Satellite Remote Sensing and GIS based Crops Forecasting & Estimation System in Pakistan

Ijaz Ahmad*, Abdul Ghafoor, Muhammad Iftikhar Bhatti ,Ibrar-ul Hassan Akhtar, Muhammad Ibrahim, Obaid-ur-Rehman

Space Applications and Research Complex, Pakistan Space and Upper Atmosphere Research Commission, Near Rawat Toll Plaza, Islamabad Highway, 44000, Islamabad, Pakistan

***ijazbhutta@hotmail.com** , **Ibrar.space@gmail.com**, **bhatti_iftikhar@hotmail.com**

ABSTRACT

Pakistan Space and Upper Atmosphere Research Commission (SUPARCO), the Space Agency of Pakistan started developing crop area estimation procedures and crop yield models, based on the application of satellite remote sensing, GIS technology, agronomy, agro-meteorology, statistics and other allied disciplines. Conventionally, Crops area estimation system traditionally is based on Village Master Sampling (VMS) from revenue department developed in late 1970s by Federal Bureau of Statistics, Pakistan. Satellite based crops monitoring system in Pakistan has been developed to forecast and estimate crops statistics of major crops which include wheat, rice, cotton, sugarcane, maize and potato since 2005. Crops area estimates are based on two approaches which are Satellite data supervised classification and area frame sampling system. Overall, classification accuracy ranged from 85-95%.

Yield modeling is based on FAO approach of yield relationship with predictor variables. Crop yield forecasting and estimation cover another important dimension of crops statistics being mostly of qualitative nature. SPOT Vegetation data is main yield predicting variable in all calibrated models. All major crops models including wheat, cotton, rice, sugarcane and maize were calibrated for yield forecasting during initial to peak growth season and estimated near harvest time. A selection criterion was the R^2 value (Co-efficient of determination) of 0.8 or more. Satellite data based crops area and yield estimation were compiled and compared later with government official statistics. Main advantage of the SUPARCO satellite based crops system is timeliness release of the data.

Keywords: Satellite Remote Sensing, Crops Monitoring, Area, yield modeling, Pakistan

1. Background:

Pakistan is a country of diverse agro-climatic regions. The climate is predominantly arid to semi-arid. The mighty Indus and its tributaries have facilitated the establishment of a network of dams, barrages and a profuse delivery system of water supplies. Despite a large territory, Pakistan's agriculture is predominantly converged in the Indus basin. Agriculture sector is facing certain challenges which require immediate and focused attention both at research and policy level. Sustainable agricultural growth based on paradigm that secure more profitable farming, high productivity of major farming systems, diversification of high value crops and demand based production. In this regard, the present government is taking various initiatives to accelerate agricultural growth and promote investment in agricultural research (Farooq, 2014).

The Government of Pakistan is in the process of upgrading and diversifying its program and capacity for an effective mechanism to ensure crop monitoring and forecasting system. MNFS&R endeavored to improve mobility, human resource development and service structure of Crop Reporting Departments in the country. The Ministry further opted to invest in cross cutting technologies as Remote Sensing and GIS for gathering spatial information on agriculture/ crops sector for timely interventions.

Conventionally, Crops area estimation system traditionally is based on Village Master Sampling (VMS) from revenue department developed in late 1970s by Federal Bureau of Statistics, Pakistan. Ground survey is carried in selected sample village and district wise crops statistics are compiled based on multiplier or raising factor. The crop production estimates are obtained by taking the product of crop acreage and the corresponding crop yield. The yield surveys are fairly extensive with plot yield data collected under a complex sampling design that is based on random sampling design. A plot of specified dimensions within a field is selected for harvesting to determine the crop yield. The sample units are randomly selected. Problems encountered concern subjectivity in responses, respondent differences and non-response. On national scale, the processing of these sample data is an expensive and time-consuming procedure. In general, there is a need for an objective, standardized and possibly cheaper and faster methodology for crop growth monitoring and yield forecasts.

Traditional methods of predicting crop yields throughout the growing season include models that assimilate climate, soils and other environmental data as response functions to describe development, photosynthesis, evapotranspiration and yield for a specific crop (Wiegand and Richardson 1990). Though based on strong physiological and physical concepts, these models are poor predictors when spatial variability in soils, stresses or management practices are present (Wiegand 1984, Wiegand and Richardson 1990). However, remote sensing of crop canopies has been promoted as a potentially valuable tool for agricultural monitoring because of its synoptic coverage and ability to 'see' in many spectral wavelengths (Hinzman et al. 1986, Quarmby et al. 1993). Numerous studies have recognized that plant development; stress and yield capabilities are expressed in the spectral reflectance from crop canopies and could be quantified using spectral vegetation indices (Jackson et al. 1986, Malingreau 1989, Weigand and Richardson 1990). Vegetation indices (VI), such as the Normalized Difference Vegetation Index (NDVI), are typically a sum, difference or ratio of two or more spectral wavelengths. They are highly correlated with photosynthetic activity in non-wilted plant foliage and are good predictors of plant canopy biomass, vigor or stress (Tucker 1979). Vegetation monitoring using the red and near infrared SPOT VGT channels has been one of the most widely used indices. The Normalized Difference Vegetation Index (NDVI) correlates closely with green biomass and the leaf area index. Despite the spatial resolution of 1 km at nadir, there are many scientific publications documenting the usefulness of SPOT VGT data as a means of monitoring vegetation conditions on a near real-time basis (Philipson and Teng, 1988; Bullock, 1992; Quarmby et al., 1993).

There was a need to develop fast track and reliable procedures to make crop forecasts and estimations early in the season or end of season. Pakistan Space and Upper Atmosphere Research Commission (SUPARCO), the Space Agency of Pakistan started developing crop area estimation procedures and crop yield models, based on the application of satellite remote sensing, GIS technology, agronomy, agro-meteorology, statistics and other allied disciplines.

2. Material and Methods

Satellite based crops monitoring system in Pakistan has been developed to forecast and estimate crops statistics of major crops which include wheat, rice, cotton, sugarcane, maize and potato since 2005 (Bussay and Akhtar 2009, Obaid ur Rehman et al., 2010 & 2011). Crops area estimates are based on two approaches which are Satellite data supervised classification and area frame sampling system. Yield modeling is based on FAO approach of yield relationship with predictor variables.

