Available online at www.sciencedirect.com
SCIENCE DIRECT®
Forest Ecology and Management 195 (2004) 221–235
Forest Ecology and Management
www.elsevier.com/locate/foreco

Geostatistical prediction of height/diameter models

Nikos Nanos[a,*], Rafael Calama[b,1], Gregorio Montero[b,2], Luis Gil[c,3]

[a]Dpto. Producción Vegetal y Silvopascicultura, UVA, Avda de Madrid 44, 34004 Palencia, Spain
[b]Department of Silviculture, CIFOR-INIA, Apdo 8111, 28080 Madrid, Spain
[c]Unit of Physiology and Genetics, ETSI Montes, Ciudad Universitaria, 28040 Madrid, Spain

Received 5 June 2003; received in revised form 3 October 2003; accepted 20 February 2004

Abstract

Height/diameter models are used in forest inventories and in stand projection systems mainly because height measurements are expensive and time-consuming. Mixed models have been frequently employed in *H*/*D* modelling assuming that model parameters can vary among stands. Using mixed models, the prediction of the random effects requires at least one additional measurement in the stand. In the present paper, we advocate the idea that empirical predictors for the random stand effects of a mixed model can be obtained using geostatistics when spatial correlation is present. Using the proposed methodology we obtained site-specific height/diameter curves without additional measurements. The case study was carried out in pure maritime pine stands of central Spain. Height/diameter measurements were available from the Second Spanish National Forest Inventory. Final predictions of random stand effects were obtained by kriging. Cokriging was also used for random stand effect prediction, using as a secondary attribute the stand density.

Keywords: *H*/*D* curves; Kriging; Variogram; Cokriging

1. Introduction

Tree-height measurements are expensive and time-consuming and often they are performed only on a small portion of sampled trees. The development of a relationship between total tree height and diameter at breast height (DBH) is considered crucial in forest inventories as well as in stand projection systems since it provides height predictions using an easily measured predictor variable such as DBH.

The construction of height/diameter (*H*/*D*) models is sometimes performed using mixed effect models, assuming that parameters of the model can vary randomly across stands. The localisation of the *H*/*D* curve has been made with the use of other stand variables (Lappi, 1991, 1997) or according to the site or region where the curve is going to be used (Fulton, 1999; Huang et al., 2000). These approaches to *H*/*D* modelling are based on the assumption that site conditions and previous silvicultural practices influence the relationship between total tree height and DBH.

In the models that use stand variables for prediction, there does not seem to be a single or universal variable to explain variation in stand-specific parameters, since these depend on the investigator's intuition, model

* Corresponding author. Tel.: +34-979108421; fax: +34-979108439.
E-mail addresses: nanos@pvs.uva.es (N. Nanos), rcalama@inia.es (R. Calama), montero@inia.es (G. Montero), lgil@montes.ump.es (L. Gil).
[1] Tel.: +34-91-3476868.
[2] Tel.: +34-91-3474198.
[3] Tel.: +34-91-3367113.

0378-1127/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.foreco.2004.02.031

specification, site productivity (Fulton, 1999), and thinning regime (Zhang et al., 1997). Fang et al. (1998), for example, used percentiles of the diameter distribution; Zakrzewski and Bella (1988) found that the quadratic mean diameter and height are the best explanatory variables, while Parresol (1992) used the basal area per hectare.

A more appealing approach for *H*/*D* modelling is the one presented by Huang et al. (2000) since their method does not require additional stand measurements for prediction. However, prior to the construction of this type of model, one needs to classify the study area into ecologically homogeneous areas and then specify a unique set of parameters for each of the areas.

Alternatively it can be assumed that *H*/*D* parameters vary continuously in space. Intuitively this assumption can be reasonable for forest stands, since site productivity is generally continuous in space while silvicultural activities are applied in spatially homogeneous areas. Geostatistics can offer an alternative approach for *H*/*D* parameter prediction if the above assumption of spatial continuity can be considered reasonable. The theory of regionalised variables and kriging offer the theoretical background over which unbiased predictions can be made by minimising and quantifying the prediction error (Matheron, 1963; Journel and Huijbregts, 1978).

In the geostatistical framework the presence of spatial correlation makes possible the prediction of a variable of interest at locations where no secondary measurements are available. On the other hand, cokriging, as a generalisation of the usual kriging system of equations, can account for secondary information when this is available. Depending on the nature of the secondary attribute, various cokriging alternative formulations have been proposed (Goovaerts, 1997).

In the present paper we propose a geostatistical approach for the estimation of stand-specific parameters of *H*/*D* models. Predictions are made by both kriging and cokriging using, in the second case, the stand density as a secondary attribute.

2. Data

The Second National Forest Inventory of Spain (NFI-2) provided the data for this study. Inventory plots were distributed systematically over Spain, on the nodes of a regular grid of 1 km in size. Plots were circular of various radius (the minimum measured diameter varied with the radius of the plot), therefore the inclusion of a tree in the sample was a function of its DBH and its distance from the centre of the plot.

The area of study is located in central Spain (Segovia province) where natural stands of maritime pine (*Pinus pinaster* Ait.) are grown on a very poor soil that consists mainly of dunes. The maritime pine stands of this area have traditionally been dedicated to resin production, therefore, silviculture had to be adapted to an increasing demand of stems to be tapped.

The plots included in the analysis (250 plots) are located in pure, even-aged, natural maritime pine stands (Fig. 1). We used only untapped trees for constructing the *H*/*D* model: resin tapping results in a strong deformation of diameter in the height of face, so measurements are not representative of the real diameter at breast height. On the other hand, it is likely that tapping reduces height growth, so there is no way to construct a unique *H*/*D* prediction system for both tapped and untapped trees.

3. Statistical analysis

3.1. Model specification

Our data are based on a sample of multiple tree measurements (DBH, total height) taken from different plots. This nested structure calls for a mixed modelling approach that allows the simultaneous estimation of fixed and random effects. This type of *H*/*D* model has been fitted to longitudinal data from re-measured plots (Hokka, 1997; Lappi, 1997) as well as to cross-sectional data obtained simultaneously from different plots (Jayaraman and Lappi, 2001).

The basic model used in this study is the two-parameter exponential equation:

$$H - 1.3 = \alpha\, e^{\beta/d} \quad (1)$$

where H is the total height (m) and d the diameter at breast height (cm). The two parameters (α, β) control the asymptotic height and the inflexion point of the model, respectively.

