Statistics and Applications
Volume 9, Nos. 1&2, 2011 (New Series), pp. 1-13

Remote sensing based wheat acreage and spectral-trend-agrometeorological Yield Forecasting : Factor Analysis Approach

Urmil Verma[1], D. S. Dabas[1], R. S. Hooda[2], M. H. Kalubarme[3], Manoj Yadav[2], M. S. Grewal[1], M. P. Sharma[2] and R. Prawasi[2]
[1]*Chaudhary Charan Singh Haryana Agricultural University, Hisar -125004, India*
[2]*Haryana Space Application Centre, Hisar-125004, India*
[3]*Space Application Centre (ISRO), Ahmedabad, India.*

Abstract

Forecasting of crop production is one of the most important aspects of agricultural statistics system. Crop production forecasting comprises crop identification, area estimation and predicting the yield of the crop. Crop acreages are estimated using current season's Indian Remote Sensing Satellite data. Stratified Random Sampling approach and supervised classification of Satellite digital data has been adopted for district-level acreage estimation. In this procedure, using ground based observations collected synchronous to satellite passes, the various crops along with other vegetation classes are identified on the satellite data and spectral signatures are generated for supervised classification. The sample segments are classified using these spectral signatures and crop acreages in the district are estimated using standard statistical aggregation procedures. The Quality Assessment (QA) of district-level acreage estimates is evaluated based on the ground truth data collected in the sample segments. To estimate wheat yields in the study districts, zonal spectral–trend-agrometeorological (agromet) model has been generated using the statistical method of factor analysis. The study has been conducted for zone-II comprising of Karnal, Kaithal, Jind, Panipat, Sonipat and Rohtak districts of Haryana state. Remote sensing based wheat acreage and model predicted yields have been compared with Department of Agriculture (DOA) estimates by computing percent relative deviation. Zonal model developed on the basis of time series data from 1978-79 to 2000-01 has been used to predict the wheat yields for the period 2001-02 to 2007-08. The overall results indicate that the integration of remote sensing data with trend based yield and weather variables provides an immense scope to improve the efficiency and reliability of wheat yield forecasts. Zonal yield models provided considerable improvement in the district-level yield forecasts by showing good agreement with the real time data.

Key words: Remote sensing; Spectral index; Trend-agromet; Eigen vectors; Factor scores; Relative deviation.

1 Introduction

India has one of the best systems in the world to collect, collate and compile data on crop production. These are based on official feedback of area and yield received from the states, which in turn gets from the districts and so on. Area is based on complete enumeration by revenue agencies and yield by crop cutting experiments. One of the limitations of conventional methods is timeliness and quality of the statistics. The final estimates are worked out some 4-5 months after the end of the season. This makes these statistics unusable for planning and management purposes. Hence, there is a considerable scope of improvement in the conventional system.

Climatic parameters are the variables which define the agricultural productivity of any region and therefore, the direct effects of the climatic factors on crop growth and development have always been a subject of detailed investigation. Weather variables affect the crop differently during different stages of growth period. On the other hand, technological advancement also affects the yield significantly. It includes the impact of increased fertilizer applications, better irrigation facilities, improved management practices, pest control and use of improved quality of seed etc. Using the statistical method of factor analysis, Chmielewski (1992) attempted to answer the question whether the observed recent climate changes have had an impact on the crop yields of winter rye in Germany. A number of studies at national/international level have been carried out to develop suitable yield forecast models for various crops using different statistical techniques; to cite just a few – Boken (2000) , Agarwal et al. (2001), Rai (2003), Dadhwal et al. (2005), Esfandiary et al. (2009) etc. have incorporated a series of weather predictors.

With the advent of Remote Sensing (RS) 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 have been widely used in the field of agriculture for estimation of area under different major crops. Sahai and Navalgund (1988) described the IRS utilization programme in agriculture. Sridhar et al. (1994) presented wheat production forecasting for pre-dominantly un-irrigated region in Madhya Pradesh. Singh and Ibrahim (1996) examined the use of multi date satellite spectral data for crop yield modeling using Markov Chain Model. Saha (1999) used satellite data and GIS technology for developing several crop yield models. Wheat yield modeling using remote sensing and weather variables was attempted on agro-climatic zone basis in Haryana (Verma et al., 2003).

