

2 **Citation\***

Aertsen, W., Kint, V., De Vos, B., Deckers, J., van Orshoven, J., Muys, B. (2012)  
Predicting forest site productivity in temperate lowland from forest floor, soil and litterfall  
characteristics using boosted regression trees. *Plant and Soil*, 354, 157-172.

DOI: [10.1016/j.ecolmodel.2010.01.007](https://doi.org/10.1016/j.ecolmodel.2010.01.007)

3  
Authors: Wim AERTSEN\*, Vincent KINT\*, Bruno DE VOS<sup>¶</sup>, Jozef  
DECKERS\*, Jos VAN ORSHOVEN\*, Bart MUYS\*

\* Department of Earth and Environmental Sciences, Katholieke  
Universiteit Leuven; Celestijnenlaan 200E Box 2411, BE-3001  
Leuven, Belgium

<sup>¶</sup> Research Institute for Nature and Forest; Gaverstraat 4, BE-  
9500 Geraardsbergen, Belgium

Corresponding author Bart Muys, Tel. +3216329721, Fax  
+3216329760, E-mail [bart.muys@ees.kuleuven.be](mailto:bart.muys@ees.kuleuven.be)

4  
5 \* Springer-Verlag authorized the authors to self-archive this article on their personal  
6 website, with full bibliographic reference to the publication and a link to the  
7 published article on Springer's website [www.springerlink.com](http://www.springerlink.com) (see DOI). The  
8 Authors ensure that the publication by Springer-Verlag is properly credited and that  
9 the relevant copyright notice is repeated verbatim.



12  
13 Creative Commons Attribution-Noncommercial-Share Alike 2.0 Belgium License

14 **Title page**

15 **a. TITLE:**

16 Predicting forest site productivity in temperate lowland from forest floor, soil and litterfall  
17 characteristics using boosted regression trees

18 **b. AUTHORS:**

19 Wim AERTSEN\*, Vincent KINT\*, Bruno DE VOS<sup>¶</sup>, Jozef DECKERS\*, Jos VAN  
20 ORSHOVEN\*, Bart MUYS\*

21 **c. AFFILIATIONS:**

22 \* Department of Earth and Environmental Sciences, Katholieke Universiteit Leuven;  
23 Celestijnenlaan 200E Box 2411, BE-3001 Leuven, Belgium

24 <sup>¶</sup> Research Institute for Nature and Forest; Gaverstraat 4, BE-9500 Geraardsbergen,  
25 Belgium

26 **d. E-MAIL ADRESSES:**

27 wim.aertsen@ees.kuleuven.be; vincent.kint@ees.kuleuven.be; bruno.devos@inbo.be;

28 seppe.deckers@ees.kuleuven.be; jos.vanorshoven@ees.kuleuven.be;

29 bart.muys@ees.kuleuven.be

30 **e. CORRESPONDING AUTHOR:**

31 Bart Muys, Department of Earth and Environmental Sciences, K.U.Leuven, Celestijnenlaan  
32 200E Box 2411, BE-3001 Leuven, Belgium.

33 Tel. +3216329721, Fax +3216329760, E-mail bart.muys@ees.kuleuven.be

34 **Keywords:** Boosted Regression Trees; *Fagus sylvatica*; litterfall nutrient concentrations; net  
35 primary production; nitrogen deposition; *Pinus sylvestris*; *Quercus robur*; site index

36 **Abstract**

37 **Aims.** The aim of this study is on the one hand to identify the most determining variables  
38 predicting the site productivity of pedunculate oak, common beech and Scots pine in  
39 temperate lowland forests of Flanders; and on the other hand to test whether the accuracy of  
40 site productivity models based exclusively on soil or forest floor predictor variables is similar  
41 to the accuracy achieved by full ecosystem models, combining all soil, vegetation, humus and  
42 litterfall composition related variables.

43 **Methods.** Boosted Regression Trees (BRT) were used to model in a climatically  
44 homogeneous region the relationship between environmental variables and site productivity.  
45 A distinction was made between soil (soil physical and chemical), forest floor (vegetation and  
46 humus) and ecosystem (soil, forest floor and litterfall composition jointly) predictors.

47 **Results.** Our results have illustrated the strength of BRT to model the non-linear behaviour of  
48 ecological processes. The ecosystem models, based on all collected variables, explained most  
49 of the variability and were more accurate than those limited to either soil or forest floor  
50 variables. Nevertheless, both the soil and forest floor models can serve as good predictive  
51 models for many forest management practices.

52 **Conclusions.** Soil granulometric fractions and litterfall nitrogen concentrations were the most  
53 effective predictors of forest site productivity in Flanders. Although many studies revealed a  
54 fertilising effect of increased nitrogen deposition, nitrogen saturation seemed to reduce  
55 species' productivity in this region.

56 **Abbreviations**

57	BRT	Boosted Regression Trees
58	GAM	Generalized Additive Models
59	LOI	Loss-On-Ignition
60	mF	mean indicator value for moisture
61	mL	mean indicator value for light
62	mN	mean indicator value for nitrogen
63	mR	mean indicator value for acidity
64	R <sup>2</sup>	coefficient of determination
65	(r)RMSE	(relative) Root Mean Squared Error
66	VIF	Variance Inflation Factor

## 67 **Introduction**

68 Site productivity is a key-indicator of forest ecosystem services like wood production and  
69 carbon sequestration and hence an important criterion for decision makers in forest  
70 management. It allows foresters to forecast growth and production and hence select the most  
71 suitable tree species for a site. However in many situations direct measurement of the site  
72 productivity is not possible.

73 Pedunculate oak (*Quercus robur* L.), common beech (*Fagus sylvatica* L.) and Scots pine  
74 (*Pinus sylvestris* L.) are three of the main tree species in north-west European temperate  
75 lowland forests (Masson 2005; Seynave et al. 2008). Adequate understanding of the species-  
76 specific site productivity is therefore of great importance in sustainable forest management. In  
77 an ecological context it is the potential net primary production of that location, whereas in  
78 practical forestry terms it is more specifically its potential wood production for a specific  
79 timber species (Skovsgaard and Vanclay 2008). Knowledge of the environmental conditions  
80 that determine species-specific site productivity may help to identify sites for which changes  
81 can be expected in species suitability, especially in the context of climate change and/or  
82 nitrogen deposition (Bergès et al. 2005).

83 One of the most widely used indicators of species-specific forest site productivity is site  
84 index, i.e. the mean height at a reference age of the dominant trees growing on a site  
85 (Skovsgaard and Vanclay 2008). In homogeneous even-aged forest stands, site index of the  
86 standing species can be directly estimated from measurements of tree height and age, using an  
87 appropriate species-specific dominant height growth model (Skovsgaard and Vanclay 2008).  
88 But in mixed or uneven-aged stands, or in the case of stand conversion to other tree species,  
89 or afforestation of land without forest, or under changing site conditions over time, direct

90 estimation is not possible. For such cases site index needs to be estimated indirectly from  
91 environmental factors like climate, topography and soil characteristics, using appropriate  
92 ecological models (Skovsgaard and Vanclay 2008).

93 Traditional site productivity modelling studies related site index to environmental variables  
94 using basic statistical methods such as linear regression. More recently, an increased use of  
95 non-linear, non-parametric and machine learning methods have been introduced to cope with  
96 the inherent complexity of ecological problems (Aertsen et al. 2010; Moisen et al. 2006).

97 Boosted Regression Trees (BRT) has been proven a promising technique for site index  
98 modelling, both in terms of accuracy and interpretability of ecosystem processes (Aertsen et  
99 al. 2011). Other advantages are the ability to handle discontinuous response (like in case of  
100 ecological thresholds) and the easy and intuitive visualisation of the results (De'ath 2007).

101 The temperate lowland of Flanders (northern Belgium) is characterised by little climatic and  
102 topographic variation. As a consequence it can be expected that the combination of soil,  
103 humus, vegetation and litterfall characteristics will explain the major part of variation in site  
104 productivity (Bergès et al. 2006; Seynave et al. 2005). Since litterfall nutrient concentrations  
105 are only available for sites where the target species is already present, and large amount of  
106 environmental variables are needed, the use of models based on all these variables is not  
107 recommended for predictive modelling (Chen et al. 1998). Simplified models based on  
108 subsets of variables, like soil or forest floor predictors are in that respect preferred. Although  
109 the combination of both floristic and abiotic models resulted in best empirical models for the  
110 estimation of sessile oak in France, Bergès et al. (2006) conclude that understory vegetation  
111 and abiotic site characteristics separately could explain the same part of variance in site index.  
112 Forest floor predictors have proven to be good site quality indicators (Bergès et al. 2006;

113 Dzwonko 2001; Gégout et al. 2003) and are moreover much easier and cheaper to collect than  
114 most abiotic counterparts.

115 In this study, soil predictors encompass the physical and chemical properties of the mineral  
116 soil, which are often in close physiological relationship with the species' critical growth  
117 factors (Chen et al. 1998). Moreover, most of these characteristics hold little temporal  
118 variability, making them representative for a stand's growing history. Forest floor factors are  
119 defined as the living biomass (vegetation) and its decomposition (humus) and are the result of  
120 both physicochemical and biological interactions (Ponge et al. 2002).

121 For common beech, Rogister (1978) developed an ecogram based on the floristic indices mR  
122 x mN (humus indicator) and mF (moisture indicator) to assess the site quality for beech in  
123 Flanders, concluding that mesophilic sites with a mull humus form were the most productive.  
124 Pedunculate oak was characterised in France by its relatively large nutrient and water  
125 requirements (Lévy et al. 1992). For Scots pine, finally, a negative effect of soil water  
126 capacity on tree growth was observed in Finland, without a significant effect of nutrient  
127 availability (Wall and Heiskanen 2009).

