimic reality but it is impossible to achieve perfect picture of the world without losing some information. This research assumes the variogram parameters estimates are unbiased with minimum variation, implying a change in measurement variations among these parameters afterwards would have come from the choice of sampling scheme and samples size selected. Recognizing this effect, the sampling scheme is then varied with the sample size from spatial autocorrelated data simulated from sequential Gaussian simulation(SGS) technique on a stochastic field. Parameters of Gaussian variogram model were chosen and by method of sequential Gaussian simulation, 10,000 data points were simulated from stochastic field. Out of the population of size 10,000, a number of sample sizes are varied with the scheme of sampling whiles the variogram estimates recorded were compared with the variogram estimates of the population to quantify the effect of sample size and sampling scheme on variogram uncertainty. The outcome of this thesis will aid in making informed decisions in many enterprises of life. A typical case is the mining industry where miners take prudent decisions regarding the number of drills and what drilling scheme(strategy) to adopt in order to hit goal deposit(s) with minimum cost.

Acknowledgment

I attribute the successful completion of this thesis to Almighty God for His traveling mercies.

Dr. S.K. Appiah and Dr. Eric Nimako Aidoo, my supervisors, who supported, encouraged, supervised and made useful suggestions throughout this study. I extend my heartfelt wishes of gratitude to them.

I am also highly thankful to my family, friends and peer reviewers for their valuable suggestions throughout this study. Lastly, to all those who contributed in diverse ways towards this thesis I say God richly bless you.

Contents

List of Tables

List of Figures

List of Abbreviations

CRS Cluster Random Sampling

DSSIM Direct Sequential Simulation

GSLIB Geostatistical Software Library

OK Ordinary Kriging

SGS Sequential Gaussian Simulation

SGSIM Sequential Gaussian Simulation

SIS Sequential Indicator Simulation

SK Simple Kriging

SRS Simple Random Sampling

STRS Stratified Random Sampling

UK Universal Kriging

Chapter 1

INTRODUCTION

1.1 Introduction

Uncertainty of variogram refers to variogram inaccuracy, error and imprecision. The impurities surrounding variogram estimates come from imperfection in spatial datasets. Stochastic fields invariably mimic reality but it is impossible to achieve perfect picture of the world without losing some information.

The research assumes that, the variogram parameters estimates are unbiased with minimum variation, implying a change in measurement variations among these parameters afterwards would have come from the choice of sampling scheme or samples size selected.

Assumptions of spatial statistical data are but slightly different from the traditional statistical data. The assumption of independence is violated in geographically referenced data (spatial data) since there will not be meaningful spatial analysis without the assumption of spatial dependence among the data points. However, the sampling scheme remains unchanged whether traditional statistical or spatial data.

1.2 Background of the study

This chapter includes explanations of the various mathematical ideas used throughout this paper. It provides a basic overview of the structures built across this page to get a better understanding of the mathematical formats here. Sample is a sub-category of a whole, and sampling include procedures for the selection of data points from people or objects for the purpose of measuring the characteristics of the general population. The characteristic may be a mean value or variance of a stochastic field Christakos (2012), or values on unsampled areas Goovaerts et al. (1997), or targets points Rogerson et al. (2004). Sampling scheme is the design for which sample selection from a population is conducted for measuring a stochastic phenomenon. Sampling design can be probabilistic or deterministic but the research is focused on only the probabilistic sampling schemes.

Some of the probabilistic sampling schemes include simple random sampling, systematic random sampling, cluster random sampling, stratified random sampling schemes. Variogram is about the graphic representation of the variance of any geographical data. The Variogram explains the structure of the spatial correlation and its significance to any geostatistical research. It also describes the variance of the difference between observations of variables as a function of separation distance. The quantification of uncertainty of any variogram is important for kriging or spatial parameter estimations.

Webster and Wiley (2001) show that variogram measurement, which is expressed as a mathematical function, is required to mimic the dependent variables performed spatially. Both interpolation and simulation techniques assume that the variogram estimate is known on the entire field when practically, the variogram is actually measured from experimental data and thus creates the inevitable uncertainty associated with variogram estimates.

For the purposes of two-dimensional sampling scheme, Let's assume to have a sample size n sampled over a study area D. If an attribute Z is sampled on n supports giving some observations $\{z(s_\alpha)\alpha = 1, 2, \cdots, n\}$. This analogy is employed in the discussion of

the following major sampling schemes.

1.2.1 Simple Random Sampling

In this type of design, a random selection of set n of sample points is made on the space D with every location in D being giving an equal chance of being included. Each size n of the samples has the same chance of being selected. The selection may be conducted with or without replacement. This type of design is easy and has a very low operational cost. However, in terms of spatial sampling, it is not robust. Griffith and Amrhein (1997) also opined that, the sampling design maybe overshadowed with under-sampling and oversampling in some sampling areas and that makes the distribution of the points not representative enough for most samples drawn.

1.2.2 Systematic Random Sampling

According to the systematic sampling, the interest is that the population or random field be divided between periods of equal proportion n from N population. In this selection process, the first unit is selected randomly or intentionally within the first interval limit, and the remaining $n-1$ units are then selected based on k^{th} intervals. This is achieved by adding periods k until the actual elements are selected, going to the end of the list and continuing to the beginning. Systematic sampling is probably the most well-known selection process. It is commonly used and easy to apply and is sometimes referred to as "pseudo-random selection". In your interval k, select k that N is greater than nk but less than $(n+1)k$.

1.2.3 Stratified Random Sampling

This type of sampling ensures that the population is geographically divided into groups and that there is homogeneity inside each group and heterogeneity between small groups or groups. The so-called "strata" or small populations are not intersected and thus, cover the entire population Cochran (2007). A simple random sample or process used for sample selection is applied in each smaller population. By dividing the population into smaller

segments or strata, the heterogeneity in each collection is reduced and it is easier to collect more closely representative samples Wang, Haining, and Cao (2010). Non-overlapping points make up the total the total entire field, that is $N_1 + N_2 + \cdots + N_K = N$. A simple random drawing with or without replacement is made in each row independently so that the sample size in the i^{th} stratum is displayed as follows: $n_1 + n_2 + \cdots + n_k = n$, where n is the total sample size. To avoid inconvenience in field stratification, convenience in spatial location is largely taken into consideration in dividing the population. For example, in the study of the abundance of certain species of fish on four farms, each category of ponds can be considered as a separate entity.

1.2.4 Cluster Random Sampling

With this sampling method, a set or set of sample frames is randomly selected a process that ensures that the entire collection is included in the sample. To illustrate a collection sample, we look at an interview in which the sample units are Oak trees in the Ghanaian forest arrangements. It is clear that if a simple random sample is used to select Oak trees, they will be distributed in several forests across this range and across the country.

However, the forest can be considered a Category(cluster) of Oak Trees. We select a few forests using simple random sampling technique and include all Oak trees in the selected forest in our sample. Such a sampling process would be a Cluster sampling model. It has many advantages over simple random sampling. Considering the example above, it will be less expensive to adopt a cluster sample than a simple random sampling sample due to the spread of spatial data points in the study area.

1.3 Problem Statement

One spatial interpolation assumption indicates that the mean is known throughout the stochastic region but in practice it is always calculate from the study region. Just as the jury selection process affects the outcome of a trial, so does the sampling scheme influences the results of the study and that, arising from where the samples were taken

are complex and irregular, leading to serious problems to be overcome in mathematical analysis, Carlson and Ripley (1997).

A typical scenario is the number of drills to be made and the drilling scheme to adopt in gold mining greatly influence the quantity of gold deposits miners will obtain with a certain margin of error. This eventual creates uncertainty on estimates from variogram models due to the presence of error. The size of the sample and the sampling scheme coupled with the unavoidable presence of spatial autocorrelation among spatial data could have some implication on uncertainty of variogram estimates. The variance error measurement may be calculated incorrectly with respect to sample design selection and sample size ((B. Ripley, 1984); (Christakos, 2012)).

The research recognizes the various concepts and arguments put up by some writers regarding the influence sample size and sampling scheme could impose on variogram estimates and will go extra mile to conduct stochastic simulation algorithm (sequential Gaussian simulation) study to quantify the effect of sample size and sampling scheme on variogram uncertainty.

1.4 Objective of the Study

The research will unearth sample size and sampling scheme effect on variogram uncertainty.

1.4.1 Specific Objectives

The specific objectives of the study are:

- To examine the effect of sample size on variogram uncertainty.
- To examine the effect of sampling scheme on the variogram uncertainty.

1.5 Justification

In other words, the Variogram is undoubtedly the result of random field variability, sample distribution and demographic figures. Could the variability emanate from different factors? Some interpolation techniques state that the location value is known throughout the learning region where practically, it is actually calculated from the experimental data sampled in the study area. This assumption creates uncertainty in estimates from variogram models. Alternatively, just as the jury selection process affects the outcome of a trial, so does sample size and scheme influence the results of the study and that, from where the samples are taken are complex and irregularly shaped, leading to difficult problems to overcome in mathematical analysis, B. D. Ripley (1977). There is therefore a need to study and model the measurement strategies and sample sizes to assess the effect of variogram variability. The result to be obtained from this simulation study may serve as a guide for future researchers in terms of sample size and sample sampling to reduce variability in variance measurement variability if not completely eliminated.

1.6 Methodology

1.6.1 Problem at Hand

Variogram uncertainty undoubtedly emanates from spatial variability at the random field, the distribution of sample and the statistic that yields the population estimate. Most importantly, the variogram uncertainty is estimated by the design technique (scheme) or the model approach of which both are deemed feasible for studies in stochastic field and population studies.

The ultimate goal of a geostatistical study is to obtain minimum variance unbiased estimate from the stochastic field or the population. Intuitive realization of the goal of spatial statistical study necessitated the need to ensure adequate allocation of sample points in the study area that will maximized precision. Identifying the choice of sample

size and sampling scheme could improve precision of variogram estimates.

1.6.2 Model for the Problem

The research employs designed based sampling approach that is based on sequential Gaussian simulation from a stochastic field or population of an area. Anyone of these licit (Gaussian, Exponential, and Spherical) variograms will be fitted into the sample points obtained from the simulation with known parameters. The effect of these sample sizes and sampling schemes are further examined on the model estimates to measure the uncertainty associated with the variogram.

Since the goal of spatial data interpolation strategies is to establish a variogram estimate that is unbiased, and has minimal error term with lower degree of uncertainty and ensures greater precision, it will be prudent and laudable simulating spatially correlated data from a random field and then vary these simulated samples from the selected designs to estimate effect the sample sizes and sample schemes have on the uncertainty of variogram estimates.

1.6.3 Source of Data

The research will adopt the Sequential Gaussian Simulation Algorithm to model spatial distribution of data samples and use appropriate design schemes for the spatial sampling. Rstudio is used for the geostatistical data simulation for the design modelling.

1. Sampling scheme is held constant whiles varying sample size on variogram model to detect its effect on the uncertainty of variogram (e.g on anyone of these models; exponential, spherical and the Gaussian models).

2. At this stage, Sample size is also held constant whiles sampling scheme is varied on variogram to ascertain the effect of the scheme of sampling on the uncertainty of variogram (e.g on anyone of these models; exponential, spherical and the Gaussian models).

1.7 Thesis Organization

The thesis spans from chapter one to chapter five, chapter one elaborates the research background, problem statement, objectives, methodology and justification. The chapter two discusses review and comparative study of literature on sample size and sampling scheme effect on variogram uncertainty whereas chapter three deals with the heuristic illustration of the design and model based methodology. Chapter four discusses the analysis of the simulated data and the results obtained whereas chapter five talks about conclusion and recommendation.

1.8 Limitation

This research adopts sequential Gaussian simulation approach to acquire random and spatially correlated data to be used to ascertain the effect of sample size and sampling scheme on variogram uncertainty. The simulation study is focused on sampling scheme and sample sizes. It will not cover variogram parameter estimations, it rather examines the behaviors of the model parameters (variogram behavior) when different sampling schemes and sample sizes are chosen from a study area.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

Sampling involves selecting or choosing part or whole of a population or a stochastic field with the aim of estimating certain characteristic of the population or the stochastic field. A proportion of objects or individuals selected from the population is call sample size. Statistical procedure employed in the selection of the sample sizes is called the sampling scheme.

These factors can be the total or mean value of a random field (Christakos (2013)), or values in the unsigned areas Goovaerts et al. (1997), or targets (s) Rogerson et al. (2004). Zhang et al. (2009) highlighted that, the effect of sampling with the accuracy of the sample can arise from fields independently and automatically generated from experimental data and hence kriging is recommended as the best way to interpret point data as error difference is reduced by the combined weight of data aggregation.

2.2 Spatial Structure and Sampling Configurations

One common goal for spatial and non-spatial designs is to configure a sampling scheme that has little or no variation associated with the estimation. Paramount in the scheme of the design is the location of the samples which largely depends on the compositional structure of the variable in question. Whenever there is an existence of certain underlying

attributes, it may be important to undertake stratification of sampling scheme of the study area whether such a design is spatial or non-spatial.

Surprisingly, the underlying variation of the attributes is most often unknown, which contradicts to some extend the objective of the design to achieve an optimal sampling scheme that will give maximum information. This scenario necessitates the call for a strategic scheme of sampling that will still retain spatial variability. The discussion seemingly agrees with Delmelle, Dony, Casas, Jia, and Tang (2014) in his handbook illustration, that spatial variability will not be captured if we under-sample and oversampling also brings redundancy in the data points, and that attention must be paid on not only quantity of the data but also the locations. Anderson, Sethajintanin, Sower, and Quarles (2008) also added that most accurate estimates largely come from most efficient sampling scheme.

Delmelle et al. (2014) also finds that, the combined sampling system is less accurate than the random samples in which the guess differences from both schemes are compared; its moderate performance is a growing function of monotonic sample size compared to a random sample. Ideally, the size of the sample area should increase in areas that exhibit high spatial variability because the values of the closest samples will show strong similarities and may not exceed the sample in those areas. The spatial autocorrelation function summarizes the numerical similarities of interest variations in different sample locations, as a function of their classification distance (Gatrell (1979); Griffith (1987)).

According to Quenouille et al. (1949), where autocorrelation is a function that reduces the distance, the spatially separated samples have less variability than the fixed range. If the decrease in autocorrelation is not linear, yet it holds high, the formal sample is more accurate than the random sample, and the fixed design, where each point falls exactly between each interval, is much more efficient than the random sample configuration (Madow (1953); Zubrzycki (1958); Dalenius, Hájek, and Zubrzycki (1961); Bellhouse (1977);Iachan (1985)). When the process is isotropic, the fixed-angle triangle formation will maintain a small difference, since it reduces the distance too far from the original sample models to the non-visited points. The square grid performs well, especially

in the case of isotropy (McBratney, Webster, and Burgess (1981); Webster and Oliver (2007);Delmelle et al. (2014)). When anisotropy is present on the other hand, the square grid pattern is favored in the hexagonal arrangement, even though the improvement is small Olea (1984).

According to Van Groenigen, Pieters, and Stein (2000), the growing model produces a suspension of point-symmetric samples similar to the linear model. However, the use of a Gaussian model often obtains sample points very close to the D boundary. According to (Delmelle et al. (2014);Olea (1984)), sample reduction in the existing space network is a problem related to sampling and is appropriate in many regions of the world where the cost of environmental monitoring is limited. The process involves reducing the number of samples needed to reach the accuracy level. Technically, it consists of selecting the available samples from the original data set that will, in conjunction with the local algorithm, generate a much better estimate of the variance of the results such as the results obtained if all sample points were used.

Usually, it is assumed that the residues are from the drying process, and that the covariogram is reduced precisely, without the nugget effect, and that the process is not isotropic. A review of studies conducted to predict soil water content, Ferreyra, Apezteguıa, Sereno, and Jones (2002) developed a similar sample reduction method, ranging from 57 observations to 10 observations. With a good order of 10 samples, more than 70% of the predicted water content has an error within $\pm 10\%$, indicating that the same level of confidence is achieved with a limited number of samples.

