Available online at www.sciencedirect.com
ScienceDirect
Forest Ecology and Management 255 (2008) 660–673
Forest Ecology and Management
www.elsevier.com/locate/foreco

An empirical ecological-type model for predicting stone pine (*Pinus pinea* L.) cone production in the Northern Plateau (Spain)

Rafael Calama [a,*], Fco. Javier Gordo [b], Sven Mutke [a], Gregorio Montero [a]

[a] *CIFOR-INIA, Ctra. Coruña, km 7.5, 28040 Madrid, Spain*

[b] *Servicio Territorial de Medio Ambiente, Junta de Castilla y León, C/Duque de la Victoria 5, 47001 Valladolid, Spain*

Received 5 March 2007; received in revised form 22 May 2007; accepted 17 September 2007

Abstract

Data from a 10 year series of cone production taken from 755 trees were used to model individual cone production in stone pine (*Pinus pinea* L.) stands in the Northern Inland Plateau of Spain following three different approaches. The first step was the construction of a silvicultural model, including typical forest growth covariates as tree size, stand density, site index and distance independent competition indices. Remaining between-plot variability was related with ecological attributes, as winter rainfall and altitude, resulting in a hybrid model. The third approach attempted to develop an ecological-type based model by considering a previous stratification of stone pine forests based on altitude, soil, geology and climate characteristics. The best model in terms of likelihood, bias and accuracy on predictions was the ecological-type one, producing unbiased marginal estimates for the main part of the territory with an efficiency reaching up to 39%.

Due to the hierarchical structure of data set, proposed model was formulated as a multilevel mixed model. Stochastic formulation allows simulating cone production under different changing scenarios and describing real distribution of cone production within a given stand. Developed model constitutes the cone yield module for PINEA2, an integrated single tree model for the management of stone pine stands within the northern Plateau of Spain.

Keywords: Non-wood forest product; Cone production; Ecological-type model; Stone pine; Mixed model

1. Introduction

The great economical and social importance of non-wood forest products (NWFP) is one of the main characteristics that distinguish Mediterranean forests from other temperate forests. Within Mediterranean NWFP we can differentiate between those which incur a reduction in timber quality or yield, such as resin or cork production, and those which are compatible with timber production, such as mushrooms, fruits or pasture. In the latter case, although the different products are compatible, their simultaneous optimisation does not necessarily. Thus, the joint production in these multifunctional stands must be optimised using specific tools and functions capable of accurately predicting timber and nontimber production under different management regimes.

On the other hand, studies about forest seeding are focused normally rather on few sample trees or seed traps in a functional-ecological approach to recruitment than on spatially exhaustive long-term inventory-like samplings, due to the lack of any market for forest seeds except as reproductive materials (Mencuccini et al., 1995; Castro et al., 1999; Piovesan and Adams, 2001; Gworek et al., 2007). This is not true for one of the main edible fruits found in Mediterranean forests, the pine nut (from the stone pine *Pinus pinea* L.), which is highly appreciated and commonly found in pastries and regional dishes. Spain accounts for more than 60% of the world's stone pine woodland and pine nut production.

Multi-objective management has been employed in the stone pine stands since the end of 19th century, focusing on timber and pine nuts as the main commercial products, whilst also considering other beneficial aspects of the forests such as protection against soil erosion, scenic beauty, biodiversity or recreational use. Understanding cone production may be key to multifunctional management in stone pine stands. Over the last 30 years, several tools have been developed to predict cone production from stone pine stands in different regions of the Mediterranean Basin. Traditional approaches to modelling stone pine cone production have evaluated the inclusion in

* Corresponding author. Tel.: +34 91 3476868; fax: +34 91 3476767.
E-mail address: rcalama@inia.es (R. Calama).

0378-1127/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.foreco.2007.09.079

the models of different tree or stand parameters such as tree size (Magini and Giannini, 1971; Cañadas, 2000; Calama and Montero, 2007), stand density (Castellani, 1989; García-Güemes, 1999; Cañadas, 2000; Piqué, 2003), stand maturity (García-Güemes, 1999) or site index (Cañadas, 2000; Calama, 2004) for predicting the average cone yield for a given period at both stand and tree level. The effect of rainfall and other climatic factors explain much of the between-year variation in yield (Mutke et al., 2005a,b), and various studies have pointed to the existence of a clear pattern of spatial correlation in average cone production (e.g. Nanos et al., 2003). However, little is known about the effect of average climate or other ecological factors on the spatial variability of cone production.

The main objective of this study was to develop and compare three different approaches for modelling single tree cone production in the Northern Plateau, one of the most important areas for the species in Spain. In the first stage, a model which included stand and tree level covariates typically used in forest growth modelling will be evaluated. The second stage involves identifying the possible ecological covariates that could explain the spatial variability in average cone production, in order to derive a hybrid model in which climatic and/or orographic information is added to the previously developed silvicultural model. Finally, as an alternative to using explicit silvicultural and ecological factors, an existing ecological stratification of stone pine forests (Gordo, 2004), will be used for explaining cone production as well as for defining the areas of highest production within the region.

Constructed models would help researchers to have a better knowledge about the effect of ecological factors and management practices on reproductive traits for this species, focusing on seed production, a limiting process in the whole Mediterranean forests. From a management point of view, the obtained model could be used as cone prediction module in the integrated PINEA2 model, currently under development. This model is a single tree, distance independent model, parameterized for different regions in Spain (see Calama et al., 2007 for additional information), and oriented to multi-functional management of stone pine.

2. Materials

2.1. Study area and permanent plots for cone production

Within the Spanish Northern Plateau, stone pine stands cover a wide area of approximately 50,000 hectares, mainly in the province of Valladolid. In fall 1995, a network of 141 sample plots was installed in even-aged stands of stone pine within the public forests of the Northern Plateau (Fig. 1). The network was established for the purposes of data collection in order to develop growth and yield models (both timber and cone) for the species. The plot selection attempted to provide a balanced representation of all the possible age, site quality and density classes identified in the region. The plots are circular, of variable size, and include 20 trees. Breast height diameter, crown diameter, total height and height to crown base were recorded for each tree at plot installation. Trees were positioned with respect to the centre of the plot. Plot age was estimated by averaging single tree age for three to five trees, computed after counting rings in cores taken at stump height using a Pressler borer. A second inventory was carried out in fall 2001.

Between 1996 and 2005, cones from the five trees nearest to the centre of the plot were manually collected every autumn (except in 10 plots where cones were collected from the ten nearest trees). The cones cropped from each tree were classified as *sound* or *damaged* (those attacked by larvae of the moth *Dioryctria mendacella* Stgr (*Lepidoptera*, *Pyralidae*) or the weevil *Pissodes validirostris* Gyll. (*Coleoptera*, *Curculionidae*)). The sound cone crop from each tree was counted and weighed. At the beginning of the trial in 1996, 755 trees were cropped annually. However, ten plots were lost completely as a result of fire or unauthorized loggings and 33 trees in other plots died of natural causes, although in the latter case, the remainder of the trees in each plot continued to form part of the trial. By

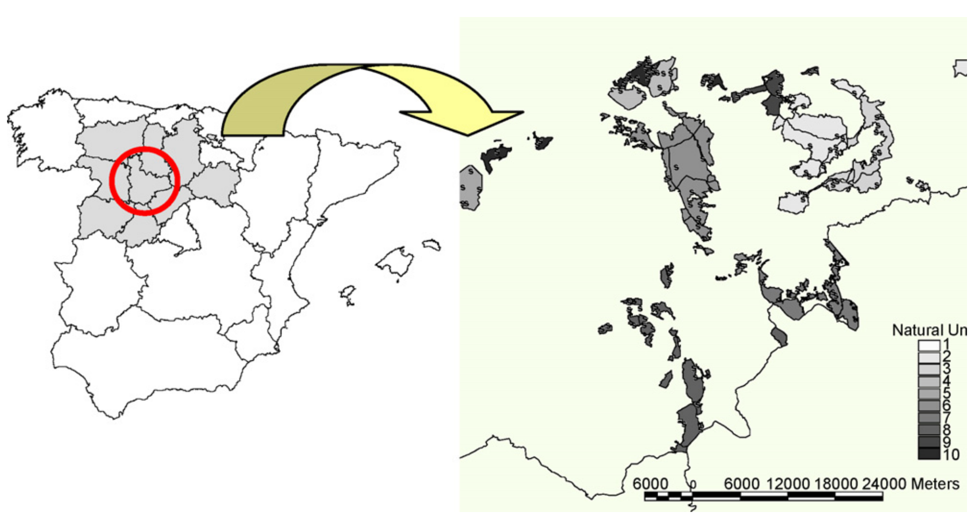

Fig. 1. Study area, natural units classification and plot location (dots).

2005, a total of 672 trees in 131 plots remained. In a few occasions, cones had been stolen from trees before measurements could be taken.

