Available online at www.sciencedirect.com

SCIENCE DIRECT®

Forest Ecology and Management 213 (2005) 117–132

Forest Ecology and Management

www.elsevier.com/locate/foreco

Mixed models to estimate tree oven-dried cork weight in Central and Southern Portugal

Javier Vázquez*, Helena Pereira

Departament of Forestry, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Tapada da Ajuda, P-1349-017 Lisboa Codex, Portugal

Received 8 June 2004; received in revised form 2 November 2004; accepted 27 March 2005

Abstract

Models of oven-dried cork weight at tree level were developed using dendrometric variables and rotation cycle of cork production (9 or 10 years) as predictors. The models were based on data obtained from permanent plots laid out in five cork regions, covering most of the cork production area in Portugal. Dendrometric variables included those related to tree size (as perimeter at breast height, stem height or crown dimensions), related to management decisions in the stripping process (as stripped length in stem and branches, stripping surface, stripping intensity or stripping coefficient) and those gathered from a cork sample (as cork thickness, cork density or cork moisture). Variables were grouped according to the complexity of measurement with the aim of developing models suitable for management, that include variables with cheaper measurement, and for research, including variables with higher measurement cost. Models were structured as mixed models with random effects decomposed in regional, plot and tree-level effects, allowing higher sensitivity and more realistic variance–covariance structures comparing with fixed effects models. The variance component at tree level reflects 80–90% of total variability and at plot level is only 1–3% The nine selected models, five for research and four for management purposes, are linear models that reflect the allometry in the relationship between cork weight and tree size.

Keywords: *Quercus suber* L.; Mixed models; Cork production; Parameter estimation; Variance components; Spatial correlation; Generalized least-squares

1. Introduction

Cork is a natural product primarily used to produce stoppers but also other manufactured goods like agglomerates for thermal and acoustic insulation or for decorative purposes. Cork oak (*Quercus suber* L.) is one of the most important forest species in Portugal, covering 713,000 ha (8% of the whole country and 21% of the forested area) (DGF, 2001), and has a high socio-economic and ecological importance. The economic importance is mainly based on cork production for wine stoppers. Portugal is the first country of the world in production,

* Corresponding author. Present adress: Department of Forest Sciences, University of Huelva, Crta, Palos-La Rábida s/n, 21819 Palos de la Frontera, Huelva, Spain. Tel.: +34 959217531; fax: +34 959217560.

E-mail address: javier.vazquez@dcaf.uhu.es (J. Vázquez).

0378-1127/$ – see front matter
doi:10.1016/j.foreco.2005.03.047

exportation and transformation of cork products (CESE, 1996).

The cork layer is originated by the continuous activity of the phellogen. Periodically (commonly at 9-year intervals, but also 10), the cork of the stem and branches with perimeter at breast height greater than 70 cm is removed. The phellogen dies in the debarking operation but a new phellogen layer is regenerated inside the inactive phloem, allowing the formation of a new cork layer. The extracted cork planks are then the raw material for cork stoppers and other less valuable cork products as agglomerates.

The development of cork weight prediction models is important in two different scenarios: (1) for the forest management, cork weight models are tools to develop integrated management models as important as diameter, cork growth and cork quality models and (2) for the economy, the assessment of cork production at local, regional and national level allows a better programming for the industry supply of raw material and the exportation of manufactured products. Several authors have developed equations to predict individual tree cork weight (Natividade, 1950; Guerreiro, 1951; Alves, 1958; Alves and Macedo, 1961; Ferreira et al., 1986; Montero, 1987; Gomes et al., 1990; Ribeiro, 1990, 1992; Ferreira and Oliveira, 1991; Costa, 1992, 1997; Costa and Oliveira, 1998; Fonseca and Parresol, 2001; Ribeiro and Tomé, 2002). The main characteristics of these models are discussed below.

- The dependent variable is fresh cork weight, measured immediately after cork extraction, excepting Ribeiro and Tomé (2002), who use air-dried cork weight (after 2-week field drying), and Costa and Oliveira (1998) who incorporate oven-dried cork weight, with cork moisture measured in a cork sample. The inter-tree and inter-regional variability in cork moisture content immediately after cork extraction is large, as referred by González-Adrados et al. (1993) and Costa (1997), thereby questioning the adequacy of using fresh cork weight as response variable. Variability in air-dried cork moisture is lower but it depends on environmental conditions, mainly temperature and humidity (González-Adrados and Calvo, 1994), and do not allow to compare cork weight in different regions. In the present work, the response variable is oven-dried cork weight, with cork moisture evaluated through a 20 cm × 20 cm sample collected at 1.30 m, that is the only way to compare cork in different regions and to calculate cork weight at different moisture contents, an important aspect in commercial transactions. There is no published research concerning the accuracy of evaluate the cork moisture content of a tree through a sample collected at 1.30 m. However, this approach will be less dependent of the humidity and temperature at the moment of cork extraction than if we consider fresh cork weight or air-dried cork weight as dependent variable.
- The independent variables in all the models are based on the measurement of variables related to tree size and stripping (tree cork weight could be simply modified varying the stripped height of the tree, that is a management decision). The main differences between the models refer to the measuring complexity of the variable or variables used, from the simplest perimeter at breast height to the most complex stripped surface (SS) in Ribeiro and Tomé (2002). Since Ferreira et al. (1986), the product of the perimeter at breast height and the maximum stripped height is the most common independent variable used, as a compromise between simplicity and approximation to the real stripped surface. One model in Ribeiro and Tomé (2002) includes also variables measured on a cork sample, such as the cork thickness.
- Most of the models are based on a nested structure of trees inside plots inside regions, or if the application area is reduced, of trees inside plots. Sometimes the plot structure does not exist and trees are grouped into stands or sites without any reference to the area or homogeneity of the stand. No model makes reference to the possible presence of spatial autocorrelation in such nested structures.
- Most models are simple linear models, or in a few cases, linear models obtained by logarithmic transformation of independent variables (Ferreira et al., 1986; Ribeiro, 1990) or transformation of both response and independent variables (Gomes et al., 1990; Ribeiro and Tomé, 2002) to consider the allometric relationship between tree size and cork weight. Fonseca and Parresol (2001) incorporate a non-linear model but Ribeiro and Tomé (2002) see no advantage of non-linear in comparison with log transformation linear models.

Considering the gaps and limitations in the referred models, new cork weight models at tree level were developed for Central and Southern Portugal. The response variable in these models is oven-dried cork weight. Sampling structure is considered to develop more realistic variance–covariance structures and to distinguish between fixed and random effects. The complexity and cost of tree measurements was also considered and models were developed with different objectives, e.g. research or management.

2. Material

2.1. Data

Data was collected in 12 permanent plots installed by the Instituto Superior de Agronomia (Lisboa, Portugal) in the regions of Azaruja, Escoural I, Escoural II, Porto Alto (Central Portugal) and S. Brás de Alportel (south). These regions are of mediterranean type climate, with summer drought period over 3 months. Situation and main climatic characteristics are shown in Table 1.

A total of 251 trees were sampled in these plots, with distribution of trees in plots and regions as indicated in Table 2. Plot area varied from 4425 to 22,790 m^2, basal area from 3.8 to 13.0 $m^2\,ha^{-1}$ and number of trees from 86.6 to 156.6 ha^{-1}. Cork-stripping cycle was 9 years in 62 trees and 10 years in 189 trees. Cork fresh weight and cork moisture for the sampled trees are shown in Table 3.

