MAIZE CROP YIELD ESTIMATION WITH REMOTE SENSING AND EMPIRICAL MODELS

Yolanda. M. Fernandez-Ordoñez[1], Member IEEE, J. Soria-Ruiz[2], Member IEEE

[1]Geomatics Laboratory, Colegio de Postgraduados, Montecillo, Mexico.
[2]Geomatics and Agroclimatology Laboratory. Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias, Zinacantepec, Mexico.

ABSTRACT

The abstract should appear at the top of the left-hand column Crop yield estimation are topics of interest in Latin-American countries, for farmers and government officers responsible of managing agricultural national policies. Besides, modern remote sensing methodologies to obtain these predictions represent important steps towards attaining the goals of precision agriculture for the 21st century. Digital data from satellite images analyzed jointly with crop modelling parameters provide information that enables crop yield estimation. The objective of this study was the estimation of yield and total volume of maize production using Spot-5 satellite images and empirical models. These models expressed a) yield (Y) as a function of LAI, and b) yield as a function of NDVI. To determine the efficiency degree of the calculated predictions at the flowering stage of the crop, yield sampling was done at the physiological maturity stage in pilot plots. Regarding yield prediction in the flowering stage, the models Y = f (LAI) reported a value of 5.96 ton.ha-1 and the model Y = f (NDVI) a value of 5.04 ton.ha-1 was obtained. These data represent 114% and 97% respectively of the true yield recorded on the field. The models are specific to the maize crop and the cultivated plots location, and that the forecasts can be acceptably accurate provided the sown areas are precisely determined

Index Terms— Maize crop, yield estimation, NDVI, LAI.

1. INTRODUCTION

In many Latin-American countries, whose economy depends in part on crop production, mainly of maize, early harvest estimations are a basic requirement to generate national statistics in order to evaluate supply and demand of grain for domestic consumption. Optical remote sensing techniques are well suited for agricultural applications, because the techniques are able to provide information on the actual status of crops at different growth stages via spectral signatures and at lower costs than extensive on-the-field surveying. Yield prediction ahead of harvest time is associated with the capacity not only to identify the crops of interest, but also to determine values for agronomic variables such as maturity, population density, vigour, disease, and weed infestation. These agronomic variables can be used as yield indicators upon which crop growth models are based. Accurate crop yield estimation therefore relies on the availability and quality of actual crop status data. During the growing stage of a crop, the Leaf Area Index (LAI) is an important variable in the modelling process and it constitutes a main factor in the determination of the reflectance values of a crop in an image. The LAI is used to correct the effects of perturbation factors in the relation crop reflectance–crop characteristics. Both the LAI and the Normalized Difference Vegetation Indices (NDVI) are related to crop vigour and to biomass production. There is recognition that an integral part of predicting yield lies in the accurate identification of growing sites and measurements of crop sown area, prior to using the various vegetation indices methods which have been proposed to measure and predict crop growth and yield [1]. The heterogeneity of corn-growing conditions in many countries makes accurate predictions of yield ahead of harvest time difficult. Such predictions are needed by the government to estimate, ahead of harvest time, the amount of staple crops such as corn that must be imported to meet expected domestic shortfalls on a yearly basis [2]. Due to the generally small size of the cultivated plots in Central Mexico originating from the land tenure structure known as minifundio, the National Research Institute for Agriculture (INIFAP) initiated since 2002, a project of yield production estimation using satellite images, together with continuous sampling in producing plots to measure variables related to yield. The aforementioned project is becoming increasingly relevant by the known fact that rich agricultural land in Mexico is quickly receding due to the pressure exerted by urbanization, industrialization, and rapid population growth. Earlier research has revealed the need to obtain better spectral signatures in order to be able to accurately determine the extent and spatial distribution of crop and vegetation species for any current agronomic cycle.

