

# Estimación de la fracción de cubierta vegetal a escala continental utilizando sensores de nueva generación

F.J. García-Haro, F. Camacho-de Coca y J. Meliá  
i.garcia.haro@uv.es

*Departamento de Termodinámica, Universitat de Valencia.  
C/ Dr. Moliner 50, 46100 Burjassot, Valencia.*

## RESUMEN

El análisis de composición espectral (SMA) usualmente determina la distribución de componentes en la escena a partir de un número finito de patrones (*endmembers*) que recogen la variabilidad espectral de la escena. En este trabajo se introduce una metodología para estimar la fracción de cubierta vegetal (FVC) a una escala continental, a partir de datos de nueva generación (VGT, SEVIRI). La metodología propuesta se ha desarrollado en el contexto del proyecto LSA SAF, orientado en la estimación de parámetros de la cubierta a partir de datos SEVIRI/MSG, y actualmente se encuentra en una fase preoperacional.

La estrategia combina las ventajas de un modelo geométrico (SMA) con un modelo probabilístico que tiene en cuenta la variabilidad natural de los componentes, y permite cuantificar el error en las estimaciones e incorporar información "a priori". Un paso previo es la extracción de atributos para hacer las estimaciones menos sensibles a factores externos (iluminación, color del suelo) y las sombras. Los resultados iniciales indican que la FVC estimada responde bien a cambios en la actividad de la vegetación y muestra un alto grado de consistencia con otros productos de nueva generación.

**PALABRAS CLAVE:** fracción de cubierta vegetal, Spectral Mixture Analysis, MSG.

## ABSTRACT

The scientific user communities have expressed the necessity of accurate maps of vegetation biophysical parameters at different scales, and suitable for long-term monitoring. IN the framework of the EUMETSAT LSA SAF our group is the responsible for the development and implementation of algorithms to obtain fractional vegetation cover (FVC) and leaf area index (LAI) products from the new EUMETSAT meteorological satellites SEVIRI/MSG and AVHRR-3/EPS.

This paper outlines a new Spectral Mixture Analysis (SMA) technique based on the Bayesian Theorem, along with the use of standardised SMA, to improve understanding of the impact of endmember variability on the derivation of subpixel vegetation fractions at a global scale. The uncertainty of FVC concerning view and sun angles is minimized by using as input an optimal configuration. Endmember signatures are no longer treated as constants, but they are represented by multi-modal probability density functions. The algorithm incorporates endmember spectral variability in the unmixing process, thus reducing misidentifications of surface components and demanding fewer endmembers. The FVC estimates obtained with the algorithm are spatially and temporally consistent and the results of preliminary validation studies are encouraging.

**KEY WORDS:** fractional vegetation cover, Spectral Mixture Analysis, MSG.

## INTRODUCCIÓN

El desarrollo tecnológico de los programas actuales de Observación de la Tierra posibilita una constelación de sensores ópticos de nueva generación (VEGETATION/SPOT, MODIS/EOS, MERIS/ENVISAT, SEVIRI/MSG) que observan el planeta simultáneamente. Sin embargo, además de los parámetros radiométricos, existe una fuerte demanda de parámetros biofísicos que sean direc-

tamente aplicables al estudio de fenómenos de la superficie. Ello ha motivado que las principales misiones espaciales hayan puesto en marcha varios proyectos para desarrollar algoritmos operacionales. En relación a los productos de la cubierta vegetal, existe un amplio consenso internacional en superar conceptos tradicionales tales como los índices de vegetación, que con frecuencia dependen de las características de la escena y del sensor.

En el contexto del proyecto LSA SAF, nuestra misión es desarrollar una metodología para estimar variables biofísicas de la cubierta, tales como la fracción de cubierta vegetal (FVC) y el índice de superficie foliar (LAI) sobre Europa y África a partir de datos SEVIRI/MSG, y con una frecuencia temporal de 10 días. Estos parámetros se consideran inputs necesarios para comprender mejor la interacción entre el ciclo del carbono, el clima y los cambios en el uso del suelo.

