https://doi.org/10.56093/JISAS.V78I2.1
Available online at http://isas.org.in/isa/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 78(2) 2024 81–87

Estimation of the Average Yield of Cotton using Outlier Robust Geographically Weighted Regression Approach

Pramod Kumar Moury[1,2,3], Tauqueer Ahmad[2], Anil Rai[4], Ankur Biswas[2], Prachi Misra Sahoo[2] and Mohammadsanaulla K. Huddar[5]

[1]The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi
[2]ICAR-Indian Agricultural Statistics Research Institute, New Delhi
[3]Amar Singh College, Lakhaoti, Bulandshahr
[4]Indian Council of Agricultural Research, New Delhi
[5]Agricultural Insurance Company of India Limited, New Delhi

Received 01 February 2023; Revised 28 May 2024; Accepted 30 May 2024

SUMMARY

The General Crop Estimation Survey (GCES) scheme requires a large number of Crop Cutting Experiments (CCEs) to be conducted to get a reliable estimate below the district level. However, conducting a large number of CCEs imposes a financial burden on Govt. agencies. Additionally, large-scale surveys like GCES often result in many outlier observations in the CCE data. To address this issue,this study was conducted to estimate the yield rate of cotton with a relatively fewer number of CCEs than the GCES scheme using the proposed Outlier Robust Geographically Weighted Regression (ORGWR) approach. Validation of the proposed methodology was done using the real CCE dataset of Amravati district for the 2012-13 agriculture year in Maharashtra. In this approach,the number of CCEs conducted for GCES scheme was reduced, and then this reduced number of the CCEs can be predicted using the proposed ORGWR approach. The predicted CCEs and the incomplete CCEs data are then combined to form a complete dataset. This complete dataset is used to calculate the crop yield accurately. The study conducted a comparison between the ORGWR approach and GCES methodology for estimating the average yield of cotton. The results showed that the ORGWR approach, when used with a lesser number of CCEs, yielded estimates that were almost equivalent to those obtained using the GCES methodology with the complete dataset. Moreover, the standard error of the estimate was reliable, indicating the validity of the results.

Keywords: Crop cutting experiments; GCES; Geographically weighted regression; Outlier robust geographically weighted regression, Spatial non-stationarity.

Statements and Declarations:

The authors declare no potential conflict of interest relevant to this article.

1. INTRODUCTION

The General Crop Estimation Surveys (GCES) Scheme is adopted by the Directorate of Economics and Statistics (DES), Ministry of Agriculture and Farmers Welfare, Govt. of India and is being implemented in all the states of the country to estimate crop yield at a higher administrative level (i.e. National, State, district). Since 1944, Crop Cutting Experiments (CCE) have been used to access crop yield. CCEs also play a very important role in crop yield estimation in crop insurance. Under GCES around 16.48 lakh CCEs were carried out in the year 2016-17. Still, these large numbers of CCEs are not adequate to provide reliable estimates at lower administrative levels below the district level. To obtain estimates below the district level, a greater number of CCEs need to be conducted than existing. This imposes an enormous additional financial burden on the government agency and increases the non-sampling

Corresponding author: Ankur Biswas
E-mail address: ankur.biswas@icar.gov.in

errors considerably resulting in further deterioration of the production statistics.

To address this problem, instead of conducting a massive number of CCEs, we may conduct a smaller number and use an appropriate prediction approach to estimate the remaining CCEs required for GCES. Linear regression is a viable method for predicting remaining CCEs. However, in large-scale surveys such as agriculture, forestry, environmental, and ecological surveys, observations are often spatially correlated. This means that the relationship between dependent and independent variables will differ across all locations in the study area, leading to spatial heterogeneity. When this happens, applying a global regression model may increase bias and the mean square error of the predicted value.

