International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:9, No:5, 2015

Using Time-Series NDVI to Model Land Cover Change: A Case Study in the Berg River Catchment Area, Western Cape, South Africa

A. S. Adesuyi, Z. Munch

***Abstract*—**This study investigates the use of a time-series of MODIS NDVI data to identify agricultural land cover change on an annual time step (2007 - 2012) and characterize the trend. Following an ISODATA classification of the MODIS imagery to selectively mask areas not agriculture or semi-natural, NDVI signatures were created to identify areas cereals and vineyards with the aid of ancillary, pictometry and field sample data for 2010. The NDVI signature curve and training samples were used to create a decision tree model in WEKA 3.6.9 using decision tree classifier (J48) algorithm; Model 1 including ISODATA classification and Model 2 not. These two models were then used to classify all data for the study area for 2010, producing land cover maps with classification accuracies of 77% and 80% for Model 1 and 2 respectively. Model 2 was subsequently used to create land cover classification and change detection maps for all other years. Subtle changes and areas of consistency (unchanged) were observed in the agricultural classes and crop practices. Over the years as predicted by the land cover classification. Forty one percent of the catchment comprised of cereals with 35% possibly following a crop rotation system. Vineyards largely remained constant with only one percent conversion to vineyard from other land cover classes.

***Keywords*—**Change detection, Land cover, NDVI, time-series.

I. Introduction

AWARENESS of spatial land cover information is essential for planning, management and monitoring of natural resources and can effectively be done using remote sensing technology [1]. Over the years remotely sensed data from traditional sources such as LANDSAT (TM+ETM), Advanced Very High Resolution Radiometer (AVHRR) and Satellite Pour l'Observation de la Terre (SPOT) has proven functional for land use and land cover (LULC) classification because of their synoptic and continuous coverage [2]. Significant progress has been attained at classifying LULC at different multispectral, medium resolution and coarse resolution data [2], [6], [9].

The use of satellite based remotely sensed data, has been widely applied to provide a cost effective means to model land use and land cover (LULC) changes over a large geographic area [3]. In the Berg River catchment, Western Cape, South Africa, a recent study [4] showed a reduction of 5% in agricultural land use between 1986 and 2007, approximately 131km^2 of food production area. In order to understand this decline (conversion to other land use) there is need to understand the current trend in agricultural land cover dynamics in the area as this can also raise the risk of erosion and flooding thereby further affecting agricultural practices.

Covering an area of approximately 9000km^2, the Berg River catchment is the largest in Western Cape Province of South Africa with landuse mostly agriculture: wheat, fruit and vineyard. It is topographically variable with mean annual temperature of between 16^{0}C and 18^{0}C, and mean annual rainfall between 300mm in the lower catchment at the coast and 1500mm at the upper mountainous catchment. This paper investigates the use of time series Moderate Resolution Imaging Spectro-radiometer (MODIS) Normalized Difference Vegetation Index (NDVI), to identify agricultural land cover change (2007 – 2012) and characterize the trend.

The paper is organized as follows. Section II describes the use of MODIS NDVI for vegetation studies. Section III presents the image pre-processing. The image analysis which involves creating NDVI signatures for training data and the different classifications are described in Section IV. Section V presents results obtained. Finally the conclusions are drawn in Section VI.

II. Modis NDVI for Vegetational Studies

MODIS provides coverage of science quality data with high temporal resolution of 1-2days and intermediate spatial resolution of 250m [5], which is well suited for crop mapping and monitoring of the study area. The MODIS NDVI product (MOD13) includes a time-series of visible red (VR) (620-670nm), near infra-red (841-876nm), normalized difference vegetation index (NDVI), enhanced vegetation index (EVI) composited at 16 day interval [6], [10]. This affords the opportunity for comprehensive broad area land use rendition. From MODIS NDVI filtered NDVI temporal profiles can be constructed to trace vegetation phenology on a time step to enhance the improvement of regional scale landscape process models [7]. Both NDVI and EVI can be accurately used for land cover mapping and detecting phenological change as they complement each other [8].

Image pre-processing generally precedes the main analysis and is intended to correct for sensor and platform-specific radiometric distortions, atmospheric effects and geometric distortions of data which in turn improves the quality of the image [11]. However, MODIS products are already corrected using the bidirectional reflectance distribution function

A. S. Adesuyi is a Masters student in Geoinformatics with Stellenbosch University. Stellenbosch 7600. South Africa (e-mail: 17180260@sun.ac.za).
Z. Munch, is a Senior lecturer with Stellenbosch University, Stellenbosch 7600. South Africa (e-mail: zmunch@sun.ac.za).

