Sampling Grid Shifting Algorithm: A Non-ergodic Spatial Bootstrap Technique for Regular and Irregular Sampling Patterns

Olatunde Mosobalaje[1], Oyinkepreye Orodu[1], and David Ogbe[2]

[1]Covenant University
[2]African University of Science and Technology

November 24, 2022

Abstract

Accounting for uncertainty in statistical model parameters is an essential part of geostatiscal reservoir characterization. While parameter uncertainty may be assessed in its ergodic form; the non-ergodic is a better characterization of the variability in the random field. Assessing non-ergodic parameter uncertainty requires re-sampling (bootstrapping) techniques. Existing techniques for such non-ergodic re-sampling are plagued with some limitations/complications. This paper therefore presents a spatial bootstrap algorithm that overcomes the limitations/complication. For a discretized field, the algorithm implements simultaneous displacements (shiftings) of all sampling points through the same distance vector. The shiftings are done across the dimensions of the field subject to the dimensionality of the sampling. In each dimension, the sampling points are shifted successively through a distance equivalent to the gridblock length in that dimension. At each shifting, a shifted sampling grid, of similar configuration as the original sampling grid, is generated. Using the shifted sampling grid, the algorithm re-samples a full-grid simulated realization of the field. The assumption of second-order stationarity implies that a sample from a shifted sampling grid is considered a repeated sample of the original sample. The algorithm has been scripted in R statistical computing environment and applied to an irregularly-sampled 3-D field with satisfactory results.

Sampling Grid Shifting Algorithm: A Non-ergodic Spatial Bootstrap Technique for Regular and Irregular Sampling Patterns

O. O. Mosobalaje [a,*], O. D. Orodu[a], D. Ogbe[b]

[a]Department of Petroleum Engineering, Covenant University, Ota, Nigeria.

[b]African University of Science and Technology, Abuja, Nigeria.

***Corresponding Author: E-mail: olatunde.mosobalaje@covenantuniversity.edu.ng**

Abstract

Accounting for uncertainty in statistical model parameters is an essential part of geostatiscal reservoir characterization. While parameter uncertainty may be assessed in its ergodic form; the non-ergodic is a better characterization of the variability in the random field. Assessing non-ergodic parameter uncertainty requires re-sampling (bootstrapping) techniques. Existing techniques for such non-ergodic re-sampling are plagued with some limitations/complications. This paper therefore presents a spatial bootstrap algorithm that overcomes the limitations/complication.

For a discretized field, the algorithm implements simultaneous displacements (shiftings) of all sampling points through the same distance vector. The shiftings are done across the dimensions of the field subject to the dimensionality of the sampling. In each dimension, the sampling points are shifted successively through a distance equivalent to the gridblock length in that dimension. At each shifting, a shifted sampling grid, of similar configuration as the original sampling grid, is generated. Using the shifted sampling grid, the algorithm re-samples a full-grid simulated realization of the field. The assumption of second-order stationarity implies that a sample from a shifted sampling grid is considered a repeated sample of the original sample. The algorithm has been scripted in R statistical computing environment and applied to an irregularly-sampled 3-D field with satisfactory results.

Keywords: Spatial bootstrap; non-ergodic, sampling; variogram; geostatistics; algorithm

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

1. Introduction

The estimation of natural resource volumes is a major objective of geological modeling. The assessment of uncertainty inherent in such volumetric estimates is in the domain of geostatistical modeling and stochastic simulation. Often, considerations for uncertainty in resource volume estimation are limited to attributes uncertainty. Several geostatistical realizations of volumetric attributes (porosity, net-to-gross thickness, saturation etc.) are simulated using fixed values of statistical model parameters such as sample mean, representative histogram and empirical variogram. Obtaining the true values of these statistical model parameters would require an exhaustive sampling of the geological field; this is prohibitive. In reality, estimates of these statistical model parameters are obtained from limited sample data. Estimates of statistical model parameters obtained from limited field sample data are necessarily uncertain (Babak and Deutsch, 2009). Random variations across samples would necessarily confer statistical uncertainty on the estimates obtained (Pardo-Iguzquiza and Dowd, 2001). More so, preferential and sparse or unrepresentative sampling introduces uncertainty into the estimates. Hence, the use of fixed values of uncertain model parameters in stochastic simulations would necessarily lead to a bias in the resource volume global uncertainty. Such bias would cause an underestimation of resource volume global uncertainty. While the subject matter of this paper is applicable to all statistical model parameters; its implementation in this paper is specifically on the uncertainty inherent in the sample mean. The sample mean is essential in defining the probability distributions of geological attributes.

While the subject matter of this paper is applicable to all statistical model parameters; it is here introduced within the context of the uncertainty in empirical (sample) variogram. Empirical variogram is very essential in geostatiscal modeling as it is required in optimal interpolation of attributes values at unsampled locations in a geological field. In a thorough assessment of uncertainty in empirical variogram, a distinction is made between ergodic variogram and non-ergodic variogram. Ergodic variogram is that which is averaged over multiple realizations of the random process (variogram model and conditioning data) that generated the field. Non-ergodic variogram is that which is averaged over multiple samples of a single realization of the random field. In essence, ergodic variogram is that of the random process that generated the field while non-ergodic variogram is that of the field itself; these two are not necessarily equal. The non-ergodic variogram is of more practical interest in geostatistical modelling because it better characterises the variability inherent in the field (Isaaks and Srivastava, 1988; cited in Brus and de Gruijter, 1994). Each of these variogram types has its associated uncertainty (fluctuations). Ergodic variogram uncertainty refers to the fluctuations of variogram estimates over multiple realizations of the random process underlying the field. On the other hand, non-ergodic variogram uncertainty refers to the fluctuations of variogram estimates over multiple samples of a single realization of the field. The uncertainty in non-ergodic variogram estimates is entirely due to sampling (i.e., random variations across samples). However, uncertainty in ergodic variogram encompasses fluctuations due to random variations of the random variables (over multiple realizations), in addition to sampling fluctuations. For both variogram measures, variogram uncertainty at a given lag interval is assessed as variance of multiple variogram estimates at that lag interval. Obtaining multiple variogram estimates at a given lag interval would necessarily require a technique for generating multiple (repeated) sample sets from the original sample set.

In statistical parlance, such techniques are known as re-sampling or bootstrap techniques. The methodology for obtaining multiple samples for ergodic variogram and associated uncertainty is straightforward. Multiple realizations of the random field are simulated using a plausible variogram model; each realization is sampled at the same set of sampling points. A plausible variogram model is a model that fits the empirical variogram of the data reasonably well. Non-ergodic multiple samples are obtained from a single realization of the random field but at multiple sets of sampling points. The existing techniques of obtaining multiple samples (re-sampling) of a single realization are plagued with some limitations/complications. The paper therefore presents a spatial bootstrap algorithm that overcomes the limitations and complications of the existing techniques. First, a concise review of the evolution of the previous re-sampling/bootstrap techniques and associated limitations/complications is presented. Thereafter, the description, functionality and procedure of the new algorithm formulated in this work are presented. Finally, the application of the algorithm to a certain irregularly-sampled 3-D geological field is presented.

2. Limitations of Existing Re-sampling Techniques

In classical statistics, a common technique for generating multiple sample sets from a given sample set is known as bootstrap (Efron and Tibshirani, 1993). The bootstrap technique assumes data independence; however, attributes data in a geological field are spatially correlated (dependent). Results presented by Deutsch (2004) show that the assumption of independence, in resampling correlated data, leads to underestimation of parameter uncertainty. Consequently, spatial bootstrap technique was developed to accommodate data correlation (Deutsch, 2004; Journel and Bitanov, 2004; Feyen and Caers, 2006). Babak and Deutsch (2009) pointed out two questionable aspects of the spatial bootstrap technique: a disregard for conditioning data and for the finite field domain. These lead to overestimation of variogram uncertainty as the unconditional simulation (underlying spatial bootstrap) give very different realizations with increasing spatial correlation (Babak and Deutsch, 2009). Moreover, the spatial bootstrap program (Deutsch, 2004) is in the league of ergodic re-sampling techniques as it entails multiple stochastic simulations of the random process in generating the several sample sets. The stochastic simulation occurs as the program draws sets of independent Gaussian values from the standard normal distribution. The random fluctuations due to the stochastic simulations also add to the variogram uncertainty. In essence, the unconditional nature of the simulation is not the only reason for the overestimation of variogram uncertainty when spatial bootstrap (Deutsch, 2004) is used. The use of stochastic realizations (even if conditional) would still necessarily lead to such overestimation; results presented in Section 4 of this paper support this assertion.

