Journal of Statistical Science and Application 2 (2014) 93-101

Application of Statistical Methods and GIS for Downscaling and Mapping Crop Statistics Using Hypertemporal Remote Sensing

Ahmed DOUAIK

Research Unit on Environment and Conservation of Natural Resources,

National Institute of Agricultural Research (INRA), Rabat, Morocco. Email: ahmed_douaik@yahoo.com

To sustain the management of natural resources, land use and land cover (LULC) should be spatially mapped and temporally monitored using GIS. For large areas, conventional methods are laborious. Alternatively, remote sensing can be used for LULC mapping and monitoring. Normalized differential vegetation index (NDVI) is the most used vegetation index for crop identification and phenology. For agricultural areas, crop statistics are estimated yearly at regional level following administrative units. However, these statistics are not informing about spatial extent of these crops within administrative units; such information is crucial for crop monitoring. The main objective of this research was to fill the gap, based on statistical methods and GIS, by adding spatial information to crop statistics by analyzing temporal NDVI profiles. The study area covers 1300 km^2. Data consist of 147 decadal Spot Vegetation NDVI images. Crop statistics were compiled on seasonal basis and aggregated to different administrative levels. Images were processed using an unsupervised classification method. A series of classification runs corresponding to different numbers of clusters were used. Using stepwise multiple linear regression, cropped areas from agricultural statistics were related to areas of each NDVI profile cluster. Estimated regression coefficients were used to generate maps showing cropped fractions by map units. The optimal number of clusters was 18. Similar profiles were merged leading to eight clusters. The results show that, for example, rice was grown, in autumn, on 50% of the area of map-units represented by NDVI-profile group 4 and 75% of the area of group 7 while it was grown, in spring, on 2, 69 and 25% of areas of NDVI-profile groups 2, 6, and 7, respectively. Regression coefficients were used to generate map of crops. This research illustrates the benefit of integrating statistical methods, GIS, remote sensing and crop statistics to delineate NDVI profile clusters with their corresponding agricultural land cover map units and to link these statistics to geographical locations. These map units can be used as a reference for future monitoring of natural resources, in particular crop growth and development and for forecasting crop production and/or yield and stresses like drought.

Keywords: Crop Statistics; GIS; Multiple Regression; NDVI; Unsupervised Classification.

Introduction

The food needs of the ever increasing world population should be satisfied quantitatively and qualitatively. Since the spatial extent of arable lands is limited, the focus is currently on a better and sustainable use and management of natural resources, including soil and land resources. In order to attain this sustainability, land

---

Corresponding author: Ahmed DOUAIK, Research Unit on Environment and Conservation of Natural Resources, National Institute of Agricultural Research (INRA), Rabat, Morocco. Email: ahmed_douaik@yahoo.com.

use and land cover should be spatially mapped and temporally monitored. As the areas to be mapped are very large, conventional methods, through aerial photo-interpretation are laborious and expensive (Tucker, 1979; Philipson, 1997; Taylor et al, 2000; Falkner and Dennis, 2002). Alternatively, satellite remote sensing tools can be beneficially used for land use and land cover mapping and monitoring (Cihlar, 2000; Lillesand et al, 2007; Giri, 2012). In this way, the normalized differential vegetation index (NDVI), initially proposed by Rouse et al (1973) and which measures the vigor and greenness of vegetation (Tarpley et al, 1984), is the most used among the vegetation indices for studying vegetation, and specifically crop, phenologies (Sarkar and Kafatos, 2004; Sakamoto et al, 2005; White et al, 2009; Atkinson et al, 2012; You et al, 2013) and yield forecasting (Das et al, 1993; Ferencz et al, 2004; Mkhabela et al, 2005). Remote sensing was also used for estimating cropped area (Campbell et al, 1987; Sheub and Atkins, 1991; Labus et al, 2002; Howard et al, 2012). Time series of NDVI were used to discriminate between vegetation and other land uses (Nordberg and Evertson, 2003; Knight et al, 2006) and, for vegetation, between different green areas specific to crops, forests, etc (Murakami and al, 2001; Balaghi et al, 2008; Xie et al, 2008). For agricultural areas, crop statistics (mainly cropped areas and production) are estimated yearly at regional level following given administrative units (USDA, 2014). However, these statistics are not informing about the spatial extent of these crops within administrative units, such information is crucial for crop monitoring in future time. Since more than two decades ago, remote sensing was helpful in determining crop acreage and productivity (Allen, 1990; Gonzalez-Alonso et al, 1997; Carfagna and Gallego, 2005). However, it is just very recently that hypertemporal remote sensing was used to describe and map variability of cropping patterns of different crops in Spain (Khan et al, 2010), rice in Vietnam (Nguyen et al, 2012), winter crops in Australia (Potgieter et al, 2013), relationship between the fraction of evergreen forests and the presence of epiphyllous liverworts in China (Jiang et al, 2013), gradient in land cover vegetation growth in Greece (Ali et al, 2013), natural landscape heterogeneity in Greece (Ali et al, 2014). The main objective of this research work was to fill the gap in the official crop statistics by adding them the spatial information through the analysis of hypertemporal remote sensing, i.e., temporal NDVI profiles using different statistical methods.