128 The aim of this study is: (1) to identify the most determining variables predicting the site  
129 productivity of pedunculate oak, common beech and Scots pine in temperate lowland forests  
130 of Flanders; and (2) to test whether the accuracy of site productivity models based exclusively  
131 on soil or forest floor predictor variables is similar to the accuracy achieved by full ecosystem  
132 models, combining all soil, vegetation, humus and litterfall composition related variables.

## 133 **Materials and methods**

### 134 Study area and sampling design

135 The study area covers the region of Flanders in the north of Belgium (13521 km<sup>2</sup>; Fig. 1).  
136 Flanders is a lowland region with a temperate and humid climate and no major topographic or  
137 climatic gradients. The pedology of the region is characterised by a pronounced north-south  
138 gradient of decreasing sand and increasing silt fractions and by the occurrence of local  
139 shallow tertiary substrates (clay/sand).

140 Spread over the territory of Flanders, 186 research plots (about 1000 m<sup>2</sup> each) were laid out in  
141 near-homogeneous, even-aged, normally stocked high-forests of the target species of at least  
142 50 years old (76 plots of pedunculate oak (*Quercus robur* L.) (54 - 183 years), 55 of common  
143 beech (*Fagus sylvatica* L.) (60-168 years) and 55 of Scots pine (*Pinus sylvestris* L.) (50-120  
144 years)) (Fig. 1). Sampling occurred according a two-step approach: The first step consisted of  
145 a stratified random selection of stands of the three species, evenly distributed over five  
146 productivity classes based on a-priori determination of the site index from height  
147 measurement and age estimation. In the second step additional plots were selected to fill gaps  
148 in common forest soil types and in geographical coverage (Kint et al. 2011). Data collection  
149 was carried out between October 2007 and April 2009.

### 150 Site productivity

151 The measured response variable was site index, a proxy for species-specific site productivity  
152 and defined as the dominant height of a homogeneous even-aged stand at the reference age of  
153 100 years (Skovsgaard and Vanclay 2008). The use of the dominant height of the stand in  
154 stead of the volume, is based on the fact that dominant height growth in pure and even-aged  
155 stands is, in contrast to standing volume, nearly independent of silvicultural techniques as

156 stand density and thinning (Assmann 1970; Skovsgaard and Vanclay 2008). Dominant height  
157 was calculated as the mean height of the five thickest trees in each research plot (Jansen et al.  
158 1996). Tree age was estimated from ring counts and adjusted by extrapolating for missing  
159 rings (due to off-pith coring and to account for rings formed before the 1 m coring height was  
160 reached). Site index was calculated from species dependent dominant height growth models  
161 (Jansen et al. 1996).

## 162 Site quality

163 Site quality was assessed by vegetation survey, humus quality description, physicochemical  
164 analysis of the mineral topsoil, description of soil horizons and litterfall analysis. The set of  
165 variables is summarized in Table 1.

## 166 Forest floor variables

167 The vegetation survey was carried out during spring and summer on a 10mx10m subplot.  
168 Based on species abundance in the herb and shrub layers, mean indicator values for moisture  
169 (mF), light (mL), acidity (mR) and nitrogen (mN) were calculated (Hill et al. 1999), as well as  
170 the combined mR x mN index of soil trophicity (Muys and Granval 1997). Hill's indicator  
171 values are a recalibration of Ellenberg's indicator values (Ellenberg et al. 1992) for the  
172 Atlantic region and are assumed to reflect more accurately the species' ecological behaviour  
173 in our study area (Godefroid et al. 2005).

174 Humus has been characterized by the humus form and the thickness of ectorganic and  
175 endorganic horizons. Humus form was determined according to Jabiol et al. (2007) and  
176 converted to the humus index of Ponge and Chevalier (2006) with values ranging from 1  
177 (eumull) to 8 (mor) (Ponge et al. 2011), and was determined as an average of 16  
178 measurements in each plot.

179           Soil variables

180   A bulked topsoil sample (0-30cm, 16 sample points mixed) has been analysed for soil  
181   granulometric fractions (laser diffraction: Clay 0-6 $\mu$ m, Silt 6-50  $\mu$ m, Sand > 50  $\mu$ m), pH-  
182   CaCl<sub>2</sub>, carbon content (LOI), total nitrogen content (Kjeldahl) and plant available phosphorus  
183   content (Bray and Kurtz 1945). Bulk density was derived using the pedotransfer function of  
184   De Vos et al. (2005). Available water capacity (AWC), air capacity (AC) and permanent  
185   wilting point (PWP) were estimated with help of the transfer functions developed by Teepe et  
186   al. (2003).

187   The deeper soil horizons were investigated in each plot by augering to a depth of 120 cm. The  
188   soil features recorded were horizon depth, occurrence of clay accumulation , occurrence of  
189   gleyic properties (rust spots, a proxy for winter groundwater table), and presence of a reduced  
190   horizon (a proxy for permanent groundwater table) or compacted horizon.

191   Rootable depth was determined with a penetrometer measurement and averaged for 16 points  
192   in each plot. Measurements were carried out before the growing season of 2008 when soils  
193   were near field capacity. A compaction limit of 3 MPa has been assumed to be the maximal  
194   resistance tree roots can overcome (Håkansson and Lipiec 2000).

195           Ecosystem variables

196   Ecosystem variables group both forest floor and soil variables completed with litterfall  
197   nutrient variables. A set of five litterfall traps was installed in a subset of 96 stands. Litterfall  
198   was collected and weighed every two weeks from September to December and analysed for  
199   several macro- and oligoelements (N, P, K, Ca, Mg, Fe, Mn, Na, Al). Different ratios were  
200   derived from the elementary concentrations. For stands encompassing more than one plot,  
201   identical litter values were attributed to these plots. Reducing sample size to these 96 plots

202 would result in a too small sample size for meaningful analysis, therefore this nestedness has  
203 been accommodated in the cross-validation procedure as described further in this section.

#### 204 Modelling techniques and performance

205 Based on our earlier experiences in modelling site index (Aertsen et al. 2011), the Boosted  
206 Regression Trees (BRT) modelling technique was selected for site index modelling. BRT is  
207 especially recommended when emphasis is not only on accuracy of the models but also on  
208 providing insight into ecosystem processes (Aertsen et al. 2010). BRT incorporates important  
209 advantages of regression tree-based methods, including the capacity of handling predictor  
210 variables of different types and distributional characteristics, the automatic detection of  
211 interactions between predictors and of incorporating missing values in an elegant fashion  
212 based on ‘surrogate’ splitting. The method is invariant to monotonic transformations of  
213 predictors and insensitive to outliers as they are isolated into a node and so have no effect on  
214 splitting. The biggest drawback of regression trees, their poor predictive performance, has  
215 been overcome by the boosting algorithm. Boosting is a numerical optimization technique for  
216 minimizing the loss function by calculating a sequence of regression trees. At each step, a  
217 new tree is added that best reduces the loss function. The final BRT model is a linear  
218 combination of all calculated trees (hundreds to thousands) that can be thought of as a  
219 regression model where each term is a tree (De'ath 2007; Elith et al. 2008; Schapire 2003).

220 For the three species, site index models were developed with a set of (i) soil predictors (soil  
221 physical and chemical variables), (ii) forest floor predictors (humus and vegetation variables)  
222 and (iii) ecosystem predictors (the combination of soil and forest floor predictors completed  
223 with litterfall variables), further termed soil, forest floor and ecosystem models. For  
224 completeness also models based on the combination of soil and forest floor predictors,  
225 without inclusion of litterfall variables, were built.

226 BRT models were built by means of R-software, version 2.10.1 (R Development Core Team  
227 2010), using the ‘gbm’ package (Ridgeway 2006) and its extensions developed by Elith et al.  
228 (2008). Models were fitted using the *gbm.step* function and a Gaussian response type (aimed  
229 at minimising squared error), with most effective settings for learning rate (0.001-0.005) and  
230 bag fraction (0.5-0.7) as found by repeated trial-and-error. Tree complexity, i.e. the number of  
231 nodes in a tree, was set to 3, according to recommendations by Elith et al. (2008) for small  
232 datasets. Models were simplified by reducing the number of explanatory variables with the  
233 *gbm.simplify* function. This simplification process is run within a 10-fold cross-validation  
234 procedure, progressively simplifying the model fitted to each fold, and using the average  
235 cross-validation error to decide how many variables can be removed from the original model  
236 without affecting predictive performance (Elith et al. 2008).

237 All reported models passed the test for acceptable level of multicollinearity (i.e. Pearson  
238 correlations between predictors  $<0.7$ , individual variance inflation factor (VIF)  $<5$  and median  
239 VIF  $<3$ ) between predictor variables, as well as for normality and homoscedasticity of  
240 residuals. Model fit and predictive performance after 10-fold cross-validation were quantified  
241 by the coefficient of determination ( $R^2$ ), the root mean squared error (RMSE) and the relative  
242 RMSE ( $rRMSE = RMSE / \text{mean site index}$ ). Predictor contributions were calculated as  
243 proportional contributions to the total error reduction over all underlying regression trees, and  
244 can be interpreted as a measure for a predictor’s importance within the model. Predictor  
245 effects indicate how much the response is affected by a certain predictor value and are  
246 visualised in partial dependence plots. To account for the nestedness of the litterfall  
247 concentration of plots located in the same stand, cross-validation groups were carefully  
248 chosen, including all or none of the observations of the same stand in a subsample to avoid an  
249 over-optimistic estimation, as proposed by De’ath (2007).

