Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 74(3) 2020 217–226

An Alternative Sampling Methodology for Estimation of Cotton Yield using Double Sampling Approach

Tauqueer Ahmad, U.C. Sud, Anil Rai and Prachi Misra Sahoo

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 13 February 2020; Revised 23 September 2020; Accepted 11 October 2020

SUMMARY

Cotton, a multiple picking crop, is grown in around nine States in India. The existing procedure of estimation of average yield of cotton is based on crop cutting experiment(CCE) approach, which utilizes data on all pickings, is cumbersome and cost prohibitive. The double sampling approach can be gainfully employed in this case by collecting data on picking which has highest correlation with the total pickings yield on a larger sample and the total pickings yield data on a smaller sample. Accordingly, a stratified two-stage two-phase sampling design has been proposed for selection of representative sample and an appropriate estimation procedure, based on double sampling regression estimator, has been developed for estimation of average yield of cotton at district level. Utilizing the data of survey conducted in the Aurangabad and Amravati district of Maharashtra State and Adilabad and Guntur district of Andhra Pradesh wherein third picking data was collected on a larger sample and total pickings yield data on a smaller sample. An expression for optimum number of villages for larger and smaller samples have been obtained by minimizing cost subject to fixed percentage standard error of the estimates. These have been worked out empirically as well.

Keywords: Yield, Cotton, Multiple picking, Sampling methodology, Double sampling, Percentage standard error, Crop Cutting Experiment (CCE).

1. INTRODUCTION

Cotton is an important fibre yielding crop of global importance. It is also known as the "White Gold" or the "King of Fibres". It is grown, world over, in tropical and subtropical regions of more than 80 countries. It is an internationally sensitive commodity. Occupying an area of around 35.72 million ha, cotton is cultivated mainly in U.S.A. (11%), China (15%) and India (34%). Current estimates of world production are about 27.25 million tonnes annually, accounting for around 2.5% of the world's arable land. China is the world's largest producer of cotton, but most of this is used domestically. India occupies highest area under cotton (12.18mha) and is the second largest country, after China, as far as production is concerned (35.20 mt). In India, the States of Gujarat (34.09%), Maharashtra (20.45%), and Andhra Pradesh (13.92%) are the leading cotton producing States having a predominantly tropical wet and dry climate.

The yield estimation of cotton in India is carried out using crop cutting experiment technique. A limitation of the crop cutting experiment technique is that it requires observation on yield data of all the pickings of sampled fields. In case, data of one or two pickings of some of the selected fields are not available, not only precision of the yield estimate will suffer but estimate of average yield will be biased. Moreover, it is very difficult to capture data for all the pickings (upto 8 pickings in some states) creating practical difficulties in implementation of the procedure. Practical feasibility of crop cutting experiment technique for multiple picking crops like cotton is a matter of concern in India. Another important consideration is cost aspect of crop yield estimation survey. The cost on conduct of crop cutting experiment involving multiple picking crops increases manifold when data are to be collected on all the pickings. Therefore, there is a need to develop an alternative sampling methodology which is easier to implement as well as capable of providing unbiased

Corresponding author: Prachi Misra Sahoo
E-mail address: prachi.iasri@gmail.com

estimates of average yield of cotton with desired precision.

2. EXISTING PROCEDURE FOR ESTIMATION OF AVERAGE YIELD OF COTTON

The existing estimation procedure followed, under the General Crop Estimation Surveys (GCES), for estimation of average yield of Cotton is as under:

At the stratum level, an estimator of the average yield of the crop is obtained as a simple arithmetic mean of plot yields (net) within a stratum.

Let

y_{hij}= The plot yield (net) in gms/plot of the j^{th} plot in the i^{th} village of h^{th} stratum

n_{hi}= Number of CCEs conducted in the i^{th} village of h^{th}stratum

m_h= Number of villages in which CCEs have been conducted in the h^{th}stratum

n_h= Number of CCEs conducted in the h^{th}stratum

L = Number of strata in a district

a_h= The area (net) under the crop in the h^{th}stratum

d = The driage ratio

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

The estimator of average green yield for the h^{th} stratum is obtained as

$$\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{m_h}\sum_{j=1}^{n_{hi}} y_{hij}$$

and the estimator of the district level average yield of the dry marketable produce per hectare is given by

$$\hat{\hat{Y}} = d.f.\frac{\sum_{h=1}^{L} a_h \bar{y}_h}{\sum_{h=1}^{L} a_h}$$

An estimator of the sampling variance of $\hat{\hat{Y}}$ is obtained as

$$\hat{V}(\hat{\hat{Y}}) = \frac{d^2.f^2\left[W\sum_{h=1}^{L}\frac{a_h^2}{n_h} + (B-W)\sum_{h=1}^{L}\frac{a_h^2\sum_{i=1}^{m_h} n_{hi}^2}{\lambda_h n_h^2}\right]}{\left[\sum_{h=1}^{L} a_h\right]^2}$$

