MODELOS MIXTOS SEMIPARAMÉTRICOS BASADOS EN SPLINES CON PENALIZACIONES: APLICACIÓN EN LA ESTIMACIÓN DE BIOMASA

Mariola Sánchez-González [1], Hortensia Sixto Blanco [1] y María Durbán Reguera [2]

[1] Dpto. Selvicultura y Gestión de Sistemas Forestales CIFOR-INIA. Ctra. de La Coruña km 7,5. 28040-MADRID (España). Correo electrónico: msanchez@inia.es
[2] Dpto. de Estadística. Universidad Carlos III de Madrid. c/Madrid 126. 28903-GETAFE (Madrid, España)

Resumen

Los modelos mixtos semiparamétricos basados en splines con penalizaciones o P-splines son modelos muy flexibles que posibilitan representar como un modelo lineal mixto una relación no lineal entre dos variables. En este trabajo se va a tratar de introducir los P-splines y la forma de formularlos como modelos mixtos, así como sus posibles aplicaciones en el ámbito forestal. Además se va a presentar una aplicación más concreta y novedosa de los P-splines como modelos mixtos, su utilización para desarrollar ecuaciones que estimen la biomasa de las diferentes fracciones de un árbol cumpliendo la propiedad de la aditividad. Dicha aplicación se va a ilustrar mediante un caso concreto en el que se ha utilizado esta metodología para desarrollar ecuaciones de biomasa para distintos híbridos de chopo en plantaciones de turno corto para producción de biomasa para energía en España.

Palabras clave: *Funciones de suavizado, Híbridos de Populus, Biomasa, Cultivos de corta rotación*

INTRODUCCIÓN

Los modelos mixtos semiparamétricos basados en splines con penalizaciones o P-splines son modelos muy flexibles que posibilitan representar como un modelo lineal mixto una relación no lineal entre dos variables.

Los modelos semiparamétricos son aquellos modelos que utilizan técnicas de suavizado tipo spline los cuales no se pueden considerar como no paramétricos ya que utilizan parámetros para realizar el suavizado. Estos modelos pueden ser simples si incluyen una única variable o aditivos, es decir, modelos que son una suma de términos, en los que algunos pueden ser lineales.

Muchos de los datos utilizados en la investigación forestal se caracterizan, entre otras cosas, por tratarse de datos longitudinales, multidimensionales y cada vez más heterogéneos que permitan estudiar distintos aspectos de la biodiversidad y de la influencia del cambio global en los ecosistemas forestales. Para el análisis de datos longitudinales, los P-splines permiten ajustar modelos más flexibles en las que las diferencias específicas individuales son una función suave del tiempo, permitiendo obtener curvas distintas para cada factor (DURBÁN et *al.*, 2005). Recientemente se ha utilizado esta metodología para modelizar la relación entre la densidad especifica de la madera y el tiempo (JORDAN et *al.* 2008).

Hoy en día recoger distintos aspectos de la biodiversidad es una parte importante tanto de los inventarios, como de los diseños experimentales. El aplicar este tipo de modelos nos permite manejar distintos tipos de datos mediante una función adecuada. Tradicionalmente, los méto-

ISSN: 1575-2410

dos de suavizado se han utilizado para detectar tendencias, y después, basándose en los resultados obtenidos, aplicar un modelo paramétrico. Con los modelos semiparamétricos podemos cumplir los dos objetivos de manera simultánea. Por ejemplo, GUAN et *al.* (2006) utilizan esta metodología para detectar el efecto de las claras en variables microclimáticas y además detectar la tendencia de dicha respuesta en el tiempo.

En el caso de los datos multidimensionales, tanto espaciales como espacio-temporales, se pueden utilizar P-splines multidimensionales para su análisis. Éstos, permiten tener una cantidad de suavizado distinta en cada una de las dimensiones (CURRIE et *al.*, 2006) y tienen también una representación como modelos mixtos (LEE & DURBÁN, 2009).

A continuación se va a tratar de introducir los P-splines y la forma de formularlos como modelos mixtos. Para terminar, se va a presentar una aplicación más concreta y novedosa de los P-splines como modelos mixtos, su utilización para desarrollar ecuaciones que estimen la biomasa de las diferentes fracciones de un árbol cumpliendo la propiedad de la aditividad. Dicha aplicación se va a ilustrar mediante un caso concreto en el que se ha utilizado esta metodología para desarrollar ecuaciones de biomasa para distintos híbridos de chopo en plantaciones de turno corto para producción de biomasa para energía en España.

