J. Ind. Soc. Agril. Statist.
62(3), 2008 : 221-230

Spatial Ranked Set Sampling from Spatially Correlated Population

Ajay Kankure and Anil Rai
Indian Agricultural Statistics Research Institute, New Delhi
(Received : October 2007)

SUMMARY

Ranked Set Sampling (RSS) as suggested by McIntyre (1952) when applied to spatially-correlated areal population fails to take into account the spatial correlation. Arbia (1990) suggested Dependent Unit Sequential Technique (DUST), a sample selection procedure for selection of areal units from spatially correlated population in which spatial correlation among the population units has been incorporated into sample selection procedure. In this article we propose a sample selection technique named as Spatial Ranked Set Sampling (SRSS) in which desirable features of both RSS and DUST have been incorporated. It has been found through a spatial simulation study that SRSS performs better in terms of efficiency with respect to SRS and there is considerable gain in efficiency with respect to RSS in case of smaller set size which is generally recommended to avoid ranking errors.

Key words: Spatial ranked set sampling, Ranked Set Sampling, Dependent unit sequential technique, Spatial simulation.

1. INTRODUCTION

In agricultural surveys, often parameters of interest such as soil properties, crop yield, insect and pest infestation, etc. are geographical in nature i.e. changes in these parameters are gradual and directional. Association of locational aspect imparts spatial dependency among the neighboring units. Therefore, techniques of traditional surveys have limitations in dealing with such spatially correlated data. Due to presence of spatial correlation, neighbouring units tend to be homogeneous. Therefore, once a particular unit is selected in a sample, selection of neighbouring units is not likely to provide additional information about the target population. It is expected that the incorporation of additional information about spatial correlation into sample selection procedure will provide more efficient sampling design. Arbia (1990) suggested Dependent Unit Sequential Technique (DUST), a sample selection procedure for selection of areal units from spatially correlated population where the locational distance between the units was used to assign the probability of selection to each unit in the population. In this technique, the first unit is selected randomly and the subsequent units are selected sequentially by assigning weights in such a manner that units nearer to earlier selected units in the sample get less probability of selection as compared to the units which are far away from earlier selected unit/units. This technique, thereby, takes into account the effect of spatial correlation and consequent homogeneity introduced in the population due to geographical nearness of the units at the selection stage itself. Arbia (1993) showed empirically that this technique leads approximately 30% gain in efficiency with respect to Simple Random Sampling (SRS). The complexity of calculation of selection probabilities at each draw through this technique can be handled through Geographical Information System (GIS) software such as ARC-GIS. Sahoo Misra *et al*. (2005) modified DUST technique through incorporation of information related to shape and size of the areal sampling units in this selection procedure.

Ranked Set Sampling (RSS) as suggested by McIntyre (1952) primarily involves the formation of ranked sets of equal size. Initially, r^2 sample units are selected randomly from the population. These r^2 selected units are allocated randomly into r sets, each of size r. Now, for quantification of any values for the variable of interest, units are ranked within each set based on a perception of relative values for this variable. This may be based on judgment or on the basis of a covariate that is highly correlated with the variable of interest. Then smallest ranked unit from first set, second smallest unit from second set, third smallest unit from third set, until largest ranked unit from the last set has been selected for actual observation on the character under study. This process is repeated for m cycles until the desired sample size, n = r * m, is obtained. In order to select a ranked set sample, it is desirable to select a set of small size to minimize the ranking errors.

The mathematical foundation for RSS was provided by Takahasi and Wakimoto (1968) and, independently, by Dell (1969). McIntyre (1952) recognized in his introduction of RSS that the effectiveness of the method was dependent upon the information gained by ranking. In practice, ranking is bound to be performed with some error. Dell and Clutter (1972) showed that the RSS estimator remains unbiased even in the presence of ranking error, and when ranking is completely random, the RSS estimator has the same precision as the SRS estimator. In case concomitant variable is used for ranking, the correlation between the concomitant variable and the variable of interest is inversely proportional to ranking error. RSS was extended to ranking on a concomitant variable by Stokes (1977). In case, these restrictive distributional and relational assumptions are satisfied and make the problems tractable, systematic, stratified estimation methods (Patil *et al*. 1993a) and regression estimation methods (Patil *et al*. 1993b) are usually shown to be more efficient than RSS. Sinha (2005) has given a detailed review of recent developments in RSS. Krishna (2002) has extended ranked set sampling under finite sampling framework while Sud and Mishra (2005) while applying RSS in double sampling have shown that the RSS is more efficient than the usual estimator.

It is expected that in case of spatially correlated finite population of areal units, a sample selection procedure based on formation of ranked sets through incorporation of spatial correlation will lead to efficient sampling design. Therefore, in this article a two-phase sampling technique is proposed in which ranked sets were formed through incorporation of spatial correlation. In this procedure m key units were selected through DUST technique to form Random Spatial Clusters (RSC) in the first phase. In the second phase, ranked sets were selected from each of these RSCs. Also, an unbiased estimation procedure for this sampling scheme for estimating population mean and its variance based on Raj (1956) procedure has been proposed. A theoretical comparison of this sampling procedure with other traditional sampling techniques such as SRS, RSS is not possible as the proposed sampling technique is a sequential method of selection at first phase. Therefore, the proposed technique has been empirically compared through a spatial simulation study based on real data. The results of this spatial simulation study indicate that Spatial Ranked Set Sampling (SRSS) is generally better than existing comparable techniques such as SRS and RSS under similar situation.