Material and Methods

Study area

The study area is situated in the western part of Nizamabad district, Hyderabad province, Andra Pradesh state, in central India. The district has an irrigation system used for rice cultivation, cotton cultivation on vertisols, dryland cropping on poor sandy soils, forests on hilly land, and degraded areas. The study area is spatially very heterogeneous. The soils are classified into four main orders: Inceptisols, Alfisols, Vertisols, and Entisols. The climate is tropical with hot summers (maximum mean monthly temperature of about 40 °C) and cool and dry winters (maximum mean monthly temperature of about 13 °C). Regarding rainfall, it is about 900 mm which occur during 2 months during the southwest monsoon. Six Mandals or sub-districts are concerned by this study covering 1300 km^2 from which 90000 Ha are agricultural lands and 18000 Ha are shrub and non cultivated area.

Data

Data consist of 147 geo-referenced and stacked Spot Vegetation composite NDVI images provided by VITO (http://www.VGT.vito.be). They have a spatial resolution of 1 km^2 and available on a decadal basis for a period ranging from April 1998 to April 2002. Two other types of data were used. Land cover map at 1/50000 scale was established from images acquired in 1994/1995 by the Indian remote sensing satellite IRS-C using the Liss-III sensor (spatial resolution of 23 m). For this work the original 18 legend entries were simplified and reduced to seven. Regarding crop statistics, they were compiled on seasonal basis and aggregated to different administrative levels (CPO, 2001).

Methods

The normalized differential vegetation index is defined by:

$$NDVI = (IR - R) / (IR + R) \quad \text{Eq (1)}$$

IR and R are the infra red (0.78 – 0.89 μm) and red (0.61 – 0.68 μm) bands, respectively, which are bands 3 and 2 for Spot Vegetation.

The NDVI values were reported as digital number (DN) values, ranging between 0 and 255, using the following equation:

$$DN = (NDVI + 0.1) / 0.004 \quad \text{Eq (2)}$$

The stacked 147 NDVI images were processed using ISODATA clustering algorithm (Mather and Koch, 2011), an unsupervised classification method, available in the Erdas Imagine software (Erdas, 2003). A series of classification runs corresponding to different number of clusters (2 to 30) were used. The ISODATA algorithm tries to minimize the Euclidian distance to form clusters. The results of the different runs are compared using the divergence separability which is a statistical measure of distance (Landgrebe, 2003); the 'best' number of clusters is the one corresponding to the run having the highest minimum and/or average divergence (Swain and Davis, 1978). The maximum number of iterations was 50 and the divergence threshold was 1. The spectral signatures of the clusters are represented graphically and similar NDVI profiles are merged to reduce the unnecessary large number of clusters.

Once the number of clusters is known, the NDVI profile clusters map is established and compared to the land cover map to match the preliminary legend of the former with the legend of the latter and also to get an idea about the land cover classes that are present in each of the NDVI profile clusters.

The NDVI profile clusters map, a raster, is converted to polygons and cropland areas are masked by using the land cover map to keep only NDVI profile clusters corresponding to agricultural land (Maselli and Rembold, 2001; Kastens et al, 2005).