2.1 Area Estimation Approach:

This describes the crops area estimation system developed at SUPARCO.

2.1.1 Development of SRS/GIS based area frame sampling system:

System has been developed based on crops peak stage satellite data of February and September. This was done through defining different stratum based on agriculture fields and cropping intensity. The land use was stratified into ten different homogenous stratum based visual interpretations (Table 1).

Table 1: Different Stratum and Definitions

S.No.	Stratum	Description
1	11	Intense Cropland (75-100 % agriculture area)
2	12	Less intense Cropland (50-75 % agriculture area)
3	21	Cropland Pasture Mixed (25-50 % agriculture area)
4	42	Mostly Pasture (<25 % agriculture area)
5	13	Un-identified seasonal vegetation
6	14	Areas rarely under vegetation

7	31	Rural area around city (Less that 50 houses / Km^2)
8	32	Inter city
9	50	Non farmland (Desert, Forest, Saline, establishments)
10	60	Water bodies (Rivers, Canals)

These strata were apportioned into Primary (about 5000 to 10000 ha each) (PSU) and Secondary (1000-2000 ha each) Sampling Units (SSU). These units were allotted serial number in a serpentine design through using tailor made software. Pakistan was divided into nine zones viz. Punjab 4, Sindh & Khyber Pakhtunkhwa 2 and 1 in Balochistan. Initially, 20 to 30 sampling units called segments, depending upon the cropping intensity, of a size of approximately 30 ha each were selected from all stratum in each zone based on probability proportional to the area. The fractional segments in each stratum were taken a whole unit. These sampling units were doubled in the subsequent years to assure synchronization of crop data with parallel techniques of image classification. The total number of segments in nine region is 379 (Table 2).

Table 2: Province and region wise number of selected ground samples (Segments)

Province	**Region**	**No of segments**
Punjab	Potohar	21
	North East	46
	Central	75
	Southern	78
Sindh	Left bank of Indus	52
	Right Bank of Indus	42
KP	North	20
	South	20
Balochistan		25
Total		379

Based on the area frame sample designing, Raising Factors (RF) was developed to estimate crop area sown in each stratum in each zone / region. These RF values helped to work out crop area sown under various crops, by a statistical design. A critical examination of the data generated was made by a team of experts in the field of Agronomy, Remote Sensing and Statistics to standardize this technique by image classification and historic trend lines. The team suggested valuable improvements in each cropping season and these changes were incorporated in the technique.

2.1.2 Satellite Image Classification Technique

Satellite data image classification is based on satellite data acquisition of specific time, ground truth surveys during cropping season, crops signature collection, lab processing, accuracy assessment and crop area estimation.

2.1.2.1 Acquisition of imagery:

Country wide acquisition of satellite imagery was done for Rabi and Kharif crops twice at the following stages.

- First at four weeks after the completion of sowing.
  (June-July for Kharif crops and December-January for Rabi crops)
- Second at eight weeks after completion of sowing.
  (August for Kharif crops and February-March for Rabi crops)

2.1.2.2 Ground Truthing Surveys (GTS):

Extensive programs were devised to undertake ground truth surveys to collect crops related information during season. Field teams visited the sampling segments through real-time navigation through GPS devices.

2.1.2.3 Satellite image classification:

The data gathered from the field were digitized. The image classification was done by developing spectral signatures of crops by using multi-date imagery. Image classification was carried out by supervised classification using Gaussian maximum likelihood method on different work units and area estimation was carried out using Image processing software.

2.2 Crop Yield Modeling For Forecasting and Estimation

The important procedural steps in crop yield modeling/forecasting and estimation are as follows (Bussay and Akhtar, 2008 & 2009),

2.2.1 Development of database

A spatial database consisting of data for the last 15 years (1998 and onward) for various variables responsible for change in crop yield was developed. These include district wise crop statistics, agro meteorological data for 36 stations covering min/max temperature, rainfall and relative humidity. The sunshine duration data was available for 8 stations. The sunshine duration data deemed to be very useful above all in the crop yield forecast as the radiation is an important limiting factor of crop production after soil moisture availability. Daily maximum and minimum temperatures were applied in the Hargreaves formula to fill the gaps in the calculated global radiation time-series. The minimum and maximum temperature was applied through the Hargreaves formula to complete the days with missing data. Hargreaves formula estimates the global radiation on the basis of daily temperature range using the maximum (Tmax) and minimum (Tmin) temperatures:

$$H = H_0 \cdot k_{RS} \cdot \sqrt{T_{\max} - T_{\min}}$$

Where Hargreaves coefficient k_{RS} which is between 0.16 (inland stations far from the sea) and 0.19 for stations at the sea-side. The (Tmax - Tmin) difference is the daily temperature amplitude.

2.2.2 Harmonization and Integration of the data

The data were harmonized for various spatial (polygons) and time scales (converted from daily to decadal). Spatial interpolation of the point data was done at a grid size of 0.05 degree for the whole country (Javid et al., 2010). The current year’s data were used to integrate with the historic data and forecast crop yields based on statistical modeling.

2.2.3 Crop Phenology and Modeling from SPOT VGT data

The important phenological stages of crop growth include: (a) time of emergence (b) time of peak growth (c) time of ripening /senescence (M. H. Khan et al., 2007). The time of emergence of a crop or more precisely the time of beginning of measurable photosynthesis on a satellite vegetation image seasonal profile is termed as starting decadal. The increment is within range of 0.01-0.05 per decadal depending on total cropped area and growth stage under the pixel of the satellite image. The time of peak growth or end of growing period and beginning of flowering is the period of maximum greenness or maximum photosynthesis and is called peak decadal. Peak decadal has the highest NDVI value of the cycle. The date of senescence or harvest (Cessation of measurable photosynthesis is called Ending Decadal. It occurs at a minimum of 3 decadal after peak decadal: The course of previous NDVI values is decreasing and the following NDVI values have increasing trend or the course is flattened.