2.2. Average climatic data

Average values for annual as well as seasonal rainfall and temperature were estimated for each plot using the climatic models developed by Sánchez-Palomares et al. (1999). These models were constructed using long-term climatic data series from the nearest weather stations of the National Meteorological Institute (INM). Since these models do not estimate annual climatic values, the average values were used as ecological covariables for each plot.

2.3. Natural unit classification

Forests in the Northern Plateau can be divided into two basic groups: *Limestone plains* and *Sandy areas*, according to geological attributes, such as rock matter or soil genesis. Using this initial division, Gordo et al. (2000) and Gordo (2004) proposed a more refined ecological stratification for the stone pine forests of the region (Fig. 1), based on soil attributes (origin, main texture, water retention), as well as physiographic (altitude and position with respect to the drainage basin) and climatic (rainfall and temperature) factors. Under this basis stone pine forests were thus classified into 10 groups, defined as *Natural units* (Table 1).

3. Methods

3.1. Response variable

The proposed response variable is the weight (in kilograms) of sound cones cropped from a tree in a year, w_{cijk}, defined as the average crop observed in the j_{th} tree located in the i_{th} plot for the k_{th} 5-year period. Since a 10-year cone production series was available (1996–2005), two average values per tree were taken, one for each 5-year series (1996–2000 and 2001–2005). If less than three measurements had been taken from a tree during one of the periods, the average value for that period was not included in the analysis. As a result, the total available values for w_{cijk} were 751 for the first period and 712 for the second, that is, a total of 1463 observations included in the analysis. Table 2 shows cone production summary statistics for each analysed period and unit.

In this study, average values for 5-year periods are analysed rather than annual values because of the large number of zero values detected every year, ranging from 30% to 83% of the total number of annual observations. In other words, 3827 observations from an annual base of 7096 (54.2%) are zeroes. An abundance of zeroes somewhat complicates the statistical analysis, since the distributions will be heavily deviated from normal, with no possible transformation. By averaging the observations, the number of zeroes is 268 out of 1463 observations (18.3%), which is a more manageable quantity. Moreover, this model was developed as a sub-model for cone production within the PINEA2 model, which bases the growth function on diameter increments at 5-year intervals, so model predictions are then carried out for 5-year periods.

Individual cone yield is a non-normal variable, clearly skewed towards low values, indicating that few trees yield a large crop, while most trees produce a smaller than average crop. In order to deal with this non-normal distribution and reduce heteroscedasticity, a logarithmic transformation $\log(w_c + 1)$ was proposed in previous works concerning cone production in this species (Cañadas, 2000; Mutke et al., 2001, 2005a; Calama and Montero, 2007) as well as for fruit production in other species (Masaka and Sato, 2002; Abrahamson and Layne, 2003; Ihalainen et al., 2003). By defining a logarithmic structure together with a linear model, we assume a multiplicative outcome for the different explanatory covariates to be included in the cone production model. When predicting real average cone productions, the expected value for the response variable $\log(w_c + 1)$ should be reverted to the arithmetic scale. Bias in anti-logarithmic transformation is corrected by including the multiplicative factor $\exp[s^2/2]$ (Flewelling and Pienaar, 1981), where s^2 is the total prediction variance.

Table 1
Main characteristics for natural units

Unit	Name	Total area (ha)	Prod. area (ha)	Altitude (m)	Annual rainfall (mm)	SI (m)	Geology and texture	Soil origin	Water retention (mm)	No. of plots
1	Torozos	1834	195	847	535	14.0	Limestone-Marl	Tertiary sediments	N.A.	2
2	Limestone plain W	4757	2338	845	485	14.9	Limestone-Marl	Tertiary sediments	200	17
3	Limestone plain E	2284	216	854	488	14.4	Limestone-Marl	Tertiary sediments	200	19
4	Valladolid	1908	1675	687	374	13.1	Quartz sands	Wind deposits	N.A.	17
5	Nava del Rey	1480	1409	710	370	11.8	Quartzitic gravels	Wind deposits	100	6
6	Viana de Cega	7211	4573	707	369	15.1	Quartz sands	Wind deposits	80–150	12
7	Iscar	2945	1057	745	380	16.7	Quartz sands-Clays	Alluvial-wind	>300	19
8	Medina	4044	1867	746	371	12.5	Quartz sands	Wind deposits	75–130	17
9	Tudela de Duero	1401	1085	712	392	12.6	Quartz sands	Wind deposits	100	10
10	River terraces	1664	1544	688	374	16.8	Alluvial meadow soils	Alluvial deposits	150–300	22

Where SI is site index, as stated by Calama et al. (2003); N.A. no available data.

Table 2
Summary statistics for cone production

Unit	Name	w_c (kg/tree)										
		1996–2000					2001–2005					1996–2005
		Min	Mean	Max	Zero	n	Min	Mean	Max	Zero	n	Mean
1	Torozos	1.500	7.195	16.220	0	10	1.787	3.150	5.515	0	10	5.172
2	Limestone plain W	0	5.337	38.980	4	85	0	3.370	35.070	9	79	4.390
3	Limestone plain E	0	1.036	8.820	12	85	0	0.701	3.200	11	85	0.869
4	Valladolid	0	0.981	6.180	28	83	0	0.217	2.800	41	79	0.609
5	Nava del Rey	0	0.308	2.980	14	30	0	0.101	0.420	10	30	0.204
6	Viana de Cega	0	1.745	13.060	15	58	0	0.571	6.970	8	48	1.214
7	Iscar	0	6.506	66.500	6	145	0	2.155	26.250	17	145	4.331
8	Medina	0	0.939	5.760	21	95	0	0.332	3.920	29	83	0.656
9	Tudela de Duero	0	1.392	7.150	3	50	0	0.197	2.100	14	49	0.801
10	River terraces	0	6.461	29.910	5	110	0	1.891	18.100	21	104	4.241

Where zero represents number of trees with average production zero; n total number of trees sampled.

3.2. Multilevel linear mixed models

Independence among observations, together with normality and homoscedasticity, is one of the basic assumptions required for ordinary statistical analysis. Due to the hierarchical structure of the data, four different levels of correlation can be identified, since an above average similarity exists between observations coming from the same tree, plot, period and plot × period. This hierarchical nested structure prevents us from using statistical methods based on ordinary least squares regression (West et al., 1984; Fox et al., 2001). Previous studies have dealt with this situation through the use of multilevel mixed models (e.g. Biging, 1985; Lappi, 1986; Gregoire, 1987). The proposed model for cone production is:

$$\begin{aligned} y_{ijk} &= \log(w_{c\,ijk}+1) = \mathbf{x}_{ijk}\boldsymbol{\beta} + u_i + v_{ij} + z_k + s_{ik} + e_{ijk} \\ &= \mathbf{x}_{ijk}\boldsymbol{\beta} + \mathbf{z}_{ijk}\mathbf{b}_{ijk} + e_{ijk} \end{aligned} \quad (1)$$

Basic multilevel mixed formulation splits the model into a fixed component ($\mathbf{x_{ijk}\beta}$), and random components acting at plot (u_i), tree within plot (v_{ij}), period (z_k) and plot × period (s_{ik}) levels; $\mathbf{x_{ijk}}$ is the design vector including explanatory covariates for $\log(w_{c\,ijk}+1)$; $\boldsymbol{\beta}$ is the vector for fixed parameters common for the whole population; u_i, v_{ij}, z_k and s_{ik} are assumed to be distributed under a normal distribution with mean zero and variance components σ_u^2, σ_v^2, σ_z^2 and σ_s^2, respectively; e_{ijk} is a residual (tree × period) error term, distributed under a normal distribution with mean zero and variance component σ_e^2. The main aim of fitting a multilevel mixed model is to obtain unbiased estimates for $\boldsymbol{\beta}$, σ_u^2, σ_v^2, σ_z^2, σ_s^2 and σ_e^2, and to predict the vector $\mathbf{b_{ijk}}$ including the EBLUPs (Empirical Best Linear Unbiased Predictors) for the random components u_i, v_j, z_k and s_{ik} specific for each sampling unit. A multilevel mixed structure allows predictions to be carried out in two different ways:

If the EBLUP's for random components are neither known nor predictable, their expected value will be zero, resulting in Marginal predictions. In this case, the expected mean value and variance prediction for a single observation y_{ijk} are given by:

$$E(y_{ijk}) = \mathbf{x_{ijk}}\boldsymbol{\beta}, \qquad \mathrm{Var}(y_{ijk}) = \sigma_u^2 + \sigma_v^2 + \sigma_z^2 + \sigma_s^2 + \sigma_e^2 \quad (2)$$

An alternative is to make predictions for the EBLUP's of the components of $\mathbf{b_{ijk}}$. In this case, the EBLUP's are directly estimated for the sampled units included in the fitting data set. For new locations, predictions can be carried out if a sample of additional response variable measurements is available, as described by Vonesh and Chinchilli (1997) or Fang and Bailey (2001). Model calibration using a sampling unit distinct from the fitting data set has frequently been applied in forest modelling (Lappi, 1991; Calama and Montero, 2004; Mehtätalo, 2004; Trincado and Burkhart, 2006). However, calibrating average cone production at tree or plot level would require observations for the cone production of a sample of trees in the subject plot over a past 5-year period.

