Semi-supervised technique to retrieve irrigated crops from Landsat ETM+ imagery for small fields and mixed cropping systems of South Asia

MSDN Gamage[1], Mobin-ud-Din Ahmad[2/*], Hugh Turral[3]

[1] Research Officer – International Water Management Institute, PO Box 2075, Colombo, Sri Lanka. n.gamage@cgiar.org

[2/*] Researcher (corresponding author) – International Water Management Institute, PO Box 2075, Colombo, Sri Lanka. a.mobin@cgiar.org

[3] Principle Researcher – International Water Management Institute, PO Box 2075, Colombo, Sri Lanka. h.turral@cgiar.org

Abstract

*Precise land cover maps and crop statistics are essential for monitoring and performance evaluation of any irrigation system. Remote sensing techniques are efficient means of obtaining land cover information with high accuracy in near real time. Many applications developed and referenced in the literature classify large tracts of homogenous cropped area, and may be less effective in highly mixed cropping systems in South Asia. This paper presents a semi-supervised (hybrid) technique to map cropped area in the Punjab Rice-Wheat and Sugarcane-Wheat cropping systems in Rechna Doab, Pakistan. Two Landsat ETM+ images (*Path 149 Row 038*) for September 2001 and March 2002 were selected to represent* Kharif *(2001) and* Rabi *(2001-02) seasons. A ground truth survey was conducted in different cropping areas that covered 0.14% of the total extent and a geo-referenced GIS coverage was prepared for image classification.*

Based on Principle Component Analysis (PCA), Red, NIR and MIR bands were selected and added to NDVI to segregate cropping area based on threshold values. GIS coverages were overlaid on a stacked image and 50% of the collected information was used to extract training signatures for all crops. The training samples were evaluated based on their location and overlap on two scatter plots between NIR:Red and NIR:MIR which yielded seven distinct crop classes. A contingency matrix was prepared using the other half of the field data to evaluate the classification accuracy.

The results show an overall accuracy of 87% in Kharif (Summer-Monsoon) compared to 83% in Rabi (Winter-Dry). The diversity in crops creates a higher spectral mixture in Kharif resulting in slightly lower accuracy. The present study shows how NDVI, when combined with individual bands improves the sensitivity and accuracy of classification.

Keywords: Remote Sensing, GIS, Crop Classification, Land use-Land cover, Spectral Mixture, Pattern Recognition, Mixed Cropping System, Irrigation, Pakistan.

Introduction

Spatio-temporal land use/land cover maps are important for judicious planning and management of limited land and water resources. Traditionally, this information is collected through laborious field surveys. Accuracy of the resulting maps largely depends on the sampling design and proportion of vegetated area surveyed. Furthermore, due to different levels of accuracy, the results of these surveys often can not be used to compare spatial and temporal land use changes. With emergence of satellite remote sensing, several attempts have been made to develop a consistent and rapid approach to map land cover and generate crop statistics (Oetter et al. 2000).

Unsupervised classification is the easiest and most common way to classify satellite images but the technique suffers from lower accuracy with mixed land cover signatures. To overcome these problems and improve classification accuracy, supervised classification approaches have been developed over the last two decades. However, most of these approaches were developed for areas with large fields of single homogenous crops and often fail to produce the desired result for small field sizes and mixed cropping systems, which are common in South Asia. According to Hurbert-Moy et al (2001), the method of classification should be based on existing landscape features, but usually it follows the same algorithm without this consideration. The scenario and outcomes differ between uniform large-scale landscapes compared to tiny land parcels with numerous different crops with resulting variation in overall accuracy. Therefore, the options are either to divide the landscape into several more homogenous units or select a classification algorithm to give the best results over the entire study area.

Ahmad et al. (1996) used SPOT XS imagery and field data to classify crops in south-eastern Punjab, Pakistan. In current study an attempt is made to further refine Ahmad et al. (1996) approach and develop a semi-supervised crop classification method for mixed cropping systems in South Asia. This approach was tested using Landsat-7 ETM+ imagery in Rechna Doab, Pakistan.

Description of Study Area

Rechna Doab is a land unit lying between the Ravi and Chenab rivers of Pakistan (Fig. 1). The gross area of the Doab is about 2.97 million hectares and currently around 80% of area is cultivated. It has a maximum length of 403 km, maximum width of 113 km and lies between longitude 71^{o} 48' to 75^{o} 20' East and latitude 30^{o} 31' to 32^{o} 51' North. Climatologically, the area is subtropical and designated as semi-arid, which is characterized by large seasonal fluctuation in temperature and rainfall. Summers are long and hot, lasting from April through September with temperatures ranging from 21^{o}C to 49^{o}C. The winter season lasts from December through February with temperatures ranging between 25^{o}C and 27^{o}C. Mean annual precipitation is about 650 mm in the upper doab, falling to 375 mm in the central and lower area. Nearly 75 percent of the annual rainfall occurs during the monsoon season - from mid-June to mid-September.

