International Journal of Science, Environment and Technology, Vol. 14, No 3, 2025, 345 – 364

ISSN 2278-3687 (O)
2277-663X (P)

ESTIMATION OF CROP PRODUCTION FOR SMALL DOMAIN USING WEATHER VARIABLES

[1]Ashish Shukla, [2]Manoj Kumar Sharma and [3]B.V.S. Sisodia

[1,3]Department of Agricultural Statistics,
Narendra Deva University of Agriculture &Technology,
Kumarganj, Faizabad, Uttar Pradesh, India
[2]Department of Statistics, Mathematics and Computer Science, SKN College of Agriculture,
Sri Karana Narendra Agriculture University, Jobner, Jaipur, Rajasthan, India
E-mails: [1]askshukl@gmail.com, [2]mksharma.stats@sknau.ac.in (*Corresponding Author*),
[3]bvssisodia@gmail.com

Abstract: The small area statistics have been derived from large scale surveys where survey statistician is interested in estimating the population parameters and also the parameters of sub-populations. Crop production statistics for a small area Block or Panchayat are now essential in view of decentralized planning process at micro-level in India. Generally, estimates of crop production or yield through crop cutting experiments are being reported at district level and these estimates are aggregated at state and country level. Singh *et al.* (2012) proposed a predictive approach of estimation of crop-production at block level using the multiple regression model fitted at district level. The same methodology has been adopted and further improved by Sharma *et al* (2015), and Sharma and Sisodia (2016). In the present paper, an attempt has been made to examine the various options of using various factors and weather variables in the regression model for better prediction of block level estimate of crop-production. A proposed methodology has been used for the more precise estimates through the use of weather variables as auxiliary variables. It is assumed that the auxiliary information available at district level is also available at block level. An empirical study on time series wheat production data of Sultanpur district of the State of Uttar Pradesh, India shows that the inclusion of weather variables in the model as an explanatory variable brought little bit reduction in the percent standard errors of the estimators. The estimator based on least squares adjustment has performed better in most of the cases than other estimators in terms of percent standard error.

Keywords: Block level estimate, Crop cutting experiment, crop production, prediction, regression model.

1. INTRODUCTION

The small area statistics have been derived from large-scale surveys where the survey statistician is not only interested in estimating the population parameters such as mean or total but also the parameters of sub-populations, called domain. A special feature of domain studies is that survey is not planned specifically for estimating domain parameters due to high resources requirement but these are developed with the help of information obtained in the large scale surveys. In the context of agriculture surveys for crop- statistics in India, the

*Received May 12, 2025 * Published June 2, 2025 * www.ijset.net*

estimates of crop production or yield are being reported at district level through scientifically designed Crop- Cutting Experiments (CCEs) under the scheme of General Crop Estimation Survey (GCES) and these estimates are aggregated at state and country level. The estimates of crop production at Block or Panchayat level are in great demand in recent times by the planners for policy formulation in the context of de-centralized planning process at micro level in India. If reasonably precise estimates of crop production are required for Block or Panchayat level then the number of CCEs is expected to increase enormously. Thus, conducting requisite number of area specific CCEs is neither operationally feasible is economically viable. An attempt, quite earlier, was made by Panse *et al.* (1966) to develop some methodologies for estimating the crop yield at block level using double sampling approach, but it could not succeed due to certain physical constraints. Stasny *et al.* (1991) have made use of a regression model to predict wheat production at county level in USA. They, in fact, developed the multiple regression models to exhibit best possible relationship between farm production and some predictor variables related with farm production for predicting wheat production at county level. Sisodia and Singh (2001) advocated a scale down approach which involved scaling down of available district level estimate of crop production to block level. Their approach did not require any additional CCEs. They attempted to establish a best possible functional relationship between time series data on crop production and its related auxiliary variables at district level, which was used to develop block level estimates of crop production assuming that the data on the same auxiliary variables were also available at block level. Sharma *et al*. (2004) developed estimators of crop yield at panchayat level using eye estimate and farmers' appraisal within the frame of design-based sample survey. Sisodia and Chandra (2012) and Sisodia and Singh (2012) have recently made further attempts to improve the estimates for block level crop-production following the approach of Sisodia and Singh (2001). Singh *et al*. (2012) proposed a predictive approach for block level estimate of crop production by fitting multiple regression model between time-series data on crop yield and related auxiliary variables.

This study investigates whether the methodologies proposed by Singh et al. (2012) can be effectively utilized to optimize the use of weather variables and other factors influencing the dependent variable y in a regression model, thereby achieving the most accurate predictions of block-level crop production. In other worlds, it can be treated as a problem of choice of explanatory variables in the model. Two options are plausible. First, one used by Singh et *al*. (2012) where available set of the auxiliary variables is used as explanatory variables.

Second, use proposed methodology of the weather variables and other factor in the regression model.

