Small Area Estimation for Crop Acreage in Remote Sensing Assisted Crop Survey

— A Case of Major Crop Acreage Estimation in Liaozhong County

Wei Zhou

Department of Rural Surveys, National Bureau of Statistics of China

75 Yuetan Nanjie, Beijing, China

weizhou@263.net

DOI: 10.1481/icasVII.2016.g44

ABSTRACT

Applying remote sensing to estimate the planted acreage is typically by using regression estimator or calibration estimator which is taking advantages of combining the sample data from ground survey with satellite image classification. For most cases of regression estimator, the crop acreage estimation for a target population such as a province or county only satisfies the precision for itself but could not be disaggregated to small areas, such as county and town level statistics.

Taking the Liaozhong county affiliated to Shenyang city, Liaoning province as a study area, the satellite images from the moderate resolution Landsat 8 OLI is classified for rice and corn as auxiliary information for population, and high resolution Chinese GF-1 and ZY-3 are visually interpreted for rice and corn which is regarded as ground truth. The whole county is segmented into grids of 100m*100m as sampling units, a simple random sampling is adopted to select samples with replicates 1000 times to do simulation of building small area models. From our simulation study, it reveals that a basic level small area model in the form of one-response multiple regression with random effects and fixed effects are both feasible to produce the estimates at town and township level. Meanwhile, the aggregate of estimates of every town could (approximately) be the estimates for the county under the assumption of linear regression. It is concluded that the small area estimation method is applicable to solve crop acreage estimation from province to county level simultaneously when targeting an entire province.

Keywords: Crop Acreage, Small Area Estimation, Remote Sensing

1. Introduction

China is a major agricultural production country as well as a large consumption and trade country in terms of farm products. Depending on the various agricultural information especially on the planted acreage statistics of major grain crops, it has been an important gist for decision making with respect to grain and food policy as well as economic development plans at national level. To acquire the planted acreage of major grain crops and their spatial distribution in a timely, accurate, quantitative way is of significance for making decision on agricultural production at various levels of governments, in order to ensure food security, steer and adjust the crop planting structure through macro-economic control, as well as improve the operational management for relevant enterprises and farms (Blaes and Vanhalle et al.; Tao and Yokozawa et al.; Chauhan and Arora et al.). The traditional approach to obtain crop acreages are usually conducted by a sample survey usually by a national statistics agency. However, this process is time-consuming, labor-intensive and lacking in spatial information (Ma et al.). Remote sensing has been used for crop acreage estimation over the last few decades and is considered to be an effective tool for detecting the crop area extents and changes at regional or global scales (Hall and Badwar; Yang et al.; Xiao et al.). Some countries has already conducted a series of operational programs which aims at the estimation for land cover/land use and crops acreage, such as the USA (LACIE,1974-1977; AGRISTARS,1980-1986; CDL,1997-2010), European Union (MARS,1998; LUCAS, 2001-2009; Geoland2, 2008(2011)) and ROK (Implement RS Application System, IRSAS, 2008-2012).

Similarly to the American and European remote sensing programs for crop acreage estimation, since year 2010 the National Bureau of Statistics (NBS) of China has collaborated with external research institute to preliminary establish an operational business mode—Remote Sensing Assisted Crop Survey (RSACS) for major provinces in terms of grain outputs in China. The RSACS basically involves (1) selecting sample segments for each individual province based on area frame, (2) conducting field survey and collecting crop data, (3) crop spatial classification based on moderate satellite images, (4) crop acreage estimation.

Many studies have been carried out on crop acreage estimation for remote sensing assisted crop survey. To produce the estimates of crops acreage for a target population, there are usually three approaches could be adopted (F.J. Gallego). The first is direct expansion estimator which is a typical sampling approach to produce the estimates by only using survey samples. The second is called calibration estimator which is based on the crop classification result from remote sensing through adjusting the classification by confusion matrix. The third is regression estimator which refers to build a linear regression model by combining the ground survey data with the classification results from remote sensing. Theoretically, the regression estimator could gain precision of estimation due to combing the ground survey data with crop classification from remote sensing, which is taken as an auxiliary information for population. The regression estimator is essential derived from linear regression method which could be widely adapted to a particular problem solving, for example, Yaozhong Pan proposed an approach to estimate winter wheat area through building regression on Crop Proportion Phenology Index (CPPI) which disaggregated from the MODIS vegetation index (VI) (Pan et al.). The Chinese RSACS has also developed a feasible solution on acreage estimation for major crops such as wheat, corn, rice at provincial level by using linear regression method. However, the current methods of crop acreage estimation mostly concentrate on producing the estimates for the target population (province) but could not be disaggregated to sub-regional (county) levels.

In the Chinese context of remote sensing assisted crop survey, the sample segments and crop classification are directly linked to meet the requirements for the target province, while the estimates for crop acreages are expected to meet the precision for both province and county level. From literature study, small area estimation is the most dominant approach to solve the multi-level estimation such as estimates for both county and province level simultaneously. In 1988 a nested-error small area model is specified for the relationship between the reported hectares of corn and soybeans within sample segments from ground survey and the corresponding satellite determination for areas under corn and soybeans, and predictions of mean hectares of corn and soybeans per segment for the 12 Iowa counties are presented (Battese G.E. et al). In 2003 small area estimation is applied in a survey conducted in the Rathbun lake Watershed in Iowa, erosions are estimated for 61 small areas within the study region (Opsomer J.D. et al). In 2006 Small area model which included sampling and model weights was proposed and applied in agriculture for predicting minor crops (Militino A.F.).

