Jour. Ind. Soc. Ag. Statistics
56(2), 2003 : 166-176

Use of Satellite Data and Farmers Eye Estimate for Crop Yield Modelling

Randhir Singh
Indian Agricultural Statistics Research Institute, New Delhi–110012
(Received : April, 2001)

SUMMARY

Forecasting of crop production is one of the most important aspects of agricultural statistics system. Yield forecasts in the country at present are based on the traditional approach of crop condition and the area under the crop. The main factors affecting crop yield are agricultural inputs and weather. Use of these factors forms one class of models of forecasting crop yields. Another approach uses crop vigor through remotely sensed satellite data. In this regard several studies have been undertaken to establish relationship between spectral parameters through vegetation indices and the crop yield. Shah [12] used satellite data and geographical information system (GIS) technology for developing crop yield models. Verma *et al.* [22] presented a study on evaluation of crop cut method and farmers reports for estimating crop production. The results of the study showed that farmers eye estimates are remarkably close to actual production figures. In the present study an attempt is made to use farmers eye estimate of crop yield more objectively as an auxiliary variable along with the spectral vegetation indices obtained from satellite data to develop an improved crop yield model. The crop yield data is obtained from the general crop estimation surveys (GCES) based on crop cutting experiments. The farmer's eye estimates of the crop yield of the fields which have been selected for GCES and the satellite data of the corresponding area are obtained at the time of maximum vegetative growth stage of crop (about 4-6 weeks before harvest) when they have the highest correlation with the crop yield. The findings of this study suggest that a reliable and timely crop yield forecast may be obtained by using the NDVI from remote sensing satellite spectral data along with the farmer's eye estimate of yield as the two explanatory variables.

1. Introduction

Agriculture is the backbone of Indian economy, contributing about 40 percent towards the Gross National Product (GNP) and providing livelihood to about 70 percent of the population. So, for a primarily agriculture based country like India, reliable, accurate and timely information on types of crops grown and their acreages, crop yield and crop growth conditions are vital components for planning efficient management of natural resources. This involves formulating and implementing appropriate prices of agricultural commodities, strengthening country's food security and distribution system and import/export policies of these commodities from time to time.

India is one of the few countries, which have a well-established system of collection of agricultural statistics and detailed statistics of land utilization are continuously available since 1884. The crop production of principal agricultural crops in the country is usually estimated as a product of area under the crop and the average yield per unit area of the crop. The estimates of the crop acreage at a district level are mostly obtained through complete enumeration whereas the average yield is obtained through general crop estimation surveys (GCES) on the basis of crop cutting experiments conducted on a number of randomly selected fields in a sample of villages in the district. Sukhatme and Panse [20] describe these surveys in detail.

The crop forecasts/advanced estimates of crops are presently developed by Ministry of Agriculture. The advance estimates of kharif crops are first prepared in July/August tentatively when behavior of South-West monsoon is clear and reports of coverage of area under crops from the states are available. The advance estimates are reviewed during December/January when estimates of area under kharif crop become available under the Timely Reporting Scheme (TRS) and results of the crop cutting experiments portion from the NSSO (normally 10%) become available. The advance estimates of rabi season are also prepared at this stage. The advance estimates are again reviewed in the month of April based on information obtained from the states giving the final forecast for kharif.

With the advent of Remote Sensing Technology during 1970s, its great potential in the field of agriculture have opened new vistas of improving the agricultural statistics system all over the world. Space borne satellite data has been widely used in the field of agriculture for estimation of area under different major crops like wheat, paddy, groundnut and sugarcane. Sahai and Navalgund [13] describe the IRS utilization programme in Agriculture. Studies have also been made to examine the relationship of crop growth parameters like leaf area index (LAI) representing crop vigour and the spectral data in the form of several vegetation indices developed from the spectral data of various bands (Tucker *et al.* [21]). Remote sensing satellite data has also been used for improving the crop yield estimation through crop cutting experiments and also for developing crop yield models.

