IDENTIFICATION AND CLASSIFICATION OF SUGARCANE BASED ON SATELLITE REMOTE SENSING

G NARCISO[1] AND E SCHMIDT[2]

[1]*Agricultural Research Council, Institute for Soil, Climate and Water, Private Bag X79, Pretoria, 0001*
[2]*South African Sugar Association Experiment Station, Private Bag X02, Mount Edgecombe, 4300*

Abstract

Remote sensing broadly implies measuring the intensity of different wavelengths of light reflected from the earth's surface. Generally satellites or low level video sensors mounted on an aeroplane are used to capture information which can be used in managing crop production. These applications include mapping, area measurement, monitoring crop condition and estimating production levels.

This paper discusses a project to investigate the use of satellite imagery for identifying and classifying sugarcane and providing timely and acceptably accurate information on the area under sugarcane production. The procedure was tested in the Eston district using a Landsat Thematic Mapper image acquired on April 22 1996.

Digital maps of farms and a database of actual area under cane were used to assess the accuracy of the classification procedures. For the area under study the satellite estimate of area under cane was within 5% of the recorded 10479 ha of sugarcane. For individual farms, the mean absolute difference in estimated sugarcane area was between 5% and 10%, depending on classification method.

Introduction

Information on the location and extent of sugarcane cultivation is important in the sugar industry. Industry reporting is based on accurate records of sugarcane area supplying each mill. While field maps are available for most farms, they are not in a computerised or digital form. Digital maps would allow an assessment using a geographical information system (GIS) of, for example, cane area falling in a selected region (such as a catchment area or climatic region), or within a selected distance from the mill or a sensitive area (such as an urban area or powerline).

Accurate digital field maps are usually prepared from Global Positioning System (GPS) surveys, digital orthophotos or by digitising field boundaries off paper maps. Satellite remote sensing can also be used to map sugarcane areas, but to a lower level of accuracy. The advantage of satellite imagery is that large areas can be captured in one image, information can be updated regularly for monitoring changes and the method is more cost effective where a high resolution is not essential. Furthermore non-visible spectrum wave bands can add further information on crop condition.

Remote sensing allows information to be obtained for a target crop by measuring the intensity of reflected electromagnetic energy. By assessing the reflectance of various wavelengths within the electromagnetic spectrum a "spectral signature" for the crop can be established. The signature will depend on the characteristics and condition of the crop. Image processing software is used to identify the combination of bands that best identify the target crop or condition. The resulting digital image of the crop surface is made up of a large number of pixels each representing a parcel of land, which can be analysed and interpreted in a GIS.

A number of satellite platforms are commonly used in South Africa for remote sensing, the most common being LANDSAT, SPOT and NOAA. The cost of data is directly related to the resolution of the image acquired.

The NOAA satellite with its AVHRR sensor has a very coarse ground resolution of 1.1 km, but its strength is that it images any point on the earth at least once a day, cloud cover permitting. There are either four or five channels sensing in the visible, near infrared and thermal infrared portions of the electromagnetic spectrum. Due to the low spatial resolution of NOAA platforms, their suitability is limited to vegetation mapping and crop condition assessment at a regional level.

The LANDSAT satellite with its TM (thematic mapper) sensor has a ground resolution of 30m and has a temporal resolution of 16 days. There are seven channels sensing the visible, near and mid infrared and thermal infrared wavelengths. LANDSAT has been widely used for mapping operations however the spatial resolution does impose accuracy limitations. The wide range of wave bands provide a great deal of information which can be used for visual or automatic classification and mapping.

The SPOT satellite has sensors, which record both multispectral (green, red and infrared bands) and panchromatic (black and white) images. The multispectral images are recorded at a 20m resolution and the panchromatic images at a 10m resolution and have a 26day temporal resolution.

Review

A limited number of satellite remote sensing studies focusing on sugarcane have been conducted internationally.

Argete (1983) explored the use of LANDSAT MSS data and weather based yield models to predict production for a Mill in the Philipines. Measured sugarcane production was shown to be closely approximated by the area and yield estimated from LANDSAT data. The combined approach using LANDSAT data and the agro-meteorological model was recommended for further testing in other parts of the industry.