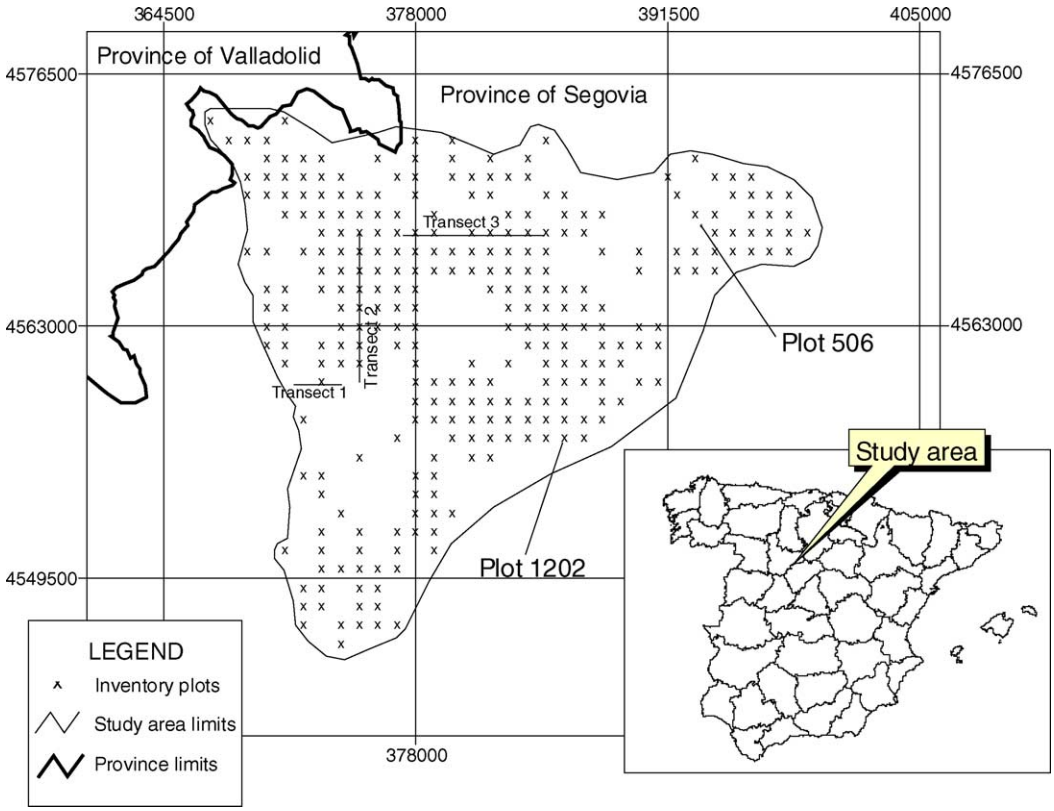

Fig. 1. Location of inventory plots. Co-ordinates are UTM (zone-30). Three transects and two inventory plots were randomly selected, in order to predict the *H*/*D* curves.

3.2. Nonlinear mixed modelling

The general expression for a nonlinear mixed effects model can be written as (Lindstrom and Bates, 1990)

$$y_{ij} = f(\boldsymbol{\Phi}_i, x_{ij}) + e_{ij} \tag{2}$$

where y_{ij} is the jth response on the plot i ($i = 1$ to s), x_{ij} the jth measurement of predictor variable x on plot i ($j = 1$ to n_i), $\boldsymbol{\Phi}_i$ is a parameter vector which can vary from plot to plot, f a nonlinear function of the predictor vector and the parameter vector, and e_{ij} a noise term. In a vectorial form

$$\boldsymbol{y}_i = f(\boldsymbol{\Phi}_i, \boldsymbol{x}_i) + \boldsymbol{e}_i \tag{3}$$

where y_i is the ith plot's entire response $[y_{i1}, y_{i2}, \ldots, y_{ij}, \ldots, y_{ini}]^{\mathrm{T}}$, $\boldsymbol{x}_i$ the ith plot's predictor vector $[x_{i1}, x_{i2}, \ldots, x_{ij}, \ldots, x_{ini}]^{\mathrm{T}}$ and $\boldsymbol{e}_i = [e_{i1}, e_{i2}, \ldots, e_{ij}, \ldots, e_{ini}]^{\mathrm{T}}$ the vector of the residual terms.

Letting $\boldsymbol{\Phi}_i$ vary from plot to plot (Davidian and Giltinan, 1995):

$$\boldsymbol{\Phi}_i = \boldsymbol{A}_i\boldsymbol{\lambda} + \boldsymbol{B}_i\boldsymbol{b}_i \tag{4}$$

where $\boldsymbol{\lambda}$ is a vector of fixed effects, common for the complete population, $\boldsymbol{b}_i$ a vector of random effects associated with the ith plot, and $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are design matrices for the fixed and random parameters, which are usually matrices containing only zeroes and ones as well as fixed and random effect covariates.

The basic assumptions for the nonlinear mixed effect model are the following:

$$\boldsymbol{b}_i \sim N(0, \boldsymbol{D}) \tag{5}$$

$$\boldsymbol{e}_i \sim N(0, \boldsymbol{R}_i(\boldsymbol{\lambda}, \boldsymbol{b}_i, \boldsymbol{\rho})) \tag{6}$$

where $\boldsymbol{D}$ is the covariance matrix for the random effects (common for all the plots), which defines between-plot variation, $\boldsymbol{R}_i(\boldsymbol{\lambda}, \boldsymbol{b}_i, \boldsymbol{\rho})$ is the within-plot

variance–covariance matrix, which defines the within-plot structure, and $\boldsymbol{\rho}$ is a vector of within-plot covariance parameters.

The main objective of the analysis in nonlinear mixed models is to estimate $\boldsymbol{\lambda}$, $\boldsymbol{D}$ and $\boldsymbol{\rho}$. Inference over these parameters is based on the maximisation of the marginal likelihood function (Palmer et al., 1991; Vonesh and Chinchilli, 1997):

$$\prod_{i=1}^{s} \int p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{\lambda}, \boldsymbol{\rho}, \boldsymbol{b}_i) p(\boldsymbol{b}_i, \boldsymbol{D})\, \mathrm{d}\boldsymbol{b}_i \tag{7}$$

The first term of this equation is the conditional density of $\boldsymbol{y}_i$ given $\boldsymbol{b}_i$, while the second term refers to the density of $\boldsymbol{b}_i$. The integral of their product is termed the marginal distribution of $\boldsymbol{y}_i$.

The maximisation of this expression requires numerical integration of the random effects, which is difficult because $\boldsymbol{b}_i$ enters in the model in a nonlinear fashion. Several approximations have been developed in order to deal with this problem, including methods such as the adaptive Gaussian quadrature (Pinheiro and Bates, 1995), which tries to numerically solve the integral. This method is available in the SAS procedure NLMIXED (SAS Institute, 2000). Most commonly used methods are those based on a linear approximation to the marginal likelihood function by expanding it with a first-order Taylor series to be linear on $\boldsymbol{b}_i$. This expansion can be either at a value of $\boldsymbol{b}_i = 0$ (Sheiner and Beal, 1980) or about a value $\boldsymbol{b}_i^*$ close to $\boldsymbol{b}_i$ (Lindstrom and Bates, 1990), defined as the EBLUP (empirical best linear unbiased predictor). The last option is available in the SAS macro NLINMIX.

Following Lindstrom and Bates (1990), after linear expansion we can obtain the estimates $\boldsymbol{D}$ and $\boldsymbol{R}_i$ of $\boldsymbol{D}$ and $\boldsymbol{R}_i$, either by maximum likelihood or restricted maximum likelihood, and also it is possible to define $\boldsymbol{y}_i$ by a given "pseudoresponse" $\tilde{\boldsymbol{y}}_i$ vector, which can be defined as a linear function over $\boldsymbol{b}_i$ and $\boldsymbol{\lambda}$:

$$\boldsymbol{y}_i = f(\boldsymbol{\Phi}_i, \boldsymbol{x}_i) + \boldsymbol{e}_i \approx \tilde{\boldsymbol{y}}_i = \boldsymbol{X}_i \cdot \boldsymbol{\lambda} + \boldsymbol{Z}_i \cdot \boldsymbol{b}_i + \tilde{\boldsymbol{e}}_i \tag{8}$$

$$E(\tilde{\boldsymbol{y}}_i|\boldsymbol{x}_i) = \boldsymbol{X}_i \cdot \boldsymbol{\lambda} \tag{9}$$

$$\boldsymbol{V}_i = \mathrm{Var}(\tilde{\boldsymbol{y}}_i|\boldsymbol{x}_i) = \boldsymbol{R}_i + \boldsymbol{Z}_i \cdot \boldsymbol{D} \cdot \boldsymbol{Z}_i^{\mathrm{T}} \tag{10}$$

$$\boldsymbol{Z}_i = \left(\frac{\partial}{\partial b_i} f(\boldsymbol{\Phi}_i, \boldsymbol{x}_i) \Big|_{b_i = b_i^*, \boldsymbol{\lambda}} \right) \cdot \boldsymbol{B}_i \tag{11}$$