En los últimos años el análisis de composición espectral, conocido en la literatura como *Spectral Mixture Analysis* (SMA) se ha aplicado con éxito para explotar la información hiper- y multi-espectral de los datos de teledetección en numerosas aplicaciones (Roberts et al. 1997, García-Haro et al. 2005a). El SMA se basa en un modelo geométrico que permite determinar la distribución de materiales –especies vegetales, rocas, minerales– presentes en la escena. Sin embargo, el problema se complica mucho cuando el número de canales en la región óptica es reducido y el IFOV es grande. En este caso, el problema se reduce a cuantificar la contribución de una única componente de interés que presenta una cierta variabilidad en presencia de materiales que constituyen el fondo o señal a filtrar.

En este trabajo se presenta una metodología para estimar de variables biofísicas de la vegetación –tales como la fracción de cobertura vegetal (FVC)– que sean adecuadas para observar la dinámica de la cubierta a escala regional y global, a partir de los datos proporcionados por la nueva generación de satélites. La metodología se basa en técnicas de SMA novedosas, adaptadas al problema planteado. A continuación se describe brevemente el fundamento del método, algunos aspectos destacados de la metodología propuesta y algún ejemplo de la aplicación del prototipo sobre datos VEGE-TATION/SPOT (VGT) y SEVIRI/MSG.

## PLANTEAMIENTO DEL PROBLEMA

### Fundamento teórico

Cada píxel de una imagen viene representado por un vector  $n$ -dimensional  $\mathbf{R}$ , siendo  $n$  el número de bandas espectrales. Tradicionalmente los modelos lineales de mezcla espectral en que se basa el SMA representan la reflectividad de cada píxel a partir de una combinación lineal de un

conjunto finito de firmas (patrones) correspondientes a componentes presentes en la escena, conocidas como *endmembers*:

$$\mathbf{R} = \mathbf{V} f_v + \mathbf{S} f_s + \varepsilon \quad (1)$$

donde  $\mathbf{V}$  y  $\mathbf{S}$  las matrices con los *endmembers* de vegetación y suelo, respectivamente,  $f_v$  y  $f_s$  son los vectores que contienen sus pesos asociados, y  $\varepsilon$  es el residuo –que habitualmente se aproxima a una distribución normal multivariante,  $\varepsilon \sim \mathbf{N}(0, \sigma^2 \mathbf{I})$ . Para que los pesos tengan un significado físico y puedan interpretarse como fracciones, deben ser positivos y sumar la unidad (condición de normalización).

En general, no sólo son desconocidas las fracciones  $\{f_v, f_s\}$  sino también los *endmembers*  $\{\mathbf{V}, \mathbf{S}\}$  que constituyen el píxel. El problema más simple consiste en suponer que  $\mathbf{V}$ , no varía (modelo determinista) y  $\mathbf{S}$  se ajusta bien a una función de distribución conocida (gausiana). Sin embargo, el problema es generalmente más complejo debido a la variabilidad en la respuesta de la vegetación y a que el suelo de fondo presenta una distribución desconocida (multi-modal), muy variable localmente. Cabe considerar dos estrategias:

(i) Modelo *probabilístico*, en que la función a determinar es una distribución estadística (generalmente normal multivariante) cuyos parámetros pueden calcularse usando técnicas estocásticas, de forma análoga a un clasificador “soft”. Cada píxel viene descrito por un grado de proximidad a cada una de las clases, asociando la probabilidad “a posteriori” con las fracciones.

(ii) Modelo *geométrico*, en el que la variabilidad de los datos se expresa de una forma determinista por medio de patrones o estructuras (*endmember*). Típicamente, el número de componentes en  $\mathbf{V}$  y  $\mathbf{S}$  es desconocido, aunque se puede estimar utilizando técnicas a partir del rango de la matriz de los datos utilizando diferentes algoritmos (e.g. *singular value decomposition*, SVD).

