Plant Archives Vol. 20, Special Issue (AIAAS-2020), 2020 pp. 538-541 ***ISSN 0972- 5210***

POST STRATIFIED CROP YIELD ESTIMATORS AT BLOCK LEVEL IN HISAR DISTRICT OF HARYANA (INDIA)

P K Muhammed Jaslam and Manoj Kumar
Department of Mathematics and Statistics, College of Basic Sciences and Humanities,
CCS Haryana Agricultural University, Hisar - 125004, Haryana

Abstract

General crop yield estimation surveys are conducted throughout the country for estimating crop yield of all major crops. The direct estimates at national as well state level are almost reliable, as the estimator's sampling error is within 5 percent, but not true at lower levels. However, demands for reliable small area statistics (district, sub district, village level) are increasing both from public and private sectors with growing concerns of governments relating to issues of distribution, equity and disparity. In the present study small area estimates at block level for the Hisar district, Haryana, India, for the period 2018-2019 were developed using crop yield data based on stratification with the use of auxiliary information's including satellite data.

Keywords: Small Area Estimation, Synthetic estimator, Composite estimator, NDVI and EVI.

Introduction

The subject 'Agricultural Statistics' primly focus on crop statistics such as crop area, production and yield and its influencing parameters. Since agriculture is a land based economic activity, the land use statistics has preeminence in the agricultural statistics. In India, crop area statistics are mainly collected by complete listing, while crop yields are measured on the basis of a sample survey method. Crop yield estimated in India on the basis of crop cutting experiments (CCEs) under the General Crop Estimation Survey (GCES). GCES were performed using survey methodology developed by Mahalanobis (1946); Sukhatme and Agarwal (1947). For yield estimates, more than 800,000 samples of CCEs are collected annually. Estimates of crop production or yield by CCE are recorded at district level and are collated at state and country level. Although this method can provide reliable estimates at district level, the method is time-consuming and time-consuming, as a result of which the CCE procedure is not followed in many places by an enumerator which leads to incorrect data reporting. In order to enhance the accuracy of the data obtained under the GCES, Department of Economics and Statistics, Ministry of Agriculture, the Government of India has initiated the 'Improvement of Crop Statistics (ICS)' program. Under this program, quality assurance of GCES' field service is conducted by overseeing about 30,000 CCEs by NSSO and supervisory officers of the State Government. The results of the ICS study indicate that CCEs are not typically carried out precisely resulting the lack of expected accuracy in the data. Given the limitations of resources and infrastructure, the GCES sample size needs to be significantly reduced so that the enumerator's work volume is minimized and, therefore, better supervision of the CCE 's operation can lead to improved data quality. But, reduction in sample size will directly impact the estimator's standard error. The reduced sample size is more alarming to produce estimates at lower administrative levels, because estimators may be unreliable based on sample data. In addition, Tikkiwal and Tikkiwal (1998); Tikkiwal and Ghiya (2000, 2004) reported that CCE direct estimates are almost accurate at both national and state levels, but are not valid at lower levels. However, the computation of crop production statistics for a small area, such as the Community Development Block, has received increasing interest in India in recent times. This is partly due to the fact that regional planning is mostly carried out on a local or regional basis and to the allocation of central funds.

The topic of the Small Areas Estimation (SAE) has become more important despite the increased need for micro-scale decision-making. The SAE techniques aim at producing reliable estimates for such small areas with small (or even no) sample sizes by borrowing strength from data of other small areas. Factors like different soil types, agricultural inputs and adoption of improved technology affect the crop yield and hence cause a lot of variability in the yield even within a stratum. Since the spectral reflectance is a manifestation of all important factors affecting the crop, a stratification of crop area on the basis of crop vigour and as reflected by the spectral data is expected to result in greater efficiency in the crop yield estimation. Singh *et al*. (1992) showed that the stratification based on Normalized Difference Vegetation Index (NDVI) at heading/ flowering stage improved the efficiency of crop yield estimation considerably.

*Corresponding author Email: jaslam.stat@hau.ac.in

Materials and Methods

The study area (Hisar district), located in northern part of India between 29.12° N, 75.81° E covers approximately 3,983 km². The net sown area of the district is 4040 Km^2 with 178.2 percent cropping intensity; of this agricultural land, Rabi wheat are key economic crops, with a total area of up to 2240 km^2. At present Hisar district consists of 9 sub-districts (blocks) *viz*., Adampur, Agroha, Barwala, Hisar-1, Hisar-2, Hansi-1, Hansi-2, Narnaund and Uklana.