250        **Results**

251        A total of twelve site index models were built, using 4 sets of potential predictor variables for  
252        each of the three species (Table 2). The ecosystem models achieved the highest performance  
253        on the fitted data as well as after cross-validation, whereas soil models achieved for all  
254        species a higher performance than forest floor models (Table 3). Models based on the  
255        combination of soil and forest floor predictors, without inclusion of the litterfall variables,  
256        were intermediate between the ecosystem models and the soil models. When the species were  
257        compared among each other, models for beech performed best, irrespective of the variables  
258        sets, while those for pine were the weakest, except for the ecosystem model (Table 3).

259        Ecosystem models

260        Litterfall nutrient concentrations and soil textural variables were the most important  
261        predictors. Especially litterfall nitrogen concentration (oak, pine) and litterfall N:P ratio  
262        (beech) had strong model effects and contributions (20-60%) (Figs 2-4). Predictors related to  
263        litter N showed a strong negative effect on the site index, *i.e.* low site index at high predictor  
264        values. For all tree species, soil texture was represented in the ecosystem models by the silt  
265        fraction. Especially for oak and beech there was a strong positive effect of silt fraction on site  
266        index. For pine this effect was weaker and showed rather a threshold behaviour, *i.e.* a higher  
267        productivity level was observed starting at a 10% threshold of silt in the soil (Fig. 4).

268        The importance of the water regime for tree growth was further indicated in the ecosystem  
269        models by the depth to a gleyic horizon for oak and the vegetation humidity indicator mF for  
270        beech and pine. Site quality of oak was negatively influenced when the winter groundwater  
271        table rose above 40cm, indicating that hydromorphic sites were unfavourable. For beech and

272 pine, based on mF, dry sites showed low productivity. It has to be noted though, that beech  
273 was not found on hydromorphic sites.

274 Both the thicknesses of ectorganic and endorganic horizons were present in the ecosystem  
275 models of oak and beech with respectively a negative and a positive threshold effect. In the  
276 pine model, no humus variables were present but a strong positive mR x mN effect was  
277 observed, which has been demonstrated to be a sensitive humus quality indicator (Godefroid  
278 et al. 2005; Rogister 1978).

#### 279 Soil models

280 While soil texture was already an important predictor in the ecosystem models, it was  
281 definitely the case in the soil models. Not only was the relative contribution of soil texture  
282 variables higher in the soil models compared to the ecosystem models, but also their effect  
283 was stronger (Table 2, Figs S4-S6). Since sand, silt and clay fractions were highly correlated,  
284 only one of these fractions should enter a model at a time. For oak, the textural effect of silt  
285 from the ecosystem model was replaced in the soil model by the sand fraction with an inverse  
286 effect.

287 A remarkable predictor in the soil models of oak and beech was the soil pH. For both species  
288 it entered the models with the same threshold value of 3.6, meaning that a sharp decrease in  
289 productivity was observed when the soil pH (CaCl<sub>2</sub>) dropped below this value. In the  
290 ecosystem model, soil pH was present for oak but not for beech.

#### 291 Forest floor models

292 Alike the soil models, the forest floor models retain most corresponding predictors of the  
293 ecosystem models and the combined soil + forest floor models. For all three species the

294 importance of water supply was indicated by the soil humidity indicator of the vegetation mF  
295 in the forest floor models (Table 2, Figs S7-S9). Similar as the negative influence of a high  
296 rising winter groundwater table in the ecosystem model, mF indicated a reduced growth of  
297 oak at hydromorphic sites.

298 The importance of humus predictor variables could already be observed in the ecosystem  
299 models, in which they were strongly present. In the forest floor models this was even more  
300 explicit. A reduced growth for oak was observed on sites from eumoder (6) onwards to mor  
301 (8) humus forms, whereas for pine a reduced site index was observed on the poorest sites with  
302 a mor humus form only. For beech, similar as for pine, a positive effect of the product mR x  
303 mN was observed. However, the threshold for beech was remarkably higher than the one for  
304 pine, indicating a higher trophicity need for beech than for pine.

305 **Discussion**

306 Species' optimal environmental conditions

307 Vegetation indicators

308 Floristic indices are often excellent indicators of site quality, as they are more informative  
309 about the plant available nutrients compared to soil analysis results. Their high correlation  
310 with various soil nutrients, soil physical and/or climatic variables makes them often cost-  
311 effective and accurate diagnostic indicators for site quality assessment (Bergès et al. 2006;  
312 Rogister 1978; Seynave et al. 2005; Wilson et al. 2001).

313 The ecogram developed for beech by Rogister (1978) is to a great extent confirmed in our  
314 study (Table 2), assigning highest productivity to sites with high mR x mN values (mull  
315 humus) with intermediate moisture content (mesophilic). The predictors of Rogister's  
316 ecogram are also the most contributing in our forest floor model of beech, and a model  
317 reduced to those two variables was only slightly weaker (results not presented). There is an  
318 increase of site productivity of beech in Flanders with increasing moisture availability  
319 reaching an optimum at mF 5.4 to 6.2 (moist to dampness sites according to Hill et al. (1999)).  
320 Wetter soils are probably less suitable but were not observed as beech is rarely planted on wet  
321 sites in Flanders because of its assumed vulnerability to fluctuating and stagnating  
322 groundwater (Masson 2005). For mR x mN an abrupt threshold value of 24 is identified,  
323 separating low from high productive sites. This threshold corresponds well with the boundary  
324 between eumoder and mull (Muys and Granval 1997) and underlines beech's preference for  
325 nutrient rich soils.

326 Humus indicators

327 Humus forms reflect the rate at which nutrients are cycled within terrestrial ecosystems. They  
328 are considered, similar to the floristic mR x mN indicator, as a strong indicator of the  
329 ecosystem's nutritional regime (Ponge and Chevalier 2006; Wilson et al. 2001). From our  
330 study it is hard to distinguish which indicator should be preferred as both predictors are  
331 present in particular models and in the forest floor model for pine they even complement each  
332 other. The considerable presence of humus predictors in the ecosystem models indicates their  
333 strength as site quality indicators.

334 Sites of lower productivity for pedunculate oak are linked to eumoder (6) to mor (8) humus  
335 forms. Similar observations have been made for sessile oak in France although for this species  
336 a productivity decrease was only noticed on dysmoder (7) and mor sites (Bergès et al. 2005;  
337 Jacquemin et al. 2000). This difference confirms the higher trophicity need of pedunculate  
338 oak compared to sessile oak described by silviculture experts (Lévy et al. 1992; Masson  
339 2005). Scots pine is obviously less nutrient demanding and lower site productivity is only  
340 noticed at mor sites. Similar observations have been made for lodgepole pine, with a  
341 significant lower site index at mor sites compared to moder sites (Szwaluk and Strong 2003).

342 Additional to the humus form were also the thickness of endorganic and ectorganic horizons  
343 already good predictors of site quality. Both variables were present in the ecosystem models  
344 of oak and beech. The thickness of both horizons is the result of a complex humus formation  
345 process and thus an indicator of the nutrient regime of the soil (Ponge and Chevalier 2006).  
346 The assignment of humus characteristics to the system of 9 humus forms proved sometimes  
347 imperfect (Ponge et al. 2011), resulting that humus thickness indicators can be a good  
348 alternative.

349 Soil indicators

350 Tree growth is mainly affected by the combined effects of soil water and nutrient availability,  
351 illustrated for pedunculate oak by Lévy et al. (1992). Texture was the most important  
352 predictor of site quality for all three species (Table 2). As soil granulometry plays an  
353 important role in both water- and nutrient holding capacity, it is not surprising that textural  
354 fractions are of major importance in prediction of site productivity. The study area is indeed  
355 characterised by a strong textural gradient along which large productivity differences occurred.  
356 Especially the silt fraction was positively correlated with tree growth, as it is a predictor in all  
357 ecosystem models. Soils with a high silt fraction have a higher available water capacity than  
358 more clay or sandy soils (Vereecken et al. 1989). But soil texture is also related to nutrient  
359 availability. Sandy soils are generally poor soils, whereas the higher cation exchange capacity  
360 and the higher weathering of silt and clay minerals result in higher nutrient availability for  
361 tree growth.

362 Soil acidity is also an important site quality factor especially for oak and beech. Soil  
363 acidification can lead to phytotoxic levels of aluminium in the rhizosphere and may disturb or  
364 inhibit nutrient uptake, leading to deficiencies of some elements (e.g., Mg, K, Ca, P) or excess  
365 of others (e.g., N) (Bakker et al. 1999). For both species the effect appears with the same  
366 threshold pH CaCl<sub>2</sub> -value of 3.6, which corresponds to a pH-H<sub>2</sub>O of around 4.2, below which  
367 growth reduction is observed. This threshold corresponds to the upper limit of the Al buffer  
368 range (Ulrich 1983) and to the transition between mull and moder humus types (Muys 1995).  
369 Below this threshold even acid tolerant tree species experience growth reduction (Ulrich  
370 1983). Similar observations were made for beech in France, where an increased growth was  
371 found when the pH ranged between 4 and 5.5 (Seynave et al. 2008).

372 Above a topsoil C:N ratio of about 19 a relatively sharp decrease in productivity for oak is  
373 observed. This is consistent with Bary-Lenger & Nebout (1993), who classified soils with

374 organic topsoil C:N ratio below 15 as good and above 20 as less favourable for pedunculate  
375 oak. High C:N values are an indication of poor litter decomposition and N-mineralisation and  
376 are related to a low soil biological activity (Muys et al. 1992). A similar negative relation  
377 between productivity and soil C:N has been observed for sessile oak in France and was  
378 explained by the higher acidity or weaker N-mineralisation at these sites (Bontemps 2006).