Derakhshan and Leuangthong (2006) generated multiple sample sets from a single simulated full-grid realization of a field by moving (shifting) a sampling grid over the field. While this shifting-like movement of the sampling grid is very easy to implement for regularly sampled fields, Derakhshan and Leuangthong (2006) noted that it could be highly intractable to implement in an irregularly-sampled field with clustered sample points. In reality, hydrocarbon reservoirs are commonly sampled irregularly with sampling bias for high porosity/permeability regions of the field. Babak and Deutsch (2009) proposed the concept of conditional finite-domain (CFD) to overcome the questionable aspects of spatial bootstrap. The CFD as a re-sampling technique is anchored on a notion similar to the movement of sampling grid as done by Derakhshan and Leuangthong (2006). However, instead of the shifting-like movement, Babak and Deutsch (2009) generated multiple sample sets by arbitrarily rotating the original sample points around an arbitrary center. While the CFD technique could work for irregularly sampled fields, its implementation by rotation of data locations has some limitations. First, the rotation of data location is very difficult to implement (Rezvandehy, 2016). This difficulty would even be more pronounced in a 3-D field as two rotation axes (azimuth and dip) would then be required. This could limit the CFD technique to 2-D fields only. Also, there are only very few number of such rotations possible before some of the sample locations are rotated outside the finite domain; hence only few re-sampling opportunities. Ensuring sample points do not fall outside the domain would increase the complexity of the technique; this is not discussed by the proponents of the technique. Increasing the re-sampling opportunities might require multiple full-grid realizations and repetition of the rotation; this would increase the variogram uncertainty. Finally, the rotation of data locations could not be implemented for anisotropic random fields as it assumes isotropy.

Among all these existing resampling techniques, the shifting-based technique of Derakhshan and Leuangthong (2006) appears to be the simplest; except for the stated difficulty in implementing it for irregularly-sampled fields. Also, the shifting-based technique is the only one that is strictly non-ergodic as it utilizes a single simulated realization of the field. It is to this end that this work now develops a simple algorithm to implement the shifting of both regular and irregular sampling grids within the context of 1-D, 2-D and 3-D random fields. This algorithm is known as Sampling Grid Shifting Algorithm (SGSA). The significance of the algorithm presented in this paper lies in its use for assessing the important non-ergodic uncertainty in statistical model parameters. In geostatiscal modelling, non-ergodic parameter uncertainty is a more realistic measure than its ergodic counterpart. Also, the non-ergodic variogram uncertainty obtained through this algorithm would be useful in assessing the adequacy of sampling designs.

3. Description of Sampling Grid Shifting

Given a sampling grid configuration (regular or irregular) and a set of sample data, the SGSA generates a suite of spatial coordinates for each of all possible shiftings of the sampling grid. Thereafter, the algorithm generates a repeated set of sample data by sampling a simulated full-grid realization of the field at each shifted sampling grid. Definitions of terms are needful here. Here, sampling grid configuration refers to the spatial position of all sample points relative to one another, in terms of separation distance and direction. In other words, the configuration of a sampling grid encapsulates both the number of sample points and the separation distance vector between the points. Shifting refers to the simultaneous movement/displacement of all sample points through same distance in the same direction. As long as all points are displaced equally in the same direction, the separation distance vector between them would not change; hence the configuration would not change. Therefore, a shifted sampling grid (obtained by shifting an original sampling grid) would have the same configuration with the original sampling grid. Truly, the spatial coordinates of sample points in the sampling grid change as a result of shifting; but the grid configuration remains the same. A shifting therefore refers to the resulting sampling grid at any instance of such displacement. The concept of sampling grid shifting is illustrated in Figure 1 with a three-point sampling grid in a 6×5 discretized 2-D field. Shifting 1 is obtained by an eastward displacement of all sample points through a distance equivalent to gridblock thickness in east direction. Shifting 2 is obtained by a southward displacement of all sample points through a distance equivalent to gridblock thickness in south direction. There are other possible shiftings for this sampling grid in this field. All sampling grids (original and shifted) have same configuration. Again, the SGSA generates the spatial coordinates for each of all such possible shiftings. Multiple samples of a single realization of the field are then obtained by sampling the simulated full-grid field at such suites of spatial coordinates. The assumption of second-order stationarity makes it possible to consider a sample obtained from a shifted sampling grid as being equivalent to a repeated sample of the original sample. By the assumption of stationarity, all samples with same configuration will have same mean, same covariance matrix, and consequently same distribution (Babak and Deutsch, 2009). Also, by the assumption of intrinsic stationarity, the resulting variogram estimates from these multiple samples would belong to the same distribution.

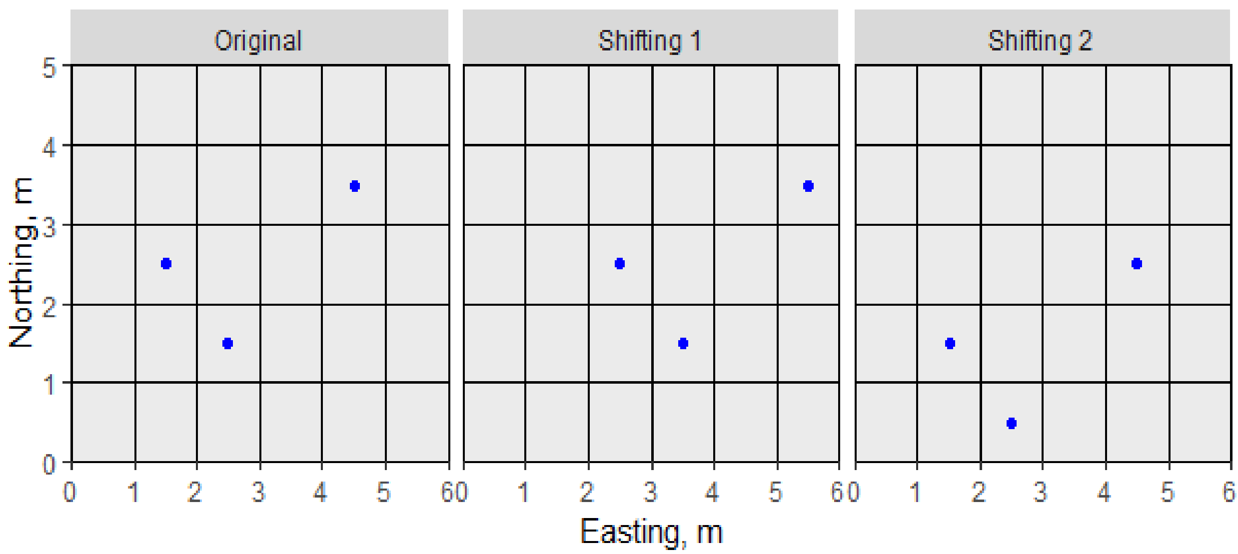

Figure 1: A depiction of sampling grid shifting

4. Functionality of the SGSA

The SGSA implements the sampling grid shiftings in a systematic and ordered sequence. The resulting suite of spatial coordinates for each shifting is stored in that ordered sequence. For a sampling grid, the suite of spatial coordinates may contain up to seven (7) items depending on dimensionality: the x-, y-, z-indices; the natural ordering indices and the easting, northing and depths (altitude) coordinates. First, the entire field is discretized into gridblocks of desired dimensions. A full-grid (exhaustive) realization of the random field is then conditionally simulated using the original sample data and the variogram model obtained from the data. The indices of gridblocks containing sample points (i.e. sample gridblocks) are thereafter obtained, both in engineering ordering (x, y, z) and natural ordering. The index of the outermost sample gridblock in each cardinal direction (westward, eastward, southward, northward, upward, downward) are identified. Based on these outermost sample gridblocks, the numbers of shifting steps permissible up to the last gridblock in each direction are determined thus:

$$WS = WSBI_x \dots 1$$

$$ES = n_x - ESBI_x \dots 2$$

$$SS = SSBI_y \dots 3$$

$$NS = n_y - NSBI_y \dots 4$$

$$US = USBI_z \dots 5$$

$$DS = n_z - DSBI_z \dots 6$$

$$S_{Total} = (WS + ES) \times (SS + NS) \times (US + DS) \dots 7$$

In the equations above, WS, ES, SS, NS, US and DS are the number of possible westward, eastward, southward, northward, upward and downward shiftings, respectively. $WSBI_x, ESBI_x, SSBI_y, NSBI_y, USBI_z$ and $DSBI_z$ are the indices of the outermost sample block in the indicated direction. S_{Total} is the total number of possible shiftings in the entire field; this includes the original sample grid. These indices are illustrated for a 10×10 discretised 2D field depicted in Figure 2. Also, n_x, n_y and n_z are the numbers of gridblocks in the discretized field along the indicated directions.

