

# Análisis espacial de los datos del Inventario Forestal Nacional utilizando técnicas geoestadísticas

N. NANOS; L. GIL  
y G. MONTERO<sup>1</sup>

## INTRODUCCIÓN

El estado selvícola de los ecosistemas forestales es resultado de la acción e interacción de una infinidad de factores ambientales, genéticos y otros de origen humano, cuyo elemento fundamental es su continuidad espacial, entendiéndose por esta expresión su homogeneidad en el espacio en el que actúan. Las formaciones forestales responden a esta acción y forman los nichos ecológicos, los hábitat, etc., es decir, pequeñas o grandes extensiones territoriales que tienen unas características comunes.

La continuidad (o correlación) espacial se refiere a la autosemejanza que puede presentar una serie de variables en el espacio. El ejemplo más representativo de este tipo de variables es la precipitación, que está reconocida como un factor que controla la distribución de las especies, su frecuencia de aparición y su productividad. Dicho factor es muy homogéneo en el espacio y se distribuye de tal manera que la precipitación que recibe un sitio localizado en el punto  $z(a)$  es muy parecida a la que recibe el punto  $z(a+h)$  localizado  $h$  metros más lejos. Adicionalmente, cuanto mayor sea la distancia  $h$ , menos se parecerá la precipitación entre los puntos  $z(a)$  y  $z(a+h)$  (Figura 1).

Capítulo

8

La presencia de autosemejanza espacial en las variables de la naturaleza tiene unas implicaciones directas en el tratamiento estadístico de las mismas, dado que los datos que se tomarán con respecto a una variable de este tipo estarán también correlacionados en el espacio en el que se han medido. La correlación espacial puede ser un problema o una solución a la hora del análisis (LEGENDRE, 1993). El análisis estadístico-clásico de datos auto-correlacionados presenta varios obstáculos, dado que una de las hipótesis fundamentales de la inferencia estadística es la independencia de las observaciones. Si se opta por métodos de análisis clásicos, la correlación espacial provoca la subestimación de los intervalos de confianza de los usuales estimadores estadísticos (como por ejemplo la media de la población) y por eso las estimaciones parecen ser más fiables de lo que verdaderamente son (CRESSIE, 1993). Durante muchos años hemos subestimado el efecto que dicha correlación puede tener sobre los resultados de nuestros análisis y nos hemos "obligado" a suponer que las observaciones son independientes entre sí, algo que en la mayoría de los casos no es cierto.

Durante los últimos años hemos visto avances estadísticos que permiten un análisis más realista de datos correlacionados espacialmente, la mayoría de los cuales se basan en fundamentos descritos bajo el título genérico de geoestadística. La geoestadística, hoy en día, forma parte de una rama de la ciencia estadística que se denomina estadística espacial<sup>1</sup>. Sus fundamentos fueron descritos por vez primera por la escuela Francesa de G. Matheron (MATHERON, 1963) y se utilizaron ampliamente en la geología (DAVIS, 1986), la ingeniería minera (JURNEL Y HUIJBREGTS, 1978) y la caracterización de reservas petrolíferas. Posteriormente, la comunidad científica

reconoció el papel que la geoestadística puede desempeñar y actualmente dicha metodología se utiliza en una gama de ciencias tan diversas como la meteorología (MARTINEZ-COB Y CUENCA, 1992), la energía nuclear (Gotway, 1994) o la ciencia forestal (HOULLIER Y PIERRAT, 1992; MATÉRN, 1994; KOHL Y GERTNER, 1997).

El principal rasgo de la geoestadística es el reconocimiento "a priori" de que los datos muestreados están correlacionados en el espacio en el que se han tomado<sup>2</sup>. Esa afirmación implica que, utilizando un ejemplo del muestreo del Segundo Inventario Forestal Nacional (IFN2), el volumen maderable de una parcela localizada en el sitio  $z(a)$  se parecerá al volumen de una parcela localizada en el sitio  $z(a+h)$ , siendo  $h$  la distancia entre parcelas. Esta hipótesis tiene unas enormes implicaciones prácticas (siempre y cuando sea cierta), dado que permite interpolar la información puntual y convertirla en mapas digitales.

