Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 74(3) 2020 237–242

Integration of Survey Data and Satellite Data for Acreage Estimation of Mango (*Mangifera indica*)

Ashis Ranjan Udgata[1], Prachi Misra Sahoo[1], Tauqueer Ahmad[1], Anil Rai[1], Ankur Biswas[1] and Gopal Krishna[2]

[1]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
[2]*RMSI Pvt. Ltd., NOIDA*

Received 16 July 2020; Revised 04 November 2020; Accepted 05 November 2020

SUMMARY

Timely and accurate estimates of crop areas are critical for enhancing agriculture management and ensuring national food security. This study aims to combine remote-sensing data and survey data to improve the accuracy of crop area estimates and decrease the cost of crop surveys at district level. In this study an estimate of mango area was obtained using Sentinel 2 satellite images of 2017 for West Godavari district of Andhra Pradesh. Further the area estimate was improved by integrating satellite data and survey data using ratio and regression estimator. Regression estimator was found to be best with lowest standard error.

Keywords: Food security, Remote sensing, Satellite data, Ratio and regression estimator.

1. INTRODUCTION

Information on crop acreage is important to formulate food policies and economic planning (Gallego 2004). Timely and accurate knowledge of agricultural acreage plays a very important role in enhancing agriculture management and ensuring national food security. Traditionally complete enumeration technique (in which agricultural acreage is measured by plot by plot) has been used to crop acreage at regional scale (Qian, Yang, and Jiao 2007). However this method is time consuming and needs large manpower and money investment. This method is also influenced by non-sampling error. Non-sampling error is an error in sample estimates which cannot be attributed to sampling fluctuations. Non-sampling errors may arise from many different sources such as defects in the frame, faulty demarcation of sample units, defects in the selection of sample units, mistakes in the collection of data due to personal variations or misunderstanding or bias or negligence or dishonesty on the part of the investigator or of the interviewer, etc. As a consequence crop acreage eastimate is poor. With development of remote sensing technique, remotely sensed images have been frequently used to identify the crop types and to estimate crop acreage (Wardlow, *et al.* 2007; Mathur and foody 2008; Udgata *et al.* 2020).However, the estimation of crop acreages via images alone at a large regional scale still suffers from certain problems. The first of these is image availability. Often, not all remote-sensing images from a survey area can be obtained due to cloudy or foggy weather. The second issue is cost. The collection of images across a large area can be very expensive. The final complication is crop identification accuracy. Due to the existence of mixed pixels, the accuracy of crop acreage estimates by remote-sensing images is still poor. To overcome this remote sensing data can be used along with survey data in the production of agricultural statistics. It can be used at the design level as well as at the estimator level (Gallego *et al.* 1998; Eitel *et al.* 2011). At the estimator level, remote sensing data are generally used as auxiliary variables in the regression estimator (Bellow,

Corresponding author: Ashis Ranjan Udgata
E-mail address: ashisu93@gmail.com

1994).The regression estimator is the most widespread way of using remote sensing data as auxiliary variable to improve the precision of estimates at the estimator level (Bellow 1994).

The aim of the present study is to integrate survey data and satellite data for acreage estimation of mango. Mango area under selected villages are calculated from both survey data and satellite data. Further ratio and regression estimation method used for improvement of overall estimate of mango area.

2. MATERIAL AND METHODS

2.1 Study Area

West Godavari district of Andhra Pradesh is taken for acreage estimation of Mango in year 2017. It is spread over 16.706041° N to 17.119720°N and 80.991595°E to 81.741431ºE with an approximate area of 7,742 km^2.Administrative headquarter of West Godavari is situated at Eluru. It receives a rainfall of 785.6 mm in South-west Monsoon amounting to 68% and 254.6 mm in North East monsoon which amounts to 21% of the total rainfall. Mainly Alluvial, Black Regur and Red Ferruginous soils along with a small portion of sandy soil in coastal belts are found in this district. Paddy, Sugarcane, Banana, Mango, Cashewnut, Coconut and Cocoa are major crops grown in this district.

2.2 data

Experimental data consists of two parts *i.e.* Satellite data and Survey data.

