STATISTICS IN TRANSITION-new series, March 2012
Vol. 13, No. 1, pp. 47—64

CROP ACREAGE AND CROP PRODUCTION ESTIMATES FOR SMALL DOMAINS - REVISITED

G.C. Tikkiwal[1], **Alka Khandelwal**[2]

ABSTRACT

For any country advance and final estimates of yield of principle crops, at National and State levels, are of great importance for its macro level planning. But, for decentralized planning and for other purposes like crop insurance, loan to farmers, etc., the reliable estimates of crop production for small domains are also in great demand. This paper, therefore, discusses and review critically the methodology used to provide crop acreage and crop production estimates for small domains, based on indirect methods of estimation, including the SICURE model approach. The indirect methods of estimation so developed use data obtained either through traditional surveys, like General Crop Estimation Surveys (GCES) data, or a combination of the surveys and satellite data.

Key words: Timely Reporting Scheme (TRS); General Crop Estimation Surveys (GCES); Simulation-cum- Regression (SICURE) model.

1. Introduction

The advance and final estimates of crop production of principle crops at national and sub-national level like districts, counties, blocks for any country are of importance for its macro and micro level planning.

In many countries, including India, the yield rate of principle crops are being estimated through crop-cutting experiments. The technique of crop-cutting experiments is mostly developed in India in early seventies. The estimation of crop yield is done under the national programme known as General Crop Estimation Surveys (GCES) using crop-cutting experiments. The GCES are being conducted through survey methodology developed mostly in 1940's [Mahalanobis (1946), Sukhatme and Agarwal (1946-47, 1947-48)].

[1] Department of Mathematics and Statistics, Jai Narain Vyas University, Jodhpur – 342005, India. E-mail: gctikkiwal@yahoo.com.
[2] Department of Mathematics and Statistics, Jai Narain Vyas University, Jodhpur – 342005, India. E-mail: alkakhandelwaljdr@gmail.com.

The estimation of crop yield involves two components viz. the estimation of crop acreage and the estimation of yield rates. As regard crop acreage estimation, a scheme known as Timely Reporting Scheme (TRS) has been in vogue since early seventies in most of the States of India. The TRS has the objective of providing quick and reliable estimates of crop acreage statistics and thereby production of the principle crops during each agricultural season on the basis of 20 percent sample villages, using direct estimators. The performance of direct estimators is satisfactory at national and state level, as the sampling error of the estimators is within 5 percent, but not at lower levels as shown by Tikkiwal and Tikkiwal (1998), Tikkiwal and Ghiya (2000, 2004). The authors developed and used synthetic and composite methods of estimation to provide crop acreage estimation for small domains. Further, it has been observed that composite method of estimation is easy to apply and this approach overcomes the limitation of synthetic estimator to some extent [cf. comments by Francisco (1998, p.254)]. Where this approach does not work, then SICURE model can be tried or other model based estimation methods may provide satisfactory results.

Apart from traditional approach the remote sensing technologies were initiated after launch of many advanced satellites, to provide crop acreage and crop production estimates for major and minor domains [Dadhwal et al. (1985)]. For example, the National Agricultural Statistics Service (NASS) of the United States of America has been using Landsat series of satellites since 1950's, and France entered the field of earth resources satellites in 1986 with the launch of SPOT-I [cf. Bellow et al. (1996)]. In India this work has been entrusted to Indian Space Research Organization. The model based estimation methods for small domains, using survey and satellite data has been developed over a period of time by authors Battese et al. (1988), Singh et al. (1992), Shaible and Casady (1994), Srivastava (2007) and others. These methods may provide efficient estimators provided a suitable model is selected and there should not be problem of mixed cropping.

This paper provides a comprehensive review of the work done on crop production estimates of small domains. In this paper Section 2 describes the methods of estimating crop acreage statistics. The Section 3 describes the crop-cutting experiments and presents method of estimation of crop yield. Section 4 discusses and review the methodology used to provide crop acreage and crop production estimates for small domains, based on survey data, whereas the model based estimation methods for small domains using survey and satellite data are discussed in Section 5.

2. Crop acreage statistics

In Temporarily Settled States (the states, in which land revenue is fixed for a definite period of years and is subject to revision at the end of this period) of India crop acreage statistics are collected on complete enumeration basis, whereas in

Permanently Settled States (the states, in which land revenue is fixed for perpetuity) they are estimated through selection of 20 percent villages. In order to provide quick and reliable advanced estimates of the crop production, in temporarily settled states also crop acreage statistics are estimated under Timely Reporting Scheme (TRS). The TRS has been in vogue since early seventies in these States of India. Under the scheme the Patwari (Village Accountant) is required to collect acreage statistics on a priority basis in a 20 percent sample of villages. These villages are selected by stratified linear systematic sampling scheme, taking Tehsil as a stratum. These statistics are further used to provide state level estimates using direct estimators viz. unbiased (based on sample mean) and ratio estimators.

