

# RESOLUCIÓN DE PROBLEMAS DEL AJUSTE SIMULTÁNEO DE SISTEMAS DE ECUACIONES: HETEROCEDASTICIDAD Y VARIABLES DEPENDIENTES CON DISTINTO NÚMERO DE OBSERVACIONES

Juan Gabriel Álvarez-González, Roque Rodríguez-Soalleiro y Alberto Rojo-Alboreca

Unidad de Gestión Forestal Sostenible. Departamento de Ingeniería Agroforestal. Escuela Politécnica Superior de la Universidad de Santiago de Compostela. Campus Universitario s/n. 27002-LUGO (España). Correo electrónico: algonjg@lugo.usc.es

## Resumen

En la modelización forestal es frecuente utilizar sistemas de ecuaciones para simular algún proceso del desarrollo de los bosques. Por ejemplo, tanto el conjunto de las ecuaciones de estimación de biomasa por fracciones arbóreas como la combinación de una tarifa de cubicación y una función de perfil son dos sistemas de ecuaciones muy utilizados. En el ajuste de estos sistemas suele ser habitual que se imponga alguna condición. Por ejemplo, cuando se ajustan ecuaciones de estimación de biomasa por fracciones arbóreas debe cumplirse la propiedad llamada “aditividad”, que consiste en que la suma de las estimaciones de biomasa de todos los componentes sea igual a la estimación de biomasa del árbol completo. En el caso del sistema constituido por la tarifa de cubicación y la función de perfil debe existir “compatibilidad”, es decir, el volumen obtenido por la tarifa debe ser igual al obtenido por integración de la función de perfil. Una solución para garantizar el cumplimiento de estas propiedades es ajustar el sistema de ecuaciones simultáneamente. Sin embargo, el procedimiento utilizado debe permitir asignar pesos diferentes a las observaciones de cada ecuación del sistema para corregir una posible falta de homogeneidad de varianzas. También puede ocurrir que las variables dependientes de las ecuaciones no tengan el mismo número de observaciones. Por ejemplo, en el caso del sistema compuesto por la tarifa de cubicación y la función de perfil, una variable dependiente es el volumen del árbol, con una observación por individuo, y la otra variable dependiente es el diámetro del árbol a una cierta altura en el tronco, con varias observaciones por individuo. En este trabajo se presentan con detalle los problemas que se pueden plantear en este tipo de ajustes y las soluciones aplicadas empleando el programa SAS/ETS.

Palabras clave: *Biomasa arbórea, Función de perfil de tronco, Datos no balanceados, Regresión SUR, Aditividad, Compatibilidad*

## DESCRIPCIÓN DE LOS PROBLEMAS

En la construcción de modelos de crecimiento y predicción de la evolución de un árbol o de un bosque es relativamente frecuente tener que realizar

ajustes simultáneos de sistemas de ecuaciones. Dos de estas situaciones muy habituales son el ajuste de sistemas de ecuaciones de estimación de biomasa por fracciones del árbol y los sistemas compatibles de estimación de volúmenes de árboles.

A continuación se describen estos dos sistemas de ecuaciones y los problemas que derivan de su ajuste:

**Ajuste de ecuaciones de biomasa por fracciones arbóreas**

Una de las propiedades más importantes que deben cumplir las ecuaciones que estiman la biomasa de las diferentes fracciones de un árbol es la llamada “aditividad” (KOZAK, 1970; CHIYENDA & KOZAK, 1984; CUNIA, 1986; PARRÉSOL, 1999). Esta característica consiste en que la suma de las estimaciones de los pesos de todos los componentes o fracciones debe ser igual a la estimación del peso del árbol completo. Existen tres procedimientos para forzar esta propiedad en un sistema de ecuaciones lineales, dependiendo de cómo se agreguen los diferentes componentes.

