

ANÁLISIS DE LA INCERTIDUMBRE EN MODELOS DE TELEDETECCIÓN

J. A. MORENO RUIZ (*) y M. CROSETTO (**)

jamoreno@aul.es

(*) Universidad de Almería. Ctra. Sacramento s/n. La Cañada de San Urbano, 04120 Almería

(**) Unitat de Teledetecció, Institut Cartogràfic de Catalunya, Parc de Montjuïc, 08038 Barcelona

RESUMEN: Normalmente se da poca importancia al problema de la incertidumbre en modelos de teledetección. En este trabajo proponemos una metodología de análisis válida para estimar la incertidumbre de productos obtenidos a partir de un modelo genérico de teledetección. En la primera parte del artículo, están descritos los fundamentos básicos de la metodología. En la segunda parte se describe una aplicación a un modelo para la detección de superficies quemadas a partir del conjunto de datos NOAA-NASA Pathfinder AVHRR 8Km.

ABSTRACT: Little emphasis is often given to the problem of uncertainty propagation in models driven by remotely sensed data. This work describes a procedure suitable to analyse the uncertainty associated with the output of a generic remote sensing model. The fundamentals of the procedure are described, and an application focused on a model for burned area detection based on the NOAA-NASA Pathfinder AVHRR 8Km data set is discussed.

Palabras clave: análisis de incertidumbre, análisis de sensibilidad, modelo de superficies quemadas, AVHRR.

INTRODUCCIÓN

La teledetección es reconocida como una importante fuente de datos que pueden servir como base para investigaciones en diferentes campos científicos y como soporte para la toma de decisión. Para llegar a productos de nivel alto, los datos "raw" de teledetección tienen que pasar una secuencia de procedimientos de procesamiento y de análisis. En este trabajo se entiende por modelo una cadena de procedimientos computacionales y algoritmos que permiten generar productos de alto nivel a partir de datos "raw" de entrada.

Un problema muy importante relacionado con los modelos de teledetección es el análisis de la incertidumbre asociada a los datos "raw" de entrada, su propagación en los modelos y su efecto en los datos de salida. La manera típica de aproximar este problema puede definirse como "aproximación determinista": una vez elegido un modelo y un conjunto de datos, ambos se consideran implícitamente como libre de errores y por consecuencia, los datos de salida se usan de manera determinista, o sea, desconociendo sus errores.

En este trabajo proponemos una metodología para la evaluación de la incertidumbre de productos de teledetección que puede ser aplicada a modelos caracterizados por cualquier grado de complejidad y orientado a cualquier tipo de datos. Dicha metodología se basa en la utilización de dos herramientas de modelización, el

Análisis de Incertidumbre (AI) y el Análisis de Sensibilidad (AS). El AI permite evaluar la incertidumbre asociada con la salida del modelo, mientras que el AS permite evaluar cual es la repercusión de la incertidumbre en los diferentes datos de entrada del modelo sobre la incertidumbre en la salida del modelo.

Análisis de Incertidumbre

El objetivo del AI es evaluar la incertidumbre asociada con la salida del modelo. El modelo, es decir, la relación entre los datos de entrada I_i ($i = 1, \dots, K$) y los de salida Y se representa mediante:

$$Y = f(p_1, p_2, \dots, p_M, I_1, I_2, \dots, I_K) \quad (1)$$

donde el operador $f(\cdot)$ es una cadena de procedimientos computacionales y algoritmos y los p_j ($j = 1, \dots, M$) son parámetros internos del modelo. El AI incluye tres importantes etapas: la identificación de todas las fuentes de incertidumbre de los datos de entrada y parámetros del modelo; la modelización de dicha incertidumbre, y la propagación de la incertidumbre a través del modelo. Existen diferentes técnicas analíticas y numéricas para estudiar la propagación de la incertidumbre (Heuvelink, 1998). En este trabajo usamos una técnica numérica, el método de Monte Carlo (MC), que no impone ninguna restricción sobre $f(\cdot)$ y por tanto constituye una herra-

mienta flexible para el AI aplicable a muchos modelos dentro del campo de la teledetección.

Para aplicar el método MC, en un primer paso, se deben identificar de una manera adecuada todos los datos de entrada y parámetros del modelo afectados por la incertidumbre. El segundo paso es la modelización de la incertidumbre: a cada dato de entrada y parámetro se tiene que asociar un modelo de error. Un modelo de error es un modelo estocástico capaz de generar una población de "versiones corruptas" de la misma realidad, que llamaremos realizaciones (Goodchild, 1992). La idea de MC es de evaluar el modelo N veces, usando "versiones corruptas" o realizaciones de los datos de entrada y de los parámetros. MC involucra los siguientes pasos:

1. Repite N veces: genera aleatoriamente una realización de los datos de entrada y parámetros, y evalúa el modelo, o sea, calcula Y con dichas realizaciones.