La manera habitual de estimar las fracciones es utilizar el método de minimización clásico con un parámetro de Lagrange para imponer la condición de normalización. La positividad de las fracciones viene garantizada por la elección de *endmembers* convenientemente situados en hipotéticos extremos del *convex hull* que contiene el hipervolumen de los datos en un subespacio de  $\mathbf{R}^n$ .

## Extracción de características

La extracción de características o atributos espectrales tiene por objeto desacoplar la información procedente de la vegetación de la de otros factores (suelo, iluminación). Ello consiste en definir una métrica que transforme los ejes originales (bandas) a lo largo de unos ejes (atributos). Un buen conjunto de atributos debería potenciar la similaridad entre patrones de la misma clase y la disimilaridad entre patrones de clases diferentes. Hemos propuesto varias técnicas para obtener atributos lo más discriminativos posible y que, a la vez, preserven la linealidad del problema, de manera que la ecuación (1) siga siendo aplicable:

(i) Técnicas estadísticas, el fundamento de las cuales consiste en maximizar:

- variancia de los datos (análisis de componentes principales)
- separabilidad entre clases (análisis discriminante)
- cociente señal/ruido (MNF).

(ii) Transformaciones lineales, que normalizan los factores no deseados:

- estandarización, o transformación cada espectro  $r$  en porcentaje de variación alrededor del valor medio,  $(r - \mu_r)/\sigma_r$ . El resultado es un vector de media cero y variancia uno.
- normalización, o cociente del espectro por su valor medio,  $r/\mu_r$
- substracción del valor medio del espectro,  $r - \mu_r$

## Estudio de sensibilidad

Para evaluar cuantitativamente las posibilidades que ofrecen los distintos métodos hemos realizado un análisis de sensibilidad a partir de distribuciones simuladas de varios tipos de suelo y vegetación, que muestra la validez de todos ellos. Es posible la utilización simultánea de varios de estos métodos de extracción de características, aumentando con ello la dimensionalidad inicial de los datos lo que, potencialmente, permite una mejor discriminación del suelo y la vegetación. No obstante, una cantidad importante de atributos aumenta la complejidad del clasificador. Además, la presencia de atributos redundantes, afecta negativamente a la fiabilidad del algoritmo.

Los resultados muestran que todos ellos son, en general, aplicables. En particular, la estandarización preserva la forma del espectro pero elimina la

mayor parte de las variaciones debidas a la iluminación, topografía, rugosidad de la superficie y color del suelo. La estandarización disminuye también significativamente la separación entre estructuras vegetales con un arquitectura muy distinta, tales como cultivos –alta reflectividad en el infrarrojo cercano (IRC)– y bosques densos –baja transmisividad y valores bajos en todas las bandas.

La utilización de las transformaciones propuestas permite disminuir la dimensionalidad de los subespacios  $V$  y  $S$  en la ecuación (1), haciéndolo compatible con la dimensionalidad de los datos; de esa manera, es posible utilizar un único *endmember* para representar las componentes de vegetación y suelo. Dado que el interés primordial no es la fracción de cada tipo de vegetación sino su suma, esta parametrización reduce los errores al hacer que el modelo sea menos sensible al ruido en los datos y factores externos. Sin embargo tanto la estandarización como la normalización no sólo alteran los espectros sino también las fracciones, de manera que en el nuevo sistema de ejes la suma de éstas ya no es la unidad. En el caso de la estandarización se tiene:

$$\sum_{i=1}^c \frac{f_i}{\sigma_i} = \frac{1}{\sigma_r} \quad (2)$$

siendo  $c$  el número de *endmembers*,  $\hat{f}_i$  los pesos de las variables estandarizadas y sus respectivas desviaciones estándar. Si no tenemos en cuenta esta condición de normalización, se produce un sesgo en las fracciones, especialmente para valores intermedios. Si se impone la condición (2) se requiere, posteriormente, transformar las fracciones:

$$f_i = \frac{\sigma_r}{\sigma_i} \hat{f}_i \quad (3)$$

Las fracciones así obtenidas,  $f_i$ , suman la unidad y carecen de sesgo. No obstante, para aplicar (3) hay que conocer la de cada componente, la cuál está muy correlacionada con su albedo, lo que implica un conocimiento del albedo del suelo y vegetación en cada píxel y, por tanto, introduce un error adicional.

