

MODELOS PARA LA GESTIÓN FORESTAL: UNA REVISIÓN DE LAS METODOLOGÍAS DE CONSTRUCCIÓN DE MODELOS DE MASA

J.G. Álvarez¹, M. Barrio¹, F. Castedo², U. Diéguez-Aranda¹, A.D. Ruiz-González¹

¹Dpto. Ingeniería Agroforestal, Universidad de Santiago de Compostela. Escola Politécnica Superior, Campus universitario, 27002 LUGO, Spain.

²Dpto. Ingeniería Agraria, Universidad de León. Escuela Superior y Técnica de Ingeniería Agraria, 24400 PONFERRADA, Spain.

Resumen. El avance de las metodologías de construcción de modelos forestales ha permitido que, en los últimos años y cada vez en mayor medida, las decisiones sobre la gestión y manejo de los bosques estén basadas en las predicciones obtenidas con dichos modelos. Los objetivos del modelo son, junto con la escala de aplicación, los factores más importantes que determinan la estrategia a seguir en su desarrollo. Del mismo modo que estos dos factores presentan una gran variabilidad, existe un gran número de metodologías estadísticas alternativas para la construcción de los modelos. En cada caso, la elección de la metodología más adecuada dependerá, entre otras cosas, de los datos disponibles, de la demanda de los usuarios finales, de la escala de representación y de las simplificaciones o relaciones que esté dispuesto a asumir el modelador.

En este trabajo se describen brevemente algunas de las metodologías más utilizadas para el desarrollo de modelos dinámicos de rodal basados en el empleo de regresión, así como algunas soluciones para los frecuentes problemas de violación de las hipótesis de partida que impone dicha metodología de estimación.

Introducción

Uno de los principales objetivos de la investigación forestal es el conocimiento de los procesos de crecimiento y producción de los árboles y los rodales forestales. Para ello, se aplican técnicas de modelización y se elaboran herramientas que permiten a los gestores comparar diferentes alternativas selvícolas, anticipándose a las consecuencias que éstas pueden tener sobre el sistema en términos de beneficios económicos, de los cambios en atributos forestales menos tangibles, en el valor estético o en la capacidad como hábitat para diferentes especies animales (Palahí, 2002).

Un modelo forestal es una abstracción de la dinámica de una masa forestal, abarcando crecimiento, mortalidad y otros cambios en la composición y la estructura de la masa. Los modelos forestales de crecimiento se pueden clasificar por las siguientes propiedades (Valsta, 1993): (1) la unidad de predicción; (2) el empleo de datos sobre la localización espacial de los árboles, es decir, modelos dependientes frente a independientes de la distancia; (3) modelos determinísticos frente a modelos estocásticos; y (4) modelos estadísticos o empíricos frente a modelos basados en procesos.

La unidad de predicción puede ser el rodal completo o un árbol individual. Algunos modelos de proceso incluso consideran mayor detalle, aunque el árbol es casi siempre el primer factor de agrupación. Los modelos estocásticos se diferencian de los modelos determinísticos en que los primeros incluyen componentes aleatorias y hacen uso de probabilidades (la componente aleatoria es la encargada de absorber las variaciones de crecimiento provocadas, entre otras, por las condiciones climáticas y estacionales y por los errores de medición). Los modelos estadísticos o empíricos se basan en medidas de longitud, masa y tiempo, mientras que los modelos de proceso, además de las mediciones anteriores, tratan de incorporar a la simulación una interpretación de la dinámica interna del rodal fisiológica y ecológicamente fundamentada, para lo que suelen incluir factores como la luz, el

ciclo de nutrientes, el balance de carbono, el índice de área foliar, etc. Su grado de complejidad varía enormemente, siendo mayor cuanto más se alejan del enfoque empírico clásico de los modelos de crecimiento. En los últimos años se han desarrollado algunos modelos híbridos entre los dos anteriores, que incorporan variables edáficas y climáticas a los modelos empíricos, que combinan las ventajas de ambos tipos de modelos (Woollons et al., 1997; Snowdon et al., 1998).

Los objetivos de la gestión forestal y los recursos disponibles determinan en gran medida el tipo de modelo de crecimiento a desarrollar, la metodología más adecuada para su elaboración, y consecuentemente los datos necesarios y la resolución para las estimaciones (Vanclay, 1994). Los primeros modelos de crecimiento que se construyeron eran de carácter “estático”, ya que representaban únicamente una de las posibles evoluciones de la densidad, en contraposición con los modelos “dinámicos”, que permiten simular diferentes evoluciones de aquella en el tiempo.

Los modelos de crecimiento estáticos pueden dar lugar a buenos resultados para rodales no aclarados o para rodales sometidos a un rango reducido de tratamientos estándar, para los cuales se dispone de un gran número de datos experimentales. Sin embargo, para poder proyectar la masa considerando un amplio rango de regímenes selvícolas, o cuando se dispone de datos procedentes de muy diversos tipos de parcelas experimentales en las que se han aplicado distintos tratamientos, es necesario elaborar modelos dinámicos, ya que son mucho más indicativos de la relación causa-efecto entre densidad de la masa y producción en volumen (Alder, 1980). Estos modelos no requieren la determinación de un número fijo de calidades, en ellos la evolución de la densidad es, hasta cierto punto, libre, y con frecuencia suele incluirse el efecto de las operaciones culturales, representando por tanto alternativas de simulación más completas y realistas que las tablas de producción clásicas, y ofreciendo mayores posibilidades en la gestión de las masas.

Los modelos de carácter dinámico predicen fundamentalmente tasas de cambio, es decir, crecimientos de alguna variable bajo condiciones distintas de densidad. A partir de estas ecuaciones se obtiene, bien por integración o por la aplicación sucesiva de incrementos, la evolución del rodal en el tiempo (Alder, 1980). El principio de este tipo de modelos debe buscarse en la teoría matemática de sistemas, cuyos fundamentos básicos son (García, 1988): (i) una descripción del estado de la masa a una cierta edad empleando variables que lo definan y a partir de las cuales se puedan determinar otras variables de interés; (ii) unas funciones de transición que definan la tasa de cambio del sistema y que dependan del estado inicial; y (iii) unas funciones de control que regulen las modificaciones entre las variables de estado debidas a cambios instantáneos originados por perturbaciones (claras, enfermedades, incendios, etc.).