2. Evalúa la incertidumbre asociada a la salida del modelo Y_i (con $i = 1, \dots, N$) representa la función de densidad de probabilidad empírica (FDP) de que permite una total caracterización de las propiedades estocásticas de. Por ejemplo, a partir de la FDP se pueden estimar la variancia y otros momentos de mayor orden.

Análisis de Sensibilidad

El AS considera la incertidumbre en el sentido contrario que el AI, desde la salida hacia la entrada, determinando como contribuye a la incertidumbre de cada una de las fuentes individuales a la incertidumbre de salida.

En la literatura se describen muchas técnicas de AS (Saltelli *et alii.* 2000). En este trabajo usamos una técnica de AS llamada "Extended FAST" basada sobre los índices de sensibilidad. El índice de sensibilidad de un cierto factor (un dato de entrada o un parámetro) representa la contribución fraccional a la variancia de la salida del modelo que es debida a este factor. "Extended FAST" es una técnica basada sobre la descomposición de la variancia que asume los errores en los factores de entrada del modelo independientes. FAST adopta la simulación MC, asociada a una particular estrategia de muestreo para la generación de realizaciones de los factores del modelo. Una descripción completa de la técnica se encuentra en (Saltelli *et alii.* 2000) y en (Crosetto *et alii.* 2001).

Un paso clave en el desarrollo de los AI y AS es la modelización de la incertidumbre. La próxima sección describe un modelo de error para datos raster, típicos datos de entrada de muchos modelos de teledetección.

Un modelo de error para datos raster

Los datos de entrada típicos de modelos de teledetección son datos raster 2D (o sea, campos bidimen-

sionales) que contienen atributos de tipo cuantitativo. Sea un campo A uno de los datos de entrada de un modelo; normalmente, solo una realización particular S de este campo es disponible. Un modelo genérico de error para este tipo de dato tendría la siguiente forma:

$$Z(\bar{x}) \equiv S(\bar{x}) + N(\bar{x}) \quad (2)$$

donde N es el campo de error de media m_N y variancia, σ_N^2 , y Z es el campo corrompido por el error. En la simulación MC, en cada ciclo se genera aleatoriamente un campo N , añadiéndolo a S y obteniendo así una "versión corrupta" Z de A . Este modelo permite introducir dos tipos de incertidumbre: un error sistemático, considerando valores de m_N distintos de cero; y un error estocástico, haciendo que N sea un campo aleatorio, con una matriz de variancia C_{NN} que permite representar la correlación espacial del campo N . En el caso de que N no tenga correlación espacial, la matriz toma la forma $C_{NN} = \sigma_N^2 I$, donde I es la matriz de identidad. Para una descripción detallada del modelo, veasé (Crosetto *et alii.* 2001).

DESCRIPCIÓN DE UNA APLICACIÓN

Se han realizado los AI y AS sobre un modelo de detección de superficies quemadas basado en el conjunto de datos NOAA-NASA Pathfinder AVHRR 8-Km. El objetivo de los análisis fue identificar, de entre todas las incertidumbres que intervienen en el modelo, aquellas que tienen mayor influencia sobre la incertidumbre de la salida. Este análisis puede ser utilizado en la etapa de construcción de un nuevo modelo, o para mejorar un modelo ya existente. En ambos casos, el análisis puede ofrecer importantes ayudas tanto para la comprensión del comportamiento del modelo como para su posible mejora.

Análisis del modelo

El modelo fue desarrollado para la extracción de superficies quemadas a partir de datos AVHRR-GAC 5-Km y fue utilizado para estimar las superficies quemadas del continente africano en el periodo 1981-1991 (Barbosa *et alii.* 1999b). El modelo se basa en un algoritmo que emplea tres tipos de umbrales: fijos, temporales y automáticos (Barbosa *et alii.* 1999a). Se realiza una primera selección mediante una secuencia de umbrales fijos cuyo objetivo es identificar los píxeles potencialmente quemados. A continuación se realiza una selección más severa mediante un umbral temporal. El último tipo de umbral actúa como una máscara, previniendo la detección de falsas superficies quemadas. Un valor umbral previo se emplea para enmascarar los píxeles nubosos. El modelo ha sido aplicado al conti-

nente africano con el conjunto de datos AVHRR GAC 8-Km (James, 1994). Junto a la incertidumbre que afecta a los datos de entrada del modelo, los canales 1 (Ch1), 2 (Ch2) y 3 (Ch3), se ha considerado también la incertidumbre de los parámetros propios del modelo.