#### 379 Litterfall nutrient concentrations

380 Litterfall nutrient concentrations appeared in all ecosystem models among the most  
381 contributing independent variables. The strength of the litterfall variables can be attributed to  
382 the fact that these observations are directly taken from the trees and can serve as indicators of  
383 nutrient availability and uptake within the rooting zone or retranslocation success, whereas  
384 soil nutrient concentrations often poorly reflect the plant available nutrient concentrations  
385 especially for N (Axmanová et al. 2011).

386 Although many studies revealed a positive relationship between site productivity and N  
387 deposition (Janssens and Luysaert 2009; Magnani et al. 2007), De Schrijver et al. (2008)  
388 indicate that N saturation of forest ecosystems is a problem in regions with high immissions  
389 like Flanders. For all three species, litter nutrient concentrations indicated that N saturation  
390 was affecting site productivity negatively.

391 For beech a lower site productivity is observed with high litterfall N:P ratios (Fig. 3). This  
392 effect has been observed before for beech in Europe and attributed to P deficiency, mainly  
393 induced by N deposition (Braun et al. 2010; Bussotti et al. 2005; Ewald 2000; Fluckiger and  
394 Braun 1998; Ilg et al. 2009). The partial dependence plot (Fig. 3) shows a clear threshold of  
395 N:P ratio in the litter equal to 17 above which growth reduction is observed, which is slightly  
396 higher than the critical value of 15, earlier described in literature and recorded in green leaves

397 (Bussotti et al. 2005; Ilg et al. 2009). This corresponds with the fact that resorption of P from  
398 senescent leaves exceeds resorption of N resulting in a higher N:P ratio in senescent  
399 compared to green leaves (Richardson et al. 2008). An increased foliar N:P ratio accompanied  
400 by a growth decline has been recorded especially in mature beech stands in different  
401 European regions. Different experiments do not attribute this phenomenon to limited P  
402 availability but rather support the hypothesis that elevated N deposition is an important cause  
403 for this development (Braun et al. 2010). Given that we did not detect a significant correlation  
404 between litter P content and the plant available P content in the soil ( $r = 0.10$ ), P availability  
405 was probably not the limiting factor. Rather a decreased uptake of P due an inhibition of  
406 mycorrhiza by high nitrogen concentrations (Braun et al. 2010; Muys 1990) seemed to be  
407 limiting for beech since mycorrhiza are important for the P uptake by forest trees (Braun et al.  
408 2010; Gusewell 2004; Wallander 2000).

409 For oak and pine critical values for P deficiency (Ilg et al. 2009) were rarely reached,  
410 indicating that P was seldom a limiting factor for these species. This corresponds with  
411 observations made in Germany where beech was affected by P deficiency whereas oak and  
412 pine did not suffer (Ilg et al. 2009). On the other hand both oak and pine showed a significant  
413 decrease in productivity with high litterfall N content (Fig 2 & 4). The negative effect of high  
414 N content in the needles of Scots pine on plant vitality has been reported for different  
415 European regions (Roskams and Neiryck 1999; Sikström et al. 1998; Thelin et al. 1998).  
416 Many forests in Belgium are subjected to high atmospheric nitrogen inputs and especially  
417 coniferous forests act as receptors (De Schrijver et al. 2007; Neiryck et al. 2007). Deficiency  
418 symptoms of P, K and Mg may appear even at relatively high concentrations in relation with  
419 an increased N content. High N supply often results in an imbalanced nutrient uptake  
420 resulting in a decreased growth (Roskams and Neiryck 1999; Sikström et al. 1998; Van den  
421 Burg and Schaap 1995). Litter N content of pine is highly correlated with N:P, N:Mg and N:K

422 ratios indicating that it is probably not a deficiency of a single nutrient, but rather a  
423 imbalanced mineral nutrition.

424 For oak the effect of N saturation through deposition is less obvious. The soil C:N content  
425 indicates a higher productivity at relatively better N supply, whereas high N content in the  
426 senescing leaves is observed in the low productive sites. Vizoso et al. (2008) postulated that  
427 the spring growth of pedunculate oak was strongly influenced by the availability of internal N  
428 resources with an almost complete dependence on N remobilization for their growth. Higher  
429 N retranslocation from senescing leaves leads to higher internal N stocks and improved  
430 remobilization and growth at the beginning of the following spring, introducing a positive  
431 feedback loop. Millard and Grelet (2010) also conclude that the N retranslocation capacity of  
432 the trees is an even better predictor of productivity than nitrogen supply as such.

433 Comparison of models based on different variable sets

434 Our results have illustrated the strength of BRT to model ecological threshold behaviour,  
435 especially in the ecosystem models (Fig. 2-4). More traditional regression techniques as  
436 multiple linear regression or generalized additive models (GAM) would not be able to handle  
437 these discontinuities in a similar way (Aertsen et al. 2011; Aertsen et al. 2010). The automatic  
438 detection of interactions between predictors is also an advantage of BRT (De'ath 2007; Elith  
439 et al. 2008). Important interactions were detected and modelled for oak between the depth of  
440 the gleyic horizon and the soil C:N ratio and also between the depth of the gleyic horizon and  
441 the silt content (Fig. 5). A small point of criticism on BRT is the fact that the gbm software  
442 package does not provide any measure of uncertainty in the partial dependence plots, as it is  
443 the case in most GAM software packages. Adding uncertainty would greatly improve the gbm  
444 software package. Although BRT is rather insensitive to outliers as they are isolated into a

445 node and so have no effect on splitting, predictions close to these outliers will enclose a  
446 higher uncertainty.

447 The ecosystem models, based on all collected variables, explained more of the variability and  
448 were more accurate than those limited to either soil or forest floor variables. Combining forest  
449 floor and soil variables and adding litterfall nutrient concentrations resulted both in improved  
450 models. Although models based on the combination of forest floor and soil variables (without  
451 the inclusion of litterfall nutrient concentrations) performed better than the separate models  
452 (Table 3), the inclusion of litterfall nutrient concentrations results in improved explanation of  
453 variability. With  $R^2$  values ranging from 0.67 to 0.77 and rRMSEs of less than 13% after  
454 cross-validation (Table 3), these models perform very well compared to results of similar site  
455 index modelling studies (Bergès et al. 2006; McKenney and Pedlar 2003; Szwaluk and Strong  
456 2003). The relatively good model performance can be attributed to the sound sampling  
457 strategy, the large amount and high quality of collected variables and the appropriate  
458 modelling technique. In comparison to other observational studies, the reported residual  
459 model error, which may be due to measurement errors, sampling bias, limitations in field data  
460 collection or genetic tree variability can be considered low.

461 As discussed in the previous section, the ecosystem models provide good insight in the  
462 ecosystem processes related to site productivity. However, the multiple dimensions of the  
463 environmental variables and their interactions also induce uncertainty about the cause-effect  
464 relationships (Chen et al. 1998). For example, nutrient related variables such as topsoil  
465 acidity, humus form, vegetation mR x mN, soil C:N ratio and litterfall nutrient concentrations,  
466 are all to some degree related and at least partly interchangeable. Moreover, the usefulness of  
467 these models for site index predictions is limited. First of all, large amounts of environmental  
468 variables have to be collected, including vegetation, humus, soil and litterfall variables, in

469 order to make predictions of site index possible. Furthermore, litterfall variables can only be  
470 collected at sites where the focal tree species is already present, making models including  
471 these variables unsuitable for afforestation or species conversion purposes. Simplified models  
472 based on a subset of (easily collectable) variables are therefore recommended for forest  
473 management practices, even if accuracy and predictive power are slightly lower.

474 As expected, the simplified models did not reach the same accuracy as the ecosystem models,  
475 with rRMSEs increasing until 10 to 17% after cross-validation (Table 3). Nevertheless, both  
476 the soil and forest floor models can serve as good predictive models for site index in Flanders.  
477 This confirms findings of Nieppola & Carleton (1991), Wang (2000) and Bergès et al. (2006)  
478 that forest floor variables as vegetation and humus can serve as good substitutes for several  
479 more difficult, expensive or integrated environmental variables, and therefore can act as good  
480 predictors for site quality.

481 Site index in Flanders were more accurately predicted by soil predictors than by forest floor  
482 ones. Bergès et al. (2006) found no difference in predictive potential between both predictor  
483 groups for sessile oak in France. The French study however covered a larger geographical  
484 area, not only characterised by larger variation in soil properties, but also by topographical  
485 and climatic variability. The predictive power of vegetation as a substitute for the complex  
486 interrelation of all these factors was therefore probably stronger than in our study, in which  
487 topographical and climate gradients are of minor influence and thus variability in site  
488 productivity was mainly characterised by soil properties. Bergès et al. (2006) also raised the  
489 hypothesis that understory vegetation is a better predictor of nutrient-related factors whereas  
490 soil variables are better predictors of water-related factors. Our results could however not  
491 confirm this hypothesis, since soil granulometry was determined as the most influencing  
492 predictor of site index for the three species in both soil and ecosystem models. This variable,

493 correlated to both water- and nutrient-related factors, was not included in the study of Berges  
494 et al. (2006).