Under this assumption, and following linear models theory (Searle, 1971), we can obtain generalised least-squares estimates of $\boldsymbol{\lambda}$ and prediction for $\boldsymbol{b}_i$:

$$\hat{\lambda} = \left(\sum_{i=1}^{s} \boldsymbol{X}_i^{\mathrm{T}} \cdot \boldsymbol{V}_i^{-1} \cdot \boldsymbol{X}_i \right)^{-1} \cdot \sum_{i=1}^{s} \boldsymbol{X}_i^{\mathrm{T}} \cdot \boldsymbol{V}_i^{-1} \cdot \tilde{\boldsymbol{y}}_i \tag{12}$$

$$\hat{b}_i = \boldsymbol{D} \cdot \boldsymbol{Z}_i^{\mathrm{T}} \cdot \boldsymbol{V}_i^{-1} \cdot (\tilde{\boldsymbol{y}}_i - \boldsymbol{X}_i \cdot \hat{\lambda}) \tag{13}$$

where $\hat{\lambda}$ is the estimate for the vector of parameters for the fixed effects $\boldsymbol{\lambda}$, and $\hat{b}_i$ the empirical best linear unbiased predictor (EBLUP) for the vector of parameters for the random effects on plot i, $\boldsymbol{b}_i$.

3.2.1. Parameter estimation

According to Fang and Bailey (2001), during the construction of a mixed model one needs to determine:

1. Which of the parameters α and β in Model (1) are going to be considered as mixed effects (both random and fixed) or purely fixed.
2. The within-plot structure (variance–covariance structure for explaining variability among trees in a same plot) as well as the between-plot variance–covariance matrix.
3. Which covariates to use for explaining among plot variability.

3.2.1.1. Parameter effects. The determination of which effects are going to vary randomly among plots is very flexible (Gregoire et al., 1995). Consideration of both parameters of Model (1) as mixed led to problems of convergence. On the other hand, a preliminary separate fit of Model (1) to each plot using ordinary least-squares, showed that parameter α was more variable than β, since estimates for β show less overlap in confidence interval across plots (Pinheiro and Bates, 1998). From a biological perspective, this means that between-plot variation is smaller in height–diameter curve inflexion point than in asymptote.

As parameters with higher variability should be preferred as mixed effects (Fang and Bailey, 2001), we decided to consider β as a fixed parameter, allowing parameter α to vary randomly from plot to plot. Designing β as fixed does not mean this parameter is truly constant across plots, but the true underlying process is approximated reasonably regarding it as fixed (Davidian and Giltinan, 1995), although it can be

explained by including stand covariates in a second stage. The general expression of Model (1) is then given by

$$H_{ij} - 1.3 = (a + v_i)\, e^{b/d_{ij}} + \varepsilon_{ij} \quad (14)$$

where a and b are fixed effects, v_i is a random effect, specific for each plot, d_{ij} the diameter at breast height for the jth tree in the ith plot and ε_{ij} the tree within-plot level error.

3.2.1.2. Within-plot variance–covariance matrix. Within a given sampling unit it is usual to find a pattern of increasing error variance as a function of the independent variable (DBH) (Huang et al., 2000). In order to avoid the violation of the basic assumption of constant variance of errors, we included a variance function in the expression of the within-plot structure (Lappi, 1997). We tested both the power and the exponential models for this reason, using different parameter values. Comparison among different models and different parameter values was done under the criterion of smaller AIC (Akaike Information Criteria). The smallest AIC value was obtained when using the inverse of the power function over the diameter with the parameter value equal to 0.5.

The expression for the within-plot variance–covariance matrix $\boldsymbol{R}_i(\boldsymbol{\lambda}, \boldsymbol{b}_i, \boldsymbol{\rho})$ is then given by

$$\boldsymbol{R}_i(\boldsymbol{\lambda}, \boldsymbol{b}_i, \boldsymbol{\rho}) = \sigma^2 \cdot \boldsymbol{G}_i \quad (15)$$

where for a plot i with n_i height–diameter measurements, $\boldsymbol{G}_i$ is an $n_i \times n_i$ diagonal matrix, with components $1/D_j^{0.5}$ and σ^2 is a scaling factor for the error dispersion (Gregoire et al., 1995), given by the value of the residual variance of the model.

3.2.1.3. Between-plot variance–covariance matrix. At this stage we defined the structure of the variance–covariance matrix for the random effects $\boldsymbol{D}$. Using a single random effect v_i, $\boldsymbol{D}$ *is a* 1×1 *matrix, whose unique component is the variance for the random effect.*

3.2.1.4. Covariate modelling. A complementary way to explain between-plot variability is to take fixed effects as a function of explanatory variables at the plot level (Hokka, 1997; Fang and Bailey, 2001). A preliminary least-squares fit for each plot showed a high linear dependence between parameter β and plot density (stems/ha):

$$\beta = b + cN \quad (16)$$

where N is the density of the plot and b and c are fixed effects.

After including stand density, final expression for the nonlinear Model (1) is

$$H_{ij} - 1.3 = (a + v_i)\, e^{(b+cN_i)/d_{ij}} + \varepsilon_{ij} \quad (17)$$

where a, b and c are fixed effects, v_i is a random effect for plot i, d_{ij} the diameter at breast height for the jth tree in the ith plot and ε_{ij} the tree within-plot level error.

In order to estimate the fixed effects as well as the components of the variance–covariance matrices for between- and within-plot variation we used SAS macro-NLINMIX. Values for the EBLUPs of the random effect were also predicted for each plot. Since some elements of the parameter vector were taken as fixed, maximisation of the marginal likelihood function was achieved using the linearisation approach (Davidian and Giltinan, 1995), with a first-order Taylor series expansion about the estimates of the random parameter, and maximum likelihood for variance components estimation.

The results of the fit using SAS macro NLINMIX are shown in Table 1. Apart from the usual parameter estimations and their standard errors we present the estimate of the unique element of the between-plot variance–covariance matrix $\boldsymbol{D}$ as well as the scaling factor for the error dispersion (σ^2) of the within-plot variance–covariance matrix $\boldsymbol{R}_i$.

3.3. Geostatistical prediction

Prediction at unsampled locations using Model (17) requires knowledge of the stand density, as well as

Table 1
Parameter estimates and fit statistics for the fixed and random effects of Model (17)

Effect	Estimate	Standard error	t-Value	Pr > |t|
a	16.3536	0.2835	57.69	<0.0001
b	−17.8424	0.4906	−36.37	<0.0001
c	0.0044	0.0009	4.56	<0.0001
Variance, v_i	8.1985			
σ^2 residual	0.2638			

knowledge of the EBLUP for the random effect v_i. We assume that stand density can be easily obtained from aerial photography or satellite imagery. To predict v_i is more complex, since at least one additional measurement is required at the specific location. If no additional measurements are available we have to consider v_i is equal to zero. Geostatistics can offer an alternative way of spatially predicting the required attributes.

3.3.1. Variogram analysis

At the first stage of the geostatistical part of the analysis we computed the experimental variograms for the plot density and for the EBLUP of the random effect v_i of Model (17). We used the classical formula:

$$\hat{\gamma}(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})}\sum_{a=1}^{N(\boldsymbol{h})}[z(u_a) - z(u_a + \boldsymbol{h})]^2 \quad (18)$$

where $\hat{\gamma}(\boldsymbol{h})$ is the semivariance for distance $\boldsymbol{h}$, $N(\boldsymbol{h})$ the number of pairs of data locations a vector $\boldsymbol{h}$ apart, while $z(u_a)$ and $z(u_a + \boldsymbol{h})$ are measurements at locations u_a and $u_a + \boldsymbol{h}$, respectively. For the geostatistical calculations we used the *spatial statistics* module of S-PLUS (1998).