El objetivo del presente capítulo es la exposición de las técnicas geoestadísticas y su aplicación al análisis de datos procedentes del IFN<sup>3</sup>. En los apartados siguientes se hace una breve descripción de la metodología geoestadística y se presenta, mediante un ejemplo, el potencial y las restricciones del análisis propuesto.

<sup>1</sup> La estadística espacial incluye además, el análisis del patrón espacial y el análisis de datos en áreas.

<sup>2</sup> Esta hipótesis distingue la geoestadística de la estadística clásica, en la cual se supone que las muestras no están de ninguna manera correlacionadas entre sí (independencia de las observaciones).

<sup>3</sup> Dada la limitada extensión del capítulo la exposición de la metodología se ha reducido sustancialmente. El lector interesado puede obtener más información en los libros de GOOVAERTS (1997) o de CHILÉS Y DELFINER (1999).

## INTRODUCCIÓN AL ANÁLISIS ESPACIAL

El análisis se distingue en dos etapas muy diferentes entre sí: *el análisis estructural y la predicción espacial*. El objetivo de la primera parte es la cuantificación de la correlación espacial. En el caso de que ésta esté presente en los datos, se procede a la segunda fase que se denomina predicción espacial. En caso contrario (ausencia de correlación espacial) el posible uso de técnicas geoestadísticas no presentará ninguna ventaja frente a técnicas de estadística clásicas.

### Análisis estructural

#### *El variograma experimental*

Esta primera etapa del análisis tiene como objetivo estudiar y cuantificar la continuidad espacial de la variable de interés. Dicho objetivo se consigue mediante la construcción del llamado variograma experimental:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{a=1}^{N(h)} [z(u_a) - z(u_a + h)]^2$$

siendo  $\hat{\gamma}(h)$  la semivarianza para una distancia  $h$ ,  $N(h)$  el número de pares de datos separados por una distancia  $h$ , mientras que  $z(u_a)$  y  $z(u_a + h)$  son las mediciones de la variable en los sitios  $u_a$  y  $u_a + h$  respectivamente.

La experiencia práctica indica que la mayoría de los variogramas de las variables de la naturaleza presentan una forma parecida al variograma de la Figura 2: cuando la distancia de separación entre dos cualesquiera parcelas muestreadas es pequeña, la semivarianza toma valores pequeños, mientras que cuanto mayor sea dicha distancia, mayor es también la

semivarianza. Se puede observar que un variograma es una medida de la disimilitud entre parcelas de muestreo en función de la distancia que las separa (contrariamente a un gráfico de correlación espacial (Figura 1) que indica la similitud).

El aumento de la semivarianza se observa hasta una cierta distancia que se denomina *alcance* del variograma, que se define como aquella distancia de separación de parcelas en la cual se pierde la autocorrelación espacial. Existen otras dos características de un variograma que merecen especial atención: el valor de semivarianza cuando  $h$  tiende a cero se denomina *efecto pepita*<sup>4</sup>, mientras que el valor máximo de semivarianza se denomina *meseta*. El efecto pepita divide la varianza total del fenómeno en dos partes (Figura 2):

1. el porcentaje de varianza por encima del valor del efecto pepita (V1), es la que se explica por la proximidad espacial.
2. la otra parte de la variabilidad del fenómeno (V2), se puede atribuir a la variabilidad en escalas menores de la del muestreo realizado o a los posibles errores de medición.

La magnitud del efecto pepita indica la presencia o ausencia de variaciones en escala pequeña (menor que la del muestreo realizado) y es un buen indicador de los errores que se cometerán durante la fase de predicción de la variable.

Cuando la semivarianza de un variograma se estabiliza (existencia de meseta) el

<sup>4</sup> El término se refiere a las pepitas de oro encontradas en las explotaciones mineras: cuando se construyen variogramas de concentración de oro en una explotación, la agregación del oro en pepitas es la razón por la que la semivarianza en distancia  $h=0$  no es igual a cero.