495 Another possible explanation for the fact that our soil models perform better than the forest  
496 floor models could be that humus and vegetation are mainly indicators of topsoil  
497 characteristics, whereas it is well known that tree roots go much deeper in their search for  
498 nutrients and water. In that perspective our set of soil variables is more complete as it  
499 represents both top- and deeper soil characteristics. Finally, the use of humus and vegetation  
500 as site descriptors can also be contested by the fact that stand characteristics like age and  
501 canopy structure have an impact on the composition of herb species and humus quality in the  
502 region (Godefroid et al. 2005; Ponge et al. 2011; Ponge and Chevalier 2006). Although we  
503 tried to eliminate these factors as much as possible in the research setup by sampling only  
504 mature, even-aged, closed-canopy and nearly pure stands, this can never be totally excluded.

#### 505 **Acknowledgements**

506 This research was supported by the Research Fund K.U.Leuven (OT/07/046) and the  
507 SimForTree project (IWT-SBO contract 060032). Special thanks go to Marc De Vrieze and  
508 Joost Malliet for the tremendous amount of field work they accomplished and the Agency for  
509 Nature and Forest for granting access permission to the forests.

510 **References**

- 511 Aertsen W, Kint V, Van Orshoven J and Muys B 2011 Evaluation of modelling techniques  
 512 for forest site productivity prediction in contrasting ecoregions using stochastic  
 513 multicriteria acceptability analysis (SMAA). *Environ. Model. Software* 26, 929-937.
- 514 Aertsen W, Kint V, Van Orshoven J, Ozkan K and Muys B 2010 Comparison and ranking of  
 515 different modelling techniques for prediction of site index in Mediterranean mountain  
 516 forests. *Ecol. Model.* 221, 1119-1130.
- 517 Assmann E 1970 *The principles of forest yield study*. Pergamon Press, Oxford.
- 518 Axmanová I, Zelený D, Li C F and Chytrý M 2011 Environmental factors influencing herb  
 519 layer productivity in Central European oak forests: insights from soil and biomass  
 520 analyses and a phytometer experiment. *Plant Soil* 342, 183-194.
- 521 Bakker M R, Nys C and Picard J F 1999 The effects of liming and gypsum applications on a  
 522 sessile oak (*Quercus petraea* (M.) Liebl.) stand at La Croix-Scaille (French Ardennes)  
 523 I. Site characteristics, soil chemistry and aerial biomass. *Plant Soil* 206, 99-108.
- 524 Bary-Lenger A and Nebout J 1993 *Le chêne*. Editions du Perron, Liège.
- 525 Bergès L, Chevalier R, Dumas Y, Franc A and Gilbert J M 2005 Sessile oak (*Quercus*  
 526 *petraea* Liebl.) site index variations in relation to climate, topography and soil in  
 527 even-aged high-forest stands in northern France. *Ann. For. Sci.* 62, 391-402.
- 528 Bergès L, Gégout J C and Franc A 2006 Can understory vegetation accurately predict site  
 529 index? A comparative study using floristic and abiotic indices in sessile oak (*Quercus*  
 530 *petraea* Liebl.) stands in northern France. *Ann. For. Sci.* 63, 31-42.
- 531 Bontemps J-D 2006 Evolution de la productivité des peuplements réguliers et  
 532 monospécifiques de hêtre (*Fagus sylvatica* L.) et de chêne sessile (*Quercus petraea*  
 533 Liebl.) dans la moitié Nord de la France au cours du XXe siècle. . In *Laboratoire*  
 534 *d'Etude des Ressources Forêt-Bois*. ENGREF, Champenoux.
- 535 Braun S, Thomas V F D, Quiring R and Fluckiger W 2010 Does nitrogen deposition increase  
 536 forest production? The role of phosphorus. *Environ. Pollut.* 158, 2043-2052.
- 537 Bray R H and Kurtz L T 1945 Determination of total, organic, and available forms of  
 538 phosphorus in soils. *Soil Sci.* 59, 39.
- 539 Bussotti F, Pancrazi M, Matteucci G and Gerosa G 2005 Leaf morphology and chemistry in  
 540 *Fagus sylvatica* (beech) trees as affected by site factors and ozone: results from  
 541 CONECOFOR permanent monitoring plots in Italy. *Tree Physiol.* 25, 211-219.
- 542 Chen H Y H, Klinka K and Kabzems R D 1998 Site index, site quality, and foliar nutrients of  
 543 trembling aspen: relationships and predictions. *Can. J. For. Res./Rev. Can. Rech. For.*  
 544 28, 1743-1755.
- 545 De'ath G 2007 Boosted trees for ecological modeling and prediction. *Ecology* 88, 243-251.
- 546 De Schrijver A, Geudens G, Augusto L, Staelens J, Mertens J, Wuyts K, Gielis L and  
 547 Verheyen K 2007 The effect of forest type on throughfall deposition and seepage flux:  
 548 a review. *Oecologia* 153, 663-674.
- 549 De Schrijver A, Verheyen K, Mertens J, Staelens J, Wuyts K and Muys B 2008 Nitrogen  
 550 saturation and net ecosystem production. *Nature* 451, E1.
- 551 De Vos B, Van Meirvenne M, Quataert P, Deckers J and Muys B 2005 Predictive quality of  
 552 pedotransfer functions for estimating bulk density of forest soils. *Soil Sci. Soc. Am. J.*  
 553 69, 500-510.
- 554 Dzwonko Z 2001 Assessment of light and soil conditions in ancient and recent woodlands by  
 555 Ellenberg indicator values. *Journal of Applied Ecology* 38, 942-951.
- 556 Elith J, Leathwick J R and Hastie T 2008 A working guide to boosted regression trees. *J.*  
 557 *Anim. Ecol.* 77, 802-813.

- 558 Ellenberg H, Weber H E, Düll R, Wirth V, Werner W and Paulissen D 1992 Zeigerwerte von  
559 Pflanzen in Mitteleuropas, 2nd ed. Scripta Geobotanica 18, 1-258.
- 560 Ewald J 2000 Does phosphorus deficiency cause low vitality in European beech (*Fagus*  
561 *sylvatica* L.) in the Bavarian Alps? Forstwissenschaftliches Centralblatt 119, 276-296.
- 562 Fluckiger W and Braun S 1998 Nitrogen deposition in Swiss forests and its possible relevance  
563 for leaf nutrient status, parasite attacks and soil acidification. Environ. Pollut. 102, 69-  
564 76.
- 565 Gégout J C, Hervé J C, Houllier F and Pierrat J C 2003 Prediction of forest soil nutrient status  
566 using vegetation. J. Veg. Sci. 14, 55-62.
- 567 Godefroid S, Massant W and Koedam N 2005 Variation in the herb species response and the  
568 humus quality across a 200-year chronosequence of beech and oak plantations in  
569 Belgium. Ecography 28, 223-235.
- 570 Güsewell S 2004 N : P ratios in terrestrial plants: variation and functional significance. New  
571 Phytol. 164, 243-266.
- 572 Håkansson I and Lipiec J 2000 A review of the usefulness of relative bulk density values in  
573 studies of soil structure and compaction. Soil and Tillage Research 53, 71-85.
- 574 Hill M, Mountford J, Roy B and Bunce R 1999 Ellenberg's indicator values for British plants.  
575 Center for Ecology and Hydrology.
- 576 Ilg K, Wellbrock N and Lux W 2009 Phosphorus supply and cycling at long-term forest  
577 monitoring sites in Germany. Eur. J. For. Res. 128, 483-492.
- 578 Jabiol B, Brêthes A, Ponge J, Toutain F and Brun J 2007 L'Humus sous toutes ses formes.  
579 AgoParisTech, Nancy.
- 580 Jacquemin B, Lacroix R, Démarcq P and Duplat P 2000 Relation entre station et fertilité du  
581 chêne sessile (*Quercus petraea* Liebl.) en forêt de Tronçais. Bulletin Technique de  
582 l'ONF 39, 33-44.
- 583 Jansen J J, Sevenster J and Faber P J 1996 Opbrengsttabellen voor belangrijke boomsoorten  
584 in Nederland. IBN-DLO, Wageningen.
- 585 Janssens I A and Luysaert S 2009 Nitrogen's carbon bonus. Nature Geoscience 2, 318-319.
- 586 Kint V, Vansteenkiste D, Aertsen W, De Vos B, Bequet R, Van Acker J and Muys B 2011  
587 Forest structure and soil fertility determine internal stem morphology of Pedunculate  
588 oak: a modelling approach using boosted regression trees. Eur. J. For. Res., In press.
- 589 Lévy G, Becker M and Duhamel D 1992 A comparison of the ecology of pedunculate and  
590 sessile oaks: radial growth in the centre and northwest of France. For. Ecol. Manag.  
591 55, 51-63.
- 592 Magnani F, Mencuccini M, Borghetti M, Berbigier P, Berninger F, Delzon S, Grelle A, Hari  
593 P, Jarvis P G, Kolari P, Kowalski A S, Lankreijer H, Law B E, Lindroth A, Loustau D,  
594 Manca G, Moncrieff J B, Rayment M, Tedeschi V, Valentini R and Grace J 2007 The  
595 human footprint in the carbon cycle of temperate and boreal forests. Nature 447, 848-  
596 850.
- 597 Masson G 2005 Autécologie des essences forestières. Tec & Doc, Paris.
- 598 McKenney D W and Pedlar J H 2003 Spatial models of site index based on climate and soil  
599 properties for two boreal tree species in Ontario, Canada. For. Ecol. Manag. 175, 497-  
600 507.
- 601 Millard P and Grelet G A 2010 Nitrogen storage and remobilization by trees:  
602 ecophysiological relevance in a changing world. Tree Physiol. 30, 1083-1095.
- 603 Moisen G G, Freeman E A, Blackard J A, Frescino T S, Zimmermann N E and Edwards T C  
604 2006 Predicting tree species presence and basal area in Utah: A comparison of  
605 stochastic gradient boosting, generalized additive models, and tree-based methods.  
606 Ecol. Model. 199, 176-187.
- 607 Muys B 1990 N-excess in the forest: effects and possible measures. Silva Gandavensis, 35-42.