Another possibility, proposed by Nanos et al. (2004), is to analyse the spatial or temporal correlation among the predicted EBLUP's for the sampled units, then make new predictions for un-sampled locations or periods by using geostatistics or time series analysis. Finally, as the distribution for random parameters is known, it is possible to carry Monte Carlo simulations, assigning random realization for the distribution of each unit (e.g. as stated by Sánchez-González et al., 2007). Whichever the case, whether the EBLUP's for $\mathbf{b_{ijk}}$ are known, predicted or simulated, *Calibrated conditional predictions* would be carried out in which the expected mean value and variance prediction for y_{ijk} are:

$$E(y_{ijk}|\mathbf{b_{ijk}}) = \mathbf{x_{ijk}}\boldsymbol{\beta} + \mathbf{z_{ijk}}\mathbf{b_{ijk}}, \qquad \mathrm{Var}(y_{ijk}|\mathbf{b_{ijk}}) = \sigma_e^2 \quad (3)$$

Intermediate alternatives consider including only a part of the random effects into vector $\mathbf{b_{ijk}}$. In this case, total prediction variance would be computed by adding the variance components for the non-considered random components to σ_e^2.

3.3. Model construction

3.3.1. Basic model

In a first stage of model construction we considered a basic structure where the only fixed component entering the model was the intercept, expressing average population value for $\log(w_c + 1)$. In this case, the EBLUP's for random components represent deviation from average population associated with each level of variability (plot, tree, period, plot $\times$ period, residual). The adequacy of the variance–covariance structure was assessed for this basic model by contrasting it against simpler models (in which some random components were removed) using likelihood ratio tests. Homoscedasticity in variance component σ_e^2 was also evaluated by plotting the variance for the empirical residual terms against different explanatory covariates.

In subsequent phases, the inclusion of explanatory covariates was evaluated to explain systematic variability in the mean response for the different levels of associated dependence. From the basic model, three different models were constructed: a *silvicultural model*, including stand and tree level covariates typically measured in or derived directly from forest inventories; a *hybrid model*, where the inclusion of climatic and orographic covariates was evaluated; and an *ecological-type model*, in which a categorical covariate defining an ecological stratification of the territory was included.

3.3.2. Silvicultural model construction

Firstly, classical silvicultural covariates, classified into five types, were tested for inclusion in the model:

- *Tree size*: breast height section g (m^2), breast height diameter d (cm), total height h (m), crown ratio cr (dimensionless). Crown width was available, but it was not evaluated since this covariate is rarely measured in forest inventories.
- *Stand density attributes*: density N (stems/ha), basal area BA (m^2/ha), stand density index SDI (dimensionless)
- *Stand maturity attributes*: age T (years), quadratic mean diameter dg (cm), dominant height Ho (m)
- *Distance-independent competition indices*: basal area for the trees larger than subject tree BAL (m^2/ha), ratio d/dg (dimensionless), ratio g/BA (dimensionless)
- *Site productivity*: site index SI (m), defined as the dominant height at index age 100, and calculated using the model by Calama et al. (2003).

A preliminary selection of covariates was effected by computing Pearson's correlation coefficient between the evaluated covariates (as well as their logarithmic transformations) and the EBLUP's for tree (tree size and competition covariates) and plot (stand density, maturity and productivity attributes) random components predicted after fitting the basic model. Pre-selected covariates were included in the model using a sequential procedure; evaluating in the first place the inclusion of tree size covariates, then stand density, competition indices, stand maturity attributes, and finally, site index. The best combination of covariates at each step was selected in terms of significant −2LL decrease (likelihood ratio test), level of significance for the fixed parameters, reduction in the value of the associated variance components, and maintenance of biological sense (covariates should enter the model with the same sign identified when testing partial correlations). Due to co-linearity among covariates within the same group, no more than two covariates per group were allowed to enter the model at each sequential phase.

3.3.3. Hybrid and ecological-type models

The silvicultural model was used as a basis to evaluate the suitability of including other climatic (annual and seasonal rainfall, average annual temperature) and orographic (altitude) factors in the construction of a hybrid model for explaining cone production. As most of the forests within the area are located on flat land, no other topographical factors such as aspect or exposition were evaluated. The sequential procedure and definite criteria for testing the inclusion of these variables were the same as those described previously. Finally, the adequacy of incorporating the ecologically based stratification in the model instead of other ecological attributes was also evaluated.

3.4. Model evaluation

Silvicultural, hybrid and ecological-type models were compared in terms of mean error (ME), associated level of significance, root mean squared error (RMSE) and modelling efficiency (EF) for conditional and marginal residuals, computed for the whole population.

$$\mathrm{ME}(y) = \frac{\sum_{ijk}(y_{ijk} - \hat{y}_{ijk})}{n},$$

$$\mathrm{RMSE}(y) = \sqrt{\frac{\sum_{ijk}(y_{ijk} - \hat{y}_{ijk})^2}{n-1}},$$

$$\mathrm{EF}(y) = 1 - \left[\frac{\sum_{ijk}(y_{ijk} - \hat{y}_{ijk})^2}{\sum_{ijk}(y_{ijk} - \bar{y})^2}\right]$$

where n represents the total number of observations; y, $\hat{y}$ and $\bar{y}$ represent observed, predicted and mean value for the response variable. The accuracy of the predictions was also evaluated by comparing the expected mean value and variance for conditional and marginal predictions with the real mean value and variance taken from the whole data set. All the statistics were computed for both logarithmic transformed and real anti-transformed scales. Once the best model had been selected, bias and possible misspecification in covariate selection was evaluated by plotting the mean value and standard error for the conditional residuals (i.e. e_{ijk}) as a function of the predicted and the predictor variables included in the model, both in transformed logarithmic as well as real (original) scale. Additionally, to detect any trend in bias associated with the different geographical areas, ME, RMSE and prediction accuracy were also contrasted separately for each *natural unit*.

In the absence of a similar experimental design with the same range of conditions containing complete data series for cone production at tree and plot level, an independent validation for

Table 3
Sequential procedure and fitting statistics for model construction

Modelling step	1	2	3	4	5	6	7	8	9
Intercept	μ	μ	μ	μ	μ	μ	μ	μ	μ
Tree size	–	–	g	g	g	g	g	g	g
Stand density	–	–	–	log(N)	log(N)	log(N)	log(N)	log(N)	log(N)
Competition	–	–	–	–	d/dg	d/dg	d/dg	d/dg	d/dg
Stand maturity	–	–	–	–	–	Ho	–	–	–
Site productivity	–	–	–	–	–	–	SI	SI	–
Climate	–	–	–	–	–	–	–	Wr	–
Orographic	–	–	–	–	–	–	–	Alt	–
Natural unit	–	–	–	–	–	–	–	–	UN
−2LL	1844.5	1730.6	1475.6	1470.4	1446.8	1438.6	1434.9	1412.0	1365.9
$p <$ LRT	–	–	<0.0001	0.0225	<0.0001	0.0041	0.0005	<0.0001	<0.0001
Contrast model	–	–	2	3	4	5	5	7	5
σ_u^2 (plot)	0.2646	0.2567	0.0640	0.0594	0.0632	0.0604	0.0502	0.02611	0
σ_s^2 (plot × period)	0.1901	0.1829	0.1865	0.1854	0.1841	0.1800	0.1839	0.1832	0.1720
σ_z^2 (period)	0.0542	0.0491	0.0634	0.0628	0.0590	0.0628	0.0585	0.0586	0.0567
σ_v^2 (tree)	0.09604	0.0724	0.0405	0.0412	0.0392	0.0403	0.0394	0.0389	0.0393
σ_e^2 (residual)	0.0573	1.0000	1.0901	1.0829	1.0732	1.0597	1.0698	1.0764	1.0723
Variance function	–	(4)	(4)	(4)	(4)	(4)	(4)	(4)	(4)

Where −2LL: −two times log-likelihood; $p <$ LRT: probability for contrast model to explain better than subject model. Rest of symbols as in the text.

the proposed model was carried out in 31 variable-size plots of 15 trees located in natural unit no. 5 (Nava del Rey forest). In these plots each year between 2001 and 2005, healthy cones were collected, counted and weighted jointly for each plot, so accuracy of the model was evaluated at plot level. Single tree diameter was also measured for all the trees, but neither plot age nor individual tree height were recorded.