2. REVIEW OF LITERATURE

The early work in the small area estimation can be dated back to 1966 when Panse *et al.* (1966) and Singh (1968) examined the feasibility of using double sampling for estimation of yield at the block level. The technique consisted in selecting a large sample of villages and fields from the block for eye estimation of yield prior to the harvest and combining it with results of crop cutting experiments conducted on a sub-sample for obtaining estimates of average yield at the block level. Gonzalez (1973) described synthetic estimates as follows: "An unbiased estimate is obtained from a sample survey from a large area; when this estimate is used to derive estimates for sub-areas under assumption that the small areas have the same characteristics as the large area, and identify these estimates as synthetic estimates". **Srivastava *et al.*** (1998) have given an overview of small area estimation technique in agriculture. In fact, in agriculture the crop yield at block level are not available and a method has been suggested by them to estimate the crop yield at block level through simulation process. However, this process requires yield at crop cutting experiments and other related information at field level in order to classify the population into homogeneous groups. Such information are, however, not readily available and even sometimes very difficult to get it from the sources responsible for it. **Battese *et al.* (1988)** proposed and applied the nested error regression model for estimation of corn and soybean in Iowa State and showed that the mean is sum of a fixed component, involving unknown parameters to be estimated and a random component to be predicted and it is called Best Linear Unbiased Predictor. They also defined variance-component estimators in the nested-error model and obtained generalized least square estimators of the parameters of the linear model. Beside this, they constructed an estimator of the variance of the error in the predictor, including terms arising from the estimation of the parameters of the model. **Rao (2004)** has discussed application of small area estimation like direct, indirect and composite method of estimation in agriculture. **Chambers and Tzavidis (2006)** described a new approach to small area estimation that is based on modelling quantile-like parameters of the conditional distribution of the target variable given the covariates. **Jiang and Lahiri (2006)** has given an excellent overview and appraisal of mixed model prediction in the context of small area estimation. **Chandra *et al.* (2007)** investigated the SAE based on linear models with spatially correlated small area effects where the neighbourhood structure is described by a contiguity matrix. Such models allow efficient use of spatial auxiliary information in SAE. In particular, they use simulation studies to compare the performances of model-based direct estimation (MBDE) and empirical best linear unbiased prediction (EBLUP) under such models. **Salvati *et al*. (2007)** investigated the use of GWR in small area estimation based on the M-quantile GWR model described in their study allows modeling of between area variability without the need to explicitly specify the area-specific random components of the model and this model does not explicitly depends on any particular small area geography, and so can be easily adapted to different geographies as the need arises. **Pratesi and Salvati (2008)** proposed the small area indirect estimators under area level random effect models when only area level data are available and the random effects are correlated. The performance of the Spatial Empirical Best Linear Unbiased Predictor (SEBLUP) has explored with a Monte Carlo simulation study on lattice data and it was applied to the results of the sample survey on Life Conditions in Tuscany (Italy). **Chandra (2009)** estimated the small area proportions under unit level spatial models. They investigate Small Area Estimation based on generalized linear mixed model (GLMM) with spatially correlated random area effects where the neighbourhood structure is described by a contiguity matrix.

Brief review of prediction approach of Singh *et al.* (2012)

Singh *et al.* (2012) proposed the following multiple regression linear model between the crop yield and auxiliary variables at district level.

$$Y_i=\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\dots+\beta_p X_{ip}+\varepsilon_i \tag{2.1}$$

where Y_i is the crop yield in the i^{th} year, (i= 1, 2, …n), X_{ij} is the value of j^{th} predictor (j=1, 2, … , P) in the i^{th} year, $\beta' = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)$ is vector of unknown parameters of the model and ε_i is error term. It is also assumed that ε_i's follow independently normal distribution with mean 0 and variance σ^2. Let the model (2.1) fitted with the data at district level by least square technique be denoted as

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip} \tag{2.2}$$

where $\hat{\beta}$ is the least square estimate of β and $\hat{Y}_i$ is the estimated value of Y_i for corresponding values of X_{ij}'s. They used the (2.2) directly at Block level to predict the Z_q, the yield at block q (q=1, 2, …Q). Let $\hat{Y}_q$ be the predicted value of the crop yield Y_q during a particular year of interest is given by

$$\hat{Y}_q = \hat{\beta}_0 + \hat{\beta}_1 x_{q1} + \dots + \hat{\beta}_p x_{qp} \tag{2.3}$$

where X_{qj}'s; j= 1, 2, …. p, are values of predictors in the q^{th} Block. Therefore an unbiased estimator of Z_q, the crop production at q^{th} Block is given by

$$\hat{Z}_q = \delta_q \hat{Y}_q \qquad (2.4)$$

where δ_q is the area under the crop at q^{th} Block, which is known through complete enumeration by State Government in India. Note that $\hat{Z}_q$ is an unbiased estimator of Z_q as $E(\hat{Z}_q) = Z_q$ because $E(\hat{Y}_q) = Y_q$ under the assumption of regression model (2.1). Following Montgomery and Peck (1982) the variance of $\hat{Z}_q$ is given by

$$V(\hat{Z}_q) = \delta_q^2 \sigma^2 C_q' (X'X)^{-1} C_q \qquad (2.5)$$

where σ^2 is the residual variance corresponding to the regression model (2.1), X is matrix of X_{ij}'s at district level given by

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & . & . & . & X_{1p} \\ 1 & X_{21} & X_{22} & . & . & . & X_{2p} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & X_{n1} & X_{n2} & . & . & . & X_{np} \end{bmatrix}$$

and X' is the transpose of X matrix. C_q is column vector of X_{qi}'s at q^{th} block level given by

$C'_q = [1 \quad X_{q1} \quad X_{q2} \quad \ldots \quad X_{qp}]$

It is obvious that in general $\sum_{q=1}^{Q} \hat{Z}_q \neq Z$, where Z is the actual crop production reported at district level through crop cutting experiment in a given year. Thus, a new scaled estimator Z_q is given by

$$\tilde{Z}_q = a_q \hat{Z}_q \qquad (2.6)$$

where a_q is constant such at $\sum_{q=1}^{Q} \tilde{Z}_q = Z$ or $\sum_{q=1}^{Q} a_q \hat{Z}_q = Z$

Three alternative choices as suggested by Sisodia and Chandra (2012) were used by Singh et al (2012) to obtain different scaled estimator $\tilde{Z}_q$. These choices along with their conditional variance/MSE are given below.