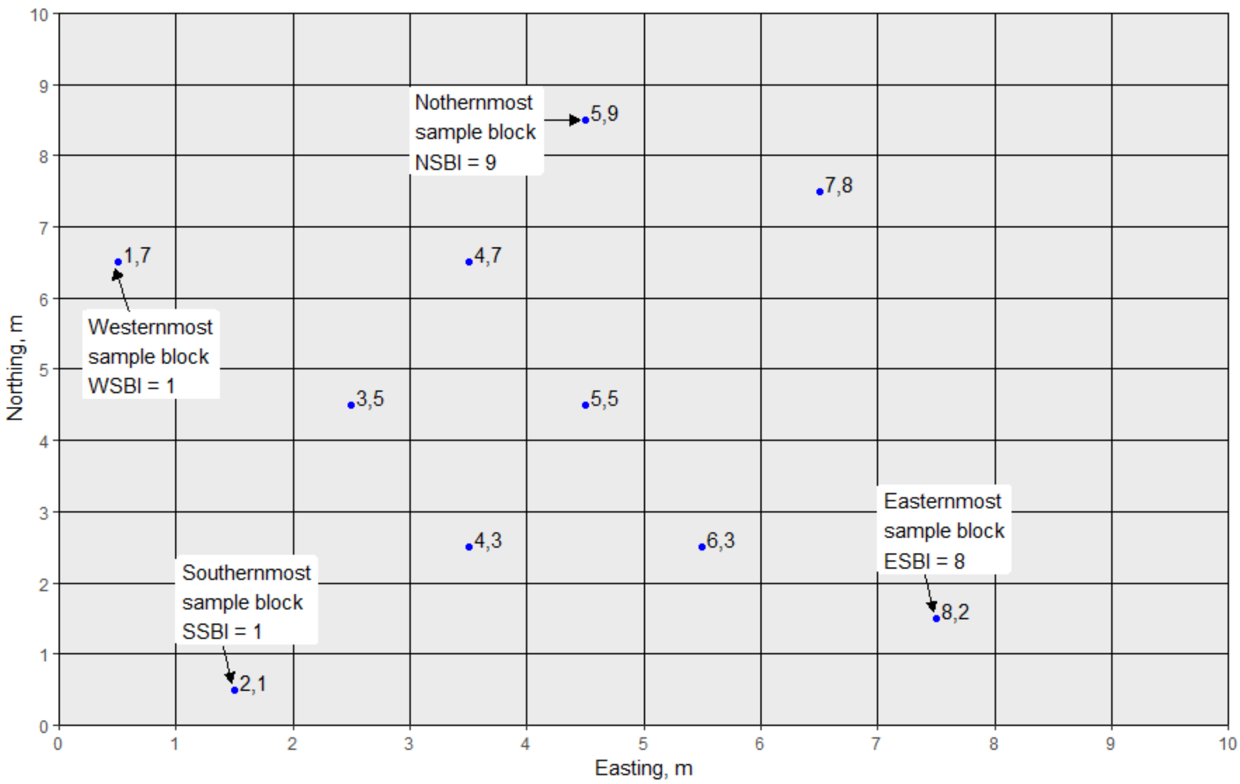

Figure 2: Depiction of outermost sample block indices

The algorithm works out these shiftings in a sequence/hierarchy of directions. In the first-tier shiftings, the westward shiftings are first implemented; the original sampling grid is taken as the first of the westward shiftings. Thereafter, the eastward shiftings are implemented. The suites of spatial coordinates resulting from all westward and eastward shiftings are then aggregated into a single unit known as the latitudinal shiftings. In the second-tier shiftings, each suite of the aggregated latitudinal shifting is subjected to longitudinal (southward and northward) shiftings. Again, the suites of spatial coordinates from these longitudinal shiftings are aggregated into a single unit known as planar shiftings. In the third-tier shiftings, the planar shiftings are subjected to vertical (upward and downward) shiftings. The suites of spatial coordinates of the planar shiftings and the vertical shiftings thereof are ultimately aggregated into a single unit known as grand shiftings. The following equations capture the number of possible shiftings in each hierarchy.

$$S_{lat} = WS + ES \quad \ldots 8$$
$$S_{longit} = SS + NS \quad \ldots 9$$
$$S_{Planar} = S_{longit} \times S_{lat} \quad \ldots 10$$
$$S_{Total} = (US + DS) \times S_{Planar} \quad \ldots 11$$

where S_{lat}, S_{longit}, S_{Planar} are the number of possible latitudinal, longitudinal and planar shiftings.

In each tier of the hierarchical shiftings, the implementation of the shifting is very straightforward. Looping through the number of possible shiftings in any direction, the algorithm successively adds/subtracts 1 to/from the directional index (i, j or k) in the concerned direction. Similarly, successive additions/subtractions of 1, n_x or $n_x n_y$ are made to/from the natural ordering indices (N_{order}) in first-, second- and third-tier shiftings, respectively. Also, gridblock dimensions $\Delta x, \Delta y \; and \; \Delta z$ are successively added/subtracted to/from the coordinates (eastings, northings or depths) in the concerned directions. The expressions in Table 1 summarise these implementations. The superscripts are the loop counter corresponding to successive shiftings in a given direction. For each shifting implemented, the algorithm samples the simulated full-grid realization at the spatial coordinates generated for that shifting. Each suite of spatial coordinates generated as well as the data sampled thereat makes up a repeated sample of the field. As a repeated sample is generated, its content (coordinates and data) is written to a file of desired format (e.g. .csv). These files are the runtime output of the algorithm. However, the grand shifting being the suites of coordinates for all shiftings, is the final output of the algorithm.

Table 1: SGSA implementation expressions

Latitudinal shiftings	Longitudinal shiftings	Vertical shiftings
$i^{c+1} = i^c \pm 1$	$i^{c+1} = i^c$	$i^{c+1} = i^c$
$j^{c+1} = j^c$	$j^{c+1} = j^c \pm 1$	$j^{c+1} = j^c$
$k^{c+1} = k^c$	$k^{c+1} = k^c$	$k^{c+1} = k^c \pm 1$
$N_{order}^{c+1} = N_{order}^{c} \pm 1$	$N_{order}^{c+1} = N_{order}^{c} \pm n_x$	$N_{order}^{c+1} = N_{order}^{c} \pm n_x n_y$
$Easting^{c+1} = Easting^c \pm \Delta x$	$Easting^{c+1} = Easting^c$	$Easting^{c+1} = Easting^c$
$Northing^{c+1} = Northing^c$	$Northing^{c+1} = Northing^c \pm \Delta y$	$Northing^{c+1} = Northing^c$
$Depth^{c+1} = Depth^c$	$Depth^{c+1} = Depth^c$	$Depth^{c+1} = Depth^c \pm \Delta z$

The shiftings are numbered and their respective suites of spatial coordinates are arranged in the grand shiftings in a logical order. The spatial coordinates of the original sampling grid is positioned at the middle (not necessarily the median location) of the arrangement. This original grid is then flanked at the left by the westward shiftings, starting with the first westward shifting. This same arrangement is implemented for the eastward shifting flanking the original grid at the right. Subsequently, the southward shiftings are made to flank the westward shiftings at the left while the northward shiftings are made to flank the eastward shiftings at the right. Finally, the upward shiftings are placed to the left of the southward shiftings, and the downward shiftings are positioned on the right of the northward shiftings. This arrangement is illustrated in Figure 3 for a three-point sampling grid in a 6×5×3 three-dimensional field. For this simple case, there are twenty seven (27) possible shiftings as computed using Equation 7. Only the nine (9) planar shiftings (Shiftings 10 – 18) are presented in the figure. Shiftings 1 – 9 are the upward shiftings while Shiftings 19 – 27 are the downward shiftings.

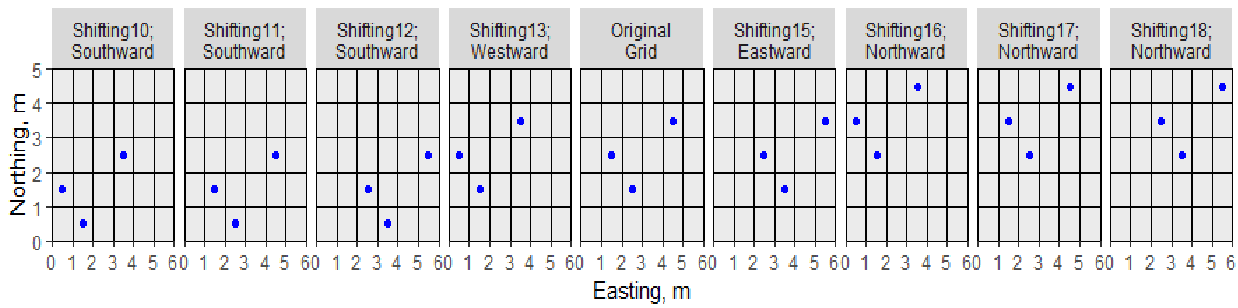

Figure 3: An array of all possible planar shiftings in a certain 3D field

5. SGSA Workflow

The foregoing functionality of the SGSA is now summarized in the following step-by-step procedure.

i. Load the file containing the coordinates and data of the original sample.

ii. Discretize the entire domain into gridblocks of desired dimensions.

iii. Simulate (conditionally) a full-grid realization using the data and a variogram model obtained from the data.

iv. Obtain indices of sample gridblocks. Append the indices to the coordinates.

v. Compute number of possible shiftings in each direction, hierarchy, and the total possible shiftings using Equations 1 – 11.

