Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 69(1) 2015 49-56

District Level Crop Yield Estimation under Spatial Small Area Model

U.C. Sud, Kaustav Aditya and Hukum Chandra
ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 10 June 2013; Revised 12 March 2015; Accepted 13 March 2015

SUMMARY

In this article we demonstrate an application of small area estimation technique to produce district level estimates of crop yield for three major crops of the State of Uttar Pradesh using the Improvement of Crop Statistics Scheme data and the auxiliary data from various secondary sources. In particular, we use a spatial model for small area estimation to improve the district level crop yield estimates. The results show improvement in the district level crop yield estimates due to use of spatial information in small area estimation.

Keywords: Crop cutting experiments, Improvement of Crop Statistics, District level estimates, Small area estimation, Spatial model.

1. INTRODUCTION

The crop yield (*i.e.*, production per hectare of land) estimates are produced on the basis of scientifically designed crop cutting experiments (CCEs) conducted under the scheme of General Crop Estimation Surveys (GCES) in India. More than 800,000 CCEs are conducted annually for this purpose. The sample size gathered through GCES is sufficient for providing precise estimates of crop yield at district level. But, the procedure of conducting the CCEs are very tedious and time consuming which makes some of the enumerators not to follow the appropriate technique for CCEs and by virtue of that the data quality of the GCES goes beyond the desired limit. To improve the quality of data collected under the GCES, a scheme titled 'Improvement of Crop Statistics (ICS)' has been introduced by the Directorate of Economics and Statistics, Ministry of Agriculture, Government of India and implemented by the National Sample Survey Office (NSSO) and the State Agricultural Statistics Authority (SASA) jointly.

Under this scheme, quality check on the field operation of GCES is carried out by supervising around 30,000 CCE by NSSO and State Government supervisory officers. The findings of the ICS results reveal that the crop cutting experiments are generally not carried out properly resulting in data which lacks desired quality. Due to limitation of infrastructure and constraints of resources, there is a felt need to reduce the sample size under GCES drastically so that volume of work of the enumerator is reduced and also better supervision of the operation of CCE becomes possible leading to improvement in data quality. But, with the reduction of sample sizes the standard error of the estimates will increase. The reduced sample size is more of concern when aim is to produce estimates at district level since estimators based on the sample data from any particular district (also referred as area or small area) can be unstable. This problem of small sample size within the districts can be solved by using small area estimation (SAE) techniques.

Corresponding author: Kaustav Aditya
E-mail address: kaustav@iasri.res.in

The SAE techniques are usually a model-based method where we use statistical models to link the variable of interest with auxiliary information, *e.g.* Census and Administrative data, for the small areas to define model-based estimators for these areas. See Pfeffermann (2002) and Rao (2003). The underlying models defining the small area estimators are referred as the small area models. These small area models are broadly classified as the area level (Fay and Herriot 1979) and unit level small area model (Battese *et al.* 1988). The area level small area models are used when auxiliary information is available only at area level. They relate small area direct estimates to area-specific covariates whereas the unit level small area models relate the unit values of a study variable to unit-specific covariates. Sud *et al.* (2012) applied SAE techniques under area level model to obtain estimates of average yield for paddy crop at small area levels in the State of Uttar Pradesh in India by linking data generated under ICS scheme by NSSO and the Population Census. They find that the estimates generated through SAE method are reliable and more efficient than the direct estimate from ICS data alone. However, they used the EBLUP estimator under area level random effect model (Fay and Herriot 1979).

In this paper we consider an application of spatial version of area level random effect (Petrucci and Salvati 2005, Petrucci *et al.* 2005 and Singh *et al.* 2005) to produce the estimates of average yield of Rice, Wheat and Sugarcane crops at district level in the State of Uttar Pradesh in India using the data under ICS scheme and the auxiliary data from Population Census 2011 and Fertilizer Statistics 2010. Section 2 introduces data used for analysis and Section 3 describes the methodology applied in analysis. In Section 4 we compare small area estimates generated by two methods namely small area estimation technique with and without spatial information. Section 5 finally presents the main conclusions.

2. DATA DESCRIPTION

We use data under ICS scheme collected during the year 2010-11 for paddy, wheat and sugarcane crops for the State of Uttar Pradesh. In the State of Uttar Pradesh there are 70 districts, however supervision, on a sub-sample, of crop cutting experiments work under ICS scheme is carried out in 42 districts for rice, 51 districts for wheat and 29 districts for sugarcane. As a result, there is no sample data for the remaining districts. These non sample districts are also referred as the out of sample districts. Fig. 1 shows the distribution of sample sizes in the sampled districts for the three crops i.e. rice, wheat and sugarcane. The area specific sample sizes for these respective sample districts for the three major crops range from minimum of 4 to maximum of 28 CCE with average of 11 CCE in case of rice, minimum of 4 and maximum of 18 CCE