Using GIS spatial analysis functions from ArcGIS (ESRI, 2009), the Mandals and the agricultural masked NDVI profile clusters map are overlaid to determine the respective areas (Ha) of each NDVI profile cluster by Mandal. These areas are further used as explanatory or independent variables, in the stepwise multiple linear regression (Neter and al, 1996), with the cropped areas (Ha) from agricultural statistics by season, crop, and Mandal as dependent variable:

$$CA = \sum_{i=1}^{n} c_i * NDVIcluster_i \quad \text{Eq (3)}$$

with CA representing cropped area (Ha) by Mandal and $NDVIcluster_i$ representing the area (Ha) of the i^{th} NDVI profile cluster.

No constant was considered in the regression and the coefficients c_i were constrained to the 0 – 1 range in order to determine the estimated fraction or percentage of total area of a given NDVI profile cluster where a given crop was grown at a given Mandal and a given season. Once the regression coefficients were estimated, the above equation was used to generate maps showing cropped fractions by map units. Statistical computations were done using the SPSS software (SPSS, 2008).

Results and Discussion

Average and minimum divergence values between clusters corresponding to the different runs (2 to 30 clusters) are reported in Figure 1.

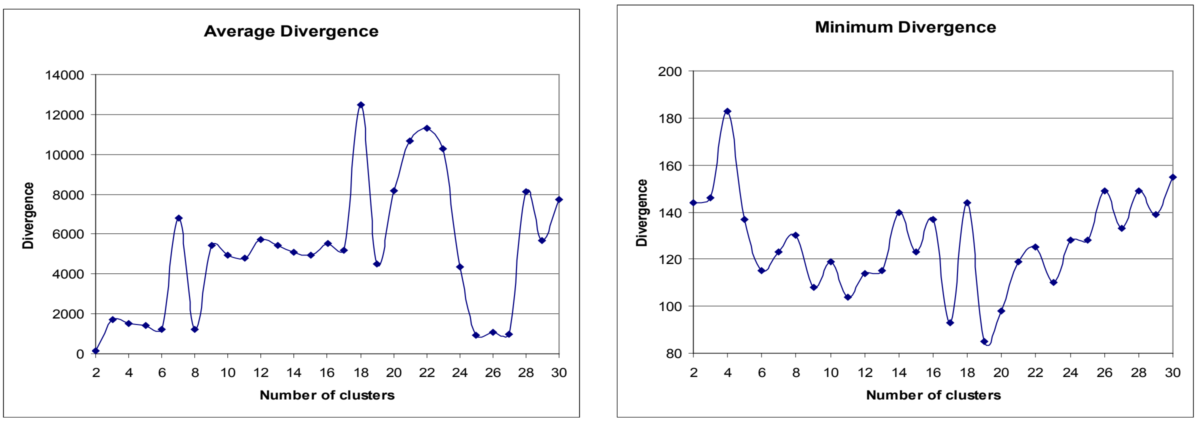

Figure 1. Average (left) and minimum (right) separability divergence values.

The highest value for the average divergence corresponds to 18 clusters whereas the highest one for minimum divergence corresponds to only 4 clusters while 18 clusters resulted in a reasonable high divergence value. So, based on these results, the optimal number of clusters given the best separability between them was taken to be 18. The corresponding average spectral signatures are displayed on Figure 2.