Satellite data

Satellite data derived from Sentinel 2 image for land cover and land use map preparation. Sentinel-2 is the latest-generation Earth observation satellite of the European Space Agency (ESA) for land and coastal applications (Drusch *et al.* 2012). It was launched in June 2015 and is part of Europe's Copernicus program aiming at independent and continued global observation (Immitzer *et al.*). Sentinel-2 offers an increased spectral and spatial resolution with 13 spectral bands, from blue to SWIR (short wave infrared), including red-edge bands of 10-m to 60-m spatial resolution, which have already proved useful for forest stress monitoring, land use and land cover mapping and biophysical variable retrieval (Vuolo *et al.*).Three sentinel 2 images (Table 1) were acquired on 28th December 2015 alongwith another image was acquired on 28th November 2015 to cover the whole district. It also enables to differentiate mango crop from other crop because this period coincide with mango flowering period.

Table 1. Scene characteristics of Sentinel 2 data

Acquisition date	Attribute name	Cloud cover	Orbit no.	UTM zone
December 28, 2015	S2A_OPER_MSI_L1C_TL_SGS__20151228T051728_20151228T102532_A002690_T44QND_N02_01_01	0	119	44N
December 28, 2015	S2A_OPER_MSI_L1C_TL_SGS__20151228T051728_20151228T102532_A002690_T44QME_N02_01_01	3	119	44N
December 28, 2015	S2A_OPER_MSI_L1C_TL_SGS__20151228T051728_20151228T102532_A002690_T44QND_N02_01_01	0	119	44N
November 28, 2015	S2A_OPER_MSI_L1C_TL_EPA__20151128T051517_20170430T185220_A002261_T44QNE_N02_04_01	0.253	119	44N

Survey data

The sampling design adopted for the survey was stratified multistage random sampling. In this design the whole state was divided into two strata. The districts covering about 70-80% of the total area under fruits in the entire State were treated as stratum one and rest of the districts *i.e.* low productive districts formed another stratum. From stratum one, about 40% districts out of the total number of districts and two districts from stratum two had been selected by simple random sampling without replacement (SRSWOR).

Taluk/tehsil-wise area figures under fruits had been also used for sub-stratifying the taluks/tehsils of the high productive districts into two groups viz. high productive taluks/tehsils and low productive taluks/tehsils. High productive taluks/tehsils were those which constitute 60-70% of the total area under fruits of the district and rest of the taluks/tehsils fall under low productive taluks/tehsils. A sample of two taluks/tehsils had been selected by SRSWOR from both the groups after rejecting taluks/tehsils contributing less than 5% of total area under fruits of the district. From each of the four selected taluks/tehsils, a sample of twenty villages had been selected by SRSWOR. Therefore in all 80 villages were selected from each district. An orchard for selection process should have minimum of 12 fruit trees of bearing age of a single fruit crop.

A sample of five orchards was to be selected from each selected village by SRSWOR. Personal Digital Assistant (PDAs)/Handsets *i.e.* Tablets was used for collection/uploading of data in one district in proposed state under study.

In this study particularly mango area in the West Godavari district was taken into consideration. The selected villages in survey were completely enumerated. House to house survey was conducted and data was collected by enquiry from all mango orchards owners. From this collected data area under mango was obtained.

2.3 Estimation procedure

Estimation procedures for estimating area under mango at district level was done in a similar manner as in CHAMAN (Coordinated programme on Horticulture Assessment and Management using geoiNformatics) project and is explained below:

Notations:

L = Total number of strata=2

T_h = Total number of blocks in h-th stratum, h = 1.2

t_h = Number of blocks selected from h-th stratum

V_{ht} = Total number of villages in t-th selected blocks of h-th stratum, h = 1, 2 and t = 1, 2, . . .,t_h

v_{ht} = Number of selected villages from t-th selected block of h-th stratum

x = Character under study

$\hat{X}_{hf}$ = Area under f-th fruit in h-th stratum

An estimate of total for the character x *i.e.* area under a fruit in the district for h-th stratum for f-th fruit is given by

$$\hat{X}_{hf} = \frac{T_h}{t_h}\sum_{t=1}^{t_h}\frac{V_{ht}}{v_{ht}}\sum_{v=1}^{v_{ht}} x_{htvf} = T\sum_{h=1}^{L} W_h \bar{y}_{hf}$$