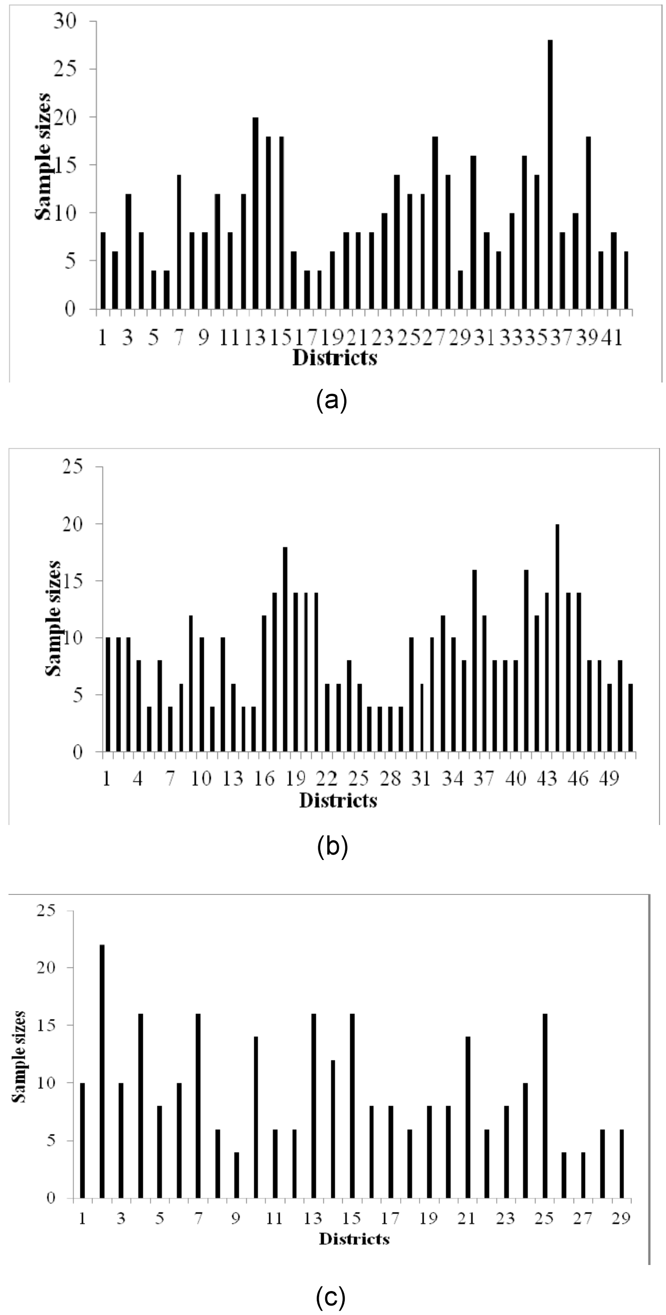

Fig. 1. District wise sample sizes for (a) Rice (b) Wheat and (c) Sugarcane under ICS of Uttar Pradesh in 2010-11

with average of 9 CCE in case of wheat and, minimum of 4 and maximum of 22 CCE with average of 10 CCE in case of sugarcane. A total of 442, 472 and 284 CCE are supervised for rice, wheat and sugarcane respectively for recording yield data in the State of Uttar Pradesh for district level. In a few districts the sample size is so small that the traditional survey estimation approaches lead to unstable district level estimates. In addition in the non-sampled districts we cannot produce estimate of crop yield due to unavailability of sample data. In SAE, covariates are taken from the Population Census 2011 and the Fertilizer Statistics 2010. There are number of covariates available from these two sources. However, we did some exploratory data analysis, for example, first we segregated group of covariates with significant correlation with crop yield and then modelling. Finally, we used population density for rice and sugarcane and fertilizer consumption during rabi season for wheat as covariates in small area estimation.

3. METHODOLOGY

The Fay–Herriot model (Fay and Herriot 1979) is widely used area level model in SAE. This model relates small area direct survey estimates to area-specific covariates. The SAE under this model is one of the most popular methods used by private and public agencies because of its flexibility in combining different sources of information and explaining different sources of errors. In this section we first elaborate SAE method under area-level Fay–Herriot model (Fay and Herriot 1979), that is, the EBLUP under this model. We then introduce the Spatial-EBLUP (Petrucci and Salvati 2005 and Singh *et al*. 2005) which takes into account the spatial structure of the data by modeling the random effects according to a SAR specification.