Crop monitoring and yield prediction ahead of harvest time is an issue for many countries. Studies entail the application of crop growth models and crop yield models. However, there is a large variety of models and there are still many problems associated with them. Studies on understanding crop yield models are exemplified by the Large Area Crop Inventory Experiment LACIE [6] and AgRISTARS [3]. Remote sensing has been used to furnish input data for the models. Spectral vegetation indices have been developed in an attempt to correct for atmospheric and soil spectral effects on remotely-sensed data [4]. The spectral vegetation index (VI), which is a measure of the total green biomass at any given time, has been related to grain yields as reported in [5]. NDVI data have been used extensively in vegetation monitoring, crop yield assessment and forecasting. The NDVI has been shown to be highly correlated with the final grain yield of cereals around the time of maximum green leaf biomass development [6]. For example, the peak value of the NDVI profile at the time of peak occurrence is analogous to the maximum leaf area index (LAI). Crop production forecasting using satellite images over large areas has been performed in different places for different crops using NDVI and time-series composites. Thus these satellite images represent a useful tool in the analysis of crop-related phenological phenomena. Successful crop yield modelling for maize has been reported for the Corn Belt in the U.S.A. [7]. Soria-Ruiz and Fernandez-Ordonez [2 op cit] presented a methodology for the estimation of corn yield in central Mexico. This method is based on the multi-temporal analysis of NOAA-AVHRR satellite images, NDVI, leaf area indices (LAI) and degree-days for yield prediction. The objectives of the research presented here are (a) to determine the spatial distribution and the cultivated surface from Spot-5 satellite imagery, and (b) to predict yield from LAI and NDVI data and empirical models adjusted in situ.

2. MATERIALS AND METHODS

A. Study Area. This work was carried out in 2010 and 2011 in the area encompassed by the Rural Development District DDR-05 in Atlacomulco, State of Mexico, Mexico (Fig.1).

B. Sampling Sites. All training samples were chosen to belong to the appropriate class; the number of samples should be large enough to provide a statistically valid class model and there should be sufficient variety in the samples to represent the full variety of appearance exhibited by the class. Ideally, samples should be selected from more than one area in the image. This sampling activity in 15 pilot parcels was carried out in field visits and the spatial plot location was supported by GPS; the reference points were located on the RGB Spot-5 composite (Table 1).

C. Spatial distribution of maize crop. Full scene Spot-5 images were used in 2010 and 2011 to determine surfaces of maize through spectral signatures. A supervised classification was carried out with the maximum-likelihood classification algorithm. It is generally accepted as a better method, since it allows for class variance provided by the maximum-likelihood rule (Maxlike).

Fig. 1. Study area in Rural Development District DDR-05 in Atlacomulco, State of Mexico, Mexico

Table 1. Selected sample plots sites in the study area.

Plot	Municipality	Geographic location (Central point)
1	Jiquipilco	19^{0} 36’ 10’’ N & 99^{0} 43’ 42’’ W
2	Ixtlahuaca	19^{0} 36’ 47’’ N & 99^{0} 43’ 25’’ W
3	Jiquipilco	19^{0} 37’ 23’’ N & 99^{0} 43’ 11’’ W
4	Jiquipilco	19^{0} 37’ 36’’ N & 99^{0} 42’ 44’’ W
5	San Felipe	19^{0} 44’ 40’’ N & 99^{0} 55’ 30’’ W
6	San Felipe	19^{0} 44’ 36’’ N & 99^{0} 55’ 15’’ W
7	Jocotitlán	19^{0} 44’ 28’’ N & 99^{0} 55’ 40’’ W
8	Jocotitlán	19^{0} 44’ 28’’ N & 99^{0} 55’ 47’’ W
9	Atlacomulco	19^{0} 47’ 40’’ N & 99^{0} 56’ 02’’ W
10	Atlacomulco	19^{0} 47’ 33’’ N & 99^{0} 55’ 47’’ W
11	Jocotitlán	19^{0} 47’ 22’’ N & 99^{0} 55’ 38’’ W
12	Acambay	19^{0} 57’ 11’’ N & 99^{0} 51’ 05’’ W
13	Acambay	19^{0} 56’ 50’’ N & 99^{0} 51’ 18’’ W
14	Acambay	19^{0} 56’ 31’’ N & 99^{0} 52’ 05’’ W
15	Acambay	19^{0} 56’ 19’’ N & 99^{0} 52’ 10’’ W