Data description

Village wheat crop yield data for 2018-2019 were collected from the Department of Agriculture & FW, Hisar. The present study aimed to develop small area crop yield estimates for wheat at block level in Hisar district of Haryana using the crop yield survey data, crop area information and the satellite spectral data.

Data description

Let $N = \sum_{i=1}^{M} N_i$ be the number of CCE plots within M small areas (blocks). The aim is to obtain information on the population characteristics of M small areas, also known as subpopulations or domains (Sarndal *et al*., 2003). The variable of interest obtained from sampled population element $(j=1,\ldots,N_i)$ within domain i is denoted by y_{ij}

$$\bar{Y}_{M,i} = \bar{y}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} y_{ij}, \quad i = 1,\ldots,M, \; j = 1,\ldots,N_i$$

An obvious estimator of block level yield is mean of the N_i samples taken from domain i

$$\sigma_i^2 = \frac{1}{N_{i-1}}\sum_{j=1}^{N_i}(y_{ij} - \bar{y}_i)^2$$

Where the estimated variance is

It only uses information from the CCE plots within the domain under consideration. Since the variance only depends on the sampled elements within the small area, it can be unstable for small Ni. In fact, this variance (CV%) determines whether or not a domain is considered a small area. If the variance is larger than appropriate or assumed to be unstable, the domain is regarded as being small (Rao, 2003). This leaves the decision on whether an area is small to the user.

Stratification procedures

Let the population (a district) consist of i small areas (block). Further, let the district area be divided into V=3 post strata's based on yield value, NDVI value and EVI values (Figure-1). The NDVI is a simple graphical indicator that can be used to analyze remote sensing measurements, assessing whether or not the target being observed contains live green vegetation. NDVI is defined as the ratio of (IR R) to (IR+R). Healthy vegetation absorbs most of the visible light that hits it and reflects a large portion of the near-infrared light. Unhealthy or sparse vegetation reflects more visible light and less near-infrared light. The value range of the NDVI is -1 to 1. The Enhanced Vegetation Index (EVI) of the crop (most advanced RVI) is an optimized vegetation index designed to enhance the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmosphere influences.

Post Stratified Direct Estimator

Let y_{ij} and x_{ij} denote the character under study, the crop yield, and the auxiliary variable, the crop acreage, respectively. Let y_{ijv} denote the crop yield for the *i*-th block in the *j*-th village and *v*-th post-strata and let n_{jv} denote the number of sample observations belonging to the *i*-th block in the *v*-th post strata

$$\bar{y}_{di} = \sum_{v} W_{iv}\bar{y}_{iv}$$

$\bar{y}_{iv} = \frac{1}{n_{iv}}\sum_{j}^{n_{iv}} y_{ijv}$ is the average yield for the v-th post stratum in the ith block; ,

X_{iv} = crop acreage for the v-th stratum in the ith block;

X_{io} = is the crop acreage for the i-th block

The approximate estimate of variance of $\bar{y}_{di}$ can be written as

Where, $\hat{V}(\bar{y}_{iv}) = \frac{s_{iv}^2}{n_{iv}}$

$$s_{iv}^2 = \frac{1}{n_{iv-1}}\sum_{j}^{n_{jv}}(y_{ijv} - \bar{y}_{iv})^2$$

Synthetic estimator

The direct estimator is based on only the number of crop cutting experiments belonging to the *i*-th block in the post strata which is quite small and hence the estimator will not be very efficient. To improve the efficiency of the direct estimator a synthetic estimator is used which makes use of the information from the whole sample. "An unbiased estimate is obtained from a sample for a large area; when this estimate is used to derive estimates for subareas on the assumption that the small areas have the same characteristics as the larger area, we identify these estimates as synthetic estimates". This is the method of "borrowing information from related subareas in order to increase the effective sample size for estimation and hence the accuracy of the resulting estimates".

$$\bar{y}_{st} = \sum_{v} W'_{iv}\bar{y}_{ov}$$

Where, $W'_{iv} = \frac{X_{iv}}{X_{io}}$

$\bar{y}_{ov}$ is the average crop yield for the v-th post stratum

$$\bar{y}_{ov} = \sum_{i} W''_{iv}\bar{y}_{iv} \qquad W''_{iv} = \frac{X_{ov}}{X_{00}}$$

X_{ov} is the crop area for the v-th post stratum

X_{00} is the total crop acreage in the district

The estimator of variance of can be $\bar{y}_{st}$ approximately written as

$$\sum_{v} W_{iv}'^{2}\hat{V}(\bar{y}_{ov}) = \sum_{v} W_{iv}'^{2} \sum_{i} W_{iv}''^{2}\hat{V}(\bar{y}_{iv})$$

Since, sample in each block has been selected independently.