2. SPATIAL RANKED SET SAMPLING (SRSS)

Let the parameter of interest be the population mean for the character under study i.e. Y. It is also assumed that the values for an auxiliary character, i.e. X, which is highly correlated with the study character, are known for the entire population. Let there be N areal units in the spatially correlated population. Let an ultimate sample of size n (= r × m) needs to be selected from this population. A sample of size n is to be selected by following method:

1. Initially m units are selected using DUST.
2. Based on the location of these m initial units all the population units are divided into m RSCs in a mutually exclusive way. Every unit of the population is assigned to a RSC formed by the key unit nearest to it, on the basis of nearest neighbourhood approach using Cartesian distances between the centroids of the areal units.
3. In each RSC, r^2 units are selected by SRS and r sets each of size r units are formed randomly.
4. The r units of each set are ranked in ascending order based on the value of the auxiliary character or by visual observation.
5. Having ranked the units in each set, the smallest unit is selected from the first set, 2^{nd} smallest unit is selected from the 2^{nd} set and in this manner the largest unit is selected from the last (r^{th}) set.
6. Thus r units are selected from each RSC independently using RSS.

3. PROBABILITY OF SELECTION

The proposed Spatial Ranked Set Sampling (SRSS) procedure involves selection of sampling units in a sequential manner. Thus, the probability of selection of each unit depends on the units previously selected in the sample. The overall probability of selection of a particular unit in a sample depends on selection technique at both the phases. Since m units were selected initially by DUST, these m units were key units for formation of the RSC. In this procedure the unit at the first draw was selected randomly. Hence, probability of selection of the ith unit of population at the 1st draw by DUST may be written as ${}^{D}p_i^{(1)} = \frac{1}{N}$ where ${}^{D}p_i^{(1)}$ is the probability of selecting i^{th} unit at the first draw by DUST, N is the number of population units and i = 1, …, N. Having selected the first unit randomly, remaining units were selected sequentially in a draw-by-draw procedure by giving differential weights to the units based on the spatial autocorrelation and distance of the unit from units selected earlier. The probability of selection of the i^{th} unit at the 2nd draw by DUST can thus be written as

$${}^{D}p_i^{(2)} = \frac{(1-\beta^{d_{iu_1}})}{\sum\limits_{\substack{i=1\\ i\neq u_1}}^{N}(1-\beta^{d_{iu_1}})}$$

where ${}^{D}p_i^{(2)}$ is the probability of selecting i^{th} unit at the second draw by DUST, u_1 is the unit selected at the 1^{st} draw, is the distance between i^{th} unit and u_1 and β is the spatial autocorrelation. Similarly, probability of selection of a unit at the m^{th} draw by DUST is

$${}^{D}p_i^{(m)} = \frac{(1-\beta^{d_{iu_1}})(1-\beta^{d_{iu_2}})...(1-\beta^{d_{iu_{m-1}}})}{\sum\limits_{\substack{i=1\\ i\neq u_1,u_2,...,u_{m-1}}}^{N}(1-\beta^{d_{iu_1}})(1-\beta^{d_{iu_2}})...(1-\beta^{d_{iu_{m-1}}})}$$

where ${}^{D}p_i^{(m)}$ is the probability of selecting i^{th} unit at the m^{th} draw by DUST, u_{m-1} is the unit selected at the $(m-1)^{th}$ draw and $d_{iu_{m-1}}$ is the distance between i^{th} unit and u_{m-1}. Thus we get,

$$\sum_{\substack{i=1\\ i\neq u_1,u_2,...,u_{m-1}}}^{N} {}^{D}p_i^{(m)} = \sum_{\substack{i=1\\ i\neq u_1,u_2,...,u_{m-1}}}^{N} \frac{(1-\beta^{d_{iu_1}})(1-\beta^{d_{iu_2}})...(1-\beta^{d_{iu_{m-1}}})}{\sum\limits_{\substack{i=1\\ i\neq u_1,\\ u_2,...,\\ u_{m-1}}}^{N}(1-\beta^{d_{iu_1}})(1-\beta^{d_{iu_2}})...(1-\beta^{d_{iu_{m-1}}})} = 1$$

A general form of the probability of selecting i^{th} unit at the t^{th} draw by DUST is thus given by

$${}^{D}p_i^{(t)} = \frac{\prod\limits_{j=1}^{t-1}(1-\beta^{d_{iu_j}})}{\sum\limits_{l\in\bar{s}}\prod\limits_{j=1}^{t-1}(1-\beta^{d_{lu_j}})}$$

where ${}^{D}p_i^{(t)}$ is the probability of selection of i^{th} unit at the t^{th} draw by DUST, u_j is the unit selected at the j^{th} draw, d_{iuj} is the distance between i^{th} unit and u_j, is the set of units not selected earlier, t = 1, … , m, j = 1, … , t – 1 and l = 1, …, N.

In order to form RSC, units selected by DUST form the key units. The distance of every remaining unit not selected at the first phase of the population is measured from each of these key units and RSCs are formed by assigning every population unit to the RSC of nearest key unit. Therefore, every unit of the population is assigned to one and only one RSC. Thus, the probability that a unit is assigned to a particular RSC depends on its distance from the key unit of that RSC. This probability increases as the distance between the unit and specified key unit selected through DUST decreases. Thus, the probability that h^{th} unit of the population is assigned to i^{th} RSC at the given t^{th} draw can be written as

$${}^{d(D)}_{\;\;i}p_h^{(t)} = \frac{\frac{1}{d_{ih}}\prod\limits_{\substack{i'\in s\\ i'\neq i}}\left(1-\frac{1}{d_{i'h}}\right)}{\sum\limits_{l\in\bar{s}}\frac{1}{d_{il}}\prod\limits_{\substack{i'\in s\\ i'\neq i}}\left(1-\frac{1}{d_{i'l}}\right)}$$