Wheat is one of the most important cereal crops in India as it forms a major constituent of the staple diet of a large part of the population. Wheat occupies today the foremost position followed by rice, not only in terms of acreage and production but also in the versatility in adopting different soils and climatic conditions. India occupies second place in terms of acreage and production among wheat growing countries of the world (Source:www.mapsofindia.com/indiaagriculture/). Haryana occupies third place for wheat production (Source: www.agricoop.nic.in/statistics) among the various states in India. Haryana has topped the states (India) in 2008-09 with highest wheat productivity of 46.14 quintals per hectare. Keeping in view the importance of the crop, an attempt has been invariably made using spectral indices, factor scores based on weather variables and trend based yield as regressors to obtain the zonal yield model. However, acreage estimation has been done by digital image processing of the remotely sensed data.

2 Study area and data used

The Haryana state comprising of 21 districts is situated between 74° 25'E to 77° 38' E longitude and 27°40'N to 30°55' N latitude . The total geographical area of the state is 44212 sq. km. Indian Remote Sensing (IRS) satellite digital data of LISS-I (spatial resolution 72.5 m), LISS-II (spatial resolution 36.25 m) and LISS-III (spatial resolution 23.5 m) sensors were used for the computation of spectral vegetation indices at Haryana State Remote Sensing Applications Centre, CCS HAU Hisar. DOA yield estimates for the past 23 years (1978-79 to 2000-01), published by Bureau of Economics and Statistics have been used for computing trend based yield. The weather data for the same period were collected from IMD, Delhi and different meteorological observatories in Haryana. Various districts in the state were grouped into four zones based on their physiography/soils and agroclimatic conditions i.e. Zone I (Ambala, Panchkula, Yamuna Nagar, Kurukshetra), Zone II (Karnal, Kaithal, Jind, Panipat, Sonipat, Rohtak), Zone III (Mahendergarh, Rewari, Jhajjar, Gurgoan, Faridabad, Mewat) and Zone IV (Sirsa, Fatehabad, Hisar, Bhiwani).

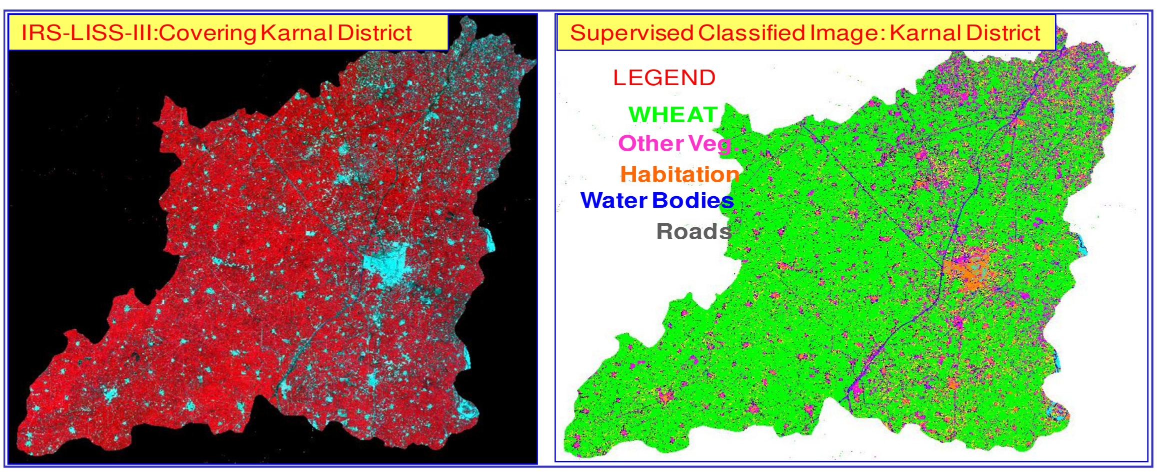
3 Acreage Estimation

Crop production forecasting comprises crop identification, area estimation and predicting the yield of the crop. Crop identification and discrimination is based on the fact that each crop has a unique spectral signature. Digital image processing techniques are used for crop acreage estimation. The procedure adopted for district-level acreage estimation is as follows:

3.1 Procedure Adopted for Acreage Estimation

3.1.1 Stratified Random Sampling

Two-step stratification of the study area was attempted which involves overlaying a sampling grid of 7.5km X 7.5km or 5km X 5km depending on the size of the districts. In each district, five homogeneous strata were delineated on the basis of vegetation density observed on previous years FCC print. These strata were identified as A: > 75% aa, B: 50-75% aa, C: 25-50% aa, D: 5-25% aa and NA: < 5% aa (NA = non agricultural area, aa= agricultural area). The density based stratification was used for acreage estimation. Since the vegetation vigour is an indicator of the growing conditions of the crop, the total variations (bright to dull red) on the False Color Composite (FCC) of the study area were categorized into good, medium and poor strata. A sampling fraction of 10 per cent or 15 per cent was adopted for selection of sample segments in each stratum. The selected sample segment locations were transferred on the Survey of India (SOI) topographical maps. The Ground truth data is collected in the selected sample segments and crop type, crop stage and vigour, soil type, irrigation etc. are recorded. These strata are taken as parameters for generation of spectral vegetation indices. Spectral indices such as ratio of Near Infra Red and Red (NIR/R) or Normalized Difference Vegetation Index = (NIR-R)/(NIR+R) were obtained for further use in the development of crop yield models. IRS satellite digital data were used in classification for district-level wheat acreage estimation. The classified data covering different districts were used for computation of spectral/vegetation indices. One of the examples of IRS LISS-III digital data used for the analysis purpose in Karnal district along with its supervised classified image is shown below.



IRS LISS-III and classified image covering Karnal district

3.1.2 Digital Data Analysis

The digital data from IRS satellite covering Haryana State, of the desired crop stage is acquired for digital data analysis. The digital analysis procedure involves:

i) Extraction of Sample Segments: Using the ground control points (GCP), and map-to-image transformation of the exact location of the sample segments on the satellite data is identified and segments of 7.5km X 7.5km or 5km X 5km in size are extracted from the satellite data for supervised classification.

ii) Spectral Signature Generation: Locations of wheat and other crops identified during ground truth data collection in the field are identified on the satellite data. The ground truth data collected in selected sample segments is used for accurate identification of the crop type and other vegetation types on the satellite data and spectral signatures are generated for classification. The spectral signatures consist of mean, variance and co-variance matrix which is unique for each crop and other vegetation types.

iii) Supervised Classification and Acreage Estimation: The spectral signatures generated after careful identification of various crops and other vegetations types are used for supervised classification. The selected sample segments are classified using the supervised classification approach. The accuracy of this classification is evaluated by generating the confusion matrix which indicates correctly classified pixels, omission and commission errors etc. If the accuracy is not acceptable, the entire procedure of crop identification, spectral signature generation and supervised classification are repeated till accuracy is within desired 90/90 accuracy goal. The results from each classified segment are aggregated using standard statistical aggregation procedures of stratified random sampling and thus, the district-level wheat acreages are estimated.

4 Crop yield modelling

Agromet models cannot account for variations in yield due to disease, pest infestation, drought etc., while the spectral data from a crop is an integrated manifestation of the crop growth condition and the environment. Remote sensing data in conjunction with meteorological data has been proved as an efficient and reliable way towards crop yield modeling. The integration of weather data with spectral data offers an approach, which combines the best features of the both. Spectral data as well as weather parameters can be easily obtained at the time of maximum vegetative stage of the crop for developing yield forecasting models well in advance of the actual harvest of the crop.