Lee-Lovick and Kirchner (1991) used LANDSAT TM data from 13 overpasses of Bundaberg, Australia, between 1986 and 1990, to assess its usefulness for monitoring sugarcane crop growth, health and yield. The remote sensed data could not be used to monitor growth or yield and crop canopy moisture levels dominated the spectral signature, masking long-term stalk elongation trends. Problems in the procedure included unreliable data due to cloud cover, equipment failure and the time consuming nature of data processing and ground control.

Johnson and Kinsey-Henderson (1997) used SPOT multispectral and panchromatic data as well as ERS radar imagery to detect land use change in the Herbert valley, Australia. SPOT panchromatic imagery was able to differentiate land use based on textural information, while SPOT multispectral data could detect land use types. Neither was effective in differentiating cane variety and crop class. It was concluded that RADAR could complement SPOT by providing information on the physical structure of the target

McDonald and Routley (1999) undertook a pilot study to determine the suitability of using satellite imagery and spatial analysis to determine the area of cane to be harvested in a particular season and the area to be harvested as the season progressed. Images acquired through the growing season were compared with mill records of area harvested. The accuracy of image classification prior to the harvest season for standing cane and fallow areas was between 91% and 97%. Relatively simple classification methods gave similar accuracy to more sophisticated procedures. The accuracy of determining harvest areas during the harvest season was very low, between 36% and 47%. A major factor was inaccuracies in the mill records of areas harvested and harvest dates.

The aim of this paper was to investigate the use of LANDSAT imagery for identifying and classifying sugarcane in the Eston mill supply area. LANDSAT was selected since it offers the best combination of spatial and temporal resolution, range in wavebands and access to and cost of images. The procedure was tested using a LANDSAT TM image acquired on April 22 1996.

Methodology

A number of methods can be used for crop identification and mapping using satellite imagery. For example LANDSAT imagery was used in South Africa in a national landcover mapping project by the CSIR and ARC where land use was manually (i.e. visually) interpreted and mapped from the image according to a predetermined classification key. Accuracy was checked in the field, or "ground truthed" (Thompson, 1996). Only contiguous land units of greater than 25ha were mapped in this project which gives rise to accuracy limitations.

The methodology used in this analysis is summarised in Figure 1. The satellite image (LANDSAT TM scene 168-81 of 22nd April 1996) acquired by SASEX from the Satellite Application Centre of the CSIR was geometrically referenced to an Albert's Equal Area Projection. GIS Arcview shape files showing farm boundaries were obtained from the Surveyor General and updated at the local extension office based on farm visits. Area under cane on the date of image acquisition was obtained for each farm from the industrial area under cane survey. The satellite image was classified into land use classes using image-processing software. Training areas of known land use were used to identify the spectral signatures corresponding to each land use. The farm boundary map was processed and overlayed on the classified image in the GIS system. Statistical analyses were performed to compare the area under sugarcane, estimated from the classified image, with the database of known area under sugarcane for each farm.

The selection of training areas was a subjective process, based on the personal interpretation by the analyst of where sugar cane (as well as the other land cover classes) occurred on the image. In order to reduce the subjectivity of this operation and to give an indication of the effects of personal interpretation on the overall results of the classification, two different and independent signature files were created, referred to hereafter as **Cla1** and **Cla2**. The spectral signatures of sugar cane as well as other land use classes determined from the training areas were stored in separate files for analysis.

The classification process is carried out using the facilities of the image processing software (ERDAS-Imagine). The operator is given a choice of three different algorithms: Maximum Likelihood, Minimum Distance and Mahlanobis Distance. These three methodologies do not differ substantially for land cover classification. The Maximum Likelihood procedure was chosen though, being the one that guarantees a relatively better land cover class discrimination.

Each signature file usually comprises of multiple signatures (training areas) for each single land use class, this in order to account for the spectral variability of that class across a scene. This approach implies that the classification process defines a number of classes (corresponding to the number of spectral signatures) with respect to each land use class. It is necessary to undertake a "re-coding" operation at the end of each classification in order to regroup the various classes.

