UNIVERSITY OF
BATH

The final publication is available at Springer https://doi.org/10.1007/s13595-017-0685-3

University of Bath

1 **Modelling sapling distribution over time using a functional predictor in a generalized additive model**

2 Daniel MORENO-FERNÁNDEZ[1,2*], Nicole H. AUGUSTIN[3], Fernando MONTES[1], Isabel CAÑELLAS[1]

3 and Mariola SÁNCHEZ-GONZÁLEZ[1]

4

5 [1]INIA-CIFOR, Ctra. A Coruña km 7.5, 28040 Madrid, Spain

6 [2]MONTES (School of Forest Engineering and Natural Resources), Universidad Politécnica de Madrid,

7 28040 Madrid, Spain

8 [3]Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK

9 *Moreno-Fernández, D. (corresponding author, danielmorenofdez@gmail.com, +34 913476853).

10

11 **Key message**

12 The effect of adult trees on sapling density distribution during the regeneration fellings is determined in a

13 *Pinus sylvestris* L. Mediterranean forest using generalized additive models

14

15 **Abstract**

16 **Context:** Spatial pattern of adult trees determines the number of new individuals after regeneration fellings,

17 which modify the light and air temperature under tree canopy.

18 **Aims:** We proposed a novel spatio-temporal model with a functional predictor in a generalized additive

19 model framework to describe non linear relationships between the size of the adult trees and the number of

20 saplings of *Pinus sylvestris* and to determine if the spatial pattern of the number of saplings remained

21 constant or changed in time.

22 **Methods:** In 2001, two plots (0.5 ha) were set up in two phases of regeneration fellings under the group

23 shelterwood method. We mapped the trees and saplings and measured their diameter and height. The

24 inventories were repeated in 2006, 2010 and 2014.

25 **Results:** We found a negative association between the diameter of adult trees and number of saplings up to

26 7 – 8 m. Beyond these distances, the diameter of adult trees was not associated with the number of saplings.

27 Our results indicate that the spatial pattern of the number of saplings remained quite constant in time.

28 **Conclusions:** The generalised additive models are a flexible tool to determine the distance range of

29 inhibition of saplings by adult trees.

30

31 **Keywords:** edge effect; intra-specific competition; mountain forest; shade tolerance; Mediterranean areas

32

33 **1 Introduction**

34   Two main types of models can be used to explain or predict the renewal of a forest after

35 regeneration fellings, seed dispersion and germination: regeneration and recruitment models. The former is

36 related to the youngest individuals, seedlings, whereas the latter is related to larger stems, saplings, which

37 reach or exceed a nominal size limit determined by the researcher (Vanclay 1992; Eerikäinen et al. 2007;

38 Miina and Heinonen 2008). Since it is both difficult and expensive to obtain suitable data for modelling the

39 regeneration, recruitment is more often modelled than regeneration. Both processes are influenced by the

40 capacity of the soil to supply water and the amount of light that reaches the young seedlings. These are the

41 most important factors for success in the establishment of new individuals (Kozlowski 2002). Hence, the

42 summer drought in dry environments cause high mortality rates of seedlings over Mediterranean areas

43 (Castro et al. 2004; Pardos et al. 2007; McDowell et al. 2008), where the water is a limited resource in the

44 vegetative period.

45   Regeneration fellings can modify the effect of summer drought on seedlings and saplings by

46 setting different target densities or spacing between remaining trees and, thus, modify the shade and the air

47 temperature (Caccia and Ballaré 1998; Pardos et al. 2007). However, not all species can tolerate the same

48 amount of shade and the shade tolerance behavior may vary with site conditions (Kobe and Coates 1997;

49 Gómez-Aparicio et al. 2006). Additionally, the light requirement of plants varies with age. Indeed, the light

50 requirement increases faster with plant age in light-demanding species than in shade tolerant species

51 (Valladares and Niinemets 2008). This determines the density and spacing between remaining trees after

52 regeneration fellings. Therefore, it is necessary to have a clear understanding of the effects of the density

53 of residual trees on new individuals over the regeneration period in order to ensure the spatial continuity of

54 the forest stand after the regeneration fellings.

55   In addition to the density and spacing between remaining trees after regeneration fellings, several

56 features should be taken into account to model the number of saplings. The age of the stand should also be

57 considered, particularly where shifts in the spatial relationship between trees and offspring over the stages

58 of the forest renewal may occur (Wada and Ribbens 1997). Changes in spatial patterns of trees over time

59 are determined by regeneration mechanisms, substrate characteristics, moisture and light availability as

60 well as intra and inter specific competition (LeMay et al. 2009). Hence, the time perspective allows us to

2

61   distinguish between competition and the initial spatial pattern of individuals (Wolf 2005; Getzin et al.

62   2006), i.e., the initial distribution of seedlings as a consequence of the dispersion and germination of the

63   seeds can vary with the development of the seedlings and competition for resources.

64   The spatial relationships between adult trees and new cohorts have previously been evaluated

65   using different approaches. The bivariate Ripley's K and related functions have been used to determine if

66   stems of two mapped cohorts of trees show spatial positive, negative or random association (see Montes

67   and Cañellas 2007; Wild et al. 2014) by testing the spatial independence between the two cohorts. Ledo et

68   al. (2014) used inhomogeneous Poisson process spatial models. These models allow the spatial distribution

69   of new individuals to be defined in function of attributes of adult trees. Other authors used distance-

70   dependent influence indices (Contreras et al. 2011) and available light under the forest canopy or the global

71   site factor as explanatory variables in different models (Pardos et al. 2007; Moreno-Fernández et al. 2015a).

72   Distance-dependent influence indices determine, at a given point, the influence of the tree size (such as

73   diameter, height, or crown variables) and the distance between trees and the studied point whereas the

74   global site factor measures the amount of light at a given point by analyzing hemispherical photographs.

75   Influence indices and site factors can easily be entered in a time-dynamic model as additional variables

76   (Eerikäinen et al. 2007; Manso et al. 2013). However, the temporal modelling of Ripley's K and related

77   functions over time is complex. LeMay et al. (2009) investigated the evolution of these functions in the

78   regeneration of *Pseudotsuga menziesii* var *glauca* (Mirb.) Franco over time using a random coefficient

79   mixed model. Furthermore, specific distance dependent models implemented using packages such as

80   SILVA or SORTIE-ND have been used in forest development simulation studies which include the

81   regeneration establishment phase (Hanewinkel and Pretzsch 2000; Ameztegui et al. 2015). These software

82   packages are compounded of several submodels for the biological processes operating at individual tree

83   level. Comas (2008) and Redenbach and Särkkä (2012) adapted the growth-interaction model proposed by

84   Renshaw and Särkkä (2001) to develop a spatio-temporal regeneration model under two regeneration

85   methods using values taken from the literature to estimate the parameters. This approach generates marked

86   point configurations changing over time.

87   Generalized additive models (GAMs) may describe a complex relationship between the response

88   and the predictors. This is especially useful in research fields such as ecology, biology or forestry in which

89   simple models cannot capture the structure of the data and more complex models may be required (Faraway

90   2006). Whereas GAMs have been used in different areas of forest science such as wood quality (Jordan et

91   al. 2008), annual radial growth (Moreno-Fernández et al. 2014), mortality (Barbeito et al. 2012) or species

92   distribution (Franklin 1998), their use in regeneration or recruitment studies is relatively scarce (Rabasa et

93   al. 2013). Augustin et al. (2009) fit spatio-temporal models within a GAM framework to monitor forest

94   health data. However, these techniques have never been used to assess the dynamics of forest regeneration.

