

ANÁLISIS DE LA APLICACION DE IMAGENES TM EN INVENTARIOS FORESTALES DE UNA AREA MEDITERRANEA

R. Salvador Civil y X. Pons Fernández

CREAF. Facultat de Ciències. Universitat Autònoma de Barcelona. 08193 Bellaterra (Barcelona). Tel: 93-5812666, fax: 93-5811312, email: ibbt0@cc.uab.es.

RESUMEN.- Diferentes modelos de regresión fueron probados con la intención de relacionar los datos ofrecidos por el sensor TM de Landsat con los datos recogidos en un inventario forestal, en una zona Mediterránea. Tanto las regresiones simples, como las múltiples con valores de reflectancia de los canales originales, no dieron ajustes suficientes de cara a hacer predicciones a nivel cuantitativo. Posteriormente se crearon nuevas variables a partir de divisiones y multiplicaciones de canales, generándose un gran número de nuevos modelos. Mediante la R^2 ajustada y la C_p de Mallows se escogieron los mejores modelos, obteniéndose unos buenos ajustes. Sin embargo la posterior realización de pruebas de test con puntos adicionales puso de manifiesto su carácter ficticio. Finalmente, algunas causas fueron propuestas de cara a explicar estos resultados encontrados.

ABSTRACT.- Different regression models were tried in order to relate satellite TM data to data from a forest inventory, in a Mediterranean area. Neither simple regressions nor multiple regressions with original bands, achieved a sufficient degree of adjustment to do quantitative predictions. Posteriorly, new variables coming from multiplications and divisions of bands were created. With these variables, all possible models were generated. Then, by means of adjusted R^2 and Mallows' C_p statistics, best models were chosen, finding quite good adjustments. Nevertheless, subsequent test trials with additional plots showed that these models were only fictitious models. Then, some possible causes were proposed in order to explain these results.

1.- INTRODUCCION

Actualmente, muchos estudios sobre la aplicación de la teledetección en inventarios forestales (Ahern *et al.*, 1991; Nemani *et al.*, 1993; Ekstrand, 1994) han sido ya realizados. Dos diferentes métodos son usualmente utilizados para procesar la información de variables forestales continuas obtenidas en el campo. Por un lado hay los métodos de clasificación (Bauer *et al.* 1994, Franklin 1994), que acaban dando un conjunto de clases discretas. Por otro lado, encontramos otros análisis que siguen dando la información de salida en valores de variables continuas, aprovechándose al máximo la información original. Los métodos de regresión están dentro de este último grupo (Zhuang *et al.*, 1993).

Históricamente, los inventarios forestales han sido realizados a partir de medidas en el campo, con una serie de limitaciones inherentes a estos métodos. Aunque los bosques ocupan superficies continuas, los muestreos son usualmente puntuales, obligando a la realización de ciegas extrapolaciones cuando información de puntos no muestreados es requerida. Además, si la extensión a muestrear es relativamente grande, extensos períodos de tiempo y grandes sumas de dinero son necesarias. Desde este punto de vista, la teledetección significa una posible solución a estos problemas (Howard, 1991).

A diferencia de los bosques homogéneos, usualmente utilizados en estudios de teledetección,

con una o pocas especies de coníferas (Ardö, 1992; Walsh, 1987; Danson and Curran, 1993) y frecuentemente situados en zonas de relieve llano (Curran, *et al.* 1992, Larsson, 1993, Rosema *et al.*, 1992), los bosques mediterráneos son mucho más heterogéneos. Esta heterogeneidad es debida a la localización marginal de los bosques en zonas de relieve abrupto, a la gestión de estos bosques, y al propio clima mediterráneo (con sotobosque visible, vegetación cambiante según orientación de las laderas, etc.).

2.- METODOLOGIA

La comarca del Solsonès, en Catalunya, fue escogida por ser una de las comarcas muestreadas por el Inventario Ecológico y Forestal de Catalunya (IEFC) con más superficie arbolada. Se utilizaron los seis canales no-térmicos de una imagen Landsat-TM con fecha del 7 de abril de 1989 (fecha incluida dentro del período de muestreo de campo). La imagen fue corregida geométricamente mediante el método propuesto por Palà y Pons (1995), obteniéndose un RMS menor que 1.0 píxeles. La corrección radiométrica se hizo con el modelo de Pons y Solé (1994), que mediante información del relieve da valores de reflectancias. Una vez las imágenes fueron corregidas, se leyeron los valores de los píxeles situados en las coordenadas de las estaciones de campo del IEFC. También se leyeron los valores de la ventana de 3 x 3 píxeles con la intención de hacer medidas de heterogeneidad de los alrededores de la estación de campo. La varianza de la ventana fue comparada con la varianza de toda la población de píxeles de las estaciones estudiadas, mediante el estadístico X^2 (Sokal and Rohlf, 1995), utilizándose como criterio de selección de estaciones.