## METODOLOGÍA

Algunos de los aspectos más destacados de la metodología propuesta son:

1.- La dimensión angular se considera como una fuente de variabilidad de la señal no deseada, la cual debemos normalizar. La manera habitual consiste en utilizar como *inputs* valores de reflectividad referidos a una configuración óptima (“nadir-cenit”), que se conocen como  $k_0$  y se derivan de la inversión de modelos lineales paramétricos “kernel-driven” (Roujean et al. 1992).

2.- Utilización de información auxiliar (productos y clasificaciones con resolución espacial mayor) para definir la distribución estadística de los *endmembers* (modelo probabilístico).

3.- Extracción de atributos (variables estandarizadas) que minimizan la variabilidad de las componentes de vegetación y suelo, reduciendo la confusión espectral entre ellas.

4.- Un proceso jerarquizado (García-Haro et al. 2005a) para determinar los *inputs* del modelo (o sea, la información “a priori” sobre la distribución de *endmembers*) de una forma local, mejorando con ello la precisión de las estimaciones.

5.- Caracterizar mejor la composición de cada píxel utilizando información adicional, como por ejemplo una imagen multitemporal de la zona, que permita identificar las fechas con mayor presencia de suelo y vegetación.

6.- Modelar la respuesta espectral asociada los posibles tipos de suelo y vegetación en la escena a partir de una mezcla de gaussianas, utilizando para ello un algoritmo numérico denominado Maximization-Expectation (McLachlan and Krishnan, 1977).

7.- Calcular la probabilidad de cada píxel de contener los posibles tipos de suelo y vegetación obtenidos en la etapa anterior. Para ello se utiliza la teoría Bayesiana e introduciendo información auxiliar “a priori” de cada componente (García-Haro et al. 2005b).

8.- Estimar la FVC a partir de un modelo probabilístico, es decir, considerando las combinaciones de suelo y vegetación más probables en cada píxel.

9.- Calcular el error asociado, teniendo en cuenta tanto la propagación de los errores asociados a los *inputs* como la incertidumbre inherente al proceso de selección de los *endmembers*.

## RESULTADOS

El algoritmo se ha implementado en el sistema LSA SAF para la obtención de la FVC a partir de datos SEVIRI/MSG. Los *inputs* utilizados son valores de reflectividad “nadir-cenit”,  $k_0$ , en tres cana-

les SEVIRI: VIS06 (0.56-0.71  $\mu\text{m}$ ), VIS08 (0.74-0.88  $\mu\text{m}$ ) y NIR16 (1.5-1.78  $\mu\text{m}$ ). Sin embargo, con el fin de evaluar la validez del algoritmo en una primera fase del prototipo, éste se ha aplicado sobre diferentes series temporales de datos de nueva generación. La figura siguiente ilustra la aplicación del mismo sobre datos decenales de  $k_0$  derivados del sensor VGT (VEGETATION/SPOT). Este sensor dispone de 4 canales espectrales: B0 (0.43-0.47  $\mu\text{m}$ ), B2 (0.61-0.68  $\mu\text{m}$ ), B3 (0.78-0.89  $\mu\text{m}$ ) y MIR (1.58-1.75  $\mu\text{m}$ ). Los tres últimos coinciden con los canales VIS-0.6, VIS-0.8 y IR-1.6 del instrumento SEVIRI.