P-SPLINES

Los splines son polinomios a trozos que se unen en puntos llamados nodos y que se utilizan para estimar funciones de suavizado (smoothing), es decir, funciones suaves de los datos a las que no se les impone ninguna forma.

Dentro de los modelos de suavizado con splines distinguimos:

- Los splines de suavizado que utilizan tantos parámetros como observaciones, lo que hace que su implementación no sea eficiente cuando el número de datos es muy elevado.
- Los splines de regresión que pueden ser ajustados mediante mínimos cuadrados una vez que se han seleccionado el número de nodos, pero la selección de los nodos se hace mediante algoritmos bastante complicados.
- Los splines con penalizaciones que combinan lo mejor de ambos enfoques: utilizan menos parámetros que los splines de suavizado, pero la selección de los nodos no es tan determinante como en los splines de regresión.

Las tres razones fundamentales para el uso de este tipo de splines son las siguientes: son splines de bajo rango (el número de nodos es mucho menor que el número de datos), la introducción de penalizaciones relaja la importancia de la elección del número y la localización de los nodos, y la correspondencia entre los P-splines y el BLUP (Best Linear Unbiased Predictor) en un modelo mixto permite utilizar la metodología existente en el campo de los modelos mixtos.

Bases y Penalizaciones

Supongamos que disponemos de pares de datos (x_i, y_i), un modelo suavizado (para datos normales) vendrá dado por:

$$y_i = f(x_i) + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma^2)$$

donde $f(.)$ es una función suave de los datos a la que no se le impone ninguna forma. Es pues una generalización de un modelo de regresión que sería $y=Ba+\varepsilon$, siendo $B = B(x)$, una base construida a partir de x. Para estimar los coeficientes de la regresión se minimiza la función de mínimos cuadrados y la curva ajustada dependerá del tamaño de la base. Cuanto mayor sea la base menos suave es la curva, de modo que cuando el número de nodos coincide con el número de datos obtenemos una curva que interpola los datos.

Para evitar este problema, se introduce una penalización que en el caso de los P-spline es discreta y que se penalizan los coeficientes directamente, en vez de penalizar la curva, lo que reduce la dimensionalidad del problema. El tipo de penalización está estrechamente ligado al tipo de base. La función de mínimos cuadrados penalizada será:

$$S(a; y, \lambda) = (y-Ba)'(y-Ba) + \lambda a'Pa$$

donde P es una matriz que penaliza los coeficientes de forma suave y λ es el parámetro de suavizado.

Resumiendo, la metodología utilizada en los P-splines es la siguiente: (a) utilizar una base para la regresión, y (b) modificar la función de verosimilitud introduciendo una penalización basada en diferencias entre coeficientes adyacentes. Las diferencias entre los distintos modelos de P-splines vienen dadas por la elección de la base *B* y de la penalización *P*.

Bases y Nodos

Hay muchas maneras de calcular la base para la regresión: polinomios truncados (Ruppert et *al.*, 2003), thin plate splines (Green & Silverman, 1994; Wood, 2003), etc. Durbán (2009) recomienda el uso de B-splines (De Boor, 1977; Dierckx, 1993), ya que son numéricamente más estables que otras bases (como es el caso de los polinomios truncados). Un B-spline está formado por trozos de polinomios conectados entre si. Un ejemplo muy simple es un B-spline de grado 1, que está formado por dos trozos de polinomio lineal que se unen en un nodo. En general un B-spline de grado p tiene las siguientes propiedades:

- Consiste en $p+1$ trozos de polinomio de orden p
- Se unen en p nodos internos
- En los nodos las derivadas son continuas hasta el orden $p+1$
- Es positivo en el dominio expandido por $p+2$ y 0 en el resto
- Se solapa con $2p$ trozos de polinomios de sus vecinos
- Para cada valor de **x**, $p+1$ B-splines son no nulos

La selección y localización de los nodos no está hecha de antemano (como ocurre en el caso de los smoothing splines). Es suficiente con elegir un número moderadamente grande de nodos equidistantes. Autores como Ruppert et *al.* (2003) aconsejan elegir los K nodos en los K-quantiles de x, es decir, cada nodo k sería el cuantil k = (K + 1) de x. En cuanto al número de nodos, la mayoría de los autores utilizan como regla:

número de nodos = min {40, valores únicos de x/4}

Penalizaciones

El tipo de penalización dependerá del tipo de base que se esté utilizando. Eilers and Marx (1996) utilizan una penalización basada en la diferencias de orden *d* entre los coeficientes adyacentes de la bases de B-splines, este tipo de penalización es más flexible ya que es independiente del grado del polinomio utilizado para construir los B-splines.