4.1 Computation of meteorological parameters

Agro-climatic indices such as Temperature Difference (TD), Growing Degree Days (GDD) and Accumulated Rainfall (ARF) have been used to develop wheat yield models for zone-II (Haryana), by extracting the factor(s) through principal component method. The Growing Degree Days (GDD) and Temperature Difference (TD) are computed as follows; TD = $\Sigma[T_{max} - T_{min}]$ and GDD = $\Sigma[\{(T_{max} + T_{min})/2\} - T_b]$; where T_{max} = maximum Temperature , T_{min} = minimum Temperature, T_b = Base Temperature (5^o). To integrate GDD, TD and weekly accumulated rainfall ARF over different growth phases; total wheat growth period has been divided into seven phenological stages viz., i) Crown Root Initiation Stage (SMW 44-46; SMW stands for standard meteorological week numbers), ii) Tillering Stage (SMW 47-49), iii) Jointing Stage (SMW 50-52), iv) Flowering Stage (SMW 1-3), v) Milking Stage (SMW 4-6), vi) Dough Stage (SMW 7-9) and vii) Maturity stage (SMW 10-14).

4.2 Time-trend analysis

Average wheat yield data of all the district have been used by considering time (year) as an independent variable and has been regressed against yield to get the trend equation the form $T_r = a+br$, where T_r = Trend yield (q/ha), a = Intercept, b = Slope and r = Year.

4.3 Spectral-trend-agromet wheat yield modelling

The regression analysis of remotely sensed and weather data (Verma et al., 2003) as such is faced with low to high multicollinearity among the explanatory variables which inflates the variances of regression coefficients, thus making these estimates less precise. Therefore, a complex statistical method like Factor Analysis being a data reduction technique for investigating the interdependence has been used in this study. This method attempts to simplify complex and diverse relationships existing among a set of observed variables, by revealing common dimensions or factors that link seemingly unrelated variables; consequently, it provides insight into the underlying structure of the data. Principal component method used for extraction of factors offered considerable improvement to overcome the problem of multicollinearity. The procedure consists of finding the eigen roots and eigen vectors of the correlation matrix of explanatory variables. First seven eigen values of correlation matrix of explanatory variables suggested seven factor solution and the remaining fourteen components accounted for a smaller amount i.e. less than 10 per cent of total variation, hence these components were not considered to be of much practical significance.

The factor score coefficient matrix (Table-1) shows the values used to compute factor scores for each case. These scores were computed by multiplying variable values with the corresponding factor score coefficients. Spectral index along with trend based yield and weather based seven factor scores were used as regressors to develop the spectral-trend-agromet yield models. The stepwise regression analysis (Draper and Smith, 1981) was performed in which variables are included or excluded one at a time with decisions at any particular step conditioned by the previous step. The zonal wheat yield relationship finalized is expressed as below:

$$Y_{est.} = 841.98 - 72.69fs_1 + 189.81fs_2 + 82.29fs_6 + 0.76T_r + 38.22V_i\ ;$$

$$R^2 = 0.81\ ,\quad Adj.R^2 = 0.80,\quad n = 88\ \ \&\ \ SE = 192.73,$$

where $Y_{est.}$ = Model predicted yield
fs_i = i^{th} factor score , i= 1,2,3…,7
T_r = Trend based yield
V_i = Spectral vegetation index

Trend yield (T_r) has been observed as an important variable appearing in the model, which is an indication of technological advancement, improvement in fertilizer/ insecticide / pesticide/ weedicide use and increased use of high yielding varieties over time.

Table 1: Factor score coefficient matrix

Variables / Components	**1**	**2**	**3**	**4**	**5**	**6**	**7**
TD_1	.120	.131	.152	-.043	-.291	.031	.039
GDD_1	.039	.049	.026	.031	.031	.051	.661
ARF_1	-.177	-.010	-.009	-.018	-.063	-.009	-.058
TD_2	.162	-.081	.103	-.041	-.002	-.170	.189
GDD_2	-.021	.047	-.325	-.028	-.024	.124	.006
ARF_2	-.177	-.008	-.008	-.013	-.073	-.016	-.061
TD_3	.146	-.048	.092	-.043	-.085	.080	.068
GDD_3	.115	.062	-.060	-.043	.123	.191	-.313
ARF_3	.042	-.063	.112	-.094	.476	-.156	-.070
TD_4	.074	.037	.368	.068	.060	-.085	.064

