Use of Remote Sensing in Estimation of Agricultural Land

Javier Vicente*, Aguilar; Patricia, Ornelas; José Luis, López; José Eduardo, Reyna; Herlay, Olva.
Servicio de Información Agroalimentaria y Pesquera (SIAP-SAGARPA)
Benjamín Franklin No. 146, col. Escandón, Miguel Hidalgo. CP 11800
Ciudad de México, México
Jaaguilar@siap.gob.mx

DOI: 10.1481/icasVII.2016.g42

ABSTRACT

Mexico's staple crops such as corn, kidney bean, wheat and sorghum are present throughout the country, therefore, it is important to identify their distribution patterns and estimate their area every agricultural cycle. For that purpose, we identify the areas that concentrate 90% of the planted area per state, then, we select samples by means of statistical and environmental criteria. Along with the field work, we acquire satellite imagery. Before the classification process, it is necessary to apply geometric and radiometric corrections and identify the crop mask. Ground truth, obtained from the field work, serves to analyse spectral signatures, histograms and scatter plots of crops or other land cover. Maximum likelihood is the classification method used and, in order to improve the agricultural parcels geometry, a post-classification is applied. Prior to the publication of these results a Kappa analysis is implemented, the results accuracy is regularly above 85%. The general results show, for the agricultural year 2015, an estimation of 5 897 499 hectares for corn, 1 545 766 for kidney bean, 1 884 748 for sorghum and 810 840 hectares for wheat.

1. Introduction

Mexico’s population growth and the increase of foreign demand for agricultural productshave resulted in a biggerrequest of farmingresources for this country. So, the knowledge about the amount of production and its localization increases in relevancy. Because of this, to get timely and reliable information about the sown area for the crops of interest in favour of the agri-food security at a national scale becomes essential.

Remote Sensing (RS), Statistics, *Geographical Information Systems* (GIS) and *Global Positioning Systems* (GPS) provide accurate information about the elementsalong the Earth's surface. Through RS, information about the land cover can be achieved without being in direct contact with it (Chuvieco, 1995). Statistics provides sampling methods applied to the natural communities, GIS processes contribute with the data analysis and output generation (Gómez y Barredo, 2005) and using the GPS technology, any object can be geographically located.

Traditionally, agricultural surfaces were estimated using methods which didn't include precision parameters. Nowadays, they are obtained through the *Area Frame Method*, where sown areas can be estimated using statistical sampling and RS (Ambrosio y Alonso, 1993; Gallego, 1995).Since the 90's decade, in Mexico, it has being planned to get the farming statistics through anAreaFrame, agricultural stratum collection and use of RS. In 2004, collegiate organizations of the *Ministry forAgriculture, Livestock, Rural Development, Fishery and Food Supply* (SAGARPA because of its name in Spanish) launched the *National Surface, Production and Efficiency Estimation Plan*, and, based on this, the *Agri-food and Fishery Information Service* (SIAP, because of its name in Spanish) adopted, in 2007, part of the European Union's *Monitoring Agriculture with Remote Sensing* (MARS) method, to generate the geo-referenced crop areas.

The MARS methodstates a sampling strategy for the evaluation. It is a stratified random area sampling, in which (i) the sampling frame is part of the territory, supported by topographical maps, (ii) the sampling unitsare 700 meters long per side squares and (iii) the stratification criteria is the land use, considering from 6 to 10 stratum or land use types (Gallego y Delince, 1991a in Deppe, 1998). The relative accuracy of this method has being estimated as 1.53 for wheat and 1.31 for barley. For the rest of the principal crops, accuracy has resulted similar (Gallego y Delince, 1991b in Deppe, 1998).

Mexican variety of ecosystems causes a heterogeneous agricultural development. Because of this, the method to estimate the crop surface has being adjusted through RS, considering the MARS method principles (Gallego, 1995). Specifically, SPOT sensor imagery is used, which has being used for diversepurposes (Valdez, González y De los Santos, 2006; Aguilar et al, 2012).

Considering all of these, we have set the objective of estimating, at national level, the area for crops of interest through Remote Sensing.

2. Method

Figure 1presents a general overview ofthis method. It consists of the following stages: sampling design, fieldwork, satellite images acquisition, digital image processing and supervised classification. It was performed for thetwoagricultural cycles considered in México: spring-summer (PV because of its name in Spanish; from April to October) and autumn-winter (OI because of its name in Spanish; from November to March).