Como hemos visto anteriormente, la penalización se añade a la función de mínimos cuadrados, dando lugar a una función de mínimos cuadrados penalizados. En general se utiliza d = 2, aunque se puede utilizar órdenes superiores o inferiores, dependiendo de la variabilidad de la curva y de la cantidad de ruido en los datos. En el caso de una penalización de orden 2 y con 5 coeficientes la matriz D es la siguiente:

$$D = \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

En la Figura 1 se ilustra lo que está haciendo la penalización: fuerza a los coeficientes a que

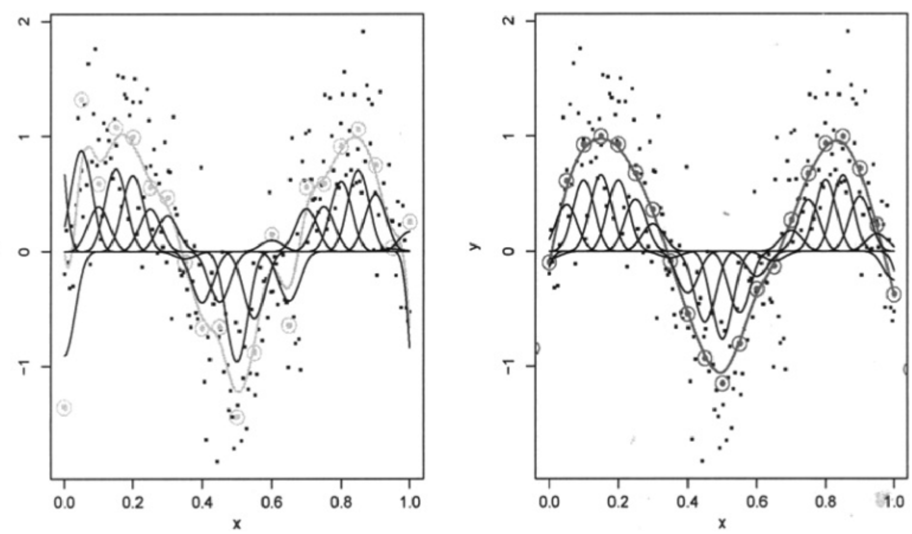

***Figura 1**. Curva estimada con 20 nodos en la base, sin penalizar los coeficientes (izquierda) y penalizando los coeficientes (derecha)*

sigan un patrón suave. Podemos ver el ajuste de una curva mediante B-splines sin y con penalización. Se muestran los funciones que forman las bases (las columnas de la matriz B) multiplicadas por los coeficientes, así como los coeficientes (en un círculo). En la parte izquierda vemos como el patrón errático de los coeficientes da lugar a una curva poco suave, en cambio en la parte derecha, cuando se les impone que se pase de un coeficiente a otro de forma suave, la curva también lo es.

Entre las propiedades de los P-splines con bases de B-splines hay que destacar que no tienen efecto de frontera (como le ocurre a los kernels), el efecto de frontera es el que hace que al extender la curva fuera del dominio de x la curva caiga rápidamente hacia 0, esto no pasa con los P-splines. Además, los P-splines ajustan de forma exacta los polinomios, es decir, si la curva es polinómica, un P-spline la reajustará exactamente. Por último, se conservan los momentos, es decir, que la media y la varianza de los valores ajustados será la misma que la de los datos sea cual sea el parámetro de suavizado, al contrario que los kernels que tienden a aumentar la varianza cuanto mayor es el suavizado.