The experimental variogram gives a measure of spatial correlation of the studied attribute. Typically, the semivariance values exhibit an ascending behaviour near the origin of the variogram and they usually level off at larger distances (the *sill* of the variogram). The semivariance value at distances close to zero is called the *nugget effect*. The distance at which the semivariance levels off is the *range* of the variogram and it represents the separation distance at which two samples can be considered to be spatially independent.

3.3.2. Anisotropy

A phenomenon is said to be anisotropic when its pattern of spatial variability changes with direction. Anisotropy can be detected by variogram maps (Isaaks and Srivastava, 1989). A variogram map is a contour map of the sample variogram surface. The directions of maximum and minimum spatial continuity are judged visually.

Once the main axes of anisotropy are identified, one proceeds to modelling the directional variograms. The construction of directional variogram models has to be done in such a way that the posterior spatial prediction can be realised: during kriging, anisotropy has to be corrected by rotating the co-ordinate system clockwise so as to identify the main axes of anisotropy and linearly transforming the rotated co-ordinates according to the anisotropic variogram model (Goovaerts, 1997). Once the anisotropic space is transformed to an isotropic one, kriging is used for prediction, and finally the transformed co-ordinates are back-transformed to their actual values.

3.3.2.1. Random effect v_i. Anisotropy was present in the case of the EBLUP for the random effect v_i, as shown in the variogram map of Fig. 2a. The axes of maximum and minimum spatial continuity lie at 150° and 60°, respectively. The directional variograms and their respective models are presented in Fig. 2b and c.

The variograms used to model the experimental counterparts are the following:

$$\gamma(h_{60}) = 3.5 + 3.2 \times \mathrm{SPHER}_{3000}(h_{60}) + 1.1 \times \mathrm{SPHER}_{6000}(h_{60}) \quad (19)$$

$$\gamma(h_{150}) = 3.5 + 3.2 \times \mathrm{SPHER}_{3000}(h_{150}) + 1.1 \times \mathrm{SPHER}_{11,000}(h_{150}) \quad (20)$$

Each directional variogram model is composed of three structures:

- a nugget component equal to 3.5 for both directions;
- a short-range spherical model (SPHER) (Goovaerts, 1997), with range equal to 3.000 m and a partial sill equal to 3.2;
- a long-range model that presents different ranges depending on the direction (6000 and 11,000 m for directions of 60° and 150°, respectively) and partial sill equal to 1.1.

3.3.2.2. Plot density. The variogram map for the plot density (not presented) showed that the spatial behaviour was isotropic. The omnidirectional experimental variogram and its model are shown in Fig. 3. Two apparent structures are present in this variogram: a short-range structure, up to a distance of 2000 m approximately and a long-range one up to 18,000 m. The variogram model is composed of two structures (also called *nested structures*):

$$\gamma(h) = 10{,}000 + 5500 \times \mathrm{SPHER}_{1600} + 4500 \times \mathrm{SPHER}_{18,000} \quad (21)$$

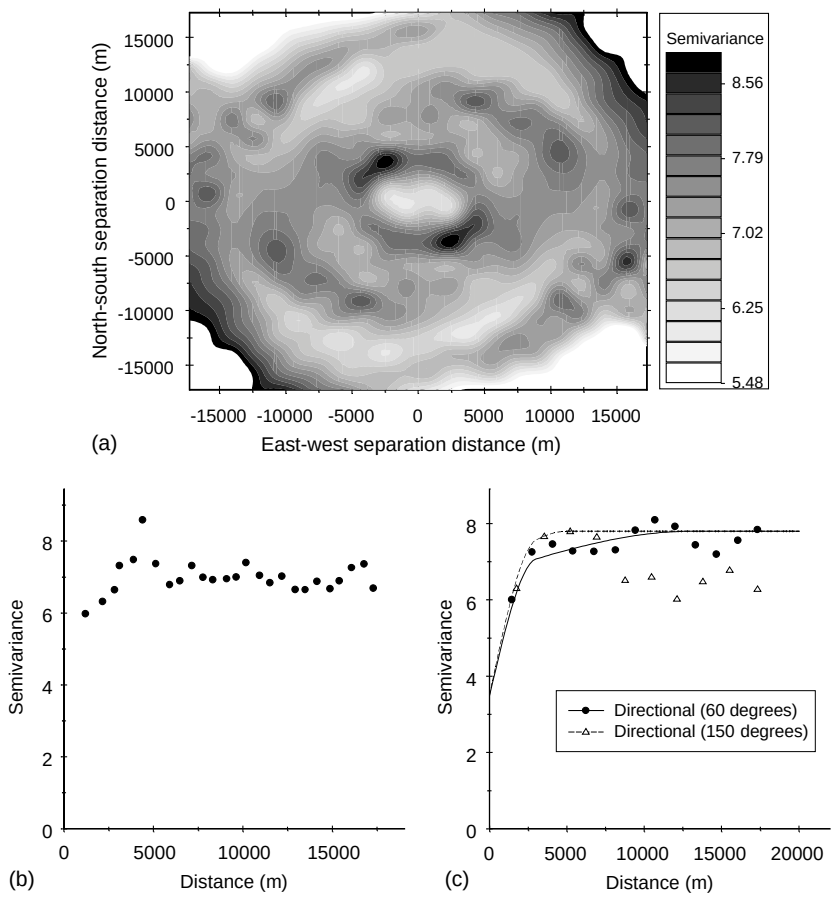

Fig. 2. Variogram map (a), omnidirectional experimental variogram (b) and directional experimental variograms and their respective models (c) for the random effect v_i of Model (17).

3.3.3. Kriging

In linear geostatistics the prediction at an unsampled location $z^*(u)$ is given using a linear combination of the values of the variable at sampled locations $z(u_a)$, $\alpha = 1, 2, \ldots, n$, with n being the total number of samples

$$z^*(u) = \sum_{a=1}^{n(u)} \lambda_a(u) z(u_a) \tag{22}$$

The kriging weights $[\lambda_a(u)]$ that minimise the error variance are obtained with a method similar to that of least-squares regression, under the condition that the prediction is unbiased. This condition (frequently called the universality condition) is satisfied by

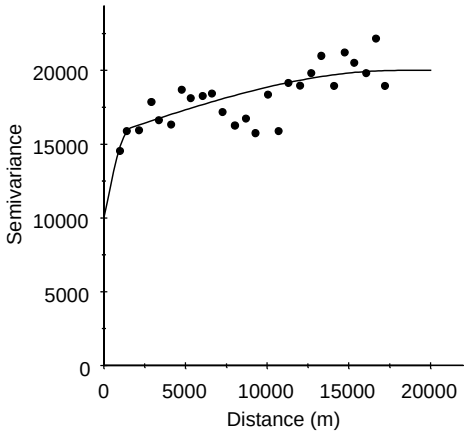

Fig. 3. Omnidirectional experimental variogram and model for the plot density.

constraining the weights $[\lambda_a(u)]$ to sum to unity. The minimisation of the error variance under the universality constraint leads to the ordinary kriging system of equations that can be written in matrix form (Journel and Huijbregts, 1978):

$$\boldsymbol{C}_{\alpha\beta}\boldsymbol{\lambda} = \boldsymbol{c} \tag{23}$$

where

$$\boldsymbol{C}_{\alpha\beta} = \begin{bmatrix} C(u_1,u_1) & \dots & \dots & C(u_1,u_n) & 1 \\ \dots & C(u_2,u_2) & \dots & \dots & 1 \\ C(u_n,u_1) & \dots & \dots & C(u_n,u_n) & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \tag{24}$$

being $C(u_a, u_b)$ the covariance[4] for the pair u_a, u_b.