Brunsdon *et al.* (1996) proposed the Geographically Weighted Regression (GWR) model to address spatial heterogeneity. Large-scale survey data often contains outliers. Several authors have discussed the outlier-robust method of geographically weighted regression (RGWR). They used different approaches such as fitting iterative GWR models (Harris *et al.*, 2010), least absolute deviation (Zhang and Mei, 2011; Afifah *et al.*, 2017), robust locally weighted least squares kernel regression method (Ma *et al*., 2014), and robust GWR models (Warsito *et al.*, 2018). Recently, few authors attempted to develop estimators using model calibration and model based/assisted approaches under finite population sampling using GWR (Saha *et al.*, 2023; Paul *et al.*, 2023a, 2023b, 2024).

Panse *et al.* (1966) adopted the double sampling approach for the estimation of block level yield by treating the farmer's eye estimate as an auxiliary character and the CCE estimate as a character under study. Raheja *et al.*(1977) developed a sampling methodology for estimating the yield of cotton and suggested a procedure for building up advanced estimates of the yield of cotton based on a few pickings. The Central Statistics Office (CSO, 2008) detailed the GCES general guidelines adopted for preparing the estimates of crop production of different crops. Ahmad *et al.* (2009) evaluated the methodologies being adopted for obtaining official and trade estimates of cotton production. Ahmad *et al.* (2013, 2020),further, developed an alternative sampling methodology for the estimation of the average yield of cotton using a double sampling technique under a stratified two-stage sampling design framework. Moury *et al.* (2020) proposed two GWR-based estimators to estimate the total population for study variables showing spatial non-stationarity.

In this study, we aimed to estimate the yield rate of cotton crops using a smaller number of CCEs than what is mandated by the GCES scheme. To achieve this, we proposed an outlier robust geographically weighted regression (ORGWR) approach. This method allowed us to estimate the remaining CCEs by fitting a GWR model and down-weighting observations with large residuals.

2. MATERIALS AND METHODS

The estimator of regression coefficients under the GWR model is similar to that of weighted least squares (WLS) of the global regression model except that the weight of a particular observation is constant over the all-regression point in the WLS method of estimation, whereas, in case of GWR parameter estimation, the weight of a particular observation is varying from location to location over all the regression points.

Let $y_i,\ i = 1, 2, ..., N,$ denotes the value of the study variable associated with the i^{th} unit of population $\boldsymbol{U}$ of size N. Let, $\boldsymbol{X} = \left(\boldsymbol{x}_1, ..., \boldsymbol{x}_i, ..., \boldsymbol{x}_N\right)^T$ is a set N x p of auxiliary variables where $\boldsymbol{x_i} = (x_{i1}, x_{i2}, ..., x_{ip})^T$ for all $i \in U$, are p auxiliary variables associated with the study variable. It is also assumed that values of auxiliary variables associated with each unit are known.Let $\left(u_i, v_i\right)$ denotes the geographical location of i^{th} unit in the space.

Considering the GWR model as

$$y_i = \beta_0\left(u_i, v_i\right) + \beta_1\left(u_i, v_i\right)x_i + e_i, \qquad i = 1, 2, ..., N \qquad ....(1)$$

whereas $\beta_0\left(u_i, v_i\right)$ and $\beta_1\left(u_i, v_i\right)$ are the location specific intercept and slope parameters of the model at the location $\left(u_i, v_i\right)$, x_i denotes the value of the auxiliary variable of the i^{th} observation, e_i is the random error component associated with i^{th} unit, which are distribution identically and independently distributed as normal with mean zero and variance σ^2.

Let the spatial weight of i^{th} observation with respect to the location $\left(u_i, v_i\right)$ is $w_i\left(u_i, v_i\right)$. For defining the geographical weighting function, let $d_{eu.i}\left(u_i, v_i\right)$ is

Euclidian distance between the regression point $\left(u_{i,}v_i\right)$ and i^{th} sampled data point and h_{gwr} is the bandwidth of geographically weighted regression analysis. We have considered the Bi-square shape of geographical weighting functions (kernel) for this study. The bi-square shape of geographical weighting functions is denoted as:

$$w_i\left(u_{i,}v_i\right)=\begin{cases}\left(1-\left(d_{eu.i}\left(u_{i,}v_i\right)\big/h_{gwr}\right)^2\right)^2 & \text{if } d_{eu.i}\left(u_{i,}v_i\right)<h_{gwr}\\ 0, & \text{otherwise}\end{cases} \quad \dots (2)$$

