International Journal of Agricultural Research and Review: ISSN-2360-7971: Vol. 2(5): pp 084-091, May, 2014. Copyright © 2014 Spring Journals

Full Length Research Paper

Estimating Tobacco Crop Area and Yield in Zimbabwe Using Operational Remote Sensing and Statistical Techniques

¹Ezekia Svotwa, ²Anxious J. Masuka, ³Barbara Maasdorp, ⁴Amon Murwira, ⁵Mhosisi Masocha

¹Department of Crop Science and Postharvest Technology, Chinhoyi University of Technology, Box 7724, Chinhoyi, Zimbabwe ²Chief Operating Officer, Ariston Holdings, Msasa, Harare, Zimbabwe. ³Professor, Department of Crop Science ,University of Zimbabwe ⁴Professor, Department of Geography and Environmental Science, University of Zimbabwe ⁵Lecturer, Department of Geography and Environmental Science, University of Zimbabwe

Corresponding Author's E-mail: esvotwa2@gmail.com

Accepted 22nd May, 2014

In this study, remotely sensed data and field measurements were used to develop a simple but robust method for estimating the tobacco hectarage and yield in Zimbabwe. The current conventional tobacco yield forecasts rely on seed purchase records, land area record and visual assessment of the crop. This is costly, time consuming and unreliable. Between 2010 and 2013, starting from September, agricultural field boundaries from a pseudo natural colour composite Landsat Thematic Mapper (TM) satellite imagery were visually interpreted and digitized. Cloud free MODIS images covering the period September to end March were downloaded and georeferenced. For each MODIS image, NDVI was estimated. Mean temporal NDVI profiles for these crops using data from sampled tobacco fields were calculated separately and compared. The results of this study indicated that, based on MODIS NDVI data, the third to fourth week of November and the third to fourth week of February are the optimal times for discriminating the irrigated from the non-irrigated tobacco. The crop areas for the three seasons were estimated and yield estimates calculated from the long-term cropped yield- area regression model. The three seasons average yield estimates were 98.8% accurate, as compared to 112% for the traditional method.

Keywords: Remotely sensed data, MODIS images georeferenced, NDVI profile yield estimates,

INTRODUCTION

Background

Crop yield estimation is necessary, particularly in countries that depend on agriculture as their main source of economy (Atzberger, 2012). Such predictions warn the decision makers about potential reduction in crop yields and allow timely import and export decision (Awosola-Olubode et al., 2008; Bouman, 1992). Knowledge of crop area at an early stage is very important in agricultural planning and policy making at both national and regional levels (Casa and Ovando, 1996). Better estimates of crop yield are obtained if crop growth is monitored during the growing season (Clevers and Leeuwen, 1996).

Traditionally, quantitative estimates of crop condition in the field, cropped area, and yield are obtained from ground-based measurements (Awosola-Olubode et al., 2008). The current conventional tobacco yield forecasts in Zimbabwe rely on seed purchase records, land area and visual assessment of the crop. Since farmers' records may not be exhaustive the current forecast may not be accurate. These conventional methods are often complicated, costly, time consuming and they cannot be run in large scale (Gallo and Flesch, 1989). Therefore it is necessary to use cheaper and faster methods for crop



Figure 1: Some digitized fields (red boundaries) based on the interpretation of the pseudo natural colour composite of the Landsat TM image of May 2010. The combination of bands used to produce the colour composite was: band 5 (red), band 4 (green) and band 3 (blue).

yield estimation. Use of yield forecasting models can be employed to avoid these problems.

Remote sensing, which provides time series data and a synoptic view of the landscape, is now widely used to assess crop condition in the field as well as estimate crop yield. The positive relationship between remotely derived vegetation indices such as the normalized difference vegetation index (NDVI) and biomass (Jackson et al., 1983; Eitel et al., 2010; Yin et al 2010) has proved to be useful for predicting crop yield. Similarly, Gomes, (2006) proved that vegetative indices calculated from spectral data strongly relate to biomass and crop yield.