En este trabajo se van a describir brevemente algunas de las metodologías más utilizadas en la construcción de modelos forestales. Dado que cada uno de los tipos de modelo descrito tiene su propia estructura que define las metodologías de construcción más adecuadas, la revisión se va a basar en los modelos de masa, puesto que, generalmente, es el paso inicial a llevar a cabo cuando se programa el desarrollo de modelos de crecimiento para cualquier especie. Además, los modelos de masa son los más adecuados para la planificación de la gestión de plantaciones forestales, ya que representan un buen compromiso entre generalidad y precisión de las estimaciones (García, 1988). En la Figura 1 se muestra el esquema básico de un modelo dinámico de masa.

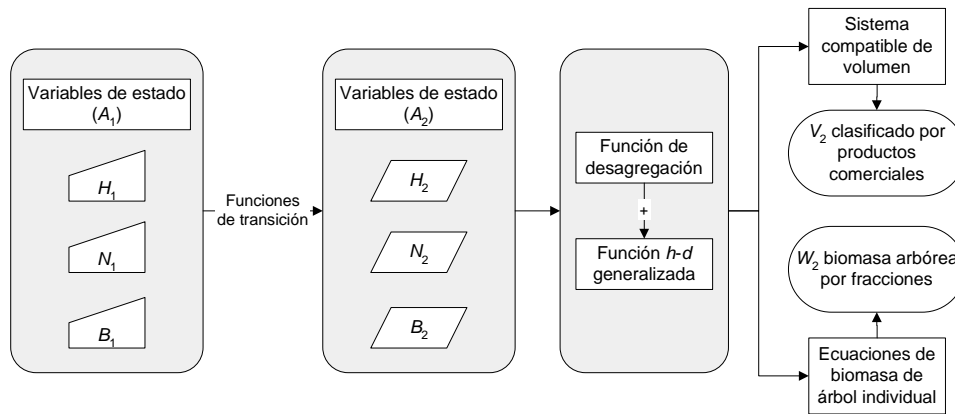


Figura 1. Estructura básica del modelo de masa a desarrollar.

De una forma sintética, en el modelo elaborado se parte de la situación de una masa en un instante determinado A_1 definido por tres variables de estado: altura dominante H_1 , número de pies/ha N_1 , y área basimétrica B_1 , y se pretende llegar a estimar el volumen por hectárea o la biomasa para una edad de proyección dada A_2 . Las tres variables de estado mencionadas se proyectan hacia el futuro mediante la utilización de sus correspondientes funciones de transición.

Desarrollo del modelo de masa

Funciones de transición

El modelo descrito en la Figura 1 se basa en el empleo de tres variables de estado (altura dominante, número de pies por hectárea y área basimétrica) y, por tanto, es necesario ajustar tres funciones de transición.

La alternativa más empleada para obtener este tipo de funciones de transición es la utilización de ecuaciones dinámicas de la forma $Y = f(Y_0, t_0, t)$, donde Y es el valor de la función a la edad t e Y_0 es el valor de la misma función a la edad t_0 . Estas ecuaciones dinámicas cumplen dos propiedades esenciales: (i) son invariantes con respecto a la edad de referencia (*base-age invariance*), es decir, cualquier par de puntos (t_0, Y_0) de una curva concreta puede emplearse para hacer predicciones a la edad t y las estimaciones obtenidas serán siempre las mismas; y (ii) son invariantes con respecto al intervalo de proyección (*path invariance*), de modo que el resultado de proyectar primero de t_0 a t_1 y luego de t_1 a t_2 es el mismo que el obtenido en un solo paso de t_0 a t_2 . Además, en el caso de que la variable de estado sea la altura dominante, una única ecuación dinámica define conjuntamente la curva de crecimiento en esta variable y el modelo de índice de sitio.

Bailey y Clutter (1974) fueron los primeros en formalizar la propiedad de invarianza con respecto a la edad de referencia, aunque en un sentido más amplio, puesto que implica que la estimación de los parámetros de la ecuación dinámica debe realizarse empleando una metodología de ajuste que sea también *base-age invariant*, es decir, que las estimaciones de los parámetros no dependan de un punto arbitrariamente seleccionado como referencia y que permita el empleo de todos los datos disponibles. Estos autores propusieron una metodología para obtener ecuaciones dinámicas con estas características y que es conocida en la literatura forestal como *Algebraic Difference Approach* (ADA). Esta metodología esencialmente se basa en la premisa de que es posible modelizar el crecimiento con una familia de curvas originadas por un modelo de crecimiento base con todos los parámetros comunes excepto uno, que es un parámetro específico de la estación forestal. Se pueden obtener, por tanto, para cada modelo de crecimiento base, tantas ecuaciones dinámicas como parámetros tenga dicho modelo.

Cieszewski y Bailey (2000) propusieron una generalización de la metodología ADA, denominada *Generalized Algebraic Difference Approach* (GADA) cuya principal ventaja radica en que se pueden obtener ecuaciones dinámicas a partir de un modelo de crecimiento base permitiendo que más de un parámetro dependa de las condiciones específicas de la estación forestal.

