<table>
<thead>
<tr>
<th>Broad Topic</th>
<th>Developing more efficient and accurate methods for using remote sensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub Topics</td>
<td>1. Developing more efficient and accurate methods for using remote sensing - Literature review</td>
</tr>
<tr>
<td></td>
<td>2. Improving methods for using existing land cover – land use data bases - Literature review</td>
</tr>
</tbody>
</table>

Under this section there are two review papers entitled ‘Developing more efficient and accurate methods for using remote sensing - Literature review’ and ‘Improving methods for using existing land cover – land use data bases - Literature review’. Both the papers are well drafted and contain comprehensive review of the literature on use of remote sensing for precise crop area and production estimation. In the first paper the review has been thoroughly discussed under six broad themes viz. New Technologies of Remote Sensing, Methods for Using Remote Sensing Data at the Design Level, Extension of the Regression or Calibration Estimators, Robustness of the Estimators Adopted for Producing Agricultural and Rural Statistics, Comparison of Regression and Calibration Estimators with Small Area Estimators and Statistical Methods for Quality Assessment of Land use/Land Cover Databases. The interesting features of this review are summarization of the new findings along with identification of the situations of their applicability, recommendations for future research and highlighting the gaps for immediate attention of the researchers. The review is very exhaustive and will serve as a useful guide to Global Strategy for Improvement of Agricultural and Rural Statistics in general and researchers in particular for pursing further research for bringing improvements in the methodologies for handling various practical situations of nonlinearity, nonstationarity, skewness, unbalancedness and missing observations. The other review is on Land Cover and Land Use. This review has outlined in very systematic manner the methodologies of remote sensing for precise assessment of land cover and land use statistics. Further the future areas of research and gaps have also been identified. There is however urgent need to develop more efficient methodologies for the countries where the farm sizes is very small with greater diversity of cultivation of different crops. The emphasis is also to be laid on the capacity building and empowering technical and scientific personnel with knowledge, physical and financial resources.
In India and at IASRI particularly, the significant research is being carried out in this area of Remote Sensing for the improvement of Agricultural Statistical System. The work is primarily carried out by the scientists and students who are involved in the research in the discipline of Sample Surveys. Some of the researches carried out at IASRI is highlighted as under:

**Applications of Remote Sensing, GIS and Spatial Modeling**

IASRI has developed an integrated methodology based on Remote Sensing, GIS and Sample Survey for Estimation of Area under paddy Crop in North-Eastern Region of the country. This part of the country is unique due to the typical problems existing in these regions. The north-eastern states particularly Meghalaya mainly consists of hilly region with thick forest cover. Besides this, the main problem is its undulating topography and non-accessibility of vast area. Further, the relative percentage area under the crops is very less. Mostly terraced farming and Jhum cultivation is practiced in these regions. Moreover, these areas particularly Meghalaya, are covered by clouds most of the time in a year. Thus it is difficult to get cloud free images of these areas. Therefore, use of remote sensing satellite data alone may not be able to provide reliable information. Further, there are no cadastral maps and village boundary maps existing for these regions. In contrary to other states, in north-eastern regions reliable information regarding total number of villages in each district/block is not available. Further, within a village total number of farmers, number of fields owned by each farmers, crops grown by the framers etc. are also not available in village records. Thus, the traditional methodology of area estimation is not applicable in these regions. Keeping all this in view, it was considered that the use of satellite data along with the ground survey data in GIS environment may be useful to obtain the reliable estimates for the area under crops. In absence of any satisfactory objective technique for crop area estimation in hilly regions, the integrated methodology using remote sensing, GIS and ground survey has been developed for crop acreage estimation (Sahoo et al. 2005, 2008). Also, a scientific methodology based on spatial stratification for acreage estimation of multiple crops including cereal crops, fruit crops and vegetables crops was further developed by Sahoo et al. (2011).
An improved spatial sampling technique known as *Contiguous Unit Based Spatial Sampling (CUBSS)* Technique was proposed by Sahoo et al. (2006). The technique incorporates size measure along with spatial contiguity of the units in the population. The spatial correlation is estimated for auxiliary character which is used along with size measure in assigning weights for selection of the sampling units. The probability of selection of any unit is governed by these weights. The principle of sample selection is the probability of selection of any unit increases as the distance from the units (area) already selected increases. The sample selection criterion is based on the weights, accounting for spatial variability and the size measure accounting for areal extent. Further, a suitable unbiased estimator which takes into account the order of the draw is suggested for this situation. The study is carried for regular lattice i.e. assuming the area to consist of regular units. In order to tackle the problem of irregularity of the sampling units, distance based neighbors are suggested. Based on these neighbors the modified formula for spatial correlation is also suggested in this study. For defining these neighbors, the concept of lagged variable and lagged series is being used. A spatial sampling technique termed as *Distance Unit Based Spatial Sampling (DUBSS)* is proposed in this study and its efficiency is compared with the existing ones and CUBBS technique by carrying out a suitable simulation study. The proposed technique performs considerably better than all the other techniques.