Grados de libertad efectivos

La estimación de los parámetros del modelo y del parámetro de suavizado se hace de forma simultánea. El otro parámetro que queda por estimar es la varianza residual, σ^2, cuya estimación depende de los grados de libertad del modelo, por lo que hablamos de grados de libertad efectivos. Si en un modelo con P-splines tomamos $\lambda = 0$, entonces los grados de libertad del modelo coinciden con la dimensión de la base B menos el numero de restricciones del modelo. Por el contrario, si tomamos un valor grande de λ el modelo es muy poco flexible y tendrá muy pocos grados de libertad. Para definir los grados de libertad en este tipo de modelos, lo más lógico es usar una definición análoga a la que se utiliza en los modelos de regresión paramétricos:

$$d.f. = traza(H) \text{ donde } \hat{y} = Hy$$

En este caso:

$$H = B(B'B + \lambda P)^{-1}B'$$

por lo que,

$$d.f. = traza(H) = traza(B'B + \lambda P)^{-1}B'B$$

Y la traza de esa matriz variará dependiendo del parámetro de suavizado (λ).

P-SPLINES COMO MODELOS MIXTOS

La gran revolución de los P-splines producida en los últimos años es debida a la posibilidad de escribir un modelo no-paramétrico o semiparamétrico donde se utilizan P-splines como un modelo mixto. La ventaja que tiene el utilizar este enfoque es que nos permite: por un lado utilizar toda la metodología desarrollada para los modelos mixtos, y por otro, utilizar el software para modelos mixtos que está disponible en la mayoría de los programas estadísticos.

La formulación como un modelo mixto se basa en modificar las bases de forma que una curva se pueda descomponer como suma de un componente polinómico (del mismo orden que la penalización) y otro no polinómico, es decir, dado el modelo:

$$y = Ba + \varepsilon,\ \varepsilon \sim N(0,\sigma^2)$$

se reformula como:

$$y = X\beta + Zu + \varepsilon,\ \varepsilon \sim N(0,\sigma^2 I)$$

donde y es el vector de valores observados, X es la matriz de diseño asociada a los efectos fijos, β es el vector de parámetros fijos, Z es la matriz de diseño asociada a los efectos aleatorios, u es el vector de parámetros aleatorios y ε es un vector de términos del error.

Según la base que haya utilizado para los P-splines, X y Z tendrán una forma diferente. En el caso de los B-splines:

$$x = [1:x] \text{ y } Z = BU\Sigma^{-1/2}$$

donde U y Σ son matrices que forman parte de la descomposición en valores singulares de la matriz de penalización.

Una vez descrita esta nueva base, es inmediato establecer la conexión con un modelo mixto, donde:

$$y = X\beta + Zu + \varepsilon, \quad u \sim N(0,\sigma_u^2 I_{c-2}), \quad \varepsilon \sim N(0,\sigma^2 I)$$

siendo c el número de columnas de la base original B, y el parámetro de suavizado $\lambda=\sigma^2/\sigma_u^2$, lo cual hace que sea posible estimar el parámetro de suavizado junto con los otros parámetros del modelo. En el contexto de los modelos mixtos, el método estándar para la estimación de los parámetros de la varianza es el método de máxima verosimilitud restringida (REML) cuya expresión viene dada por:

$$l_R(\sigma_u^2\sigma_\varepsilon^2) = -\frac{1}{2}log|V| - \frac{1}{2}log|X'V^{-1}X| - \frac{1}{2}y'\left(V^{-1} - V^{-1}X\left(X'^{V^{-1}}X\right)^{-1}X'^{V^{-1}}\right)y$$

donde $V = \sigma_u^2 ZZ' + \sigma^2 I$. El vector de parámetros β y el vector de coeficientes aleatorios u son estimados como:

$$\hat{\beta} = \left(X'\hat{V}^{-1}X\right)^{-1} X'\hat{V}^{-1}y$$
$$\hat{u} = \hat{\sigma}_u^2 Z'^{\hat{V}^{-1}}\left(y - X\hat{\beta}\right)$$

En la Figura 2 se ilustra como al formular el P-spline como un modelo mixto el suavizado de la curva se incluye en la parte aleatoria del modelo. Podemos ver la relación entre valores ajustados y observados de un modelo, al tener en cuenta sólo la parte fija del modelo (gráfico de la izquierda), el modelo no capta la relación no lineal que existe entre las variables relacionadas en el modelo, lo cual se corrige al tener en cuenta el modelo mixto completo (gráfico de la derecha).