, where ${}^{d(D)}_{\;\;i}p_h^{(t)}$ is the distance based probability that h^{th} unit of the population is assigned to i^{th} RSC at the t^{th} draw, d_{ih} is the distance between h^{th} unit of the population and the i^{th} key unit selected earlier by DUST, s is the set of units previously selected in (t – 1) draws and i′ = 1, … , t – 1. Having formed the RSCs, a ranked set sample of size r is selected independently from each RSC. Let the i^{th} RSC have N_i number of sampling units. Further, let r number of sets each of size r units be selected randomly in each RSC. The r units in each set are ranked in an increasing order. It can be seen that the sample consists of r places to be filled by units selected from the population of N_i units. In case of ordered set k^{th} place of the set corresponds to the k^{th} rank. There are (k – 1) places before the k^{th} place and (r – k) places after the k^{th} place in an ordered set of size r. Let be the rank of a unit in the i^{th} RSC out of N_i units which has been selected in the j^{th} set and it has rank k in that set. The (k – 1) places before the k^{th} place in the j^{th} ordered set can be filled by (${}^{j}g_i^k$ – 1) places in the ordered population and the (n-k) places after rank k can be filled by (N_i – ${}^{j}g_i^k$) units of the ordered population. Thus, the total number of ways in which (k – 1) places of the j^{th} ordered set can be filled by (${}^{j}g_i^k$, – 1) units of the ordered population is given by $\binom{{}^{j}g_i^k - 1}{k_j - 1}$.

Similarly, the number of ways in which remaining (n-k) places after rank k can be filled by the (N_i – ${}^{j}g_i^k$) units of the ordered population is given by $\binom{N_i - {}^{j}g_i^k}{r - k_j}$. The total number of SRS sets of size r possible from the population of N_i units is $\binom{N_i}{r}$. Therefore, the probability that a unit with rank ${}^{j}g_i^k$ in the population of N_i units has k^{th} rank in the j^{th} SRS set of size r is given by $\dfrac{\binom{{}^{j}g_i^k - 1}{k_j - 1}\binom{N_i - {}^{j}g_i^k}{r - k_j}}{\binom{N_i}{r}}$. In order to get a ranked set sample of size r units, smallest ranked unit is selected from the first set; second ranked unit is selected from the second set; and in this manner the highest ranked unit is selected from the last set for measurement of the unit for the character under study. Hence, k^{th} unit of the i^{th} RSC is selected in the ranked set sample if it has either 1^{st} rank in the 1^{st} set or 2^{nd} rank in the 2^{nd} set, and in this manner, if it has r^{th} rank in the r^{th} set. The probability of selecting a unit by RSS in a RSC is obtained by summing the probabilities of all the ranks from 1 to r in the respective sets. This probability can be written as

$$^{R(D)}p_{jk} = \sum_{k=j=1}^{r} \frac{1}{r} \frac{\binom{{}^{j}g_i^k - 1}{k_j - 1}\binom{N_i - {}^{j}g_i^k}{r - k_j}}{\binom{N_i}{r}}$$

where N_i is the number of units in the i^{th} RSC, ${}^{i}g_i^k$ is the rank in the i^{th} RSC of the k^{th} ranked unit in the j^{th} set, r is the number of sets as well as the set size, k_j is the k^{th} rank of unit in the j^{th} set, k = 1,…, r, $\sum_{i=1}^{m} N_i = N$. Thus, selection of a unit in the ultimate sample depends on the RSC formation due to the random selection of the key units of each spatial cluster, size of the RSC and rank of the sampling units within a RSC. Thus, the probability of selection of h^{th} unit which has rank k in the j^{th} random set selected in the i^{th} RSC at the t^{th} draw is given by [Probability of selecting i^{th} key unit] × [Probability that h^{th} unit belongs to i^{th} RSC / i^{th} key unit has been selected] × [Probability of selecting h^{th} unit which has rank in the i^{th} RSC and has rank k_j in the j^{th} random set/ i^{th} key unit has been selected and h^{th} unit belongs to i^{th} RSC formed by the i^{th} key unit]. Thus, $p_{hijk}^{(t)}$ is the probability of selection h^{th} unit which has rank k in the j^{th} random set selected in the i^{th} RSC at the t^{th} draw is given by

$$p_{hijk}^{(t)} = \frac{\prod_{j=1}^{t-1}(1-\beta^{d_{iu_j}})}{\sum_{l\in\bar{s}}\prod_{j=1}^{t-1}(1-\beta^{d_{lu_j}})} \frac{\frac{1}{d_{ih}}\prod_{\substack{i'\in s\\ i'\neq i}}\left(1-\frac{1}{d_{i'h}}\right)}{\sum_{l\in\bar{s}}\frac{1}{d_{il}}\prod_{\substack{i'\in s\\ i'\neq i}}\left(1-\frac{1}{d_{i'l}}\right)} \sum_{k=j=1}^{r}\frac{1}{r}\frac{\binom{{}^{j}g_i^k - 1}{k_j - 1}\binom{N_i - {}^{j}g_i^k}{r - k_j}}{\binom{N_i}{r}} \tag{1}$$

