Estimation of site index from environmental variables for eucalypt stands in Portugal

Nguyen Thanh Phuong

Dissertation to obtain a Master's Degree in

Mediterranean Forestry and Natural Resources Management

Supervisor: Professor Margarida Tomé

Supervisor: Professor Susana Miguel Barreiro

Draft Version

2021

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my supervisors, Professor Margarida Tomé and Professor Susana Miguel Barreiro, who read a lot of my revisions and gave me advice. Without their guidance and persistent support, this dissertation would not have been possible.

I also would like to extend my appreciation to Jean Magalhaes. He supported me a lot in the process regarding QGIS. Thanks to him, I have learned and known more about using QGIS and ArcGIS.

It is my pleasure to learn from the amazing professors at the University of Padova, University of Valladolid and University of Lisbon, from whom I acquire more knowledge and skills. I would like to thank the MEDfOR Program's secretary and co-coordinator, Catarina Tavares, for her patience and support for me during the long period of time.

Last but not least, I want to dedicate this dissertation to my families, who stand by my side and encourages me to finish this work.

Abstract

Eucalypt (*Eucalyptus globulus* Labill.) is a fast-growing species native to Australia. It has been used for plantations in Portugal for several decades, where it has become one of the most widespread tree species. Due to the increasing wood demand by the pulp and paper companies, the eucalypt plantation areas have expanded across Portugal since the '60s. To support the strategic planning process and sustainable forest management, the site index concept is used to assess the eucalypt's productivity. This dissertation aims to develop four models, from the environmental factors classified into topographical, edaphic and climatic predictors and from the best subset of all predictors, to estimate the site index. The dataset consisted of 20,879 records, coming from permanent plots, continuous inventory plots, and trial plots established in eucalypt plantations and measured for several rotations, up to three rotations. Multiple linear regression was used to determine the key factors affecting the site index and develop the models. The model developed with the best subset of all predictors was called the integrated model, which was subsequently applied to quantile regression. Quantile regression provided a comprehensive picture of the range of model coefficients at different quantiles of the outcomes. Finally, the integrated model, which explained 32.22% of the site index variation, was used to create the potential site index distribution map across Portugal for a 500 x 500m grid used in the Portuguese National Forest Inventory. The models developed in this study provided important information about the spatial distribution of eucalypt productivity and a helpful tool to assist forest management and planning.

Keywords: site index, *Eucalyptus globulus* Labill., Portugal, multiple linear regression, quantile regression.

Abstrato

O eucalipto (*Eucalyptus globulus* Labill.) é uma espécie de crescimento rápido nativa da Austrália. É utilizada para plantações em Portugal há várias décadas, onde se tornou uma das espécies arbóreas mais difundidas. Devido ao aumento da procura de madeira pelas empresas de pasta e papel, as áreas de plantação de eucalipto têm vindo a expandir-se em Portugal desde os anos 60. Para apoiar o processo de planeamento estratégico e gestão florestal sustentável, o conceito de índice de qualidade da estação é utilizado para avaliar a produtividade do eucalipto. Esta dissertação tem como objetivo desenvolver quatro modelos, a partir dos fatores ambientais classificados em preditores topográficos, edáficos e climáticos e do melhor subconjunto de todos os preditores, para estimar o índice de qualidade da estação. O conjunto de dados consistiu em 20.879 registos, provenientes de parcelas permanentes, parcelas de inventário florestal continuo e parcelas de ensaios experimentais estabelecidas em plantações de eucalipto e medidas ao longo de várias rotações, até três rotações. A regressão linear múltipla foi usada para determinar os principais fatores que afetam o índice de qualidade da estação e desenvolver os modelos. O modelo desenvolvido resultante do melhor subconjunto de todos os preditores foi denominado modelo integrado, o qual foi posteriormente utilizado emregressão quantílica. A regressão de quantis forneceu uma imagem abrangente da faixa de coeficientes do modelo em diferentes quantis dos resultados. Por fim, o modelo integrado, que explicou 32,22% da variação do índice de qualidade da estação, foi utilizado para criar o mapa de distribuição do índice de qualidade da estação do eucalipto em Portugal para a grelha de 500 x 500m utilizada no Inventário Florestal Nacional Português. Os modelos desenvolvidos neste estudo forneceram informações importantes sobre a distribuição espacial da produtividade do eucalipto e são uma ferramenta útil para auxiliar o planeamento e gestão florestal.

Palavras-chave: índice de qualidade da estação, *Eucalyptus globulus* Labill., Portugal, regressão linear múltipla, regressão quantílica.

Resumo alargado

O *Eucalyptus globulus* Labill., inicialmente cultivado para fins ornamentais (Goes, 1962), é hoje a espécie mais plantada em Portugal (ICNF, 2019). Este facto deve-se a duas razões principais. Em primeiro lugar, as condições naturais e climáticas em Portugal são adequadas para o desenvolvimento do eucalipto (Rejmánek e Richardson, 2011). Em segundo lugar, devido à crescente procura de madeira por parte das empresas de pasta e papel, a área de eucalipto tem aumentado desde a década de 60. Para apoiar o processo de gestão florestal, o índice de qualidade da estação tem sido usado para avaliar a produtividade do eucalipto.

A estimativa precisa do índice de qualidade da estação pode ser usada como um indicador útil e confiável do crescimento da floresta. Os modelos que estimam o índice de qualidade da estação I exclusivamente a partir de variáveis ambientais são necessários para estimar o índice de qualidade da estação antes de estabelecer uma plantação num local onde antes não existia eucalipto. Estes modelos usam geralmente regressão linear múltipla (RLM) para ajustar a relação entre o índice de qualidade da estação e as variáveis ambientais. No entanto, A RLM tem muitas suposições estritas sobre normalidade e heterogeneidade e baseia-se principalmente na estimativa do valor médio da variável dependente para diferentes valores das variáveis dependentes. Os coeficientes de correlação obtidos com modelos de índice de qualidade da estação baseados em RLM não são geralmente muito altos. Por outro lado, a regressão quantílica é mais flexível para descrever a distribuição da relação entre as variáveis dependentes e independentes e permite a estimativa do índice de qualidade da estação, para um conjunto de valores para qualquer percentil.

O principal objetivo desta dissertação foi desenvolver modelos com base em RLM e RQ para estimar o índice de qualidade da estação de eucalipto para Portugal utilizando como preditores variáveis ambientais disponíveis em mapas digitais. Além disso, foi criado um mapa da distribuição potencial do índice de qualidade da estação em Portugal utilizando os modelos ajustados.

Os dados utilizados nesta dissertação foram selecionados a partir da base de dados EGLOBdata, gerida pelo grupo ForChange do Centro de Estudos Florestais (CEF) do Instituto Superior de Agronomia (ISA). Esta base de dados contém informação de várias fontes, como ensaios experimentais, parcelas permanentes estabelecidas e medidas juntamente pelo grupo ForChange e pelas empresas. O conjunto de dados usado para ajustar o modelo consistia em 20.879 registos e foi medido durante várias rotações, até três rotações. Para estimar o índice de qualidade da estação para cada parcela e rotação, foi dada preferência ao par (altura dominante do povoamento, idade do povoamento) para o qual a idade era mais próxima da idade padrão do eucalipto (10 anos).

As variáveis ambientais utlizadas neste estudo foram categorizadas em três grupos principais: climáticas, topográficas e edáficas. Foram utilizadas nove variáveis climáticas: evapotranspiração (calculada pela quantidade de evapotranspiração de água), duração da estação de geada no ano (número de dias de geada), humidade relativa do ar, horas de insolação por ano, número de dias com mais de 1 mm de precipitação por ano, precipitação total anual, radiação solar (quantidade total de radiação global por ano) e temperatura. Em relação às variáveis topográficas, os preditores foram longitude, latitude, altitude, declive e aspeto. As variáveis edáficas incluíram tipo de solo, litologia, fluxo de água no solo, quantidade de recursos hídricos subterrâneos e pH do solo.

Aa correlação entre o índice de qualidade da estação e cada variável ambiental foi avaliada numa análise preliminar. Para as variáveis contínuas, aplicou-se a correlação de Pearson. Para as variáveis categóricas, fez-se uma análise de *“box-plot”* para avaliar como é que cada variável em cada categoria pode influenciar o índice de qualidade da estação. Em seguida, o modelo para estimar o índice de qualidade da estação para eucalipto foi ajustado utilizando regressão linear múltipla. As variáveis categóricas foram transformadas em variáveis dummy para entrarem na equação. O processo de modelação foi realizado utilizando o software estatístico R dentro do software R (R Development Core Team, 2013). O melhor modelo - o modelo integrado foi encontrado utilizando todas as variáveis e aplicando regressão stepwise com o Akaike Information Criterion (AIC).

No modelo integrado aplicou-se também a regressão quantilica para observar como o sinal e o valor de cada preditor mudaram para os diferentes quantis. Por fim, o modelo integrado foi utilizado para criar um mapa de produtividade potencial para Portugal.

O resultado da análise preliminar mostrou que a evapotranspiração teve a maior correlação estatística com o índice de qualidade da estação. Essa relação foi positiva, o que significa que quanto maior a taxa de evapotranspiração, maior o valor do índice de qualidade da estação correspondente. Em termos de aplicação da regressão linear múltipla, o modelo integrado foi o melhor modelo explicando 32,22% da variação do índice de qualidade da estação. Ao aplicar a regressão quantilica ao modelo integrado, os *outliers* foram observados principalmente abaixo do quantil inferior (de 0 a 25 percentil), uma vez que os intervalos de confiança dos coeficientes foram mais expandidos do que no quantil posterior. Em relação ao mapa de produtividade potencial, a melhor localização para o crescimento do eucalipto é, junto à costa. O índice de qualidade da estação mais alto (> 23m) foi encontrado no Noroeste de Portugal No centro de Portugal, os valores do índice de qualidade da estação situaram-se no intervalo médio de 14 a 20m. Em geral, de oeste (áreas costeiras) para leste (interior) de Portugal, o valor do índice de qualidade da estação diminuiu gradualmente.

No entanto, existem algumas questões que requerem um estudo mais aprofundado. Para o mapa de variância do índice de qualidade da estação, em comparação com o valor do índice de qualidade da estação das parcelas de estudo observadas, havia muitas parcelas onde os valores observados eram muito diferentes do valor estimado. Além disso, havia também alguns locais onde existem parcelas, onde em teoria, não deveria haveria plantações de eucalipto devido às condições ambientais inadequadas (altitude elevada, baixa precipitação A limitação deste estudo pode vir do número limitado de variáveis independentes usadas para prever o índice de qualidade da estação. Por exemplo, este estudo não considerou os componentes químicos do solo, como nitrogênio e fósforo, que desempenham um papel importante na produtividade do eucalipto (Almeida, 1996). Além disso, a profundidade do solo ou a distância ao mar também não foram incluídas neste estudo. Assim, essas variáveis deveriam ser consideradas num estudo futuro. Uma sugestão para trabalhos futuros é o desenvolvimento de uma metodologia para refinar a previsão média fornecida pelo modelo obtido nesta tese, utilizando informações mais detalhadas de um determinado local, como localização na encosta e profundidade do solo.

Table of Contents

List of tables

List of figures

1. Introduction

Eucalyptus globulus Labill. is planted in more than 90 countries globally (Jacobs, 1981). Eucalypts, originating in Australia, were first grown in Portugal for ornamental purposes (Goes, 1962). Until now, eucalypt has become the most planted species in Portugal to produce timber and fibre (ICNF, 2019). Portugal's natural and climatic conditions are suitable for the genuine establishment of eucalypt in many regions, as eucalypt can grow well in various environments, such as temperate, tropical and subtropical (Rejmánek and Richardson, 2011). Due to climate change, the development of eucalypt plantations has also been affected. With the increase in global temperature and atmospheric CO2 (Solomon, 2007), eucalypt productivity can increase or decrease depending on specific locations (Booth, 2013).

Due to the increasing wood demand for pulp and paper companies, the eucalypt plantation areas have expanded across Portugal since the ’60s. However, this led to the land conflict among eucalypt and other species. Additionally, as the eucalypt is grown in homogeneous forests occupying large continuous areas, becoming flammable material for wildfire, which is one of the main problems for forestry in Portugal, careful planning of land use at the landscape level has been pointed out as essential to minimize possible harmful environmental impacts (Tomé et al., 2021). To support the strategic planning process, site index is used to assess the productivity of eucalypt. For this reason, it is necessary to develop a method to estimate site productivity, often evaluated through site index, from Portugal's environmental variables.

The relationship between dominant height at a certain age and site productivity, known as site index, plays a crucial role in forest management and is often applied to estimate forest productivity (e.g. Burkhart and Tomé, 2012). The site index concept, suitable for even-aged forestry, assumes that dominant height is not affected by silvicultural treatments. Many studies have been undertaken to develop site index models or models that predict the evolution of dominant height with age for different quality sites. The set of curves representing dominant height growth for the range of possible site indices is known as site index curves and is usually based on stand age and dominant height. Site index curves can be developed and computed mostly by applying difference equations, the guide-curve method, and the parameter prediction method (Burkhart and Tomé, 2012). Over time, to improve model accuracy, climatic variables were included in the models (Burkhart and Tomé, 2012).

Predicting site index accurately can be used as a helpful and reliable indicator of forest growth. The models estimating site index exclusively from environmental variables are needed when estimating site index before establishing a plantation. Those models mostly use multiple linear regression (MLR) to fit the relationship between site index and the environmental variables. However, MLR has many strict assumptions regarding normality and heterogeneity and is mainly focused on estimating the conditional mean of the dependent variable across different values of the regressors. The correlation coefficients obtained with MLR site index models are usually not very high. For example, the study of Marques (1991) on estimating the site index of maritime pine (*Pinus pinaster* Air.) from site variables (namely edaphic, climatic and topographical variables) explained 54.4% of the variation. Or the study of Paulo et al. (2015) on cork oak (*Quercus suber* L.) from the climatic and edaphic variables explained 70% of the total variation if detailed variables on soil were available, but just 49% if using variables available from digital maps. However, in the detailed model most of the variation was explained by only one factor (positively related with soil variables and negatively with evaporation), which itself explained 61.1% of the variation. The correlation coefficients of these studies when based on digital information were not very high because micro factors such as soil depth, position in the slope or azimuth (sometimes not obtained with precision from digital maps) are not taken into account.

Quantile regression (QR), on the other hand, is more flexible to describe the distribution of the relationship between dependent variables and independent variables and allows the prediction of the site index, for a set of values for the regressors, for any percentile. A user may therefore estimate the mean site index from the MLR model and use the quantile regression to estimate some extreme percentiles (for instance, 95 and 5 percentiles) to obtain the range of site index that can be obtained on that location. With this information, he/she can then use his/her local knowledge to adjust the estimated mean for a particular site using the extreme percentiles as guidelines. Recently, QR has also been applied to other analysis in forestry (Xie et al., 2020, Hinko-Najera et al., 2019; Paulo et al., 2021).

1.1 Objectives

This dissertation's main objective was to develop MLR and QR models to predict the eucalypt site index in Portugal using environmental variables available in digital maps as predictors. Furthermore, a map of the potential site index distribution across Portugal was created using the fitted models. These goals were completed according to the following specific objectives:

1. Identifying the relationship between site index of eucalypt and environmental variables across Portugal:
2. Developing models to estimate site index for eucalypt by applying MLR. The MLR model was developed in 4 stages, each one corresponding to a "best" model: (M_1): model including just topographical variables; (M_2): model including just climatic

variables; 3) (M_3): model including just edaphic variables; (M_i): model including a subset of environmental variables from all groups:

3. Applying QR to the M_i model at different quantiles (95%, 75%, 50%, 25%, 5%);
4. Use the M_i resulting from the second objective to create the site index variation map for a spatial unit of 500 x 500m (the one used in the Portuguese National Forest Inventory).

1.2 Hypothesis

The main driving questions in this dissertation were:

1. Which environmental factors were most significantly correlated to the site index in eucalypt?
2. Which group of predictors best explained the variation of site index when applying MLR?
3. Can QR be used to estimate the range of site indices that can occur, due to micro-variation of environmental factors, at a particular location?
4. Is the MLR model developed suitable for estimating the geographic areas with the highest and lowest site indices across Portugal?

1.3 Thesis structure

The thesis is henceforth structured as follows. Chapter 2 contains a review of the relevant knowledge. The methods used in the study are then described in Chapter 3, then the results are presented and discussed in Chapter 4. Finally, Chapter 5 outlines the main conclusions and identifies both limitations to the study and recommendations for further research.

The thesis ends with the References' list.

2. Background

In this chapter, the main overview and the definitions relating to the topic of research are presented. This chapter is divided into three subchapters. The first subchapter presents a brief history of forestry in Portugal. It focused on forest area and structure change over time and the forest ownership distribution, among other things. The second subchapter describes eucalypt plantations in Portugal, how these are managed, and their main threats. The final subchapter presents some concepts relating to site index and a review of other authors' previous works in this field.

2.1. Brief history of forestry in Portugal

Due to the edaphic and climatic characteristics, Portugal has suitable conditions for forest growth. Nevertheless, at the beginning of the 19th century, the forest area in Portugal was not high (Mendes et al., 2002). In the 1830s, the primary issue relating to land use distribution was the high percentage of uncultivated land, which accounted for 62.3% of the land use. Back then, the forest land and the agricultural lands were only 14.1% and 21.5%, respectively (Mendes et al., 2002). As the uncultivated land was high, afforestation was considered the top priority of forest policy (Mendes 2008). The first primary mission of afforestation was to prevent soil erosion at the coastal dunes (Mendes, 2008). The General of Administration of the Crown forest (Administração Geral das Matas da Real Coroa) was established to manage the afforestation process (Mendes, 2008). Initially, the increase was mostly achieved with maritime pine (*Pinus pinaster*) plantations. For an 80-year-period, from 1875, the total area of the forest in Portugal increased from 640 thousand hectares (mainly maritime pine and evergreen oak) (Pery, 1875) to approximately 3.3 million hectares (DGF, 2001). Afforestation, as a priority of forest policy, lasted until the end of the 20th century (DGF, 2001). From then on, instead of continuing afforestation, it was replaced by the replanting and protecting existing stands (Cesaro et al., 2008).