Composite estimator

When small area samples are relatively small, the synthetic estimators outperform the simple direct estimators; however, when small area sample sizes are large, the direct estimators outperform the synthetic estimators. Thus, it was concluded that a weighted sum of these two (2) estimators would be better than choosing one over the other - The composite estimator

$$\bar{y}_{ct} = \varphi_t \bar{y}_{dt} + (1-\varphi_t)\bar{y}_{st}$$

The approximate optimal weight $\varphi_t^{*} = \frac{1}{1+f_t}$

Where, $f_t = \frac{\hat{V}(\bar{y}_{dt})}{\hat{V}(\bar{y}_{st})}$

The estimator of variance of $\bar{y}_{ct}$ can be approximately written as

$$\hat{V}(\bar{y}_{ct}) \cong \varphi 2_t \hat{V}(\bar{y}_{dt}) + (1-\varphi_t)^2 \hat{V}(\bar{y}_{st})$$

Result and Discussion

Hisar district in Haryana was chosen for the study of wheat crop yield estimation during 2018-2019 using the crop yield data obtained from general crop yield estimation surveys based on crop cutting experiments and the satellite spectral data of MOD13Q1 v006 satellite for 18 February 2019. At this point the crop was at the heading/flowering stage and the spectral indices had the highest correlation with crop yield.

A total of 275 villages from 9 blocks of the Hisar district were considered for CCEs and the sample size of the block (Number of village) ranged from 12 to 46. Based on yield the 275 villages are stratified into 3 strata (Figure-2 and Table-1). Figure-3 and Table-2 shows the post stratification based on NDVI values also the Figure-4 and Table-3 displays stratification based on EVI values of the wheat crop in the villages.

All the developed estimators (Post stratified Direct, Synthetic and Composite) provide more efficient estimation of crop yield compared with the usual estimator of crop yield. Table-4 shows the coefficient of variation (CV%) of different wheat crop yield estimators, as expected the CV value is very less for composite estimates. Figure-5 reveals the best composite estimate for particular blocks. Crop yield estimator for the blocks based on NDVI stratification is found more efficient (Based on CV value) for Hansi 1, Hansi 2 and Adampur whereas the estimator based on yield stratification is found the best for the rest of blocks.

The more reliable wheat yield estimate is given in the figure-6. This study shows that reliable estimates of crop yield at the block level can be obtained from the existing sampling design of general crop estimation surveys and the existing number of crop cutting experiments provided that we make use of auxiliary information such as spectral data for stratification of crop area and the improved small area estimators.

References

Mahalanobis PC (1946). Sample surveys of crop yields in India. *Sankhyā*, **7:** 269-280.

Sukhatme PV and Agarwal OP (1947). Crop-cutting survey on wheat by the random sampling method. In Report: Indian Council of Agricultural Research, New Delhi.

Tikkiwal BD and Tikkiwal GC (1998). Small area estimation in India - Crop yield and acreage statistics. Paper presented at the International Conference on Agricultural Statistics - 2000, Washington, D.C., USA.

Tikkiwal GC and Ghiya A (2004). A generalized class of composite estimators with application to crop acreage estimation for small domains. *Stat. Trans.*, **1:** 697-711.

Tikkiwal GC and Ghiya A (2000). A generalized class of synthetic estimators with application to crop acreage estimation for small domains. *Biom. J.*, **42:** 865-876.

Singh R, Semwal DP, Rai A and Chhikara RS (2002). Small area estimation of crop yield using remote sensing satellite data. *Int. J. Remote Sens.*, **23:** 49-56.

Singh R, Goyal RC, Saha SK and Chhikara RS (1992). Use of satellite spectral data in crop yield estimation surveys. *Int. J. Remote Sens.*, **13:** 2583-2592.

Table 1: classification of villages based on yield group

1st Strata < 4144.21	26 villages
2nd Strata 4144.21 to 5738.66	217 villages
3rd Strata > 5738.66	32 villages

Table 2: classification of villages based on NDVI value

1st Strata< 0.66 (moderately healthy class)	16 villages
2nd Strata 0.66-0.8 (healthy class)	101 villages
3rd Strata > 0.8 (Very healthy class)	158 villages

Table 3: Classification of villages based on EVI value

1[st]Strata < 0.485	40 Villages
2[nd]Strata 0.485 to 0.684	197 villages
3[rd]Strata >0.684	38 villages

Table 4: Coefficient of variation (CV%) of the different estimators of wheat crop yield