During 1990-92 a study was conducted at the Indian Agricultural Statistics Research Institute (IASRI) to examine the usefulness of satellite spectral data for stratification of crop area based on vegetation indices for improving crop yield estimation from the GCES. This study showed that the efficiency of crop yield estimation could be increased considerably by using the satellite data along with the survey data. The results of this study are given in Singh *et al.* [14].

Another similar study was undertaken during 1996-98 for improved estimation of wheat crop yield in district Rohtak for 1995-1996 using the IRS 1B - LISS II satellite data and the crop yield data for rabi 1996. The results from this study presented in Singh *et al.* [17] also showed that satellite data in the form of vegetation indices greatly improves the efficiency of crop yield estimator.

2. *Use of Satellite Data in Crop Yield Modelling*

The main factors affecting crop yield are inputs and weather. Use of these factors forms one class of models for forecasting crop yields. The other approach uses plant vigour measured through plant characters. It is assumed that plant characters are integrated effects of all the factors affecting yield. Yet another approach is measurement of crop vigor through remotely sensed data. These approaches are being tried by various organizations. Box and Jenkins [2] used time series models for forecasting yield where the variation in yield during different years is explained using historical data through trend analysis and presented the well-known technique of auto regressive integrated moving averages (ARIMA).

The approach using weather parameters is normally based on time series data. The major work in this regard has been attempted at IMD. Their studies involve identification of significant weather parameters in different periods and utilizing these parameters in the regression model along with trend. At IASRI, studies have been carried out at district level using weekly weather parameters. Various composite weather variables were derived as weighted accumulations of weekly weather parameters up to the time of forecast and were used as regressors in the model along with trend. Principal components of weather variables were also tried for developing the model (Agarwal *et al.* [1]; Jain *et al.* [5]). The problem associated with meteorological model is assumption of same weather prevailing in a larger area, as observatories are sparsely located. These models also require long series of data, which are generally not available for most of the locations.

The other approach using plant characters collected at farmers' fields has been attempted through pilot studies at IASRI, New Delhi. Mainly two types of models, between year and within year models have been used. Between year models are based on historic data and involve an assumption that present year is a part of the composite population of the previous years. These models utilize the plant characters at some suitable phenological stage of crop growth either as such or their suitable transformations through multiple regression technique (Jha *et al.* [6]).

In case of crop yield modelling using satellite data, several studies have been undertaken to establish relationship between spectral parameters through vegetation indices and the crop yield. Sridhar *et al.* [19] presented wheat production forecasting for a predominantly un-irrigated region in Madhya

Pradesh. Singh and Ibrahim [16] examined the use of multi date satellite spectral data for crop yield modeling using Markov Chain Model. Saha [12] used satellite data and GIS technology for developing several crop yield models.

A study on "Evaluation of crop cut method and farmers reports for estimating crop production" (Verma *et al.* [22]) was undertaken at Longacre Agricultural Development Centre, UK. This study was carried out in 5 countries in Africa during 1987 with the objective of comparing crop estimates based on crop cut methods with estimates obtained by asking farmers directly to state their production. The results of the study showed that farmers eye estimates are remarkably close to actual production figures in all the countries and they also show considerably small variance compared to the estimates based on crop cutting experiments. After the publication of this report considerable interest is again focused on using farmers estimates which are much cheaper to obtain and easier to conduct. However, farmers estimate being subjective in nature are not considered very reliable.

In the present study, therefore an effort is made to use the farmers eye estimate more objectively as a auxiliary variable along with the spectral indices to improve the efficiency of crop yield forecasting models.

2.1 Study Area and Extent of Data Used in the Study

The study was conducted for district Rohtak, Haryana which is one of the major wheat growing areas having an acreage of more than 66 percent under wheat crop during rabi season. The data used in the study are

(a) *General crop yield estimation survey data :* The yield data for the rabi season for the years 1995-96 and 1997-98 from GCES for wheat crop for district Rohtak, Haryana.

(b) *Satellite data :* The satellite data for 1995-96 from IRS-1B, LISS-II of path 30 and Row 47 of 17th Feb., 1996. The total area of Rohtak district is covered in one sub scene B_2 of 30-47. For 1997-98 IRS-1D LISS-III data of path 95 and row 51 for 4th Feb., 1998.