Besides afforestation, over time, due to the forest fires and the pinewood nematode's attack (*Bursaphelenchus xylophilus*) and the subsequent policy to cut down or harvest all of the trees having the disease symptoms and those around a specific buffer, the area of pine forest was reduced, reaching 713.3 thousand hectares by 2015 (ICNF, 2019). On the contrary, from 1995 to 2010, the eucalypt plantation area was expanded by 95.5 thousand hectares by using the harvested pine forests, abandoned agricultural land, and shrublands and pastures, with the areas of 70 thousand, 12 thousand, and 13.5 thousand hectares, respectively (IFN 2010).

According to the latest National Forest Inventory (ICNF, 2019), forest cover is estimated at 3224.2 thousand ha, which occupies 36.2% area of mainland Portugal (Table 1). Forests include two components: the forested areas (representing the forest stands) and the areas temporarily deforested (including burnt areas, harvested areas, and areas under regeneration where forest cover is expected to return in the short run) (Table 2). In general, forested regions increased by 0.2% per year from 1995-2015 (ICNF, 2019).

Table 1- Land use area in mainland Portugal. (ICNF, 2019; Inventory reference year 2015).

Land use	Area of soil use						
	1995	2005	2010	2015			(2005-2015)
	10^3 ha	10^3 ha	10^3 ha	10^3 ha	%UT	%Error	10^3 ha
Forest	2205.6	3215.9	3164.2	3224.2	36.2	± 0.4	+8.3
Shrubland and pastures	2539.6	2716.7	2832.1	2766.2	31	± 0.5	+49.5
Unproductive	190.3	195.8	185.4	191.7	2.2	± 2.2	-4.1
Water inland	151.9	178.2	185.2	192.8	2.2	± 2.2	+14.7
Agriculture	2407.3	2204.7	2117.2	2092.9	0.6	± 0.6	-111.8
Urban	315.5	399.0	427.2	442.4	1.4	± 1.4	+43.4
Mainland Portugal	8910.2	8910.2	8910.2	8910.2	100	-	-

Table 2- Forest area by type of occupation (ICNF, 2019; Inventory reference year 2015).

	Forest area by type of occupation						
	1995	2005	2010	2015			(2005-2015)
Occupation	10^3 ha	10^3 ha	10^3 ha	10^3 ha	%UT	%Error	10^3 ha
Wood areas (stands)	2792.8	2910.7	2948.8	2987.1	92.6	± 0.2	+85.4
Temporarily deforested surface:	512.8	314.1	215.3	237.0	7.4	± 1.9	-77.1
Harvested	15.6	28.3	38.0	98.7	3.1	± 3.1	+70.4
Burnt	44.3	104.6	30.0	12.6	0.4	± 8.7	-92.0
Under regeneration	453.0	181.2	147.4	125.7	3.9	± 2.7	-55.5
Total: forest	3305.6	3215.9	3164.2	3224.2	100.0	± 0.4	+8.3

Based on the structure, function, and landscape terms, Portugal's forested areas could be classified into four main groups (Table 3): evergreen oaks, pine forests, silvo-industrial hardwoods, and deciduous hardwoods (ICNF, 2019). Specifically, evergreen oaks, or in other words, "montados", including cork oak and holm oaks, are the leading forest group. They account for 1/3 of the total forest area, with an estimated area of 1 million hectares. Cork and holm oak forests are managed as agroforestry systems, with 719.9 million hectares and 349.4 million hectares, respectively (ICNF, 2019). "Montados" is followed by the pine forests group, including maritime pine and stone pine covering an area estimated to be nearly 1 million hectares.

Table 3- Forest area by group of species and type of forest formation (ICNF, 2019; Inventory reference year 2015).

Forest area by a group of species and type of forest formation							
Occupation	1995	2005	2010	2015			(2005-2015)
	10^3 ha	10^3 ha	10^3 ha	10^3 ha	%UT	%Error	10^3 ha
Tree species formation							
Hardwoods	2028.4	2038.1	2066.5	2054.8	63.7	±0.4	+16.6
Softwoods	1041.7	920.7	846.1	744.8	23.1	±1.0	-176.0
Mixed hardwoods and softwoods	214.8	229.3	243.4	418.9	13.0	± 1.4	+189.6
Temporary deforested areas with unidentified tree species	20.6	27.6	8.1	5.7	0.2	± 13.0	-22.0
Total: forest	3305.6	3215.9	3164.2	3224.2	100	± 0.4	+8.3
Forest formations							
Pine forests and other softwoods	1159.5	1044.4	975.1	959.1	29.7	± 0.8	-85.4
Eucalyptus	717.2	785.9	810.8	845	26.2	± 0.9	+59.1
Decidous hardwoods	279.9	274.2	286	320.2	9.9	± 1.6	+46.1
Evergreen hardwoods	1125.8	1079.0	1078.6	1085.8	33.7	± 0.8	+6.8
Acacia	2.7	4.7	5.5	8.4	0.3	± 10.6	+3.7
Temporary deforested areas with unidentified tree species	20.6	27.6	8.1	5.7	0.2	± 13.0	-22.0
Total: forest	3305.6	3215.9	3164.2	3224.2	100	± 0.4	+30.3

Concerning the geographic distribution of the species, maritime pine forests are mainly located in the North and Center of Portugal, while eucalypt plantations are usually found along the coast in the North and Centre of the country. Evergreen oaks can be seen in the Center and Southern areas of the country (Figure 1) (ICNF, 2019).

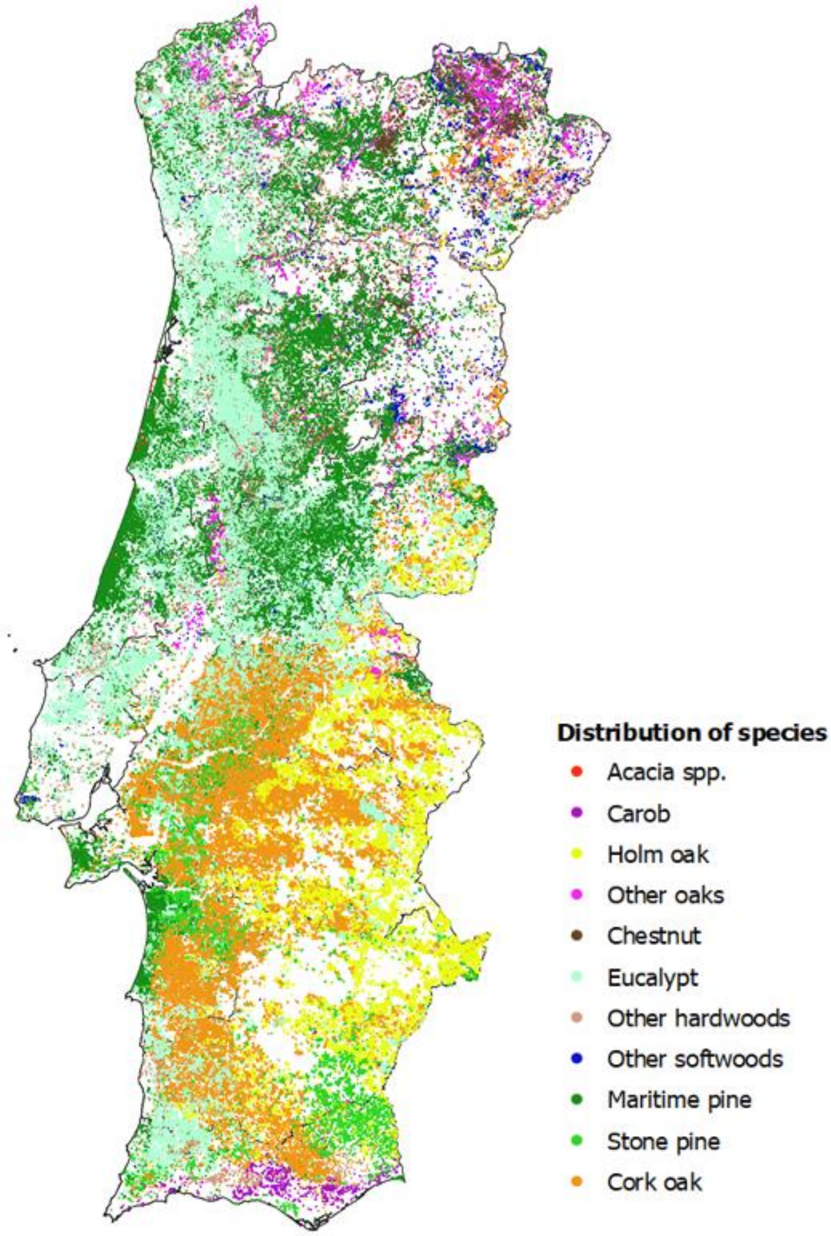


Figure 1- Distribution of species across Portugal (ICNF, 2019; Inventory reference year 2015).

Forest ownership can be divided into three main groups: 1) public ownership; 2) private ownership; 3) /other types of ownership. In the world, public forests, owned mainly by the State at the central, regional, or council level, generally account for approximately 73% (FRA, 2006). Private ownership, which comprises forest properties belonging either to non-industrial (such as the small scale forest owners) or industrial private forest owners (Mendes et al., 2014), represent 22% (FRA, 2006). In Europe, although varying among countries, up to 90% of the forests are public. However, in Portugal, this pattern is reversed. Based on the FRA (2006), Portugal was considered one of the top 10 countries in Europe with the highest private forest ownership percentage. Up to 1995, private owners constituted 93.4% of the forest area (Table 4) (adapted from Mendes et al. 2004).

Table 4- Distribution of area of forests and other wooded lands by types of ownership from 1928 to 1995 (adapted from Mendes et al. 2004).

Forest by ownership	1928		1959		1974/82		1995	
Type	Area (ha)	%	Area (ha)	%	Area (ha)	%	Area (ha)	%
Public	53662	2.3	58000	2.0	78000	2.6	40000	1.2
Communal	55954	2.4	145000	5.0	380000	12.4	180000	5.4
Private	2221824	95.3	2697000	93.0	2598000	85.0	3129000	93.4
Total	2331440	100	2900000	100	3056000	100	3349000	100

Specifically, most of the forestland in Portugal is owned by non-industrial private forest owners (Mendes et al., 2014). About 61% of private owners have less than 5 hectares, accounting for 26% of the total forest area (Santos et al., 2013). Moreover, there is a geographic contrast of land ownership structures in Portugal (Baptista and Santos, 2005). In early 2000, the small-scale forest estates (below 10 ha) were mainly distributed in the northern and central regions, where the main species are pine and eucalypt. These small privately-owned areas led to difficulties in forest management, policy and protective measures implementation. One solution was to gather the forest owners into associations to create a minimum area for sustainable forest management (DGRF, 2007). Conversely, the large-scale forest holdings (>100 hectares) are mainly found in the southern regions, where the "montados" agroforestry system is found.

2.2. Eucalypt plantations in Portugal

Eucalypt species, native to Australia, can be considered one of the hardwood species most planted outside its natural ranges (Rejmanek et al., 2011). It can be found in many different regions, including temperate, tropical, and subtropical areas, with more than 110 species planted in over 90 countries (Booth, 2013). Eucalypts have been planted outside of their natural territory because of their fast growth and survival ability in dry and low nutrient environments (Florence, 1996). Even so, the yield of eucalypt species dramatically depends on the soil, water, and nutrient availability (Florence, 2004; Pereira et al., 1989). It is estimated that 90% of eucalypt plantations worldwide come from "big nine" species, which are *E. camaldulensis, E. grandis*, *E. teretecornis, E. globulus, E. nitens, E. urophylla, E. saligna, E. dunnii, E. pellita* and their hybrids (Harwood, 2011). Eucalypts are planted for different biomass uses: pulp, fiberboard, fuelwood, and charcoal (Turnbull, 1999). Not only that, but eucalypt can also be found in smaller-scale areas, such as in small woodlots, and be used as windbreaks (Jacobs, 1981).

Eucalypt was introduced in Portugal in the middle of the 19th century and was initially used for ornamental purposes (Goes, 1962). Later, due to its socio-economic benefits and raw material provision, the area of eucalypt plantations expanded in the mid-20th century (Jacobs, 1979; Potts et al., 2004). The eucalypt species mostly grown in Portugal is the *Eucalyptus globulus* Labill., also known as Tasmanian blue gum.

So far, eucalypt plantations have experienced a long journey in Portugal, from introduced species to the most widespread species. Before eucalypt was planted in Portugal, the main species grown were native species, including maritime pine and evergreen oaks, such as cork oak (*Quercus suber*) and holm oak (*Quercus rotundifolia*). In the beginning, when first introduced in Portugal, eucalypt areas were negligible. After World War II, due to the increasing demand for pulp, the area of eucalypt plantations increased, leading to more than 200 thousand hectares in the late seventies (DGOGF, 1978). According to the latest Portuguese National Forest Inventory (IFN6 - reference year 2015), the total area of eucalypt amounted to 26% of the entire forest cover (about 845 thousand hectares) (Figure 2), becoming the most prevalent species across the country (ICNF, 2019). Conversely, the area of maritime pine decreased. This trend became more and more evident from 1995 onward.

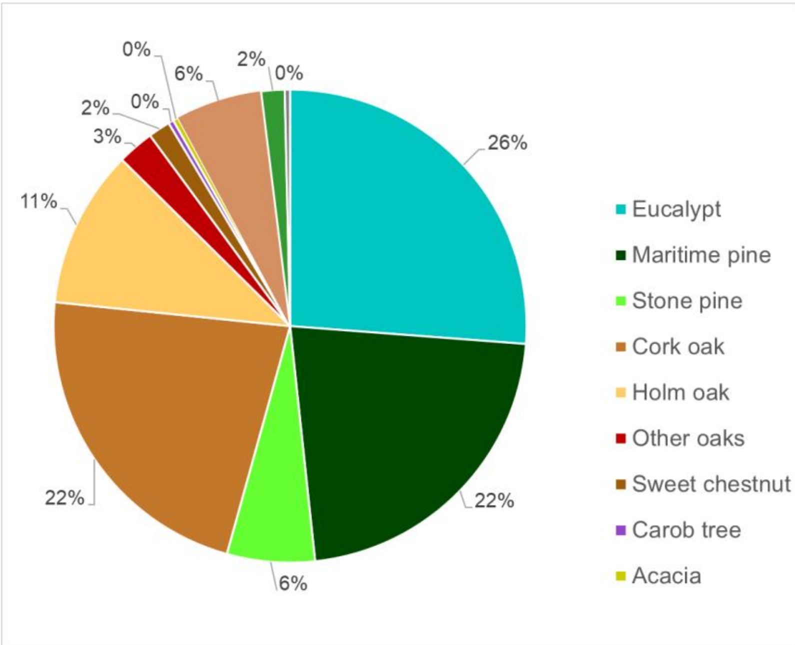


Figure 2- Share of forest area by species in Portugal (ICNF, 2019: Inventory reference year 2015).

Unlike most Portuguese forests planted and managed as high forests, *Eucalypt* spp. is managed as a short-rotation coppice (a rotation of 10-12 years) (Turnbull, 1984). Eucalypt is managed by pulp and paper companies and non-industrial private owners or the respective forest owner associations. By 2018, the Portuguese pulp and paper industries were directly responsible for managing 150 thousand hectares of eucalypt plantations (CELPA, 2018), accounting for 18% of the eucalypt plantation area in the country.

Eucalypt plays an essential role in the Portuguese economy, as it is a fast-growing species and provides good quality fibre (e.g., Jorge et al., 2000, Arau et al., 2003). The parameter used to assess the pulpwood quality is fibre length (Hudson et al., 1995, Sandercock et al., 1995). In terms of eucalypt, fibres account for 64-68% of the wood (Dadswell, 1972). The pulp and paper sectors are currently responsible for 4.9% of the national exports and 0.6% of the Gross Domestic Product (CELPA, 2018). Due to the economic benefits, pulp and paper companies invest in eucalypt plantations and manage them from an industrial perspective (Oliveira et al., 2017).

The economic importance of eucalypt in Portugal has led to considerable research about plantations' establishment to maximize productivity that depends significantly on the site quality. In particular, in high-quality sites, eucalypts can be harvested at the age of 10, while in low-quality sites, the harvest age can be delayed until 14. However, extending the harvest age is less common because the owners are afraid of losing their wood to forest fires. Managing eucalypt for three rotations, including one plantation followed by two coppices, is usually recommended. However, due to the high cost of site establishment, rotations can go up to the 4^{th} cycle. Not only the rotation length and the number of coppices but also site conditions affect wood yields. In the best quality sites, the wood yields can surpass 30 m^3/hectare/year (Tomé, 2000).

Besides being focused on the plantation's establishment, improving the genetic of the eucalypt has also received attention. Due to the need to improve vitality and productivity, genetic improvement programs have been conducted under pulp and paper companies' sponsorship. Eucalypt clones or improved genetic materials have been used since 1995 (RAIZ, 2000). Until recently, clones were primarily used by pulp and paper companies because these are produced in the pulp and paper company's nurseries to supply the areas under industrial management, leaving non-industrial owners with limited access to these materials. But there is a tendency to enhance the availability of genetically improved material to the private owners.

Although eucalypt can survive in various climatic environments, such as tropical, subtropical, and temperate regions (Iglesias and Wilstermann, 2008), climatic factors influence yield, with temperatures below 0^oc in winter being considered the primary limiting climatic factor to eucalypt's development (Pereira et al., 1996 ; Ribeiro and Tomé, 2000; Alves et al., 2012).