El hecho de que un P-spline se pueda escribir como un modelo mixto nos proporciona ventajas en dos ámbitos distintos: por un lado hace que se pueda flexibilizar la hipótesis de linealidad en infinidad de modelos, y por otro, es posible incluir estructuras complejas en los modelos usuales de suavizado. Por ejemplo, en el caso de datos correlados, si intentamos ajustar una curva sin tener en cuenta la correlación que hay en los datos, veremos que los métodos de selección del parámetro de suavizado van a elegir un valor del parámetro menor al que corresponde y por lo tanto la curva no va a ser suave. Si quisiéramos estimar tanto el parámetro de suavizado como los parámetros que determinan la estructura de correlación, necesitaríamos hacerlo de forma iterativa, lo que haría que el resultado final fuera muy sensible a la elección de los parámetros iníciales. Sin embargo, con los P-splines como modelos mixtos es inmediato el introducir una estructura de correlación, por ejemplo mediante la expansión del error a través de una estructura autoregresiva, y estimarla simultáneamente a la curva suave (CURRIE & DURBÁN, 2002).

La formulación como modelos mixtos permite además calcular los intervalos de confianza de la curva ajustada teniendo en cuenta tanto la varianza del error como el sesgo al cuadrado.

Por último mencionar que todo lo explicado hasta ahora se puede aplicar al caso de modelos lineales generalizados con penalizaciones en el caso de que trabajemos con datos no gaussianos.

APLICACIÓN: ADITIVIDAD BIOMASA

Una aplicación más concreta y novedosa de los P-splines como modelos mixtos es su utilización para desarrollar ecuaciones que estimen la biomasa de las diferentes fracciones de un árbol cumpliendo la propiedad de la aditividad. En concreto lo hemos utilizado para obtener ecuaciones de biomasa para 9 clones de híbridos de chopo en plantaciones de turno corto para producción de biomasa para energía en España.

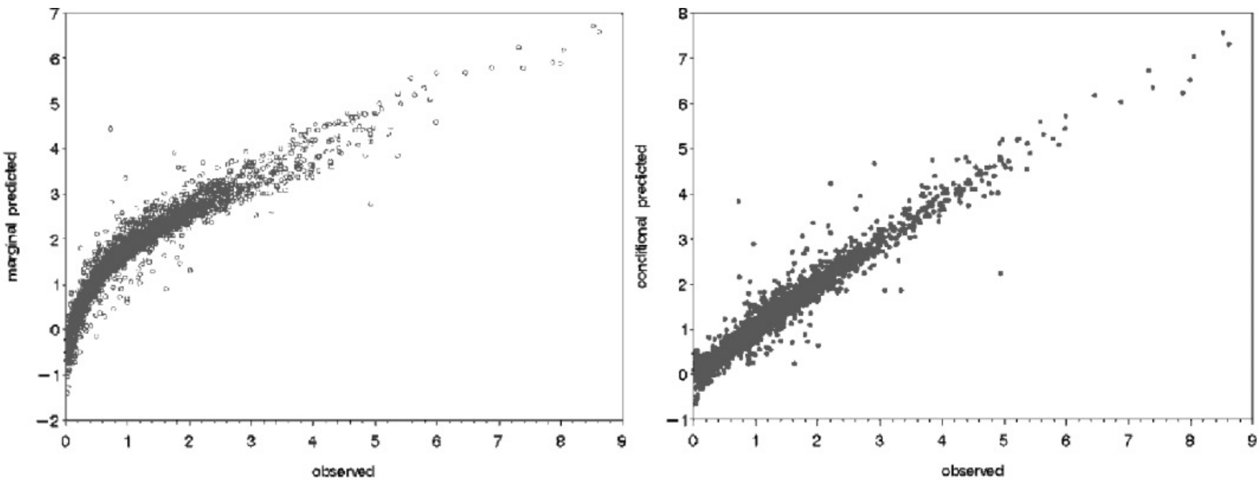

***Figura 2**: Relación entre los valores observados y los valores estimados teniendo sólo en cuenta la parte fija del modelo (izquierda) y los valores estimados teniendo en cuenta tanto la parte fija como la parte aleatoria (derecha)*

Material y Métodos

Se han establecido 3 ensayos en tres sitios localizados en las provincias de León, Madrid y Soria, con 4 bloques cada uno de ellos y 3 replicaciones en cada bloque. En total contamos con 2916 observaciones. Al acabar el turno de 3 años se midieron el diámetro normal y la altura total del árbol, y se pesaron los árboles separando el tronco de las ramas y pesándolos por separado. En la Tabla 1 se muestra las características de los datos utilizados para el desarrollo de las ecuaciones de biomasa.