95   *Pinus sylvestris* L. is the most widely distributed pine species in the world (Mason and Alía 2000).

96   It can be found throughout Eurasia, stretching from Spain in the South-West to the far east of Russia

97   (Houston Durrant et al. 2016). This pine species is commonly considered to be a light-demanding species

98   in Central and northern Europe (Mátyás et al. 2003). However, it has a half-shade tolerant behavior in

99   southern locations like Spain (Montes and Cañellas 2007), partially due to the high temperatures and

100   drought conditions present during the summer months. Whereas during the early stages *P. sylvestris*

101   seedlings prefer moderate light conditions (Pardos et al. 2007; Barbeito et al. 2009), the later development

102   of saplings is inhibited by competition from the adult trees (Montes and Cañellas 2007). The variation on

103   shade tolerance and climate conditions across its distribution condition the regeneration method; while seed

104   tree and clear cutting are the main methods used in Central and Northern Europe, different alternatives of

105   the shelterwood method are commonly used in Southern Europe (Mason and Alía 2000). In general, 2 000

106   seedlings per hectare are considered to be a sufficient natural regeneration density (Rodríguez-García et al.

107   2010; Hyppönen et al. 2013).

108   In this work, we propose a methodology to describe non-linear relationships between the size of

109   the adult trees and number of saplings of *P. sylvestris* in Mediterranean mountains as a smooth function.

110   We carried it out analyzing data from repeated measurements of two large plots at two stages of the

111   regeneration period where all the stems were mapped. We modelled the spatio-temporal distribution of the

112   number of saplings using a functional predictor (see for example Wood 2011) in a GAM framework (Hastie

113   and Tibshirani 1989; Wood 2006). The functional predictor allowed us to weight the effect of every adult

114   tree on the number of saplings per quadrat based on the distance between adult trees and saplings. In

115   addition, the approach can deal with spatial correlation and a spatio-temporal trend, i.e. changes in the

116   spatial pattern of number of saplings during the development of the stand. In this regard, we fitted two

117   models with different spatio-temporal structures to determine if the spatial pattern of the number of saplings

118   remained constant or changed in time.

119

2 Material and methods

2.1 Study area and data

The study was carried out in a Scots pine forest (Pinar de Valsaín) located on the north facing slopes of the Central Range of Spain (40º 49′N, 4º 01′W). The elevation ranges from 1200 to 1600 m, the annual rainfall is about 1000 mm and the mean temperature is around 9.8 ºC. Regeneration is achieved using the group shelterwood method over a 40-year regeneration period. The regeneration fellings create small gaps (0.1-0.2 ha) for the establishment of the regeneration. As regeneration appears, subsequent harvests are carried out over the regeneration period to widen the gaps. The final fellings under the group shelterwood method take place at 120-140 years but some legacy trees are left for biodiversity conservation reasons at the end of the regeneration period.

In 2001, we set up a chronosequence of six plots (0.5 ha) covering all the rotation period (see Moreno-Fernández et al. 2015b for details) to study the dynamics and structure in Mediterranean forests of *P. sylvestris*. This chronosequence represents the management of *P. sylvestris* in the study area from the beginning to the end of the rotation period (Fig. 1) and it contains six plots. The plots were as homogeneous as possible in terms of altitude, exposure and site quality. Since we aim to address the influence of the adult trees on the saplings, we selected two plots at different stages of the regeneration period: at an intermediate stage of the regeneration period (100 x 50 m, Fig. 2, ca. 19-years-old) and at the end of the regeneration period (58.82 x 85 m, Fig. 3, ca. 32-years-old). Young individuals with different size were spread over the youngest studied plot. In this plot, regeneration fellings were done from 2010 to 2014 removing mainly trees located in the corners of the plot (Fig. 2). At the end of the regeneration period, the arrival of new individuals has almost been completed and the crown cover is getting closer. Additionally, some legacy trees (larger trees) appear in this plot (Fig. 3). Another plot, at the first stages of the regeneration period, was available. However, the arrival of new individuals has started as consequence of the natural dynamics but the number of saplings was still quite low (Fig. 1). Therefore, we did not include this plot in the analysis.

At the time the plots were set up, we carried out the first inventory in which all the stems higher than 1.30 m were labelled individually and their diameter at breast height (dbh) and height were measured. We numbered and classified the stems into: trees (dbh≥10 cm) and saplings (height≥1.30 m and dbh<10 cm). We distinguished two cohorts of trees: adult trees (dbh≥20 cm) and small trees (10≤dbh<20 cm). We mapped the position of every tree (adult and small trees) in each plot and additionally, we grouped the

149   saplings into a 2 x 2 m quadrat grid. The coordinates of the center of each quadrat were used to determine

150   the position of each quadrat. These measurements were repeated in 2006, 2010 and 2014.

151   In order to model the sapling distribution, we used the number of saplings per quadrat (Ns) in each

152   plot as response variable. We expected Ns to be highly related to the density of surrounding trees and

153   distance to the surrounding trees, as well as to the time since the beginning of the regeneration fellings.

154   However, the spatial dependence between the saplings and the two cohorts, adult and small trees, varies

155   over stand development (Montes and Cañellas 2007). Thus, we considered as predictors the dbh of the adult

156   trees (dbh≥20 cm), the distance in meters from adult tree to each sapling quadrat (considering all the adult

157   trees within a maximum radious of 30 m from each sapling quadrat; Montes and Cañellas 2007) and the

158   number of small trees ($Nsmall$; 10≤dbh<20 cm) surrounding every sampling quadrat within a radius of 10

159   m and the inventory year. The distribution of Ns, number of small and adult trees over inventories is shown

160   in Figs. 1, 2 and 3. We assume that at a given distance, larger dbh of the adult trees entails greater

161   competition between adult trees and saplings. Furthermore, we consider that this competition effect

162   between adult trees and Ns decreases with distance. Therefore, a model in which the coefficient of the dbh

163   depends on the distance between adult trees and the sapling quadrat would be very suitable. These

164   requirements can be taken into account using a linear functional predictor in a GAM. Thus, this approach

165   allowed us to weight the effect of every adult tree on the number of saplings per quadrat based on the

166   distance between adult trees and saplings.

167   **2.2 Edge effect correction**

168   The quadrats close to the boundaries of the plots are affected by the edge effect and this must be

169   corrected (Ledo et al. 2014). Thus, the number of adult and small trees which surround a quadrat within 30

170   m and 10 m, respectively, can be underestimated because some of them may be located outside the plot

171   (Goreaud and Pélissier 1999). Several authors (Lancaster and Downes 1998; Perry et al. 2006;

172   Pommerening and Stoyan 2006) have investigated the edge effect and have analyzed the suitability of

173   different edge-corrections for the calculation of the indices of spatial forest structure, Ripley´s K and related

174   second order functions.

175   In order to take account of the edge effect on the number of small and adult trees we used values

176   per unit area, i.e., density. For each quadrat, we estimated the area of the 10 m radius circle within the plot

177   ($AreaIn10$ in m^2). Therefore, $AreaIn10$ changes with the distance between the quadrat and plot border, i.e,

178     $AreaIn10$ is smaller in the quadrats closer to the plot border. Then, we obtained the density of small trees

179 as *Nsmall/AreaIn10*. We corrected the edge effect on adult trees by using the dbh density as *dbh/AreaIn30*.

180 *AreaIn30* is the area (in m^2) of the 30 m radius circle within the plot. Thus, we assume that the surrounding

181 shelter trees outside the plot would be of similar density than within the area. *AreaIn30* is the area (in m^2)

182 of the 30 m radius circle within the plot.