2.3 Sequential Gaussian Simulation (SGS)

SGS is an algorithm that works with node sequences, and later uses the same values as status data. It is necessary to use standard Gaussian values in the SGS method; Thus, the information is transformed into a Gaussian space. The basic steps in the SGS algorithm according to Deutsch, Journel, et al. (1992) are listed below:

1. Calculate the histogram of the raw data, and the statistical parameters.

2. Convert the data into a Gaussian space.

3. Also calculate the model of variance in Gaussian data.

4. Define a grid.

5. Choose a random method.

6. Calculate the value for each location from all other values (known and performed) and define Gaussian.

7. Draw a random value from a Gaussian distribution known as the generated value.

8. Imitating other places in order.

9. . Price Backtransform pricing (in this step refocus recognition).

10. To make some more sense, steps 1 through 9 are repeated.

According to Soltani, Afzal, and Asghari (2013), sequential estimation is a stochastic algorithm that gains more information based on the same input data. This data may be continuous or fragmented. Regarding the type of data, the linear regression, the Gaussian linear regression (SGS) or the linear regression model will be used. The most precise algorithm for reconstructing a multi-dimensional Gaussian field is given by the sequence principle. SGS requires standard Gaussian data with zero zero and unit variance, so at SGS, the data is converted to Gaussian using the quantile transformation. Each variation is made in proportion to its standard ccdf with a simple Kriging program. The positioning data contains all the original data and all initial values are obtained from the neighboring location of the generated state. A conditional simulation of a continuous z (u) transition in a Gaussian space. Backtransform standard pricing created for the original unit. With respect to the transition to Gaussian and then back to the original unit, mathematical fluctuations are subject to mimicry but variables must be discrete and quantitative and quantitative. The following tests should be performed after all nodes have been made: reproducibility of (1) data values in data areas, (2) original histogram, (3) original summary statistics, and (4) covariance input model.

There are many stochastic simulation algorithms but Sequential Gaussian Simulation (SGS) among them is widely used because of its speed and precision in selecting spatial data. The basic stochastic simulation is designed to overcome the smooth prediction effect of kriging, which is especially important when focusing on equations in sharp or excessive map construction. Simulation algorithms look for both spatial and temporal variability of real data in sampled locations and variance estimates in unmixed locations. It means that stochastic simulation reproduces the sample statistics (histogram and semi-variogram model) and respects the sample data in their original locations. Therefore, the stochastic simulation map represents the spatial distribution of the logical attribute over the selected map ? (?).

Stochastic simulation is a widely accepted tool in various geostatistics. The goal of the stochastic simulation is to produce a surface texture with a set of "as much as possible drawn" insights made. Estimates are called global precision by using one-by-two, or multiple mathematical points in the study area. Its counterpart, kriging, is accurate in your area in the slightest sense of diversity, yet it offers inaccurate representations of local diversity Caers (2000a). Subsequent simulations in their various geostatistical tastes have grown to be one of the most well-known and relevant tools for obtaining comparisons of histogram of a particular type and variogram (s) obtained from the data (Isaaks (1992),? (?)). In mathematical terms, Gaussian sequential sequencing is the simplest form of simulation (Deutsch and Journel, 1998) because of conditional sequences from which the applied values are drawn by Gaussian with parameters determined by the solution of a simple kriging scheme. However various limitations and limitations may be caused by the following Gaussian simulation: SGSIM relies on various Gaussianity assumptions, which are thought to be completely un-tested in practice, but seem to be taken for granted. Multi-Gaussianity leads to a more uniformly limited cross-sectional understanding (higher penetration), a property that often conflicts with local facts.

SGSIM requires the transformation to be a Gaussian spatial pre-simulation and the associated backtransformation after the simulation completion. However, the frequency

with which the primary variable is to be generated must be expressed in a second-order variance or non-linear average of the basic variable Caers, Journel, et al. (1998). Normal transformations of marks are nonlinear transformations, so they destroy the existing linear relationship between the primary and secondary variables, or, they change the nonlinear if the relation is nonlinear. SGSIM reproduces, in a sense, only the standard school variogram, not the original variogram model. Usually the production of a common scoring variation graph involves the redesigning of the original data dialog when the data histogram is not overly tied. However, in the case of high skewness, the reprogramming of the variogram model after the back-conversion is not guaranteed.

It is therefore much easier to model directly in the space of the principal variable without the change and without relying on general Gaussian assumptions. Proposed direct sequencing (DSSIM) has been proposed Journel (1994a) to target the exact data point and is not based on many Gaussian assumptions. In fact, DSSIM can produce a better knowledge of low-throughput (more connected) signals than traditional SGSIM. DSSIM relies on an important theoretical result Journel (1994b) that, in order to generate a given covariance model, the sequential distribution used in the sequential model can be of any kind as long as it determines the meaning of kriging and variance.

There are two important limitations to the current DSSIM algorithm: DSSIM does not validate the histogram reconstruction, which is the only non-linear regression for the global sample and variance. Therefore, background labeling of the histogram may be required, which may affect the variogram reconstruction. DSSIM does not allow for its reconstruction of the reference pointers as much as possible with the use of a sequential reference interface Deutsch and Journel (1998). The re-disclosure of the indicators is important when the excess values are removed from the data and need to be recalculated in the given understanding.

Theorem 2.3.1 *A sequential algorithm to generate a specific covariance model is sufficient that all ccdfs point to sample means of some of the variances from that covariance model.*

2.4 Essence of Backtransformation

Sequential sequencing identifies only the variogram Caers (2000b). The histogram seen in the derivative will generally depend on the type of distribution of the conditions used, on the global mean value and the global variance (as used in simple kriging) and on the amount of data available. The theorem 1 does not guarantee the reproducibility of the data histogram, which is why backtransformation should be used to identify any target histogram. The GSLIB-program transformation Shimazaki and Shinomoto (2007) can be used to perform this function. However, it would be more convenient to have a method that reconstructs the histogram and variogram at the same time, without relying on the histogram view that could potentially damage the variogram output. Soares (2001a) in his journal pointed out that, sequential modeling has become one of the most common and most popular algorithm for generating the global distribution and uncertainty of variables of different resources in Earth science because of its simplicity. Different versions of sequential sequences require the evolution of original fluctuations and alternative methods for measuring local distribution functions. Direct sequence equations, which imply that without real-time convergence, are widely used in the spatial direction of phase transition Soares (1998). For sequential estimation of continuous variables, sequential indicator simulation (SIS) and Sequential Gaussian simulation (SGS) -qualify first-order variables into a set of variables or standard Gaussian variables. Local estimation of the probability distribution is performed within the indicator or multGaussian order, respectively. The advantages and disadvantages of the use of one of these algorithms for continuous variance symptoms have been reported extensively in Goovaerts et al. (1997). It would be fair to say that the major barriers to these approaches are directly or indirectly related to the need to evolve real change. Journel (1994a) introduced the first step of the use of continuous variables without any previous modifications: he showed that direct estimation of continuous variables was successful in proposing a covariance model, as long as the values used were derived from a local distribution based on simple kriging equations with variances consistent with simple variance estimation. of kriging. This is called the Travery theorem.Caers (2000b) confirms that the spatial covariance of the original vari-

able is reproduced but not the histogram, which is one of the main requirements of any estimation algorithm. This has been a severe limitation on the actual use of the direct measurement method.Caers et al. (1998) proposed the use of post-processing to convert "pseudo" values out of another set of values, deriving approximately the histogram of the original variance and maintaining the data specification. However, this recent change, in some cases, can be detrimental to variogram production. Caers (2000b) suggested a straightforward directional direction by introducing a set of linear constraints after decomposition and diffusion, thereby avoiding background deformation. Soares (2001b), also proposes a new method of measuring precision directly based on the principle introduced by Journel (1994a). This simple algorithm is effective to generate variogram and histogram of continuous variables. The main advantage of his newly proposed algorithm is that it allows the replication process without costing any source variable.

Consider the continuous variation of $z(x)$ with cdf $F_Z(z) = P(z(x) < Z)$ and a constant variogram $\gamma(h)$. Our key interest is reproducing both $z(x)$ and $\gamma(h)$ in the final map created. A sequential algorithm for continuous flexibility according to (Goovaerts et al. (1997) and Soares (2001b)) follows a sequence:

1. Select a random node x_μ in the location area instead of the standard grid of the to-do areas.

2. Measure the local distribution function of x_μ, given the original $z(x_u)$ data and the previous replicated values $z^s(x_i)$.

3. Draw the number of $z^s(x_i)$ from the local cdf.

4. Return to stage (1) until all points are visited in an informal way.

In stage (2) local cdf measurements are usually done with indicator method (SIS) or in the form of a multiGaussian (SGS), both of which require a change in actual variability. A goal that is directly supported by direct sequence can be summarized as follows: If local cdfs are focused on simple kriging point $z(x_\mu)^* = m + \sum_\alpha(\lambda_\alpha)[Z(x_\alpha) - m], x_\alpha$ status data; that is, the original and earlier data points for the conditional difference identified

by the simple variation of the frame $\sigma^2_{SK}(x_\mu)$; It doesn't matter what possible distribution we choose, geographical model or reproduced variograms on the final map created.

Bourgault (1997) and Caers et al. (1998) provided an effective indication of this statement in the various forms of transmission. The problem is that, with the exception of a few basic distributions (e.g., Gaussian), this simulation method does not produce histograms. The main reason for this problem is that the local cdf cannot be fully recognized in terms of location definition and geographical diversity. The idea proposed by Soares (2001b) is to apply a limited understanding of space and diversity, not to define a local cdf but a sample from a global cdf. It is the same process with consecutive indicator measurement: in this algorithm the global histogram remains the same number of classes in each successive step; in place the data for determining data determines which classes will be sampled to produce the new value generated. For example, with the given step of the SIS process, let's say your local x_u values are just in the first two classes of a ten-grade histogram. As a result, a realization value $z^s(x_\mu)$ is usually deducted from those two categories.

In the proposed algorithm the function of the integrated distribution $F_Z(z)$ is the same in the following sequence. The z distances in space are selected from $F_Z(z)$, defining the new $F'_Z(Z)$ and then sampling $z^s(x_\mu)$ are taken from the sample from the selected $F'_Z(Z)$ stream. The intervals in space are "focused" on a simple kriging estimate $z(x_u)^*$, space interval range depending on the SK measured variation $\sigma^2_{SK}(x_u)$. One way to describe these space distances is to select the lower $Z(x_i)$ set of test histogram (which you wish to reproduce on the last map generated) in the sense that the means and variations of the selected values of $Z(x_i)$ are equal to the local SK values of $z(x_u)^*$ and the SK variations of $\sigma^2_{SK}(x_u)$ respectively: $z(x_\mu)^* = \frac{1}{n}\sum_{i=1}^{n} Z(x_i)$ and $\sigma^2_{SK}(x_\mu) = \frac{1}{n}\sum_{i=1}^{n}[Z(x_i) - z(x_\mu)^*]^2$.

Thereafter, the estimated value of $z^s(x_\mu)$ is deducted from $F'_Z(Z)$ of the selected values. But we need to keep in mind that the most critical situations of lack of space (for space as a result of nugget), those spatial intervals are of equal magnitude, or equal distribution $F'_Z(Z)$ of equal variation. It means that since the spatial intervals are chosen, or

$F_Z^{'}(Z)$ is a random selection from $F_Z(z)$, this corresponds to the design method B. Ripley (1987) of drawing values from the global cdf. Alternatively, the easiest to use, is to define the scope of such global cdf sampling distances, according to the value $\sigma^2_{SK}(x_\mu)$, using the Gaussian distribution.

Suppose ϕ the standard deviation of the $z(x)$ values $y(x) = \phi(z(x))$ with $G(y(x)) = F_Z(z(x))$. The local value of SK $z(x_\mu)^*$ has the same value as the Gaussian, $y(x_\mu)^*$, $y(x_\mu)^* = \phi(z(x_\mu)^*)$ which, together with standard SK $\sigma^2_{SK}(x_\mu)$, may describe a Gaussian cdf $G(y(x_\mu)^*, \sigma^2_{SK}(x_\mu))$. This p from the division of the uniform $U(0,1)$ should collected. Generate the value y^s from $G(y(x_\mu)^*, \sigma^2_{SK}(x_\mu)$. $y^s = G^{-1}(y(x_\mu)^*, \sigma^2_{SK}(x_\mu), p)$. Finally, the estimated value of $z^s(x_\mu)$ is obtained by inverse transform $\phi^{-1} : z^s(x_\mu) = \phi^{-1}(y^s)$, implying that, $z^s(x_\mu)$ is a sample from the intervals of $F_Z(z)$ defined by the local values of $z(x_\mu)^*$ and $\sigma^2_{SK}(x_\mu)$.

Importantly; (1) The Gaussian modification is used only for the distribution samples of $F_Z(z)$ intervals of distribution. No role in local cdf rating; therefore, no Gaussian concept of converted values is considered. The whole sequence process is done with Original $z(x_\mu)$. (2) Because we use offline(nonlinear) function ϕ to find global cdf spaces, we cannot guarantee that the expectations of $z^s(x_\mu) = z(x_\mu)^* : E\{y^s\} = y(x_\mu)^* = \varphi(z(x_\mu))$ but $E(z^s(x_u)) \neq z(x_u)^*$.

This theoretical limitation of the sampling method has not shown a significant effect in most cases, and should be measured in terms of ease of use. However, in those cases where the histogram of $z(x)$, which is used for standard deviation (2), has very little data, especially in lower frequency classes, the following partiality may be possible: $E(z^s(x_\mu)) \neq z(x_\mu)^*$. This means we are trying to make relation (4) between Gaussian cdf focusing on $y(x_u)^*$ and a $z(x_\mu)$ cdf not focusing on $z(x_\mu)$ given the lack of data.

The value $y(x_\mu)^*$ relates to the area value of $z(x_\mu)^*$. The value used for $z^s(x_\mu)$ is deducted from the output of $F_Z(z)$ defined by $G(y(x_\mu)^*, \sigma^2_{SK}(x_\mu))$. After calculating the

normal change of $z(x)$, Goovaerts et al. (1997). DSSIM can be defined in the following steps:

1. Specify a random path across the grid of places $x_\mu, \mu = 1, N_s$ to be made.

2. Measure the mean area and the variance of $z(x_\mu)$, identified sequentially, with the simple kriging $z(x_\mu)^*$ and the $\sigma^2_{SK}(x_\mu)$ measurement variance included in the test data $z(x_i)$ and previous calculated $z^s(x_\mu)$.

3. Specify the output interval for $F_Z(z)$, using Gaussian cdf: $G(y(x_\mu)^*,\ \sigma^2_{SK}(x_\mu))$, where $y(x_u)^* = \varphi(z(x_\mu))$.

4. Draw the value of $z^s(x_\mu)$ from cdf $F_Z(z)$. Generate the p value from the uniform $U(0,1)$, Generate a value y^s from $G(y(x_\mu)^*, \sigma^2_{SK}(x_\mu)) : y = G^{-1}(y(x_\mu)^*, \sigma^2_{SK}(x_\mu), p)$. Return the approximate value $z(x_\mu)^* = \varphi^{-1}(y^s)$.

5. Loop until all N_s locations are visited and made.

Soares (2001a) in his paper pointed out that sequential imitation has become the most common and most popular algorithm to produce land redistribution and the uncertainty of resource variables that differ from Earth science because of its simplicity. Various versions of the sequential imitation require the modification of the original variables and different methods for measuring local distribution functions. Direct sequential measurement, meaning that in addition to the initial variable dynamics, it is widely used in the spatial orientation of the dynamics of the categories Soares (1998).

As for sequential imitation of continuous variable, its alternatives-sequential indicator imitation and sequential Gaussian simulation —require the initial variables be changed into a set of discrete variables or standard Gaussian variables.