The following measurements and variables were computed for each tree (Table 4):

- Field measurements: Previous to cork extraction, the perimeter at breast height over cork (PBOC) was measured and a cork sample of 20 cm × 20 cm was taken at 1.3 m, facing west exposition. The cork sample was introduced in a plastic bag to maintain cork moisture. Cork was then extracted as cork planks and weighed immediately. Variables that reflect tree size (as perimeter at breast height under cork, stem height or crown radius) and variables focusing exclusively on the area of the tree stem and branches that yields cork (such as the stripping lengths and perimeters under cork of stem and branches), with the aim of approximate or calculate exactly the surface of the tree that yield cork, were then gathered.
- Calculated variables: Include variables that account for the stripping length of the tree (sum of stripping lengths of stem and the mean, maximum or total value of branches); the stripped surface itself (SS) or variables that reflect, as relative values, the stripping

Table 1
Location and main climatic characteristics of plots

Region	Plot	UTM Coordinates	Altitude (m)	P (mm)	D (months)	T (°C)	TMMH (°C)	TMH (°C)	Tmc (°C)	Tmmc (°C)
Azaruja	1	29SPC04738779	319	604	4.2	15.8	30.2	23.2	9.5	6.1
	2	29SPC04718757	318							
	3	29SPC05848851	321							
Escoural I	4	29SNC75606860	371	817	3.6	15.8	30.2	23.2	9.5	6.1
	5	29SNC75956839	369							
	6	29SNC76976851	347							
Escoural II	7	29SNC78257767	289	817	3.6	15.8	30.2	23.2	9.5	6.1
	8	29SNC78136892	310							
	9	29SNC78196909	316							
Porto Alto	10	29SNC12889717	23	574	4.1	16.5	28.8	22.6	10.3	5.9
	11	29SNC12939726	23							
S. Brás de Alportel	12	29SPB01302097	453	984	4.1	16.3	30.9	24.3	10.4	6.7

Climatic variables are referred to the period 1961–1990; P, mean annual precipitation (mm); D, drought period (number of months in which $P < 2T$); T, mean annual temperature (°C); TMMH, mean maximum temperature of the hottest month; TMH, mean temperature of the hottest month; Tmc, mean temperature of the coldest month; Tmmc, mean minimum temperature of the coldest month.

Table 2
Distribution of trees per plots and regions and main characteristics of plots

Region	Plot	Area (m^2)	Number of sampled trees	Cork-stripping cycle (year)	Basal area ($m^2\ ha^{-1}$)	Number of trees (ha^{-1})
Azaruja	1	7726.4	21	10	9.0	113.2
	2	14325.0	21	9	7.7	86.6
	3	4727.8	20	10	10.6	145.9
Escoural I	4	11581.5	16	10	11.6	93.6
	5	10602.2	15	10	12.6	128.3
	6	8879.9	15	10	11.0	86.7
Escoural II	7	12491.3	18	10	13.0	151.3
	8	12376.1	17	10	11.5	99.8
	9	8731.3	19	10	10.3	120.8
Porto Alto	10	5493.4	32	9	11.3	156.6
	11	4424.8	19	9	8.7	126.6
S. Brás de Alportel	12	22789.7	50	10	3.8	103.6

pressure the tree is receiving (stripping intensity (SI) = SS/basal area or stripping coefficient (SC) = stripping length/perimeter at breast height). In the 20 cm × 20 cm cork sample, the value of cork thickness, surface density (cork sample weight/cork sample surface), density and moisture content was measured.

The response variable, oven-dried cork weight (DW), was computed with the cork weight after extraction (*W*) and the moisture measured in the cork sample. Table 5 shows the characterisation of the main variables collected in the 251 sampled trees.

3. Methods

3.1. Model and variance–covariance structure

The general model structure is:

$$y_{ijk} = \beta_0 + \sum_{s=1}^{h} \beta_s x_{ijks} + r_i + p_{ij} + e_{ijk}$$

where y_{ijk} is the oven-dried cork weight of *k*th tree within *j*th plot in *i*th region, β_0 the intercept, x_{ijks} the value of the *s*th selected variable in *k*th tree within *j*th plot in *i*th region, β_S the unknown coefficient of the *s*th

Table 3
Cork moisture and cork fresh weight of the sampled trees with indication of the coefficient of variation (CV%) and maximun (Max) and minimum (Min) values

Plot	*n*	Cork moisture				Cork fresh weight			
		Mean (%)	CV (%)	Max	Min	Mean (kg)	CV (%)	Max	Min
1	21	38.12	22.21	57.81	25.10	42.60	53.40	85.00	10.00
2	20	8.51[a]	13.01	10.19	6.71	40.83	45.24	74.00	7.00
3	19	28.18	16.11	35.85	20.04	37.39	57.13	111.00	20.00
4	16	45.82	23.19	60.84	28.00	37.06	69.37	115.00	15.00
5	15	47.39	32.44	81.17	32.18	34.20	76.78	100.00	15.00
6	15	36.38	14.17	46.23	25.51	33.33	75.29	95.00	10.00
7	18	45.65	21.79	61.91	28.15	43.83	81.05	140.00	15.00
8	17	33.24	18.01	48.36	23.61	35.24	75.94	101.00	12.00
9	19	35.47	13.05	42.29	25.60	27.37	87.75	120.00	12.00
10	32	22.14	34.02	50.43	4.41	23.78	59.51	62.00	7.00
11	19	29.05	94.32	123.81	9.43	28.88	77.95	80.00	8.00
12	50	32.16	41.96	68.15	13.95	24.26	106.07	125.50	4.50

[a] Cork weight was measured several hours after cork extraction.

Table 4
Variables considered in the sampled trees

No.	Variable	Definition and description	Unit
1	W	Cork weight measured immediately after cork extraction	kg
2	PBOC	Perimeter at breast height over cork	m
3	PBIC	Perimeter at breast height under cork	m
4	R_i	Crown radius at azimuth i, with $i = 0°$, $90°$, $180°$ and $270°$	m
5	SH	Stem height	m
6	SHS	Stripped height in the stem	m
7	PMIC	Stem perimeter inside cork measured at the middle of the stem stripped height	m
8	NB	Number of stripped main boughs	–
9	SLS_i	Stripped length of main bough i measured from the insertion into the stem	m
10	SLB_j	Stripped length of bough j measured from the insertion into other bough	m
11	PB_j	Perimeter at breast height of bough j under cork measured at the middle of the bough stripped length	m
12	BAIC	Basal area under cork. $BAIC = PBIC^2/4\pi$	m^2
13	CB	Calculated cork thickness. $CB = 10^3(PBOC - PBIC)/2\pi$	mm
14	SLMAX	Maximum stripping length. $SLMAX = SHS + \max(SLS_i)$	m
15	SLMEAN	Mean stripped length. $SLMEAN = SHS + \frac{\sum_{i=1}^{NB} SLS_i}{NB}$	m
16	SLTOT	Total stripped length. $SLTOT = SHS + \sum_{i=1}^{NB} SLS_i$	m
17	SCMAX	Maximum stripping coefficient. SCMAX = SLMAX/PBOC	–
18	SCMEAN	Mean stripping coefficient. SCMEAN = SLMEAN/PBOC	–
19	SCTOT	Total stripping coefficient. SCTOT = SLTOT/PBOC	–
20	SS	Stripped surface. $SS = SHS \cdot PMIC + \Sigma SLB_j \cdot PB_j$	m^2
21	ASS	Stripped surface approximation, proposed by Costa (1993), unpublished. $ASS = PBIC \cdot SHS + PBIC \cdot SLMEAN \cdot \sqrt{NB}$	m^2
22	SIN	Stripping intensity. SIN = SS/BAIC	–
23	WSS	Fresh cork weight by unit of stripped surface. WSS = W/SS	kg m^{-2}
24	RM	Mean crown radius. $RM = \Sigma R_i/4$	m
25	CCS	Circular crown surface. $SCC = \pi \cdot (\Sigma R_i/4)^2$	m^2
26	ECS	Elliptical crown surface	m^2
27	CTB	Sample cork thickness before boiling	mm
28	CTA	Sample cork thickness after boiling	mm
29	KGM2	Sample weight by unit of surface. Weight measured 15 days after boiling	kg m^{-2}
30	KGM3	Density of the cork sample	kg m^{-3}
31	SFW	Fresh cork weight of the cork sample	kg
32	SDW	Oven-dried (5 days at 103 °C) cork weight of the cork sample	kg
33	H	Cork sample moisture. $H = 100 \times (SFW - SDW)/SDW$	%
34	DW	Tree oven-dried cork weight. $DW = W \times 100/(100 + H)$	kg
35	DWSS	Oven-dried cork weight by unit of stripped surface. DWSS = DW/SS	kg m^{-2}