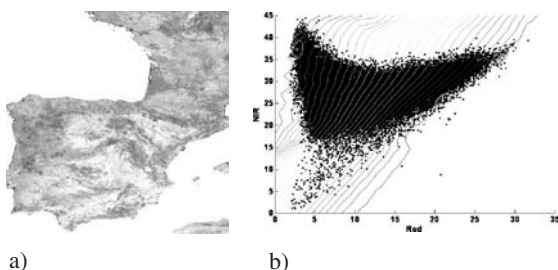
Un paso previo que facilita la identificación del albedo de suelo/vegetación consiste en utilizar una serie temporal larga. La Figura 1a muestra una composición en la que para cada píxel hemos seleccionado la escena con una contribución de suelo mayor durante enero-septiembre 2003, en base a un algoritmo simple, que retiene el valor medio de los píxeles con valor mínimo de NDVI. La Figura 1b muestra el bien conocido “triángulo de reflectividades” en el espacio rojo-IRC. La base del mismo está constituida por áreas de suelo desnudo. Las diferencias graduales se deben a cambios en el color, brillo, humedad, rugosidad de la superficie a lo largo de la llamada “línea de suelo”. Este es un claro ejemplo en el que no puede definirse de una forma determinista la respuesta utilizando *endmembers* puros (‘hard’) o gaussianos para representar a los suelos presentes en la escena, ya que éstos forman una distribución continua.

El vértice opuesto corresponde a áreas de vegetación verde y vigorosa que, en la figura, corresponden mayoritariamente bosques cerrados con hoja ancha y perenne. La variabilidad de la signatura de vegetación incluye, además, otras cubiertas se han filtrado en la composición temporal (cultivos, bosques de coníferas, etc.) y que introducen una gran variabilidad espectral. Se han superpuesto las isóneas correspondientes a la fracción de cubierta vegetal obtenida con el algoritmo propuesto, así como las elipses correspondientes a distribuciones de probabilidad asociadas a componentes puras de vegetación y suelo.

A diferencia de muchos de los métodos habituales, el método propuesto proporciona una solución regularizada, es decir, responde de una forma gradual y continua a los cambios en la reflectividad. Además, tiende a mitigar los errores debidos al ruido en los datos y a la presencia de artefactos en los datos, que generalmente corresponden a super-

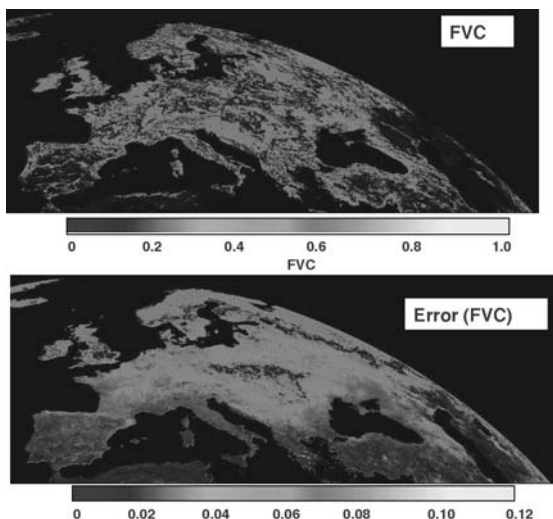


ficies incorrectamente representadas por el conjunto posible de *endmembers*.



**Figura 1.** a) Falso color (B2, B3, MIR) de la imagen de k0 del sensor VGT correspondiente a una cobertura vegetal mínima obtenida mediante una composición del periodo enero-septiembre de 2003. b) Diagrama bidimensional B2-B3 de la imagen.

Finalmente, la Figura 2 muestra el producto de FVC obtenido para una imagen SEVIRI del 19 de Agosto de 2005. Lo más destacable es la adecuada consistencia espacial del producto y el grado de coherencia con la información verdad-terreno. En general, la incertidumbre del producto es mayor en latitudes elevadas, lo cual se debe principalmente a que el error de los *inputs* es conside-



**\*Figura 2.** Estimaciones de la FVC y de su error asociado a partir de datos SEVIRI/MSG sobre la región de Europa. El producto corresponde a una ventana temporal de 10 días centrada en el 19 de Agosto de 2005.

rablemente mayor a consecuencia de la elevada cobertura nubosa.