Crop yield estimation surveys based on crop cutting experiments are conducted throughout the country for obtaining precise estimates of average yield for all major crops. Sukhatme and Panse (1951) gave the estimation procedure of estimating average yield and crop production based on crop cutting experiments under general crop yield estimation surveys. Use of satellite data along with survey data of crop yield from General Crop Estimation Surveys (GCES) based on crop cutting experiments for obtaining improved estimators of crop yield has been undertaken at IASRI since 1990. Singh et al. (1992) and Singh and Goyal (1993) have used spectral vegetation indices like Normalized Difference Vegetation Index (NDVI) to obtain improved crop yield estimators. Singh et al. (2000) have developed post-stratified estimators of crop yield.
using spectral data in the form of vegetation indices for stratification of cropped area. Global Positioning System (GPS) has been used for collecting the data for identification of crop plots of the survey data. Singh et al. (2002) have given small area estimates of crop yield at Tehsil/Block level. Two small area estimators namely Direct and Synthetic estimators have been developed from general crop yield estimation surveys on crop cutting experiments. Rai et al (2007) have developed spatial models for crop yield estimation. This procedure is also extended for estimation of crop yield at small area levels.

Many studies have been carried out to relate plant biometrical characters and spectral parameters. By relating the reflectance data of individual crops, in specific wavelength regions to canopy growth or vigor, it is possible to predict yield estimates using appropriate modeling techniques backed up by adequate ground truths. The availability of regular space observations has simulated the development of computerized agriculture information systems in several countries. Development of reliable crop-yield models with minimal data has become a major thrust area in India also. Studies related to use of data from different spectral regions, generation of models using growth-profile parameters and incorporation of agro-meteorological information in the simple yield models have been carried on widely.

Over the past decade, a number of crop yield forecasting models using remote sensing inputs have been developed and used in making forecasts. These include (a) Single date RS-based models (b) Spectral profile related growth parameters derived from multi-date data (c) Combinations of different parameters like trend, RS and meteorological parameters either by including all in a multiple linear regression equation or by optimal combination of different estimates. Singh and Ibrahim (1996) examined the use of multi-date satellite spectral data for crop yield modeling using Markov Chain model. Saha (1999) used satellite data and GIS for developing several yield models for forecasting. Singh et al. (2000 b) have given spectral and integrated models using spectral data and farmers eye estimate for forecasting crop yield.
Agricultural fields are spatial in nature, if we consider productivity of the field with respect to certain crop provided other factors as constant. It does not change abruptly from one field to another. The change is very gradual and the neighboring fields have more or less same structure. In general crop-cutting experiments (CCE) are carried out for yield estimation in selected villages. It may be noted that from crop cutting experiment the estimation of production at lower level like Tehsil's/villages etc is difficult and prone to large error due to small sample size. The application of spatial statistics in agriculture to improve the prediction and estimation may be a useful attempt for small area levels. Spatial characteristics and Crop Cutting Experiment (CCE) will help us in giving the better estimate at lower level also. With the help of available literature in the field of spatial statistics, it is possible to apply suitable spatial models to predict the production surfaces i.e. values of production at each point of the map, of the target region. Field sizes of our country are fairly small; therefore it is more appropriate to utilize the spectral data as auxiliary information.