Variables numbered 1–11 are of direct measurement in the field, with 2–5 reflecting tree size; 12–26 are computed from the field variables; 27–33 are measured in or computed from a 20 cm × 20 cm cork sample extracted at 1.3 m; 34–35 need information from both field and sample measurements; 34 is the response variable.

of h dendrometric variables selected, r_i the random regional effect, p_{ij} the random plot effect and e_{ijk} is the tree random effect or pure error. Hypothesis related to distribution properties of random effects were: $r_i \sim$ iid $N(0,\ \sigma_r^2)$, $p_{ij} \sim$ iid $N(0,\ \sigma_p^2)$, $e_{ijk} \sim$ iid $N(0,\ \sigma_e^2)$, independence between different hierarchical levels (that is, $cov\ (r_i, p_{ij})$=0, $cov\ (p_{ij}, e_{ijk})$=0, $cov\ (r_i, e_{ijk})$=0) and independence also between different values inside a hierarchical level (that is, $cov\ (r_i,\ r_{i'}) = 0$ for $i \neq i'$, $cov\ (p_{ij},\ p_{ij'}) = 0$ for $j \neq j'$, $cov\ (e_{ijk},\ e_{ijk'}) = 0$ for $k \neq k'$). The model is a mixed linear model; it is linear in the variables and coefficients and includes fixed effects ((μ and βs) and random effects (r, p) other than pure error. In matrix form:

$$\boldsymbol{y} = \boldsymbol{Xa} + \boldsymbol{Zb} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is an $N = 251 \times 1$ matrix with values of the dependent variable, $\boldsymbol{X}$ the fixed $N \times (h + 1)$ matrix containing the values of the s selected independent variables with first column vector of 1, $\boldsymbol{a}$ the

Table 5
Characterization of the data collected in the 251 sampled trees, showing the most important variables related to cork production

Variable	Mean	S.D.	Minimum	1st quartile	Median	3rd quartile	Maximum
PBOC (m)	1.25	0.30	0.65	1.05	1.22	1.41	2.26
PBIC (m)	1.07	0.30	0.45	0.89	1.05	1.25	2.00
SHS (m)	1.90	0.55	0.80	1.54	1.85	2.20	4.10
SLMAX (m)	2.39	0.92	0.80	1.62	2.20	3.00	6.00
CB (m^2)	27.59	6.63	6.37	22.28	27.06	31.83	47.75
SCMAX (–)	1.89	0.54	0.85	1.49	1.83	2.22	3.75
SS (m^2)	3.10	2.29	0.50	1.45	2.30	3.75	11.27
SIN (–)	30.35	9.55	12.54	23.36	29.92	35.94	56.97
RM (m)	4.07	1.10	1.35	3.36	4.09	4.80	7.36
CTB (mm)	30.29	7.18	16.85	25.00	29.50	34.13	61.85
KGM2 (kg m^{-2})	8.07	1.86	4.39	6.76	7.80	9.05	15.31
KGM3 (kg m^{-3})	246.60	44.10	167.23	218.65	239.69	268.44	429.16
DW (kg)	24.96	18.31	3.27	12.27	19.41	30.93	99.42

S.D., standard deviation.

$(h + 1) \times 1$ vector containing the coefficients of fixed effects, $\boldsymbol{Z}$ the random regional and plot effects design matrix $N \times (m + t)$ with m, number of regions and t total number of plots, $\boldsymbol{b}$ the random coefficients $(m + t) \times 1$ vector and $\boldsymbol{e}$ is the residual error $N \times 1$ vector. The variance–covariance matrix of vector $\boldsymbol{y}$, according to the previous hypothesis, is:

$$\begin{aligned}\mathrm{var}(\boldsymbol{y}) &= \mathrm{var}(\boldsymbol{Xa} + \boldsymbol{Zb} + \boldsymbol{e}) = \mathrm{var}(\boldsymbol{Zb} + \boldsymbol{e}) \\ &= \boldsymbol{Z}\mathrm{var}(\boldsymbol{b})Z' + \mathrm{var}(\boldsymbol{e}) = \boldsymbol{ZG}Z' + \boldsymbol{R} = \boldsymbol{V}\end{aligned}$$

This mixed model structure takes into account the possible spatial correlation, varying covariances between trees depending on their spatial situation (higher covariance if they are in the same plot, smaller if they belong to different plots inside the same region and null if they belong to different regions).

3.2. Selection of covariates

As matrix $\boldsymbol{Z}$ depends only on sampling structure, the models are completely defined after selecting the covariates that form the $\boldsymbol{X}$ matrix. The selection process attends the following three criteria:

(1) Absorption of variability: Tree size and stripping pressure (a term that indicates if a tree has or not a high stripped surface according to its size) are obviously the main factors that affect total tree cork weight and are the first considered in the process. After a pre-selection of models based on the dendrometric variables, the influence of cork-stripping rotation is tested.
(2) Multipurpose objective: As it is intended to develop models suitable for research and management, variables were classified according to their measurement difficulty. Three classes with several sub-classes were formed totalling 15 different kinds of models (Table 6).
(3) Statistical criteria: The choice between two models of the same type is based on statistical criteria presented in Section 3.2.1.

3.2.1. Selection of dendrometric variables

Selection of these variables was made through adjustment using the following model structures: (1) the ‘‘pure’’ linear model,

$$y_{ijks} = \beta_0 + \sum_{s=1}^{h} \beta_s x_{ijks} + e^{*}_{ijk}$$

and the non-linear model with multiplicative error

$$y_{ijk} = \beta_0 + x_{ijk1}^{\beta_1} x_{ijk2}^{\beta_2} \cdots x_{ijkh}^{\beta_h} e^{**}_{ijk}$$

that with double logarithmic transformation yields the linear model (2):

$$\begin{aligned}\log(y_{ijk}) &= \log(\beta_0) + \sum_{s=1}^{h} \beta_s \log(x_{ijks}) + \log\left(e^{**}_{ijk}\right) \\ &= \log(\beta_0) + \sum_{s=1}^{h} \beta_s \log(x_{ijks}) + e^{***}_{ijk}\end{aligned}$$

Table 6
Classification of cork oak models in classes and sub-classes according to the measuring complexity of the variables included and main model application

Class	Sub-class	Variables measured	Model application
A	1	Several variables in cork sample + SS + crown dimensions	Research
	2	Several variables in cork sample + SS	A cork sample is required
	3	KGM2 in sample + SS	
	4	CTB in sample + SS	
B	5	SS + CB + crown dimensions	Research
	6	SS + CB	No cork sample is required
	7	SS	It is necessary to measure all stripping lengths and,
	8	All stripping lengths + CB + crown dimensions	in models B-5–B-7, the circumference of all
	9	All stripping lengths + CB	stripped branches to calculate SS
	10	All stripping lengths	
C	11	Maximum stripping length + CB + crown	Forest management and inventory
	12	Maximum stripping length + CB	No cork sample is required
	13	Maximum stripping length + crown dimensions	At most, the measurement of the maximum
	14	Maximum stripping length	stripping length is required
	15	Only PBOC or PBIC	

A-1 model types are the most difficult to implement as they use the most number of variables with difficult measurement. C-15 model types are those in which variables are easiest to measure.

This last approach reflects the fact that the relationship between cork weight and tree size is allometric. However, ‘‘pure’’ linear models were also estimated, as most of previous cork oak weight models at tree level followed this approach.