Local estimates of possible distributions are made within the index or multGaussian format, respectively. The advantages and disadvantages of using one of these algorithms for the nature of continuous symptom variability have been extensively reported in Goovaerts et al. (1997). It would be fair to say that the major obstacles to these approaches are directly or indirectly related to the need to change real flexibility. Journel

(1994b) introduced the first step in the simulation of continuous variables without any previous modifications: he has shown that direct continuous variable simulations are effective in elevating the covariance model, as long as the deducted values are taken from a local shift focused on simple kriging equations with simple variance of kriging. This is called the Travery theorem. Caers (2000a) confirms that the earth's completeness of the actual variables is recreated but not histogram, which is one of the key requirements of any simulated algorithm.

This has been a key obstacle of the actual use of the direct sequential simulation method. Caers et al. (1998) proposes the use of postal mail to convert "pseudo" values out of another set of values, obtain approximately the histogram of the original variance and maintain data specifications. However, these recent modifications, in some cases, could impair variogram production. Caers (2000b) suggested direct indicator imitation by introducing a set of specific limitation after the definition of kriging average and variance, thus avoiding background reversal.Soares (1998), also proposes a new method of direct sequential simulation based on the principle presented by Journel (1994a). This simple algorithm is effective in generating variogram and histogram for continuous flexibility. The great advantage of his newly updated algorithm is that it allows the cosimulation process without costing any real flexible change.

A goal that is directly supported by direct sequential simulation can be summarized as follows: If local cdfs are focused on simple kriging scales. $z(x_\mu)^* = m + \sum_\alpha (\lambda_\alpha)[Z(x_\alpha) - m], x_\alpha$ data being dependent; that is, the original and earlier data points for the conditional difference identified by the simple kriging variation of estimate $\sigma^2_{SK}(x_\alpha)$; It doesn't matter what possible distribution we choose, geographical model or variogram models are reproduced on the final map created. Bourgault (1997) and Caers et al. (1998) provided an effective indication of this statement in the various forms of transmission. The problem is that, with the exception of a few parametric distributions (e.g., Gaussian), this simulation method does not produce histograms. The main reason for this problem is that the local cdf cannot be fully recognized in terms of location definition and geographical diversity.

The idea proposed by Soares (2001a) is to apply a limited understanding of spatial location and diversity, not to define a local cdf but a sample from a global cdf. It is the same process with consecutive indicator simulation: in this algorithm the global histogram remains the same number of classes in each successive step; in this place, the conditional data determine which classes will be sampled to produce the new value generated. For example, with the given step of the SIS process, let's say your local x_μ values are just in the first two classes of a ten-grade histogram.As a result, a simulated value of $z^s(x_\mu)$ is usually deducted from those two categories.

In the proposed algorithm the function of the integrated distribution $F_Z(z)$ is the same in the following sequence. The z intervals are selected from $F_Z(z)$, defining the new $F'_Z(Z)$ and then sampling $z^s(x_\mu)$ are taken from the samples from the selected $F'_Z(Z)$ stream. These intervals are "focused" on a simple kriging estimate $z(x_\mu)^*$, the time range depending $\sigma^2_{SK}(x_\mu)$ on the SK variance. One way to describe the intervals is to select the lower $Z(x_i)$ set of test histogram (which you wish to reproduce on the last map generated) in in the sense that the means and variations of the selected values of $Z(x_i)$ are equal to the local SK values of $z(x_u)^*$ and the variations of SK $\sigma^2_{SK}(x_\mu)$ respectively: $z(x_\mu)^* = \frac{1}{n}\sum_{i=1}^{n} Z(x_i)$ and $\sigma^2_{SK}(x_\mu) = \frac{1}{n}\sum_{i=1}^{n}[Z(x_i) - z(x_\mu)^*]^2$.

Thereafter, the realization value of $z^s(x_\mu)$ is deducted from $F'_Z(Z)$ of the selected values. But we need to keep in mind of some critical situations of lack of spatial structure (covariance model due to complete nugget effect), intervals of that kind are of same magnitude, or equal distribution $F'_Z(Z)$ of equal variation. It indicates that as soon as, or $F'_Z(Z)$, random selection from $F_Z(z)$, it relates to the design method B. Ripley (1987) of drawing values from the global cdf. Alternatively, the easiest to use, is to define the scope of such global cdf sampling intervals, in line with $sigma^2_{SK}(x_\mu)$ value, using the Gaussian distribution.

Assume *phi* is the standard deviation of the $z(x)$ values $y(x) = \phi(z(x))$ with $G(y(x)) = F_Z(z(x))$. The local value of SK $z(x_\mu)^*$ has the same value as the Gaussian $y(x_\mu)^*, y(x_\mu)^* = \phi(z(x_\mu)^*)$ which, together with the standard SK $\sigma^2_{SK}(x_\mu)$ value, may define a Gaussian cdf $G(y(x_\mu)^*, \sigma^2_{SK}(x_\mu))$. The Gaussian cdf is then used to establish the spatial intervals of

the cdf of $z(x)$ value for sampling: figure p value from the uniform distribution $U(0,1)$. Generate the value y^s from $G(y(x_\mu)^*, \sigma^2_{SK}(x_\mu)$. $y^s = G^{-1}(y(x_\mu)^*, \sigma^2_{SK}(x_\mu), p)$. Finally, the estimated value of $z^s(x_\mu)$ is obtained by inverse transform φ^{-1}: $z^s(x_\mu) = \varphi^{-1}(y^s)$ and that shows that, $z^s(x_\mu)$ is a sample from the $F_Z(z)$ intervals defined by the local values of $(x_\mu)^*$ and $\sigma^2_{SK}(x_\mu)$.

It is important to know that, the Gaussian modification is only used for the $F_Z(z)$ distribution intervals. No role in local cdf rating; therefore, no Gaussian concept of converted values is considered. The whole sequence process is done with Original $z(x_u)$. Since we use the nonlinear function of φ to find global cdf intervals, we cannot guarantee that the expectations of global cdf, $z^s(x_\mu) = z(x_\mu)^* : E\{y^s\} = y(x_u)^* = \varphi(z(x_u))$ but $E(z^s(x_\mu)) \neq z(x_\mu)^*$ This theoretical limitation of the sampling method has not shown a significant effect in most cases, and should be measured in terms of ease of use. However, in those cases where the histogram of $z(x)$, which is used for standard deviation, has very little data, especially in lower frequency classes, the following may be possible: $E(z^s(x_\mu)) \neq z(x_\mu)^*$. This means we are trying to make relationship between the Gaussian cdf focusing on $y(x_\mu)^*$ and the $z(x_\mu)$ cdf not focusing on $z(x_\mu)$, given the lack of data.

Chapter 3

METHODOLOGY

3.1 Introduction

Kriging estimator is one of the most salient spatial interpolation techniques every Geostatistician would employ on spatial data distribution. The kriging estimator could be largely influenced by factors such as sample size, sampling scheme, type of data etc. That is to say, the sanctity or accuracy of the kriging estimate could largely depend on at least one of these factors mentioned. This thesis therefore seeks to measure the impact of the size of sample and design of sampling on uncertainty of variograms from which the kriging estimate is deduced.

3.2 Variogram

Variogram is a function of separation distance, it gives a measure of spatial dependence by figuring out how sampled data are related to distance and direction. The separation distance refers to the distance between two spatial locations $z(u_\alpha)$ and $z(u_\alpha + h)$ which may be differentiated by lag h. Mathematically, variogram is defined by

$$2\hat{\gamma}(h) = \frac{1}{n(h)} \sum_{\alpha=1}^{n(h)} [z(u_\alpha + h) - z(u_\alpha)]^2, \tag{3.1}$$

with $n(h)$ being number of sample points disjointed by spatial locations $z(u_\alpha)$ and $z(u_\alpha + h)$.

3.3 Semivariogram

Semivariogram as the name implies simply represents half of the variogram function. In some cases, the two functions are used interchangeably. Mathematically, Semivariogram model may be defined by

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{\alpha=1}^{n(h)} [z(u_\alpha + h) - z(u_\alpha)]^2, \tag{3.2}$$

with $n(h)$ being number of sample points disjointed by spatial locations $z(u_\alpha)$ and $z(u_\alpha + h)$.

It is worth noting that an empirical variogram function is fitted into a licit theoretical variogram model in the data computational process. The three licit basic variogram models are described below;

3.4 Spherical model(Exist in $\mathbf{R}^d, d \geq 1$)

Spherical model is defined by

$$\gamma_{Sph}(h;\theta) = \begin{cases} 0, & h = 0 \\ c_0 + c_1\{(\frac{3}{2})(\frac{h}{r}) - (\frac{1}{2})(\frac{h}{r})^3\}, & h \geq r \\ c_0 + c_1, & 0 < h \leq r \end{cases}$$

where

$$\theta = (c_0, c_1, r)^T, \quad c_0, c_1, r \geq 0. \tag{3.3}$$

This model type exhibits around the neighborhood of the origin linearity behavior whiles establishing spatial independence beyond r.

3.5 Gaussian model (Exist in $\mathbf{R}^d, 1 \leq d \leq 3$)

The Gaussian model is defined by

$$\gamma_{Gau}(h;\theta) = \begin{cases} 0, & \|h\| = 0 \\ c_0 + c_1\{1 - e^{-3(\|\frac{h}{r}\|)^2}\}, & 0 \leq \|h\| \leq r \end{cases}$$

where

$$\theta = (c_0, c_1, r)^T, \quad c_0, c_1, r \geq 0. \tag{3.4}$$

Gaussian model produces feature similar to quadratic function near the origin of the variogram function with short range spatial correlation that is higher than any of valid variogram models (second order stationary models) with same practical range.

3.6 Exponential model (Exist in $\mathbf{R}^d, d \geq 1$)

The exponential model as one of most salient second order stationary (SOS) models is defined by

$$\gamma_{Exp}(h;\theta) = \begin{cases} 0, & h = 0 \\ c_0 + c_1\{1 - e^{-\frac{\|h\|}{r}}\}, & 0 < h \leq r \end{cases}$$

where

$$\theta = (c_0, c_1, r)^T, \quad c_0, c_1, r \geq 0. \tag{3.5}$$

This model also exhibits linearity behavior at its origin and with spatial correlation decaying to zero as separation distance increases. The range will become the lowest value of h for $\gamma_{Exp}(h;\theta)$ to be equal to the sill whenever the sill exist. Whenever the sill exist not, then a practical range is defined for which $\gamma_{Exp}(h;\theta)$ approaches 95% of the sill.

3.7 Variogram parameters

Most variograms come with parameters that play a major role in construction variogram which gives the closest estimate of the global autocorrelation make-up of the stochastic

phenomenon in question. These parameters include the nugget, range and the sill.

3.7.1 The nugget($\gamma_{Nugg}(h;\theta) = c_0 \quad as \quad h \to \infty$)

This is sometimes called the measurement error or small-scale variation leading to spatial discontinuity around the origin. On an empirical variogram, it is the value of $\gamma(h)$ whenever $h = 0$. Mathematically, it is defined by

$$\gamma_{Nugg}(h;\theta) = \begin{cases} 0, & h = 0 \\ c_0, & h \neq 0 \end{cases}$$

where

$$\theta = c_0 \geq 0. \tag{3.6}$$

3.7.2 The Range

It is the distance at which location data show autonomy among themselves. At this point spatial autocorrelation does not exist any longer in data.

3.7.3 The Sill

This is the limiting value of $\gamma(h)$ as $h \to \infty$ (as h increases indefinitely). In other words, the total variance that allows the semivariogram to level off is known as the sill. The partial sill is the difference between the sill and the nugget.

3.7.4 Stationarity Assumption

Certain stationarity assumptions about the Random Function are made to enable Geostatistician to say something about the joint and marginal distributions of the random variables. If random function is second-order stationary, then following must be seen;

- The expected value $E\{Z(\mu)\}$ exists and be independent of the lag h.

- The covariant function

$$C(h) = E\{Z(\mu+h)Z(\mu)\} - E\{Z(\mu)\}E\{Z(\mu+h\} \tag{3.7}$$

exists and varies with lag h.

From these assumption, the following formulations are deduced;

$$\gamma(h) = C_0 - C(h); \quad \rho(h) = 1 - \frac{\gamma(h)}{C(h)}, \tag{3.8}$$

this implies that $C(h) \to 0$ as $\|h\| \to \infty$ and $\gamma(h) = C_0$ for $\|h\| \to \infty$.

3.7.5 Intrinsic Stationarity

If the increment in $Z(\mu+h) - Z(\mu)$ follows the second order stationarity condition, then the increment(Random Function) becomes intrinsic stationary. This implies

$$2\hat{\gamma}(h) = var\{Z(\mu+h) - Z(\mu)\} = E\{[Z(\mu+h) - Z(\mu)]^2\} \tag{3.9}$$

3.8 Kriging

Kriging is an artificial geostatistics method used to predict unconfirmed values in unknown locations based on observable values in known locations. Dr. G. Krige, a South African mining engineer who was the first to develop this method of speculation in predicting the true distribution of ore-grade from a sample based on ore-grade Krige (1952). But the construction of the most appropriate linear estimation methods came from others,Cressie (2015). There are different types of kriging methods; simple kriging (SK), ordinary kriging (OK), universal kriging (UK) among others.

3.8.1 Simple Kriging

Let $\{Z(\mu_\alpha), \alpha = 1, 2, 3, \cdots, n\}$ be sample data, the kriging formulation can written as $Z^*(\mu) - m(\mu) = \sum_{\alpha=1}^{n(u)} \lambda_\alpha(\mu)[Z(\mu_\alpha) - m(\mu_\alpha)]$, where $Z^*(\mu)$ is an estimate of a true

(but unknown) value $Z(\mu)$, $n(\mu)$ is the number of sample values at point μ, $\lambda_\alpha(u)$ is the weight assigned to $z(\mu_\alpha)$, $z(\mu_\alpha)$ is the realization of $Z(\mu_\alpha)$, $m(\mu) = E\{Z(\mu)\}$ and $m(\mu_\alpha) = E\{Z(\mu_\alpha)\}$ with estimation error $Z^*(\mu) - z(\mu)$. The goal to find an unbiased estimate that minimizes the error variance $\sigma_E^2 = var\{Z^*(\mu) - z(\mu)\}$ subject to the unbiased estimate $E\{Z(\mu) - Z(\mu)\} = 0$ but for simple kriging(SK), let the kriging formulation be $Z^*(\mu) = \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha(\mu)[Z(\mu_\alpha) - m(\mu_\alpha)] + m(\mu)$, where $m(u)$ is a known constant throughout the entire study area.

$$\begin{aligned} Z^*_{SK}(\mu) &= \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{SK}(\mu)[Z(\mu_\alpha) - m] + m \\ &= \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{SK}(\mu) Z(\mu_\alpha) - m \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{SK}(\mu) + m \\ &= \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{SK}(\mu) Z(\mu_\alpha) + [1 - \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{SK}(\mu)]m \end{aligned} \tag{3.10}$$

$E\{Z^*_{SK}(\mu) - Z(\mu)\} = 0$, since SK estimator is automatically an unbiased estimator, the unbiased constraint is redundant here. Consider $R(\mu) = Z(\mu) - m$ and $R(\mu_\alpha) = Z(\mu_\alpha) - m$.

$$\begin{aligned} Z^*_{SK}(\mu) - Z(\mu) &= [Z^\mu_{SK}(\mu) - m] - [Z(\mu) - m] \\ &= \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{SK}(\mu) R(\mu_\alpha) - R(\mu) = R^*_{SK}(\mu) - R(\mu), \end{aligned} \tag{3.11}$$

hence

$$\begin{aligned} \sigma^2_{SK}(\mu) &= Var\{Z^*_{SK}(\mu) - Z(\mu)\} \\ &= Var\{R^*_{SK}(\mu) - R(\mu)\} \\ &= Var\{R^*_{SK}(\mu)\} + Var\{R(\mu)\} - 2Cov\{R^*_{SK}(\mu), R(\mu)\} \\ &= Q(\lambda_\alpha^{SK}(\mu), \alpha = 1, 2, 3, \cdots, n(u)) \end{aligned} \tag{3.12}$$

