

ESTIMACION DE UN PERFIL TEMPORAL DEL NDVI

P.C. Alvarez Esteban y C. Rueda Sabater

Dpto. de Estadística e Investigación Operativa. Facultad de Ciencias. Universidad de Valladolid. Tfno: (983) 423111. Fax: (983) 423013. E-mail: pedroc@dali.eis.uva.es

J. Gallego Pinilla

Institute for Remote Sensing Applications. J.R.C. Ispra (Italia)

E-mail: javier.gallego@cen.jrc.it

RESUMEN.- Este trabajo describe un procedimiento para estimar el perfil temporal del NDVI (Normalized Difference Vegetation Index), basado en métodos de regresión con pesos, que tiene en cuenta el carácter asimétrico del ruido asociado a los datos y la posible existencia de "depresiones" en dicho perfil. El ajuste final obtenido con este método parece mejorar en varios aspectos el obtenido con otros métodos basados en la cápsula convexa.

ABSTRACT.- This paper describes a procedure to estimate time profiles of the Normalized Difference Vegetation Index (NDVI) based on weighted least squares that take into account the biased noise associated with the data and the possible troughs in the profile. The final estimator obtained with this method seems to improve in several aspects the one obtained with an upper envelope approach.

1.- INTRODUCCION

Uno de los objetivos del proyecto MARS de la Unión Europea es la obtención de forma periódica de boletines que muestren el estado de la vegetación en Europa. Esta información será usada, entre otros fines, para detectar tan pronto como sea posible las zonas en las que la vegetación no se está desarrollando normalmente. Para conseguir esto se utiliza el índice de vegetación NDVI obtenido a partir de imágenes del sensor NOAA-AVHRR después de una serie de correcciones y filtros (de posición, angulares, etc...). Es sabido que, después de estas correcciones, el NDVI observado está aún afectado por un ruido (debido a la transparencia atmosférica y aerosoles principalmente) que hace que el valor que observamos tienda en general a ser menor que el verdadero valor del NDVI. Otra característica importante de los datos es que hay muchos días en los que debido a una presencia considerable de nubes, la observación tomada no es válida, por lo que se desecha.

Actualmente, el ajuste que se hace de la nube de puntos que forma el NDVI a lo largo de un año, es el obtenido mediante la cápsula convexa calculada cada 60 días. Este procedimiento, que puede considerarse más o menos adecuado para visualizar la tendencia general, resulta ineficaz para detectar los cambios de tendencia en el NDVI que se producen dentro del período de cálculo y que interesan por lo dicho anteriormente (ver la figura 1, entre los días 250 y 300 aproximadamente). Además presenta otras deficiencias como que el ajuste es muy dependiente del período de 60 días elegido.

Existen varias publicaciones en las que podemos encontrar una descripción más detallada del

problema (Sharman and Millot, 1993) así como del proceso de captación y manipulación de las imágenes (Taylor, 1993).

Lo que aquí presentamos forma parte de un trabajo para la U.E. en el que se presentan varias alternativas a los métodos existentes, que mejoran algunos aspectos de estos.

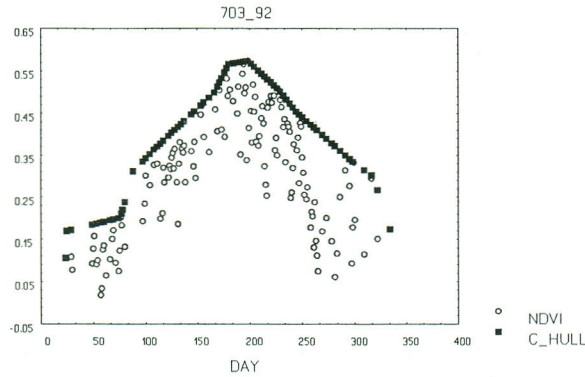


Figura 1.- Ajuste de la nube de puntos del NDVI por la cápsula convexa.