The classification process provided two main outputs, for each of the two signature files used:

- A land-cover map, where the single pixels are labeled according to a specific land use.
- A report file indicating the number of pixels and thus the surface occupied by a specific land cover class over the whole image.

The classification procedure is a statistical process, in the sense that it labels the pixel on the base of its similarity to a model defined by the "spectral signature", regardless of the context or the location of the pixel itself. Given the high variability of surface appearance, the mapping process can result in a "salt and pepper" effect with a random distribution of differently classified pixels, even over homogeneous land covers. In order to eliminate the local anomalies (within a homogeneous land cover), a filtering process can be applied. This process makes use of a "majority statistical filter" that reassigns isolated pixels to the surrounding class, thus artificially reconstructing a map that more closely reassembles the "real field" layout. The results after applying the statistical filter are referred to as **Cla1f** and **Cla2f** below.

Results

The application of the above methods to the Eston area had several advantages and a few drawbacks. The main advantage was the fact that sugar cane is by far the most cultivated crop in the scene and it is grown over large fields on commercial farms. This greatly reduces the possibility of confusion among different land uses and makes classification procedures simpler than would be the case with small cane plots. A major drawback is the steep topography of the area, which induces a distortion on the actual shape and size of the pixels thus introducing an error factor in the estimation of the surface and also results in shadow effects.

In Table 1 the results for the two signature data set (i.e. two training areas) are provided both for the raw (Cla1 and 2) and filtered option (Cla1f and 2f).

Table 1 indicates that the total area estimated under sugarcane over the whole image (comprising 1071 square kilometers), ranged between 29098ha (Cla1) and 31665ha (Cla2) depending on training area, representing a difference of approximately 9%. Application of the statistical filter resulted in a significant reduction in estimated area under cane for Cla1 (640ha) but little change for Cla2. It should be noted that the actual area under sugarcane for the whole image was not known.

The classified image was thus superimposed on the GIS map of 46 farms, which contained records of area under cane on the date of image acquisition. The aggregate results for the 46 farms are given in Table 2. The results show that unfiltered estimates of total area under sugarcane were within 2% of grower records and the filtered estimates within 5% of grower records. A scatter plot for individual farms is given in Figure 2 for Cla1 and Cla2.

Two statistical indicators were used to ascertain the accuracy of the various classifications: the "Mean Absolute Difference" (MADif) and the "Mean Absolute Deviation" (MADev). Absolute differences are used to ensure over and under estimation did not compensate for each other.

The Mean Absolute Difference is less conservative in establishing the accuracy of the classification for low values (i.e. small areas), when compared with the "Mean Absolute Deviation", becoming more conservative as values increase. In this study, area under cane per farm is relatively large so that the MADif will be the most conservative test of the two.

Mean Absolute Difference

$$\mathbf{MADif}_{(i)} = [1/n \sum_{i=1}^{n} | \mathbf{Clas}_{(i)} - \mathbf{M}_{(i)} |] / [1/n \sum_{i=1}^{n} \mathbf{M}_{(i)}]$$

Where:

MADif: Mean Absolute Difference between the classifications and the measured value of the Grower's Area. Value expressed in percentage.
Clas(i): Satellite classified area per farm (i).
M (i): Grower's record area per farm (i).
n: Total of farms

Mean Absolute Deviation

$$\mathbf{MADev}_{(i)} = 1/n [\sum_{i=1}^{n} | \mathbf{Clas}_{(i)} - \mathbf{M}_{(i)} | / \mathbf{M}_{(i)}]$$

Table 1. Estimated area sugarcane from the satellite image.

Land use	Cla1 (Ha)	Cla1f (Ha)	Cla2 (Ha)	Cla2f (Ha)
Sugar cane	29,098	28,458	31,665	31,767

Cla1 - Raw using the first spectral signature file; Cla1f - Filtered using the first spectral signature file
Cla2 - Raw using the second spectral signature file; Cla2f – Filtered using the second spectral signature file

Table 2. Observed and estimated area under sugarcane for 46 farms.