(c) *Farmers yield appraisal data :* The data of the farmers eye estimates has been collected for the years 1995-96 and 1997-98 for wheat crop yield from the farmers for the fields which have been selected for crop cutting experiments in the GCES. The data has been collected at the time of maximum vegetative growth stage where satellite data has highest correlation with yield.

Further a Global Positioning System (GPS) was also used to identify the locations of the plots selected for crop cutting experiment in terms of their latitudes/longitudes and also the locations of ground control points (GCP's) which were later used to rectify the raw digital spectral data.

(d) *Topographic maps :* A topographic map is the best tool to supply ground truth information for visual interpretation and identification of various features on satellite imageries for identifying locations of villages along with related features like continuous roads, canals, railway tracks etc. on FCC's. Survey of India topographical maps of Rohtak district on 1:50,000 scale were used to identify the location of villages selected for the GCES.

(e) *Stratification of imageries using density slicing technique :* Spectral response characteristics of healthy vegetation, can easily be characterized in different parts of the electromagnetic spectrum. To further enhance the discrimination between different spectral vegetation classes, computation of different vegetation indices using infrared and red band data in the electromagnetic spectrum, for describing the crop growth conditions, are commonly used. Two most commonly used vegetation indices are

The Normalized Difference Vegetation Index (NDVI) defined as

$$\text{NDVI} = \frac{\text{IR} - \text{R}}{\text{IR} + \text{R}}, \text{ and}$$

The Ratio Vegetation Index (RVI) defined as

$$\text{RVI} = \frac{\text{IR}}{\text{R}}$$

Where IR and R refer to radiance in infrared (band-4) and red (band-3) bands of the satellite. These two indices have been used in the present study to generate the images for post-stratification of the study area on the basis of vegetation vigor.

3. Spectral Yield Models

Spectral yield models are empirical models, which directly relate the crop yield to the multi-spectral satellite data or derived parameters in the form of spectral vegetation indices (SVI). In this procedure SVI at the time of maximum vegetative stage of the crop is related to final crop yield through regression techniques and pre harvest crop yield is forecasted. In India district level yields of major crops like wheat, paddy, sorghum etc. have been developed under crop acreage and production estimation (CAPE) project undertaken by National Remote Sensing Agency (NRSA), Hyderabad, Deptt. of Space (Dadhwal *et al.* [3], Parihar *et al.* [11]). However these models could explain about 60% variation in yield and hence are not very efficient.

In the present study we consider post-stratification based on vegetation indices to develop regression coefficient of y (yield) on the spectral response parameter (x) which may lead to improvement in efficiency of the regression model.

The concept of density slicing was used to divide the RVI and NDVI imageries into different vegetation classes. The RVI and NDVI grey level values were linearly stretched over the total range (0-255) of grey level values and were divided into 3 classes named as

(i) Non-vegetation class
(ii) Average vegetation class, and
(iii) High vegetation class

Assigning different colors to different class range values, the stratified imageries were generated and area falling under different strata could be obtained which have been used as the strata weights. Table 1 gives the distribution of grey values and the area of different strata.

Table 1. Distribution of grey values and area of different strata based on RVI and NDVI for district Rohtak for rabi 1995-96 and 1997-98

		NDVI			RVI		
Sl.No.	Stratum/Veg classes	Range of grey values	No. of villages selected	Area Sq.Km.	Range of grey values	No. of villages selected	Area Sq.Km.
			1995–96				
1.	Non vegetation	0–167	–	934.986	0–67	–	851.412
2.	Av.-vegetation	168–217	39	1263.129	68–187	42	1360.522
3.	High-vegetation	218–255	36	1240.109	188–255	33	1226.468
			1997–98				
1.	Non Vegetation	0–167	–	1525.33	0–78	–	1440.07
2.	Av.-vegetation	168–214	30	861.62	79–186	31	908.64
3.	High-vegetation	215–255	42	1121.85	187–255	41	1160.14

Now we define the regression model as

$$y_{hi} = \beta_o + \beta_h \, x_{hi} + e_{hi}$$

where y_{hi} and x_{hi} denote the yield and the spectral response of the i^{th} unit in the h^{th} stratum, h =1, 2, ..., L; i =1, 2,.., m_h . m_h is the number of sampling units in the h-th stratum.