<u>Choice- I</u>: Ratio Adjustment

The simplest choice of a_q is to take $a_q = a$ for all q.it result to $a = Z/\sum_{q}^{Q} \hat{Z}_q$ and a new scaled estimator of Z_q is

$$\tilde{Z}_q^{(1)} = \hat{Z}_q \left[Z/\sum_{q=1}^{Q} \hat{Z}_q \right] \tag{2.7}$$

With conditional MSE

$$MSE\left[\tilde{Z}_q^{(1)}\right] = \frac{\left(Z - \sum_{q=1}^{Q} \hat{Z}_q\right)^2 Z_q^2}{\left(\sum_{q=1}^{Q} \hat{Z}_q\right)^2} + \frac{Z^2}{\left(\sum_{q=1}^{Q} \hat{Z}_q\right)^2} V\left(\hat{Z}_q\right) \tag{2.8}$$

Note that the Estimator $\tilde{Z}_q^{(1)}$ is not an unbiased estimator of Z_q

Choice- II: Least square adjustment

Another choice of a_q is obtained by minimizing the sum of squared difference between $\tilde{Z}_q$ and $\hat{Z}_q$ subject to condition that $\sum_{q=1}^{Q} a_q \hat{Z}_q = Z$. This results to

$$a_q = \frac{Z - \sum_{q=1}^{Q} \hat{Z}_q}{Q\hat{Z}_q} + 1 \tag{2.10}$$

Thus, another new scaled estimator of Z_q is given by,

$$\tilde{Z}_q^{(2)} = \hat{Z}_q + \frac{Z - \sum_{q=1}^{Q} \hat{Z}_q}{Q} \tag{2.11}$$

with conditional variance

$$V\left(\tilde{Z}_q^{(2)}\right) = \frac{Q-2}{Q} V\left(\hat{Z}_q\right) + \frac{1}{Q^2} \sum_{q=1}^{Q} V\left(\hat{Z}_q\right) \tag{2.12}$$

Note that $\tilde{Z}_q^{(2)}$ is unbiased estimator of $\hat{Z}_q$.

Choice- III: Relative least square adjustment

The third choice of a_q could be one that minimizes the sum of square of relative differences $\left(\tilde{Z}_q - \hat{Z}_q\right)/\hat{Z}_q$ subject to condition that $\sum_{q=1}^{Q} a_q \hat{Z}_q = Z$.This results to

$$a_q = 1 + \frac{Z - \sum_{q=1}^{Q} \hat{Z}_q}{\sum_{q=1}^{Q} \hat{Z}_q^2} . \hat{Z}_q \qquad (2.13)$$

and a new scaled estimator of Z_q is

$$\tilde{Z}_q^{(3)} = \hat{Z}_q + \frac{\hat{Z}_q^2}{\sum_{q=1}^{Q} \hat{Z}_q^2} \left(Z - \sum_{q=1}^{Q} \hat{Z}_q \right) \qquad (2.14)$$

with conditional MSE

$$\text{MSE}\left[\tilde{Z}_q^{(3)}\right] = \left[1 + \frac{Z - \sum_{q=1}^{Q} \hat{Z}_q}{\sum_{q=1}^{Q} \hat{Z}^2{}_q} \hat{Z}_q \right]^2 V\left(\hat{Z}_q\right) \qquad (2.15)$$

Comparison for relative efficiency of $\tilde{Z}_q^{(1)}, \tilde{Z}_q^{(2)}$ and $\tilde{Z}_q^{(3)}$ over, $\hat{Z}_q$

Taking difference of $V\left(\hat{Z}_q\right)$ and $\text{MSE}\left(\tilde{Z}_q^{(1)}\right)$, i.e. $V\left(\hat{Z}_q\right)$ - $\text{MSE}\left(\tilde{Z}_q^{(1)}\right)$ and after simplification, we get the inequality

$$V\left(\hat{Z}_q\right) \geq \frac{\left(z - \sum_{q=1}^{Q} \hat{z}_q \right)^2 z_q^2}{\left[\left(\sum_{q=1}^{Q} \hat{z}_q \right)^2 - z^2 \right]} \qquad (2.16)$$

for $\tilde{Z}_q^{(1)}$ to be efficient than $\hat{Z}_q$. However, it seems that inequality (2.16) may not hold true in general. Taking difference of $V\left(\hat{Z}_q\right)$ and $V\left(\tilde{Z}_q^{(2)}\right)$, i.e. $V\left(\hat{Z}_q\right)$- $V\left(\tilde{Z}_q^{(2)}\right)$, and after simplification, we get the following inequality

$$V\left(\hat{Z}_q\right) \geq \frac{1}{2Q} \sum_{q=1}^{Q} V\left(\hat{Z}_q\right) \qquad (2.17)$$

for $\tilde{Z}_q^{(2)}$ to be efficient than $\hat{Z}_q$. It is obvious from the inequality (2.17) that the $V\left(\hat{Z}_q\right)$must be greater than the half of the average of variances of $\hat{Z}_q$ (q=1, 2, … Q), which is quite possible in general.

Taking difference of $V\left(\hat{Z}_q\right)$ and $\text{MSE}\left(\tilde{Z}_q^{(3)}\right)$, i.e. $V\left(\hat{Z}_q\right)$ - $\text{MSE}\left(\tilde{Z}_q^{(3)}\right)$, we get the following inequality

$$1-\left\{1+\frac{z-\sum_{q=1}^{Q}\hat{z}_q}{\sum_{q=1}^{Q}\hat{z}_q^2}\hat{z}_q\right\}^2 \geq 0 \tag{2.18}$$

Case1: If $Z-\sum_{q=1}^{Q}\hat{Z}_q > 0$, then RHS of (2.18) is negative indicating thereby $\hat{Z}_q$ will be always efficient than $\tilde{Z}_q^{(3)}$

Case2: If $Z-\sum_{q=1}^{Q}\hat{Z}_q < 0$, then RHS of (2.18) is always positive indicating thereby $\tilde{Z}_q^{(3)}$ will always be efficient than $\hat{Z}_q$.