Where,

$$Q() = \sum_{\alpha=1}^{n(\mu)} \lambda_{\alpha}^{SK}(\mu) \sum_{\beta=1}^{n(\mu)} \lambda_{\beta}^{SK} C_R(\mu_\alpha - \mu_\beta) + C_R(0) - 2 \sum_{\alpha=1}^{n(\mu)} \lambda_{\alpha}^{SK}(\mu) C_R(\mu_\alpha - \mu). \tag{3.13}$$

Calculating the optimal $\lambda_{\alpha}^{SK}(\mu)$, consider the following;

$$\frac{1}{2} \frac{\delta Q(\mu)}{\delta \lambda_{\alpha}^{SK}(\mu)} = \sum_{\beta=1}^{n(\mu)} \lambda_{\beta}^{SK} C_R(\mu_\alpha - \mu_\beta) + C_R(\mu_\alpha - \mu) = 0 \tag{3.14}$$

$K_{SK}\lambda_{SK}(\mu) = k_{SK}$, where $K_{SK} = n(\mu) \times n(\mu)$ covariance matrix between sample points, $\lambda_{SK}(\mu)$ and k_{SK} are $n(\mu) \times 1$ matrices.

$$K_{SK} = \begin{bmatrix} C(\mu_1 - \mu_2) & \cdots & C(\mu_1 - \mu_{n(\mu)}) \\ \vdots & \ddots & \vdots \\ C(\mu_{n(\mu)} - \mu_1) & \cdots & C(\mu_{n(\mu)} - \mu_{n(\mu)}) \end{bmatrix}, \quad k_{SK} = \begin{bmatrix} C(\mu_1 - \mu_2) \\ \vdots \\ C(\mu_{n(\mu)} - \mu_1) \end{bmatrix} \tag{3.15}$$

and

$$\lambda_{SK}(\mu) = \begin{bmatrix} \lambda_1^{SK}\mu \\ \vdots \\ \lambda_{n(\mu)}^{SK}\mu \end{bmatrix} \tag{3.16}$$

with $\lambda_{SK}(\mu) = K_{SK}^{-1} k_{SK}$ and $\sigma_{SK}^2(\mu) = C(\theta) - K_{SK}^T K_{SK}^{-1} k_{SK}$. A unique solution will be obtained if K_{SK} is positive definite.

3.8.2 Ordinary Kriging

Similarly, formulation of ordinary kriging is also deduced as below:

$$\begin{aligned} Z^*(\mu) &= \sum_{\alpha=1}^{n(u)} \lambda_\alpha(\mu)[Z(\mu_\alpha) - m(\mu_\alpha)] + m(\mu) \\ &= \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha(\mu)[Z(\mu_\alpha) - m] + m \\ &= \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha(\mu) Z(\mu_\alpha) - m \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha(\mu) + m \\ &= \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha(\mu) Z(\mu_\alpha) + [1 - Z(\mu_\alpha)]m, \end{aligned} \tag{3.17}$$

for ordinary kriging (OK), $m(\mu)$ is assumed to constant yet unknown. $Z^*_{OK}(\mu) = \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{OK}(\mu) Z(\mu_\alpha)$ with $\sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{OK}(\mu) = 1$. The unbiased condition is satisfied through; $E\{Z^*_{OK}(\mu) - Z(\mu)\} = (\sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{OK}(\mu))m(\mu) - m(\mu) = m(\mu) - m(\mu) = 0$. For us to be able to reduce error variance $\sigma^2_{OK}(\mu)$, the method of Lagrange multipliers will be adopted as follows;

$$L(\lambda_\alpha^{OK}(\mu); 2\mu_\alpha^{OK}(\mu)) = \sigma_E^2(\mu) + 2\mu_\alpha^{OK}(\mu)(\sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{OK}(\mu) - 1) \tag{3.18}$$

where $2\mu_\alpha^{OK}(\mu)$ is Lagrange parameter. We obtain an optimal kriging weights as follows;

$$\begin{aligned} \frac{1}{2}\frac{\delta L(\mu)}{\delta \lambda_\alpha^{OK}(\mu)}\frac{\delta L(\mu)}{\delta \mu_\alpha^{OK}(\mu)} &= \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{SK}(\mu) \sum_{\beta=1}^{n(\mu)} \lambda_\beta^{SK} C_R(\mu_\alpha - \mu_\beta) + C_R(0) \\ &\quad -2\sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{SK}(\mu) C_R(\mu_\alpha - \mu) + 2\mu_\alpha^{OK}(\mu)(\sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{OK}(\mu) - 1) \\ \frac{1}{2}\frac{\delta L(\mu)}{\delta \lambda_\alpha^{OK}(\mu)} &= \sum_{\beta=1}^{n(\mu)} \lambda_\beta^{SK} C_R(\mu_\alpha - \mu_\beta) - C_R(\mu_\alpha - \mu) + \mu_{OK}(\mu). \end{aligned} \tag{3.19}$$

This leads to;

$$\sum_{\beta=1}^{n(\mu)} \lambda_\beta^{SK} C_R(\mu_\alpha - \mu_\beta) + \mu_{OK}(\mu) = C_R(\mu_\alpha - \mu), \alpha = 1, 2, 3, \cdots, n(u) \tag{3.20}$$

and

$$\frac{1}{2}\frac{\delta L(\mu)}{\delta \mu_\alpha^{OK}(\mu)} = \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{OK}(\mu) - 1$$
$$\sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{OK}(\mu) = 1 \tag{3.21}$$

This provides us with a minimum error variance

$$\sigma_E^2(\mu) = C(0) - \sum_{\alpha=1}^{n(\mu)} \lambda_\alpha^{OK}(\mu) - \mu_{OK}(\mu). \tag{3.22}$$

The OK model is written in semivariogram notation as

$$\sum_{\beta=1}^{n(\mu)} \lambda_\beta^{OK}(\mu)[C(0) - \gamma(\mu_\alpha - \mu_\beta)] + \mu_{OK}(\mu) = C(0) - \gamma(\mu_\alpha - \mu_\beta) \tag{3.23}$$

subject to constraint $\sum_{\beta=1}^{n(\mu)} \lambda_\beta^{OK}(\mu) = 1$.

3.8.3 Universal Kriging

This kriging type involves both visual surface analysis (drift) with ordinary kriging in a way quantifying trends, the measurement method (interpolation technique) is being seen as an old-fashioned when the intrinsic random function was adopted. Model consideration is given as

$$Z(\mu) = \sum_{j=1}^{p+1} f_{j-1}(\mu)\beta_{j-1} + \delta(\mu), \tag{3.24}$$

for $\mu \in D$ where $\beta \equiv (\beta_0, \cdots, \beta_p)^T \in \mathbf{R}^{p+1}$ is unknown vector of parameters and $\delta(.)$ is a zero-mean intrinsic stationary random model. The equation $Z(\mu) = \sum_{j=1}^{p+1} f_{j-1}(\mu)\beta_{j-1} + \delta(\mu)$ can be transformed to $Z = X\beta + \delta$. where X is an $n \times (p+1)$ matrix whose $(i-j)^{th}$ element is $f_{j-1}\mu_i$ Universal kriging as used by Matheron (1969) refers to the intolerance of the forecaster when the line of prediction is an unobservable linear combination of known functions.

3.9 Method of Solution

The following formulation is deduced in realizing the goal of thesis;

1. Make a dummy spatial coordinate of length 10000 in R.
2. Simulate 1000 simulations of length 10000 in R using unconditional Gaussian simulation.
3. Extract the 1000 attributes simulated and take their average to form a single attribute.
4. Combine the spatial coordinates with the attribute (the average value from1000 simulations) to form a data frame.
5. Select sample sizes of 50, 150 and 500 from population of size 10000.
6. Observe the impact of each of the selected sample sizes on the uncertainty of simple random sampling, cluster sampling and stratified sampling schemes.
7. Observe all these sampling schemes on each of the selected sample sizes again to quantify their impact on the uncertainty of the variogram maps.

Chapter 4

ANALYSIS AND RESULTS

4.1 Introduction

The data at hand is simulated from unconditional Gaussian simulation using gstat package in R. The dataset has dummy coordinates of length 10000 whiles the simulated attribute is 1000. These datasets are all Gaussian since the simulation originated from unconditional Gaussian technique as illustrated on following figures.

4.2 variogram Analysis

The concentration of spatial attributes on the stochastic fields on the graphs ranges from blue to orange, as indicated by the scale. Blue represents a low concentration of spatial attributes, whereas orange represents a high concentration of spatial phenomena. These characteristics or phenomena could include the presence of disease hotspots, gold deposits, heavy metals, criminal hotspots, and so on. The "plus signs" on the stochastic field represent sample locations where spatial attributes were sampled.

Figure 4.1: Comparing sample structures of fixed samples size 50 for different sampling schemes

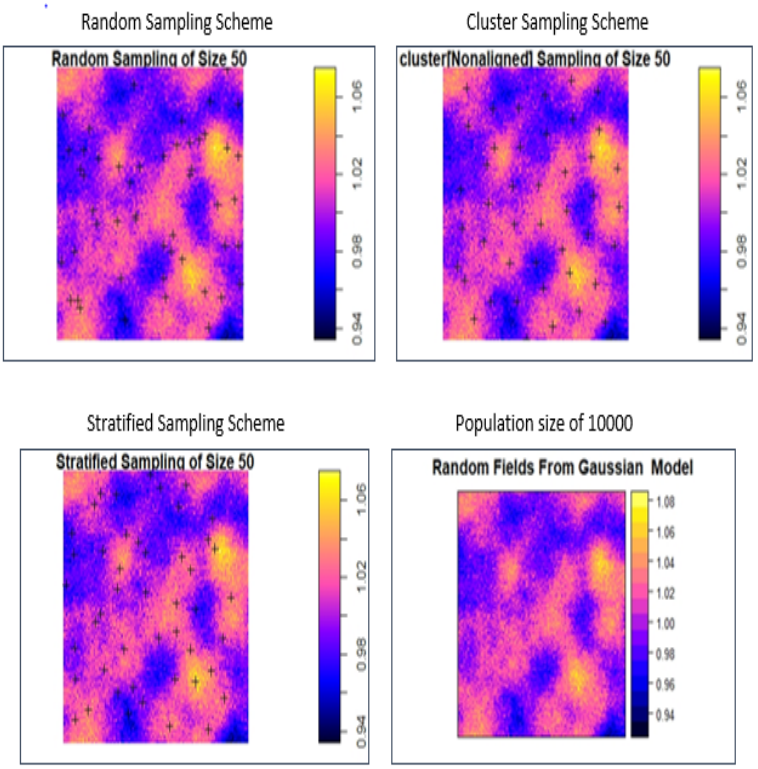


Figure 4.2: Comparing sample structures of fixed samples size 150 for different sampling schemes

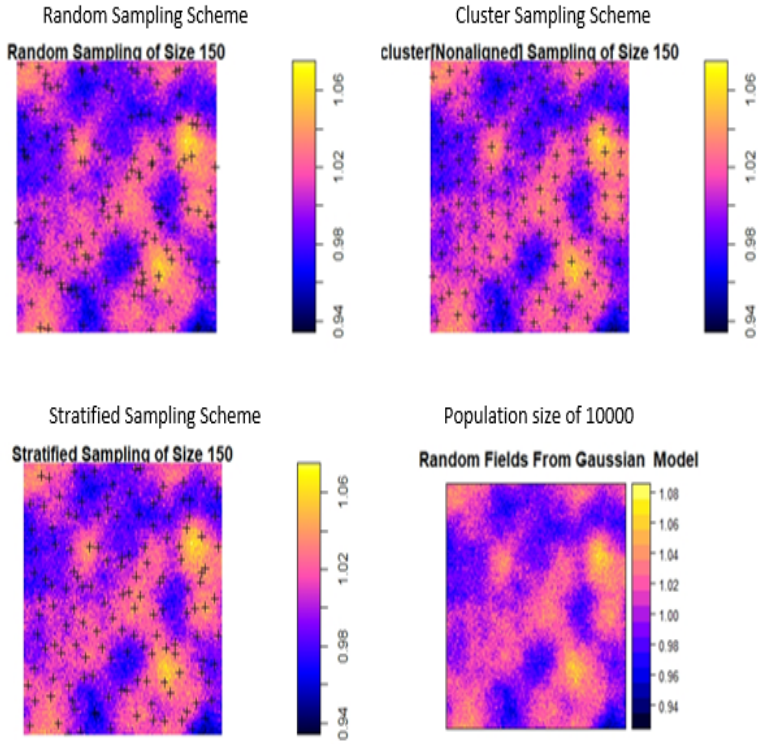
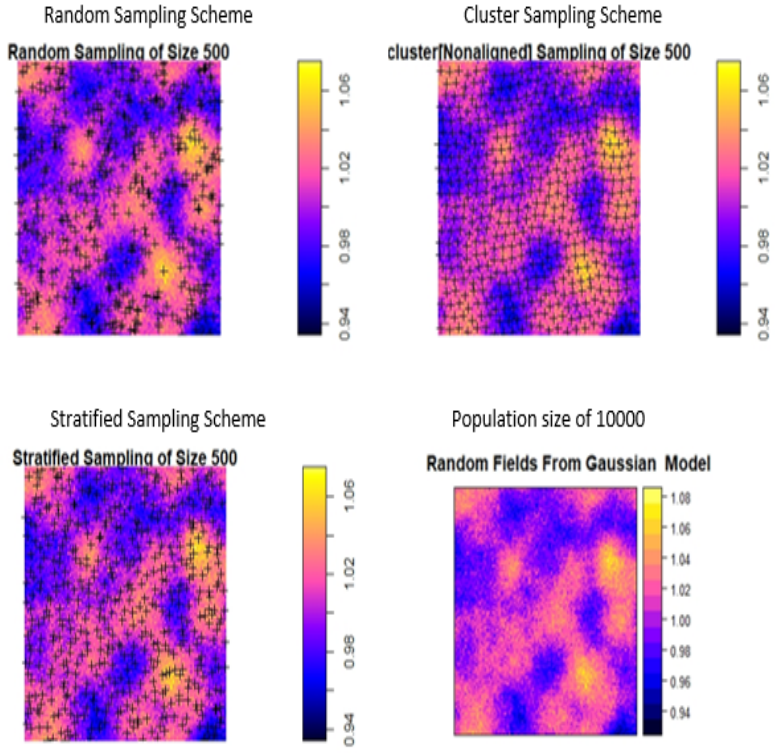


Figure 4.3: Comparing sample structures of fixed samples size 500 for different sampling schemes



Histograms were use to check normality of distribution of the spatial attributes over the stochastic field as illustrated on Figure 4.4 to Figure 4.6 below.