2.2.4 Development of calibration matrices and Model Development

The matrices were developed for all variables responsible for change in crop yield (Akhtar, 2011),

2.2.4.1 Principal Component Analysis

This is one of the main components of the model calibration which reduces the dimensional aspects of all independent variables defining the crop yield.

2.2.4.2 Correlation Matrix

This analysis is carried out to find the possible co-linear relationship within PCA derived variables to reduce the biasness in final model. The co-linear variables are identified and only those one with moderate to high independency nature are used in model calibration.

2.2.4.3 Outlier detection

This step is necessary to remove the suspicious observations with the help of statistical test mainly Whisker Box plot or cook distance techniques. This improves the model accuracy and eliminates the bias extremes cases.

2.2.4.4 Multiple Regression Analysis

Multiple regression analysis is carried out between the selected independent variables which are significantly responsible for change in yield. At the end of Model calibration, the model based error in yield/production forecast/estimation is carried out to define the confidence interval of the forecast/estimates (Variance, Average Absolute error, Average error etc). Validation is based on model output at different spatial scales like the production.

2.3 <u>Forecasting and Estimation of Agricultural Statistics</u>

Crop area estimates are made available after the ground survey campaign and image processing of the seasonal acquired data, crop yield modeling and quality assessment.

3. Results and Discussion

Satellite based crops monitoring system in Pakistan has been flourished after 2005 due to its timeliness and reliability of crop statistics. Crops area estimation through area frame sampling system mainly relies on the quality of ground data collected during season in sample segments (Figure 1). This field information on crops sown in each segment was digitized in ArcGIS software. Digitization of the samples was carried out at 1:3000 scales to avoid any field size impact on crops area estimation (Figure 2). The segments based crops information summarized by the stratum and Raising factor were used to estimate the sample based crops estimates.

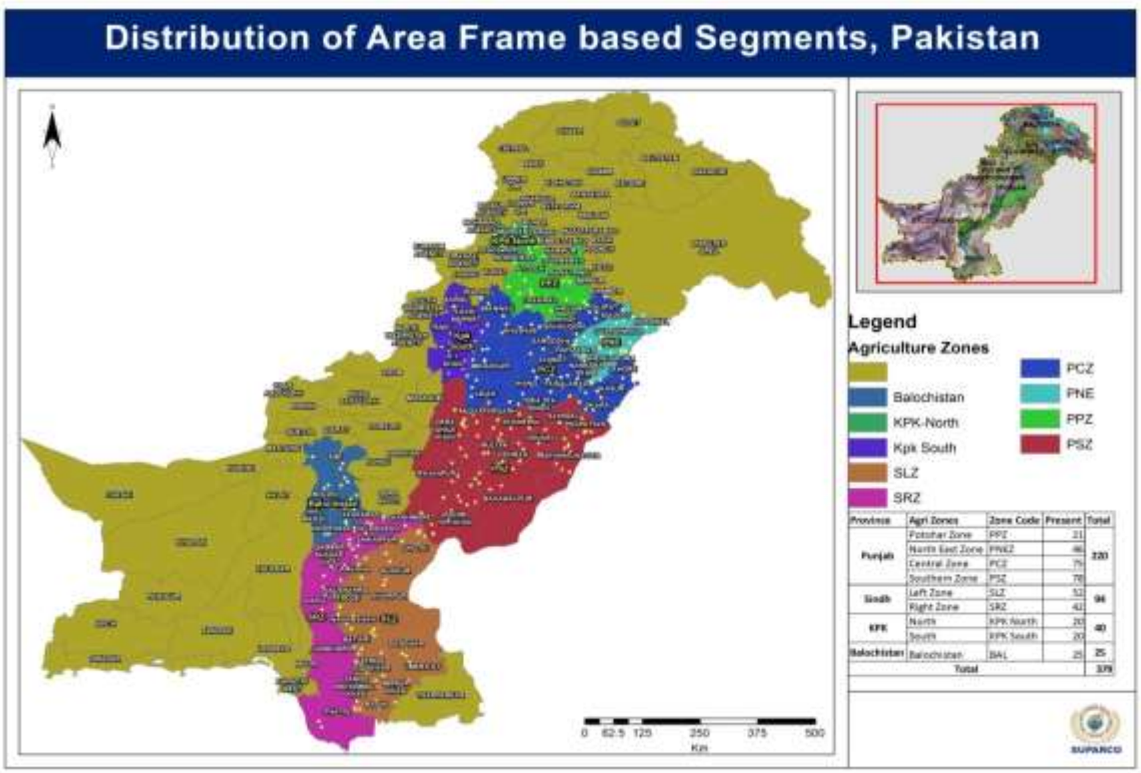

Figure 1: Satellite based Area Frame Sampling System showing distributed Ground Survey Samples (Segments)

Figure 2: Ground Surveyed Segments and Digitized Information

Beside Area frame sampling, satellite image classification was used to estimate the crops acreage using SPOT-5 satellite data of the different time during growing season including early growth and peak growth of crops (Figure 3). Early season image shows that majority of crop is still under sowing stage whereas peak season satellite data reflects fields with actively growing crops.