In this study, we attempt to produce the crop acreage estimation at multi-levels by using method of small area estimation. The study area is chosen in Liaozhong county which is affiliated to Shenyang city, Liaoning province of China. Two kinds of spatial resolution satellite images has been acquired and processed for year 2014. The visual interpretation results from high resolution GF-1 and ZY3 for rice and corn was regarded as ground truth, while extracted crop classification for rice and corn from the moderate resolution Landsat 8 OLI was taken as the auxiliary information for population. We explored the small area model of basic unit level with random effects and fixed effects respectively by combining the ground truth data with auxiliary information of population. Given different sampling fraction under simple random sampling by using Monte Carlo simulation, the result of coefficient of variation (C.V) derived both from the model based mean square error (MSE) and simulation based MSE are computed and make comparison.

2. Experimental Area and Procedures

2.1 Experimental area

We choose the Liaozhong county, which is affiliated to Shenyang city of Liaoning province, as experimental area or target population in sampling terminology. There are altogether 20 towns and townships within the Liaozhong county, regarded as small areas or domains in our study. Liaozhong located in 122°28~123°6′ longitude east and 41°12′~41°47′ latitude north, having the total territory area around 1460 square kilometers. Liaozhong is situated in lower tier of Liaohe river watershed, belong to a washed plain impacted by Laiohe river and Hunhe river. On annual average, the cumulative sunshine hours are 2575 hours and cumulative precipitation is 640 millimetre, it is suitable for agricultural production. The major species for grain crops are rice, corn and soybean.

2.2 Moderate spatial resolution images

For large scale crop monitor program by using remote sensing, usually the moderate resolution satellite images are adopted to extract optical spectrum to distinguish the crops. In this study, Landsat 8 OLI satellite imagery on June 4, August 7 and September 8 are acquired respectively, and geometry correction as well as radiation correction are processed. Based on post-processed standard orthophoto images, a maximum likelihood classifier (MLC) is adopted to extract the rice and corn on pixel-base.

The three temporal images are classified for rice and corn individually by using MLC, and then the final distinguished crops of rice and corn are obtained by detecting the dynamic change of three temporal results. In the process of MLC applied to distinguish different crops, the real crop survey data in 2014 from Liaoning Survey Organization of National Bureau of Statistics(NBS) had been used as training set of samples when applying MLC. From the ongoing crop sample surveys conducted in Liaozhong county by the survey organization of NBS, we had 15 sample villages and each village allocated 5 sample segments in the size of 2 hectares. Altogether there are 75 sample segments of survey data could be used as ground truth in the process of satellite image classification. Finally the classified results and spatial distribution for rice and corn are obtained in format of vector data, which covers the whole county and serves as auxiliary information of population.

2.3 High spatial resolution images

We acquired the Chinese GF-1 and ZY-3 satellite images (high spatial resolution imagery) to be visually interpreted to distinguish corn and rice corresponding to each arable field or plot respectively. Subject to the limitation of acquisition of qualified satellite images, we have acquired one temporal GF-1 images dated on October 7, 2014 and one temporal ZY-3 images dated on June 4 and June 14, 2014 before the autumn harvest. The acquired high resolution images are pre-processed by geometry and radiometric correction as well as a fusion of 2m panchromatic and 8m multi-spectral images. Taking the advantages of the coverage of delineated cropland for this county being obtained from a previous land survey, it facilitated the visual interpretation to distinguish of rice and corn within the cropland effectively. The result of visual interpretation is used as an approximate ground truth for corn and rice, and is also applied to assess the accuracy of crop classification from moderate resolution images.

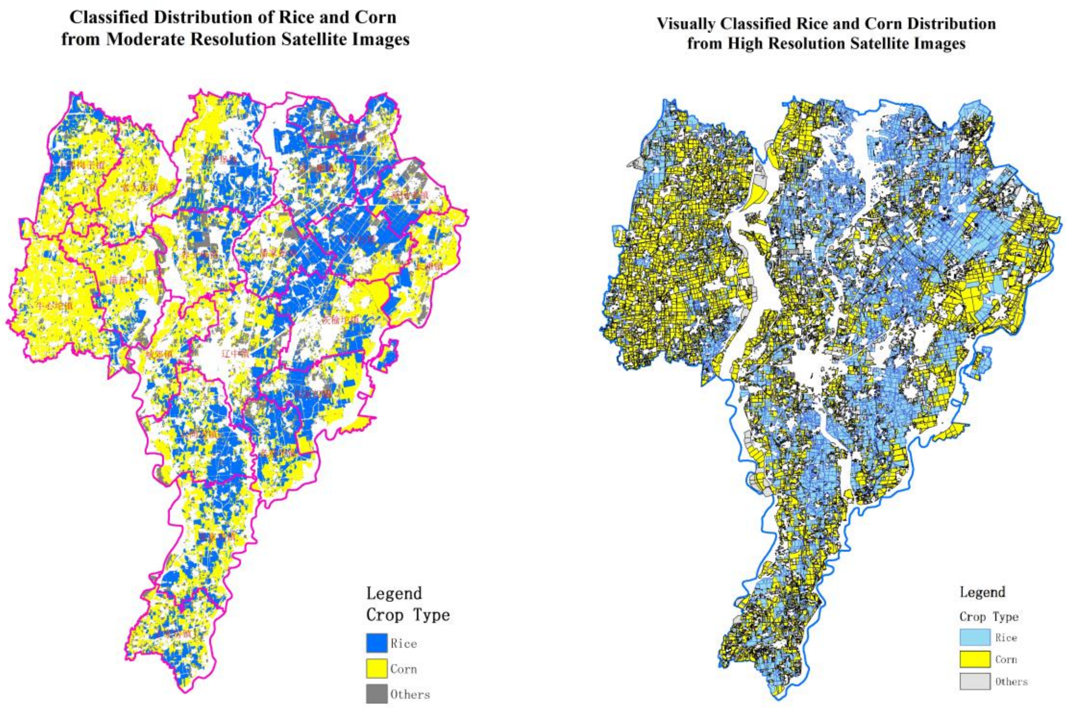

Figure 1: *Planted distribution of rice and corn from moderate and high resolution satellite images respectively.*

2.4 Sampling design

With the facility of GIS system, we delineated the whole territory of Liaozhong county into squared grids in size of 100 meters×100 meters. There are altogether 138279 grids, which are regarded as sampling units. A simple random sampling is adopted to select samples, and sampling fraction are chosen as 0.07%, 0.14%, 0.25%, 0.5%, 1%, 2%, 5% respectively to make comparison. In our study, the crop classification from the satellite images of Landsat 8 is taken as the auxiliary information for population, while the crop interpretation from the GF-1 and ZY-3 is regarded as data of ground truth.