Outliers are often present in sample survey data. To deal with them, M-estimation (Huber, 1981) is used to lower the weight of observations with high residuals. First, we estimated the model and then assigned $w_{i.r}\left(u_{i,}v_i\right)$ weight to the i^{th} observation with respect to the regression point location $\left(u_{i,}v_i\right)$. This created a diagonal residual weight matrix, W_r, with its diagonal element as

$$w_{i.r}\left(u_{i,}v_i\right)=\begin{cases}1, & \text{if } \left|\hat{e}_i\right|\le 2\hat{\sigma}\\ \left[1-\left(\left|\hat{e}_i\right|-2\right)^2\right]^2, & \text{if } 2\hat{\sigma}<\left|\hat{e}_i\right|\le 3\hat{\sigma}\\ 0, & \text{otherwise}\end{cases} \quad \dots (3)$$

Now, this residual weight is multiplied with the spatial weight. Thus, a new combined weight matrix, W_c, is thus created, which is the term-by-term product of each element of the spatial weight matrix and residual weight matrix which is defined as:

$$w_{i.C}\left(u_{i,}v_i\right)=w_i\left(u_{i,}v_i\right).w_{i.r}\left(u_{i,}v_i\right). \quad \dots (4)$$

Thus,

$$\boldsymbol{W}_C\left(u_{i,}v_i\right)=\begin{pmatrix} w_{1.C}\left(u_{i,}v_i\right) & 0 & 0 & \cdots & 0\\ 0 & w_{2.C}\left(u_{i,}v_i\right) & 0 & . & .\\ . & . & w_{i.C}\left(u_{i,}v_i\right) & . & .\\ . & . & . & . & .\\ 0 & 0 & 0 & \cdots & w_{n.C}\left(u_{i,}v_i\right)\end{pmatrix} \quad \dots (5)$$

Then, the GWR model parameter can be estimated as

$$\hat{\boldsymbol{\beta}}^{gwr}\left(u_{i,}v_i\right)=\left(\boldsymbol{X}_s^T\boldsymbol{W}_C\left(u_{i,}v_i\right)\boldsymbol{X}_s\right)^{-1}\boldsymbol{X}_s^T\boldsymbol{W}_C\left(u_{i,}v_i\right)\boldsymbol{y}_s. \quad \dots (6)$$

The process involved multiple cycles, where the latest set of residuals was computed, and a new model fit after each cycle. The weights were then recalculated, and the model was updated again using the new weights. This automated approach helped to handle outliers while calibrating the GWR model. Ultimately, this led to an outlier-robust estimate of the parameter for the GWR model, i.e., $\hat{\boldsymbol{\beta}}^{rgwr}\left(u_{i,}v_i\right)$ as

$$\hat{\boldsymbol{\beta}}^{rgwr}\left(u_{i,}v_i\right)=\begin{bmatrix}\hat{\beta}_0^{rgwr}\left(u_{i,}v_i\right)\\ \hat{\beta}_1^{rgwr}\left(u_{i,}v_i\right)\end{bmatrix} \quad \dots (7)$$

Robust geographically weighted regression coefficients are obtained for non-sampled point as

$$\hat{\beta}_0^{rgwr}\left(u_j,v_j\right)=\frac{\sum_{i=1}^{n}w_i\left(u_j,v_j\right)\beta_0^{rgwr}\left(u_i,v_i\right)}{\sum_{i=1}^{n}w_i\left(u_j,v_j\right)}, \quad \dots (8)$$

$$\hat{\beta}_1^{rgwr}\left(u_j,v_j\right)=\frac{\sum_{i=1}^{n}w_i\left(u_j,v_j\right)\beta_1^{rgwr}\left(u_i,v_i\right)}{\sum_{i=1}^{n}w_i\left(u_j,v_j\right)} \quad \dots (9)$$

for all $j=1,2,3,\dots,N-n$.