Las ecuaciones se desarrollaron aplicando la metodología explicada. Según GOICOA et *al.* (2010) la propiedad de aditividad, es decir, que la suma de la estimación de los componentes del árbol sea igual a la predicción de la biomasa total, se puede conseguir utilizando la misma base y el mismo parámetro de suavizado. Para ello primero se ajusta la ecuación correspondiente a la biomasa total de árbol y a continuación el modelo es ajustado para los distintos componentes fijando el parámetro de suavizado.

En todos los modelos probados se ha incluido el clon como efecto fijo y el sitio y el bloque por sitio como efectos aleatorios. En un análisis preliminar se encontró que los efectos aleatorios asociado a la pendiente eran triviales por lo que finalmente se consideró que sólo el termino independiente variaba aleatoriamente entre sitio y bloque.

Los modelos que hemos probado han sido los siguientes. En primer lugar se ha probado un modelo semiparamétrico simple incluyendo sólo una variable. Pero como uno de los objetivos del trabajo era comparar la producción de biomasa de los distintos clones se ha introducido una interacción de curva por factor, de forma que el factor clon no afecta solo al termino independiente sino también a la función de suavizado de forma que tenemos una curva de suavizado distinta para cada uno de los clones.

$$Bt_{ijk} = \beta clon + f_{clon}(x_{ijk}) + u_i + v_{ij} + e_{ijk}$$

donde Bt_{ijk} biomasa total medida en árbol k situado dentro del bloque j del sitio i, β es el vector de parámetros asociado al efecto clon, $f_{clon}(x_{ijk})$ son nueve funciones de suavizado distintas en función del clon, construidas a partir del vector de valores de la variable predictora x para el árbol k dentro del bloque j en el sitio i, u_i and v_{ij} son los efectos aleatorios comunes a las observaciones tomadas en el mismo sitio y en el mismo bloque por sitio respectivamente, y e_{ijk} es el termino residual del error. Para simplificar se asume un parámetro de varianza común a las nueves funciones, pero los efectos aleatorios son independientes en cada una de ellas, traducido quiere decir que las curvas son diferentes pero con la misma cantidad de suavizado (DURBÁN et *al.*, 2005). Este modelo se ha probado incluyendo como variables explicativas diámetro normal (dbh), y las variables combinadas dh (dbh por altura total) y d2h (dbh al cuadrado por altura total).

Además se ha probado un modelo aditivo que incluye diámetro normal y altura como variables predictoras, la cuestión es que ambas variables no son independientes, hay una interacción entre ambas variables que hay que incluir de algún modo en el modelo. La manera más sencilla consiste en incluir una nueva variable que es el producto de las dos.

$$Bt_{ijk} = \beta clon + f_{clon}(d_{ijk}) + f_{clon}(h_{ijk}) + f_{clon}(dh_{ijk}) + u_i + v_{ij} + e_{ijk}$$

donde dh_{ijk} es el producto del diámetro normal y la altura para el árbol k dentro del bloque j en el sitio i.

Para el ajuste de los modelos se ha elaborado una macro de SAS (SAS INSTITUTE INC., 2004). La evaluación y comparación de los modelos se ha basado en un análisis de los residuos tanto gráfico como numérico. Los estadísticos analiza-

Variable	Media	Mínimo	Máximo	Desviación Típica
Diámetro (cm)	0,20	7,60	2,99	1,16
Altura (m)	1,31	9,60	5,40	1,63
d2h	0,05	532,55	65,65	64,51
Biomasa total (kg)	0,01	8,61	1,22	1,08
Biomasa fustes (kg)	0,01	7,07	1,04	0,91
Biomasa ramas (kg)	0,00	1,71	0,18	0,18

***Tabla 1**. Características de los datos utilizados*

dos fueron el sesgo, la raíz cuadrada del error cuadrático medio (RMSE) y el criterio de información de Akaike (AIC) (BURHAM & ANDERSON, 2002). Además se han elaborado unos gráficos que muestran la función de suavizado estimada con unos intervalos de confianza del 95%, junto a los residuos parciales.