2.5 Analysis for multi-level estimation

We probe the multi-level estimation for major crop acreage in the context of remote sensing assisted survey in Liaozhong county, which aims to produce the crop acreage estimates at town level and simultaneously generate the estimates for the county. The method of small area estimation is adopted, and a basic unit level model for small area is specified. The small area model is set in two scenarios for the effects of small area, one is the random effects and the other is fixed effects. Given the sampling scheme of simple random, a replicate of 1000 samples are selected for crop acreage estimation and consequent precision assessment. For each town and township, the estimate precision derived from the small area model is compared with the precision from direct domain estimation. In our study, we present simulation results as the following: (1) Computing the MSE of the model according to parameter estimation such as Empirical Best Linear Unbiased Prediction (EBLUP) and its CVs, (2) Based on the estimates from each individual model from replicates, computing the MSE and its CVs by simulation, (3) Based on the direct domain expansion, computing the mean squared error (MSE) and its coefficient of variation (CVs).

3. Multi-level Estimation by Small Area Model

3.1 Small area model with random effects

3.1.1 Model setting

For the crop of interest y_{ij}, assuming the random effects for small area, the adopted unit level small area model is as follows:

$$y_{ij} = \mu + X_{ij}{}^T \beta + v_i + e_{ij}, \quad j = 1, 2, ..., n_i \ , \ i = 1,2,...,m \tag{1}$$

Where, y_{ij} is a specified crop (rice or corn) acreage for the jth sample grid in the ith town and township, m is the total number of towns and townships, in our case m=20, while n_i is the sample size in the ith town and township, μ is the intercept in the model, $X_{ij} = (x_{ij1}, ..., x_{ijp})'$ refers to the crop classification results from the moderate satellite which is composed of the variables $x_{ij1}, ..., x_{ijp}$. Random effects v_i and error term e_i are independent and identically distribution, subject to normal distribution with mean 0 and identical variance σ_v^2 and σ_e^2 respectively.

From the generalized ordinary least square method, we have

$$\hat{\beta} = \left(\sum_{i=1}^{m} X_i' V_i^{-1} X_i \right) \left(\sum_{i=1}^{m} X_i' V_i^{-1} Y_i \right) \tag{2}$$

$$\hat{v}_i = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2 / n_i} (\bar{y}_i - \bar{x}_i \hat{\beta}) \tag{3}$$

Where, $V_i=\sigma_v^2 1_{n_i}' 1_{n_i}+\sigma_e^2 I_{n_i}$, $X_i=(X_{i1},...,X_{in_i})'$, $Y_i=(y_{i1},...,y_{in_i})'$, $\bar{x}_i$ is the sample mean for the auxiliary information of the ith town and township.

Hence, the estimate for the sub-population total of the ith town and township is as follows:

$$\hat{Y}_i = N_i[\bar{X}_{ui}\hat{\beta}+\frac{\sigma_v^2}{\sigma_v^2+\sigma_e^2/n_i}(\bar{y}_i-\bar{x}_i\hat{\beta})] \tag{4}$$

Where, $\bar{X}_{ui}$ is the sub-population mean of classified result of rice or corn from moderate satellite images of the ith town, N_i is the total number of grids in the ith town and township.

For the variance estimation of random effects v_i and error item e_i, we adopt the moment method introduced by Fuller and Battese (1973):

$$\hat{\sigma}_e^2=SSE(1)/(n-m-p_1),\quad \hat{\sigma}_v^2=\max\{[SSE(2)-(n-p)\hat{\sigma}_e^2]/\eta,0\}$$

Where, $SSE(1)$ is the sum of squared residuals from the regression $y_{ij}-\bar{y}_i$ on $X_{ij}-\bar{x}_i$, $SSE(2)$ is the sum of squared residuals from the regression y_{ij} on X_{ij}, p_1 is the number of non-zero in $X_{ij}-\bar{x}_i$, $\eta=\sum_{i=1}^{m} n_i - n_i^2\bar{x}_i(\sum_{i=1}^{m} X_i' X_i)^{-1}\bar{x}_i'$. If the estimates of variance plugging into formula (3) and (4), then we have the total estimate $\hat{Y}_i^E$ for the ith town and township by EBLUP estimation:

$$\hat{Y}_i^E = N_i[\bar{X}_{ui}\hat{\beta}^E+\frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2+\hat{\sigma}_e^2/n_i}(\bar{y}_i-\bar{x}_i\hat{\beta}^E)] \tag{5}$$

Where, $\hat{\beta}^E$ is the estimate of β when plugging into the estimates of variance in formula (2). Alternatively the parameter estimates could be estimated by the ML or REML method, there are no significant difference among the estimates.

3.1.2 MSE of EBLUP at town level

For each small area (town) of the model, the MSE of Empirical Best Linear Unbiased Prediction depends on the model, Rao(2003) gave the formula as follows.

$$MSE(\hat{Y}_i)=g_{1i}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)+g_{2i}(\hat{\sigma}_v^2,\hat{\sigma}_e^2)+2g_{3i}(\hat{\sigma}_v^2,\hat{\sigma}_e^2) \tag{6}$$

Where the definition of $g_{1i}(.)$, $g_{2i}(.)$ and $g_{3i}(.)$ are referred to Rao's literature Small Area Estimation.

