

# Estadísticas agrarias y teledetección. Una nueva forma de aplicar el estimador de regresión

F. González y J. M. Cuevas

Sección de Teledetección Agraria

Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria

Apdo. 8111. 28080 Madrid

## RESUMEN

La estimación de superficies cultivadas a partir de encuestas de campo e imágenes de satélite va a constituir el método normal de realización de las estadísticas agrarias en los países del sur de la Comunidad Europea.

En el presente trabajo se propone una metodología alternativa en la aplicación del estimador de regresión que parte de unos supuestos más realistas y necesita un esfuerzo de cálculo menor.

**PALABRAS CLAVE:** estadísticas agrarias, Landsat TM, estimador de regresión.

## ABSTRACT

The estimation of crop areas performed by integration of sampling techniques and satellite images is the method utilized for doing the agricultural statistics in the countries of the south of Europe.

In this paper it is presented a new methodology for the application of the regression estimator that departs from more realistic suppositions and needs lesser computation effort.

**KEY WORDS:** agrarian statistics, Landsat TM, regression estimator.

## INTRODUCCIÓN

La estimación de superficies cultivadas a partir de la integración de encuestas de campo e imágenes de satélite de alta resolución espacial se realiza de modo operacional en Estados Unidos y Europa desde 1980 (Allen y Hanuschack, 1988; Delince, 1990).

La metodología utilizada tradicionalmente en este tipo de aplicaciones está basada en el empleo del estimador de regresión (Cochran, 1977; Ozga et al., 1977; Moro et al., 1984; Commission de Communautés Européenes, 1987; González et al., 1988; González et al., 1990; González et al., 1991).

El uso tradicional del estimador por regresión supone realizar la clasificación supervisada del área en estudio a partir de una imagen de satélite de alta resolución espacial, por ejemplo Landsat TM o SPOT HRV. La realización de este proceso suele ser bastante costosa desde el punto de vista del esfuerzo de cálculo y no está además exenta de unos errores debidos, fundamentalmente, a la existencia de píxeles heterogéneos en sus características espectrales.

Así pues, el proceso de clasificación supervisada de la imagen es una simplificación, ya que en la realidad pueden existir muchos píxeles que no correspondan a un único tipo de cubierta sino a varios.

El objetivo fundamental de este trabajo consiste en asignar a cada píxel una determinada probabilidad de pertenecer a cada tipo de cubierta y utilizar estas probabilidades en el cálculo de las propor-

ciones que corresponden a cada tipo de cultivo o cubierta, tanto a nivel de unidad de muestreo o segmento como a nivel de toda el área en estudio.

La realización de un estudio comparativo respecto a la aplicación tradicional del método de regresión permitirá extraer las conclusiones prácticas oportunas respecto al interés de utilizar la novedad metodológica que se propone.

## MATERIAL Y MÉTODOS

### Metodología utilizada

La metodología para realizar la estimación de superficies cultivadas a partir de imágenes de satélite y encuestas de campo se basa en la utilización del estimador de regresión (Cochran, 1977).

En González-Alonso et al. (1991) se encuentra una descripción detallada de dicha metodología, así como las fórmulas empleadas para realizar la estimación de superficies cultivadas por el método de expansión directa y por el método de regresión.

Para realizar la estimación de superficies cultivadas por el método de expansión directa solamente se consideran los datos obtenidos mediante una encuesta de segmentos en el campo, mientras que para realizar la estimación de superficies por el método de regresión es necesario ajustar el modelo de regresión:

$$Y_i = b_0 + b_1 X_i \quad i=1, n \quad (1)$$

Siendo

n = Número de segmentos encuestados.

$Y_i$  = Porcentaje determinado por digitalización que ocupa el cultivo de interés en el segmento  $i$ .

$X_i$  = Porcentaje que ocupa el cultivo de interés en el segmento  $i$  calculado a partir de la clasificación de píxeles que corresponden al segmento  $i$ .

$b_0, b_1$  = Coeficientes de regresión.

Una medida de la calidad del ajuste de regresión existente entre  $Y_i$  y  $X_i$  viene dada por el coeficiente de determinación:

$$R^2 = \frac{\left[ \sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X}) \right]^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2 \sum_{i=1}^n (X_i - \bar{X})^2} \quad (2)$$

Siendo  $Y$  y  $X$  las medias muestrales de  $Y_i$  y  $X_i$  respectivamente.