Due to scanty and erratic rainfall, successful agriculture is only possible with irrigation. Rechna Doab is one of the oldest and most intensively developed irrigated areas of the Punjab Province of Pakistan. Rice, cotton and forage crops dominate the summer season (*Kharif*) whereas wheat and forage are the major crops in winter (*Rabi*). In central Rechna Doab, sugarcane is the dominant crop, which is a long season annual crop of around 11 months duration. Minor crops include oil seeds, vegetables and orchards. Land distribution is highly fragmented and holdings are variable in size. About 56% land of Rechna Doab is owned by small farmers with farm sizes of 5 hectares or less and only 25% land is with large farmers owning 20 hectares per farm or more (GoP, 2001). Due to low canal discharges, farms have small fields of about 0.3 ha or so. To maximize gross margins, farmers adopt mixed cropping patterns. Both low flow rate and attempts to maximize farm gross margin result in small field sizes and a mixed cropping pattern.

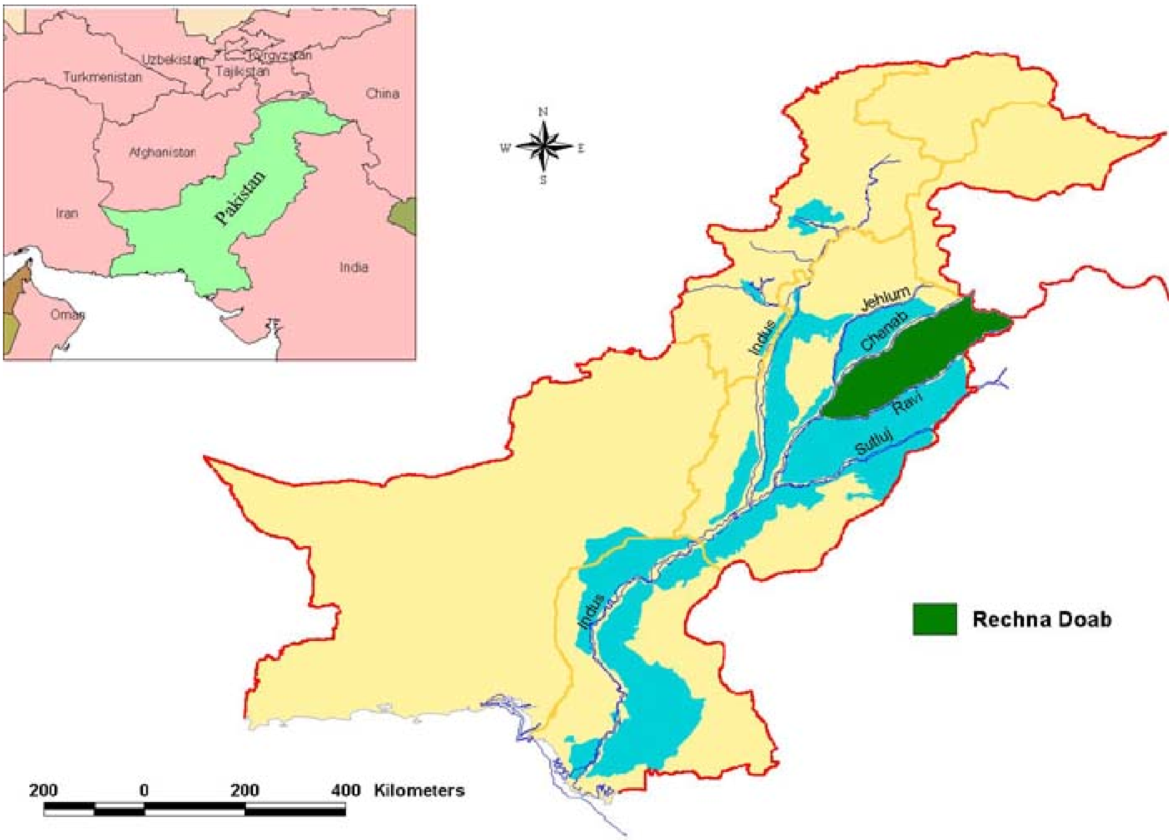


Figure 1 The location of Rechna Doab in Pakistan

Data Collection and Management

Satellite Imagery

Two 30m resolution Landsat7 ETM+ images (*Path 149* and *Row 038*) of 30th September 2001 and 25th March 2002 were selected to represent *Kharif* (2001) and *Rabi* (2001/2002) seasons for crop classification. According to crop calendar these images were acquired when the crop in field were fully matured. A detailed GPS survey was performed to geo-reference satellite images. The digital numbers (DN) of geo-referenced images were converted into reflectance values according to the methods outlined in the Landsat7 Science Data Users Handbook (2004). Since the gain on each of the Landsat7 channels can vary from high to low between successive overpasses, it is essential to convert apparent digital numbers to reflectance for consistency between different images.

Field Data

Information collected through two crop surveys covering sample watercourses in the Punjab rice-wheat and Punjab sugarcane-wheat areas was used in this study (Figure 2). Crop type, sowing date, and harvesting date were collected at *Killa* (67m x 60 m cell size) level. This information was collected in previous studies conducted by the International Water Management Institute (IWMI) for rice-wheat consortium and decision support projects. Although the selected watercourses are not well distributed across Rechna Doab (as required under ideal scenario) but still sufficiently represent crops in Rechna Doab. For image processing, *Killa* grids are generated for each sample watercourse and geo-referenced in ARCGIS. This process allowed the display of crop information directly on the satellite imagery, in order to obtain training samples for classification. and the net sample area covers about 0.14 % of whole study area.