3.1.3 Population total at county level

In the random effects model, the sum of acreage estimate Y_i for the total number of small area $i=1$ to m is approximately equal to the total acreage at county level in the assumption of general linear regression. In fact, sum of the expectation of each Y_i is equal to expectation of total $\hat{Y}$. Due to the $E(v_i)=0$, we have the follows.

$$E(\hat{Y})=E(\hat{Y}_1)+E(\hat{Y}_2)+...+E(\hat{Y}_m) \quad (7)$$

3.2 Small Area model with fixed effects

3.2.1 Model setting

The formation of prediction model with fixed effects of small area is the same as the model in equation (1). The only difference is that v_i is assuming as fixed effects instead of random effects, that is to say for each small area (town) there is a fixed value of v_i in the model.

By differential method for the parameter estimation, we have:

$$\hat{\beta}=\left(\sum_{i=1}^{m} X_i^T Q_i X_i\right)\left(\sum_{i=1}^{m} X_i^T Q_i Y_i\right) \quad (8)$$

$$\hat{v}_i=\bar{y}_i-\bar{x}_{\mathbf{i}}^{\mathbf{T}}\hat{\beta}-\bar{y} \quad (9)$$

Where $Q_i=I_{\mathbf{n}_i}-\frac{1}{n_i}1_{\mathbf{n}_i}1_{\mathbf{n}_i}^T$, $X_i=(x_{i1}^T,...,x_{in_i}^T)^T$, $Y_i=(y_{i1},...,y_{in_i})^T$, $\bar{x}_i$ is the sample mean of auxiliary information from the ith town and township.

The estimate for the population mean of small area is as follows:

$$\hat{\hat{Y}}_i=\bar{X}_{ui}^T\hat{\beta}+(\bar{y}_i-\bar{x}_i^T\hat{\beta}) \quad (10)$$

Where $\bar{X}_{ui}$ is the sub-population mean of classified rice or corn from moderate satellite images of the ith town. $\bar{y}_i$ is the sample mean of the ground truth for rice or corn of the ith town. $\bar{x}_i$ is the sample mean of the classified rice or corn of the ith town.

Then the estimate for the population total of small area is $N_i\hat{\hat{Y}}_i$.

The estimate for the error term e_i is as follows:

$$\hat{\sigma}_e^2=SSE(1)/(n-m) \quad (11)$$

Where, $SSE(1)$ is the sum of squared residuals from the regression $y_{ij}-\bar{y}_i$ on $X_{ij}-\bar{x}_i$.

3.2.2 MSE of model prediction at town level

For each small area (town), the MSE which depends on the model is as follows:

$$MSE(\hat{Y}_i)=E(\hat{Y}_i-Y_i)^2=n_i^2(\bar{X}_i-\bar{x}_i)^T(\sum_{i=1}^{m}X_i^TQ_iX_i)^{-1}(\bar{X}_i-\bar{x}_i)+n_i\hat{\sigma}_e^2 \quad (12)$$

Where, $\hat{Y}_i$ is the estimate of crop acreage from the ith town, Y_i is the approximately real planted acreage which is visually interpreted from the high resolution satellite imagery. $\bar{X}_i$ and $\bar{x}_i$ are sub-population mean and sample mean of the crop acreage which are classified from moderate satellite imagery. The definition of X_i is the same as in formula (8) , and the definition of $\hat{\sigma}_e^2$ is the same as in formula (11).

3.2.3 Population total at county level

In the fixed effects model, for total estimates of each town and township level, we have:

$$\hat{Y}_i = N_i \bar{X}_{\mathbf{ui}}^{\mathbf{T}} \hat{\beta} + N_i (\bar{y}_i - \bar{\mathbf{x}}_{\mathbf{i}}^{\mathbf{T}} \hat{\beta}) \tag{13}$$

To sum up Y_i over the number of m towns and townships, we have the total estimates for entire county as follows:

$$\begin{aligned} \hat{Y} &= \sum_{i=1}^{m} [N_i \bar{X}_{\mathbf{ui}}^{\mathbf{T}} \hat{\beta} + N_i (\bar{y}_i - \bar{\mathbf{x}}_{\mathbf{i}}^{\mathbf{T}} \hat{\beta})] \\ &= \sum_{i=1}^{m} N_i \bar{y}_i + \sum_{i=1}^{m} N_i (\bar{X}_{\mathbf{ui}}^{\mathbf{T}} - \bar{\mathbf{x}}_{\mathbf{i}}^{\mathbf{T}}) \hat{\beta} \end{aligned} \tag{14}$$

Therefore, the sum of acreage estimates Y_i for the total number of small area $i=1$ to m is exactly equal to the total acreage at county level in the assumption of general linear regression. Similarly, under the context of sampling weight to be applied, the sum of acreage estimates over all of small areas is also exactly equal to the total acreage at county level in the assumption of general linear regression.

4. Simulation Results

4.1 Experimental Data

For the experimental area Liaozhong county, we have delineated the whole territory into grids in size of 100 meters ×100 meters. There are altogether 138279 grids which is used as the sampling units (PSU). The information used in small area model building are as follows (Table 1), among them column (5) and (6) are used as dependent variables, column (7) and (8) are used as independent variables. In our study, the Liaozhong county is the target population, and its affiliated 20 towns and townships are regarded as small areas (domains). Two major crops of rice and corn are our variables of interest.