4. THE PROPOSED ESTIMATOR

The probabilities of selection of units by SRSS vary at each draw and it depends on the selection of earlier units selected in the sample as discussed in the above section. An unbiased estimator for estimating population mean under these circumstances was given by Raj (1956). Let a sample selected by SRSS be { $y_{hijk}^{(1)}$, $y_{hijk}^{(2)}$, $y_{hijk}^{(3)}$, ..., $y_{hijk}^{(n)}$ }, where $y_{hijk}^{(t)}$ is the value of the main character pertaining to the h^{th} unit selected from j^{th} set having rank in i^{th} RSC at t^{th} draw and { $p_{hijk}^{(1)}$, $p_{hijk}^{(2)}$, $p_{hijk}^{(3)}$, ... $p_{hijk}^{(n)}$ } are corresponding probabilities of selection obtained from equation (1). Now define, $d_t = \frac{1}{N}\left(y_{hijk}^{(1)} + y_{hijk}^{(2)} + \ldots + y_{hijk}^{(t-1)} + \frac{y_{hijk}^{(t)}}{p_{hijk}^{(t)}}\right)$. It can be seen that d_t is an unbiased estimator of the population mean $\bar{Y}$ for all t, t =1,...,n. Therefore, the proposed estimator of population mean $\bar{Y}$ is given by

$$\bar{y}_{SRSS} = \frac{1}{n}\sum_{t=1}^{n} d_t \qquad (2)$$

Although, the expression for $V(\bar{y}_{SRSS})$ is too complex to write out explicitly, an unbiased estimator of the variance can be written in a much simpler form. The unbiased estimate of variance of the estimator $\bar{y}_{SRSS}$ is given by

$$\hat{V}(\bar{y}_{SRSS}) = \frac{1}{n(n-1)}\sum_{t=1}^{n}\left[d_t - \frac{1}{n}\sum_{t=1}^{n} d_t\right]^2 \qquad (3)$$

This variance estimator is known to be always positive.

5. SIMULATION STUDY

The proposed sampling procedure takes into account the spatial correlation while forming ranked sets of RSS procedure. Since, theoretically, it is not possible to compare this sampling technique to respective comparable sampling designs, an illustrative simulation study was carried out, in order to facilitate the comparison of the proposed SRSS methodology with the existing SRS and RSS methods. In this simulation study, yield data for the rabi season for the year 1997-98 from General Crop Estimation Surveys (GCES) based on Crop Cutting Experiment (CCE) of wheat crop for Rohtak district of Haryana state has been used to generate desired wheat yield data points over two-dimensional space. Satellite data of February 4, 1998 from Indian Remote Sensing Satellite (IRS-1D) captured through Linear Imagine Self Scanner (LISS-III) sensor has been utilized for this study to simulate spatial population. The False Color Composite (FCC) image of Rohtak district has been shown in Fig.1. The details of sensor characteristics used for the study area are provided in Table 1. Wheat yield per plot, y, obtained from CCE in 1997-98 from Rohtak district of Haryana state and variable corresponding to Normalized Difference Vegetation Index (NDVI) generated from digital data of IRS-1D satellite, LISS-III sensor, for each village were obtained using the classified image of Rohtak district shown in Fig. 2. The NDVI feature was derived with the help of the georeferenced satellite image using the transformation: $NDVI_i = \frac{IR_i - R_i}{IR_i + R_i}$ where, IR_i is Digital Number (DN)-value of i^{th} pixel corresponding to infrared band of the image whereas R_i denotes the DN-value of the respective red band.

Table 1. LISS-III sensor characteristics

Sensor	LISS -III
Resolution	23.7 m
Swath	127 km (bands 2, 3, 4) 134 km (band 5 –MIR)
Temporal Resolution	24 days
Spectral Bands	0.52 – 059 microns (B2) 0.62 – 0.68 microns (B3) 0.77 – 0.86 microns (B4) 1.55 – 1.7 microns (B5)
Path/Row	95/51

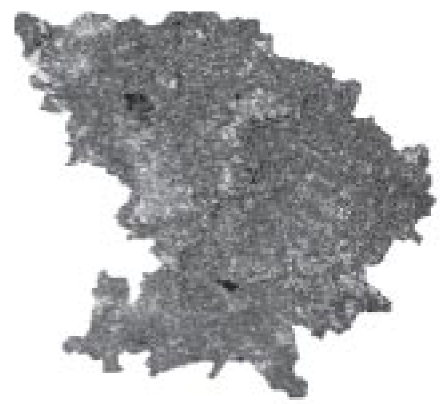


Fig. 1. FCC of Rohtak district derived from IRS-1D, LISS-III sensor as on Feb.04, 1998

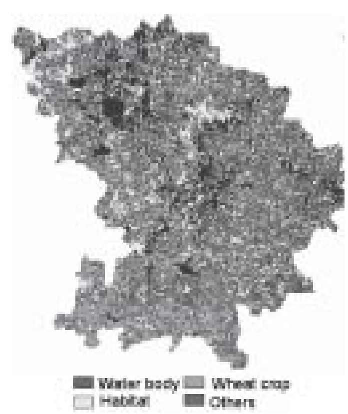


Fig. 2. Classified image of the Rohtak district

The yield data obtained from CCE and the corresponding locations of plots of CCEs in terms of latitude and longitude was identified. The spatial models i.e. variograms were applied to the data of yields and locations to determine the spatial model of best fit along with its parameters. With the help of best-fitted spatial model, ordinary kriging method was used for spatial prediction at unsampled locations. Kriging is a technique of spatial interpolation of variable at a particular location based on values of the same variables at sampled locations through incorporating spatial variation using spatial models. Ordinary kriging gives both yield prediction and standard error of prediction at sampled as well as unsampled locations. Also, production surface was obtained using ordinary kriging method in the form of grids of appropriate size Rai *et. al.* (2007). With the help of spatial model, the yield values were generated corresponding to each wheat pixel of the image. In order to get the yield of individual village, which is required for this simulation study, the village map was overlayed over the district map and the estimate of yield and corresponding standard error for 396 villages was obtained by taking average of all the grids falling in the village boundary. The yield statistics generated was attached to the centroid of each village. In order to estimate different moments of different statistics considered for this study 500 samples of sample sizes (10, 12, 20, 30) of the villages were selected by SRS, RSS and SRSS for comparison. The calculation of the probabilities of selection, which involved very complex calculations, were carried out in SAS software using various PROC commands. Once the selection probabilities were calculated, Des Raj estimator, Raj (1956), was used to obtain unbiased estimate of the population mean for the main character of interest, yield (y) and the auxiliary character, NDVI values (x). Mean, variance, Skewness and Kurtosis were calculated on the basis of estimates obtained from 500 samples for each combination of the sample size (n) for each case of SRSS estimators. To compare the desired statistical properties of the proposed sampling schemes with the existing sampling schemes, the percentage gain in efficiency as compared to the estimator based on SRS and RSS has been calculated using the formula:

$$GE_T = \left[\frac{V(\bar{y}_T) - V(\bar{y}_{SRSS})}{V(\bar{y}_{SRSS})}\right] \times 100 \quad \text{where, } V(\bar{y}_{SRSS})$$

is the variance for the SRSS estimator, $V(\bar{y}_T)$ is the variance for estimator T, where T is SRS or RSS. Percent relative bias has also been calculated for each estimator using the expression, $\% \ \text{Bias} = \left[\frac{\bar{y}_{SRSS} - \bar{Y}}{\bar{Y}}\right] \times 100$.

The sample selection procedure for SRSS was carried out in two stages. At the first stage key units were selected for the formation of RSC and then ranked set sampling was carried out in these RSC as explained above in Section 2. This procedure of sample selection may give effective sample size being smaller than the expected sample size. Hence, an average of the effective sample size obtained for the 500 samples was calculated. This has been called Average Effective Sample Size (ASS).

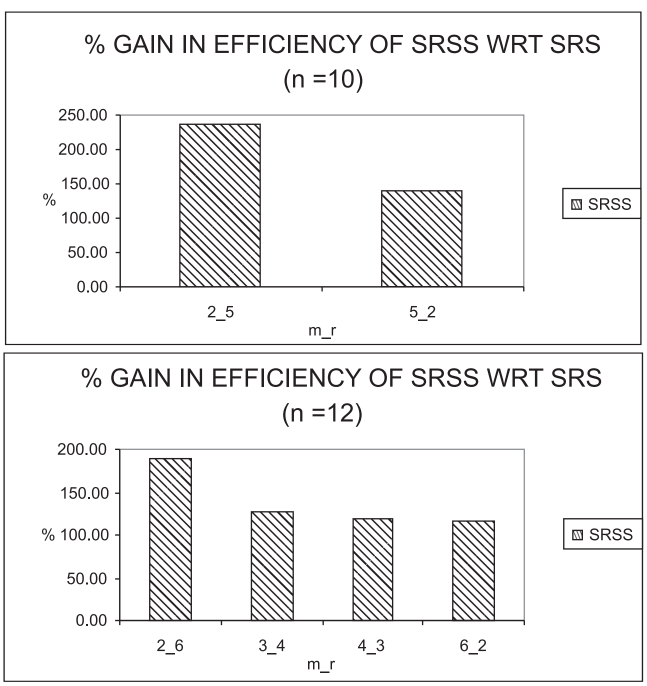
Table 2 gives the estimates of mean, variance, skewness, kurtosis, % Bias, GESRS, GERSS and ASS obtained for SRSS for the different combinations of set size and number of sets for each sample size. Different combinations of a given sample size are formed due to the fact that the ultimate sample is a product of two integers; the number of RSC formed at the first stage of sample selection and the number of units selected by RSS from each RSC. An ultimate sample of n = 10 units results in two combinations of m × r, viz., (m = 5, r = 2) and (m = 2, r = 5). Here m is the number of RSC formed

Table 2. Estimates of Mean, Variance, Skewness, Kurtosis, % Bias, Gain in Efficiency (GE) with respect to SRS and RSS and ASS obtained for SRSS

n=r×m	m_r	Mean	Variance	Skewness	Kurtosis	% Bias	GE_{SRS}	GE_{RSS}	ASS
10	2_5	19.61	0.00237	1.28	3.71	-0.23	237.61	1.80	9.99
10	5_2	19.66	0.00335	0.59	2.11	-0.01	138.48	19.53	10.00
12	2_6	19.62	0.00175	1.29	3.78	-0.23	188.31	-32.22	11.94
12	3_4	19.64	0.00223	0.62	0.37	-0.13	126.12	-5.11	11.98
12	4_3	19.64	0.00231	0.22	0.06	-0.10	118.15	25.31	11.99
12	6_2	19.66	0.00233	0.25	1.24	-0.03	116.55	36.04	12.00
20	2_10	19.61	0.00090	1.13	2.74	-0.25	372.33	-21.49	18.80
20	4_5	19.64	0.00137	0.38	0.08	-0.10	210.71	-18.13	19.70
20	5_4	19.65	0.00147	0.31	-0.07	-0.06	189.41	18.46	19.86
20	10_2	19.67	0.00225	1.72	4.32	0.07	89.46	38.22	19.99
30	3_10	19.62	0.00112	0.61	1.68	-0.22	107.21	-87.41	21.56
30	5_6	19.64	0.00105	-0.01	-0.12	-0.09	119.28	-30.30	27.97
30	6_5	19.65	0.00107	0.17	0.13	-0.06	115.29	-7.08	28.71
30	10_3	19.66	0.00091	0.15	-0.30	0.00	104.32	14.61	29.56
30	15_2	19.68	0.00150	1.53	2.63	0.12	53.88	26.85	29.75

at the first stage of sample selection and r is the number of units selected by RSS, independently from each RSC. These combinations have been written as m_r in column (2) of the table.