183 **2.3 Statistical analysis**

184 For each of the two plots we modelled the expected number of saplings $E(Ns_{ij}) = \mu_{ij}$ in quadrat i

185 and j-th inventory ($j=1,\ldots,4$) using the following GAM:

186
$$\log(\mu_{ij}) = \alpha + \beta \cdot \frac{Nsmall_{ij}}{AreaIn10_i} + \sum_{n=1}^{N} \left(f_1(Dist_{in}) \cdot \frac{dbh_{jn}}{AreaIn30_i} \right) + f_2(X_i, Y_i) + Time_j \qquad (1)$$

187 with Ns_{ij} following a negative binomial distribution. This distribution is suitable for overdispersed counts

188 such as those we are dealing with here. The variance function is $V(\mu_{ij}) = \mu_{ij} + \mu_{ij}^2/\theta$, involving the extra

189 parameter θ to be estimated. The greater θ is, the more similar the negative binomial distribution is to the

190 Poisson distribution. Small values for θ indicate aggregation. The parameter α is the intercept of the model,

191 β is the unknown but estimable parameter of the number of small trees. $Dist_{in}$ is a matrix which contains

192 the distances (in m) from the adult tree ($n=1, \ldots, N$) to the i-th quadrat, whereas dbh_{jn} is the matrix of the

193 dbh of the adult tree ($n=1,\ldots, N$). When the distance of the n-th adult tree to the i-th quadrat was greater than

194 30 m, the dbh was set to 0. $\sum_{n=1}^{N} \left(f_1(Dist_{in}) \cdot dbh_{jn}/AreaIn30_i \right)$ is functional predictor where $f_1(Dist_{in})$ is the

195 smooth coefficient of dbh_{jn}. The function $f_2(X_i, Y_i)$ is a spatial smooth term to account the spatial trend and

196 spatial correlation of the number of saplings. Any spatial trend will caused by other unmeasured

197 environmental variables and hence the spatial smooth term is a proxy for other unmeasured environmental

198 effects. X_i and Y_i are the coordinates of the i-th quadrat and $Time_j$ is the temporal factor referred to the j-th

199 inventory.

200 Model 1 above separates the effects of space and time, i.e. the two effects are additive. The model

201 can be made more flexible by allowing the spatial smooth to change in time, i.e., this model contains a

202 spatial smooth per j-th inventory:

203
$$\log(\mu_{ij}) = \alpha + \beta \cdot \frac{Nsmall_{ij}}{AreaIn10_i} + \sum_{n=1}^{N} \left(f_1(Dist_{in}) \cdot \frac{dbh_{jn}}{AreaIn30_i} \right) + f_2^j(X_i, Y_i) \qquad (2)$$

204 We used Akaike´s Information Criterion (AIC) to select the variables by using backward stepwise

205 procedure and choosing the best spatio-temporal structure.

206   Functions f_1 and f_2 were represented using thin plate regression splines (Wood 2003). Thin plate

207   regression splines keep the basis and the penalty of the full thin plate splines (Duchon 1977) but the basis

208   is truncated to obtain low rank smoothers. This avoids the problems of the knot placement of the regression

209   splines and reduces the computational requirements of the smoothing splines (Wood 2003). Penalized

210   regression smoothers such as thin plate regression splines are computationally efficient because their basis

211   have a relatively modest size, k. In practice, k determines the upper limit on the degrees of freedom

212   associated with the smooth function, hence k must be chosen when fitting models. However, the actual

213   effective degrees of freedom of the smooth function are controlled by the degree of penalization selected

214   during fitting. The degree of penalization determines how smooth the function is. So, k should be chosen

215   to be large enough to represent the underlying process reasonably well, but small enough to ensure

216   reasonable computational efficiency. The exact choice of k is not critical (Wood 2006).

217   The spatial smooth $f_2(X_i, Y_i)$ is confounded with the functional predictor term,

218        $\sum_{n=1}^{N}\left(f_1(Dist_{in})\cdot dbh_{jn}\big/AreaIn30_i\right)$, since both terms describe, in some way, the spatial pattern in the

219   response. To avoid further confounding, we decided to include the effect of small trees in a linear form

220   rather than a functional predictor. We used $k=10$ for f_1 since it was enough to represent the variation of the

221   coefficient of dbh as the actual effective degrees of freedom for f_1 was between 3 and 4 - well below 10.

222   As we are ultimately interested in estimating the f_1 of the functional predictor and f_2 is entered to eliminate

223   the spatial correlation, we selected the smallest basis dimension (k) in f_2 that eliminated the spatial

224   correlation. For the different values of k in f_2, we checked whether the spatial correlation had been

225   eliminated in the model by plotting semivariograms of the model residuals per inventory with envelopes

226   from 99 permutations under the assumption of no spatial correlation (see Augustin et al. 2009 for a

227   description).

228   The statistical analyses were carried out in R 3.3.3. (R Core Team 2017) using the "gam" function

229   of the package "mgcv" (Wood 2011) for fitting the models where we used the restricted maximum

230   likelihood option. This means that the smoothness parameters are estimated using restricted maximum

231   likelihood estimation and a penalized iterative re-weighted least squares algorithm is used to find all other

232   parameters, i.e. the coefficients of basis functions and coefficients of linear terms. See Wood (2011) for the

233   theory and Augustin et al. (2015) for a functional predictor example. For model checking we used the

234   functions "variog" and "variog.mc.env" of the package "geoR" (Ribeiro and Diggle 2016) for estimating

235   the semivariograms and the envelopes.

237 **3 Results**

238 **3.1 Intermediate stages of the regeneration period**

239   The total number of saplings was inversely related to the time whereas the mean size (dbh and

240 height) of the saplings increased with time (Table 1). Saplings were spread around the plot except in the

241 center and the bottom left corner (Fig. 2). Unlike the saplings, we found that the number of trees, both small

242 and adult trees, increased with time, the mean dbh and size of this stratum decreasing with time due to

243 ingrowth of individuals from the previous class (Table 1). During the study period, we found a great

244 increase in small trees, especially in the lower right part of the plot (Fig. 2).

245   Both model 1 and model 2 explained a similar amount of deviance, almost 41%. However, the

246 AIC of model 1 was lower than that of model 2 (Table 2). Therefore, we selected model 1, the more

247 parsimonious model, with additive effects of space and time. This entails that the spatial distribution of the

248 saplings remained constant over the time. The spatial smooth function (f_2) and the temporal factor (*Time*)

249 improved the model in terms of AIC (Table 3). Fig. 4 shows the estimated spatial smooth function f_2 on the

250 scale of the linear predictor. The estimate of the aggregation parameter θ of the negative binomial

251 distribution is 1.5 and 1.4 in model 1 and 2, respectively. Our results show that we have chosen k large

252 enough for both functions f_1 and f_2, as we see that the effective degrees of freedom given in Table 2 are

253 below k-1; the same applies to results for the other plot. The coefficients of *dbh* were robust to changes in

254 k. This also applies to results of the other plot. Fig. 5 shows that the spatial correlation was eliminated.

255   Removing the term relating to the density of small trees increased the AIC (Table 3). The β of the

256 density of small trees was negative (β =-0.0004) pointing towards competition between small trees and

257 saplings. Furthermore, Fig. 6 shows the smooth coefficient of *dbh* (f_1) of adult trees over *Dist*. More saplings

258 are expected to be found when the product of the smooth coefficient and the *dbh* is large, that is, the model

259 predicts the greatest number of saplings for the largest trees located at the distances to which f_1 is highest.

