# EBLUP estimate of crop yield at sub-district level in Hisar district, Haryana, India using MODIS/Terra data

**P. K. Muhammed Jaslam[1,*], Manoj Kumar[1], Anurag[2] and Manjeet[2]**

[1]Department of Mathematics and Statistics, and
[2]Department of Agricultural Meteorology, CCS Haryana Agricultural University, Hisar 125 004, India

**The present study was carried out to develop improved crop yield estimates for rice and wheat crops through the Empirical Best Linear Unbiased Prediction (EBLUP) procedure via the Fay–Herriot area level model at sub-district level in Hisar district. Village-wise crop cutting data and auxiliary remote sensing data (satellite imaginaries) derived from the MODIS Vegetation Indices (MOD13Q1) version 6 were used for model construction. It is noteworthy that the coefficient of variation of the developed EBLUP estimates was below 10% for almost all sub-districts. The study revealed a significant enhancement in the efficiency of the yield estimator in comparison to the direct estimator, which recommended that with the use of remote sensing data together with crop cutting experiment data, crop yield estimates can be obtained on a smaller scale than the district using existing crop cutting experiments in the district.**

**Keywords:** Crop yield estimation, Fay–Herriot area level model, MODIS/Terra, NDVI, small area estimation.

Agricultural production is subjected to various uncertainties, hazards and unforeseen extreme climatic situations which surge the risk of agriculture production. Many threats directly affect the agricultural production, which in turn impact the economic condition of the farmers. According to the National Crime Records Bureau statistics, a total of 12,602 farmers (8,007 cultivators; 4,595 farm workers) committed suicides in 2015. For these unforeseen circumstances, many governmental and non-governmental organizations have sought to lessen the farmer's financial loss. The Pradhan Mantri Fasal Bima Yojana (PMFBY) is one such initiative of the Government of India (GoI). The insurance scheme was introduced in 2016 with the goal of providing farmers with insurance against crop losses. The impact of these initiatives is reflected in the 2018 figures, a total of 10,349 (5763 farmers/cultivators and 4586 farm workers) which are less in comparison to previous years. This is partly due to the progress made in the approach taken to measure the yield and the damage that has occurred. Earlier in India, crop yields were estimated solely based on crop cutting experiments under the national programme known as the General Crop Estimates Survey (GCES), which was performed using the survey methodology developed earlier[1,2]. Crop cutting experiments (CCEs) are conducted in the field by identifying a given area in the field, harvesting the crop in the area and weighing the yield. Every year 20 per cent districts are chosen for these experiments. The direct estimates at national as well state level are almost reliable, as the estimator's sampling error is within 5 per cent, but not true at lower levels as demonstrated in refs 3–5. 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.

The Ministry of Agriculture and Farmers Welfare, GoI has begun to use innovative technologies such as remote sensing, drones, online data transmission, artificial intelligence, modelling tools, etc. to address the problem of the reliability and speed of the CCEs. This will ensure the accurate assessment and timely payment of claims of farmers. The KISAN (C[K]rop Insurance using Space Technology and GeoiNformatics) project, as part of the use of technology in PMFBY, envisages the use of high-resolution remote sensing data from satellites and unmanned aerial vehicles to optimize crop cutting experiment planning and improve yield estimation. The government also uses satellite imagery to assess crop area, crop condition and crop yield at district level under various programmes such as Coordinated Horticulture Assessment and Geo-Informatics Management and Forecasting Agricultural Output Using Space, Agrometeorology & Land Based Observations (FASAL). In addition, an expression of interest has been voiced by GoI with a view to migrate into a technology-based yield estimation with lesser number of CCEs for the *kharif* 2019 season at gram panchayat level.

The topic of small area estimation (SAE) has gained importance in view of growing needs of micro level planning. In many SAE problems, the unit level small area model cannot always be used mainly due to inaccessible unit level data. Rao[6] also inferred that direct estimation of small sample sizes, specific to the domain can lead to

*For correspondence. (e-mail: jaslam.stat@hau.ac.in)

estimates with significant sampling error. In such circumstances, SAE is implemented under area level. The term small domain or area usually refers to the subset of a population for which due to certain data limitations accurate statistics of interest cannot be generated. Area level SAE is the most common method under SAE as it is more flexible in coalescing various information sources and identifying various error sources. The model Fay–Herriot[7] is commonly used in SAE at area level. This model links direct survey estimates of small area to area-specific auxiliary variables. In India, the Crop Acreage and Production Estimation (CAPE) project has been carried out over the last decades for various major crop statistics using satellite/remote sensing data as covariates. A detailed review of the works is given elsewhere[8,9]. To list a few related works, Patel *et al.*[10] and Hooda *et al.*[11] estimated more accurate regional wheat yields in Haryana with the aid of remotely sensed images. Singh *et al.*[12,13] utilized satellite data along with crop yield survey data to construct reliable post-stratified crop yield estimators at district level as well as small area estimators of tehsil yield in Rohtak district, Haryana.

Remote sensing perceived reflection of terrestrial vegetation can be decoded to specific environmental parameters, like productivity and vegetation indices[14]. These are valuable parameters for various applications in science, policy and land management[15]. The empirical Normalized Difference Vegetation Index (NDVI) models have often been used in the literature to forecast crop yields due to their simplicity and linkage to photosynthesis activity[16]. In this study, satellite data were computed to test its association with crop yield (rice and wheat) followed by the Empirical Best Linear Unbiased Prediction (EBLUP) of sub-district crop yield via the Fay–Herriot area level model using selected satellite data as an auxiliary variable.