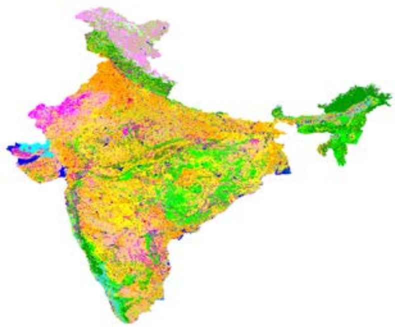
The performance of both direct estimators in the state of Rajasthan, like in other states, is satisfactory at state level, as the sampling error is within 5 percent. However, the sampling error of both direct estimators increases considerably, when they are used for estimating acreage statistics of various principle crops even at district level, what to speak of levels lower than a district. Tikkiwal and Ghiya (2000, 2004) notice that the sampling error of direct ratio estimator for Kharif crops of Jodhpur district (of Rajasthan state) for the agricultural season 1991-92 varies approximately between 6 to 68 percent. Therefore, there is a need to use indirect estimators at district and lower levels for decentralized planning and other purposes like crop insurance, bank loan to farmers. As regards estimation of yield rates, it is being done through crop-cutting experiments.

It may be noted here that for administrative purposes India is divided into the number of states, each state consists of a number of districts, each district consists of a number of tehsils and further each tehsil consists of a number of villages.

The crop acreage statistics are also collected by the Indian Space Research Organization under its Crop Acreage and Crop Production (CAPE) project, through remote sensing technology. But due to mixed cropping pattern, prevailed in India, this technique of the crop acreage statistics are not so reliable.

Land Use and Land Cover statistics of India and the state of Rajasthan are shown here using the satellite data obtained from Regional Remote Sensing Centre-West, Indian Space Research Organization, Department of Space (India).

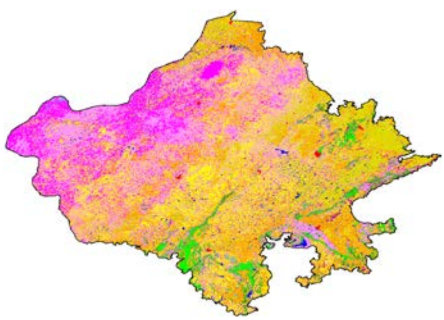
LAND USE / LAND COVER STATISTICS OF INDIA (2010-11)



Categories	Area (Sq Km)
Built-up	20705.73
Kharif only	540441.5
Rabi only	258442.8
Zaid only	9387.03
Double / tripple	544436.1
Current fallow	385835.11
Plantation/orchard	67798.33
Evergreen forest	167870.67
Deciduous forest	337899.13
Scrub/Deg. forest	145263.53
Littoral swamp	4659.24
Grassland	74915.82
Other wasteland	289810.76
Gullied	9986.8
Scrubland	186970.02
Water bodies	79362.73
Snow covered	40645.31
Shifting Cultivation	1769.02
Rann	19274.94
Total	**3185474.57**

Source: AWIFS Satellite data (2010-11)

LAND USE / LAND COVER STATISTICS OF RAJASTHAN (2010-11)



Categories	Area (Sq Km)
Built-up	855.74
Kharif only	46773.01
Rabi only	40726.94
Zaid only	869.21
Double/Triple crop	52818.26
Current fallow	56693.91
Deciduous forest	8406.61
Shrub/scrub/degraded forest	9792.06
Grassland/grazing land	121.39
Other wastelands	67727.72
Gullied/ravines land	1169.01
Scrub land	39627.74
Waterbodies	4684.84
Total	330266.43

Source: AWIFS Satellite data (2010-11)

In Indian system there are mainly three agricultural season's viz. Kharif, Rabi and Zaid.

(i) Kharif crops – the crops sown in June-July and harvested in October-November every year.

(ii) Rabi crops – the crops sown in November-December and harvested March-April every year.

(iii) Zaid crops – the crops grown between March and June are known as Zaid.

3. Estimation of crop yield

Final estimates of crop production based on complete enumeration of area and yield become available much after the crops are actually harvested. However, the Government may require advance estimates of production for taking various policy decisions relating to pricing, marketing, export/import, distribution, etc. Considering the genuine requirement of crop estimates much before the crops are harvested for various policy purposes, a time schedule of releasing the advance estimates has been evolved under a national programme known as "General Crop Estimation Surveys (GCES)". The GCES uses the technique of crop-cutting experiments.