260 f_1 varied smoothly across the distances with significantly negative values from 0 m up to 7 m. From 7 m, f_1

261 is not statistically different from zero as the 95 % confidence intervals contained zero. This suggests

262 competition between adult trees and saplings at shorter distances (<7 m) and no relationship at larger

263 distances between these two cohorts. From 13 to 20 m, the mean value of f_1 turned positive and significant

264 reaching the largest values of the smooth function. Beyond 20 m, the smooth function f_1 started decreasing

265 and it was not statistically different from zero.

3.2 End of the regeneration period

The dynamics of the saplings and the trees followed the same trends as at intermediate stages of the regeneration period: the number of saplings decreased and their mean size increased with time. The number of trees decreased but the mean dbh and height increased over the four inventories because of fellings. However, in this plot there were less saplings and their mean size was larger than in the youngest studied plot. Additionally, there were more trees overall at the end of the regeneration period than in the previous stages. Nevertheless, Table 1 shows that the number of small trees reduced with time whereas the number of adult trees increased.

As in the youngest studied plot, the model with the additive spatio-temporal structure (model 1), which assumes a constant spatial distribution of the saplings over the studied period, showed a lower AIC than model 2 (Table 2). We also found a significant effect of the spatio-temporal terms (f_2 and *Time*) in terms of AIC (Table 3). The map of the contour lines (Fig. 7) represents well the spatial distribution of the saplings during the last stages of the regeneration period (Fig. 3). The semivariograms showed that the spatial structure eliminated the spatial correlation (Fig. 8). The estimate of the aggregation parameter θ of the negative binomial distribution is smaller than in the other plot, it is around 0.8 (Table 2). This indicates that saplings were more aggregated at the end of the regeneration period than in the previous stage, which is also confirmed by the visual inspection of the spatial distribution of saplings (Fig. 2 and 3) showing a more homogenous spread of saplings in earlier stages of the regeneration process.

In this plot, the β of density of small trees did not reduce the AIC whereas the rest of the terms reduced the AIC significantly (Table 3). Table 2 shows the effective degrees of freedom of the basis functions. The coefficient of the *dbh* of adult trees (f_1) took significant negative values from 0 to 8 m (Fig. 6). From 8 m, the 95 % confidence intervals contained the zero, and therefore we can state that the coefficient is not statistically different from 0. This suggests competition between saplings and adult trees at very small distances and no effect beyond 8 m.

4 Discussion

We fitted a GAM with a functional predictor in the model to describe the influence of size of trees on the number of saplings by distance in two stages of the regeneration period. We have confirmed that the functional predictor is useful to achieve this aim. In GAMs explanatory variables may enter the model in many different forms: as variables with linear effects, smooth terms, tensor products of several variables,

296 with varying coefficients or as functional predictors. Additionally, alternative response distribution families

297 and link functions can be selected (see for example Wood 2006). Therefore, all this makes the approach

298 employed suitable to be used in other fields of forestry or ecology in which the response variable depends

299 on the size and distance of the neighbors. For instance, this approach could be useful to fit growth or

300 mortality models instead of using competition indices in parametric models (Contreras et al. 2011).

301   In this work, we have studied the last stages of the renewal of a forest after regeneration fellings.

302 Other authors have modelled the whole renewal of the forest using multistage models. For instance, Manso

303 et al. (2014) proposed a multistage model based on partial studies or submodels in order to predict the

304 regeneration occurrence of *Pinus pinea* L. in space and time. They considered different stages such as seed

305 dispersal, seed germination, post-dispersal predation and seedling survival. Multistage models provide

306 deeper ecological understanding than ours but the implementation is harder and requires stronger ecological

307 hypotheses. However, our approach shows great flexibility and might be used to determine the effects of a

308 limited number of factors on sapling distribution without making any assumptions about other factors

309 involved on dispersion and survival processes.

310   As mentioned above, this methodology allows different types of variables to be included in the

311 model. In this work, we included the density of small trees as a linear term and the spatio-temporal structure.

312 It might be useful to use variables driving the regeneration as predictors in the model, like shrub cover, soil

313 characteristics, cover and depth of litter or grass in each quadrat. However, gathering this data on large

314 plots requires a great effort and the influence of these variables on the seedlings of *P. sylvestris* has already

315 been studied at smaller scales (González-Martínez and Bravo 2001; Pardos et al. 2007; Barbeito et al. 2011;

316 Moreno-Fernández et al. 2015a). On the other hand, new individuals of *P. sylvestris* are expected to be

317 more affected by soil moisture than by other microsite characteristics (Barbeito et al. 2009; Moreno-

318 Fernández et al. 2015a). However, because youngest individuals – seedlings – are less resistant to drought

319 than older – saplings (Maseda and Fernández 2006; Rodríguez-García et al. 2011; Manso et al. 2014), it

320 seems it is more necessary to include environmental variables in models dealing with seedlings rather than

321 in those dealing with older individuals – saplings. Moreover, it is likely the distance to adult trees is

322 confounded with other local factors. Any residual spatial trend in a model without a spatial smooth term is

323 caused by missing (unmeasured) environmental variables. Furthermore, the residual spatial trend could be

324 due to the seedling spatial structure that would result from past dispersal events from adjacent mother trees.

325 We have included the spatial smooth term as a proxy for effects of unmeasured environmental variables

11

and for the spatial pattern of the new individuals during previous stages of the forest renewal. We have investigated goodness-of-fit thoroughly, and found that we did not have any spatial trend in residuals or residual spatial correlation. This means that the models fit well and there was no model mis-specification. Although we have only results from two plots, it is striking that the estimated functions f_1 (of the effect of dbh) shown in Fig. 6 are very similar.

Our approach allows to test whether the spatial pattern remained constant over time by comparing model 1 which assumes a constant spatial pattern with model 2 which allows for a spatial pattern changing in time in the model selection. In our case, model 1 was selected suggesting that the spatial pattern of the saplings remained constant over time. If model 2 had been selected, the spatial pattern of saplings would have changed over the time. Due to the gradual low intensity fellings regime, which avoids damaging the established saplings clumps, these clumps persist and the spatial structure of the saplings remains fairly constant in each plot during the 15-year measurement period. On the other hand, if we were dealing with faster-growth species the saplings could move to the next cohort faster and then change the spatial pattern. Our results are consistent with LeMay et al. (2009) who reported that the spatial pattern of the new individuals of *P. menziesii* did not change very much over time.

Although we only analyzed data from two large plots (0.5 ha) we re-measured the plots four times, leading to four observations per plot. Large-sized plots with few sampling over time are common in regeneration studies describing spatial processes. These kind of plots have been used in tropical (Ledo et al. 2015), temperate (McDonald et al. 2003) and Mediterranean forests (Montes and Cañellas 2007; Ledo et al. 2014). Additionally, we modelled the number of saplings per 4 m^2 quadrat, i.e., we used 5 000 quadrats covering different competition conditions to fit every model. Moreover, the models presented in this work were fitted for explanatory purposes rather than predictive purposes. If we had aimed to fit a predictive model, we would have needed more temporal measurements to cover all the regeneration period.

The underlying process studied here is the competition between trees and saplings. Our findings are in concordance with other studies: the saplings of *P. sylvestris* require high light conditions for successful development (Montes and Cañellas 2007). In Mediterranean areas, *P. sylvestris* seedlings require microsites with moderate light conditions (Pardos et al. 2007). These microsites ensure higher soil moisture than in open canopies but conserve enough level of sun radiation. In this regard, Castro et al. (2005) analyzed the growth of *P. sylvestris* seedlings in southern Spain under different light and water conditions concluding that the effects of water addition on seedlings growth are more evident in lightly microsites.