En el método de expansión directa la superficie total ocupada será:

$$T = D\bar{Y} \quad (3)$$

siendo  $D$  la superficie en hectáreas del área en estudio. La varianza de esta estimación será:

$$V(T) = D^2 V(\bar{Y}) \quad (4)$$

En el método de estimación por regresión la superficie total ocupada por el cultivo de interés,  $T_{reg}$  será:

$$T_{reg} = D\bar{Y}_{reg} \quad (5)$$

$$\text{siendo } \bar{Y}_{reg} = \bar{Y} + b_1(\bar{X}_{pob} - \bar{X}) \quad (6)$$

donde  $b_1$  es el coeficiente de regresión determinado anteriormente y  $X_{pob}$  la proporción poblacional obtenida para el cultivo de interés a partir de la clasificación de todos los píxeles que componen el área en estudio. En estas condiciones,

$$V(T_{reg}) = V(T) (1 - R^2) \quad (7)$$

siendo  $R^2$  la estimación del coeficiente de determinación entre  $Y_i$  y  $X_i$ .

La novedad metodológica que se propone en este trabajo consiste en emplear en lugar de la media muestral  $\bar{X}$  y la media poblacional  $\bar{X}_{pob}$  una nueva media muestral  $\bar{X}^*$  y una nueva media poblacional  $\bar{X}_{pob}^*$ . Así pues, en lugar de calcular la media poblacional como

$$\bar{X}_{pob} = N_k / N \quad (8)$$

donde  $N_k$  es el número de píxeles clasificados, en la clasificación supervisada de la imagen, como

pertenecientes a la clase  $k$  y  $N$  el número total de píxeles que componen el área en estudio, se calcula el estimador bayesiano de la media poblacional  $\bar{X}_{k,pob}^*$  de la forma siguiente:

$$\bar{X}_{k,pob}^* = E [ P(k/\text{datos radiométricos}) ] = \frac{1}{N} \sum_{i=1}^N P_i(k/\text{datos radiométricos del píxel } i) \quad (9)$$

Teniendo en cuenta que por el teorema de Bayes:

$$P_i(k/\text{d.r. pix. } i) = \frac{P_i(\text{d.r. pix. } i/k)}{\sum_{i=1}^{NC} P_i(\text{d.r. pix. } i/l)} \quad (10)$$

Siendo  $P_k$  la probabilidad a priori de la clase  $k$ ,  $P_i(\text{d.r. pix. } i/k)$  el valor que toma la función de densidad multinormal para la clase  $k$  en el caso del píxel  $i$  y  $NC$  el número de clases de cubierta.

De análoga forma se calcularían los valores de  $X_i^*$  para cada segmento, así como de:

$$X^* = \frac{\sum_{i=1}^n X_i^*}{n} \quad (11)$$

Así pues, en el método que se propone:

$$\bar{Y}_{reg} = \bar{Y} + b_1^* (\bar{X}^* - \bar{X}_{pob}^*) \quad (12)$$

$$V(T_{reg}) = V(T) (1 - R^{*2}) \quad (13)$$

Siendo  $b_1^*$  el coeficiente de regresión y  $R^{*2}$  la estimación del coeficiente de determinación entre  $Y_i$  y  $X_i^*$ .

### El área en estudio, inventario realizado e imagen utilizada

El área en estudio está situada al sur de la Comunidad Autónoma de Navarra y tiene una superficie aproximada de 40.000 ha. Se trata de una zona principalmente agrícola en secano, donde abundan los cereales (cebada y trigo), aunque el viñedo tiene una presencia importante.

El número de segmentos o áreas de verdad terreno investigados en el terreno fue 13. La selección de esta muestra de segmentos se hizo de un modo aleatorio y no se realizó estratificación del área en estudio. Los segmentos tuvieron la forma de un cuadrado de 700 m de lado, con una superficie de, aproximadamente, 50 ha. La intensidad del muestreo realizado fue del 1,5 %.

Los trabajos de campo se realizaron entre el 15 y el 30 de marzo de 1988, distinguiendo nueve clases de cubierta: trigo, cebada, maíz, espárrago, viñedo, barbecho, coníferas, encinares y ríos.

La imagen de satélite utilizada ha sido una imagen Landsat-TM captada el 15 de marzo de 1988.

Se ha utilizado el sistema ERAFIS (Moro et al., 1984) para realizar el proceso de la imagen.

La estimación de superficies solamente se ha realizado para los principales cultivos presentes en el área en estudio: cebada y barbecho. Para estos cultivos se han obtenido y comparado tres diferentes métodos de estimación: expansión directa, método de regresión tradicional y método de regresión propuesto.

## RESULTADOS Y DISCUSIÓN

En la Tabla 1 aparecen los resultados obtenidos en la estimación de superficies para los tres métodos ensayados: expansión directa, regresión tradicional y regresión propuesta.