Consequently, eucalypts prefer the areas close to the sea, with a milder temperature, lower thermal amplitude, and usually lower number of frost days (Catry et al., 2015).

Along with the precipitation and temperature, soil characteristics are among the most critical factors affecting eucalypt development (Boland et al., 2006; Kirkpatrick, 1975; Jacobs, 1981). For example, if planted in soil with insufficient depth, poor drainage, salinity, high alkalinity, and assimilable carbonates, eucalypt development will be limited (Jacobs, 1981). Furthermore, the topographic factors, such as aspect, slope, and topographic position, affect the incoming solar radiation, water flow, and soil erosion, which later influence local temperature, radiation, water, and nutrient availability (Águas et al., 2014).

Climate change, motivated by the decrease of precipitation, may substantially impact the development and wood productivity of the eucalypt in Portugal, expressed by a lower amount of wood supply (Garcia-Gonzalo et al., 2014). The climate change situation has two possible impacts on the growth and development of the eucalypt. On the one hand, increasing CO_2 enhances photosynthesis process speed and increases water use efficiency. On the other hand, higher CO_2 levels affect temperature and rainfall regimes. To assess how vulnerable eucalypts are to climate change in climatic and atmospheric changes, a study by Booth (2013) accepted a vulnerability function from a European study (Schröter and ATEAM Consortium, 2004). In this study, vulnerability is a function of potential impact and adaptive capacity (Schröter and ATEAM Consortium, 2004). Regarding potential impact, a modified model of 3-PG, which included and simulated the amount of CO_2 for photosynthesis in case the amount of CO_2 in the atmosphere increased, showed an increase in productivity (Booth 2013). Eucalypts also had a high ability to adapt to the change, so generally, eucalypts were at a medium level to react with the climate change (Booth 2013).

Portugal has the most significant percentage of annual burnt forest area in Europe and is one of the world's largest (San-Miguel-Ayanz et al., 2017). The annual burnt rate comprised an average of 3% of the forest and shrubland cover (Mateus et al., 2014). Thus, it is not surprising that the Portuguese National Forest Strategy (DGRF, 2007) considers wildfires as the most relevant hazard affecting sustainable forest management. Many studies have shown that fire occurrence increased because agricultural and grazing areas were replaced by forests (Duguy et al., 2007; Pausas and Fernández-Muñoz, 2012). Figure 3 shows that pasture and shrublands were the highest percentages of the burned area from 1996 to 2014. Fire hazard is often determined by the timing, intensity and frequency of fuel treatments under the occurrence of fire-prone climate conditions (Mirra et al., 2017). However, fuel load reduction treatments will likely affect understory biodiversity; therefore, sustainable forest management is a difficult task that requires compromise solutions that consider all these factors (Maia et al., 2014).

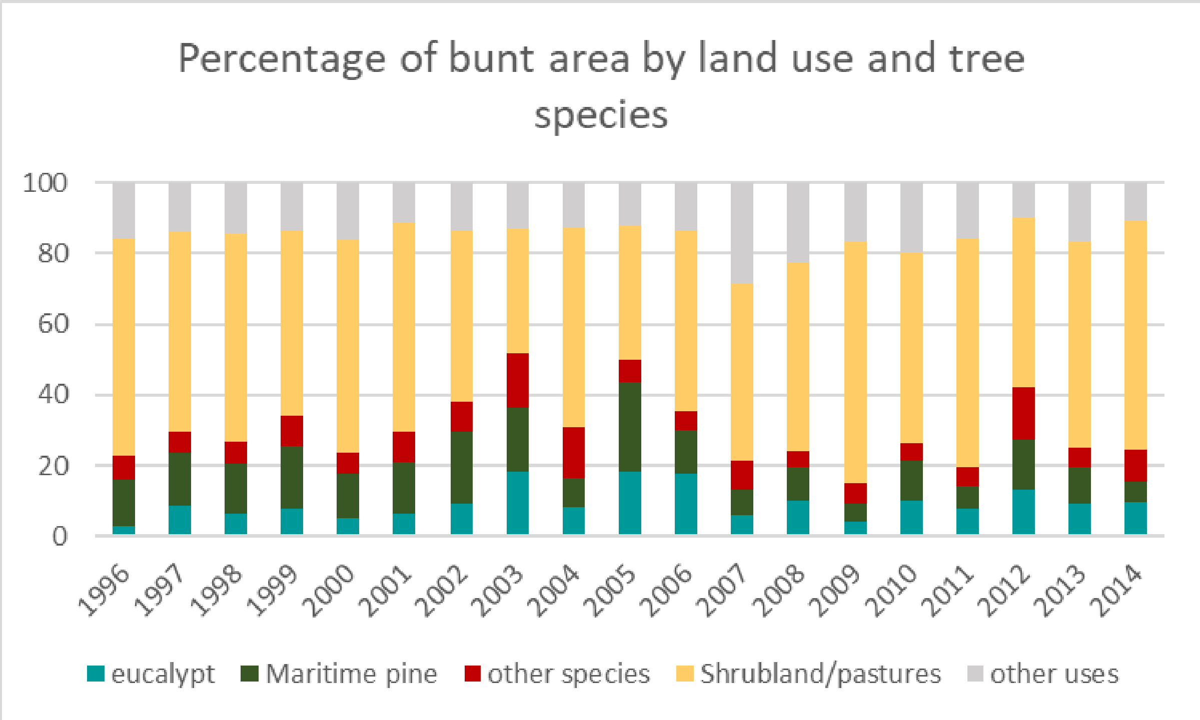


Figure 3- Distribution of total burned area in mainland Portugal by land-use class (ICNF, 2019; Inventory reference year 2015).

Portuguese studies have shown that fire prefers shrubs to pines and eucalypts (Fernandes et al., 2019), and the share of burnt areas by land use and tree species supports these findings (Figure 3). Nevertheless, there are several reasons, some related to the species while others concerning its management and landscape planning, that can explain why eucalypt plantations are at high risk of wildfires in Portugal. First, the eucalypt itself is considered a flammable material (Luke and McArthur, 1978). Even if in small quantity, its volatile oils can enhance fires (Ganteaume et al., 2010). Second, the long streamers of semi-detached bark produced can originate long-distance spot fires (Cruz et al., 2015). Finally, because eucalypt is planted in large homogeneous areas, this can intensify fire activity due to the lack of horizontal discontinuity of fuel across the landscape (Fernandes et al., 2019). However, these plantations rapidly achieve vertical discontinuity given eucalypt fast-growth ability, particularly in industrially managed plantations where understory fuel loads are removed more or less often depending on the site, rotation, and stand age. Due to either post-fire abandonment or poor management, it often leads to irregular stands and/or mixtures of maritime pine and eucalypt, which are considered the most susceptible to fire (Moreira et al., 2009) with the consequent loss of vertical discontinuity.

Besides the undeniable economic benefits, enlarging the eucalypt plantation areas may adversely affect the land. These effects include the change in local micro-climate, soil depletion, competition with native species, and soil water availability (Almeida and Riekerk, 1990). Intensive eucalypt plantations can deplete the soil nutrients and break down the soil structure (Madeira et al., 1989; Hopmans et al., 1993; Ludwig et al., 1997; Khanna, 1997). Gradually, this will affect the soil's ability to hold water, and in the long-term, it will affect the sustainability of the plantation site (Jones et al., 1999). Moreover, soil can also be eroded between clear-cutting and the next plantation, especially in steep slopes (Jones et al., 1999).

Knowing about the environmental factors affecting the development of the eucalypt and how eucalypt plantations react back to the environment may support managers in prioritizing the most appropriate locations for eucalypt plantations. Thus, it is crucial and necessary to monitor the best place for the plantations of this species.

2.3. The importance of site index and the different approaches to model it

A forest site is defined as a specific area comprising all the necessary environmental factors for a tree's development. These environmental factors not only affect the development of the trees but also interact with each other. To assess the forest site's quality for a specific species, site quality and site productivity are used (Menéndez-Miguélez et al., 2015). Site quality and site productivity refers to the essential ability of the site to provide the resources for tree growth and the potential growth of the species that the site can support or the expected productivity and can be changed if modified (e,g, site preparation, fertilization, irrigation) (Clutter et al., (1983).

Accurate estimation of site quality has always been given much attention by forest managers. The main reason is that foresters want to choose the most appropriate species, which will provide the highest volume growth for a specific site (Ringius et al., 1997). Indeed, assessing the site potential by estimassting wood productivity is one of the most critical aspects of land management (Monserud et al., 1990; Carmean, 1996; Tyler et al., 1996)

Both site quality and site productivity can be estimated indirectly using the site index concept. Site index (S) is defined as the dominant height at a species-specific reference age (e.g. Burkhart and Tomé, 2012; Carmean, 1975; Vanclay, 1992; Skovsgaard and Vanclay, 2008). Dominant height is usually defined as the average height of the 100 thickest (in diameter at breast height) trees per hectare. But in other parts of the world, it can also be estimated as the average height of the dominant and co-dominant trees or the 100 tallest trees per ha. The largest trees in diameter at breast height are used because these are the least affected by thinning intensity when thinning from below is applied (Skovsgaard and Vanclay, 2008).

Moreover, the large trees are less dependent on stand density (Iles, 2003). Several methods are applied to estimate the site index based on dominant height at the corresponding stand reference age, which can be grouped as the independent fitting of site index curves, guide-curve method, parameter prediction method, difference equations approach, and mixmodelling techniques (Burkhart and Tomé, 2012). Site index is commonly used in plantations for three main reasons: (1) it is easy to interpret the results; (2) it has a significant relationship with site quality; and (3) it is easy to estimate for any existing even-aged stand (Bravo-Oviedo et al., 2004).

To increase the dominant height modelling's explanatory ability, environmental factors can be added as predictors when modelling dominant height growth (Burkhart and Tomé, 2012). Specifically, the existing statistical growth equations can be expanded by refining the environmental variables' parameters (Weiskittel et al., 2011). For example, some studies have incorporated the environmental variables to predict site index from the dominant height of eucalypt when applying difference equation methods. For instance, Scolforo et al. (2016) applied the polymorphic modified Von Bertalanffy-Richards model to predict the site index of *Eucalyptus* spp. in Brazil. Additionally, the study of Scolforo et al. (2020) estimated the dominant height of clonal eucalypt by using the modified Chapman-Richards growth function. The asymptote parameter in the equation was expressed as a function of the annual soil water deficit. This method has been proven to be more accurate for estimating the dominant height of the species and describing how climatic variation can affect eucalypt as it helped increase the fitting performance and decrease the heteroscedasticity. Regarding eucalypt, in Portugal, GLOBULUS 3 model, which was upgraded from model GLOBULUS 2.1 (Tomé et al., 2001) by including environmental variables, such as altitude and number of precipitation days, was also used to estimate the site index curve using a difference equations' approach (Tomé et al., 2006).

There are also limitations when using age-dependent approaches to estimating site index (S). First, in the field, determining the stand age exactly may not be easy, and a small error can lead to significant differences in site index estimates (Burkhart and Tomé, 2012). Second, although computing site index from the stand's dominant height and age is not difficult, it may not be feasible in many cases. In particular, when the stand age is not known (e.g., uneven-aged stands) or when tree heights are difficult to measure. Finally, if the location did not have the species before, no age or height data would be available. This prompted more research on developing models, which are independent of the stand age and dependent on the environmental variables, to assess the potential productivity.

The relationship between S and environmental variables may provide useful information for sustainable forest management planning. Based on the difference in water availability and the number of frost days, Amaro et al. (1998) presented the first classification of eucalypt productivity across Portugal, including seven site classes. Later, these site classes were re-classified into two main productivity classes. The first, representing the high productivity class, can be generally found in the coastal areas being characterized by high water availability and low frost risk; whereas the second, the lower productivity class, including the inland regions of Portugal characterized by increased risk of frost occurrence no matter the water availability is high or not. Several works have been conducted to estimate different species' site index as a function of environmental explanatory variables by using statistical methods independent of the stand age. Specifically, some Portugal. For example, Paulo et al. (2015) assessed the site index of cork oak (*Quercus suber* L.) as a function of soil and climatic variables using Partial Least Squares regression. Fontes et al. (2003) used site factors (including soil, climatic and topographic factors) to model the productivity of Douglas-fir. The model created by digitized variables was then applied to develop the productivity distribution map across Portugal. In addition, Marques (1991) applied stepwise regression to predict the site quality of maritime pine (*Pinus pinaster* Air.) by using site factors, including temperature in autumn and soil characteristics (amount of potassium, porosity and the content of fine sand).

Apart from studies in Portugal, similar researches were applied to other woody species also found in other countries. To cite just some examples, Bravo-Oviedo and Montero (2005) used edaphic variables as predictors to estimate the potential productivity in stone pine (*Pinus pinea* L.) stands located in sandy areas in Spain by using Contingency tables and Correspondence analysis. Corona et al. (1998) applied MLR to examine the correlation between the site index of Douglas-fir in Italy and a group of environmental variables, including climatic, topographic, and soil variables. Regarding eucalypt, in Australia, the study of Grant et al. (2010) used some site variables (including available water storage capacity, mean annual rainfall, soil, altitude) and applied backward stepwise MLR to estimate the site index of *Eucalyptus dunnii*.

There are several obstacles when using environmental explanatory variables to assess the site index. First, some environmental variables show high multicollinearity among them (Weiskittel et al., 2011). This multicollinearity leads to bias when combining them in one equation. Second, as the site's factors can be synergistic or analogistic (Battaglia and Sands, 1998), their interaction is difficult to predict and integrate into one equation. The multiple relationships among factors may lead to developing a descriptive model adapted to the dataset used to create it, but not a general good predictive model for site index. Finally, by using statistical techniques to find the correlation between S and environmental factors, in some cases the model with the highest R^2 is not correct from a biological point of view studies had been conducted on the prevalent species in.

Moreover, some environmental factors that influence the dominant height growth may not be relevant in a multiple regression model that uses a combination of factors in one model to predict the site index. Therefore, it is necessary to have more studies on the species' biological traits and the interaction of the environmental characteristics. In Portugal, many researchers and pulp and paper companies have focused more on eucalypt plantation's growth and yield studies (Fontes et al., 2006; Tomé et al., 2006). A similar tendency is also found for eucalypt plantations in other parts of the world, such as Australia and Brazil.

3. Materials and methods

3.1. Materials

3.1.1 Site index

Data covering most regions where the pulp and paper companies' have eucalypt plantations under their management were selected from the EGLOBdata database. The EGLOBdata, managed by the ForChange group of the Forest Research Centre (CEF) of the Instituto Superior de Agronomia (ISA), contains information from various sources such as experimental trials, permanent and forest inventory plots established and measured by the ForChange Group, each company, or jointly. The location of the sample points was overlaid with the distribution of eucalypt pure and dominant stands according to the 2015 photo-interpretation (Figure 4) available at the Institute for Nature Conservation and Forests (ICNF) website (https://geocatalogo.icnf.pt/catalogo_tema3.html). It can be seen that the plots cover most of the area where the eucalypt plantations occur.

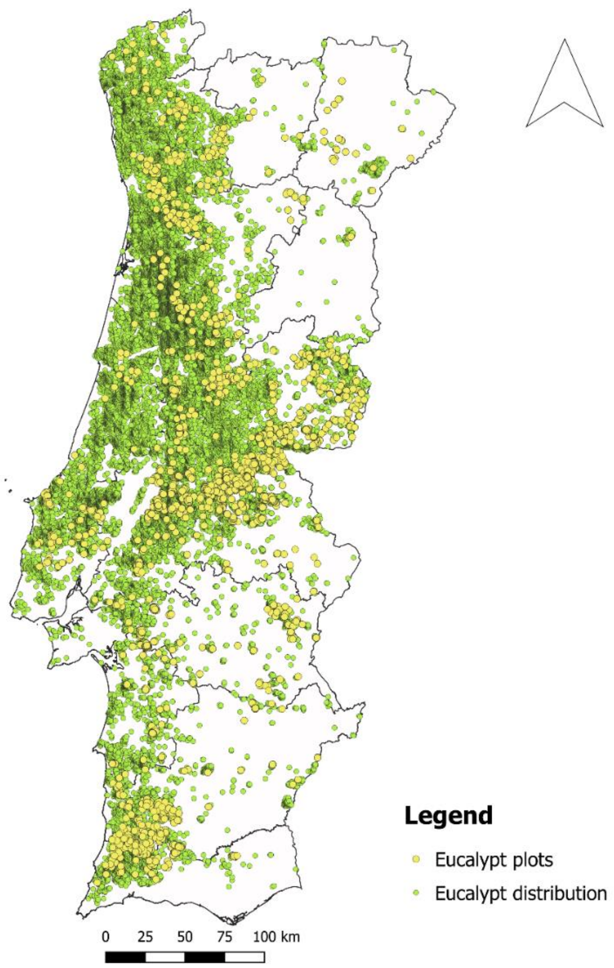


Figure 4- Distribution of pure and dominant Eucalyptus according to the photo-interpretation of 2015 (IFN6 grid of 500 x 500m) overlaid with the plots available for this research.

This study used a total of 166 permanent plots (PP), some of them measured for more than one rotation, 56 experimental plots (Alto do Vilão (AV), Quinta do Paço (QP) and Vilar de Luz (VL) spacing trials) and 28,504 forest inventory plots. Alto do Vilão trial combines spacing and thinning trial plots (Soares et al., 2001), Vilar de Luz and Quinta do Paço just spacing trial plots. The dominant height was calculated as the average height of the 100 thickest trees per hectare for all the dataset plots.