2.- DESCRIPCION DEL METODO

Para la obtención de un modelo adecuado proponemos un algoritmo basado en la realización de forma iterativa de regresiones polinómicas a trozos con pesos (Seber and Wild, 1989). Así, en cada etapa ajustamos el modelo dado por:

$$\begin{aligned}
 & f_1(x|\beta_1), \quad \text{si} \quad x \leq \alpha_1 \\
 & f_2(x|\beta_2), \quad \text{si} \quad \alpha_1 \leq x \leq \alpha_2 \\
 f(x|\beta, \alpha) = & \text{SCALESYM700} \begin{bmatrix} \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \end{bmatrix} \\
 & f_k(x|\beta_k), \quad \text{si} \quad \alpha_{k-1} \leq x
 \end{aligned} \tag{1}$$

donde cada f_i es un polinomio cúbico en el intervalo definido por los extremos de los intervalos a los que se llama nodos.

El método propuesto consiste en dividir cada año en un número de intervalos predeterminado ajustando en cada uno de ellos un polinomio cúbico e imponiendo la restricción de que en los nodos la función ajustada sea al menos continua. En la primera iteración se consideran todos los pesos iguales, pero en las siguientes, los pesos de las observaciones se obtienen a partir de una función de pesos que se aplica a los residuos estudentizados externos (Myers, 1990) de la iteración anterior. Esto significa que a partir de la segunda iteración no todas las observaciones tienen la misma importancia.

2.1.- Uso de una función cúbica a trozos

El uso de una curva cúbica a trozos en vez de un polinomio general está motivado por el

hecho de que en este último el comportamiento local determina completamente el comportamiento global, lo que podría dar lugar a que algunos cambios en la tendencia del NDVI no pudiesen ser reflejados por el estimador final.

2.2.- Elección de la función de pesos

El criterio de mínimos cuadrados está especialmente diseñado para distribuciones de la perturbación que sean simétricas respecto de cero, en cambio en nuestro caso nos encontramos con un ruido claramente asimétrico. Esta es la razón por la que en vez de plantear un modelo de regresión usual introducimos una regresión con pesos. Puesto que el ruido tiende a disminuir el verdadero valor del NDVI, las observaciones que después de una primera regresión sin pesos se encuentren por encima de la curva ajustada, observaciones con residuo positivo, serán las que más cerca estén del verdadero valor y por ello son las que mayor información dan sobre por dónde debe ir el perfil estimado del NDVI. Así, la función de peso que se seleccione debe ser una función que de mayor peso a las observaciones con residuos positivos y menor a las que tienen residuos negativos. Además cuanto mayor sea el residuo positivo mayor ha de ser el peso y cuanto menor sea el residuo negativo menor su peso, con la precaución de no hacer pesar demasiado observaciones extremadamente más grandes que las demás (outliers por arriba), seguramente erróneas, que pueden falsear el verdadero perfil del NDVI. Varias han sido las familias de funciones utilizadas en los ensayos, de las cuales la que mejores resultados proporciona es el arco tangente debidamente escalado para que sea positiva (ecuación (2) con $k=3$ y $m=0$). Las principales ventajas de esta función de pesos sobre otras con las que se probó son, aparte de proporcionar el resultado esperado, la rapidez de convergencia hacia el perfil final (con dos iteraciones es suficiente) y la menor sensibilidad hacia posibles outliers por arriba.

$$g(r)=2 \arctan(kr+m) + \pi \quad (2)$$

2.3.- Elección de la localización y del número de nodos

Se puede comprobar (ver figuras 2 y 4) que tanto el número de nodos como su posición influye en cierta medida en el perfil final. Podemos pensar en una serie de recomendaciones ideales (Wegman and Wright, 1983) de acuerdo a la forma funcional de la curva que estamos ajustando en cada trozo: (a).- poner tantos nodos como sea posible, asegurándose de que al menos haya 6 ó 7 puntos en cada intervalo (peligro: superajuste); (b).- no colocar más de un extremo (máximo o mínimo) y un punto de inflexión en cada intervalo; (c).- los extremos han de estar cerca del centro del intervalo; (d).- los puntos de inflexión han de estar cerca de los nodos. Sin embargo, es extremadamente difícil, además de costoso, implementar estas recomendaciones en un algoritmo automático. Por ello, proponemos una serie de alternativas que han demostrado funcionar bien y que se pueden resumir en dos puntos: 1º.- Número de nodos fijo (cinco ha demostrado ser un número suficiente) igualmente espaciados. 2º.- Añadir restricciones de derivabilidad en los nodos si se desea una curva más suave, y en este caso aumentar ligeramente el número de nodos.

