Identifying Sugarcane Plantation using LANDSAT-8 Images with Support Vector Machines

Sidik Mulyono[1], Nadirah[1]

[1]Agency for the Assessment and Application of Technology (BPPT), Republic of Indonesia

Email: sidik.mulyono@bppt.go.id, nadirah@bppt.go.id

Abstract. The use of remote sensing has been highly beneficial in the identification and also mapping and monitoring of plantations. The identification of plantations includes the physiology, disease, environmental conditions, and also the production and time of harvesting. It can be done by doing satellite imagery classification. However, to reach the final result of identification, it could be carried out by getting the solid ground truth information. This paper will discuss about detection of sugarcane plantation in Magetan district of East Java province area by using LANDSAT-8 image with specific approach of phenology profile using EVI (Enhanced Vegetation Index) value from satellite data, as an alternative vegetation index to address some of the limitation of the NDVI (Normalized Difference Vegetation Index). Method of classification used for detecting sugarcane plantation is Support Vector machines (SVM), which is a promising machine learning methodology. It has the ability to generalize well even with limited training samples and complex data. A number of samples of phenology profile for training purpose using SVMs are obtained from the area that identified as sugarcane plantation during field campaign in 2015. The same manner is also done for the objects instead of sugarcane plantation with relatively the same number of samples. The result of the research shows that Remote Sensing is able to detect the sugarcane plantation cross the district with good accuracy.

1. Introduction

Sugarcane (*Saccharum officinarum*) is a special kind of grass that allows multiple cuts throughout several years and is considered as a semi-perennial. It grows quickly with an abundant reproduction and allows for economic exploitation of many parts of the plant. It has a 4-phases during its growing, i.e. germination phase, tillering phase, stem elongation phase, maturity and ripening phase. From planting until the first cut it is called cane plant, whose cycle lasts between 12 to 18 months, depending on the season and region of planting. After the first cut, the regrowth of sugarcane is replaced by the ratoon cane, with a normal cycle of 12 months. This planting cycletypically happens 4 to 5 times before it completely ploughed out and replanted by a new plant crop.

In Indonesia, sugar which is one of the basic needs for the community and industry, is still continuing to be a problem due to lack of domestic production, whilst the demand of sugar tends to be increased continuously. A large proportion of sugar produced by the sugarcane is imported from overseas countries. The condition of existing sugarcane mills that mostly located on Java and Sumatera island is too old, low productivity, and highly dependent on sugarcane farmers that only have very limited area for sugarcane crop on Java and Sumatera island. The rapid increment of sugar demand compared to the production improvement is relatively not balanced, makes Indonesia be a big importer country for raw crystal sugar as well as refined sugar. The development of sugar industry must be conducted integrally which starts from plantation, processing, marketing, and distribution supported by stakeholders, supporting agencies such as human resources development, finance and banking, and transport.

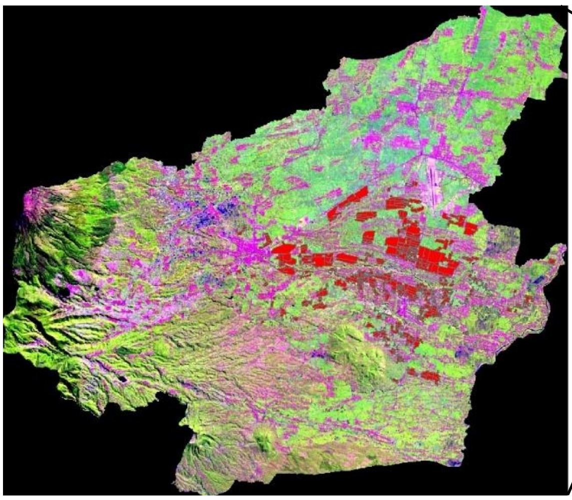
Since sugar production capacity in Indonesia is still not sufficient to fulfil the national demand on sugar, therefore the Ministry of Agriculture had been revised the target of production capacity in 2014 from 5.7 million tons to 3.1 tons (agroindonesia.co.id, 2012). The target decline was triggered by the difficulty to acquire new land for expansion of sugarcane field. To achieve the production of 5.7 million tons, it needs a new sugarcane field about 350,000 hectares. But in reality, extending program for expansion of sugarcane field does not run smoothly, and until now the required land has not been able to be acquired S. Sinunggila, 2015). However, this paper will discuss about detection of sugarcane plantation in Magetan district of East Java province area by using LANDSAT-8 images with specific approach of phenology profile using EVI (Enhanced Vegetation Index) value from satellite data, as an alternative vegetation index to address some of the limitation of the NDVI (Normalized Difference Vegetation Index).