where

$$\lambda_h = \frac{n_h^2 - \sum_{i=1}^{m_h} n_{hi}^2}{n_h - (m_h - 1)},$$

$$B = \frac{\sum_{i=1}^{m_h}\frac{\left(\sum_{j=1}^{n_{hi}} y_{hij}\right)^2}{n_{hi}} - \frac{\left(\sum_{i=1}^{m_h}\sum_{j=1}^{n_{hi}} y_{hij}\right)^2}{n_h}}{\sum_{h=1}^{L}(m_h - 1)},$$

the mean square between villages and

$$W = \frac{\sum_{h=1}^{L}\left[\sum_{i=1}^{m_h}\sum_{j=1}^{n_{hi}} y_{hij}^2 - \sum_{i=1}^{m_h}\frac{\left(\sum_{j=1}^{n_{hi}} y_{hij}\right)^2}{n_{hi}}\right]}{\sum_{h=1}^{L}(n_h - m_h)},$$

the mean square within villages.

An estimator of percentage standard error of $\hat{\hat{Y}}$ is given by

$$\% S\hat{E}(\hat{\hat{Y}}) = \frac{\sqrt{\hat{V}(\hat{\hat{Y}})}}{(\hat{\hat{Y}})} \times 100$$

3. SAMPLING DESIGN USED FOR SELECTION OF SAMPLE

Stratified multistage random sampling design is generally adopted for carrying out General Crop Estimation Surveys (GCES) with taluks/revenue inspector circles as strata, revenue villages within a stratum as a first stage units of sampling, survey number/field within each selected village as a second stage sampling unit and experimental plot of a specified shape and size as the ultimate stage sampling unit.

Sampling design for GCES

Stratum – Taluk

↓

First stage sampling unit – Revenue Village

↓

Second stage sampling unit – Survey Number/Field

↓

Ultimate stage sampling unit – Experimental Plot

(Specified shape and size)

Total number of experiments at district level are finalized as per the following norms:

1. 130 to 150 CCEs are planned at district level (65 to 75 villages having two CCEs per village) if area under experimental crop in a district is above 80,000 hectares.
2. 100 to 120 CCEs are planned at district level (50 to 60 villages having 2 CCEs per village) if area under experimental crop in a district is between 20,000 to 80,000 hectares.
3. 40 to 80 CCEs are planned at district level (20 to 40 villages having two CCEs per village) if area under experimental crop in a district is below 20,000 hectares.

The number of villages allotted to a district is allocated among the strata, within a district, roughly in proportion to the area under the crop in different stratum. Generally, Crop insurance scheme is implemented at taluk/circle level in India. To determine compensation to be paid to the farmer in the event of crop loss, average yields are calculated on 18/12 crop cutting experiments at taluk/circle level respectively. The number of villages is increased up to 9/6, if the allocated number of villages to the strata within a district are less than 9/6 at taluk/circle level respectively. Hence, due to implementation of crop insurance scheme, number of CCEs planned at district level increased up to 700.

In each taluk/circle, in Maharashtra, a number of villages are selected every year for CCEs as per above norms at the time of district level-training programs conducted every year during Kharif, Rabi and Summer seasons. The list containing the names of the selected villages are made available by State Statistical Authority, Department of Agriculture, Maharashtra State, Pune to the concerned officers.

In each selected village, two fields growing the particular crop are selected according to the random sampling method. For the above purpose, two survey numbers are first selected out of all survey numbers for experimental crop in the village according to the said method.

In each selected survey number, where there are more than one field sown with the crop, one such field nearest to the South-West corner of the selected survey number should be selected. In each field thus selected, one plot of specified size is to be located at random.

4. PROPOSED PROCEDURE OF ESTIMATION OF CROP YIELD OF COTTON

A detailed examination of the data pertaining to five districts of Andhra Pradesh (A.P.) and five districts of Maharashtra showed that the yield of cotton at different pickings were highly correlated with the total yield. Hence, it may not be statistically proper to collect data in respect of all the pickings from the same sample of fields since the additional information gained would be only marginal. Therefore, it would be desirable to examine the possibility of using other sampling design such as double sampling which may not only be more efficient but also operationally more convenient resulting in more reliable data for estimation of crop yield of multiple picking crop like cotton.