A continuación se presenta un ejemplo de obtención de ecuaciones dinámicas empleando la metodología GADA basadas en el modelo base de Bertalanffy-Richards, cuya forma integral simplificada es:

$$Y = a_1(1 - \exp(-a_2t))^{a_3} \quad (1)$$

siendo a_i los parámetros a estimar. Como hay tres parámetros se pueden obtener tres ecuaciones dinámicas suponiendo que sólo uno de ellos depende de las condiciones de la estación forestal (en este caso las metodologías GADA y ADA proporcionan los mismos modelos). Si se considera que la asíntota horizontal (a_1) es el parámetro específico de la estación se obtiene el siguiente desarrollo:

$$Y = X(1 - \exp(-a_2t))^{a_3} \quad (2)$$

La solución de X en la ecuación (2) con los valores iniciales t_0 e Y_0 ($X_0 = f(t_0, Y_0)$) para t e Y , respectivamente, es:

$$X_0 = Y_0 / (1 - \exp(-a_2t_0))^{a_3}$$

Sustituyendo el valor de X en la ecuación (2) se obtiene la siguiente ecuación dinámica:

$$Y = Y_0 \left(\frac{1 - \exp(-a_2t)}{1 - \exp(-a_2t_0)} \right)^{a_3} \quad (3)$$

Del mismo modo, considerando los parámetros a_2 y a_3 como específicos de la estación forestal se llega a las siguientes ecuaciones dinámicas, respectivamente:

$$Y = a_1 \left(1 - \left(1 - \left(\frac{Y_0}{a_1} \right)^{1/a_3} \right)^{t/t_0} \right)^{a_3} \quad (4)$$

$$Y = a_1 \left(\frac{Y_0}{a_1} \right)^{\frac{\ln(1 - \exp(-a_2t))}{\ln(1 - \exp(-a_2t_0))}} \quad (5)$$

Las dos últimas ecuaciones tienen una única asíntota para toda la familia de curvas. Sin embargo, en el caso de la ecuación dinámica para la variable de masa altura dominante, una propiedad importante es que la familia de curvas sea polimórfica y con múltiples asíntotas. Para ello, si se emplea nuevamente la ecuación de Bertalanffy-Richards, es necesario que al menos dos de los parámetros sean específicos de la estación forestal. En este caso se puede derivar una ecuación GADA considerando que tanto el parámetro a_1 como el parámetro a_3 dependen de la estación X . El modelo base se puede reparametrizar para obtener una solución para la variable X . Para ello la asíntota (a_1) se considera que es una función exponencial de X , y el parámetro a_3 es una función lineal de la inversa de X (Krumland y Eng, 2004). La elección de la expresión más adecuada de cada parámetro en función de X depende de los datos y del dominio de edades de aplicación de la ecuación (Cieszewski y Bailey, 2000):

$$Y = \exp(X) (1 - \exp(-a_2t))^{(a_4 + a_5/X)} \quad (6)$$

Para resolver la ecuación anterior en función de X es necesario calcular las raíces de una ecuación cuadrática y seleccionar la más adecuada para obtener la ecuación dinámica. La solución de X en la ecuación (6) con los valores iniciales t_0 e Y_0 es

$$X_0 = \frac{1}{2} \left((\ln Y_0 - a_4 L_0) \pm \sqrt{(\ln Y_0 - a_4 L_0)^2 - 4a_5 L_0} \right) \text{ siendo } L_0 = \ln(1 - \exp(-a_2t_0))$$

Seleccionando la raíz real y positiva (i.e., la que va precedida por el signo positivo en la raíz cuadrada) y sustituyendo en la ecuación (6) se obtiene la siguiente ecuación dinámica que da lugar a una familia de curvas polimórficas con múltiples asíntotas:

$$Y = Y_0 \left(\frac{1 - \exp(-a_2 t)}{1 - \exp(-a_2 t_0)} \right)^{(a_4 + a_5 / X_0)} \quad (7)$$

Las variables de estado que se modelizan con ecuaciones dinámicas (altura dominante, número de pies por hectárea o área basimétrica) son medidas en campo y están sujetas a errores de medición. Al ajustar la ecuación dinámica se asume que la variable Y no tiene error cuando está a la derecha de la igualdad pero tiene error cuando está en la parte izquierda de la igualdad, lo que plantea un problema puesto que, en esta situación, las técnicas de estimación convencionales aplicadas a modelos lineales dan lugar a estimaciones de los parámetros sesgadas (Myers, 1990). Por extensión, el problema es similar en modelos no lineales como las ecuaciones dinámicas expuestas, por lo que los valores de la variable de estado que aparecen en la parte derecha de la igualdad deben ser valores predichos que no pueden ser conocidos hasta que los parámetros hayan sido estimados (Krumland y Eng, 2004).

Se han utilizado diferentes metodologías para resolver este problema (p.ej., Tait et al., 1988; Gregoire et al., 1995; Gregoire y Schabenberger, 1996; Cieszewski et al., 2000). Uno de los métodos más sencillos de aplicar es el propuesto por Cieszewski et al. (2000) basado en el empleo de variables *dummy*, que en realidad es una simplificación de los modelos de efectos mixtos propuestos por Lindstrom y Bates (1990). Información detallada de esta metodología y de sus posibles aplicaciones se puede encontrar en Cieszewski (2003).

En la metodología *dummy* los valores iniciales de la ecuación dinámica son los mismos para todos los datos de una unidad de muestreo determinada (árbol o parcela). La edad inicial puede elegirse arbitrariamente para cada individuo, dentro de unos límites (p.ej., no puede ser cero). El valor de la variable de estado correspondiente a la edad inicial para cada unidad de muestreo se estima simultáneamente al resto de los parámetros globales de la ecuación dinámica. A diferencia de las metodologías de estimación tradicionales que requieren una elección de la edad base previa al ajuste para forzar a la ecuación a pasar por el punto variable de estado/edad, el método de variables *dummy* tiene en cuenta que cada medición tienen error y, por tanto, no fuerza a que la ecuación pase a través de ninguna medición concreta. En lugar de eso, la curva se ajusta a las tendencias individuales de los datos originales. Para aplicar este método se necesita contar con, al menos, dos mediciones para cada unidad de muestreo (árbol o parcela), y el número de unidades de muestreo debe ser mayor que el número de parámetros globales de la ecuación dinámica.

Usando como ejemplo de aplicación la ecuación (3), el valor inicial de la variable de estado Y_0 se sustituye por una suma de términos que contienen un parámetro específico y una variable para cada unidad de muestreo:

$$Y = (Y_{01}I_1 + Y_{02}I_2 + \dots + Y_{0n}I_n) \left(\frac{1 - \exp(-b_2 t)}{1 - \exp(-b_2 t_0)} \right)^{b_3} \quad (8)$$

donde I_i es la variable *dummy* que vale 1 para la unidad de muestreo i y 0 para las demás.