## 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 sq. km. The net sown area of the district is 4040 sq. km with 178.2 per cent cropping intensity; *kharif* rice and *rabi* wheat are the major economic crop which occupy up to 700 and 2240 sq. km of the total area respectively (Figure 1). 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-wise rice and wheat crop yield data for 2017–2018 were collected from the Department of Agriculture and FW, Hisar. The present study aimed to incorporate spectral indices in statistical models to assess the spatial pattern of yield of different crops. Various space agencies now provide remote sensing (indices) data as an open source for academic and research applications. MODIS (NASA, USA) is also one of those data sets commonly used globally in geospatial science. MOD13Q1 provides 16-day global spatial datasets with 250 m resolution as a gridded level-3 product in the Sinusoidal projection. MODIS data is more preferred for regional vegetation monitoring due to its improved swath and repetitiveness, which in turn allows large areas to be covered on the same date. The spectral indices (NDVI) are dependent on the spectral resolution, i.e. the number of spectral bands recorded by the sensor. MODIS captures radiance across 36 bands, helping to obtain a more accurate NDVI estimate. For the present study, MOD13Q1 v006 satellite images were downloaded from the US Geological Survey website for various crop stage period of the rice and wheat crops.

Crop masking is a process of stratifying a region into different crop types, which is an important step in developing earth observations (EO)-based yield assessment and forecasting models. However, one of the difficulties in monitoring and forecasting crop yields using RS images is the availability of timely seasonal detailed crop type masks that can be used to identify the crop of interest prior to the end of the growing season. A general cropland mask is often used to distinguish cropped areas from other types of land use rather than a crop-specific mask[17–19]. For example, Maselli *et al.*[20] used the NDVI threshold to isolate cropland pixels of interest in the Sahel region. On the same pattern, here in the present study crop pixels were isolated by thresholding the NDVI values. Crop pixel so classified were verified on ground at some sites as can be seen in Figure 2.

Satellite data was interpreted and for minimum NDVI value, maximum NDVI and mean NDVI value of the crop in respective sub-district along with NDVI value of the sub-districts as a whole at the time of maximum flowering/heading, i.e. for rice 13 August 2017 and for wheat 18 February 2018. Data set of the annual integral of NDVI (iNDVI) averaged over the different crop stages from seedling stage to the grain filling stage of the respective crops was also prepared. The vegetation indices were classified based on the NDVI values range (–1 to +1) into 4 classes using the threshold range technique, i.e. dense vegetation (≥0.6), moderate vegetation (0.4–0.6), sparse vegetation (0.2–0.4) and non-vegetation cover (≤0.2). Figure 3 represents the NDVI images for Hisar district 2017–18.

### *Statistical approach*

*Selection of suitable covariate*. For model building and diagnosis, 10 K thumb rule suggests selecting $k$ number of independent variable (in this case covariates) when having 10 $k$ observations[21]. So here, we can take a single covariate at a time. Karl Pearson's product moment correlation coefficient is worked out to pick the finest