Westward Shiftings (starting with the original grid)

vi. Determine the beginning and end of the range of positions for the westward shiftings.

$$West_{ends} = [(US - 1)S_{Planar}] + [(SS - 1)S_{lat}] + WS$$

$$West_{begins} = West_{ends} - WS + 1$$

vii. Implement the following steps in a repetitive loop counting from 1 to WS. Let w be the loop counter.

a) Determine the position for the w^{th} westward shifting of the original sampling grid:

$$Position_{west} = West_{ends} - (w - 1)$$

b) Generate the suite of spatial coordinates for the w^{th} westward shifting using equations in Table 1

c) Assign the generated spatial coordinates to the appropriate position on the latitudinal and planar shiftings placeholders.

d) Sample the full-grid realization at the generated spatial coordinates.

e) Write the generated spatial coordinates and the sampled data to a .csv file.

Eastward Shiftings

viii. If $ES > 0$, perform Steps ix – x; else, move to Step xi

ix. Determine the beginning and end of the range of positions for the eastward shiftings.

$East_{begins} = [(US - 1)S_{Planar}] + [(SS - 1)S_{lat}] + WS + 1$

$East_{ends} = East_{begins} + ES - 1$

x. Implement the following steps in a repetitive loop counting from 1 to ES. Let e be the loop counter.

a) Determine the position for the e^{th} eastward shifting:

$Position_{east} = East_{begins} + (e - 1)$

b) Generate the suite of spatial coordinates for the e^{th} eastward shifting using equations in Table 1.

c) Assign the generated spatial coordinates to the appropriate position on the latitudinal and planar shiftings placeholders.

d) Sample the full-grid realization at the generated spatial coordinates.

e) Write the generated spatial coordinates and the sampled data to a .csv file.

Southward Shiftings

xi. If $SS > 1$, perform Steps xii – xiii; else, move to Step xiv

xii. Determine the beginning and end of the range of positions for the southward shiftings.

$South_{ends} = [(US - 1)S_{Planar}] + [(SS - 1)S_{lat}]$

$South_{begins} = South_{ends} - [(SS - 1)S_{lat}] + 1$

xiii. Implement the following steps in a two-tier nested repetitive loop. The outer loop counts from 1 to $SS - 1$ while the inner loop counts from 1 to S_{lat}. Let s and l be the outer and the inner loop counters, respectively.

a) Determine the position for the s^{th} southward shifting of the l^{th} latitudinal shifting:

$Position_{south} = South_{ends} - [(s - 1)S_{lat}] - (l - 1)$

b) Generate the suite of spatial coordinates for the s^{th} southward shifting of the l^{th} latitudinal shifting using equations in Table 1

c) Assign the generated spatial coordinates to the appropriate position on the planar shiftings placeholder.

d) Sample the full-grid realization at the generated spatial coordinates.

e) Write the generated spatial coordinates and the sampled data to a .csv file.

Northward Shiftings

xiv. If $NS > 0$ perform Steps xv – xvi; else, move to Step xvii

xv. Determine the beginning and end of the range of positions for the northward shiftings.

$North_{begins} = [(US - 1)S_{Planar}] + SS \times S_{lat} + 1$

$North_{ends} = North_{begins} + NS \times S_{lat} - 1$

xvi. Implement the following steps in a two-tier nested repetitive loop. The outer loop counts from 1 to NS while the inner loop counts from 1 to S_{lat}. Let n and l be the outer and the inner loop counters, respectively.

a) Determine the position for the n^{th} northward shifting of the l^{th} latitudinal shifting:

$$Position_{north} = North_{begins} + [(n - 1)S_{lat}] + (l - 1)$$

b) Generate the suite of spatial coordinates for the n^{th} northward shifting of the l^{th} latitudinal shifting using equations in Table 1

c) Assign the generated spatial coordinates to the appropriate position on the planar shiftings placeholder.

d) Sample the full-grid realization at the generated spatial coordinates.

e) Write the generated spatial coordinates and the sampled data to a .csv file.

Upward Shiftings

xvii. If $US > 1$ perform Steps xviii – xix; else, move to Step xx

xviii. Determine the beginning and end of the range of positions for the upward shiftings.

$Up_{ends} = [(US - 1)S_{Planar}]$

$Up_{begins} = 1$

xix. Implement the following steps in a three-tier nested repetitive loop. The outermost loop counts from 1 to $US - 1$; the middle loop counts from 1 to S_{longit}; and the innermost loop counts from 1 to S_{lat}. Let u, p and l be the outermost, middle and the innermost loop counters, respectively.

a) Determine the position for the u^{th} upward shifting of the p^{th} longitudinal shifting of the l^{th} latitudinal shifting:

$Position_{up} = Up_{ends} - [(u - 1)S_{Planar}] - [(p - 1)S_{lat}] - (l - 1)$

b) Generate the suite of spatial coordinates for the u^{th} upward shifting of the p^{th} longitudinal shifting of the l^{th} latitudinal shifting using equations in Table 1.

c) Assign the generated spatial coordinates to the appropriate position on the upward shiftings placeholder.

d) Sample the full-grid realization at the generated spatial coordinates.

e) Write the generated spatial coordinates and the sampled data to a .csv file.

Downward Shiftings

xx. If $DS > 0$ perform Steps xxi – xxii; else, move to Step xxiii

xxi. Determine the beginning and end of the range of positions for the downward shiftings.

$Down_{begins} = US \times S_{Planar} + 1$

$Down_{ends} = S_{Total}$

xxii. Implement the following steps in a three-tier nested repetitive loop. The outermost loop counts from 1 to DS; the middle loop counts from 1 to S_{longit}; and the innermost loop counts from 1 to S_{lat} Let d, p and l be the outermost, middle and the innermost loop counters, respectively.

a) Determine the position for the d^{th} downward shifting of the p^{th} longitudinal shifting of the l^{th} horizontal shifting:

$Position_{down} = Down_{begins} + (d-1)S_{Planar} + (p-1)S_{lat} + (l-1)$

b) Generate the suite of spatial coordinates for the d^{th} downward shifting of the p^{th} longitudinal shifting of the l^{th} horizontal shifting using equations in Table 1.

c) Assign the generated spatial coordinates to the appropriate position on the downward shiftings placeholder.

d) Sample the full-grid realization at the generated spatial coordinates.

e) Write the generated spatial coordinates and the sampled data to a .csv file.

Final Aggregation

xxiii. Aggregate the upward shiftings, planar shiftings and downward shiftings into the grand shifting and write to a .csv file.

The flowcharts for this procedure are here presented in Figures 4 – 7.