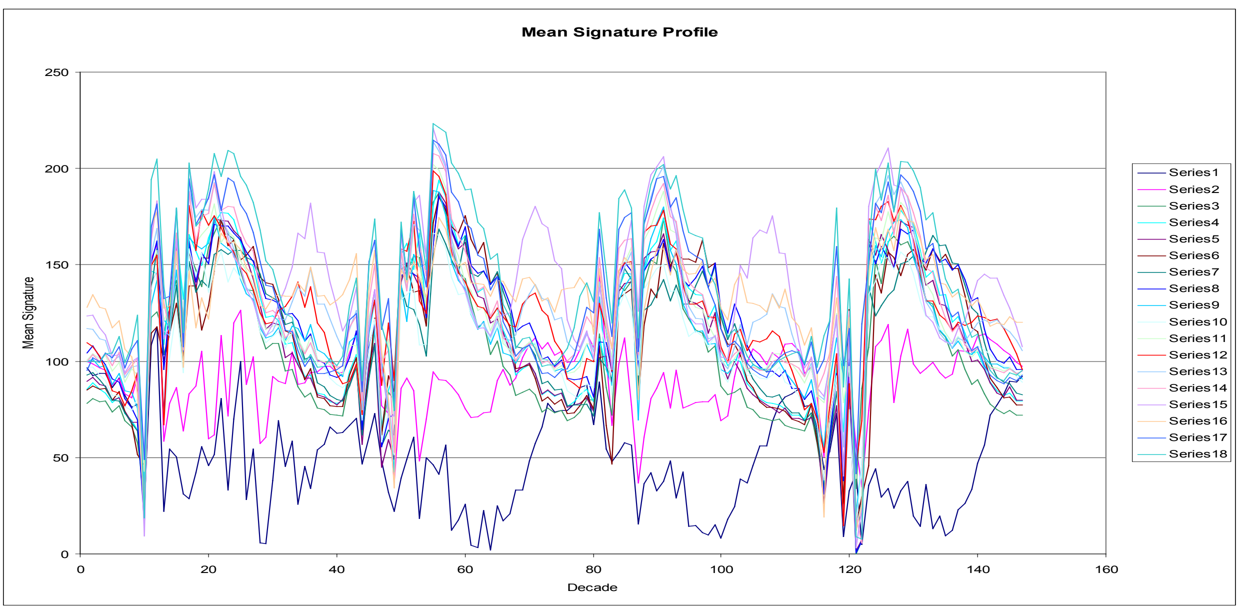

Figure 2. Average spectral signatures of the 18 NDVI profile clusters.

From this figure, some profiles (mainly 1, 2, and 15) have a distinct pattern while most of the others have a more or less similar pattern. Most similar profiles were merged: profiles 3 to 7 and 10; profiles 8, 9, 11 and 14; profiles 12 and 13; and profiles 17 and 18. This combining of NDVI profiles resulted finally in only eight clusters (Figure 3) and the corresponding NDVI-unit map is displayed in Figure 4.

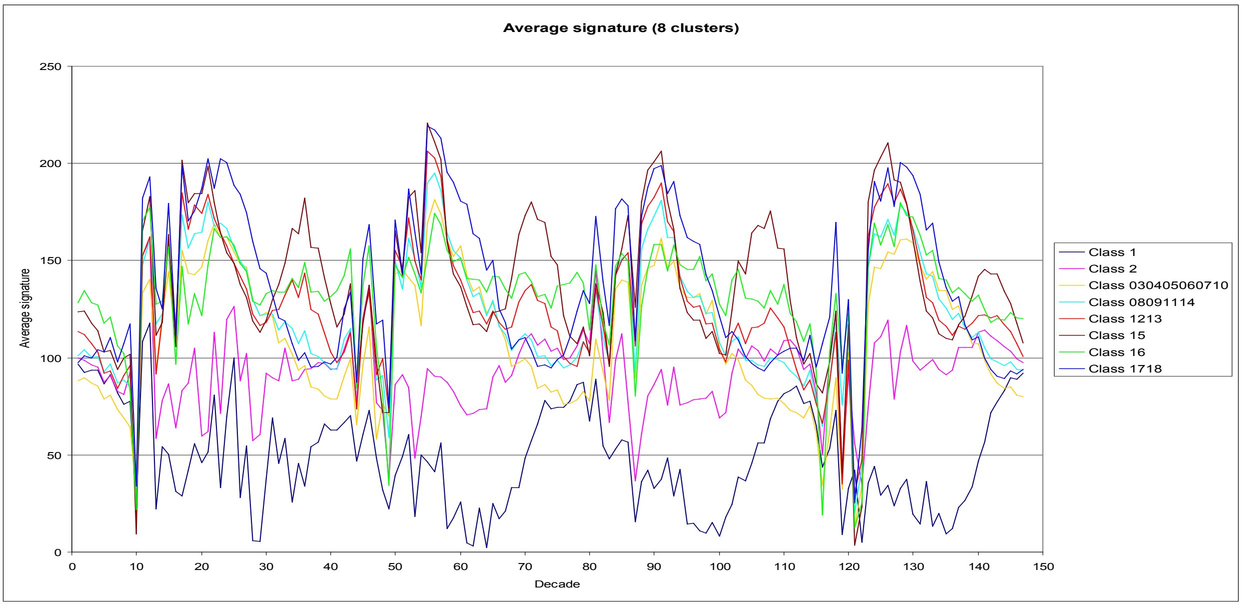

Figure 3. Average spectral signatures of remaining 8 NDVI profile clusters after merging similar ones.