Una vez ajustada la ecuación dinámica, las estimaciones de la variable de estado para una nueva unidad de muestreo se obtienen sustituyendo los valores de los parámetros específicos por la medición real obtenida en la nueva unidad de muestreo.

A modo de ejemplo, en la Figura 2 se comparan las curvas de calidad de estación obtenidas para *Pinus radiata* D. Don en Galicia empleando las ecuaciones dinámicas descritas (Diéguez-Aranda et al., 2005) y la metodología de variables *dummy*. Se puede observar como la ecuación con dos parámetros específicos para cada estación forestal muestra una mejor adaptación a los datos reales que las ecuaciones con un único parámetro específico.

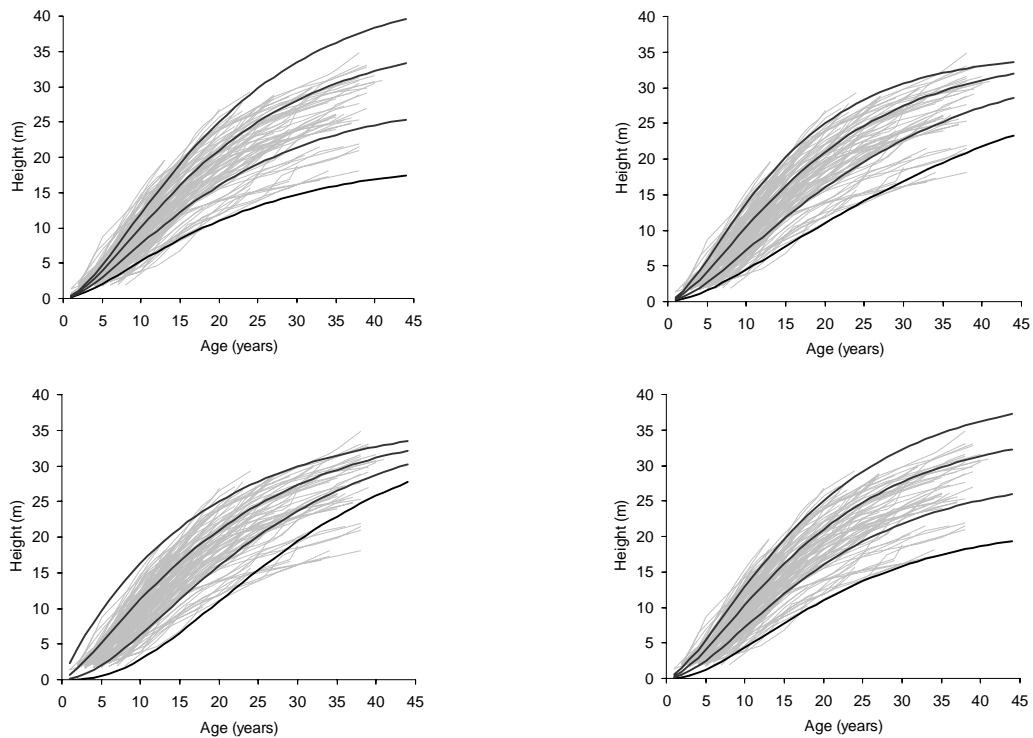


Figura 2. Curvas para índices de sitio de 11, 16, 21 y 25 m a la edad de referencia de 20 años para modelos dinámicos obtenidos a partir del modelo base de Bertalanffy-Richards (ecuación 1) ajustados con el método de variables *dummy* superpuestas sobre los perfiles de los árboles empleados en el ajuste (Diéguez-Aranda et al., 2005). Arriba-izq.: a_1 depende de la estación (ecuación 3). Arriba-der.: a_2 depende de la estación (ecuación 4). Abajo-izq.: a_3 depende de la estación (ecuación 5). Abajo-der.: a_1 y a_3 dependen de la estación (ecuación 7).

En la Figura 3 se comparan las curvas de calidad de estación obtenidas con una ecuación estática (Sánchez et al., 2003), con las proporcionadas mediante la ecuación dinámica (7) ajustada empleando el método de las variables *dummy*. Se observa claramente que la segunda opción permite incluso corregir la caída de la asíntota que se produce con frecuencia en la construcción de curvas de calidad de estación como consecuencia de la falta de datos de unidades de muestreo a edades avanzadas y en las mejores calidades de estación.

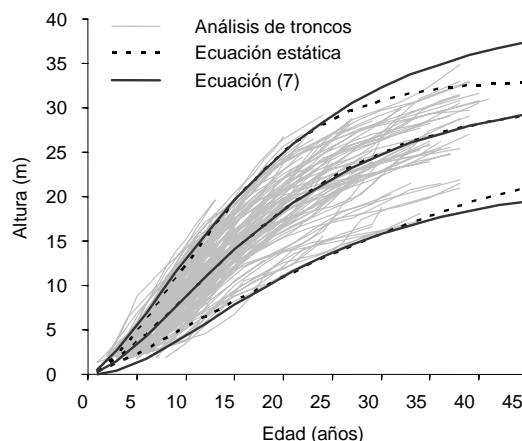


Figura 3. Comparación de las curvas con índices de sitio de 11, 18.5 y 25 m a la edad de referencia de 20 años obtenidas con un modelo que no es invariante con respecto a la edad de referencia (ecuación estática obtenida por Sánchez et al., 2003) y la ecuación dinámica (7) ajustada con el método de variables *dummy*.

El método de las variables *dummy* funciona bien para pequeños conjuntos de datos (p.ej. 100-200 unidades de muestreo). Para bases de datos mayores (más de 800 unidades de

muestreo), esta metodología puede tardar horas en converger e incluso no llegar a converger (Krumland y Eng, 2004). En estos casos es más adecuado emplear métodos como el denominado *nested iterative procedure* propuesto por Tait et al. (1988), que conceptualmente persigue los mismo objetivos que el método de las variables *dummy* (Cieszewski et al., 2000; Krumland y Eng, 2004).