Para este estudio se escogieron solamente estaciones puras de *Pinus nigra*, ya que esta es la especie forestal más abundante en la zona estudiada. De las 175 estaciones originalmente escogidas, 47 estaciones seleccionadas por la prueba de homogeneidad espacial fueron finalmente utilizadas. De todas las variables aportadas por el IEFC, el recubrimiento de copas de la estación (CC) y el índice de área foliar (LAI) fueron escogidas por ser las variables con más probabilidad de ser observadas desde el sensor. Para el LAI, se esperó una disminución en la variación de su respuesta espectral al aumentar sus valores, por tanto se le aplicó una transformación logarítmica. Para CC una simple relación lineal fue supuesta (en este caso, las estaciones con más del 100% de recubrimiento no fueron utilizadas).

Diferentes modelos de regresión fueron propuestos de cara a relacionar los datos del inventario (variables dependientes) con los datos del satélite (variables independientes):

1.- Regresión simple con los seis canales, uno a uno, y con valores del NDVI como variables independientes.

2.- Regresión múltiple con los seis canales a la vez como variables independientes.

3.- Regresión múltiple con nuevas variables obtenidas a partir de divisiones y multiplicaciones de los canales originales (25 variables fueron escogidas como posibles variables independientes). Todos los posibles modelos de una variable independiente, hasta las 25 variables a la vez, fueron creados de forma exhaustiva. Luego, de cara a elegir los mejores modelos, dos estadísticos fueron utilizados:

a) La R^2 ajustada, definida por la fórmula siguiente:

$$R^2_{adj} = 1 - \frac{(n-1)(1-R^2)}{n-p}$$

Donde n es el número de observaciones, p es el número de términos (incluyendo el

independiente) del modelo creado, y R^2 es el coeficiente de determinación. El modelo a elegir es el que tenga el valor mas alto de este estadístico (Ott, 1988).

b) La C_p de Mallows, definida por la fórmula siguiente:

$$C_p = \frac{SS_E}{\sigma_{exp}^2} - (n-2p)$$

Donde SS_E es la suma de cuadrados residual del modelo y σ_{exp}^2 es la varianza residual paramétrica esperada (usualmente obtenida del modelo formado por todas las posibles variables independientes). El modelo a elegir es el que tenga la p mas pequeña dentro de los modelos que cumplen $C_p \approx p$ (Ott, 1988).

Finalmente, a partir de 10 submuestras obtenidas al azar del conjunto total de 47 variables, se crearon modelos de 5 variables (escogidas entre las 25 propuestas). Estos modelos fueron testados con los puntos adicionales no incluidos en las submuestras, mediante el cálculo del coeficiente de correlación (r) entre valores observados y predichos.

3.- RESULTADOS

Ajustes muy pobres fueron encontrados en los modelos de regresión simple. Ni para CC ni para $\ln(LAI)$ en ningún caso se obtuvieron valores de R^2 superiores a 0.1, y ninguno de ellos fue significativo estadísticamente.

Mejores resultados, en comparación a los anteriores, se obtuvieron con las regresiones múltiples con los seis canales originales (Tabla I). Para las dos variables dependientes, modelos ligeramente significativos ($p < 0.05$) fueron alcanzados, aunque estos no fueron suficientemente ajustados para poder realizar buenas predicciones a nivel cuantitativo.

n	variable	R^2	p
41	CC	0.3404	0.0208
47	$\ln(LAI)$	0.2646	0.0450

Tabla I.- Resultados de las regresiones múltiples con los seis canales originales. Las diferentes columnas dan: el número de puntos, la variable dependiente, el coeficiente de determinación y la significación estadística.

Cuando se eligieron los mejores modelos formados con la combinación exhaustiva de nuevas variables, mediante el estadístico R^2 ajustada, mejores resultados fueron incluso alcanzados (Tabla II). Estos relativamente buenos ajustes fueron claramente apoyados por una sólida significación estadística.