GDD_4	.084	.113	.021	.046	.429	.172	.145
ARF_4	-.006	-.058	-.096	-.021	-.067	.511	.030
TD_5	-.015	.065	.020	.283	-.110	.050	-.147
GDD_5	.069	.249	.052	.025	-.012	-.216	-.060
ARF5	-.006	.010	-.083	-.393	-.012	.050	-.075
TD_6	.097	-.241	.141	.023	.048	-.098	-.140
GDD_6	-.023	-.266	-.054	.351	.097	-.151	.317
ARF_6	.000	.286	.052	.054	-.005	-.030	.140
TD_7	.018	-.023	-.161	-.042	-.162	-.343	-.032
GDD_7	.116	-.081	-.051	.203	.139	-.144	.007
ARF_7	-.122	-.074	.127	.021	-.024	.054	.095

Table 2: Comparison of RS based wheat acreage with DOA estimates.

		Acreage(000'ha)		
	Year of Forecast	**RS**	**DOA**	**RD(%)**
Rohtak	2001-02	91.6	92.0	-0.43
	2003-04	91.0	89.0	2.25
	2004-05	86.0	87.4	-1.60
	2005-06	86.7	88.0	-1.48
	2006-07	88.7	95.0	-6.63
	2007-08	92.6	99.0	-6.46
Karnal	2001-02	161.2	162.9	-1.04
	2003-04	169.4	167.2	1.32
	2004-05	169.4	170.7	-0.76
	2005-06	171.6	173.0	-0.81
	2006-07	173.5	167.0	3.89
	2007-08	173.2	169.0	2.49

Kaithal	2001-02	169.7	168.7	0.59
	2003-04	172.7	173.6	-0.52
	2004-05	174.0	174.3	-0.17
	2005-06	176.4	173.0	1.97
	2006-07	172.6	167.0	3.35
	2007-08	172.8	169.0	2.25
Jind	2001-02	201.6	206.8	-2.51
	2003-04	205.1	206.9	-0.87
	2004-05	207.2	206.9	0.14
	2005-06	205.6	205.0	0.29
	2006-07	203.4	213.0	-4.51
	2007-08	210.4	213.0	-1.22
Panipat	2001-02	78.1	81.7	-4.41
	2003-04	84.7	81.3	4.18
	2004-05	81.5	83.8	-2.74
	2005-06	84.2	83.0	1.45
	2006-07	82.3	82.0	0.37
	2007-08	83.7	84.0	-0.36
Sonipat	2001-02	135.8	135.9	-0.07
	2003-04	133.6	132.7	0.68
	2004-05	130.9	137.0	-4.45
	2005-06	137.8	142.0	-2.96
	2006-07	142.9	140.0	2.07
	2007-08	142.6	141.0	1.13

2002-03 - No cloud-free satellite data was available for any of the districts
RD%= [(RS based acreage - DOA acreage) / DOA acreage] *100

Table 3: Comparison of spectral-trend-agromet model based yield forecasts with DOA yield estimates.

	Year of Forecast	DOA Yield (q/ha)	Model predicted yield (q/ha)	RD(%)
Rohtak	2001-02	38.39	40.26	4.87
	2002-03	39.64	38.59	-2.65
	2003-04	38.35	37.74	-1.59
	2004-05	36.80	39.48	7.28
	2006-07	39.58	42.11	6.39
Karnal	2001-02	45.80	44.20	-3.49
	2002-03	43.63	43.57	-0.14
	2003-04	41.38	41.30	-0.19
	2004-05	41.83	43.52	4.04
	2005-06	43.67	44.39	1.65
	2006-07	44.23	46.17	4.39
	2007-08	46.29	44.16	-4.60
Kaithal	2001-02	45.48	43.11	-5.21
	2002-03	43.98	41.00	-6.78
	2003-04	41.52	39.92	-3.85
	2004-05	41.15	42.15	2.43
	2005-06	41.97	42.61	1.52
	2006-07	43.73	44.27	1.23
	2007-08	44.39	42.12	-5.11
Jind	2001-02	42.36	42.71	0.83
	2002-03	42.11	40.93	-2.80
	2003-04	39.56	40.15	1.49
	2004-05	41.48	42.60	2.70
	2005-06	40.18	43.19	7.49

