INT. J. REMOTE SENSING, 1997, VOL. 18, NO. 2, 467–470

# Remote sensing and agricultural statistics: crop area estimation in north-eastern Spain through diachronic Landsat TM and ground sample data

F. GONZÁLEZ-ALONSO, J. M. CUEVAS

Remote Sensing Unit, National Institute for Agricultural and Food Research and Technology (INIA), P.B. 8111, Madrid 28080, Spain

R. ARBIOL and X. BAULIES

Remote Sensing Service, Cartographic Institute of Catalonia (ICC), Parc de Montjuïc, Barcelona 08038, Spain

(*Received 21 May 1996; in final form 11 September 1996*)

**Abstract.** Landsat Thematic Maper (TM) imagery was used to estimate crop area using the regression method. Two different kind of estimates have been made; one using a Landsat TM image of the same year as the ground sample and the other using an image taken the year before. The results show that if the rate of land cover change is not great the use of an image of the year previous to the ground survey is a way of reducing the costs of the images acquisition without a great loss in the accuracy of the area estimation of the main crops compared to the conventional regression method.

## 1. Introduction

Objective methods of producing crop statistics based on frame and ground sampling techniques are widely used today (Gallego 1995). The combining of ground sampling with the results obtained from the supervised classification of fine spatial resolution satellite images, more or less of the same date, by the use of the regression estimator (Cochran 1977) considerably improves the estimation accuracy of the crop areas.

This methodology is performed in many regions and continents under different agrarian conditions (Allen and Hanuschak 1988, Delince 1990, González-Alonso *et al.* 1991).

The objective of this study has been to analyse and compare the results obtained in the crop area estimation in the cases of using an image contemporaneous with the ground survey and one image taken the year before.

## 2. Methodology

The methodology for estimating crop areas using satellite images and ground sample data is based on the use of regression estimator (Allen and Hanuschak 1988, Delince 1990, González-Alonso *et al.* 1991; González-Alonso and Cuevas 1993 a, 1993 b).

To estimate crop areas using the direct expansion method only the ground data sample is used, while in the regression method it is necessary to fit the regression

model:

$$Y_i = b_0 + b_1 X_i \qquad i = 1, \ldots, n \tag{1}$$

Where $Y_i$ is the percentage of the area occupied by the crop of interest in the segment $i$ (obtained from the ground survey) $X_i$ the percentage of the area occupied by the crop of interest in the segment $i$ (calculated from the classification of the pixels corresponding to the segment $i$), $b_0, b_1$ are the simple regression coefficients and $n$ the number of segments in the sample.

If $\bar{Y}$ is the sample mean of $Y_i$, in the direct expansion method the area occupied by the crop of interest, $T$, will be

$$T = D\bar{Y} \tag{2}$$

$D$ being the area of the study zone.

In the estimation by the regression method the surface occupied by the crop of interest, $T_{reg}$, will be

$$T_{reg} = D\bar{Y}_{reg} \tag{3}$$

Being $\bar{Y}_{reg} = \bar{Y} + b_1(\bar{X}_{pob} - \bar{X})$.

Where $b_1$ is the simple regression coefficient previously calculated, $\bar{X}_{pob}$ the proportion of pixels classified as the crop of interest, obtained from the supervised classification of all the pixels in the study zone, and $\bar{X}$ the sample mean of $X_i$.

In these conditions the variance of the total crop area estimation by the regression method is

$$V(T_{reg}) = V(T)(1 - r^2) \tag{4}$$

$V(T)$ being the variance of the total crop area estimation by the direct expansion method and $r$ the simple correlation coefficient between $Y$ and $X$. Logically, the greater the value of $r$, the lower the variance of the total crop area estimation by the regression method.

In order to determine the success associated with the regression estimator, its relative efficiency ($RE$) is calculated (Allen and Hanuschack 1988). The $RE$ of the regression estimator compared to the direct expansion estimator is defined as the ratio of their respective variances

$$RE = V(T)/V(T_{reg}) = 1/(1 - r^2) \tag{5}$$

If $RE = 2$ then the same precision would have been obtained if the ground data sample size had been doubled and the satellite image had not been used.

The interest of the regression estimator consists of the correction of the estimated mean of a variable $Y$, known from a sample of size $n$, as a function of the results obtained from an auxiliary variable $X$ that has some degree of linear correlation with the interest variable $Y$. In our case, for a given crop, for each segment in the sample, $Y$ is the proportion occupied by the crop as deduced from the digitization of the ground survey and $X$ is the proportion of pixels of the satellite image classified as being of the given crop.

In the traditional application of the regression method the variable $X$ is deduced form the supervised classification of an image contemporaneous with the ground survey. If the analysis of the ground surveys of two consecutive years shows that the structure of land uses and the kind of crops present in the study zone have not suffered very severe modifications, using the results of the classification of an image

from the year before, or previous years, could be considered. This negates the necessity to acquire the complete satellite image cover for the study area every year decreasing in an important way the costs of the images acquisition involved using the regression method in large regions.