Estimation method was OLS with the usual assumptions. As it has been previously discussed, the diagonal variance–covariance matrix that is assumed in OLS is not a proper initial structure but just variable selection is intended in this step and not final fitting, and selection-oriented fitting by OLS facilitates the automation of the selection process. As the initial number of variables was too high to use ‘‘all possible regression methods’’ (25 variables in the case of logarithmic transformation that means 2^{25} regressions), a previous partition of variables was made grouping variables that present high internal correlation so that no more than one of them would be present in the final model (Table 7). The automation algorithms included this restriction and the number of all possible combinations was affordable (294,912 in the case of logarithmic transformation).

Pre-selection of models was based on (Myers, 1986):

- Fitting statistics: adjusted R^2 (AdjR^2), mean squared error (MSE), Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC).
- Predictive statistics: sum of squared prediction residuals (PRESS) and sum of absolute prediction residuals (APRESS).
- Multicollinearity statistics: variance inflation factors (VIF) and condition number (NCOND). These statistics were combined in three punctuation algorithms with 0–1 range with different weights to fitting, predictive and multicollinearity criteria, respectively: PUNT1, with 30, 50, 20%; PUNT2, with 30, 60, 10% and PUNT3, with 20, 70, 10%. The aim of selecting three punctuation algorithms is that models with particularly good performance accord-

Table 7
Partition of variables in the estimation of logaritmic models

Number of partition	Variables included in partition
1	L (NB + 1)
2	LPBOC, LPBIC, LPMIC
3	LSLMAX, LSLMEAN, LSLTOT
4	LWSS, LDWSS, LKGM2
5	LSCMAX, LSCMEAN, LSCTOT
6	LSIN
7	LCB, LCTB, LCTA
8	LRM, LCCS, LECS
9	LSHS, LSH
10	LSS, LASS
11	LKGM3

Two variables belonging to the same partition cannot enter the same model. Letter L before the variable indicates that that variable has been log transformed.

ing fitting criteria but with worse predictive statistics properties were not discarded in the first step. As the aim of the models is to predict cork weight, more importance is giving to predictive statistics (it represent 50, 60, and 70% of the total value of PUNT1, PUNT2 and PUNT3, respectively) but it seems interesting to maintain in this first step models with relevant fitting statistics for further analysis. For instance, PUNT1 algorithm was:

$$\text{PUNT1} = \frac{3}{10}\left(\frac{1}{4}\left(\frac{\min(\text{MSE})}{\text{MSE}} + \frac{\text{Adj}R^2}{\max(\text{Adj}R^2)} + \frac{\min(\text{AIC})}{\text{AIC}} + \frac{\min(\text{SBC})}{\text{SBC}}\right)\right) + \frac{5}{10}\left(\frac{1}{2}\left(\frac{\min(\text{PRESS})}{\text{PRESS}} + \frac{\min(A\text{PRES})}{\text{APRES}}\right)\right) + \frac{2}{10}\left(\frac{1}{2}\left(\frac{1}{\max(\text{VIF})} + \frac{1}{\max(\text{NCOND})}\right)\right)$$

In log transformation linear models, all criteria were calculated in original units. To obtain unbiased estimations in original units the applied transformation was (Flewelling and Pienaar, 1981): $\hat{y} = \exp(\log(\hat{y}) + 0.5\text{MSE})$. All models with VIF > 5 were eliminated. In next steps, models of classes A–C (Table 6) were treated separately and the first 50 models in any of the three punctuation algorithms in each model class were pre-selected and tested for normality (Shapiro and Wilk, 1965) and homoscedasticity (White, 1980). Models with no-normal error structure or heteroskedasticity were discarded. The final selection was done discarding that models that presented worse values of the punctuation algorithms than any other model with less or equal measurement cost.

3.2.2. Influence of stripping rotation

The influence of stripping rotation was analysed introducing dummy variables. In the model $y_{ijk} = \beta_0 x_{ijk1}^{\beta_1} x_{ijk2}^{\beta_2} \cdots x_{ijkv}^{\beta_v} e_{ijk}^{**}$, the following $(v + 1)$ dummy variables were included:

- T, with value = 1 if rotation = 9 years and e if rotation = 10 years;
- D_z with value = $1/x_{ijkz}$ if rotation = 9 years and 1 if rotation = 10 years.

The new structure is:

$$y_{ijk} = T^a \beta_0 x_{ijk1}^{\beta_1} x_{ijk2}^{\beta_2} \cdots x_{ijkv}^{\beta_v} (D_1 x_{ijk1})^{\alpha_1} (D_2 x_{ijk2})^{\alpha_2} \cdots \times (D_v x_{ijkv})^{\alpha_v} e_{ijk}^{**}$$

Taking logarithms, the model is:

$$\log(y_{ijk}) = a \log T + \log \beta_0 + \beta_1 \log(x_{ijk1}) + \cdots + \beta_v \log(x_{ijkv}) + \alpha_1 \log(D_1 x_{ijk1}) + \cdots + \alpha_h \log(D_v x_{ijkv}) + e_{ijk}^{***}$$

For the 9-year rotation period:

$$\log(y_{ijk}) = \log\beta_0 + \sum_{s=1}^{v} \beta_s \log(x_{ijks}) + e_{ijk}^{**},$$

and for the 10-year rotation period:

$$\log(y_{ijk}) = (a + \log \beta_0) + \sum_{s=1}^{v} (\beta_s + \alpha_s) \log(x_{ijks}) + e_{ijk}^{***}.$$

The new variables $\log(D_s x_{ijks})$ will be denoted as DX_s. The null hypothesis H_0: $a = \alpha_1 = \alpha_2 = \cdots = \alpha_v = 0$ was tested through a F-test (α value = 0.05). If H_0 is rejected, different sub-groups of v, $v - 1$, $v - 2$, elements are tested up to find the largest sub-group, if any exist, in which H_0 is accepted. Particular attention was paid to multicollinearity relationships between dummy and continuous variables calculating VIF for each variable, the condition number and the variance proportions. In the presence of collinearity, the dummy variable with less type II sum of squares was suppressed. All calculations were implemented in SAS/IML® programs.

3.3. Final estimation and selection in the mixed model

Estimation and final selection was made resolving the mixed model equations (Searle, 1971). Fixed effects vector estimation is $\hat{\boldsymbol{a}}^{\text{o}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$ (that is the GLS estimation) and the random effects vector prediction is $\hat{\boldsymbol{b}}^{\text{o}} = \hat{\boldsymbol{G}}\boldsymbol{Z}'\hat{\boldsymbol{V}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{a}}^{\text{o}})$. Hence, a previous estimation of $\boldsymbol{V}$ is needed and, then, the estimation of $\boldsymbol{R}$ and $\boldsymbol{G}$ ($\boldsymbol{V} = \mathbf{ZGZ}' + \boldsymbol{R}$). The only unknown elements of $\boldsymbol{R}$ and $\boldsymbol{G}$ are σ_{r}^2, σ_{p}^2, σ_{e}^2 (variance components).

The method used to estimate the variance components was restricted maximum likelihood (REML) (Henderson, 1984). The Newton–Raphson algorithm of procedure MIXED of SAS/ETS® was used to minimise the −2 times the log likelihood function. Mixed model estimations were then used to compare pre-selected models from A to C classes. Following Littell et al. (1996) the comparison criteria used was: (a) value of −2 times the log likelihood function; (b) Akaike's Information Criterion; (c) Schwarz's Bayesian Criterion.

4. Results

4.1. Selection of dendrometric variables

Residuals in all ''pure'' linear models did not follow normal distribution and presented strong heteroskedasticity. Hence, only linear models originated from a double log transformation, which take into account the allometric relationship between the dendrometric variables and the response variable, were considered in next steps.

The covariates selection and the multipurpose objective of model building led to the choice of nine different models (Table 8). They include four type A, two type B and four type C models. According to the selection rules expressed in methods, models 7 and 9 should have been discarded. Model 7 was maintained, because the presence of variable LPBIC, that allows to predict tree cork weight after cork extraction, versus model 8. Model 9 is one of the most used in cork prediction and it was maintained to allow comparisons with pre-existing models. The type A models, for research, show better values of fitting and predictive criteria. The difference between types B and C models is not so stressed.