Gupta (2002) and Rai et al. (2007) have developed an integrated model for wheat crop yield estimation using the survey data on wheat crop yield from CCE along with the satellite spectral data in the form of vegetation indices i.e. Normalized Difference Vegetation index (NDVI). The study proved that the use of remote sensing satellite data along with the crop yield data based on CCE can greatly improve the efficiency of crop yield estimators at small area level.

Gupta (2007) has developed spatial prediction model under Bayesian framework considering four different situations i.e. (i) using prior information about parameters which is non-informative for known and unknown variance (ii) using prior information about parameters which is informative for known and unknown variance, (iii) using prior information about parameters as natural conjugate prior for known and unknown variance and (iv) using fuzzy approach for linear interval model for vague characters under study. It was shown in this study through simulation that Bayesian spatial prediction models are always better than usual spatial models. This may be due to the
fact that information contained in the sample as well as about the parameter of the model has been utilized in the estimation procedure. Further, there is significant gain in the precision in case of geographical variables when spatial effects were taken into account in the estimation procedure under Bayesian framework.

A study was undertaken by Rai et al (2004) in district of Lalitpur in UP state of India due to the fact that this district has been observed to have considerable area under most of the land use classification categories. It has been observed in this study that quality of revenue records in the study area i.e. Lalitpur district is quite reliable for most of the usual nine fold classified land use classes. The statistics of land use classes were restricted to five broader classes, which can be identified by using single time digital data of Remote Sensing out of the above nine-fold classification. This can be easily obtained using RS. These statistics of land use classes obtained through RS could be used as auxiliary information in spatial /non-spatial models to get reliable statistics of different classes. The above models can be used to predict the statistics related to these classes for non-surveyed area/villages of the districts. Hence, it is possible to develop reliable land use statistics at any smaller level i.e. panchayat/block/tehsils using above models. To take care of spatial dependence of the neighboring units, the classical sampling technique approach is being modified such that the probabilities of selecting neighboring units, once a particular unit is selected in the sample, becomes less as compared to distant areal units. The best fitted spatial model for each class of land use was found to be different, depending on the spatial distribution of the land use class patches of land in the district. The prediction of area under different land use categories covered under nine fold classification based on satellite data using spatial model seems to be quite satisfactory.

A study was conducted (Ahmad et al. 2003) for identification of potential area for agroforestry in Yamuna Nager district of Haryana state of India. A new Objective Analytical Hierarchy Procedure (OAHP), which is based on the statistical techniques, has been proposed. Under this procedure important factors responsible for growth of agro-forestry have been identified and the Suitability Index of agro-forestry for each village
using Spatial-Analytic Hierarchy Process has been constructed. This index was compared with the corresponding composite development index for each village. All the villages of Yamuna Nagar district of Haryana State have been categorized as High, Moderate and Low potential villages for agro-forestry using this technique. This procedure can also be helpful to study the social, economic and environmental development.

Further, Jha (2009) has proposed a Functional Spatial Regression (FSR) model under Measurement Errors (ME) and a Structural Spatial Regression (SSR) model under ME for estimation of regression coefficients in case of spatial variables. Explanatory variables under FSR model are assumed to be fixed while it is random in case of SSR model. Modified estimates of spatial regression coefficients were proposed following Ordinary Least Squares (OLS), Generalized Least Squares (GLS), Maximum Likelihood Estimation (MLE) and Method of Moment Estimation (MME) approaches in both model structures. It has been shown through spatial simulation that proposed estimators are unbiased, consistent and comparatively more efficient than corresponding usual estimators. Four different approaches were followed to incorporate spatial effects in the proposed models.

In case of spatially correlated population it is expected that neighboring units are likely to be more homogeneous then distant units. Kankure and Rai (2007) 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 than 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. 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.