Desagregación del modelo de masa

Una vez obtenidas las funciones de transición de las tres variables de estado fundamentales se puede estimar el estado de una masa cualquiera a la edad t conocido su estado inicial a la edad t_0 . El modelo se puede completar añadiendo las denominadas funciones de estado, que no son más que ecuaciones estáticas que permiten determinar el valor de otras variables de estado a la edad t a partir de los valores de las variables fundamentales a dicha edad (p.ej. el volumen total, la altura media, la biomasa por fracciones).

En el caso de que se pretenda obtener información más detallada de la estructura del rodal es necesario añadir un módulo de desagregación al modelo. Esencialmente, un módulo de desagregación permite determinar el número de árboles del rodal que pertenecen a una determinada clase de dimensión (habitualmente clases de diámetro, clases de área basal, clases de volumen o clases combinadas de diámetro y altura). Habitualmente, el proceso de desagregación se realiza empleando métodos paramétricos basados en el empleo de una función de densidad de probabilidad (*probability density function*, PDF) univariante o bivariante, dependiendo de si se va a estimar el número de árboles para una única categoría de clases de dimensión (p.ej. diámetro o área basal o volumen) o para dos categorías combinadas (p.ej. diámetro y altura simultáneamente).

Existen muchas funciones de densidad de probabilidad que permiten describir la estructura de una masa en clases de dimensión (p.ej. Charlier, Normal, Beta, Gamma, Johnson S_B , Weibull). De todas ellas, la más empleada en el campo forestal ha sido la propuesta por Weibull, cuya función de densidad es:

$$f(x) = \left(\frac{c}{b}\right) \left(\frac{x-a}{b}\right)^{c-1} e^{-\left(\frac{x-a}{b}\right)^c} \quad (9)$$

siendo x la variable aleatoria que define las clases de dimensión, a un parámetro de situación que marca el origen de la función, b un parámetro de escala y c un parámetro de forma.

Los parámetros de esta función se pueden estimar empleando distintas metodologías que pueden clasificarse en dos grandes grupos (Vanclay, 1994): estimación de parámetros (*parameter estimation*) y recuperación de parámetros (*parameter recovery*). En general, la segunda metodología permite obtener estimaciones más precisas incluso para proyecciones a largo plazo (p.ej. Borders y Patterson, 1990; Torres-Rojo et al., 2000).

El método de recuperación de parámetros se basa en relacionar variables de estado (normalmente altura dominante, área basimétrica o número de pies por hectárea) con los percentiles o los momentos de la distribución de la variable que define las clases de dimensión (habitualmente el diámetro) para, posteriormente, obtener los parámetros a partir de dichos percentiles o momentos. La ventaja del método de los momentos radica en que directamente garantiza la compatibilidad entre el área basimétrica obtenida con la función de transición y el área basimétrica obtenida por suma de las áreas basales de cada árbol del rodal al desagregar.

Como ejemplo se muestra el sistema de dos ecuaciones que permite obtener los parámetros de la función de Weibull (ecuación 9) en el caso concreto de que la variable que define las clases de dimensión sea el diámetro y asumiendo que el parámetro de situación a es igual a cero:

$$d_g^2 = \left(\bar{d}^2 / \Gamma^2 \left(1 + \frac{1}{c} \right) \right) \left(\Gamma \left(1 + \frac{2}{c} \right) - \Gamma^2 \left(1 + \frac{1}{c} \right) \right) + \bar{d}^2 \text{ y } b = \bar{d} / \Gamma \left(1 + \frac{1}{c} \right) \quad (10)$$

donde \bar{d} es el diámetro medio, d_g es el diámetro medio cuadrático y $\Gamma(i)$ es la función Gamma definida como $\Gamma(i) = \int_0^\infty x^{i-1} e^{-x} dx$ y cuyos valores están tabulados.

El sistema anterior se resuelve por procedimientos iterativos una vez que se conocen el diámetro medio cuadrático y el diámetro medio aritmético. El primero se puede obtener a partir de los valores de área basimétrica y número de pies por hectárea del rodal. El segundo debe estimarse mediante una ecuación ajustada para tal fin. Puesto que el diámetro medio aritmético es siempre menor o igual que el cuadrático, el modelo debe tener en cuenta esta relación. Una posible expresión que garantiza tal condición es la siguiente (Frazier, 1981):

$$\bar{d} = d_g - e^{\mathbf{X}\boldsymbol{\beta}} \quad (11)$$

donde \mathbf{X} es un vector de variables dependientes cuyos valores deben poder obtenerse de las funciones de transición o de estado (p.ej. altura dominante, número de pies, edad) y $\boldsymbol{\beta}$ es un vector de parámetros a estimar.

Función h-d generalizada

Una vez desagregado el modelo por clases de dimensión se pueden utilizar ecuaciones estáticas de árbol individual para estimar volúmenes comerciales y/o biomasa arbórea por fracciones. Lo habitual es que estas ecuaciones estáticas dependan tanto del diámetro del árbol como de su altura (p.ej. funciones de perfil, tarifas de cubicación, tarifas de biomasa) por lo que, en el caso de que se hayan utilizado funciones de densidad univariantes (como es el caso de la función Weibull descrita) es necesario estimar las alturas de cada árbol.