Table 1: *Crop area from visual interpretation and classification by town and township*

Unit:Hectare

Town ID (1)	Town Code (2)	Town Name (3)	Number of PSU (4)	Rice Area: Truth (5)	Corn area: Truth (6)	Rice Area from OLI8 (7)	Corn Area from OLI8 (8)
1	210122100	Liaozhong	6647	1750.008	2129.939	1352.919	1990.517
2	210122101	Ciyutuo	8434	3020.741	2931.661	2646.791	2244.159
3	210122102	Yujiafang	6307	1466.566	2458.064	1277.625	1535.788
4	210122103	Lengzipu	9796	3714.180	3016.737	3256.117	2997.363
5	210122104	Manduhu	7761	1072.259	4020.944	793.871	3767.203
6	210122105	Zhujiafang	8781	2642.468	3605.698	2330.282	2354.021

7	210122106	Liuerpu	5990	2856.171	1038.542	2519.282	785.550
8	210122107	Xinmintun	3792	1623.280	647.918	1337.879	698.411
9	210122108	Yangshigang	6335	3020.798	813.376	2870.082	712.628
10	210122109	Xiaojiamen	6492	2608.219	2228.829	2337.943	2175.943
11	210122110	Changtanzheng	5627	930.406	3164.420	839.103	2677.139
12	210122111	Sifangtai	6229	3216.909	1263.475	2985.907	981.072
13	210122200	Chengjiaozhen	4780	757.650	2209.059	542.686	2130.448
14	210122201	Liujianfang	8470	3246.437	2481.732	2885.688	2193.679
15	210122202	Yangshipu	6342	1692.727	1906.360	1380.257	1951.681
16	210122203	Panjiapu	5381	2236.938	1198.193	1901.378	1083.401
17	210122204	Laoguantuo	5235	2218.991	1433.240	1970.074	1585.276
18	210122205	Laodafang	7179	965.132	3844.526	478.141	3835.990
19	210122206	Dahei Gangzi	7917	2257.346	3573.115	1469.630	3720.634
20	210122207	Niuxintuo	10784	1201.306	6568.248	789.193	6389.981
Total			138279	42498.532	50534.077	35964.847	45810.884

We adopted simple random sampling to select samples, and designated seven different sampling fraction respectively as follows: (1) 0.07%, (2)0.14%, (3)0.25%, (4)0.5%, (5)1%, (6) 2%, (7) 5%.

With respect to each sampling fraction, we do the simulation by selecting the samples k=1000 times. In our case, we adopted small area model to estimate the rice and corn acreage for each town and township. We attempted two scenarios for the small area with random effects and fixed effects respectively. Meanwhile, In order to assess the estimate precision from the method of small area model, we compared the results with that from the direct expansion estimation for small area.

For the 1000 replicate samples, in order to compare the estimate precision of crop acreage, we computed the MSE and C.Vs with respect to rice and corn acreage for each town and township by the following three approaches: (a) direct expansion for domain estimates; (b) model prediction (EBLUP) from the small area model; (c) the estimates (prediction) for each town and township based on each individual small area model from replicates.

For the above approaches (a) and (c), the mean square error (MSE) for each small area i could be computed by replicates.

$$MSE_s(\hat{Y}_i) = E(\hat{Y}_{ir} - Y_{ir})^2 = \sum(\hat{Y}_{ir} - Y_{ir})^2 \Big/ k \qquad (15)$$

Where, $\hat{Y}_{ir}$ is the rice or corn planted acreage estimated from the rth replicate of the ith town and township. While Y_{ir} is the rice or corn truly planted acreage from the rth replicate of the ith town and township. Converting to coefficient of variation (CV), we have:

$$C.V(\hat{Y}_i) = \frac{\sqrt{MSE_s(\hat{Y}_i)}}{Y_i} * 100\% \qquad (16)$$

For the above approach (b), for each replicate we got $MSE_r(\hat{Y}_i)$ and $C.V_r(\hat{Y}_i)$ from model itself, then taking the average of all these $C.V_r(\hat{Y}_i)$ as a result of $C.V(\hat{Y}_i)$.

4.2 Simulation results for rice and corn

Referring to the formation of equation (1), y_{ij} refers to rice or corn real planted acreage from approximate ground truth, X_{ij} refers to rice and corn acreage which is classified from image. There are m=20 towns and townships in Liaozhong county. We illustrate the simulation results from the small area model with random effects and fixed effects.