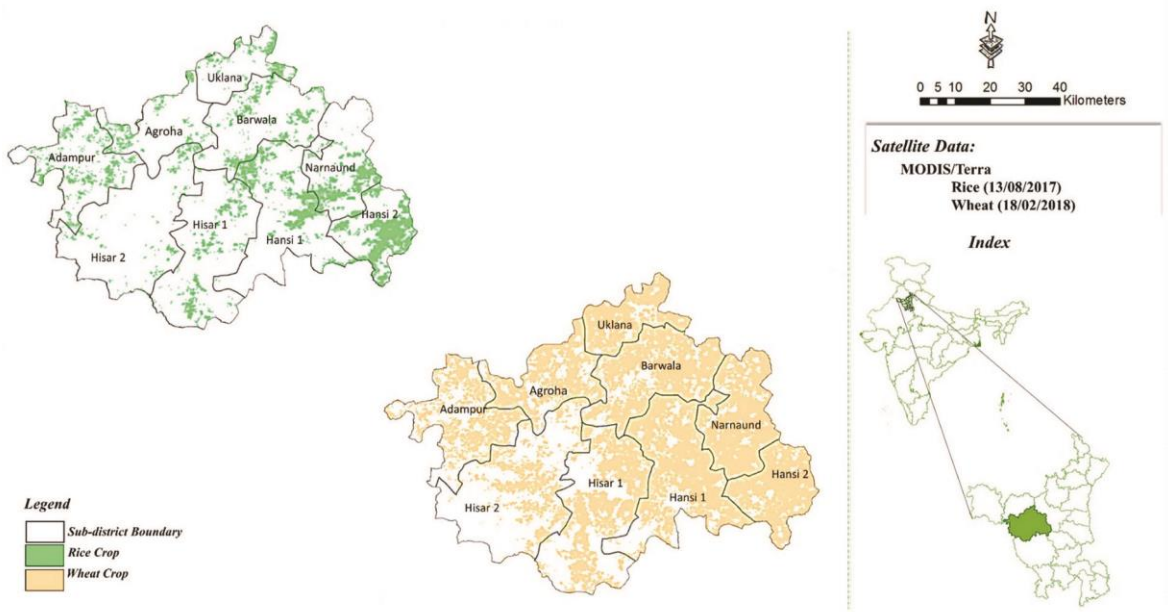


**Figure 1.** Crop map for rice and wheat (2017–2018).

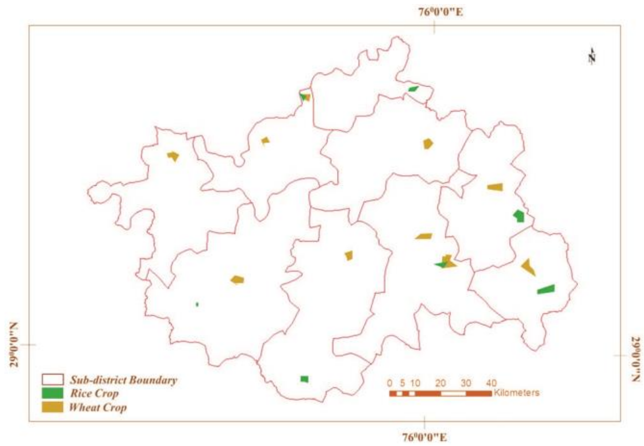


**Figure 2.** Sites for crop identification/validation.

covariate for yield data among the NDVI data set. Correlation coefficient is denoted by $r$, working formula for $r$ is given by

$$r = \frac{D\sum XY - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left(D\sum X^2 - \left(\sum X\right)^2\right)\left(D\sum Y^2 - \left(\sum Y\right)^2\right)}}. \quad (1)$$

In the above expression, $X$ and $Y$ denote the measurements on variables $X$ and $Y$ and $D$ is the number of pairs of observations, i.e. number of sub districts. The variables are said to be positively correlated if $r$ is positive and negatively correlated if $r$ is negative.

*Area level random effect model to derive EBLUP estimate of crop yield.* When direct estimation is not feasible, alternative model-based methods for developing small area estimates must be used. One common approach uses mixed (random) effect models for the estimation of small areas[7,22]. The mixed effects model includes a fixed effects part and a random effects part, the latter accounting for area variations beyond that explained by the auxiliary variables included in the fixed model part[23].

Let $y_i$ denote the observed direct estimate of the unobservable population-level quantity (e.g. average yield) $Y_i$ of variable of interest $y$ for area (or sub district) $i$. Let $X_i$ be the known auxiliary variable, obtained from NDVI data set, related to the population mean $Y_i$. The area specific two-stage model given by Fay and Herriot[7] is described as

$$y_i = Y_i + e_i \text{ and } Y_i = X_i^{\mathrm{T}}\beta + u_i. \quad (2)$$

The first stage in this model accounts for the sampling variability of the direct estimates $y_i$ of true area means $Y_i$ and the second stage links the true area means $Y_i$ to a known covariate $X_i$. Alternatively, model (2) may be expressed as

$$y_i = X_i^{\mathrm{T}}\beta + u_i + e_i;\ i = 1, 2 \ldots D. \quad (3)$$

Here $\beta$ is a vector of unknown fixed effect parameter, $u_i$'s are independent and identically distributed normal random errors with $E(u_i) = 0$ and $\mathrm{var}(u_i) = \sigma_u^2$, and $e_i$'s are independent sampling errors normally distributed with $E(e_i|Y_i) = 0$, $\mathrm{var}(e_i|Y_i) = \sigma_i^2$. The two errors are independent of each other within and across areas. Usually $\sigma_i^2$, is known while $\sigma_u^2$ is unknown and it has to be estimated from the data. Estimation methods of $\sigma_u^2$ include

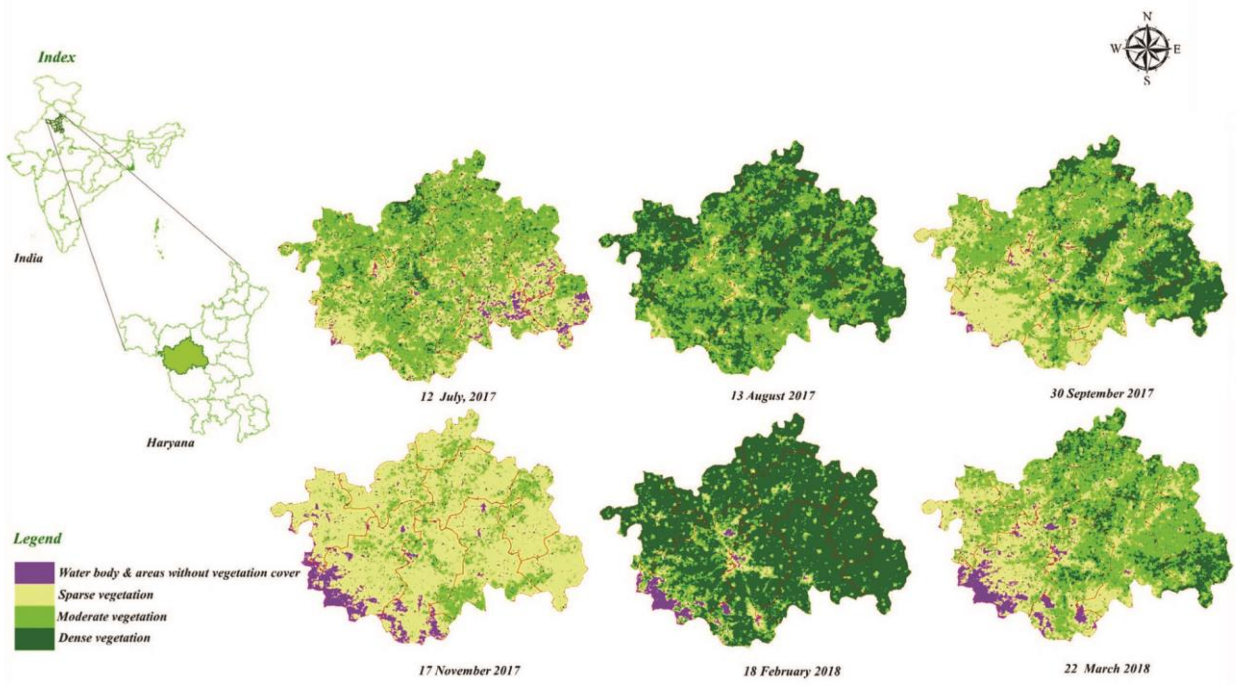


**Figure 3.** NDVI images (MODIS/Terra) for Hisar district (2017–2018).

Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML) under normality[9]. Let $\hat{\sigma}_u^2$ denote estimate of $\sigma_u^2$. Then under model (3), the EBLUP of $Y_i$ given by