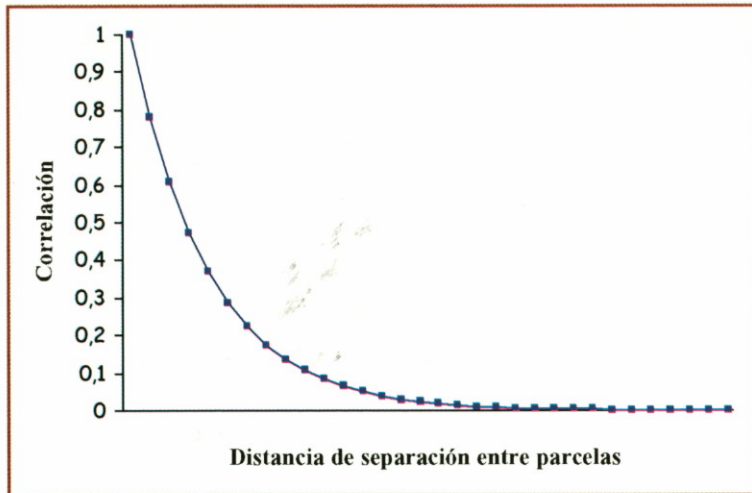


Figura 1. Gráfico de autocorrelación espacial

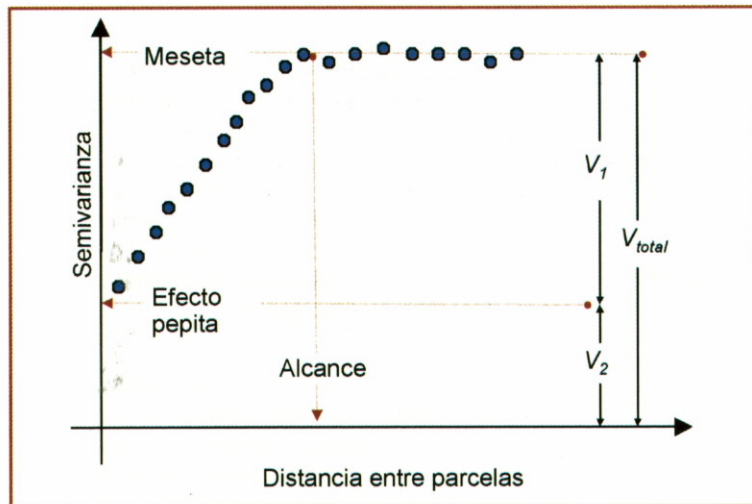


Figura 2. Ejemplo de un variograma

fenómeno se denomina *transitorio*. Estos fenómenos, los mas frecuentes en la naturaleza, se producen cuando en el área de estudio existen discontinuidades que separan compartimentos de altos y bajos valores según se expone en la Figura 3. En estos casos el alcance de correlación

espacial es aproximadamente igual a longitud media de los compartimentos. Cuando la semivarianza no se estabiliza en altas distancias de separación (ausencia de meseta) el fenómeno que se estudia no es estacionario y el tratamiento de los datos se tiene que basar en métodos de geoestadística no estacionaria.

### Modelización del variograma

Una vez construido el variograma experimental, se procede a su modelización con uno de los modelos de variogramas existentes. Para la predicción espacial es necesario ajustar un modelo de variograma, que proporcione los valores de semivarianza en distancias donde no existen valores experimentales. Un modelo de variograma tiene que cumplir la condición de ser una función positivamente semi-definida, necesaria para garantizar la existencia y singularidad de la solución del sistema de krigado (presentado en el apartado siguiente).

Entre una gran variedad de modelos, el esférico es el que mas se ha utilizado en las aplicaciones. Su fórmula viene dada por la siguiente expresión [ver por ejemplo, GOOVAERTS (1997)]:

$$\gamma = C[(3r/2a)+(r^3/2\alpha^3)]$$

donde  $r$  es la distancia,  $C$  la meseta, y  $\alpha$  el alcance del variograma.

El variograma experimental proporciona una medida de la disimilaridad de la variable, en función de la distancia que separa los puntos en los que ésta se ha medido. Su construcción proporciona información valiosa sobre el comportamiento espacial del fenómeno mientras que su modelización es imprescindible para el proceso de predicción que se presenta en el apartado siguiente.