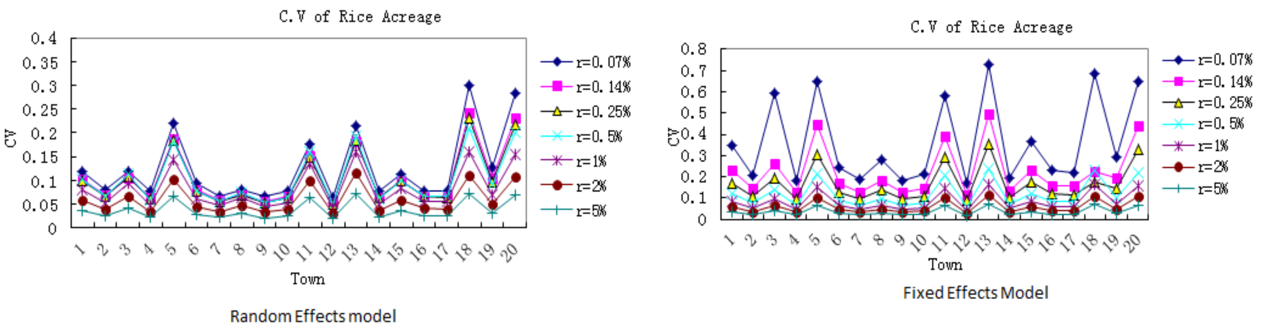

Figure 2: *C.Vs of rice acreage from MSE of Prediction for small area model*

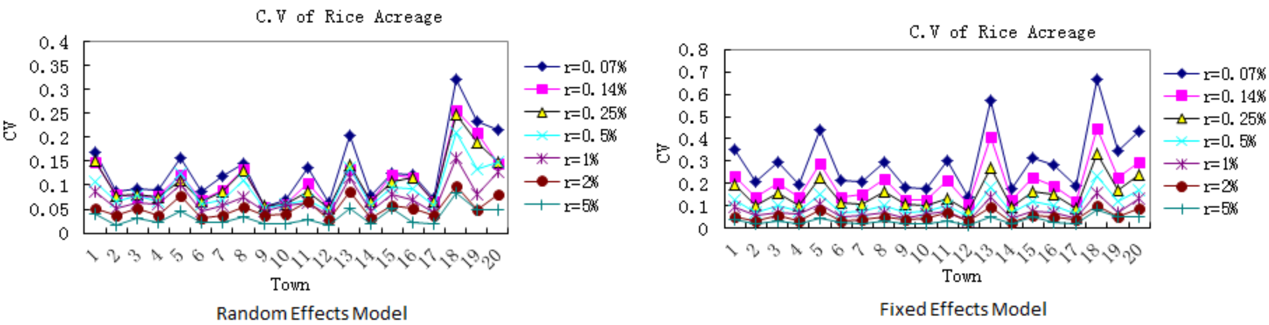

Figure 3: *C.Vs of rice acreage from MSE based on individual estimates derived from models*

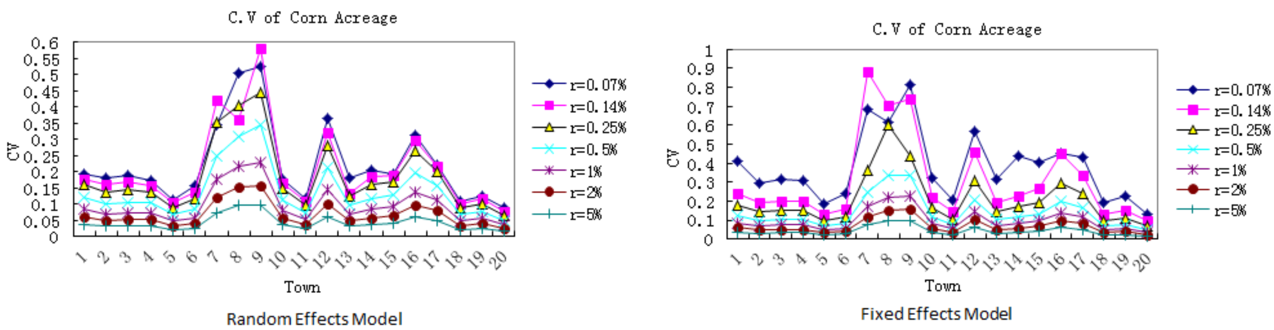

Figure 4: *C.Vs of corn acreage from MSE of Prediction for small area model*

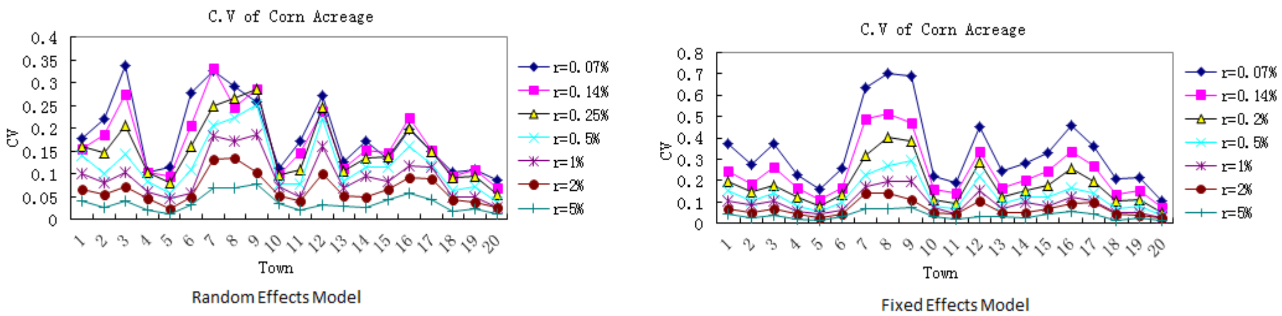

Figure 5: *C.Vs of corn acreage from MSE based on individual estimates derived from models*

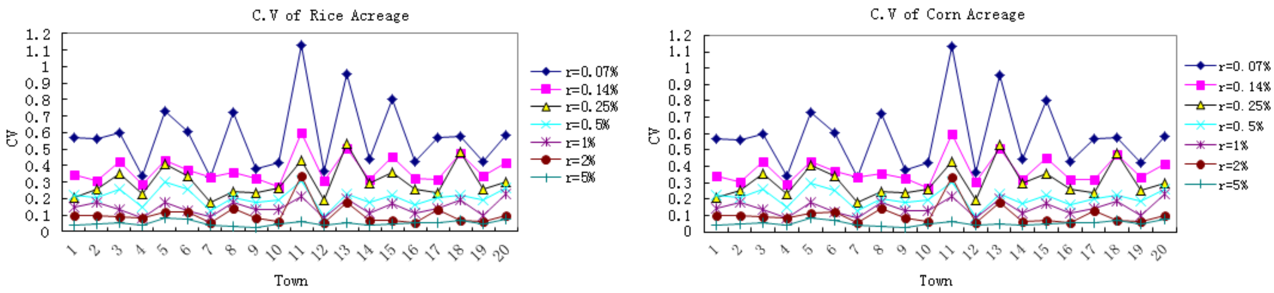

Figure 6: *C.Vs from the direct expansion for domain estimates*

The above Figure 2 and Figure 3 are C.Vs of rice acreage computed from model based MSE (approach b) and simulation based MSE (approach c) respectively. Similarly, Figure 4 and Figure 5 are C.Vs of corn acreage computed from model based MSE (approach b) and simulation based MSE (approach c) respectively. For small area model, if model fitness is good enough, then the MSE derived from the model should be very close to the MSE calculated from the simulation. To simplify the assessment of estimate precision for small areas, we could focus on the C.Vs results of Figure 3, Figure 5 and Figure 6 to make comparison.