3.1. Crop – cutting experiments under GCES

The most important factor of Crop production statistics is the estimation of yield rates. Presently the yield rates are estimated through crop-cutting experiments under GCES. The GCES covers 68 crops (52 foods and 16 non foods) in 22 states and 04 union territories. Such surveys are conducted twice a year to cover different types of crops.

About five hundred thousand crop-cutting experiments, for major crops throughout the country, are conducted annually under this programme. The sampling design adopted for the GCES is a multistage stratified random sampling with tehsils/inspector land revenue circles/community development blocks, etc. as strata, the villages selected randomly form the primary stage sampling unit, the fields from each selected village formed the second stage sampling unit and the experimental plot within the field form the ultimate stage of sampling. A sample of villages is selected from different strata in proportion to the area under crop. From each selected village, two fields are selected randomly and from each field a plot of fixed shape and size usually measuring (5meter x 5meter) is selected for recording the green yield by actual harvesting the crop.

3.2. Estimation Procedure

Estimation procedure for estimating of crop yield through Crop Estimation Surveys:

The methodology generally adopted for estimating the average yield of crop is as below:

At the stratum (tehsil) level, the estimated average yield of the crop is obtained as a simple arithmetic mean of plot yields. For this, let

Y_{ijk} – The green yield (net in gms/plot) of the k-th plot in the j-th village in the i-th stratum.

n_{ij} – Number of experiments analyzed in the j-th village of i-th stratum.

m_i – Number of villages in which experiments are analyzed in the i-th stratum.

n_i – Number of experiments analyzed in the i-th stratum.

S – Number of strata in a district.

a_i – The area (net) of the crop in the i-th stratum.

f – The conversion factor for converting the green yield per plot into the yield of dry marketable produce per hectare.

Stratum level average of the green yield for the i-th stratum is

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{m_i} \sum_{k=1}^{n_{ij}} Y_{ijk}$$

and further, District level estimated average yield of the dry marketable produce per hectare is given by

$$\bar{Y} = \frac{\sum_{i=1}^{s} a_i.\bar{Y}_i}{\sum_{i=1}^{s} a_i}. f$$

Then, $\bar{Y}$ is to be multiplied by the District level crop acreage estimates of that particular crop, to have an estimate of the yield.

4. Estimation for small domains using survey data

For decentralized planning and other purposes like crop insurance, loan to farmers, etc., the Governments need reliable agricultural statistics for small domains like district, CD block, counties, etc. But the estimates provided by National Agencies such as NSSO, TRS and EARAS are generally reliable at the state level and not at district level. In such situation "Small Area Estimation" methods hold out a promising solution.

4.1. Synthetic and composite methods

Tikkiwal, B.D. and Tikkiwal, G.C. (1998) in their invited paper presents an excellent review of the landmarks in the development of crop yield and acreage statistics in India and other developing countries. As regards providing estimates of average crop yield at small area (Assistant Agricultural Officer (AAO) circle) level the authors use direct methods only, because the sample size was sufficiently large. In the absence of such information/sufficient data the SICURE model can be helpful. The authors further demonstrate the use of synthetic and composite estimators to provide reliable acreage statistics at small area levels. The small areas in this study are Inspector Land Revenue Circles (ILRC's), the sub-groups of Tehsils. The study suggests the use of composite estimators, if the synthetic assumption closely meets. When this assumption does not meet, they suggest the use of other types of estimators such as those obtained through the SICURE model (1993). The following discussant Francisco Juvier Gallego's (1998, p. 254) comments on this paper show applicability of the results.

"…The approach might overcome some limitations of synthetic estimators and looks easier to apply than other small area estimation procedures that have been used in agricultural statistics [For example, Battese et al. (1988)]. Some additional clarifications would be of interest on the computation of variance from a single sample. If the results presented are confirmed in other countries, the method would be of interest, and not only for developing countries, as stated in the paper. Actually, India is a developed country if we speak about statistics".

Tikkiwal, G.C. and Ghiya, A. (2000) define and discuss a generalized class of synthetic estimators with application to crop acreage estimation for small domains (ILRC's), using auxiliary information, under different sampling schemes. The generalized class of synthetic estimators, among others, includes the simple, ratio and product synthetic estimators. The proposed class of synthetic estimators gives consistent estimators if the synthetic assumption holds. Further, the authors compare the relative performance of a number of synthetic estimators with direct estimators, empirically, through a simulation study using live data. The study reveals that for the domains where synthetic estimators do not deviate considerably from their corresponding assumption, performance of the synthetic estimator is satisfactory. When the synthetic estimators deviate considerably from their corresponding assumption, then the authors suggest to look for other types of estimators such as those obtained through the SICURE model [Tikkiwal (1993)].