Now, the predicted value of yield of CCEs plots at non-sampled points can be denoted as

$$\hat{y}_j=\hat{\beta}_0^{rgwr}\left(u_j,v_j\right)+\hat{\beta}_1^{rgwr}\left(u_j,v_j\right)x_j, \quad j=1,2,\dots,(N-n). \quad \dots (10)$$

For evaluating the performance of the proposed method, we considered all CCE data points used in the GCES scheme as a population and sampled a smaller portion of data points from it. We used the sampled data points to fit the ORGWR approach, which was then used to predict the remaining data points of the CCEs under GCES. We obtained ORGWR model parameter estimates from Eq. (6) at the sampled locations. Using these estimates, we predicted the intercept and slope parameters at non-sampled locations using Eq. (8) and Eq. (9), respectively.

Now, we estimated cotton yield using combined observed and predicted datasets, following a stratified three-stage random sampling methodology as recommended in CSO (2008). The sampling design followed for crop cutting experiments in the above-mentioned manual was a stratified three-stage random sampling with mandals/taluks/blocks/tehsils as strata, villages within the stratum as first-stage units (FSU), fields/survey numbers in the selected village as the second-stage units (SSU), and plots of specified size within the selected field as third-stage units. In this study, we used the data available in the Division of Sample Surveys, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, under the project entitled "Study to develop an alternative methodology for estimation of cotton production". In Amravati district, there are 14 tehsils in the Amravati district which are considered as strata in this study. The design based estimation of the survey data is described below.

Let,

y_{hij}= The plot yield in kg/plot of the j^{th} plot in the i^{th} village in the h^{th} stratum

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

m_h= Number of villages in which experiments are conducted in the h^{th} stratum

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

L = Number of strata in a district

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

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

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

and the district-level average yield per hectare is given by

$$\hat{\hat{Y}} = \frac{\sum_{h=1}^{L} a_h \bar{y}_h}{\sum_{h=1}^{L} a_h}. \qquad \dots (11)$$

The sampling variance of $\hat{\hat{Y}}$ is obtained as

$$\hat{V}\left(\hat{\hat{Y}}\right) = \frac{\left[W\sum_{h=1}^{L}\frac{a_h^2}{n_h} + \left(B-W\right)\sum_{h=1}^{L}\frac{a_h^2\sum_{i=1}^{m_h} n_{hi}^2}{\lambda_h n_h^2}\right]}{\left[\sum_{h=1}^{L} a_h\right]^2}, \qquad \dots (12)$$

where, $\lambda_h = \dfrac{n_h^2 - \sum_{i=1}^{m_h} n_{hi}^2}{n_h - \left(m_h - 1\right)}$;

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

mean square between villages and

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

mean square within villages.

The ORGWR approach was used to predict the yield of a portion of CCE's observations required for GCES. However, this predicted yield of CCE plots could introduce bias in estimating the average cotton yield. Therefore, the mean squared error (MSE) of the estimator of the average yield was calculated as

$$MSE\left(\hat{\hat{Y}}\right) = \hat{V}\left(\hat{\hat{Y}}\right) + \left[Bias\left(\hat{\hat{Y}}\right)\right]^2.$$

We assumed that the prediction bias at non-sampled values followed the same trend as at sampled data points obtained using the ORGWR approach. Thus,

$$\text{Bias}\left(\hat{\hat{Y}}\right) = \left[\hat{\hat{Y}}\right]_A - \left[\hat{\hat{Y}}\right]_P,$$

where

$\left[\hat{\hat{Y}}\right]_A$ was obtained using actual cotton yield data at the sampled CCE plots and

$\left[\hat{\hat{Y}}\right]_P$ was calculated using predicted cotton yield data of GCES's CCEs at the same locations using the ORGWR approach