Figure 4.4: Comparing histograms of fixed sample size 50 with different sampling schemes

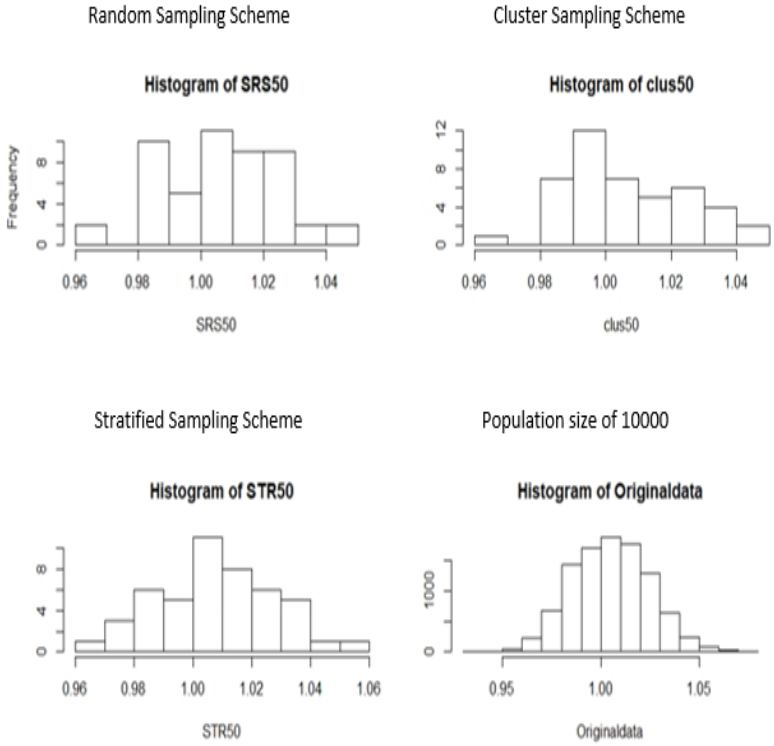


Figure 4.5: Comparing histograms of fixed sample size 150 with different sampling schemes

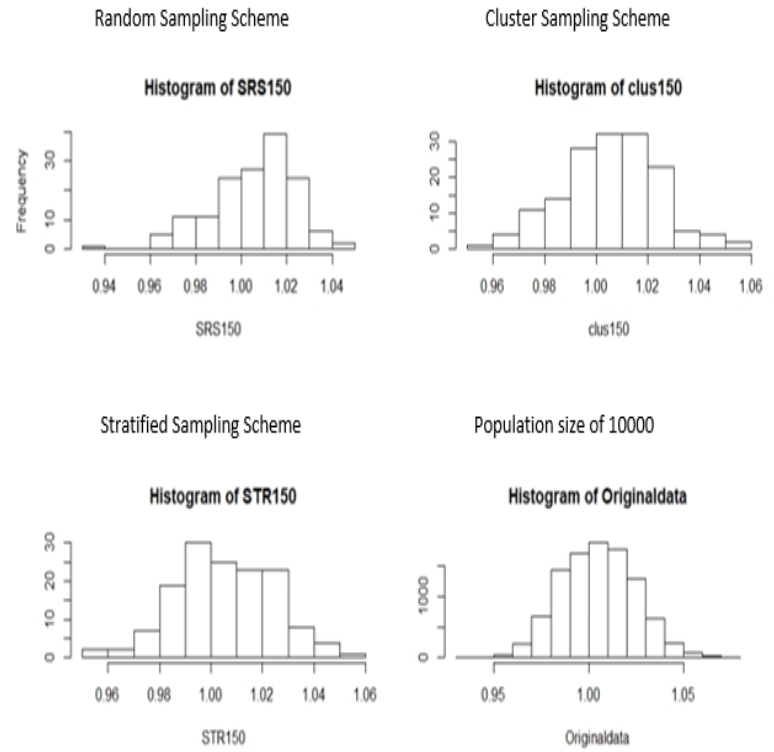
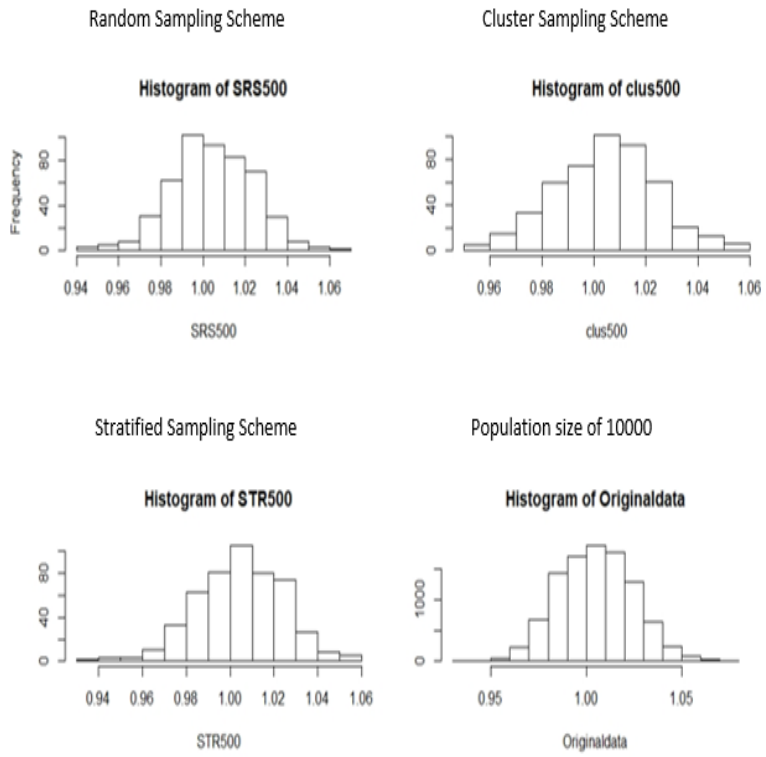


Figure 4.6: Comparing histograms of fixed sample size 500 with different sampling schemes



4.3 Stages in analysis of data

The data simulation and analysis procedures are outlined as follows;

1. Two different variables (var1 and var2) containing values up to size 10000 were generated from a base package in R called “expand.grid”.

2. These variables were renamed as spatial points “x” and “y”.

3. Gstat object was created in R containing Gaussian model parameters with values.

4. Combining the gstat and the base objects, a dismo object “predict” was used for predicting(simulating) values all over the entire study area.

5. An average of all the 1000 variables simulated was taken to form a univariate dependent variable.

6. Simple random sampling, cluster sampling and stratified sampling schemes together with samples such as 50(small sample), 150(medium sample) and 500(large sample) were calculated in R.

7. The sampled data was exported to excel to create data frame to much corresponding coordinates and their dependent variables with values using excel packages "concatenate" and "vlookup".

8. The data is imported back to R for variogram fitting and mapping.

4.4 Results

4.4.1 Effect of Sample Size on Variogram Uncertainty

Figure 4.7: Effect of varying sample size on semivariogram models of SRS scheme

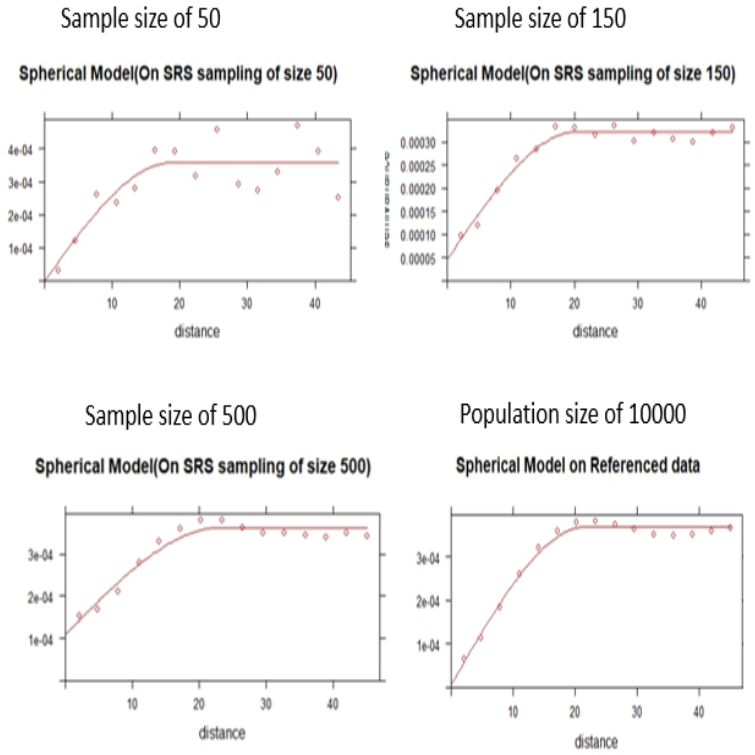


Figure 4.8: Effect of varying sample size on semivariogram models of cluster random sampling scheme

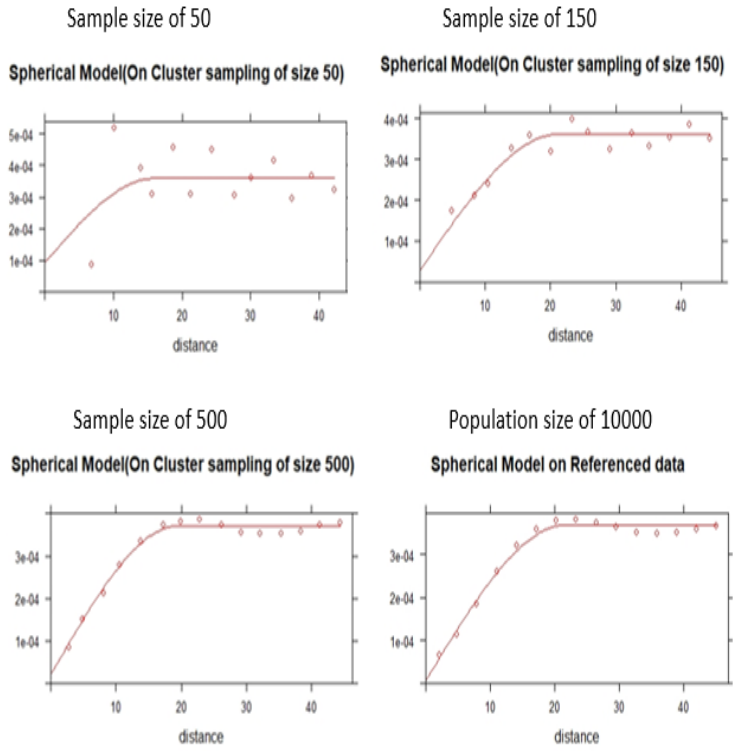
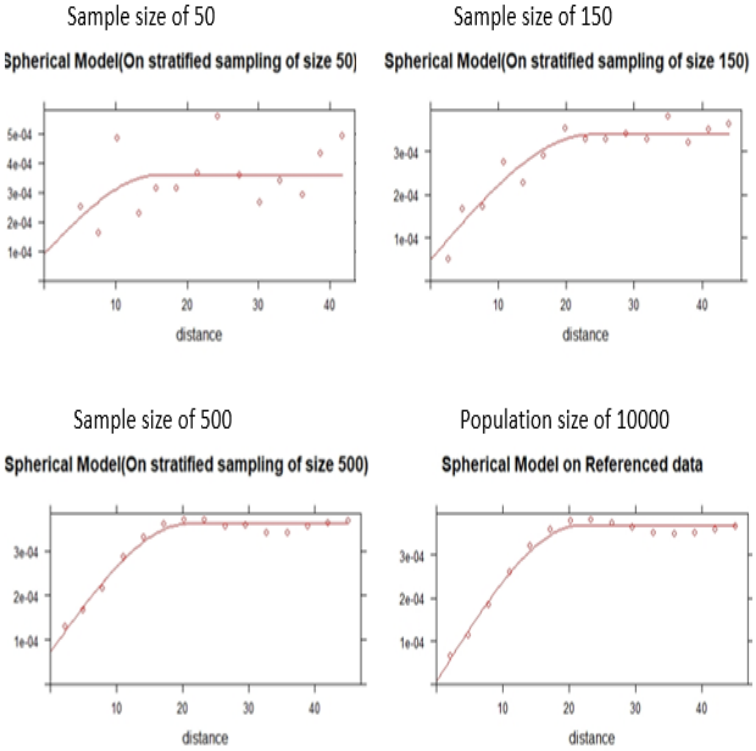


Figure 4.9: Effect of varying sample size on semivariogram models of stratified random sampling Scheme



4.4.2 Effect of Sampling Scheme on Variogram Uncertainty

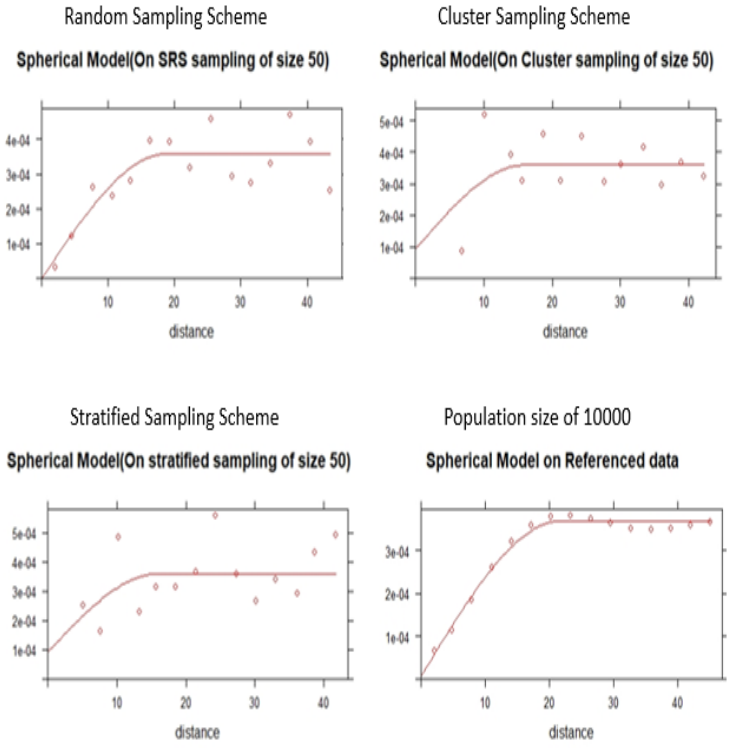
Model	Psill	Range
Nug	0.000006686	0.00000
Sph	0.0003628	21.91134

Table 4.1: Spherical Model Parameters for Global variogram

Model	Psill	Range
Nug	0.0000527614	0.00000
Gau	0.0003184058	10.60762

Table 4.2: Gaussian Model Parameters for Global variogram

Figure 4.10: Effect of varying sampling schemes for fixed sample size 50 on variogram uncertainty



Model	Psill	Range
Nug	0.00000	0.00000
Sph	0.0003575995	18.82379

Table 4.3: Model Parameters for variogram on simple random sampling of size 50

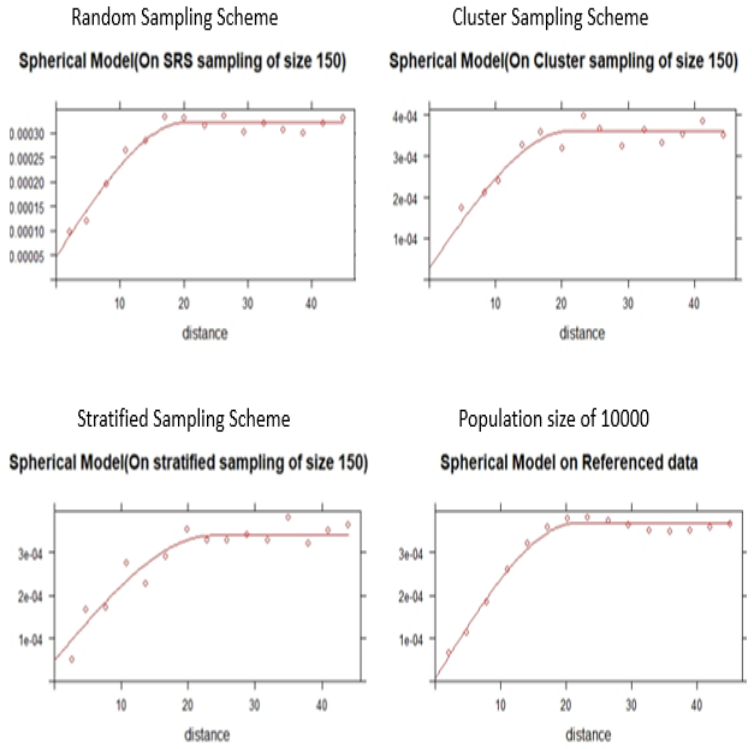
Model	Psill	Range
Nug	0.00000	0.00000
Sph	0.00035707	13.98952