De la observación de dicha tabla se desprende que los resultados obtenidos con los dos métodos de regresión ensayados son muy parecidos y en consecuencia la aplicación de uno u otro método debería estar justificada en base a su mayor simplicidad de aplicación o economía de coste. En este sentido hay que señalar que el método de regresión propuesto es más eficiente que el tradicional desde el punto de vista del esfuerzo de cálculo. Esta economía en el tiempo de cálculo se debe a que las operaciones de suma consumen menos tiempo de cálculo informático que la selección del máximo de un *array* para cada uno de los píxeles comprendidos en el área en estudio.

Así pues, si el interés en utilizar la imagen de satélite en la estimación de superficies es exclusivamente estadístico y no se está interesado en la obtención de una cartografía temática relativa a la distribución de los tipos de cubierta, el empleo del método propuesto estaría justificado y supondría una mejora en la aplicación del método de estimación por regresión.

## CONCLUSIONES

La estimación de las proporciones que ocupa cada cultivo o tipo de cubierta a partir del cálculo de probabilidades bayesianas en lugar de a partir de los resultados de un proceso de clasificación supervisada y su consideración en la aplicación del estimador por regresión, proporciona unos resultados análogos a los obtenidos mediante la aplicación tradicional del método de regresión, pero con una mayor economía de cálculo y un planteamiento más realista respecto a la homogeneidad interna

de cada píxel en relación a los tipos de cubierta que se encuentran en él.

## BIBLIOGRAFÍA

- ALLEN, J.D., HANUSCHAK, G.D. (1988), *The remote sensing applications program of the National Agricultural Statistics Service: 1980-1987*, Washington, D.C., NASS Staff Report SRB-88-08, National Agricultural Statistical Service.
- COCHRAN, W.G. (1977). *Sampling Techniques*, New York, Wiley.
- COMMISSION DES COMMUNAUTES EUROPEENNES (1987), *Programme décennal de recherche et développement pour l'application de la télédétection aux statistiques agricoles*, SP.1.87.39/FR. Ispra, Centre Commun de la Recherche.
- DELINCE, J. (1990), «Un premier bilan de l'action 1 «Inventaires Régionaux» du project agriculture après deux annés d'activité». *Proceedings of the Application of Remote Sensing to Agricultural Statistics*, EUR 12581 EN, Commission of fue European Communities, Ispra, Joint Research Centre, pp. 53-58.
- GONZALEZ-ALONSO, F., CUEVAS, J.M., MORO, J. (1988). *Aplicación del Sistema Informático ERAFIS al reconocimiento y discriminación de los usos del suelo en la Provincia de Toledo a partir de imágenes digitales MSS-Landsat 4*. Madrid, Instituto Nacional de Investigaciones Agrarias.
- GONZÁLEZ-ALONSO, F., LÓPEZ, S., CUEVAS, J.M. (1990). *Influencia de la información contenida en imágenes de satélite sobre la precisión de la estimación de la superficie cultivada de cereal en una zona piloto de la Comunidad Autónoma de Navarra*. Madrid, INIA.
- GONZÁLEZ-ALONSO, F., LÓPEZ, S., 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, pp. 29-35.
- HANUSCHAK, G., SIGMAN, R., CRAIG, M., OZGA, M., LUEBE, R., COOK, P., KLEWEND, MILLER, C. (1979). "Obtaining timely crop area estimates using ground-gathered and Landsat data», *ESCS Technical Bulletin*, No. 1609. Washington, D.C. United States Department of Agriculture.
- MORO, J., GONZÁLEZ-ALONSO, F., CUEVAS, J.M. (1984). *ERAFIS. Un sistema informático para la estimación de recursos agrícolas y forestales mediante imágenes de satélite*. Madrid, INIA.
- OZGA, M., DONOVAN, W., GLEASON, C. (1977). «An interactive system for agricultural estimates using Landsat data». En *Proc. of the 1977 Symposium of Remotely Sensed Data*, Indiana, West Lafayette, pp. 113-123.

Cultivo	Método de estimación	Superficie total	Desviación típica	Coefficiente de variación	R <sup>2</sup>	Eficiencia Relativa	Medias muestr.	Medias poblac.
Cebada	Expansión directa	12128,6	2884,8	23,78			0,3053	
	Regresión tradicional	11573,1	947,4	8,18	0,892	9,27	0,3368	0,3203
	Regresión propuesta	11445,1	936,5	8,18	0,895	9,49	0,3284	0,3107
Barbecho	Expansión directa	13753,5	2125,5	15,45			0,3462	
	Regresión tradicional	18622,6	1230,1	6,60	0,665	2,98	0,3216	0,4777
	Regresión propuesta	17830,6	1233,2	6,91	0,663	2,97	0,3248	0,4207

**Tabla 1.** Resultados obtenidos en la estimación de superficies cultivadas mediante la aplicación de los tres métodos ensayados (expansión directa, regresión tradicional y regresión propuesta).