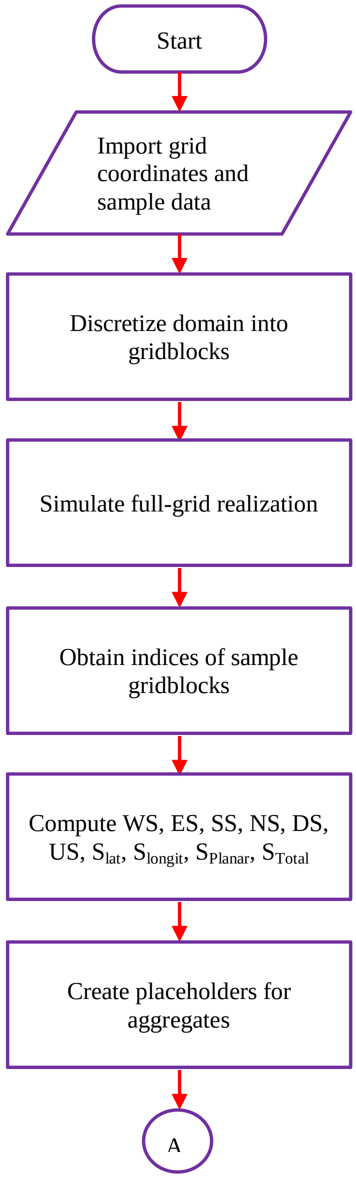

Figure 4: SGSA flowchart – preliminary steps

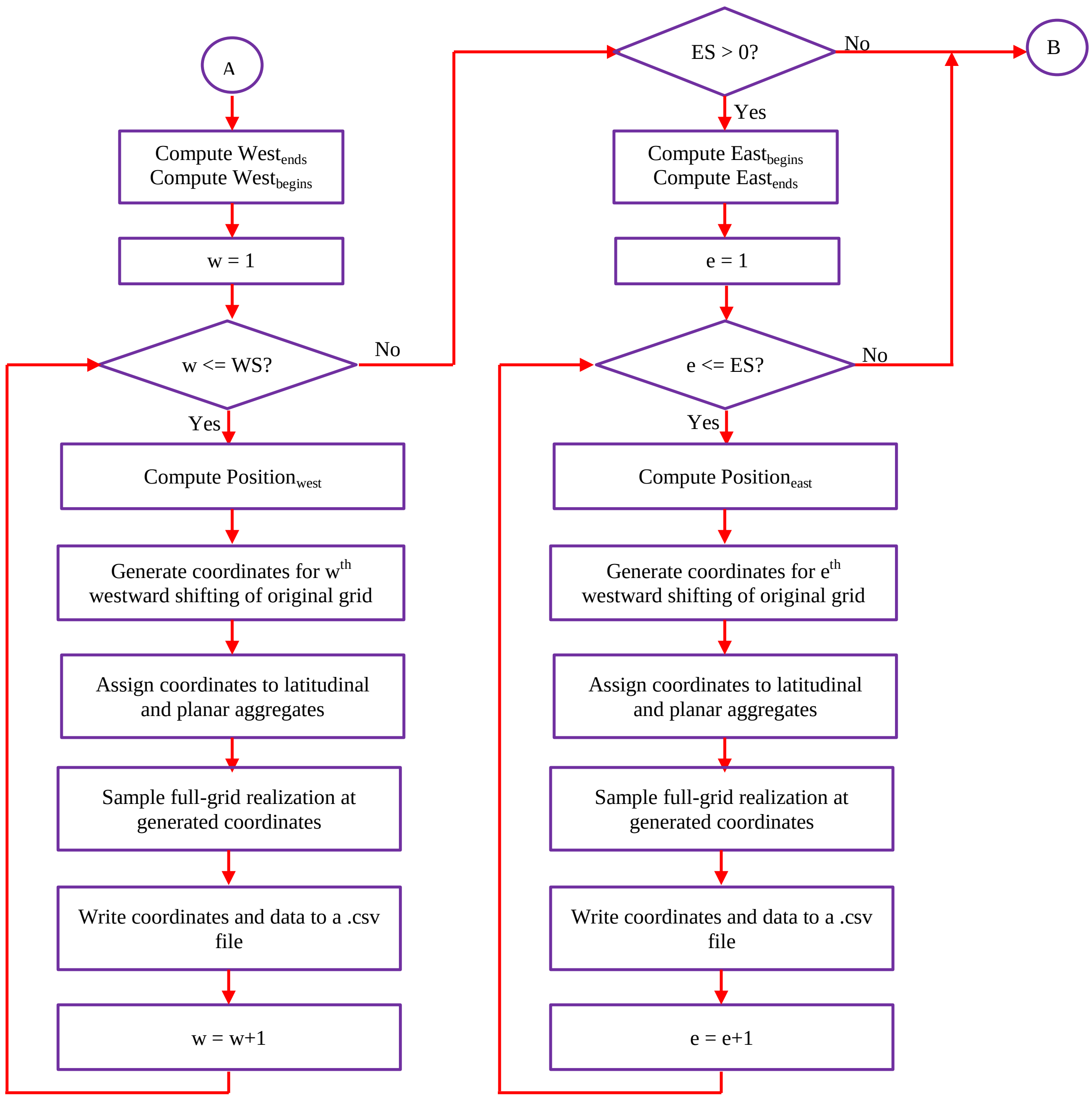

Figure 5: SGSA flowchart – first-tier shiftings

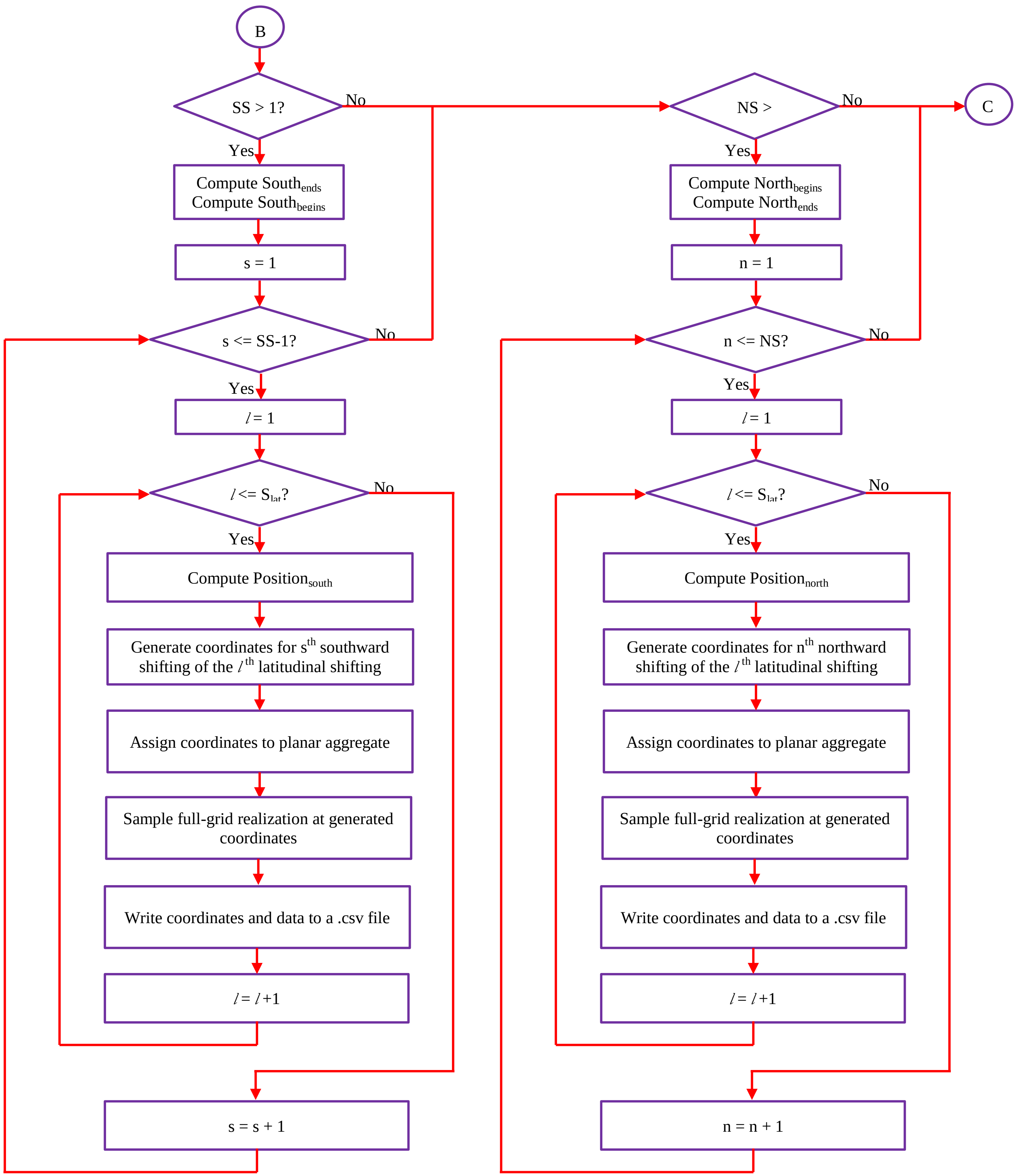

Figure 6: SGSA flowchart – second-tier shiftings

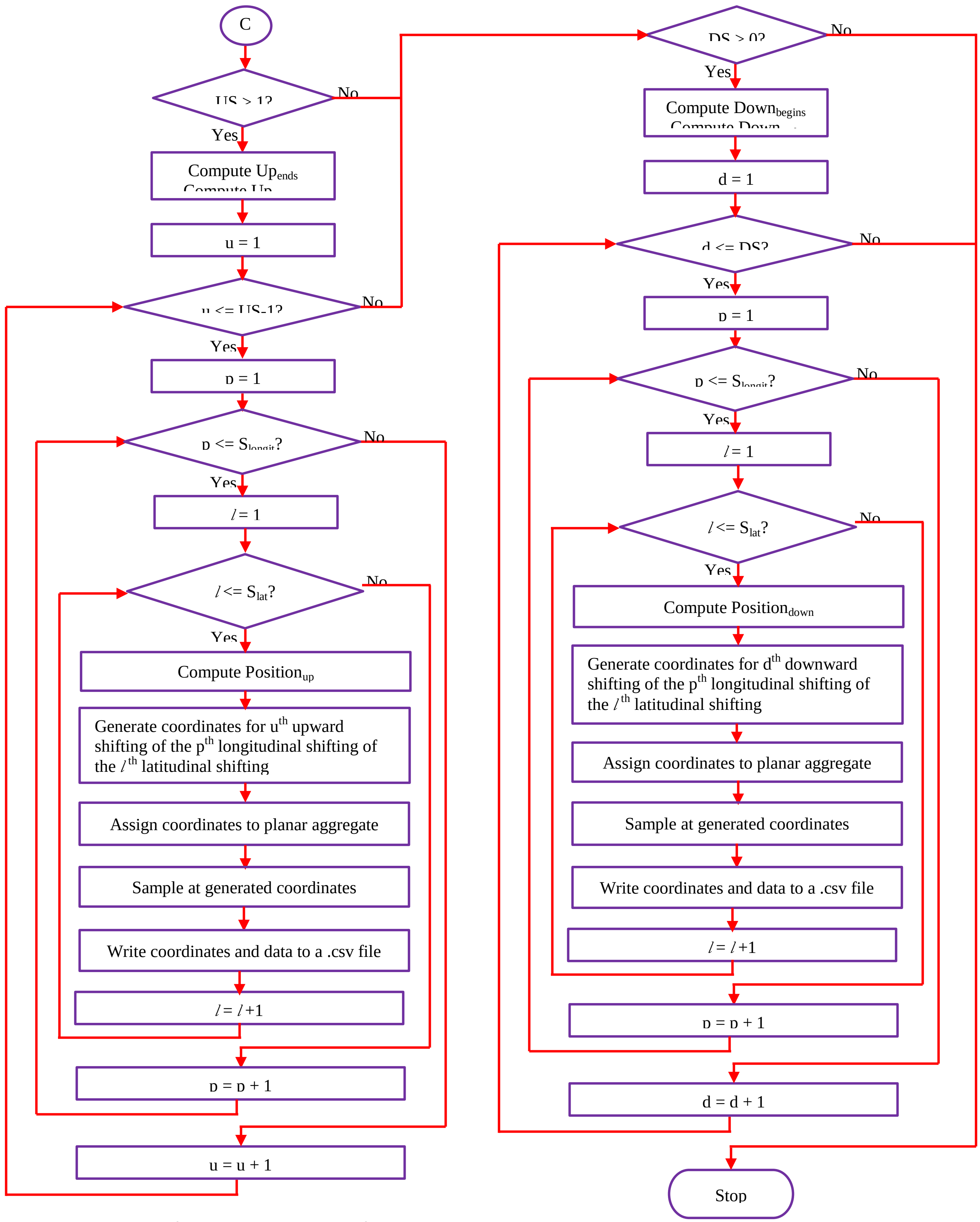

Figure 7: SGSA flowchart – third-tier shiftings

6. SGSA Implementation in R

The SGSA has been scripted in version 3.6.3 of R (R Core Team, 2020), a language and environment for statistical computing and graphics. The source code is presented as a function named '`sgsa`' and is available for download at the GitHub repository of the primary author (https://github.com/TTOWG/sgsa). The function file (`function_sgsa.R`) can be downloaded into users' R workspace and then be loaded into the calling environment with the command `source("function_sgsa.R")`. Once loaded, the function can be called using the command `sgsa()` with required arguments (input parameters) listed in the parenthesis. The input parameters and the output files of `sgsa` are here described.

The first input of `sgsa` is `sampledata`. It is a data frame (data table) containing the sample point coordinates and the attribute values of the original sample data to be bootstrapped. Depending on the dimensionality at hand, `sampledata` may contain up to four columns (4). Relevant coordinate values should occupy the first set of columns in the order x-, y-, z-coordinates. Attributes values should occupy the last column. The column headers should be named as `x_coord`, `y_coord`, `z_coord`, and `attribute`, respectively. Any direction not relevant to the field (e.g., z-direction in an x-y 2D field) should simply be omitted in `sampledata`. The function accommodates cases of redundant directions – the field being of higher dimensionality than the sampling grid. For example, a field may be 3-D (x-, y-, z-directions) in space while the attribute is invariant in z-direction and therefore sampled only in x-y plane. This makes the z-direction relevant to the field but redundant in the context of bootstrapping the sample data. In such cases, `sampledata` should still be prepared as a 3-D case with the redundant coordinate values set to a constant value for all the data points. The constant value may be set as the value of the redundant direction corresponding to the sampling plane. The next three input parameters of `sgsa` are `x_origin`, `y_origin` and `z_origin`. These specify the coordinates of the origin of the field. The default values of these origin coordinates have been set to zero. Hence, users may omit any or all of these inputs in a call to `sgsa`, if the default values are appropriate. Irrelevant or redundant direction(s) should also be omitted. The numbers of gridblocks into which the field should be discretized are to be specified using inputs `nx`, `ny`, `nz`. Each of these has been set to a default value of 1. The default value, 1, is appropriate for any direction that is not relevant to the field at hand and/or any direction that is redundant in the context the bootstrapping task at hand. Hence, in calling `sgsa`, users should omit any of `nx`, `ny`, `nz` corresponding to an irrelevant or redundant direction. The gridblock sizes are specified in the next three inputs: `deltaX`, `deltaY` and `deltaZ`. These also have been set to default values of zero. The default value of zero is appropriate for a direction that is not relevant to the field. Hence, users should specify non-default values for directions relevant to the field being considered and simply omit the irrelevant direction(s). Concisely, leaving the number and size of gridblock at default values (of 1 and 0, respectively) for any direction would render that direction irrelevant to the field being considered. However, leaving the number of gridblock at default value (of 1) but specifying a non-default value (> 0) for gridblock size in any direction would render that direction relevant to the field but redundant in the context of sampling and bootstrapping.