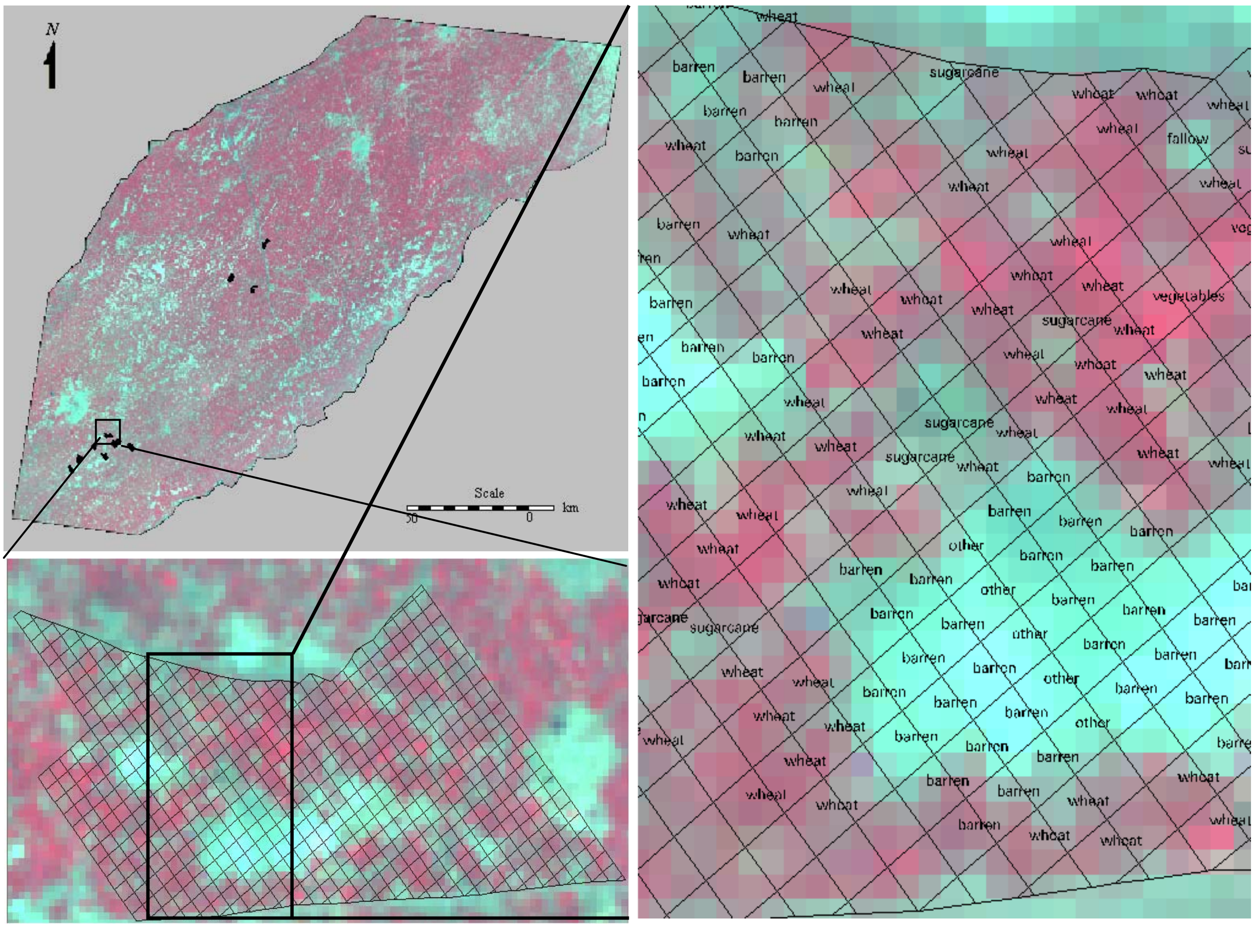


Figure 2 Location of sample watercourse in the Rechna Doab. Each grid represents a *Killa* and crop information collected within that unit.

Classification Approach

A semi-supervised classification methodology is developed for crop classification using high spatial resolution imagery (shown schematically in Fig. 3). The main steps of proposed methodology are:

1. Segregate vegetative area from non-vegetative areas. For this a commonly used vegetation index, NDVI, was calculated using the reflectance of near infrared (NIR) λ_{NIR} and red band λ_{red},

$$NDVI = (\lambda_{NIR} - \lambda_{red}) / (\lambda_{NIR} + \lambda_{red})$$

2. The collected GIS data was overlaid on calculated NDVI and, based on field information; a threshold value was estimated for each image to segregate vegetative/irrigated areas from non-vegetative areas. Reflectance from vegetative areas was used to perform Principal Component Analysis (PCA) to select best band combinations for vegetation classification. To further enhance the vegetation signatures, NDVI was stacked with selected bands selected by PCA.