To estimate the site index for each plot and rotation, preference was given to the pair (stand dominant height, stand age) for which the age was closest to the eucalypt base age (10 years). The site index for each sample plot was estimated for a base age of 10 years using the difference equation developed by Tomé et al. (2006) (equation 1). Plots with the stand ages younger than five years or older than 12.5 years were discarded from the data set unless they presented unique soil and climate characteristics that were not covered by any other plot in the dataset.

$$S = (29.0669 + 0.2880 * \mathrm{DaysPrec.}) \times \left(\frac{\mathrm{hdom}}{29.0669 + 0.2880 \times \mathrm{DaysPrec.}} \right)^{\left(t/10 \right)^{0.4890}} \qquad \text{(equation1)}$$

In which:

S: site index value (m)

DaysPrec: annual average number of days with more than 1 mm of precipitation over the period of 1931 and 1960

hdom: dominant height of the eucalypt at the age t (m)

Figure 5 depicted the range of site index value of all the plots and the distribution by S classes.

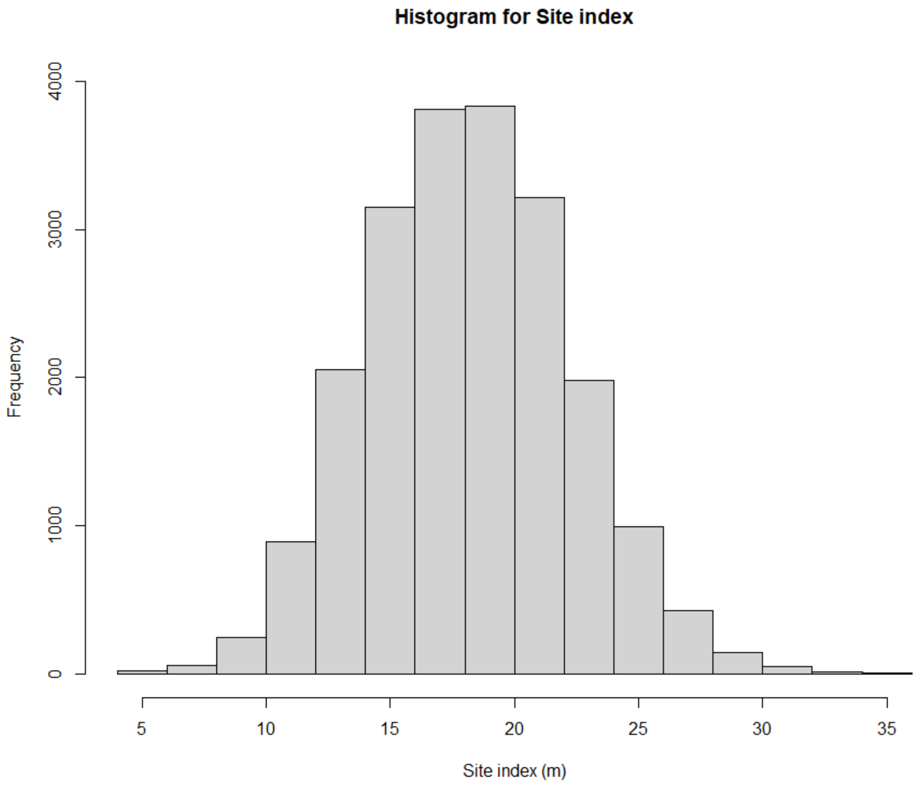


Figure 5- Distribution of research and inventory plots by site index class

From Figure 5, one can see the values of site index are a little skewed to the left. The site index estimates varied between 11.8 and 25 m at 5% and 95% percentile, respectively, with an average value of 18.2 m

3.1.2. Environmental variables:

The environmental variables in this study were categorized into three main groups: climatic, topographic, and edaphic. Table 5 summarizes all the environmental variables used.

Climatic variables were obtained from the digital cartography (Digital Environmental Atlas) available at the Portuguese Environment Agency (APA) website (https://apambiente.pt/index.php?ref=x178). Their values represent the annual average values for 30 years: 1931 to 1960. A total of nine climatic variables were used: evapotranspiration (calculated by the amount of water evapotranspirated), duration of frost season in the year (number of frost days), relative humidity, hours of insolation per year, number of days with more than 1 mm of precipitation per year, total annual precipitation, solar radiation (total amount of global radiation per year) and temperature.

Regarding topographic variables, the predictors were longitude, latitude, elevation, slope, and aspect. The information on slope and aspect in each plot were extracted from the Digital Elevation Model (DEM) of Portugal, with a resolution of 30 meters.

Edaphic variables in this study included soil type, lithology, water flow in soil, amount of underground water resources, and soil pH. Soil types were obtained from Land cover/use Statistics (LUCAS database). In this study, the soil types observed in the sample plots resulted in Cambisol, Podzol, Luvisol, Fluvisols, Lithosols, Planasols, and Podzol soils. These different soil groups were based on the classification of Food and Agriculture Organization (FAO). The remaining edaphic variables were collected from the digital cartography (Digital Environmental Atlas).

Table 5- The abbreviation of each environmental variable by variable group.

	Environment variable category	Abbreviation	Site variables	Unit	Data source
1	Topographic	LAT	Latitude	degree	EGLOBDatabase
2		LONG	Longtitude	degree	EGLOBDatabase
3		ELEV	Elevation	m	Atlas do Ambiente
4		SLOP	Slope	%	Atlas do Ambiente
5		ASPECT	Aspect	Categorial variables	Atlas do Ambiente
6	Edaphic	SOIL	Soil types	Categorial variables	LUCAS database
7		SOILPH	Soil pH		Atlas do Ambiente
8		RUNOFF	Water flow in soil	mm	Atlas do Ambiente
9		WATER	Underground water resources	$m^3/km^2.day$	Atlas do Ambiente
10		LITHO	Lithology	Categorial variables	Atlas do Ambiente
11	Climatic	ET	Evapotranspiration	mm	Atlas do Ambiente
12		FROST	Number of frost days	day	Atlas do Ambiente
13		FROSTM	Number of frost months	month	Atlas do Ambiente
14		HUMI	Humidity	%	Atlas do Ambiente
15		INSO	Insolation	hour	Atlas do Ambiente

16		PRECD	Number of precipitation days	day	Atlas do Ambiente
17		PRECT	Total precipitation	mm	Atlas do Ambiente
18		RADI	Global radiation	kcal/cm^2	Atlas do Ambiente
19		TEMP	Temperature	ºC	Atlas do Ambiente

The categorical variables (including soil type, lithology, and aspect) were grouped into smaller groups, represented in Table 6.

Table 6- Sub-grouping of categorical variables.

Categorial variables	Group	Plots in the category (%)
Soil type	Cambisols	20.35
	Fluvisols	0.29
	Lithosols	45.11
	Luvisols	12.92
	Planosol	0.24
	Podzol soil	21.09
Lithology	Sandstone (Arenitos)	0.10
	Alluvial (Aluvioes)	25.70
	Limestone (Calcarios)	1.03
	Granite (Granitos)	8.14
	Xitos	60.53
	Outros (Others)	4.17
Aspect	N (North)	11.33
	NE (Northeast)	9.55
	E (East)	10.18
	SE (Southeast)	12.26
	S (South)	9.31
	SW (Southwest)	16.27
	W (West)	18.31
	NW (Northwest)	12.80

3.2. Model building

3.2.1. Site index estimation and preliminary data analysis

Before fitting the model, it was necessary to understand more about the whole dataset. The initial data set, which included information about the age and the dominant height of the eucalyptus plots, was carefully assessed, and some plots were eliminated. Firstly, the data of a few clonal plots were removed, the future model will apply to seminal plantations. Secondly, due to the wildfires that occurred in 2008, many plots were affected. The last age and dominant heights collected in some plots were far from the base age (some plots were only 4 years old). Thus, if these data were used in the model, it could affect the final result of the model.

For this reason, these plots with age out of the range from 5 to 12.5 were eliminated. Moreover, plots with no coordinates data (274 values) and missing aspect (139 values) were eliminated from the dataset.

The remaining plots were used to estimate the S value using equation 1 from the Tomé et al. (2006).

The correlation between the site index and each environmental variable was also analyzed. For the continuous variables, Pearson's correlation was applied. Pearson correlation was used to evaluate the relationship between two continuous variables. The value of Pearson correlation ranges from -1 to 1. If the absolute value of Pearson correlation is closer to 1, the greater the relation between two variables. When Pearson correlation is equal to 0, no correlation exists. The sign of Pearson correlation implies the relationship is positive or negative. For the categorical variables, a box plot analysis was used to assess how each variable in one category can influence the site index.

3.3. Multiple linear regression

Multiple linear regression is a statistical technique that assumes a linear relationship between a dependent variable and one or more predictor variables. Indeed, MLR is the ordinary least-squares. It has been applied in various study fields, such as mathematics, economics, and forestry. Regarding using MLR statistical technique to estimate the site index of eucalypt, in Australia, the study of Grant et al. (2010) used some site variables and applied backwards stepwise MLR to model the site index of *Eucalyptus dunnii*. Using this approach, the model developed does not require a growth function developed using dominant height and stand age. This study found a reasonable correlation, as 35% of the model's observed variability was explained. Until now, there have been just a few studies for modelling site index for eucalypt in Portugal exclusively from environmental variables and independent of dominant height and stand age (PROF of Centro Litoral, Centro Interior and Algarve).

One major issue when applying MLR is that this method only depicts the response variables at the conditional mean instead of all the quantiles. In some cases, this is not enough to represent a comprehensive relationship between predictors and response variables. For this reason, in this study, QR, a useful statistical tool that allows calculating the regression functions for different conditional quantiles (Liu and Wu, 2009), was combined with MLR.

The model to estimate the site index for eucalypt was fitted by using multiple linear regression. The categorical variables became dummy variables in the equation. The modelling process was conducted using the R statistical software within the R software (R Development Core Team, 2013). The relationship between site index and other environmental variables was then assumed to follow the equation two:

$$S = \varepsilon + \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \qquad \text{(equation 2)}$$

In this equation, S stands for site index; X_i stands for the environmental variables; β_i stands for the regression coefficients, and ε is the model error.

As mentioned before, firstly, the model development was made in 4 stages, the first three looking for the best models with each group of variables: climatic (M_1), topographic (M_2) and edaphic (M_3). This way, we got an idea of which variables within each group were more relevant to explain site index. Then, the 4th stage implied the search for a model with all the variables (M_i)

Secondly, with all variables, the best model - the integrated model - to predict the site index was found using stepwise regression and the Akaike Information Criterion (AIC). These models were then compared by these criteria:

- Adjusted R^2 and the residual sum of squares, which were used to assess the fitting ability.
- Akaike Information Criterion (AIC) that, in a set of models, estimates the quality of each model, relative to each of the other models, being therefore very useful in model selection.
- The regression assumption – normality of the residuals and heteroscedasticity, which was assessed through graphical residual analysis. The normality of the residuals and the homogeneity of the variance of studentized residuals were evaluated, respectively, by normal QQ-plots and plots of studentized residuals over predicted values.
- Variance Inflation Factor (VIF). The collinearity among predictors was checked by using the VIF. The higher the value of VIF, the higher the collinearity among the predictors. When using a set of predictors with high collinearity, the coefficients of the models are dependent of the values of the chosen predictors. Subsequently, when applying the model to predict the site index at a new point out of the data range (extrapolation), the results can be very poor (Myers, 1990).

- Analysis of the biological meaning of the model. Even avoiding the collinearity among the predictors, some coefficients might have an opposite sign to the biological rationale, which is not desirable.
- Model validation: The bias and precision of the models must be assessed with an independent data set or using resampling techniques or leave-one-out techniques. Here we opted by using the press residuals or prediction residuals, a leave-one-out technique. In calculating press residuals, each observation is deleted in turn, and then the model fitted with the remaining observations. Then, each obtained regression equation is used to calculate the predicted value, subsequently the press residual (Myers, 1990). Several statistics can be computed with the press residuals (rp_i). Usually, the bias is evaluated with the average of the press residuals (equation 3):

$$bias = \frac{1}{n}\sum_{i=1}^{n} rp_i \qquad \text{(equation 3)}$$

  And the precision with the average of the absolute value of the press residuals (equation 4):

$$precision = \frac{1}{n}\sum_{i=1}^{n} |rp_i| \qquad \text{(equation 4)}$$

- Modelling efficiency (ME_p) presents the model performance through a relative scale, ranging from 0 (showing that the model is not better than the simple average model) to 1 (perfect fit) (Soares et al., 1995; Jerome K Vanclay and Skovsgaard, 1997). The equation of ME_p is presented below (equation 5):

$$ME_p = 1 - \frac{\sum_{i=1}^{n} rp_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad \text{(equation 5)}$$

3.4. Quantile regression (QR)

QR was first introduced by Koenker and Bassett in 1978 (Koenker and Bassett Jr, 1978). It has then been studied and applied in various fields, such as medicine (Cole and Green, 1992; Heagerty and Pepe, 1999) and economics (Hendricks and Koenker, 1992; Koenker and Hallock, 2001). Recently, it was applied in forest growth modelling to model height-diameter relationships (Xie et al., 2020) and study the relationship between stem growth and climate (Hinko-Najera et al., 2019).

Quantile Regression supports and improves the usual linear regression. For example, when linear regression conditions are violated, such as lack of homogeneity, QR can be used to measure the heterogeneous effects of covariates to the outcome at different conditional quantiles (e.g. 0.05, 0.25, 0.5, 0.75, 0.95 quantiles) (Huang et al., 2017). Quantiles are defined as points collected at regular intervals from the cumulative distribution function of a random variable. By doing this, QR can present a complete depiction of the whole distribution of the outcome. Moreover, when the sample is not symmetric and has heavy tails, the mean value is no longer sufficient to depict centrality. In this case, the sample median (the 0.5 percentile) can represent the centrality better than linear regression.

Additionally, QR is more effective when the sample has many outliers. Moreover, the conditions to use QR are less strict than those applied in traditional mean regression modelling because there are no assumptions for distributing the outcome. Thus, when linear regression assumptions are violated, QR is a robust alternative that extends and completes traditional regression. Due to its advantages, the number of studies applying QR has been increasing recently.

Multiple linear regression was first conducted to estimate the mean site index, which was not thought as enough to describe the relationship between site index and other predictors thoroughly. Thus, the predictors obtained from the integrated model of multiple linear regression were used in QR that allowed estimating the value of S at the median (50%) and other quantiles (95%, 75%, 25%, 5%). The equation for the site index is now (equation 6):

$$S = \varepsilon + \beta_{0T} + \beta_{1T} X_1 + \beta_{2T} X_2 + \cdots + \beta_{nT} X_n \qquad \text{(equation 6)}$$

In this equation, S stands for site index; X_i stands for the environmental variables; β_{iT} stands for the regression coefficients; and ε is the model error.

QR modelling was conducted in the R software (R Development Core Team, 2014) by using the "quantreg" package (Koenker, 2013).

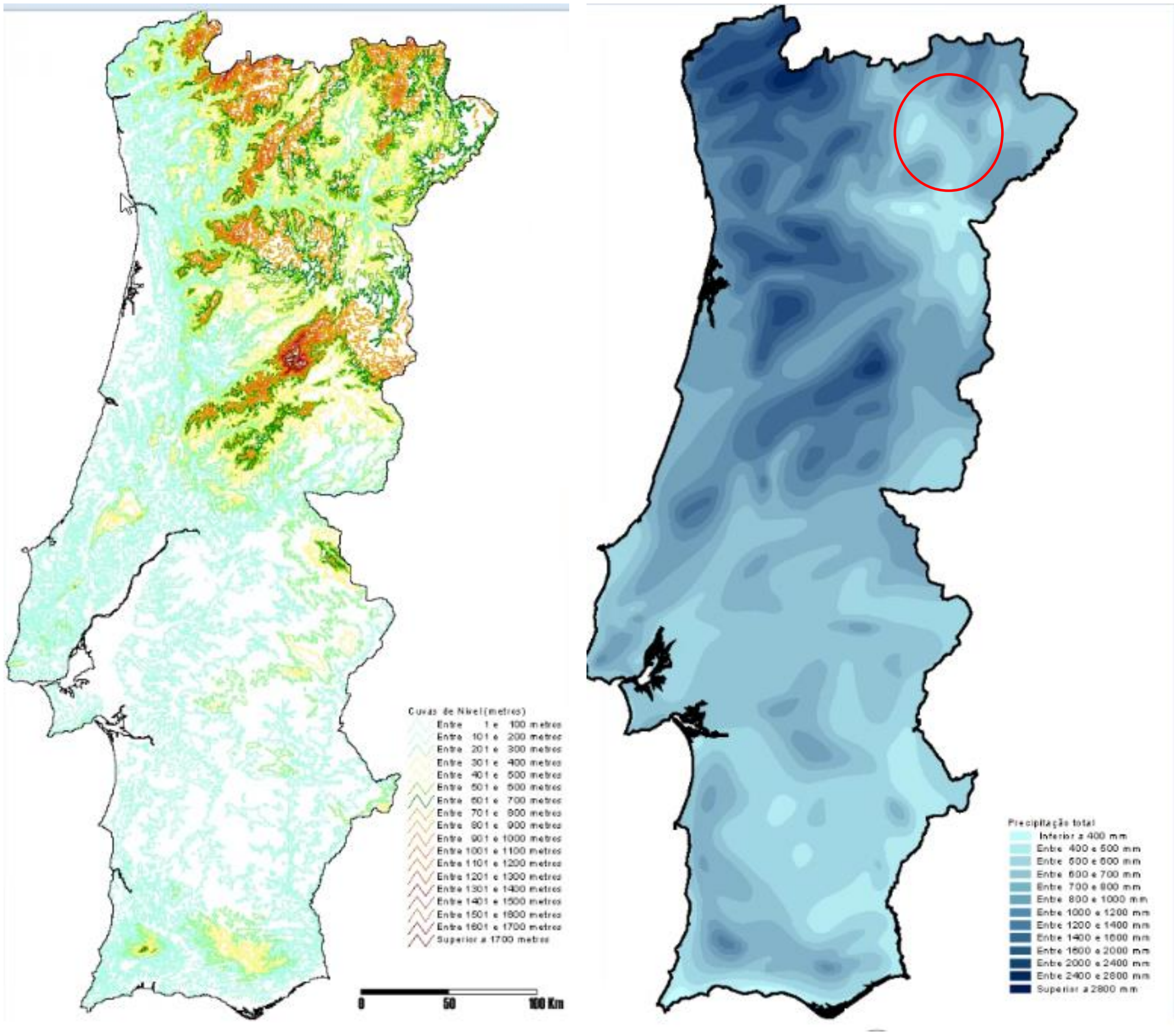
The integrated model (M_i) was used to conduct the quantile regression to observe how the sign and the value of each predictor changed for the different quantiles.