The vector $\boldsymbol{\lambda}$ contains the unknown kriging weights and the Lagrange multiplier μ, while $\boldsymbol{c}$ is the vector of covariances between the point where the prediction is required (u) and the sampled plots (u_a):

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix}, \qquad \boldsymbol{c} = \begin{bmatrix} C(u_1,u) \\ C(u_2,u) \\ \vdots \\ C(u_n,u) \\ 1 \end{bmatrix} \tag{25}$$

The solution of the kriging system provides the estimation $z^*(u)$ as well as the kriging variance that is usually used as a measure of the uncertainty of the prediction.

3.3.4. Filtering systematic errors

Kriging is known to be an exact interpolator when the nugget effect of the variogram can be attributed to short-range variability. On the contrary, if part of the nugget effect is interpreted as measurement (or estimation) error, then this component of variability should be filtered out since one is usually interested in knowing the real value of the attribute at location $z(u)$ and not the value distorted by measurement errors (Cressie, 1993).

[4] Under the assumption of stationarity, the kriging system can be expressed in terms of semivariograms or covariances. Covariances are preferred for reasons of computational efficiency (Goovaerts, 1997).

In our case the estimation of the EBLUP v_i of Model (17) is distorted by two types of error: a standard error of estimation and possible errors of tree-height measurements. The plot density, as an estimator of the true stand density, is also associated with an estimation error. In order to filter out this systematic error and obtain an error-free estimation, the variance of the estimation error of both variables (random effect and plot density) has to be added to the diagonal terms of $\boldsymbol{C}_{\alpha\beta}$. Kriging is no longer an exact interpolator (Chilés and Delfiner, 1999).

We performed kriging on the nodes of 100 m × 100 m grid, using a global search strategy, since the experimental variograms do not indicate deviations from stationarity. The kriged maps for the random effect v_i and for the stand density are shown in Fig. 4.

3.3.5. The use of auxiliary variables

During this step of the analysis we investigated the possibility of including a secondary variable into the prediction of the random effect v_i. The variable chosen for this reason was the stand density, which is already used in Model (17). The inclusion of the stand density into the spatial prediction of v_i can be realised if spatial cross-correlation between them is present in the data. This type of cross-correlation is examined with the use of cross-variograms.

3.3.5.1. Cross-variograms: inference and modelling. Let $Z_f(x)$ denote a multivariate random function indexed by the subscript $f = 1, 2, \dots, p$ defined over the domain D of R^2 and sampled over a set of N_f points ($N_f > 0$). In the simplest case where $f = 2$ the cross-variogram between Z_1 and Z_2 is computed using the formula

$$\hat{\gamma}_{12}(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})}\sum_{a=1}^{N(\boldsymbol{h})}[z_1(u_a) - z_1(u_a+\boldsymbol{h})][z_2(u_a) - z_2(u_a+\boldsymbol{h})] \tag{26}$$

where the symbols are as in (18), but with two variables implicated instead of one.

In the multivariate case one needs to infer $f(f+1)/2$ direct and cross-variograms. The difficulty lies on the fact that these variograms cannot be

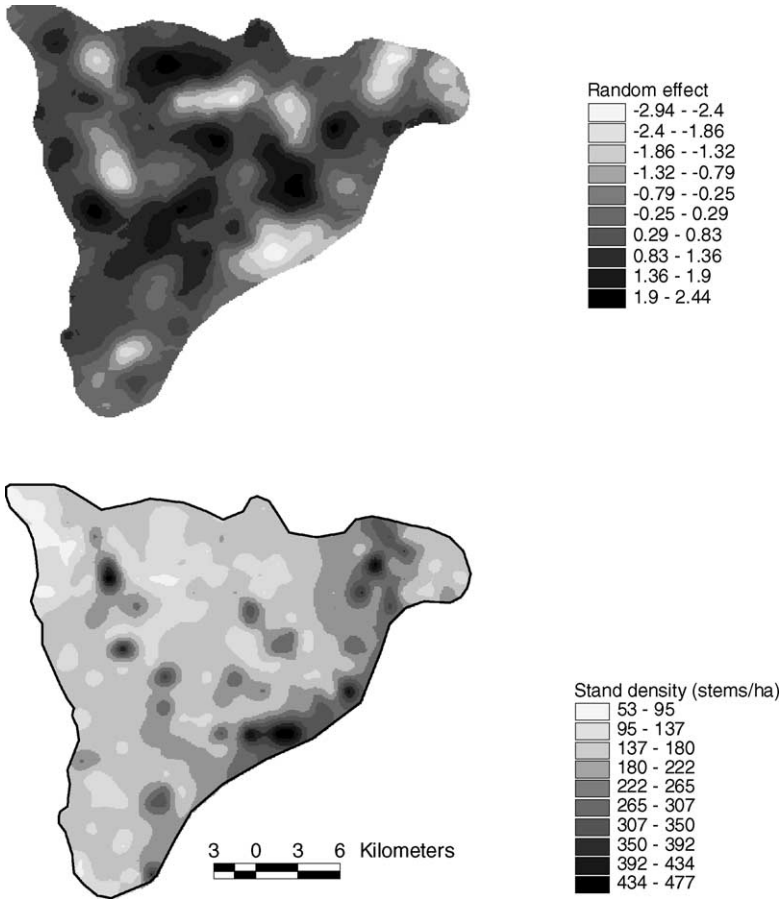

Fig. 4. Kriging maps for the random effect v_i of Model (17) and for the stand density.

inferred independently from one another, but as whole must satisfy certain conditions that guarantee the uniqueness and existence of a solution of the cokriging system. The linear model of coregionalisation (LMC) is used for this reason since it provides sufficient flexibility while being a permissible model. The LMC is a set of direct and cross-variogram models whose coregionalisation matrices are positive definite (Goovaerts, 1997).

Letting $\Gamma(h)$ denote the matrix of all direct and cross-variograms, the LMC is written in matrix form

$$\Gamma(h) = \sum_{k=1}^{s} \boldsymbol{B}_k \gamma_k(h) \tag{27}$$

where $\boldsymbol{B}_k$ is the coregionalisation matrix for the kth structure.

In this model all direct and cross-variograms must be linear combinations of the same basic structures (indexed by k). Its permissibility can be easily checked by simply checking the positive definite condition for each coregionalisation matrix.