Further let

$$(e_{hi}) = 0$$

$$V(e_{hi}) = \sigma_h^2 \text{, and}$$

$$Cov(e_{hi} \, e_{hj}) = 0 \text{ for } i \neq j$$

Here β_h may be estimated separately for each stratum. Under the given assumption, the Ordinary Least Square (OLS) estimator of the regression coefficient β_h for the h^{th} stratum may be given by $\hat{\beta}_h$ as

$$\hat{\beta}_h = \frac{\sum_{i=1}^{m_h}(x_{hi} - \bar{x}_h)(y_{hi} - \bar{y}_h)}{\sum_{i=1}^{m_h}(x_{hi} - \bar{x}_h)^2}$$

with $$\bar{x}_h = \sum_{i=1}^{m_h} x_{hi} / m_h, \ \bar{y}_h = \sum_{i=1}^{m_h} y_{hi}/m_h$$

The standard error of $\hat{\beta}_h$ is given by

$$SE\ (\hat{\beta}_h) = \frac{\sigma_h}{\sqrt{\sum_{i=1}^{m_h}(x_{hi} - \bar{x}_h)^2}}$$

An unbiased estimator of σ_h^2 is given by

$$s_h^2 = \frac{\sum_{i=1}^{m_h}(y_{hi} - \hat{\beta}_0 - \hat{\beta}_h\, x_{hi})^2}{(m_h - 2)}$$

The fitted regression equation can be used to predict the value of yield corresponding to a chosen value x'_{hi} of the spectral response as given by

$$\hat{y}_{hi} = \hat{\beta}_0 + \hat{\beta}_h\, x'_{hi}$$

The standard error of the predicted value is given by

$$SE(\hat{y}_{hi}) = \left[V(\bar{y}_h) + (x'_{hi} - \bar{x}_h)^2\, V(\hat{\beta}_h)\right]^{1/2}$$

$$= \sqrt{\frac{\sigma_h^2}{m_h} + \frac{(x'_{hi} - \bar{x}_h)^2\, \sigma_h^2}{\sum_{i=1}^{m_h}(x'_{hi} - \bar{x}_h)^2}}$$

The regression coefficient estimator for the population may be defined as

$$\hat{\beta} = \sum_{h=1}^{L} w_h\, \hat{\beta}_h$$

where w_h is the weight of the h^{th} stratum.

4. *Integrated Yield Model Using Spectral Data and Farmers Eye Estimate of Crop Yield*

Most of the crop yield models developed so far could not be adopted in practice either because of delay in the availability of data on different variables to be used in the model or the high cost in collecting the data and in analyzing the results.

For any operational yield model to be successful for adoption, it is necessary that data should be available much before the harvest of the crop and it should be cost effective. Spectral data in the form of vegetation indices have been proved to be very useful for explaining variability of the crop yield which can be easily available for use in yield forecasting models. Further Verma *et al.* [22] showed that farmers eye estimates are remarkably close to actual production figures. But, eye estimates being subjective and amenable to several non-sampling errors, it is desirable that these estimates are not used directly for estimation of crop yield. However, this information can be used as auxiliary variable along with the spectral vegetation indices to improve the efficiency of the crop yield models. An earlier such attempt on using eye appraisal of crop yield of a large number of sample fields as auxiliary information had been made by Panse *et al.* [10].

In the present study, therefore suitable models using spectral vegetation indices in the form of NDVI, RVI and the farmers eye estimate of the crop yield as explanatory variables in the regression model have been developed for improved crop yield forecasting models. Both these variables can be easily obtained at the time of maximum vegetative stage of crop and have been proved very effective for developing suitable yield forecasting models.