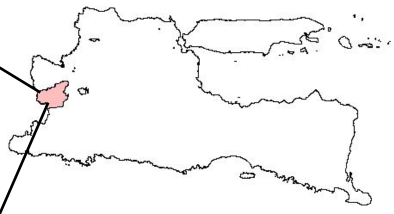
Recent studies have used remote sensing images with both passive sensor (reflectance based measurement) and active sensor (backscatter based measurement) to map sugarcane fields through a various classification methods applied to remote sensing image time series, since spatial information can be provided in detail by satellite imagery. The amount of detail contained in the imagery is dependent on the attributes of the satellite. The intensity of the reflected electromagnetic radiation from agricultural vegetation such as sugarcane can be measured and recorded by satellite remote sensing instruments as a digital image. EO-1 Hyperion Hyperspectral imagery has been used to develop prediction model for detecting variety and plant cycle of sugarcane crop with support vector machines and random forests [1]. Understanding the backscattering behavior of sugarcane over its planting cycle and its relationships with sugarcane growth-related parameters is prerequisite for developing effective methods to map sugarcane plantation area and monitor sugarcane growth using ENVISAT Advanced SAR data. The alternating polarization precision (APP) data was proposed for mapping sugarcane growth area and retrieve LAI of sugarcane crop derived from the polarization ratio of ENVISAT ASAR HV/HH data [2]. In addition, Antunes et al. (2011) proposed the potential of data mining techniques to analyze MODIS/Terra remote sensing data is proposed for classifying sugarcane crop using the decision tree induction technique with occupying some rules [3]. The constellation data of both SPOT-4 and SPOT-5 are applied to predict a productivity of sugarcane crop using an empirical relationship with a growing season-integrated Normalized Difference Vegetation Index NDVI, which is compared with the Kumar-Monteith efficiency model and a forced-coupling method [4]. Meanwhile, NDVI derived from LANDSAT-8 has great ability for detecting crop type, crop conditions (harvested or growing) and mapping sugarcane crop field for medium sized farms, and normalized difference water index (NDWI) derived from LANDSAT-8 also has great potential to distinguish for green and burnt harvest by segregating dry and humid surfaces [5]. This paper addresses on how to identify sugarcane plantation through several time series of EVI derived from LANDSAT-8 images that represents phenology profile of 1 planting cycle for sugarcane crop. Support vector machines (SVM) is adopted to distinguish phenology profile for non-sugarcane field and sugarcane plantation area.

Figure 1. Sugarcane crop

2. Study area

The study area is located in Magetan district of East Java Province (Figure 1), which centered at 7° 38' 30" S and 111° 20' 30" E and altitude between 660 - 1,660 meters above sea level. Geographically, Magetan district is suitable for the development of agricultural products, therefore a superior product of this district is agricultural products, both food crops and other horticulture. Plantation crops in this district is mostly dominated by sugarcane, which are used to supply materials for the sugar crystal to sugar mills in Magetan district. Sugarcane plantation area in this district is obtained from GPS field survey during fiscal year of 2015 managed by PTPN-XI, which is shown in figure 2 (a) indicated by red shading. This area is then used as a reference to evaluate the accuracy of the model for identifying sugarcane plantation in all over Magetan district.





(b) East Java Province

(a) Magetan District

Figure 2. Study area

3. Data and Methodology

3.1 Data samples

LANDSAT-8 satellite launched on February 11, 2013, belongs to U.S government is dedicated for earth observation. This eighth satellite in the LANDSAT program is built by a collaboration between National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS), which provides moderate-resolution (15 m -100 m, depending on spectral frequency) measurements of the Earth's terrestrial and polar regions in the visible, near-infrared, short wave infrared, and thermal infrared. The images used in this study are obtained by downloading them from website http://glovis.usgs.gov/ for free.