3. Thereafter, Feature Space Images (FSI) were plotted for all band combinations. Fifty percent of geo-referenced crop data were used to obtain training samples for all land use classes. These training samples were evaluated based on their location on the various scatter plots (Figure 4a & b, Figure 5a & b) and finally merged into broad crop categories. These final classes were used for supervised classification using the parametric rule of maximum likelihood.

4. Finally, using the remaining 50% of field data, a contingency matrix were computed for accuracy assessment.

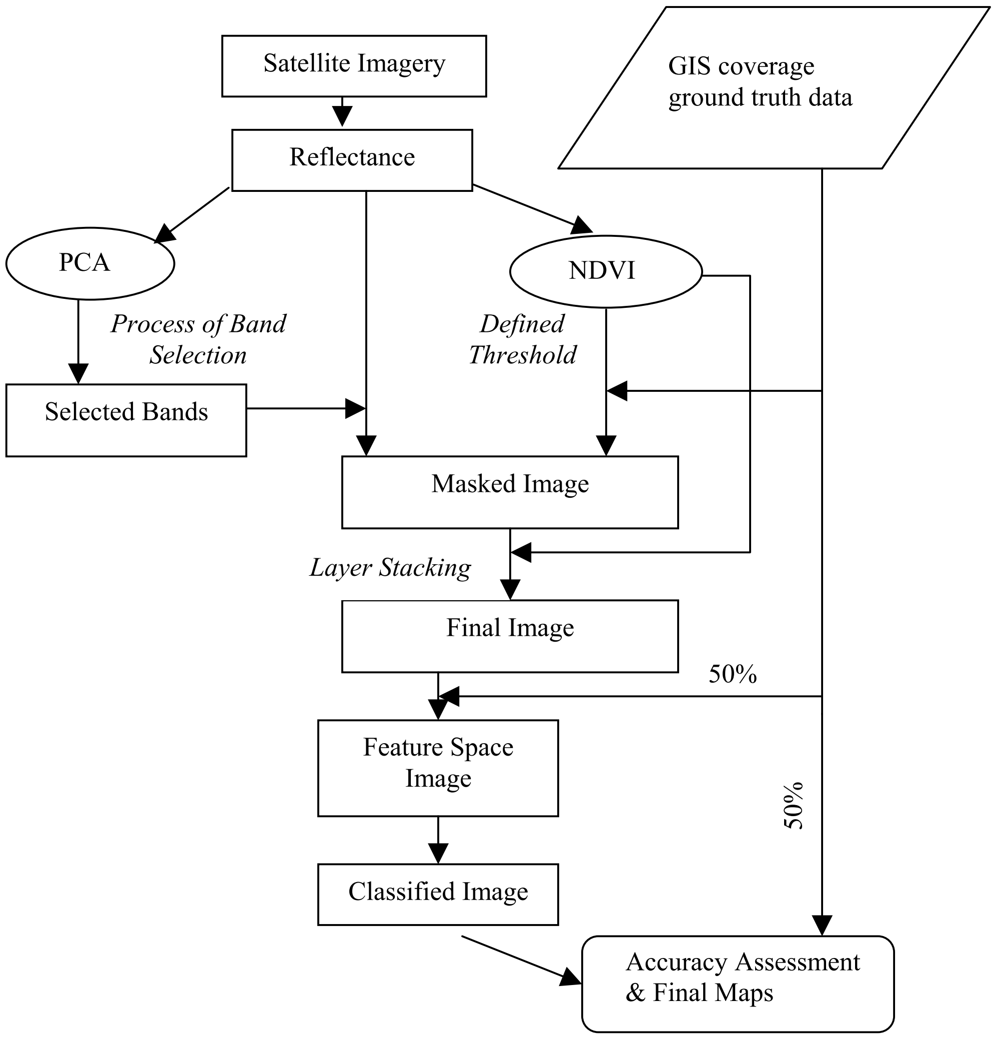


Figure 3 Schematic diagram of the semi-supervised crop classification methodology