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

$$\%S.E.\left(\hat{\hat{Y}}\right) = \frac{\sqrt{M\hat{S}E\left(\hat{\hat{Y}}\right)}}{\left(\hat{\hat{Y}}\right)} \times 100. \qquad \dots (13)$$

3. RESULTS AND DISCUSSIONS

In this study, the geo-spatial locations of the CCEs village of the Amravati district of Maharashtra were obtained using the GIS map (Fig. 1). The CCE data of the Amravati district of Maharashtra state consists of

341 villages under the cotton crop during the agriculture year 2012-2013. In Amravati district, there are 14 tehsils in the Amravati district which are considered as strata in this study. However, from the GIS map, we could only associate the geo-spatial locations to 316 villages. Therefore, the current study assumes that the CCEs data of the cotton crop of the Amravati district consists of 316 villages only for which we obtained geo-spatial locations. It was assumed that the village locations obtained from the GIS map correspond to the average yield of 2 CCE plots located in the same sampled village. It has been observed during the previous study done by Ahmad *et al.* (2013) that crop yield at the third picking of cotton crop has the highest correlation with the total yield of cotton crop. Hence, crop yield at the third picking of the cotton crop is used as an auxiliary variable in this study.

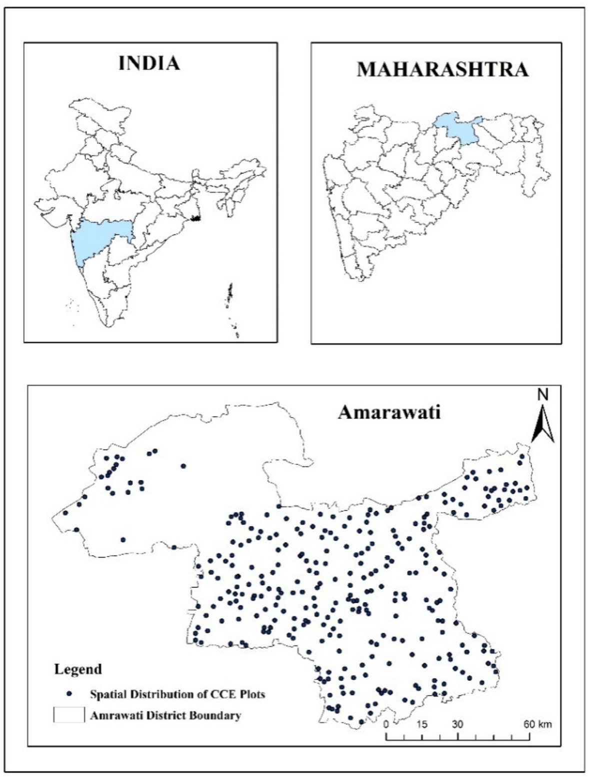

Fig. 1. Distribution of CCE villages selected under GCES for obtaining the yield rate of cotton in the Amravati District, Maharashtra for the year 2012-13

Further, we tested whether the GWR model fits the dataset better than the OLS model using the F_2 statistic method by Leung *et al.* (2000). Table 1 shows the statistical comparison between the GWR and OLS models in explaining the cotton crop yield dataset.

Table 1. Statistical indicator for comparing the GWR model over the OLS model for explaining the yield dataset of the cotton crop

F_2 statistic	Numerator DF	Denominator DF	Pr(>F)
2.2172	82.2709	314	4.731e-07 *

*: Represent that the test statistic found significant

Based on the findings presented in Table 1, the GWR model performed better than the OLS model. To test for spatial non-stationarity in the data, we employed F_3 statistics (Leung *et al.*, 2000), and the results are reported in Table 2 and Fig. 2.