El procedimiento más sencillo (Procedimiento 1) consiste en ajustar de forma individual los modelos de cada fracción ( $\hat{w}_i$ ) y obtener la estimación del peso total como suma de los modelos de todas las fracciones (SRIVASTAVA & GILES, 1987):

$$\begin{aligned} \hat{w}_1 &= f_1(x_1) \\ \hat{w}_2 &= f_2(x_2) \\ &\vdots \\ \hat{w}_k &= f_k(x_k) \\ \hat{w}_{total} &= \sum_{i=1}^k \hat{w}_i = \hat{w}_1 + \hat{w}_2 + \dots + \hat{w}_k \end{aligned} \quad [1]$$

donde  $x_1, x_2, \dots, x_k$  son las variables independientes usadas para cada fracción. Si se emplea este procedimiento, cada modelo puede incluir diferentes variables explicativas. Además, se pueden plantear modelos lineales y no lineales en el sistema. Otra ventaja es que no es necesario tener datos de todas las fracciones de todos los árboles. Sin embargo, no considera la dependencia entre los errores de los componentes del árbol y es muy difícil estimar la suma de los errores. De esta forma, la varianza de la estimación de la biomasa total,  $s^2(\hat{w}_{total})$ , es muy elevada, puesto que su valor depende de las varianzas de las estimaciones de cada fracción,  $s^2(\hat{w}_i)$ , así como de las covarianzas entre estimaciones, con  $(\hat{w}_i, \hat{w}_j)$ , tal y como se refleja en la siguiente expresión:

$$s^2(\hat{w}_{total}) = \sum_{i=1}^k s^2(\hat{w}_i) + 2 \cdot \sum \sum Cov(\hat{w}_i, \hat{w}_j) \quad [2]$$

En consecuencia, el procedimiento 1 podría dar lugar a una acumulación de errores en la predicción de la biomasa total, lo que normalmente no será deseable, por ser la estimación de esta variable con frecuencia de mayor importancia que el propio desglose en componentes de la misma.

Una alternativa sería ajustar la ecuación de biomasa total y las ecuaciones de todas las fracciones excepto aquella para la que se requiera una menor exactitud (Procedimiento 2). Las estimaciones de esta fracción se obtienen restando a las estimaciones de biomasa total las de las restantes fracciones. Sin embargo, en estos casos es frecuente que se obtengan valores negativos al realizar estas restas.

El procedimiento 2 garantiza la aditividad empleando las mismas variables explicativas en las ecuaciones de regresión de los diferentes componentes del árbol y en la de biomasa total, tal y como refleja el sistema [3]. De este modo, los valores de los parámetros del modelo de estimación del peso total resultan de la suma de los parámetros obtenidos para cada variable explicativa al ajustar el modelo a cada fracción por separado. En el caso de emplear mínimos cuadrados ponderados se han de usar las mismas ponderaciones en las regresiones de todos los componentes. Pese a la obligación de emplear el mismo modelo en los  $k$  componentes y en la biomasa total, algunos autores han conseguido salvar este condicionante empleando mínimos cuadrados restringidos (CHIYENDA & KOZAK, 1984).

$$\begin{aligned} \hat{w}_1 &= \beta_{01} + \beta_{11} \cdot x_1 + \beta_{21} \cdot x_2 + \dots + \beta_{j1} \cdot x_j \\ \hat{w}_2 &= \beta_{02} + \beta_{12} \cdot x_1 + \beta_{22} \cdot x_2 + \dots + \beta_{j2} \cdot x_j \\ &\vdots \\ \hat{w}_k &= \beta_{0k} + \beta_{1k} \cdot x_1 + \beta_{2k} \cdot x_2 + \dots + \beta_{jk} \cdot x_j \end{aligned} \quad [3]$$

$$\hat{w}_{total} = (\beta_{01} + \beta_{02} + \dots + \beta_{0k}) + (\beta_{11} + \beta_{12} + \dots + \beta_{1k}) \cdot x_1 + (\beta_{21} + \beta_{22} + \dots + \beta_{2k}) \cdot x_2 + \dots + (\beta_{j1} + \beta_{j2} + \dots + \beta_{jk}) \cdot x_j$$