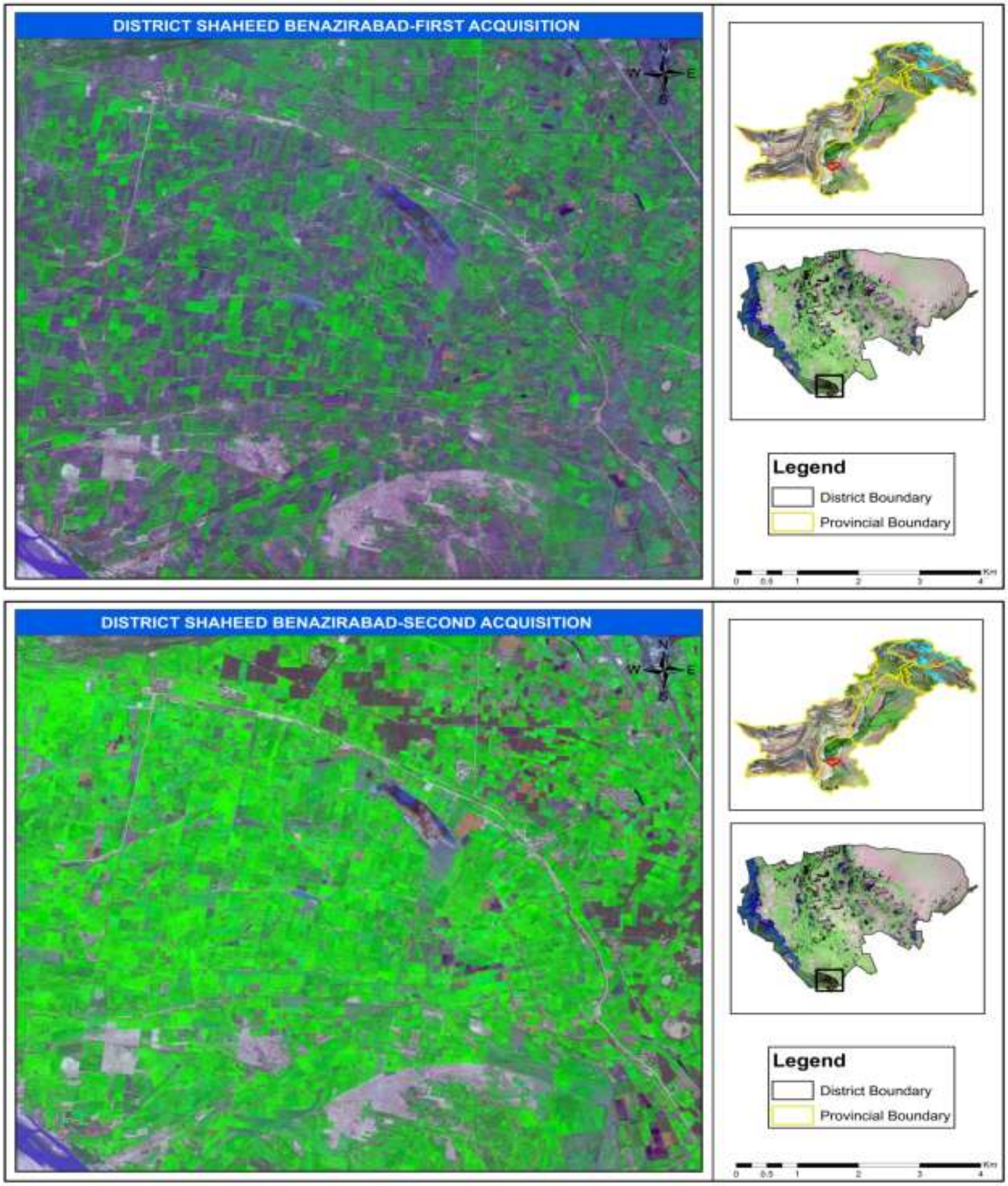

Figure 3: SPOT5 satellite data (Ist and 2^{nd} acquisition)