ISSN: 2517-942X
Vol:9, No:5, 2015

(BRDF). Since MODIS data possesses sub-pixel geo-locational accuracy of ±50m at nadir, geometry inaccuracies will not largely affect the vegetation index (VI) changes between observations in a time-series [6], [8].

In the following section, pre-processing of the MODIS satellite imagery is described.

III. Image Pre-Processing

Since MODIS products are already radiometrically and atmospherically corrected, but provided in Hierarchical Data Format (HDF), the first pre-processing step was the conversion of the images from the HDF file format to GeoTIFF format using the MODIS Resampling Tool (MRT).

The MRT tool was used to select the bands and subset the image to the extent of the study area. The Nearest Neighbour resampling method was chosen with an output cell size of 250m, giving the output image the desired projection (Transverse Mercator, Central Meridian 19 with Datum: WGS84) to suit the study area. A script was used to process the 138 input images (23 images per year for 6 years) in an automated fashion.

IV. Image Analysis

The image analysis comprised unsupervised classification, creation of NDVI signatures, decision tree model training, decision tree classification, post-processing and change detection.

A. Unsupervised Classification

Unsupervised classification was carried out in ENVI 4.8 while training samples were created in ArcMap 10. The 23 MODIS NDVI data sets for each year were merged into a single image file per year using layer stacking and masked to the study area extent Output for each year was saved as a Geotiff file, for accessibility in ArcMap.

ISODATA image classification was carried out on the masked MODIS imagery for year 2010 only in ENVI [2], since this correspond with the field data collected [4]. The land cover classification system (LCCS) used, conforms to the LCCS described by Chief Directorate: National Geo-spatial Information's (CD: NGI) given in Table [12]. Training samples for each class were selected based on field data gathered.

TABLE I
Land Cover Classes and Representation

Land cover classes	Representation
Agriculture	Crops, wheat, Vineyard, orchard
Bare Soil	Grounds left to fallow
Semi Natural	Weeds, grasses, shrubs, fynbos
Trees	Eucalyptus, Gum Tree, Acacia
Impervious surface	Houses, Industrial Area
Erosion Scar	Areas of Gully Erosion
Plantation	Pine Plantation,
Water	Dams, Rivers

The classification outputs were tested against the training samples to validate the class with the highest pixel seperability. Visual inspection is not enough to conclude that a classification output is acceptable, thus class validation was carried out by generating ROIs for each class in order to create a statistical table for the classification.

From the statistics derived ISODATA 40 classes was chosen (85% classification accuracy) to use in further processing. To ensure that areas of mixed pixels between different vegetation types are accounted for, additional data filtering was carried out. For this study, urban areas were excluded and only three (3) class groups namely: agriculture, semi-natural and agriculture/semi-natural were considered for further processing. The next section explains the method to create NDVI signatures for the time-series.

B. Creating NDVI Signatures for the Selected Classes

In order to differentiate between small grain (cereals), vineyard and other agriculture, the reference data and additional field sample data were classified as *Cereals*, *Vineyard* and *Others*, validated using pictometry and used to create the NDVI signatures from the annual time-series images.

Using the selected points, the NDVI data for each of the 23 time steps per year for period 2007 to 2012 were extracted. After removing poor quality data (cloud cover, water), a median value for each of the 23 periods for each year was calculated for classes *Cereal* and *Vineyard*. The median values calculated per year were plotted on a graph to identify the unique NDVI signature curve for *Cereals* (Fig. 1) and *Vineyard* (Fig. 2) with NDVI values plotted on the vertical axis against the day of acquisition of the imagery on the horizontal axis. The graphs also indicated the 5th & 95th percentile identifying the VI tail threshold.

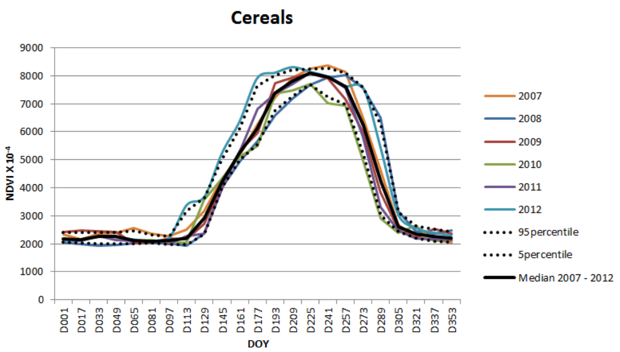

Fig. 1 NDVI Signature Curve for class *Cereal* (2007 – 2012).