Let the population is divided into D small areas (district in our application) or areas and we use a subscript d to index the quantities related to district d (d = 1, 2,..., D). Let $\hat{\theta}_d$ denotes the direct survey estimate of unobservable population value θ_d for district d (d = 1, 2,..., D). Let $\mathbf{x}_d$ be the p-vector of known auxiliary variables, often obtained from various administrative and census records, related to the population mean θ_d. The simple area specific two stage model suggested by Fay and Herriot (1979) is,

$$\hat{\theta}_d = \theta_d + e_d \text{ and } \hat{\theta}_d = \mathbf{x}_d^T\boldsymbol{\beta} + u_d,\ d = 1, 2, ..., D. \quad (1)$$

We can express model (1) as an area level linear mixed model of the form

$$\hat{\theta}_d = \mathbf{x}_d^T\boldsymbol{\beta} + u_d + e_d,\ d = 1, 2, ..., D. \quad (2)$$

Here β is a p-vector of unknown fixed effect parameters u_d's are independent and identically distributed normal random errors with E(u_d) = 0 and Var (u_d) = σ_u^2, and e_d's are independent sampling errors normally distributed with $E(e_d|q_d) = 0$ and Var $(e_d|q_d)$ = σ_d^2. The two errors are independent of each other within and across areas. Usually, σ_d^2 is known while σ_u^2 is unknown and it is estimated from sample data. Methods of estimating σ_u^2 include maximum likelihood (ML) and restricted maximum likelihood (REML) under normality and the method of fitting constants without normality assumption (Rao 2003). Let $\hat{\sigma}_u^2$ denotes estimate of σ_u^2. Then under model (2), the Empirical Best Linear Unbiased Predictor (EBLUP) of θ_d is given by

$$\hat{\theta}_d^{EBLUP} = \mathbf{x}_d^T\hat{\boldsymbol{\beta}} + \hat{\gamma}_d(\hat{\theta}_d - \mathbf{x}_d^T\hat{\boldsymbol{\beta}}) = \hat{\gamma}_d\hat{\theta}_d + (1-\hat{\gamma}_d)\mathbf{x}_d^T\hat{\boldsymbol{\beta}}, \quad (3)$$

where $\hat{\gamma}_d = \dfrac{\hat{\sigma}_u^2}{(\sigma_u^2 + \hat{\sigma}_u^2)}$ and $\hat{\boldsymbol{\beta}}$ is the generalized least square estimate of β. Note that $\hat{\theta}_d^{EBLUP}$ is a linear combination of direct estimate $\hat{\theta}_d$ and the model based regression synthetic estimate $\mathbf{x}_d^T\hat{\boldsymbol{\beta}}$ with weight $\hat{\gamma}_d$. Here $\hat{\gamma}_d$ is called the "shrinkage factor" since it 'shrinks' the direct estimator, $\hat{\theta}_d$ towards the synthetic estimator, $\mathbf{x}_d^T\hat{\boldsymbol{\beta}}$.

An approximately model unbiased estimate of mean squared error (MSE) of the EBLUP (3) is given by Prasad and Rao (1990) as follows.

$$MSE(\hat{\theta}_d^{EBLUP}) = g_{1d}(\hat{\sigma}_u^2) + g_{2d}(\hat{\sigma}_u^2) + 2g_{3d}(\hat{\sigma}_u^2)V\hat{a}r(\hat{\sigma}_u^2), \quad (4)$$

where,

$$g_{1d}(\hat{\sigma}_u^2) = \hat{\gamma}_d\hat{\sigma}_d^2$$

$$g_{2d}(\hat{\sigma}_u^2) = (1-\hat{\gamma}_d)^2\mathbf{x}_d^T\hat{V}(\hat{\boldsymbol{\beta}})\mathbf{x}_d$$

$$g_{3d}(\hat{\sigma}_u^2) = \left[\frac{\hat{\sigma}_d^4}{(\hat{\sigma}_d^2 + \hat{\sigma}_u^2)^3}\right]\sum_{d=1}^{D}\hat{V}ar(\hat{\sigma}_u^2)$$

with $V\hat{a}r(\hat{\sigma}_u^2) = 2D^{-2}\sum_{d=1}^{D}(\hat{\sigma}_d^2 + \hat{\sigma}_u^2)^2$ when estimating $\hat{\sigma}_u^2$ by the method of fitting constants.

In Section 2 we noticed that there are many out of sample districts in the data and the conventional approach for estimating small areas in this case is synthetic estimation, based on a suitable model fitted to the data from the sampled areas. This is equivalent to setting the area effect for out of sampled area to zero. Under model (2), the synthetic EBLUP predictor for θ_d is

$$\hat{\theta}_d^{SYN} = \mathbf{x}_d^T \hat{\boldsymbol{\beta}}. \tag{5}$$

This predictor is referred as the Synthetic EBLUP (hereafter denoted by SYN). Under model (2), the MSE estimate for the synthetic predictor (5) is

$$MSE(\hat{\theta}_d^{SYN}) = \mathbf{x}_d^T Var(\hat{\boldsymbol{\beta}})\mathbf{x}_d + \hat{\sigma}_u^2. \tag{6}$$