Incertidumbre en los datos de entrada

Ch1 y Ch2 (Ch1/2) son fuertemente dependientes de la variación de las condiciones geométricas en la adquisición de los datos. Los satélites NOAA 7, 9 y 11, utilizados para la adquisición de los datos Pathfinder, fueron puestos en órbita para mantener fijo el sincronismo de las respectivas pasadas con respecto a la posición del sol. Debido a una degradación de las órbitas, se encontraron desfases en dichas pasadas de 1-3 horas (Price *et alii.* 1991). Estos desfases afectan sistemáticamente a los valores de Ch1/2, y por tanto a cualquier análisis comparativo interanual sobre la detección de superficies quemadas. El efecto sistemático sobre Ch1/2 fue evaluado mediante simulación con el software 6S (Vermote *et alii.* 1998), considerando las diferentes geometrías de adquisición de los datos y en particular la variación en los valores de los ángulos cenitales solares entre los diferentes años. Como un ejemplo, cinco perfiles anuales de Ch2 sobre una zona de sabana se muestran en Figura 1.

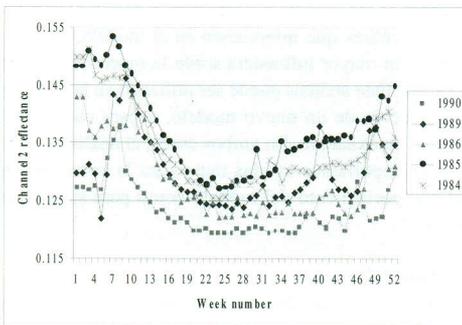


Figura 1. Evolución anual del Ch2 sobre una zona de sabana en cinco diferentes años.

A partir de esas simulaciones se ha estimado el error sistemático, que en el análisis MC se ha considerado uniformemente distribuido en el rango $[-7, 7 \text{ \%}]$ de los valores de reflectancia nominal (Tabla 1). El error estocástico de Ch1/2 se estimó mediante análisis estadístico de una serie de ventanas en el desierto del Sahara; se escogió el desierto por su estabilidad radiométrica. La desviación estándar del error σ_N obtenida fue de 1.6%.

Datos de entrada	Incertidumbre asociada	Distribución
Ch1/2	Err. sistemático	U: $[-0.07; 0.07 \text{ \%}]$
Ch1/2	Err. estocástico	N: $[0; 1.6 \text{ \%}]$
Ch3	Err. sistemático	U: $[-2; 2 \text{ K}]$
Ch3	Err. estocástico	N: $[0; 2 \text{ K}]$
Umbral por nubes	Máscara	Proporción P de píxeles
Umbral de Ch2	Error	U: $[-1; 1 \text{ \%}]$
Umbral de Ch3	Error	U: $[-2; 2 \text{ K}]$

Tabla 1. Modelización de la incertidumbre en los datos de entrada y en los parámetros del modelo. La distribución uniforme se indica con U:[rango] y la normal con N:[media, desviación estándar].

Para el canal Ch3 se ha considerado un error sistemático uniformemente distribuido en el rango $[-2, 2 \text{ K}]$ y se ha utilizado un error estocástico normalmente distribuido de media nula y desviación estándar de 2 K (Tabla 1).

Incertidumbre en los parámetros del modelo

Los dos umbrales más importantes en el modelo son el umbral de Ch2 ($Ch2 < 12,5 \text{ \%}$, llamado T2) y el umbral de Ch3 ($Ch3 > 312 \text{ K}$, llamado T3). T2 y T3 fueron determinados empíricamente, y de aquí su incertidumbre asociada, que ha sido modelada con errores uniformemente distribuidos en el rango $[-1, 1 \text{ \%}]$ y $[-2, 2 \text{ K}]$ para T2 y T3 respectivamente.

Una tercera incertidumbre, asociada con el umbral relacionado con la máscara de nubes (TC), también ha sido considerada en este trabajo. Más detalles se encuentran en (Crosetto *et alii.* 2001).

Análisis de resultados

Los siete factores considerados en la simulación MC, están en la Tabla 1. La simulación MC se realizó para el año 1990 con un total de 2190 ejecuciones. Este número ha demostrado ser suficiente para estimar con adecuada precisión los índices de sensibilidad en este modelo (Crosetto *et alii.* 2001). Los resultados del AS se muestran en la Tabla 2. Los dos factores que afectan de forma más considerable a la incertidumbre del modelo son el valor umbral T2 y el error sistemático de Ch1/2.

Factores de entrada	Índice Total
Error sistemático Ch1/2	0.37
Error estocástico Ch1/2	0.08
Error sistemático Ch3	0.05
Error estocástico Ch3	0.01
Umbral TC	0.09
Umbral sobre el canal 2	0.43
Umbral sobre el canal 3	0.09

Tabla 2. Resultados del AS: índices de sensibilidad de orden total.