International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:9, No:5, 2015

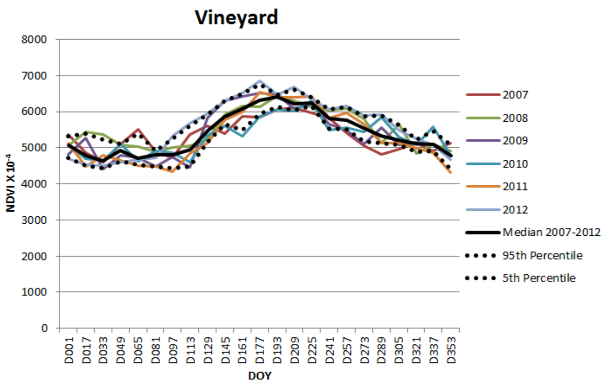

Fig. 2 NDVI Signature Curve for class *Vineyard* (2007 – 2012).

C. *WEKA Decision Tree Model Training*

The NDVI signature curves were used to create a decision tree model in WEKA 3.6.9. Two decision tree models were tested: in Model 1, the output class from the ISODATA classification was added to the NDVI signature of the training data as an additional input to creating the rule-sets, while in Model 2, the ISODATA class value was ignored. Training data with negative NDVI values or null class values were deleted to arrive at the final 690 samples which were used as training data in WEKA.

The training samples were processed in WEKA to generate a decision tree classifier using the J48 algorithm for classification [2], [3]. After extensive out-of-sampling testing, a Certainty Factor value of 0.25%, a minimum leaf size of 2, and number of folds (allowable sub-trees) of 3 were selected. The Certainty Factor (CF) value enables error-based pruning of the decision tree and helps to decrease over fitting thereby increasing the tree's predictability when applied to unseen data [2], [3], [8].

Two separate decision tree classification models were generated from the training samples using training sets as the test option after which an accuracy assessment was derived for both models. The decision trees for Model 1 and Model 2 can be seen in Figs. 3 and 4 respectively. Each point of the 690 samples was used in the training set.

These two decision tree classification models were then applied to the 2010 stacked data of the study area using the simple command line (SCL) interface within WEKA. The SCL is a programming interface, which allows for java commands to directly interface with the J48 decision tree classifier module within WEKA. These commands produced a prediction for each cell in the study area using the trained decision tree models.