4. Results

4.1. Model construction

4.1.1. Basic model

The first column in Table 3 shows the result after fitting the basic model with the intercept as the only fixed parameter and a complete variance structure (plot, period, plot × period and tree random components). Simpler variance–covariance structures were evaluated in terms of likelihood ratio tests, but none of them explained variability in cone production more effectively. Conditional residuals e_{ijk} for the basic model were analysed to identify possible deviations from basic inference assumptions. A clear pattern of heteroscedasticity was detected when plotting the variance of the residuals against different classes of breast height diameter (Fig. 2). Therefore, the following variance function was proposed:

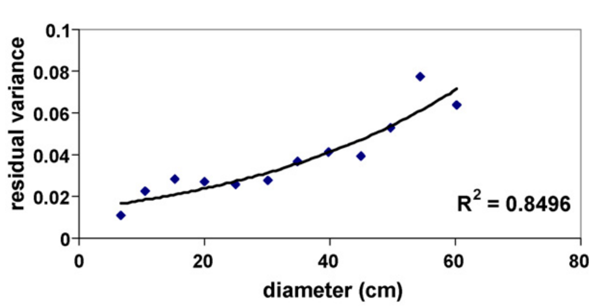

Fig. 2. Residual variance for basic model (dots) and fitted variance function (solid line).

$$\sigma_e^2 = 0.0138\exp^{(0.0274\,\mathrm{d})} \quad (4)$$

The results after fitting the basic model under residual variance structure (4) are shown in column 2, Table 3. A larger pattern of systematic unexplained variability in cone production is associated with between-plot variance (42% of unexplained variance), followed by plot × period (30%) and tree within plot (12%) variance, while period and residual (tree × period) components each explain approximately 8%.

4.1.2. Silvicultural model

Table 4 shows the correlation among the EBLUP's and the different tree and stand level covariates evaluated. Those

Table 4
Correlation between EBLUP's for basic model and evaluated covariates

Level of variabilty	Covariate type	Covariate	r	p-Value
Stand level	Stand density	log(N)	−0.5727	<0.0001
		BA	0.1777	0.0350
		SDI	0.1221	0.1646
	Stand maturity	Age	0.3990	<0.0001
		dg	0.6955	<0.0001
		Ho	0.6268	<0.0001
	Stand productivity	SI	0.2809	0.0007
Tree level	Tree size	d	0.2230	<0.0001
		g	0.2160	<0.0001
		h	0.1345	<0.0001
		Crown ratio	0.0074	0.7740
	Competition	d/dg	0.4335	<0.0001
		g/BA	0.1744	<0.0001
		BAL	0.0421	0.1283

Symbols as stated in text.

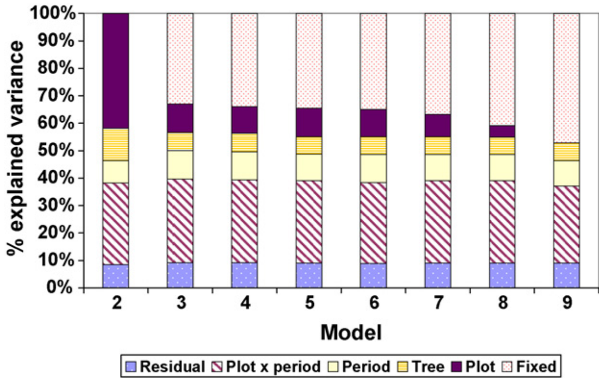

Fig. 3. Percentage of variability explained by the fixed part of the model, random components and residual unexplained, for the sequential phases of model construction (number of modeling step as in Table 3).

variables largely correlated at plot level were: quadratic mean diameter dg, dominant height Ho and the natural logarithm for the number of stems per hectare log(N). The tree level variables most correlated with tree EBLUP's were ratio d/dg and tree section g.

Table 3, columns 3–7 show the sequential process for constructing the silvicultural model. Only the best model fit identified at each phase of covariate type inclusion is presented. The likelihood ratio tests indicate significant improvement when comparing the fit of the model at each phase with the best model for prior phases. The best tree size variable entering the model was tree section g, the inclusion of which substantially reduced both tree and plot level variability. In subsequent steps, log(N), d/dg and SI were included in the model. The stand maturity attribute included was dominant height Ho, although this attribute became nonsignificant when site index was proposed for inclusion in the subsequent step. The combined presence of site index and other stand maturity attributes (quadratic mean diameter or age) was evaluated but no improvement to the model was detected. It was therefore decided that none of the stand maturity covariates would be included.

Fig. 3 shows the variation in the percentage of variability explained by the fixed part of the model and the random levels of variability throughout the process of model construction. The inclusion of stand and tree level covariates explains 37% of initial variability, reducing plot-level variability to 9% and tree level variability to 6%. Period, plot × period and residual (tree × period) unexplained variability remain almost constant, indicating that the covariates included do not explain systematic variability at those levels. The final expression for the *silvicultural model*, once covariate structure is selected and fitted using restricted maximum likelihood (REML) methods, is given by:

$$\begin{aligned}\log(w_c+1) =\,&0.7408+4.7407\,\mathrm{g}+0.5081\,\mathrm{d/dg}\\&-0.2611\log(N)+0.0350\,\mathrm{SI}\\&+u+v+z+s+e\end{aligned} \tag{5}$$

All fixed and random parameters in the model are significant at a α-level = 0.05; u, v, z and s are plot, tree, period and plot × period random components, with univariate normal distribution, mean zero and variances of $\sigma_u^2 = 0.0520$, $\sigma_v^2 = 0.0377$, $\sigma_z^2 = 0.1168$, and $\sigma_s^2 = 0.1827$, respectively. The residual error e has a univariate normal distribution with mean zero and heteroscedastic variance defined by:

$$\sigma_e^2 = 0.0089\exp(0.0606\,d) \tag{6}$$

4.1.3. Hybrid model

EBLUP's for random plot components predicted after fitting the silvicultural model (5) were used to identify possible climate and orography covariates which might explain the remaining between-plot variance. Table 5 shows the correlation between-plot level EBLUP's and these covariates. Significantly associated covariates were evaluated for inclusion in the silvicultural model. Column 8, Table 3 shows the results for the best *hybrid model*, including winter rainfall wr (average value for long-term data series, in mm per year) as a positively associated covariate and stand altitude alt (meters a.s.l.) as negatively associated. The inclusion in the model of both covariates significantly improved it, reducing between-plot variability to 5% of the original unexplained variance. The final expression for the *hybrid model*, after the REML fit is:

$$\begin{aligned}\log(w_c+1) =\,&1.2745+4.9892\,\mathrm{g}+0.4821\,d/\mathrm{dg}\\&-0.2636\log(N)+0.0357\,\mathrm{SI}+0.0177\,\mathrm{wr}\\&-0.0034\,\mathrm{alt}+u+v+z+s+e\end{aligned} \tag{7}$$

With all fixed and random parameters significant at α-level = 0.05; u, v, z and s as previously defined, but in this case with variances of $\sigma_u^2 = 0.0296$, $\sigma_v^2 = 0.0377$, $\sigma_z^2 = 0.1174$, and

Table 5
Correlation between EBLUP's for silvicultural model and evaluated ecological covariates

Level of variability	Covariate type	Covariate	r	p-Value
Stand level	Climate and orography	Altitude	0.1966	0.0194
		Winter rainfall	0.3116	0.0002
		Spring rainfall	0.2776	0.0009
		Summer rainfall	0.2386	0.0044
		Autumn rainfall	0.2946	0.0004
		Total rainfall	0.2925	0.0004
		Average temperature	−0.2428	0.0037

$\sigma_s^2 = 0.1824$; the residual error e now has a univariate normal distribution with mean zero and heterocedastic variance:

$$\sigma_e^2 = 0.0094\,\exp^{(0.0590\,d)} \tag{8}$$

4.1.4. Ecological-type model

Natural units were defined based on average climate covariates, geological factors and soil attributes. Therefore, we hypothesized on the possibility that categorical classification in natural units could explain variability in cone production, complementing or even substituting site index and/or ecological covariates in the model. When *natural units* were included in the model, the parameters for site index, winter rainfall and altitude became nonsignificant, indicating that natural units jointly express the effects of climate, soil and orographic factors on cone production. Column 9, Table 3 shows how the inclusion of natural unit stratification in the model leads to a significant improvement in comparison to previous silvicultural and hybrid models, reducing the between-plot variability down to a null value. The final expression for the *Ecological-type model* (after REML fit) is:

$$\log(w_c + 1) = 1.4796 + 4.2383\,g + 0.5539\,d/\mathrm{dg} - 0.2320\log(N) + \mathrm{UN} + v + z + s + e \tag{9}$$

