

## Predicting site index from climate and soil variables for cork oak (*Quercus suber* L.) stands in Portugal

Joana Amaral Paulo · João H. N. Palma · Alberto Azevedo Gomes ·  
Sónia Pacheco Faias · José Tomé · Margarida Tomé

Received: 3 April 2014 / Accepted: 20 October 2014 / Published online: 26 October 2014  
© Springer Science+Business Media Dordrecht 2014

**Abstract** Site productivity, assessed through site index, was modelled using partial least squares regression as a function of soil and climatic variables. Two alternative models were developed: a full model, considering all available explanatory variables, and a reduced model, considering only variables that can be obtained without digging a soil pit. The reduced model was used for mapping the site index distribution in Portugal, on the basis of existing digital cartography available for the whole country. The developed models indicate the importance of water availability and soil water holding capacity for site index value distribution. Site index was related to climate, namely evaporation and frost, and soil characteristics such as lithology, soil texture, soil depth, thickness of the A horizon and soil classification. The variability of the estimated values within the map (9.5–16.8 m with an average value of 13.4 m) reflects the impact of soil characteristics on the site productivity estimation. These variables should be taken into consideration during the establishment of new plantations of cork oak, and management of existing plantations. Results confirm the potential distribution of cork oak in coastal regions. They also suggest the existence of a considerable area, located both North and South of the Tagus river, where site indices values of medium ([13;15]) to high ([15;17]) productivity classes may be expected. The species is then expected to be able to have good productivity along the northern coastal areas of Portugal, where presently it is not a common species but where, according to historical records, it occurred until the middle of the sixteenth century. The present research focused on tree growth. Cork growth and cork quality distribution needs to be further researched through the establishment of long term experimental sites along the

---

J. A. Paulo (✉) · J. H. N. Palma · S. P. Faias · J. Tomé · M. Tomé  
Forest Ecosystem Management under Global Change Research Group (ForChange), Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisbon, Portugal  
e-mail: joanaap@isa.ulisboa.pt

A. A. Gomes  
Instituto Nacional de Investigação Agrária e Veterinária, Quinta do Marquês, 2780-159 Oeiras, Portugal

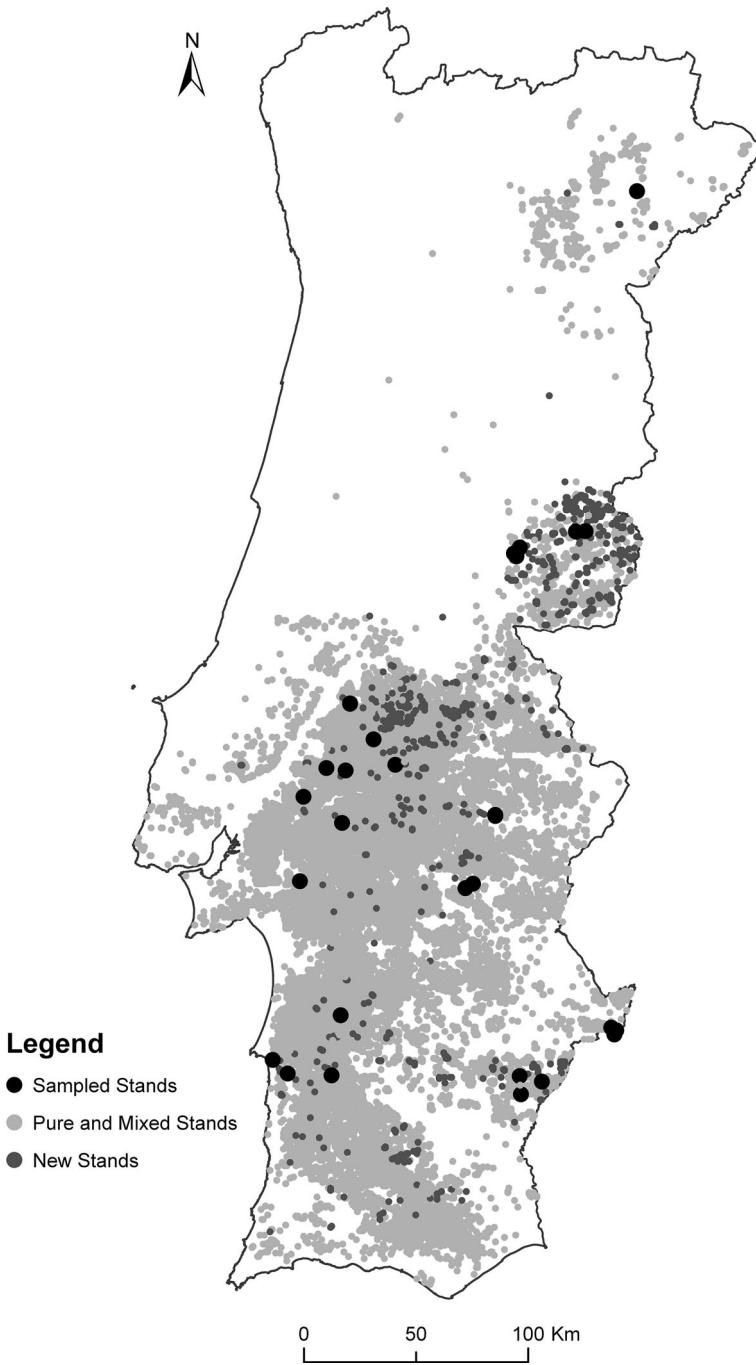
distribution area of cork oak, namely in the central and northern coastal areas of the country.