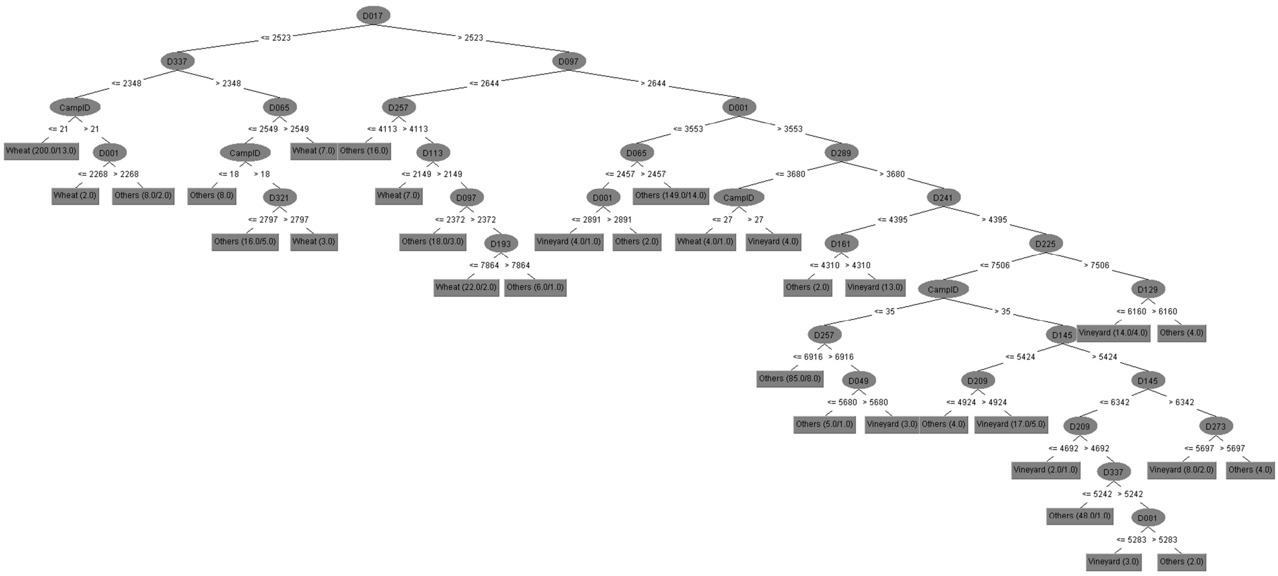

Fig. 3 Pruned Decision Tree for Model 1

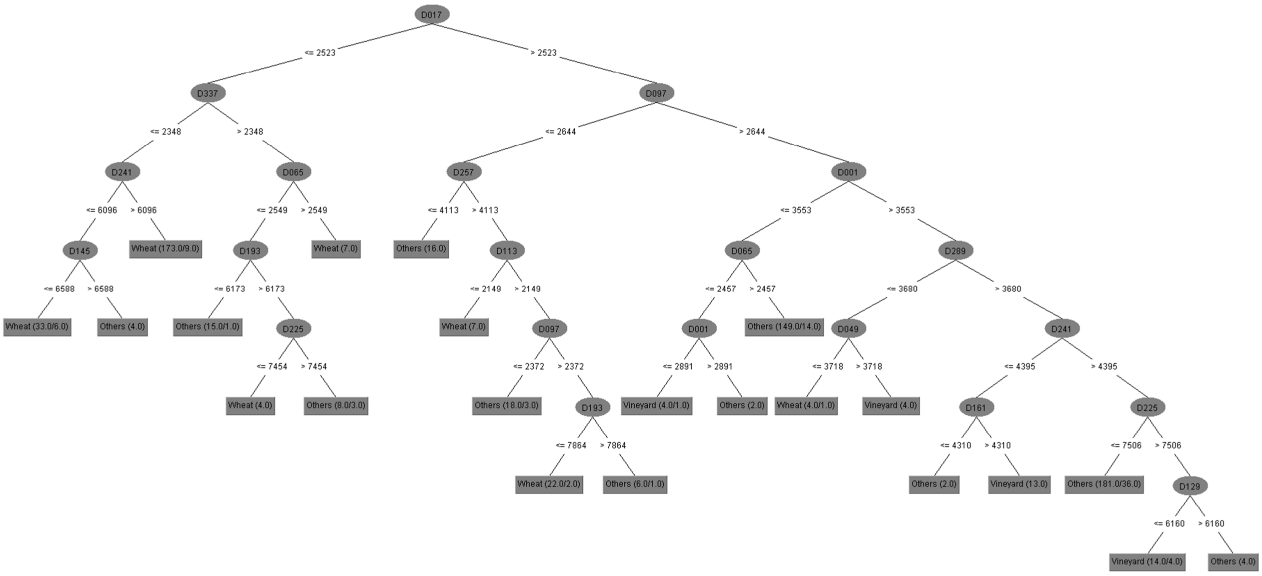

Fig. 4 Pruned Decision Tree for Model 2

D. WEKA Classification Post-Processing

The WEKA classification output for each of the two models for 2010 was transformed back into GEOTIFF format via Excel and ArcMap. These data were used in the accuracy assessment. WEKA predictions for the other years (2007 – 2009, 2011-2012) were carried out using only Model 2, due to the fact that ISODATA classification was not performed on these years due to absence of reference data. The WEKA classification output was transformed to GEOTIFF format for change analysis purposes.

V. Implementation and Results

When training the decision tree classifier in WEKA, Model 1 which included the ISODATA classification output gave a better result with overall accuracy of 90.7% and a kappa of 0.8373, whereas Model 2 only had an overall accuracy of 88.3% (kappa of 0.7869). Since Model 1 could not be replicated for each of the other years, due to lack of reference data and time required for each ISODATA classification, Model 2 was selected with the aim of predicting the other years using the training samples generated from the median of all years' data.

Both the decision tree classification models generated (Model 1 and Model 2), were used to create a land cover map based on the three vegetative classes under agriculture for 2010. The *Vineyard* NDVI signature (Fig. 2) is constantly high (NDVI~0.5) throughout the year. This can be attributed to growth during the summer months, often with irrigation applied, and grass or cover crops appearing in the vineyard during the winter rainy season even though leaves go brown (autumn) or are lost (winter). In contrast the NDVI signature for *Cereals* shows the pronounced growing season in winter (NDVI>0.7), with nearly bare soil (NDVI ~0.2) in the dry summer months.