4.2. Influence of stripping rotation

Table 9 shows the results of the F-tests to test the influence of the stripping rotation on cork weight. In all A, and in two C models, the null hypothesis is rejected. As not all the variables included in these models could be logically influenced by the stripping rotation, a survey to find the largest sub-group in which the null hypothesis is accepted was followed (Table 10). Models 2, 6 and 7 that included two dummy variables each, revealed strong multicollinearity (statistics not shown), so in models 2, the dummy variable DSIN, with less type II sum of squares, was not considered, as also DCB in models 6 and 7. As DRM was not significant in models 6 and 7 after the suppression of DCB, no dummy variables were finally included in these two models.

Hence, only type A models are able to reflect changes in cork weight when the stripping rotation

Table 8
Pre-selected models in the covariates selection phase, with indication of fitting criteria and punctuation

No.	Variables in model	Fitting criteria									Punctuation		
		Class	MSE	AdjRSQ	AIC	SBC	PRESS	APRES	VIF	NCON	PUNT1	PUNT2	PUNT3
1	LKGM2 LCTB LRM LSS	A1	13.35	0.96	655.46	673.09	13.80	2.43	2.47	8.04	0.85 (6)	0.93 (1)	0.92 (1)
2	LPBIC LKGM2 LSIN LCTB	A2	14.18	0.96	670.58	688.20	14.63	2.47	2.49	8.16	0.83 (36)	0.90 (48)	0.89 (48)
3	LPBIC LSIN LCTB	A4	16.49	0.95	707.39	721.49	16.88	2.61	1.06	1.64	0.87 (2)	0.88 (117)	0.87 (133)
4	LPBOC LSIN LSHS	B7	26.83	0.92	829.59	843.70	27.83	3.29	1.98	5.83	0.60 (2700)	0.63 (2853)	0.62 (2848)
5	LPBIC LSIN LCB	B6	28.70	0.91	846.51	860.61	29.92	3.25	1.03	1.40	0.69 (1147)	0.67 (2422)	0.66 (2443)
6	LPBOC LSLMAX LCB LRM	C11	31.43	0.91	870.32	887.95	32.82	3.38	4.09	15.72	0.53 (4143)	0.57 (3994)	0.56 (3987)
7	LPBIC LSLMAX LCB LRM	C11	31.74	0.91	872.80	890.43	33.11	3.39	4.24	16.27	0.53 (4247)	0.57 (4060)	0.55 (4050)
8	LPBOC LSCMAX	C14	34.16	0.90	889.31	899.89	34.95	3.61	1.03	1.44	0.64 (2030)	0.61 (3157)	0.60 (3194)
9	LPBOC LSLMAX	C14	34.20	0.90	889.54	900.11	34.98	3.61	2.04	6.00	0.54 (3851)	0.56 (4173)	0.55 (4202)

Numbers in brackets besides the fitting punctuation indicate the rank of the model according to the punctuation algorithm; MSE, mean squared error; CV, coefficient of variation; AdjRSQ, adjusted R-squared; AIC, Akaike's Information Criterion; SBC, Schwarz's Bayesian Criterion; PRESS, sum of squared prediction residuals; APRES, sum of absolute prediction residuals; VIF, variance inflaction factors; NCON, condition number.

Table 9
F of Snedecor value and probability in the contrast: H_0: $a = \alpha_1 = \alpha_2 = \cdots = \alpha_i = 0$ against H_1: any a or $\alpha \neq 0$ in the analysis of the influence of stripping rotation on cork weight

No.	Independent variables	F-value	Probability ($F > F$-value)
1	LKGM2 LCTB LRM LSS	9.85	0.0001
2	LPBIC LKGM2 LSIN LCTB	10.95	0.0001
3	LPBIC LSIN LCTB	6.51	0.0001
4	LPBOC LSIN LSHS	2.22	0.0669
5	LPBIC LSIN LCB	2.35	0.0551
6	LPBOC LSLMAX LCB LRM	2.60	0.0259
7	LPBIC LSLMAX LCB LRM	2.62	0.0248
8	LPBOC LSCMAX	0.57	0.6363
9	LPBOC LSLMAX	0.58	0.6309

Degrees of freedom are $(v + 1, n - 2v - 2)$ with v, number of variables in the model and n, number of observations. Letter L before the variables indicates that the variable has been log transformed.

Table 10
Sub-groups of variables in which null hypothesis H_0: coefficient a and αs corresponding to the shown variables = 0 is not rejected in those models that, considering all variables, null hypothesis is rejected (Table 9)

No.	Group of variables in which H_0 is not rejected	F-value	Probability ($F > F$-value)	Variables with significant αs
1	DKGM2 DRM DSS	1.79	0.1322	DCTB
2	DPBIC DKGM2	0.26	0.8524	DSIN DCTB
3	DPBIC DSIN	2.15	0.0935	DCTB
6	DPBOC DSLMAX	2.02	0.1109	DCB DRM
7	DPBIC DSLMAX	1.68	0.1713	DCB DRM

Variable DKGM2, as example, indicates the variable log (D_{KGM2} KGM2), with D_{KGM2} = 1/KGM2 if stripping rotation is 9 years and 1 if stripping rotation is 10 years.

changes. The variables whose estimation is affected by this change is the cork thickness measured in a sample before boiling (CTB).

4.3. Final estimation with mixed model structure

The final fitting of the nine models as mixed models led to the results of Table 11. The variance component at tree level represents the highest component of variation (between 78% and 91% of total variance), followed by regional effects component (5–17%). Plot random effect accounts only for 1.1–3% of total variation.

The improvement in the more complex models (the smaller the value of −2 times Res Log Likelihood the better; the larger the values of AIC and SBC, the better) in which more variables are measured is due to

Table 11
Fitting criteria and variance components in the nine mixed models

Model	Class	−2 Res Log Likelihood	AIC	SBC	Variance components			
					Region (σ_r^2)	Plot (σ_D^2)	Tree (σ_e^2)	Total
1	A1	−316.090	155.045	149.793	0.003159	0.000676	0.013971	0.017807
2	A2	−315.284	154.642	149.390	0.002971	0.000285	0.014190	0.017447
3	A4	−278.743	136.371	131.113	0.003018	0.000633	0.016584	0.020236
4	B7	−201.153	97.576	92.312	0.003621	0.000304	0.023557	0.027483
5	B6	−205.045	99.522	94.258	0.003931	0.000567	0.022995	0.027493
6	C11	−177.064	85.532	80.274	0.001440	0.000843	0.025805	0.028089
7	C11	−176.851	85.425	80.167	0.001820	0.001069	0.025648	0.028538
8	C14	−170.033	82.016	76.746	0.002344	0.000935	0.027037	0.030316
9	C14	−169.961	81.980	76.710	0.002315	0.000967	0.027037	0.030319

Table 12
Estimation of fixed effects and prediction of regional random effects in the nine selected mixed models

	Model number								
	1	2	3	4	5	6	7	8	9
Variables									
Intercept	0.050	-1.859^{***}	-1.915^{***}	−0.076	-0.898^{**}	1.269^{**}	0.984^{**}	1.992^{***}	1.992^{***}
PBIC		1.940^{***}	2.003^{***}		2.018^{***}		1.020^{***}		
PBOC				2.329^{***}		1.197^{***}		2.324^{***}	1.397^{***}
CB					0.305^{***}	0.153^{**}	0.302^{***}		
SLMAX						0.931^{***}	0.934^{***}		0.927^{***}
SCMAX								0.926^{***}	
SHS				0.144^{**}					
SS	0.889^{***}								
SIN		0.857^{***}	0.861^{***}	0.752^{***}	0.840^{***}				
RM	0.173^{***}					0.185^{**}	0.178^{*}		
CTB	0.332^{***}	0.348^{***}	−0.019						
KGM2	0.411^{***}	0.386^{***}							
DCTB	-0.026^{*}	-0.027^{***}							
Regional random effects									
Azaruja	0.020	0.026	0.027	−0.019	−0.019	−0.013	−0.014	−0.032	−0.032
Escoural I	−0.063	−0.075	−0.079	−0.088	−0.088	−0.039	−0.043	−0.048	−0.048
Escoural II	−0.038	−0.021	−0.008	0.036	0.038	0.011	0.009	0.032	0.032
Porto Alto	0.033	0.035	0.021	0.011	0.024	0.002	−0.001	0.004	0.004
Alportel	0.048	0.036	0.039	0.059	0.045	0.039	0.049	0.044	0.043

All variables are log transformed.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

a reduction of the error component at tree level, because regional and plot-variance components increase in some of them. As it is logical, the more precise measurements at tree level affect mainly the variance component at the same hierarchical level. The fitting criteria (−2 times Res Log Likelihood, AIC, SBC) also follow a logical performance; more complex models show better values and differences inside one class are usually of little importance.