Pooling of strata estimates to get district level estimate:

The estimates of both the strata in a district are added to get the estimate at district level.

$$\hat{X}_f = \sum_{h=1}^{L}\hat{X}_{hf} = T\sum_{h=1}^{L} W_h \bar{x}_{hf}$$

where $T = \sum_{h=1}^{L} T_h$, $W_h = \frac{T_h}{T}$ and

$$\bar{x}_{hf} = \frac{1}{t_h}\sum_{t=1}^{t_h}\frac{V_{ht}}{v_{ht}}\sum_{v=1}^{v_{htf}} x_{htvf}$$

Estimate of variance

An estimate of variance of $\hat{X}_f$ is given by

$$\hat{V}\left(\hat{X}_f\right) = T^2\sum_{h=1}^{L} W_h^2\ \hat{V}\left(\bar{x}_{hf}\right)$$

where

$$\hat{V}\left(\bar{x}_{hf}\right) = \frac{1}{t_h}\left(1-\frac{t_h}{T_h}\right)\frac{1}{t_h-1}\sum_{t=1}^{t_h}\left(V_{ht}\ \bar{x}_{htf} - \bar{x}_{hf}\right)^2 + \frac{1}{T_h \times t_h}\sum_{t=1}^{t_h}\frac{1}{v_{ht}}\left(1-\frac{v_{ht}}{V_{ht}}\right)s_{tf}^2$$

$$\hat{V}\left(\bar{x}_{hf}\right) = \frac{1}{t_h}\left(1-\frac{t_h}{T_h}\right)\frac{1}{t_h-1}\sum_{t=1}^{t_h}\left(V_{ht}\ \bar{x}_{htf} - \bar{x}_{hf}\right)^2 + \frac{1}{T_h \times t_h}\sum_{t=1}^{t_h}\frac{1}{v_{ht}}\left(1-\frac{v_{ht}}{V_{ht}}\right)s_{tf}^2$$

$$s_{tf}^2 = \frac{1}{v_{ht}-1}\sum_{v=1}^{v_{htf}}\left(x_{htvf} - \bar{x}_{htf}\right)^2 \text{ and } \bar{x}_{htf} = \frac{1}{v_{ht}}\sum_{v=1}^{v_{htf}} x_{htvf}$$

Estimate of standard error of $\hat{X}_f$ is given by

$$\hat{S.E.}(\hat{X}_f) = \sqrt{\hat{V}(\hat{X}_f)}$$

Estimate of percentage standard error of $\hat{X}_f$ is given by

$$\%\hat{S.E.} = \frac{\hat{S.E.}(\hat{X}_f)}{\hat{X}_f}\times 100$$

2.4 Land use and land cover map preparation

Field surveys were conducted in November 2015 to December 2015 using hand held GPS device. GPS coordinates at each site were recorded and all types of land covers in the district were studied. A sufficient number of points has been collected from which seventy percent (70%) points were used for selecting training samples and remaining were used as validation of sample points. As per the demographic features and the characteristics of study area, as well as research objectives, a classification system with eight categories was designed. The eight land categories were mango orchards, water bodies, forest, built up, dense vegetation areas, wetland, and agricultural land. Agricultural lands are the areas excluding mango areas. The latitude and longitude of various classes like mango, water bodies, forest, built up, dense vegetation areas, wetland, agricultural land were collected during ground survey period. By considering the data on the above mentioned classes Region of Interest (ROI) was established by choosing one or more polygons for each class. land use and land cover map is prepared by using Support vector machine Radial Basis Function (SVMRBF), a supervised classification technique (Pal 2008). In supervised classification technique we have the prior information about the number of classes to be included in the land use and land cover map. For detail classification one can refer to Udgata *et al.* 2020.

2.5 Improved area estimator integrating survey data and satellite data

Estimated area from the satellite data further improved by integrating with survey data. Ratio estimator and regression estimator were used for improving the area using satellite data as auxiliary variable.