Sisodia and Singh (2001) develop three synthetic estimators of total crop production Y_i of i-th block (small area) level using crop production and other relative information at district level, as given below:

Estimator (1)

$$\hat{Y}_i = \left(\sum_{j=1}^{P} W_j X_j \right) \hat{\bar{Y}} \qquad (4.1.1)$$

where, $\hat{\bar{Y}} = \dfrac{\hat{Y}}{A}$; $\hat{Y}$ is obtained through multiple linear regression model.

A = Area under the crop in a given year

X_j = Value of j-th predictor at the block level in a given year

W_j = Weight assign to each predictor.

Estimator (2) & Estimator (3) are of the form

$$\tilde{Y}_i = b_i \ \hat{Y}_i$$

where, b_i are constants such that $\sum_{i=1}^{a} \tilde{Y}_i = \sum_{i=1}^{a} b_i \hat{Y}_i = Y$

For $b_i = b$ (constant) second estimator of Y_i is, i.e.

Estimator (2)

$$\tilde{Y}_i^{(1)} = \hat{Y}_i \frac{Y}{\sum_{i=1}^{a} \hat{Y}_i} \tag{4.1.2}$$

Y = actual crop production reported at district level through crop-cutting experiments in a given year.

For $b_i = 1 + \dfrac{\left(Y - \sum_{i=1}^{a} \hat{Y}_i \right)}{a \hat{Y}_i}$ third estimator of Y_i is, i.e.

Estimator (3)

$$\tilde{Y}_i^{(2)} = \hat{Y}_i + \frac{\left(Y - \sum_{i=1}^{a} \hat{Y}_i \right)}{a}$$

(4.1.3)

a = Number of blocks in the district.

Further, the authors carried out an empirical study for rice crop in Faizabad district of Uttar Pradesh during the years 1981-82 and 1982-83 to compare the relative efficiency of these estimators under multiple linear regression model. The relative efficiency of $\tilde{Y}_i^{(1)}$ and $\tilde{Y}_i^{(2)}$ over $\hat{Y}_i$ comes out to be same for all the blocks, i.e. 88.84% and 105.88% respectively during the year 1981-82. Similarly, during the year 1982-83 it comes out to be 110.80% and 105.88% respectively.

Thus $\tilde{Y}_i^{(2)}$ is found to be most efficient when comparing with $\hat{Y}_i$ and $\tilde{Y}_i^{(1)}$ in case 1 when weights are given to be more than 1. In case 2 when weights are less than 1 both estimators $\tilde{Y}_i^{(1)}$ and $\tilde{Y}_i^{(2)}$ are found to be more efficient than $\hat{Y}_i$. But $\tilde{Y}_i^{(2)}$ need not be the most efficient estimator. The results presented in the Table 4 and Table 5 (p. 313 & 315) does not correlate with the findings, when the estimated values are compared with the actual estimates based on crop-cutting experiments.

All the three estimators considered by the authors are nothing but synthetic, regression type, estimators and, therefore, their efficiency depends on the validity of the assumption of the corresponding synthetics estimator under use. Also, the three estimators are design-biased; therefore, ignoring the bias remains a serious limitation. But these estimators can be further improved upon by the technique of composite estimation. [cf. Tikkiwal and Ghiya (2004)].

Tikkiwal and Ghiya (2004) define and discuss a generalized class of composite estimators for small domains, using auxiliary information, under different sampling schemes. The proposed estimator of population mean $\overline{Y}_i$, based on auxiliary variable 'x' under SRSWOR design is defined as:

$$\overline{y}_{c,i} = w_i \overline{y}_i \left(\frac{\overline{x}_i}{\overline{X}_i} \right)^{\beta_1} + (1 - w_i)\overline{y} \left(\frac{\overline{x}}{\overline{X}_i} \right)^{\beta_2} ; \quad (0 \le w_i \le 1) \qquad (4.1.4)$$

where, β_1 and β_2 are suitably chosen constants.

The estimator $\overline{y}_{c,i}$ is a weighted sum of the generalized direct estimator [Srivastava (1967)] and the generalized synthetic estimator [Tikkiwal and Ghiya (2000)].

The proposed estimator has desirable consistency property (in traditional sense), when the following assumption is satisfied.

$$\overline{Y}_i \left(\overline{X}_i\right)^{\beta_2} \cong \overline{Y} \left(\overline{X}\right)^{\beta_2} \qquad (4.1.5)$$

It is to be noted that the synthetic estimator may be heavily biased unless the above assumption is satisfied [cf. Tikkiwal and Ghiya (2000), Eq. (4.1)].

The proposed generalized class of composite estimators includes a number of direct, synthetic and composite estimators as special cases. Here follows a list of such estimators with corresponding choice of values of the different constants.