In model (2) the random area effects are considered to be independent. However, it is often reasonable to assume that the effects of neighbouring areas (defined, for example, by a contiguity criterion) are correlated, with the correlation decaying to zero as the distance between these areas increases. Consequently, small area models should allow for spatial correlation of area random effects. In order to take into account the correlation between neighbouring areas we consider the use of spatial models for random area effects. We consider a linear regression model with spatial dependence in the error structure. In particular, we assume a Simultaneous Autoregressive (SAR) error process, where the vector of random area effects $\boldsymbol{v} = (v_d)$ satisfies $\boldsymbol{v} = \rho \mathbf{W}\boldsymbol{v} + \boldsymbol{u}$ and ρ is a spatial autoregressive coefficient, $\mathbf{W}$ is a proximity matrix of order D and $\boldsymbol{u} \sim N(0, \sigma_u^2 \boldsymbol{I}_D)$. Since $\boldsymbol{v} = (\boldsymbol{I}_D - \rho\mathbf{W})^{-1} \boldsymbol{u}$ with $E(\boldsymbol{u}) = 0$ and $Var(\boldsymbol{u}) = \sigma_u^2 \boldsymbol{I}_D$, we have $E(\boldsymbol{v}) = 0$ and $Var(\boldsymbol{v}) = \sigma_u^2 [(\boldsymbol{I}_D - \rho\mathbf{W})(\boldsymbol{I}_D - \rho\mathbf{W}^T)]^{-1} = \Omega$. The $\mathbf{W}$ matrix describes how random effects from neighbouring areas are related, whereas ρ defines the strength of this spatial relationship. The simplest way to define $\mathbf{W}$ is as a contiguity matrix. The elements of $\mathbf{W}$ take non-zero values only for those pairs of areas that are adjacent. Then the model (2) with spatially correlated errors is

$$\theta = \mathbf{x}\beta + \mathbf{z}(\boldsymbol{I}_D - \rho\mathbf{W})^{-1}\boldsymbol{u} + \mathbf{e} = \mathbf{x}\beta + \mathbf{z}\boldsymbol{v} + \mathbf{e}. \tag{7}$$

The covariance matrix of the vector θ is $\mathbf{V} = \mathbf{z}\Omega\mathbf{z}^T + \mathbf{R}$. In practice, the vector of parameters $\psi = (\sigma_u^2, \rho)^T$ is unknown. Assuming normality of the random effects, the parameter vector σ_u^2 and ρ can be estimated via ML as well as REML methods. Numerical approximations to either the ML or REML estimators $\hat{\sigma}_u^2$ and $\hat{\rho}$ can be obtained via a two-step procedure. At the first step, the Nelder-Mead algorithm is used to approximate these estimates. The second step then uses these approximations as starting values for a Fisher scoring algorithm. See Petrucci *et al.* (2005), Petrucci and Salvati (2005) for computational details. Replacing $\psi = (\sigma_u^2, \rho)^T$ with an asymptotically consistent estimator $\hat{\psi} = (\hat{\sigma}_u^2, \hat{\rho})^T$, and assuming that (7) holds, the spatial Empirical Best Linear Unbiased Predictor (Spatial-EBLUP or SEBLUP) of θ_d is

$$\hat{\theta}_d^{Spatial-EBLUP} = \mathbf{x}_d^T \hat{\boldsymbol{\beta}}^s + a_d^T \hat{v}, \text{ with}$$
$$\hat{v} = \hat{\Omega} z^T \hat{\mathbf{V}}^{-1}(\boldsymbol{y} - \mathbf{x}\hat{\boldsymbol{\beta}}^s), \tag{8}$$

where $\hat{\boldsymbol{\beta}}^s = \left(\mathbf{x}^T\hat{\mathbf{V}}^{-1}\mathbf{x}\right)^{-1}\left(\mathbf{x}^T\hat{\mathbf{V}}^{-1}\mathbf{y}\right)$ is the EBLUE of β under model (7), a_d is the D-vector $(0, \ldots, 1, \ldots, 0)^T$ with the 1 in the d^{th} position, $\hat{\Omega} = \hat{\sigma}_u^2[(\boldsymbol{I}_D - \hat{\rho}\mathbf{W})(\boldsymbol{I}_D - \hat{\rho}\mathbf{W}^T)]^{-1}$ and $\hat{\mathbf{V}} = \{z\hat{\sigma}_u^2[(\boldsymbol{I}_D - \hat{\rho}\mathbf{W})(\boldsymbol{I}_D - \hat{\rho}\mathbf{W}^T)]^{-1}\mathbf{z}^T + diag(\sigma_{ed}^2)\}$. For out of sampled areas, spatial Synthetic EBLUP (hereafter denoted by Spatial-SYN) of θ_d is

$$\hat{\theta}_d^{Spatial-SYN} = \mathbf{x}_d^T \hat{\boldsymbol{\beta}}^s. \tag{9}$$