Obviously, the precision of estimates gain significant improvement for small areas either in random effects model or fixed effects model especially for a smaller sampling fraction, compared with the direct domain estimation approach. For small area model with fixed effects, it holds on an additive property under the assumption of linear model that the aggregate of all estimates of each town and township is strictly equal to the population total for the whole county. While for the small area model with random effects, the above equilibrium is approximately hold on.

In our simulation, we also examined the C.Vs of aggregate at county level from summation of individual estimates of towns and townships in either random effects or fixed effects model. The C.Vs are even outperformed than that of a direct expansion at county level (Table 2).

Table 2: *C.V of rice and corn acreage at county level*

Sampling Fraction	Direct Expansion: Ground truth		Model Simulation Random Effects		Model Simulation Fixed Effects	
	C.V of Rive	C.V of Corn	C.V of Rice	C.V of Corn	C.V of Rice	C.V of Corn
r=0.07%	14.60%	12.28%	8.22%	10.74%	6.06%	5.77%
r=0.14%	8.62%	6.65%	7.21%	8.69%	4.20%	4.12%
r=0.25%	7.20%	5.26%	6.76%	6.75%	3.04%	3.32%
r=0.5%	4.82%	3.62%	4.78%	3.83%	2.14%	2.53%
r=1%	4.21%	3.12%	2.55%	2.20%	1.35%	1.78%
r=2%	2.25%	1.81%	1.53%	1.25%	1.15%	1.02%
r=5%	1.18%	1.08%	0.79%	0.71%	0.71%	0.63%

5. Discussion and Conclusion

5.1 Precision Efficiency

Especially in the scenario of small sampling fraction, our study illustrates that model fitness of small area model with random effects is better than that of model with fixed effects, which is reflected in the results of statistical test for parameters as well as the efficiency gain of precision. To examine the precision of estimates for each individual town and township, it is obviously that the C.Vs (defines as in section 4.1) is relatively lower when applying random effects model compared to fixed effects model especially for a smaller sampling fraction. When sampling fraction is 0.07%, taking the average C.Vs of rice acreage and corn acreage for all the 20 towns as indicators, the ratio of the C.Vs from random effects over the C.Vs from fixed effects is around 44% and 60% respectively, this implies that model fitness of random effects in this scenario is better. When sampling fraction is increased to 5%, the ratio of the C.Vs from random effects over the C.Vs from fixed effects is around 98% and 99%, this implies there is no significant difference between random or fixed effects model. On average, the ratios corresponding to all the other scenarios of sampling fraction lies in between of lower level 44% and upper level 99%.

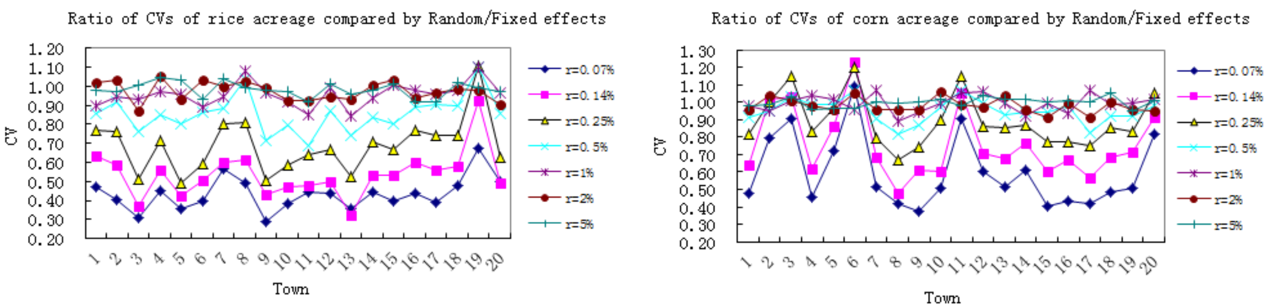

Figure 7: *The ratio of CVs from MSE under model with random effects versus fixed effects*

5.2 Model Fitness

In real situation, the setting of small area model either with random effects or fixed effects mostly depends on the user's assumption and understanding to the problem solving. In econometrics, usually the Hausman test is used to check whether there is a fixed or random effects preferred to the underlined component. But the validity of the Hausman test not always guaranteed, that means sometimes it is difficult to determine whether to choose fixed effects or random effects. In this study, even if the small area model with fixed effects does not have a sufficient model fitness compared with that of model with random effects especially when a smaller sampling fraction, it is also robust to predict a reasonable results for each town and township when we scrutinize the estimates and its C.Vs. Therefore, it implies that small area model with either random effects or fixed effects are mostly feasible to produce estimates for domains like town and township in this case.

With regard to the multi-level estimation both for the town and county level simultaneously, for the model with fixed effects the summation of each estimate of towns exactly equals the estimate for the whole county under the assumption of general linear regression. While for the model with random effects the summation of each estimate of towns approximately equals the estimate for the whole county under the assumption of general linear regression.

5.3 Model Sensitivity of Classification Accuracy

As we have seen, small area model of basic unit level with random effects or fixed effects is constructed by combining the ground truth data with classified image data regarded as auxiliary information of population. Since the satellite image classification affected by many factors such as the image quality and classification methods, the accuracy of classification may varied from good to medium. Given the certain sampling fraction, the model precision for estimates at town level could be impacted by the classification accuracy.