3. MATERIALS & METHODS

A brief review by Singh *et al.* (2012) followed by Sharma *et al.* (2015) and Sharma & Sisodia (2016) for estimation of crop production at Block level in the preceding section. They used only input factors as independent variables in the regression model. The environment plays an important role in crop production. It the case of wheat, it requires a cool climate with optimal temperature between 16-22^{0}C for growth, through temperature below 3-4^{0}C can be harmful, and excessive heat during flowering and grain filing causes stress, reducing yield. Sudden or prolonged temperature changes also negative impact health. Wind plays a critical role in wheat cultivation by dispersing pollen for pollination and increasing carbon dioxide supply to enhance photosynthesis (Satorre and Slafer, 1999). Therefore, it is proposed to include weather variables also besides input factors as independent variables in the regression model. The proposed model at district level is given below.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip} + \lambda_1 I_{i1} + \lambda_2 I_{i2} + ... + \lambda_m I_{im} + \varepsilon_i \tag{3.1}$$

where Y_i is the yield of the crop during i^{th} year (i=1,2,3,..n), X_{ij} is the value of j^{th} input factor (j=1,2,….p) corresponding to i^{th} year, I_{im} is the Index value of m^{th} weather variable (m=1,2,…M) corresponding to i^{th} year. $\beta_0, \beta_1, \beta_2, \ldots, \beta_P, \lambda_1, \lambda_2, \ldots, \lambda_M$ are model parameters and ϵ_i is error term assumed to follow independently normal distribution with mean zero and variance σ^2.The weather index is constructed following the procedure laid down in Agrawal *et al.* (1986). It is described below.

$$I_m = \frac{\sum_{w=1}^{k} r_{mw} X_{mw}}{\sum_{w=1}^{k} r_{mw}} \quad (3.2)$$

Where X_{mw} is the value of m^{th} weather variable in w^{th} week and r_{mw} is the correlation coefficient between de-trended crop yield and m^{th} weather variable in w^{th} week . I_m is in fact weighted average of m^{th} weather variable.

The model (3.1) is fitted with data by ordinary least square technique

Let the fitted model be denoted as

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + ...\hat{\beta}_p X_{ip} + \hat{\lambda}_1 I_{i1} + \hat{\lambda}_2 I_{i2} + ... + \hat{\lambda}_m I_{im} \quad (3.3)$$

The fitted model (3.3) is used directly to predict the Block yield of the crop. The predicted crop yield for q^{th} Block (q=1,2,…..Q) in a given year is given by

$$\hat{Y}_q = \hat{\beta}_0 + \hat{\beta}_1 X_{q1} + \hat{\beta}_2 X_{q2} + ... + \hat{\beta}_p X_{qp} + \hat{\lambda}_1 I_{q1} + \hat{\lambda}_2 I_{q2} + ... + \hat{\lambda}_M I_{qM} \quad (3.4)$$

where $\hat{Y}_q$ is the predicted crop yield for q^{th} Block in the given year, X_{qp} and I_{qm} is the value of I_m in q^{th} Block in a given year, and $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p, \hat{\lambda}_I, ..., \hat{\lambda}_M$ are estimated model parameters.

An estimator of crop production, Z_q, for q^{th} Block in a given year, is therefore, given by

$$\hat{Z}_q = \delta_q . \hat{Y}_q \quad (3.5)$$

where δ_q is the area under the crop in q^{th} Block.

In general, $\sum_{q=1}^{Q} \hat{Z}_q \neq Z,$ where Z is crop production reported by CCEs at district level in that given year.

Following Singh *et al.* (2012) and others, We develop a sealed estimator of Z_q as

$$\tilde{Z}_q = a_q \hat{Z}_q \quad (3.6)$$

such that $\sum_{q=1}^{Q} \tilde{Z}_q = \sum_{q=1}^{Q} a_q \hat{Z}_q = Z,$ where a_q is the sealed factor.

The choices for a_q are similar to those described by Singh *et al.* (2012). The resultant estimators, i.e. $\tilde{Z}_q^{(1)}, \tilde{Z}_q^{(2)}$ and $\tilde{Z}_q^{(3)}$ will also be similar to described by Singh *et al.* (2012). The expressions for conditional variances and MSE of the scaled estimators will also be similar to those given in section-2. The only difference we get in the magnitude of $\hat{Z}_q, \tilde{Z}_q^{(1)}, \tilde{Z}_q^{(2)}, \tilde{Z}_q^{(3)}$ and in the variance of $\hat{Z}_q$. The variance of $\hat{Z}_q$ can be obtained by following the Montegomery & Peck as follows.

$$V(\hat{Z}_q) = \delta_q^2 \hat{\sigma}^2 C'_q (X'X)^{-1} C_q$$

where X is the matrix of independent variables in the model, given by

$$X = \begin{bmatrix} 1 & X_{11} & X_{12}\ldots\ldots & X_{1p} & I_{11} & I_{12}\ldots\ldots & I_{1M} \\ 1 & X_{21} & X_{22}\ldots\ldots & X_{2p} & I_{21} & I_{22}\ldots\ldots & I_{2M} \\ . & . & \ldots\ldots\ldots. & . & . & \ldots\ldots\ldots & . \\ . & . & \ldots\ldots\ldots & . & . & \ldots\ldots\ldots & . \\ . & . & \ldots\ldots\ldots. & . & . & \ldots\ldots\ldots & . \\ 1 & X_{n1} & X_{n2}\ldots\ldots & X_{np} & I_{n1} & I_{n2}\ldots\ldots & I_{nM} \end{bmatrix}$$

X' is the transpose of X and C'_q is column vector of X'_{qp} and I_{qM} corresponding to q^{th} Block in a given year as follows.

$$C'_q = \begin{bmatrix} 1 X_{q1} & X_{q2}\ldots\ldots & X_{qp} & I_{q1} & I_{q2}\ldots\ldots & I_{qM} \end{bmatrix}$$

$\hat{\sigma}^2$ is the estimated residual variance which can be obtained from the Table of analysis of variance of regression model.