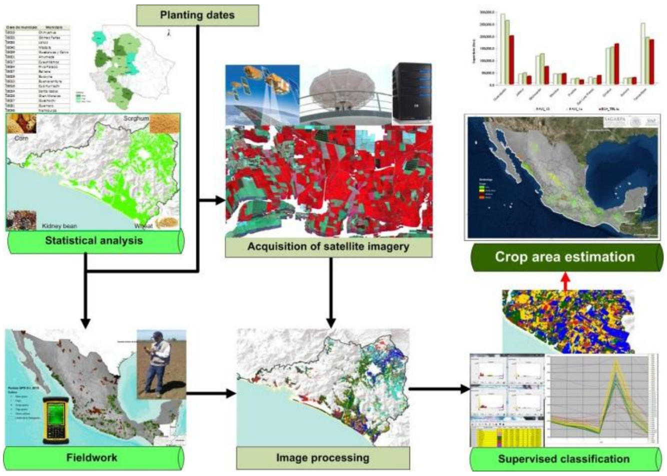

Figure 1: *General overview of the method*

2.1 Sampling Design

The sampling universe was the list of municipalities (4,475) in the Web Agriculture and Livestock Network (RAW, because of its name in Spanish) database. It contains information at national, state and municipality level about sowing and harvesting progress corresponding to both cycles, for irrigation and seasonal regimes. Specifically, the reported information for 2014 was used.

The sampling method is targeted, stratified by conglomerate in a single stage and systematized with a distribution proportional to the whole of the sowed surface for each conglomerate. It is targeted because the information is obtained in sampling units which the field technician assumes as representative. It is stratified by the land use and crop type; and it is conglomerated by the geographical location of each federal entity. It is considered as in one stage because the primary sampling unit for each conglomerate is the municipality, and the sampling unit is a geographical point registered inside an area sowed of the crop of interest. It is systematized because the basic units are organized from most to less sowed surface, so the bigger amount of the sampling units will be placed where most of the sowed surface is located, in a proportional distribution.

2.2 Sampling Selection and Calculation

For each strata (corn, kidney bean, sorghum and wheat) and federal entity, the municipalities which together represent 90% of the sowed surface, are selected. The sampling space for each municipality considered 3% of the sowed surface. The MARS method (Gallego, 1995) was modified as a response to the specific heterogeneity and crop diversity conditions and the available resources. The number of sampling points was determined with the following formula:

(1) $$n = \frac{Z^2 pqN}{(N-1)\varepsilon^2 + Z^2 pq}$$

Where

n= Number of points.
Z= Confidence level (90%).
p= Positive variability (90%).
q=Negative variability (10%).
N= Population size.
ε= Accuracy error (15%)

2.3 Fieldwork

22,573 GPS points were considered for the 2015 exercise, being 17,805 for the PV cycle and 4,768 for the OI cycle. Table 1 presents their distribution.

Table 1:*Stratified samplingdistribution*

Cycle	Corn		Kidney Bean		Sorghum		Wheat	
	Surface (ha)	Points	Surface (ha)	Points	Surface (ha)	Points	Surface (ha)	Points
PV	172,232	11,855	39,564	3,317	24,651	1,989	2,625	644
OI	30,673	2,435	5,584	1,007	22,676	569	19,430	757
Total	202,905	14,290	45,148	4,324	47,327	2,558	22,055	1,401

Once the number of samples was defined for each crop and municipality, the sowing dates are consulted. Considering this, field work and the acquisition of satellite images are scheduled, which is planned to be when crops are in flowering and fruiting(Fernandez, Gepts y López, 1986).

Each GPS point should be 50 m inside the plot, being each one about 1 km far from the nearest. Each plot should be at least 0.5 ha big. The information collected at each pointwas: crop, crop's height (in meters), approximated property's surface, phenological stage, variety, water regime, percentage of ground covered by the crop, weed percentage, sinister, north, south east and west bordering crops, and 2 geo-referenced pictures.Besides local particularities, from July to September, for the PV cycle; and from January to March, for the OI cycle, have resulted the adequate months for the fieldwork.

2.4 Satellite Image Acquisition

Considering the sowing dates, the satellite image acquisition is requested to the *Mexico Receiving Station* (ERMEX because of its name in Spanish), which is the Mexican facility for the management of the SPOT images. For this study, they should be ortho-rectified, a maximum cloud cover of 10% and an incidence angle range of 20° to 25°.In the PV cycle, there were used 1,475 images, and in the OI cycle 1,064 images were used.

2.5 Digital Images Processing

In order to perform the supervised classification, the digital levels (DL) of the image should be expressed in reflectance values. For this, a radiometric correction should be performed.