For the construction of an LMC between v_i and plot density we assumed that the direct and cross-variograms are isotropic. This assumption can be questionable for the random effect v_i, but it is useful in order to keep the models as simple as possible. Both variables were transformed to zero mean and unit variance prior to the construction of direct and cross-variograms. Using this transformation the cross-variogram is not

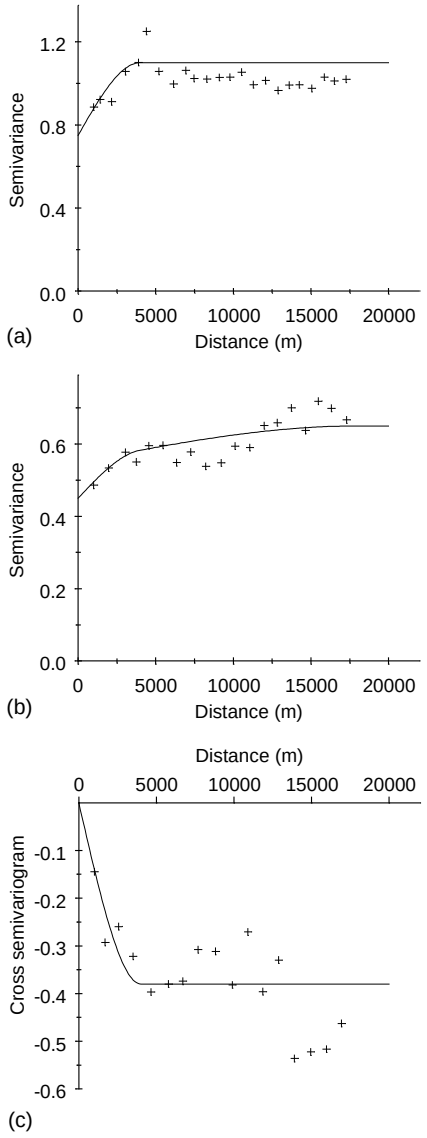

Fig. 5. Direct and cross-variograms for the linear model of coregionalisation between random effect v_i and plot density: (a) direct variogram of the random effect v_i of Model (17), (b) direct variogram for plot density and (c) cross-(semi)variogram between random effect and plot density. Variables were standardised to zero mean and unit variance prior to the variogram computation.

altered by the differences of magnitude between variables.

The LMC is graphically displayed in Fig. 5. All direct or cross-variograms are composed of three basic structures namely a nugget component and two spherical variograms with ranges of 4000 and 18,000 m. The LMC can be written in matrix form

$$\begin{bmatrix} \gamma_{11}(\boldsymbol{h}) & \gamma_{21}(\boldsymbol{h}) \\ \gamma_{12}(\boldsymbol{h}) & \gamma_{22}(\boldsymbol{h}) \end{bmatrix} = \begin{bmatrix} 0.75 & 0 \\ 0 & 0.45 \end{bmatrix} g_0(\boldsymbol{h}) + \begin{bmatrix} 0.35 & -0.38 \\ -0.38 & 0.1 \end{bmatrix} \times \mathrm{SPHER}_{4000}(\boldsymbol{h}) + \begin{bmatrix} 0 & 0 \\ 0 & 0.1 \end{bmatrix} \times \mathrm{SPHER}_{18{,}000}(\boldsymbol{h}) \tag{28}$$

where $\gamma_{11}(\boldsymbol{h})$ and $\gamma_{22}(\boldsymbol{h})$ are the direct variograms for random effect v_i and plot density, respectively, $\gamma_{12}(\boldsymbol{h})$ is their cross-variogram, $g_0(\boldsymbol{h})$ the nugget component, and SPHER the spherical variogram model (the subscript indicates the range of the model). It can be easily checked that all three coregionalisation matrices are positive definite, therefore the above coregionalisation model is an admissible one.

3.3.5.2. Cokriging. Once a permissible model has been constructed the cokriging estimator at an unsampled location $z^*_{\mathrm{CK}}(u)$ is defined as an optimum linear combination of all available data. In the simplest case of two variables ($f = 2$):

$$z^*_{\mathrm{CK}}(u) = \sum_{a_1=1}^{n_1(u)} \lambda_{a_1}(u) z_1(u_{a_1}) + \sum_{a_2=1}^{n_2(u)} \lambda_{a_2}(u) z_2(u_{a_2}) \tag{29}$$

The weights of the primary variable $[\lambda_{a_1}(u)]$ have to sum to unity while for the secondary variable $[\lambda_{a_2}(u)]$ sum to zero. The minimisation of the error variance using the same technique as in ordinary kriging, leads to the ordinary cokriging system of equations, which can be written in matrix form

$$\boldsymbol{C}^{\mathrm{CK}}_{\alpha\beta} \boldsymbol{\lambda}^{\mathrm{CK}} = \boldsymbol{c}^{\mathrm{CK}} \tag{30}$$

where

$$\boldsymbol{C}^{\mathrm{CK}}_{\alpha\beta} = \begin{bmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} & 1_1 & 0_1 \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} & 0_2 & 1_2 \\ 1_1^{\mathrm{T}} & 0_1^{T} & 0 & 0 \\ 0_1^{\mathrm{T}} & 1_1^{T} & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{\lambda}^{\mathrm{CK}} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \mu_1 \\ \mu_2 \end{bmatrix}, \qquad \boldsymbol{c}^{\mathrm{CK}} = \begin{bmatrix} \boldsymbol{c}_{11} \\ \boldsymbol{c}_{21} \\ 1 \\ 0 \end{bmatrix} \tag{31}$$

where $\boldsymbol{C}_{11}(n_1 \times n_1)$ and $\boldsymbol{C}_{22}(n_2 \times n_2)$ are the covariance matrices for the primary and secondary variable, respectively, and $\boldsymbol{C}_{12}$ and $\boldsymbol{C}_{21}(n_1 \times n_2)$ the cross-covariance matrix between primary and secondary variable. Systematic errors have to be filtered out similarly to the ordinary kriging system of equations, by adding the variance of the estimation error to the diagonal terms of the covariance matrices. The vector $\boldsymbol{\lambda}^{\mathrm{CK}}$ contains the unknown weights for primary and secondary data (λ_1, λ_2) and the Lagrange multipliers (μ_1, μ_2) while $\boldsymbol{c}^{\mathrm{CK}}$ is the vector of direct and cross-covariance between sample points and the point where the estimation is required. Vectors $\mathbf{1}$ and $\mathbf{0}$ are the unit and null vectors, respectively, of length n_f.

The left-hand matrix $\boldsymbol{C}^{\mathrm{CK}}_{\alpha\beta}$ of (30) is a symmetric matrix with dimension $(n_1 + n_2 + 1) \times (n_1 + n_2 + 1)$. In our case study with 250 plots for each of the primary and the secondary variable $\boldsymbol{C}^{\mathrm{CK}}_{\alpha\beta}$ is a 501×501 matrix. When dealing with large data sets it is convenient to reduce the dimension of $\boldsymbol{C}^{\mathrm{CK}}_{\alpha\beta}$ for reasons of computational efficiency and stability in the results. For this reason we used a search strategy prior to the computation of the weights $\lambda_{a_1}(u)$ and $\lambda_{a_2}(u)$ by selecting the 20 nearest plots to the location where the estimation was required.

3.4. Spatial prediction of H/D curves

Using the maps of Fig. 4 the prediction of the site-specific *H*/*D* curve can be obtained for all locations within the study area. We have chosen arbitrarily three transects and two plots within the area (Fig. 1) in order to predict and illustrate the *H*/*D* curves.

In Fig. 6 we present the estimated (by kriging) EBLUP for the random effect and stand density along the transects 1–3 (Fig. 1). Additionally, we present a third estimation which corresponds to the estimated (by cokriging) random effect v_i assuming that the stand density at the location where the prediction being made equal to 300 stems/ha.

The first transect, located in the western part of the study area, has a length of 2500 m (Fig. 6a). The stand density varies from 150 to 160 stems/ha and the random effect v_i from 1 to 2 (both predictions were made by kriging). Under the assumption that stand density equals 300 stems/ha the random effect v_i of Model (17) is substantially smaller than the value predicted by kriging.

The second transect has a length of 8000 m and it passes from eight locations that coincide exactly with the spatial locations of the inventory plots. The kriging estimation of stand density and random effect presents some discontinuities when the location to be estimated coincides with data locations (Fig. 6b). This is due to the exactitude property of kriging, although in this case the discontinuities are less pronounced than usual, since part of the nugget variance was interpreted as estimation error. The cokriging prediction for v_i (if stand density is assumed equal to 300 stems/ha) does not always lie below the one given by kriging. Similarly for transect 3 the effect of a higher density on the random effect v_i is positive or negative depending on the location of the prediction (Fig. 6c).