El principal inconveniente de este procedimiento 2 radica en asumir que los  $k$  componentes de un mismo árbol son independientes, esto es, que no hay dependencia entre los errores, lo que no es cierto en la mayoría de los casos. Además, las ecuaciones ajustadas para cada fracción no tienen porqué ser las mejores desde el punto de vista estadístico, e incluso algunos coeficientes de

la regresión se incluyen en la expresión final del modelo aunque no sean significativamente distintos de cero. De forma contraria a lo comentado para el procedimiento 1, en este caso la aditividad no está garantizada si no se dispone de datos para todas las fracciones de todos los árboles, por lo que sería necesario filtrar la información y considerar en el ajuste únicamente los pies de los que se dispusiese de valores de peso de todos los componentes arbóreos.

Asumiendo la hipótesis de independencia entre componentes, la varianza de la estimación de la biomasa total,  $s^2(\hat{w}_{total})$ , es la suma de las varianzas de las  $k$  fracciones arbóreas, .

$$s^2(\hat{w}_{total}) = \sum_{i=1}^k s^2(\hat{w}_i) \quad [4]$$

Por otro lado, la necesidad de emplear las mismas variables regresoras en todas las ecuaciones de biomasa implica en muchos casos la aparición de multicolinealidad.

Forzar el ajuste del mismo modelo para cada una de las componentes podría dar lugar a una falta de significación de alguno de ellos, o de la estimación de alguno de los parámetros. Por ejemplo, es frecuente que el diámetro de copa sea una buena variable independiente en las ecuaciones de estimación de las fracciones de copa, sin embargo, no aporta mejoras en las ecuaciones de madera o corteza. Además, el número de parámetros a estimar se incrementa mucho, lo que va en contra de la necesaria parsimonia del sistema.

El procedimiento 3 es el más flexible, pero es también la de más difícil utilización. Se basa en el ajuste de un sistema de ecuaciones aparentemente no relacionadas y formado por las funciones de regresión de los  $k$  componentes arbóreos junto con el de biomasa total (Sistema [5]).

$$\begin{aligned} \hat{w}_1 &= f_1(x_1) \\ \hat{w}_2 &= f_2(x_2) \\ &\vdots \\ \hat{w}_k &= f_k(x_k) \\ \hat{w}_{total} &= f_{total}(x_1, x_2, \dots, x_k) \end{aligned} \quad [5]$$

En este sistema, no es necesario que las ecuaciones de cada fracción arbórea presenten la misma expresión matemática ni las mismas variables independientes. Las variables explica-

tivas del modelo de biomasa total son todas las variables regresoras que aparecen en los modelos de cada componente.

Por otro lado, en este caso no es necesario disponer de datos de peso de todas las fracciones consideradas en todos los árboles muestreados. Se pueden emplear además diferentes ponderaciones en las ecuaciones de cada fracción resueltas mediante regresión por peso. Este sistema, sin relaciones analíticas entre ecuaciones, se suele resolver empleando la regresión SUR (*Seemingly Unrelated Regression*), conocida también como mínimos cuadrados generalizados conjuntos. Se trata de una generalización del método de regresión por mínimos cuadrados ordinarios (*Ordinary Least Squares, OLS*) para un sistema de ecuaciones. Como OLS, el método SUR asume que todas las variables regresoras son variables independientes, pero SUR usa la correlación entre los errores de diferentes ecuaciones (es decir,  $cov(e_i, e_j) \neq 0$  para los pares de  $i$  y de  $j$ ) para mejorar la eficiencia de las estimaciones. El sistema de ecuaciones resultante del ajuste es siempre el mejor posible, aunque el ajuste de las ecuaciones de cada fracción de forma individual no sea necesariamente el mejor.