Where natural unit UN is a categorical fixed variable (value given in Table 6). All the parameters in the model are significant; v, z and s are as previously defined but with variances $\sigma_v^2 = 0.0392$, $\sigma_z^2 = 0.1148$, and $\sigma_s^2 = 0.1792$; e is a residual error term which has a univariate normal distribution, mean zero and heterocedastic variance of:

$$\sigma_e^2 = 0.0102\,\exp^{(0.0561\,d)} \tag{10}$$

4.2. Model evaluation

After sequential procedure for model construction, it was found that the best model in terms of likelihood and level of original variability explained by fixed effects was the ecological-type model. Table 7 shows a comparison of the models in terms of modelling efficiency, mean error, root mean square error and the average and variance of the observed and predicted values. As can be seen, the logarithmic conditional and marginal predictions for the three models are largely unbiased (p-values ranging between 0.4208 and 0.8907). The real, anti-transformed conditional predictions are unbiased, while real-scale marginal predictions show a slight downward bias (p-values 0.0186 for the hybrid model, 0.0245 for the ecological-type model and 0.0565 for the silvicultural model), indicating a certain limitation in bias correction. The best model in terms of modelling efficiency EF is the ecological-type model, although for conditional predictions, the EF for all three models reaches values of up to 92% (log-scale) and 87% (real-scale), indicating that only a small part of total variability is associated with the residual tree × period effect. With respect to marginal predictions, which express the approximate accuracy of the model for practical use, EF values for the ecological-type model reach values of 45.5% for log-scale and 38.8% for real scale, surpassing the values obtained for the rest of models. The hybrid model performs slightly better than the silvicultural model, especially for log-scale marginal predictions. If the RMSE are compared, the best results are again attained by the ecological-type model, which shows the most

Table 6
Natural units EBLUP value

Unit	Name	UN value
1	Torozos	0.4457
2	Limestone plain W	0
3	Limestone plain E	−0.4620
4	Valladolid	−0.6850
5	Nava del Rey	−0.6287
6	Viana de Cega	−0.5579
7	Iscar	−0.3241
8	Medina	−0.6903
9	Tudela de Duero	−0.5907
10	River terraces	−0.0283

Table 7
Evaluation statistics for silvicultural, hybrid and ecological-type model (whole population)

Prediction type	Scale	Statistic	Silvicultural	Hybrid	Ecological-type
Conditional	Log-scale	ME	−0.0012	−0.0012	−0.0008
		p-Value	0.8405	0.8419	0.8907
		RMSE	0.2358	0.2358	0.2313
		Avge Pred	0.7665	0.7675	0.7671
		Avge Obs	0.7663		
		Var Pred	0.5779	0.579	0.5776
		Var Obs	0.7012		
		EF (%)	92.07	92.07	92.37
	Real-scale	ME	0.0423	0.0440	0.0805
		p-Value	0.3799	0.3551	0.0701
		RMSE	1.8429	1.8442	1.6913
		Avge Pred	2.3979	2.3957	2.3602
		Avge Obs	2.4409		
		Var Pred	25.800	25.653	23.027
		Var Obs	26.157		
		EF (%)	87.02	87.00	89.06
Marginal	Log-scale	ME	0.004	0.0136	0.0069
		p-Value	0.8159	0.4208	0.6689
		RMSE	0.6648	0.6485	0.6181
		Avge Pred	0.7622	0.7526	0.7594
		Avge Obs	0.7663		
		Var Pred	0.2513	0.2646	0.3116
		Var Obs	0.7012		
		EF (%)	36.98	40.03	45.52
	Real-scale	ME	0.202	0.249	0.2352
		p-Value	0.0565	0.0186	0.0245
		RMSE	4.0530	4.0511	4.0024
		Avge Pred	2.2383	2.1912	2.2050
		Avge Obs	2.4409		
		Var Pred	7.764	7.688	7.225
		Var Obs	26.157		
		EF (%)	37.20	37.26	38.76

Where Avge Pred and Avge Obs represents mean predicted and observed values. Var pred and Var Obs represents variance for predicted and observed values. Rest of symbols as in text.

consistent reduction for log-scale predictions. In marginal real-scale predictions, all three models reach a RMSE close to 4 kg/tree.

The differences between conditional and marginal predictions with respect to average and variance values are small for all three models. Common to the three models is the large difference between observed and predicted variance detected between marginal and conditional predictions. For example, the marginal predictions for the silvicultural, hybrid and ecological-type models give mean values of 2.2383, 2.1912 and 2.2050 kg/tree, with prediction variances of 7.764, 7.688 and 7.225, while the real observations show a mean value of 2.4409 kg/tree and a variance of 26.157. Marginal predictions show a slight downward bias, but the variance is clearly underestimated. This means that cone yield predictions per hectare would be close to the actual figures, but inter-individual variability for the predicted values would be smaller than the real variance.

After considering likelihood, percentage of original variability explained, bias behaviour and predictive ability, the ecological-type model (9) was selected as the best for explaining cone production. No noticeable trends were detected when plotting conditional residuals (e) for this model against the explanatory stand and tree level covariates evaluated (Fig. 4), the mean error being nonsignificant for the main range of explanatory covariate values. When plotting the residuals against predicted values, significant underestimation was detected for those trees with the greatest production, indicating the failure of the model to accurately predict cone production in these cases. When the selected model (9) is analysed separately for each natural unit (Table 8), unbiased estimates are obtained for all the natural units as well as for each prediction type

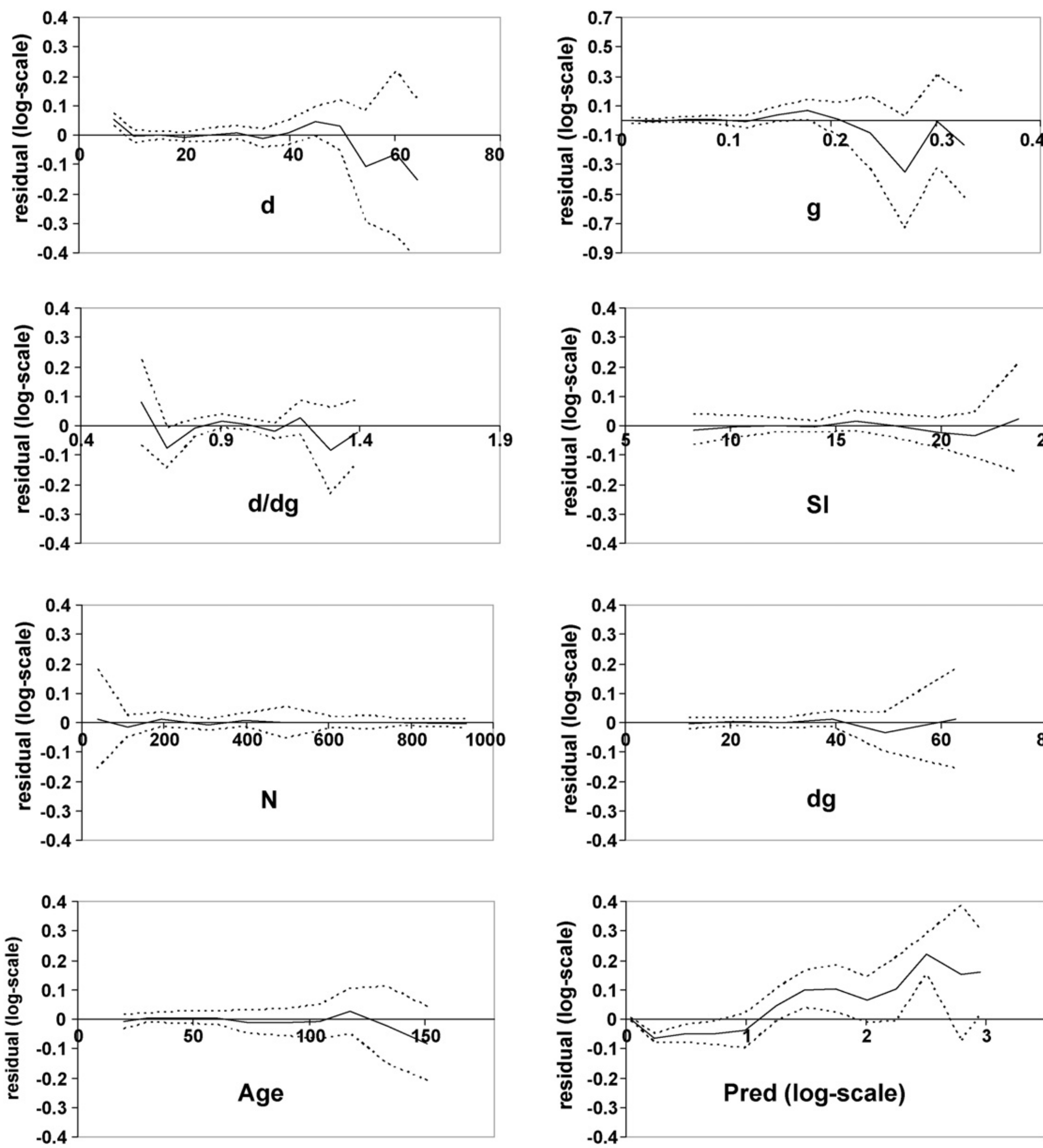

Fig. 4. Mean conditional residual in log-scale (solid line) as a function of predicted value and explanatory stand and tree level covariates. Dotted lines indicate standard error for the mean.