Finally we present in Fig. 7 the predicted *H*/*D* curves for plots 506 and 1202 (Fig. 1) assuming different stand densities. The reduction of stand density to 160 stems/ha has different effects on the predicted curve, being positive for plot 506 and negative for plot 1202.

4. Discussion

4.1. Variogram analysis for the random stand effect and stand density

The range of spatial correlation for v_i was equal to 6000 and 11,000 m for directions of 150° and 60°, respectively. Therefore compartments of low and high values for the random effect are of the same magnitude. The variability of v_i reflects the variability of the relationship between total tree height and DBH and, intuitively, derives mainly from differences in water availability, soil richness (Fulton, 1999) and the time and intensity of thinning (Zhang et al., 1997).

On the other hand, the variogram analysis for the stand density revealed the existence of two nested structures. Stand density is a variable with a hypothetical range of spatial correlation equal to the average size of forest blocks (Mandallaz, 2000). In our study area blocks have a size of 2000 m approximately, therefore the range of the first nested structure of Model (21) is in accordance with this data. The

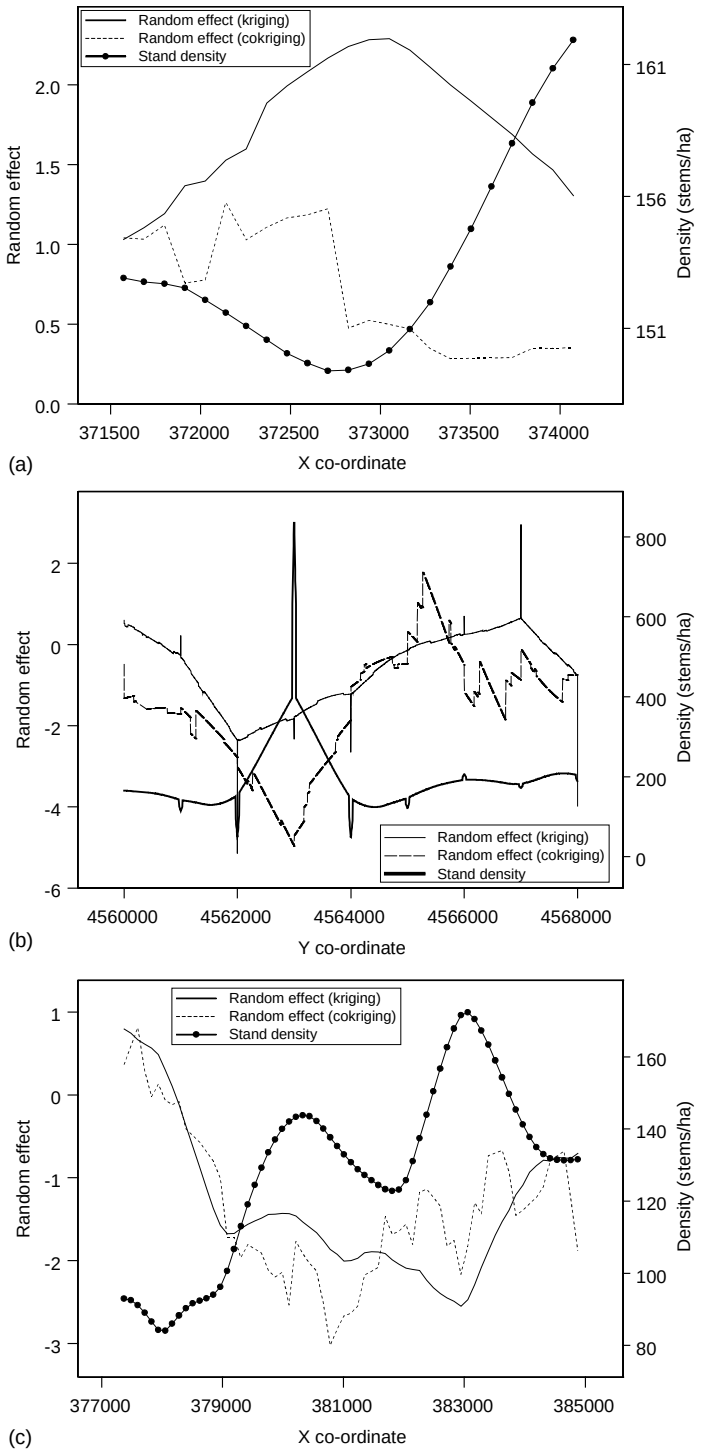

Fig. 6. Prediction of the random effect v_i of Model (17) and of stand density along the three transects of Fig. 1. For the cokriging prediction of v_i it was assumed that the stand density at the location where the prediction was required was equal to 300 stems/ha.

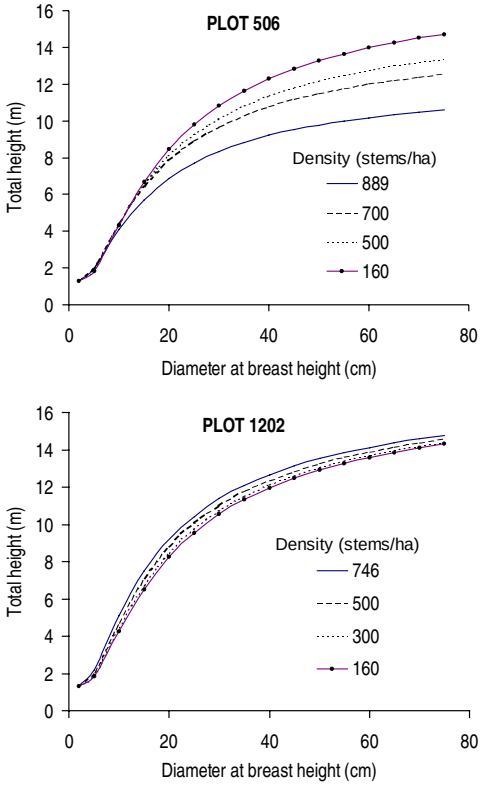

Fig. 7. Predicted *H*/*D* curves at plots 506 and 1202 (see Fig. 1) for different stand densities. The first curve for each plot (corresponding to 889 stems/ha for plot 506 and 746 stems/ha for plot 1202) is the actual stand density.

long-range structure observed in the same variogram is possibly due to differences in silvicultural practices (forest units located far away have different densities).

The large nugget effect in the raw variogram of both the studied variables can be attributed, according to Chilés and Delfiner (1999), to three causes:

- Structures with a range shorter than the smallest inter-point distance (short-range variability).
- Estimation errors of random effects or stand density.
- Micro-structures that is a component with a range shorter than the sampling support. In our case the sampling support refers to the inventory plot, therefore micro-structures should be absent if one is willing to assume that within an inventory plot there could be only one *H*/*D* curve or that the stand density is constant.

4.2. Geostatistical prediction

Frequently parameters of *H*/*D* models vary across stands and when sufficient data are available it is convenient to construct models for them in an attempt to localise the model and give more accurate predictions. *H*/*D* parameters are influenced by a great number of factors, the thinning regime and the site quality being the most important ones (intuitively). The combined action of these and many other (less important) factors causes a large spatial continuity for random stand effects, which in turn can be used for spatial prediction.

H/*D* curves derived with the use of a random coefficients model provide more accurate predictions if random effects can be predicted. To do this, at least one additional height measurement from a tree within the stand is required (Lappi, 1997). Alternatively, prediction of random stand effects can be done with geostatistics without the necessity of additional measurements. This property is particularly interesting when dealing with large inventory areas (national forest inventories, for example) since the only available data are that collected at inventory plots. However, geostatistical predictions are associated with a certain scale of spatial variation and their use should be restricted to the same scale at which data were collected.