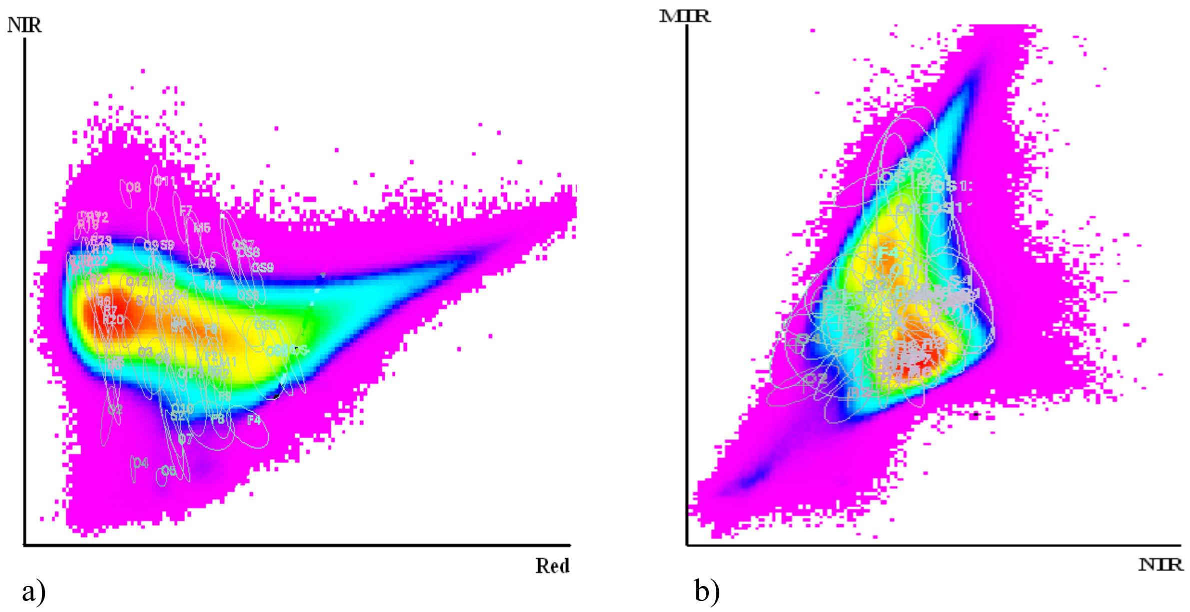
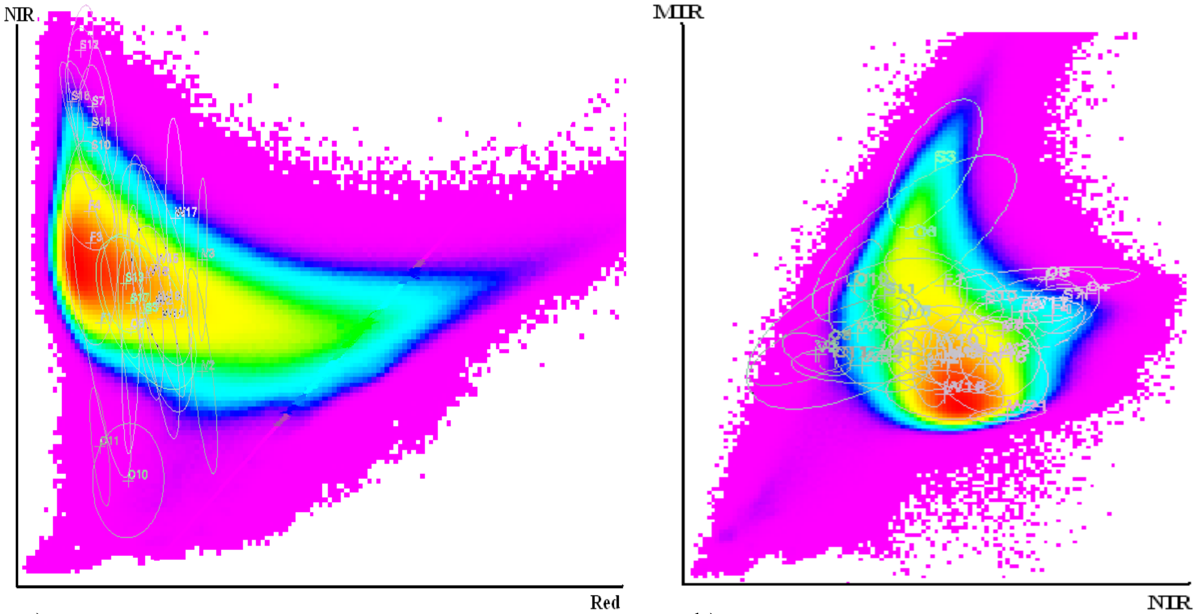


Figure 4 Feature Space Image between Red, Near Infrared (NIR) and Mid Infrared (MID) bands of Landsat image of 30th September 2001. The polygon names starting "R" represent rice, "O" for orchard, "M" Maize, "F" Fodders, "OS" Oil seed, and "SC" for Sugarcane.



a) b)

Figure 5 Feature Space Image between Red, Near Infrared (NIR) and Mid Infrared (MID) bands of Landsat image of 25th March 2002. The polygon names starting "W" represent wheat, "O" for orchard, "F" Fodders, "V" vegetables and "SC" for Sugarcane.

Results and Discussion

The classified images for *Kharif* 2001 and *Rabi* 2001-02 are presented in Fig. 6 and Fig. 7, respectively. Rice is prominent in the *Kharif* season and it represents more than 42% of the total land of the classified image (Table 1). In contrast, wheat is the dominant crop in *Rabi* and covers 54.5% of the total area. The Punjab rice-wheat and Punjab sugarcane-wheat zones can easily be distinguished in both images and there is clear transition between these two agro-climatic zones. In both seasons and zones all crops are planted but the areal extent of rice is more than 60% in Punjab rice-wheat zone and only 20% in Punjab sugarcane-wheat zone. It is noticeable that sugarcane is prominent in the southern part of the Doab. Sugarcane represents 10.8% of *Kharif* crop area and 10.6% in *Rabi* season. This difference could be attributed to intercropping of wheat or fodder with sugarcane planting. Moreover, in *Kharif*, sugarcane has a similar reflectance to rice in NIR and Red bands (Fig. 4a), but separation was achieved by analyzing the feature space plot of NIR versus MIR (Fig. 4b). *Kharif* fodder (sorghum & maize) has a similar signature to Maize grown for grain. This difference could be also differentiated in all feature space images. Perhaps this might be the reason that area under fodder is comparatively greater in *Kharif* (16.6%) than *Rabi* (14.0%).