As long as the surface estimation should be focused on the agricultural area, the satellite images were cut with the geographical delimitation of said areas, which SIAP has available (Figure 2).

a) b)

Figure 2: *Satellite image before (a) and after (b) cutting of non-agricultural areas*

2.6 Supervised Classification

A supervised classification is performed to the satellite images, supported by the fieldwork, to get the above stated classes, including the “other coverings” one (other crops, highways, roads, canals and water bodies). The band configuration to make this identification is false color, which is obtained by placing the near infra-red, red and green bands in the red, green and blue order, which is the 4,3,2 combination for SPOT 6/7. These bands are chosen because they have the greater original variance content (Chuvieco, 1995).

The training phase is fundamental in the supervised classification. In it, several factors should be considered, since they influence in the reflectance for each selected band. For example, this can result in a category expressed in several spectral classes, two or more categories expressed in a single spectral class, or several categories expressed in several classes. For each training field, homogeneous samples or variable surfaces for the same category are sampled. Each one of these should be composed of a number of pixels between 10m and 100 m for each category (Jensen, 1986& Mather, 1987 in Chuvieco, 1995), considering the physical variations of the Earth surface.

These variations are evaluated through a spectral signature and a separability analysis. The advantage of the spectral signature analysis is to highlight the pixels that belong to a certain category, according to the parallelepiped decision rule. Meanwhile, the separability analysis calculates the statistical distance between signatures, which can be used to determine the best group of elements that will be used in the classification.

In the assignation phase, using the maximum likelihood method each element is assigned to a category assuming it has a normal distribution. Considering this, knowing the mean and standard deviation for each category, the probability of association to it can be estimated. The post classification process consists in a generalization according to the variability of the categories, according the neighbor ones and, if applicable,integrates them. When there are isolated pixels, they are assigned to the dominant category. Also, when a pixel is smaller than 0.25 ha, it is assigned to the neighbor category.

To evaluate the classification, 3% of the field samples are used in a confusion matrix, which expresses the conflicts between classes. Due to its configuration, it can be identified what was correctly classified according to what was sampled in the fieldwork, and mistakes of classifying what is not present in field or elements in reality which are not represented in the classification. The proportion of the correctly classified points and the overall is the reliability of the map (Chuvieco, 1995).

3. Results

In the 2015 exercise, geo-referenced information was sampled for 25,392 plots, being 18,157 for the PV cycle and 7,235 for the OI cycle (Table 2).

Table 2: *Plots sampled during 2015*

Cycle	Corn	Kidney Bean	Sorghum	Wheat
PV	12,690	2,528	2,511	428
OI	3,450	807	1,020	1,958
Total	16,140	3,335	3,531	2,386

About the separability analysis, the "other crop", "forest", "soil" and "protected agriculture" are easily distinguished form the crops of interest. Separability between the crops of interest defines that it is easily identifiable from sorghum, sorghum from wheat and kidney bean from sorghum. In order to validate the sowed surface for the crops of interest, 3% of the points of each one were used. They were 2% more than those stated by Congalton (1988b) in Chuvieco (1995).

Based on the fieldwork information, agriculture coverage, satellite images and these processed, the spatial distribution of the sowed surface of the crops of interest is presented. Specifically about the PV cycle, corn is present almost in all the agricultural area of the country, principally in the Transversal Volcanic System, meanwhile kidney bean has is more present in Zacatecas and Durango, and sorghum and wheat are dominant principally in Tamaulipas, Sinaloa and Guanajuato.

This evaluation assigned correctly 484 of 542 GPS points (Table 3). Corn and sorghum were confused, because in their development stage they have a similar spectral behavior. Also corn and kidney bean were confused, that is because, especially in the central and southern part of the country, they are sowed associated. Other factor that affect is the continuous presence of clouds. The overall reliability of this cycle is of 89.3%

Table 3: *Confusion matrix for the PV cycle*

	Map class	Corn	Kidney bean	Sorghum	Wheat	Total	User's accuracy	Commission error
Field class	Corn	366.0	21.0	9	0	396.0	92.4	7.6
	Kidney bean	4.0	50.0	0	1	55.0	90.9	9.1
	Sorghum	15.0	3	61.0		79.0	77.2	22.8
	Wheat	3	2	0.0	7.0	12.0	58.3	41.7
	Total	388.0	76.0	70.0	8.0	**542.0**		
	Producer's accuracy	94.3	65.8	87.1	87.5			
	Omission error	5.7	34.2	12.9	12.5			

For the OI cycle, corn was located principally in Sinaloa, wheat in Sonora and Baja California, sorghum in Tamaulipas and kidney bean is relatively low in presence.