The results from the accuracy assessment of the two land cover maps generated for 2010, carried out on a hundred random points, are noted in Tables II and III. The Model 1 land cover data set has an overall accuracy of 77% (kappa 0.55) (Table II). Even though Model 1 gave a higher accuracy when training data the classifier, it was noted that Model 1 did not have a higher accuracy than Model 2 in the final classification.

TABLE II
Confusion Matrix for Model 1 Classification

Reference data	Map Data			
	O	V	C	Total
Others (O)	48	2	12	62
Vineyard (V)	4	1	0	5
Cereal (C)	5	0	28	33
Total	57	3	40	100
%Error of Omission	22.58	80.00	15.51	
%Error of Commission	15.79	66.67	30.00	
Producer Accuracy	77.42	20.00	84.85	
User Accuracy	84.21	33.33	70.00	
Overall Accuracy	77%			
Kappa	0.55			

TABLE III
Confusion Matrix for Model 2 Classification

Reference data	Map Data			
	O	V	C	Total
Others (O)	49	2	10	61
Vineyard (V)	1	1	0	2
Cereal (C)	7	0	30	37
Total	57	3	40	100
%Error of Omission	19.67	50.00	18.92	
%Error of Commission	14.04	66.67	25.05	
Producer Accuracy	80.33	50.00	33.33	
User Accuracy	85.96	33.33	75.00	
Overall Accuracy	80%			
Kappa	0.6			

International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:9, No:5, 2015

Model 2 has a higher accuracy of 80% (Table III) with kappa of 0.6. This outcome supported the decision to use Model 2 only for classifying data for the other years.

However in both models, the *Vineyard* class was not well represented, having a percentage error of omission of 80% in Model 1 and 50% in Model 2. This could be attributed to the problem of mixed pixel or misclassification of the *Vineyard* class as many vineyards are smaller than the MODIS pixel resolution of 6.25 hectares.

Model 1 and 2 outputs were also compared to each other to determine the transferability of the selected model to the other years. The overall accuracy of Model 1 compared to Model 2 was 91% and the difference in classification is reflected in Table IV. Noticeable in the comparison (Table IV) was the high number of pixels (7835) classified as *Others* in Model 2 but *Vineyard* in Model 1, which explains the low representation in the *Vineyard* class.

TABLE IV
COMPARISON FOR MODEL 1 AND 2

		Model 2			
		Others	**Vineyard**	**Cereal**	**Row Total**
Model 1	**Others**	75028	0	2039	77067
	Vineyard	7835	5761	450	14046
	Cereal	1764	161	43580	45505
	Column Total	84627	5922	46069	136618

Using the Model 2 decision tree, data for all other years (2007-2009, 2011, 2012) were classified and compared. Table V summarizes these changes.

TABLE V
SUMMARY OF DOMINANT LAND COVER CLASSES 2007 - 20012

Predominant class (2007-2012)	Pixels	Area (km^2)	Percentage of catchment
Classification error	972	61	<1
Others (all years)	55967	3498	41
Others with one / two years misclassified as vineyard in any of the years	11129	696	8
Possible conversion to vineyard from Others	1629	102	1
Vineyard (all years)	2207	138	2
Cereal (all years)	16724	1045	12
Cereal-fallow agricultural practice	47990	2999	35

Forty seven percent of the catchment comprises of class *Cereals* with 35% following a crop rotation system where fields are left to lie fallow for soil to recover, or an alternative crop, classified as *Others* in this study, is planted. *Vineyards* which take a longer time to establish, remained constant over the period 2007 - 2012, however, some conversion to class *Vineyard* (1%) took place as this is regarded as a higher income crop. Fig. 5 shows the year-on-year variations in the agricultural classes (2007 – 2012) identified through the WEKA model.

While some of the changes might be as a result of misclassification as indicated in Table V, much of the study area has remained the same, with agricultural land use dominating the landscape. The changes might also be as a result of the crop rotation system and new crop being introduced, such as vegetables which was not targeted in this study.