4. RESULTS AND DISCUSSION

An empirical study has been conducted to illustrate the relative merits of these options under consideration. The time series data on yield of wheat, percent irrigated area under wheat (X_1), fertilizer consumption for wheat (N, P, K) in kg/ha (X_2) and percent relative area under wheat as percentage of gross cropped area (X_3) for Sultanpur district of Utter Pradesh in India pertaining to the period 1990-91 to 2008-09 were used for the study. The Block wise time series data on three auxiliary variables, X_1, X_2 and X_3 for Sultanpur district pertaining to the same period and the weekly time series data of weather variables viz. Minimum temperature, Maximum temperature, Wind velocity from Standard Metrological Week 44 to the next year of SMW 15 for the aforesaid period which were utilized in the present investigation. The Block wise actual data of wheat production based on crop cutting experiments in Sultanpur district during the year 2006-07 are also available in District Statistical Bulletins as part of some pilot studies conducted by Govt. of UP, which were used for validation.

The block estimates of wheat production for the year 2006-07 based on various estimators developed using four statistical methodology have been obtained along with their standard error *etc.* The results are presented in tow sections as follows.

Results based on the methodology suggested by Singh *et al*. (2012)

The model (2.1) has been fitted with three sets of data separately by ordinary least square technique. The results of the fitted model are summarized in Table 4.1.1.

Table 4.1.1. Details of Fitted Model

Set	Fitted Model	R^2(%)	SE($\hat{Y}$)	($\hat{\sigma}^2$)
Set 1	$Y = -544.53+5.74X_1+0.764X_2-0.261X_3$ **	84.67	1.33	1.78

*P <0.05, ** P<.01

The fitted model (2.1) for the sets of data at district level has been directly used to predict the block estimates of wheat yield using block wise data on X_1, X_2 and X_3 for the year 2006-07. In fact, the block level wheat yield based on CCEs for the year 2006-07 was available from a Pilot-Survey conducted by Directorate of Agricultural Statistics and Crop Insurance, Govt. of U.P., India. So, the data on wheat yield of block level of the year 2006-07 were used to validate the block level estimates of wheat production obtained from the proposed estimators.

The blocks estimates of wheat production based on four estimators and their per cent standard error were computed and are presented in Table 4.3. To measure how far the adjusted estimates $\tilde{Z}_q^{(i)}$ are from the Z_q, an average distance between adjusted estimates and Z_q has been computed by following formula.

$$D_i = \sqrt{\frac{\sum_{q=1}^{Q}\left(Z_q - \hat{\theta}_{qi}\right)^2}{Q}} \qquad (i= 1, 2, 3, 4) \quad (4.1.2)$$

Where $\hat{\theta}_{qi} = \hat{Z}_q, \hat{\theta}_{qi} = \tilde{Z}_q^{(1)}, \hat{\theta}_{qi} = \tilde{Z}_q^{(2)}, \hat{\theta}_{qi} = \tilde{Z}_q^{(3)}$ and Q is number of blocks. The values of D_i's are also presented in Table 4.1.2.

$$\text{Percent standard error (PSE)} = \sqrt{\frac{\text{MSE/ Variance of the estimator}}{\text{Estimate}}} \times 100 \qquad (4.1.2)$$

The results presented in Table 4.1.2 showed that the estimates $\hat{Z}_q$, $\tilde{Z}_q^{(1)}$, $\tilde{Z}_q^{(2)}$ and $\tilde{Z}_q^{(3)}$ are almost at par in terms of percent standard error of block estimate. However, the percent standard error of block estimates ranged in general to the maximum 9% with exception of Waldirai and Shahgarh which yielded about 20% and 24% standard error.

The block estimates obtained from various estimators have been found to be close to the actual yield. However, in few blocks the estimates are not close to actual production, probably because of local effects of the blocks, which may not uniformly similar to other block. The average distance (D_i) has been found to be minimum, i.e. 20743 in case of $\tilde{Z}_q^{(2)}$.

The percent errors of the estimates based on $\tilde{Z}_q^{(2)}$ were almost at par with that of $\hat{Z}_q$ and $\tilde{Z}_q^{(3)}$. Therefore, on the basis of percent standard error of the block estimates and the average distance it can be recommended that $\tilde{Z}_q^{(2)}$ could be preferred in practice for estimating the block estimate among all the other estimators.

Results based on the proposed methodology using weather variables including input factors.

To illustrate the proposed statistical methodology developed in preceding chapter an empirical study has been carried out. It is proposed to estimate wheat production at block level in Sultanpur district of Uttar Pradesh, India. The time series data on Y_i (wheat yield), I_{im} (Index value of mth weather variable) and X_{ij} (covariates) pertaining to the period 1990-91 to 2008-09 were used for the fitting the model at district level. These data are given in Appendix-I (Table I.2). The block wise data on Y_q (wheat yield based on CCEs), Z_q (wheat production based on CCEs) and X_j's for the year 2006-07 are given in Appendix-II. The block wise data on X_j's and I_{im} have been used for developing block level estimate of Y_q from the proposed estimators in the year 2006-07. The weighted weather Index I_m developed at district level was also used at Block level.

Table 4.1.2: Block estimates of wheat production based on different estimators and their per cent standard error during the year 2006-07 using multiple regression model based on only input factors