En el modelo, la funcionalidad de T2 es la de separar en una primera aproximación los píxeles potencialmente quemados del resto. Por lo tanto, se esperaba un pobre impacto de su incertidumbre sobre el modelo (o sea un bajo índice de sensibilidad). Sorprendentemente, su índice de sensibilidad es alto, con un valor del 0.43, indicando que T2 no se ha comportado como se esperaba. Esto puede ser debido a las diferencias entre el conjunto de datos AVHRR 8Km, considerado en este trabajo, y el conjunto de datos AVHRR 5km considerado en (Barbosa *et alii*. 1999a). De cualquier forma, este resultado claramente indica que la atención del modelador debe estar centrada sobre el valor de T2.

El error sistemático de Ch1/2 es la segunda mayor fuente de incertidumbre en la salida: para mejorar la precisión del modelo se debe reducir este error. El bajo índice de sensibilidad asociado con el umbral TC, de valor 0.09, indica que TC tiene una baja influencia sobre la salida del modelo. Entonces, el esfuerzo para las futuras mejoras del modelo no debería ser utilizado en analizar dicho umbral. Igualmente, no sería aconsejable desarrollar un módulo de pre-procesamiento para reducir el error estocástico sobre Ch3, ya que su impacto sobre la salida es insignificante, el índice es de 0.01.

CONCLUSIONES

Se ha descrito una metodología para evaluar la propagación de la incertidumbre en un genérico modelo de teledetección. Esta metodología ha sido utilizada para analizar un modelo de detección de superficies quemadas sobre el conjunto de datos Pathfinder AVHRR 8-km. El objetivo de este trabajo ha consistido en evaluar el comportamiento del modelo en orden de identificar las fuentes más importantes de incertidumbre. Los errores asociados con la entradas fueron descompuestos en los componentes sistemáticos y estocásticos. Los errores característicos fueron obtenidos mediante análisis estadístico y simulaciones, o asumiendo unos niveles de error en los datos. La incertidumbre de los parámetros fue modelada transformando los valores umbrales en variables aleatorias.

Los resultados dan una importante ayuda para entender el comportamiento del modelo y para eventualmente mejorarlo. Con el objetivo de mejorar dicho modelo, se debería prestar una especial atención en los dos factores más influyentes: el umbral T2 y el error sistemático sobre Ch1/2. Al contrario, se debería dar poca importancia a los factores con baja influencia sobre la salida del modelo.

BIBLIOGRAFÍA

- BARBOSA P., GREGOIRE J. M., CARDOSO PEREIRA J. M. (1999a). An algorithm for extracting burned areas from time series of AVHRR GAC data applied at a continental scale. *Remote Sensing of Environment* 69: 253-263.
- BARBOSA P., STROPIANA D., GREGOIRE J. M., CARDOSO PEREIRA J. M. (1999b). An assessment of vegetation fire in Africa (1981-1991): Burned areas, burned biomass, and atmospheric emissions. *Global Biogeochemical Cycles* 13(4): 933-950.
- CROSETTO M., MORENO RUIZ J. A., CRIPPA B. (2001). Uncertainty propagation in models driven by remotely sensed data. *Remote Sensing of Environment* (in press).
- GOODCHILD M. F., GUOQING S., SHIREN Y. (1992). "Development and test of an error model for categorical data." *Int. Journal of Geographic Information Systems* 6: 87-104.
- HEUVELINK, G. B. M. (1998). Error propagation in environmental modelling with GIS. London, Taylor and Francis.
- JAMES M. E. (1994). "The Pathfinder data set: an improved coarse resolution data set for terrestrial monitoring." *International Journal of Remote Sensing* 15(17): 3347-3363.
- PRICE J. C. (1991). "Timing of NOAA afternoon passes." *International Journal of Remote Sensing* 12 (1): 193-198.
- SALTELLI A., CHAN K., SCOTT M. (Editors) (2000). Sensitivity analysis. New York, Wiley.
- VERMOTE E., TANRÉ D., DEUZÉ J. L., HERMAN M., MORCRETTE J. J. (1998). Second simulation of the satellite signal in the solar spectrum: an overview. *IEEE Transactions on Geoscience Remote Sensing* 35 (3): 675-686.

AGRADECIMIENTOS

Los autores agradecen al Dr. A. Saltelli y al Dr. S. Tarantola del JRC (Ispra, Italia) por el software de AS, y a la Unidad de Global Vegetation Monitoring del JRC (Ispra, Italia) por el soporte durante este trabajo.