Land use	Grower Records	Cla1 (Ha)	Cla1f (Ha)	Cla2 (Ha)	Cla2f (Ha)
Sugar cane	10,479	10,272	10,974	10,557	10,913

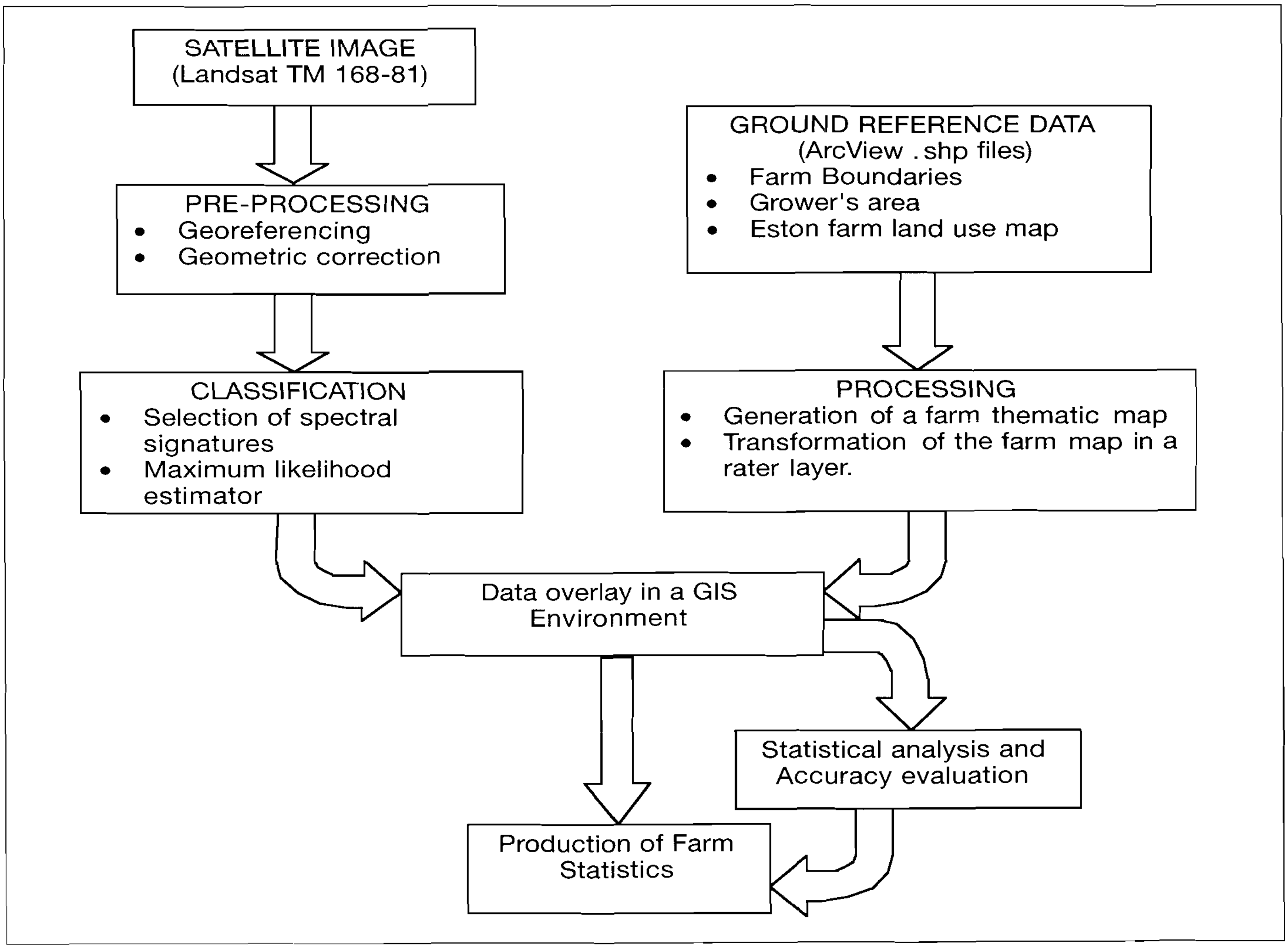

Figure 1. Outline of the methodological approach.