Table 12 indicates the estimation of fixed effects and the prediction of regional random effects for the finally nine selected models.

5. Discussion and conclusions

5.1. Dendrometric covariates included in models

Cork production at tree level depends strongly on two sources of variation: tree size and the management criteria. In cork oak forests, management has an extraordinary importance, since increasing or reducing the stripped surface of the tree, without further consideration, can easily change the amount of cork that a certain tree produces. Hence, the impact of management decisions at tree level is much more important than in species, where wood volume estimation is the main goal. The independent variables should, therefore, reflect the management criteria. The selected nine models support this reasoning; no model includes only variables related to tree size and there is always at least a variable related to stripping. Also, the variable that absorbs most of the variability (highest type III sum of squares, data not shown) in all the models in which it appears is the stripping surface (SS); this is a variable that reflects both the tree size and the stripping management. In models in which there are separate variables to express tree size (i.e. perimeter at breast height under or over cork) and stripping management (i.e. the maximum stripping length), more variability is absorbed by the last ones.

5.2. Sources of error in cork oak weight estimation

The cork weight of a tree can be calculated considering the following identities (modified from Montero (1987):

$$W = \text{SSe} \cdot \text{CTh} \cdot \text{DENS} = \frac{\text{PUC}^2}{4} \cdot \text{SINe} \cdot \text{CTh} \cdot \text{DENS} = \text{SSe} \cdot \text{WS} = \frac{\text{PUC}^2}{4} \cdot \text{SINe} \cdot \text{WS} \quad (1)$$

where W is cork weight (fundamental unit M), SSe the exact stripping surface (L^2), CTh the cork thickness (L), DENS the cork density (ML^{-3}), PUC the perimeter under cork at 1.3 m (L), SINe the stripping intensity (adimensional) and WS is the cork weight per unit of stripped surface (ML^{-2}). The error in the estimation of cork weight at tree level can originate from different sources:

- Error in the estimation of SSe, CTh, DENS or WS. The problem is focused on A and B models, that include at least one of these kind of variables. The error in the estimation of the SSe through the measured stripped surface (SS) results mainly from the approximation of the tree stripped surface to a sum of cylinders. When irregularities in stem and branches and number of stripped branches increase, the error in the estimation also increases. In the case of the estimation of CTh, DENS or WS through variables like cork thickness, cork density and cork surface density (CTB, KGM3, KGM2), the source of error is that these variables are measured at the same height (samples are taken at 1.3 m), but their value is not constant along the tree; cork thickness, for example, decreases from the lower to the upper part of the stem and branches (Montero and Vallejo, 1992). Knowledge about the distribution profile of cork thickness, density and surface density along the tree could help to calculate the stem height, where the sample should be collected to increase the precision in estimation. After SS and SIN, the variable that explains more variability is CTB so this could improve the precision in the estimation of cork weight in research studies.
- Error due to the estimation of parameters in identity (1) with simpler variables. That is the case of the selected C models (6–9), in which stripping surface is approached through the joint measurement of variables like perimeter at breast height (PBOC or PBIC) and maximum stripping length (SLMAX), or maximum stripping coefficient (SCMAX). SLMAX and SCMAX are selected instead of more complex variables like total stripping length (SLTOT) or coefficient (SCTOT), or mean stripping length (SLMEAN) or coefficient (SCTOT). However, this depends on the tree characteristics of the sample; in this case most of the trees (63%) are only stripped in the stem and the differences between these variables will be higher when more branches are stripped.
- Error due to not considering one or several variables included in Eq. (1). This is the case of the A model number three and of all B and C models that are clearly sub-specified. This is the main source of error, as could be seen in the large variation of −2 times Res log Likelihood, SBC and AIC values when moving from types A to B and C models.

5.3. Functional form

Most of the A models can be recognized in Eq. (1). As an example, model 2 is close to $W = \frac{\text{PUC}^2}{4} \cdot \text{SINe} \cdot \text{CTh} \cdot \text{DENS}$. As we estimate the parameters with error, the cork weight of a certain tree will be $W_j = a\,\text{PBIC}^b\,\text{SIN}^c\,\text{CTB}^d\,\text{KGM3}^e\,e_j$, with ej the error term; so, there is an allometric relationship between the independent variables and the response variable. The use of this functional form in cork weight modelling has been scarce – only Gomes et al. (1990) and Ribeiro and Tomé (2002) – and much more attention has been paid to "pure" linear models.

Strict linear models show strong heteroskedasticity: when tree size increases the error in the estimation of the stripped surface of the tree also increases, especially if estimated through simple measurements and variability in stripping intensity and number of stripped boughs is larger in bigger trees. This indicates that heteroskedasticity is more a rule than an exception. Hence, the use of strictly linear models with parameter estimation trough OLS should be an exception and only applicable when the cork-stripping management is homogeneous and the population is near to regular.

5.4. Influence of stripping rotation

The introduction of variables related to cork thickness (CB, CTA or CTB), cork density (KGM3) or surface density of cork (KGM2) in some models could lead to think that, in such cases, there is no need to take into account the stripping rotation, as it will be reflected in the value of these variables. Nevertheless, these values are always gathered from a sample taken at 1.3 m, but the distribution of the value of such variables is not uniform along the stem, as previously discussed. Influence of stripping rotation on cork weight at tree level is reflected in the introduction of dummy variables related with the cork thickness (DCTB, in models 1–3) measured in the cork sample. All the models that introduce these variables are correctly specified; that is, they include all the variables of Eq. (1). Hence, these dummy variables are expressing the change in the distribution profile of the variables CTB from the lower to the upper part of the tree when the cork-stripping operation is delayed 1 year (from 9 to 10 years). With information about the distribution profile of these variables and its change along the years, the height at which the cork sample should be obtained could be modified and this correction would not be necessary.

The other models do not include any correction due to stripping rotation, because they are sub-specified and do not include any variable in the fixed part of the model with capability to absorb this effect. In these cases (B and C models), the effect of stripping rotation is included in the error term.

5.5. Covariance structure

The mixed model structure allows to distinguish random component at regional, plot and tree level and to adopt a more realistic variance–covariance structure in the case of nested sampling designs. When trees are grouped in plots, the hypothesis of independence of residuals and the corresponding diagonal variance–covariance matrix should not be a prior.

In the cork weight models that have been analysed, most of variability is focused at tree level (80–90%) and just 1–3% at plot level. The highest tree level variability is in accordance with the fact that cork oak is a heterozigotic species and shows a high inter-tree variability in other aspects like cork quality (Macedo et al., 1998) or cork moisture (Costa, 1997). The low value of the variance component at plot level could result from the few number of plots per region and their deficient spatial distribution inside the region. Also, the regional division of Portugal focused on cork weight analysis is not strictly defined and a delimitation based on ecological classification would lead probably to a small number of regions with specific locations. In that case, region should be considered as fixed and not a random effect. If plot-level variance obtained in future works designed with a better distribution and higher number of plots per region is similar, a final structure of the model could be a fixed effect model considering only a regional fixed effect.