	2006-07	42.12	44.99	6.81
	2007-08	41.93	43.34	3.36
Panipat	2001-02	43.58	43.20	-0.87
	2002-03	43.78	42.71	-2.44
	2003-04	42.34	40.15	-5.17
	2004-05	43.11	42.28	-1.93
	2005-06	43.97	42.92	-2.39
	2006-07	44.77	44.50	-0.60
	2007-08	43.75	42.55	-2.74
Sonipat	2001-02	43.70	41.95	-4.00
	2002-03	43.42	40.17	-7.49
	2003-04	37.86	39.24	3.65
	2004-05	39.79	41.50	4.30
	2005-06	40.24	42.14	4.72
	2006-07	45.47	43.87	-3.52
	2007-08	42.94	42.20	-1.72

2005-06, 2007-08- Data , not available for Rohtak district

Spectral-trend-agromet model - weather based factor scores + spectral index + trend based yield used as regressors

RD%= [(Model predicted yield - DOA yield estimate) / DOA yield estimate] *100

4.4 Performance evaluation of wheat acreage and yield estimation

A comparison of RS based wheat acreage with DOA acreage estimates in Table-2 presents a good agreement between the real time data and RS based estimates. Spectral-trend-agromet yield model developed on agro-climatic zone basis has been used to predict wheat yields of Rohtak, Karnal, Kaithal, Jind, Panipat and Sonipat districts of the state. Average wheat yield estimates obtained have been presented in Table-3. It is inferred that the district-level yield predictions for the years 2001-02 to 2007-08 have improved significantly. Model based yields are showing good agreement with DOA yield estimates, favouring the use of the zonal model for pre-harvest wheat yield forecasting. The model predicted yields and their percent deviations from DOA yield estimates are within acceptable limits for all the districts which indicates that the developed zonal model can operationally be used to predict the pre-harvest wheat yields in the study area.

References

Agarwal, R., Jain, R.C. and Mehta, S.C. (2001). Yield forecast based on weather variable and agricultural inputs on agro-climatic zone basis, *Ind. J. Agric. Sci.* **71**(**7**), 487-490.

Boken, V.K. (2000). Forecasting spring wheat yield using time series analysis. A case study for the Canadian prairies. *Agro. J.* **92**, 1047-1053.

Chmielewski, Frank-M. (1992). Impact of climate changes on crop yields of winter rye in Halle (Southeastern Germany). Climate Research **2**, 23-33.

Dadhwal, V.K., Sehgal, V.K., Singh, R.P. and Rajak, D.R. (2005). Wheat yield modeling using satellite remote sensing with weather data: Recent Indian experience. *Mausam* **54**, 253-262.

Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis*, 2nd ed. New York: John Wiley.

Esfandiary, F., Aghaie, G. and Mehr, A.D. (2009). Wheat yield prediction through agro meteorological indices for Ardebil district. *World Academy of Science: Engineering and Technology* **49**, 32-35.

Rai, T. (2003). Use of discriminant function in yield forecasting, *IASRI Report*, Library Avenue, IARI, New Delhi.

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.

Sahai, B. 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.

Singh, R. and Ibrahim , A.E.I.(1996). Use of spectral data in Markov Chain model for crop yield forecasting. *J. Ind. Soc. of Rem. Sens.* **24** (**3**), 145 -152.

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 unirrigated region in Madhya Pradesh (India). *Int. J. Rem. Sens*. **15**(**6**), 1307-1316.

Verma, U., Ruhal, D.S., Hooda, R.S., Yadav. M., Khera, A.P., Singh, C.P., Kalubarme, M.H. and Hooda, I.S.(2003). Wheat Yield Modelling Using Remote Sensing and agrometeorological data in Haryana State. *J. Ind. Soc. Agric. Statist*. **56**(**2**), 190-98.

Author for correspondence

Dr.(Mrs.) Urmil Verma
Professor (Statistics)
Department of Mathematics and Statistics
CCS Haryana Agricultural University
Hisar- 125001, Haryana
E-mail: Vermas21@hotmail.com