$$\hat{Y}_i^{\text{EBLUP}} = X_i^T \hat{\beta} + \hat{u}_i = \hat{\gamma}_i y_i + (1-\hat{\gamma}_i) X_i^T \hat{\beta}, \tag{4}$$

where $\hat{\gamma}_i = (\hat{\sigma}_u^2 / (\sigma_i^2 + \hat{\sigma}_u^2))$ and $\hat{\beta}$ is the generalized least square estimate of $\beta$. It may be noted that $\hat{Y}_i^{\text{EBLUP}}$ is a linear combination of direct estimate and the model based regression synthetic estimate with weight $\hat{\gamma}_i$. Prasad and Rao[24] suggested an approximately model unbiased (i.e. with bias of order $o(1/D)$) estimate of mean squared error (MSE) of the EBLUP (3) given by

$$\text{MSE}\,(\hat{Y}_i^{\text{EBLUP}}) = g_{1i}(\sigma_u^2) + g_{2i}(\sigma_u^2) + 2g_{3i}(\sigma_u^2)\,\text{var}(\hat{\sigma}_u^2), \tag{5}$$

where $g_{1i}(\sigma_u^2) = \hat{\gamma}_i \sigma_i^2$, $g_{2i}(\sigma_u^2) = (1-\hat{\gamma}_i)^2 X_i^T \,\text{var}(\hat{\beta}) X_i$, and $g_{3i}(\sigma_u^2) = \left\{\frac{\hat{\sigma}_u^4}{(\sigma_i^2+\hat{\sigma}_u^2)^3}\right\} \text{var}(\hat{\sigma}_u^2)$ with

$$\text{var}(\hat{\sigma}_u^2) \approx 2D^{-2} \sum_{i=1}^{D} (\sigma_i^2 + \hat{\sigma}_u^2)^2.$$

### *Analytical software's used*

ArcGIS and ENVI softwares were used for processing and analysing geospatial imagery and Fay–Herriot model-based EBLUP estimator was developed using SAE package[25] in R software interface.

## Results and discussion

The first objective of the study was to identify the best auxiliary variables for small-scale crop yield estimation. Among the vegetation indices (Supplementary Tables 1 and 2), the maximum iNDVI and the maximum NDVI of the rice crop in the respective sub-district were found to have a significant negative correlation ($r = -0.79$ and $-0.65$ respectively) with the yield of rice. The other indices had a negligible correlation with rice yield. Drastic changes in NDVI values have occurred, from tillering to filling stage. NDVI's average value increased from tillering to jointing, then decreased to filling stage. Maximum NDVI's higher value reflects low grain filling rate, hence less yield. Similar results were reported by Liu *et al.*[26] and Zhao *et al.*[27], where they found the highest N amount/maximum NDVI value associated with reduced grain yield. For instance, in wheat crop, the mean NDVI value and mean iNDVI value in the respective sub-district were found to have a comparatively better and positive correlation ($r = +0.39$ and $+0.51$ respectively) with wheat yield. Chandel *et al.*[28] also observed that NDVI value is positively correlated with grain yield and can be used to predict wheat yield. These variables have therefore been used as covariate data for the respective crops. Figure 4 shows correlogram of the analysis.

As a percentage estimate, the CVs show the sample's unpredictability. While there are no globally accepted tables for determining which CV is too small, estimates are considered incorrect for large CVs. Tables 1 and 2 show the direct estimate and the different EBLUP estimates developed using selected covariates along with the