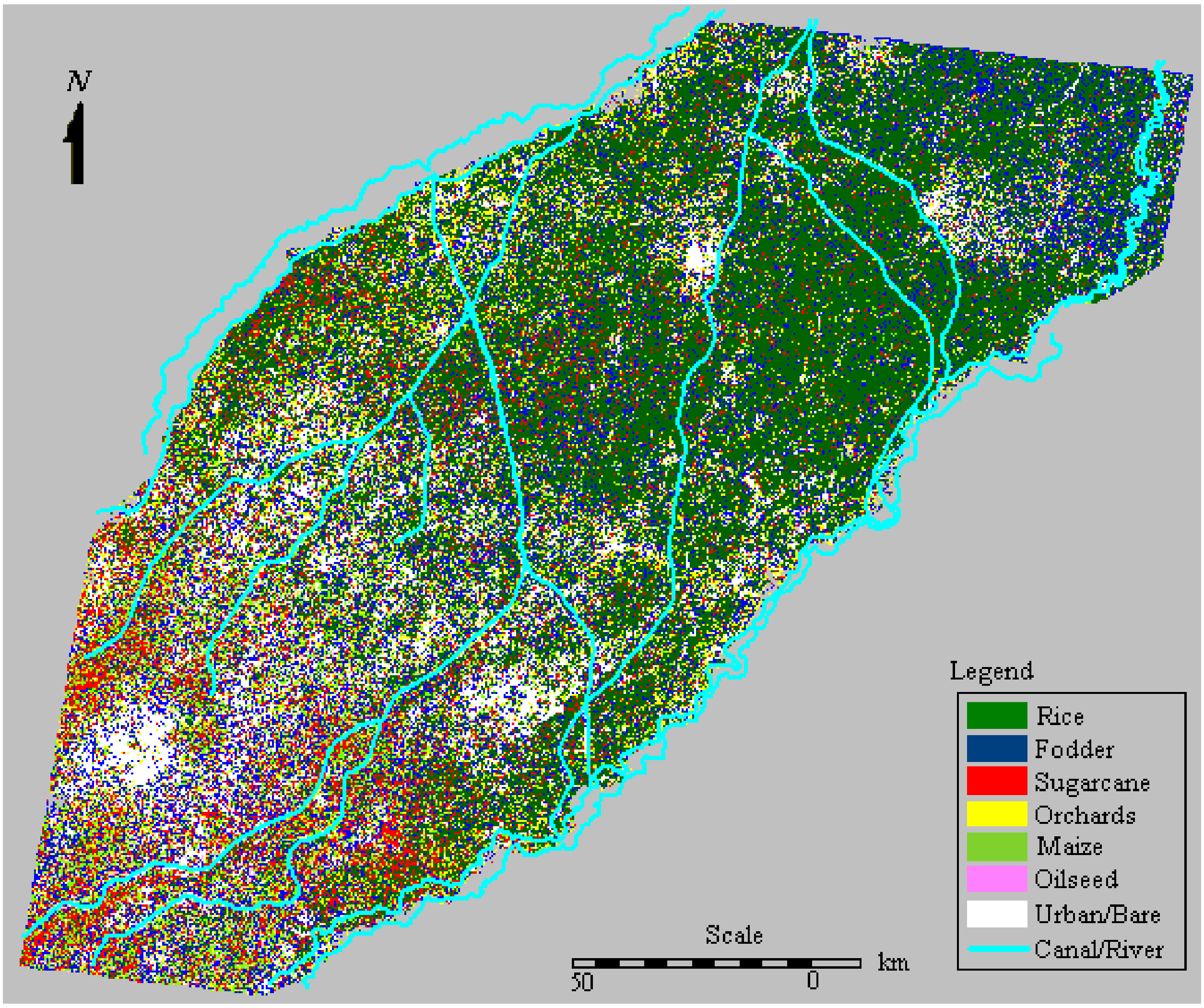


Figure 6 Land use/Land cover classification based on Landsat ETM+ image 30th September 2001 (*Kharif*-2001)

Orchard areas (6.4%) in *Rabi* are under-estimated compared to *Kharif* (8.6%) due to intercropping. This results in a mixed signature of pure orchards and orchards with wheat. But healthy dense orchards are always clearly distinguishable since its thick prominent canopy layer has a high NDVI. Sparse orchards could not be segmented as a pure crop in feature space and are thus classified mostly as wheat since wheat was the most prominent land cover, and also because there is little leaf cover in deciduous orchards in winter. This reduces the estimated percentage of orchards in *Rabi* season by 2 %.