**Keywords** *Quercus suber* L. · Cork oak · Site index · Site quality · Partial least squares · Potential distribution

## Introduction

Cork oak forests in Portugal have high economic and social importance. This species is also a key element in a multifunctional agroforestry system (*montado*), which usually combines the production of several non-wood forest products (e.g. cork, pasture, livestock), and is important for its biodiversity, water retention and soil conservation capacities. Recently its importance for carbon stocking, both in existing stands (Paulo et al. 2013) and new plantations (Coelho et al. 2012) has been recognized. According to the last Portuguese national forest inventory (AFN 2010), 715,922 ha are currently occupied by cork oak forests/agroforests in Portugal (23 % of the total forest area in the country), mainly in the South of the country (Fig. 1). Historical records show, however, that this distribution is smaller than the potential distribution area in the country, since until the middle of the sixteenth century cork oak was distributed in a wider area in Portugal (Costa 1997), and the area of distribution of this species (APA 1984; Correia and Oliveira 2003) is larger than the area where it is presently found. The reduction of the distribution area of this species, related to a complex interaction of several anthropogenic and environmental variables, has been recently studied and discussed by several authors, using different methodologies, in several Mediterranean countries such as Portugal (Coelho et al. 2012), Italy (Vessella and Schirone 2013) and Spain (e.g. Lacambra et al. 2010). In Portugal, areas of new cork oak plantations have been established in recent decades. In addition to plantings in forest areas, afforestation of agricultural land has also been relevant since specific measures of financial support have been established for this proposal. New cork oak plantations are established not only in the traditional area of distribution of the species, but also in the northern area of the country (Fig. 1) where historical records show that the species was formerly present (Costa 1997) and where climatic thresholds suggest a possible occurrence. This has also been recently demonstrated by research in other Mediterranean countries such as Spain (Lacambra et al. 2010) and Italy (Vessella and Schirone 2013). As a result, a total of 46,815 ha of cork oak stands are identified as recent cork oak plantations (<20 years old) in the last Portuguese national forest inventory (AFN 2010).

The study of the importance of recent plantations for carbon sequestration and cork production is an important subject, requiring the use of forest growth models implemented as simulators, such as the one developed by Coelho et al. (2012). Different site productivity evaluation methods may be used by these growth models. Site index is one of these methods, used both for species usually pruned to a specific form of stem, or other species where this operation is not required. For example, site index is used in the individual tree diameter models developed by Tomé et al. (2006) and Sánchez-González et al. (2006) for even-aged cork oak stands in Portugal and Spain, respectively, and by Gea-Izquierdo et al. (2008) for *Quercus ilex* in Iberian open oak woodlands. Site index (SI) refers to the dominant height of a stand at some specified reference age or base age, commonly selected to lie close to the average rotation age. One of the reasons for the frequent use of this



**Fig. 1** Location of sampled stands included in the data set (*black*), along the current distribution of cork oak stands (*light grey*), represented by the location of the National Forest Inventory (NFI) photo points (500 m × 500 m grid) with the presence of cork oak. NFI plots identified as new cork oak plantations are shown in *dark grey*

method is the fact that when site index curves exist for a given species, the determination of site index only requires one measurement of some of the trees in the stand, taken at any known age (e.g. Bravo-Oviedo et al. 2004; Sánchez-González et al. 2005). According to Clutter et al. (1983) classification, site index curves are considered a direct method for site index evaluation. When the stand age is not known, or when tree measurements are not available for the species, indirect methods of evaluating the site index from climatic and/or soil variables are needed (Clutter et al. 1983). Several statistical approaches have been used and compared (Aertsens et al. 2010) for the development of these indirect methods, such as multiple regression (e.g. Corona et al. 1998; Fontes et al. 2003), spline smoothing functions (McKenney and Pedlar 2003), tree based regression (McKenney and Pedlar 2003), contingency tables and correspondence analysis (Bravo-Oviedo and Montero 2005), regression kriging or partial least squares (Palmer et al. 2009).

Partial least squares (PLS) is an advanced multivariate statistical approach that allows a set of response variables to be predicted from a set of predictors. It has proven to be a powerful tool for versatile data analysis, to simplify data and to analyse the correlation between groups of variables at the same time. It has been largely used in several scientific areas such as remote sensing (e.g. Bikindou et al. 2012), wood technology (e.g. Alves et al. 2012), chemometrics (e.g. Wold 1994) or ecology (Carrascal et al. 2009; Cheng-liang et al. 2012), in cases when data sets are characterized by a large number of correlated predictor variables and/or a relatively small number of observations. Compared with multiple linear regression, the predictor variables are replaced by a certain number of factors (also called components, latent vectors, or latent variables), estimated in order to simultaneously describe the variation of the space created by the prediction variables and the response variable. As a result, PLS regression can avoid collinearity problems frequently encountered when using ordinary least squares, by compressing these collinear variables into orthogonal and non-correlated factors. More details on this statistical methodology can be found, for instance, in Geladi and Kowalski (1986) or Abdi (2010).

The objective of the present work was to develop a model, based on climate and soil variables, to predict site index for cork oak stands in Portugal within the distribution area of the species. Additionally, the model was used to spatially determine the distribution of the site index value of cork oak in Portugal.

## Materials and methods

### Site index and predictor variables data collection

Data from 100 plots established in 42 even-aged cork oak stands of known age were used for this study. Due to the low density of trees typical in adult cork oak stands, the characterization of several stand variables demands the measurement of plots with areas larger than the ones typically used in forest inventories (Almeida and Tomé 2010; Tomé 2005). For this reason, plots with an area of 2827 m<sup>2</sup> were defined, either circular with 30 m of radius or rectangular with 53.2 m × 53.2 m. This set of plots was obtained from two different sources: 40 permanent research plots from ForChange research group established during several research projects, and 60 plots established for this research. Stands included in this last subset were selected in order to, in addition to the first one, best cover the current distribution of cork oak stands. Also, the location of the plots established in the same stand was defined in order to account for the variability of the stand due to topographic and soil features, and therefore different site productivity. The location of the

plots is shown in Fig. 1, as well as the current distribution of cork oak stands (young plantations and adult stands in different colours).