The last four input parameters of `sgsa` are `vargmodel`, `beta`, `nmin` and `nmax`. These respectively correspond to `model`, `beta`, `nmin` and `nmax` parameters of `gstat` function in gstat – an open-source R package for geostatistical simulations (Pebesma, 2004). In the `sgsa` function script, the `gstat` function is called to create an object that contains all information required for the full-grid simulation of the field. Parameter `vargmodel` should be the variogram model obtained from the data and should be defined by calling `vgm` in the gstat package. Parameter `beta` is only applicable if the full-grid simulation of the field is to be based on simple kriging. In that case, it should be specified as the expected value of the attribute. Parameters `nmin` and `nmax` are the minimum and maximum number of nearest data observations to be used in simulation, respectively. The values of these fourteen (14) arguments of `sgsa` could be listed in the order of their positions as discussed here, if none is to be omitted in the `sgsa` call. In such positional listing, the parameter names need not be indicated as the function would assign the listed values to their respective names. However, the arguments should be listed with names and values, if some are to be omitted in the call.

The runtime output files of function `sgsa` are delimited data files of the *.csv* format. Each of these files contain the sample gridblocks indices, sample points coordinates and attribute values for a particular repeated sample generated. In other words, there are as many of these files as there are repeated sample generated. The name of each file indicates the order (position) of the sampling grid shifting that generated the file; for example, Sample_1.

7. Application to Sample Mean Uncertainty in an Irregularly-sampled 3-D Field

In order to demonstrate its applicability, the SGSA is here implemented to assess the uncertainty in the sample mean of porosity attribute of a bitumen field. The field, located at Agbabu, south-west Nigeria, has been the subject of recent spatial correlation studies (Mosobalaje et al., 2019a, 2019b, 2019c). The application of the SGSA to non-ergodic variogram uncertainty has been earlier reported in Mosobalaje et al. (2019c) and is therefore not duplicated here. The sample mean and empirical variogram are the parameters required to fully define the probability distribution of spatially-correlated attributes that follow multivariate Gaussian distribution. Incorporating the uncertainty in these parameters into geostatistical simulations would reduce bias in the resource volumes global uncertainty.

7.1 Field and Sample Data Description

Agbabu field is part of the vast deposits of heavy oil and natural bitumen in the Dahomey Basin. The Dahomey basin is a costal sedimentary basin that spans from Ghana-Ivory Coast border to western Nigeria. Figure 8 is the geologic map of the outcrop sections of the deposits showing the Agbabu field. A comprehensive review of the geographical extent, geology, lithology and stratigraphy of the Dahomey basin and of the Agbabu field is presented in Mosobalaje et al. (2019a). In the Agbabu area, sand/shale sequences deposited in the Afowo formation and in the lower parts of Araromi formation are bitumen-saturated. The bitumen-saturated sand deposits (tar sands) have been observed to occur in both Horizon X and Horizon Y. These two horizons are separated by an organic-rich shale layer (oil shale). Adegoke et al. (1980) drilled forty (40) wells on the $17km^2$ study area from which some 583 tar sand and oil shale core samples were obtained at various depths. The investigation proceeded to determine the weight percent bitumen and water saturations of each core sample. Figure 9 shows the locations of the wells within the Agbabu field. Mosobalaje et al. (2019a) deployed basic principle of volumetric proportions to compute and generate reservoir porosity database from the existing Adegoke et al. (1980) raw database. The descriptive analyses by Mosobalaje et al. (2019a) were conducted only on bituminous sand Horizons X and Y 443 data points. Consequent on the exclusion of certain spurious data points, only 408 data points from 33 wells were included in the analyses. Furthermore, exploratory spatial data analysis conducted by Mosobalaje et al. (2019b) detected some spatial outlier pairs in the 408-points porosity database. These spatial outlier pairs were excluded from the estimation and modeling of porosity variogram. The resulting database containing 362 core porosity data as well as the sampling points' coordinates is the subject of the application of SGSA reported in this paper.