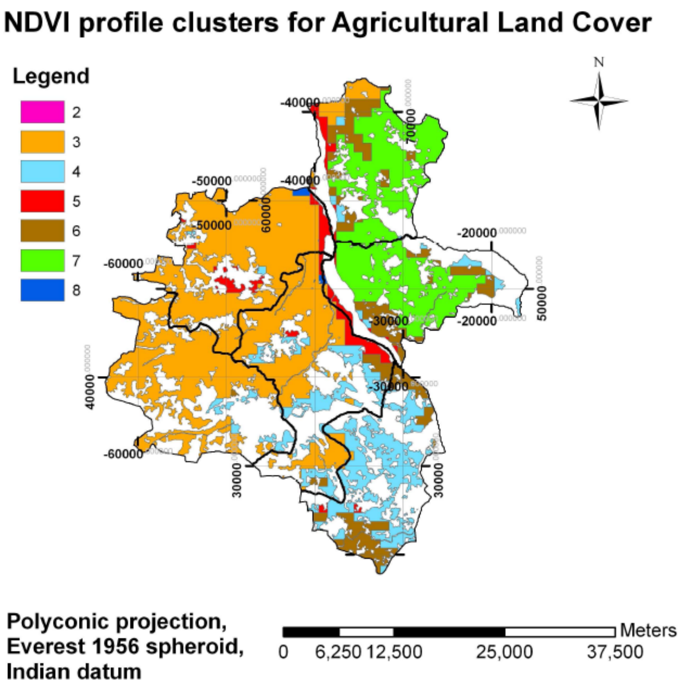

Figure 4. NDVI profile cluster map. N.B: cluster 1 is not present in the six Mandals.

White zones correspond to non agricultural areas. NDVI profile cluster 1 was present in the whole image of the Nizamabad district but not in the six Mandals or sub-districts corresponding to the study area. Also, clusters 2 and 8 are largely under-represented whereas clusters 3, 4, and 7 are much more present. These results are confirmed by statistics provided in Table 1 which show the relation between agricultural land cover (either in Kharif, Rabi or both seasons) and the NDVI profile clusters and their corresponding areas and fractions. Cluster 3 is present in almost half of the total agricultural area, cluster 7 in fifth of this area and clusters 4 and 6 in 15 and 10 %, respectively. In contrast, clusters 2 and 8 are present in less than 1% while cluster 5 is present in 4% of the total agricultural area.

The results of stepwise multiple linear regression, for the main crops by season, are reported in Table 2.

This table shows that clusters 2, 5, and 8 are not involved in the regression models, at least, for the crops used at this step. This relates directly to the very limited extent of these clusters (see Table 1). Rice was grown, in the Kharif season, on 50% of the area of map-units represented by the NDVI-profile group 4 and 75% of the area of group 7 while it was grown, in Rabi season, on 2, 69 and 25% of areas of NDVI-profile groups 2, 6, and 7, respectively.

Table 1

Areas (Ha) and percentage of agricultural land cover by season corresponding to each NDVI-unit

NDVI unit / Season	Kharif	Rabi	Both seasons	Area (Ha)	Percentage
2	1	8	0	9	0.01
3	21602	19806	1001	42409	48.92
4	9875	2199	1414	13488	15.56
5	1289	1104	985	3378	3.90
6	3067	2426	3426	8920	10.29
7	240	3393	14583	18216	21.01
8	164	93	7	264	0.30
Total Area (Ha)	36239	29029	21417	86685	100
Percentage	41.81	33.49	24.71	100	