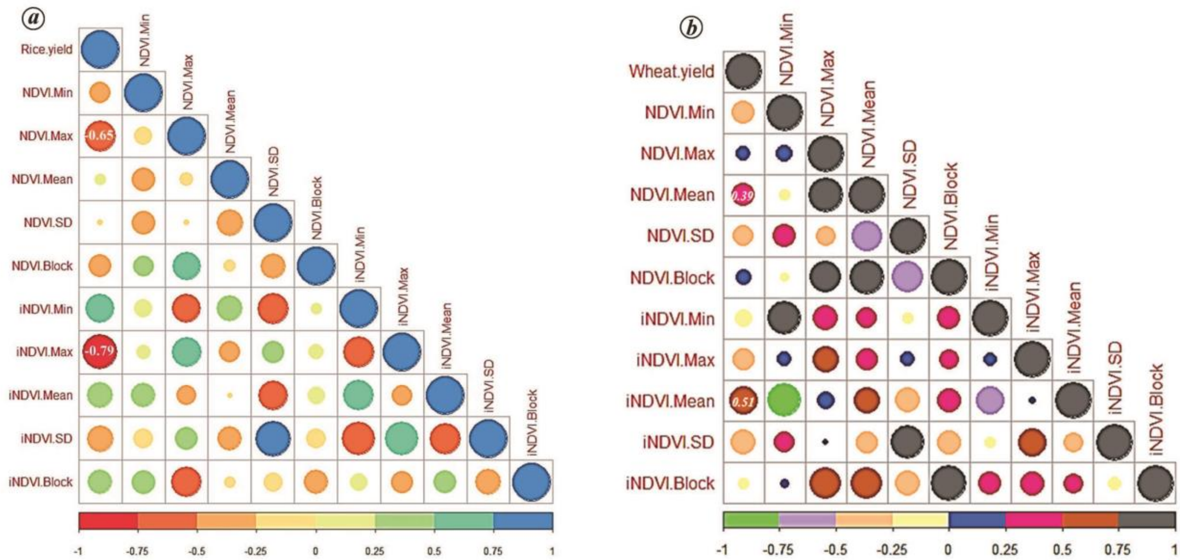


**Figure 4.** Correlogram of (***a***) rice yield and (***b***) wheat yield satellite spectral data set.

**Table 1.** Block level yield estimates (kg/ha) of rice crop for the Hisar district (2017–2018)

| | Rice crop | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Block (Sub-district) | Sample size | Direct estimate | CV (%) | EBLUP estimate ($y_i$ ~ NDVI Max.) | CV (%) | RMSE | EBLUP estimate ($y_i$ ~ iNDVI Max.) | CV (%) | RMSE |
| Adampur | 14 | 3420.91 | 14.84 | 3048.09 | 10.07 | 307.09 | 3818.48 | 9.62 | 367.28 |
| Agroha | 17 | 2849.72 | 10.18 | 3092.30 | 9.00 | 278.18 | 3018.37 | 7.76 | 234.37 |
| Barwala | 36 | 3068.55 | 8.89 | 3042.70 | 8.09 | 246.22 | 2834.20 | 7.80 | 221.02 |
| Hisar1 | 36 | 2662.48 | 9.03 | 2687.11 | 8.27 | 222.21 | 2511.55 | 8.78 | 220.62 |
| Hisar2 | 11 | 3451.85 | 5.58 | 3325.93 | 5.76 | 191.51 | 3325.06 | 6.09 | 202.47 |
| Uklana | 12 | 2538.80 | 13.51 | 2565.18 | 10.64 | 273.01 | 2493.60 | 9.51 | 237.07 |
| Narnaund | 31 | 1827.91 | 12.76 | 2070.24 | 10.77 | 222.91 | 2127.52 | 10.40 | 221.34 |
| Hansi1 | 36 | 2620.21 | 8.55 | 2577.83 | 8.33 | 214.86 | 2695.14 | 7.68 | 207.03 |
| Hansi2 | 22 | 2615.90 | 7.81 | 2534.32 | 8.06 | 204.24 | 2574.73 | 7.81 | 201.05 |

**Table 2.** Block level yield estimates (kg/ha) of wheat crop for the Hisar district (2017–2018)

| | Wheat crop | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Block (Sub-district) | Sample size | Direct estimate | CV (%) | EBLUP ($y_i$ ~ NDVI Mean) | CV (%) | RMSE | EBLUP ($y_i$ ~ iNDVI Mean) | CV (%) | RMSE |
| Adampur | 24 | 4863.61 | 8.58 | 4697.84 | 6.69 | 314.11 | 4813.19 | 6.32 | 304.22 |
| Agroha | 23 | 4866.90 | 6.21 | 4941.07 | 6.34 | 313.04 | 4966.83 | 6.29 | 312.47 |
| Barwala | 38 | 5077.00 | 9.71 | 5037.23 | 3.81 | 191.76 | 5076.50 | 4.01 | 203.58 |
| Hisar1 | 46 | 4902.53 | 6.71 | 4975.01 | 5.74 | 285.51 | 4927.90 | 6.02 | 296.85 |
| Hisar2 | 39 | 4773.34 | 13.95 | 4505.52 | 8.49 | 382.66 | 4778.34 | 5.87 | 280.32 |
| Uklana | 12 | 5304.65 | 4.31 | 5167.33 | 8.19 | 423.21 | 5079.92 | 8.17 | 414.90 |
| Narnaund | 31 | 5497.33 | 8.81 | 5263.49 | 5.03 | 264.77 | 5247.68 | 6.05 | 317.29 |
| Hansi1 | 40 | 4825.95 | 9.26 | 4912.79 | 4.49 | 220.78 | 4926.10 | 4.61 | 227.29 |
| Hansi2 | 22 | 4138.58 | 12.58 | 5065.51 | 3.66 | 185.18 | 5169.64 | 4.79 | 247.50 |

CV and for the crop yield. The estimated CVs for model-based estimates are much more precise than direct estimates. Similar results have been reported elsewhere[29–31].

Ensuing figures present the direct estimate and EBLUP estimators of rice yield (Figure 5 *a*), direct estimate and EBLUP estimators of wheat yield (Figure 5 *b*), CVs of direct and EBLUP estimators for rice yield (Figure 6 *a*)