356 Moreover, once the seedlings have stepped into saplings, the maintenance costs increase with size (Falster

357 and Westoby 2003) and higher minimum light levels are required for survival (Williams et al. 1999).

358 Additionally, their roots can reach deeper soil layers with more water availability (Ritchie 1981).

359 Considering this, it seems it is necessary to reduce the canopy to favor the development of the saplings after

360 seedling establishment under moderate light conditions in Mediterranean areas. However, the shade

361 tolerance of *P. sylvestris* differ among regions. In northern locations where the summer drought is not a

362 limiting factor for seedling development, natural regeneration takes place in open canopies by using the

363 seed tree method (Hyppönen et al. 2013). In these latitudes, the negative spatial association between *P.*

364 *sylvestris* adult trees and saplings may be even more pronounced than in our study.

365 The establishment of the new stand has been achieved successfully at the end of the regeneration

366 period, the number of saplings decreased and the arrival of new individuals is no longer expected. Hence,

367 the mean dbh of the saplings is getting close to 10 cm, the lower limit for small trees. In this plot, the

368 number of small trees decreased over time due to the mortality as well as the growth and consequent

369 reclassification of trees as adult trees. Most of the trees in this plot were not mother trees of the saplings

370 but rather new cohorts of trees established at the first and intermediate stages of the regeneration period,

371 such as those in our youngest studied plot. Therefore, the spacing between saplings and adult trees is a

372 consequence of the competition between trees of different sizes.

373

374 **5 Conclusions**

375 We show that functional predictor in GAMs is a useful tool for modelling these kind of data as

376 they allow to model nonlinear and linear relationships. In addition they allow to take account of the spatio-

377 temporal structure of the data by inclusion of spatial and spatio-temporal smooth predictors. The

378 methodology proposed has not been employed in forestry or ecology and can be broadly used in

379 regeneration studies or in other fields of forestry or ecology dealing with spatio-temporal data. Therefore

380 this methodology is potentially applicable in future ecological studies because of its flexibility.

381 Additionally, this model can be used as a first step for a predictive model when more temporal data is

382 available. We found that once the seedlings have become established, the density of the adult trees must be

383 reduced heavily to allow the saplings to grow under high light conditions. In Mediterranean stands of *P.*

384 *sylvestris*, the radius of the gaps created during the regeneration fellings under the group shelterwood should

385   be always larger than 7-8 m in order to minimize the competition between adult trees and saplings; whereas

386   if the radius is between 13 – 20 m the number of saplings will be maximized

387

388   **References**

389   Ameztegui A, Coll L, Messier C (2015) Modelling the effect of climate-induced changes in recruitment

390   and juvenile growth on mixed-forest dynamics: The case of montane–subalpine Pyrenean ecotones.

391   Ecol Modell 313:84–93. doi: 10.1016/j.ecolmodel.2015.06.029

392   Augustin NH, Mattocks C, Faraway JJ, et al (2015) Modelling a response as a function of high-frequency

393   count data: The association between physical activity and fat mass. Stat Methods Med Res 1–20. doi:

394   10.1177/0962280215595832

395   Augustin NH, Musio M, von Wilpert K, et al (2009) Modeling spatiotemporal forest health monitoring

396   data. J Am Stat Assoc 104:899–911. doi: 10.1198/jasa.2009.ap07058

397   Barbeito I, Dawes MA, Rixen C, et al (2012) Factors driving mortality and growth at treeline: a 30-year

398   experiment of 92 000 conifers. Ecology 93:389–401.

399   Barbeito I, Fortin M-J, Montes F, Cañellas I (2009) Response of pine natural regeneration to small-scale

400   spatial variation in a managed Mediterranean mountain forest. Appl Veg Sci 12:488–503. doi:

401   10.1111/j.1654-109X.2009.01043.x

402   Barbeito I, LeMay V, Calama R, Cañellas I (2011) Regeneration of Mediterranean *Pinus sylvestris* under

403   two alternative shelterwood systems within a multiscale framework. Can J For Res 41:341–351. doi:

404   10.1139/X10-214

405   Caccia FD, Ballaré CL (1998) Effects of tree cover, understory vegetation, and litter on regeneration of

406   Douglas-fir (*Pseudotsuga menziesii*) in southwestern Argentina. Can J For Res 28:683–692. doi:

407   10.1139/cjfr-28-5-683

408   Castro J, Zamora R, Hodar JA, Gomez JM (2004) Seedling establishment of a boreal tree species (*Pinus*

409   *sylvestris*) at its southernmost distribution limit: consequences of being in a marginal Mediterranean

410   habitat. J Ecol 92:266–277. doi: 10.1111/j.0022-0477.2004.00870.x

411   Castro J, Zamora R, Hódar JA, Gómez JM (2005) Alleviation of summer drought boosts establishment

412   success of *Pinus sylvestris* in a Mediterranean mountain: An experimental approach. Plant Ecol

413   181:191–202. doi: 10.1007/s11258-005-6626-5

414   Comas C (2008) Modelling forest regeneration strategies through the development of a spatio-temporal

415   growth interaction model. Stoch Environ Res Risk Assess 23:1089–1102. doi: 10.1007/s00477-008-

416   0282-y

417 Contreras MA, Affleck D, Chung W (2011) Evaluating tree competition indices as predictors of basal area

418   increment in western Montana forests. For Ecol Manage 262:1939–1949. doi:

419   10.1016/j.foreco.2011.08.031

420 Duchon J (1977) Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In: Schempp W,

421   Zeller K (eds) Constructive Theory of Functions of Several Variables: Proceedings of a Conference

422   Held at Oberwolfach April 25 -- May 1, 1976. Springer Berlin Heidelberg, Berlin, Heidelberg, pp

423   85–100

424 Eerikäinen K, Miina J, Valkonen S (2007) Models for the regeneration establishment and the development

425   of established seedlings in uneven-aged, Norway spruce dominated forest stands of southern Finland.

426   For Ecol Manage 242:444–461. doi: 10.1016/j.foreco.2007.01.078

427 Falster DS, Westoby M (2003) Plant height and evolutionary games. TRENDS Ecol Evol 18:337–342. doi:

428   10.1016/S0169-5347(03)00061-2

429 Faraway JJ (2006) Extending the linear model with R: generalized linear, mixed effects and nonparametric

430   regression models.

431 Franklin J (1998) Predicting the distribution of shrub species in southern California from climate and

432   terrain-derived variables. J Veg Sci 9:733–748. doi: 10.2307/3237291

433 Getzin S, Dean C, He F, et al (2006) Spatial patterns and competition of tree species in a Douglas fir

434   chronosequence on Vancouver Island. Ecography (Cop) 29:671–682. doi: 10.1111/j.2006.0906-

435   7590.04675.x

436 Gómez-Aparicio L, Valladares F, Zamora R (2006) Differential light responses of Mediterranean tree

437   saplings: linking ecophysiology with regeneration niche in four co-occurring species. Tree Physiol

438   26:947–958. doi: 10.1093/treephys/26.7.947

439 González-Martínez SC, Bravo F (2001) Density and population structure of the natural regeneration of

440   Scots pine (*Pinus sylvestris* L.) in the High Ebro Basin (Northern Spain). Ann For Sci 58:277–288.

441   doi: 10.1051/forest:2001126

442 Goreaud F, Pélissier R (1999) On explicit formulas of edge effect correction for Ripley's K -function. J

443   Veg Sci 10:433–438. doi: 10.2307/3237072

444 Hanewinkel M, Pretzsch H (2000) Modelling the conversion from even-aged to uneven-aged stands of

445   Norway spruce (*Picea abies* L. Karst.) with a distance-dependent growth simulator. For Ecol Manage

446   134:55–70. doi: 10.1016/S0378-1127(99)00245-5

447 Hastie T, Tibshirani R (1989) Generalized additive models. Stat Sci 10:297–318. doi:

448   10.1214/ss/1177013604

449 Houston Durrant T, de Rigo D, Caudullo G (2016) *Pinus sylvestris* in Europe: distribution, habitat, usage

450   and threats. In: San-Miguel-Ayanz J, de Rigo D, Caudullo G, et al (eds) European Atlas of Forest

451   Tree Species. Publication Office of the European Union, Luxembourg,

452 Hyppönen M, Hallikainen V, Niemelä J, Rautio P (2013) The contradictory role of understory vegetation

453   on the success of Scots pine regeneration. Silva Fenn 47:1–19.