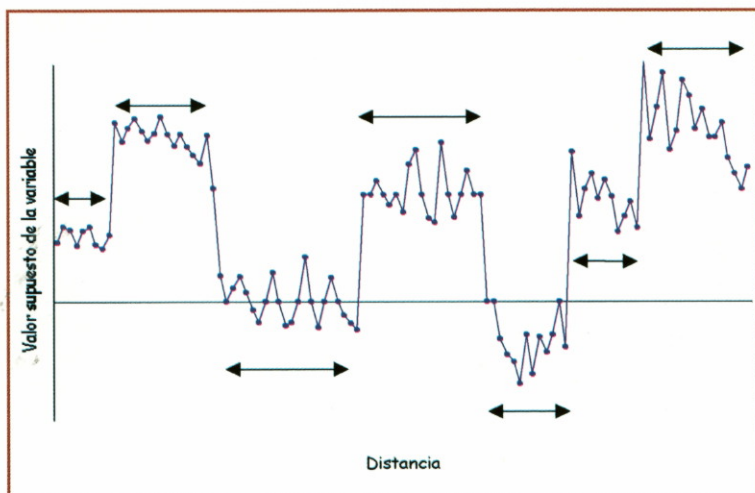


Figura 3. Representación gráfica de un fenómeno transitorio. Las flechas indican la longitud de los compartimentos, cuyo valor medio tiene que ser igual al alcance del variograma.

### Predicción espacial

La predicción espacial consiste en estimar la variable de interés en todos los puntos del área de estudio, utilizando para ello la información contenida en las parcelas muestreadas, junto con el variograma que anteriormente se ha construido. La predicción se realiza con un método denominado "kriging" [krigeado o krigeaje en Castellano (OLEA, 1991)] en honor del ingeniero de minas Dr. D. Krige quien fue el primero en utilizarlo.

Los estimadores del krigeado pertenecen a la clase de estimadores BLUE (*Best Linear Unbiased Estimator*) y se obtienen con métodos muy parecidos a los de la regresión lineal. Tienen dos propiedades importantes<sup>5</sup>:

1. La predicción es insesgada.
2. La varianza del error de predicción es la mínima posible, propiedad que garantiza que la precisión de la estimación va a ser la mayor posible.

En la geoestadística lineal la estimación  $z^*(\mathbf{u})$  en un punto  $\mathbf{u}$  no muestreado, se hace utilizando una combinación lineal de los valores de la misma variable en los puntos muestreados  $\{z(\mathbf{u}_a), a = 1, 2, \dots, n\}$ :

$$z^*(\mathbf{u}) = \sum_{a=1}^{n(\mathbf{u})} \lambda_a(\mathbf{u}) z(\mathbf{u}_a)$$

siendo  $\lambda_a(\mathbf{u})$  el peso asignado al dato  $z(\mathbf{u}_a)$ .

Utilizando esta formulación, el problema de encontrar el mejor estimador, se reduce a determinar cuales son los ponderadores ( $\lambda_a(\mathbf{u})$ ) óptimos. Imponer la condición de que el estimador  $z^*(\mathbf{u})$

<sup>5</sup> Estas dos propiedades son las que convierten el krigeado en el mejor método de estimación (lineal) espacial conocido hasta el momento, contrariamente a otros métodos de predicción espacial basados en la distancia inversa o los llamados "splines".

sea insesgado, equivale a imponer que la suma de los ponderadores  $\lambda_a(\mathbf{u})$  sea igual a uno (condición de universalidad):

$$\sum_{a=1}^{n(u)} \lambda_a(u) = 1$$

Además de cumplir la condición de ser insesgados, los estimadores de krigado tienen que ser de varianza mínima. Para conseguir dicha condición se define el error de la predicción<sup>6</sup>:  $R(\mathbf{u}) = Z^*(\mathbf{u}) - Z(\mathbf{u})$  y su varianza se iguala a cero:

$$\text{Var}\{R(\mathbf{u})\} \rightarrow 0$$

Se puede comprobar que la varianza del error en [5] depende solamente de la covarianza espacial, la cual se estima con el uso del variograma. La minimización de [5] se hace mediante la técnica de Lagrange bajo la condición [4] de universalidad, obteniéndose el siguiente sistema de ecuaciones, cuya solución proporciona los ponderadores óptimos:

$$\sum_{\beta=1}^{n(u)} \lambda_a \gamma(u_a - u_b) + \mu = \gamma(u_a - u) \quad a = 1, \dots, n(u)$$

$$\sum_{a=1}^{n(u)} \lambda_a(u) = 1$$

siendo  $\gamma(u_a - u_b)$  la semivarianza entre los puntos  $u_a$  y  $u_b$ , mientras que  $\mu$  es el multiplicador Lagrange.

<sup>6</sup> El error modelizado pero no el error real, ya que nunca se conoce el valor de  $Z(u)$

## CASO DE APLICACIÓN

Para la exposición de la metodología se han elegido 1072 parcelas del IFN2 localizadas en el norte de la provincia de Segovia (Figura 4). La variable que se analiza es el Volumen Maderable con Corteza (VCC) de las parcelas.

El variograma experimental obtenido para el VCC y el modelo de variograma ajustado se presentan en la Figura 5. Se puede observar que el variograma experimental cambia de pendiente a partir de una distancia de 4 km, algo que implica que existen dos escalas de variación espacial. La primera, con alcance igual a 4 km se debe atribuir a la similitud que presentan las parcelas localizadas en un mismo monte. La segunda con alcance de 40 km, representa la variabilidad a nivel provincial y se debe a diferencias en la productividad potencial forestal.

El efecto pepita es igual a 2000 y la meseta 4300, por consiguiente el porcentaje de la variabilidad total que se explica por la proximidad espacial es igual a 53,49%. El resto de la variabilidad del fenómeno se debe atribuir a las variaciones en una escala mayor que la del muestreo realizado o a posibles errores de medición y estimación del VCC. El elevado porcentaje de variabilidad debida al efecto pepita implica que existen factores a escala local que influyen en la cantidad de volumen maderable y que

estos factores tienen un rango de influencia menor que 1000 m (la distancia mínima entre parcelas). Esta parte de la variabilidad no se puede modelizar usando el muestreo del IFN2 pero sí usando parcelas en una malla más fina. La presencia del efecto pepita condiciona hasta un cierto punto el ámbito de utilización de las

predicciones; en concreto, dado que los datos utilizados para la predicción provienen de una escala nacional y además se observa (mediante el alto porcentaje del efecto pepita) variabilidad a escalas grandes, las predicciones que se van a dar se tienen que utilizar para la planificación forestal a nivel nacional pero no a nivel local (nivel de un monte por ejemplo).

En la Figura 6 se presenta el mapa de predicción del VCC para una parte de la región de estudio. Cabe destacar que para la misma región que se presenta en este mapa (los montes de Coca y Cuéllar), los mapas que se adjuntan en el IFN2 de la provincia de Segovia (ICONA, 1994) indican que el VCC es igual a 21-60 m<sup>3</sup>/ha. El mapa de la Figura 6 indica la existencia de variaciones significativas en cuanto al volumen maderable.

## DISCUSIÓN

La metodología geoes-tadística puede tener muchas aplicaciones en el campo forestal, una de las cuales es sin duda la extrapolación de los datos del IFN. Los mapas construidos con el método de kri-