Remote sensing data has the potential and the capacity to provide spatial information at global scale; of features and phenomena on earth on an almost real-time basis (Green, and Invins, 1985; Hatfield and Pinter-Jr, 1993). With the use of remote sensing technology, plant physiological and morphological differences can be distinguished within fields in real time, cost efficiently and timely (Huete, et al., 2002). Remote sensing has facilitated a deeper understanding of the environment because it has many processes over a broad range of spatial and temporal scales, which can provide information on the actual status of agricultural crops (Kustas et al., 1994.). Best results are obtained by using remote sensing data in estimating biophysical values regularly during the growing season and subsequently calibrating the growth models based on these estimates (Maas and Dunlap, 1989; Moulin et al., 1998).

Multispectral sensors such as LANDSAT TM and MODIS are able to view more than one particular band of energy selected in various regions of the electromagnetic spectrum (Prasad et al., 2006). Presently there are several systems that can provide regular coverage of the Earth's surface. These can supply us regularly with remotely sensed data of the whole area of the country or selected regions of economic interest to farmers such as farm land (Prasad et al., 2006). With the help of these data, not only the spatial distribution of the crops can be determined but their status and vigour in the growing period can be monitored (Reynolds and Yittayew, 2000).

In Zimbabwe, timely and reliable estimates of potential

tobacco yield are important because tobacco has a huge economic impact on the country's economy. Thus, monitoring the growth and phenological development of the tobacco in the field is critical for obtaining early estimates of yield (Rizzi and Rudorff, 2005). However, the lack of an objective and robust method for estimating tobacco yield has often led to contradicting estimates being provided by different stakeholders. This compromises national planning and marketing of the crop. A more objective and simple method for yield estimation could assist tobacco stakeholders by providing precise data on tobacco growth characteristics, land area under tobacco and potential yield available for the export market.

In this study, remotely sensed (satellite) data and field measurements were used to develop a simple but robust method for estimating the hectarage under tobacco. Estimates are provided for both the irrigated and rainfed tobacco crops planted throughout Zimbabwe. Remote sensing was used because it provides observations over large areas at regular intervals. This reduces the costs for obtaining crop estimates. It is hypothesized that from the Global Positioning System coordinates of the sampled tobacco cropped fields of different planting dates crop growth can be monitored throughout the season. Seasonal crop profiles can also be developed from the NDVI values calculated from the satellite imagery. The derived signal can be used to develop and validate a remotely sensing based algorithm for operational estimation of tobacco cropped area in Zimbabwe. The computed area can be used as input data for tobacco yield estimation model developed from a long-term tobacco area-vield data.

MATERIALS AND METHODS

Determining the agricultural field and tobacco soil masks

In September and October 2010, we visually interpreted and digitized agricultural field boundaries from a pseudo natural colour composite (bands 5, 4 and 3) of Landsat Thematic Mapper (TM) satellite imagery (Figure 1). The



Figure 2: Distribution of sandy soils suitable for tobacco in Zimbabwe



Figure 3: Spatial distribution of the tobacco growing districts of Zimbabwe (TRB, 2012)

images covered the major tobacco growing regions in Mashonaland East, Mashonaland West, Mashonaland Central and Manicaland provinces of Zimbabwe. Landsat TM imagery was used to calculate the cultivation mask for two main reasons. Landsat images were be downloaded from the internet free of charge at www.glovis.usgs.gov. The high spatial resolution of 30 m made it possible to accurately digitize fields of different sizes as illustrated in Figure 1. The purpose of this activity was to obtain an accurate map of fields within which tobacco field area estimation would be undertaken.

The crop field mask produced using steps described above, included all crops, yet tobacco is grown on sandy soils. To ensure that estimates provided covered the targeted crop, tobacco, the FAO based soil map of Zimbabwe was digitized in a Geographic Information System (GIS), Integrated Land and Water Information System (ILWIS) 3.3, and sandy soils which are suitable for tobacco were extracted (Figure 2). This soil map was overlaid with the cultivated field mask to extract all fields that were on sandy soils, which are potential tobacco fields.