Following Petrucci and Salvati (2005), an approximately unbiased estimator of the MSE of the SEBLUP (6) is

$$MSE(\hat{\theta}_d^{SEBLUP-SYN})$$
$$= g_{1d}^{(s)}(\hat{\psi}) + 2g_{2d}^{(s)}(\hat{\psi}) + 2g_{3d}^{(s)}(\hat{\psi}) - \mathbf{B}_d^{(s)T}(\hat{\psi})\nabla g_{1d}^{(s)}(\hat{\psi}), \tag{10}$$

where the first term $g_{1d}^{(s)}(\hat{\psi})$ is due to the estimation of random area effects and is of order $O(1)$ while the second term $g_{2d}^{(s)}(\hat{\psi})$ is due to the estimation of β and is of order $O(D^{-1})$ for large D. The third term $g_{3d}^{(s)}(\hat{\psi})$ is due to the estimation of the variance component. Finally, the last term $\mathbf{B}_d^{(s)T}(\hat{\psi})\nabla g_{1d}^{(s)}(\hat{\psi})$ is bias when ML method of estimation is used for variance component. This term is negligible and thus ignored when REML or method of moment is used for parameter estimation. Various terms of (7) are:

$$g_{1d}^{(s)}(\hat{\psi}) = a_d^T(\hat{\Omega} - \hat{\Omega}\mathbf{z}^T\hat{\mathbf{V}}^{-1}\mathbf{z}\hat{\Omega})a_d,$$
$$g_{2d}^{(s)}(\hat{\psi}) = (\mathbf{x}_d^T - \mathbf{c}_d^T\mathbf{x})(\mathbf{x}^T\hat{\mathbf{V}}^{-1}\mathbf{x})(\mathbf{x}_d^T - \mathbf{c}_d^T\mathbf{x})^T, \text{ and}$$
$$g_{3d}^{(s)}(\hat{\psi}) = tr\{(\nabla\mathbf{c}_d^T)\hat{\mathbf{V}}(\nabla\mathbf{c}_d)\hat{V}(\hat{\psi})\},$$

with $\mathbf{c}_d^T = a_d^T \hat{\Omega} z^T \hat{\mathbf{V}}^{-1}$, $\nabla \mathbf{c}_d^T = \partial \mathbf{c}_d^T / \partial \psi, \hat{V}(\hat{\psi})$ is the estimate of the asymptotic covariance matrix of $\hat{\psi}$ defined by the inverse of the relevant observed information matrix and $\mathbf{B}_d^{(s)T}(\hat{\psi}) \nabla g_{1d}^{(s)}(\hat{\psi})$ bias correction due to ML estimator of ψ.

4. EMPIRICAL STUDY

In this section we report the results from analysis carried out to produce the district level crop yield estimates. We compare the spatial EBLUP (SEBLUP), the EBLUP and the direct estimator used to generate the district level crop yield estimates. We examine the usefulness spatial information in producing the small area estimates. The analysis is carried out for three major crops (rice, wheat and sugarcane) using the ICS data of the State of Uttar Pradesh. We used SAE package of R-Software for our analysis. The values of yield estimates generated by using direct survey estimator, EBLUP and SEBLUP along with their percentage standard errors (%SE) are given in Table 1, 2 and 3 for rice, wheat and sugarcane crops respectively. The percentage standard error (%SE) of the estimator $\hat{\theta}_d$ in district d is calculated as

$$\%SE_d = 100 \times \frac{SE(\hat{\theta}_d)}{\hat{\theta}_d}; d = 1, ..., D.$$

These results in Tables 1-3 clearly indicate that the SEBLUP method is providing better estimates than the usual EBLUP and the direct survey estimator. It can also be seen that there is a significant improvement in the %SE of the SEBLUP than the EBLUP and the direct estimates. Two points emerged from this analysis, (i) the small area estimate provides efficient and better estimates for crop yield as compared to the direct survey estimates, (ii) the use of spatial information improve the efficiency of small area estimates. For out of sample districts we produced the SEBLUP estimates. These out of sample districts are 28, 19 and 41 for rice, wheat and sugarcane respectively. The district level yield estimates for these out of sample districts produced using SEBLUP are reported in Table 4, 5 and 6 for rice, wheat and sugarcane respectively. It is noteworthy that in some districts %SE is high, in

Table 1. District level yield estimates (gms/CCE plot area) of rice crop for Uttar Pradesh for 2010-11.