3.5. Site index mapping

The integrated model M_i was used to create a potential productivity map for Portugal. Each predictor's data were extracted separately to the shapefile containing points with the distance 500 x 500m (grid of photo interpretation in the Portuguese National Forest Inventory), then each shapefile was rasterized. Based on the equation of models M_i, raster calculations were applied to each raster file to reproduce 5 quantile, integrated model, and 95 quantile maps. The outcome maps indicated an interval of potential site index in Portugal.

However, not all of the area is suitable for the development of eucalypt. Based on expert knowledge of restrictive conditions to the development of eucalypt were added as masks to the productivity maps. Specifically, eucalypt is not often found at an elevation higher than 600m. Thus, these regions were assigned a value of zero in the mask. Moreover, the suitable precipitation threshold for eucalypt development was assumed to be above 600mm. For this reason, and because climatic data for precipitation was available in classes, two thresholds were considered when setting up the mask: < 600 and ≤ 600, with locations having precipitation below these thresholds being assigned a value of 0. Two maps based on the combination of the thresholds set for elevation and precipitation were prepared. Map 1 with a mask where altitude > 600m was combined with precipitation ≤ 600mm and Map 2 with a mask where altitude > 600m was combined with precipitation < 600mm. Both maps were validated by the comparison with the altitude and the total precipitation maps from Atlas do Ambiente (Figure 6). Additionally, a map with the distribution of mean annual increment of eucalypt plots produced by one of the Portuguese pulp and paper companies was also used to support the mask selection and validate the model results (map available at http://www.thenavigatorcompany.com/var/ezdemo_site/storage/original/application/0cd66a2916b723fb55dd179edf75c723.pdf).

After that, this chosen mask was used to create a map of the integrated model after applied quantile regression at 5 and 95 percentiles. Finally, these two maps were used to compare with the final map of the integrated model M_i.



a) Altitude map for Portugal

b) Precipitation map for Portugal

Figure 6- Total precipitation and altitude maps for Portugal. Source: Atlas do Ambiente.

4. Results

4.1. Preliminary analysis for all variables

4.1.1. For the continuous variables

The results of Pearson's correlation coefficients for pairs of continuous variables are summarized in Figure 7.

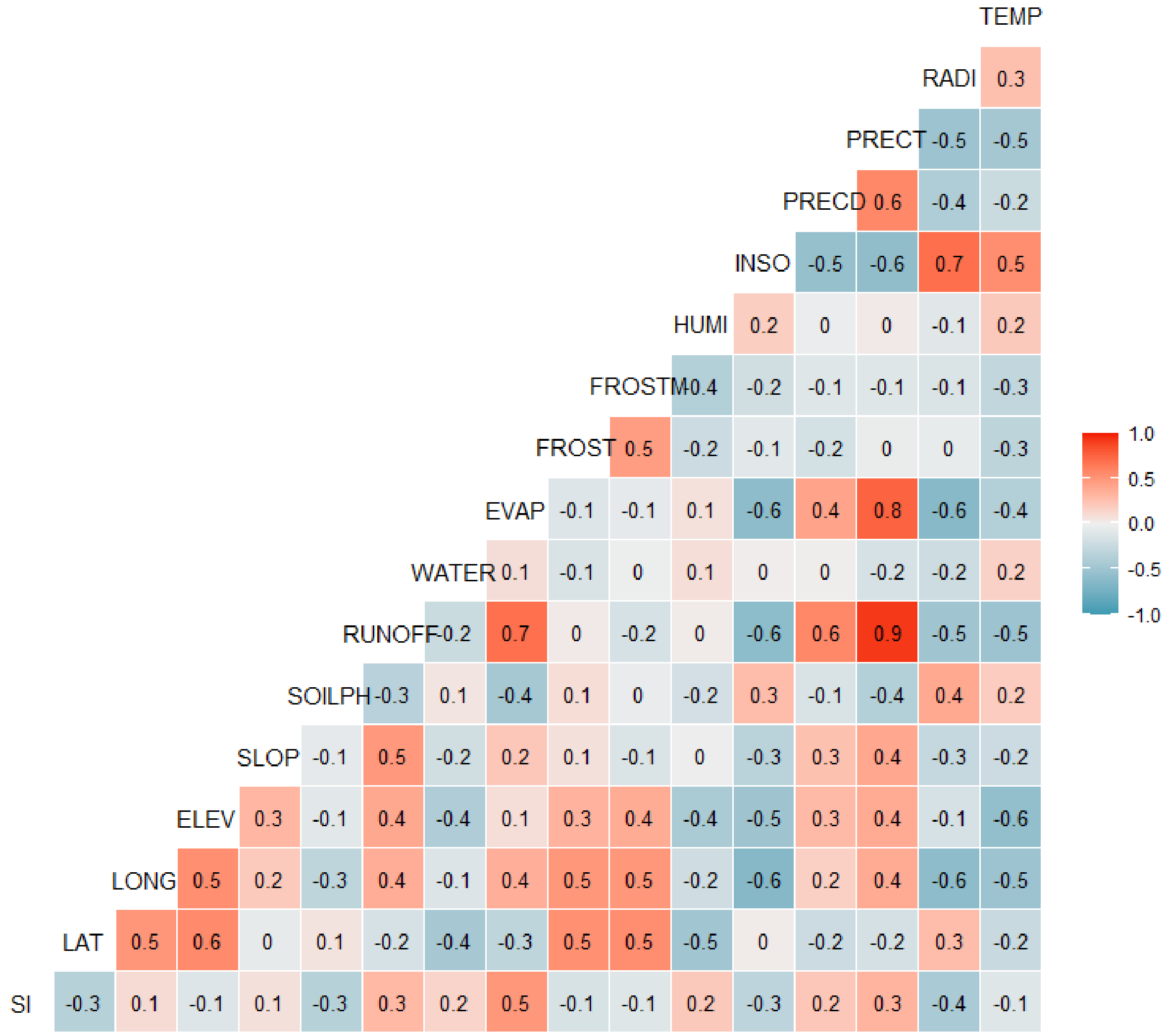


Figure 7- Correlation between S and each of the continuous environmental variables

From Figure 7, evapotranspiration (ET) had the highest statistical correlation with the site index (0.5). This relationship was positive, meaning that the higher the evapotranspiration rate, the higher the value of the corresponding site index. On the other hand, longitude (LONG), elevation (ELEV), Slope (SLOP), number of frost days (FROST), number of frost months (FROSTM), and temperature (TEMP) showed the lowest correlation with site index. Among the continuous variables, total precipitation (PRECT) and run-off (RUNOFF), total precipitation (PRECT) and evapotranspiration (ET) had the highest correlation with each other.

4.1.2. For the categorical variables:

Figure 8, Figure 9, and Figure 10 describe the correlation between categorical variables and site index. In Figure 8, site index was not influenced much by different azimuths, as the mean value of SI in each group was not significantly different, around 20m.

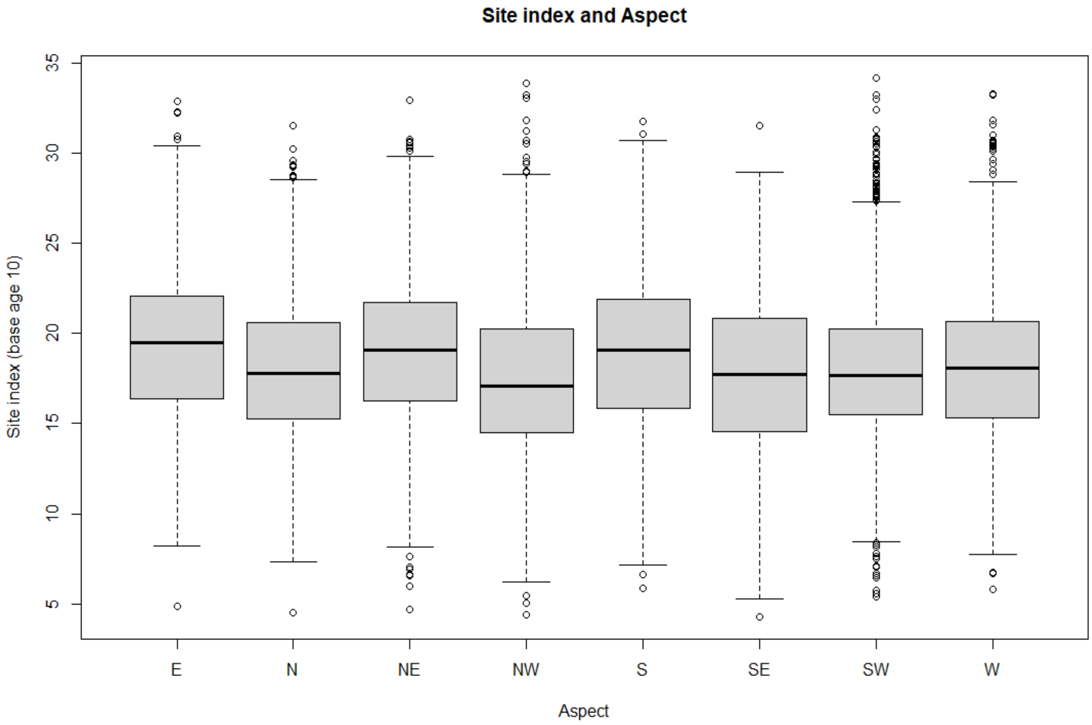


Figure 8- Box-plot between S and each categorical variable describing aspect.

However, site index was affected by different lithology and soil types. Regarding lithology, the highest mean value of site index was observed in alluvial soil, and the lowest one was in the “others” group, which were approximately 22 and 16m, respectively (Figure 9).

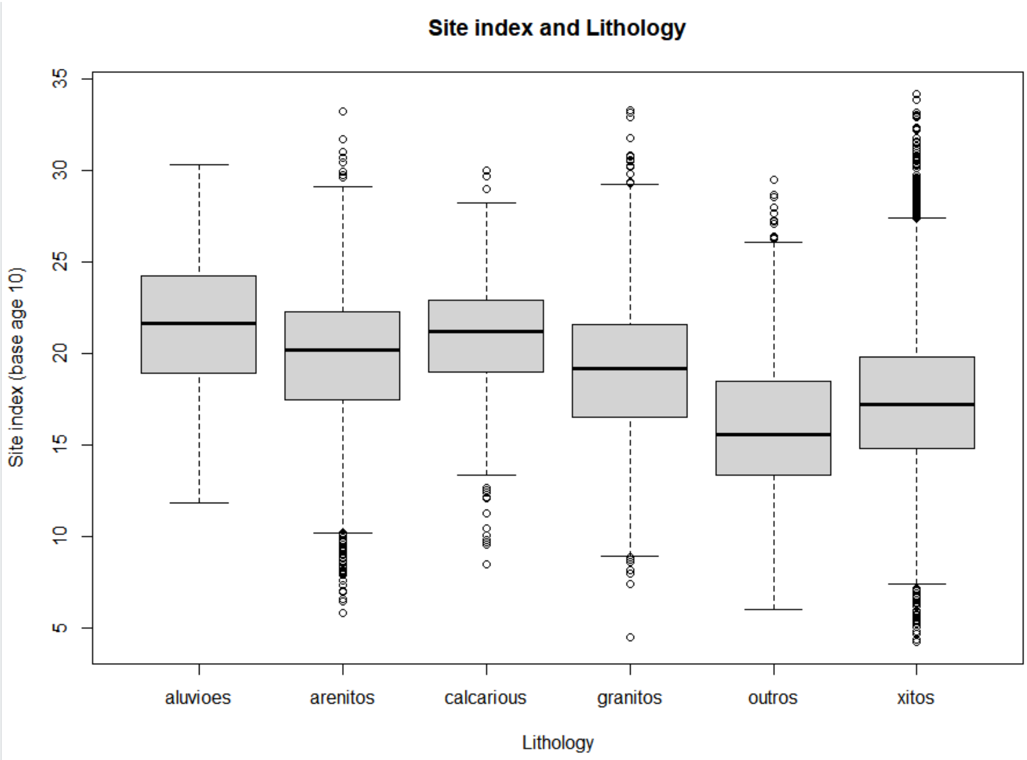


Figure 9- Box-plot between S and each categorical variable describing lithology.

Regarding soil type, cambisols, fluvisols, and podzols had a similar highest value of site index, whereas lithosols, luvisols and planosols had nearly the same lowest value of site index.

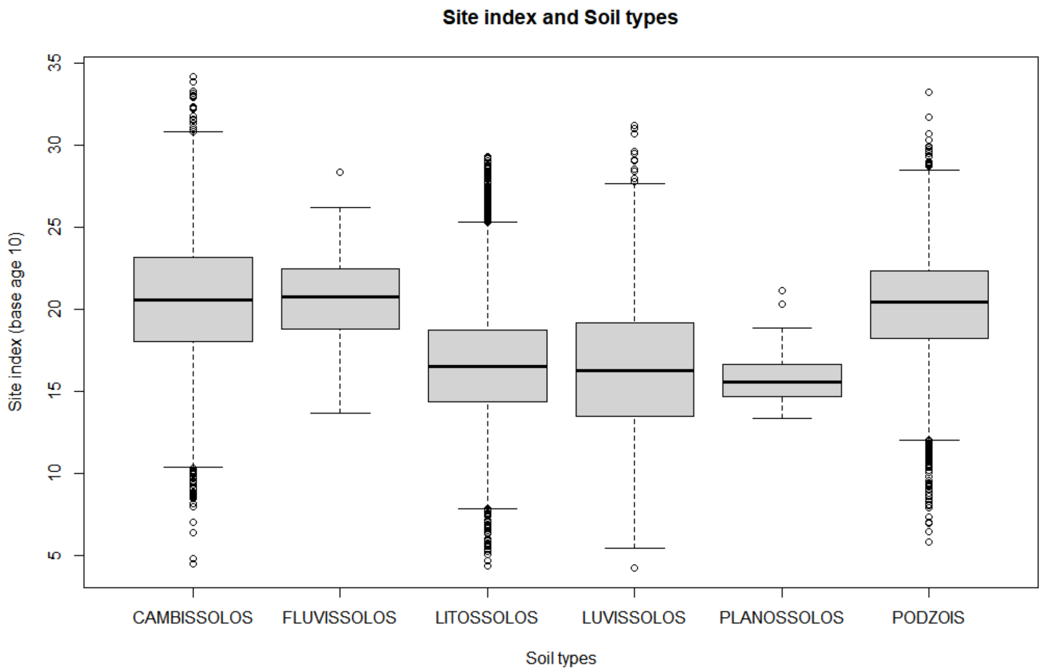


Figure 10- Box-plot between S and each categorical variable describing soil types

4.2. Multiple linear regression

Several alternative models can be used to predict the site index in the study area with a similar accuracy. However, according to the selection criteria for stepwise regression (AIC), the best model was chosen for each group of variables and considered all the evaluation criteria presented in section 3.3.

All the coefficients in the four models presented in Table 7 are statistically significant at the p-value <0.05. The VIF values for all the variables in the four models were smaller than 5. The integrated model was the best model that explained 32.22% of the site index variation among the four models. In comparison among the three groups of predictors topographic, climatic, and edaphic), climatic variables (model M_1) had the highest effect on S (explaining 29.93% of the variance), then followed by the edaphic variables (M_3) describing 23.69% of the variance. The topographic variables explained only 22.69% of the variance. However, the integrated model included variables from all three groups.

Table 7- Result of multiple linear regression for the model with each group of variables

	Coefficients	R-square	Adjusted R-square	Residual sum of squares	AIC	VIF max	ME	PRESS
Climatic model (M_1)		29.93	29.92	240395.4	110282.7	1.67	0.299	240517.3
Intercept	22.201							
ET	0.018							
HUMI	0.130							
RADI	-0.156							
FROST	-0.019							
Topographic model (M_2)		22.69	22.67	265244.7	112338.6	1.34	0.226	265415.2
Intercept	24.66							
LAT	-4.46×10^{-5}							
LONG	1.27×10^{-5}							
SE	-0.672							
NW	-0.621							
S	0.369							
Edaphic model (M_3)		23.69	23.67	261830	112066	1.60	0.236	261983.9
Intercept	18.053							
SOILPH	-0.365							
RUNOFF	0.006							
arenitos	1.942							
LITOSSOLOS	-1.247							
Integrated model (M_i)		32.24	32.22	232479.3	109587.6	1.99	0.322	232646.9
Intercept	18.693							
ET	0.017							
LITOSSOLOS	-1.488							
LUVISSOLOS	-1.535							
HUMI	0.078							
RADI	-0.097							
ELEV	-0.002							

The plot of quantile-quantile (QQ-plots) of the four models was used to assess the normality of the residuals (Figure 11).