gado se pueden utilizar para visualizar los datos que actualmente están almace-

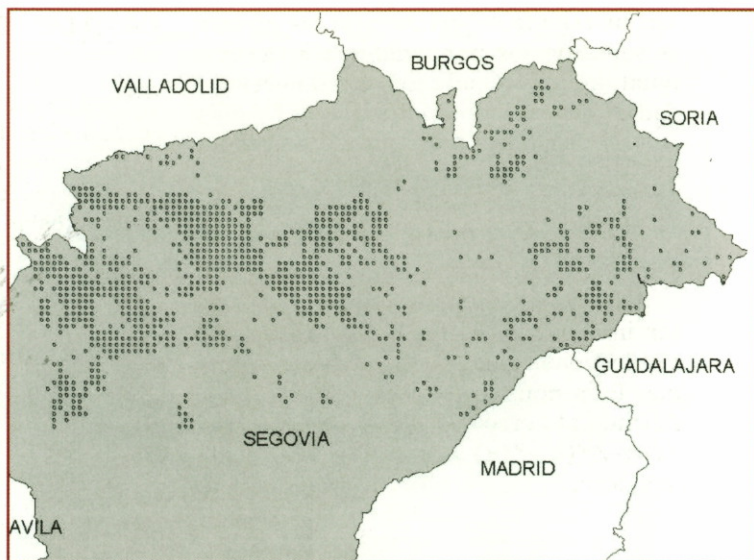


Figura 4. Localización de las 1072 parcelas del 2IFN utilizadas en el análisis

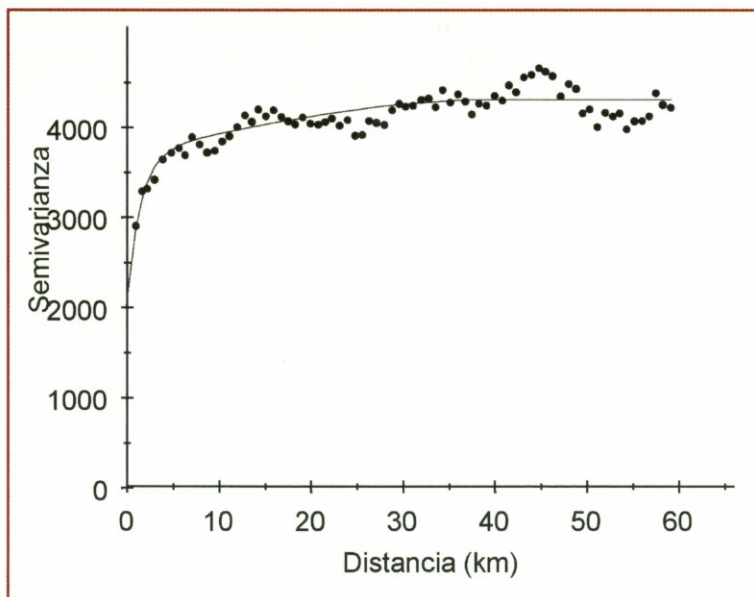


Figura 5. Variograma experimental y modelo ajustado para el VCC.