GPS measurements of tobacco fields

Between September and January of the 2010-2011, 2011-2012 and 2012 - 2013 seasons, fieldwork was undertaken to measure the location of tobacco fields and obtain data on the dates of tobacco planting. At least twelve (12) fields were sampled from each of the 15 traditional tobacco growing districts of Zimbabwe (Figure 3)



Figure 4: The spatial distribution of the tobacco fields sampled for yield forecasting

A total of 203 tobacco fields located in Mashonaland East, Mashonaland West, and Manicaland provinces of Zimbabwe, and their locations measured using global positioning system (GPS) unit. To minimize edge effects, GPS measurements were made at the centroids of tobacco fields. The average GPS error was 6 m. The sampled tobacco fields were classified based on the dates of planting into irrigated; planted in September and October, and rainfed; planted November and December. The spatial distribution of sample fields surveyed during fieldwork is shown in Figure 4.

Temporal MODIS NDVI data for tobacco fields

To monitor tobacco development in the field using remote sensing, cloud free MODIS images covering the period September to end of March were downloaded freely at http://modis-land.gsfc.nasa.gov/ and georeferenced. For each MODIS image, NDVI, which correlates positively with plant biomass was estimated as follows:

NDVI = $(\rho_{\text{NIR}} - \rho_{\text{RED}}) / (\rho_{\text{NIR}} + \rho_{\text{RED}})$

where: ρ_{NIR} is reflectance in the Near Infrared band (846–885 η m) and ρ_{RED} is the reflectance in the Red (600–680 η m) band. NDVI is widely used to accurately estimate vegetation greenness, i.e., the total concentration of chlorophyll in plants.

The mean temporal NDVI profiles for the September/ October planted crops and those for the November/ December were calculated separately using data from sampled tobacco fields and these were assessed in order to determine the optimum date for discriminating the irrigated and the rainfed crops.

Mapping irrigated and non-irrigated tobacco

Logistic regression was applied on the 29th November 2010 image to estimate the probability that different fields in the tobacco field mask was under an irrigated or a rainfed tobacco. The area estimates of the September-October and November-December crops were derived using a Geographical Information System (GIS) programme, Environment for Visualising Images (ENVI) 4.3, by extracting all pixels within the tobacco soil and cultivation field masks where the probability of belonging to either the September-October or November-December planting dates was equal to or greater than 95%.

Correcting for potential MODIS areal underestimates

The spatial resolution of MODIS imagery, which is 250 m, may lead to underestimation of the area under tobacco as fields which are smaller than 6.25 ha in area may be missed. To address this problem, 19 fields were selected and their area estimated using both MODIS and Landsat TM. The fields were visually interpreted and digitized them into a GIS a regression function relating Landsat TM area estimate with MODIS area estimate was fitted (Figure 5). Equation 1:

Y = 1.154x



Figure 5: Relationship between MODIS derived and Land sat TM-derived field area estimates.

 $R^2 = 0.894$

Where Y is Landsat TM area estimate (ha) Modis area estimate (ha)

The resultant regression function was then applied to the MODIS-derived tobacco area estimates for both irrigated and rainfed crops to correct for potential tobacco area underestimation.

A classification method for the September-October tobacco crop based on late November satellite data was used for area determination, taking advantage of the high contrast between irrigated tobacco and bare fields. Due to the spectral confusion associated with the November-December planted crop resulting from weeds and other crops, the longer MODIS time series (September to April) and a Maximum Entropy Method (Maxent) algorithm to map the November-December crop was also applied. The Maxent approach seeks to extract as much information from a measurement as is justified by the data's signal-to-noise ratio. The approach was applied to calculate the probability that the field adjacent to the sampled November-December, had a tobacco crop of the same planting time as this training sample. One major advantage of MAXENT is that it only requires positive data for training and reduces the labour and time involved in manually collecting training (Foody et al. 2006) Finally, the MODIS-Landsat regression relationship was applied to correct for the potential area underestimation by MODIS. Furthermore, high resolution Landsat images of October 2011 and November 2011 were acquired to further validate the September-October area estimates.

Crop Survey Estimates

The Department of Agriculture and Extension (Agritex) collects data on tobacco production activities at village level through the Ward based extension officers. The information is mainly on area planted/ week and general crop status. These statistics are sent to the National office through the Districts and the Provinces, and are compiles into weekly report. At the end of the month, a monthly report is sent out to all tobacco Industry stakeholders. In this study, the crop area estimates from

the Agritex were used, as standards against which the remote sensing estimates were compared.