Table 2. Statistical indicators for testing the presence of spatial non-stationarity in the dataset

Parameter	Value of F_3 statistic	Numerator DF	Denominator DF	Pr(>F)
Intercept	2.8510	121.5096	269.13	6.033e-13 *
Slope parameter	2.0568	57.2825	269.13	7.001e-05 *

Table 3. Summary of parameter estimates obtained using the GWR model

Parameter	Minimum	1st Quartile	Median	3rd Quartile	Maximum
Intercept	1.0318	5.4948	11.8060	16.3625	25.7648
Slope	1.2180	2.4333	3.1134	3.4336	5.1916

Tables 2 and 3 indicate that the parameters of the GWR model differ across the study area, suggesting the presence of spatial non-stationarity in the cotton yield data. In Table 4, we compare the performance of the GWR and OLS models in fitting the cotton crop yield dataset in the Amravati district of Maharashtra. The GWR model outperforms the OLS model with higher R_2 and Adjusted R_2 values and a lower residual sum of squares. Therefore, the GWR model provides a more accurate explanation of the cotton crop yield rate in the Amravati district of Maharashtra than the OLS model.

Table 4. Statistical indicators explaining the fitting of the GWR and the OLS model on the dataset

Parameter	GWR model	OLS Model
R^2	0.808	0.660
Adjusted R^2	0.759	0.659
Residual sum of squares	9702.542	17138.82

In this study, we aimed to estimate the yield rate of cotton crops using a smaller number of CCEs than the GCES scheme. To do this, we used the proposed ORGWR approach, which down-weights observations