Table 3. LANDSAT-8 images used in this study

No. image	Date of image	No. image	Date of image	No. image	Date of image
1	2015-01-14	9	2015-05-22	17	2015-09-15
2	2015-01-30	10	2015-06-07	18	2015-09-27
3	2015-02-15	11	2015-06-23	19	2015-10-13
4	2015-03-03	12	2015-07-09	20	2015-10-29
5	2015-03-19	13	2015-07-25	21	2015-11-14
6	2015-04-04	14	2015-08-10	22	2015-11-30
7	2015-04-20	15	2015-08-26		
8	2015-05-06	16	2015-09-11		

A number of LANDSAT-8 images used in this study are 22 series acquired from January 14 until November 30, 2015 with having 7-bands spectral respectively. After doing radiometric, geometric, and atmospheric correction, the calculation of enhanced vegetation index (EVI) is applied to all images to perform EVI images using IDL code program in ENVI. Finally, the entire images of EVI is merged into a cubic image of EVI by using layer stacking technique in ENVI, whereas one pixel will present phenology profile during 2015 in all over Magetan district.

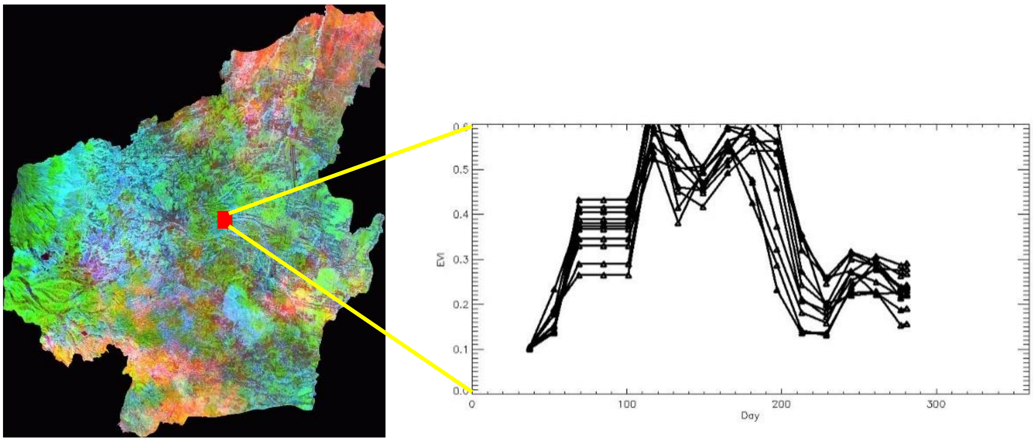


Figure 3. Original EVI cubic image and crop phenology profiles

Since LANDSAT-8 image is sensitively affected by cloud, then the profile looks like having irregular fluctuations, as shown in figure 3, which is it will lead difficulties in advance analysis. In order to avoid something that are not desirable in the analysis, it should be done by smoothing technique called Savitzky-Golay digital filter method with 2 order of polynomial and 5 points of subsets. This method is defined for the purpose of smoothing the data, that is, to increase the signal-to-noise ratio greatly distorting the curves in general.

In this study, instead of conducting the field survey, we selected training sample areas from smoothed EVI cubic image for 6 classes of the object surrounding Magetan district area. One class for sugarcane crop area is derived from the sugarcane plantation area managed by PTPN XI, whereas the rest of them (5 classes) are for the objects represent instead of sugarcane crop that are derived arbitrary from the images, as shown in figure 5. From this figure, we can understand that through phenology profile derived from EVI, each object can be determined as a unique profile.

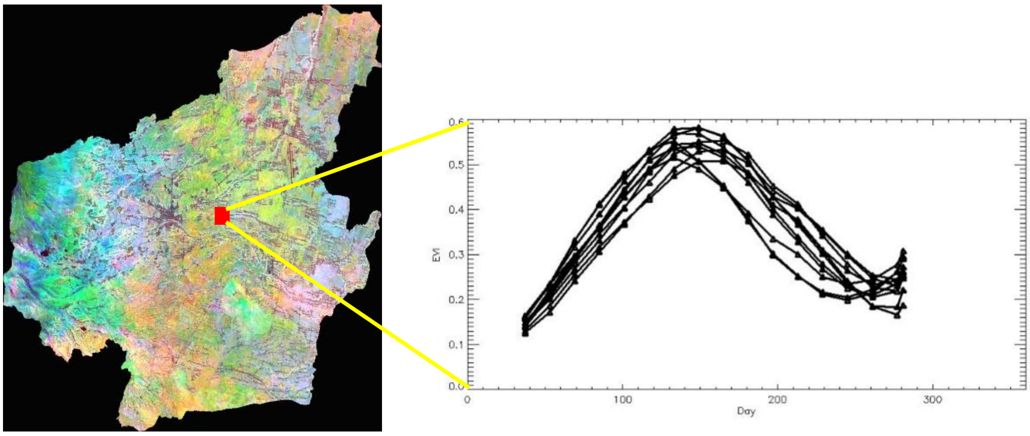
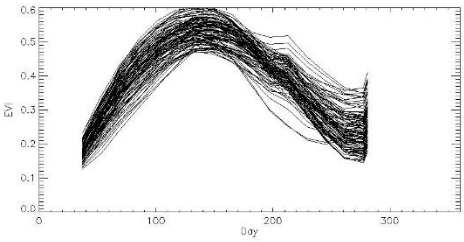
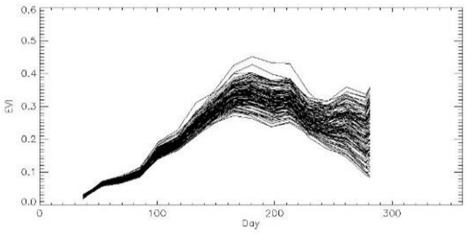