The site index for each sample plot was estimated using the model developed by Sánchez-González et al. (2005), for a base age of 80 years. The dominant height was calculated using the 25 trees per hectare with the largest diameter, instead of the usual number of 100 trees per hectare, because of the typically low density of cork oak stands that have an average value of 66 trees ha<sup>-1</sup> in pure cork oak stands according to the last Portuguese national forest inventory (AFN 2010). The estimated values for the site index of the sample plots varied from 10.0 to 18.4 m, with an average of 15.1 m.

In each plot, a soil pit was dug until the R/C or C/R horizon was reached. The depths of the pits varied between 45 cm and 200 cm. The following variables were used for this study (Table 1): FAO soil group according to the IUSS Working Group WRB (2006) classification, lithology according to Silva (1983) classification, coarse elements percentage, soil textural class (fine, medium and coarse), thickness of the A horizon and soil depth. The last five variables can only be obtained through observation of the soil profile. FAO soil group and lithology may be obtained through digital cartography such as the environmental atlas available at the Portuguese Agency for the Environment website (<http://sniamb.apambiente.pt/webatlas/>).

The slope, aspect and position in the slope of the sample plots (Table 1) were determined locally, but these characteristics can also be obtained with reasonable precision using available topographic maps (1:25,000) or a digital elevation model. The topographic wetness index developed by Bevan and Kirkby (1979) was computed for each plot (Table 1) using the Jarvis et al. (2008) digital terrain model and according to Sorensen et al. (2005).

Climatological variables recognized as relevant for cork oak environmental distribution were obtained from the 83 meteorological stations network of the Portuguese Meteorological Service (Table 1). The variables considered were computed from average monthly data over a 30-year period (1961–1990): mean minimum temperature (°C), mean temperature (°C), mean maximum temperature (°C), mean monthly precipitation (mm), mean number of days with precipitation per month, mean monthly evaporation measured with Piche evaporimeter (mm) and mean number of days with frost per month. The Martonne climatic index (De Martonne 1925), a measure of the aridity of region, was computed as followed at the plot level:  $M = P_m / (T_m + 10)$ , where  $M$  is the Martonne climatic index,  $P_m$  is mean annual precipitation in mm and  $T_m$  is mean annual temperature in °C.

## Modelling approach

Two alternative models were developed. Model 1 ( $M_1$ ), the full model, was developed by considering all the available explanatory variables as candidates: FAO soil group, soil lithology, soil coarse elements percentage, soil textural class, thickness of the A horizon, soil depth, slope, aspect and position in the slope of the plot, topographic wetness index, climatic variables and Martonne climatic index. Model 2 ( $M_2$ ), the reduced model, was developed by removing variables that can only be obtained through observation of the soil profiles, which required the digging of a soil pit (coarse elements percentage, textural class, thickness of the A horizon and soil depth). Climatic variables were candidate predictors for both alternative models.

Although the PLS method allows the development of a model that includes a very large number of independent variables, it is important to identify which predictor variables are most useful for predicting the response variable and which variables make small and uncertain contributions to the model. For this determination, we used the variable

**Table 1** Environmental and land variables used in the data analysis

Variable	Source	Minimum	Average	Maximum
FAO soil group (IUSS Working Group WRB 2006)	Soil pit analysis (M <sub>1</sub> ) and Portuguese Agency for the Environment digital cartography (M <sub>2</sub> )			
Lithology (Silva 1983)	Soil pit analysis (M <sub>1</sub> )			
Soil textural class				
Coarse elements percentage (% m <sup>-2</sup> )		0°	20	60
Thickness of the A horizon (cm)		10.0	21.5	40.0
Soil depth (cm)		1.5	84.1	160
Slope (°)	Determined locally (M <sub>1</sub> ) or with topographic maps 1:25,000 (M <sub>2</sub> )	0	5°	20°
Aspect				
Position in the slope				
Topographic wetness index	Computed according to Sorensen et al. (2005)	3.4	7.2	15.2
Martonne index	Computed according to De Martonne (1925)	19.2	24.9	39.1
Mean minimum temperature (°C)	National meteorological station network of the Portuguese Meteorological Service (83 stations)	7.9	9.8	13.2
Mean temperature (°C)		14.1	15.6	16.4
Mean maximum temperature (°C)		18.3	21.4	22.7
Mean monthly precipitation (mm)		41.1	52.6	78.6
Mean number of days with precipitation per month		5	8	10
Mean monthly evaporation (mm)		78.2	120.8	163.2
Mean number of days with frost per month		0	2	5

Climatic variables data set refer to mean values, computed using monthly averages of the 30-year period of 1961–1990

importance projection (VIP) statistic, which evaluates the contribution of each predictor variable to the PLS regression model for both predictor and response variables. Predictor variables that had relatively small absolute values for the regression coefficients, based on the centred and scaled predictors, and VIP values of  $<0.8$  were removed from the model (Wold 1994). The test proposed by van der Voet (1994) was applied to choose the number of factors. For each model, the number of factors chosen was the fewest that presented residuals that were not significantly larger (at 0.05 level) than the residuals of the model with the minimum predicted residual sum of squares statistic value. SI modelling was carried out using partial least squares (PLS) regression, using the PLS procedure of the SAS 9.2 software (SAS Institute Inc. 2000). Plots of the studentized residuals versus the predicted values and QQ-plots of the studentized residuals were used to look for inadequacies of the models and for outliers. Plots of the observed response values versus predicted values were used in the diagnostic review of the adequacy of the models.