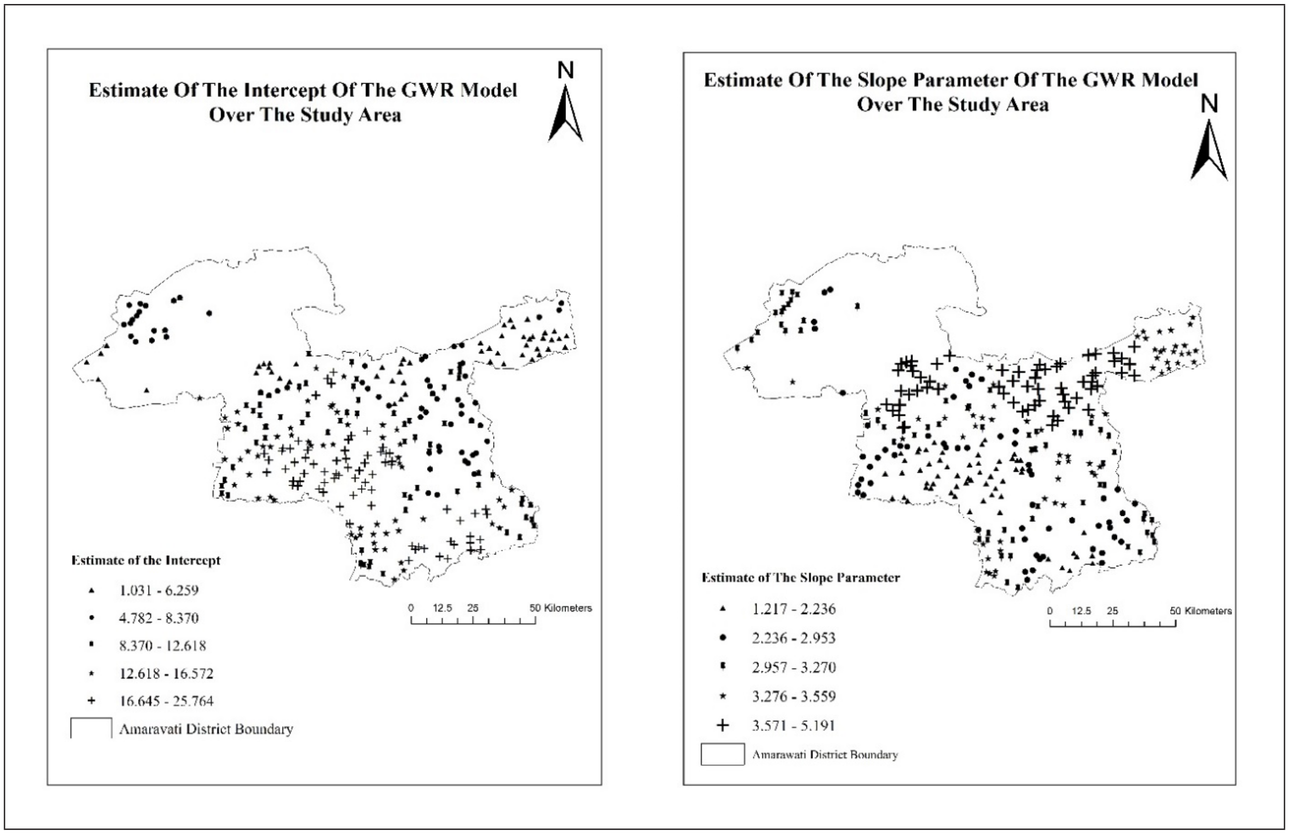

Fig. 2. Distribution of intercept and slope estimates in the GWR model for predicting cotton yield in unsampled CCE villages of Amravati District, Maharashtra

with large residuals when a GWR model is fitted using the CCE data of the Amravati district of Maharashtra for the agricultural year 2012-13. The Amravati district has a total of 316 CCE villages, and it consists of Tehsil as a stratum. To create our working datasets, we randomly retained 50% and 75% of CCE villages from each stratum. We then predicted the yield rate of the deleted observations using the proposed method (Eqn. 10) on the working data. After that, we obtained the combined dataset by combining the respective working data with predicted observations. The average yield of the cotton crop was estimated along with its MSE of the Amravati district using each combined dataset as well as the complete dataset of CCE villages. We then compared the result obtained from the combined dataset with that of the complete CCE dataset of the Amravati district. The estimates of average cotton yield (kg/ha) along with their percentage Standard Error (%SE) are presented in Table 5.

Table 5. Comparison of estimated average cotton crop yield using the GCES procedure and a combined dataset of original and predicted observations using the ORGWR approach

Sample Size (% of the original sample)	Average yield (kg/ha)	%SE
158 (50%)	512.707	0.868
237 (75%)	511.360	0.744
316 (100%)	510.871	0.714

Table 5 presents the average cotton yield estimates obtained through two different approaches i.e. the GCES procedure with a complete dataset and the ORGWR approach using a combination of incomplete datasets and predicted datasets. We used the ORGWR approach with a reduced number of CCE villages (50% and 75% of the complete dataset) to obtain the estimates of average cotton yield. The findings showed that the estimates obtained through the ORGWR approach are comparable to those obtained through the traditional GCES approach using the entire dataset, with reliable standard errors. We also calculated the standard errors of the average cotton yield estimates, which are reliable and decreasing as the sample size increases.

4. CONCLUSIONS

This study aimed to estimate the yield rate of cotton crop using a method that requires fewer CCEs than the GCES scheme. The proposed approach is called the ORGWR method. The study found that the estimates of the average yield of cotton using the ORGWR approach are comparable to the estimate obtained using the GCES methodology with the complete dataset, and it provides reliable standard errors. The study suggests that with a stratified three-stage random sampling framework with mandals/taluks/blocks/tehsils as strata, villages within the stratum as first-stage units (FSU), fields/survey numbers in the selected village as the second-stage units (SSU), and plots of specified size within the selected field as third-stage units, we can reduce the number of CCEs up to 50% of the CCEs needed for the GCES scheme and predict the rest of the CCEs using the proposed ORGWR approach for estimating crop yield by using 3rd picking of the cotton crop as an auxiliary variable. This method can produce an estimate of the average yield of cotton with a reliable degree of precision.

ACKNOWLEDGMENTS

The first author gratefully acknowledges the fellowship provided by the ICAR-Indian Agricultural Research Institute, New Delhi for pursuing the Ph.D. degree. The authors also acknowledge the suggestions provided by the anonymous reviewer which led significant improvement of the article.