Table 4.1. Picking-wise correlation with total yield in Maharashtra for the year 2006-07

District	Pick-1	Pick-2	Pick-3	Pick-4	Pick-5	Pick-6	Pick-7	Pick-8
Aurangabad	0.55	0.65	0.82	0.84	0.83	0.74	0.51	0.24
Jalgaon	0.44	0.73	0.82	0.86	0.83	0.70	0.49	-0.40
Jalna	0.65	0.77	0.86	0.86	0.77	0.73	0.60	0.77
Buldhana	0.66	0.69	0.82	0.81	0.72	0.61	0.44	0.30
Yavatmal	0.58	0.76	0.87	0.91	0.86	0.75	0.51	0.35

Table 4.2. Picking-wise correlation with total yield in Andhra Pradesh (A.P.) for the year 2005-06

District	Pick-1	Pick-2	Pick-3	Pick-4	Pick-5	Pick-6	Pick-7
Adilabad	0.59	0.61	0.82	0.82	0.81	0.48	0.16
Guntur	0.30	0.60	0.86	0.81	0.71	0.50	0.30
Khammam	0.65	0.81	0.87	0.56	0.66	0.37	-0.24
Karimnagar	0.65	0.80	0.80	0.45	0.43	0.41	-
Warangal	0.70	0.81	0.81	0.23	0.27	0.26	0.20

Results of both the Tables 4.1 and Table 4.2 reveal that 3rd and 2nd pickings in A.P. and 4th, 3rd, 5th and 6th pickings in Maharashtra have high correlation with total yield. It is, therefore, desirable to use the yield of picking having highest correlation with total yield as auxiliary variable. Thus, there is an opportunity to use double sampling approach in this situation.

4.1 Estimation procedure using double sampling approach under stratified two stage random sampling design framework

The estimation procedure for estimation of average yield of cotton using double sampling approach under stratified two stage random sampling design framework is as under:

Let

L = number of strata (mandals/taluks) in a district

N_h = total number of fsu's (villages) in h-th stratum (h = 1, 2, …, L)

n_h' = number of fsu's selected randomly in h-th stratum for observing yield for the p-th picking (p = 1, 2, …, P)

n_h = size of sub-sample selected randomly in h-th stratum for observing yield for the remaining pickings

m' = number of ssu's (fields) selected for observing yield for the p-th picking in i-th village of h-th stratum (i = 1, 2, …, n_h')

m = size of sub-sample for observing yield for remaining pickings of these m' ssu's

$y_{hij}(p)$ = yield of cotton in j-th field of i-th village (j = 1, 2, …, m') in h-th stratum corresponding to p-th picking.

An estimator of average yield corresponding to p-th picking, $\bar{y}_{nm}(p)$ for a district is given by

$$\bar{y}_{nm}(p) = \frac{1}{n_0 m}\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m} y_{hij}(p)$$

where $n_0 = \sum_{h=1}^{L} n_h$

An estimator of average yield, $\bar{y}_{nm}$ for a district is given by

$$\bar{y}_{nm} = \sum_{p=1}^{P} \bar{y}_{nm}(p)$$

This is a biased estimate of the population average discussed by Panse *et al.* (1966). However, investigations by Sukhatme and Panse (1951) into the magnitude of the bias from the data collected in district crop-cutting surveys has shown that the bias is negligible. A double sampling regression estimator of average yield of cotton under the proposed framework can be written as

$$\bar{y}_{ld}(2) = \bar{y}_{nm} + \hat{\beta}[\bar{y}_{n'm'}(p) - \bar{y}_{nm}(p)]$$

where

$$\bar{y}_{n'm'}(p) = \frac{1}{n_0' m'}\sum_{h=1}^{L}\sum_{i=1}^{n_h'}\sum_{j=1}^{m'} y_{hij}(p),$$

$$n_0' = \sum_{h=1}^{L} n_h' \quad \hat{\beta} = \frac{\left(\frac{1}{n_0} - \frac{1}{n_0'}\right)s_{by(p)y} + \frac{1}{n_0'}\left(\frac{1}{m} - \frac{1}{m'}\right)s_{wy(p)y}}{\left(\frac{1}{n_0} - \frac{1}{n_0'}\right)s^2_{by(p)} + \frac{1}{n_0'}\left(\frac{1}{m} - \frac{1}{m'}\right)s^2_{wy(p)}}$$

where

$$s_{by(p)y} = \frac{1}{L}\sum_{h=1}^{L}\frac{1}{(n_h - 1)}\sum_{i=1}^{n_h}(\bar{y}_{hi.(p)} - \bar{y}_{h..(p)})(\bar{y}_{hi.} - \bar{y}_{h..})$$

$$s_{wy(p)y} = \frac{1}{n_o(m-1)}\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m}(y_{hij(p)} - \bar{y}_{hi.(p)})(y_{hij} - \bar{y}_{hi.})$$

where

$$\bar{y}_{hi.}(p) = \frac{1}{m}\sum_{j=1}^{m} y_{hij}(p) \text{ and } \bar{y}_{h..}(p) = \frac{1}{n_h}\sum_{i=1}^{n_h}\bar{y}_{hi.}(p),$$

$$s^2_{by(p)} = \frac{1}{L}\sum_{h=1}^{L}\frac{1}{(n_h-1)}\sum_{i=1}^{n_h}(\bar{y}_{hi.}(p)-\bar{y}_{h..}(p))^2$$

$$s^2_{wy(p)} = \frac{1}{n_O(m-1)}\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m}(y_{hij}(p)-\bar{y}_{hi.}(p))^2$$