Los modelos seleccionados inicialmente por la C_p de Mallows fueron erróneos, debido a que, el modelo construido con la 25 variables con objetivo de estimar la varianza paramétrica tenía una relación de n° de observaciones / n° de variables demasiado baja (Montgomery y Peck, 1992). Con la intención de evitar este problema, las variables escogidas por la R^2 ajustada fueron utilizadas como variables iniciales en el proceso. Entonces, los modelos obtenidos fueron casi idénticos a los seleccionados por la R^2 ajustada, con sus mismos ajustes y su misma significación estadística.

Finalmente, para las pruebas de test de modelos mediante puntos adicionales, se encontraron

unos buenos ajustes para los modelos (Tabla III), siempre acompañados de una alta significación estadística ($p < 0.01$). Sin embargo, los valores de correlación para los puntos de test, inesperadamente, fueron comparativamente mucho mas bajos, llegando incluso a obtener valores negativos.

n	variable	R ² ajus.	R ²	p	vars. en modelo
41	CC	0.4948	0.6337	0.0005	11
47	ln(LAI)	0.5514	0.6782	0.0000	13

Tabla II.- Algunas características de los modelos seleccionados por tener los mas altos valores de R² ajustada. Las diferentes columnas dan: el número de estaciones, la variable dependiente, la R² ajustada, la significación estadística del ajuste y el numero de variables independientes aceptado en el modelo.

4.- DISCUSION Y CONCLUSIONES

Los resultados de las correlaciones realizadas sobre los puntos adicionales de test ponen en evidencia la creación de modelos ficticios. Dos razones pueden haber llevado a la obtención de estos resultados.

Por un lado el reducido número de variables puede haber afectado el proceso de creación de estos modelos. Aunque la división entre puntos para el modelo y puntos para el test fue realizada al azar para las diez pruebas, es posible que los puntos del modelo no describiesen todas las condiciones ambientales donde la especie podía ser encontrada. En consecuencia, puede que las variables seleccionadas y los coeficientes ajustados a los modelos, no hubiesen podido simular las condiciones concretas de algunos puntos del test.

Por otro lado, como sugieren Sen y Srivastava (1990), si se crean muchas variables independientes al azar, es posible que algunas de ellas, por casualidad, se ajusten a una variable dependiente concreta. En nuestro caso, con 25 posibles variables a elegir, es posible que este fenómeno haya sucedido, al menos en cierto grado.

Una vez analizados los resultados del test con modelos de 5 variables, parece claro que los modelos con mas de 10 variables, elegidos por la R² ajustada y la C_p de Mallows son también ficticios. En realidad creemos que los valores de los ajustes reales deben ser mas modestos, debido a una serie de limitaciones de diferente ámbito.

Por ejemplo, provenientes del satélite pueden haber habido restricciones espectrales (p.e. bandas muy anchas), radiométricas (p.e. rango dinámico bajo) o de resolución espacial. El modelo de corrección radiométrica también tiene una serie de premisas que podrían haber sido restrictivas (Pons y Solé, 1994).

Los procesos de captación y procesado de los datos del inventario forestal también pueden haber sido fuentes de dispersión de los modelos. Por un lado, la localización inexacta de las estaciones de campo, en bosques altamente heterogéneos, puede haber afectado de forma significativa. También el uso de ecuaciones alométricas para predecir las variables del inventario parece haber marcado unos límites para los posibles ajustes.

prueba	variable	n modelo	R ² modelo	n test	r test
1	CC	32	0.5573	9	0.1237
	ln(LAI)	35	0.5853	12	-0.1816
2	CC	30	0.5420	11	0.1007
	ln(LAI)	35	0.5984	12	0.0891
3	CC	30	0.4875	11	0.3649
	ln(LAI)	35	0.5215	12	0.5383
4	CC	31	0.5961	10	0.2171
	ln(LAI)	35	0.5319	12	0.1276
5	CC	32	0.4995	9	0.5309
	ln(LAI)	35	0.4963	12	0.6073
6	CC	31	0.6252	10	-0.1117
	ln(LAI)	35	0.5227	12	0.2129
7	CC	30	0.5572	11	0.5968
	ln(LAI)	35	0.5301	12	0.4161
8	CC	32	0.4682	9	0.4572
	ln(LAI)	35	0.4871	12	-0.3378
9	CC	32	0.4750	9	0.1260
	ln(LAI)	35	0.5550	12	0.4978
10	CC	30	0.6025	11	0.2872
	ln(LAI)	35	0.5898	12	0.3467

Tabla III.- Resultados de los modelos creados a partir de diez submuestras, y de las pruebas de test realizadas con los puntos sobrantes. Las diferentes columnas dan: el número de prueba, la variable dependiente, el número de puntos para el ajuste, el coeficiente de determinación del modelo, el número de puntos sobrantes para el test y los coeficientes de correlación de los tests.