Yield Estimates

data (T.I.M.B, 2012) was used to develop a cropped area-yield linear regression model. Y = 3.511X - 96929R2 = 0.9415

Where

Y is the crop yield in tons X is the estimated crop area

The yield estimates for the experimental period were then used as input data for estimating the yield for the 2010-2011 and the 2011-2012 seasons. The calculated estimates were compared with the actual volume of tobacco delivered at the Auction floors as reported by the Tobacco Industries Marketing Board (T.I.M.B, 2012; T.R.B, 2012).

RESULTS

The results for the 2012 - 2013 were not conclusive because of the prevailing weather conditions during the middle and latter parts of the two seasons, which prevented the complete collection of satellite data. There was so much spectral confusion between the November and the December planted crops and, as a result, the 2011 - 2012 data (Figure 6) was used in the description of the season NDVI profile for the sampled tobacco fields.

The NDVI generally increased from 0.2 - 0.25 early in the season to reach a peak of 0.6 - 0.7 at 8 -12 weeks after planting (Figure 6). The September and the October NDVI in the entire sampled crop field were statistically similar (p < 0.05). During the third week of December, the September crop NDVI was greater than that for October (p < 0.05). Both NDVI for the September and the October crops were statistically greater than those extracted from both the November and the Longterm (38 year) tobacco cropped area and yield



Figure 6: Temporal MODIS NDVI profiles from September 2011 to April 2012 grouped by planting month. Data are means while whiskers represent 95% confidence intervals calculated from more than 150 sampled tobacco fields per planting date. The NDVI profiles show spectral confusion with minimal chances of separability.

December fields. Around mid January the NDVI's for all the four crops were similar, and at the same time the NDVI for the September and the October crops were declining, while those for the November and the December crops were still on an increasing pathway. Between mid and early March, the September and the October crops were similar (p < 0.05) and however, lower (p < 0.05) than the November and the December crops. The latter two were also similar (p < 0.05).

The NDVI profile for September/ October irrigated crops were not separable and so were those for the November/ December planted crops (Figure 6). As a result, the area estimates for the September/ October crops were combined and so were those for the November and December crops.

Area estimates

The final area estimates for the irrigated and rainfed tobacco were derived after applying a correction factor based on the linear relationship between Landsat area estimates and MODIS area estimates. The September and October tobacco crops were combined into one class and the November and December tobacco crops into another, based on exploratory data analysis in Figures 6. The final area estimates for the 2010-2013 irrigated and rainfed tobacco crops are presented in table 6.1, and these were compared with the crop survey area estimates from the Department of Agriculture and extension of the Ministry of Agriculture. From the crop

survey area estimates, the Department of Agriculture and Extension also issues out yield estimate as presented in Table 6.1.

The results for the 2012-2013 season were however not conclusive. From the 69268.85 hectares for the September-November crop estimated using a supervised classification method and verified with Landsat TM imagery, the December crop area was estimated by assuming that it was in the same proportion of 25 % as in the previous two seasons. The three year average accuracy level for the remote sensing estimate (98.8 %) appeared more reliable, as compared to the crop survey accuracy level (113.6 %) (Table 6.1).

DISCUSSION

The optimal times for discriminating the irrigated tobacco from the non-irrigated tobacco does agree with findings from similar experiments carried out at Kutsaga Research Station (Tobacco T.R.B, 2012). Also similar to this earlier work is the difficulty associated with separating the December crop from the three earlier planted.

The level of accuracy of 98.8 % for this model was comparable the wheat estimation model developed by Zhang et al (2003; 2012). However, the two were quadratic model, relating directly NDVI between the wheat re-greening period and aboveground biomass. The overall error minus 1.2 - 1.7 was lower than the plus

Season	RS area(ha)	Crop survey area (ha)	RS Yield (mkg)	Crop survey (mkg)	Actual tobacco sold (mkg)	RS (%)	Accuracy	Crop survey accuracy
2010-2011	72944	57180	130.718	160.000	133.000	98.3		120.3
2011-2012	76712	67318	141.438	170.000	143.886	98.3		118.1
2012-2013	86586	107000	165.600	170.000	166.000	99.8		102.4
Average	78747	77166	145.919	163.333	147.629	98.8		113.6

Table 6.1: The final area (ha) and yield (mkg) estimates for the 2010-13 seasons, after applying a long-term crop yield-area relationship

13.6 % three year average (Table 6.1) from the department of Agriculture. Future work in tobacco should also focus on a more direct NDVI-above ground biomass relationship.