An estimator of $MSE(\bar{y}_{ld}(2))$ is given as

$$\hat{MSE}(\bar{y}_{ld}(2)) = \left\{\left(\frac{1}{n_0}-\frac{1}{N_0}\right)s^2_{by}+\frac{1}{N_0 m}s^2_{wy}\right\}(1-r^{*2})+ r^{*2}\left\{\left(\frac{1}{n'_0}-\frac{1}{N_0}\right)s^2_{by}+\left[\frac{1}{N_0 m}-\frac{1}{n'_0}\left(\frac{1}{m}-\frac{1}{m'}\right)s^2_{wy}\right]\right\}$$

where $N_0 = L\bar{N}$, and $\bar{N}$ is the harmonic mean of N_{hs}. This is the usual estimator in sub-sampling design.

$$s^2_{by} = \frac{1}{L}\sum_{h=1}^{L}\frac{1}{(n_h-1)}\sum_{i=1}^{n_h}(\bar{y}_{hi.}-\bar{y}_{h..})^2$$

$$s^2_{wy} = \frac{1}{n_0(m-1)}\sum_{h=1}^{L}\sum_{i=1}^{n_h}\sum_{j=1}^{m}(y_{hij}-\bar{y}_{hi.})^2$$

$$r^{*2} = \frac{q^2_{y(p)y}}{q_{y(p)y(p)}q_{yy}}$$

with

$$q_{y(p)y} = \left(\frac{1}{n_0}-\frac{1}{n'_0}\right)s_{by(p)y}+\frac{1}{n_0}\left(\frac{1}{m}-\frac{1}{m'}\right)s_{wy(p)y}$$

$$q_{y(p)y(p)} = \left(\frac{1}{n_0}-\frac{1}{n'_0}\right)s^2_{by(p)}+\frac{1}{n_0}\left(\frac{1}{m}-\frac{1}{m'}\right)s^2_{wy(p)}$$

$$q_{yy} = \left(\frac{1}{n_0}-\frac{1}{n'_0}\right)s^2_{by}+\frac{1}{n_0}\left(\frac{1}{m}-\frac{1}{m'}\right)s^2_{wy}$$

4.2 Validation of proposed methodology for estimation of average yield of cotton

For the validation of proposed methodology, surveys were planned for data collection in Amravati and Aurangabad districts of Maharashtra State while in Andhra Pradesh (A.P.) State, planning was done in Adilabad and Guntur districts. In Maharashtra State, two additional fields were selected in 114 villages selected for CCE in Aurangabad district and 118 villages selected for CCE in Amravati district while in A.P., two additional fields were selected in 77 villages selected for CCE in Warangal district and 75 villages selected for CCE in Guntur district.

In addition to CCE data for the third picking from two additional fields from the selected CCE villages of the selected districts of both the States, the regular CCE data of selected districts of both the States was also required for validation of the proposed double sampling procedure under stratified two stage random sampling design framework.

The estimate of average yield of cotton along with % S.E. for the year 2012-13 was obtained using proposed double sampling approach under stratified two stage sampling design framework. In order to compare the results, the estimate along with % S.E. was also obtained using existing GCES procedure.

For implementation purpose, the selected CCE villages in the district have been treated as preliminary sample villages (n′) for the third picking yield and a sub-sample of 40% and 30% CCE villages (n) were selected by SRSWOR from the preliminary sample villages for the remaining pickings and out of the four fields (m′) for the third picking, in each of these sub-sample villages, two fields (m) were selected for observing yield pertaining to the remaining pickings from CCE plot. The estimate of average yield of cotton along with % S.E. was obtained for both the districts of Maharashtra State and both the districts of A.P. State. The results of the district-wise analysis are presented in the Tables 4.3 to 4.6.