In these conditions the diachronic estimator of the regression mean for the year $n$ will be

$$\bar{Y}_{reg,n} = \bar{Y}_n + b_1(\bar{X}_{n-1} - \bar{x}_{n-1}) \tag{6}$$

Being $\bar{X}_{n-1}$ and $\bar{x}_{n-1}$ obtained from the classification of the satellite image of the year $n-1$.

## 3. Application and results

The pilot area where this methodology has been tested is formed by the strata 1 (irrigated herbaceous crops) and 2 (unirrigated herbaceous crops) of the Lleida province (Autonomous Community of Catalonia, northeastern Spain). The study area surface is 293 511 ha. The ground sample is composed of 54 square segments of 700 m by 700 m (49 ha), which were ground surveyed in June 1994. The sampling rate is about 1 per cent. The used Landsat-5 Thematic Mapper (TM) images were recorded in June 1993 and 1994.

In order to analyse the results of this methodology, barley, the most abundant crop in the study area was chosen. According to the Landsat TM images supervised classifications it occupied 39·44 per cent of the study area in 1993 and 44·33 per cent in 1994.

Table 1 shows the results obtained for the barley applying the direct expansion method (using only the 1994 ground survey), the conventional regression method (using the 1994 ground survey and the classification of the 1994 Landsat TM image) and the diachronic regression method (using the 1994 ground survey and the classification of the 1993 Landsat TM image).

Both regression methods produce a very similar reduction of the variance, being consequently the coefficients of variation quite similar, although the estimation by the conventional regression method is slightly better, 4·09 per cent against 4·39 per cent. The relative efficiency of the conventional regression method is higher than the diachronic regression method, 9·60 against 9·26, but the difference is quite small.

The use of the 1994 ground survey combined with the results of the classification of the 1993 Landsat TM image produces a reduction of 9·26 times in the variance of the estimation of the barley area without any additional cost, as the 1993 image is supposed to have been previously acquired and classified.

Table 1. Obtained results for the barley with the different methods of area estimation.

| | SUR | STE | CV | $r^2$ | *RE* | $b_0$ | $b_1$ |
|---|---|---|---|---|---|---|---|
| Direct expansion (1994 ground survey) | 113 235 | 13 576 | 11·98 | | | | |
| Regression (1994 Landsat TM image) | 107 045 | 4379 | 4·09 | 0·895 | 9·60 | 0·023 | 0·776 |
| Diachronic regression (1993 Landsat TM image) | 101 550 | 4460 | 4·39 | 0·892 | 9·26 | 0·032 | 0·860 |

SUR = surface in hectares, STE = standard error, CV = coefficient of variation, $r^2$ = coefficient of determination, *RE* = relative efficiency, $b_0$ and $b_1$ = regression coefficients.

## 4. Conclusion

The application of the regression estimator using classified images of the year before the ground survey allows obtaining reductions in the variance of the crop area estimation very similar to those obtained fusing contemporaneous imagery in the regions where the structure of land uses and the kind of crops do not suffer extreme changes from one year to the following. This methodology overcomes the necessity of acquiring satellite image cover each year and consequently produces reductions in the costs involved in the use of the regression method in the estimation of crop areas.

## References

ALLEN, J. D., and HANUSCHAK, G. D., 1988, The remote sensing applications program of the National Agricultural Statistics Service: 1980–1987. NASS Staff Report SRB-88-08 (Washington, D.C.: National Agricultural Statistical Service).

COCHRAN, W. G., 1977, *Sampling Techniques* (New York: Wiley).

DELINCE, J., 1990, Un premier bilan de l'action 1 'Inventaires Régionaux' du project agriculture aprés deux années d'activité. *Proceedings of the Application of Remote Sensing to Agricultural Statistics*, EUR 12581 EN (Luxembourg: Commission of the European Communities, Joint Research Centre), pp. 53–58.

GALLEGO, F. J., 1995, *Sampling frames of square segments.* Joint Research Centre. European Commission. EUR 16317EN.

GONZÁLEZ-ALONSO, F., LÓPEZ, S., and CUEVAS, J. M., 1991, Comparing two methodologies for crop area estimation in Spain using Landsat TM images and ground-gathered data. *Remote Sensing of Environment*, **35**, 29–35.

GONZÁLEZ-ALONSO, F., and CUEVAS, J. M., 1993 a, Comparing Landsat TM and MSS images for crop area estimation in Spain. *Proceedings of 12th EARSeL Symposium, Remote Sensing for Monitoring the Changing Environment of Europe, Eger, Hungary, 8–11 September 1992* (Rotterdam: Balkema), pp. 219–221.

GONZÁLEZ-ALONSO, F., and CUEVAS, J. M., 1993 b, Remote sensing and agricultural statistics: crop area estimation through regression estimators and confusion matrices. *International Journal of Remote Sensing*, **14**, 1215–1219.