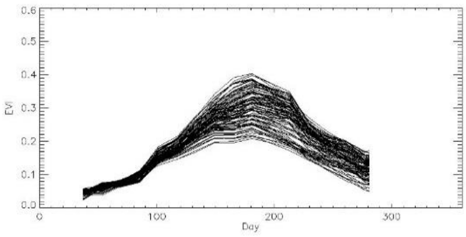
Figure 4. Smoothed EVI cubic image and crop phenology profiles



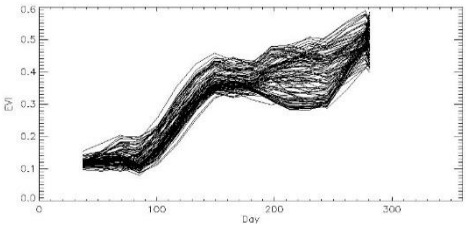
(a) Sugarcane: 2,506 pixels



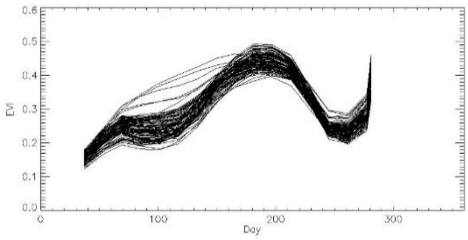
(b) Non Sugarcane 01: 1,567 pixels



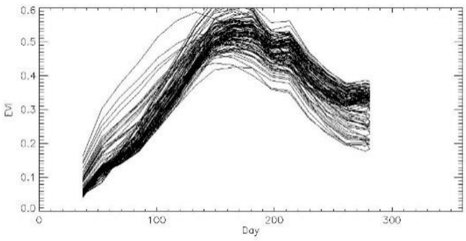
(c) Non Sugarcane 02: 1,640 pixels



(d) Non Sugarcane 03: 1,618 pixels



(e) Non Sugarcane 04: 2,559 pixels



(f) Non Sugarcane 05: 2,437 pixels

Figure 5. Training sample data through phenology profile

3.2. Support Vector Machines

The support vector machine (SVM) is a popular technique and useful for data classification. Basically, SVM classifies binary data by determining the separating hyper-plane or decision surface, which maximizes the margin between the two classes in the training data [6]. Kernel function provides SVMs with powerful ability of efficiently determining the nonlinear decision surfaces after transforming the input space into kernel space. In addition, SVM is able to transform a non-convex problem into a convex problem with a single global minima and thanks to its quadratic optimization scheme. Given a training set of instance-label pairs (x_i, y_i), where $i = 1, \dots, l$, $\mathrm{x}_i \in R^n$, and $y \in \{1, -1\}^l$, the SVM requires the solution of the following optimization problem:

$$\begin{aligned} &\min_{w,b,\xi} && \frac{1}{2} w^T w + C \sum_{i=l}^{l} \xi_i \\ &\text{Subject to} && y_i(\mathrm{w}^T \phi(\mathrm{x}_i) + b) \geq 1 - \xi_i \\ &&& \xi_i \geq 0 \end{aligned} \tag{1}$$

Here training vectors x_i are mapped into a higher dimensional space by the functionϕ. SVM finds a linear separating hyper-plane with the maximal margin this higher dimensional space, whereas $C > 0$ is the penalty parameter of the error term.

The development of SVMs classification for sugarcane field detection is done by using a library for support vector machines (LIBSVM) based on sequence minimal optimization (SMO) [7] installed in ENVI software. The kernel trick used in this paper is radial basis function (RBF).