REFERENCES

Afifah, R., Andriyana, Y. and Jaya, I.M. (2017). Robust geographically weighted regression with least absolute deviation method in case of poverty in Java Island. *InAIP Conference Proceedings*, **1827(1)**, 020023, AIP Publishing LLC.

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

Ahmad, T., Bhatia, V.K., Sud, U.C., Rai, A. and Sahoo, P.M. (2013). Study to develop an alternative methodology for estimation of cotton production. Project Report, ICAR-IASRI, New Delhi Publication.

Ahmad, T., Sud, U.C., Rai, A. and Sahoo, P.M. (2020). An alternative sampling methodology for estimation of cotton yield using double sampling approach. *Journal of the Indian Society of Agricultural Statistics*, **74(3)**, 217-226.

Brunsdon, C., Fotheringham, A.S. and Charlton, M. (1996). Geographically weighted regression: a method for exploring spatial non-stationarity. *Geographical Analysis*, **28(4)**, 281-298.

CSO. (2008). *Manual on Area and Crop Production Statistics*. CSOM-AG-01. Central Statistical Organization, New Delhi.

Harris, P., Fotheringham, A.S. and Juggins, S. (2010). Robust geographically weighted regression: a technique for quantifying spatial relationships between freshwater acidification critical loads and catchment attributes. *Annals of the Association of American Geographers*, **100(2)**, 286-306.

Leung, Y., Mei, C.L. and Zhang, W.X. (2000). Statistical tests for spatial non-stationarity based on the geographically weighted regression model. *Environment and Planning A*, **32(1)**, 9-32.

Ma, J., Chan, J.C.W. and Canters, F. (2014). Robust locally weighted regression for super resolution enhancement of multi-angle remote sensing imagery. *InIEEE journal of selected topics in applied earth observations and remote sensing*, **7(4)**, 1357-1371.

Moury, P.K., Tauqueer Ahmad, T., Rai, A., Biswas, A. and Sahoo, P.M. (2020). Outlier Robust FinitePopulation Estimation under Spatial Non-stationarity. *International Journal of Agricultural and Statistical Sciences,* **16(2)**, 535-545.

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

Paul, N.C., Rai, A., Ahmad, T., Biswas, A. and Sahoo, P.M. (2023a). GWR-assisted integrated estimator of finite population total under two-phase sampling: a model-assisted approach. *Journal of Applied Statistics*. DOI: 10.1080/02664763.2023.22808792024

Paul, N.C., Rai, A., Ahmad, T., Biswas, A. and Sahoo, P.M. (2024). Spatial approach for the estimation of average yield of cotton using reduced number of crop cutting experiments. *Current Science*, 125(5), 518-529. DOI: 10.18520/cs/v125/i5/518-529

Paul, N.C., Rai, A., Ahmad, T., Biswas, A. and Sahoo, P.M. (2024). Spatially integrated estimator of finite population total by integrating data from two independent surveys using spatial information. *Journal of the Korean Statistical Society*, **53**, 222-247.

Raheja, S.K., Goel, B.B.P.S., Mehrotra, P.C. and Rustogi, V.S. (1977). Pilot sample survey for estimating yield of cotton in Hissar (Haryana). Project Report, IASRI, New Delhi Publication.

Saha, B., Biswas, A., Ahmad, T. and Paul, N.C. (2023). Geographically weighted regression-based model calibration estimation of finite population total under geo-referenced complex surveys. *Journal of Agricultural, Biological and Environmental Statistics,* https://doi.org/10.1007/s13253-023-00576-9.

Warsito, B., Yasin, H., Ispriyanti, D. and Hoyyi, A. (2018). Robust geographically weighted regression of modeling the Air Polluter Standard Index (APSI). *Journal of Physics: Conference Series*, **1025(1)**, 012096: IOP Publishing.

Zhang, H. and Mei, C. (2011). Local least absolute deviation estimation of spatially varying coefficient models: robust geographically weighted regression approaches. *International Journal of Geographical Information Science*, **25(9)**, 1467-1489.