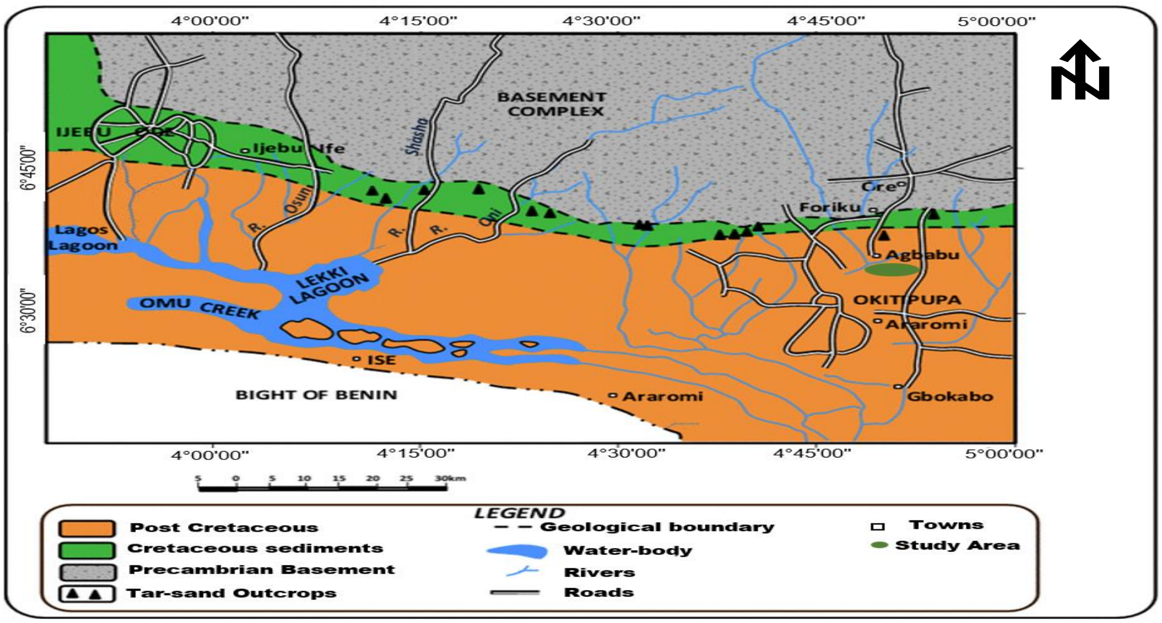

Figure 8: Geologic map of the outcrop sections of the Nigerian bitumen deposits.
Adapted from Enu (1985)

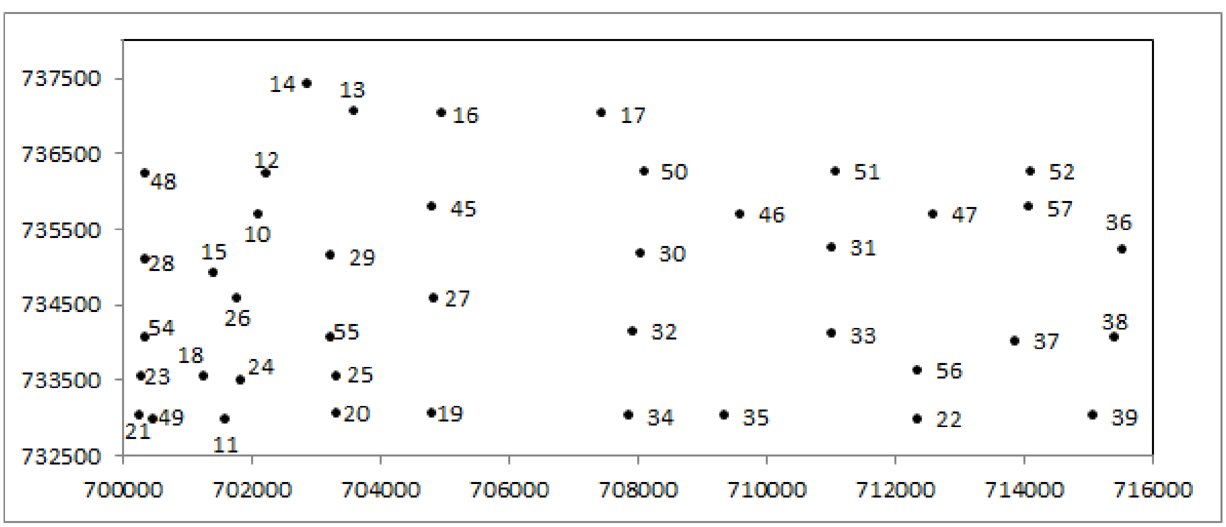

Figure 9: Location map of the study area showing well locations Adapted from Falebita *et al*. (2014)

7.2 Re-sampling and Sample Mean Computation

Non-ergodic multiple samples of the Agbabu field porosity attribute was generated by calling function `sgsa`. The original 362-point sample data (coordinates and porosity values) was prepared as a .csv file and imported into the R workspace via the following command:

`sampledata = read.csv(file.choose(),header=T)`.

The gridblock dimension intended for this call was 100ft, 100ft. and 1ft. in the x-, y-, and z-direction, respectively. This dimension would give rise to a 160 × 52 × 100 3-D grid on the study area. The plausible variogram model for the original sample data has been reported in Mosobalaje et al. (2019b) as follows:

$$\gamma = 0.0024 + 0.0016Sph_{\substack{a_{h_{max}}=3500 \\ a_{h_{min}}=1500 \\ a_{vert}=15}}(\vec{h}) + 0.0005Sph_{\substack{a_{h_{max}}=3500 \\ a_{h_{min}}=1500 \\ a_{vert}=\infty}}(\vec{h}) - - - - - - 12$$

This variogram model was set up in R workspace using the `vgm` function in gstat package thus:

```
library(gstat)
library(sp)
zonalanisomodel = vgm(psill = 0.0005, "Sph", range = 1000000000, anis = c(0,90,0,3.5
e-6,1.5e-6))
Integrated3Dmodel = vgm(nugget = 0.0024, psill = 0.0016, range = 3500, model = "Sph"
, anis=c(90,0,0,0.4286,4.286e-3), add.to = zonalanisomodel)
```

On the basis of the foregoing, the call to function `sgsa` was made thus:

```
Allshifts = sgsa(sampledata = sampledata, x_origin = 700000, y_origin = 732500, z_or
igin = 0, nx = 160, ny = 52, nz = 100,  deltaX = 100, deltaY = 100, deltaZ = 1, varg
model = Integrated3Dmodel, nmin = 20, nmax = 50)
```

The indices of the outermost sample blocks in this discretized field warranted a total of 1,001 possible shiftings of the sampling grid. Accordingly, the call to `sgsa` generated 1,001 non-ergodic repeated samples of the original porosity sample data. Each repeated sample contains 362 spatially correlated data points. According to the functionality built into function `sgsa`, each repeated sample is generated as .csv file domiciled in the working directory.

Once the 1,001 .csv files were generated, the following lines of codes were implemented to access each file and compute the mean of the repeated sample contained therein. All the computed mean values were stored in a data frame pre-created for that purpose. Similar lines can be written for other statistics of interest such as empirical variogram.

```
TotalPossibleshifts = 1001

# Pre-creating data frame for sample means
```

```r
AllShiftingsmeanporoHolder= data.frame(matrix(0,TotalPossibleshifts,2))
names(AllShiftingsmeanporoHolder) = c("Shifting ID", "MeanPoro")

for (i in 1:TotalPossibleshifts){
# Accessing each file

ShiftSamplePointsandData = read.csv(paste0("Sample_",i,".csv"),header=T)


# Sample mean computation
  Shift_meanporo = mean(ShiftSamplePointsandData$attribute)

  #Sending the computed mean to pre-created data frame.
  AllShiftingsmeanporoHolder[i,1]=i
  AllShiftingsmeanporoHolder[i,2]=Shift_meanporo
}
```

7.3 Uncertainty Assessment

The sampling distribution of the computed sample means is here presented as Figure 10. First, the histogram of the sample mean values is observed to exhibit reasonable degrees of symmetry. Hence, the sampling distribution of the sample mean can be approximated as Normal (Gaussian) distribution. This observation fits well into theoretical expectations. The porosity attribute of the Agbabu field has been observed to follow Normal distribution (Mosobalaje et al., 2019a). Expectedly therefore, the mean porosity follows the Normal distribution. Furthermore, considering the size of each sample (362); the central limit theorem warrants that the sample means be normally distributed (Walpole et al., 2012). The uncertainty in the sample mean is here quantified to be 0.005 being the standard deviation of the sampling distribution of the mean porosity. This standard deviation value as well as the mean of the mean porosity (0.24) could be used to define the distribution of mean porosity for the Agbabu field. Such distribution captures the uncertainty in the mean porosity. Values drawn from such distribution could therefore be deployed in stochastic simulations of the attributes, instead of using a single fixed value.