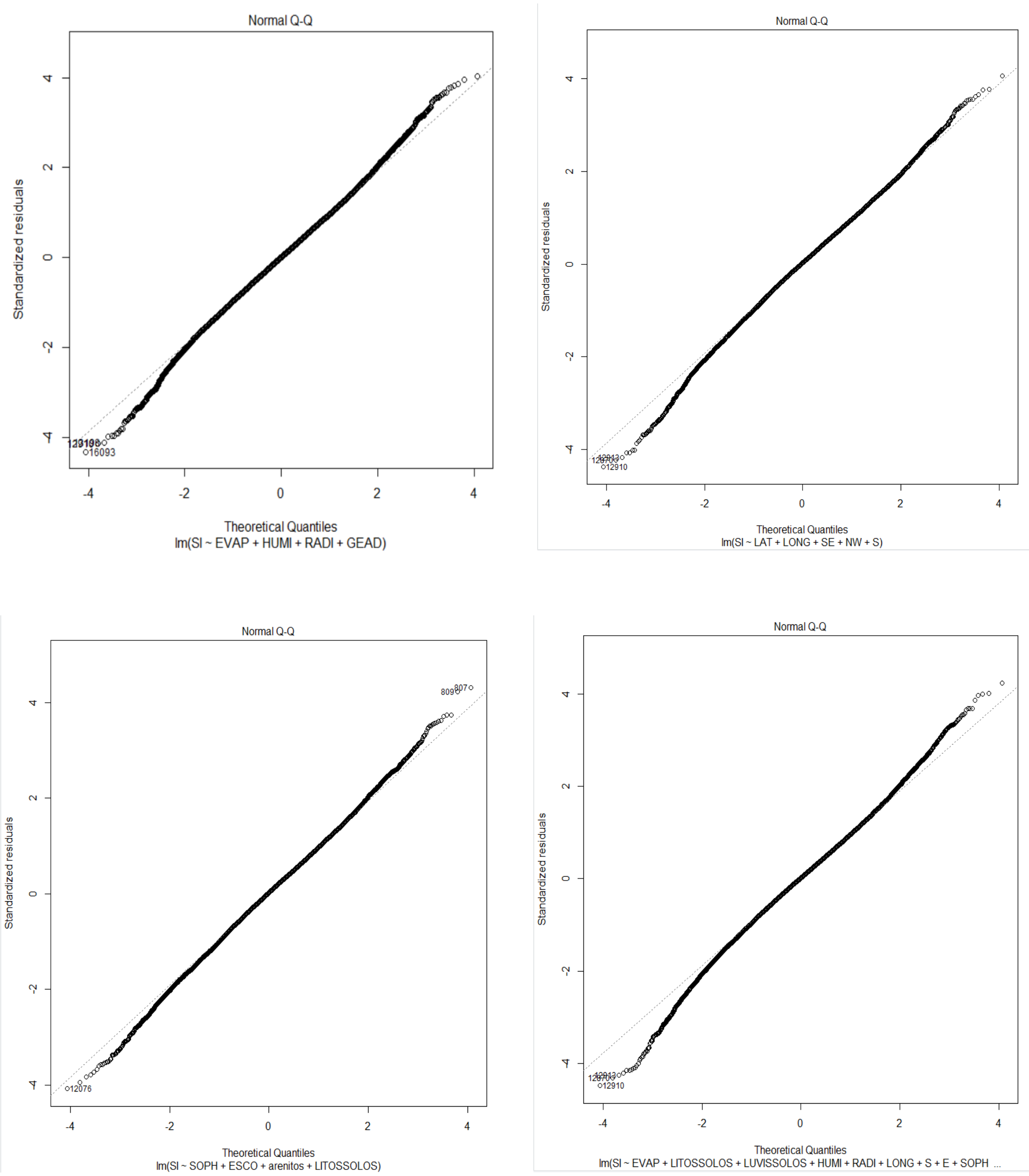


Figure 11- QQ-plots of the fitted models: a) Model including climatic variables (M_1); b) Model including topographic variables (M_2); c) Model including edaphic variables (M_3); and d) Model including all types of variables (M_i).

From the QQ-plot of the four models, it can be concluded that the residuals of all models are slightly skewed left, but the distribution of the error does not strongly deviate from the normal distribution.

The plots of studentized residuals against fitted values were used to assess the homogeneity of variance of studentized residuals (Figure 12). None of the plots showed any violations of this statistical assumption.

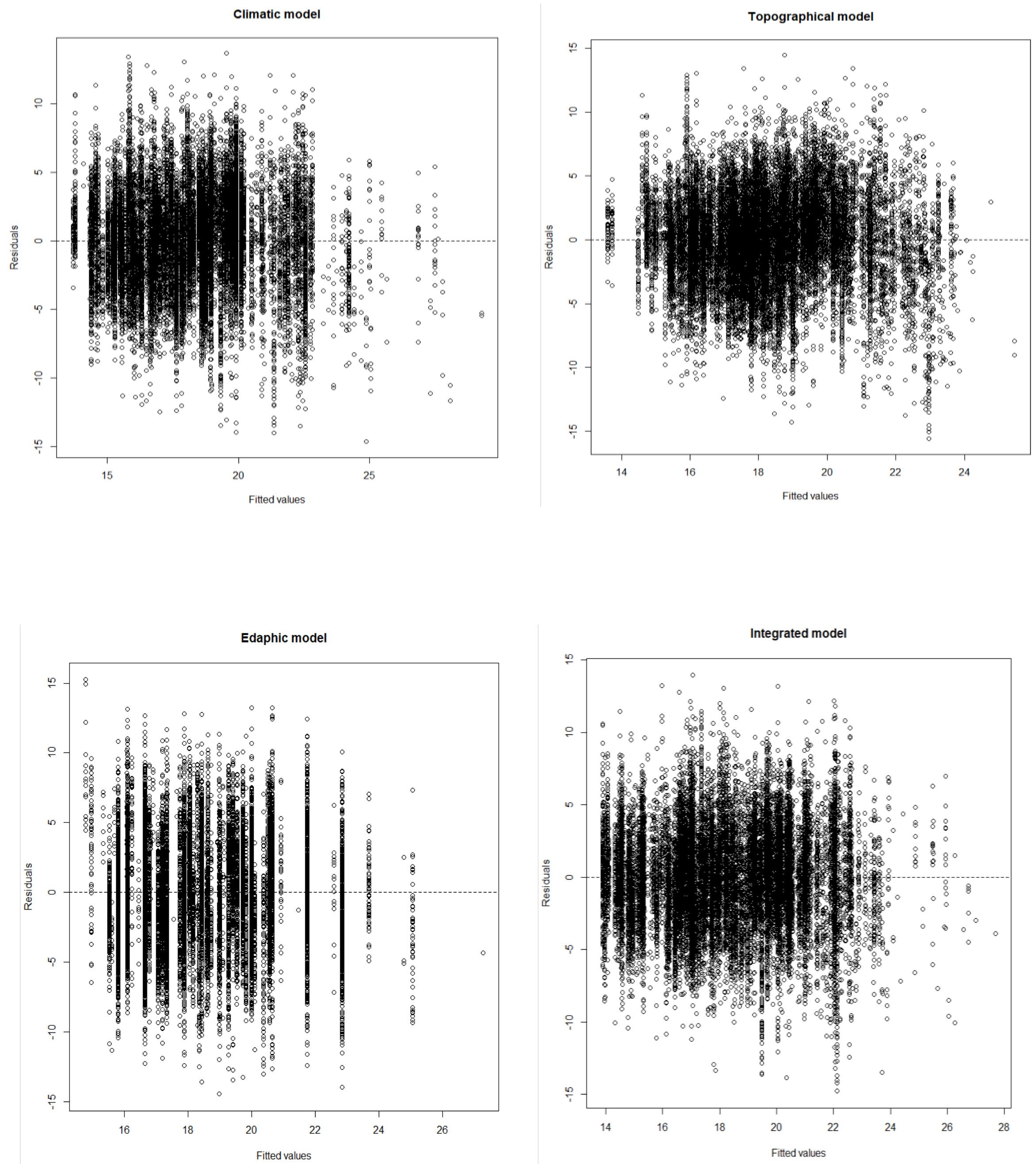


Figure 12- Plots of studentized residuals and fitted values for all models

4.3. Quantile regression

Table 8 showed the results of all coefficients at the different quantiles. The bold values were the coefficients that were statistically significant at the t-test with α < 0.05.

Table 8- The result of quantile regression applied to the integrated model.

τ (tau)	0.05	0.25	0.5	0.75	0.95
Intercept	1.052	**15.461**	**18.962**	**25.081**	**25.280**
ET	**0.016**	**0.016**	**0.017**	**0.018**	**0.019**
LITOSSOLOS	**- 0.832**	**-1.622**	**-1.759**	**-1.588**	**-1.189**
LUVISSOLOS	**-1.216**	**-2.140**	**-1.838**	**-1.373**	**-0.849**
HUMI	**0.079**	**0.069**	**0.074**	**0.086**	**0.111**
RADI	-0.020	**-0.080**	**-0.096**	**-0.130**	**-0.131**
ELEV	**-0.001**	**-0.002**	**-0.002**	**-0.002**	**-0.001**

Figure 13 showed the plots of the quantile regression coefficients for the quantiles τ of the site index. The parameters include the intercept of the model, evaporation (ET), lithosols (LITOSSOLOS), luvisols (LUVISSOLOS), humidity (HUMI), radiation (RADI), and elevation (ELEV). The regression coefficients represent the change of the coefficients of the predictors in each quantile. The grey shading depicts 95% confidence intervals of the coefficient. The solid red line is the estimate from MLR.

For the integrated model, outliers were observed mostly below the lower quantile (from 0 to 25 percentile) since the confidence intervals of the coefficients were more expanded than this at the onwards quantile. In most of the variables (except lithosols and luvisols), the value and the range of confidence, intervals were similar in MLR and 0.5 quantiles. The sign of the coefficients of all variables remained the same compared with the sign of the coefficients in MLR in all quantiles. In specific, evapotranspiration and humidity were positively correlated with the site index and an increasing value of the slope is visible through all the quantiles. These results showed the higher impact of these variables for the best sites. On the other hand, the coefficients for lithosols, luvisols, radiation, and elevation were negatively correlated with the site index, showing that they are more important in the poor sites.

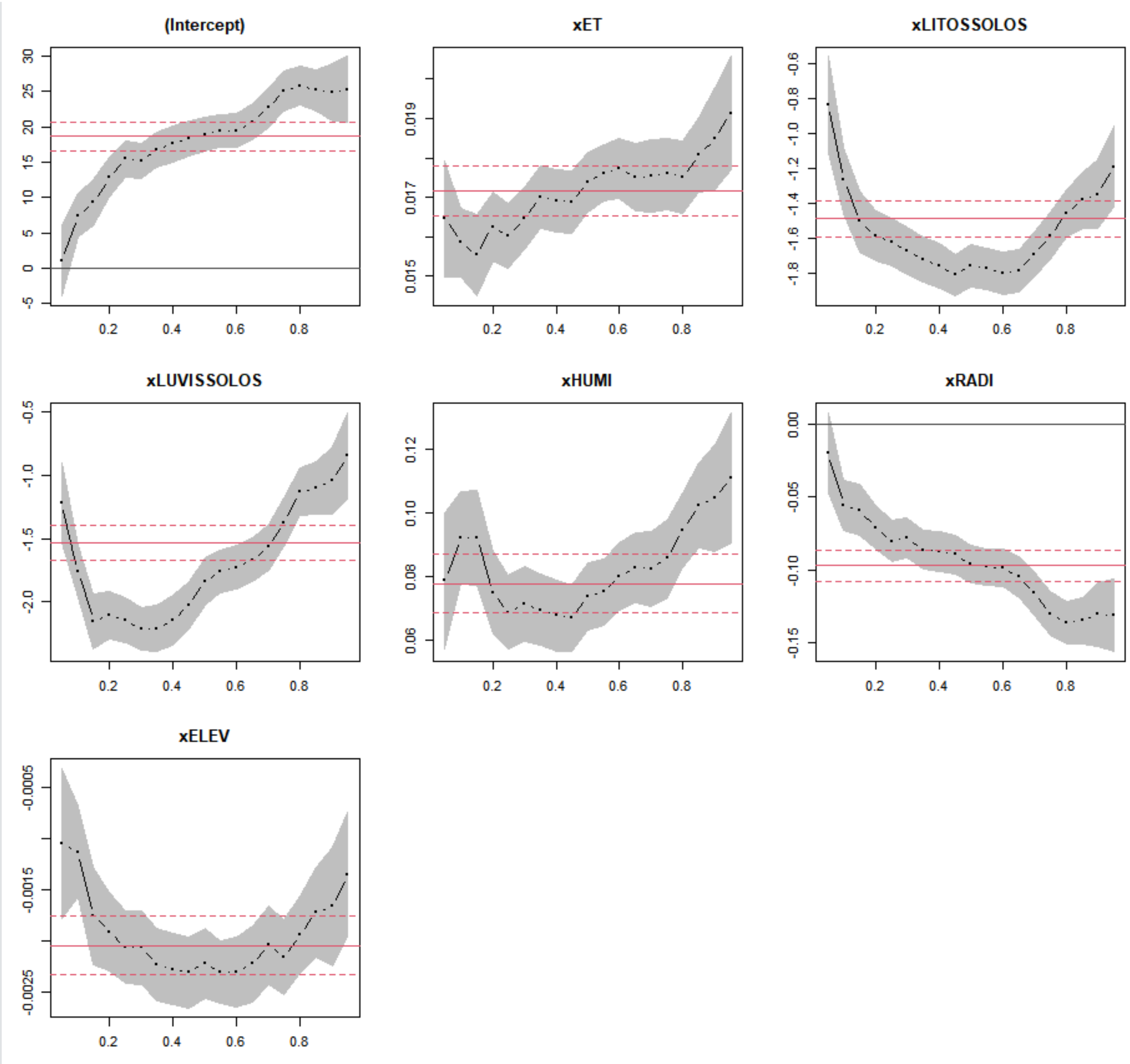


Figure 13- Confidence intervals for the parameter estimates by quantile regression for the variables in model M_i

4.4. Site index map for Portugal

The best-fitted model obtained with MLR, the integrated model, was used to estimated S within QGIS for each pixel of 500 x 500m NFI6 grid. The sample plots were represented by dots, with different colors indicating the different observed site index values. The areas where the elevation was higher than 600m combined with the total precipitation was less than 600 mm or less than and equal to 600mm were used as a mask for the eucalypt distribution area. The restricted regions were compared in Figure 4.

Compared with the distribution of eucalyptus map in Figure 4, the restricted area considering a precipitation ≤ 600m is more similar than the one with a precipitation < 600mm. For this reason, the map in Figure 14a was chosen as the final map. The site index values were classified as “not appropriate” if outside the mask, “very poor” (site index value <=14 m), “poor” (with the site index value ranging from 14-17m), “average” (site index value ranged from 17-20m), “good” (site index value ranged from 20-23m), and finally, “very good” (site index value was more than 23m).

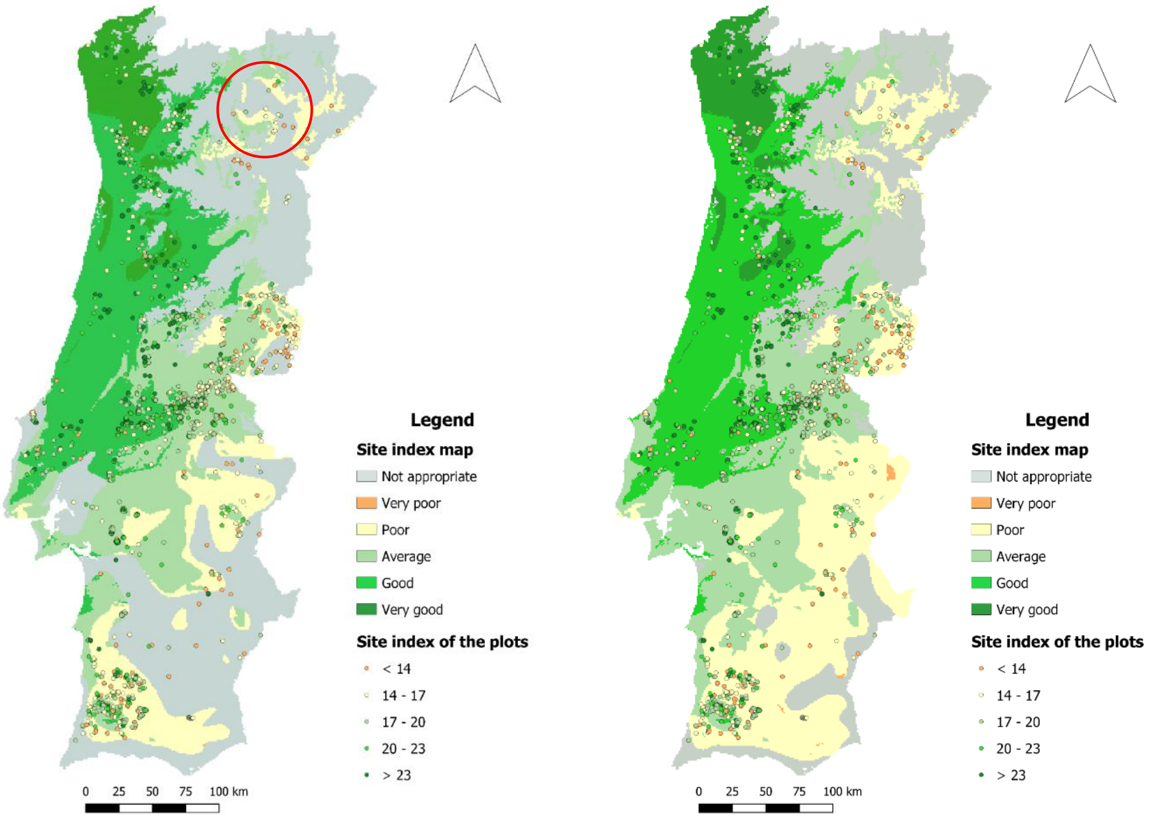


Figure 14- a) Site index variation map of eucalypt for Portugal with a mask resulting of the combination of altitudes >600 and precipitation ≤ 600m; b) Site index variation map of eucalypt for Portugal with a mask resulting of the combination of altitudes >600 and precipitation < 600m.