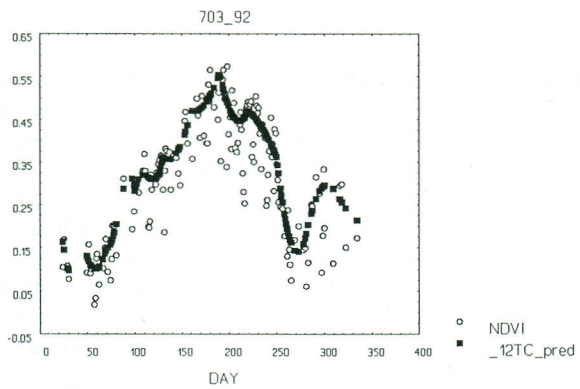
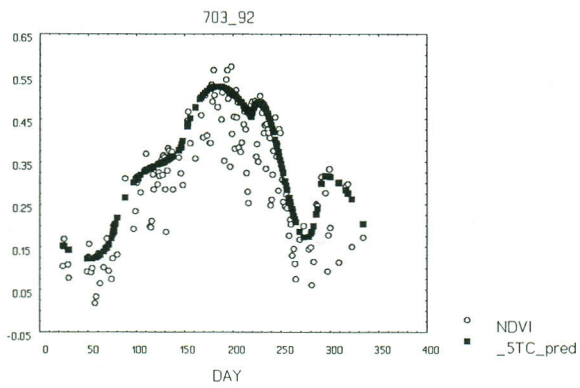
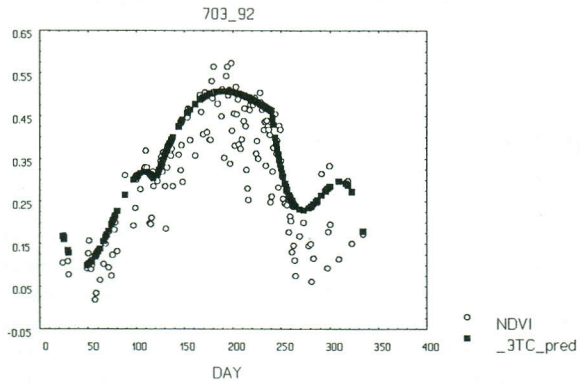


Figura 2.- Ajuste en la zona 703 con 2, 4 y 11 nodos respectivamente y con restricciones de continuidad.

3.- RESULTADOS

Como resultado, habiendo probado el algoritmo descrito en la sección 2 sobre varios conjuntos de datos proporcionados por la U.E., obtenemos unos ajustes más finos que los obtenidos mediante la cápsula convexa, puesto que recogen los cambios, incrementos o decrementos, 'locales' del NDVI a lo largo del año (comparar p.e. figuras 1 y 2). Además el método presenta otras ventajas como son la simplicidad computacional (con dos regresiones para cada zona basta), requisito de gran valor en un procedimiento que será automatizado, y la posibilidad de hacer ajustes de forma periódica (p.e.: cada 15 días).

En la figura 2 comprobamos el efecto del aumento del número de nodos. La figura 3 muestra ajustes con restricciones de derivabilidad en los que se ve claramente una mayor suavidad. En el primer plot de la figura 4, correspondiente a otra zona, comprobamos que el aumento del NDVI entre los días 150 y 200 (aprox.) no es recogido por el ajuste (con 2 nodos), mientras que si aumentamos el número de nodos (a 5, i.e., 6 trozos), este defecto es subsanado.