Table 2

Adjusted R^2 and coefficients (%) for stepwise multiple linear regression with total areas (Ha) for main crops in both seasons

		NDVI units				
Kharif	Adjusted R^2	3	4	6	7	Area (Ha)
Cotton	87.5	15.6				6860
Maize	81.3		4.1			482
Pulses	96.9	48.0	64.1			29121
Rice	95.0		50.3		75.3	22774
Sugarcane	89.9			26.0		2395
Rabi						
Groundnut	80.3			53.2		5942
Pulses	80.9	5.5				2824
Rice	99.8	1.8		69.1	25.0	11481
Sorghum	86.1	32.5				15454
Sugarcane	85.9			21.6		1960
Total Area (Ha) for both seasons		42409	13488	8920	18216	

The above regression coefficients were used to generate map of crops. For illustration, the map of rice, for both seasons, is displayed in Figure 5.

The comparison of these two maps shows that rice is cropped in both seasons mainly in 3 Mandals: the southernmost one and the two located in the North-East of the study area. Regarding the southernmost Mandal, rice is not cropped in the same, but different, locations during the two seasons. For the two North-East Mandals, the non cropped areas (0%) in Kharif were intensively cropped (69%) in Rabi whereas the areas intensively cropped (75%) in Kharif were moderately cropped (25%) in Rabi.

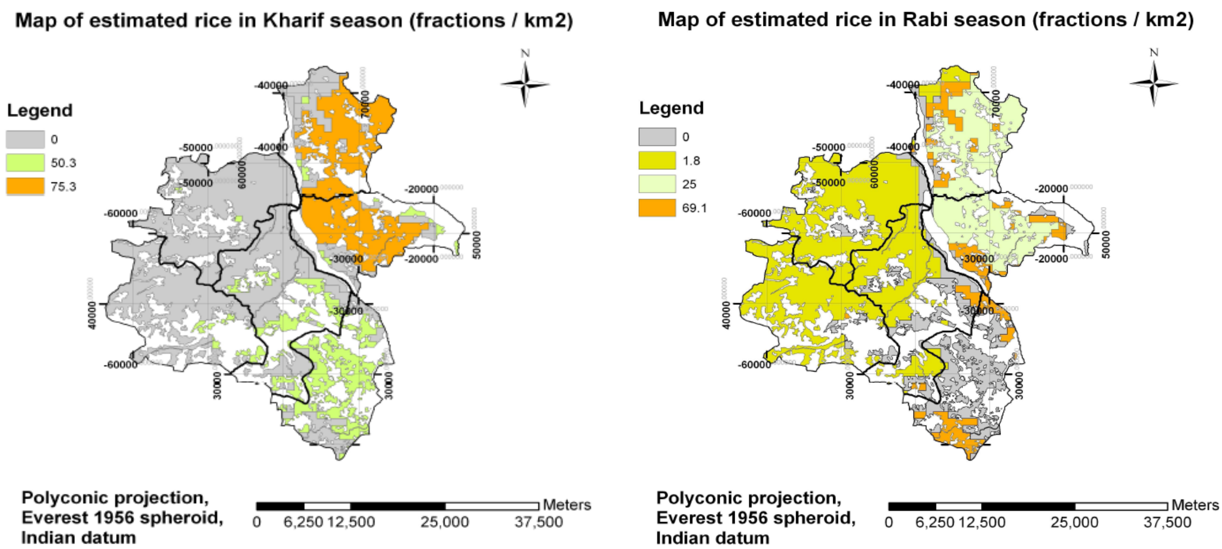

Figure 5. Estimated maps for rice grown in Kharif (left) and Rabi (right) seasons.

Conclusions

This research work illustrated the benefit of integrating hypertemporal remote sensing data with crop statistics to delineate NDVI profile clusters with their corresponding agricultural land cover map units and to link these statistics to geographical locations (mainly administrative units). These map units can be used as a reference for future monitoring of natural resources, in particular crop growth and development and consequently for forecasting crop production and/or yield and stresses like drought.