Generalmente dos árboles con el mismo diámetro dentro de un mismo rodal no tienen la misma altura, por lo que, para simular la variabilidad natural de las alturas reales observadas, es necesario añadir a los modelos determinísticos altura-diámetro una componente estocástica basada en una función de distribución de probabilidad (Parresol y Lloyd, 2004). La estimación estocástica utiliza el error estándar de una nueva estimación de forma similar a la empleada para la construcción del intervalo de predicción en un modelo de regresión. La diferencia radica en que, en vez de usar el valor correspondiente a un límite fijo, por ejemplo -1,96 y 1,96 respectivamente para una probabilidad del 2,5% y el 97,5% e infinitos grados de libertad, el valor de t se sustituye por el obtenido aleatoriamente a partir de la inversa de la función de distribución normal. Por tanto, la expresión utilizada para asignar las alturas estocásticamente a cada árbol de la muestra es:

$$\hat{y}_{i(estoc)} = \hat{y}_i + F_U^{-1} s_{\hat{y}_{i(nuevo)}} \quad (12)$$

donde $\hat{y}_{i(estoc)}$ es la estimación estocástica de la altura, \hat{y}_i es la estimación determinística (obtenida mediante el modelo ajustado), F_U^{-1} es el valor de la inversa de la función de distribución normal estándar para la variable aleatoria U generada uniformemente en el intervalo (0,1), y $s_{\hat{y}_{i(nuevo)}}$ es el error estándar de la predicción para un nuevo individuo, cuyo valor se obtiene como:

$$s_{\hat{y}_{i(nuevo)}} = \sqrt{\text{var}(\hat{y}_{i(nuevo)})} = \sqrt{EMC_{\psi} \boldsymbol{\psi}_i(\hat{\boldsymbol{\theta}}) + \mathbf{z}(\mathbf{b})'_i \mathbf{S}^2(\mathbf{b}) \mathbf{z}(\mathbf{b})_i} \quad (13)$$

siendo EMC_{ψ} el error medio cuadrático obtenido en el ajuste del modelo determinístico teniendo en cuenta los pesos empleados si se ha realizado una corrección de heterocedasticidad; $\boldsymbol{\psi}_i(\hat{\boldsymbol{\theta}})$ es el i -ésimo elemento de la diagonal de la matriz de pesos empleados en el ajuste del modelo determinístico; $\mathbf{S}^2(\mathbf{b})$ es la matriz de varianzas-covarianzas

del estimador de mínimos cuadrados \mathbf{b} del vector de parámetros $\boldsymbol{\beta}$, y $\mathbf{z}(\mathbf{b})_i$ es el vector de derivadas parciales del modelo determinístico con respecto a cada uno de sus parámetros.

Esta metodología se puede emplear siempre que se requieran diferentes estimaciones de una variable dependiente para los mismos valores de las variables independientes. En el caso de la relación altura-diámetro descrita se ha utilizado la inversa de una distribución normal para generar variabilidad; no obstante, en cada situación se debe analizar previamente cual es la distribución específica más conveniente (ver p.ej. Parresol y LLOYD, 2004).

En la Figura 4 se muestran las alturas observadas (blanco), las estimadas con un modelo determinístico (gris claro) y las estimadas con la metodología estocástica descrita (gris oscuro) por clases diamétricas para un rodal de *Pinus sylvestris* L. en Galicia (Barrio et al., 2005). Se observa claramente como las alturas del modelo estocástico presentan una variabilidad mucho más semejante a la real que las estimadas con el modelo determinístico.

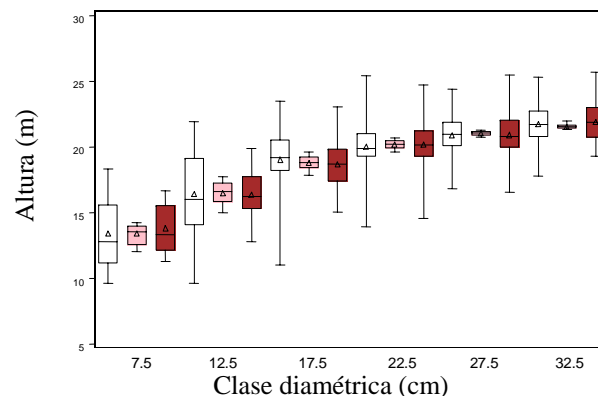


Figura 4. Box-plots de las distribuciones de alturas por clases diamétricas. El triángulo representa la altura media, el rectángulo representa el rango intercuartil y los valores máximos y mínimos están representados por los extremos superior e inferior de las líneas verticales. En blanco la distribución observada, en gris claro la distribución estimada por un modelo determinístico y en gris oscuro la estimada con un modelo estocástico (Barrio et al., 2005).

Una solución alternativa al empleo de la relación altura-diámetro estocástica es el uso de funciones de densidad bivariantes (p.ej. Normal bivalente, Beta bivalente, Johnson S_{BB}) para desagregar las variables de estado de masa por clases de diámetro y de altura, simultáneamente. Una descripción más detallada de estas funciones bivariantes se puede encontrar en los trabajos de Tewari y Gadow (1999), Zucchini et al. (2001) o Li et al., (2002).

Autocorrelación y heterocedasticidad

El ajuste de ecuaciones de regresión que describen el comportamiento de unidades muestrales (rodales o árboles) a lo largo del tiempo presenta una serie de problemas estadísticos. A menudo, cuando se trabaja con series temporales correspondientes a la misma unidad muestral, los errores están correlacionados entre sí, existiendo un problema de autocorrelación. Por otro lado, cuando se emplean datos de diferentes unidades muestrales con distintas características podría darse la circunstancia de que la varianza de los errores no fuese constante (heterocedasticidad). En el caso de que se combinen los dos tipos de datos, situación bastante frecuente en la construcción de modelos forestales, ambos problemas pueden aparecer simultáneamente.

Aunque las metodologías de detección y corrección de los problemas asociados al empleo de datos longitudinales como los descritos anteriormente son ampliamente conocidas, la modelización de la estructura del error en los modelos forestales no ha sido frecuente (Gregoire et al. 1995), en parte debido a que los parámetros estimados y los valores predichos siguen siendo insesgados aún existiendo autocorrelación. Sin embargo, la autocorrelación tiene gran importancia cuando el objetivo primordial es identificar las variables independientes que son estadísticamente significativas para el modelo.

Una de las metodologías más empleadas para corregir problemas de autocorrelación en datos longitudinales es la modelización de la estructura del error, y uno de los modelos más habituales es el que representa una estructura autorregresiva continua (*continuous autoregressive error structure CAR(x)*). Esta estructura se puede incluso utilizar con datos de una misma unidad muestral obtenidos a intervalos no constantes en el tiempo o en el espacio y con un número diferente de mediciones en cada unidad muestral (Zimmerman y Núñez-Antón, 2001), características muy comunes en los datos empleados en modelización forestal.