Table 4.3. Estimates of average yield of cotton (Kg/ha) along with % S.E. using different approaches for Aurangabad district of Maharashtra State for the year 2012-13

Existing procedure			
No. of sampled villages	**Average yield (Kg/ha)**	**% S.E.**	**Total no. of pickings involved**
Total GCES villages (83)	202.027	5.49	1328
Double sampling regression procedure under stratified two stage sampling design framework			
n′=83 (for one picking) n=33 (40% of n′) (for remaining pickings)	189.711	5.32	794
n′=83 (for one picking) n=26 (30% of n′) (for remaining pickings)	188.966	6.39	696

A close perusal of the above Table 4.3 reveals that under the double sampling regression procedure involving stratified two stage sampling design framework, the estimates of average yield of cotton was obtained with less than 7% standard error for Aurangabad district which is fairly reliable. The estimate is almost at par with the estimate obtained using existing procedure. Further, as the sample size decreases *i.e.* from 40% to 30%, percentage standard error increases for the proposed procedure.

The alternative procedure is operationally more convenient than the existing GCES procedure and is expected to reduce the workload of the field staff significantly which in turn will lead to good quality CCE data from these limited number of plots.

Table 4.4. Estimates of average yield of cotton (Kg/ha) along with % S.E. using different approaches for Amravati district of Maharashtra State for the year 2012-13

Existing procedure			
No. of sampled villages	**Average yield (Kg/ha)**	**% S.E.**	**Total no. of pickings involved**
Total GCES villages (111)	637.055	2.81	1776
Double sampling regression procedure under stratified two stage sampling design framework			
n′=111 (for one picking) n=43 (40% of n′) (for remaining pickings)	529.168	2.85	1046
n′=111 (for one picking) n′=34 (30% of n’) (for remaining pickings)	511.038	2.92	920

It may be seen from the above Table 4.4 that in case of the proposed alternative procedure the estimate of average yield of cotton is obtained with less than 3% standard error for Amravati district which is fairly reliable. Further, as the sample size decreases *i.e.* from 40% to 30%, percentage standard error increases.

Table 4.5. Estimates of average yield of cotton (Kg/ha) along with % S.E. using different approaches for Guntur district of A.P. State for the year 2012-13

Existing procedure			
No. of sampled villages	**Average yield (Kg/ha)**	**% S.E.**	**Total no. of pickings involved**
Total GCES villages (65)	629.399	3.95	910
Double sampling regression procedure under stratified two stage sampling design framework			
n′=65 (for one picking) n=52 (for remaining pickings)	608.882	3.61	884

The analysis of data in respect of Guntur district reveals a similar trend as earlier *i.e.* the proposed estimator scores over the estimator currently used in terms of the criterion of percentage standard error. It may be observed from the Table 4.5 that the estimate of average yield of cotton in case of existing GCES procedure is obtained based on data on 65 villages. But double sampling regression procedure under stratified two stage sampling design framework requires CCE villages data for all the mandals in the district. In most of the mandals only two villages data was available and at least two villages data is necessary for the analysis. Hence, there was not much scope for sub-sampling and accordingly, the 40% and 30% sub-sampling of villages could not be done. The results are presented based on the maximum possible reduced number of villages *i.e.* 52 villages. The improved performance of double sampling approach based estimator is evident.

Table 4.6. Estimates of average yield of cotton (Kg/ha) along with % S.E. using different approaches for Warangal district of A.P. State for the year 2012-13

Existing procedure			
No. of sampled villages	**Average yield (Kg/ha)**	**% S.E.**	**Total no. of pickings involved**
Total GCES villages (59)	371.034	1.63	708
Double sampling regression procedure under stratified two stage sampling design framework			
n′=59 (for one picking) n=48 (for remaining pickings)	371.891	5.69	716

The results of analysis of data in Warangal district are bit different than earlier *i.e.* although the percentage standard errors are within permissible limits, the existing estimator scores over the proposed estimator in terms of the criterion of percentage standard error of the estimator. This highlights the need to compare the proposed estimator over the existing estimator using optimum sample sizes determined by minimizing expected cost for a fixed value of percentage standard error of the estimator.

5. ESTIMATION OF SAMPLE SIZE (NUMBER OF VILLAGES TO BE SELECTED IN A DISTRICT FOR THE SURVEY)

The guiding principle in the determination of optimum sample size is either to fix variance of the estimator and minimize cost or to fix cost and minimize variance of the estimator. We, in this paper, aimed at minimizing the cost by fixing the variance of the estimator. The optimum number of villages to be sampled in a district was determined by fixing the percentage standard to 5 and 7.