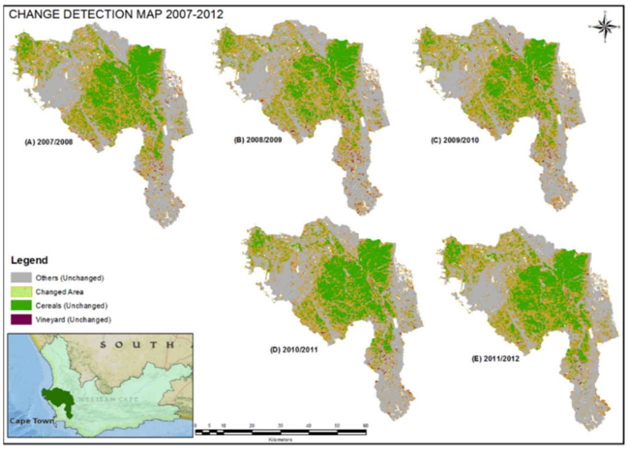

Fig. 5 Change detection between years (2007 - 2012)

In this year-on-year comparison of changes from *Cereal* to *Other* and *Other* to *Cereal*, it was found that about 10% of the catchment was subject to land use change associated with different crop practices.

VI. CONCLUSION

Through a combination of unsupervised and decision tree classification generated by machine learning tool WEKA, land cover maps were extracted for six consecutive years (2007 – 2012) and agricultural land use investigated. Ancillary data and reference data aided in validating and enhancing the classification during and post classification. Change detection was carried out to evaluate the level of change in the study area within the time steps which was noticeable and attributed to the cropping practices carried out by farmers.

Contrary to [4] who found a decrease in agricultural land cover in the catchment in favor of natural vegetation over the period 1986 to 2007 using a single LANDSAT image per time-step, this study identified differences in land use practices, rather than conversion.

This underlines the strength of using multi-temporal imagery, such as the MODIS NDVI products for this type of analysis. Using only the agricultural classes *Cereal* and *Vineyard*, the change detection indicated that, farmers in the study area either tend to leave the land fallow for a given period or change the type of crop practice. However an increase in gully erosion (<1% of agricultural land) was detected, which are however being managed and channeled by the farmer to suit their farming practices.

Further studies should investigate the land use change for the catchment area using a greater variety of agricultural land cover classes, which could include orchards, pastures and vegetables. With the availability of LANDSAT 8 at higher temporal resolutions, the study could be repeated using higher spatial resolution data to resolve the mixed pixel challenge for small fields. The opportunity exists to automate and facilitate this process over the multiple software packages used for processing (ArcMap, ENVI and WEKA) which would allow more effective analysis. This tool could also help land managers predict and model the seasonal fluctuations in farming crop practices.

References

[1] D. Yacouba, H. Guangduo and W. Xingping, Assessment of land cover / land use changes using NDVI and DEM in Puer & Sima Counties Yuman Province China. *World Rural Observations* 1(2): 1-11, 2009.
[2] B. D. Wardlow and S. L. Egebert. Large area crop mapping using time-series MODIS 250m NDVI data: An assessment for the US Central Great Plains. *Remote Sensing of Environment* 112: 1096-1116, 2008.
[3] J. C. Brown et al., Classifying multilayer agricultural land use data from mato grosso using time series MODIS vegetation index data. *Remote Sensing of Environment* 130: 39-50, 2013.
[4] T. Stuckenberg, Land cover change in the Berg river catchment area, 2012.
[5] C. O. Justice and J. R. G. Townshend JRG, Special issue on Moderate Resolution Imaging Spectroradiometer (MODIS): A new generation of land surface monitoring. *Remote Sensing of Environment* 83:1-2, 2002.
[6] B. D. Wardlow, S. L. Egebert and J. H. Kastens JH, Analysis of time-series MODIS 250m vegetation index data for crop classification in the US Central Plains. *Remote Sensing of Environment* 108: 290-310, 2007.
[7] R. S. Lunneta et al., Land cover change detection using multi-temporal MODIS NDVI data. Remote Sensing of Environment 105:142-154, 2006.
[8] B. D. Wardlow and S. L. Egebert, A comparison of MODIS 250m EVI and NDVI data for crop mapping: Case study for south west Kansas. *International Journal of Remote Sensing* 31(3): 805-830, 2010.
[9] R. R. Colditz, Time series generation and classification of MODIS data for land cover Mapping, 2007.
[10] X. Zhang et al., Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment* 84: 471-475, 2003.
[11] J. B. Campbell and R. H. Wynne, Introduction to remote sensing 5th ed. The Guilford Press. New York, 2011.
[12] W. Lück,. *CSIR Satellite Application Centre: Earth Observation Service Centre Land cover class definition report*. Pretoria: CSIR, 2006.