Table 4.1. Various Direct, Synthetic and Composite Estimators as Special Cases of the Generalized Composite Estimators

S No	Estimator	w_i	$(1-w_i)$	β_1	β_2
1	Simple Direct $(\bar{y}_i)$	1	0	0	-
2	Simple Synthetic $(\bar{y})$	0	1	-	0
3	Simple Ratio $\left[\left(\bar{y}_i / \bar{x}_i\right)\bar{X}_i\right]$	1	0	-1	-
4	Ratio Synthetic $\left[\left(\bar{y} / \bar{x}\right)\bar{X}_i\right]$	0	1	-	-1
5	Simple Product $\left[\left(\bar{x}_i / \bar{X}_i\right)\bar{y}_i\right]$	1	0	1	-
6	Product Synthetic $\left[\left(\bar{y} / \bar{X}_i\right)\bar{x}\right]$	0	1	-	1
7	Composite : Combining simple direct with simple synthetic $w_i\bar{y}_i+(1-w_i)\bar{y}$	w_i	$(1-w_i)$	0	0
8	Composite: combining simple direct with ratio synthetic $w_i\bar{y}_i+(1-w_i)\frac{\bar{y}}{\bar{x}}\bar{X}_i$	w_i	$(1-w_i)$	0	-1
9	Composite : combining simple ratio with ratio synthetic $w_i\frac{\bar{y}_i}{\bar{x}_i}\bar{X}_i+(1-w_i)\frac{\bar{y}}{\bar{x}}\bar{X}_i$	w_i	$(1-w_i)$	-1	-1

Further, the authors comparing the various empirical results of Absolute Relative Bias (ARB) and Simulated relative standard error (Srse), draw the conclusion that if the synthetic estimators do not deviate considerably from their corresponding assumptions (describe in Eq. 4.1.5), then performance of the composite estimators (given at S.No.9 in the Table 4.1), based on a sample of 20% villages, is satisfactory at the level of ILRCs. Therefore, these estimators will certainly perform better up to the level of district. When the given condition is not satisfied we should look for other methods of estimation. One of such method is to use SICURE model (1993) or the methods presented in Ghosh & Rao (1994).

Sharma, Srivastava and Sud (2004) consider two different synthetic estimators based on auxiliary variables for providing crop yield estimate at Gram Panchayat (small area) level. The proposed estimators for i-th Gram Panchayat (GP) are defined as follows:

$$\hat{T}_i = \frac{\bar{x}_i}{\sum_{i=1}^{a} A_i \bar{x}_i} A \hat{\bar{Y}}, \quad \text{i = 1, 2,..., a} \qquad (4.1.6)$$

and,

$$\hat{T}_i^{'} = \bar{x}_i + \hat{\beta}_{iopt} \left[\hat{\bar{Y}} - \frac{\sum_{i=1}^{a} A_i \bar{x}_i}{A} \right] \qquad (4.1.7)$$

$$\hat{\beta}_{iopt} = \frac{\left(A_i \big/ A \right) \hat{\sigma}_{x_i}^2}{n_i \left[\hat{V}\left(\hat{\bar{Y}} \right) + \frac{1}{A^2} \sum_{i=1}^{a} \frac{A_i^2 \sigma_{x_i}^2}{n_i} \right]}$$

a = number of GP in a block.

A_i = Area under a particular crop for the i-th GP.

$A = \sum_{i=1}^{a} A_i$ = Total area under the crop in the block.

N_i = Number of farmers in the i-th GP.

n_i = Number of farmers selected in the i-th GP for obtaining information about the expected yield of the crop grown in the field.

x_{ij} = expected yield as obtained from j-th farmer in the i-th GP; j = 1,2,...,n_i.

$\hat{\bar{Y}}$ = block level estimate of crop yield as obtained through the method of crop-cutting experiment.

$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$, average of expected yield of i-th GP.

In the empirical study the proposed estimators are based on crop yield estimates obtained through crop-cutting experiments under General Crop Estimation Surveys and estimate of crop production obtained through data collected from a fresh selected random sample of 10 farmers from each of all the GP's in a district. Analysis of data obtained from a survey carried out on wheat crop in the Basti district of the state of Uttar Pradesh in India in the year 2000 revealed that both the estimators perform satisfactorily in terms of the criterion of percentage root mean squared error as it varies from 1% to 10% in most of the cases. The biases of both the estimators are also negligible. Here, it may be noted that the estimators proposed by Sharma et al. (2004) depend on the estimate of crop production obtained through crop-cutting experiments and on the basis of fresh samples selected independently from each Gram Panchayat (GP) of the block. In the case study, for example, a block roughly consists of 90 GP's which results in a selection of an additional sample of 900 farmers. This method, therefore, does not fall within the domain of small area estimation methods. Also basis of the proposed estimator is the assumption that "over estimation and under estimation", with respect to estimates obtained through method of crop-cutting experiments, behave in similar way in all the domains of a block of interest; which is not realistic. Apart from this there are errors in the formulae of Bias and Mean Square Errors [Eq. (6.2), (6.3) of the paper].