Gharde et al. (2012) has obtained small area estimates for spatial unit level model under Hierarchical Bayes (HB) framework. Three different weighting approaches were used to incorporate the spatial effects in the model in the form of elements in the weight matrix. Study shows that posterior mean for all three weighting approaches are almost same in all small areas but variances are different. Variances are least for all small areas in case of spherical variogram approach as compare to other two methods. It can easily be seen that neighbourhood criteria method performs well as compare to Gaussian-decay function method of incorporating spatial-effects.

**Small Area Estimation**

In small area estimation literature, use of unbiased direct estimators for small areas are generally not recommended and are considered too variable to be of any practical use. Chandra and Chambers (2009) have proposed a class of model based direct estimators (MBDE) for small areas that appear to overcome this objection, in the sense that these estimators are comparable in efficiency to the indirect model-based small area estimators that are now widely used. These MBDE estimators are weighted linear combinations of the actual sample data from the small areas of interest. As a result, there are many practical advantages associated with such MBDE estimators, for example, mean squared error estimation. In fact, in this case the weights borrow strength via a model that explicitly allows for small area effects. In general surveys, normally there are more than one variable (i.e., multivariate in nature). Chandra and Chambers (2009) have introduced the multipurpose sample weights for small area estimation of multivariate surveys. That is, they extended the MBDE approach to multipurpose weighting based small area
estimation. In many surveys, e.g. agricultural, environmental and business surveys, data are typically skewed and linear model assumptions are questionable. Salvati et al. (2010) have further extended the MBDE approach of small area estimation based on Nonparametric model to produce estimate when underlying model is not known or mean function is incorrectly specified. Commonly used methods for small area estimation are based on assumption that a linear mixed model can be used to characterize the relationship between the survey variable Y and an auxiliary variable X in the small areas of interest. In particular, Empirical Best Linear Unbiased Prediction (EBLUP) and model-based direct (MBDE) estimation (Chambers and Chandra, 2009) are typically based on linear model assumptions. However, when the data are skewed, the relationship between Y and X may not be linear in the original (raw) scale, but can be linear in a transformed scale, e.g. the logarithmic scale. In such cases we would expect estimation based on a linear model for Y to be inefficient, and an appropriate technique for small area estimation should then be based on a linear mixed model for a transformed version of Y. Chandra and Chambers (2011) have proposed small area estimation techniques for variables that can be modeled linearly following a non-linear transformation. In particular, they extended the model-based direct estimation approach to small area estimation for skewed data that are consistent with a linear mixed model in the logarithmic scale. Recently for such data, Berg and Chandra (2012) have proposed an empirical best predictor for small area means in the sense that it has minimum mean squared error in the class of unbiased predictors.

Presence of excess zeros in the data is a well-known problem in survey estimation. When the focus is on small area estimation using survey data, presence of excess zeros within a small area are clearly much more influential than they are in the larger overall sample. Standard approaches of small area estimation are inefficient for zero-inflated data situations. Chandra and Sud (2012) have developed an appropriate methodology for such data situations. In particular, a method was developed for small area estimation using the mixture model that accounts for presence of excess zeros in the data. Many often population level auxiliary information is not available and sometimes this information exist but are not accessible for use in survey estimation. In some cases, population and
sample data can have inconsistency with respect to definition of auxiliary information since they are collected from different agencies/sources. In such situations, we are not even able to make full use of auxiliary information available in the survey data. Chandra et al. (2013) have discussed a small area estimator for small area means for the situation when population level auxiliary information is not available. This method of small area estimation uses estimated population level auxiliary information using survey weights.