Ideally, model validation should be carried out with a data set independent from the one used to fit the model. When such data are not available, resampling techniques such as boot-strapping and jackknife methods are used to mimic a validation data set (e.g., Davison and Hinkley 1997). In our case validation was made by withholding one plot at a time and fitting the model with the remaining observations. This guaranties that the observations used for the computation of the residuals are independent from the ones used to fit the model. The parameters resulting from this fit were used to estimate the site index value for the plots from the deleted stand. The resulting prediction residuals  $r_i$ , called PRESS residuals (e.g. Myers 1990), were computed as  $r_i = y_i - \text{yest}_i$ , where  $y_i$  and  $\text{yest}_i$  are the observed and the predicted site index values, respectively. The following validation statistics were used to evaluate and compare the models predictive performances: the mean of the press residuals ( $\bar{r}_i$ ) was used to evaluate bias, the mean of the absolute value of the press residuals ( $|\bar{r}_i|$ ) was used as an evaluation of model precision, the 5th and 95th percentile values of the press residuals (p5 and p95 respectively) were used to complement the assessment of model precision. In addition, model efficiency (ef), or the proportion of variation explained by the model, was computed using the press residuals as:

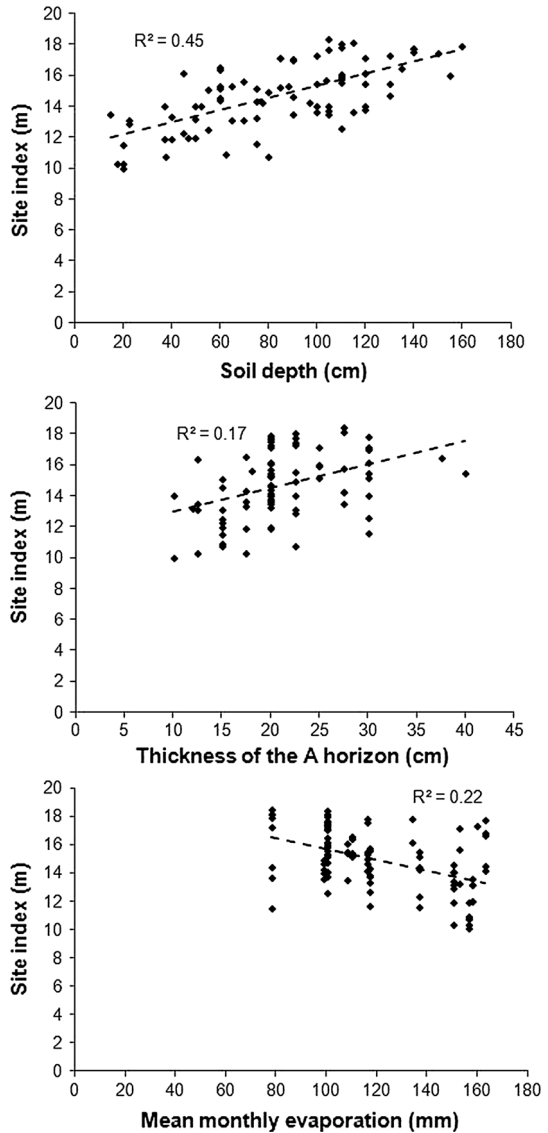
$$\text{ef} = 1 - \frac{\sum_{i=1}^n r_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

where  $n$  is the number of observations ( $n = 100$ ),  $\bar{y}$  is the mean value of the site index dependent variable, and other symbols are as previously defined.

### Mapping site index distribution for cork oak stands in Portugal

Model  $M_2$ , or reduced model, was used for mapping the site index values that can be expected along the distribution area of cork oak in Portugal. The map was restricted to areas where digital cartographic values from soil, lithology and climatological variables are included within the range of the data set used for fitting the model. A soil map (Cardoso et al. 1973) and a lithologic map (Silva 1983), available for the whole country, albeit at a relatively coarse scale (1:1.000.000), were used to gather data at the national level. The previously mentioned maps are available in a digital format at the Portuguese Agency for the Environment website (<http://sniamb.apambiente.pt/webatlas/>). Maps with 1 km resolution for mean monthly evaporation and mean number of days with frost per month were built with inverse distance weight on the 83 stations observed monthly data from 1961 to 1990 (Miranda et al. 2002). All the geographic datasets were converted to grids and the

**Fig. 2** Relation between site index and soil depth (cm), thickness of the A horizon (cm) and mean monthly evaporation (mm)



reduced model equation ( $M_2$ ) was formulated in Raster Calculator ArcGIS© (Spatial Analyst Extension), producing a map with the resulting site index.

## Results

### Model development

The relation between the site index value and some of the climatic and soil variables was evident by the linear relation they presented (Fig. 2), suggesting that these variables are relevant for the model development.



**Table 2** Percent variation accounted for by partial least squares factor in each of the fitted models

Factor	Full model ( $M_1$ )		Reduced ( $M_2$ )	
	Dependent variable		Dependent variable	
	Current	Total	Current	Total
1	61.1	61.1	41.7	41.7
2	7.4	68.5	6.5	48.3
3	1.5	70.0	0.4	48.7
4			0.4	49.1

By applying the van der Voet (1994) test, the final number of extracted factors for each of the models was three and four for the full ( $M_1$ ) and reduced ( $M_2$ ) models respectively. The percent variation accounted for by PLS factors of these two models on the dependent variable are presented in Table 2.

Regarding the full model (Table 3), the three extracted factors explained a total of 70.0 % of the site index variable variation. This value is mostly associated with the first factor, which alone is responsible for 61.1 % of the total variation, while the second and third factors are responsible only for 7.4 and 1.5 % respectively. Correlation loading plots for the two first factors (Fig. 3a) suggest a large and positive effect of an increase of soil depth and thickness of the A horizon, in opposition to a negative effect of mean monthly evaporation, on site index value. The first factor was also related to soil textural classes and soil groups. Coarse texture class is positively differentiated in opposition to fine texture class, while medium texture class is associated to a close to zero loading value. Also Arenosols and Leptosols are have an antagonist relation to site index, the first associated with a positive effect, while the second with a negative effect on the site index value.