Con el fin de dotar a los productos de la significación e incertidumbres necesarias, se ha iniciado la intercomparación con algunos productos actuales (VGT, POLDER, MODIS, MERIS), en el contexto del proyecto GEOLAND (<http://www.gmes-geoland.info/>). En general, los resultados muestran una elevada consistencia con el producto FVC level 1 de la misión VGT, en fase de validación, y con el producto LAI MOD15A2 versión 4.1 de la misión MODIS/TERRA (García-Haro et al. 2005b). Asimismo, muestran una adecuada correspondencia con estimaciones realizadas a partir de medidas *in situ* en una zona piloto de validación, Barrax (Albacete) (Camacho-de Coca et al. 2005).

## CONCLUSIONES Y DIRECCIONES FUTURAS

La comunidad científica requiere de estimaciones precisas de variables biofísicas de la vegetación –tales como la fracción de cobertura vegetal que sean adecuadas para observar la dinámica de la cubierta a escala regional y global. Se presenta aquí un algoritmo adecuado para ello, el cual ofrece ventajas importantes tales como incorporar la variabilidad de suelo y vegetación, estimar la incertidumbre asociada y reducir su error. Los resultados iniciales indican que la FVC estimada responde bien a cambios en la actividad de la vegetación. Asimismo, ha mostrado un alto grado de consistencia con otros productos actuales. En el futuro se pretende realizar grandes esfuerzos para validar este producto, a través de programas de validación como VALERI (<http://www.avignon.inra.fr/valeri/>) del CNES, bigfoot (<http://www.fsl.orst.edu/larse/bigfoot/index.html>) de la NASA para validar los productos MODLAND (Morissette et al 2002), y campañas como DAISEX o SPARC financiadas por la ESA (<http://gpds.uv.es/sparc/>) o la más reciente iniciativa METLAND para validar los productos LSA SAF.

## AGRADECIMIENTOS

Este trabajo ha sido subvencionado por los proyectos LSA SAF de EUMETSAT, IDEAS (REN2002-

Todas las figuras precedidas de asterisco se incluyen en el cuadernillo anexo de color

01495) de CICYT y TEDECVA de la Generalitat Valenciana (GV2004-B-191). Los datos VGT han sido facilitados por el proyecto CYCLOPES, Avignon/INRA.

## BIBLIOGRAFÍA

- CAMACHO-DE COCA, F., MARTINEZ, B. y GARCÍA-HARO, F.J. 2005. Validación directa de productos de vegetación de baja resolución en la zona agrícola de Barrax, Proc. XI Congreso Nacional de Teledetección, Puerto de la Cruz, Tenerife.
- GARCÍA-HARO, F.J., SOMMER, S. and KEMPER, T. 2005a. Variable multiple endmember spectral mixture analysis (VMESMA), *International Journal of Remote Sensing*. 26:2135-2162.
- GARCÍA-HARO, F.J., CAMACHO-DE COCA, F. MELIÁ, J. and MARTÍNEZ, B. 2005b. Proceedings of the Operational derivation of vegetation products in the framework of the LSA SAF project, EUMETSAT Meteorological satellite conference, Dubrovnik, Croatia (en prensa).
- MCLACHLAN, G. J., and KRISHNAN, T. 1997. The EM algorithm and extensions. New York: Wiley.
- McLachlan, G. J., & Peel, D. 2001. Finite mixture models. New York: Wiley.
- MORISSETTE, J.T., PRIVETTE, J.L., and JUSTICE, C.O. 2002. A framework for the validation of MODIS Land products. *Remote Sensing of the Environment*. 83: 77-96.
- ROBERTS, D. A., GARDNER, M., and CHURCH, R. 1998. Mapping chaparral in the Santa Monica mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment*, 65. 267-279.
- ROUJEAN, J.L., LEROY, M., and DESCHAMPS, P.Y. 1992. A Bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *Journal of Geophysical Research*. 97 (18): 20.455-20.468.