El empleo de regresión SUR en el ajuste simultáneo de las ecuaciones de biomasa de las distintas fracciones arbóreas planteadas implica considerar este conjunto de funciones como un sistema de ecuaciones aparentemente no relacionadas, que responde a la siguiente forma:

$$\begin{aligned} y_1 &= \beta_{10} + \beta_{11} \cdot x_1 + \beta_{12} \cdot x_2 + e_1 \\ y_2 &= \beta_{20} + \beta_{21} \cdot x_3 + \beta_{22} \cdot x_4 + e_2 \\ y_3 &= \beta_{30} + \beta_{31} \cdot x_5 + \beta_{32} \cdot x_6 + e_3 \end{aligned} \quad [6]$$

Las variables que se encuentran en la parte izquierda de la ecuación se denominan variables endógenas o dependientes, ya que están determinadas dentro del sistema de ecuaciones. Las variables que aparecen en la parte derecha de la ecuación se denominan predeterminadas, y pueden ser también variables endógenas cuyos valores se han determinado en el propio modelo (variables instrumentales), o variables cuyos valores se han medido directamente y por tanto no han sido estimadas por el sistema (variables exógenas).

El ajuste de ecuaciones de biomasa presenta problemas de heterocedasticidad. Una de las metodologías utilizadas para la corrección de

este problema es el empleo de pesos o regresión ponderada (SCHLAEGEL, 1982; CLUTTER et al., 1983; CUNIA, 1986; PARRÉSOL, 1999). En el análisis por pesos o regresión ponderada se asocia a cada observación un peso igual a la inversa de su varianza ( $\sigma_i^2$ ). Habitualmente, esta varianza está relacionada con una o más de las variables independientes, por lo que la aplicación de esta metodología implica conocer dicha relación, lo que resulta bastante complicado. 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 aumenta el valor de la variable independiente ( $x$ ). 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 empleando la metodología de optimización propuesta por HARVEY (1976), que consiste en usar los errores del modelo ajustado sin pesos ( $\hat{\epsilon}_i$ ) como variable dependiente en el modelo potencial de varianza del error, es decir:

$$\hat{\epsilon}_i^2 = x_i^k \text{ o bien } \ln \hat{\epsilon}_i^2 = a + k \ln(x_i) \quad [7]$$

### Ajuste de sistemas compatibles de estimación de volúmenes de árboles

Idealmente, un sistema de estimación de volumen debería ser “compatible”, es decir, el volumen estimado al integrar la función de perfil desde la base hasta el ápice del árbol debería coincidir con el estimado por una tarifa de cubicación para la misma especie y área geográfica (DEMAERSCHALK, 1972; CLUTTER, 1980).

La forma más comúnmente empleada para garantizar esta compatibilidad ha sido expresar el coeficiente  $\beta$  del modelo de variable combinada de Spurr (SPURR, 1952) sin término independiente ( $V = \beta \cdot d^2 \cdot h$ , donde  $V$  es el volumen total,  $d$  el diámetro normal y  $h$  la altura total del árbol) en función de los parámetros estimados en una función de perfil, o viceversa, usando una relación de compatibilidad entre ambas ecuaciones. Un ejemplo de este tipo de compatibilidad es el sistema de estimación de volumen derivado de la función de perfil propuesta por KOZAK et al. (1969):

$$d_i^2 = \left[ b_1 \cdot \left( \frac{h_i^2}{h^2} - 1 \right) - b_2 \cdot \left( \frac{h_i}{h} - 1 \right) \right] \cdot d^2 \quad [8]$$

donde  $d_i$  es el diámetro del tronco del árbol a la altura  $h_i$ ;  $d$  es el diámetro normal;  $h$  la altura total y  $b_i$  ( $i = 1, 2$ ) son los parámetros del modelo a estimar. La integración de la ecuación [8] entre la base y el ápice del árbol da lugar a una expresión similar al modelo de variable combinada de Spurr:

$$v = \int_0^h \frac{\pi}{4} \cdot d_i^2 \cdot dh_i = \int_0^h \frac{\pi}{4} \cdot \left[ b_1 \cdot \left( \frac{h_i^2}{h^2} - 1 \right) - b_2 \cdot \left( \frac{h_i}{h} - 1 \right) \right] \cdot d^2 \cdot dh_i = \frac{\pi}{4} \cdot d^2 \cdot h \cdot \left( \frac{b_2}{2} - \frac{2 \cdot b_1}{3} \right) [9]$$

$$= \beta \cdot d^2 \cdot h \quad \text{con} \quad \beta = \frac{\pi}{4} \cdot \left( \frac{b_2}{2} - \frac{2 \cdot b_1}{3} \right)$$