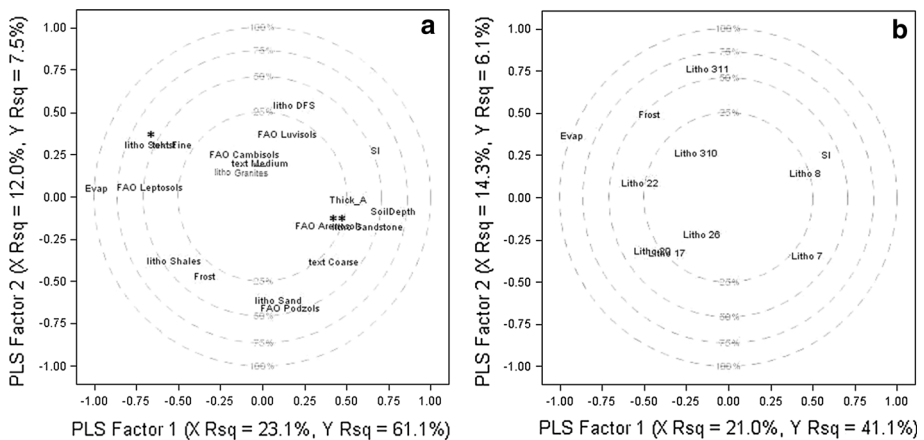
For the reduced model (Table 4), the total variation of the dependent variable is related more to the first factors (41.7 % of a total value of 49.1 % explained by the four factors). Correlation loading plots (Fig. 3b) showed that in the first factor, higher values of mean annual evaporation and mean annual number of days with frost are both related to lower site index values. This factor also suggests the differentiation between sands and sandstone soil lithology classes from shales and shists. The first two present positive loading values while the last two are associated with negative values. Other soil lithology classes present near to zero values of loading values from the first factor.

None of the models included climatic variables related to temperature, precipitation, Martonne's climatic index, topographic wetness index, soil coarse elements percentage, slope, aspect or position in the slope, since the estimated VIP statistic values for these variables were smaller than 0.8, suggesting their removal from the model.

Results concerning model validation (Table 5) show that best results were obtained with the full model ( $M_1$ ), using information from soil profiles and/or local surveys, but this information requires digging soil pits or collecting soil samples using a probe. The costs and time associated with such operations led to the search for an alternative model ( $M_2$ ) which was shown to have lower predictive abilities but nevertheless an acceptable predictive ability to be used for mapping site index. The reduced model ( $M_2$ ), with fewer explanatory variables, provides lower precision ( $|\bar{r}_i|$ , p5, p95 values) and has less efficiency (ef values), as expected. In terms of bias ( $r_i$  values), the behaviour of the two models is

**Table 3** Estimated coefficient values for the full model ( $M_1$ )

Variable	Coefficient
Intercept	14.420
Soil group (FAO soil classification)	
Arenosols	0.468
Cambisols	0.150
Leptosols	-0.492
Luvisols	0.233
Podzols	-0.596
Soil lithology	
Sand	-0.602
Sandstone	0.256
Detritic formations on schist	0.817
Shales	-1.965
Schist	0.074
Granites	0.406
Soil depth (cm)	0.013
Soil textural class	
Fine	-0.518
Medium	-0.028
Coarse	0.495
Thickness of the A horizon	0.037
Mean monthly evaporation (mm)	-0.012
Mean number of days with frost per month	-0.284



**Fig. 3** Correlation loading plots for the two first PLS factors. **a** Full model ( $M_1$ ); **b** reduced model ( $M_2$ ). SI—site index; Frost—mean number of days with frost per month; Evap—mean monthly evaporation; Thick\_A—thickness of the A horizon; SoilDepth—soil depth; text—texture (classes according to Table 3); litho—soil lithology (classes according to Tables 3, 4 for  $M_1$  and  $M_2$  respectively). Asterisk read: litho Schist (circle 50 % approximately) and text Fine (between circles 25 and 50 % approximately), double asterisk read: FAO Arenosols (circle 25 % approximately) and litho Sandstone (between circles 25 and 50 % approximately)

**Table 4** Estimated coefficient values for the reduced model ( $M_2$ )

Variable	Coefficient
Intercept	16.046
Soil lithology*	
Sands, rounded pebbles, sandstones poorly consolidated, clays (7)	0.574
Sandstone, more or less argillaceous limestones, sands, gravels, clays (8)	1.073
Shales, graywackes, sandstones (17)	-4.303
Shales, quartzites, amphibolites (20)	-3.364
Shales, graywackes (22)	-1.237
Schist, amphibolic schists, amphibolites, micaschists, graywackes, quartzites, carbonated rocks, gneisses (26)	-0.951
Quartzdiorites (311)	2.125
Granites and similar rocks (310)	0.115
Mean monthly evaporation (mm)	-0.006
Mean number of days with frost per month	-0.291

\* Numbers in brackets represent the soil lithology group (codlito) according to Silva (1983) in pp 5–6, for the national scale digital map available at the Portuguese Agency for the Environment website (<http://sniamb.apambiente.pt/webatlas/>)