District	Yield			% SE			District	Yield			% SE		
	Direct	EBLUP	SEBLUP	Direct	EBLUP	SEBLUP		Direct	EBLUP	SEBLUP	Direct	EBLUP	SEBLUP
Saharanpur	17256	16984	17397	7.71	7.68	7.17	Kaushambi	17400	17045	16419	8.32	8.29	7.54
M.Nagar	19033	18918	18816	4.25	4.24	4.17	Allahabad	14830	14648	14611	13.84	13.36	10.88
Bijnor	15233	15033	15109	7.74	7.72	7.08	Barabanki	13743	13198	13824	17.09	16.72	11.53
Moradabad	16613	16506	15741	11.95	11.51	9.96	Faizabad	16021	15780	15738	10.33	10.16	8.61
J.P. Nagar	12050	12044	12113	2.61	2.60	2.57	Ambedkarnagar	18696	18608	18277	3.83	3.82	3.76
Ghaziabad	21833	27652	19985	15.01	10.92	14.61	Sultanpur	17438	16672	15843	12.07	12.01	9.32
Buland Shahar	14321	14125	13890	8.77	8.73	8.09	Bahraich	13543	13417	13490	7.89	7.86	7.27
Aligarh	6539	7111	10015	31.35	27.50	16.14	Shrawasti	13131	11816	12859	22.58	22.76	15.23
Etah+Mainpuri	15325	15385	15441	8.05	7.88	7.26	Gonda	11441	11388	12212	11.84	11.64	9.60
Badaun	15879	15521	15059	8.47	8.49	7.73	Sidharthnagar	13950	13663	13809	13.36	13.11	10.73
Shahjahanpur	18225	17475	16529	13.72	13.35	10.44	Sant Kabir Nagar	15592	15092	15140	17.06	16.30	12.31
Khiri	13833	13184	13854	13.85	13.93	10.97	Maharajganj	16010	15873	15635	6.28	6.26	5.97
Sitapur	13060	12739	13434	14.10	13.91	11.06	Gorakhpur	13688	13973	14202	15.78	14.68	10.98
Hardoi	14717	14001	14586	14.53	14.50	10.77	Kushinagar	13979	14009	14287	8.65	8.49	7.74
Rai Bareilly	14125	13920	14407	8.66	8.63	7.75	Azamgarh	11311	11327	11480	4.38	4.37	4.24
Farrukhabad	21967	16448	15941	26.44	25.99	13.67	Mau	12469	12843	12476	18.27	16.75	12.97
Kannauj	14950	14253	14982	15.36	15.18	11.00	Balia	9195	9311	9878	11.70	11.40	10.13
Etawah	24588	20101	17034	15.18	16.02	11.96	Ghazipur	12376	12388	12291	7.25	7.18	6.86
Auraiya	18242	14866	15769	23.69	24.03	13.64	S.R.Nagar	12700	12982	13181	10.87	10.41	9.35
Kanpur®	18081	15324	14946	20.11	20.61	14.55	Mirzapur	8763	8245	10238	41.72	38.45	20.50
Kanpur(u)	11319	11845	13120	16.43	15.11	11.42	Sonbhadra	6308	6038	8259	29.98	30.06	19.55

Table 2. District level yield estimates (gms/CCE plot area) of wheat crop for Uttar Pradesh for 2010-11.

District	Yield			% SE			District	Yield			% SE		
	Direct	**EBLUP**	**SEBLUP**	**Direct**	**EBLUP**	**SEBLUP**		**Direct**	**EBLUP**	**SEBLUP**	**Direct**	**EBLUP**	**SEBLUP**
Saharanpur	15560	14999	15684	12.75	12.65	10.60	Hamirpur	9148	9148	10474	0.45	0.45	10.61
M.Nagar	16985	16950	16611	2.70	2.70	6.38	Banda	11284	11241	11315	4.30	4.30	4.18
Bijnor	14865	14272	13046	13.03	13.01	12.79	Chitrakoot	7506	7501	9662	2.75	2.75	16.47
Moradabad	13906	13870	13843	9.29	9.14	13.98	Fatehpur	13220	12995	13216	8.68	8.69	9.13
J.P Nagar	10175	10081	10895	13.99	13.80	7.78	Kaushambi	12544	12083	12302	19.44	18.89	10.97
Meerut	16700	16694	15077	1.72	1.72	9.14	Allahabad	10600	10610	11965	9.02	8.92	9.69
Bagpat	15788	15744	15533	3.43	3.42	3.37	Barabanki	15892	15728	15063	5.52	5.52	7.76
Ghaziabad	14050	17542	13886	17.33	13.21	10.42	Faizabad	13280	13160	12778	10.29	10.16	8.54
Bulandshahar	16919	16723	16761	5.44	5.45	2.68	Ambedkarnagar	12044	11866	12679	17.74	17.11	11.91
Aligarh	13154	12923	13649	14.19	13.51	9.41	Sultanpur	9351	9319	9382	14.85	14.57	2.19
Hathras	13363	11248	13398	41.75	37.35	6.42	Bahraich	11513	11367	11300	9.57	9.56	9.41
Agra	14335	14156	14714	10.30	10.17	10.24	Balrampur	8988	8791	9240	20.29	19.97	7.32
Firozabad	16883	16379	15512	10.76	10.69	8.36	Siddrathnagar	10281	10274	10564	4.37	4.36	16.58
kansiramnagar	14550	14553	14390	4.46	4.44	4.27	Maharajganj	14256	14228	13680	3.23	3.23	8.43
Mainpuri	18388	17428	14627	10.12	10.26	10.95	Gorakhpur	11938	12011	11981	11.26	10.96	2.38
Badaun	16038	15518	15939	9.71	9.76	2.80	Kushinagar	11679	11698	11767	8.42	8.31	7.67
Shshjahanpur	12729	12723	13753	5.54	5.51	9.53	Deoria	11786	11797	11871	7.04	6.98	6.90
Sitapur	10539	10519	10739	5.08	5.08	4.09	Azamgarh	11949	11947	11938	3.65	3.64	9.25
Hardoi	12675	12555	12731	7.14	7.14	6.30	Balia	13321	13283	12266	5.96	5.93	17.62
Unnao	13504	13049	13503	12.50	12.52	0.30	Ghazipur	9943	9958	11252	7.87	7.80	16.34
Raibareilly	10557	10524	10864	6.14	6.13	5.67	Chandauli	8900	8876	9319	10.47	10.39	9.03
Etawah	11717	11124	12543	19.89	19.71	10.55	Varanasi	10681	10704	11421	2.38	2.37	12.43
Aurriya	13200	12945	12888	9.65	9.66	6.57	S.R.Nagar	12067	12094	11951	4.52	4.50	4.42
Kanpur(D)	17713	16358	15591	12.67	12.96	8.61	Mirzapur	7713	7534	9172	25.38	24.87	16.25
Jhansi	12892	10866	12586	26.18	27.49	8.54	Sonbhadra	4967	4883	5216	24.76	24.75	8.24
Lalitpur	11500	10448	11483	19.54	20.31	3.91							