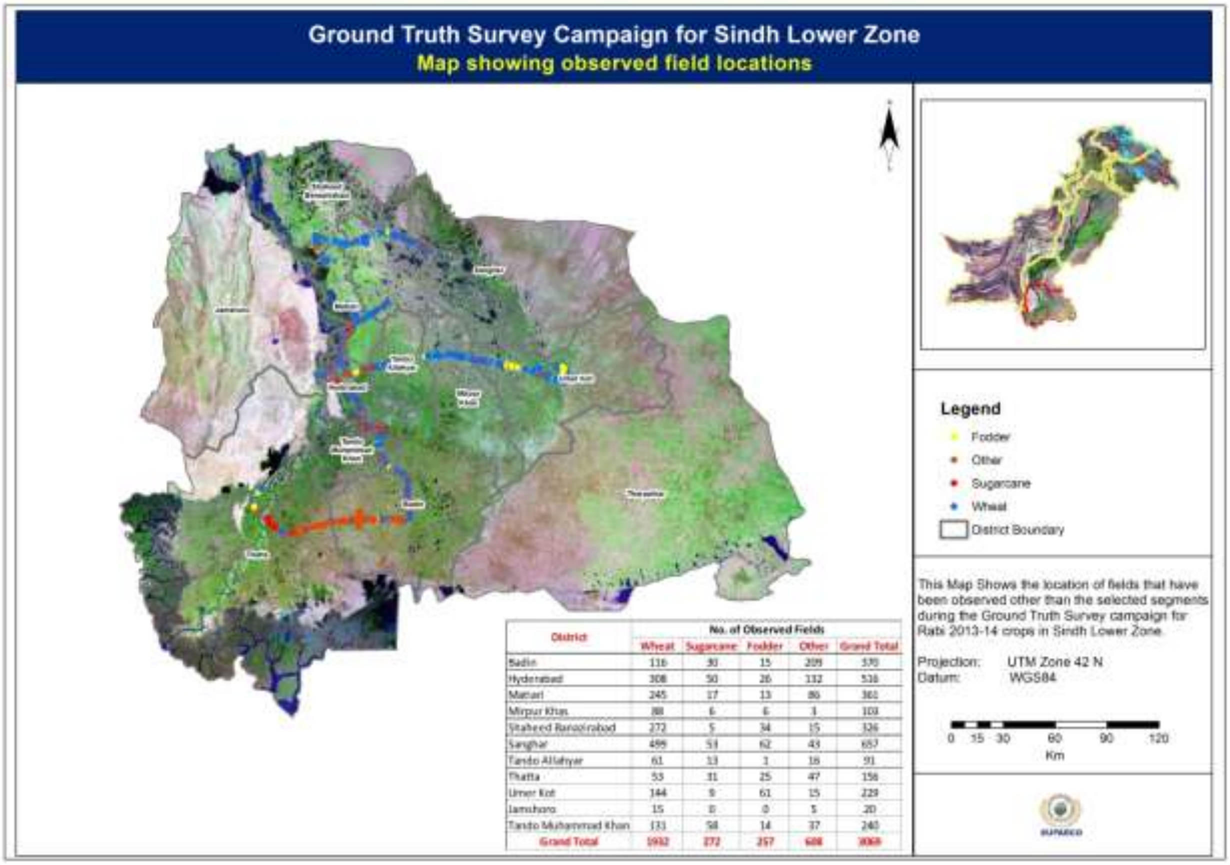

Figure 4: Crops Fields on two different time SPOT5 satellite data

Supervised classification with Gaussian Maximum Likelihood method was adopted. Information on agriculture and non-agriculture were collected out during ground truth survey. Random independent crops signatures were also collected during survey to compensate the spatial context in Segments information (Figure 4). These marked fields points were divided into training (70%) and testing (30%) data through random selection tool in ArcGIS software. Training data was used to train the supervised classifier and classified data was produced as an output (Figure 5). Overall, classification ranged between 85-95% depending on the satellite data quality, number of crops grown, crop type and topography of the area. Quality of classified data was assessed through confusion matrix analysis by using independent testing data. Overall, quality test proved to be useful method to revisit less quality classified data. Accuracy assessment of data ranged from 85-95%. The classified data were subsetted at administrative level to compile the district wise crops estimates. These estimates were compared with those released by government of Pakistan like wheat (Table 3).

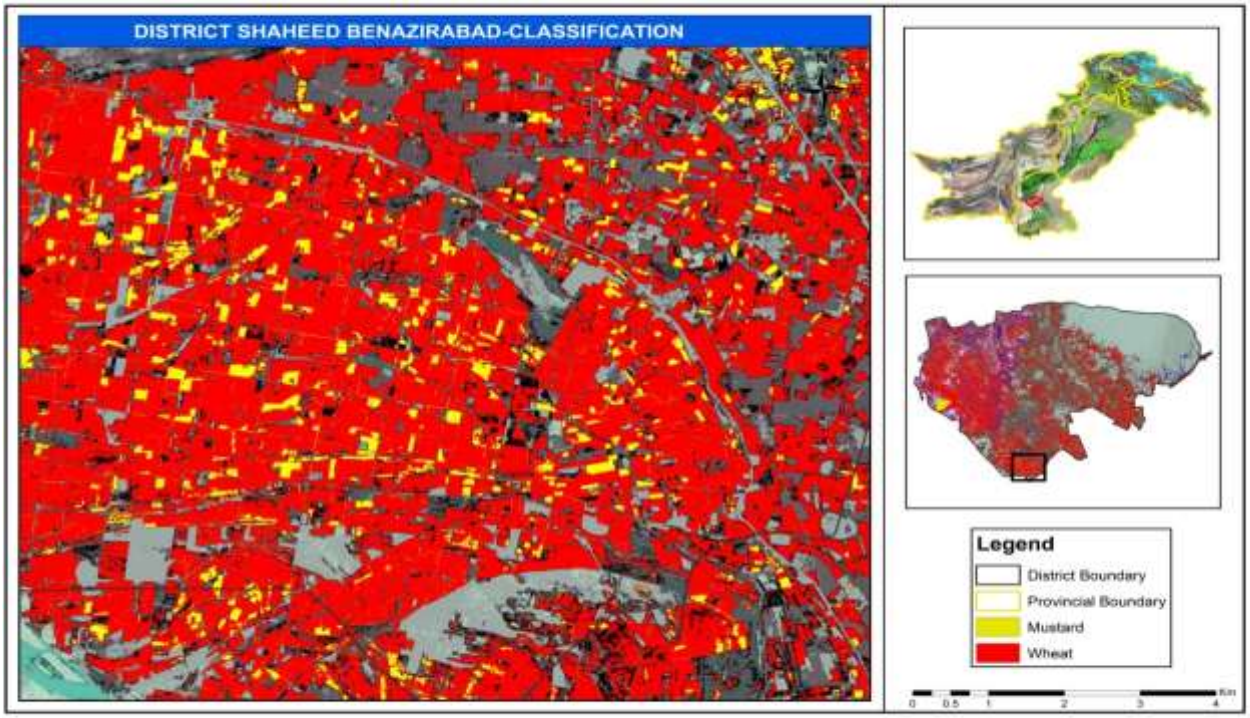

Figure 5: Classified SPOT5 satellite data

Crop yield forecasting and estimation cover r important dimension of crops statistics being of mostly qualitative nature. SPOT Vegetation data is main yield predicting variable in all calibrated models. Crop phenology was mapped along with the related NDVI values to find out the direct relationship with crop yield. All spatial database was developed into a model calibrating matrix. NDVI profile helped to differentiate the crop performance during different years (Figure 6&7).