**Table 5** Validation statistics values for the fitted models

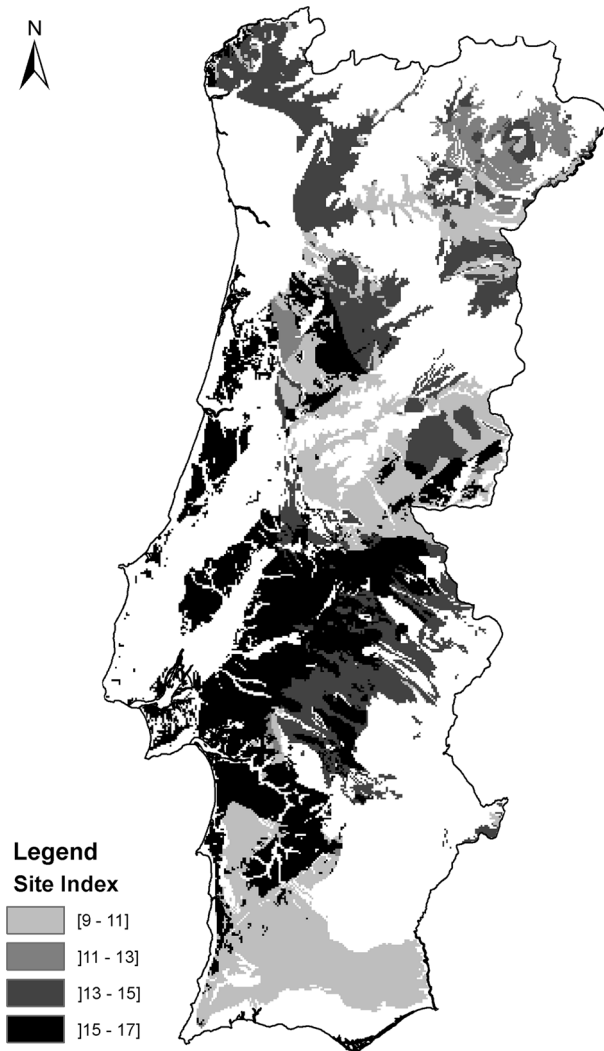
Model	$\bar{r}$	$ \bar{r}_i $	p5	p95	ef
Full ( $M_1$ )	-0.007	0.972	-2.301	1.818	0.745
Reduced ( $M_2$ )	-0.001	1.259	-2.494	2.310	0.442

$\bar{r}_i$ —mean of the value of the press residuals;  $|\bar{r}_i|$ —mean of the absolute value of the press residuals; p5—5th percentile value of the press residuals; p95—95th percentile value of the press residuals; ef—model efficiency

similar. The observation of the plots of the studentized residuals *versus* the predicted values and of the QQ-plots of the studentized residuals (not shown) did not reveal any inadequacies of the selected models.

#### Mapping predictions of the site index distribution in Portugal

The map of the site index values for Portuguese stands within the distribution area of cork oak in Portugal (Fig. 4) was obtained using the reduced model ( $M_2$ ). Site index values were grouped into 2 m classes. The predicted site index distribution (values extending from 9.5 to 16.8 m with average value of 13.4 m) was found to be in relatively close agreement with our knowledge of cork oak stands along the territory. The occurrence of higher site index values along the coast is not limited to the southern part of the country, since these values are also present north of the Tagus River.



**Fig. 4** Site index predicted values by the fitted  $M_2$  model within the distribution area of cork oak in Portugal

## Discussion

Fisher and Binkley (2000) consider that parent material, soil depth, effective rooting depth, drainage class, soil aeration and nutrient availability are the soil factors that most strongly influence forest development and tree growth. Although not all of these variables are present in the dataset used for this study, it was assumed that the variables collected from the soil pits serve as proxies for these important soil characteristics. The preference of cork oak for coarse and medium soil textures was expressed by the values of the parameters associated to these classes, in opposition to negative parameter values associated to fine textures suggesting, also through empirical knowledge, that cork oak prefers higher hydraulic conductivity soils (well drained). The presented results are in accordance with

recent research, where the importance of soil water holding capacity on the potential distribution of the species was demonstrated (Lacambra et al. 2010), as well as the importance of several soil related variables, particularly lithology (Hidalgo et al. 2008; Vessella and Schirone 2013) and soil groups (Costa et al. 2008), for the distribution of cork oak in several Mediterranean countries.

The climatic variables included in both models and the parameter values associated with each one suggest that water availability, e.g. the water cycle, lays a key role for cork oak productivity, in accordance to results found for other Mediterranean species (Bravo-Oviedo and Montero 2005) and recently demonstrated for cork oak in several Mediterranean regions (David et al. 2007; Besson et al. 2014). The water availability relevance was shown through the inclusion of mean monthly evaporation and soil variables. It demonstrated that when evaporation increases the reduction of water accessibility for trees determines a decrease in site productivity.

The variability of the estimated values within the map reflects the impact of soil characteristics on the site productivity estimation. When comparing these predictions with the ones from Sánchez-González et al. (2005) for cork oak species in two Spanish regions, noticeable differences were found. Portuguese stand predictions using model  $M_2$  ranged from 9.5 to 16.8 m (average value of 13.4 m), while with the Sánchez-González et al. (2005) estimates for the two Spanish regions range from 6 to 14 m. These results mean that no predictions are included in classes 6 and 8 as defined by Sánchez-González et al. (2005), corresponding to values with a lower limit of 5 m and an upper limit of 9 m. Additionally, the highest class predicted by this model has an upper limit of 15 m, while predictions made by the selected models of this study reach values of the class ]15;17]. It is important to note that these differences result nevertheless from the usage of two different methods: a direct site index estimation method (site index curves) by Sánchez-González et al. (2005), and the indirect method developed in this research.

## Conclusions

The models developed using partial least squares allow the prediction of a site index based on soil and climate variables. Variables related to evaporation and frost events throughout the year were the climate variables most closely related to site index values. Site index is also related to soil variables such as soil lithology, soil texture, soil depth, thickness of the A horizon and soil classification, as these variables have significant influences on soil water availability throughout the year.