Table 8
Evaluation statistics for ecological-type model

Prediction type	Scale	Statistic	Natural Unit UN 1	2	3	4	5	6	7	8	9	10
Conditional	Log-scale	ME	−0.0071	0.0071	0.0042	−0.0043	−0.0012	0.0097	0.0061	−0.0117	−0.0062	−0.0113
		p-Value	0.8701	0.7393	0.6828	0.7303	0.9209	0.6288	0.7227	0.4177	0.7590	0.5759
		RMSE	0.1927	0.2723	0.1378	0.1602	0.0912	0.2065	0.2911	0.1878	0.1997	0.2959
		Avge Pred	1.6756	1.1934	0.4731	0.3225	0.1477	0.5157	1.1092	0.3813	0.4373	1.2298
		Avge Obs	1.6685	1.2005	0.4773	0.3182	0.1465	0.5254	1.1152	0.3695	0.4311	1.2185
	Real-scale	ME	−0.0189	0.2102	0.0362	0.0087	−0.0110	0.1126	0.0740	−0.0090	0.0246	0.1936
		p-Value	0.9494	0.2087	0.3205	0.7964	0.6696	0.2162	0.5915	0.7966	0.7018	0.2961
		RMSE	1.3186	2.1430	0.4888	0.4299	0.1998	0.9382	2.3475	0.4543	0.6385	2.7114
		Avge Pred	5.1913	4.1797	0.8319	0.6003	0.2154	1.1012	4.2567	0.6533	0.7762	4.0470
		Avge Obs	5.1723	4.3899	0.8681	0.6090	0.2043	1.2137	4.3307	0.6442	0.8008	4.2407
Marginal	Log-scale	ME	−0.0071	0.0173	0.0176	−0.0094	−0.0012	0.0314	0.0310	−0.0157	−0.0053	−0.0155
		p-Value	0.9510	0.7595	0.5574	0.7966	0.9659	0.4742	0.4584	0.6731	0.9104	0.8019
		RMSE	0.5125	0.7234	0.4029	0.4642	0.2118	0.4518	0.7114	0.4808	0.4716	0.9036
		Avge Pred	1.6756	1.1832	0.4597	0.3276	0.1477	0.4939	1.0843	0.3852	0.4364	1.2340
		Avge Obs	1.6685	1.2005	0.4773	0.3182	0.1465	0.5254	1.1152	0.3695	0.4311	1.2185
	Real-scale	ME	−0.7011	0.5389	−0.1312	−0.1103	−0.2048	0.0227	0.9563	−0.2651	−0.1481	0.4817
		p-Value	0.4185	0.2039	0.1081	0.2087	0.0002	0.8918	0.0068	0.0030	0.2527	0.2050
		RMSE	3.8583	5.4366	1.0976	1.1180	0.4448	1.7177	6.0552	1.1723	1.2893	5.5633
		Avge Pred	5.8734	3.8511	0.9993	0.7194	0.4091	1.1910	3.3744	0.9093	0.9489	3.7590
		Avge Obs	5.1723	4.3899	0.8681	0.6090	0.2043	1.2137	4.3307	0.6442	0.8008	4.2407

Where Avge Pred and Avge Obs represents mean predicted and observed values. Rest of symbols as in text.

(conditional or marginal) or scale (log or real), except in the case of marginal real-scale estimates for units 5, 7 and 8. The RMSE for real-scale marginal predictions ranges from 0.4448 kg/tree for the least productive unit (no. 5, Nava del Rey unit) to 6.0552 kg/tree for one of the most productive areas (no. 7, Iscar unit).

4.3. Case study: practical application of the ecological-type model for predictions in natural unit no. 5

4.3.1. Independent data-set validation

The ecological-type model (9) was validated by contrasting observed versus predicted average cone production between 2001 and 2005 for the 465 trees in the 31 plots of the validation data set (natural unit no. 5). The average statistics for the validation plots are shown in Table 9. Comparison were carried using conditional predictions including in model (9) the value for the predicted EBLUP for whole regional period 2001–2005 effect $z_k = -0.259$. In the simulation, we assumed an expected marginal value of zero for tree, plot × period and residual random components, v_{ij}, s_{ik} and e_{ijk}. The correction factor for antilogarithmic transformation is given by:

$$\text{Correction factor} = \exp(0.5 \times [0.0392 + 0.1792 + 0.0102\exp^{(0.0561\,d)}])$$

Observed versus predicted plot-level cone yield values are charted in Fig. 5. The mean error for the 31 analysed plots was −1.534 kg/plot, *p*-value 0.0293, while the root mean square error reached 4.046 kg/plot (equivalent to 0.262 kg/tree). The predicted annual value for total production of the 31 plots was 266 kg (8.586 kg/plot), while the observed real value was 218 kg (7.051 kg/plot).

4.3.2. Stochastic simulations

Predictions for future cone production should take into account between-period variability. Conditional simulations

Table 9
Summary statistics for validation data set in natural unit 5, *Nava del Rey*

	N (stem/ha)	dg (cm)	BA (m^2/ha)	Average cone yield 2001–2005 kg/plot	kg/tree	kg/ha
Mean	138	28.2	8.38	7.051	0.470	56.727
Min	65	17.7	2.93	1.777	0.118	12.443
Max	367	39.5	14.79	21.099	1.406	144.393

Data from 31 plots containing 465 trees.

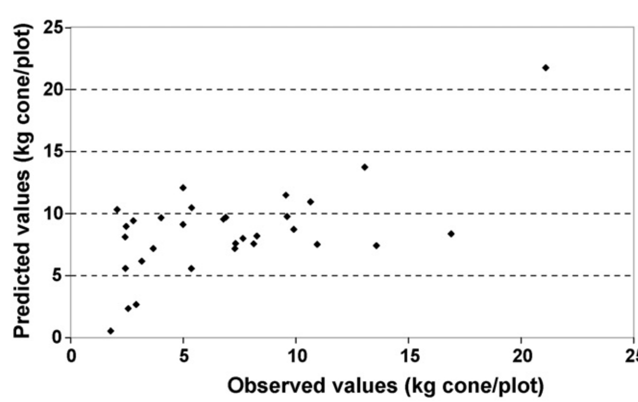

Fig. 5. Observed vs. conditional predicted cone production values for the 31 plots in validation data set (natural unit no. 5). Period 2001–2006.

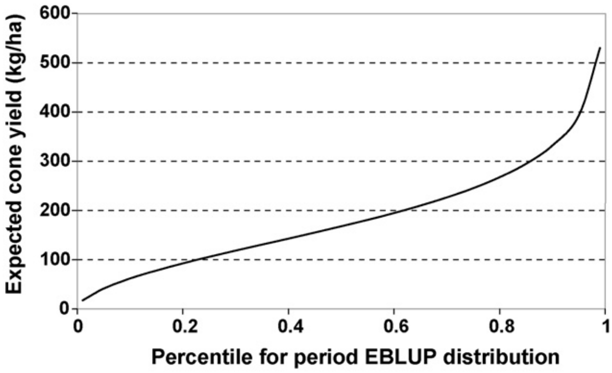

Fig. 6. Expected cone yield (kg/ha) in validation plot no. 26 as a function of the percentile of the distribution for random period effects (mean zero, variance 0.1142).

would be useful for defining and simulating scenarios based on favourable or unfavourable periods (probably related to climatic events, such as rainfall reduction). If we focus on validation plot no. 26, in which $N = 135$ stem/ha; dg = 33.4 cm; area = 1110 m^2, and marginal prediction (EBLUP for $z_k = 0$) for cone production = 18.631 kg/plot (167.8 kg/ha), it can be seen from Fig. 6 how by simulating random period effects, ranging from the 1st to the 99th percentile of the random period distribution (normal with mean zero and variance 0.1142) the conditional predictions for cone production per hectare are affected. If we consider that viable areas in terms of crop production are those with an average cone crop above 125 kg/ha (approximately 33rd percentile), then there is a 67% probability that plot 26 will reach this value, whereas the probability of average cone crops above 400 kg/ha is less than 5%. Likewise, if we define a scenario in which the expected period EBLUP is reduced from zero down to −0.1, the expected crops per hectare would fall below a value of 139 kg/ha.

A final possibility for stochastic prediction is to simulate the real cone production distribution for the trees within the plot, by adding to each tree a random realization from the distribution of tree effects v_{ij} (normal with mean zero and variance 0.0392). For plot 26, in an average period, marginal tree level prediction reaches an average expected value of 1.242 kg/tree, with minimum and maximum values of 0.483 kg/tree and 2.187 kg/tree (predicted within plot variance 0.394). The expected commercially viable crop, defined as that obtained from trees with an average crop greater than 1 kg, should reach 12.628 kg/plot, equivalent to 114 kg/ha. After 100 Monte Carlo simulations, carrying out conditional predictions by assigning to each tree a random realization from the tree random effects distribution, the expected average value for tree level production is similar (1.256 kg/tree), but the average minimum, maximum and variance values are now 0.258 kg/tree, 2.848 kg/tree and 0.620 kg/tree, whilst the expected cone crop collected for trees with an average production >1 kg reaches 14.712 kg/plot, which is equivalent to 132 kg/ha. In this way, by simulating real distribution it is possible to obtain more accurate estimates for commercially viable crops.