5. *Results and Discussion*

The usual linear regression based models have been developed with the crop yield (y) as the dependent variable and three independent variables, namely RVI (x_1), NDVI (x_2) and the farmer's eye estimate of crop yield of the corresponding plot (x_3). The models have been developed using the data for the rabi 1995-96. These models have been used to forecast the crop yield for rabi 1997-98 using the independent variables for 1997-98. The results are given in Table 2. From this table, it is seen that R^2 value is 0.45 and 0.54 respectively when only RVI and NDVI alone are used in the model and it increases to 0.59 when both these variables are used. However the R^2 value is 0.86 when only farmers eye estimate is used as the explanatory variable and the R^2 value increases to around 0.90 when it is used along with RVI or along with NDVI or along with both RVI and NDVI together. The partial correlation coefficients of the yield (y) with RVI (x_1) and yield with NDVI (x_2) after adjusting for the effect of farmer's eye estimate (x_3) are respectively 0.4690 and 0.4924. The deviation of the predicted yield from the actual yield is very low. In almost all the cases it is less than 2%. The standard error of the predicted value is also small in all the cases and as expected it is smallest when all the three variables are used together but it is not much different in the case when only farmers eye estimate and NDVI are used.

The results suggest that a reliable and timely forecast may be obtained using NDVI from the satellite spectral data along with the farmers eye estimate as the two explanatory variables. Both of these variables can be obtained at the time of maximum vigor of the crop and objective reliable forecast may be made about 6-8 weeks before actual harvest of crop.

Table 2. Wheat crop yield forecasting model using RVI (x_1), NDVI (x_2) and the farmers eye estimate (x_3) as independent variables for district Rohtak for forecasting crop yield for rabi 1997-98 (using the model based on data for rabi 1995-96)

	R^2	a	b	Predicted value (Q/hac.) $\hat{y}$	%S.E.	Percentage Deviation
$\hat{y} = a + b_1x_1$	0.451596 (3.102028)	3.3445 (0.5724)	4.251948 (1.998869)	35.86 (4.8568)	13.5434	0.1653
$\hat{y} = a + b_1x_2$	0.543511 (2.830157)	-6.18267 (2.721178)	44.87417 (5.024239)	35.86 (4.7004)	13.1071	0.1652
$\hat{y} = a + b_1x_3$	0.867496 (1.124314)	2.036448 (1.52888)	0.216212 (0.021125)	34.85 (2.1200)	6.0828	3.0619
$\hat{y} = a + b_1x_1 + b_2x_2$	0.59259 (2.03613)	-3.798801 (8.212242)	b_1 = - 2.530032 (5.094414) b_2 = 55.993241 (46.451635)	35.22 (5.2418)	14.8829	1.9872
$\hat{y} = a + b_1x_1 + b_2x_3$	0.90009 (1.00829)	0.785252 (1.479713)	b_1 = 1.146923 (0.510314) b_2 = 0.1797795 (0.024910)	34.86 (1.8352)	5.2647	3.0424
$\hat{y} = a + b_1x_2 + b_2x_3$	0.90345 (0.99122)	-1.049277 (1.865161)	b_1 = 11.219523 (1.679861) b_2 = 0.175972 (0.025054)	34.86 (1.8046)	5.1771	3.0430
$\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3$	0.90406 (1.02274)	-2.144346 (4.132285)	b_1 = -0.770271 (2.572191) b_2 = 18.257945 (23.994514) b_3 = 0.175182 (0.025984)	34.86 (1.7992)	5.1613	3.0433

Actual crop yield for rabi 1997-98 = 35.92 (Q/hac)
(Figures in brackets give the corresponding standard error)

REFERENCES

[1] Aggarwal, Ranjana, Jain, R.C. and Jha, M.P. (1986). Models for studying rice crop - weather relationship. *Mausam*, **37(1)**, 67-70.

[2] Box, G.E.P. and Jenkins, G.M. (1976). Time series analysis: Forecasting and control. Holden-Day, San Franscisco, 575.

[3] Dadhwal, V.K., and Parihar, J.S. (1985). Estimation of 1983-84 wheat acreage of Karnal district (Haryana) using Landsat MSS digital data. Technical note, IRS-UP/SAC/CPF/TN/09, Space Application Centre, Ahmedabad.

[4] Hubback, I.A. (1927). Sampling for rice yield in Bihar and Orissa. Pusa Agric. Res. Inst., Bull. No. 166.