454 Jordan L, Clark A, Schimleck LR, et al (2008) Regional variation in wood specific gravity of planted

455   loblolly pine in the United States. Can J For Res 38:698–710. doi: 10.1139/X07-158

456 Kobe RK, Coates KD (1997) Models of sapling mortality as a function of growth to characterize

457   interspecific variation in shade tolerance of eight tree species of northwestern British Columbia. Can

458   J For Res 27:227–236. doi: 10.1139/x96-182

459 Kozlowski TT (2002) Physiological ecology of natural regeneration of harvested and disturbed forest

460   stands: implications for forest management. For Ecol Manage 158:195–221. doi: 10.1016/S0378-

461   1127(00)00712-X

462 Lancaster J, Downes BJ (1998) Spatial point pattern analysis of available and exploited resources.

463   Ecography (Cop) 27:94–102.

464 Ledo A, Cañellas I, Barbeito I, et al (2014) Species coexistence in a mixed Mediterranean pine forest:

465   Spatio-temporal variability in trade-offs between facilitation and competition. For Ecol Manage

466   322:89–97. doi: 10.1016/j.foreco.2014.02.038

467 Ledo A, Cayuela L, Manso R, Condés S (2015) Recruitment patterns and potential mechanisms of

468   community assembly in an Andean Cloud Forest. J Veg Sci 26:876–888. doi: 10.1111/jvs.12287

469 LeMay V, Pommerening A, Marshall P (2009) Spatio-temporal structure of multi-storied, multi-aged

470   interior Douglas fir (*Pseudotsuga menziesii* var. *glauca*) stands. J Ecol 97:1062–1074. doi:

471   10.1111/j.1365-2745.2009.01542.x

472 Manso R, Fortin M, Calama R, Pardos M (2013) Modelling seed germination in forest tree species through

473   survival analysis: The *Pinus pinea* L. case study. For Ecol Manage 289:515–524. doi:

474   10.1016/j.foreco.2012.10.028

475   Manso R, Pukkala T, Pardos M, et al (2014) Modelling *Pinus pinea* forest management to attain natural

476   regeneration under present and future climatic scenarios. Can J For Res 44:250–262. doi:

477   10.1139/cjfr-2013-0179

478   Maseda PH, Fernández RJ (2006) Stay wet or else: three ways in which plants can adjust hydraulically to

479   their environment. J Exp Bot 57:3963–77. doi: 10.1093/jxb/erl127

480   Mason W, Alía R (2000) Current and future status of Scots pine (*Pinus sylvestris* L.) forests in Europe.

481   Investig Agrar Sist y Recur For Fuera de s:317–335.

482   Mátyás C, Ackzell L, Samuel C (2003) EUFORGEN Technical guidelines for genetic conservation and use

483   for Scots pine (*Pinus sylvestris*).

484   McDonald RI, Peet RK, Urban DL (2003) Spatial pattern of *Quercus* regeneration limitation and *Acer*

485   *rubrum* invasion in a Piedmont forest. J Veg Sci 14:441–450. doi: 10.1658/1100-

486   9233(2003)014[0441:SPOQRL]2.0.CO;2

487   McDowell N, Pockman WT, Allen CD, et al (2008) Mechanisms of plant survival and mortality during

488   drought: Why do some plants survive while others succumb to drought ? New Phytol 178:719–739.

489   doi: 10.1111/j.1469-8137.2008.02436.x

490   Miina J, Heinonen J (2008) Stochastic simulation of forest regeneration establishment using a multilevel

491   multivariate model. For Sci 54:206–219. doi: http://dx.doi.org/10.1016/j.csda.2012.08.006

492   Montes F, Cañellas I (2007) The spatial relationship between post-crop remaining trees and the

493   establishment of saplings in *Pinus sylvestris* stands in Spain. Appl Veg Sci 10:151. doi:

494   10.1658/1402-2001(2007)10[151:TSRBPR]2.0.CO;2

495   Moreno-Fernández D, Cañellas I, Barbeito I, et al (2015a) Alternative approaches to assessing the natural

496   regeneration of Scots pine in a Mediterranean forest. Ann For Sci 72:569–583. doi: 10.1007/s13595-

497   015-0479-4

498   Moreno-Fernández D, Díaz-Pinés E, Barbeito I, et al (2015b) Temporal carbon dynamics over the rotation

499   period of two alternative management systems in Mediterranean mountain Scots pine forests. For

500   Ecol Manage 348:186–195. doi: 10.1016/j.foreco.2015.03.043

501   Moreno-Fernández D, Sánchez-González M, Álvarez-González JG, et al (2014) Response to the interaction

502   of thinning and pruning of pine species in Mediterranean mountains. Eur J For Res 133:833–843. doi:

503   10.1007/s10342-014-0800-z

504   Pardos M, Montes F, Aranda I, Cañellas I (2007) Influence of environmental conditions on germinant

505   survival and diversity of Scots pine (*Pinus sylvestris* L.) in central Spain. Eur J For Res 126:37–47.

506   doi: 10.1007/s10342-005-0090-6

507   Perry GLW, Miller BP, Enright NJ (2006) A comparison of methods for the statistical analysis of spatial

508   point patterns in plant ecology. Plant Ecol 187:59–82. doi: 10.1007/s11258-006-9133-4

509   Pommerening A, Stoyan D (2006) Edge-correction needs in estimating indices of spatial forest structure.

510   Can J For Res 36:1723–1739. doi: 10.1139/X06-060

511   Rabasa SG, Granda E, Benavides R, et al (2013) Disparity in elevational shifts of European trees in response

512   to recent climate warming. Glob Chang Biol 19:2490–2499. doi: 10.1111/gcb.12220

513   R Core Team (2017) R: A language and environment for statistical computing.

514   Redenbach C, Särkkä A (2012) Parameter estimation for growth interaction processes using spatio-

515   temporal information. Comput Stat Data Anal 57:672–683.

516   Renshaw E, Särkkä A (2001) Gibbs point processes for studying the development of spatial-temporal

517   stochastic processes. Comput Stat Data Anal 36:85–105. doi: 10.1016/S0167-9473(00)00028-1

518   Ribeiro PJ, Diggle PJ (2016) geoR: Analysis of Geostatistical Data. www.leg.ufpr.br/geoR.

519   Ritchie J (1981) Soil water availability. Plant Soil 58:357–338.

520   Rodríguez-García E, Bravo F, Spies TA (2011) Effects of overstorey canopy, plant-plant interactions and

521   soil properties on Mediterranean maritime pine seedling dynamics. For Ecol Manage 262:244–251.