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right), \gamma > 0 \tag{2}$$

Compared to other kernel tricks (linear, polynomial, and sigmoid), the γ parameter of RBF kernel provides greater flexibility in controlling the desired classification accuracy. Such accuracy control mechanism is highly important when we want to combine the SVM as weak classifiers in boosting learning to overcome the over-fitting problem.

3.3. Multi class strategy for SVM

One major disadvantage of SVM, however, is that SVM classification is binary. This problem can be resolved in a simple and robust procedure by training several SVMs simultaneously in any multiclass strategy. In this study, we use pairwise classification strategy that performed in pyramid structure to classify 6 classes of object, which will consist of at least 15 classifiers. This strategy is already involved in ENVI software platform, and we set value for γ and C as the default, i.e. 0.045 and 1.0 respectively.

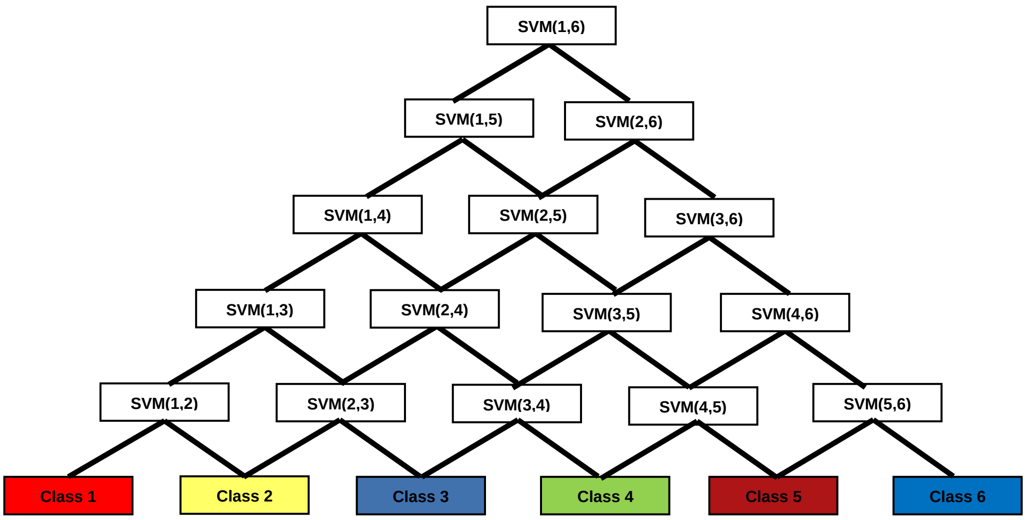
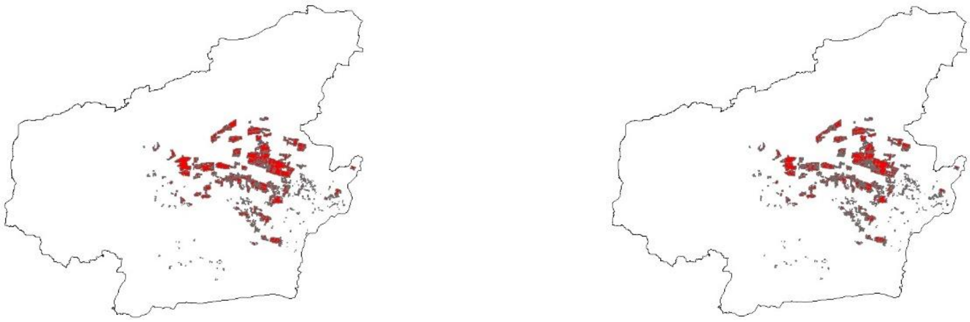


