**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. **ahmed\_douaik@yahoo.com**

1. **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 use and land cover should be spatially mapped and temporally monitored. As the areas to be mapped are very large, conventional methods are laborious and expensive. Alternatively, aerial photographs and satellite remote sensing tools can be beneficially used for land use and land cover mapping and monitoring. In this way, the normalized differential vegetation index (NDVI) is the most used among the vegetation indices for studying vegetation, and specifically crop, phenologies (Sarkar and Kafatos, 2004). Time series of NDVI were used to discriminate between vegetation and other land uses and, for vegetation, between different green areas specific to crops, forests, etc (Murakami and al, 2001). For agricultural areas, crop statistics (mainly cropped areas and production) are estimated yearly at regional level following given administrative units. 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. The main objective of this research work was to fill the gap by adding the spatial information to the crop statistics by analyzing hypertemporal remote sensing, i.e., temporal NDVI profiles using different statistical methods.

1. **Material and methods**
   1. ***Study area***

The study area is situated in the western part of Nizamabad district, Hyderabad province, Andra Pradesh state, in central India. Six Mandals or sub-districts are concerned by this study covering 1300 km2 from which 90000 Ha are agricultural lands and 18000 Ha are shrub and non cultivated area.

* 1. ***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 km2 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.

* 1. ***Methods***

The normalized differential vegetation index is defined by:

*NDVI = (IR- R) / (IR + R) 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 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; the ‘best’ number of clusters is the one corresponding to the run having the highest minimum and/or average divergence. 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.

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:

 *Eq (3)*

with *CA* representing cropped area (Ha) by Mandal and *NDVIclusteri* representing the area (Ha) of the *i*th NDVI profile cluster.

No constant was considered in the regression and the coefficients *ci* 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).

1. **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.

*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** |  |

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 2. Adjusted R2 and coefficients (%) for stepwise multiple linear regression with total areas (Ha) for main crops in both seasons.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **NDVI units** | | | |  |
| *Kharif* | **Adjusted R2** | **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.

 

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

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.

1. **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.

1. **References**

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.

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

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(7): 1335-1348.

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

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.

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