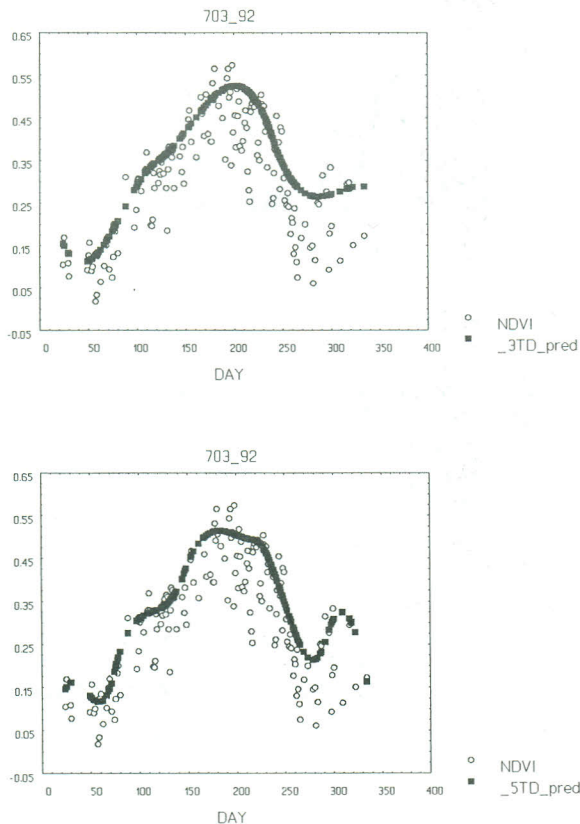


Figura 3.- Ajustes de la zona 703 con 2 y 4 nodos, y restricciones de derivabilidad.

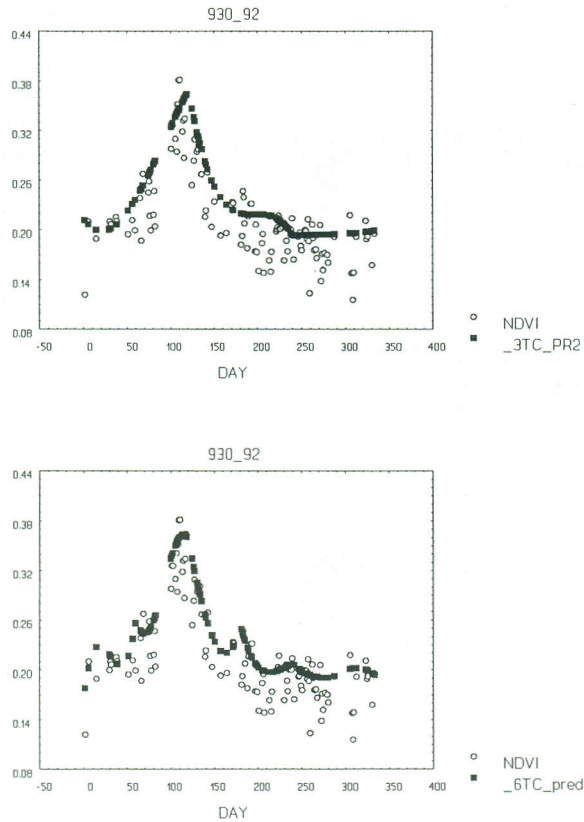


Figura 4.- Ajustes de la zona 930 con 2 y 5 nodos, y con restricciones de continuidad.

4.- REFERENCIAS

- Myers, R.H. 1990. *Classical and Modern Regression with Applications*. 2nd Edition PWS-KENT. Belmont, California 488 p.
- Taylor, J. 1993. Basic ideas of Image Analysis and some available Satellite Images. Imanal 1 - Imanal 2. En *EuroCourses: Remote Sensing Applied to Agricultural Statistics: Regional Inventories*. Atenas (Grecia).
- Seber, G.A.F. and Wild, C.J. 1989. *Nonlinear Regression*. John Wiley & Sons. New York.
- Sharman, M.J. and Millot, M. 1993. Comparing Time Profiles: Problems in Monitoring Vegetation. En *6th AVHRR Data Users' Meeting*. Belgirate (Italia).
- Wegman, E.J. and Wright, I.W. 1983. Splines in Statistics. *JASA* 78: 351-365.