S.No.	Block	Actual Prod. (qt) Z	Block Estimates				% Standard Error			
			$\hat{Z}_q$	$\tilde{Z}_q^{(1)}$	$\tilde{Z}_q^{(2)}$	$\tilde{Z}_q^{(3)}$	$\hat{Z}_q$	$\tilde{Z}_q^{(1)}$	$\tilde{Z}_q^{(2)}$	$\tilde{Z}_q^{(3)}$
1	Shukul bazaar	188293	186441	183623	183268	184092	1.83	2.39	2.45	1.83
2	Jagdishpur	198054	200259	197233	197086	197549	4.25	4.52	4.41	4.25
3	Musafhirkhana	166349	161401	158961	158227	159640	7.32	7.48	7.39	7.32
4	Waldi Rai	142614	151343	149055	148169	149795	20.39	20.44	19.96	20.39
5	Jamo	230371	228137	224689	224963	224620	3.76	4.06	3.89	3.76
6	Shahgarh	136072	179539	176825	176365	177361	24.05	24.10	23.41	24.05
7	Gauriganj	242798	232773	229255	229599	229112	5.95	6.15	5.91	5.95
8	Amethi	168959	197332	194350	194159	194701	7.15	7.31	7.11	7.15
9	Bhetua	146587	171435	168844	168261	169449	2.79	3.18	3.28	2.79
10	Bhadar	138111	169024	166469	165850	167093	8.70	8.83	8.65	8.70
11	Sangrampur	112059	131548	129560	128374	130379	8.75	8.89	8.89	8.75
12	Dhanpatganj	233104	214005	210770	210831	210910	1.85	2.41	2.32	1.85
13	Kurebhar	217956	224842	221443	221668	221426	1.69	2.28	2.15	1.69
14	Jai Singh Pur	232049	235704	232141	232530	231950	2.94	3.32	3.14	2.94
15	Kurwar	236005	203760	200680	200586	200954	1.93	2.47	2.43	1.93
16	Dube Pur	214739	211437	208241	208263	208416	3.45	3.78	3.66	3.45
17	Bhadaiyeea	176209	214941	211692	211767	211819	1.82	2.38	2.29	1.82
18	Dostpur	229495	215650	212390	212476	212507	2.91	3.29	3.17	2.91
19	Akhand Nagar	314681	274881	270727	271708	269775	4.09	4.37	4.11	4.09
20	Lambhua	234987	236447	232873	233273	232669	3.38	3.71	3.53	3.38
21	Pratap Pur Kamaicha	177156	172890	170277	169716	170870	7.88	8.03	7.87	7.88
22	Kadipur	412974	405653	399522	402480	394534	3.49	3.82	3.45	3.49
	Total	4549621	4619441	4549621	4549621	4549621				
Average distance							20985	21066	20743	21460

Source: Sharma *et al.* (2015)

Table 4.2.1. Details of Fitted Model

Set	Fitted Model	R^2(%)	SE($\hat{Y}$)	($\hat{\sigma}^2$)
Set 1	$Y = -91.74+1.05X_1+0.07X_2+0.02X_3-0.19I_{11}+0.41I_{21}+0.52I_{31}$ **	92.81	1.0216	1.0438

*P <0.05, ** P<.01

The estimates of regression coefficient, their standard error and value of coefficient of determinations (R^2) *etc.* have been presented in Table 4.2.1. The fertilizer consumption showed positive and significant effect.I_{11} (maximum temperature) has shown negative and significant effect on the yield. The value of coefficient of determination (R^2) was found to be quite high i.e. 92.81% which is indicative of the fact that these variables included in the model have been quite sufficient to explain the variability in the data of wheat yield at district level. The analysis of variance for regression analysis of the aforesaid model is presented in Table 4.2.1. This shows the overall significance of the model fitted.

The model fitted at district level was directly used to predict the block estimates of wheat yield using block wise data on X_1, X_2, X_3, I_{11}, I_{21} and I_{31} for the year 2006-07. The blocks estimates of wheat production based on four estimators and their per cent standard error were computed and are presented in Table 4.2.2. Di's have been computed by formula given in equation 4.1.2. The values of Di's are also presented in Table 4.2.2

The results presented in Table 4.2.2 showed that the estimates $\hat{Z}_q$, $\tilde{Z}_q^{(1)}$, $\tilde{Z}_q^{(2)}$ and $\tilde{Z}_q^{(3)}$ are almost at par in terms of percent standard error of block estimate. However, the percent standard error of block estimates ranged in general to the maximum 7% with exception of Waldirai and Shahgarh which yielded about 20% and 17% standard error.

The block estimates obtained from various estimators have been found to be close to the actual yield. However, in few blocks the estimates are not close to actual production, probably because of local effects of the blocks, which may not uniformly similar to other block. The average distance (D_i) has been found to be minimum, i.e. 24456 in case of $\tilde{Z}_q^{(2)}$. The percent errors of the estimates based on $\hat{Z}_q$, $\tilde{Z}_q^{(2)}$ and $\tilde{Z}_q^{(3)}$ were almost at par. Therefore, on the basis of percent standard error of the block estimates and the average distance it can be recommended that $\tilde{Z}_q^{(2)}$ could be preferred in practice for estimating the block estimate among all the other estimators.

Table 4.2.2: Block estimates of wheat production based on different estimators and their per cent standard error during the year 2006-07 using multiple regression model on the basis of input factors and weather variables.

S.No.	Block	Actual Prod. (qt) Z	Block Estimates				% Standard Error			
			$\hat{Z}_q$	$\tilde{Z}_q^{(1)}$	$\tilde{Z}_q^{(2)}$	$\tilde{Z}_q^{(3)}$	$\hat{Z}_q$	$\tilde{Z}_q^{(1)}$	$\tilde{Z}_q^{(2)}$	$\tilde{Z}_q^{(3)}$
1	Shukul bazaar	188293	184163	183299	183189	183446	2.09	2.14	2.47	2.09
2	Jagdishpur	198054	195222	194306	194247	194416	3.86	3.89	3.94	3.86
3	Musafhirkhana	166349	149678	148976	148704	149205	6.82	6.84	6.79	6.82
4	Waldi Rai	142614	128156	127555	127181	127809	20.48	20.48	19.79	20.48
5	Jamo	230371	225439	224381	224464	224364	4.10	4.13	4.10	4.10
6	Shahgarh	136072	203612	202657	202637	202735	17.29	17.30	16.62	17.29
7	Gauriganj	242798	237627	236512	236652	236433	4.69	4.71	4.62	4.69
8	Amethi	168959	199683	198747	198709	198840	5.80	5.81	5.71	5.80
9	Bhetua	146587	168678	167887	167703	168076	2.63	2.67	2.98	2.63
10	Bhadar	138111	170599	169799	169625	169984	7.17	7.19	7.05	7.17
11	Sangrampur	112059	130264	129653	129289	129905	7.51	7.53	7.50	7.51
12	Dhanpatganj	233104	211326	210335	210352	210382	2.34	2.39	2.57	2.34
13	Kurebhar	217956	219559	218529	218585	218540	2.24	2.29	2.46	2.24
14	Jai Singh Pur	232049	233189	232095	232214	232039	2.69	2.73	2.81	2.69
15	Kurwar	236005	201002	200060	200028	200148	2.61	2.65	2.83	2.61
16	Dube Pur	214739	209941	208956	208967	209009	3.00	3.04	3.14	3.00
17	Bhadaiyeea	176209	209895	208911	208921	208964	2.39	2.43	2.61	2.39
18	Dostpur	229495	207420	206447	206445	206510	3.28	3.32	3.40	3.28
19	Akhand Nagar	314681	274146	272860	273171	272556	4.02	4.05	3.97	4.02
20	Lambhua	234987	238921	237800	237946	237714	2.93	2.96	3.01	2.93
21	Pratap Pur Kamaicha	177156	174430	173612	173456	173787	6.44	6.46	6.36	6.44
22	Kadipur	412974	398110	396243	397136	394759	3.15	3.18	3.08	3.15
	Total	4549621	4571062	4549621	4549621	4549621				
Average distance							24475	24519	24456	24607