The results from the site index mapping for the area in which cork oak stands are currently distributed confirms the optimum distribution of cork oak in coastal regions. This result suggests that the species may have good productivity along the northern coastal areas of Portugal, where currently it is not a common species but where, according to records concerning the collection of wood for ship building, it occurred until the middle of the sixteenth century. Since nowadays the main extractible product from cork oak is not wood but cork, cork growth and quality distribution and influencing variables needs to be further researched through the establishment of long term experimental sites along the distribution area of cork oak, namely in the central and northern coastal areas.

**Acknowledgments** This work was supported by Fundação para a Ciência e a Tecnologia (Portugal) under contracts SFRH/BD/23855/2005 and SFRH/BPD/96475/2013, and under projects CarbWoodCork (POCI/AGR/57279/2004 and PPCDT/AGR/57279/2004) and Pest-OE/AGR/UI0239/2011. Support was also given

by EU projects MOTIVE (Grant Agreement 226544) and StarTree (Grant Agreement 311919), both financed under the Seventh Framework Program for Research and Technological Development. Authors acknowledge the collaboration in data collection from: Associação de Produtores Florestais do Concelho de Coruche e Limitrofes (APFC), Associação dos Agricultores de Charneca (ACHAR), Associação dos Produtores Florestais de Ponte de Sôr (AFLOSOR), Associação de Produtores Florestais do Vale do Sado (ANSUB), Câmara Municipal de Barrancos, Empresa Municipal Herdade da Contenda, Câmara Municipal de Alfândega da Fé. Authors acknowledge manuscript revision made by Dr. Jo Smith.

## References

- Abdi H (2010) Partial least squares regression and projection on latent structure regression (PLS regression). *Comput Stat* 2:97–106
- Aertsen W, Kinta V, van Orshoven J, Özkanb K, Muys B (2010) Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecol Model* 221(8):1119–1130
- AFN (2010) Inventário Florestal Nacional Portugal Continental IFN5, 2005–2006. Autoridade Florestal Nacional, Lisboa
- Almeida A, Tomé M (2010) Field sampling of cork value before extraction in Portuguese ‘*montados*’. *Agrofor Syst* 79(3):419–430
- Alves A, Santos A, Rozenberg P, Pâques LE, Charpentier J, Schwanninger M, Rodrigues J (2012) A common near infrared based partial least squares regression model for the prediction of wood density of *Pinus pinaster* and *Larix X eurolepis*. *Wood Sci Technol* 46:157–175
- APA (1984) Carta Ecológica de Portugal—Atlas Digital de Portugal. Agencia Portuguesa do Ambiente, Lisboa. <http://sniamb.apambiente.pt/webatlas/>
- Besson CK, Lobo-do-Vale R, Rodrigues ML, Almeida P, Herd A, Grant OM, David TS, Schmidt M, Otieno D, Keenan TF, Gouveia C, Mériaux C, Chaves MM, Pereira JS (2014) Cork oak physiological responses to manipulated water availability in a Mediterranean woodland. *Agric For Meteorol* 184:230–242. doi:10.1016/j.agrformet.2013.10.004
- Bevan KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. *Hydrol Sci Bull Sci Hydrol* 24:43–69
- Bikindou FDA, Gomat HY, Deleporte P, Bouillet J, Moukini R, Mbedi Y, Nguoua E, Brunet D, Sita S, Diazenza J, Vouidibio J, Mareschal L, Ranger J, Saint-André L (2012) Are NIR spectra useful for predicting site indices in sandy soils under Eucalyptus stands in Republic of Congo? *For Ecol Manage* 266:126–137
- Bravo-Oviedo A, Montero G (2005) Site index in relation to edaphic variables in stone pine (*Pinus pinea* L.) stands in south west Spain. *Ann For Sci* 62:61–72
- Bravo-Oviedo A, Río MD, Montero G (2004) Site index curves and growth model for Mediterranean maritime pine (*Pinus pinaster* Ait.) in Spain. *For Ecol Manage* 201:187–197. doi:10.1016/j.foreco.2004.06.031
- Cardoso JC, Bessa MT, Marado MB (1973) Carta de solos de Portugal (1:1.000.000). *Agron Lusit XXXIII(I–IV)*:481–602
- Carrascal LM, Galván I, Gordo O (2009) Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos* 118(5):681–690
- Cheng-liang L, Cun-yong J, Ti-jiu C, Xia J, Xiao-hua W, Xuc-ying D (2012) Estimating canopy closure density and above-ground tree biomass using partial least square methods in Chinese boreal forests. *J For Res* 23(2):191–196
- Clutter JL, Fortson JC, Pienaar LV, Brister GH, Bailey RL (1983) Timber management: a quantitative approach. Krieger, Florida
- Coelho MB, Paulo JA, Palma JHN, Tomé M (2012) Contribution of cork oak plantations installed after 1990 in Portugal to the Kyoto commitments and to the landowners economy. *For Policy Econ* 17:59–68
- Corona P, Scotti R, Tarchiani N (1998) Relationship between environmental factors and site index in Douglas-fir plantations in central Italy. *For Ecol Manage* 110:195–207
- Correia AV, Oliveira AC (2003) Principais espécies florestais com interesse para Portugal. Ministério da Agricultura, Desenvolvimento Rural e Pescas - Direcção Geral das Florestas, Lisboa
- Costa LF (1997) Naus e galeões na ribeira de Lisboa. A construção naval no século XVI para a rota do cabo. Patrimonia – Associação de Projectos Culturais e Formação Turística, Cascais
- Costa A, Madeira M, Oliveira AC (2008) The relationship between cork oak growth patterns and soil, slope and drainage, in a cork oak woodland in Southern Portugal. *For Ecol Manage* 255:1525–1535