In order to estimate sample size in case of double sampling regression procedure, a simple cost function given by

$C = cn_0 m + c' n_0' m'$

was considered, where

c = costs involved in conducting crop cutting experiment for all the pickings of a selected plot (e.g. travel, equipment, harvesting and drying, stationery) and

c′ = costs involved in conducting crop cutting experiment for one picking (3rd picking) of a selected plot (e.g. travel, equipment, harvesting and drying, stationery)

= {1/(Total number of pickings)} x c

By minimizing the cost subject to a fixed variance V_0, the optimum values of n_0 were obtained from the following relations

$$\frac{n_0}{n_0'} = \sqrt{\frac{c'(1-r^{*2})}{cr^{*2}}} = \theta$$

and

$$n_0 = \left\{(1-r^{*2}) + \theta r^{*2}\right\}\left(s_{by}^2 + \frac{1}{m}s_{wy}^2\right)\Bigg/\left[\hat{V}_0 + \frac{s_{by}^2}{N_0}\right]$$

The proposed regression estimator, involving auxiliary information, was accordingly compared with the estimator which does not make use of auxiliary information on the basis of optimum sample sizes determined by minimizing the cost function subject to a fixed value of the percentage standard error (the optimum sample sizes were determined by fixing the percentage standard error to 5 and 7).

The estimated sample size with the desired percentage standard error of the district level estimates of average yield of cotton for Maharashtra and A.P. States are presented in the following Table 5.1.

The optimum sample size is maximum of the estimated sample size used for estimation of average yield of cotton in both the States.

A close perusal of Table 5.1 reveals that the district level average yield of cotton can be estimated with less than or equal to 5% standard error, if a preliminary sample of 137 villages are selected for third picking and a sub-sample of 48 villages are selected for the remaining pickings from each of the selected districts. Therefore, the recommended sample size *i.e.* number of preliminary sample villages and sub-sample villages to be selected from a district are 137 (one hundred thirty seven) and 48 (forty eight) respectively. Further, the percentage standard error of the double sampling regression estimator is directly related to "r" *i.e.* higher the "r" value smaller the percentage standard error of the estimator.

The analysis of results given in Table 5.2 reveals that at district level, the average yield of cotton can be estimated with less than or equal to 7% standard error, if a preliminary sample of 71 villages are selected for third picking and a sub-sample of 25 villages are selected for the remaining pickings from each of the selected districts. Therefore, the recommended sample size *i.e.* number of preliminary sample villages and sub-sample villages to be selected from a district are 71 (seventy-one) and 25 (twenty-five) respectively with minimum two sub-sample villages per mandal/ taluka of the district as per requirement of the proposed alternative sampling methodology. As earlier, higher

Table 5.1. Sample size (number of villages per district) for district level estimates of average yield of cotton for 5% standard error

State	Districts	Calculated S.E. (%)	No. of villages (n)	No. of villages (prel. Sample) (n′)	Desired S.E. (%)	Estimated sample size (n_0)	c′/c	r	Estimated sample size (n_0')
Double sampling regression procedure under stratified two stage sampling design framework									
Maharashtra	Aurangabad	6.39	26	83	5.00	48	0.125	0.7109	137
	Amravati	2.92	34	111		11	0.125	0.7867	30
Andhra Pradesh	Guntur	3.61	52	65		34	0.143	0.7681	102
	Warangal	5.69	48	59		45	0.167	0.7992	127
Recommended optimum sample size i.e maximum of estimated sample size						48			137

Table 5.2. Sample size (number of villages per district) for district level estimates of average yield of cotton for 7% standard error

State	Districts	Calculated S.E. (%)	No. of villages (n)	No. of villages (prel. Sample) (n′)	Desired S.E. (%)	Estimated sample size (n_0)	c′/c	r	Estimated sample size (n_0')
Double sampling regression procedure under stratified two stage sampling design framework									
Maharashtra	Aurangabad	6.39	26	83	7.00	25	0.125	0.7109	71
	Amravati	2.92	34	111		6	0.125	0.7867	17
Andhra Pradesh	Guntur	3.61	52	65		18	0.143	0.7681	55
	Warangal	5.69	48	59		25	0.167	0.7992	70
Recommended optimum sample size i.e maximum of estimated sample size						25			71

the r value, smaller the percentage standard error of the estimator.

The optimum values of sample size in the context of simple linear estimator (existing estimator) are as given in Table 5.3 and Table 5.4:

Table 5.3. Sample size (number of villages per district) for district level simple estimates (GCES estimates) of average yield of cotton for 5% standard error

State	Districts	Calculated S.E. (%)	No. of villages (n)	Desired S.E. (%)	Estimated sample size (n_0)
Maharashtra	Aurangabad	5.49	83	5.00	100
	Amravati	2.81	111		35
Andhra Pradesh	Guntur	3.95	65		41
	Warangal	1.63	59		6
Recommended optimum sample size *i.e.* maximum of estimated sample size					100

Table 5.4. Sample size (number of villages per district) for district level simple estimates (GCES estimates) of average yield of cotton for 7% standard error