D. Leaf Area Indices Sampling. In the sampled plots, which were used as training sites, leaf area index (LAI) measurements were taken and recorded with high precision equipment: AccuPAR Linear Par Ceptometer, Model PAR-80®. The LAI samplings were taken at 15 day intervals in the 2010; data collection started during the vegetative development stage of the crop [8].

E. Yield prediction models. The prediction models were obtained from the plots in the first year (2010). LAI data was measured directly at these plots. NDVI data was obtained for the same plots through the Spot-5 images. This process was repeated during the second year (2011). All the data for the second year were used to calibrate the prediction models generated in the previous year.

3. RESULTS AND DISCUSSION

A. Maize crop area. The cultivated area was obtained from two scenes of Spot-5 images, one scene of July 2010 and other of August 2011. Supported by data from the training areas (pilot plots) and under a supervised classification technique, the spatial distribution maps for 2010 and 2011 were obtained. Fig. 2 shows these maps.

(a) Scene of July/2010 (b) Scene of August/2011

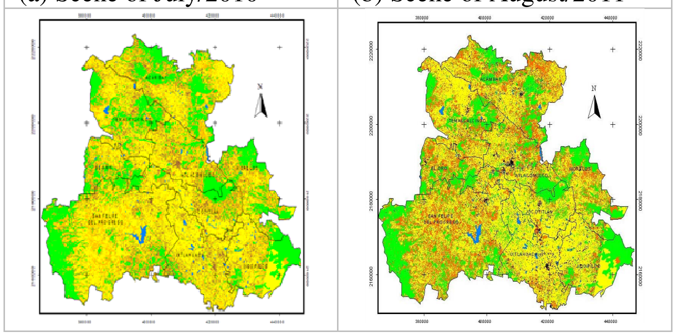

Fig. 2. Spatial distribution of maize in 2010 (a) and 2011 (b), obtained from Spot-5 images.

The cultivated areas surfaces obtained were 141,726 ha in 2010 and 143,498 ha in 2011. There are differences between the cultivated maize areas obtained from satellite images versus officially data reported by the local ministry of agriculture (SEDAGRO). The largest difference occurs in the second year (Table 2).

Table 2. Surface of maize obtained by SEDAGRO and SPOT-5 images.

SOURCE (data)	MAIZE SUFACES (hectare)	
	2010	2011
Local ministry of agriculture (SEDAGRO)	149,243	172,449
Spot-5	141,726	143,498

B. LAI and NDVI data. Table 3 shows the LAI and NDVI values corresponding to the phenological flowering stage of maize during the 2011. It can be seen that for all pilot plots a mean value of 3.96 with a standard deviation of 0.3 was obtained; for the NDVI the mean value was 0.41 with a standard deviation of 0.03.

C. Yield estimation. With the LAI and NDVI data obtained from the pilot plots, the prediction models derived for the same zone in the previous year were fed. With this information, the estimation of crop yields were: a) yield as a function of LAI and b) yield as a function of NDVI.

Substituting the average value of LAI = 3.96 with $\sigma \pm 0.3$ (3.66 – 4.26) corresponding to the flowering stage in the model gives: $Y = 0.3178\,(X)^2 - 0.7796\,(X) + 4.0662$

where: X = LAI average corresponding to the flowering stage of 2011, Y = Yield (ton.ha^{-1}).

Table 3. LAI and NDVI in flowering stage for maize.