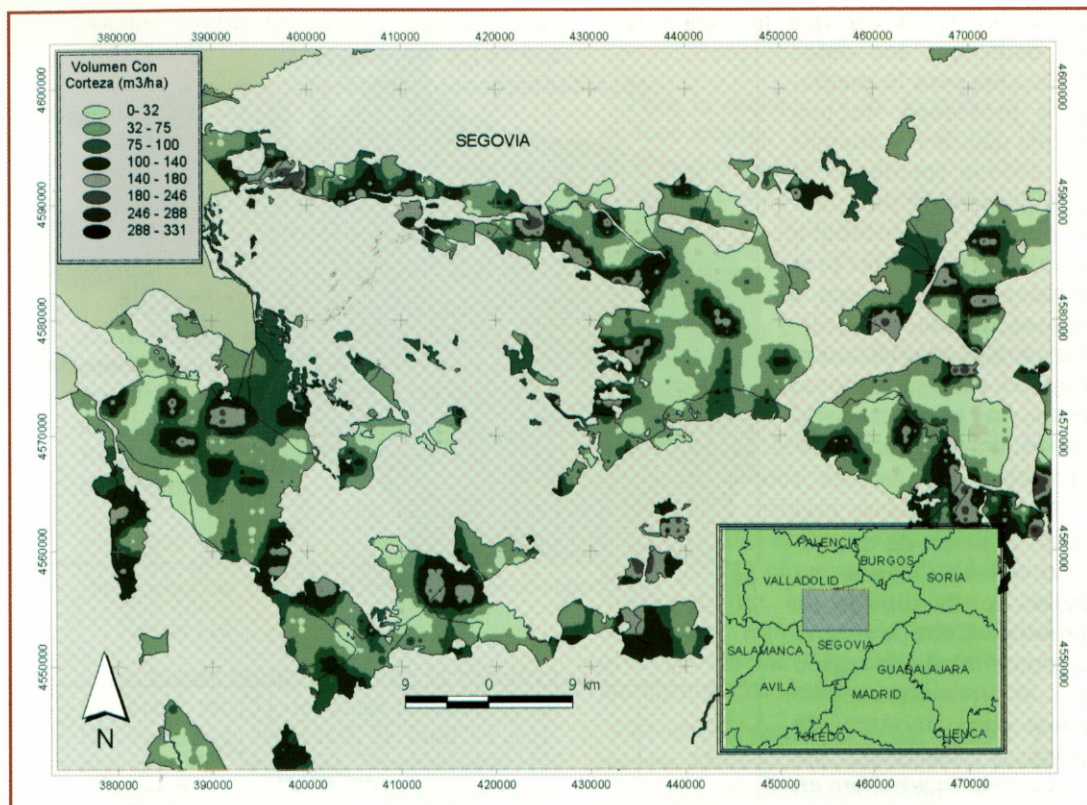


Figura 6. Mapa de predicción espacial del VCC para los montes de Coca y Cuéllar

nados en una serie de tablas que son de limitada utilización práctica.

El análisis geoestadístico da mejores resultados cuando los datos están distribuidos de manera sistemática en el espacio, condición que cumplen de manera absoluta los datos del IFN2. Las variables que se pueden analizar con éstas técnicas incluyen casi todas las variables de interés selvícola, siempre y cuando éstas presenten correlación espacial en la escala en la que se ha realizado el muestreo (escala regional o nacional). Una de estas variables puede ser, por ejemplo, la presencia ó ausencia de una cierta especie o el diámetro medio de la masa o como se ha indicado el volu-

men maderable por hectárea.

Sin embargo a la hora de decidir la variable que se va a analizar, uno tiene que tener en cuenta dos factores condicionantes de la presencia o ausencia de correlación espacial:

- el alcance de correlación que intuitivamente presenta la variable de interés (a)
- la distancia media entre parcelas de muestreo (d)

Se pueden distinguir dos situaciones:

1.  $a > d$ : la correlación espacial se puede detectar y es aconsejable el uso de geoestadística
2.  $a < d$ : la escala en la que varía el fenó-



meno es mayor que la escala del muestreo realizado y entonces el variograma indicará la ausencia de correlación espacial. Por ejemplo, si se quiere analizar la variable diámetro medio de la parcela, uno tiene que tener en cuenta que la división dasocrática condiciona la distribución espacial de la misma, de manera que cabe esperar que el alcance de correlación espacial de esta variable será igual a la longitud media de los tramos forestales. Si la distancia entre parcelas de muestreo es mayor que la longitud media de los tramos, la correlación espacial no se va a detectar.

Los mapas construidos con el krigado no representan la realidad del fenómeno que se estudia, sino la versión más probable de la misma. La construcción de un mapa de predicción espacial se basa en la información contenida en las parcelas muestreadas y en la correlación espacial existente en los datos. La predicción realizada con el método de krigado, utiliza un criterio de optimización que en este caso es la minimización de la varianza del error de predicción. Este tipo de estimación contiene errores que deben ser cuantificados con el fin de que el usuario de los mapas esté informado de la incertidumbre asociada a este tipo de estimación. Hoy día existen varios métodos de simulación geoestadística que se utilizan para cuantificar la incertidumbre espacial (CHILÉS Y DELFINER, 1999).

Por otra parte existen varias aproximaciones metodológicas que permiten incorporar en el proceso de predicción la información de variables secundarias, que a menudo se estiman con menos precisión que la variable primaria pero se pueden obtener para grandes extensiones territoriales (GOOVAERTS, 1997). La infor-

mación vía satélite, las fotografías aéreas o los modelos digitales del terreno son ejemplos típicos de este tipo de información. La incorporación de variables secundarias suele aumentar la precisión de la estimación final, siempre y cuando exista una buena correlación entre la variable primaria y la(s) secundaria(s).