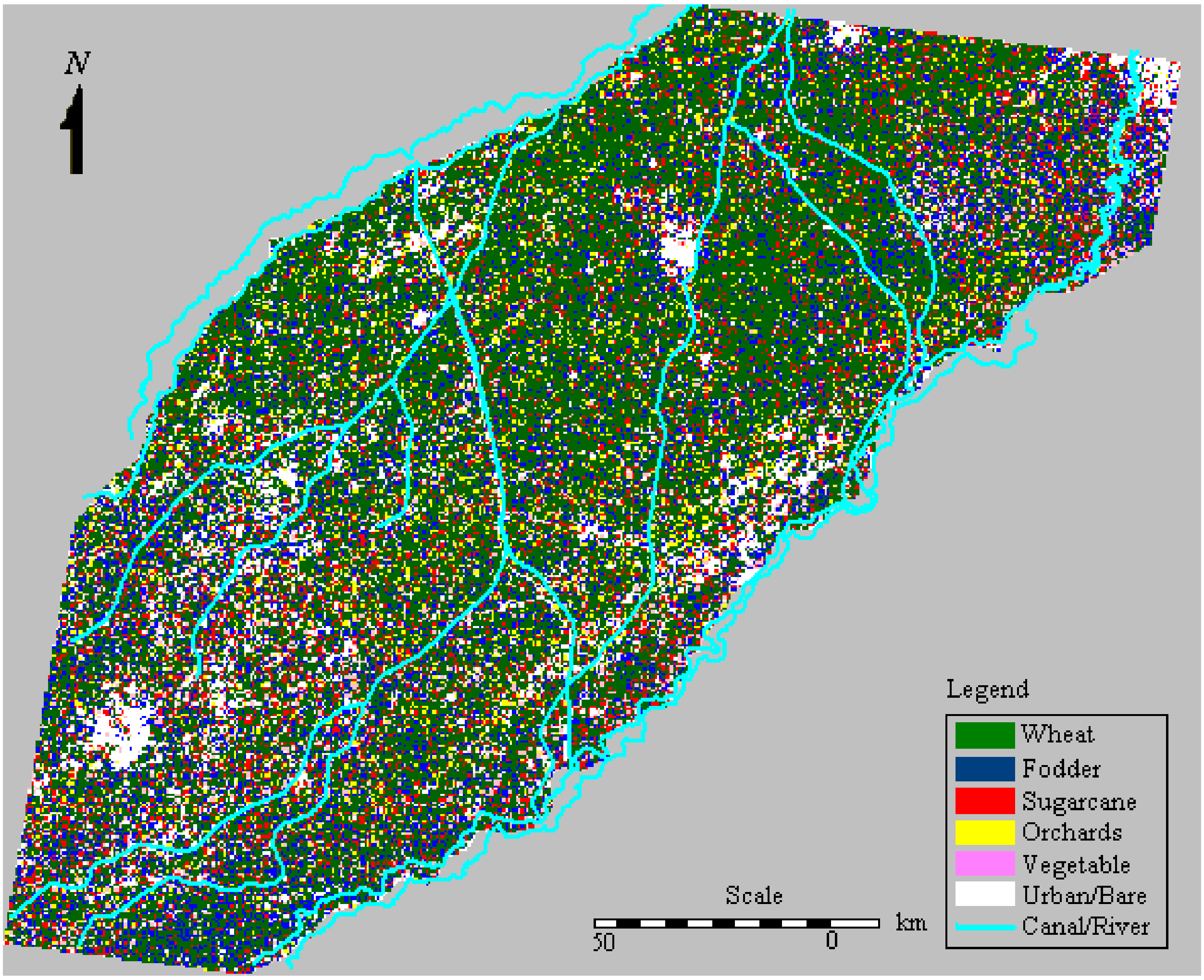


Figure 7 Land use/Land cove classification based on Landsat ETM+ image 25th March 2002 (*Rabi* 2001/2002)

Due to a lack of field information, oil seeds, cotton and vegetables were the most difficult to identify in both images. Cotton is not a major crop in the classified image and hence resulted in no field information from selected watercourses. For oil seeds and vegetables, this was mainly due to very small plot size and more complex mixing especially of vegetable types. The sample size of oil seeds and vegetables that was used to classify the image was also very small; hence identification of crops was more difficult. Especially in *Kharif*, there were no pure *Killa* of *Kharif* vegetables and therefore we were unable to identify vegetables in this season. The masked Bare/Urban area of *Kharif* (10.9%) is higher than *Rabi* (9.3%) and because the area of bare/fallow land in *Kharif* season increases slightly due to insufficient water available for farming.

Table 1. The results of land cover classification in *Kharif* 2001 and *Rabi* 2001/2002.

***Kharif* 2001**			***Rabi* 2001/02**		
Crop	**Area (ha)**	**%**	**Crop**	**Area (ha)**	**%**
Rice	825930	42.4	Wheat	1061676	54.5
Fodder	324191	16.6	Fodder	272099	14.0
Sugarcane	210199	10.8	Sugarcane	206069	10.6
Orchards	168175	8.6	Orchards	123998	6.4
Maize	148046	7.6	Vegetable	102803	5.3
Oil seed	57677	3.0	Bare/Urban	180547	9.3
Bare/Urban	212975	10.9			
Total	1947192	100	Total	1947192	100

For accuracy assessment, contingency matrixes were computed. According to the contingency tables those were prepared using the 50% field data, not used for training samples. (Table 2 and 3), the overall classification accuracy was calculated by dividing the number of correctly classified samples by the total number of pixels checked. Two ways of measuring the accuracy of an individual class were used: the *producer's accuracy* and the *user's accuracy*. Producer's accuracy measures the probability of a reference site being correctly classified whereas User's accuracy is an indication of the probability that a site classified in the image actually represents that class on the ground. The user's accuracy is calculated by taking the number of correctly classified samples and dividing that number by the classified total. The producer's accuracy is found by dividing the number of correctly classified samples by the reference total for that class.