Ratio estimator

The notations used in the formulae for estimators, MSE, SEetc are given below:

Let

y_i = area under i[th] village obtained from survey data, i= 1,2,…,n

x_i = area under i[th] village obtained from satellite data, i= 1,2,…,n

$\sum_{i=1}^{N} X_i$ = total area under mango in the district obtained from satellite data

$\hat{Y}_R$ = Area estimate at district level using ratio estimator

$$R_n = \frac{\bar{y}_n}{\bar{x}_n}$$

MSE ($\hat{Y}_R$) = mean square error of total area estimate

SE ($\hat{Y}_R$) = standard error of total estimate

%SE = percentage standard error

n = number of village selected =40

N = total number of villages in West Godavari = 903

The improved estimator for obtaining total area under mango of area under mango in the district is given as:

$$\hat{Y}_R = N\,\bar{y}_R = N\,\frac{\bar{y}_n}{\bar{x}_n}\,\bar{X}_N = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}\sum_{i=1}^{N} X_i$$

The mean square error of the estimator is given as

$$\widehat{\text{MSE}}(\hat{Y}_R) = N^2\left(\frac{1}{n} - \frac{1}{N}\right)\left(s_y^2 + R_n^2 s_x^2 - 2R_n s_{xy}\right)$$

The standard error is given as

$$\text{SE}(\hat{Y}_R) = \sqrt{\text{MSE}(\hat{Y}_R)}$$

And the percent standard error is given as

$$\%\text{SE} = \frac{SE(\hat{Y}_R)}{\hat{Y}_R} \times 100$$

Where $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2$,

$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$ and

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)$$

Regression estimator

The symbols used for denoting estimation stage are as follows:

Let,

y_i = area under i[th] village obtained from survey data, i= 1,2,…,n

x_i = area under i[th] village obtained from satellite data, i= 1,2,…,n

$\sum_{i=1}^{N} X_i$ = total area under mango in the district obtained from satellite data

$\bar{X}$ = mean area obtained from satellite data

$\hat{Y}_{lr}$ = total area estimate using regression estimator

n = number of village selected =40

N = total number of villages in West Godavari=903

The improved estimator for obtaining total area under mango of area under mango in the district is given as:

$$\hat{Y}_{lr} = N\bar{y}_{lr} = N(\bar{y} + b(\bar{X} - \bar{x}))$$

where $b = \frac{s_{xy}}{s_x^2}$

$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$ and

$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)$

The mean square error of the estimator is given as

$$\text{MSE}\,(\hat{Y}_{lr}) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)(1-r^2)s_y^2$$

where $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y}_n)^2$

The standard error is given as

$$\text{SE}\,(\hat{Y}_{lr}) = \sqrt{\text{MSE}(\hat{Y}_{lr})}$$

And the percent standard error is given as

$$\%\text{SE} = \frac{SE(\hat{Y}_{lr})}{\hat{Y}_{lr}} \times 100$$

3. RESULT AND DISCUSSION

There are four stages of land-cover classification such as designing a proper classification system, collection of sufficient and well representative reference data, selection of suitable variables, and classification algorithm and evaluation the classification result (Lu and Weng 2007). In area of remote sensing our main focus should be on selection of classification system and classification algorithm. On the basis of landscape characteristics and availability of remote sensing data it is important to determine the variables that can be used in separation of different land cover types. Landcover classification accuracy cannot be improved by selection of improper combination of variables (Li *et al.*2012).