Resultados y Discusión

Para modelizar la biomasa aérea total se ha utilizado como base un B-spline con 40 nodos. Como es habitual se han utilizado polinomios de orden tres como base y una penalización de grado 2.

En un análisis preliminar se ha comparado el ajuste de curvas específicas por clon frente a un suavizado común de todos los clones, así como el ajuste de curvas paralelas donde el factor clon sólo afecta al término independiente con las curvas de suavizado específicas por clon, encontrando en ambas comparaciones diferencias significativas en el valor de -2LL.

Entre los modelos semiparamétricos simples (MSS) el que mejor funciona es el que incluye d2h como variable explicativa. El modelo aditivo que incluye la interacción de diámetro normal y altura como un producto de ambas variables fue finalmente descartado ya que los parámetros relacionados con la altura no son significativos indicando que hay una sobreparametrización del modelo. Esto indica que la interacción entre diámetro normal y altura no se modeliza adecuadamente, lo que sugiere que podría ser más adecuado un modelo aditivo multidimensional que incluya la interacción diámetro-altura. En la Tabla 2 se muestran los resultados obtenidos al ajustar los modelos probados. Todos los parámetros son significativos a un nivel del 1%.

En la Figura 3 se muestra, como ejemplo, la relación entre la biomasa total aérea estimada y la variable d2h con intervalos de confianza al 95% junto a los residuos parciales (puntos), para uno de los clones estudiados. Los grados de libertad efectivos del modelo son 6.5, lo que viene a reforzar la hipotesis de no linealidad, ya que se puede considerar un modelo como no lineal cuando los grados de libertad efectivos son mayores que 2 (CURRIE & DURBÁN, 2002).

	Biomasa total			Biomasa fuste		Biomasa ramas	
Modelo	**sesgo**	**RMSE**	**AIC**	**sesgo**	**RMSE**	**sesgo**	**RMSE**
MSS (d)	0,0096	0,2744	-1.021,4	0,0079	0,2434	0,0017	0,0796
MSS (dh)	-0,0019	0,2601	-613,0	-0,0013	0,2009	-0,0006	0,0924
MSS (d2h)	0,0009	0,2103	-1.699,0	0,0010	0,1694	0,0000	0,0839

***Tabla 2**. Valores de los estadísticos de bondad de ajuste para los modelos semiparamétricos simples probados (MSS)*

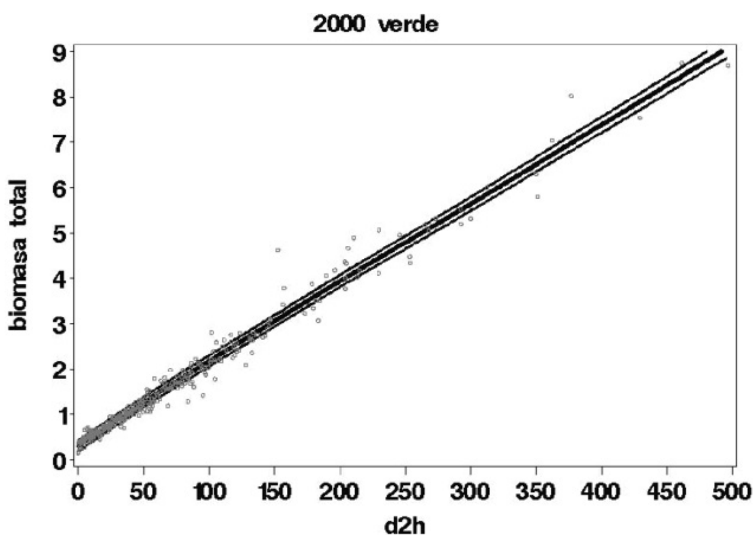

***Figura 3**. Relación entre la biomasa total aerea estimada en kg y la variable d2h con intervalos de confianza al 95% junto a los residuos parciales (punto) para el clon 2000 verde*

Entre la ventajas de esta metodología para desarrollar funciones de biomasa que cumplan la propiedad de aditividad destacan que se trata de modelos muy flexibles en los que la propiedad aditiva se consigue automáticamente considerando el mismo parámetro de suavizado sin ninguna dificultad extra en la estimación. Si los comparamos con otros procedimientos para conseguir la aditividad, esto modelos son fáciles de implementar y proporcionan buenos resultados. Mientras que por ejemplo el non-linear seemingly unrelated regression puede producir matrices de covarianzas singulares lo que no garantiza un resultado único (GOICOA et *al.*, 2010).