In conclusion, nine models, five for research purposes and four for management and practical purposes, were selected to predict oven-dried cork weight in Portugal. All models include at least one variable related with the cork-stripping operation, i.e. stripping surface, stripping intensity, stripping coefficient or stripping length, and they are those that absorb a higher variability. The models with higher precision include also a variable measured in a cork sample, i.e. cork thickness or cork surface density. The effect of the cork cycle (9 or 10 years) on cork production can only be noticed in those models that include variables with a higher measurement cost, such as cork thickness.

6. Model application: inference at different spatial levels and comparison with OLS estimation

In this point, application of the mixed model 9 (one of the most useful for management purposes) will be developed for particular values of the independent variables. The general model structure is $\boldsymbol{y} = \boldsymbol{Xa} + \boldsymbol{Zb} + \boldsymbol{e}$. The response variable is LDW (log transformation of tree oven-dried cork weight in kilograms). The covariates are: PBOC (log transformation of perimeter at breast height over cork in metres) and SLMAX (log transformation of maximum stripping length in metres). The matrix of fixed effects $\boldsymbol{X}$ is then: $\boldsymbol{X} = [\mathbf{1\ PBOC\ SLMAX}]_{251\times 3}$. REML estimation of variance components, needed for $\boldsymbol{V}$ and fixed effects estimation and prediction of random effects is (Table 11):

$$\hat{\sigma}_r^2 = 0.0023157; \quad \hat{\sigma}_p^2 = 0.00096690;$$
$$\hat{\sigma}_e^2 = 0.02703706$$

Solution of mixed model equations gives the estimation of fixed effects, $\hat{\boldsymbol{a}}^o$:

$$\hat{\boldsymbol{a}}^o = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}) - \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} = \begin{bmatrix} 1.9918 \\ 1.3968 \\ 0.6267 \end{bmatrix}$$

and the prediction of random effects $\hat{\boldsymbol{b}}^o$:

$$\hat{\boldsymbol{b}}^o = \hat{\boldsymbol{G}}Z'\hat{\boldsymbol{V}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{a}}^o) = \begin{bmatrix} \hat{r}_{\text{Azaruja}} \\ \hat{r}_{\text{Escoural I}} \\ \hat{r}_{\text{Escoural II}} \\ \hat{r}_{\text{Porto Alto}} \\ \hat{r}_{\text{Alportel}} \\ \hat{p}_1 \\ \hat{p}_2 \\ \hat{p}_3 \\ \hat{p}_4 \\ \hat{p}_5 \\ \hat{p}_6 \\ \hat{p}_7 \\ \hat{p}_8 \\ \hat{p}_9 \\ \hat{p}_{10} \\ \hat{p}_{11} \\ \hat{p}_{12} \end{bmatrix} = \begin{bmatrix} -0.032 \\ -0.048 \\ 0.032 \\ 0.004 \\ 0.043 \\ -0.024 \\ 0.029 \\ -0.018 \\ -0.012 \\ -0.012 \\ 0.004 \\ 0.025 \\ -0.003 \\ -0.007 \\ 0.004 \\ -0.002 \\ 0.018 \end{bmatrix}_{17\times 1}$$

According to the information available of random effects, mixed models can be used to make estimations or predictions at different inference spaces. As an example, if we are interested in the estimation of oven-dried cork weight in a tree with a perimeter at breast height over cork = 1.35 m and a maximum stripping length = 3.0 m (log transformed values: PBOC = 0.3001 and SLMAX = 1.0986) and we do not have information about its spatial situation, or we are interested in the mean value of cork weight for the population of trees with these characteristics, then we are interested in the unconditional expectation: $\boldsymbol{E}(\boldsymbol{y}) = \boldsymbol{E}(\boldsymbol{Xa} + \boldsymbol{Zb} + \boldsymbol{e}) = \boldsymbol{Xa}$. This is the broad inference space (McLean et al., 1991) and estimation includes only the fixed part of the model. The best linear unbiased estimator (BLUE) of a lineal combination of $\boldsymbol{a}$, $\boldsymbol{K}'\boldsymbol{a}$, if estimable, is $\boldsymbol{K}'\hat{\boldsymbol{a}}^0$ (Henderson, 1984). In this case:

$$\boldsymbol{K}' = [1,\ 0.3001,\ 1.0986]; \qquad \boldsymbol{a}^o = \begin{bmatrix} 1.9918 \\ 1.3968 \\ 0.9267 \end{bmatrix}$$

and BLUE is 3.4291 = LDW.

If we have additional information about the spatial situation of the tree, a better estimation of cork weight can be obtained. If the tree is in region 1 (Azaruja) we are interested in the conditional expectation for a given value of the random regional effects, that is $\boldsymbol{E}(\boldsymbol{y}|\boldsymbol{b}) = \boldsymbol{E}(\boldsymbol{Xa} + \boldsymbol{Zb} + \boldsymbol{e}) = \boldsymbol{Xa} + \boldsymbol{Zb}$. In this case, the best lineal unbiased predictor (BLUP) of a lineal combination of fixed and random effects ($\boldsymbol{K}'\boldsymbol{a} + \boldsymbol{M}'\boldsymbol{b}$), is $\boldsymbol{K}'\hat{\boldsymbol{a}}^o + \boldsymbol{M}'\hat{\boldsymbol{b}}^o$ (Littell et al., 1996), with $\boldsymbol{K}'$ = [1 0.3001 1.0986] and $\boldsymbol{M}'$ = [1 0 0 0 0|0 0 0 0 0 0 0 0 0 0 0 0] and BLUP is 3.3968 = LDW. If we know additionally that tree is in plot 1 (in Azaruja region) then $\boldsymbol{M}'$ = [1 0 0 0 0|1 0 0 0 0 0 0 0 0 0 0 0] and BLUP is 3.3726 = LDW.

To obtain unbiased estimations in original units the estimation error is needed. Following the mixed model equations (Henderson, 1984):

$$\begin{bmatrix} \boldsymbol{X}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{X} & \boldsymbol{X}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{Z} + \hat{\boldsymbol{G}}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}^o \\ \boldsymbol{b}^o \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{y} \\ \boldsymbol{X}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{y} \end{bmatrix}$$

the variance–covariance matrix of $(\hat{\boldsymbol{a}}^o - \boldsymbol{a},\ \boldsymbol{b}^o - \boldsymbol{b})$ is:

$$\hat{\boldsymbol{C}} = \begin{bmatrix} \boldsymbol{X}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{X} & \boldsymbol{X}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{Z} + \hat{\boldsymbol{G}}^{-1} \end{bmatrix}^{-}$$

and the variance of a lineal combination

$$\boldsymbol{K}'\hat{\boldsymbol{a}}^o + \boldsymbol{M}'\boldsymbol{b}^o = \boldsymbol{L}\begin{bmatrix} \boldsymbol{a}^o \\ \boldsymbol{b}^o \end{bmatrix}$$

is:

$$\text{var}\left(\boldsymbol{L}\begin{bmatrix} \boldsymbol{a}^o \\ \boldsymbol{b}^o \end{bmatrix}\right) = \boldsymbol{L}\hat{\boldsymbol{C}}\boldsymbol{L}'$$

Table 13 shows the estimated values and variance of oven-dried cork weight in logarithmic units for the

Table 13
Estimated values and variance of oven dried cork weight in logarithmic units (LDW) and arithmetic units (DW) with prediction intervals at different inference levels for a tree with PBOC = 1.35 m and SLMAX = 3.0 m