Table 2 Contingency table for the classified image of 30th September 2001.

Crops	Rice	Fodder	Sugarcane	Orchards	Maize	Oilseed	*Total*	User's Accuracy (%)
Rice	**580**	6	16	20	0	5	*627*	**92.50**
Fodder	3	**123**	14	2	2	6	*150*	**82.00**
Sugarcane	14	10	**151**	3	2	0	*180*	**83.89**
Orchards	3	8	2	**58**	2	0	*73*	**79.45**
Maize	0	6	1	0	**7**	0	*14*	**50.00**
Oilseed	6	7	3	3	0	**60**	*79*	**75.95**
Total	*606*	*160*	*187*	*86*	*13*	*71*	*1123*	-
Producer's Accuracy (%)	**95.71**	**76.88**	**80.75**	**67.44**	**53.85**	**84.51**	-	**87.18**

Table 3 Contingency table for classified image of 25th March 2002

Crops	Wheat	Fodder	Sugarcane	Orchards	Vegetable	*Total*	User's Accuracy (%)
Wheat	**978**	28	116	16	24	*1162*	**84.17**
Fodder	23	**52**	0	8	0	*83*	**62.65**
Sugarcane	4	0	**62**	5	1	*72*	**86.11**
Orchards	12	0	1	**54**	0	*67*	**80.60**
Vegetable	1	0	1	1	**16**	*19*	**84.21**
Total	*1018*	*80*	*180*	*84*	*41*	*1403*	-
Producer's Accuracy (%)	**96.07**	**65**	**34.44**	**64.29**	**39.02**	-	**82.82**

The overall accuracy of the *Kharif* season supervised classification yielded 87 % pixels correctly identified (Table 2). Except orchard and maize, all other crops were classified with high accuracies. Maize classification accuracy was 54% producer's and 50% user's. The apparent reason for this low accuracy is intercropping with orchards and similarity in signatures of *Kharif* fodder (sorghum & maize) with maize grown for grain. In *Rabi* season, the overall accuracy was 83% (Table 3). Wheat is the dominant crops in *Rabi* season and classified with higher accuracy: 96% producer's accuracy and 84% user's accuracy. Sugarcane and fodder have lower producer's accuracy, which could be attributed to intercropping of wheat with sugarcane planting. Vegetables were also classified with low accuracy, and as discussed in previous section, this could be mainly due to very small plot size and more complex mixing especially of vegetable types.

Conclusions and future directions

The objective of the work was to develop a semi-supervised classification approach for crop classification which is suitable for small and mixed cropping patterns. Results reveal that the methodology is sound enough to calculate land use land-cover statistics with sufficient accuracy within the existing complex situation. NDVI worked well in demarcating vegetation and non-vegetated area. An attempt was made to clearly identify intercropping but this needs at least two images per season since intercropping consists of perennials inter-sown with seasonal crops. Moreover, the ground truth data used in this study was not collected for the purpose of classification and therefore better accuracy can be achieved with better geographically sampled data, especially for crops with low reference samples, such as vegetables and maize. Further to very fine demarcation of crop types, knowledge of cropping pattern is essential for the selection of appropriate timings for image selection.

Further work will compare the results of the semi-supervised classification with classification derived from a 3-date Landsat Pan composite image (R: date 1; G: date 2; B: date 3) at 15m resolution. In this technique, the image is segmented using object oriented techniques and then classified using unsupervised classification ISO Cluster algorithm with subsequent class identification and naming.

Acknowledgement

The authors wish to acknowledge Remote Sensing and GIS team at IWMI Pakistan for their contribution of collecting ground data and preliminary image processing work. Further thanks to all support staff at IWMI Pakistan with field assistance.

References

Ahmad, M.D., J. Jaminson, S. Asif and P. Strosser, 1996. Use of Satellite imagery for Land use mapping and Crop Identification in Irrigated Systems in Pakistan, 17^{th} Asian Conference of Remote Sensing (*ACRS*), Colombo.

Government of Pakistan, 2001. Agricultural Census 2000, Province Report Vol. II Part-2, Punjab. Government of Pakistan Statistics Division, Agriculture Census Organization, Lahore Pakistan.

Hurbert-Moy, L., A. Cotonnec, L.Le Du, A. Chardin and P. Perz, 2001. A Comparison of Parametric Classification Procedures of Remotely Sensed Data Applied on Different Landscape Units, *Remote Sensing of Environment*, 75:174-187, Elsevier Science Inc, New York.

Oetter, D.R., W.B. Cohen, M. Berterretche, T.K. Maiersperger and R.E. Kennedy, 2000. Land cover mapping in an agricultural setting using multiseasonal thematic mapper data, *Remote Sensing of Environment* 76: 139-155, Elsevier Science Inc, New York.

URL: Landsat 7 Science Data Users Handbook, Chapter 11-Data Products, 2004. Available at http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_htmls/chapter11/chapter11.html