The ecoregion-based *H*/*D* models reported by Huang et al. (2000) can predict the *H*/*D* curve at locations where no secondary information is available, but their construction is based on a prior classification of the study area into ecoregions, something that is automatically done when using a geostatistical approach.

The spatial prediction of *H*/*D* curves should be used in conjunction with spatial models for other stand attributes, such as the diameter distribution (Nanos and Montero, 2002) or nontimber products (Nanos et al., 2001), and integrated into a spatial model of forest production. Spatial models can be particularly interesting in our study area since the two main products (resin and timber) are considered to be incompatible (resin extraction diminishes the timber quality); therefore their joint production should be based on a spatially optimised forest plan.

4.3. Prediction by cokriging

Traditionally, geostatistics have been applied in the field of mining and the petroleum industry where the secondary variables included in the cokriging systems cannot be altered by the manager of the resource. However, in forestry and in our case study especially, the inclusion of the stand density as a secondary variable may be needed for the prediction of the primary variable (in this case the random effect v_i) when stand density decreases due to thinning. In that case, cokriging could be used for predicting the primary variable in the future, rather than providing a more reliable spatial estimation.

In that sense, the inclusion of the stand density in a cokriging system does not necessarily require a denser sampling of this attribute, as it is usually reported in the geostatistical literature (see, for example Goovaerts, 1997). In the present study the stand density is measured at the same number of plots as the primary variable (usually reported as the *isotopic case*), but due to the above reasons this should not be interpreted as a disadvantage. Cokriging will probably not improve over kriging, but it can be used for predicting random stand effects at future stages of the stand development.

An interesting future step would be the prediction of *H*/*D* curves in the space–time domain (Houllier and Pierrat, 1992; Stein et al., 1998) when multiple measurements per individual are available. The inclusion of the time dimension into a spatio-temporal prediction system will provide a new geostatistical growth-and-yield modelling methodology. Geostatistical modelling when incorporated into in a forest growth-and-yield model will provide (theoretically) some advantages:

- Geostatistical models are site-specific: future predictions will be made based on nearby observations (nearby both in space and time) therefore no site-index modelling will be necessary.
- The stochastic nature of geostatistics will give a stochastic component to forest modelling: the future projection of two stands located some distance apart and having the same site quality and density does not have to be identical. It will probably be, but this will depend on the nearby stands, their evolution, and dynamics.

5. Conclusions

In this analysis we tried to derive a method for predicting spatially the height/diameter relationship. The proposed modelling strategy combines mixed models and geostatistical methodology. The analysis can be divided into two steps: at the first stage a height/diameter model is fitted to data of each plot, using mixed models and letting some of the model parameters to vary across plots. At the second stage we examined the spatial correlation of random stand effects (variogram analysis) and we used the observed correlation in order to localise the model and predict random stand effects all over the study area (kriging). Finally, we examined the spatial cross-correlation between random stand effects and stand density and we used this correlation in order to predict spatially the random stand effects for different stand densities.

The advantage of the proposed methodology over traditional prediction systems is that it is possible to predict random stand effects of a height/diameter model without additional stand measurements.

Acknowledgements

The authors wish to thank Dr. Ricardo Alía and Dr. Javier Moro, for their comments on the manuscript. This study was partially funded by a grant from Sofia Chlorou foundation and the National Technical University of Athens, Greece.

References

Chilés, J.P., Delfiner, P., 1999. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York, 695 pp.

Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley, New York, 900 pp.

Davidian, M., Giltinan, D.M., 1995. Nonlinear Models for Repeated Measurement Data. Chapman & Hall, London, 359 pp.

Fang, Z., Bailey, R.L., 2001. Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. For. Sci. 47 (3), 287–300.

Fang, Z., Bailey, R.L., Fang, Z.X., 1998. Height–diameter models for tropical forests on Hainan Island in southern China. For. Ecol. Manage. 110, 315–327.

Fulton, M., 1999. Patterns of height–diameter relationships for selected tree species and sites in eastern Texas. Can. J. For. Res. 29, 1445–1448.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, 483 pp.

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.

Hokka, H., 1997. Height–diameter curves with random intercepts and slopes for trees growing on drained peatlands. For. Ecol. Manage. 97, 63–72.

Houllier, F., Pierrat, J.C., 1992. Application des modéles statistiques spatio-temporels aux échantillonnages forestiers successifs. Can. J. For. Res. 22, 1988–1995.

Huang, S., Price, D., Titus, S., 2000. Development of ecoregion-based height–diameter models for white spruce in boreal forests. For. Ecol. Manage. 129, 125–141.

Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York, 561 pp.

Jayaraman, K., Lappi, J., 2001. Estimation of height–diameter curves through multilevel models with special reference to even-aged teak stands. For. Ecol. Manage. 142, 155–162.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London, 600 pp.

Lappi, J., 1991. Calibration of height and volume equations with random parameters. For. Sci. 37, 781–801.

Lappi, J., 1997. A longitudinal analysis of height/diameter curves. For. Sci. 43, 555–570.

Lindstrom, M.J., Bates, D.M., 1990. Nonlinear mixed effects for repeated measures data. Biometrics 46, 673–687.

Mandallaz, D., 2000. Estimation of the spatial covariance in universal kriging: application to forest inventory. Environ. Ecol. Statist. 7, 263–284.

Matheron, G., 1963. Principles of geostatistics. Econ. Geol. 58, 1246–1266.

Nanos, N., Montero, G., 2002. Spatial prediction of diameter distribution models. For. Ecol. Manage. 161, 147–158.

Nanos, N., Tadesse, W., Montero, G., Gil, L., Alía, R., 2001. Spatial stochastic modelling of resin yield from pine stands. Can. J. For. Res. 31, 1140–1147.

Palmer, M.J., Phillips, B.F., Smith, G.T., 1991. Application of nonlinear models with random coefficients to growth data. Biometrics 47, 623–635.

Parresol, B.R., 1992. Baldcypress height–diameter equations and their confidence intervals. Can. J. For. Res. 22, 1429–1434.

Pinheiro, J.C., Bates, D.M., 1995. Approximation to the log-likelihood function in the nonlinear mixed-effects model. J. Comput. Graph. Statist. 4, 12–35.

Pinheiro, J.C., Bates, D.M., 1998. Model building for nonlinear mixed effects model. Technical Report of Department of Statistics, University of Wisconsin, 11 pp.

SAS Institute, 2000. SAS/STAT User's Guide, Version 8. SAS Institute Inc., Cary, NC.

Searle, S.R., 1971. Linear Models. Wiley, New York, 532 pp.

Sheiner, L., Beal, S.L., 1980. Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis–Menten model: routine clinical pharmacokinetic data. J. Pharmacokinet. Biopharmaceut. 8, 553–571.

S-PLUS, 1998. User's Guide. Data Analysis Products Division, MathSoft, Seattle, WA.

Stein, A., Van Groeningen, J.W., Jeger, M.J., Hoosbeek, M.R., 1998. Space–time statistics for environmental and agricultural related phenomena. Environ. Ecol. Statist. 5, 155–172.

Vonesh, E.F., Chinchilli, V.M., 1997. Linear and Nonlinear Models for the Analysis of Repeated Measurements. Marcel Dekker, New York.

Zakrzewski, W.T., Bella, I.E., 1988. Two new height models for volume estimation of lodgepole pine stands. Can. J. For. Res. 18, 195–201.

Zhang, S., Burkhart, H.E., Amateis, R.L., Zhang, S.A., 1997. The influence of thinning on tree height and diameter relationships in loblolly pine plantations. South. J. Appl. For. 21, 199–205.