Table 4.4: Model Parameters for variogram on cluster random sample of size 50

Model	Psill	Range
Nug	0.0001902389	0.00000
Gau	0.0001828241	10.98444

Table 4.5: Model Parameters for variogram on stratified random sample of size 50

Figure 4.11: Effect of varying sampling schemes for fixed sample size 150 on variogram uncertainty



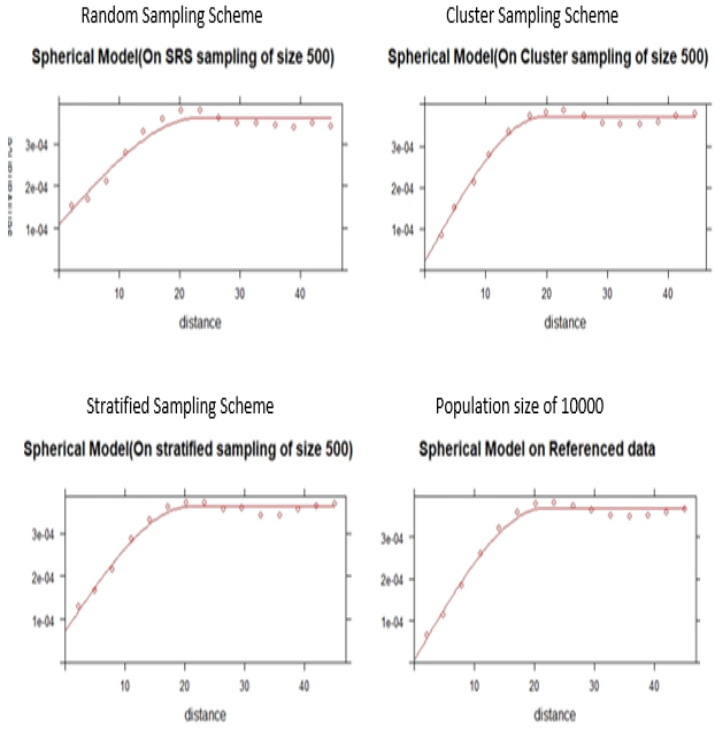
Model	Psill	Range
Nug	0.0001902389	0.00000
Gau	0.0001828241	10.98444

Table 4.6: Model Parameters for variogram on systematic random sample of size 50

Model	Psill	Range
Nug	0.00004678	0.00000
Sph	0.000275738	20.50441

Table 4.7: Model Parameters for variogram on simple random sampling of size 150

Figure 4.12: Effect of varying sampling schemes for fixed sample size 500 on variogram uncertainty



Model	Psill	Range
Nug	0.0001065656	0.00000
Gau	0.0002598078	9.117232

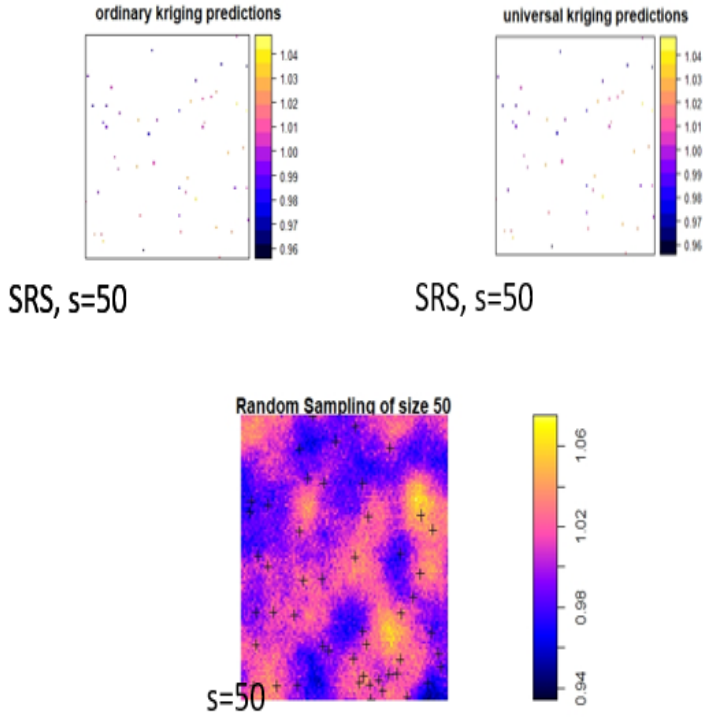
Table 4.8: Model Parameters for variogram on cluster random sample of size 150

Model	Psill	Range
Nug	0.00008087383	0.00000
Gau	0.0002480031	10.24303

Table 4.9: Model Parameters for variogram on stratified random sample of size 150

4.4.3 Maps of Kriging Predictions and Kriging Variance on Ordinary and Universal Kriging

Figure 4.13: Structure of ordinary and universal kriging predictions on SRS of size 50



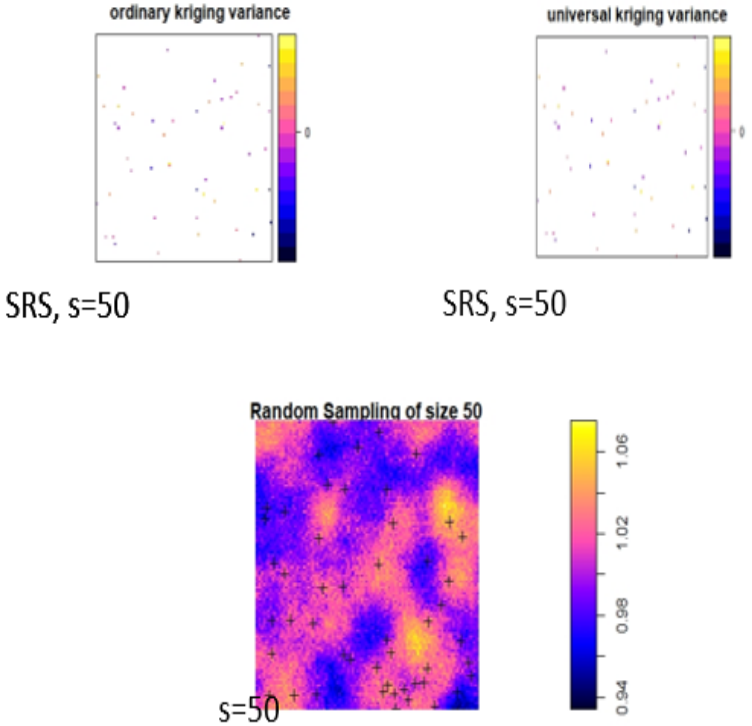
Model	Psill	Range
Nug	0.00004678335	0.00000
Sph	0.0002757383	20.50441

Table 4.10: Model Parameters for variogram on systematic random sample of size 150

Model	Psill	Range
Nug	0.0001084369	0.00000
Sph	0.0002538804	23.35208

Table 4.11: Model Parameters for variogram on simple random sampling of size 500

Figure 4.14: Structure of ordinary and universal kriging variance on SRS of size 50



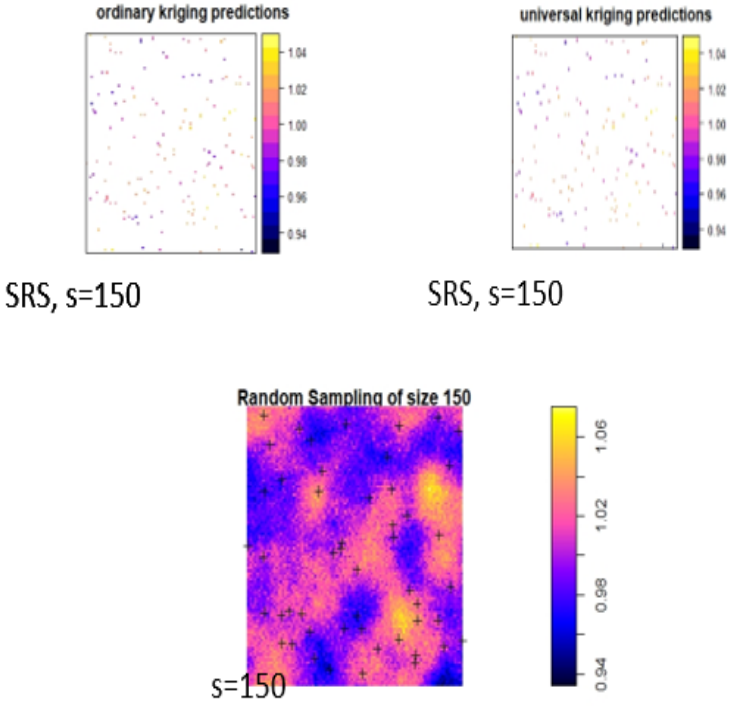
Model	Psill	Range
Nug	0.0000614719	0.00000
Gau	0.0003126234	6.628231

Table 4.12: Model Parameters for variogram on cluster random sample of size 500

Model	Psill	Range
Nug	0.00007351414	0.00000
Sph	0.0002888536	20.95157

Table 4.13: Model Parameters for variogram on stratified random sample of size 500

Figure 4.15: Structure of ordinary and universal kriging predictions on SRS of size 150



Model	Psill	Range
Nug	0.0001084369	0.00000
Sph	0.0002538804	23.35208

Table 4.14: Model Parameters for variogram on systematic random sample of size 500

Figure 4.16: Structure of ordinary and universal kriging variance on SRS of size 150

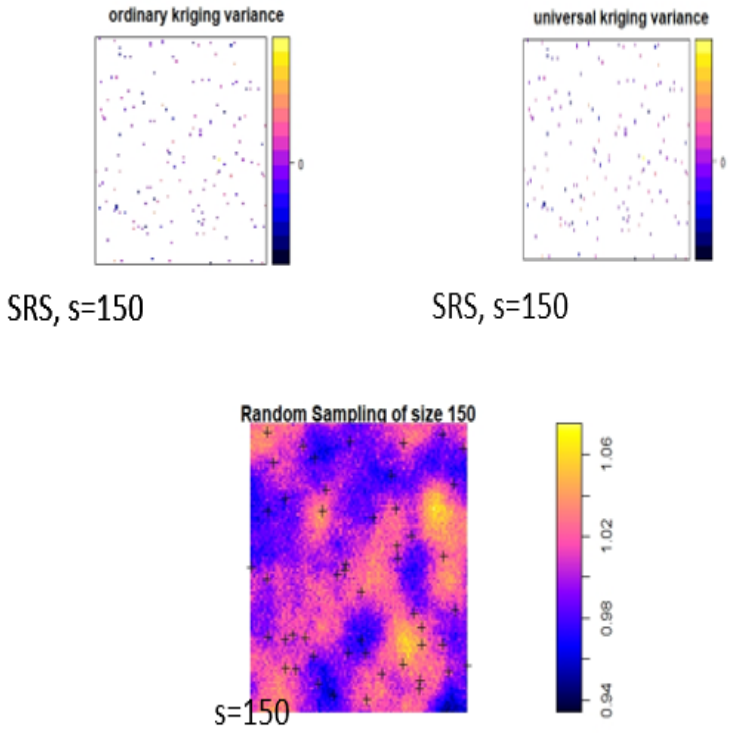


Figure 4.17: Structure of ordinary and universal kriging predictions on SRS of size 500

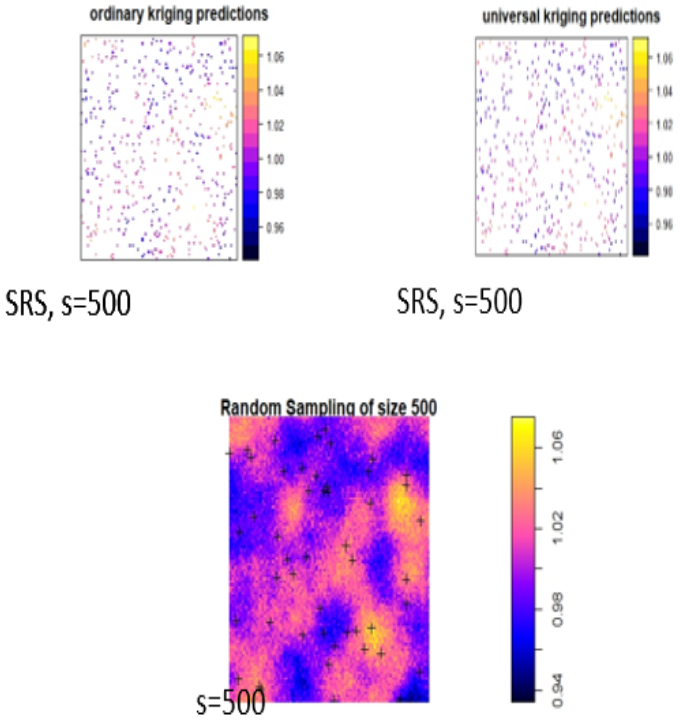


Figure 4.18: Structure of ordinary and universal kriging variance on SRS of size 500

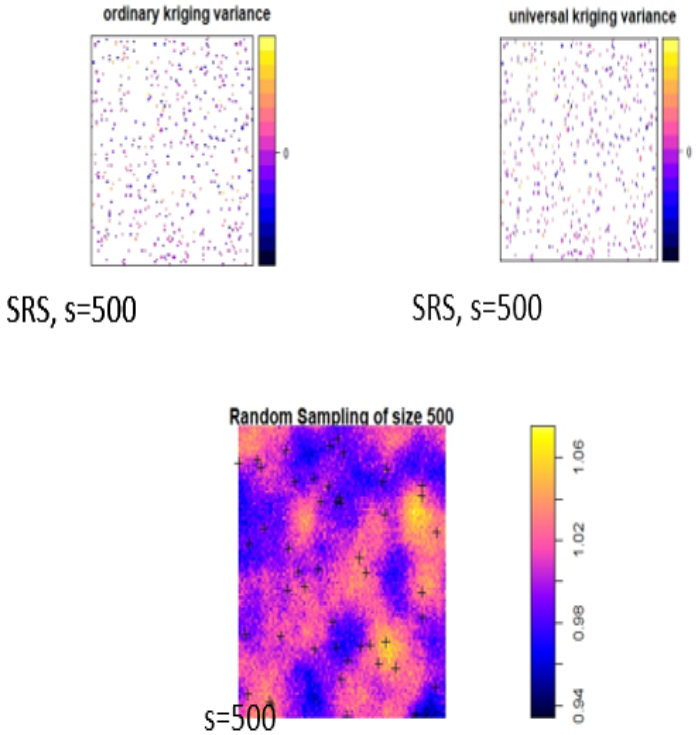


Figure 4.19: Structure of ordinary and universal kriging predictions on CRS of size 50

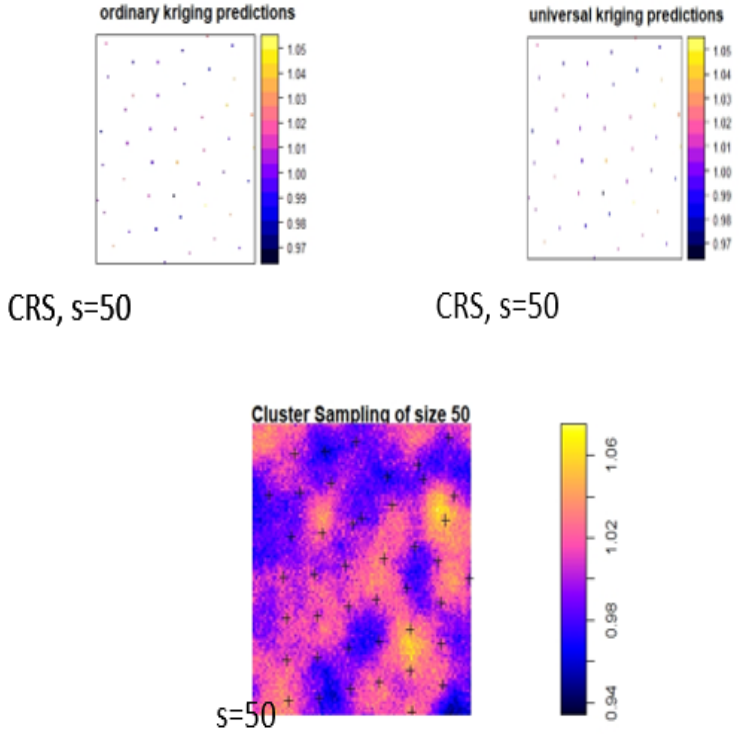


Figure 4.20: Structure of ordinary and universal kriging variance on CRS of size 50