En la bibliografía forestal se pueden encontrar varios casos de aplicación de la geoestadística a la estimación de variables relacionadas con la producción forestal como, por ejemplo, la calidad de la estación (HOCK *et al.*, 1993), el volumen maderable (HOLMGREN Y THURESSON, 1997), la distribución diamétrica (NANOS Y MONTERO, en prensa), la producción de resina (NANOS *et al.*, 2001) u otras relacionadas con caracteres como la diversidad genética (BUCCI Y VENDRAMIN, 2000; CORRE *et al.*, 1998) y estructural (WEBER, 2000). 🌲

## REFERENCIAS BIBLIOGRÁFICAS

- BUCCI, G.; VENDRAMIN, G. 2000. Delineation of genetic zones in the European Norway spruce natural range: preliminary evidence. *Molecular Ecology* **9**, 923-934.
- CHILÉS, J.P.; DELFINER, P. 1999. *Geostatistics: modeling spatial uncertainty*. John Wiley & Sons, New York.
- CORRE, V.L.; ROUSSEL, G.; ZANETTO, A.; KREMER, A.; LE CORRE, V. 1998. Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. III. Patterns of variation identified by geostatistical analyses. *Heredity* **80**, 464-473.
- CRESSIE, N.A.C. 1993. *Statistics for spatial data*. Revised Edition ed. John Wiley & Sons, New York.
- DAVIS, J.C. 1986. *Statistics and data analysis in geology*. Second edition ed. John Wiley & Sons, Chichester.
- GOOVAERTS, P. 1997. *Geostatistics for natural resources evaluation*. Oxford University Press, New York.

- GOTWAY, A.C. 1994. The use of conditional simulation in nuclear-waste-site performance assessment. *Technometrics* **36**, 129-141.
- HOCK, B.K.; PAYN, T.W.; SHIRLEY, J.W. 1993. Using a geographic information system and geostatistics to estimate site index of *Pinus radiata* for Kaingaroa Forest, New Zealand. *N. Z. J. For. Sci.* **23**, 264-277.
- HOLMGREN, P.; THÜRESSON, T. 1997. Applying objectively estimated and spatially continuous forest parameters in tactical planning to obtain dynamic treatment units. *For. Sci.* **43**, 317-326.
- HOULLIER, F.; PIERRAT, J.C. 1992. Application des modèles statistiques spatio-temporels aux échantillonnages forestiers successifs. *Can. J. For. Res.* **22**, 1988-1995.
- ICONA. 1994. Segundo Inventario Forestal Nacional, 1986-1995, Castilla y León, Segovia. MAPA, ICONA, Madrid.
- JOURNAL, A.G.; HUIJBREGTS, C.J. 1978. *Mining geostatistics*. Academic press, London.
- KOHL, M.; GERTNER, G. 1997. Geostatistics in evaluating forest damage surveys: considerations on methods for describing spatial distributions. *For. Ecol. Manage.* **95**, 131-140.
- LEGENDRE, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**, 1659-1673.
- MARTINEZ-COB, A.; CUENCA, H.R. 1992. Influence of elevation on regional evapotranspiration using multivariate geostatistics for various climatic regimes in Oregon. *Journal of Hydrology* **136**, 353-380.
- MATÉRN, B. 1994. On spatial statistics in forestry. *Rev. Ciencia Forestal en México* **19**, 141-152.
- MATHERON, G. 1963. Principles of geostatistics. *Econ. Geol.* **58**, 1246-1266.
- NANOS, N.; MONTERO, G. in press. Spatial prediction of diameter distribution models. *For. Ecol. Manage.*
- NANOS, N.; TADESSE, W.; MONTERO, G.; GIL, L.; ALÍA, R. 2001. Spatial stochastic modeling of resin yield from pine stands. *Can. J. For. Res.* **31**, 1140-1147.
- OLEA, R. 1991. *Geostatistical glossary and multilingual dictionary*. Oxford University Press.
- WEBER, J. 2000. *Geostatistische Analyse der Struktur von Waldbeständen am Beispiel ausgewählter Bannwälder in Baden-Württemberg*, FVA Baden-Württemberg.