PLOT	LAI	NDVI
1	4.81	0.43
2	3.70	0.40
3	4.38	0.39
4	3.41	0.41
5	3.91	0.44
6	4.01	0.41
7	3.98	0.43
8	3.73	0.49
9	4.22	0.42
10	3.93	0.41
11	3.89	0.43
12	3.98	0.43
13	3.89	0.41
14	3.78	0.33
15	3.81	0.37
$\sum$	59.43	6.2
$\overline{X}$	3.96	0.41
St. Deviation (σ)	0.3	0.03

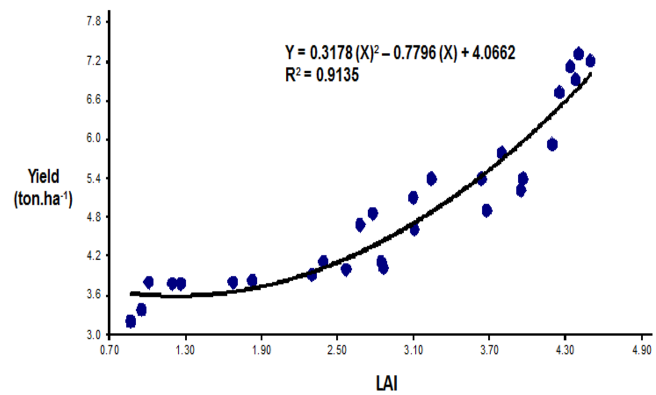

Fig. 3. Yield estimation model for maize yield using LAI data in the flowering stage.

1) Yield as a function of the LAI. (Fig. 3).

If LAI = 3.96 with standard deviation = ± 0.3 (3.66 – 4.26)

Then: $Y = 0.3178\,(3.96)^2 - 0.7796\,(3.96) + 4.0662$

$Y = 4.98 - 3.08 + 4.0662$. Therefore: Y = 5.96 ton.ha^{-1}

2) Yield as a function of the NDVI (Fig. 4).

By substituting the mean value of the NDVI = 0.41 with σ = ± 0.03 (0.38 - 0.44) in the prediction model:

$Y = -13.098\,(X)^2 + 23.302\,(X) - 2.3006$

Where: X = average NDVI corresponding to the maize flowering stage and Y = Yield (ton.ha^{-1}).

By substituting the mean value of the NDVI = 0.41 with σ = ± 0.03 (0.38 - 0.44) in the prediction model:

$Y = -13.098\,(X)^2 + 23.302\,(X) - 2.3006$

Where: X = average NDVI corresponding to the maize flowering stage and Y = Yield (ton.ha^{-1}).

If NDVI = 0.41, with standard deviation = ± 0.03 (0.38 - 0.44). Then: $Y = -13.098\,(0.41)^2 + 23.302\,(0.41) - 2.3006$

Therefore Y = 5.04 ton.ha^{-1}

D. Accuracy of estimation. In order to determine the accuracy (effectiveness) of calculated predictions in the flowering stage of the maize crops in 2011, a sampling of yield in the physiological maturity stage (14% of grain humidity) was carried out in the pilot plots (Table 5).

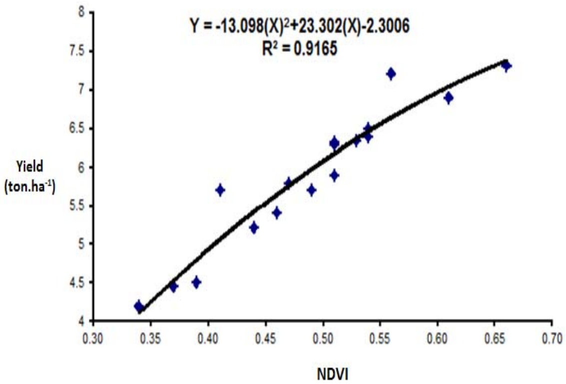

Fig. 4. Yield estimation model for maize using NDVI data in the flowering stage

Table 4. Yield estimation and production volume by two sources for two years.