522   doi: 10.1016/j.foreco.2011.03.029

523   Rodríguez-García E, Juez L, Bravo F (2010) Environmental influences on post-harvest natural regeneration

524   of *Pinus pinaster* Ait. in Mediterranean forest stands submitted to the seed-tree selection method. Eur

525   J For Res 129:1119–1128. doi: 10.1007/s10342-010-0399-7

526   Valladares F, Niinemets Ü (2008) Shade tolerance, a key plant feature of complex nature and consequences.

527   Annu Rev Ecol Evol Syst 39:237–257. doi: 10.1146/annurev.ecolsys.39.110707.173506

528   Vanclay JK (1992) Modelling regeneration and recruitment in a tropical rain forest. Can J For Res 22:1235–

529   1248. doi: 10.1139/x92-165

530   Wada N, Ribbens E (1997) Japanese maple (*Acer palmatum* var. *matsumurae*, Aceraceae) recuitment

531   patterns: seeds,seedlings and salings in relation to conspecific adult neighbors. Oikos 84:1294–1300.

532   Wild J, Kopecký M, Svoboda M, et al (2014) Spatial patterns with memory: tree regeneration after stand-

533   replacing disturbance in Picea abies mountain forests. J Veg Sci 25:1327–1340. doi:

534   10.1111/jvs.12189

535 Williams H, Messier C, Kneeshaw DD (1999) Effects of light availability and sapling size on the growth

536   and crown morphology of understory Douglas-fir and lodgepole pine. Can J For Res 29:222–231.

537   doi: 10.1139/x98-189

538 Wolf A (2005) Fifty year record of change in tree spatial patterns within a mixed deciduous forest. For Ecol

539   Manage 215:212–223. doi: 10.1016/j.foreco.2005.05.021

540 Wood SN (2003) Thin-plate regression splines. J R Stat Soc 65:95–114. doi: 10.1111/1467-9868.00374

541 Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of

542   semiparametric generalized linear models. J R Stat Soc Ser B Stat Methodol 73:3–36. doi:

543   10.1111/j.1467-9868.2010.00749.x

544 Wood SN (2006) Generalized Additive Models: an Introduction with R. CRC Press, United States of

545   America

546

547 **Tables**

548 **Table 1** Summary of the mean forest features in each plot during the four inventories. Trees (dbh>= 10cm),

549 adult trees (dbh≥20 cm), small trees (10≤dbh<20 cm), saplings (dbh<10cm and height ≥1.30m). Standard

550 deviation is within brackets

Feature	2001	2006	2010	2014
Intermediate stage of the regeneration period				
Number of saplings	1 861	1 625	1 498	1 347
Mean diameter of saplings (cm)	3.68 (2.31)	4.42 (2.29)	4.61 (2.38)	4.77 (2.36)
Mean height of saplings (m)	4.20 (1.78)	4.86 (2.03)	5.38 (2.21)	5.76 (2.36)
Number of adult trees	80	62	75	102
Number of small trees	152	250	351	399
Number of trees	232	312	426	501
Mean diameter of trees (cm)	23.07 (14.75)	17.89 (11.30)	16.68 (8.89)	16.96 (8.72)
Mean height of trees (m)	15.00 (6.90)	12.46 (5.32)	12.31 (4.23)	13.26 (4.16)
End of the regeneration period				
Number of saplings	558	364	208	117
Mean diameter of saplings (cm)	5.54 (2.66)	6.43 (2.32)	6.55 (2.24)	6.97 (2.19)
Mean height of saplings	7.03 (3.27)	7.82 (3.32)	8.61 (3.44)	8.70 (3.55)

Number of adult trees	174	233	283	317
Number of small trees	568	492	434	366
Number of trees	742	725	717	683
Mean diameter of trees (cm)	16.88 (6.04)	17.77 (5.99)	19.15 (6.45)	20.30 (6.78)
Mean height of trees (m)	14.72 (3.06)	16.21 (3.44)	17.16 (3.13)	18.85 (3.33)

551

552

553

554

555

556

557 **Table 2** Percentage of deviance explained, AIC (Akaike´s Information Criterion), θ parameter in the

558 variance of the negative binomial distribution, basis dimension (k) and effective degrees of freedom (e.df)

559 of the functional linear predictor and the spatial smooth according to model 1 and model 2 in both plots.

560 Inventory 2001, Inventory 2006, Inventory 2010 and Inventory 2014 indicate the effective degrees of

561 freedom of the spatial smooth during the four inventories in Model 2

Feature	Intermediate stage of the regeneration period		End of the regeneration period	
	Model 1	Model 2	Model 1	Model 2
Deviance explained (%)	40.7	41.0	34.8	41
AIC	1 3120.98	1 3432.74	5 083.48	5 184.80
θ of variance	1.53	1.42	0.83	0.76
k of f_1	10	10	10	10
e.df of f_1	3.47	3.599	4.33	4.40
k of f_2	100	100	30	30
e.df of f_2 in model 1	90.42	-	24.50	-
Inventory 2001	-	69.60	-	19.47
Inventory 2006	-	65.64	-	17.30
Inventory 2010	-	63.42	-	16.28
Inventory 2014	-	63.11	-	13.15

562   f_1: linear predictor. f_2: spatial smoother

563

564

565

566

567

568

569

570

571

572   **Table 3** Summary of the backward stepwise variables selection process according to the Akaike´s

573   Information Criterion (AIC). In bold, the selected model

| Variables included in the alternative models during backward stepwise selection process	AIC
Intermediate stage of the regeneration period	
(1) $\alpha + \beta \cdot \dfrac{Nsmall_{ij}}{AreaIn10_i} + \sum\limits_{n=1}^{N}\left(f_1(Dist_{in}) \cdot \dfrac{dbh_{jn}}{AreaIn30_i} \right) + f_2\left(X_i, Y_i\right) + Time_j$	**1 3120.98**
(2) $\alpha + \sum\limits_{n=1}^{N}\left(f_1(Dist_{in}) \cdot \dfrac{dbh_{jn}}{AreaIn30_i} \right) + f_2\left(X_i, Y_i\right) + Time_j$	1 3126.06
(3) $\alpha + \beta \cdot \dfrac{Nsmall_{ij}}{AreaIn10_i} + f_2\left(X_i, Y_i\right) + Time_j$	1 3125.11
(4) $\alpha + \beta \cdot \dfrac{Nsmall_{ij}}{AreaIn10_i} + \sum\limits_{n=1}^{N}\left(f_1(Dist_{in}) \cdot \dfrac{dbh_{jn}}{AreaIn30_i} \right)$	1 500.13
(5) $\alpha + \beta \cdot \dfrac{Nsmall_{ij}}{AreaIn10_i} + \sum\limits_{n=1}^{N}\left(f_1(Dist_{in}) \cdot \dfrac{dbh_{jn}}{AreaIn30_i} \right) + f_2^j\left(X_i, Y_i\right)$	1 3432.74
End of the regeneration period	
(1) $\alpha + \beta \cdot \dfrac{Nsmall_{ij}}{AreaIn10_i} + \sum\limits_{n=1}^{N}\left(f_1(Dist_{in}) \cdot \dfrac{dbh_{jn}}{AreaIn30_i} \right) + f_2\left(X_i, Y_i\right) + Time_j$	5 083.94
(2) $\alpha + \sum\limits_{n=1}^{N}\left(f_1(Dist_{in}) \cdot \dfrac{dbh_{jn}}{AreaIn30_i} \right) + f_2\left(X_i, Y_i\right) + Time_j$	**5 083.48**