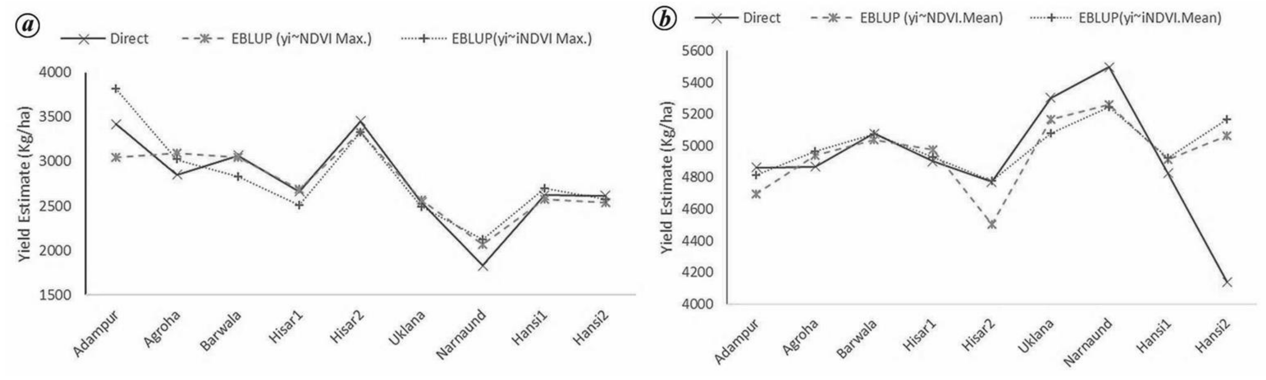


**Figure 5.** Sub-district wise direct and EBLUP estimators of (***a***) rice yield (kg/ha) and (***b***) wheat yield (kg/ha).

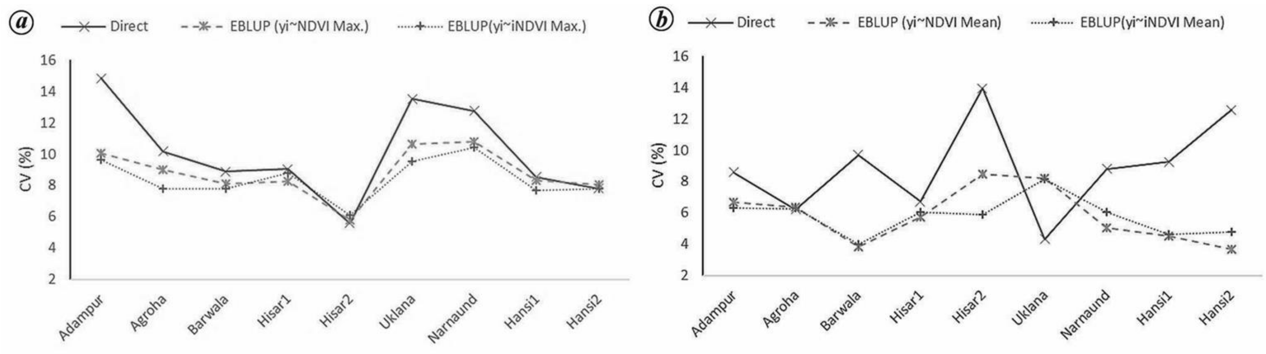


**Figure 6.** Sub-district wise CVs for the direct and EBLUP estimators of (***a***) rice yield and (***b***) wheat yield.

and wheat yield (Figure 6 *b*). In some of the cases direct estimators were having CVs around 5 per cent and in those sub-districts EBLUP estimation found to be futile. However, for other sub-districts the direct estimators were not so efficient and gains due to EBLUP estimation were substantial. These results clearly illustrate spatial unequal distribution of rice and wheat yields in the different sub-districts of the Hisar district.

The most adequate model among the developed competing small area models is accessed through goodness of fit measures such as minimal root mean square error (RMSE), log likelihood, Akaike information criterion (AIC), Bayesian information criterion (BIC) and Kashyap information criterion (KIC). Table 3 reveals that the model developed using the maximum iNDVI as an auxiliary variable is optimal for estimating the yield of rice, while the area-level model with mean NDVI as auxiliary variable is optimal for estimating the yield of wheat in the study area.

### *Bias diagnostic*

A comparison of the best model-based and the direct survey results on their degree of extremity are determined by the application of bias diagnostic. Besides, if the direct estimates are unbiased, their regression will be linear to the true values and correspond to the identity line. When model-based estimates are close to true values, the regression of direct estimates will be analogous to model-based estimates[28,32]. The direct estimates on $Y$-axis and model-based estimates on the $X$-axis and look for divergence of regression line from $Y = X$ were plotted. Figure 7 shows the bias scatter plots of the direct estimates against the model-based estimates. The plots show that the direct rice yield estimator is more or less unbiased, and the model estimates are also less extreme compared to the direct estimates.

## Conclusion and future thrust

The SAE techniques described earlier were applied to the rice and wheat yield data of the Hisar district of Haryana, India. Empirical results showed that the sub-district crop production estimates obtained by the use of remote sensing data together with survey data were reasonably good. It should be noted that the coefficient of variation of the EBLUP estimates was below 10% for almost all sub-districts. These estimates can be useful for the resource allocation and for making of agricultural policy decisions. Such yield estimates are also helpful in identifying sub-districts with lower crop yield to draw planner's attention. Furthermore, SAE offers estimates for those

**Table 3.** Model comparison based on goodness of fit criterion

| | Small area model | Log-likelihood | AIC | BIC | KIC |
|---|---|---|---|---|---|
| Rice | EBLUP ($y_i$ ~ NDVI Max.) | –65.795 | 137.590 | 138.182 | 140.590 |
| | EBLUP ($y_i$ ~ iNDVI Max.) | –64.154 | 134.308 | 134.900 | 137.309 |
| Wheat | EBLUP ($y_i$ ~ NDVI Mean) | –64.604 | 135.208 | 135.800 | 138.208 |
| | EBLUP ($y_i$ ~ iNDVI Mean) | –65.153 | 136.305 | 136.897 | 139.305 |