Por tanto, en este caso la compatibilidad entre la función de perfil y la tarifa de cubicación es que el parámetro  $\beta$  de esta última viene dado por la siguiente relación con los parámetros  $b_i$  de la función de perfil:

$$\beta = \frac{\pi}{4} \cdot \left( \frac{b_2}{2} - \frac{2 \cdot b_1}{3} \right) \quad [10]$$

La compatibilidad entre la función de perfil y la tarifa de cubicación no depende del procedimiento de estimación de los parámetros, por lo que se pueden emplear tres diferentes metodologías: (i) estimar los parámetros de la función de perfil usando los datos de árboles tipo con mediciones de diámetro a diferentes alturas y, posteriormente, calcular el valor del parámetro  $\beta$  usando la relación de compatibilidad dada en la ecuación [10]; (ii) estimar el parámetro  $\beta$  de la tarifa de cubicación utilizando los datos de diámetro normal, altura y volumen total de árboles tipo y, posteriormente, estimar los parámetros de la función de perfil usando los datos de diámetros a diferentes alturas de árboles tipo pero imponiendo en el ajuste la condición de compatibilidad dada por la ecuación [10]; (iii) ajustar ambas ecuaciones (la función de perfil y la tarifa de cubicación) simultáneamente, empleando en la tarifa de cubicación la expresión completa en la que el valor del parámetro  $\beta$  se sustituye por la expresión de compatibilidad (ecuación [10]).

Las dos primeras opciones permiten obtener la mejor estimación de diámetros a una altura dada o del volumen total, dependiendo de si se ha priorizado en el ajuste a la función de perfil o a la tarifa

de cubicación, respectivamente; es decir, si se ha ajustado primero la función de perfil y ese ajuste define la expresión matemática de la tarifa de cubicación o, por el contrario, ha sido a la inversa. La tercera opción permite obtener un mínimo en la suma de cuadrados del error del sistema en su conjunto, es decir, simultáneamente minimiza los errores del diámetro a diferentes alturas y los errores en la estimación de volúmenes.

El principal problema metodológico que presenta el ajuste simultáneo de este tipo de sistemas compatibles de volumen es que el número de observaciones a utilizar en cada una de las dos ecuaciones no es el mismo. Para cada árbol sólo hay una observación de volumen total para ajustar la tarifa de cubicación; sin embargo, hay más de una observación por árbol del diámetro a diferentes alturas (en el caso más simple se cuenta con tres observaciones por árbol: diámetro a la altura del tocón, diámetro a la altura normal y diámetro a la altura total). Por ejemplo, en la tabla 1 aparecen las observaciones de dos árboles tipo con 5 y 6 mediciones de diámetro a diferentes alturas del tronco y un único valor de volumen total.

Para resolver este problema se propone crear una estructura especial para la base de datos de partida: la observación de volumen total de cada árbol se asigna repetidamente a cada una de las observaciones de diámetro a diferentes alturas del mismo árbol. Para evitar el problema que supone la adición de nuevos datos en un número diferente a cada árbol (puesto que no todos tie-

nen el mismo número de observaciones de diámetros a lo largo del tronco), a cada observación de volumen se le asigna un peso en el ajuste igual a la inversa del número de observaciones de cada árbol ( $n_i$ ). Continuando con el ejemplo anterior, la estructura de datos empleada para el ajuste sería la mostrada en la tabla 2.