5. Estimation for small domains using satellite and survey data

National Agricultural Statistics Service (NAAS) of the U.S. Department of Agriculture has been a user of remote sensing data since 1950's when it began using mid-altitude aerial photography to construct sampling frames for the 48 states of the continental United States. A new era in remote sensing began in 1972 with the launch of the Landsat I earth-resource monitoring satellite. Four Landsats have been launched since 1972 with Landsat IV and V which are still in operation. A regression estimator was developed which related the ground-gathered area frame data to the computer classification of Landset MSS (multi-spectral scanner images). The basic regression approach used to produce state estimates does not produce reliable county (small area) estimates. Three domain indirect regression estimators have been used or considered for producing small area county estimates using ancillary satellite data by NAAS. From 1972 to 1982 the Huddleston-Ray estimator was used, from 1983 to 1990 Battese-Fuller family of estimators was used and since 1991, the Battese-Fuller model has been used to produce country estimates with Landsat TM (Thematic Mapper) data. The details of these models have been discussed in detail by Bellow et al. (1996).

In India, as mentioned above, at present crop area statistics are based on complete enumeration of all fields and crop yield statistics based on GCES. With the advent of remote sensing technology satellite data has been widely used by many countries including India for obtaining various crop statistics. Several studies have been conducted during the past decade by the India's Department of Space under the Crop Acreage and Production Estimation (CAPE) project for crop acreage and production estimation for various major crops using satellite spectral data. Recently some studies have been taken at the Indian Agricultural Statistics Research Institute, New Delhi, India to develop more efficient estimator of crop yield using satellite data along with survey data of crop yield based on crop-cutting experiments. [cf. Singh et al. (1992), Goel et al. (1994), Shaible and

Casady (1994), Singh et al. (1999), Singh and Goel (2000), Singh (2004), Srivastava (2007)].

Singh and Geol (2000) used synthetic estimators to provide the yield estimates at Tehsil and Block levels, using crop yield data for Rabi crops 1997-98 obtained from GCES and the satellite spectral data of IRS-1D LISS-III. The study shows that the standard error of synthetic estimator is less than the corresponding direct estimator. The study also developed yield estimates at District level, using the direct estimator under post-stratification. The standard error of the direct estimator at district level is very small (around 5%). This confirms the results of earlier study by Singh et al. (1999).

Singh (2003) used the farmer's eye estimate of crop yield corresponding to the crop plots selected for crop-cutting experiment as an auxiliary variable along with the vegetation indices for improving the crop yield forecasting models. The yield data pertains to wheat crop yield data for district Rohtak for the year 1995-96 based on crop-cutting experiments. Spectral data in the form of vegetation indices RVI and NDVI has been obtained from IRS 1B-LISS II dated February 17, 1996 for the region. The farmer's eye estimate is obtained from the selected farmers for the fields in which crop-cutting experiments were conducted. Singh (2004) reviewing the earlier work also developed regression estimates using RVI (x_1), NDVI (x_2) and farmers eye estimate of crop yield of the corresponding plot (x_3) as auxiliary variables for forecasting cop yield at district level.

In all the above studies the performance of the synthetic estimators are measured in terms of standard errors. However, ignoring the bias remains a serious limitation.

In country like India almost 70% population is dependent on agriculture. The farm sizes in India are very small with diversified crops in each season. The practice of mixed cropping is quite dominant. Therefore, it may not be possible to prepare accurate area frame using remote sensing technology due to limitations of satellite sensor in detecting and differentiating small fields and crops grown, both for major as well as for minor domains.

Rao, J. N. K. (2004) provides an appraisal of indirect estimates, both traditional and model based. He provides a brief account of small area estimation in the context of agriculture surveys. He presents model based small area estimation under a basic area level model and a basic unit level model. He reviews work of Fuller (1981), Battese et al. (1988), Stasny et al. (1991) and Singh and Goel (2000).

Fuller (1981) applies the mixed area level model

$$\hat{\theta}_i = z_i^T \beta + v_i + e_i , \quad \text{i=1,2,...,m} \tag{5.1}$$

to estimate mean soybean hectares per segment in 1978 at the county level.

This model is combination of a **basic area level model**

$$\hat{\theta}_i = \theta_i + e_i , \qquad i=1,2,...,m$$

and a **linear regression model**

$$\theta_i = z_i^T \beta + v_i , \qquad i = 1,2,...m$$

where, sampling error e_i' s are assumed to be independent across area with mean 0 and known variance ψ_i, and

model error $v_i^{'}$s are assumed to be independent and identically distributed with mean 0 and variance σ_v^2, $z_i = (z_{1i},..., z_{Pi})^T$ area specific auxiliary variates.