- David TS, Henriques MO, Kurz-Besson C, Nunes J, Valente F, Vaz M, Pereira JS, Siegwolf R, Chaves MM, Gazarini LC, David JS (2007) Water-use strategies in two co-occurring Mediterranean evergreen oaks: surviving the summer drought. *Tree Physiol* 27(6):793–803. doi:10.1093/treephys/27.6.793
- Davison AC, Hinkley DV (1997) Bootstrap methods and their application. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge
- De Martonne E (1925) *Traite de Geographie Physique*, vol 1. Paris, pp 1925–1927
- Fisher RF, Binkley D (2000) Ecology and management of forest soils, 3rd edn. Wiley, New York
- Fontes L, Tomé M, Thompson F, Yeomans A, Luís JS, Savill P (2003) Modelling the Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) site index from site factors in Portugal. *Forestry* 76:491–507
- Gea-Izquierdo G, Cañellas I, Montero G (2008) Site index in agroforestry systems: age-dependent and age-independent dynamic diameter growth models for *Quercus ilex* in Iberian open oak woodlands. *Can J For Res* 38(1):101–113
- Geladi P, Kowalski B (1986) Partial least squares regression: a tutorial. *Anal Chim Acta* 185:1–17
- Hidalgo PJ, Marín JM, Quijada J, Moreira JM (2008) A spatial distribution model of cork oak (*Quercus suber*) in Southern Spain: a suitable tool for reforestation. *For Ecol Manage* 255:25–34
- IUSS Working Group WRB (2006) World reference base for soil resources 2006, 2nd edn. World Soil Resources reports no. 103, FAO, Rome
- Jarvis A, Reuter HI, Nelson A, Guevara E (2008) Hole-filled SRTM for the globe, Version 4. CGIAR–CSI SRTM 90m Database. International Center for Tropical Agriculture, Cali, Columbia. <http://srtm.csi.cgiar.org>
- Lacambra LCJ, Andray AB, Francés FS (2010) Influence of soil water holding capacity on the potential distribution of forest species. A case study: the potential distribution of cork oak (*Quercus suber* L.) in central-western Spain. *Eur J For Res* 129:111–117
- McKenney DW, Pedlar JH (2003) Spatial models of site index based on climate and soil properties for two boreal tree species in Ontario, Canada. *For Ecol Manage* 175:497–507
- Miranda P, Coelho FES, Tomé AR, Valente MA, Carvalho A, Pires C, Pires VC, Ramalho C (2002) 20th century Portuguese climate and climate scenarios. In: Santos FD, Forbes K, Moita R (eds) *Climate change in Portugal. Scenarios, impacts and adaptation measures—SIAM Project*. Gradiva Publishers, Lisbon, pp 23–83
- Myers RH (1990) *Classical and modern regression with applications*, 2nd edn. Duxbury classic series. Pacific Grove, California, p 357
- Palmer DJ, Hock BK, Kimberley MO, Watt MS, Lowe DJ, Payn TW (2009) Comparison of spatial prediction techniques for developing *Pinus radiata* productivity surfaces across New Zealand. *For Ecol Manage* 258:2046–2055
- Paulo JA, Palma JHN, Tomé M (2013) Impact of forest policies to promote plantation and tree density increase in cork oak stands for future cork availability. The 3rd International Congress on Planted Forests. 16th to 21st May 2013. [http://www.efiatlantic.efi.int/portal/events/past\\_events/2013/icpf/scientific\\_workshops/porto\\_portugal/](http://www.efiatlantic.efi.int/portal/events/past_events/2013/icpf/scientific_workshops/porto_portugal/)
- Sánchez-González M, Tomé M, Montero G (2005) Modelling height and diameter growth of dominant cork oak trees in Spain. *Ann For Sci* 62:633–643
- Sánchez-González M, Río M, Cañellas I, Montero G (2006) Distance independent tree diameter growth model for cork oak stands. *For Ecol Manage* 225:262–270
- SAS Institute Inc. (2000) *SAS/STAT user's guide: version 8*, vol 1–3. SAS Institute Inc., Cary
- Silva AMS (1983) *Carta Litológica/Notícia Explicativa (I.13) de Portugal*. Atlas do Ambiente. Comissão Nacional do Ambiente, Lisboa. <http://sniamb.apambiente.pt/webatlas/>
- Sorensen R, Zinko U, Seibert J (2005) On the calculation of the topographic wetness index: evaluation of different methods on field observations. *Hydrol Earth Syst Sci Dis* 2:1807–1834
- Tomé M (2005) *Demonstração da gestão do montado de sobre apoiada em inventário florestal e modelos de crescimento e produção*. PROJECTO AGRO no 81. Relatório Final – Relatório de Execução Material (Volume I). Publicações GIMREF RFP 1/2005. Universidade Técnica de Lisboa. Instituto Superior Agronomia. Centro de Estudos Florestais, Lisboa. <http://hdl.handle.net/10400.5/1845>
- Tomé J, Tomé M, Barreiro S, Paulo JA (2006) Age-independent difference equations for modelling tree and stand growth. *Can J For Res* 36:1621–1630
- van der Voet H (1994) Comparing the predictive accuracy of models using a simple randomization test. *Chemom Intell Lab Syst* 25:313–323
- Vessella F, Schirone B (2013) Predicting potential distribution of *Quercus suber* in Italy based on ecological niche models: conservation insights and reforestation involvements. *For Ecol Manage* 304:150–161
- Wold S (1994) PLS for multivariate linear modeling. QSAR: chemometric methods in molecular design. In: van de Waterbeemd H (ed) *Methods and principles in medicinal chemistry*. Verlag-Chemie, Weinheim, pp 195–218