Figure 6. Pairwise multiclass strategy

4. Result and discussion

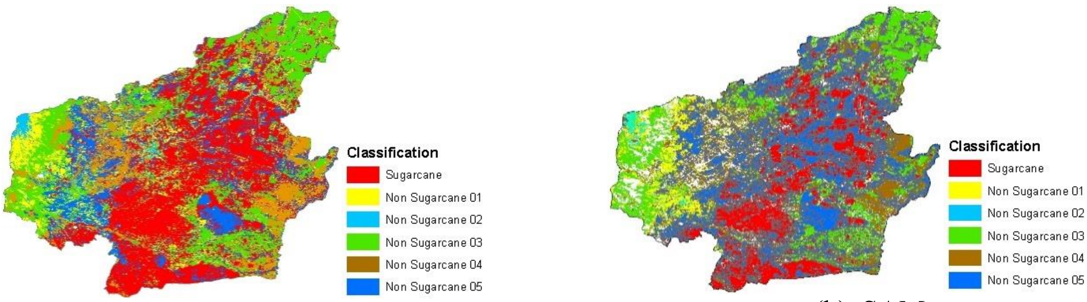
We compared the result obtained from SVM with conventional classification method in remote sensing that is called spectral angle mapper (SAM) that provided in ENVI software as well. In case of the pairwise multiclass strategy of SVM, the entire training sample data are employed as an input for learning. Meanwhile, in case of SAM, each training sample data are used as reference vectors with threshold set at 0.2 radian against the target in general. Both classification results are evaluated against the sugarcane plantation area managed by PTPN-XI as a reference area. Total accuracy can be determined by calculating a total number of sugarcane's pixel predicted by both classification methods allocated in reference area, divided by entire pixel of reference area. Total accuracy for SVM reached about 90.78%, whereas SAM reached only at 78.92% (figure 7).

By using the same manner, we then can identify the sugarcane plantation in entire area of Magetan district as shown by red shading in figure 4.2. The total sugarcane plantation can be detected by SVM is reached about 20,410 ha (37.40%), whereas in case of SAM is reached about 11,143 ha (20.95%). Both methods give the promising results on identification of sugarcane plantation in Magetan district that can provide important information regarding to the potential of sugarcane material for supplying to sugarcane mill, beside the existing area that have been managing.



(a) SVM (b) SAM

Figure 7. Classification result in reference area of PTPN-XI



(a) SVM (b) SAM

Figure 8. Sugarcane plantation in Magetan district

5. Conclusion

In this paper we proposed the novelty method for identifying sugarcane plantation area through crop phenology profile of EVI time series derived from LANDSAT-8 images.In term of classification, the data set of phenology profile derived from vegetation index have much better approach instead of we use only spectral data set, because they represent the whole of characteristic of an object in a certain time range.

In addition, this method is also suitable applied for classifying other crops/plantations or objects that have phase change, such as palm oil plantation, acacia plantation, etc. Using this classification model, we can identify another sugarcane plantation outside area managed by PTPN-XI.By detecting more sugarcane plantation, it will give an information on sugarcane crop stock in order to increase the material supply to the sugarcane mill to achieve national need on sugar.

6. Acknowledgement

The authors would like to thank the people from the PTPN-XI for their efforts in collecting the *in situ* data for sugarcane plantation used as reference area in this study, and colleagues from Center of Technology for Natural Resources Development, Indonesia Agency for the Assessment and Application of Technology for theirhelp for receiving and processing LANDSAT-8 images.

Reference

[1]. YL Everingham, KH Lowe, DA Donald, DH Coomans, and J Markley 2007Advanced Satellit Imagery to Classify Sugarcane Crop Characteristic *Agron. Sustain. Dev.***27** 111-117

[2]. Hui Lin, Jinsong Chen, Zhiyuan Pei, Songling Zhang, and Xianzhi Hu 2009Monitoring Sugarcane Growth Using ENVISAT ASAR Data *IEEE Transactions on Geoscience and Remote Sensing* **47**(8) August 2009

[3]. JFG Antunes, LHA Rodrigues, and S R de M Oliveira 2011 Data Mining for Sugarcane Crop Classification Using MODIS Data *EFITA/WCCA* '11

[4]. Julien Morel, Pierre Todoroff, AgnèsBégué, Aurore Bury, Jean-François Martiné 4and Michel Petit 2014Toward a Satellite-Based System of Sugarcane Yield Estimation and Forecasting in Smallholder Farming Conditions: A Case Study on Reunion Island *Remote Sens*.**6** 6620-6635; doi:10.3390/rs6076620

[5]. Betty Mulianga, AgnèsBégué, Pascal Clouvel, and Pierre Todoroff 2015Mapping Cropping Practices of a Sugarcane-Based Cropping System in Kenya Using Remote Sensing *Remote Sens*.**7** 14428-14444; doi:10.3390/rs71114428

[6]. Rick Archibald and George Fann October 2007Feature Selection and Classification of Hyperspectral Images with Support Vector Machines *IEEE Geoscience and Remote Sensing Letter* **4**(4) October 2007

[7]. Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin 2010 *A Practical Guide to Support Vector Classification*, Software available at http://www.csie.ntu.edu.tw/~cjlin, last updated April 15, 2010