5. Discussion and conclusions

Single tree cone production, averaged over a 5-year period, was defined as a stochastic process, where different sources of variability can be recognized at plot, tree, period, plot × period and residual (tree × period) levels. After comparing the three different modelling alternatives it was found that the *ecological-type* model performed best, while the *hybrid* approach, despite a more explicit inclusion of climatic and orographic conditions, only performed slightly better than the traditional *silvicultural* approach.

The proposed silvicultural model includes tree section, stand density, site index and the ratio between tree diameter and the quadratic mean diameter of the plot as covariates. These covariates are similar to those proposed in previous works to explain single tree cone or nut production for species in other regions (Cappelli, 1958; Cañadas, 2000; Montero et al., 2001), or in controlled conditions such as clonal banks (Mutke et al., 2005a). The covariates included indicate that the bigger or more dominant the trees, the lower the density and the higher the site quality of the stands where they grow, the larger their average cone crops. From a management viewpoint, this means that by maintaining low densities through the application of low intensity thinning from the early phases of development, individual cone production can be increased, thus reducing collection costs.

The level of unexplained between-plot variability after fitting the silvicultural model (9%), is similar to that found by Calama and Montero (2007) when modelling annual cone and seed production for stands in the Central Range of Spain. This points to the spatial dependence of cone production, as stated by Nanos et al. (2003) in their analysis of average cone production for the same region. Assuming an ecological origin for this spatial variability, climatic and orographic covariates were evaluated to explain the remaining spatial variability. The average values for winter rainfall were found to be significant and positively associated with cone production, whilst altitude was found to be negatively related. Unexplained between-plot variability was thus reduced down to 5%. This result implies that the warmer (lower altitude) or more humid areas within the region are those which attain the larger cone crops. This result agrees with previous findings, i.e. that the most productive regions within the area of natural distribution are those free of extreme winter temperatures where water availability is not compromised (e.g. oceanic-influenced stands in Portugal or Spanish Central Range). Levels of rainfall in winter and spring are closely related to water availability at the time of primordial formation and pollination, being the meteorological factor which best explains temporal variability in cone production (Mutke et al., 2005b). The results of this study indicate that winter rainfall levels are of great importance in explaining not only temporal dynamics but also the spatial variability of flowering and fruiting processes in Mediterranean forests. The inference that water is the main climatic factor ruling flowering and fruiting is supported by findings for other xeric and Mediterranean environments (Koenig et al., 1996; Abrahamson and Layne, 2003), contrasting with the results obtained for boreal (e.g. Henttonen et al., 1986; Sirois, 2000) or temperate (e.g. Woodward et al., 1994; Masaka and Sato, 2002) forests, where temperature was found to be the main factor controlling the reproduction process.

The ecological based model, developed including natural units explains a major part of the original between-plot variability in cone production, making it unnecessary to include site index, rainfall or altitude in the model. Marginal predictions carried out using the ecological based model account for 45% of the original unexplained variability in cone production (log-scale, 38.8% in real-scale), achieving unbiased or slightly downward-biased estimates for the main part of the studied territory. Since Burkhart's seminal work (1977), ecological-type stratifications have been widely used as an alternative to traditional climatic, soil or orographic covariates for modelling spatial variability in site index equations (Beland and Bergeron, 1996; Beaumont et al., 1999; Álvarez-González et al., 2005), height–diameter relationships (Huang et al., 2000), diameter or basal area increment (Barrio-Anta et al., 2006), stem curve equations (Lappi, 1986), NWFP such as cork (Vazquez, 2002) or even for constructing nationally applicable growth and yield models (Dzierzon and Mason, 2006). Among the few studies devoted to modelling forest fruit production, both Ihalainen et al. (2003) and Reynolds-Hogland et al. (2006) included site type indicators as explanatory covariates for predicting spatial and temporal variability in berry yields. Our result shows that an adequate expertise-based ecological classification of forest terrains allows factors such as climate or soil attributes to be considered indirectly in the modelling process without the necessity for direct inclusion as covariates which would be both complex and expensive. Stratification facilitates provincial management of forest stands under a basis of cone production, identifying the more viable areas in terms of cone yield or on the contrary, areas where there is a high probability of a non-profitable cone crop. In this case, the more exploitable cone production units were those with greater water availability due to either: higher rainfall (limestone areas, units 1 and 2), a superficial water table (river terraces, unit 10) or a larger retention capacity associated with more developed soils (unit 7). On the other hand, smaller productions are associated with highly permeable sandy or gravel soils (water retention under 100 mm), in areas with lower annual rainfall (units 4, 5, 6, 8 and 9). As can be seen, natural unit stratification, that reflects also soil drainage conditions, is better for explaining water availability than the hybrid model in which only rainfall input is considered.

Approximately 6% of the remaining variability after fitting the ecological based model is accounted for by tree-within plot variance, a figure similar to that found by Calama and Montero (2007). The remaining tree-within plot variability is related to genotype (Mutke et al., 2005a) as well as microsite factors. Improving marginal predictions appears to be a difficult task since the main part of the remaining unexplained variability is associated with periodical effects (period, plot $\times$ period, tree $\times$ period). Even by identifying the whole set of factors explaining tree and plot level variability (soil attributes, tree level distance dependent competition, aggregation indices, microsite factors), or by calibrating models using additional observations, the percentage of explained variability would not reach 50%.

The identification of climatic factors controlling periodical effects will increase scientific understanding of the ecology of the species and will facilitate the definition of future simulation scenarios such as reduction in rainfall, temperature increment, pest affection, etc. Rainfall has been postulated as the main factor explaining long-term variability (Gordo, 2004; Mutke et al., 2005a,b). Incidentally, in our study, the EBLUP for the 1996–2000 period is +0.259, while for 2001–2005 the value is -0.259, coinciding with a 10% reduction in average rainfall detected between those periods (492 mm versus 448 mm, data from Central INM station in Valladolid). With regard to the unexplained plot $\times$ period variability, possible explanatory factors may be related with short spatial range meteorological events (storms, freezing), pest affection or certain plot responses (due to specific soil attributes) to climatic conditions for a given period. Such attributes are not easily to register accurately in time and space nor to simulate, hence they are difficult to integrate within a model.

Together with the increased amount of explained variability through marginal predictions, stochastic simulation for changing future scenarios, and spatial arrangement of the region based on potential cone productivity, the proposed ecological-type model has a main advantage compared with the silvicultural and hybrid developed models: the fact that neither age nor climatic measurements are required for making predictions (since neither age nor site index enter the model), meaning that it can be used directly from typical management inventory data (usually including only diameter distribution for a given plot). Main disadvantage is related with the application of the ecological model out of the geographical range defined by the natural units' stratification. In these areas, hybrid (if climatic data are available) and silvicultural models can be considered capable tools for estimating cone production. A second possibility is to extend the concept of natural units to the rest of regions where the stone pine is actually growing, leading to a homogenous classification of the species' territory based on cone production criteria. Whatever the case, the three proposed models are directly incorporable as an independent module within the integrated PINEA2 model, since the predictor covariates are obtained directly from inventory data and proposed management schedules. After integrating the cone production module, PINEA2 provides a powerful tool for the sustainable management of stone pine landscapes orientated towards cone production, allowing development and yield to be simulated under different scenarios of non-stationary and erratic environmental conditions and different silvicultural schedules.

Acknowledgements

Authors wish to thank Carlos García-Güemes, Angel Bachiller and Enrique Garriga for their dedication to plot installation, maintenance and data collection. Research was carried out in the context of INIA funded project CPE-03-001-C5.2.

References

Abrahamson, W.G., Layne, J.N., 2003. Long-term patterns of acorn production for five oak species in xeric Florida uplands. Ecology 84 (9), 2476–2492.

Álvarez-González, J.G., Ruíz, A.D., Rodríguez-Soalleiro, R., Barrio-Anta, M., 2005. Ecoregional site index models for *Pinus pinaster* in Galicia (NW Spain). Ann. For. Sci. 62, 115–127.

Barrio-Anta, M., Castedo, F., Diéguez-Aranda, U., Álvarez-González, J.G., Parresol, B.R., Rodríguez-Soalleiro, R., 2006. Development of a basal area growth system for maritime pine in northwestern Spain using the generalized algebraic difference approach. Can. J. For. Res. 36, 1461–1474.

Beaumont, J.F., Ung, C.H., Bernier-Cardou, M., 1999. Relating site index to ecological factors in black spruce stands: tests of hypotheses. For. Sci. 45 (4), 484–491.

Beland, M., Bergeron, Y., 1996. Height growth of jack pine (Pinus banksiana) in relation to site types in boreal forests of Abatibi, Quebec. Can. J. For. Res. 26, 2170–2179.