The classification for the OI cycle assigned correctly 201 GPS points of the 215 of the total (Table 4). As happened in the PV cycle, corn still presents confusion with kidney bean and sorghum. The reliability of this classification was of 93.5%. This is because most of the agriculture is of irrigation regime and also this classification has improved because there were fewer clouds in the images.

Table 4: *Confusion matrix for the OI cycle*

	Map class	Corn	Kidney bean	Sorghum	Wheat	Total	User's accuracy	Commission error
Field class	Corn	98.0	1.0	2	0	101.0	97.0	3.0
	Kidney bean	5.0	20.0	0	0	25.0	80.0	20.0
	Sorghum	3.0	0	30.0	2	35.0	85.7	14.3
	Wheat	0	1	0.0	53.0	54.0	98.1	1.9
	Total	106.0	22.0	32.0	55.0	**215.0**		
	Producer's accuracy	92.5	90.9	93.8	96.4			
	Omission error	7.5	9.1	6.3	3.6			

Based on the analysis of these results, the surfaces inTable5 were obtained. It can be appreciated that the PV cycle has more sowed surface, since all the country is benefited by the rainy season, from May to August.

Table5: *Estimatedsowedsurface*

Cycle	Sowed Surface (Millions hectares)			
	Corn	Kidney Bean	Sorghum	Wheat
PV	5.0	1.1	1.2	0.2
OI	0.9	0.2	0.7	0.7

4. Discussion and Conclusions

The results obtained by the modified MARS method and those obtained by a traditional agriculture surface estimation are presented inTable 6. In the PV cycle, the difference is 0.1 ha (1.3%) and for the OI cycle is 0.2 ha (8%).

Table 6: *Estimated surfaces by both methods*

	Surface (Million Ha)	
Method	PV	OI
Modified MARS	7.5	2.5
Traditional	7.4	2.3
Difference	0.1	0.2

Both methods were different in the estimated amount of surfaces. These differences may have being caused by the physical configuration of the terrain or the specific parameters the sensor had at the moment the image was taken.

The use of a method supported by RS and Statistics provides various benefits to the agricultural land estimation, such as the possibility of performing it on big surfaces. Also, due to the stock images, comparisons along time can be done. Also, the possibility of coordination between the dates of the fieldwork and capture of the satellite image allow a more detailed estimation. In this scenario, the satellite information will match to that reported in the RAW, so some specific variations can be ignored.Applying specific statistical methods, the sampling process presents a concrete focus. In this way, integrating the RAW information will provide important improvements to the sample design.

It is important to establish that merging the benefits of the incorporation of these two disciplines, big optimizations of economical and human resources will be achieved. Although, it has to be considered that this innovations is in a development stage and it has to be subject to different evaluation process.

REFERENCES

Aguilar-Rivera N., Galindo G., Fortanelli J. (2012) Evaluación agroindustrial del cultivo de caña de azúcar (*Saccharumofficinarum L.)* mediante imágenes SPOT 5 HRV en la Huasteca México. Revista de la Facultad de Agronomía, La Plata 111 (2), 64-74.

Ambrosio L., Alonso, R. (1993) Estimación de superficies cultivadas por muestreo de áreas y teledetección. Estadística Española 35(132): 91.103

Chuvieco, E., (1995) Fundamentos de Teledetección espacial. RIALP S.A. 2Ed. Madrid, España, 242-390.

Deppe, F. (1998) Forest Area Estimation Using Sample Surveys and Landsat MSS and TM Data. Photogrammetric Engineering & Remote Sensing. 64 (4), April 1998. Pp. 285 – 292.

Fernández F., Gepts P., López M. (1986) Etapas de desarrollo de la planta de frijol común (*Phaseolusvulgaris L*.). Centro Internacional de Agricultura Tropical, CIAT, 7-28

Gallego J. (1995) Sampling frames of square segments. Ispra, 4-44.

Gómez D. M., Barredo, J. (2005) Sistemas de Información Geografía y Evaluación Multicriterio aplicados al Ordenamiento del Territorio. Edit. Ra-Ma. Madrid, 276 p.

Valdez J. R., González M. J., De los Santos H. M. (2006) Estimación de cobertura arbórea mediante imágenes satelitales multiespectrales de alta resolución. Agrociencia 40, 383-394.