- 608 Muys B 1995 The influence of tree species on humus quality and nutrient availability on a  
609 regional scale (Flanders, Belgium). In Nutrient uptake and cycling in forest  
610 ecosystems. Eds. L O Nilsson, R F Hüttl and U T Johansson. pp 649-660. Kluwer  
611 Academic Publishers, Dordrecht.
- 612 Muys B and Granval P 1997 Earthworms as bio-indicators of forest site quality. *Soil Biol.*  
613 *Biochem.* 29, 323-328.
- 614 Muys B, Lust N and Granval P 1992 Effects of grassland afforestation with different tree  
615 species on earthworm communities, litter decomposition and nutrient status. *Soil Biol.*  
616 *Biochem.* 24, 1459-1466.
- 617 Neiryneck J, Kowalski A S, Carrara A, Genouw G, Berghmans P and Ceulemans R 2007  
618 Fluxes of oxidised and reduced nitrogen above a mixed coniferous forest exposed to  
619 various nitrogen emission sources. *Environ. Pollut.* 149, 31-43.
- 620 Nieppola J J and Carleton T J 1991 Relations between understorey vegetation, site  
621 productivity, and environmental factors in *Pinus sylvestris* L. stands in southern  
622 Finland. *Vegetatio* 93, 57-72.
- 623 Ponge J-F, Jabiol B and Gégout J-C 2011 Geology and climate conditions affect more humus  
624 forms than forest canopies at large scale in temperate forests. *Geoderma* 162, 187-195.
- 625 Ponge J F and Chevalier R 2006 Humus Index as an indicator of forest stand and soil  
626 properties. *For. Ecol. Manag.* 233, 165-175.
- 627 Ponge J F, Chevalier R and Loussot P 2002 Humus index: An integrated tool for the  
628 assessment of forest floor and topsoil properties. *Soil Sci. Soc. Am. J.* 66, 1996-2001.
- 629 Richardson S J, Allen R B and Doherty J E 2008 Shifts in leaf N : P ratio during resorption  
630 reflect soil P in temperate rainforest. *Funct. Ecol.* 22, 738-745.
- 631 Ridgeway G 2006 Generalized boosted models: a guide to the gbm package.
- 632 Rogister J 1978 De ekologiske mR- en mN-waarden van de kruidlaag en de humuskwaliteit  
633 van bosplantengezelschappen. Proefstation van Waters en Bossen, Groenendaal.
- 634 Roskams P and Neiryneck J 1999 De voedingstoestand van Grove den (*Pinus sylvestris* L.) in  
635 het Level II-proefvlak in Brasschaat In Mededelingen. Instituut voor Bosbouw en  
636 Wildbeheer, Geraardsbergen.
- 637 Schapire R E 2003 The boosting approach to machine learning: An overview. *Nonlinear*  
638 *Estimation and Classification* 171, 149-171.
- 639 Seynave I, Gégout J C, Hervé J C and Dhôte J F 2008 Is the spatial distribution of European  
640 beech (*Fagus sylvatica* L.) limited by its potential height growth? *J. Biogeogr.* 35,  
641 1851-1862.
- 642 Seynave I, Gégout J C, Herve J C, Dhote J F, Drapier J, Bruno E and Dume G 2005 *Picea*  
643 *abies* site index prediction by environmental factors and understorey vegetation: a  
644 two-scale approach based on survey databases. *Can. J. For. Res./Rev. Can. Rech. For.*  
645 35, 1669-1678.
- 646 Sikström U, Nohrstedt H O, Pettersson F and Jacobson S 1998 Stem-growth response of  
647 *Pinus sylvestris* and *Picea abies* to nitrogen fertilization as related to needle nitrogen  
648 concentration. *Trees* 12, 208-214.
- 649 Skovsgaard J P and Vanclay J K 2008 Forest site productivity: a review of the evolution of  
650 dendrometric concepts for even-aged stands. *Forestry* 81, 12-31.
- 651 Szwaluk K S and Strong W L 2003 Near-surface soil characteristics and understory plants as  
652 predictors of *Pinus contorta* site index in southwestern Alberta, Canada. *For. Ecol.*  
653 *Manag.* 176, 13-24.
- 654 Teepe R, Dilling H and Beese F 2003 Estimating water retention curves of forest soils from  
655 soil texture and bulk density. *J. Plant Nutr. Soil Sci.* 166, 111-119.

- 656 Thelin G, Rosengren-Brinck U, Nihlgard B and Barkman A 1998 Trends in needle and soil  
657 chemistry of Norway spruce and Scots pine stands in South Sweden 1985-1994.  
658 Environ. Pollut. 99, 149-158.
- 659 Ulrich B 1983 Soil acidity and its relations to acid deposition. In Effects of accumulation of  
660 air pollutants in forest ecosystems. Eds. B Ulrich and J Pankrath. pp 127-146. Reidel  
661 Publishing Company, Dordrecht.
- 662 Van den Burg J and Schaap W 1995 Richtlijnen voor mineralentoediening en bekalking als  
663 effectgerichte maatregelen in bossen. p. 63. Informatie- en KennisCentrum  
664 Natuurbeheer, Wageningen.
- 665 Vereecken H, Maes J, Feyen J and Darius P 1989 Estimating the soil-moisture retention  
666 characteristic from texture, bulk-density, and carbon content. Soil Sci. 148, 389-403.
- 667 Vizoso S, Gerant D, Guehl J M, Joffre R, Chalot M, Gross P and Maillard P 2008 Do  
668 elevation of CO<sub>2</sub> concentration and nitrogen fertilization alter storage and  
669 remobilization of carbon and nitrogen in pedunculate oak saplings? Tree Physiol. 28,  
670 1729-1739.
- 671 Wall A and Heiskanen J 2009 Soil-water content and air-filled porosity affect height growth  
672 of Scots pine in afforested arable land in Finland. For. Ecol. Manag. 257, 1751-1756.
- 673 Wallander H 2000 Uptake of P from apatite by *Pinus sylvestris* seedlings colonised by  
674 different ectomycorrhizal fungi. Plant Soil 218, 249-256.
- 675 Wang G G 2000 Use of understory vegetation in classifying soil moisture and nutrient  
676 regimes. For. Ecol. Manag. 129, 93-100.
- 677 Wilson S M, Pyatt D G, Malcolm D C and Connolly T 2001 The use of ground vegetation and  
678 humus type as indicators of soil nutrient regime for an ecological site classification of  
679 British forests. For. Ecol. Manag. 140, 101-116.
- 680

**Table 1.** Summary of the site characteristics stratified according to the three studied tree species

Variable	Pedunculate oak (n=76)			Common beech (n=55)			Scots pine (n=55)		
	Mean (S.D.)	Min.	Max.	Mean (S.D.)	Min.	Max.	Mean (S.D.)	Min.	Max.
Site index (m)	28.98 (3.93)	20.61	36.61	37.77 (4.45)	26.21	43.57	24.07 (4.24)	15.81	35.32
mF	5.9 (0.6)	5.0	7.9	5.4 (0.4)	5.0	6.3	5.8 (0.7)	5.0	7.9
mN	5.0 (1.0)	2.2	6.1	4.9 (1.0)	3.0	7.0	3.7 (1.0)	2.1	6.0
mR	5.0 (0.9)	2.5	6.5	4.7 (0.7)	3.0	6.0	3.4 (1.0)	2.1	6.0
mL	5.6 (0.6)	4.2	6.9	4.9 (0.8)	3.0	6.0	6.0 (0.4)	5.2	7.0
mR x mN	26.2 (8.6)	7.0	39.6	23.5 (6.2)	9.0	35.9	13.2 (7.5)	6.2	35.6
Humus form	5.8 (1.0)	4.0	8.0	6.5 (0.7)	5.0	8.0	7.1 (0.8)	5.3	8.0
Ectorganic thickness (cm)	5.1 (2.8)	1.0	12.0	4.6 (3.3)	0.0	20.0	5.7 (3.3)	1.0	15.0
Endorganic thickness (cm)	4.8 (4.7)	0.0	21.0	2.9 (2.0)	0.1	11.7	4.0 (3.9)	0.0	18.8
Clay (%)	16.6 (7.9)	4.2	31.1	17.2 (8.7)	4.8	30.8	7.9 (5.6)	1.9	21.7
Silt (%)	28.7 (20.6)	3.2	64.7	34.5 (20.5)	4.9	65.8	12.2 (11.3)	1.9	50.4
Sand (%)	54.7 (27.8)	8.4	92.7	48.3 (29.1)	5.3	89.6	79.9 (16.7)	27.9	96.2
Soil pH	3.59 (0.17)	3.24	4.14	3.64 (0.11)	3.31	3.93	3.53 (0.21)	3.1	3.95
Soil Organic Matter (mg g <sup>-1</sup> )	41.8 (13.8)	12.4	81.3	38.1 (9.4)	17.1	63.8	28.4 (11.4)	9.7	52.0
Soil N (mg g <sup>-1</sup> )	1.2 (0.4)	0.4	2.8	1.0 (0.2)	0.5	1.8	0.7 (0.3)	0.3	1.5
Soil P (mg kg <sup>-1</sup> )	9.42 (7.47)	1.0	28.6	7.37 (5.89)	1.0	24.5	6.38 (6.32)	1.0	32.6
Soil C:N	19.07 (4.12)	9.91	39.6	22.39 (3.11)	17.22	29.67	24.28 (7.86)	13.5	54
Depth of clay horizon (cm)	76.08 (42.54)	0	120	76.18 (42.78)	0	120	96.89 (37.99)	7	120
Depth of gleyic horizon (cm)	51.88 (34.04)	0	120	69.05 (29.06)	18	120	88.31 (34.13)	10	120
Depth of reduction horizon (cm)	106.74 (22.02)	40	120	116.04 (15.02)	32	120	116.44 (14.23)	30	120
Bulk density (g cm <sup>-3</sup> )	1.41 (0.07)	1.23	1.58	1.43 (0.05)	1.30	1.55	1.18 (0.06)	1.36	1.59
Available water capacity (cm <sup>3</sup> cm <sup>-3</sup> )	0.19 (0.01)	0.16	0.20	0.19 (0.01)	0.16	0.21	0.17 (0.01)	0.16	0.20
Air capacity (cm <sup>3</sup> cm <sup>-3</sup> )	0.14 (0.05)	0.04	0.22	0.13 (0.06)	0.04	0.21	0.18 (0.03)	0.08	0.22
Permanent wilting point (cm <sup>3</sup> cm <sup>-3</sup> )	0.14 (0.05)	0.06	0.22	0.15 (0.05)	0.07	0.23	0.09 (0.03)	0.05	0.18
Litterfall N (mg g <sup>-1</sup> )	11.6 (0.9)	10.1	14.0	9.5 (1.1)	7.8	12.7	7.3 (1.5)	5.2	11.3
Litterfall P (mg kg <sup>-1</sup> )	970 (284)	395	1451	501 (182)	263	951	331 (83)	225	531
Litterfall K (mg kg <sup>-1</sup> )	4325 (894)	2606	6505	4477 (1466)	1856	8359	1617 (395)	1068	2491
Litterfall Ca (mg kg <sup>-1</sup> )	10165 (2674)	5707	15581	6694 (1839)	3533	9994	5027 (1557)	2737	11069
Litterfall Mg (mg kg <sup>-1</sup> )	1153 (182)	837	1511	907 (237)	365	1293	504 (78)	319	696