Using the data $\left\{\left(\hat{\theta}_i, z_i\right), i = 1,..., m\right\}$ we can obtain estimates, θ_i^*, of the realized values of θ_i from the mixed model.

It may be noted that empirical best linear unbiased prediction (EBLUP) method is applicable for mixed linear models and its estimates do not require normality assumption on the random errors v_i and e_i. EBLUP estimate of θ_i is a composite estimate of the form

$$\theta_i^* = \hat{w}_i \hat{\theta}_i + \left(1 - \hat{w}_i\right) z_i^T \hat{\beta} ; \; \hat{w}_i = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \psi_i} \qquad (5.2)$$

which is a weighted combination of direct estimate $\hat{\theta}_i$ and a regression synthetic estimate $z_i^T \hat{\beta}$.

$\hat{\beta}$ is the weighted least square estimate of β with weights $\left(\hat{\sigma}_v^2 + \psi_i\right)^{-1}$.

$\hat{\sigma}_v^2$ is an estimate of the variance component σ_v^2.

Fuller obtained model based estimates of the population means, $\overline{Y}_i$ for the sampled county (m=10) as well as the non sampled counties. His model is given by

$$\bar{y}_i - z_{3i} = \beta_0 + \beta_1 z_{2i} + \beta_2 z_{3i} + v_i + e_i \qquad (5.3)$$

with known error variance σ_v^2 and σ_e^2.

z_{2i} = mean number of pixels of soybeans per area segment ascertained by satellite imaginary.

z_{3i} = mean soybean hectares from the 1974 U.S. Agricultural Census, as county (area) level covariates.

Note that z_{2i} and z_{3i} are known for all the 16 counties.

The model (5.3) is a special case of (5.2) with $\hat{\theta}_i = \bar{y}_i - z_{3i}$ and $\psi_i = \psi = \sigma_e^2$. Fuller's estimate of $\bar{Y}_i$ for sampled counties is obtained from (5.2.2) as

$$\bar{y}_{iF}^* = g^{-1}\left(\theta_i^*\right) = \theta_i^* + z_{3i}$$

$$= z_{3i} + z_i^T \hat{\beta} + w\left(\bar{y}_i - z_{3i} - z_i^T \hat{\beta}\right), \quad i \in s$$

where, $$w = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2}$$

For the non sampled counties, $\bar{y}_{iF}^* = z_{3i} + z_i^T \hat{\beta}, \quad i \notin s$

He concludes that the model based estimates, $\bar{y}_{iF}^*$, outperform in term of total MSE. They are also better than the direct estimates $\bar{y}_i$ in terms of total MSE for the sampled counties.

Battese et al. (1988) also consider the problem of crop acreage estimation using farm interview data in conjunction with LANDSAT satellite data. The authors use the nested error linear regression model $y_{ij} = x_{ij}^T \beta + v_i + e_{ij}$; j = 1,2,..., N_i ; i = 1,...,m to estimate area under corn and area under soybeans for i-th small area (counties) in north-central Iowa.

y_{ij} is variable under study related to unit-specific auxiliary data $x_{ij} = \left(1, x_{1ij}, x_{2ij}\right)^T$ and normally distributed errors v_i and e_{ij}.

Authors present the EBLUP estimates of small area means for both crops. Estimated standard errors of the EBLUP estimates and the survey regression estimates $\left[\bar{y}_i + \left(\bar{X}_i - \bar{x}_i\right)^T \hat{\beta}\right]$ are also given. The ratio of the estimated standard error of the EBLUP estimate to that of the survey regression estimate decreases as the size of sample decreases.

Rao (2004) further uses Hierarchical Bayes approach to test the fitness of the model given by Battese et al. with auxiliary data x_{ij}. Under the criterion of posterior probabilities use, it is noted that for values of such probabilities close to 0.5 it indicated good fit but for probabilities close to 0 and 1, it suggests poor fit of the model.

The major problem with the model based approach is of selection a suitable model. Therefore, selection and validation play a vital role in model based estimation. If the assumed models do not provide a good fit to the sample data, the model based approach can lead to erroneous estimates.

Acknowledgement

The research was supported by the grant received from Department of Science and Technology, New Delhi under Women Scientist Scheme.

The authors are thankful to the Referee for his constructive suggestions which helped to improve the paper.

REFERENCES

BATTESE, G.E., HARTER, R.M. and FULLER, W.A. (1988): An error-components model for prediction of crop areas using survey and satellite data. J. Amer. Statist. Assoc., 83, 28-36.