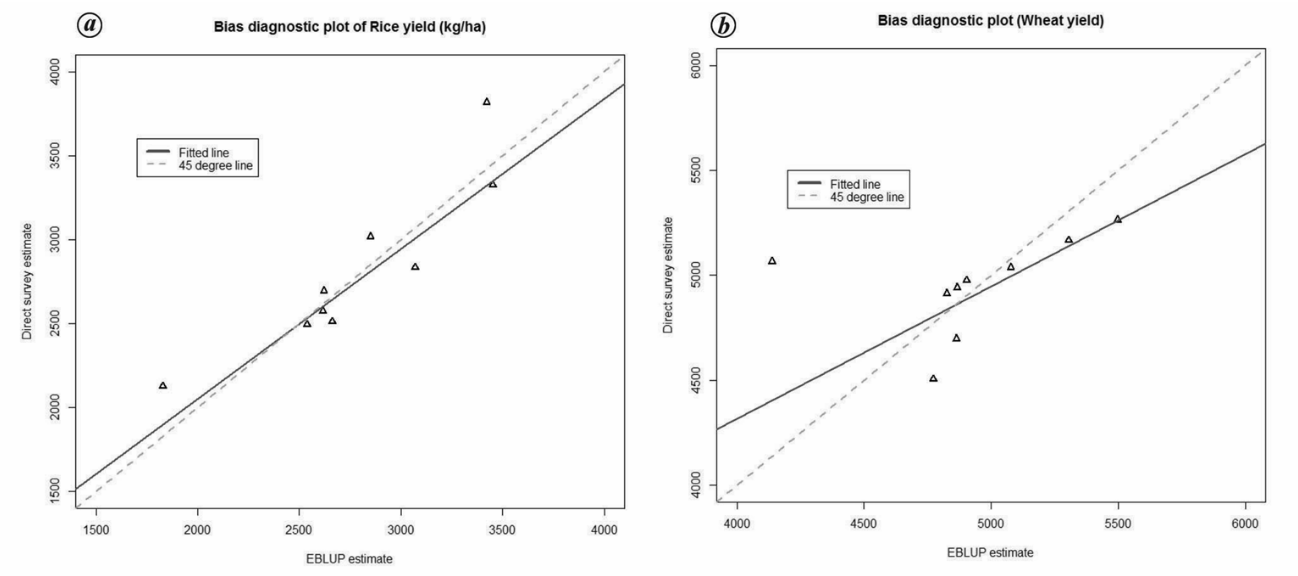


**Figure 7.** Bias diagnostic plot for (***a***) rice yield estimation and (***b***) wheat yield estimation.

districts where there is no sample information under ICS, so direct estimates cannot be determined. Therefore, wherever a sufficient number of CCEs cannot be performed due to cost or infrastructure constraints or both, the SAE technique may be used to produce accurate crop yield estimates based on a smaller sample. Also, spatial association (or spatial dependence) effects can be used to boost disaggregate-level estimates.

The GoI is currently placing a lot of emphasis on micro-level planning. Generating the gram panchayat level estimates are crucial in view of agricultural policy planning in the country. To the best of our knowledge, no studies are reported on the application of unit level SAE in Indian agricultural data so far. Further, different robust method of small area estimation approaches have also been developed recently[33,34], which is useful for limiting the influence of outliers on small area estimators. These methods can be widely adapted to other data sets from different districts and to several crops for the generation of yield estimates at micro level.

1. Mahalanobis, P. C., Sample surveys of crop yields in India. *Sankhyā*, 1946, **7**, 269–280.
2. Sukhatme, P. V. and Agarwal, O. P., Crop-cutting survey on wheat by the random sampling method, Indian Council of Agricultural Research, New Delhi, 1947.
3. Tikkiwal, B. D. and Tikkiwal, G. C., Small area estimation in India – Crop yield and acreage statistics. Paper presented at the International Conference on Agricultural Statistics–2000, Washington, DC, USA, 1998.
4. Tikkiwal, G. C. and Ghiya, A., A generalized class of composite estimators with application to crop acreage estimation for small domains. *Stat. Trans*., 2004, **1**, 697–711.
5. Tikkiwal, G. C. and Ghiya, A., A generalized class of synthetic estimators with application to crop acreage estimation for small domains. *Biom. J*., 2000, **42**, 865–876.
6. Rao, J. N. K., *Small Area Estimation*, John Wiley, New York, 2003, p. 344.
7. Fay, R. E. and Herriot, R. A., Estimates of income for small places: An application of james-stein procedures to census data. *J. Am. Stat. Assoc*., 1979, **74**, 269–277.
8. Navalgund, R. R., Parihar, J. S., Ajai and Nageshwar Rao, P. P., Crop inventory using remotely sensed data. *Curr. Sci*., 1991, **61**, 162–171.
9. Dadhwal, V. K., Singh, R. P., Dutta, S. and Parihar, J. S., Remote sensing-based crop discrimination and area estimation: a review of Indian experience. *Trop. Ecol*., 2002, **43**, 102–122.
10. Patel, N. R., Bhattacharjee, B., Mohammed, A. J., Tanupriya, B. and Saha, S. K., Remote sensing of regional yield assessment of wheat in Haryana, India. *Int. J. Remote Sens*., 2006, **27**, 4071–4090.
11. Hooda, R. S., Yadav, M. and Kalubarme, M. H., Wheat production estimation using remote sensing data: An Indian experience. In Proceedings of ISPRS Archives XXXVI-8/W48 Workshop on Remote Sensing Support to Crop Yield Forecast and Area Estimates. Stresa, Italy, 2006.
12. Singh, R., Semwal, D. P., Rai, A. and Chhikara, R. S., Small area estimation of crop yield using remote sensing satellite data. *Int. J. Remote Sens*., 2002, **23**, 49–56.