Litterfall Na (mg kg <sup>-1</sup> )	353 (83)	196	620	317 (101)	204	629	288 (119)	167	621
Litterfall Al (mg kg <sup>-1</sup> )	107 (39)	54	243	127 (28)	80	207	219 (46)	149	358
Litterfall Fe (mg kg <sup>-1</sup> )	186 (41)	124	328	207 (50)	139	309	173 (27)	131	261
Litterfall Mn (mg kg <sup>-1</sup> )	1492 (1204)	322	3990	1351 (958)	263	3838	503 (441)	131	1866
Litterfall N:P	13.59 (6.03)	8.39	30.94	21.62 (7.87)	9.93	37.06	23.49 (8.29)	13.89	43.28

681 **Table 2.** Predictor effects and their ranking in the different models. Effect legend: ↑/↓  
682 increasing/decreasing threshold (with value), ↗/↘ gradual increase/decrease (with range); double  
683 arrows and arrows between brackets indicate strong/weak effects (i.e. covering >50% or <20% of the  
684 common range of the fitted function) respectively. The ranking of predictors in each model is based on  
685 predictor contributions. Predictors of the ecosystem models are plotted in Figs 2-4, whereas those of  
686 the soil and forest floor models are found in the supporting information.

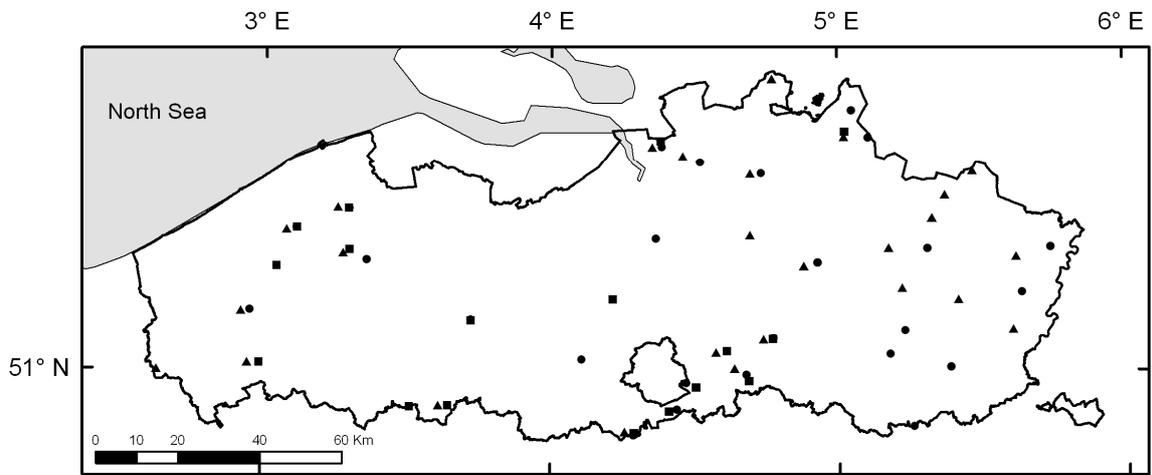
Variable set	Species	Predictor	Contribution (%)	effect	
				direction & magnitude	range or value
Ecosystem model (All variables)	Oak	Silt (%)	22.3	↗↗	[10-50]
		Litter N (mg g <sup>-1</sup> )	19.4	↘↘	[1.12-1.24]
		Ectorganic thickness (cm)	16.5	↘↘	[4-7]
		Soil pH	11.4	↑↑	3.6
		Endorganic thickness (cm)	11.0	↑	4
	Beech	Soil C:N	10.0	↓	19
		Depth of gleyic horizon (cm)	9.4	↑	40
		Litter N:P	51.3	↘↘	18
		Silt (%)	15.1	↗	[20-55]
		mF	14.9	↗	[5.0-5.4]
	Pine	Endorganic thickness (cm)	7.5	↑	2.5
		Ectorganic thickness (cm)	6.1	↓	5
		Depth of clay acc. (cm)	5.1	↘	[90-120]
		Litter N (mg g <sup>-1</sup> )	63.5	↘↘	0.75
		mR x mN	14.9	↗	[9-17]
	mF	7.8	(↑)	5.8	
	Silt (%)	7.6	(↑)	8	
	Soil OM (mg g <sup>-1</sup> )	6.2	(↑)	2.5	
Soil + forest floor model (Soil + humus + vegetation variables)	Oak	Sand (%)	28.1	↘↘	[20-80]
		Ectorganic thickness (cm)	24.5	↘↘	[4-7]
		Depth of gleyic horizon (cm)	14.9	↑	40
		Humus form	12.5	↘↘	5.5
		Soil C:N	10.3	↗	[17-22]
	Beech	mF	9.7	↘	[5.5-6.0]
		Silt (%)	31.6	↗↗	[20-55]
		mF	18.6	↗↗	[5.0-5.4]
		mR x mN	13.7	↓	24
		Endorganic thickness (cm)	11.0	↑	2.5
	Pine	Soil pH	9.8	↑↑	3.6
		Ectorganic thickness (cm)	7.9	↓	5
		Depth of clay acc. (cm)	7.5	↘	[90-120]
		Silt (%)	25.3	↗↗	[6-24]
		mR x mN	22.6	↗↗	[9-17]
		Depth of gleyic horizon (cm)	17.0	↑	100
		Soil OM (mg g <sup>-1</sup> )	15.6	↑	2.5
		mF	9.7	(↑)	5.8
Humus form		9.7	(↘)	[7-7.5]	
Soil model (Soil variables)	Oak	Sand (%)	57.5	↘↘	[20-80]
		Depth of gleyic horizon (cm)	19.8	↑	40
		Soil pH	11.7	↗	[3.4-3.8]
	Beech	Soil C:N	11.0	↓	19
		Silt (%)	69.2	↗↗	[25-55]
		Soil pH	16.4	↑	3.6
	Pine	Depth of clay acc. (cm)	14.4	↘	[40-120]
		Silt (%)	45.4	↗↗	[6-24]
		Soil OM (mg g <sup>-1</sup> )	28.0	↑	2.5
	Depth of gleyic horizon (cm)	26.6	↑	100	
Forest floor model (Humus-vegetation var.)	Oak	Ectorganic thickness (cm)	46.1	↘↘	[4-7]
		Humus form	27.9	↘↘	5.5
		mF	26.0	↘↘	[5.5-6.0]
	Beech	mF	33.1	↗↗	[5.0-5.4]
		mR x mN	25.3	↑	24
		Ectorganic thickness (cm)	21.2	↘↘	5
	Pine	Endorganic thickness (cm)	20.5	↑	2.5
		mR x mN	61.7	↗↗	[9-17]
		mF	23.1	↑	5.8
	Humus form	15.2	↘	[7-8]	

687  
688

**Table 3.** Model performance indicators: coefficient of determination ( $R^2$ ), root mean squared error (RMSE, in m) and relative RMSE (rRMSE, in %) for model fit and for 10-fold cross-validation.