Based on the originally classified vector data from Landsat 8 OLI, which overall accuracy for rice and corn are around 85% and 90% respectively. In order to simulate different accuracies for classification, we added a disturbing term which is a certain constant multiply a random number subject to the standardized normal distribution for each grids. Then we calculated the accuracy of classification again by comparing the pseudo classified data with ground truth from visually image interpretation.

For the random effects model, the average C.Vs of rice acreage and corn acreage at town level corresponding to various accuracies are shown in Figure 8. Taken rice as an example, when sampling fraction is 0.07% and classification accuracy is 80%, the average C.Vs at town level is around 10.7%. Given the classification accuracy at 65%, in order to obtain the same average C.Vs at around 10.7%, we need to have the sampling fraction at 0.14% which means a doubled sample size. It reveals that both classification accuracy and sample size determines an expected CVs at town and township level. In practice, we need to consider the trade-off of more accurate classification or field work of samples.

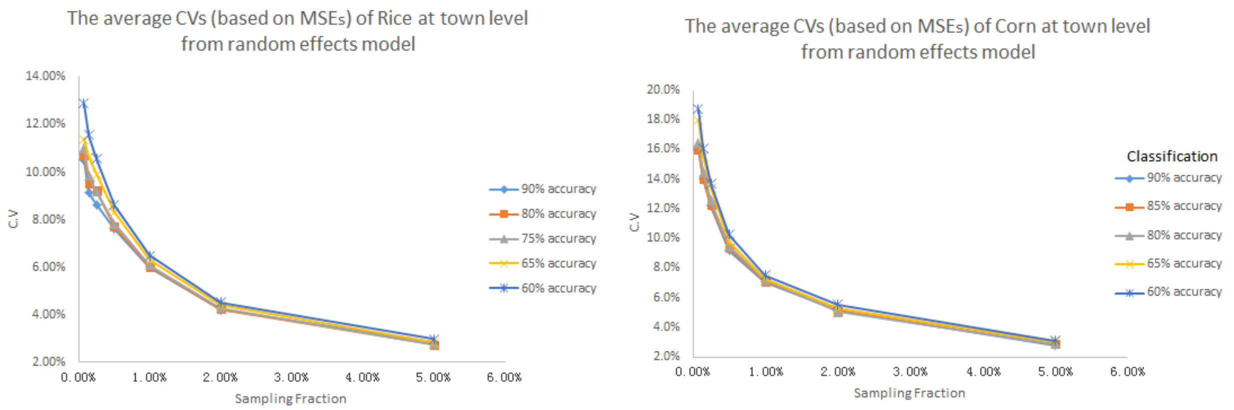

Figure 8: *The average CVs at town level under random effects model with various accuracy*

5.4 Conclusion

In this study, it is illustrated that a basic level small area model in the form of one-response multiple regression with random effects and fixed effects are both feasible to produce the estimates at town and township level. Compared with the fixed effects model, the random effects model gains more efficiency in terms of its C.Vs for town and townships level. Taking the advantages of small area model which combines the sample data with auxiliary information of population, this method could produce the estimates for each small area even if there is rare or none sample within the domain. Given a certain sampling fraction, a more accurate classification for crops which is taking as auxiliary information for population will benefit the estimate precision for each town and township from small area model.

Small area model provide a solution to produce estimates not only for the towns and townships but also for the entire county simultaneously. For fixed effects model, the aggregate of estimates of every town and township is exactly equal to the estimates for the county under the assumption of linear regression. While for random effects model, this additive property is approximately hold on. Therefore, it is provide a solution for multi-level estimation by applying small area model.

In practice, for the crop surveys which samples are selected from province as target population like the business mode of China's agricultural statistics, there would be a significant precision gains of applying small area model which could produce the estimates for every county as well as province itself in a coherent way.

REFERENCES

[1] Blaes, X. and Vanhalle, L. et al. Efficiency of crop identification based on optical and SAR image time series[J]. *Remote sensing of environment,* 2005,96 (3-4): 352-365.

[2]Chauhan, H. J. and Arora, M. K. et al. Estimating land cover class area from remote sensing classification[J]. *Journal of Applied Remote Sensing*, 2008, 2: 1-12.

[3] Battese, G.E. Harter, R.M. and Fuller, W.A. An error-components model for prediction of county crop areas using survey and satellite data[J]. *Journal of the American Statistical Association,* 1988(83): 28-36.

[4]Carfagna, E., & Gallego, F. J. Using remote sensing for agricultural statistics[J]. *International Statistical Review*, 2005(73): 389–404.

[5] Fay, R.E. and Herriot, R.A. Estimation of income from small places: an application of James-stein procedures to census data[J]. *Journal of the American Statistical Association*, 1979(74): 269-277.

[6] Gallego F.J. Remote sensing and land cover area estimation[J]. *International Journal of Remote Sensing*, 2004, 25(15): 3019-3047.

[7] Gallego, J. and C. Bamps. Using CORINE land cover and the point survey LUCAS for area estimation[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2008, 10 (4): 467-475.

[8]Smith, J.H. Stehman, S.V. Wickham, J.D. et al. Effects of landscape characteristics on land-cover class accuracy[J]. *Remote Sensing of Environment*, 2003, 84(3): 342-349.

[9] Pan Y.Z., Li L., Zhang J.S. et. al. Winter wheat area estimation from MODIS-EVI time series data using the Crop Proportion Phenology Index [J]. *Remote Sensing of Environment*, 2012(119): 232–242.

[10] Rao, J.N.K. Small Area Estimation[M], New York:Wiley, 2003.

[11] Opsomer J.D., Botts C., and Kim J. Y. Small area estimation in a watershed erosion assessment survey[J]. *Journal of Agricultural, Biological, and Environmental Statistics*, 2003, 8(2): 139-152.

[12] Militino A.F., Ugarte M.D., and Goicoa T. Combining sampling and model weights in agriculture small area estimation [J]. *Environmetrics*, 2007(18): 87-99.