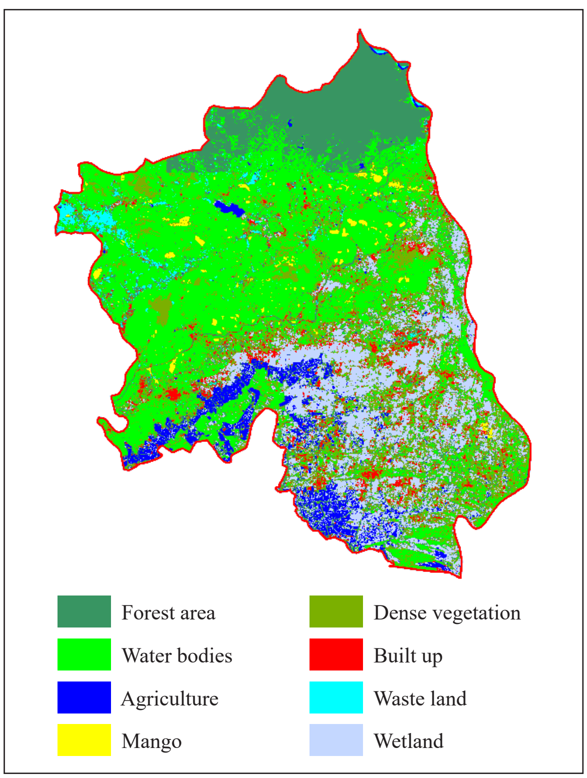

Fig. 1. Landuse and land cover map using SVMRBF

Another important feature in classification is time of data acquisition, *i.e.* vegetation types have different spectral signatures in different time periods.

Land use land cover map of West Godavari using SVMRBF is obtained using ROI (Fig 1). Once land use land cover classified image is obtained, area under any class can be obtained by pixels counting method. A vector layer of 40 village boundary map is prepared in ARC-GIS software. These vector layer was overlayed over the classified image to get the mango area from selected villages from satellite data (Fig 2).

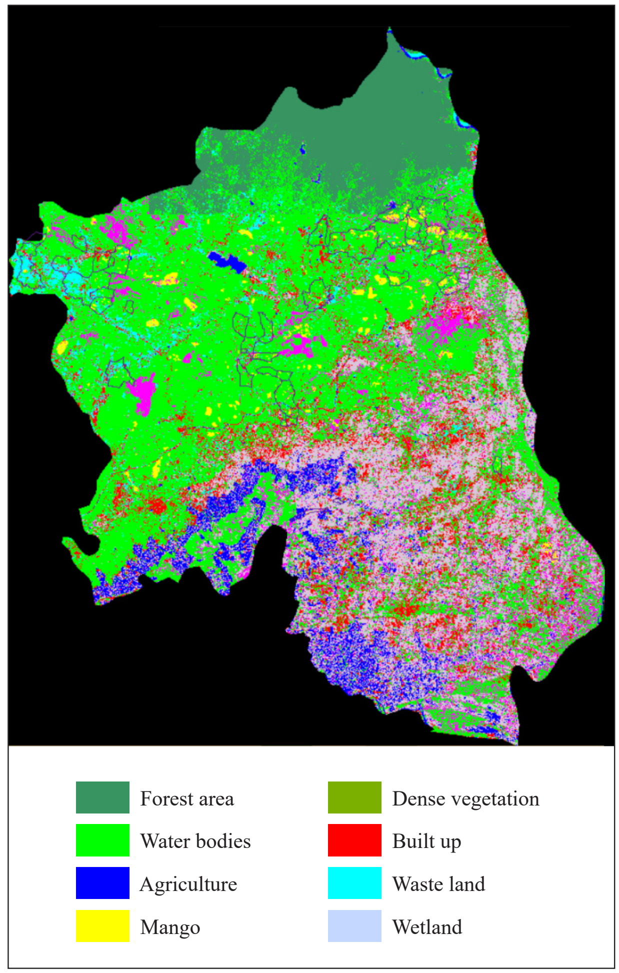

Fig. 2. Overlayed vector layer of selected village

The ratio and regression estimation methods were used as mentioned in the previous section. The sample mean and standard error of ratio and regression estimator was calculated using MS excel software. The results are summarized in table 2. The ratio and regression estimator are seem to be quite efficient as compared with the estimator based solely on survey data with less than 15% standard error at district level.

Table 2. Estimates of area under mango using various methods

	Estimate of area under mango (heactare)	**% S.E.**
Survey data	14258.15	18
Ratio estimator	13571.63	14
Regression Estimator	15215.32	11

As per the thumb rule the estimate with least percentage of standard error is considered to be the most efficient estimator. From Table 2 it is evident that the regression estimator gives an estimate of 15215.32ha area under mango with least *i.e.* 11% standard error as compared to ratio estimator with percentage standard error of 14%.