	Tree situation		Estimation (log units)		Estimation (arithmetic units)			
	Region	Plot	LDW	var (LDW)	DW (kg)	95% prediction intervals		
						Lower limit (kg)	Upper limit (kg)	Range (kg)
Mixed model	–	–	3.4291	0.0007459	30.9	28.7	33.2	4.5
	Azaruja	–	3.3968	0.0006278	29.9	28.3	31.6	3.3
	Azaruja	1	3.3726	0.0007587	29.2	27.5	30.9	3.4
OLS	–	–	3.4238	0.0001871	30.7	29.9	31.5	1.6

Comparison between mixed model approach and ordinary least-square (OLS) estimation.

different inference levels. Prediction intervals can be calculated as:

$$\boldsymbol{L}\begin{bmatrix}\boldsymbol{a}^{\mathrm{o}}\\ \boldsymbol{b}^{\mathrm{o}}\end{bmatrix} \pm t_{\hat{\nu},\alpha/2}\sqrt{\boldsymbol{L}\hat{\boldsymbol{C}}\boldsymbol{L}}$$

with $\hat{\nu}$ degrees of freedom (according to Satterthwaite approximation). To obtain unbiased estimations in original units, the following transformation is needed:

$$\mathrm{DW} = \exp\left\{\log(\mathrm{DW}) + \tfrac{1}{2}\mathrm{var}(\log(\mathrm{DW})\right\}$$

Unbiased estimations in original units are shown in Table 13.

If we consider now the model as a fixed effects model without considering random effects, the structure will be: $\boldsymbol{y} = \boldsymbol{Xa} + \boldsymbol{e}$ with $\boldsymbol{X} = [\mathbf{1}\ \mathbf{LPOBC}\ \mathbf{LSLMAX}]_{251\times3}$. If estimation is made with OLS with the usual assumptions, then var ($\boldsymbol{y}$) = var ($\boldsymbol{e}$) = $\boldsymbol{R} = \sigma_e^2 \boldsymbol{I}_N$ and estimation of $\hat{\boldsymbol{a}}^{\mathrm{o}}$:

$$\hat{\boldsymbol{a}}^{\mathrm{o}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}^{-1}\boldsymbol{y} = \begin{bmatrix}2.0205\\ 1.3913\\ 0.8935\end{bmatrix}$$

The residual variance estimation is $\hat{\sigma}_e^2 = 0.02955$ and the BLUE of oven-dried cork weight in logarithmic units is LDW = 3.4238. We can construct also confidence intervals and obtain unbiased estimations in original units. Table 13 shows these values and allow to compare the OLS estimation with the estimation considering a mixed model structure in the case of broad inference space. It can be noticed that a fixed effects approach underestimates the error in cork weight estimation.

Acknowledgements

This work is part of the investigation supported by projects AIR3-CT92-0135 (''Silvicultural Management of Cork Oak Stands Towards Improved Cork Production and Quality''), FAIR3-CT96-1438 (''CORKASSES'' Field Assessment and Modelling of Cork Production and Quality''), PRAXIS XXI (''Conhecimento fundamental da cortiça e detecção precoce da qualidade da cortiça'') and a Grant of Program PG-ING94 from the Spanish Ministry of Science for the first author.

References

Alves, A.M., 1958. Tabelas de previsão de peso de cortiça para o sobreiro nos xistos do carbónico. Boletim da Junta Nacional da Cortiça 237, 233–243.

Alves, A.M., Macedo, A.A.M., 1961. Sobre tabelas de peso de cortiça. Boletim da Junta Nacional da Cortica 23 (275), 165–168.

CESE, 1996. O sector florestal português. Conselho Ensino Superior Empresa. Ministerio da Educação.

Costa, A., 1992. Modelos de previsão de produção de cortiça. 2° Encontro dos montados de sobro e azinho. Evora 150–159.

Costa, A., 1993. Avaliação da superficie geradora do sobreiro. Unpublished.

Costa, A., 1997. Sobreiro. Caracterização do crescimento e avaliação da produção de cortiça. M.Sc. Thesis. Instituto Superior de Agronomia, Lisboa, Portugal.

Costa, A., Oliveira, A.M.C., 1998. Estudo da variação da produção de cortiça por árvore individual entre duas tiragens consecutivas em parcelas permanentes. In: Pereira, H. (Ed.), Cork Oak and Oak (Sobreiro e cortiça). European Conference on cork oak and oak. 5–7 May 1997, Centro de Estudos Florestais, Lisboa, Portugal, pp. 54–59.

DGF, 2001. Direcção Geral das Florestas website: www.dgf.min-agricultura.pt.

Ferreira, M.C., Oliveira, A.M.C., 1991. Modelling cork oak production in Portugal. Agroforestry Syst. 16, 41–54.

Ferreira, M.C., Tomé, M., Oliveira, A.M.C., Alves, A.M., 1986. Selecção de modelos de previsão de peso de cortiça. 1° Encontro sobre os montados de sobro e azinho. Evora 65–83.

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

Fonseca, T., Parresol, B.R., 2001. A new model for cork weight estimation in Northern Portugal with methodology for construction of confidence intervals. For. Ecol. Manage. 152, 131–139.

Gomes, A.A., Amorin, I.E., Bernardo, A., 1990. Análise comparativa da produção em peso de cortiça em três unidades-solo, na mata de Valverde-Grándola. II Congresso Florestal Nacional. Sociedade Portuguesa de Ciências Florestais. Porto 350–359.

González-Adrados, J.R., Calvo, R.M., 1994. Variación de la humedad de equilibrio del corcho en plancha con la humedad relativa. Modelos de regresión no lineal para las isotermas de adsorción. Investigación Agraria. Sistemas y Recursos Forestales 3 (2), 199–209.

González-Adrados, J.R., Montero, G., Ortega, C., 1993. Caracterización productiva de los alcornocales catalanes. Investigación Agraria. Sistemas y Recursos Forestales 2 (1), 55–69.

Guerreiro, M.G., 1951. Previsão do peso da cortiça explorável de um montado. Publicações da Direcção Geral dos Serviços Florestais e Aquicolas 18 (1), 57–83.

Henderson, C.R., 1984. Applications of Linear Models in Animal Breeding. University of Guelp, Canada.

Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., 1996. SAS System for Mixed Models. SAS Institute, Cary, USA.

Macedo, J, Lopes, F., Pereira, H., 1998. Influencia do metodo de amostragem na avaliacao da qualidade da cortiça no mato. In: Pereira, H. (Ed.), Cork Oak and Cork (Sobreiro e cortiça). European Conference, 5–7 May 1997, pp. 93–98.

McLean, R.A., Sanders, W.L., Stroup, W.W., 1991. A unified approach to mixed linear models. American Statistician 45, 54–64.

Montero, G., 1987. Modelos para cuantificar la producción de corcho en alcornocales (*Quercus suber* L.) en función de la calidad de estación y los tratamientos selvícolas. Ph.D. Thesis. INIA, Madrid.

Montero, G., Vallejo, R., 1992. Variación del calibre del corcho medido a distintas Alturas. Investigación Agraria. Sistemas y Recursos Forestales 1 (2), 182–188.

Myers, R., 1986. Classical and Modern Regression with Applications. Duxbury Press, Boston.

Natividade, J.V., 1950. Subericultura. Direcção Geral dos Serviços Florestais e Aquicolas. Lisboa.

Ribeiro, F., Tomé, M., 2002. Cork weight prediction at tree level. For. Ecol. Manage. 171, 231–241.

Ribeiro, N.A., 1990. Comparação de métodos de amostragem para a avaliação do potencial produtivo dos montados de sobro da Celbi. Relatório final de Estágio de licenciatura em engenharia florestal. Universidade Técnica do Alto Douro, Vila-Real, Portugal.

Ribeiro, N.A., 1992. Caracterização do potencial produtivo de uma área de sobro na regiao de Coruche. 2° Encontro dos montados de sobro e azinho. Evora 134–149.

Searle, S.R., 1971. Linear Models. John Wiley and Sons, New York.

Shapiro, S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591–611.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.