The numbers of seasons were also not sufficient to improve the results of the average deviation between the crop survey e and the remote sensing estimate.

The application of regression techniques has been for years applied in establishing trends and strengths of relationships among variables. Garvin (1986; Garvin 1985) applied this in predicting tobacco yield from biophysical parameters, while Gomes (2006) applied the technique when he sought the relationship between environmental factors of growth and yield of newly developed maize varieties.

The overall model was multiplied by factor 0.8, which Garvin (1980) experimentally established as the proportion that reached the market after handling losses are incurred. Such a generalization however, could be misleading as it may conceal extremes. In general losses depend on the level of fertilisation during the crop growth phase and later on correct condition of the crop before handling (Garvin, 1986). In farms with poor handling facilities losses can be as high as 30 %, while losses for a correctly conditioned system can be as low as 15 %. With the increase in the number of fly-by-night tobacco growers who are generally not prepared for the season, the proportion of tobacco that finally reaching the floor could be seriously understated.

CONCLUSION

The results of this study indicated that, based on MODIS NDVI data, the last week of November and the period between late January and early March are the optimal times for discriminating the irrigated from the non-irrigated tobacco.

A simple but robust method for estimating the hectarage under tobacco was developed. From the results of this study the irrigated can be discriminated from the rainfed tobacco crop using satellite remote sensing and area estimates for 2010-2011 and 2011-2012 seasons were made. A procedure for estimating tobacco cropped area was developed and a linear model derived from longterm cropped area-yield data, field-based input data and freely available multi-temporal MODIS data can be applied to estimate tobacco yield in Zimbabwe. The modelled yield figures had a good correspondence with the actual yield values provided by the Tobacco Industries Marketing Board. The spatial crop distribution coincided with the overall pattern of the traditional tobacco growing regions of Zimbabwe.

Future work should focus on providing yield distribution maps as basis for efficient land use planning and to make direct estimation of crop yield from the spatial NDVI distribution. Thus, this yield estimation approach could be taken as an economical and feasible way to achieve spatially distributed crop monitoring and yield estimation on a regional scale. The advantage in the process of deriving the necessary input data, it will offer a description of crop development and, naturally, indications of crop yield. However, the actual derivation cropped area from remotely sensed data is of cumbersome. The computation of NDVI values from the sampled field and the ground-truth background which is necessary to achieve a good reliability of the results is complicated and expensive.

ACKNOWLEDGMENT

The authors are grateful to the Tobacco Research Board/Kutsaga Research Station for funding this series of experiments. The authors declare that there is no conflict of interests. The Tobacco Research Board is a research institution that strives to publish all the research work that is carried out and not for financial gain.

REFERENCES

- Atzberger C (2012). Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs". Remote Sens. 2013, 5, 949-981
- Awosola-Olubode OO, Chilonda P, Minde I, Bhatt Y (2008). Indicators for Monitoring and Evaluation of Agriculturak Performance and Shared Goals. Regional Strategic Analysis and Knowledge support Syatem (ReSAKSS) Working Paper No. 24.
- Bouman BAM (1992). "Accuracy of estimating the leaf area index from vegetation indices derived from crop reflectance characteristics, a simulation study". Int. J. Remote Sens. 16:3069–3084.
- Casa A, Ovando G (1996). Development of a tool to monitor