It can be observed from Table 2 that the percent bias is very low, from -0.25 to 0.12 which suggests that the estimator is unbiased. Considering the gain in efficiency with respect to SRS given by GE_{SRS}, it can be observed that SRSS is more efficient than SRS, as there is positive gain in efficiency for all combinations of m and r. Fig. 3 shows GE_{SRS} against the different sample sizes for their various combinations of r and m for SRSS. For a given sample size, the bars are arranged in increasing order of RSC denoted by m. *e.g.* for sample size 12, the group of four bars are arranged in increasing order of RSC denoted by m as 2, 3, 4 and 6. In case of given sample size, an increase in the number of RSCs results in a consequent decrease in the size of the ranked set sample selected from each RSC. In selection of a usual ranked set sample in a single cycle, the sample size is equal to the number of same-sized sets selected randomly from the population which is again equal to the set size of each set. In SRSS, ranked set sampling is carried out in each RSC in a single cycle. So it can be observed that an increase in the number of RSCs results in a consequent decrease in the size of the ranked set sample and consequently a decrease in the set size, denoted by r, and vice-versa.



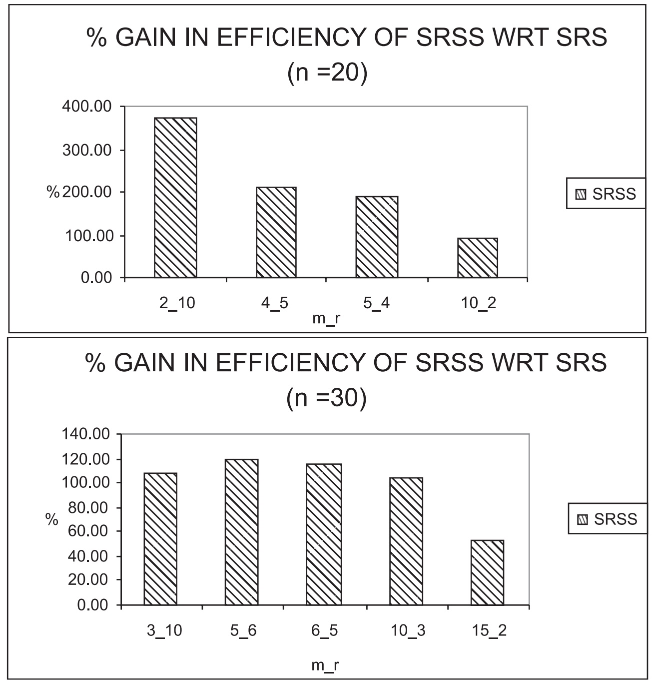
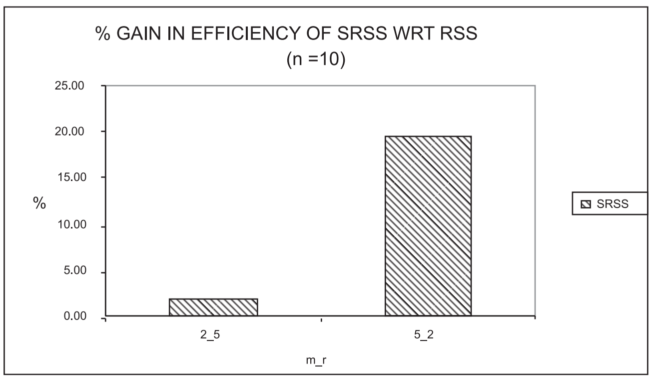
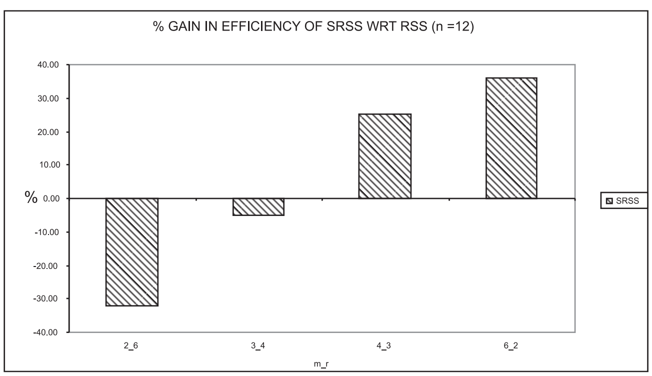
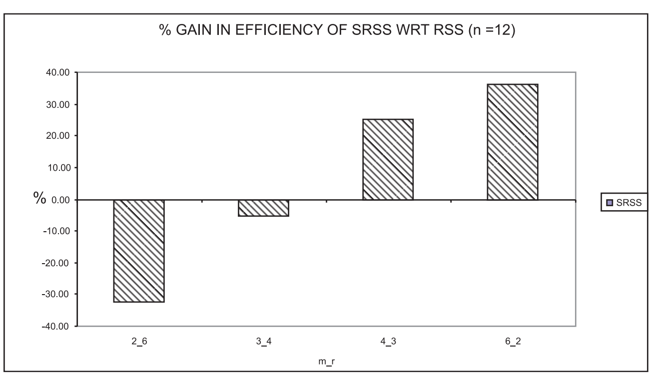