References

Ali A, De Bie CAJM, Skidmore AK, Scarrott RG, Hamad A, Venus V, and Lymberakis P. (2013). Mapping land cover gradients through analysis of hyper-temporal NDVI imagery. *International Journal of Applied Earth Observation and Geoinformation*, 23: 301–312

Ali A, De Bie CAJM, Skidmore AK, Scarrott RG, and Lymberakis P. (2014). Mapping the heterogeneity of natural and semi-natural landscapes. *International Journal of Applied Earth Observation and Geoinformation*, 26: 176–183

Allen JD. (1990). A look at the remote sensing applications program of the National Agricultural Statistics Service. *Journal of Official Statistics*, 6: 393-409

Atkinson PM, Jeganathan C, Dash J, and Atzberger C. (2012). Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of the Environment*, 123: 400–417

Balaghi R, Tychon B, Eerens H, and Jlibene M. (2008). Empirical regression models using NDVI, rainfall and temperature data for early prediction of wheat grain yields in Morocco. *International Journal of Applied Earth Observation*, 10: 438-452

Campbell NA, De Boer ES, and Hick PT. (1987). Some observations on crop profile modeling. *International Journal of Remote Sensing*, 8: 193-201

Carfagna E and Gallego FJ. (2005). Using remote sensing for agricultural statistics. *International Statistical Review*, 73: 389–404

Cihlar J. (2000). Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing*, 21: 1093 – 1114

CPO. (2001). Handbook of statistics – Nizamabad District, Chief Planning Officer: Nizamabad, Andhra Pradesh, India

Das DK, Mishra KK, and Kalra N. (1993). Assessing growth and yield of wheat using remotely sensed canopy temperature and spectral indices. *International Journal of Remote Sensing*, 14: 3081– 3092

ERDAS (2003). ERDAS field guide. 7th Ed. GIS and Mapping, LLC, Atlanta, Georgia and Leica Geosystems GIS and Mapping, LLC.

ESRI. (2009). ArcGIS: ArcMap 9.2. ESRI: Redlands, California, USA.

Falkner E and Dennis M. (2002). Aerial mapping: methods and applications. 2nd Edition, CRC Press: Boca Raton, Florida, USA

Ferencz C, Bognar P, Lichtenberger J, Hamar D, Tarcsai G, and Timar G. (2004). Crop yield estimation by satellite remote sensing. *International Journal of Remote Sensing*, 20: 4113–4149

Giri CP. (2012). Remote sensing of land use and land cover: Principles and applications. CRC Press: Boca Raton, Florida, USA

Gonzalez-Alonso F, Cuevas JM, Arbiol R, and Baulies X. (1997). Remote sensing and agricultural statistics: Crop area estimation in north-eastern Spain through diachronic Landsat TM and ground sample data. *International Journal of Remote Sensing*, 18: 467-470

Howard DM, Whlie BK, and Tieszen LL. (2012). Crop classification modeling using remote sensing and environmental data in the Greater Platte River Basin, USA. *International Journal of Remote Sensing*, 33: 6094-6108

Jiang Y, De Bie CAJM, Wang T, Skidmore AK, Liu X, Song S, and Shao X. (2013). Hyper-temporal remote sensing helps in relating epiphyllous liverworts and evergreen forests. *Journal of Vegetation Science* 24 (2013) 214–226

Kastens JH, Kastens TL, Kastens DLA, Price KP, Martinko EA, and Lee RY. (2005). Image masking for crop yield forecasting using AVHRR NDVI time series imagery. *Remote Sensing of Environment*, 99: 341 – 356

Khan MR, De Bie CAJM, Van Keulen H, Smaling EMA, and Real R. (2010). Disaggregating and mapping crop statistics using hypertemporal remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 12: 36–46

Knight JF, Lunetta RS, Ediriwickrema J, and Khorram S. (2006) Regional scale land cover characterization using MODIS-NDVI 250m multi-temporal imagery: a phenology-based approach. *GIScience and Remote Sensing*, 43: 1–23

Labus MP, Nielsen GA, Lawrence LR, Engel R, and Long DS. (2002). Wheat yield estimates using multi-temporal NDVI satellite imagery. *International Journal of Remote Sensing*, 23: 4169-4180

Landgrebe DA. (2003). Signal theory methods in multispectral remote sensing. Wiley: Hoboken, New Jersey, USA