Table 3. District level yield estimates (gms/CCE plot area) of sugarcane crop for Uttar Pradesh for 2010-11.

District	**Yield**			**% SE**			**District**	**Yield**			**% SE**		
	Direct	EBLUP	SEBLUP	Direct	EBLUP	SEBLUP		Direct	EBLUP	SEBLUP	Direct	EBLUP	SEBLUP
Saharanpur	15560	128546	138556	23.22	21.22	19.69	Barabanki	16038	75366	102398	35.55	30.63	22.54
M.Nagar	16985	136228	141221	10.87	10.61	10.23	Faizabad	12729	155410	152544	11.76	11.48	11.69
Bijnore	14865	110182	110586	23.45	21.88	21.80	Bahraich	10539	101962	102690	32.80	30.51	30.29
Muradabad	13906	75216.1	76210	38.95	27.81	27.45	Balrampur	12675	82256	102835	15.02	14.73	11.78
Rampur	10175	133211	134296	8.41	8.30	8.24	Gonda	13504	108348	109281	13.12	12.78	12.67
Meerut	16700	119689	133674	40.36	25.97	23.25	Basti	10557	86976	103127	24.50	21.94	18.50
Baghpat	15788	160367	165735	11.08	10.94	10.58	Maharajganj	11717	141447	145278	3.29	3.29	3.20
Bulandshar	14050	123911	125889	7.44	7.41	7.30	Gorakhpur	13200	122220	133102	23.29	19.24	17.67
Mathura	16919	136580	139892	5.95	5.95	5.81	Kushinagar	17713	136393	139429	8.76	8.54	8.36
Bareilly	13154	139917	144012	10.41	10.15	9.86	Deoria	12892	126231	145755	8.60	8.39	7.26
Pilibhit	13363	111760	121665	22.82	23.17	21.28	Mau	11500	148889	152082	3.36	3.35	3.28
Shahjahanpur	14335	89277.5	92747	38.74	33.63	32.38	Jaunpur	9148	178756	179878	19.05	18.16	18.04
Khiri	16883	135532	136961	5.64	5.67	5.61	Varanasi	11284	153410	207675	18.36	13.91	10.27
Sitapur	14550	72145.9	102106	23.17	21.52	15.21	J.P.Nagar	7506	102673	102998	45.84	35.16	35.05
Hardoi	18388	163052	167252	6.90	6.96	6.78							

Table 4. District level yield estimates (gms/CCE plot area) of rice crop for out of sample districts using Spatial EBLUP (SEBLUP).

Districts	Yield	% SE	Districts	Yield	% SE
Rampur	14244	42.41	Jaunpur	14464	42.67
Mathura	13796	43.84	Agra	14426	41.88
Bareilly	14420	41.89	Firozabad	14366	42.05
Pilibhit	13531	44.77	Bagpat	14242	42.42
Unnao	13690	44.20	Mahamaya nag	13997	43.18
Lucknow	15747	38.74	Baharich	13859	43.61
Banda	13193	46.16	Chandauli	13834	43.69
Fatehpur	13618	44.46	Kanshiram nagar	13767	43.91
Pratapgarh	14003	43.16	Mainpuri	13685	44.18
Balrampur	13658	44.32	Jhansi	13175	46.13
Basti	14115	42.80	Jalaun	13117	46.58
Varanasi	16808	37.00	Chitrakut	13011	47.20
Meerut	14878	40.67	Mahboba n	12959	47.02
Deoria	14678	41.18	Hamirpur	12942	46.70
G B Nagar	14667	41.21	Lalitpur	12891	47.31

Table 5. District level yield estimates (gms/CCE plot area) of wheat crop for out of sample districts using Spatial EBLUP (SEBLUP).