Variable set	Species	Fit			Cross-validation		
		$R^2$	RMSE	rRMSE	$R^2$	RMSE	rRMSE
Ecosystem model (All variables)	Oak	0.67	2.33	8.05	0.32	3.23	11.13
	Beech	0.77	2.11	6.08	0.58	2.85	8.21
	Pine	0.68	2.41	10.00	0.47	3.06	12.71
Soil + forest floor model (Soil, humus and vegetation variables)	Oak	0.56	2.68	9.24	0.27	3.32	11.45
	Beech	0.75	2.22	6.38	0.51	3.09	8.88
	Pine	0.60	2.78	11.55	0.22	3.73	15.48
Soil model (Soil variables)	Oak	0.53	2.81	9.68	0.25	3.39	11.69
	Beech	0.60	2.82	8.12	0.37	3.52	10.11
	Pine	0.47	3.15	13.10	0.21	3.73	15.50
Forest floor model (Humus-vegetation var.)	Oak	0.41	3.07	10.60	0.22	3.45	11.91
	Beech	0.53	3.12	8.98	0.29	3.72	10.70
	Pine	0.33	3.54	14.71	0.08	4.02	16.70

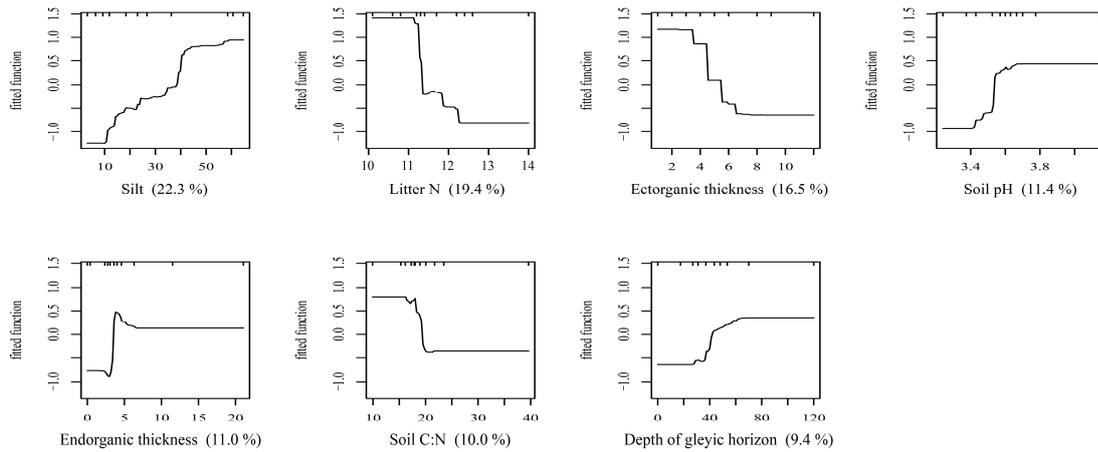
691 **Figures**



692

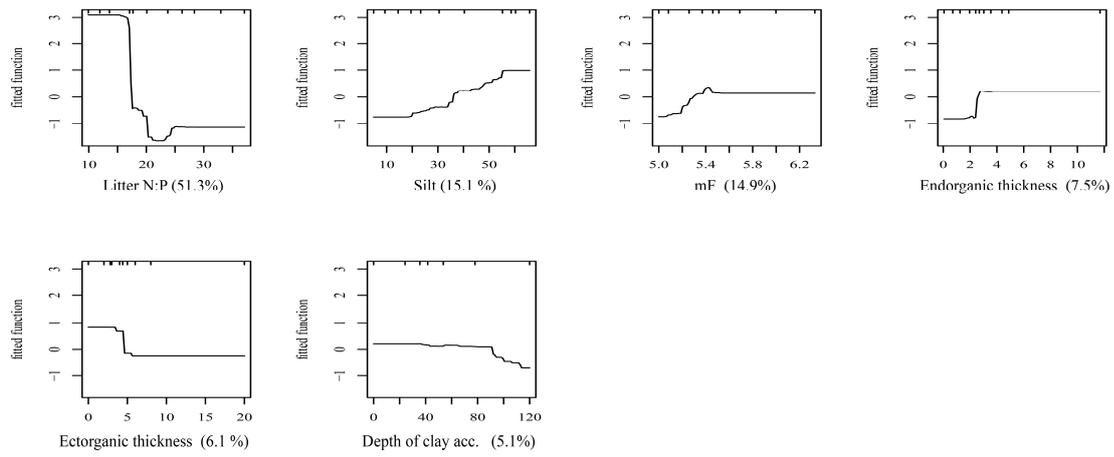
693 **Figure 1.** Location of the research plots in Flanders, Belgium. Research plots are labelled  
694 according to the dominant tree species (■ common beech, ● pedunculate oak, ▲ Scots pine).

695 Some locations contain more than one sample plot.



696

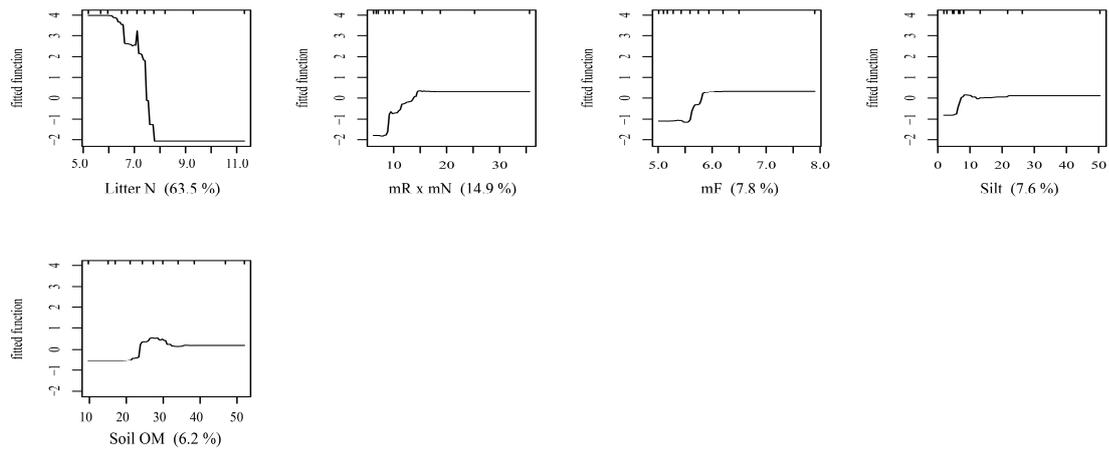
697 **Figure 2.** Partial dependence plots of the predictor variables in the ecosystem model of  
 698 pedunculate oak. Fitted functions are centred around the mean site index and plotted on a  
 699 common scale The relative contribution of each predictor is reported between brackets. Rug  
 700 plots at inside top of plots show distribution of sample plots along that variable, in deciles.



701

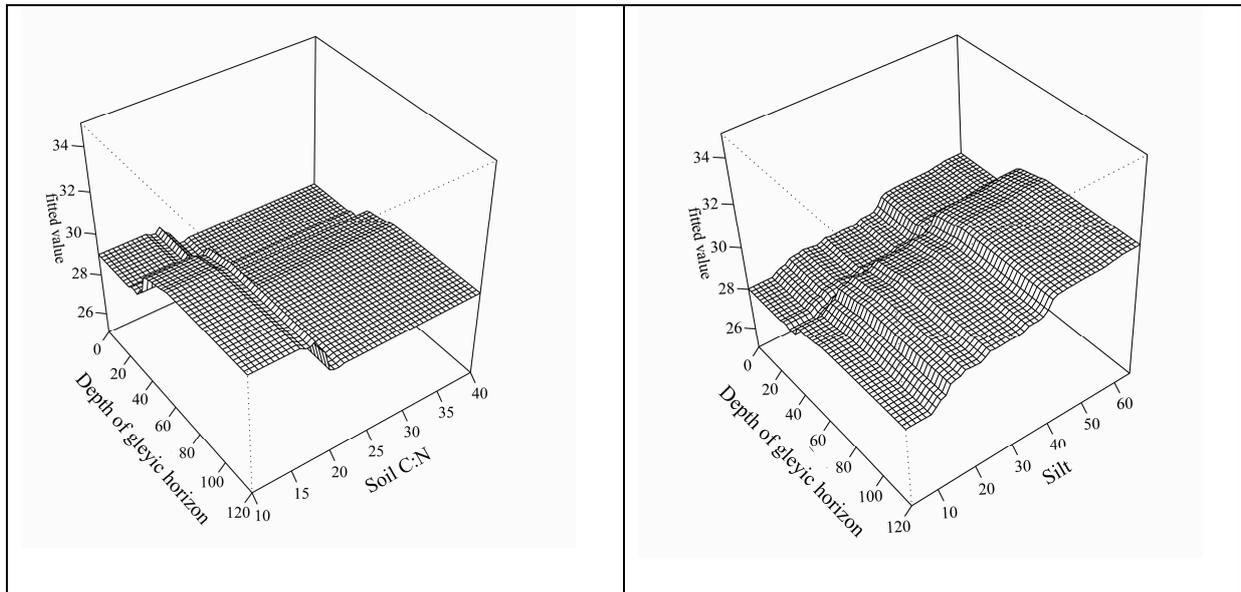
702 **Figure 3.** Partial dependence plots of the predictor variables in the ecosystem model of

703 common beech (cf. Fig. 2 for details).



704

705 **Figure 4.** Partial dependence plots of the predictor variables in the ecosystem model of Scots  
 706 pine (cf. Fig. 2 for details).



709 **Figure 5.** Three-dimensional partial dependence plots representing the strongest interactions  
 710 in the ecosystem site index model of pedunculate oak: (a) between the depth of the gleyic  
 711 horizon and the soil C:N ratio and (b) between the depth of the gleyic horizon and the silt  
 712 content.