From the map, the best location for eucalypt growth is, as it is well known, along the coastal areas. In the North west of Portugal, site index value was highest (> 23m). In the centre of Portugal, the site index values were in the middle range, from 14 to 20m. In general, horizontally, from the west (coastal areas) to the east (inland) of Portugal, the site index value decreased gradually.

Using the GLOBULUS model we can calculate the mean annual increment for each site index (Figure 15).

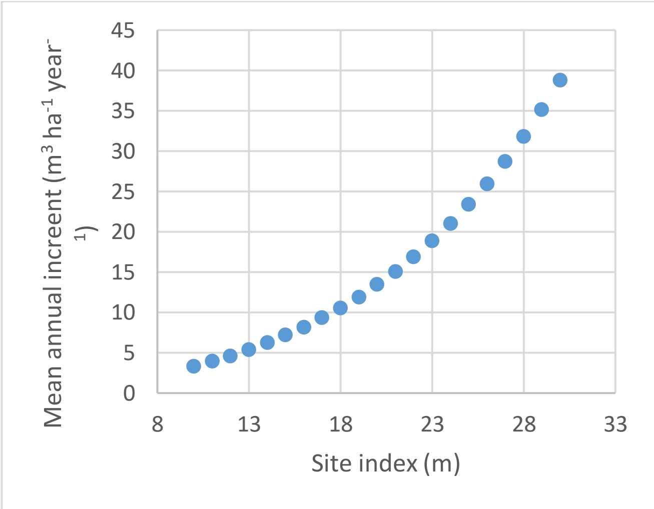
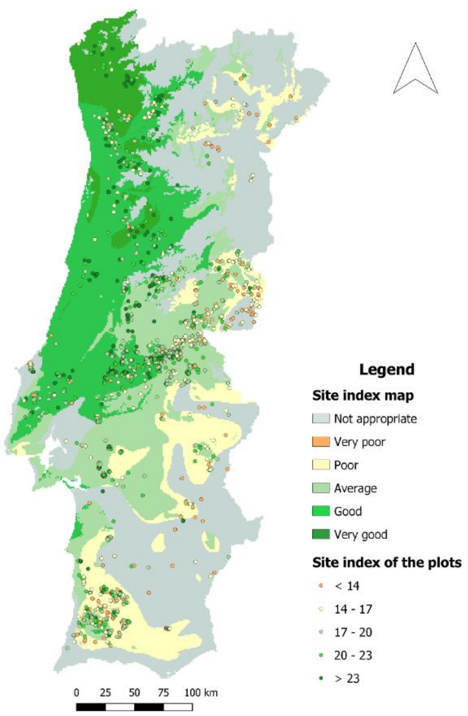
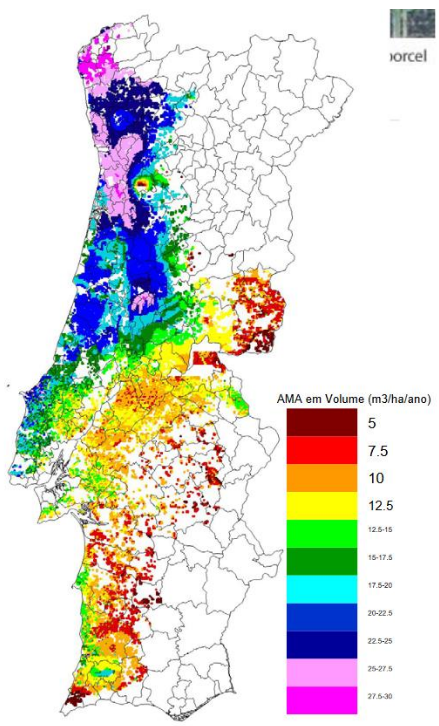


Figure 15- Mean annual increment for each site index

The pulp and paper company’s map with the distribution of plots by mean annual increment (MAI) in volume support the map produced in this research (Figure 16). Not only the areas evidencing the highest MAI correspond to the areas classified as “Good” and “Very good”, considerable correspondence was also found for the poorer site index estimates and the lowest MAI values.



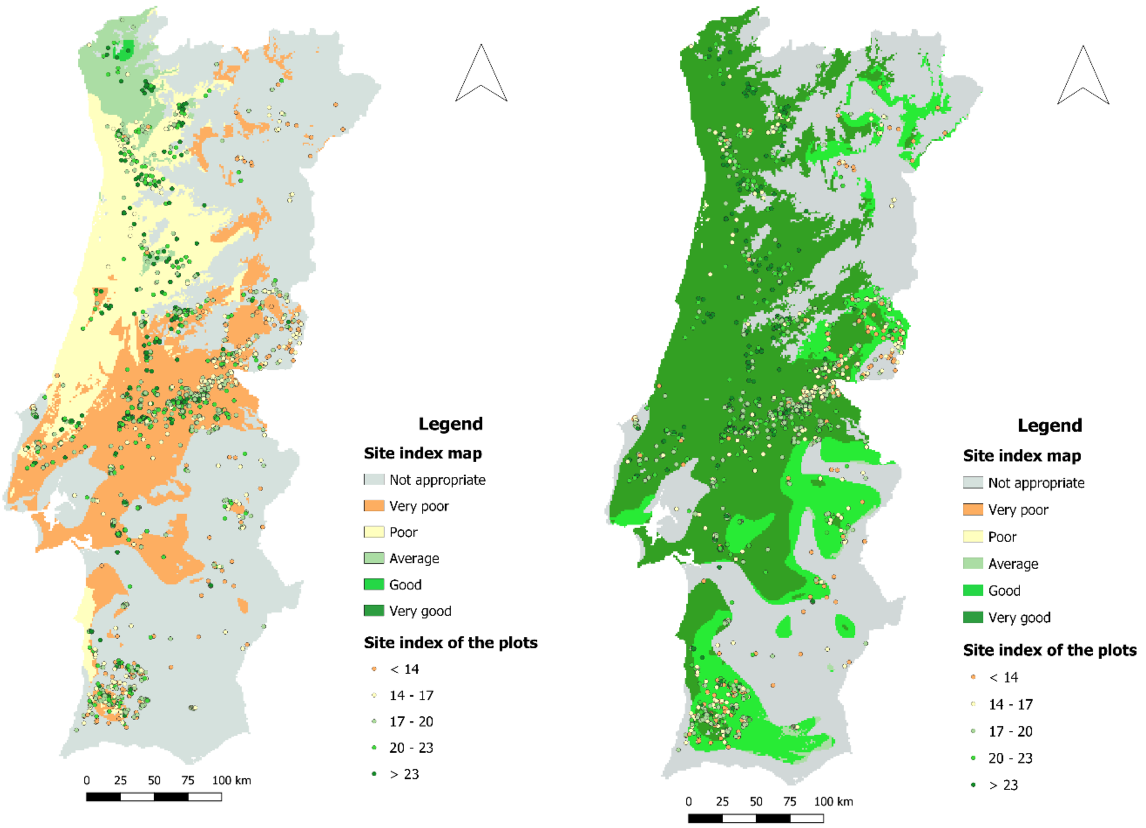
a) Final eucalypt distribution map



b) Distribution of plots by mean annual increment volume

Figure 16- Comparison of the estimated distribution of eucalypt productivity using multiple linear regression with the distribution of the set of sample plots of a pulp and paper company with the mean annual increment in volume in classes shown in different colors.

When applied the quantile regression to the integrated model, the site index variation map at 5 percentile and 95 percentile were also illustrated to compare with the site index variation map (Figure 17).



a) Site index variation map at five percentile b) Site index variation map at 95 percentile

Figure 17- Site index variation map at a) 5 percentile, and b) at 95 percentile.

It is noticeable that the site index variation map at 5 percentile mostly presented the site index classified from “Not appropriate” to “Average” categories; whereas, in the map at 95 percentile, most of the classification were from “Good” to “Very good”. As can be seen, there is a very high variability in the same local that depends on variables that were not considered to develop the models or

5. Discussion

In the model using climatic variables (model M_1), evapotranspiration (ET), humidity (HUMI), radiation (RADI), and number of frost days (FROST) were used to predict the site index. Evapotranspiration and humidity presented a positive correlation, whereas radiation and the number of frost days showed an inverse trend. Three out of the four climatic variables (HUMI, RADI, ET) were also included in the integrated model (M_i) and showed the same correlation. This result indicated that when humidity and evapotranspiration increased, the site productivity of the eucalypt increased, whereas increasing radiation can affect negatively the development of the eucalypt. The result is reasonable as evapotranspiration also showed the highest positive correlation with site index in the Pearson correlation coefficient test. The positive role of moisture availability (here represented by evapotranspiration and humidity) was also found in the study of Grant et al. (2010b) for *Eucalyptus dunnii* in Australia.

In model M_2 (the one including topographic variables) the five selected variables, including longitude (LONG), latitude (LAT), Southeast (SE), Northwest (NW), and South (S), were used to predict the site index. In Pearson's correlation coefficient tests, only longitude had a positive correlation with site index. However, this positive correlation was not in the integrated model when applying MLR. In addition, even though elevation did not appear in the topographic model, it showed a negative correlation with the site index in the integrated model.

In the model with edaphic variables (M_3), soil pH (SOILPH), water flow in the soil (RUNOFF), sandstone (arenitos), and lithosols (LITOSSOLOS) were selected to predict the site index. Water flow in soil and sandstone showed a positive correlation with the site index in the model, whereas soil pH and luvissols showed an inverse trend. The negative coefficient of lithosols remained in the integrated model and when applying QR at different quantiles.

Even though MLR played an essential role in statistical modeling, it is not enough to explain the relationship among variables, mainly when outlier or the dependent variables' extreme values existed (Zhang et al., 2020). On the other hand, QR can be seen as a tool for researchers to assess the distribution of the S values around any point on the conditional distribution. Thanks to the QR, the distribution of dependent variables was described entirely (Zhang et al., 2020). The QR results at the median quantile range were similar to the mean regression of MLR for all parameters in the edaphic variables. From the QQ plots of the four MLR models (Figure 11), it is clear that the dataset was slightly skewed left due to the outliers. By applying QR, the study's extreme values were described clearly at the 0.05 and 0.95 quantiles, where the confidence interval at most of the parameters was the largest.

For the site index variance map, compared with the site index value of the observed study plots, there were many plots where the observed values were far different from the predicted value. Moreover, there were also some locations where, in theory, there would be no eucalypt plantation due to the inappropriate environmental conditions (high elevation, too less precipitation). However, in reality, these locations still had eucalypt plantations. It is also noticeable to mention that when comparing the site index variation map with the total precipitation map (Figure 6), the North West regions had the highest site index value also overlapped with the regions that had the highest amount of total precipitation. However, the regions in North East of Portugal (circled in red in Figure 14) with total precipitation less than 600mm still witnessed the development of eucalypt in the site index variation map (under the class “very poor”).

Regarding the site index variation maps at the 5 and 95 percentile (Figure 16), these maps emphasized the range of the outliers in the model. At the 5 percentile, outliers were mostly witnessed at the smallest ranges of the site productivity. In contrast, at the 95 percentile, outliers were found on the highest range of the site productivity.

This study's limitation can come from the limited number of independent variables used to predict the site index. For example, this study did not consider the soil's chemical components, such as nitrogen and phosphorus, which play an important role in eucalypt productivity (Almeida, 1996). Moreover, the soil depth or the distance to the sea were also not included in this study. Thus, these variables should be considered in a future study. Additionally, to extending the set of predictors to be considered it would be desirable to complete the dataset extending the coverage of sample plots to areas of eucalypt occurrence which have not been covered by the current dataset. This could help refine model estimates. Still related to the dataset, a search and elimination of oversampling in some particular areas might also improve results.

Besides, this study can be developed intensively by considering the influence of eucalypt genetics and silviculture methods applied in each site. This is because growth can be improved using genetically improved stock (Henson and Vanclay, 2004). Thus, the site index map for eucalypt across Portugal could become more accurate.

Furthermore, results could be further improved if more disaggregated climatic data was available, namely total precipitation since the data available was grouped in classes with a range of 100mm and the differences observed in Figures 14 and 16 derive of total precipitation values.

A suggestion for future research is the development of a methodology to refine the mean prediction provided by the model obtained in this thesis by using more detailed information from a particular site, such as location in the slope and soil depth.

6. Conclusion

This dissertation's results provided a broader description of the relationship between environmental variables and site index for eucalypt in Portugal not only at the mean value by applying MLR but also for different quantiles by applying QR. As a result, a final map with a potential distribution of eucalypt productivity areas was additionally produced which can be used as a helpful tool regarding the location of new eucalypt plantations for the future eucalypt plantations management and planning strategy. However, because these models only explained a relatively small portion of site index variance, it is necessary to introduce more environmental variables to the fitting model. Moreover, more data points should be gathered to be used as an independent dataset for evaluating and validating processes. A refinement in the elaboration of the mask is also required.

References

Águas, A., Ferreira, A., Maia, P., Fernandes, P. M., Roxo, L., Keizer, J., Silva, J. S., Rego, F. C., & Moreira, F. (2014). Natural establishment of Eucalyptus globulus Labill. in burnt stands in Portugal. Forest Ecology and Management, 323, 47–56. https://doi.org/10.1016/j.foreco.2014.03.012

Almeida, A. P., & Riekerk, H. (1990). Water balance of Eucalyptus globulus and Quercus suber forest stands in south Portugal. Forest Ecology and Management, 38(1–2), 55–64. https://doi.org/10.1016/0378-1127(90)90085-P

Almeida, M. H., Chaves, M. M., & Silva, J. C. (1994). Cold acclimation in eucalypt hybrids. Tree Physiology, 14(7-8–9), 921–932.

Alves, A., Santos, A., Rozenberg, P., Pâques, L. E., Charpentier, J.-P., 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× eurolepis. Wood Science and Technology, 46(1–3), 157–175.

Amaral, J., Paulo, P., Firmino, N., Pacheco, S., & Margarida, F. (2021). Quantile regression for modelling the impact of climate in cork growth quantiles in Portugal. European Journal of Forest Research, 0123456789. https://doi.org/10.1007/s10342-021-01379-8

Amaro, A., Reed, D., Tomé, M., & Themido, I. (1998). Modeling dominant height growth: Eucalyptus plantations in Portugal. Forest Science, 44(1), 37–46. https://doi.org/10.1093/forestscience/44.1.37

Baptista, F. O., Santos, R. T. (2005). Os Proprietários Florestais. 44 (Ciências Sociais da Universidade de Lisboa. 2009), 654–656.

Battaglia, M., & Sands, P. J. (1998). Process-based forest productivity models and their application in forest management. Forest Ecology and Management, 102(1), 13–32.

Boland, D. J., Brooker, M. I. H., Chippendale, G. M., Hall, N., Hyland, B. P. M., Johnston, R. D., Kleinig, D. A., McDonald, M. W., & Turner, J. D. (2006). Forest trees of Australia. CSIRO publishing.

Booth, T. H. (2013). Eucalypt plantations and climate change. Forest Ecology and Management, 301, 28–34. https://doi.org/10.1016/j.foreco.2012.04.004

Bravo-Oviedo, A., Del Río, M., & Montero, G. (2004). Site index curves and growth model for Mediterranean maritime pine (Pinus pinaster Ait.) in Spain. Forest Ecology and Management, 201(2–3), 187–197.

Bravo-Oviedo, A., & Montero, G. (2005). Site index in relation to edaphic variables in stone pine (Pinus pinea L.) stands in south west Spain. Annals of Forest Science, 62(1), 61–72.

Carmean, W. H. (1975). Forest site quality evaluation in the United States. In Advances in agronomy (Vol. 27, pp. 209–269). Elsevier.

Carmean, W. H. (1996). Site-quality evaluation, site-quality maintenance, and site-specific management for forest land in northwest Ontario (Issue 105). Thunder Bay: Ontario Ministry of Natural Resources, Northwest Science and ….

Catry, F. X., Moreira, F., Deus, E., Silva, J. S., & Águas, A. (2015). Assessing the extent and the environmental drivers of Eucalyptus globulus wildling establishment in Portugal: results from a countrywide survey. Biological Invasions, 17(11), 3163–3181. https://doi.org/10.1007/s10530-015-0943-y

CELPA. (2015). Associação da Indústria Papeleira. Boletim Estatístico, Industria Papeleira Portuguesa, 1, pp98. Available 30 June 2021. http://www.celpa.pt/wp-content/uploads/2016/09/Boletim_WEB_2015.pdf

Cesaro, L., & Gatto, P. (2008). The Multifunctional Role of Forests - Policies , Methods and Case Studies. In The Multifunctional Role of Forests - Policies , Methods and Case Studies (Issue EFI Proceedings 55).

Clutter, J. L., Fortson, J. C., Pienaar, L. V, Brister, G. H., & Bailey, R. L. (1983). Timber management: A quantitative approach. John Wiley & Sons, Inc.

Cole, T. J., & Green, P. J. (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. Statistics in Medicine, 11(10), 1305–1319.

Corona, P., Scotti, R., & Tarchiani, N. (1998). Relationship between environmental factors and site index in Douglas-fir plantations in central Italy. Forest Ecology and Management, 110(1–3), 195–207. https://doi.org/10.1016/S0378-1127(98)00281-3

Cruz, M. G., Gould, J. S., Alexander, M. E., Sullivan, A. L., McCaw, W. L., & Matthews, S. (2015). Empirical-based models for predicting head-fire rate of spread in Australian fuel types. Australian Forestry, 78(3), 118–158.