El principal inconveniente de esta metodología es que todos los componentes de la biomasa deben tener las mismas variables y la misma función de ponderación para corregir la heteroscedasticidad. Sin embargo en el caso que nos ocupa no consideramos que sea un problema, ya que normalmente la ecuación de biomasa en ramas es la que suele precisar de alguna variable más además del diámetro normal y la altura para mejorar las estimaciones. (DILLEN et *al*, 2007). Pero en nuestro caso, debido a que el turno es muy corto el porcentaje de biomasa en ramas es muy pequeña por lo que creemos que no merece la pena realizar la medición de una variable extra.

Agradecimientos

Los datos utilizados en la aplicación mostrada han sido financiados con el proyecto del Plan Sectorial INIA RTA2008-00025-C02-00 titulado "Cultivos forestales como productores de biomasa con fines energéticos".

BIBLIOGRAFÍA

BURNHAM, K.P. & ANDERSON, D.R.; 2002. *Model Selection and Multimodel Inference: a practical information-theoretic approach*, 2nd edition. Springer-Verlag. New York.

CURRIE, I.D. & DURBÁN, M.; 2002. Flexible smoothing with P-splines: a unified approach. *Stat. Model.* 4 : 333–349.

DE BOOR, C.; 1978. *A practical guide to splines.* Springer. Berlin.

DILLEN, S.Y.; MARRON, M.; BASTIEN, C.; RICCIOTTI, L.; SALANI, F.; SABATTIC, M.; PINEL, M.P.C.; RAE, M.A.; TAYLOR, G. & CEULEMANS, R.; 2007. Effects of environment and progeny on biomass estimations of five hybrid poplar families grown at three contrasting sites across Europe. *Forest Ecol. Manage.* 252: 12-23.

DIERCKX, P.; 1993. *Curve and surface fitting with splines.* Claredon. Oxford.

DURBÁN, M.; 2009. An introduction to smoothing with penalties: P-splines. *Bol. Estad. Invest. Op.* 25(3): 195-205.

DURBÁN, M.; HAREZLAK, J.; WAND, M.P. & CARROLL, R.J.; 2005. Simple fitting of subject-specific curves for longitudinal data. *Statist. Med.* 24: 1153-1167

EILERS, P.H.C.; 1999. Comment on "The analysis of designed experiments and longitudinal data by using smoothing splines", by Verbyla AP, Cullis BR, Kenward MG, Welham JS. *Appl. Stat.* 48: 269–311

EILERS, P.H.C. & MARX, B.D.; 1996. Flexible smoothing with B-splines and penalties. *Stat. Sci.* 11: 89–1216

GOICOA, T.; MILITINO, A.F. & UGARTE, M.D.; 2010. Modelling aboveground tree biomass while achieving the additivity property. *Environ. Ecol. Stat.* 18: 367-384.

GREEN, P. & SILVERMAN, B.; 1994. *Nonparametric regression and Generalized Linear Models. Monographs on Statistics and Applied Probability.* Chapman & Hall. London.

GUAN, B.T.; WENG S.; KUO, S.; CHANG, T.; HSU, H. & SHEN, C.; 2006. Analyzing the effects of stand thinning on microclimates with semiparametric smoothing splines. *Can. J. For. Res.* 36: 1641-1648.

JORDAN, L.; CLARK, A.; SCHIMLECK, L.R.; HALL, D.B. & DANIELS, R.F.; 2008. Regional variation in wood specific gravity of planted loblolly pine in the United States. *Can. J. For. Res.* 38: 698-710.

LEE, D-J. & DURBÁN, M.; 2009. Smooth-CAR mixed models for spatial count data. *Comput. Stat. Dat. Anal.* 53: 2968-2979.

RUPPERT, D.; WAND, M.P. & CARROLL, R.J.; 2003. *Semiparametric regression.* Cambridge University Press. New York.

SAS INSTITUTE INC.; 2004. *SAS/ETS 9.1 User's Guide.* SAS Institute Inc. Cary, NC.

WAND, M.P.; 2003. Smoothing and mixed models. *Comput. Stat.* 18: 223–249

WOOD, S.; 2003. Thin plate splines regression. *J.R. Stat. Soc.* 65(1): 95-114.