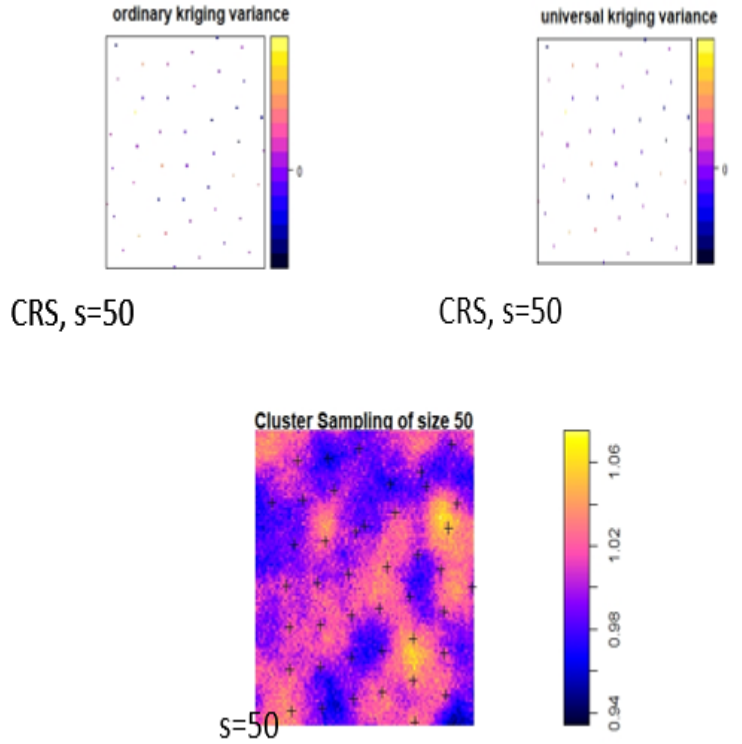


Figure 4.21: Structure of ordinary and universal kriging predictions on CRS of size 150

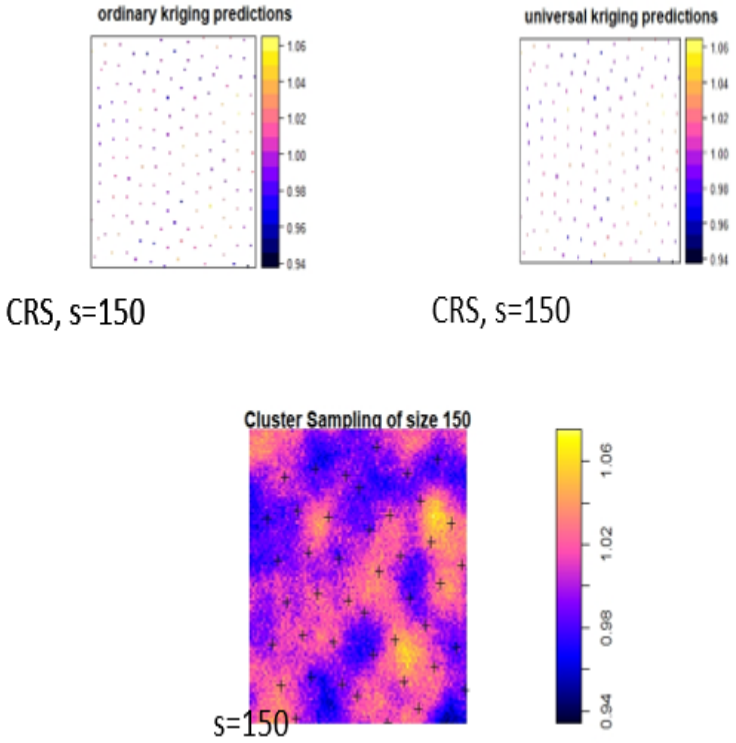


Figure 4.22: Structure of ordinary and universal kriging variance on CRS of size 150

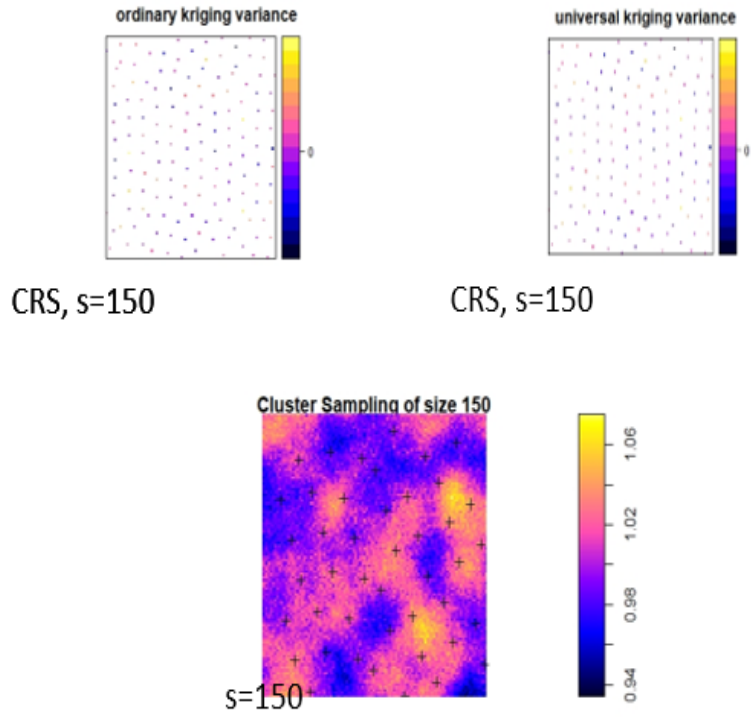


Figure 4.23: Structure of ordinary and universal kriging predictions on CRS of size 500

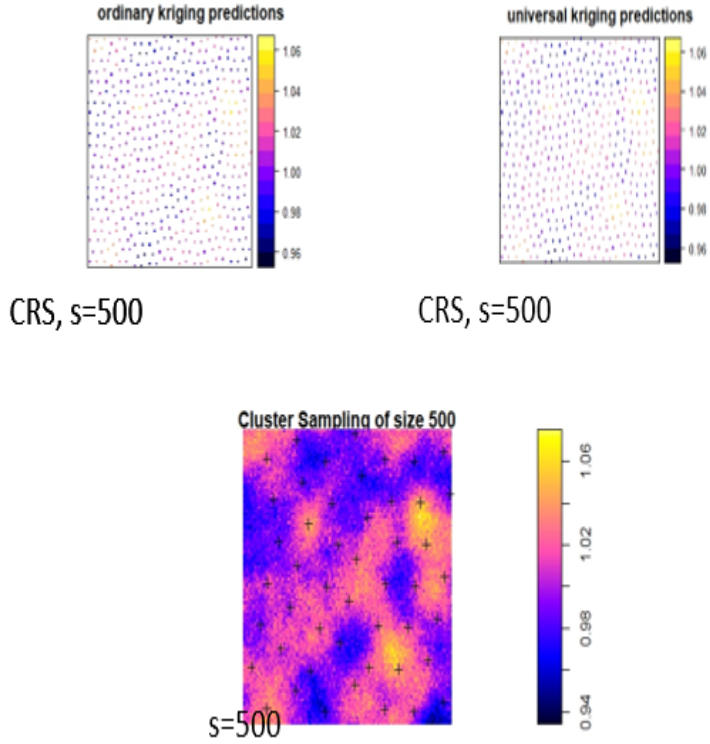


Figure 4.24: Structure of ordinary and universal kriging variance on CRS of size 500

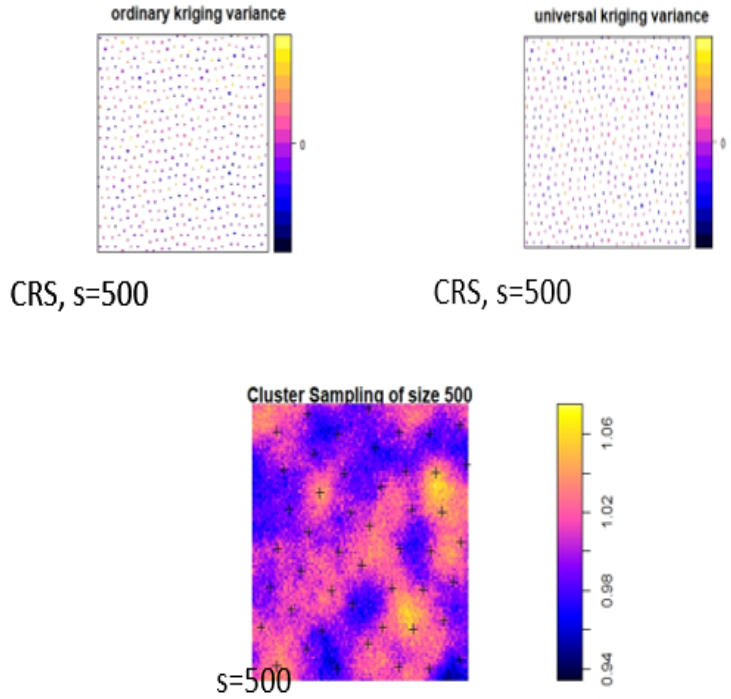


Figure 4.25: Structure of ordinary and universal kriging predictions on STRS of size 50

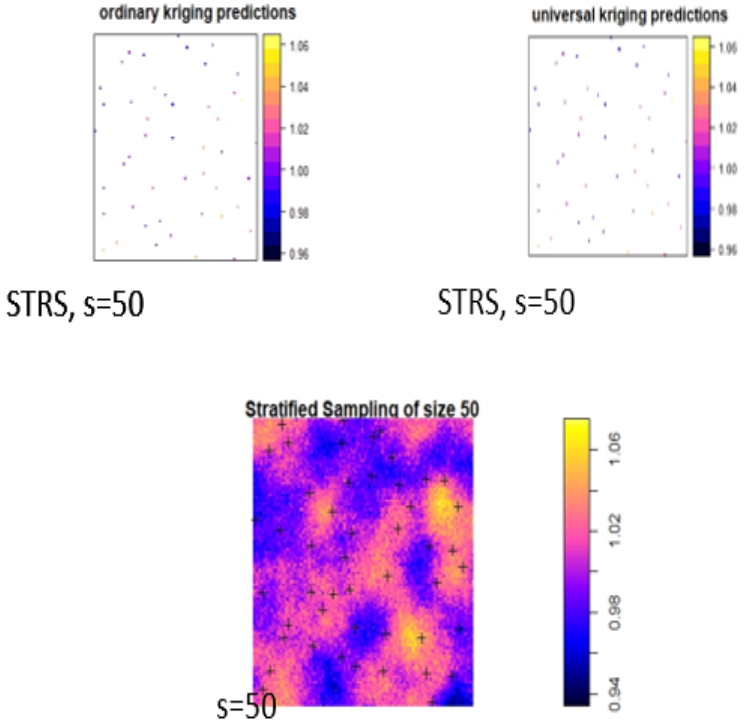


Figure 4.26: Structure of ordinary and universal kriging variance on STRS of size 50

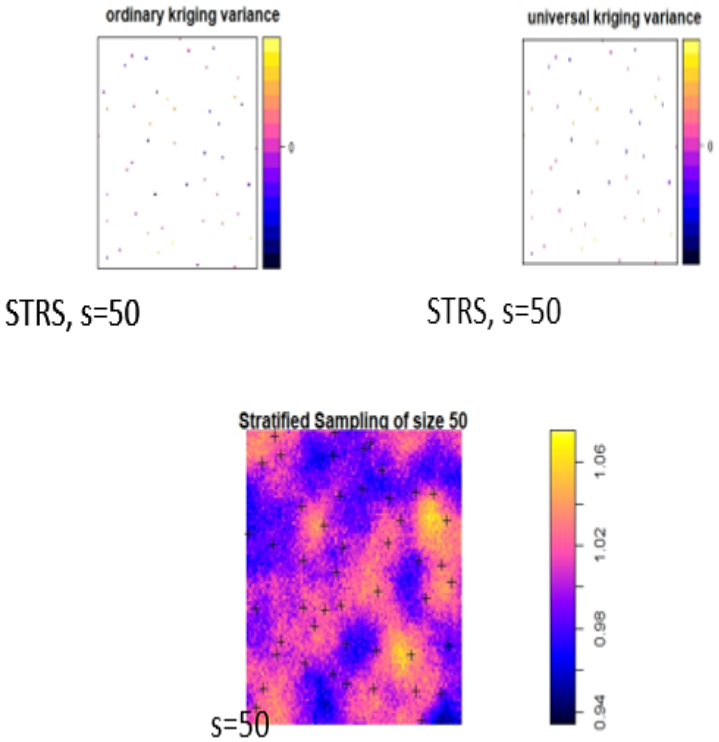


Figure 4.27: Structure of ordinary and universal kriging predictions on STRS of size 150

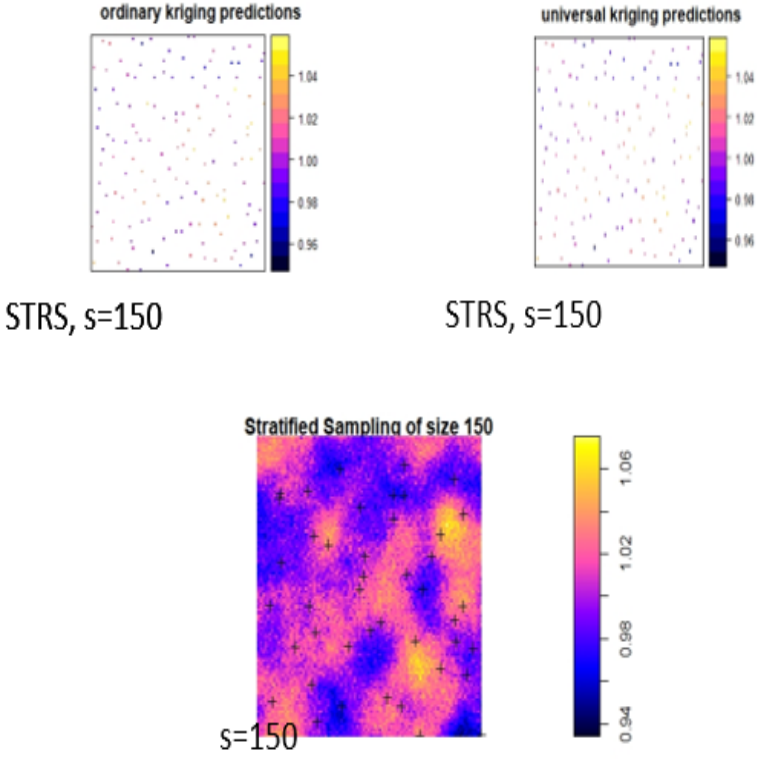


Figure 4.28: Structure of ordinary and universal kriging variance on STRS of size 150