Commonly used method of small area estimation assumes that the relationship between the variable of interest (Y) and auxiliary information (X) is same over the study space. However, in many data sets this is not true and the process is referred to as the spatial non-stationary. Chandra et al. (2012) have proposed a Geographical Weighted Pseudo Empirical Best Linear Unbiased Predictor (GWPEBLUP) for small area averages, and developed a conditional approach for estimating its mean squared error. A crucial aspect of small area estimation is estimation of the mean squared error of the resulting small area estimators. Chambers et al. (2011) have introduced robust mean squared error estimation for linear predictors of finite population domain means. Their approach represents an extension of the well known ‘sandwich’ type variance estimator used in population level sample survey inference, and appears to lead to a mean squared error estimator that is simpler to implement, and potentially more robust, than alternatives suggested in the small area literature. They have also demonstrated the usefulness of this approach through both model-based as well as design-based simulation.

Much of the small area estimation literature focuses on population totals and means. However, users of survey data are often interested in the finite population distribution of a survey variable, and the measures (e.g. medians, quartiles, percentiles) that characterize the shape of this distribution at small area level. Salvati et al. (2012) have described MBDE estimator of the small area distribution function. The MBDE is defined as weighted sum of sample data from the area of interest, with weights derived from the calibrated spline-based estimate of the finite population distribution function, under an appropriately specified regression model with random area effects. They have also illustrated the MSE estimation of the MBDE.
Recently proposed outlier robust small area estimators can be substantially biased when outliers are drawn from a distribution that has a different mean from that of the rest of the survey data. This naturally leads to the idea of an outlier robust bias correction for these estimators. Chambers *et al.* (2013) have developed this idea and also proposed two different analytical mean squared error estimators for the ensuing bias corrected outlier robust estimators. Chandra *et al.* (2011a) have illustrated this in the Debt-Investment Survey 2002-03 of NSSO and the Population Census 2001 and the Agriculture Census 2003 and estimated the proportion of indebted households at district as well as at district by land holding levels in the State of Uttar Pradesh. Similarly, Chandra *et al.* (2011b) have employed small area estimation approach to derive the estimates of proportion of poor households at district level in the State of Uttar Pradesh in India by linking data from the Household Consumer Expenditure Survey 2006-07 of NSSO's 63rd round and the Population Census. Both these studies show that the small area estimates are precise and representative for the area they belong. Moreover, they have acceptable value of coefficient of variation. Chandra (2013) has applied small area estimation method for crop yield estimation at district level using Improvement of Crop Statistics Scheme Data and Linking with Census Data.

*Calibration Studies*

Sud *et al.* (2012) have developed a calibration approach based product type estimator which performs better than the usual product estimator. The double sampling case has also been dealt with. A higher calibration approach based estimator of variance has been developed (the estimator is used when there is negative correlation between study and the auxiliary variable i.e. hours of sleep and age). A simulation study has also been carried out to demonstrate the theoretical results. (Sud *et al.* 2012) The calibration approach has been used to develop a regression type estimator for inverse relationship between study and auxiliary variable. The proposed estimator performs better than the usual regression estimator. The double sampling case has also been dealt with. Higher order calibration based estimator is also developed. Results of theory are illustrated through simulation study. (Sud *et al.* 2013). The non-response situation has also been developed using
calibration approach. The Hansen and Hurwitz estimator has been extended to the case when auxiliary information is available for non-response stratum, entire population and for the case of double sampling. Theory is illustrated through simulation. (Raman et al. 2013).

**Balanced Sampling Plans Excluding Adjacent Units**

Balanced sampling plans excluding adjacent units \( \text{BSA}(m) \) plans are the sampling plans in which any pair of units which are separated by a distance of \( m \) units do not appear together in a sample whereas all other pairs of units appear together equally often in the samples, given that the population units are arranged in circular order. In BSA\((m)\) plans, first order inclusion probabilities are constant and second order inclusion probabilities are zero for pairs of adjacent units and constant for other pairs of units. These plans are useful for sampling from populations in which the nearer units provide similar observations due to natural ordering of the units in time or space.

An important series of binary, proper, incomplete block designs called polygonal designs are useful for obtaining BSA\((m)\) plans. A polygonal design is an arrangement of \( v \) treatments in \( b \) blocks of size \( k (v \) with \( r \) replications and distance \( m \) such that any two treatments with distance less than or equal to \( m \) do not appear together in a block and any other pair of treatments with distance more than \( m \) appear together in same number of blocks. If population size is equal to number of treatments in a polygonal design and sample size is equal to block size then the requisite BSA\((m)\) plan can be obtained by assigning equal probability of selection to each of the blocks of a polygonal design.