Como ejemplo, la estructura de segundo orden CAR(2) modeliza el error del modelo empleando la siguiente expresión:

$$e_{ij} = d_1 \rho_1^{t_{ij}-t_{ij-1}} e_{ij-1} + d_2 \rho_2^{t_{ij}-t_{ij-2}} e_{ij-2} + \varepsilon_{ij} \quad (14)$$

donde e_{ij} es el j -ésimo residuo de la unidad muestral i , $d_k = 1$ para $j > k$ y es 0 para $j = k$ ($k = 1, 2$), ρ_k es el parámetro autorregresivo de orden k a estimar, $t_{ij} - t_{ij-k}$ es el tiempo o la distancia que separa la medición j -ésima de la medición $j-k$ en cada unidad muestral, $t_{ij} > t_{ij-k}$, y ε_{ij} es el error aleatorio. En el caso de las curvas de calidad de estación la autocorrelación es temporal (t_{ij} es la edad de la medición j en el árbol i) mientras que en las funciones de perfil la autocorrelación es espacial (t_{ij} es la altura de la medición j en el árbol i).

En la Figura 5 se muestra el efecto de incluir la estructura del error en el ajuste de una función de perfil para *Pinus sylvestris* L. en Galicia.

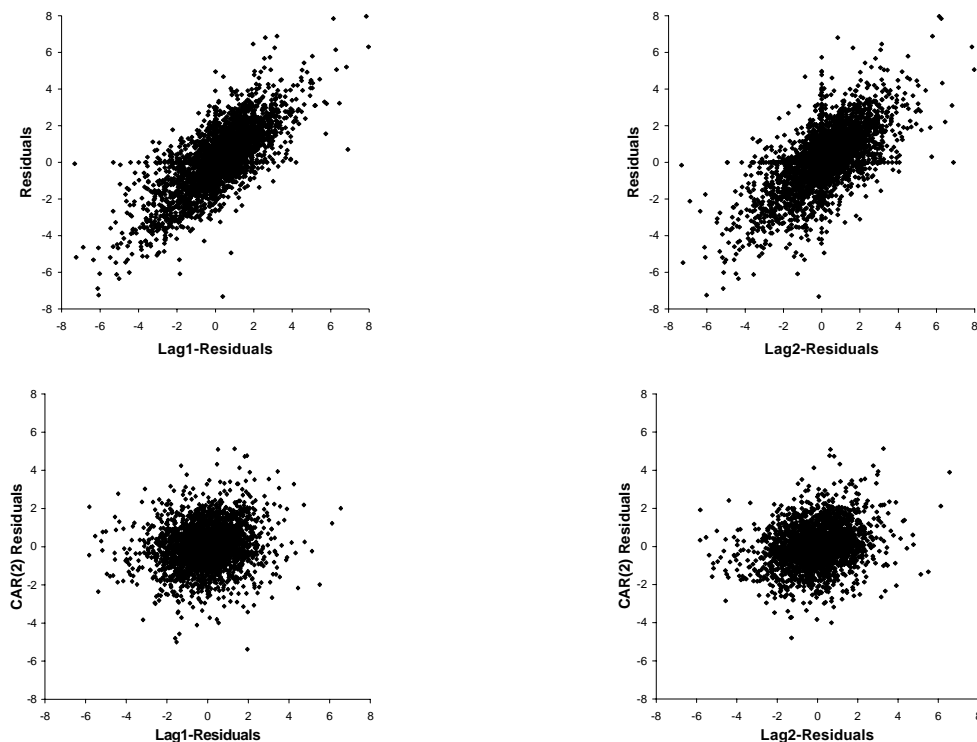


Figura 5. Residuos frente a: residuos de la medición anterior (columna izquierda), residuos de dos mediciones anteriores (columna derecha) para ajustes de una función de perfil de *Pinus sylvestris* L. en Galicia sin tener en cuenta la estructura del error (fila superior) y asumiendo una estructura CAR(2) (fila inferior).

El principal objetivo al tener en cuenta la estructura del error en los modelos es obtener estimaciones de los parámetros insesgadas y eficientes, por lo que, desde el punto de vista práctico, los parámetros de autocorrelación no se emplean en la estimación de un nuevo valor de la variable dependiente, puesto que los valores e_{ij-1} , e_{ij-2} , etc., serán desconocidos a menos que se trabaje con datos de medidas repetidas de la misma unidad muestral.

Otro problema frecuente en la modelización forestal es la falta de homogeneidad de varianza de los errores o heterocedasticidad. En estas situaciones a menudo ocurre que la

varianza del error está relacionada con una o más de las variables independientes del modelo. Una posible solución es asociar a cada observación un peso igual a la inversa de su varianza (σ_i^2). Si la varianza es desconocida, el problema es estimar el valor adecuado del peso. Con bastante frecuencia, al examinar los gráficos de residuos frente a las variables independientes se observa que la varianza se incrementa de forma notable cuando se incrementa al valor de la variable independiente x_i . Este tipo de heterocedasticidad se puede corregir empleando una función potencial como peso $\sigma_i^2 = x_i^k$ (Neter et al., 1996). El valor k del exponente se puede determinar de modo que se obtenga la mayor reducción en el valor de índice de Furnival, o bien se puede emplear la metodología de optimización propuesta por Harvey (1976) y que consiste en emplear los errores del modelo ajustado sin pesos ($\hat{\epsilon}_i$) como variable dependiente en el modelo potencial de varianza del error, es decir: $\hat{\epsilon}_i^2 = x_i^k$ o $\ln \hat{\epsilon}_i^2 = k \ln(x_i)$.