5.- REFERENCIAS

- Ahern, F.J., Erdle, T., Maclean, D.A. y Kneppek, I.D. 1991. A quantitative relationship between forest growth rates and Thematic Mapper reflectance measurements. *Int. J. Remote Sensing*. **12**: 387-400.
- Ardö, J. 1992. Volume quantification of coniferous forest compartments using spectral radiance recorded by Landsat Thematic Mapper. *Int. J. Remote Sensing*. **13**: 1779-1786.
- Bauer, M.E., Burk, T.E., Ek, A.R., Coppin, P.R., Lime, S.D., Walsh, T.A., Walters, D.K., Befort, W. y Heinzen, D.F. 1994. Satellite Inventory of Minnesota Forest Resources. *Photogrammetric Engineering & Remote Sensing*. **60**: 287-298.
- Curran, P.J., Dungan, J.L. y Gholz, H.L. 1992. Seasonal LAI in Slash Pine Estimated with Landsat TM. *Remote sens. environ*. **39**: 3-13.
- Danson, F.M. y Curran P.J. 1993. Factors Affecting the Remotely Sensed Response of Coniferous Forest Plantations. *Remote sens. environ*. **43**: 55-65.
- Ekstrand, S. 1994. Assessment of Forest Damage with Landsat TM: Correction for Varying Forest Stand Characteristics. *Remote Sens. Environ*. **47**: 291-302.
- Franklin, S.E. 1994. Discrimination of Subalpine Forest Species and Canopy Density Using Digital CASI, SPOT PLA, and Landsat TM Data. *Photogrammetric Engineering & Remote Sensing*. **60**: 1233-1241.
- Howard, J.A. 1991. *Remote Sensing of Forest Resources*. Chapman & Hall. London. pp. 13-23.

- Larsson, H. 1993. Linear regressions for canopy cover estimation in Acacia woodlands using Landsat-TM, -MSS and SPOT HRV XS data. *Int. J. Remote Sensing*. **14**: 2129-2136.
- Montgomery, D.C y Peck, E.A. 1992. *Introduction to Linear Regression Analysis*. John Wiley & sons. USA.
- Nemani, R., Pierce, L. y Running, S. 1993. Forest ecosystem processes at the watershed scale: sensitivity to remotely-sensed Leaf Area Index estimates. *Int. J. Remote Sensing*. **14**: 2519-2534.
- Ott, L. 1988. *An introduction to statistical methods and data analysis*. PWS-Kent publishing company. Boston. 835 p.
- Palà, V. y Pons, X. 1995. Incorporation of Relieve into Geometric Correction Based on Polynomials. *Photogrammetric Engineering and Remote Sensing*. **7**: 935-944.
- Pons, X., Solé-Sugrañes, L. 1994. A Simple Radiometric Correction Model to Improve Automatic Mapping of Vegetation from Multispectral Satellite Data. *Remote sens. environ.* **48**: 191-204.
- Rosema, A., Verhoef, W., Noorbergen, H. y Borgesius, J.J. 1992. A New Forest Light Interaction Model in Support of Forest Monitoring. *Remote sens. environ.* **42**: 23-41.
- Sen, A. y Srivastava, M. 1990. *Regression Analysis, Methods, and Applications*. Springer-Verlag. New York. pp 233-234.
- Sokal, R.R. y Rohlf, F.J. 1995. *Biometry*. 3^a ed. W.H. Freeman and company. New York. 859 pàgs.
- Walsh, S.T. 1987. Variability of Landsat MSS spectral responses of forests in relation to stand and site characteristics. *Int. J. Remote Sensing*. **8**: 1289-1299.
- Zhuang, H.C., Shapiro, M. y Bagley, C.F. 1993. Relaxation vegetation index in non-linear modelling of ground plant cover by satellite remote-sensing data. *Int. J. Remote Sensing*. **14**: 3447-3470.