Districts	Yield	% SE	Districts	Yield	% SE
G B Nagar	9565	11.41	Jalaun	9676	25.60
Mathura	9692	27.68	Mahoba n	9523	6.050
Bareilly	9825	44.27	Pratapgarh	9660	23.610
Pilibhit	9742	33.98	Shrawasti	9523	6.060
Shahjahanpur	9884	51.38	Gonda	9738	33.39
Lucknow	9664	24.08	Basti	9691	27.57
Rai Bareilly	9692	27.71	S.K.Nagar	9616	18.01
Farukhabad	9783	39.05	Maunathbhanjan	9644	21.55
Kannauj	9658	23.33	Jaunpur	9727	32.07
Kanpur(u)	9836	45.59			

particular, for sugarcane. We observed that the number of out of sample districts in this case is 41, that is, more than sample districts. We used model fitted using data from 29 districts to predict yield for 41 districts. Similar problem was also observed in rice and wheat crops but in this number of sample districts are more than the out of sample district and hence prediction are little better.

Table 6. District level yield estimates (gms/CCE plot area) of sugarcane crop for out of sample districts using Spatial EBLUP (SEBLUP).

Districts	Yield	% SE	Districts	Yield	% SE
Ghaziabad	11249	12.16	G B nagar	9718	34.58
Etah	9817	17.87	Agra	9643	37.48
Badaun	9417	23.84	Balia	9640	35.98
Unnao	9416	27.46	Gazipur	9635	63.77
Lucknow	10051	24.56	Firozabad	9625	65.00
Rai Bareilly	9447	32.46	Mahamaya Nag.	9511	53.31
Farrukhabad	9519	29.87	Baharich	9468	47.92
Kannauj	9471	35.93	Chandauli	9461	48.75
Etawah	9417	38.14	Kanshiram nagar	9440	45.11
Auraiya	9416	44.83	Mainpuri	9415	44.46
Kanpur®	9389	42.12	Mirzapur	9351	43.04
Kanpur(u)	9846	44.13	Jhansi	9257	41.61
Banda	9263	47.06	Jalaun	9239	39.98
Fatehpur	9394	43.01	Chitrakut	9207	38.29
Pratapgarh	9513	42.80	Mahboba n	9191	36.50
Kaushambi	9537	41.76	Sonbhadra	9188	34.58
Allahabad	9641	40.57	Hamirpur	9186	32.56
Ambedkarnagar	9606	39.59	Lalitpur	9170	30.42
Sultanpur	9513	38.66	Aligarh	9598	27.60
Shrawasti	9327	37.84	Sidharthnagar	9528	25.17
S K Nagar	9615	41.60	Ajamgarh	9672	22.25
S R Nagar	9892	41.79			

5. CONCLUSIONS

This paper demonstrated an application of small area estimation technique to produce reliable district level estimates of crop yield using CCE supervised under ICS scheme data combined with covariates from secondary sources. Although the ICS supervised crop cutting experiments number only 30,000 in the entire country *i.e.* the sample size is very low, the collected data is of very high quality. The estimates generated using this data are expected to be relatively free from various sources of non-sampling errors. Hence, it is, recommended that wherever it is not possible to conduct adequate number of crop cutting experiments due to constraints of cost or infrastructure or both, small area estimation technique can be gainfully used to generate reliable estimates of crop yield based on a smaller sample to obtain more precise estimates than the direct survey estimates. The precision of these small area estimates can further enhance by using spatial small area estimation techniques.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the valuable comments and suggestions of the Editor and the referee. These led to a considerable improvement in the paper.

REFERENCES

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error component model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.*, **83**, 28-36.

Census of India (2011). Registrar General and Census Commissioner, New Delhi, India.

Fay, R.E. and Herriot, R.A. (1979). Estimation of income from small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74,** 269-277.

Fertilizer Statistics (2010-11). The Fertilizer Association of India, New Delhi.

Petrucci, A. and Salvati, N. (2005). Small area estimation for spatial correlation in water-shed erosion assessment. *J. Agric. Biol. Environ. Statist.*, **11(2),** 169-182.

Petrucci, A., Pratesi, M. and Salvati, N. (2005). Geographic information in small area estimation: Small area models and spatially correlated random area effects. *Stat. Trans.*, **3(7)**, 609-623.

Pfeffermann, D. (2002). Small area estimation: New Developments and directions. *Intern. Statist. Rev.*, **70**, 125-143.

Rao, J.N.K. (2003). *Small Area Estimation.* Wiley, New York.

Singh, B.B., Shukla, G.K. and Kundu, D. (2005). Spatio-Temporal models in small area estimation technique. *Survey Meth.,* **31,** 183-195.

Sud, U.C., Chandra, H. and Srivastava, A.K. (2012). Crop yield estimation at district level using Improvement of Crop Statistics Scheme data - An application of small area estimation technique. *J. Ind. Soc. Agric. Statist.*, **66(2)**, 321-326.