Líneas futuras de trabajo

En los próximos años será necesario realizar un esfuerzo en el desarrollo de modelos más precisos y detallados que den respuesta a las nuevas demandas que los gestores requieren para poder llevar a cabo una planificación forestal sostenible. Las conclusiones a las que se llegó en la el congreso “Reality, models and parameter estimation” celebrado en Sesimbra en el año 2002 (<http://www.isg.pt/proyectos/cigest/files/Papers/sesimbrasummary-1.doc>) pueden servir de punto de partida para todos aquellos interesados en profundizar en el desarrollo futuro de los modelos forestales.

Referencias

- Alder, D., 1980. Estimación del volumen forestal y predicción del rendimiento. Estudios FAO Montes 22/2, Roma.
- Bailey, R.L., Clutter, J.L., 1974. Base-age invariant polymorphic site curves. For. Sci. 20, 155-159.
- Barrio, M., Diéguez-Aranda, U., Castedo, F., Álvarez-González, J.G., Rojo, A., 2005. Mimicking natural height variability using a stochastic height-diameter relationship. IUFRO conference “Sustainable Forest in Theory and Practice”. Edinburgh.
- Borders, B.E., Patterson, W.D., 1990. Projecting stand tables: a comparison of the Weibull diameter distribution method, a percentile-based projection method and a basal area growth projection method. For. Sci. 36, 413-424.
- Cieszewski, C.J., 2003. Developing a well-behaved dynamic site equation using a modified Hossfeld IV function $Y^3 = (ax^m)/(c + x^{m-1})$, a simplified mixed-model and scant subalpine fir data. For. Sci. 49, 539-554.
- Cieszewski, C.J., Bailey, R.L., 2000. Generalized Algebraic Difference Approach: Theory based derivation of dynamic site equations with polymorphism and variable asymptotes. For. Sci. 46, 116-126.
- Cieszewski, C.J., Harrison, M., Martin, S.W., 2000b. Practical methods for estimating non-biased parameters in self-referencing growth and yield models. University of Georgia PMRC-TR 2000-7.
- Diéguez-Aranda, U., Burkhart, H.E., Rodríguez, R., 2005. Modeling dominant height growth of radiata pine (*Pinus radiata* D. Don) plantations in north-western Spain. For. Ecol. Manage. (en prensa)

- Frazier, J.R., 1981. Compatible whole-stand and diameter distribution models for loblolly pine. Unpublished Ph. D. Thesis. VPI & SU.
- Furnival, G.M., 1961. An index for comparing equations used on constructing volume tables. *For. Sci.* 7(4), 337-341.
- García, O., 1988. Growth modelling – a (re)view. *N. Z. For.* 33, 14-17.
- Gregoire, T.G., Schabenberger, O., 1996. A non-linear mixed-effects model to predict cumulative bole volume of standing trees. *Journal of Applied Statistics* 23, 257-271.
- Gregoire, T.G., Schabenberger, O., Barrett, J.P., 1995. Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. *Can. J. For. Res.* 25, 137-156.
- Harvey, A.C., 1976. Estimating regression models with multiplicative heterocedasticity. *Econometrica*, 44, 461-465.
- Krumland, B., Eng, H., 2004. Site index systems for major young-growth forest and woodland species in northern California. California Forestry Report No. 4. Cal. Dep. Forestry and Fire Protection, Sacramento, CA., in press.
- Li, F., Zhang, L., Davis, C.J., 2002. Modeling the joint distribution of tree diameters and heights by bivariate generalized beta distribution. *For. Sci.* 48(1), 47-58.
- Lindstrom, M.J., Bates, D.M., 1990. Nonlinear mixed effects models for repeated measures data. *Biometrics* 46, 673-687.
- Myers, R.H., 1990. Classical and modern regression with applications. Second edition. Duxbury Press, Belmont, California.
- Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W. 1996. Applied linear statistical models. 4th edition. McGraw-Hill, New York.
- Palahí, M., 2002. Modelling the stand development and optimizing the management of even-aged Scots pine forests in north-east Spain. Research Notes. Faculty of Forestry, University of Joensuu, 58.
- Parresol B.R., Lloyd F.T., 2004. The stochastic tree modelling approach used to derive tree list for the GIS/CISC identified stands at the Savannah River Site. Internal Report. USDA For Serv SRS, Asheville, NC.
- Sánchez, F., Rodríguez, R., Rojo, A., Álvarez, J.G., López, C., Gorgoso, J., Castedo, F., 2003. Crecimiento y tablas de producción de *Pinus radiata* D. Don en Galicia. *Invest. Agrar.: Sist. Recur. For.* 12, 65-83.
- Snowdon, P., Jovanovic, T., Booth, T.H. 1998. Incorporation of indices of annual climatic variation into growth models for *Pinus radiata*. *For. Ecol. Manage.*, 117, 187-197.
- Tait, D.E., Cieszewski, C.J., Bella, I.E., 1988. The stand dynamics of lodgepole pine. *Can. J. For. Res.* 18, 1255-1260.
- Tewari, V.P., Gadov, K.v., 1999. Modelling the relationship between tree diameters and heights using SBB distribution. *For. Ecol. Manage.* 119, 171-176.
- Torres-Rojo, J.M., Magaña-Torres, O.S., Acosta-Mireles, M., 2000. Metodología para mejorar la predicción de parámetros de distribuciones diamétricas. *Agrociencia* 34, 627-637.

- Valsta, L.T., 1993. Stand management optimization based on growth simulators. The Finnish Forest Research Institute, Research Papers 453, 51.
- Vanclay, J.K., 1994. Modelling forest growth and yield. Applications to mixed tropical forests. CAB International, Wallingford, UK.
- Wollons, R.C., Snowdon, P., Mitchell, N.D., 1997. Argumenting empirical stand projection equations with edaphic and climatic variables. *For. Ecol. Manage.* 98, 267-275.
- Zimmerman, D.L., Núñez-Antón, V. 2001. Parametric modelling of growth curve data: An overview (with discussion). *Test* 10, 1-73.
- Zucchini, W., Schmidt, M., Gadow, K.v., 2001. A model for the diameter-height distribution in an uneven-aged beech forest and a method to assess the fit of such models. *Silva Fennica* 35(2), 169–183.