13. Singh, R., Goyal, R. C., Saha, S. K. and Chhikara, R. S., Use of satellite spectral data in crop yield estimation surveys. *Int. J. Remote Sens*., 1992, **13**, 2583–2592.
14. Jaafar, H. H. and Ahmad, F. A., Crop yield prediction from remotely sensed vegetation indices and primary productivity in arid and semi-arid lands. *Int. J. Remote Sens*., 2015, **36**, 4570–4589.
15. Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M. and Hashimoto, H., A continuous satellite-derived measure of global terrestrial primary production. *BioScience*, 2004, **54**, 547–560.
16. Reeves, M. C., Zhao, M. and Running, W., Usefulness and limits of MODIS GPP for estimating wheat yield. *Int. J. Remote Sens.*, 2005, **26**, 1403–1421.
17. Rasmussen, M. S., Operational yield forecast using AVHRR NDVI data: Reduction of environmental and inter-annual variability. *Int. J. Remote Sens.*, 1997, **18**, 1059–1077.
18. Genovese, G., Vignolles, C., Negre, T. and Passera, G., A methodology for a combined use of normalized difference vegetation index and CORINE land cover data for crop yield monitoring and forecasting. A case study on Spain. *Agronomie*, 2001, **21**, 91–111.
19. Kogan, F., Guo, W., Yang, W. and Shannon, H., Space-based vegetation health for wheat yield modeling and prediction in Australia. *J. Appl. Remote Sens*., 2018, **12**, 1–12.
20. Maselli, F., Romanelli, S., Bottai, L. and Maracchi, G., Processing of GAC NDVI data for yield forecasting in the Sahelian region. *Int. J. Remote Sens.*, 2000, **21**, 3509–3523.
21. Troutt, M. D., Regression, 10 k rule of thumb for. In *Encyclopedia of Statistical Sciences* (eds Kotz, S. *et al.*), John Wiley, New Jersey, 2004, p. 9686.
22. Battese, G. E., Harter, R. M. and Fuller, W. A., An error component model for prediction of county crop areas using survey and satellite data. *J. Am. Stat. Assoc.*, 1998, **83**, 28–36.
23. Sud, U. C., Bhatia, V. K., Chandra, H. and Srivastava, A. K., Crop yield estimation at district level by combining improvement of crop statistics scheme data and census data. Paper presented at 4th meeting, Wye city group on statistics on rural development and agricultural household income, Rio de Janeiro, Brazil, 2011.
24. Prasad, N. G. N. and Rao, J. N. K., The estimation of the mean squared error of small area estimators. *J. Am. Stat. Assoc*., 1990, **85**, 163–171.
25. Molina, I. and Marhuenda, Y., SAE: An R package for small area estimation. *R. Journal*, 2015, **7**, 81–98.
26. Liu, K., Li, Y., Hu, H., Zhou, L., Xiao, X. and Yu, P., Estimating rice yield based on normalized difference vegetation index at heading stage of different nitrogen application rates in southeast of china. *J. Environ. Agric. Sci*., 2015, **2**, 1–9.
27. Zhao, L. M., Wu, L. H., Li, Y. S., Lu, X. H., Zhu, D. F. and Uphoff, N., Influence of the system of rice intensification on rice yield and nitrogen and water use efficiency with different N application rates. *Exp. Agric*., 2009, **45**, 275–286.
28. Chandel, N. S., Tiwari, P. S., Singh, K. P., Jat, D., Gaikwad, B. B., Tripathi, H. and Golhani, K., Yield prediction in wheat (*Triticum aestivum* L.) using spectral reflectance indices. *Curr. Sci*., 2019, **116**, 272–278.
29. Chandra, H., Salvati, N. and Sud, U. C., Disaggregate-level estimates of indebtedness in the state of Uttar Pradesh in India – an application of small area estimation technique. *J. Appl. Stat*., 2011, **38**, 2413–2432.
30. Erciulescu, A. L., Cruze, N. B. and Nandram, B., Model-based county level crop estimates incorporating auxiliary sources of information. *J. R. Statist. Soc. A.*, 2019, **182**, 283–303.
31. Srivastava, A. K., Small area estimation – A perspective and some applications**.** *J. Ind. Soc. Agril. Statist.*, 2007, **61**, 295–309.
32. Ambler, R., Caplan, D., Chambers, R., Kovacevic, M. and Wang, S., Combining unemployment benefits data and LFS data to estimate ILO unemployment for small areas: An application of a modified Fay–Herriot Method. In Proceedings of the International Association of Survey Statistician. Meeting of the ISI, Seoul, 2001.
33. Sinha, S. K., Robust small area estimation in generalized linear mixed models. *METRON*, 2019, **77**, 201–225.
34. Sugasawa, S. and Kubokawa, T., Small area estimation with mixed models: a review. *Jpn. J. Stat. Data Sci*., 2020; https://doi.org/10.1007/s42081-020-00075-y.

ACKNOWLEDGEMENTS. We thank Department of Agriculture & FW, Hisar, for providing village-wise crop yield data for the study. We also thank the Department of Agricultural Meteorology, College of Agriculture, CCS HAU, Hisar for providing the Geoinformatics laboratory facility. We also thank anonymous reviewers for valuable comments.

Received 25 March 2020; revised accepted 12 May 2020

doi: 10.18520/cs/v119/i12/1982-1989