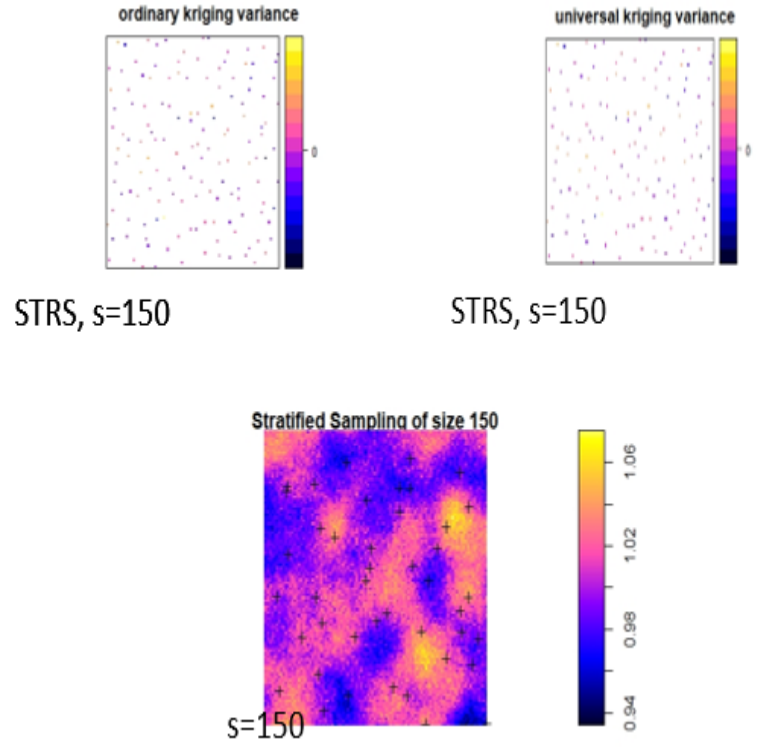


Figure 4.29: Structure of ordinary and universal kriging predictions on STRS of size 500

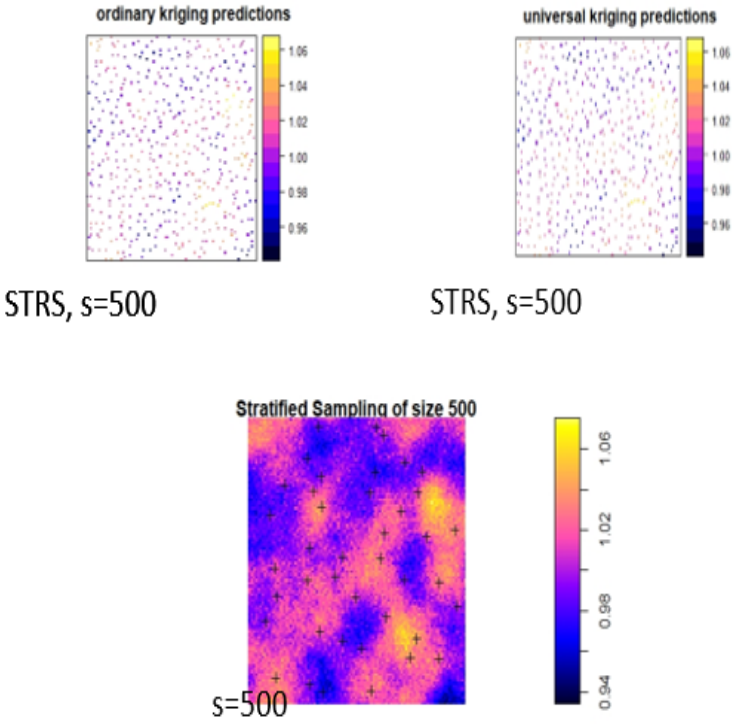
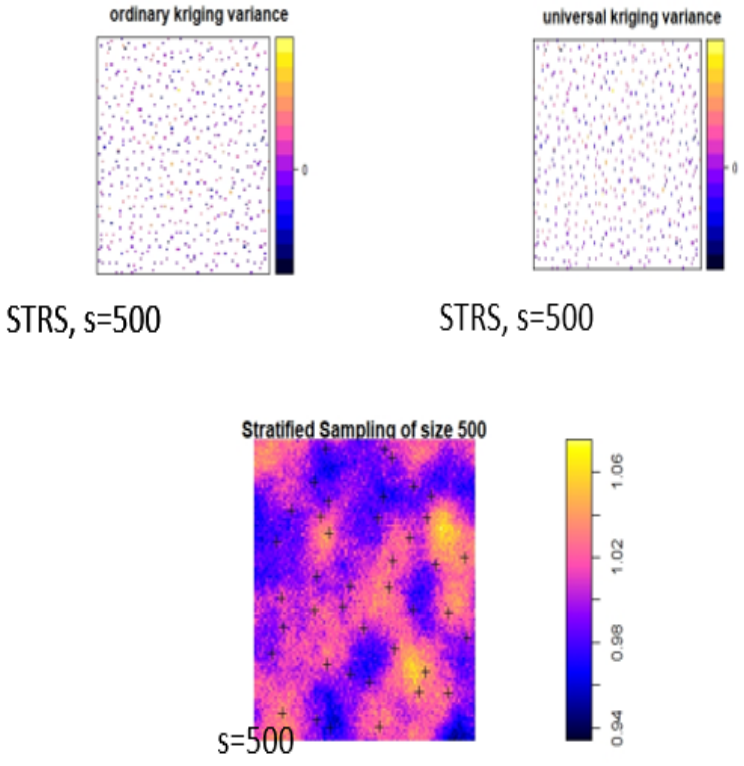


Figure 4.30: Structure of ordinary and universal kriging variance on STRS of size 500



4.5 Discussion

On figure 4.7 above, SRS scheme with sample size 50 has a poor variogram fit whiles sample 150 has better fit than sample 50 and 500. SRS scheme with sample size 50 did not give an accurate fit due to the scanty number of spatial attributes compared to the original fit. Though the fit under sample 500 is better than that of 50, but the fit is not too good possibly due to data redundancy. The discussion seemingly agrees with Delmelle et al. (2014) in his handbook illustration, that spatial variability will not be captured if we under-sample and oversampling also brings redundancy in the data points, and that attention must be paid on not only quantity of the data but also the locations.

However, on figure 4.8 where cluster random sampling scheme was used, variogram uncertainty obviously reduced with increased in sample size from 50 to 150 and 500 as the variogram fit under the sample 500 was closer to the original variogram fit shown above. Figure 4.9 illustrates similar features as that of figure 4.8 signifying the high level of variogram uncertainty surrounding variograms with small samples than those with large samples. Just as sample size, sampling scheme essentially plays pivotal role on variogram uncertainty. On figure 4.10, where the sample size was 50, no scheme of sampling was able to produce an accurate variogram fit however, simple random sampling produced a better variogram fit than stratified random sampling and cluster random sampling schemes when all the three schemes of sampling were compared to the original variogram fit.

These findings do not agree with Delmelle et al. (2014) in his study in which he found that, the combined sample system is at least more accurate than the random sample when comparing the observed differences in the two systems; its moderate performance is a function that increases sample size by sample size compared to random sample. Ideally, the size of the sample area should increase in areas that show significant area variability because the values of the closest samples will show strong similarities and care must be taken not to oversample as well.

It is worth noting that, Delmelle et al. (2014) focused on comparing relative variances of these sampling schemes other than comparing the efficiency of their variograms fit with the variogram fit of the whole population data. His study could not produce a control variogram fit against sample variograms, hence rendering a different method of determining the most efficient sampling scheme in minimizing variogram uncertainty. That could account for the difficulty in determining the most efficient sampling scheme in his work. On figure 4.11 above, where the sample size was raised to 150, each one of the variograms fits; simple random sampling, cluster sampling and stratified sampling improved significantly however, simple random sampling design turned to be the most efficient sampling design when their variograms were compared to the global variogram fit. Simple random sampling design appeared to be the most accurate sample technique followed by cluster sampling design and then stratified sampling design when the data was relatively large.

Considering figure 4.12 above, only cluster random sampling and stratified random sampling techniques were efficient. At a very large sample size 500, only simple random sampling scheme did not produce a variogram fit that was equally as accurate as the global variogram fit, given a signal that uncertainty associated with a variogram could drastically be minimized when the sample size is large as well as the type of scheme of sampling adopted.

Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

From the above analysis and discussion, it is obvious that uncertainty or certainty of variogram largely depends on the size of spatial sample. Spatial variability of the geographical surface can be reduced if the sample size is large whiles taken into consideration the cost of acquiring the data. It is worth noting that spatial redundancy can also occur in areas where sample size is increased when spatial variability is very low. Moreover, the denser the spatial points, the better the estimates of the variogram. This accounted for variograms of higher sampled points being almost same the as the global variogram fit. On the other hand, the three schemes of sampling produced poor variogram fits at a point when sample size was 50 but at sample size 150, all the three schemes of sampling improved in their fitted variograms and the best fitted variogram came from simple random sampling scheme. However, when the sample size was increased to 500, all three except the fitted variogram from simple random sampling scheme produced a consistent and an accurate variogram fit. It is therefore obvious to conclude that the uncertainty of variogram also largely depends on the scheme of the sampling Geostatistician adopts for studies.

5.1.1 Findings

Variograms for systematic sampling scheme were not plotted by gstat package in R due to the fact that its data points are characterized by irregular intervals.

5.2 Recommendations

Based on the findings from this study, the following recommendations were made; Firstly, from the results and discussion of this research it is clear the choice of sample size and sampling scheme undermine informed decisions in many enterprises of life; disease mapping, crime combating, fishing, oil and gas exploration, mining among others. For instance, in mining industry, miners will need to take prudent decisions regarding the number of drills and what drilling scheme(strategy) to adopt in order to hit goal deposit(s) with minimum cost.

Secondly, researchers who have interest in this field of studies should think of finding out an effective and efficient method of using gstat in estimating variogram of irregular intervals between sample locations. Finally, further research should focus on comparing standard errors of kriging estimates in a way of identifying most appropriate sample size and sampling scheme for informed decision making with minimum cost.

References

Anderson, K., Sethajintanin, D., Sower, G., & Quarles, L. (2008). Field trial and modeling of uptake rates of in situ lipid-free polyethylene membrane passive sampler. *Environmental Science & Technology*, *42*(12), 4486–4493.

Bellhouse, D. (1977). Some optimal designs for sampling in two dimensions. *Biometrika*, *64*(3), 605–611.

Bourgault, G. (1997). Using non-gaussian distributions in geostatistical simulations. *Mathematical geology*, *29*(3), 315–334.

Caers, J. (2000a). Adding local accuracy to direct sequential simulation. *Mathematical Geology*, *32*(7), 815–850.

Caers, J. (2000b). Direct sequential indicator simulation. *Geostats*, 39–48.

Caers, J., Journel, A. G., et al. (1998). Stochastic reservoir simulation using neural networks trained on outcrop data. In *Spe annual technical conference and exhibition*.

Carlson, T. N., & Ripley, D. A. (1997). On the relation between ndvi, fractional vegetation cover, and leaf area index. *Remote sensing of Environment*, *62*(3), 241–252.

Christakos, G. (2012). *Random field models in earth sciences*. Courier Corporation.

Christakos, G. (2013). *Random field models in earth sciences*. Elsevier.

Cochran, W. G. (2007). *Sampling techniques*. John Wiley & Sons.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Dalenius, T., Hájek, J., & Zubrzycki, S. (1961). On plane sampling and related geometrical problems. In *Proceedings of the 4th berkeley symposium on probability and mathematical statistics* (Vol. 1, pp. 125–150).

Delmelle, E., Dony, C., Casas, I., Jia, M., & Tang, W. (2014). Visualizing the im-

pact of space-time uncertainties on dengue fever patterns. *International Journal of Geographical Information Science*, *28*(5), 1107–1127.

Deutsch, C. V., & Journel, A. G. (1998). Gslib. *Geostatistical software library and userŠs guide*, *369*.

Deutsch, C. V., Journel, A. G., et al. (1992). Geostatistical software library and userŠs guide. *New York*, *119*(147).

Ferreyra, R., Apezteguıa, H., Sereno, R., & Jones, J. (2002). Reduction of soil water spatial sampling density using scaled semivariograms and simulated annealing. *Geoderma*, *110*(3-4), 265–289.

Gatrell, A. (1979). Autocorrelation in spaces. *Environment and Planning A*, *11*(5), 507–516.

Goovaerts, P., et al. (1997). *Geostatistics for natural resources evaluation*. Oxford University Press on Demand.

Griffith, D. A. (1987). Spatial autocorrelation. *A Primer. Washington DC: Association of American Geographers*.

Griffith, D. A., & Amrhein, C. (1997). *Multivariate statistical analysis for geographers*.

Iachan, R. (1985). Plane sampling. *Statistics & probability letters*, *3*(3), 151–159.

Isaaks, E. H. (1992). The application of monte carlo methods to the analysis of spatially correlated data.

Journel, A. G. (1994a). Modeling uncertainty: some conceptual thoughts. In *Geostatistics for the next century* (pp. 30–43). Springer.

Journel, A. G. (1994b). Resampling from stochastic simulations. *Environmental and Ecological Statistics*, *1*(1), 63–91.

Krige, D. G. (1952). A statistical approach to some basic mine valuation problems on the witwatersrand, by dg krige, published in the journal, december 1951: interim reply by the author to the discussion. *Journal of the Southern African Institute of Mining and Metallurgy*, *52*(11), 264–266.

Madow, W. G. (1953). On the theory of systematic sampling, iii. comparison of centered and random start systematic sampling. *The Annals of Mathematical Statistics*,

101–106.

McBratney, A., Webster, R., & Burgess, T. (1981). The design of optimal sampling schemes for local estimation and mapping of of regionalized variablesŮi: Theory and method. *Computers & Geosciences*, *7*(4), 331–334.

Olea, R. A. (1984). Sampling design optimization for spatial functions. *Journal of the international Association for Mathematical Geology*, *16*(4), 369–392.

Quenouille, M. H., et al. (1949). Problems in plane sampling. *The Annals of Mathematical Statistics*, *20*(3), 355–375.

Ripley, B. (1984). *198 1. spatial statistics.* Wiley, New York.

Ripley, B. (1987). *Stochastic simulation. new york: John viley and sons.* Inc.

Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(2), 172–192.

Rogerson, P. A., Delmelle, E., Batta, R., Akella, M., Blatt, A., & Wilson, G. (2004). Optimal sampling design for variables with varying spatial importance. *Geographical Analysis*, *36*(2), 177–194.

Shimazaki, H., & Shinomoto, S. (2007). A method for selecting the bin size of a time histogram. *Neural computation*, *19*(6), 1503–1527.

Soares, A. (1998). Sequential indicator simulation with correction for local probabilities. *Mathematical geology*, *30*(6), 761–765.

Soares, A. (2001a). Direct sequential simulation and cosimulation. *Mathematical Geology*, *33*(8), 911–926.

Soares, A. (2001b). Direct sequential simulation and cosimulation. *Mathematical Geology*, *33*(8), 911–926.

Soltani, F., Afzal, P., & Asghari, O. (2013). Sequential gaussian simulation in the sungun cu porphyry deposit and comparing the stationary reproduction with ordinary kriging. *Universal Journal of Geoscience*, *1*(2), 106–113.

Van Groenigen, J., Pieters, G., & Stein, A. (2000). Optimizing spatial sampling for multivariate contamination in urban areas. *Environmetrics: The official journal of the International Environmetrics Society*, *11*(2), 227–244.

Wang, J., Haining, R., & Cao, Z. (2010). Sample surveying to estimate the mean of a heterogeneous surface: reducing the error variance through zoning. *International Journal of Geographical Information Science*, *24*(4), 523–543.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists*. John Wiley & Sons.

Webster, R., & Wiley, M. O. (2001). Geostatistics for environmental scientists.

Zhang, H., Lan, Y., Lacey, R. E., Huang, Y., Hoffmann, W. C., Martin, D., & Bora, G. (2009). Analysis of variograms with various sample sizes from a multispectral image. *International Journal of Agricultural and Biological Engineering*, *2*(4), 62–69.

Zubrzycki, S. (1958). Remarks on random stratified and systematic sampling in a plane. In *Colloquium mathematicae* (Vol. 6, pp. 251–264).