BELLOW, M., GRAHAM, M. and IWIG, W. C. (1996): Country estimation of crop acreage using satellite data. Indirect Estimation in U.S. Fedral Programs, Springer, 104-123.

CSO (1990): Working group on small area development programme statistics. Report, Department of Statistics, Ministry of Planning, Government of India, New Delhi.

DADHWAL, V.K. and PARIHAR, J.S. (1985): Estimation of 1983-84 wheat acreage of Karnal district (Haryana) using landsat MSS digital data. Technical note, IRS-UP/SAC/CPF/TN/09, Space Application Centre, Ahmedabad.

FRANCISCO, J.G. (1998): Comments on the invited paper entitled “Small area estimation in India – Crop yield and acreage statistics” by Tikkiwal, B.D. and Tikkiwal, G.C. Proceedings of the International Conference “Agricultural Statistics – 2000”, 238 – 254.

FULLER, W.A. (1981): Regression estimation for small areas. Rural America in Passage: Statistics for Policy, National Academy Press, Washington, D.C., 572-586.

GHOSH, M. and RAO, J.N.K. (1994): Small area estimation: An appraisal. Statist. Sci., 9, 55-93.

MAHALANOBIS, P.C. (1946): Sample surveys of crop yields in India. Sankhya, 269-280.

RAO, J.N.K. (2004): Small area estimation with applications to agriculture. J. Ind. Soc. Agril. Statist., 57, 159-170.

SCHAIBLE, W.L. and CASADY, R.J. (1994): The development, application, and evaluation of small area estimators. Statistics in Transition, 1 (6), 727-746.

SHARMA, S.D., SRIVASTAVA, A.K. and SUD, U.C. (2004): Small area crop estimation methodology for crop yield estimates at Gram Panchayat level. J. Ind. Soc. Agril. Statist., 57, 26-37.

SINGH, D. (1968): Double sampling and its application in agriculture. J. Ind. Soc. Agril. Statist. (Panse Memorial volume).

SINGH, R. (2003): Use of satellite data and farmers eye estimate for crop yield modeling. J. Ind. Soc. Agril. Statist., 56 (2), 166-176.

SINGH, R. (2004): Application of remote sensing technology for crop yield estimation. J. Ind. Soc. Agril. Statist, 57, 226-246.

SINGH, R. and GOEL, R.C. (2000): Use of remote sensing satellite data in crop surveys. Technical Report, Indian Agricultural Statistics Research Institute, New Delhi.

SINGH, R., GOEL, R.C., PANDEY, L.M. and SAHA, S.K. (1999): Use of remote sensing technology in crop yield estimation surveys- II, Project Report, IASRI, New Delhi.

SINGH, R., GOYAL, R.C., SAHA, S.K. and CHHIKARA, R.S. (1992): Use of satellite spectral data in crop yield estimation surveys. Int. J. Rem. Sens., 13 (14), 2583-2592.

SISODIA, B.V.S. and SINGH, A. (2001): On small area estimation – An empirical Study. J. Ind. Soc. Agril. Statist., 54 (3), 303-316.

SRIVASTAVA, A.K. (2007): Small area estimation – A perspective and some applications. J. Ind. Soc. Agril. Statist., 61 (3), 295-309.

STASNY, E.A., GOEL, P.K. and RAMSEY, D.J. (1991): County estimates of wheat production. Survey Methodology, 17, 211-255.

SUKHATME, P.V. and AGARWAL, O.P. (1946-47, 1947-48): Crop-cutting survey on wheat by the random sampling method. Report, Indian Council of Agricultural Research, New Delhi.

TIKKIWAL, B.D. (1991): Modeling through survey data for small domains. Keynote Address at the Symposium on Modeling held at Kurukshetra University, March 7 – 9, 1991.

TIKKIWAL, B.D. (1993): Modeling through survey data for small domains. Proc. Scientific Conference on Small Area Statistics and Survey Design held in September 1992 at Warsaw, Poland.

TIKKIWAL, B.D. and TIKKIWAL, G.C. (1998): Small area estimation in India – Crop yield and acreage statistics. Invited paper in the Proceedings of the International Conference "Agricultural Statistics – 2000" (along with the comments of Dr. J.G. Francisco, the discussant of the paper), 238 – 254. The Conference was held in March 1998 at Washington, D.C., USA.

TIKKIWAL, G.C. and GHIYA, A. (2000): A generalized class of synthetic estimators with application to crop acreage estimation for small domains. Biometrical Journal, 42 (7), 865-876.

TIKKIWAL, G.C. and GHIYA, A. (2004): A generalized class of composite estimators with application to crop acreage estimation for small domains. Statistics in Transition, 1(6), 697-711.