4. CONCLUSIONS

One way of integration of satellite data and survey data has been explored and used for mango acreage estimation at district level. From the results, it has been observed that the regression estimator is found to be most efficient with minimum percentage standard error followed by ratio estimator. In our study both regression and ratio estimator are quite reliable as they possess standard error below 15 percent at district level. The study therefore concludes that by integrating survey data and remote sensing data as auxiliary variable, improved estimates of area under mango can be obtained at district level.

REFERENCES

Bellow, M.E. (1994). Application of satellite data to crop area estimation at the county level. Research report no. STB-94-02, NASS, UDA. Washington D.C.

Carfagna, E. (1998). Area frame sample designs: a comparison with the MARS project. Proceedings ofAgricultural Statistics 2000. International Statistical Institut. Voorburg.

Drusch, M., Del B., U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P. and Martimort, P. (2012) Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing and Environment*, **120**, 25-36.

Eitel, J.U., Vierling, L.A., Litvak, M.E., Long, D.S., Schulthess, U., Ager, A.A., Krofcheck, D.J. andStoscheck, L.(2011) Broadband red-edge information from satellites improves early stress detection in a New Mexico coniferwoodland. *Remote Sensing and Environment*, **115**, 3640-3646.

Gallego, F. J. (2004) Remote Sensing and Land Cover Area Estimation. *International Journal of Remote Sensing,* **25(15)**, 3019-3047.

Immitzer, M., Vuolo, F. and Atzberger, C. (2016) First experience with Sentinel-2 data for crop and tree species classifications in Central Europe. *Remote Sensing*, **166(8)**, 1-27.

Li G, Lu D, Moran E and Sant'Anna S J S. (2012). A Comparative Analysis of ClassificationAlgorithms and Multiple Sensor Data for Land Use/Land Cover Classification in the Brazilian Amazon. *Journal of Applied Remote Sensing,* **6(1)**, 0617061-23.

Lu D and Weng Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing,* **28(5)**, 823-70.

Mathur, A., and G. M. Foody. (2008). Crop Classification by Support Vector Machine with Intelligently Selected Training Data for an Operational Application. *International Journal of Remote Sensing,* **29(8)**, 2227–2240.

Pal M. (2008)Ensemble of support vector machines for land cover classification. *International Journal of Remote Sensing,* **29(10)**, 3043-3049.

Qian, Y. L., B. J. Yang, and X. F. Jiao. (2007) Accuracy Assessment on the Crop Area Estimating Method Based on RS Sampling at National Scale: A Case Study of Northeast China's Rice Area Estimation Assessment. *Transactions of the Chinese Society of Agricultural Engineering,* **23(11)**, 180-187.

Schuster, C., Förster, M. and Kleinschmit, B. (2012) Testing the red edge channel for improving landuse classifications based on high-resolution multi-spectral satellite data. *International Journal of Remote Sensing,* **33(17)**, 5583-5599.

Taylor J., Sannier C., Delincé J, Gallego F.J. (1997). Regional Crop Inventories in Europe Assisted byRemote Sensing: 1988-1993. Synthesis Report. Office for Publications of the EC. Luxembourg.

Udgata, A.R., Sahoo, P. M., Ahmad, T., Rai. A. and Krishna, G. (2020) Remote sensing and machine learning techniques for acreage estimation of mango, *Indian Journal of Agricultural Sciences*, **90(3)**.

Vuolo, F.,Zółtak, M., Pipitone, C., Zappa, L., Wenng, H., Immitzer, M., Weiss, M., Baret, F. and Atzberger, C. (2016) Data service platform for Sentinel-2 surface reflectance and value- added products: System use and examples. *Remote Sensing,* **8(11)**, 938.

Wadlow, B.D., Egbert, S.L. and Kastens J. H. (2007) Analysis of time series MODIS 250 m Vegetation index data for crop classification in the U.S central great plains. *Remote sensing and environment,* **108(3)**, 290-310.