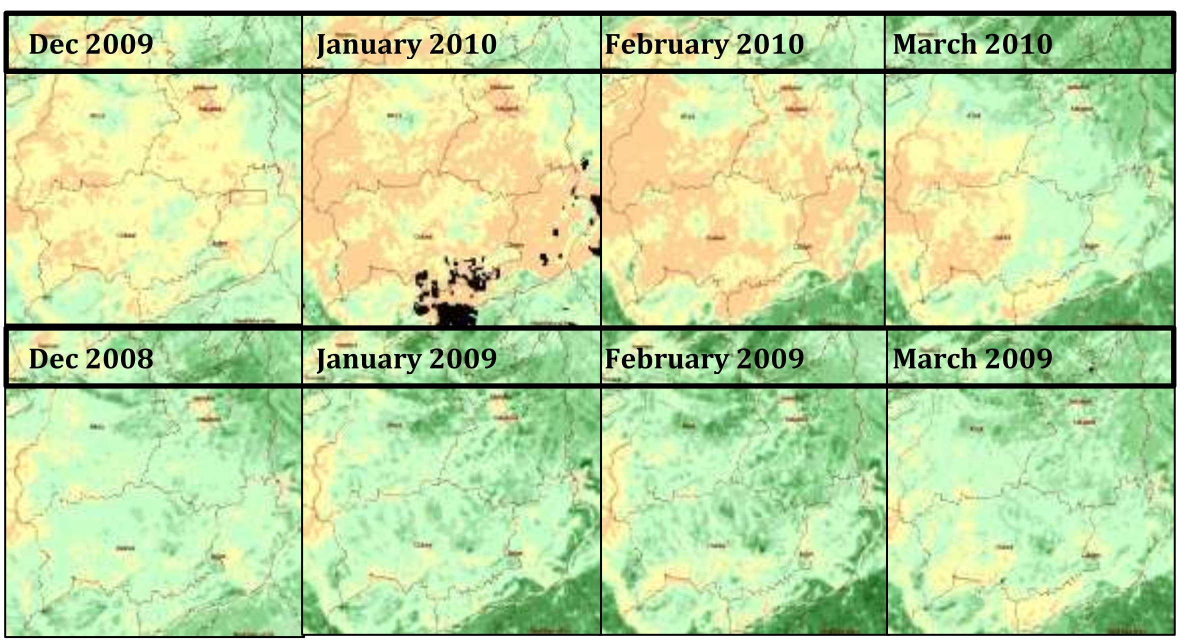

Figure 6: Monthly SPOT NDVI behavior in Rainfed area of Punjab.

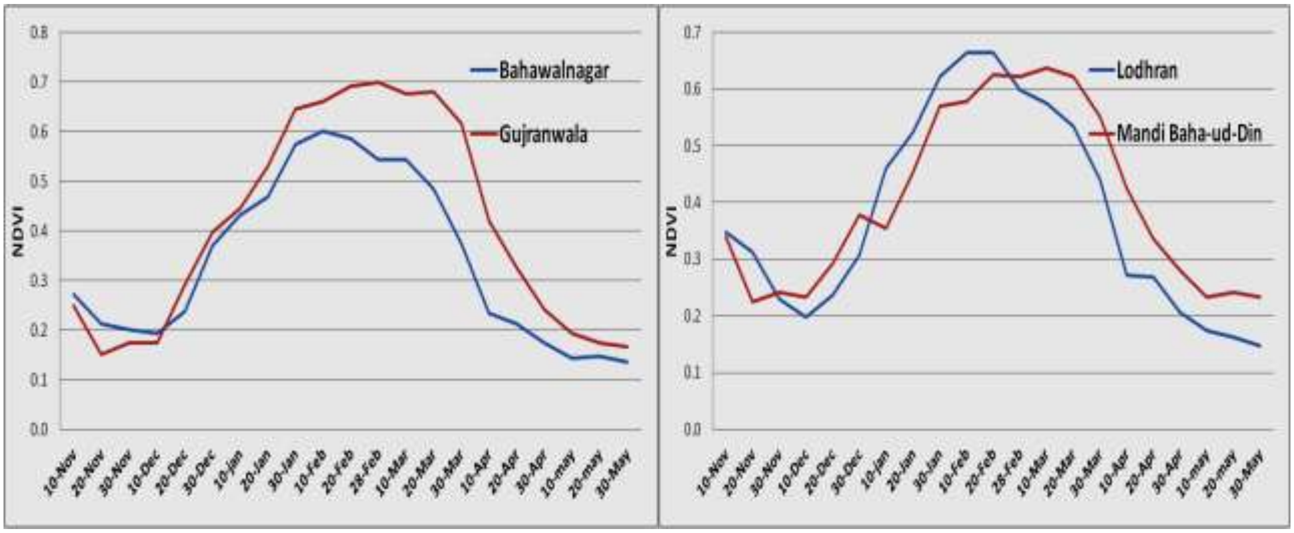

Figure 7: NDVI growth profile at district level.

PCA was applied to reduce the dimensionality in the predictor variables (Figure 8). Significant variables explaining the variance of 99% were selected. These selected variables were tested with multi-colinear test to identify the false relationship among predictor variables to reduce the biasness in multiple regression crop yield model.

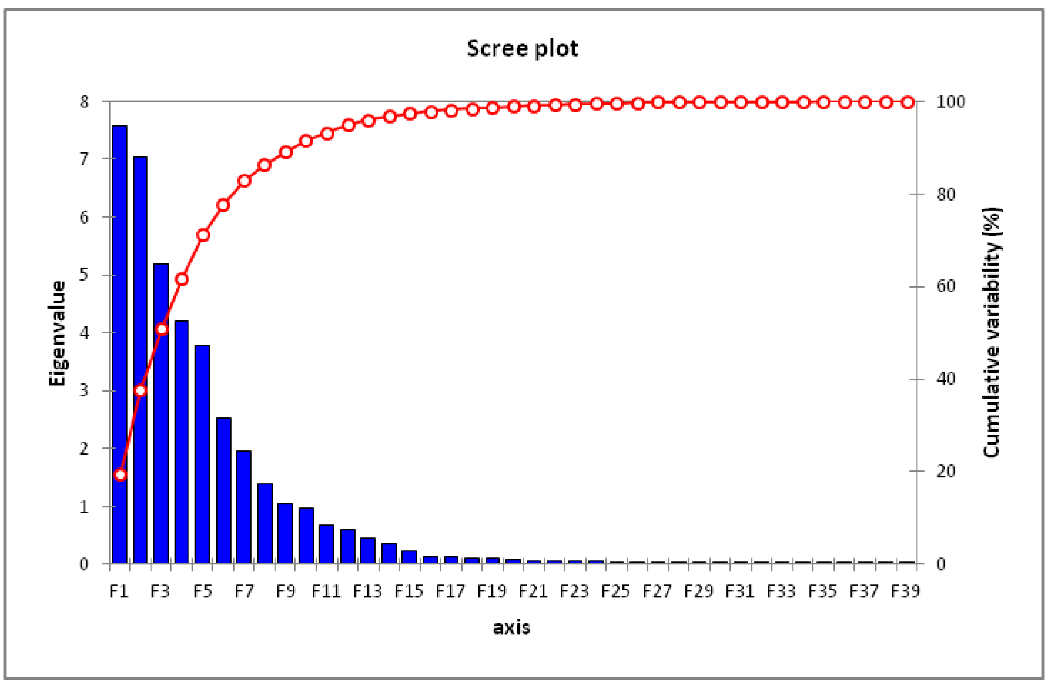

Figure 8: PCA Analysis and cumulative variation.