Lillesand T, Kiefer RW, and Chipman J. (2007). Remote sensing and image interpretation. 6th Ed. Wiley: New York, USA

Maselli F and Rembold F. (2001). Analysis of GAC NDVI data for cropland identification and yield forecasting in Mediterranean African countries. *Photogrammetric Engineering and Remote Sensing*, 67: 593– 602

Mather PM and Koch M. (2011). Computer processing of remotely-sensed images: an introduction. 4th ed. Wiley-Blackwell: Chichester, UK

Mkhabela MS, Mkhabela MS, and Mashinini NN. (2005). Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data derived from NOAA's-AVHRR. *Agricultural and Forest Meteorology*, 129: 1–9

Murakami T, Ogawa S, Ishitsuka N, Kumagai K, and Saito G. (2001). Crop discrimination with multi-temporal SPOT/HRV data in the Saga Plains, Japan. *International Journal of Remote sensing*, 22: 1335-1348

Neter J, Kutner MH, Nachtsheim CJ, and Wasserman W. (1996). Applied linear statistical models. 4th ed. McGraw-Hill: Boston, Massachusetts, USA.

Nguyen TTH, De Bie CAJM, Ali A, Smaling MA, and Chu TH. (2012). Mapping the irrigated rice cropping patterns of the Mekong delta, Vietnam, through hyper-temporal SPOT NDVI image analysis. *International Journal of Remote sensing*, 33: 415-434

Nordberg ML and Evertson J. (2003). Vegetation index differencing and linear regression for change detection in a Swedish mountain range using Landsat TM and ETM+ imagery. *Land Degradation and Development*, 16:139–149

Philipson W. (1997). Manual of photographic interpretation. 2nd Ed. Bethesda. MD: American Society for Photogrammetry and Remote Sensing

Potgieter AB, Lawson K, and Huete AR. (2013). Determining crop acreage estimates for specific winter crops using shape attributes from sequential MODIS imagery. *International Journal of Applied Earth Observation and Geoinformation*, 23: 254–263

Rouse JW, Haas RH, Schell JA, and Deering DW. (1973). Monitoring vegetation systems in the Great Plains with ERTS. Third Earth Resources Technology Satelite-1 Symposium (pp. 309-317). Washington, D.C.

Sakamoto T, Yokozawa M, Toritani H, Shibayama M, Ishitsuka N, and Ohno H. (2005). A crop phenology detection method using time-series MODIS data. *Remote Sensing of the Environment*, 96: 366–374

Sarkar S and Kafatos M. (2004). Inter-annual variability of vegetation over India sub-continent and its relation to the different meteorological parameters. *Remote sensing of the Environment*, 90: 268-280.

Sheub SS and Atkins PJ. (1991). Crop area estimation: a comparison of remote sensing and census methods. *Geography*, 76: 235-239

SPSS. (2008). SPSS Statistics for Windows, Version 17.0. SPSS: Chicago, Illinois, USA.

Swain PH and Davis SM. (1978). Remote sensing: the quantitative approach. McGraw-Hill: New York, USA

Tarpley J, Schneider S, and Money R. (1984). Global vegetation indices from NOAA-7 meteorological satellite. *Journal of Climate and Applied Meteorology*, 23: 491–494

Taylor JC, Brewer TR, and Bird AC. (2000). Monitoring landscape changes in the National Parks of England and Wales using aerial photo-interpretation and GIS. *International Journal of Remote Sensing*, 21: 2737-2752

Tucker CJ. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of the Environment*, 8: 127-150

USDA. (2014). Crop production 2013 summary. United States Department of Agriculture: National Agricultural Statistics Service

White MA, De Beurs KM, Didan K, Inouye DW, Richardson AD, Jensen OP, O'Keefe J, Zhang G, Nemani RR, and Van Leeuwen WJD. (2009). Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006. *Global Change Biology*, 15: 2335–2359

Xie Y, Sha Z, and Yu M. (2008). Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*, 1: 9-23

You X, Meng J, Zhang M, and Dong T. (2013). Remote sensing based detection of crop phenology for agricultural zones in China using a new threshold method. *Remote Sensing*, 5 : 3190-3211