	Usual	D-YIELD	S-YIELD	C-YIELD	D-NDVI	S-NDVI	C-NDVI	D-EVI	S-EVI	C-EVI
Hansi 1	10.15	7.13	5.45	4.33	6.49	5.32	4.12	7.16	5.95	4.58
Hansi 2	7.59	6.31	2.57	2.38	5.09	2.31	2.11	5.18	3.88	3.11
Narnaund	20.97	2.64	2.65	1.87	9.49	8.82	6.49	5.48	2.94	2.59
Hisar 1	6.97	5.19	1.40	1.35	4.76	1.92	1.78	6.35	2.17	2.05
Hisar 2	18.06	5.21	3.76	3.06	6.74	7.20	4.92	8.97	6.91	5.48
Adampur	16.62	5.87	1.32	1.28	4.61	1.07	1.04	6.80	7.40	5.01
Barwala	20.76	3.70	1.75	1.58	11.07	8.46	6.75	12.30	8.12	6.78
Uklana	12.51	5.49	1.20	1.17	6.84	1.72	1.67	7.49	3.17	2.92
Agroha	13.09	5.15	1.51	1.45	6.26	2.20	2.08	8.45	1.98	1.93

*usual = simple average, D = post stratified Direct, S = synthetic and C = Composite

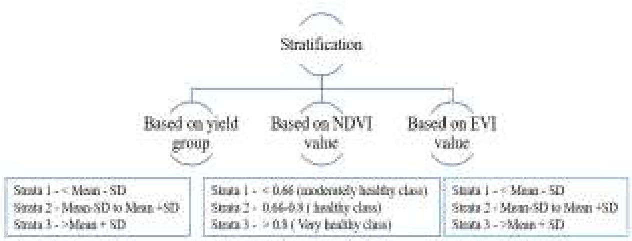

Figure 1: Stratal classification based on yield, NDVI and EVI

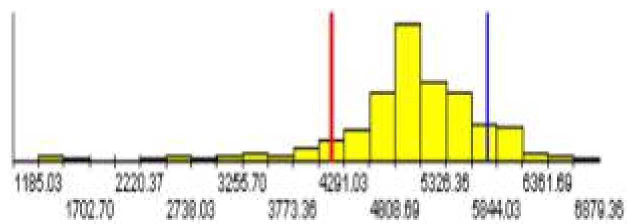

Figure 2: Classification of villages based on yield group

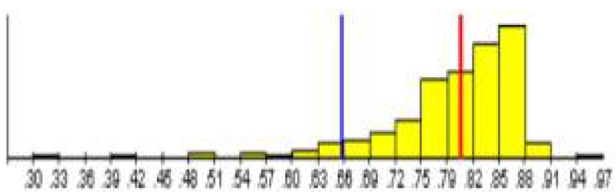

Figure 3: Classification of villages based on NDVI value

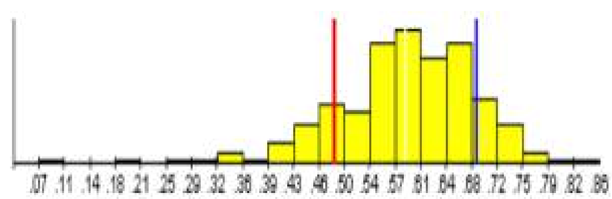

Figure 4: Classification of villages based on EVI value

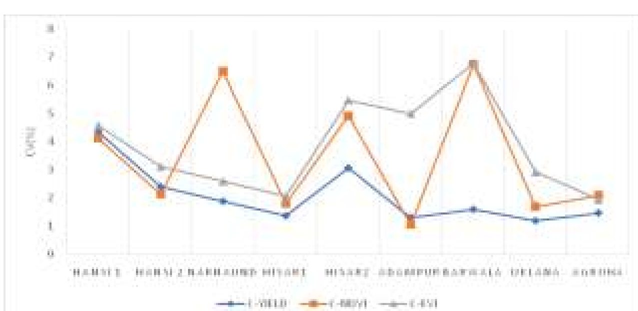

Figure 5: Coefficient of variation (CV%) of composite estimators of wheat crop yield

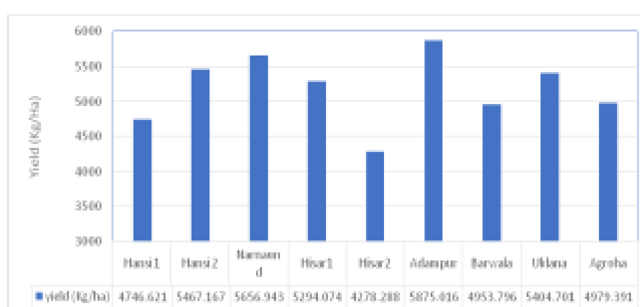

Figure 6: Reliable estimate of block level wheat yield for Hisar District