La estimación de los parámetros de los dos sistemas de ecuaciones se puede realizar con el procedimiento MODEL del programa SAS/ETS® (SAS INSTITUTE INC., 2004), utilizando la metodología SUR (*Seemingly Unrelated Regression*) (ZELLNER, 1962). El peso asociado a cada observación "peso<sub>i</sub>" se debe incluir en el ajuste especificando en el programa la opción  $resid.V = resid.V / peso_i$ . Con esta opción los residuos ( $resid.V$ ) son multiplicados por la raíz cuadrada del peso porque esta función afecta al residuo antes de que sea elevado al cuadrado para determinar la suma de cuadrados (SAS INSTITUTE INC., 2004).

## PROGRAMAS EMPLEADOS EN CADA AJUSTE

### Ajuste de ecuaciones de biomasa por fracciones arbóreas

En la figura 1 se muestra el esquema del programa que se puede utilizar para ajustar un sistema de ecuaciones de estimación de biomasa por fracciones con el programa SAS/ETS® (SAS INSTITUTE INC., 2004).

árbol	$h_i$ (m)	$d_i$ (m)	$d$ (m)	$h$ (m)	volumen (m <sup>3</sup> )
1	0,00	0,15	0,11	8,01	0,0359
1	1,30	0,11			
1	4,00	0,06			
1	6,51	0,04			
1	8,01	0,00			
2	0,00	0,17	0,13	12,35	0,0605
2	1,30	0,13			
2	4,00	0,08			
2	7,16	0,04			
2	10,01	0,02			
2	12,35	0,00			

**Tabla 1.** Ejemplo de estructura de datos de árboles tipo para el ajuste de sistemas compatibles de volumen, siendo  $d_i$  el diámetro a diferentes alturas,  $h_i$ ,  $d$  el diámetro normal y  $h$  la altura total

árbol	$h_i$ (m)	$d_i$ (m)	$d$ (m)	$h$ (m)	volumen (m <sup>3</sup> )	Peso
1	0,00	0,15	0,11	8,01	0,0359	1/5
1	1,30	0,11	0,11	8,01	0,0359	1/5
1	4,00	0,06	0,11	8,01	0,0359	1/5
1	6,51	0,04	0,11	8,01	0,0359	1/5
1	8,01	0,00	0,11	8,01	0,0359	1/5
2	0,00	0,17	0,13	12,35	0,0605	1/6
2	1,30	0,13	0,13	12,35	0,0605	1/6
2	4,00	0,08	0,13	12,35	0,0605	1/6
2	7,16	0,04	0,13	12,35	0,0605	1/6
2	10,01	0,02	0,13	12,35	0,0605	1/6
2	12,35	0,00	0,13	12,35	0,0605	1/6

**Tabla 2.** Ejemplo de estructura de datos de árboles tipo modificada para el ajuste simultáneo de sistemas compatibles de volumen

### Ajuste de sistemas compatibles de estimación de volúmenes de árboles

En la figura 2 se muestra el esquema del programa que se puede utilizar para ajustar un sistema de ecuaciones compatible formado por una función de perfil y una tarifa de cubicación empleando el programa SAS/ETS® (SAS INSTITUTE INC., 2004).