[5] Jain, R.C., Aggarwal, Ranjana and Jha, M.P. (1980). Effects of climatic variables on rice yields and its forecast. *Mausam*, **31(4)**, 591-596.

[6] Jha, M.P., Jain, R.C. and Singh, D. (1981). Pre-harvest forecasting of sugarcane yield. *Indian J. Agric. Sci.*, **51(11)**, 757-761.

[7] Mahalanobis, P.C. (1945). A report on the Bihar crop survey, 1943-44. *Sankhya*, **7**, 29-118.

[8] Mahalanobis, P.C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *J. Roy. Statist. Soc.*, **190**, 325-370.

[9] Manual of nationwide land use/land cover mapping using digital techniques part-II (1999). RRSSC Nagpur, Deptt. of space.

[10] Panse, V.G., Rajgopalan, M. and Pillai, S.S. (1966). Estimation of crop yield for small domain. *Biometrics*, 374-384.

[11] Parihar, J.S., Panigrahi, S., Dadhwal, V.K., Bhatt, H.P., Dass, N.K., Ghosh, B.K. and Behra, D. (1987). Rice acreage estimation in South Orissa using stratified random sampling approach. Scientific Note, IRS-UP/SAC/CPR/SN/09/87, 24.

[12] Saha, S.K. (1999). Crop yield modeling using satellite remote sensing and GIS-current status and future prospects. *Proc. of Geoinformatics - Beyond 2000*, an International Conference on Geoinformatics for Natural Resources Assessment, Monitoring and Management, held at IIRS Dehradun during 9-11 March, 1999.

[13.] Sahai, Baldev and Navalgund, R R. (1988). Indian Remote Sensing Satellite utilization programe for Agriculture, Remote Sensing in Agriculture (ed. Baldev sahai). *Ind. Soc. of Rem. Sen.*, Ahmedabad, Chapter 41.

[14] Singh, R., Goyal, R.C., Saha, S.K. and Chhikara, R.S. (1992). Use of Satellite Spectral data in Crop Yield Estimation Surveys. *Int. J. Remote Sensing*, **13(14)**, 2583-2592.

[15] Singh, Randhir and Goyal, R.C. (1993). Use of remote sensing technology in crop yield estimation surveys. Project Report, IASRI, New Delhi.

[16] Singh, Randhir and Ibrahim, A.E.I. (1996). Use of spectral data in Markov Chain model for crop yield forecasting. *Jr. Indian Socy. of Remote Sensing*, **24(3)**, 145-152.

[17] Singh, Randhir, Goyal, R.C., Pandey, L.M., and Saha, S.K. (1999). Use of remote sensing technology in crop yield estimation surveys – II, Project Report IASRI, New Delhi.

[18] Singh, R., Semwal, Dinesh P., Rai Anil and Chhikara, Raj S. (2002). Small area estimation of crop yield using remote sensing satellite data. *Int. Jr. of Remote Sensing*, **23(1)**, 49-56.

[19] Sridhar,V.N., Dadhwal, V.K., Choudhary, K.N., Sharma, R., Bairagi, G.D. and Sharma, A.K., (1994). Wheat production forecasting for a predominantly unirrgated region in Madhya Pradesh (India). *Int. J. Rem. Sens.*, **15(6)**, 1307-1316.

[20] Sukhatme, P.V. and Panse, V.G. (1951). Crop surveys in India - II. *Jour. Ind. Soc. Agril. Stat.*, **2,** 95-168.

[21] Tucker, C.J., Holban, B.N., Elgin, J.H., and Mc Murtrey, J.E. (1980). Relationship of spectral data to grain yield variations. Photogramm. *Engg. Remote Sensing*, **46,** 657-662.

[22] Verma, Vijay, Marchant, Tun and Scott., C. (1988). Evaluation of crop cut methods and farmers reports for estimating crop production. Longacre Agricultural Development Centre Ltd., London, U.K.

[23] Forecasting Agricultural Output using Space, Agrometeorology and Land based observation (FASAL) (1999). Scientific Note, Space Application Centre, Ahmedabad.