Dadswell, H. E., & HE, D. (1972). The anatomy of eucalypt woods. CSIRO (Aust.).

DGF. (2001). Direcção Geral das Florestas. Inventário Florestal Nacional Portugal Continental, 3ª Revisão, 1995-1998, Lisboa. Portugal.

DGOGF. (1979). Direcçao Geral do Ordenamento e Gestão Florestal. Distribuição Da Floresta Em Portugal Continental - Áreas Florestais Por Concelhos, 1978, Lisboa, Portugal

DGRF. (2007). Direcção-Geral dos Recursos Florestais. Estratégia Nacional para as Florestas. Imprensa Nacional-Casa da Moeda, Lisboa. Portugal

Duguy, B., Alloza, J. A., Röder, A., Vallejo, R., & Pastor, F. (2007). Modelling the effects of landscape fuel treatments on fire growth and behaviour in a Mediterranean landscape (eastern Spain). International Journal of Wildland Fire, 16(5), 619–632.

Fernandes, P. M., Guiomar, N., & Rossa, C. G. (2019). Analysing eucalypt expansion in Portugal as a fire-regime modifier. Science of the Total Environment, 666, 79–88. https://doi.org/10.1016/j.scitotenv.2019.02.237

Florence, R. G. (2004). Ecology and silviculture of eucalypt forests. CSIRO publishing.

Fontes, L., Landsberg, J., Tomé, J., Tomé, M., Pacheco, C. A., Soares, P., & Araujo, C. (2006). Calibration and testing of a generalized process-based model for use in Portuguese eucalyptus plantations. Canadian Journal of Forest Research, 36(12), 3209–3221.

Fontes, L., Tomé, M., Thompson, F., Yeomans, A., Luis, J. S., & Savill, P. (2003). Modelling the Douglas-fir (Pseudotsuga menziesii (Mirb.) Franco) site index from site factors in Portugal. Forestry, 76(5), 491–507. https://doi.org/10.1093/forestry/76.5.491

Ganteaume, A., Lampin-Maillet, C., Guijarro, M., Hernando, C., Jappiot, M., Fonturbel, T., Pérez-Gorostiaga, P., & Vega, J. A. (2010). Spot fires: fuel bed flammability and capability of firebrands to ignite fuel beds. International Journal of Wildland Fire, 18(8), 951–969.

Garcia-Gonzalo, J., Borges, J. G., Palma, J. H. N., & Zubizarreta-Gerendiain, A. (2014). A decision support system for management planning of Eucalyptus plantations facing climate change. Annals of Forest Science, 71(2), 187–199. https://doi.org/10.1007/s13595-013-0337-1

FRA. (2006). Global Forest Resources Assessment 2005. Food and Agricultue Organization of the United Nations, FAO Forestry Paper 147. FAO Rome, Italy. Available 30 June 2021: http://www.env-edu.gr/Documents/Global%20Forest%20Resources%20Assessment%202005.pdf

Goes, E. (1962). Os eucaliptos em Portugal, ecologia cultura e produções. Direcção-Geral

Dos Serviços Florestais e Aquícolas, Lisboa.

Grant, J. C., Nichols, J. D., Smith, R. G. B., Brennan, P., & Vanclay, J. K. (2010b). Site index prediction of Eucalyptus dunnii Maiden plantations with soil and site parameters in sub-tropical eastern Australia. Australian Forestry, 73(4), 234–245. https://doi.org/10.1080/00049158.2010.10676334

Harwood, C. (2011). New introductions–doing it right. Developing a Eucalypt Resource: Learning from Australia and Elsewhere: University of Canterbury. Christchurch, New Zealand: Wood Technology Research Centre, 43–54.

Heagerty, P. J., & Pepe, M. S. (1999). Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. Journal of the Royal Statistical Society: Series C (Applied Statistics), 48(4), 533–551.

Hendricks, W., & Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. Journal of the American Statistical Association, 87(417), 58–68.

Henson, M., & Vanclay, J. K. (2004). The value of good sites and good genotypes: an analysis of Eucalyptus dunnii plantations in NSW.

Hinko-Najera, N., Najera Umaña, J. C., Smith, M. G., Löw, M., Griebel, A., & Bennett, L. T. (2019). Relationships of intra-annual stem growth with climate indicate distinct growth niches for two co-occurring temperate eucalypts. Science of the Total Environment, 690, 991–1004. https://doi.org/10.1016/j.scitotenv.2019.07.024

Hopmans, P., Stewart, H. T. L., & Flinn, D. W. (1993). Impacts of harvesting on nutrients in a eucalypt ecosystem in southeastern Australia. Forest Ecology and Management, 59(1–2), 29–51.

Huang, Q., Zhang, H., Chen, J., & He, M. (2017). Quantile Regression Models and Their Applications: A Review. Journal of Biometrics & Biostatistics, 08(03). https://doi.org/10.4172/2155-6180.1000354

Hudson, I., Wilson, L., Sandercock, C., & Sands, R. (1995). Within-ring variability of wood microstructure in Eucalyptus nitens. Proc. CRCTHF-IUFRO Conference. Eucalypt Plantations: Improving Fibre Yield and Quality, 110–115.

ICNF (2019). Instituto da Conservação da Natureza e Florestas. 6º Inventário Florestal Nacional – Relatório Final. Instituto da Conservação da Natureza e Florestas, Lisboa, Portugal. Available 30 June 2021: http://www2.icnf.pt/portal/florestas/ifn/resource/doc/ifn/ifn6/IFN6_Relatorio_completo-2019-11-28.pdf

Iglesias, G.T., Wistermann, D. (2008). Eucalyptus universalis. Global cultivated Eucalyptus forest map. Eucalyptologics: Information on Eucalyptus cultivation worldwide. GIT Forestry Consulting.

Jacobs, M.R. (1981). Eucalypts for planting. In Forestry Series (Vol. 11). https://doi.org/10.1017/CBO9781107415324.004

Jones, H. E., Madeira, M., Herraez, L., Dighton, J., Fabiâo, A., González-Rio, F., Fernandez Marcos, M., Gomez, C., Tomê, M., Feith, H., Magalhâes, M. C., & Howson, G. (1999). The effect of organic-matter management on the productivity of Eucalyptus globulus stands in Spain and Portugal: Tree growth and harvest residue decomposition in relation to site and treatment. Forest Ecology and Management, 122(1–2), 73–86. https://doi.org/10.1016/S0378-1127(99)00033-X

Jorge, F., Quilhó, T., & Pereira, H. (2000). Variability of fibre length in wood and bark in Eucalyptus globulus. IAWA Journal, 21(1), 41–48.

Khanna, P. K. (1997). Nutrient cycling under mixed-species tree systems in southeast Asia. Agroforestry Systems, 38(1), 99–120.

Kirkpatrick, J. B. (1975). Natural distribution of Eucalyptus globulus Labill. Australian Geographer, 13(1), 22–35.

Koenker, R. (2013). Quantreg: Quantile Regression. R package version 5.05. R Foundation for Statistical Computing: Vienna) Available at: Http://CRAN. R-Project. Org/Package= Quantreg.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. Econometrica: Journal of the Econometric Society, 33–50.

Koenker, R., & Hallock, K. F. (2001). Quantile regression. Journal of Economic Perspectives, 15(4), 143–156.

Liu, Y., & Wu, Y. (2009). Stepwise multiple quantile regression estimation using non-crossing constraints. Statistics and Its Interface, 2(3), 299–310. https://doi.org/10.4310/sii.2009.v2.n3.a4

Ludwig, B., Khanna, P. K., Raison, R. J., & Jacobsen, K. (1997). Modelling changes in cation composition of a soil after clearfelling a eucalypt forest in East Gippsland, Australia. Geoderma, 80(1–2), 95–116.

Luke, R. H. . M. (1978). Bushfires in Australia. CSIRO Division of Forestry Research, Canberra.

M, R., & Richardson DM. (2011). Eucalypts. In: Simberloff D, Rejmanek M (eds) Encyclopedia of biological invasions. University of California Press, 203–209.

Madeira, M. V. A., Melo, M. G., Alexandre, C. A., & Steen, E. (1989). Effects of deep ploughing and superficial disc harrowing on physical and chemical soil properties and biomass in a new plantation of Eucalyptus globulus. Soil and Tillage Research, 14(2), 163–175.

Maia, P., Keizer, J., Vasques, A., Abrantes, N., Roxo, L., Fernandes, P., Ferreira, A., & Moreira, F. (2014). Post-fire plant diversity and abundance in pine and eucalypt stands in Portugal: Effects of biogeography, topography, forest type and post-fire management. Forest Ecology and Management, 334, 154–162. https://doi.org/10.1016/j.foreco.2014.08.030

Marques, C. P. (1991). Evaluating site quality of even-aged maritime pine stands in northern Portugal using direct and indirect methods. Forest Ecology and Management, 41(3–4), 193–204.

Mateus, P., & Fernandes, P. M. (2014). Forest fires in Portugal: dynamics, causes and policies. In Forest Context and Policies in Portugal (pp. 97–115). Springer.

Mendes, A.M.S.C. & Dias, R. . (2002). Financial Instruments of Forest Policy in Portugal in the 1980s and 1990s. In Andreas Ottitsch, Ilpo Tikkanen & Pere Riera Eds. Financial Instruments of Forest Policy, No 42(EFI Proceedings. Joensuu. Finland. European Forest Institute), 95–116.

Mendes. A.M.S.C., Feliciano, D., Tavares, M., Dias, R. (2004). The Portuguese Forests.Working Papers de Economia (Economics Working Papers). Faculdade de Economia e Gestão, Universidade Católica Portuguesa (Porto).

Mendes, A. (2008). The role of institutions in forest development: the case of forest services and forest owners' associations in Portugal. The Multifunctional Role of Forests Policies, Methods and Case Studies, 105.

Menéndez-Miguélez, M., Álvarez-Álvarez, P., Majada, J., & Canga, E. (2015). Effects of soil nutrients and environmental factors on site productivity in Castanea sativa Mill. coppice stands in NW Spain. New Forests, 46(2), 217–233. https://doi.org/10.1007/s11056-014-9456-2

Mirra, I. M., Oliveira, T. M., Barros, A. M. G., & Fernandes, P. M. (2017). Fuel dynamics following fire hazard reduction treatments in blue gum (Eucalyptus globulus) plantations in Portugal. Forest Ecology and Management, 398, 185–195. https://doi.org/10.1016/j.foreco.2017.05.016

Monserud, R. A., & Rehfeldt, G. E. (1990). Genetic and environmental components of variation of site index in inland Douglas-fir. Forest Science, 36(1), 1–9.

Moreira, F., Vaz, P., Catry, F., & Silva, J. S. (2009). Regional variations in wildfire susceptibility of land-cover types in Portugal: implications for landscape management to minimize fire hazard. International Journal of Wildland Fire, 18(5), 563–574.

Myers, R. H., & Myers, R. H. (1990). Classical and modern regression with applications (Vol. 2). Duxbury press Belmont, CA.

Oliveira, T. M., Guiomar, N., Baptista, F. O., Pereira, J. M. C., & Claro, J. (2017). Is Portugal's forest transition going up in smoke? Land Use Policy, 66, 214–226.

Paulo, J. A., Palma, J. H. N., Gomes, A. A., Faias, S. P., Tomé, J., & Tomé, M. N. (66). ri. (2015b). Predicting site index from climate and soil variables for cork oak (Quercus suber L.) stands in Portugal. New Forests, 46(2), 293–307. https://doi.org/10.1007/s11056-014-9462-4

Pausas, J. G., & Fernández-Muñoz, S. (2012). Fire regime changes in the Western Mediterranean Basin: from fuel-limited to drought-driven fire regime. Climatic Change, 110(1), 215–226.

Pereira, J. S., Linder, S., Araújo, M. C., Pereira, H., Ericsson, T., Borralho, N., & Leal, L. C. (1989). Optimization of Biomass Production in Eucalyptus Globulus Plantations.—A Case Study. In Biomass production by fast-growing trees (pp. 101–121). Springer.

Pereira, J. S., Tomé, M., Madeira, M., Oliveira, A. C., Tomé, J., & Almeida, M. H. (1996). Eucalypt plantations in Portugal. Nutrition of the Eucalypts. CSIRO, Australia, 371–387.

Pery, G. A. (1875). Geographia e estatistica geral de Portugal e colonias: com um atlas. Impr. nacional.

Resende, R. T., Soares, A. A. V., Forrester, D. I., Marcatti, G. E., dos Santos, A. R., Takahashi, E. K., e Silva, F. F., Grattapaglia, D., Resende, M. D. V., & Leite, H. G. (2018). Environmental uniformity, site quality and tree competition interact to determine stand productivity of clonal Eucalyptus. Forest Ecology and Management, 410(December 2017), 76–83. https://doi.org/10.1016/j.foreco.2017.12.038

Ribeiro, F., & Tomé, M. (2000). Classificação Climática de Portugal Continental, baseada em informação do Atlas do Ambiente. Revista Ciências Agrárias, 23.

Ringius, G. S., Meades, S., & Sims, R. A. (1997). Indicator plant species in Canadian forests. Presses Université Laval.

San-Miguel-Ayanz, J., Durrant, T., Boca, R., Libertà, G., Branco, A., De Rigo, D., Ferrari, D., Maianti, P., Vivancos, T. A., & Costa, H. (2017). Forest fires in Europe. Middle East and North Africa, 10(663443), 2017.

Schröter, D., & Consortium, A. (2004). Global change vulnerability—Assessing the European human-environment system. Potsdam Institute for Climate Impact Research.

Scolforo, H. F., de Castro Neto, F., Scolforo, J. R. S., Burkhart, H., McTague, J. P., Raimundo, M. R., Loos, R. A., da Fonseca, S., & Sartório, R. C. (2016). Modeling dominant height growth of eucalyptus plantations with parameters conditioned to climatic variations. Forest Ecology and Management, 380, 182–195. https://doi.org/10.1016/j.foreco.2016.09.001

Scolforo, H. F., McTague, J. P., Burkhart, H., Roise, J., Alvares, C. A., & Stape, J. L. (2020). Site index estimation for clonal eucalypt plantations in Brazil: A modeling approach refined by environmental variables. Forest Ecology and Management, 466(November 2019), 118079. https://doi.org/10.1016/j.foreco.2020.118079

Skovsgaard, J. P., & Vanclay, J. K. (2008). Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. Forestry: An International Journal of Forest Research, 81(1), 13–31.

Soares, P., Tomé, M., Skovsgaard, J. P., & Vanclay, J. K. (1995). Evaluating a growth model for forest management using continuous forest inventory data. Forest Ecology and Management, 71(3), 251–265.

Soares, P., Tomé, M. (2001). A tree crown ratio prediction equation for eucalypt plantations To cite this version : HAL Id : hal-00884196 Original article A tree crown ratio prediction equation for eucalypt plantations.

Solomon, S. (2007). IPCC (2007): Climate change the physical science basis. AGUFM, 2007, U43D-01.

Team, R. C. (2013). R development core team. RA Lang Environ Stat Comput, 55, 275–286.

Tomé, M., Ribeiro, F., & Soares, P. (2001). O modelo GLOBULUS 2.1. Relatórios Técnico–Científicos do GIMREF, nº1/2001. Dep. Eng. Florestal, ISA, Lisboa.

Tomé, M., Soares, P., & Oliveira, T. (2006). O modelo GLOBULUS 30 Dados e equações

Publicações GIMREF RC2/2006Universidade Técnica de Lisboa. Instituto Superior de Agronomia, Centro de Estudos Florestais, Lisboa.

Tomé, M. (2000). Wood and non-wood production from plantation forests. Ecological and Socio-Economic Impacts of Close-to-Nature Forestry and Plantation Forestry: A Comparative Analysis. Scientific Seminar of the 7th Annual EFI Conference. European Forest Institute, Lisbon, 37–55.

Tomé, M., Almeida, M. H., Barreiro, S., Branco, M. R., Deus, E., Pinto, G., Silva, J. S., Soares, P., & Rodríguez-Soalleiro, R. (2021). Opportunities and challenges of Eucalyptus plantations in Europe: the Iberian Peninsula experience. European Journal of Forest Research, 1–59. https://doi.org/10.1007/s10342-021-01358-z

Turnbull, J. W. (1984). Choice of specific and seed sources. Eucalyptus for Wood Production, 6–65.

Turnbull, J. W. (1999). Eucalypt plantations. In Planted forests: Contributions to the quest for sustainable societies (pp. 37–52). Springer.

Tyler, A. L., MacMillan, D. C., & Dutch, J. (1996). Models to predict the General Yield Class of Douglas fir, Japanese larch and Scots pine on better quality land in Scotland. Forestry: An International Journal of Forest Research, 69(1), 13–24.

Vanclay, J. K. (1992). Assessing site productivity in tropical moist forests: a review. For Ecol Manag 54: 257–287.

Vanclay, J. K., & Skovsgaard, J. P. (1997). Evaluating forest growth models. Ecological Modelling, 98(1), 1–12.

Weiskittel, A. R., Hann, D. W., Kershaw Jr, J. A., & Vanclay, J. K. (2011). Forest growth and yield modeling. John Wiley & Sons.

Xie, L., Widagdo, F. R. A., Dong, L., & Li, F. (2020). Modeling height-diameter relationships for mixed-species plantations of Fraxinus mandshurica rupr. and Larix olgensis henry in Northeastern China. Forests, 11(6). https://doi.org/10.3390/F11060610

Zhang, B., Sajjad, S., Chen, K., Zhou, L., Zhang, Y., Yong, K. K., & Sun, Y. (2020). Predicting Tree Height–Diameter Relationship from Relative Competition Levels Using Quantile Regression Models for Chinese Fir (Cunninghamia lanceolata) in Fujian Province, China. Forests, 11(2), 183.