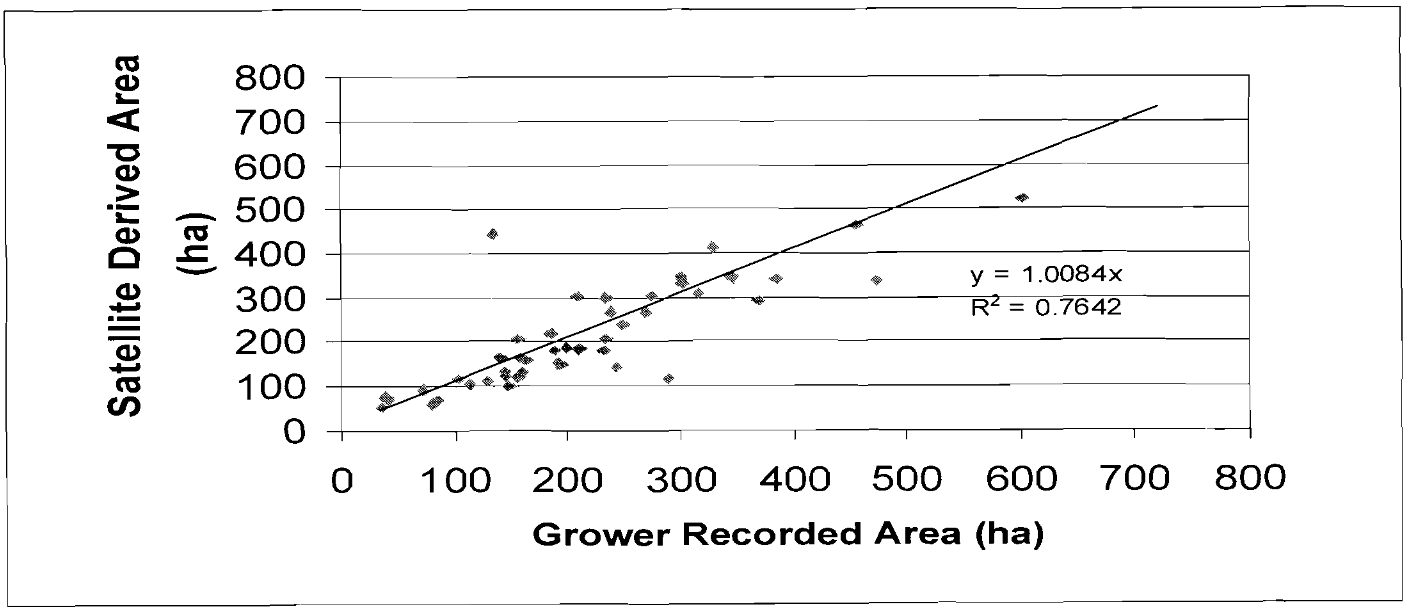

Figure 2. Scatter plot of are under sugarcane from grower records and satellite data (Cla1 and Cla2).

Table 3. Mean Absolute Difference and Deviation accuracy analysis of the classification.

	Absolute difference (%)			
	Cla1	Cla1f	Cla2	Cla2f
Mean Absolute Difference	0.07	0.10	0.05	0.07
Mean Absolute Deviation	0.06	0.07	0.03	0.05

Where:

MADev (i): Mean Absolute Deviation between the classifications and the measured value of the Grower's Area per farm. Value expressed in percentage.

Clas (i): Satellite classified area per farm (i).

M (i): Grower's record area per farm (i).

The result of this statistical analysis is illustrated in Table 3, which indicates that the best performing classification is Cla2 using the second set of spectral signatures with no filtering.

Conclusions

The above analysis demonstrates the feasibility of using LANDSAT satellite imagery for identifying and classifying sugarcane. Estimates of sugarcane area, derived from the classification of satellite data, were within 5% of the total area recorded by growers on 46 farms. A slight difference in estimated area resulted, depending on the "training area" used to define the spectral signature of sugarcane. The use of a filter to eliminate possible anomalies within a homogenous land cover resulted in less accuracy in area determination.