- crop growth and grain yield. Application Manual. Center for Surveying and Assessment of Agriculture and Natural Resources (CREAN). Agricultural Sciences Faculty (Department) of the Córdoba National University (Argentina) (Undated)
- Clevers JPGW, Leeuwen HJC (1996). Combined use of optical and microwave remote sensing data for crop growth monitoring. Remote Sensing of Environment 56:42-51.
- Eitel JUH, Keefe RF, Long DS, Davis AS, Vierling LA (2010). "Active ground optical remote sensing for improved
- monitoring of seedling stress in nurseries," Sensors, vol. 10, no. 4, pp. 2843–2850,.
- Foody GM, Mathur A, Sanchez-Hernandez C, Boyd DS (2006). Training set size requirements for the classification of a specific class. Remote Sensing of Environment, 104, pp. 1– 14.
- Gallo KP, Flesch TK (1989). Large-area crop monitoring with the NOAA-AVHRR: Estimating the silking stage of corn development". Remote Sens. Environ. 27:73–80.
- Garvin RT (1980). Losses from Reaping to the floors = What are they worth?. Zimbabwe Tobacco Today. 3 (12): Pp 7–11
- Garvin RT (1985). Flue Cured Tobacco Yield Estimation in Crop research". Zimbabwe Science News. 19 (5/ 6): May-June 1985.pp 65-67
- Garvin RT (1986). Yield Estimation- A tool for Reducing Losses". Zimbabwe Tobacco Today. 9 (12): December 1986. pp 32-35.
- Gomes R (2006). Non-parametric crop yield forecasting, A didactic case study for Zimbabwe. ISPRS Archives XXXVI-8/W48 Workshop proceedings: Remote sensing support to crop yield forecast and area estimates. Commission VIII, WG VIII/10.
- Green CF, Invins DJ (1985). Time of Sowing and Yields of Winter Wheat. J. Agric. Sci. Cambridge 104:235-238.
- Hatfield JL, Pinter-Jr, PJ (1993). Remote sensing for crop protection. Crop Prot. 12 (6):403–413.
- Huete A, Didan KT, Miura EP, Rodriguez X, Gao, Ferreira LG (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83 (1-2): 195-213 (2002).
- Jackson RD, Slater PN, Pinter Jr. PJ (1983). "Discrimination of growth and water stress in wheat by various vegetation indices through clear and turbid atmospheres," Remote Sensing of Environment, vol. 13, no. 3, pp. 187–208, 1983.
- Kustas, W.P. Perry, E.M., Doraiswamy, P.C. and Moran, M.S, 1994. Using satellite remote sensing to extrapolate evapotranspiration estimates in time and space over a semiarid Rangeland basin. Remote Sensing of Environment 49:275-286
- Maas SJ, Dunlap JR (1989). Reflectance, Transmittance, and Absorptance of light by normal, etiolated and albino corn leaves. Agron. J. 81:105-110.
- Moulin S, Bondeau A, Delecolle R (1998). Combining agricultural crop models and satellite observations from field to regional scales. Int. J. Remote Sens. 19:1021–1036.
- Prasad AK, Lim CB, Ramesh PA, Singh B, Menas K (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. Int. J. Appl. Earth Observat. Geoinformat. 8 26–33.
- Reynolds M, Yittayew DC (2000). Slack, Estimation crop yields and production by integrating the FAO Crop Speci. C Water Balance model with real-time satellite data and groundbased ancillary data. Int. g. remote sensing, 21 (18):3487-3508.

Rizzi R, Rudorff BFT (2005). Estimativa da produtividade de soja por meio de um modelo agrometeorológicoespectral: estudo de caso no Rio Grande do Sul. Anais XII Simpósio Brasileiro de Sensoriamento Remoto, Goiânia, Brasil, 16-21 Abril 2005, INPE, p. 237-244.

- Tobacco Industries Marketing Board (T.I.M.B), (2011). Annual report and Accounts for the Year Ended 30 June, 2010. T.I.M.B, Harare.
- Tobacco Industries Marketing Board (T.I.M.B), (2012). Annual report and Accounts for the Year Ended 30 June, 2011. T.I.M.B, Harare.
- Tobacco Research Board (T.R.B), (2012). Annual report and Accounts for the Year Ended 30 June, 2011. TRB, Harare.
- Yin X, McClure A, Tyler D (2010). "Relationships of plant height and canopy NDVI with nitrogen nutrition and yields of corn,"in Proceedings of the 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia, August
- Zhang H, Chen H, Zhou G (2012). The Model Of Wheat Yield Forecast Based On Modis-NDVI — A Case Study Of Xinxiang. ISPRS Annals of the Photogrammetry. Remote Sensing and Spatial Information Sciences, Volume I-7 (2012). XXII ISPRS Congress, 25 August–01 September 2012, Melbourne, Australia.
- Zhang MA, Fried J AH (2003). Strahler, F. Hodges, and B. C. Reed, "Monitoring vegetation phenology using MODIS". Remote Sensing of Environment 84:471-475.