SOURCE	YIELD ($ton.ha^{-1}$)		PRODUCTION VOLUME (thousands of ton)	
	2010	2011	2010	2011
Local ministry of agriculture (SEDAGRO)	3.56	3.3	531.3	569.08
Yield estimation models: $Y= 0.3178(LAI)^2 – 0.7796 (LAI) + 4.0662$	5.09	5.96	721.4	855.2
$Y=-13.098(NDVI)^2+ 23.302 (NDVI) – 2.3006$		5.04		723.2

Table 5. Sample yield values for maize pilot plots in 2011.

PLOT	OBSERVED YIELD ($ton.ha^{-1}$)
1	5.34
2	5.23
3	4.83
4	5.38
5	6.14
6	5.31
7	4.89
8	5.23
9	5.87
10	5.73
11	5.34
12	4.61
13	4.86
14	4.78
15	4.82
$\overline{X}$	5.22

The general average observed value in table 5, when compared with the calculated predictions, provides a measure of the effectiveness of predictions calculated at the flowering stage. This is shown in (Table 6). The Y = f (IAF) model gave an over-estimation of 14%, while the model Y = f (NDVI) gave a 97% effectiveness, when compared with the real mean value obtained on the field.

Table 6. Effectiveness of yield prediction with two models in the second year.

PREDICTION MODEL	PREDICTED YIELD ($ton.ha^{-1}$)	ACCURACY (%)
$Y = 0.3178 (LAI)^2 – 0.7796 (LAI) + 4.0662$	5.96	114
$Y = -13.098 (NDVI)^2 + 23.302 (NDVI) – 2.3006$	5.04	97
Average yield with field sampling:	5.22	----

4. CONCLUSIONS

Yield estimation at the flowering stages of maize obtained in the agricultural cycles of 2010 and 2011, show high accuracy with respect to the observed true field values in the study area. The LAI derived prediction model over-estimated yield by 14%, whereas the NDVI attained 97% of accuracy. The LAI derived model behaviour is due to variation in field data collection, which occurred at different hours during the day. This affects directly the incidence angle of sun light in the plant canopy, and thus the foliar response measured by the instrument. This methodology is currently being used in the State of Mexico to estimate the yield and volume of corn production. These estimates are used by government officers to implement grain imports policies in relation to domestic demand.

5. REFERENCES

[1] Gommes, R., “Agro-meteorological Models and Remote Sensing for Crop Monitoring and Forecasting in Asia and the Pacific,” *Report of the FAO Asia-Pacific Conference, Annex VII APDC/01/3*, FAO-SDRN. 2001.

[2] Soria-Ruiz, J. and Y. Fernandez-Ordonez, “Prediction of corn yields using vegetation indices from NOAA-AVHRR satellite images and degree-days.” *Geocarto International*, Vol. 18, No. 4, p.33-42, 2003.

[3] Ferguson, M. C., “Evaluation of trends in yield models: Agristars supporting research,” SR J1-04157, JSC-17428, December, 1982.

[4] Broge, N.H. and J. V. Mortensen. “Deriving green crop area index and canopy chlorophyll density of winter wheat from spectral reflectance data,” *Remote Sensing of Environment*, Vol. 81, no. 1, pp. 45-57, 2002.

[5] Kalubarme, M.H., M. B. Potdar, K. R. Manjunath, R. K. Mahey and S. S. Siddhu, “Growth profile based crop yield models: a case study of large area wheat yield modelling and its extendibility using atmospheric corrected NOAA AVHRR data,” *International Journal of Remote Sensing,* Vol. 24 No. 10, p 2037 – 2054, 2003.

[6] Tucker, C. J., Holven, B. N., Elgin, J. A., and McMurtrey, J. E., “Remote sensing of total dry matter accumulation in winter wheat,” *Remote Sensing of Environment*, 11, 171–189, 1991.

[7] Hayas, M.J. and W.L. Decker.”Using NOAA AVHRR data to estimate maize production in the United States Corn Belt,” *Remote Sensing,* Vol. 17 No. 16, 3189-3200, 1996.

[8] Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., & Leavitt, B., “Remote estimation of leaf area index and green leaf biomass in maize canopies,” *Geophysical Research Letters*, Vol. 30 No. 5, 2003.