The procedure achieved significant accuracy for individual farms, ranging between 3% (MADev%) and 5% (MADif%) of the actual grower's area when using classification 2, not filtered (Cla2). Furthermore MADif% and MADev% are below 10% for all classification methods. Likely sources of error in estimated area under cane are poor classification of harvested fields (although this is not likely to be significant since the mill had only been open 109 days the date of image capture), flowering cane and ploughout land.

Significant errors were encountered on certain farms (see Figure 2). This can be ascribed to inaccurate information on farm boundaries, inaccurate grower records of area under cane.

The use of the Landsat TM sensor thus has potential for mapping areas under sugarcane on a broad scale for the purposes of determining cane area falling in a selected region or within a certain distance from a point or region of interest. The method is not accurate enough to be used for field mapping and field area determination. The method is less applicable on areas where cane is grown in small plots remote from one another.

The methodology has the following advantages:

- The statistical figures can be obtained virtually in real time following the acquisition of the satellite image
- It is possible to derive from the satellite data set ancillary information on crop condition or phenological stages.
- It is possible to obtain statistics on other land use classes besides the main target crop.

Recommendations

Possible ways to improve the accuracy of sugarcane classification and address shortcomings experienced during the implementation of this project can be summarised as follows:

1) Use of an Area Frame Approach for the stratification of the area and the definition of the sampling scheme.
   a) The application of the full theory of Area Frame implies the stratification of the study area. This means excluding all those surfaces, which are of no interest to the scope of the statistic (non-agricultural areas) and grouping the area object of the study into broad clusters, which reduce the internal variability of the sample.
   b) Stratification according to farm and field size is another item to be considered as it is often related to the cultivation patterns and the technological level of the cultivation. Stratification could also help in optimizing the sample size and to improve accuracy for more sparsely cultivated crops.
2) Use of randomized sampling for the selection of the training areas and for the extraction of the spectral signatures.
3) Use of a multi temporal approach: with the acquisition of multiple satellite coverage during the season. This approach would greatly enhance the discrimination capabilities for the single crops or land use classes, allowing the exclusion of natural vegetation, or the early identification of different ground covers when they are more diversified due to the specific phenological development.
4) Quality control on all of the process phases: Given the relatively complicated setup of the whole procedure it is fundamental that each phase is thoroughly controlled in terms of correctness of application and quality of the data produced. Wrong or incomplete data, feeding the procedure, especially from the ground surveys, can easily hamper the whole outcome of the procedure as these errors are carried along up the final production of the estimate. A strategic factor in guaranteeing the quality of the data is a thorough coordination and quality control of ground surveys and the maximum possible automatisation of all the following data processing tasks.
5) Application of advanced classification procedures: Several options are possible to improve the quality of the classification procedures, these may refer either to statis-

tical algorithms or to improved data integration using the available databases (ground surveys, land cover and soil maps, meteorological data). This all refers to a more advanced use of GIS technology and filtering techniques applied to the accomplished classification. Probably the most promising field of research on this subject is provided by the integrated application of neural networks and "expert systems", especially for classes, which do not have a Gaussian distribution of signatures.

Acknowledgements

Messers D Hellmann and D Hampson of the Eston extension office are acknowledged for his assistance in obtaining grower production areas and mapping of farm boundaries.

References

Argete, JC (1983). Sugarcane production estimates using LANDSAT MSS data and agrometeorological models. *Proc. Phillippines Sug Technol Ass* : 583-596.

Johnson, AKL and Kinsey-Henderson, AE (1997). Satellite-based remote sensing for monitoring land use in the sugar industry. *Proc Australian Soc Sug Cane Technol* 19: 237-245.

Lee-Lovick, G and Kirchner, L (1991). Limitations of LANDSAT TM data in monitoring growth and predicting yields in sugarcane. *Proc Australian Soc Sug Cane Technol* 13:124-129.

McDonald, L and Routley, S (1999). Use of satellite imagery, remote sensing and spatial analysis to determine the area under cane prior to and during the harvest season. Report to the Herbert Cane Protection and Productivity Board. Cooperative Research Centre Report No 03/99 : 28pp

Thompson, M (1996). A standard land-cover classification scheme for remote-sensing applications in *South Africa. South African Journal of Science* 92: 34 – 42.

Factory