All major crops models including wheat, cotton, rice, sugarcane and maize were calibrated for yield forecasting during initial to peak growth season and estimated near harvest time. Yield historical data was regressed with multiple predictor variable and only most significant variables were selected. (Figure 9)

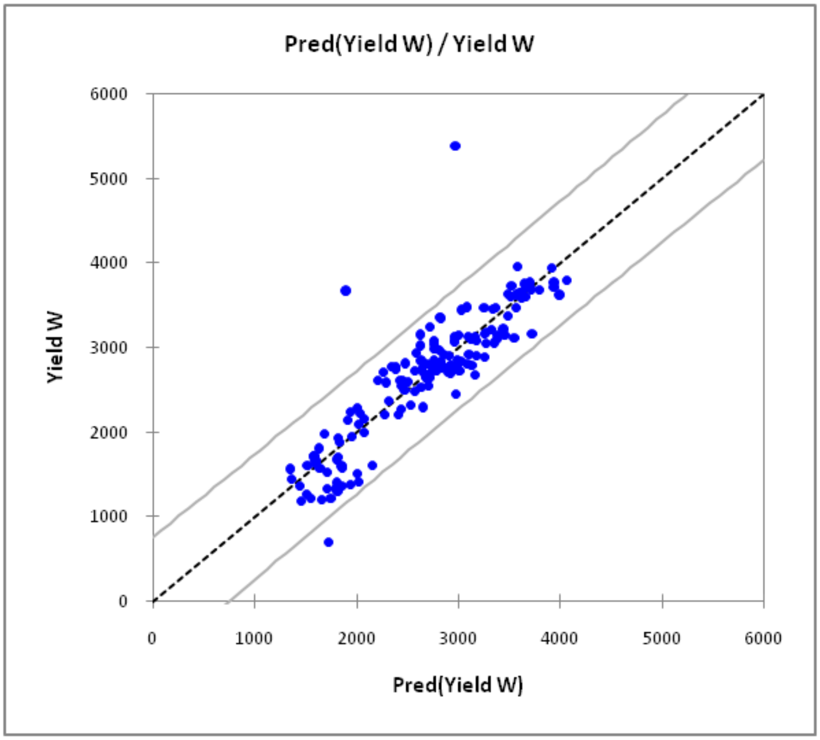

Figure 9: Model performance showing model predicted yield against observed wheat yield (kg/ha)

Model parameters with their coefficients were used to estimate yield for each crop for current season (Table 3).

Table 3: Calibrated Wheat crop yield model parameters and coefficients for yield estimation.

Model parameters:

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Intercept	1549.003	769.794	2.012	0.052	-12.211	3110.217
ppt_nov-mar	1.554	0.219	7.103	< 0.0001	1.110	1.997
Mxt10	62.588	11.400	5.490	< 0.0001	39.467	85.709
ppt_jan	5.669	0.853	6.646	< 0.0001	3.939	7.399
Amp15	-112.037	20.551	-5.452	< 0.0001	-153.716	-70.357
Mxt5	-112.256	23.547	-4.767	< 0.0001	-160.011	-64.500
Mnt4	129.066	24.143	5.346	< 0.0001	80.103	178.030
Yavg	0.825	0.149	5.525	< 0.0001	0.522	1.127

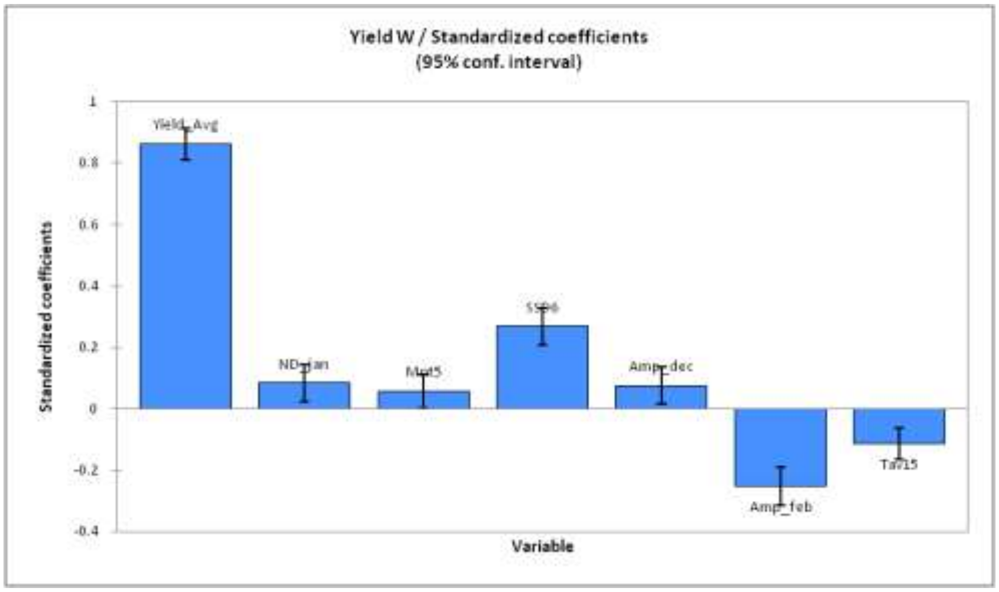

Figure 10: Model parameters co-efficient value with standard error

Satellite data based crops area and yield estimation were compiled and compared later with government official statistics (Table 4). This comparison was to outline the significance of the data being produced through satellite technology system. SUPARCO has developed crop monitoring system based on satellite data whereas agriculture department estimation system is based on revenue department village census data. Main advantage of the SUPARCO satellite based crops system is the timeliness release of data. Wheat and other Rabi crops statistics are released by end of March to mid of April every year whereas official statistics are released after 4-6 months of crop harvest from October to November.