Fig. 3. % Gain in Efficiency of SRSS over SRS for n = 10, 12, 20 and 30

In general, it can be seen that, for a given sample size, GE_{SRS} decreases with an increase in the number of RSC, denoted by m (or consequently, with decrease in set size, denoted by r). GE_{SRS} is 237.61 for the combination 2_5 while it decreases to 138.48 for the combination 5_2 as the number of RSCs increase from 2 to 5 (or the set size decreases from 5 to 2) for the sample of size 10 units. A similar trend has been observed for samples of size 12, 20 and 30. In case of sample size 12, GE_{SRS} is 188.30, 126.12, 118.15 and 116.55 for combinations 2_6, 3_4, 4_3 and 6_2 respectively, thereby, indicating that as the number of RSCs increase from 2 to 6, there is consistent decrease in GE_{SRS}. Similarly, in case of sample size 20, GE_{SRS} is 372.33, 210.71, 189.41 and 89.46 for combinations 2_10, 4_5, 5_4 and 10_2 respectively, indicating again that as the number of RSCs increase from 2 to 10, there is consistent decrease in GE_{SRS}. In case of sample size 30, the trend of decrease in GE_{SRS} with increase in the number of RSCs is maintained, except for the combination 3_10 which has smaller GE_{SRS} than the combination 5_6 inspite of the prior having lesser number of RSCs than the latter. This may be attributed to ASS being less than the expected sample size. This happens due to the fact that, having formed the m random spatial clusters, each cluster should have atleast r^2 units to get a RSS of r units from each RSC. For large-sized sample, some clusters have less than r^2 units in them as a result of which we do not get any RSS from that cluster. Hence, the actual sample obtained is smaller than expected. ASS is only 21.56 for the combination 3_10 instead of the expected sample size 30, while ASS for the combination 5_6 is 27.97. Considering the % Gain in efficiency with respect to RSS given by GE_{RSS}, it can be observed from Fig. 4 that SRSS is more efficient than RSS when the set size is small and consequently for greater number of RSCs. There is positive gain in efficiency for a combination in which m is greater and r is smaller. GE_{RSS} is 1.80 for the combination 2_5 while it increases to 19.53 for the combination 5_2 as the number of RSCs increase from 2 to 5 (or the set size decreases from 5 to 2) for the sample of size 10 units thereby indicating that increase in the number of clusters (or consequently a decrease in the set size) has resulted in increasing efficiency of SRSS as compared to RSS. A similar trend has been observed for samples of size 12, 20 and 30.







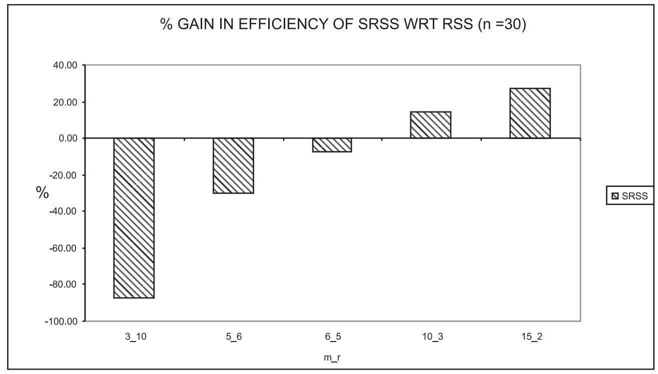
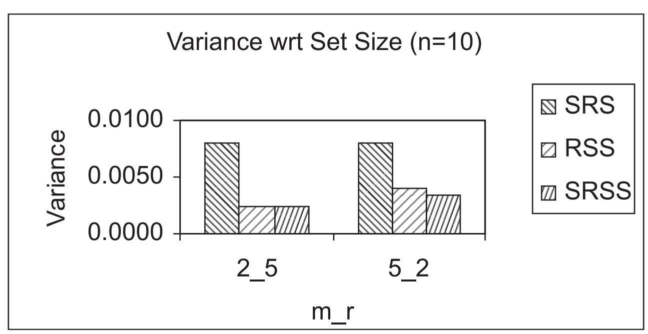


Fig. 4. % Gain in Efficiency of SRSS over RSS for n = 10, 12, 20 and 30

In case of sample size 12, GE_{RSS} is -32.22 and -5.11 for combinations 2_6 and 3_4 while it is 25.31 and 36.04 for combinations 4_3 and 6_2 respectively, thereby indicating that as the number of RSCs increase from 2 to 6, there is consistent increase in GE_{RSS}. Similarly, in case of sample size 20, GERSS is -21.49 and -18.13 for combinations 2_10 and 4_5 while it is 18.46 and 38.22 for combinations 5_4 and 10_2 respectively, indicating again that as the number of RSCs increase from 2 to 10, there is consistent increase in GE_{RSS}. In case of sample size 30, the trend of increase in GE_{RSS} with increase in the number of RSCs is maintained.

The results discussed so far show that GE_{SRS} and GE_{RSS} have inverse trend with respect to the number of RSCs. GE_{SRS} decreases with an increase in RSCs whereas GE_{RSS} increases with an increase in RSCs. This shows that as the set size is decreased (or consequently, as the number of RSCs increased) RSS shows faster increase in variance (and consequently faster decrease in efficiency) as compared to SRSS which is evident from Fig. 5. SRSS is more efficient than RSS only so long as the set size is small. As the set size is increased (or consequently, as the number of RSCs in decreased) progressively, RSS turns out to be more efficient than SRSS indicated by the negative GE_{RSS}.