State	Districts	Calculated S.E. (%)	No. of villages (n)	Desired S.E. (%)	Estimated sample size (n_0)
Maharashtra	Aurangabad	5.49	83	7.00	51
	Amravati	2.81	111		18
Andhra Pradesh	Guntur	3.95	65		21
	Warangal	1.63	59		3
Recommended optimum sample size *i.e.* maximum of estimated sample size					51

Using the optimum values of sample sizes, the percentage reduction in cost by using the double sampling regression estimator over an estimator which does not use auxiliary information are given in Table 5.5 and Table 5.6 by fixing the percentage standard error values to 5 and 7 percent respectively. Keeping in view the Indian scenario *i.e.* the manpower available, costs etc., the values of m and m′ are fixed at 2 and 4 respectively.

Table 5.5. Percentage reduction in cost using different approaches based on optimum sample size of villages required per district for obtaining district level estimates of average yield of cotton for Maharashtra and A.P. States for 5% standard error

Existing GCES procedure				
Optimum number of sample villages required per district		Average number of pickings based on four districts	Total number of pickings involved	Percentage reduction in cost
100		7	1400	
Double sampling regression procedure under stratified two stage sampling design framework				
Optimum no. of villages (preliminary sample) required per district	Optimum no. of villages (sub- sample) required per district	Average number of pickings based on four districts	Total number of pickings involved	Percentage reduction in cost
137	48	7	1124	19.71

Table 5.6. Percentage reduction in cost using different approaches based on optimum sample size of villages required per district for obtaining district level estimates of average yield of cotton for Maharashtra and A.P. States for 7% standard error

Existing GCES procedure				
Optimum number of sample villages required per district		Average number of pickings based on four districts	Total number of pickings involved	Percentage reduction in cost
51		7	714	
Double sampling regression procedure under stratified two stage sampling design framework				
Optimum no. of villages (preliminary sample) required per district	Optimum no. of villages (sub- sample) required per district	Average number of pickings based on four districts	Total number of pickings involved	Percentage reduction in cost
71	25	7	584	18.21

The analysis of results given in Table 5.5 and Table 5.6 reveal that there is considerable reduction in cost using the double sampling procedure as compared to an estimator which does not make use of information on auxiliary variable.

More and less the entire cotton area in the country has now been covered under Bt. Cotton that requires only 2-3 pickings for the Crop Cutting Experiments. Therefore, in order to determine which picking should be used as an auxiliary variable, the CCE data of cotton pertaining to two major cotton growing districts namely, Mansa and Muktsar Sahib of Punjab State and two districts namely, Hisar and Sirsa of Haryana State for the year 2012-13 were acquired. The data from both the States was available for four (4) pickings only. Picking-wise correlation with the total yield for the CCE data for all the four selected districts, two each Punjab and Haryana States, was obtained and are presented in the Tables 5.7 and Table 5.8.

Table 5.7. Picking-wise correlation with total yield for plots in the selected districts of Punjab State for Desi and American Cotton (Combined) for the year 2012-13

State	Districts	Pick-1	Pick-2	Pick-3	Pick-4
Punjab	Mansa	0.652751	0.753949	0.80976	0.142816
	Muktsar Sahib	0.769401	0.731496	0.79637	0.364979

Table 5.8. Picking-wise correlation with total yield for plots in the selected districts of Haryana State for the year 2012-13

State	Districts	Pick-1	Pick-2	Pick-3	Pick-4
Haryana	Hisar	0.467793	0.821971	0.651065	0.305694
	Sirsa	0.668537	0.679208	0.350899	0.004528

It can be inferred from the Table 5.7 and Table 5.8 that 3rd picking has the highest correlation with total yield in both the districts of Punjab State while in Haryana State, 2nd picking has the highest correlation with total yield in both the districts. Thus, **3rd picking data is recommended for use as auxiliary variable in Punjab State while 2nd picking data is recommended for use in Haryana State** for estimation of average yield of cotton using double sampling regression estimator.

The CCE data of other cotton growing States were analyzed to determine the picking having highest correlation with total yield. Accordingly, CCE data on cotton pertaining to two districts of each of the five States namely, Gujarat, Karnataka, Madhya Pradesh, Rajasthan and Tamil Nadu for the year 2012-13 was acquired. Picking-wise correlation with the total yield for the irrigated and unirrigated plots (combined) was obtained and results are presented in the Table 5.9.