Z_q Crop production (qt) based on CCEs

5. SUMMARY AND CONCLUSION

The percent standard errors of the estimators using methodology of Sing *et al* (2012) and the proposed methodology are presented in the Table 5.1.

Table 5.1: Percent standard errors of the estimators

S. No.	Block	Singh *et al.* (2012)				Proposed methodology			
		$\hat{Z}_q$	$\tilde{Z}_q^{(1)}$	$\tilde{Z}_q^{(2)}$	$\tilde{Z}_q^{(3)}$	$\hat{Z}_q$	$\tilde{Z}_q^{(1)}$	$\tilde{Z}_q^{(2)}$	$\tilde{Z}_q^{(3)}$
1	Shukul bazaar	1.83	2.39	2.45	1.83	2.09	2.14	2.47	2.09
2	Jagdishpur	4.25	4.52	4.41	4.25	3.86	3.89	3.94	3.86
3	Musafhirkhana	7.32	7.48	7.39	7.32	6.82	6.84	6.79	6.82
4	Waldi Rai	20.39	20.44	19.96	20.39	20.48	20.48	19.79	20.48
5	Jamo	3.76	4.06	3.89	3.76	4.10	4.13	4.10	4.10
6	Shahgarh	24.05	24.10	23.41	24.05	17.29	17.30	16.62	17.29
7	Gauriganj	5.95	6.15	5.91	5.95	4.69	4.71	4.62	4.69
8	Amethi	7.15	7.31	7.11	7.15	5.80	5.81	5.71	5.80
9	Bhetua	2.79	3.18	3.28	2.79	2.63	2.67	2.98	2.63
10	Bhadar	8.70	8.83	8.65	8.70	7.17	7.19	7.05	7.17
11	Sangrampur	8.75	8.89	8.89	8.75	7.51	7.53	7.50	7.51
12	Dhanpatganj	1.85	2.41	2.32	1.85	2.34	2.39	2.57	2.34
13	Kurebhar	1.69	2.28	2.15	1.69	2.24	2.29	2.46	2.24
14	Jai Singh Pur	2.94	3.32	3.14	2.94	2.69	2.73	2.81	2.69
15	Kurwar	1.93	2.47	2.43	1.93	2.61	2.65	2.83	2.61
16	Dube Pur	3.45	3.78	3.66	3.45	3.00	3.04	3.14	3.00
17	Bhadaiyeea	1.82	2.38	2.29	1.82	2.39	2.43	2.61	2.39
18	Dostpur	2.91	3.29	3.17	2.91	3.28	3.32	3.40	3.28
19	Akhand Nagar	4.09	4.37	4.11	4.09	4.02	4.05	3.97	4.02
20	Lambhua	3.38	3.71	3.53	3.38	2.93	2.96	3.01	2.93
21	Pratap Pur Kamaicha	7.88	8.03	7.87	7.88	6.44	6.46	6.36	6.44
22	Kadipur	3.49	3.82	3.45	3.49	3.15	3.18	3.08	3.15

It can be observed from results of the Table 5.1 that by adding weather variables in the model as explanatory variables the percent standard errors of the estimators have reduced in general. However, the reduction in the percent standard errors have been found to be quite marginal although it was expected to be substantial reduction. Probably it might be because of variation between blocks in respect of weather variables but in the present investigation the magnitude of weather variable are considered constant for each Block. Moreover, the effects of weather variables have been considered as fixed effect in the model, where the weather variables effects are random in general over years.

Results based on the study conducted by Sisodia & Chandra (2012), Singh *et al*. (2012), Sisodia and Singh (2012), Sharma *et al*. (2015) and Sharma & Sisodia (2016) are similar to the results of the present investigation.

In view of the above discussion and the results in the preceding section, the following specific conclusions have been presented below.

1. The inclusion of weather variables in the model as an explanatory variable brought little bit reduction in the percent standard errors of the estimators.
2. On the basis of results, it is very obvious that Blocks estimators based on the estimators and actual estimate based on CCEs are almost close to each other in most of the Blocks.
3. Among the four estimators (an unadjusted one and three adjusted) the adjusted estimator $\tilde{Z}_q^{(2)}$ has been found to be precise for the producing the reliable estimates of wheat production at block level in view of per cent standard error and minimum average distance.
4. The Block estimates could be further improved if Block variations in terms of soil type and other factors can be accounted for in the model.
5. It may be including that the larger sets of data could be used for better precision of the estimates. If a greater number of auxiliary variables are available at district and block levels, the use of larger set of the auxiliary variables may further improve the estimates of crop production at block level.

REFERENCES

[1] Battese, G. E., Harter R. M. and Fuller, W. A. (1988), An error components model for prediction of county crop area using survey and satellite data. *J. Amer. Stat. Assoc*., 83: 28-36.

[2] Chambers, R. and Tzavidis, N. (2006), M-quantile models for small area estimation. *Biometrika*, 93(2): 255-268.