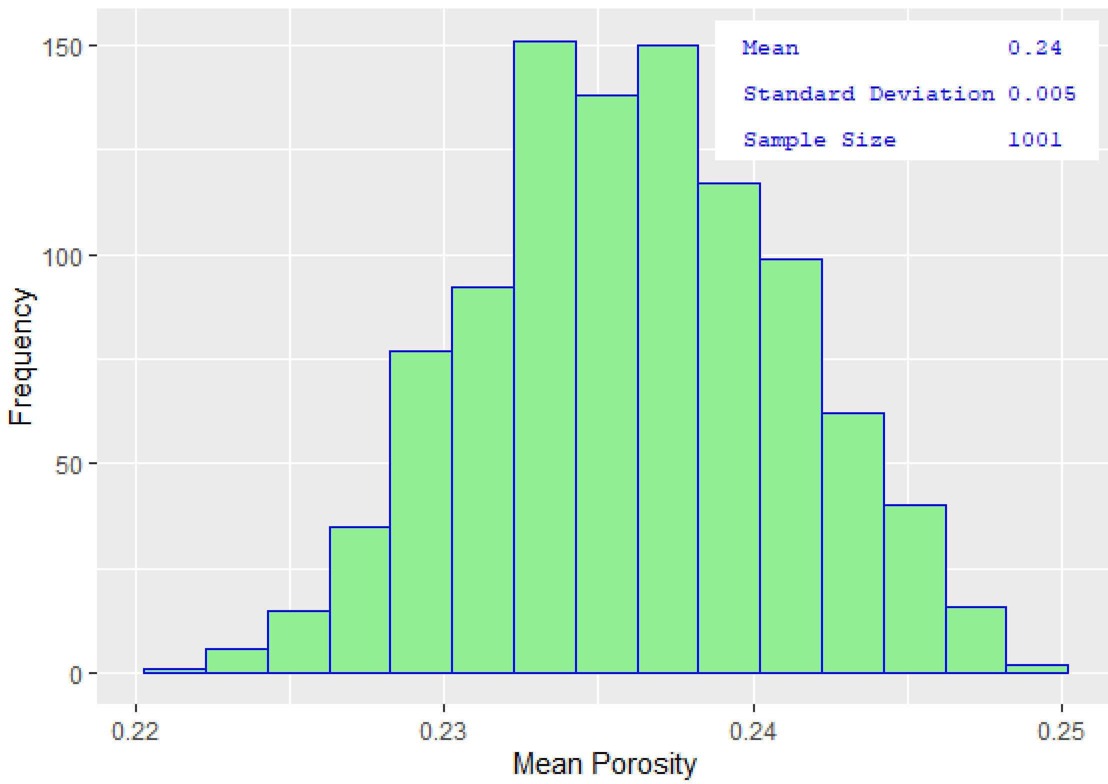

Figure 10: Sampling distribution of the computed sample means

8. Conclusions

A new spatial bootstrap algorithm that overcomes the limitations and complications of existing techniques has been developed and scripted as a function in R statistical computing environment. The algorithm is based on the translation (shifting) of sampling grid over the entire domain of the field to be sampled. It is applicable to 1-D, 2-D and 3-D isotropic/anisotropic fields being sampled with regular or irregular sampling grid. The functionality of the algorithm ensures each shifted sampling grid retains the configuration of the original grid. Consequently, the sample obtained from each shifted grid is deemed a repeat of the sample from the original grid. Given that a field is adequately discretized, the algorithm is capable of generating a number of repeated samples sufficient to yield realistic statistical inference on the sampling distribution. The function script has been written to simplify the art of specifying its input arguments. When applied to an irregularly-sampled 3-D field, the algorithm yielded results considered satisfactory and useful.

Acknowledgements

First, inspirations from *the only wise God* are acknowledged. Dr. A. A. Adepelumi facilitated access to previous report of the Geological Consultancy Unit of Obafemi Awolowo University. Publication support from Covenant University Centre for Research, Innovation and Discovery (CUCRID) is also appreciated.

Nomenclature

$a_{h_{max}}$	Major horizontal correlation range
$a_{h_{min}}$	Minor horizontal correlation range
a_{vert}	Vertical correlation range
c	Shifting counter
DS	Number of possible downward shifting of a sampling grid
$DSBI_z$	z-index of the upward outermost sample block in a discretized random field
ES	Number of possible eastward shifting of a sampling grid
$ESBI_x$	x-index of the eastward outermost sample block in a discretized random field
i, j, k	Gridblock indices; engineering ordering
N_{order}	Natural ordering index of a gridblock in a discretized random field
NS	Number of possible northward shifting of a sampling grid
$NSBI_y$	y-index of the northward outermost sample block in a discretized random field
n_x, n_y, n_z	Number of gridblock in x-, y- and z-direction in a discretized random field
SGSA	Sampling Grid Shifting Algorithm
S_{Planar}	Number of possible planar shifting of a sampling grid
SS	Number of possible southward shifting of a sampling grid
$SSBI_y$	y-index of the southward outermost sample block in a discretized random field
S_{Total}	Total number of possible shiftings
S_{lat}	Number of possible latitudinal shifting of a sampling grid
S_{longit}	Number of possible longitudinal shifting of a sampling grid
US	Number of possible upward shifting of a sampling grid
$USBI_z$	z-index of the downward outermost sample block in a discretized random field
WS	Number of possible westward shifting of a sampling grid
$WSBI_x$	x-index of the westward outermost sample block in a discretized random field
$\Delta x, \Delta y, \Delta z$	Gridblock dimensions
$\gamma(\vec{h})$	Variogram, squared unit of attribute

References

Adegoke, O. S., Ako, B. D., Enu, E. I., Afonja, A. A., Ajayi, T. R., Emofurieta, W. O., … Soremekun, O. A. (1980). Geotechnical investigations of the Ondo State bituminous sands – Volume I: Geology and reserves estimate. *Unpublished*. Geological Consultancy Unit, Department of Geology, University of Ife (now Obafemi Awolowo University), Nigeria, 257.

Babak, O. and Deutsch, C. V. (2009). Accounting for parameter uncertainty in reservoir uncertainty assessment: The conditional finite-domain approach. *Natural Resources Research,* **18 (1):** 7-17.

Brus, D. J. and de Gruijter, J. J. (1994). Estimation of non-ergodic variograms and their sampling variance by design-based sampling strategies. *Mathematical Geology*, **26(4)**: 437 – 454.

Derakhshan SH, Deutsch CV, Leuangthong O. Methodology for Choosing the Optimal Tolerance Parameters for Variogram Calculation. Centre for Computational Geostatistics Annual Report 9 2007a.

Derakhshan, H. and Leuangthong, O. (2006). Impact of data spacing on variogram uncertainty. CCG Annual Report 8-117, Centre for Computational Geostatistics, University of Alberta, 1 – 19

Deutsch, C. V. (2004). A statistical resampling program for correlated data: Spatial Bootstrap. CCG Annual Report 6-401, Centre for Computational Geostatistics, University of Alberta.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York City, NY: Chapman and Hall

Falebita, D. E., Oyebanjo, O. M., and Ajayi, T. R. (2014) A Geostatistical Review of the Bitumen Reserves of the Upper Cretaceous Afowo Formation, Agbabu Area, Ondo State, Eastern Dahomey Basin, Nigeria. *P etroleum and Coal* 56(5), pp 572 – 581.

Feyen, L. and Caers, J. (2006). Quantifying geological uncertainty for flow and transport modeling in multi-modal heterogeneous formations. *Advances in Water Resources*, **29(6)**: 912–929.

Journel, A. G. and Bitanov, A. (2004). Uncertainty in N/G ratio in early reservoir development. *Journal of Petroleum Science and Engineering*, **44(1-2)**: 115–130.

Mosobalaje, O. O., Ajia, M. V., Orodu, O. D. and Ogbe, D. (2019c). Ergodic and Non-ergodic Variogram Uncertainty in Agbabu Bitumen Field Porosity Data. Paper presented at the 2019 International Conference on Sustainable Infrastructural Development.

Mosobalaje, O. O., Orodu, O. D. and Ogbe, D. (2019a). "Descriptive Statistics and Probability Distributions of Volumetric Parameters of a Nigerian Heavy Oil and Bitumen Deposit", *Journal of Petroleum Exploration and Production Technology*, 9, 645 - 661. Available Online: https://doi.org/10.1007/s13202-018-0498-4

Mosobalaje, O. O., Orodu, O. D. and Ogbe, D. (2019b). Estimating and Modeling of Spatial Variability of Volumetric Attributes of a Nigerian Heavy Oil and Bitumen Deposit. Paper presented at the 2019 SPE Nigerian Annual International Conference and Exhibition (NAICE).

Pardo-Iguzquiza, E. and Dowd, P. A. (2001). Variance-covariance matrix of the experimental variogram: Assessing variogram uncertainty. *Mathematical Geology*, **33(4)**: 397–419.

Pebesma E. J. (2004). Multivariable geostatistics in S: the gstat Package. *Computers and Geosciences*, **30**: 683–691. http://dx.doi.org/10.1016/j.cageo.2004.03.012.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Rezvandehy, M. (2016). Geostatistical reservoir modeling with parameter uncertainty in presence of limited well data. PhD Thesis, University of Alberta, Department of Civil and Environmental Engineering.

Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K. (2012). Probability and Statistics for Engineeers and Scientists. Boston: Pearson Educational Incorporated.