### BIBLIOGRAFÍA

- CHIYENDA, S.S. & KOZAK, A.; 1984. Additivity of component biomass regression equations when the underlying model is linear. *Can. J. For. Res.* 14: 441-446.
- CLUTTER, J.L.; 1980. Development of taper functions from variable-top merchantable volume equations. *For. Sci.* 26: 117-120.
- CLUTTER, J.L.; FORSTON, J.C.; PIENAAR, L.V.; BRISTER, G.H. & BAILEY, R.L.; 1983. *Timber management: a quantitative approach*. Wiley. New York.
- CUNIA, T.; 1986. Construction of tree biomass tables by linear regression techniques. In: *Estimating tree biomass regressions and their error. Proceedings of the Workshop on tree biomass regression functions and their contribution to the error of forest inventory estimates*: 27-37. Syracuse. New York.
- DEMAERSCHALK, J.; 1972. Converting volume equations to compatible taper equations. *For. Sci.* 18: 241-245.
- HARVEY, A.C.; 1976. Estimating regression models with multiplicative heteroscedasticity. *Econometrica* 44: 461-465.
- KOZAK, A.; 1970. Methods of ensuring additivity of biomass components by regression analysis. *For. Chron.* 46: 402-404.
- KOZAK, A.; MUNRO, D. & SMITH, J.; 1969. Taper functions and their application in forest inventory. *For. Chron.* 45: 278-283.
- NETER, J.; KUTNER, M.H.; NACHTSHEIM, C.J. & WASSERMAN, W.; 1996. *Applied linear statistical models*. 4th edition. McGraw-Hill. New York.
- PARRESOL, B.R.; 1999. Assessing tree and stand biomass: a review with examples and critical comparisons. *For. Sci.* 45: 573-593.
- SAS INSTITUTE INC.; 2004. *SAS/ETS® 9.1 User's Guide*. Cary, NC: SAS Institute Inc.
- SCHLAEGEL, B.E.; 1982. Acer negundo biomass component regression analysis for the Mississippi Delta. *For. Sci.* 28: 355-358.
- SPURR, S.H.; 1952. *Forest inventory*. The Ronald Press Co. New York.
- SRIVASTAVA, V.K. & GILES, D.E.A.; 1987. *Seemingly Unrelated Regression Equations Models: Estimation and Inference*. Marcel Dekker. New York.
- ZELLNER, A.; 1962. An efficient method of estimating seemingly unrelated regressions and test for aggregation bias. *J. Am. Stat. Assoc.* 57: 348-368.

#En primer lugar se ajusta cada ecuación por separado para determinar la expresión más adecuada. En este caso se prueba un modelo alométrico para ajustar la fracción de madera a los datos que están en un fichero SAS cuya localización viene indicada por el nombre de la biblioteca en la que está ubicado el fichero y por su nombre asignado en SAS (*nombre\_de\_biblioteca.nombre\_de\_archivo*). Los parámetros del modelo son b11, b12 y b13 a los que se debe asignar un valor numérico de partida en el comando *fit* de la sintaxis del procedimiento *model*. Se incluye la opción *out=residuos* en el comando *fit* para crear un fichero que contenga los errores del modelo. La variable que contiene estos errores se denomina, por defecto, con el mismo nombre de la variable dependiente del modelo, es decir, en este caso se denomina *madera#*

```
proc model data = nombre_de_biblioteca.nombre_de_archivo;
  parms b11 b12 b13; madera = b11*d**b12*h**b13;
  fit madera start = (b11 = valor_de_partida
  b12 = valor_de_partida b13 = valor_de_partida)/out = residuos;
run;
```

#A continuación se determina el valor k del exponente del peso empleando la metodología de optimización propuesta por HARVEY (1976) y que consiste en usar los errores del modelo ajustado sin pesos ( $\hat{\epsilon}_i$ ) como variable dependiente en el modelo potencial de varianza del error de la ecuación [7]. El fichero que contiene los residuos#:

```
data peso;
  set residuos;
  lres = log(madera);
  ld2h = log(d**2*h);
run;
proc reg data = peso;
  model lres = ld2h;
run;
```

#Por último, una vez aplicada la misma metodología a cada fracción de biomasa, se ajustan todas simultáneamente empleando el comando *resid.fracción* para indicar los pesos pero con la expresión incluida dentro de una raíz cuadrada (*sqrt*), tal y como se indicó en el apartado de material y métodos. El término *vp* que aparece en el comando *fit* para cada parámetro indica que se debe incluir un valor numérico de partida del procedimiento iterativo de estimación de parámetros#

```
proc model data = nombre_de_biblioteca.nombre_de_archivo;
  var madera corteza ramas acicula total d h;
  parms b11 b12 b13 b21 b22 b23 b31 b41;

  madera = b11*d**b12*h**b13;
  resid.madera = resid.madera / sqrt(d2h**parámetro_ld2h);
  corteza = b21*d**b22*h**b23;
  resid.corteza = resid.corteza / sqrt(obtener_peso);
  ramas = b31*d**2;
  resid.ramas = resid.ramas / sqrt(obtener_peso);
  acicula = b41*d**2;
  resid.acicula = resid.acicula / sqrt(obtener_peso);
  total = b11*d**b12*h**b13+b21*d**b22*h**b23+b31*d**2+b41*d*2;
  resid.total = resid.total / sqrt(obtener_peso);

  fit madera corteza ramas acicula total
  start = (b11 vp b12 vp b13 vp b21 vp b22 vp b23 vp b31 vp b41 vp) / sur out=salida;
run; quit;
```