The analysis reveals that 2nd picking has the highest correlation with total yield in both the districts of Gujarat and Karnataka States while in Madhya Pradesh State, 3rd picking has the highest correlation with total yield in both the districts. Further, in Rajasthan State, 2nd picking has the highest correlation with total yield in Ganganagar district and 1st picking has the highest correlation with total yield in Hanumangarh district followed by 2nd picking. Since the correlation coefficient for 2nd picking is significant in Hanumangarh district, **therefore, 2nd picking is recommended to be used as auxiliary variablein Rajasthan State for the sake of uniformity.**

Similarly, in Tamil Nadu State, the correlation coefficient for 3rd picking is significant. Therefore, 3rd

Table 5.9. Picking-wise correlation with total yield for irrigated and unirrigated plots (combined) in the selected districts of five States for the year 2012-13

State	District	Pick-1	Pick-2	Pick-3	Pick-4	Pick-5	Pick-6	Pick-7	Pick-8
Gujarat	Patan	0.6444	0.8942	0.9081*	0.8938*	0.7601*	0.4618*	0.2088*	**
	Mehsana	0.3514	0.6915	0.7967*	0.7423*	0.4863*	0.3832*	0.0559*	
Karnataka	Yadgir	0.6498	0.8384	0.8020*	0.2960*	0.0706*	**	**	**
	Haveri	0.6581	0.7973	0.8894*	0.8517*	0.4383	0.4931*	**	**
Madhya Pradesh	Khrgon	0.5849	0.8195	0.8398	0.8368	0.7248*	0.0473*	0.0467*	0.0529*
	Dhar	0.5735	0.7948	0.8144	0.7048	0.5895*	0.5577*	**	**
Rajasthan	Ganganagar	0.6532	0.8647	0.6417*	0.7873*	**	**	**	**
	Hanumangarh	0.7626	0.6945	0.7225*	**	**	**	**	**
Tamil Nadu	Perambalur	0.4903	0.8081	0.8004	0.7257*	**	**	**	**
	Salem	0.3781	0.6307	0.8003	0.8301	0.7828	0.5595*	0.3492*	0.2481*

Note: *Insufficient data available
**Data not available

picking should be treated as auxiliary variable in Tamil Nadu State. The recommended picking, to be taken as auxiliary variable, in the nine States is given in Table 5.10:

Table 5.10. Recommended picking to be used as auxiliary variable for analysis of cotton data using developed methodology in nine major cotton growing States of India

S. No.	State	Picking to be used as auxiliary variable
1	Andhra Pradesh	3rd Picking
2	Gujarat	2nd Picking
3	Haryana	2nd Picking
4	Karnataka	2nd Picking
5	Madhya Pradesh	3rd Picking
6	Maharashtra	3rd Picking
7	Punjab	3rd Picking
8	Rajasthan	2nd Picking
9	Tamil Nadu	3rd Picking

6. CONCLUSIONS

The present study has revealed very encouraging results as shown from the analysis of results and demonstrated the feasibility of estimating cotton production with limited number of pickings using the double sampling approach based regression estimator. In view of the above, it is recommended that the alternative sampling methodology using double sampling regression procedure under stratified two stage sampling design framework may be adopted in all the cotton growing States of the country for estimation of average yield of cotton which will not only provide reliable estimates of average yield of cotton but will significantly reduce cost of the survey and will also be operationally more convenient than the GCES procedure. Further, the workload of the field staff will be significantly reduced which in turn will lead to good quality CCE data from limited number of plots.

REFERENCES

Agricultural Research Data Book (2012, 2013, 2014, 2015, 2016, 2017). IASRI publication.

Ahmad, T., Bathla, H.V.L., Rai, A., Sahoo, P.M., Gupta, A.K., Jain, V.K. and Mhadgut, D.V. (2009). Study to investigate the causes of variation between official and trade estimates of Cotton production. Project Report, IASRI publication.

Indian Cotton, A Profile (2010, 2011, 2012). The Cotton Corporation of India Ltd. publication.

Kajale, J. (2006). Study to investigate the causes of variation between official and trade estimates of Cotton production in Maharashtra. AERC Report., Gokhale Institute of Politics and Economics, Pune.

Panse, V.G., Rajagopalan, M. and Pillai, S.S. (1966). Estimation of crop yields for smaller areas. *Biometrics*, **22(2)**, 374-384.

Ryerson, R.A., Dobbins, R.N. and Thibault, C. (1985). Timely crop area estimates from Landsat. *Photo. Engg. Rem. Sens.*, **51**, 1735-1743.

Sukhatme, B.V. and Koshal, R.S. (1959). A contribution to double sampling. *J. Ind. Soc. Agril. Statist.*, **11**, 128-144.

Sukhatme, P.V. and Panse, V.G. (1951). Crop surveys in India II. *J. Ind. Soc. Agril. Statist.*, **3**, 97-168.

Watt, G. (1972). A Dictionary of the Economic Products of India. Cosmo Publication, New Delhi, **4**, 45-47.