Biging, G.S., 1985. Improved estimates of site index curves using a varying-parameter model. For. Sci. 31, 248–257.

Burkhart, H.E., 1977. Site index equations for radiate pine in New Zealand. N.Z. J. For. Sci. 81, 232–234.

Calama, R., 2004. Modelo interregional de selvicultura para *Pinus pinea* L. Aproximación mediante 508 funciones con componentes aleatorios. Ph. Dr. Thesis. Universidad Politécnica de Madrid.

Calama, R., Cañadas, N., Montero, G., 2003. Inter-regional variability in site index models for even-aged stands of stone pine (*Pinus pinea* L.) in Spain. Ann. For. Sci. 60, 259–269.

Calama, R., Montero, G., 2004. Interregional non-linear height–diameter model with random coefficients for Stone Pine in Spain. Can. J. For. Res. 34, 150–163.

Calama, R., Montero, G., 2007. Cone and seed production from stone pine (*Pinus pinea* L.) stands in Central Range (Spain). Eur. J. For. Res. 126 (1), 23–35.

Calama, R., Sánchez-González, M., Montero, G., 2007. Integrated management models for Mediterranean multifunctional forests: the case of stone pine (*Pinus pinea* L.) EFI Proceedings.

Cañadas, M.N., 2000. *Pinus pinea* L. en el Sistema Central (valles del Tiétar y del Alberche): desarrollo de un modelo de crecimiento y producción de piña. Ph. Dr. Thesis. Universidad Politécnica de Madrid.

Cappelli, M., 1958. Note preliminari sulla produzione individuale de stroboli in *Pinus pinea* L. L'Italia Forestale e Montana 13 (5), 181–203.

Castellani, C., 1989. La produzione legnosa e del fruto e la durata economico delle pinete coetanee di pino domestico (*Pinus pinea* L.) in un complesso assestato a prevalente funzione produttiva in Italia. Annali ISAFA 12, 161–221.

Castro, J., Gómez, J.M., García, D., Zamora, R., Hódar, J.A., 1999. Seed predation and dispersal in relict Scots pine forests in southern Spain. Plant Ecol. 145, 115–123.

Dzierzon, H., Mason, E.G., 2006. Towards a nationwide growth and yield model for radiate pine plantations in New Zealand. Can. J. For. Res. 36, 2533–2543.

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.

Flewelling, J., Pienaar, L.V., 1981. Multiplicative regression with lognormal errors. For. Sci. 27 (2), 281–289.

Fox, J.C., Ades, P.K., Bi, H., 2001. Stochastic structure and individual-tree growth models. For. Ecol. Manage. 154, 261–276.

García-Güemes C., 1999. Modelo de simulación selvícola para *Pinus pinea* L. en la provincia de Valladolid. Ph. Dr. Thesis. Universidad Politécnica de Madrid.

Gordo, F.J., Mutke, S., Gil, L., 2000. La producción de piña de *Pinus pinea* L. en los montes públicos de Valladolid. First Internacional Meeting on stone pine (*Pinus pinea* L.). Valladolid. Junta de Castilla y León. Issue 2, 269–277.

Gordo, F.J., 2004. Selección de grandes productores de fruto de *Pinus pinea* L. En la Meseta Norte. Ph. Dr. Thesis. Universidad Politécnica de Madrid.

Gregoire, T.G., 1987. Generalized error structure for forestry yield models. For. Sci. 33, 423–444.

Gworek, J.R., Vander Wall, S.B., Brussard, P.F., 2007. Changes in biotic interactions and climate determine recruitment of Jeffrey pine along an elevation gradient. For. Ecol. Manage. 239, 57–68.

Henttonen, H., Kanninen, M., Nygren, M., Ojansuu, R., 1986. The maturation if *Pinus sylvestris* in relation to temperature climate in northern Finland. Scand. J. For. Res. 1, 243–249.

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

Ihalainen, M., Salo, K., Pukkala, T., 2003. Empirical prediction models for Vaccinium myrtillus and V. vitis-idaea berry yields in North Karelia, Finland. Silva Fennica 37 (1), 95–108.

Koenig, W.D., Knops, J.M.H., Carmen, W.J., Stanback, M.T., Mumme, R.L., 1996. Acorn production by oaks in central California: influence of weather at three levels. Can. J. For. Res. 26, 1677–1683.

Lappi, J., 1986. Mixed linear models for analyzing and predicting stem form variation of scots pine. Communicationes Instituti Forestalis Fenniae 134, 1–69.

Lappi, J., 1991. Calibration of height and volume equations with random parameters. For. Sci. 37 (3), 781–801.

Magini, E., Giannini, R., 1971. Prime osservazioni sulla produzione di strobili e semi di un parco di cloni di pino domestico (*Pinus pinea* L.). Italia Forestale E Montana 26 (2), 63–78.

Masaka, K., Sato, H., 2002. Acorn production by Kashiwa oak in a coastal forest under fluctuating weather conditions. Can. J. For. Res. 32, 9–15.

Mehtätalo, L., 2004. A longitudinal height–diameter model for Norway spruce in Finland. Can. J. For. Res. 34, 131–140.

Mencuccini, M., Piussi, P., Zanzi Sulli, A., 1995. Thirty years of seed production in a subalpine Norway spruce forest: patterns of temporal and spatial variation. For. Ecol. Manage. 76, 109–125.

Montero, G., Candela, J.A., Ruiz-Peinado, R., Gutiérrez, M., Pavón, J., Bachiller, A., Ortega, C., Cañellas, I., 2001. Age and density influence in flowering and kernel quality in *Pinus pinea* L. forests in the south of Huelva, International Field Trip Meeting Mediterranean Silviculture.In: Sevilla. IUFRO-INIA. pp. 99–116.

Mutke, S., Gordo, J., Gil, L., 2001. Modelo Individual de producción de piñón de Pinus pinea L. como criterio de selección fenotípica. Third Spanish National Forest Congress. Granada. Junta de Andalucía - TRAGSA - SECF. Issue 3, 172–178.

Mutke, S., Gordo, F.J., Gil, L., 2005a. Characterization of a stone pine (*Pinus pinea* L.) clone bank. Silvae Genet. 54 (4–5), 189–197.

Mutke, S., Gordo, F.J., Gil, L., 2005b. Variability of Mediterranean Stone pine cone production: yield loss as response to climatic change. Agric. For. Met. 132, 263–272.

Nanos, N., Calama, R., Cañadas, N., García, C., Montero, G., 2003. Spatial stochastic modelling for cone production from stone pine (*Pinus pinea* L.) stands in Spanish Northern Plateau. In: Amaro, A., Reed, D., Soares, P. (Eds.), Modelling Forest Systems. CABI Publishing, Wallingford, pp. 131–141.

Nanos, N., Calama, R., Gil, L., Montero, G., 2004. Geostatistical prediction of height/diameter models: a case study. For. Ecol. Manage. 195, 221–235.

Piovesan, G., Adams, J.M., 2001. Masting behaviour in beech: linking reproduction and climatic variation. Can. J. Bot. 79, 1039–1047.

Piqué, M., 2003. Modelos de producción para las masas de *Pinus pinea* L. en Catalunya: orientaciones para la gestión y el aprovechamiento sostenible de madera y piña. Ph. Dr. Thesis. Universidad de Lleida.

Reynolds-Hogland, M.J., Mitchell, M.S., Powell, R.A., 2006. Spatio-temporal availability of soft mast in celarcuts in the Southern Appalachians. For. Ecol. Manage. 237, 103–114.

Sánchez-González, M., Calama, R., Cañellas, I., Montero, G., 2007. Variables influencing cork thickness in Spanish cork oak forests: a modelling approach. Ann. For. Sci. 64, 301–312.

Sánchez-Palomares, O., Sánchez-Serrano, F., Carretero, M.P., 1999. Modelos y catografía de estimaciones climáticas y termopluviométricas para la España Peninsular. MAPA-INIA, Madrid.

Sirois, L., 2000. Spatiotemporal variation in black spruce cone and seed crops along a boreal forest—tree line transect. Can. J. For. Res. 30, 900–909.

Trincado, G., Burkhart, H., 2006. A generalized approach for modelling and localizing stem profile curves. For. Sci. 52 (6), 670–682.

Vazquez, J., 2002. Modelos predictivos de producción de corcho y detección precoz de la calidad. Ph. Dr. Thesis. Universida de Técnica de Lisboa. Lisboa.

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

West, P.W., Ratkowsky, D.A., Davis, A.W., 1984. Problems of hypothesis testing of regressions with multiple measurements from individual sampling units. For. Ecol. Manage. 7, 207–224.

Woodward, A., Silsbee, D.G., Schreiner, E.G., Means, J.E., 1994. Influence of climate on radial growth and cone productionin subalpine fir (Abies lasocarpa) and mountain hemlock (Tsuga mertensiana). Can. J. For. Res. 24, 1133–1143.