**Figura 1.** Esquema del programa de SAS/ETS® para el ajuste de un sistema de ecuaciones de estimación de biomasa por fracciones arbóreas. “Madera”, “corteza”, “ramas”, “acicula”, “total”, “d” y “h” son variables del fichero de datos que contienen el peso seco de la biomasa de madera, corteza, ramas, acículas y total, así como el diámetro normal del árbol y su altura total, respectivamente

#En primer lugar se ajusta la tarifa de cubicación (variable *vm3arbol*) para determinar el peso a incluir para evitar problemas de falta de homogeneidad de varianzas. Se aplica también la metodología propuesta por HARVEY (1976) para estimar el valor del exponente *k* de la ecuación [7]. De nuevo el fichero SAS con los datos se indica por el nombre de la biblioteca en la que está ubicado y por su nombre de archivo asignado en SAS (*nombre\_de\_biblioteca.nombre\_de\_archivo*). El parámetro del modelo es *alpha* y se le debe asignar un valor numérico de partida en el comando **fit** de la sintaxis del procedimiento **model**. Se incluye la opción *out=residuos* en el comando **fit** para crear un fichero que contenga los errores del modelo. La variable que contiene estos errores se denomina, por defecto, con el mismo nombre de la variable dependiente del modelo, es decir, en este caso se denomina *vm3arbol#*

```
proc model data = nombre_de_biblioteca.nombre_de_archivo;
  parms alpha;
  vm3arbol=alpha*d**2*h;
  fit vm3arbol start=(alpha = valor_de_partida) /
  out=residuos;
run; quit;
```

```
data peso;
  set residuos;
  lres=log(vm3arb);
  ld2h=log(d**2*h);
run;
```

```
proc reg data=peso;
  model lres=ld2h;
run;
```

#Finalmente, se ajustan simultáneamente la tarifa de cubicación (variable *vm3arbol*) y la función de perfil (variable *di*) empleando el comando **resid.vm3arbol** para indicar la ponderación de los datos de la tarifa. La ponderación tiene dos términos: *Ni*, que es una variable del fichero de datos que indica el número de veces que se repite cada observación de volumen (según la modificación de la estructura de datos que se comentó en el apartado de Descripción de los problemas) y el término que corrige la heterocedasticidad con el valor del exponente antes calculado según la metodología de HARVEY (1970). Ambos términos deben estar multiplicados e incluidos dentro de una raíz cuadrada (*sqrt*), tal y como se indicó en el apartado de material y métodos. El término *vp* que aparece en el comando **fit** para cada parámetro indica que se debe incluir un valor numérico de partida del procedimiento iterativo de estimación de parámetros#

```
proc model data = nombre_de_biblioteca.nombre_de_archivo;
  parms b1 b2;
  k=(3.1415926/4);
  z=(h-hi)/h;
  q=hi/h;
  vm3arbol=d**2*h*(k*((b2/2)-(2*b1/3)));
  resid.vm3arbol=resid.vm3arbol/sqrt(Ni*(d**2*h)**parametro_ld2h);
  di=sqrt(b1*(q**2-1)-b2*(q-1))*d;
  fit vm3arbol di start=(b2=vp b1=vp)/SUR
  out=salida;
```

**Figura 2.** Esquema del programa de SAS/ETS® para el ajuste de un sistema compatible formado por una función de perfil y una tarifa de cubicación. Las variables del fichero de datos son: “*Vm3arbol*”, que contiene el volumen de cada árbol tipo; “*di*” que es el diámetro del tronco a la altura “*hi*”; “*d*” que es el diámetro normal del árbol y “*h*” que es su altura total