Table 4: Comparison of wheat estimate of SUPARCO data with Official statistics

Season	Province	SUPARCO Wheat Estimates			Official Statistics of Wheat			Difference (%)		
		Area (000 ha)	Yield (kg/ha)	Production (000 tons)	Area (000 ha)	Yield (kg/ha)	Production (000 tons)	Area	Yield	Production
2010-11	**Punjab**	6695.0	2764.0	18505.0	6691.0	2845.8	19041.0	0.1	-2.9	-2.8
	Sindh	1509.0	2585.0	3900.8	1144.4	3746.8	4287.9	31.9	-31.0	-9.0
	K.P	645.2	2015.0	1300.1	724.5	1595.4	1155.8	-10.9	26.3	12.5
	Balochistan	305.0	1967.9	600.2	340.8	2139.6	729.1	-10.5	-8.0	-17.7
	Pakistan	**9154.2**	**2655.2**	**24306.1**	**8900.6**	**2832.8**	**25213.8**	**2.8**	**-6.3**	**-3.6**
2011-12	**Punjab**	6621.0	2270.0	18340.2	6482.9	2736.2	17738.9	2.1	-17.0	3.4
	Sindh	1482.2	2519.0	3733.7	1049.2	3585.2	3761.5	41.3	-29.7	-0.7
	K.P	757.9	1599.0	1211.9	729.3	1549.9	1130.3	3.9	3.2	7.2
	Balochistan	349.0	2133.0	744.4	388.4	2169.7	842.7	-10.1	-1.7	-11.7
	Pakistan	**9210.1**	**2609.0**	**24030.2**	**8649.8**	**2713.7**	**23473.3**	**6.5**	**-3.9**	**2.4**

4. Conclusion

Remote sensing based agriculture monitoring is an important component of food security information system which provides reliable and timely crop area estimates and crop production forecasts at national, regional and global scale. The System contributes to support policy making to ensure food security. To develop fast track and reliable procedures to make crop forecasts and estimations early in the season or end of season, Pakistan Space and Upper Atmosphere Research Commission (SUPARCO), the Space Agency of Pakistan started to develop crop area estimation procedures and crop yield models, based on the application of satellite remote sensing, GIS technology, agronomy, agro-meteorology, statistics and other allied disciplines. System has been developed based on 2.5 to 10 meter high resolution and SPOT Vegetation data of one square kilometer. The image acquisition was carried twice during each cropping season at a time span of 4 weeks and 8 weeks after sowing of crops.

Satellite data based crops area and yield estimation were compiled and compared later with government official statistics. SUPARCO satellite based crops system provides fast track and reliable crop forecasts and estimates.

References

Akhtar, I. H. (2011). Crop Yield Forecast Modelling: A case study of Sugarcane in Pakistan from Combined Remote Sensing and Agrometeorological Techniques. VDM Verlag Dr. Müller, Germany,112 pages.

Benedetti, R., and Rossini, P., (1993). On the use of NDVI pro. les as a tool for agricultural statistics: The case study of wheat yield estimate and forecast in Emilia Romagna. Remote Sensing of Environment, 45, 311–326.

Bullock, P.R.. (1992). Operational Estimates of Western Canadian Grain Production Using NOAA AVHRR LAC Data. Canadian Journal of Remote Sensing, 18(4), pp. 23-28.

Bussay, A and I.H. Akhtar. (2008). Wheat yield/production forecasting and estimation Technology. Pakistan Space and Upper Atmosphere Research Commission, Islamabad, Pakistan. 25 Pp.

Bussay, A and I.H. Akhtar. (2009). Crop yield/production forecasting and estimation Technology for Kharif Crops (Cotton, Rice & Sugarcane). Pakistan Space and Upper Atmosphere Research Commission, Islamabad, Pakistan. 62 Pp.

Doraiswamy, P. C., and Cook, P. W., (1995). Spring wheat yield assessment using NOAA AVHRR data. Canadian Journal of Remote Sensing, 21, 43–51.

Farooq, O. (2014). Agriculture, In: Economic Survey of Pakistan. Ministry of Finance, Government of Pakistan. p 23-41.

Hinzman, L. D., Bauer,M. E., and Daughtry, C. S. T., (1986). Effects of nitrogen fertilization on growth and re • ectance characteristics of winter wheat. Remote Sensing of Environment, 19, 47–61.

Javid, M., Obaid-ur-Rehman, M. Hanif and I. Iqbal. (2010) Pakistan: Spatial Interpolation technique of Temperature Estimation for Crop Forecasting and food Security. Second APSCO

Symposium on Food Security & Monitoring of Agriculture through Satellite Technology Islamabad, Pakistan. 21-24 September, 2010

Khan, M.H., I. Ahmad, M. I. Bhatti, M. Asif and I. H. Akhtar. (2007). International FAO Training in Crop Yield Forecasting Modelling at University de Liege, Belgium, 30 Pp. http://www.hoefsloot.com/Downloads/pakistan_UTF-101-crop_forecasting_training.pdf

Obaid-ur-Rehman, M. Hanif, M. Javid and I. Iqbal. (2010) Pakistan: Monitoring of Wheat Crop for Food Security, Using Satellite Remote Sensing and GIS Application. Second APSCO Symposium on Food Security & Monitoring of Agriculture through Satellite Technology Islamabad, Pakistan. 21-24 September, 2010

Obaid-ur-Rehman, M. Hanif, I. H. Akhtar, I. Sofia, M. Javid. (2011) RS-GIS based Crop Monitoring and Forecasting System. National Conference on "Sustainable Agriculture in Changing Climate", Bara Gali, Pakistan, 7-9 July, 2011

Philipson, W.R., and W.L.Teng. (1988). Operational Interpretation of AVHRR Vegetation Indices for World CropInformation. Photogrammetric Engineering and Remote Sensing, 54(1), pp. 55-59.

Quarmby, N. A.,Milnes, M, Hindle, T. L., and Silleos, N., (1993). The use of multi-temporalNDVI measurements from AVHRR data for crop yield estimation and prediction.International Journal of Remote Sensing, 14, 199–210.

Tucker, C. J., (1979). Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment, 8, 127–150.

Wiegand,C. L., (1984). The value of direct observations of crop canopies for indicating growing conditions and yield. T he 18th International Symposium on Remote Sensing of Environment, Paris, France, October 1–5, pp. 1551–1560.

Wiegand, C. L., and Richardson, A. J., (1990). Use of spectral vegetation indicies to infer leaf area, evapotranspiration and yield: I. Rationale. Journal of Agronomy, 82, 623–629.