[3] Chambers, R. L. and Finney, G. A. (1977), Log-linear models for small area estimation. Unpublished paper, Aus. Bure. of Stat.

[4] Chandra, H. (2009), Estimation of small area proportions under unit level spatial models. *J. Ind. Soc. Agril. Stat.,* 63(3): 117-122.

[5] Chandra, H. and Chambers, R. (2009), Multipurpose weighting form small area estimation. *J. Off. Stat.*, 25(3): 379-395.

[6] Chandra, H. and Chambers, R. L. (2005), Comparing EBLUP and C-EBLUP for small area estimation. *Stat. Trans.*, 7: 637-648.

[7] Chandra, H., Salvati, N. and Chambers, R. (2007), Small area estimation for spatially correlated population- A comparison of direct and indirect model-based methods. *Stat. Trans*., 8: 887-906.

[8] Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment. *Environmental Research Letters*, 13(11), 114003. DOI: https://doi.org/10.1088/1748-9326/aae159.

[9] D. (1968), Double sampling and its application in agriculture, contribution in statistics and agricultural science. Dr. V.G. Panse lecture, Indian Society of Agricultural Statistics, New Delhi.

[10] Gonzalez, M. E. (1973), Use and evaluation of synthetic estimators. Proceedings of the Social Sciences Section, *Amer. Stat. Assoc*., 33-36.

[11] Jeet, M., Anurag, Niwas, R., & Tomar, D. (2022). Integrating weather model & remote sensing indices for wheat yield prediction in Haryana, India. *Environment Conservation Journal*, 23(1&2), 124–130. DOI: https://doi.org/10.36953/ECJ.021811-2118.

[12] Jiang, J. and Lahiri, P. (2006), Mixed model prediction and small area estimation. *Test*, 15: 1–96.

[13] Khaki, S., Wang, L., & Archontoulis, S. V. (2020). Crop yield prediction integrating genotype and weather variables using deep learning. *PLOS ONE*, 15(6), e0233949. DOI: https://doi.org/10.1371/journal.pone.0233949.

[14] Mistry, M. N., Wing, I. S., & De Cian, E. (2017). Simulated vs. empirical weather responsiveness of crop yields: US evidence and implications for the agricultural impacts of climate change. *Environmental Research Letters*. DOI: https://doi.org/10.1088/1748-9326/aa788c.

[15] Montgomery, D. C. and Peck, E. A. (1982), Introduction to linear regression analysis. John Willy & Sons, Canada.

[16] Panse, V. G., Rajagopalan, M. and Pillai, S. (1966), Estimation of crop yields for small areas. *Biometrics*, **66**: 374-388.

[17] Panse, V. G., Rajagopalan, M. and Pillai, S. (1966), Estimation of crop yields for small areas. *Biometrics*, 66: 374-388. Singh,

[18] Pratesi, M. and Salvati, N. (2008), Small area estimation: the EBLUP estimator with autoregressive random area effects. *Statist. Methods Appl*., 17: 113-141.

[19] Rao, J.N.K. (2003), Small Area Estimation. Wiley- Interscience, Wiley Series in Surv. Methodo.

[20] Salvati, N., Pratesi, M., Tzavidis, N. and Chambers, R. (2007), Spatial M-quantile models for small area estimation. Working Paper (15-08), Centre for Statistical and survey Methodology, The University of Wollongong, pp: 1-16.

[21] Satorre, E.H., and Slafer, G.A. (1999). Wheat: Ecology and Physiology of Yield Determination. CRC Press.

[22] Sharma, M.K. and Sisodia, B.V.S., (2016), Estimation of Crop Production for smaller geographical area- An application of discriminant function analysis, *Journal of Statistical Theory and Practice*, *10*(2):444-455.

[23] Sharma, M.K., Sisodia, B.V.S., and Kumar, Sunil (2015), On choice of explanatory variables in linear model for estimation of crop production at smaller geographical area. *J. Ind. Soc. Agril. Stat.,* **69**(3): 315-326.

[24] Sharma, S. D., Srivastava, A.K. and Sud, U.C. (2004), Small area crop estimation for crop yield estimates at Gram Panchayat level. *J. Ind. Soc. Agril. Stat.,* **57**: 26-37.

[25] Singh, B., Sisodia, B.V.S., Singh, A. and Kaushal, R. P. (2012), A note on estimation methods of crop production at block level. *Journal of Appl. Stat.,* **39** (9): 2015-2027.

[26] Singh, R., Semwal, D. P., Rai, A., & Chhikara, R. S. (2002). Small area estimation of crop yield using remote sensing satellite data. *International Journal of Remote Sensing*, 23(1), 49–56. DOI: https://doi.org/10.1080/01431160010014756.

[27] Sisodia, B.V.S. and Chandra, H. (2012), Estimation of crop production at geographic level in India. *J. Ind. Soc. Agril. Stat*., **66** (2):313-320.

[28] Sisodia, B.V.S. and Singh, A. (2012), On small area estimation- An application in agriculture. *Journal of Ind. Soc. Agril. Stat*., **66** (3): 391-400

[29] Sisodia, B.V.S., Chandra, Hukum and Shukla, D. (2015), An application of area level model for estimation of crop production at smaller geographical level in India. *Int. J. Agril. And Stat. Science*, **11**(1): 97-104.

[30] Srivastava, A.K., Ahuja, D.L., Bhatia, D.K. and Mathur, D.C. (1998), Small area estimation techniques in agriculture, Souvenir, Advances in Agricultural Statistics and Computer Application. I.A.S.R.I., New Delhi, 32-45.

[31] Stasny, E. A., Geol, P. K. and Rumsey, D. J. (1991), County estimates of wheat production. *Survey methodology*, 17 (2): 211-225.

[32] Sukhatme, P. V. and Sukhatme, B. V. (1970), Sampling Theory of surveys with Applications, Iowa State University Press, Ames, IA.

[33] Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709. DOI: https://doi.org/10.1016/j.compag.2020.105709.