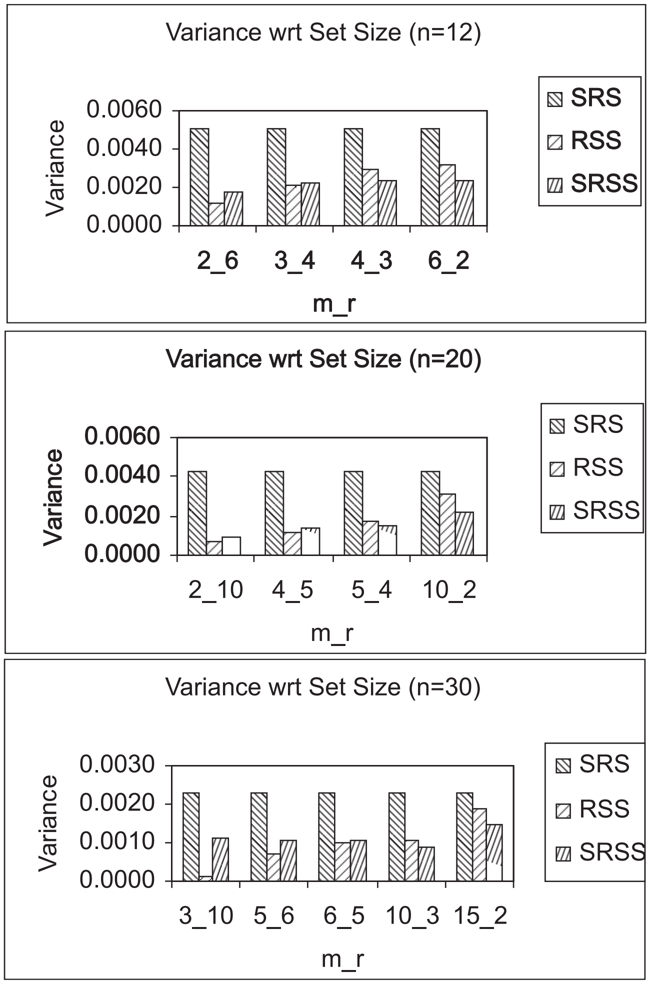


Fig. 5. Variance of SRS, RSS and SRSS for n=10, 12, 20 and 30

From the population of 396 units, it was only possible to select samples of size 10, 12 and 20 completely without the ASS being too small. It can be observed from Fig. 3 that an increase in the sample size has resulted in an increase in the gain in efficiency, only so long as the complete sample could be selected. Maximum GE_{SRS} was observed for sample of size 20.

6. CONCLUSION

Ranked Set Sampling (RSS) is always more efficient then Simple Random Sampling (SRS) as randomization procedure in RSS provides better representation of the population in the selected sample. Further, additional information about ranking of the selected units in the selection process improves the efficiency of RSS over SRS. In case of spatially correlated population it is expected that neighboring units are likely to be more homogeneous then distant units. Therefore, in the proposed Spatial Ranked Set Sampling (SRSS) in which additional information about spatial relationship among neighboring units is being incorporated in the sample selection process through spatial correlation is found to be better then RSS design. A spatial simulation study has been performed to empirically study the statistical properties of the proposed sampling strategy with respect to SRS and RSS. It was found that proposed SRSS is always better then SRS and there is considerable gain in precision. Also, there is gain in efficiency of proposed SRSS over RSS in most of practical situation in which RSS is applicable i.e. smaller set sizes. Other statistical properties based on moments, of proposed SRSS are found to be similar with respect to SRS and RSS. This procedure has number of practical applications in agricultural surveys. For example, in case of yield estimation of a crop at district level, this selection procedure can provide more efficient sampling strategies. It is well known that yield of a particular crop has spatial relationship with yield of the same crop in neighboring villages. Therefore, it is always desirable to selected villages for conducting CCE through proposed techniques using satellite digital data as auxiliary variable as demonstrated in simulation study. Presently these villages within a stratum are selected by SRS. In other words, sample size can be considerably reduced for same level precision.

REFERENCES

Arbia, G. (1990). GIS-based sampling design procedures. *Proc. of EGIS*, **1,** EGIS Foundation, Utrecht.

Arbia, S. (1993). The use of GIS in spatial statistical surveys. *Int. Stat. Rev.*, **61(2)**, 339-359.

Cochran, W. G. (1977). *Sampling Techniques*, 3rd edition. Wiley, New York.

Dell, T.R. (1969). The Theory and Some Applications of Ranked set Sampling. Ph.D. Thesis, Department of Statistics, University of Georgia, Athens. Georgia.

Dell, T.R. and Clutter, J.L. (1972). Ranked set sampling theorywith order statistics background. *Biometrics*, **28,** 545-553.

Krishna, Pravin (2002). Some Aspects of Ranked Set Sampling from Finite Population. Unpublished M. Sc. Thesis of I.A.R.I., New Delhi.

McIntyre, G.A. (1952). A method of unbiased selective sampling using ranked sets. *Austr. J. Agri. Res.* **3,** 385-390.

Patil, G.P., Sinha, A.K. and Taillie, C. (1993a). Relative precision of ranked set sampling: Comparison with the regression estimator, *Environmetrics*, **4(4),** 399-412.

Patil, G.P., Sinha, A.K. and Taillie, C. (1993). Estimation of the population variance using ranked set samples. Technical Report No. 93-0501. Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, University Park, PA, USA.

Rai, A., Gupta, N.K. and Singh, R. (2007). Small area estimation of crop production using spatial models. *Mod. Asst. Stat. Apl.*, **2(2),** 89-98.

Raj, D. (1956). Some estimators in sampling with varying probabilities without replacement. *J. Amer. Statist. Assoc.*, **51,** 269-284.

Sahoo, P.M, Singh, R. and Rai, A. (2006). Spatial sampling procedures for agricultural surveys using geographical Information system. *J. Ind. Soc. Agril. Statist.*, **60(2),** 134-143.

Sinha, A.K. (2005). On some recent developments in ranked set sampling. *Bull. Inf. Cybernetics*, **37,** 137-160.

Stokes, S.L. (1977). Ranked set sampling with concomitant variables. *Comm. Statist. –Theory Methods*, **A6**, 1207-1211.

Sud, U.C. and Mishra, D.C. (2006). Estimation of finite population mean using ranked set two stage sampling designs. *J. Ind. Soc. Agril. Statist*, **60(2)**, 108-117.

Takahashi, K. and Wakimoto, K. (1968). On biased estimates of the population mean based on the sample stratified by means of ordering. *Ann. Inst. Stat. Math.*, **20**, 1-31