Mandal, Parsad and Gupta (2008a) developed two approaches for computer-aided construction of BSA\((m)\) plans. In the first approach, an algorithm for construction of polygonal designs/BSA\((m)\) plans is developed using method of symmetrically repeated differences. In the second approach, BSA\((m)\) plans are obtained using linear programming to minimize the probability of selecting the samples which contain contiguous units subject to the constraints of specified first and second order inclusion probabilities of BSA\((m)\) plans. The methods are illustrated with numerical examples. A
catalogue of 75 new polygonal designs in the range $v \leq 40, b \leq 400, k \leq 7$ and $m \leq 4$ is also given.

Linear programming approach used above may yield rational values for the decision variables in the range of 0 and 1. To overcome this problem, a linear integer programming approach may be used permitting only non-negative integral values for the decision variables. Linear integer programming approach has two major advantages. Firstly, it is more efficient than linear programming approach since it uses much less number of constraints and less number of variables in the objective function and secondly the support of the generated plan is expected to be much smaller than the plans obtained through linear programming approach. Mandal, Gupta and Parsad (2011) generated BSA ($m$) plans or PDs for given $v$, $k$, $\lambda$ and $m \leq \left\lfloor \frac{v}{2} \right\rfloor$ using linear integer programming approach.

In some situations, the units may be arranged in a linear order rather than in circular fashion. Mandal, Parsad and Gupta (2008a) have also obtained BSA plans for linear ordering of units. It may be noted here that first order inclusion probabilities are not equal for all units under linear BSA plan.

BSA ($m$) plans suffers from the drawback that the unbiased estimation of variance of Horvitz-Thompson estimator of population mean is not possible. To tackle this problem, Mandal, Parsad and Gupta (2009), introduced a family of distance balanced sampling plans (DBSP) with the property that the second order inclusion probabilities are non-decreasing function of distance between the two concerned units. Unbiased estimation of variance of Horvitz-Thompson estimator of population mean for DBSP is considered. The conditions for DBSP to be more efficient than simple random sampling without replacement (SRSWOR) and BSA ($m$) plans have been obtained. DBSP can give rise to a large number of sampling plans depending on choice of the distance function. Three particular members of the family DBSP namely two points, three points and $\left\lfloor \frac{N}{2} \right\rfloor$-points
DBSP were investigated for suitable choice of distance function. It has been shown that if the assumption of decreasing correlation between the units as the distance between the units increases holds then the proposed plans are more efficient than other alternative sampling plans such as SRSWOR and BSA \((m)\) plans. A class of incomplete block designs, called as distance balanced incomplete block (DSBIB) designs is introduced, whose blocks can act as a support of the DBSP. Existence and construction of DSBIB designs has also been given. Mandal, Parsad and Gupta (2010) used linear programming approach to obtain DBSP and provided a list of all DBSP with sample size for population sizes upto 100.

Mandal, Parsad and Gupta (2008b) developed IPPS plans excluding adjacent (IPPSEA plans) units for sampling from populations where there is natural ordering of the units along with variability in sizes of the units. IPPSEA plans have been obtained by making use of binary, proper and unequireplicated block designs and linear programming approach. The performance of proposed IPPSEA plans using Horvitz-Thompson estimator was compared with other alternative sampling plans such as SRSWOR, BSA \((m)\) plans, probability proportional to size with replacement, Hartley and Rao’ strategy, Rao, Hartley and Cochran’s strategy and Sampfords’ IPPS plan using the a real life population. It was seen that the proposed plan performs quite better compared to alternative plans. Unbiased estimation of Horvitz-Thompson estimator of population total is not possible in these types of plans because some of the second order inclusion probabilities are zero. To resolve this problem, one approximate variance estimation technique has been suggested.
References:


<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
</tr>
</thead>
</table>