$$(3)\ \alpha + f_2\left(X_i, Y_i\right) + Time_j \qquad\qquad 5\ 120.71$$

$$(4)\ \alpha + \sum_{n=1}^{N}\left(f_1(Dist_{in}) \cdot \frac{dbh_{jn}}{AreaIn30_i}\right) \qquad\qquad 5\ 387.18$$

$$(5)\ \alpha + \beta \cdot \frac{Nsmall_{ij}}{AreaIn10_i} + \sum_{n=1}^{N}\left(f_1(Dist_{in}) \cdot \frac{dbh_{jn}}{AreaIn30_i}\right) + f_2^{\,j}\left(X_i, Y_i\right) \qquad 5\ 184.80$$
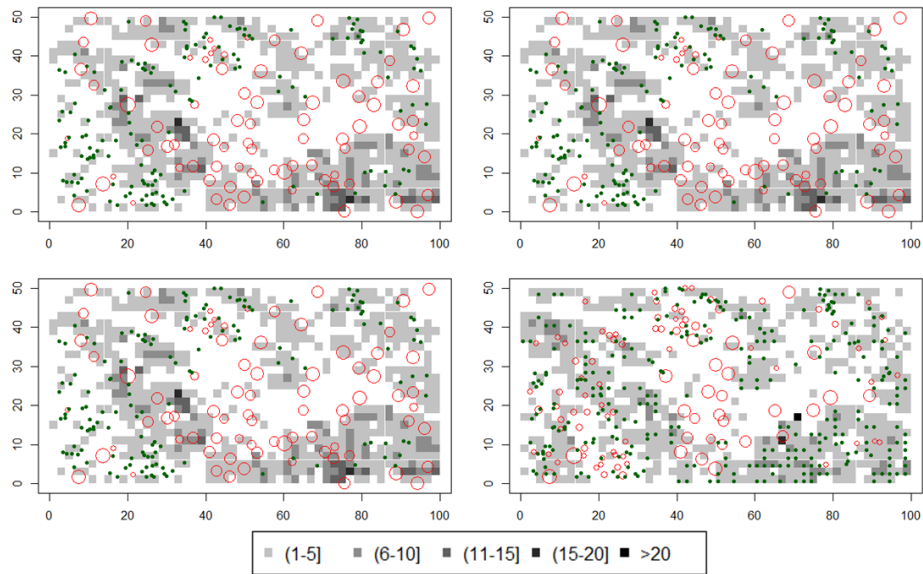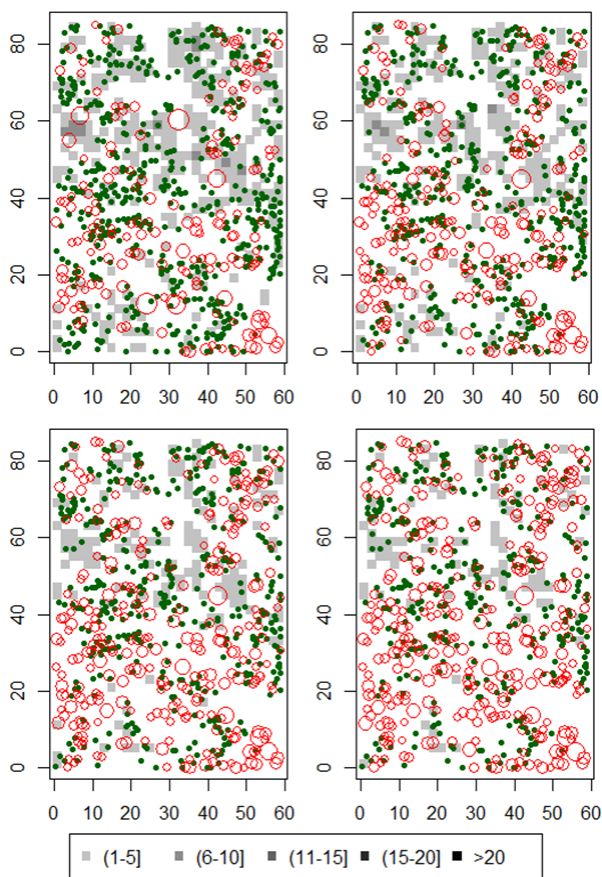
574

575

576

577

578

579 **Caption of figures**
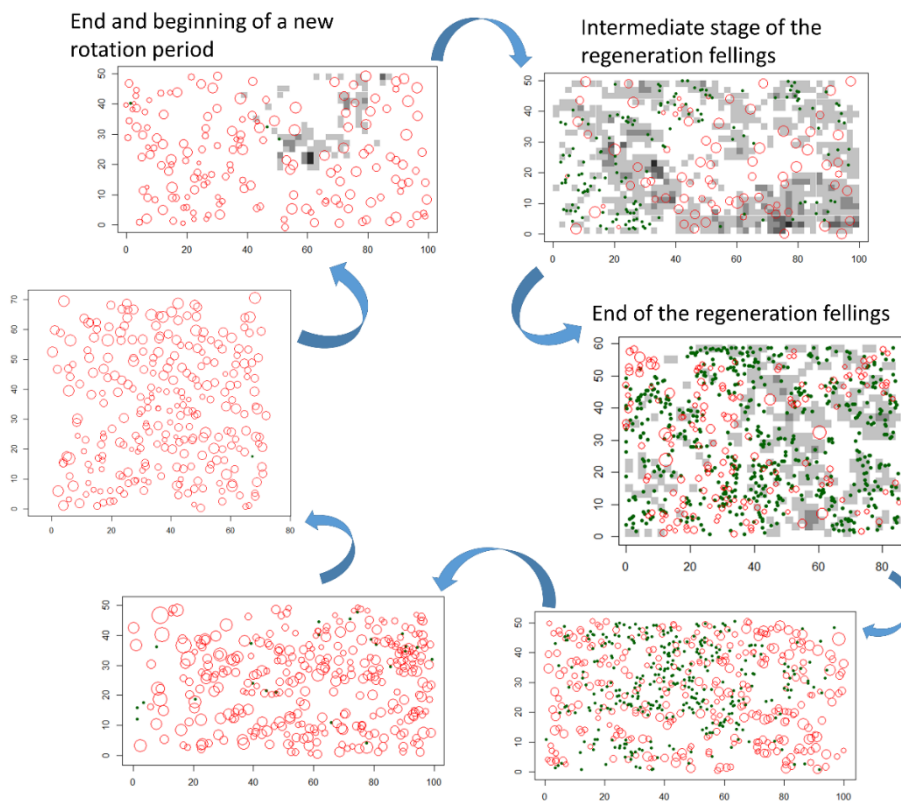
580

581 **Fig. 1** Position of adult trees (dbh≥20 cm; red circles), small trees (10≤dbh≤20 cm; green dots) and number

582 of saplings per quadrat (darker tones indicate larger number of saplings) of the six plots of the

583 chronosequence in 2001. Size of adult trees is proportional to dbh

584

Fig. 2 Position of adult trees (dbh≥20 cm; red circles), small trees (10≤dbh≤20 cm; green dots) and number of saplings per quadrat (black and gray squares) at intermediate stages of the regeneration period in 2001 (upper left), 2006 (upper right), 2010 (bottom left) and 2014 (bottom right). Size of adult trees is proportional to dbh


589

590 **Fig. 3** Position of adult trees (dbh≥20 cm; red circles), small trees (10≤dbh≤20 cm; green dots) and number

591 of saplings per quadrat (black and gray squares) at the end of the regeneration period in 2001 (upper left),

592 2006 (upper right), 2010 (bottom left) and 2014 (bottom right). Size of adult trees is proportional to dbh
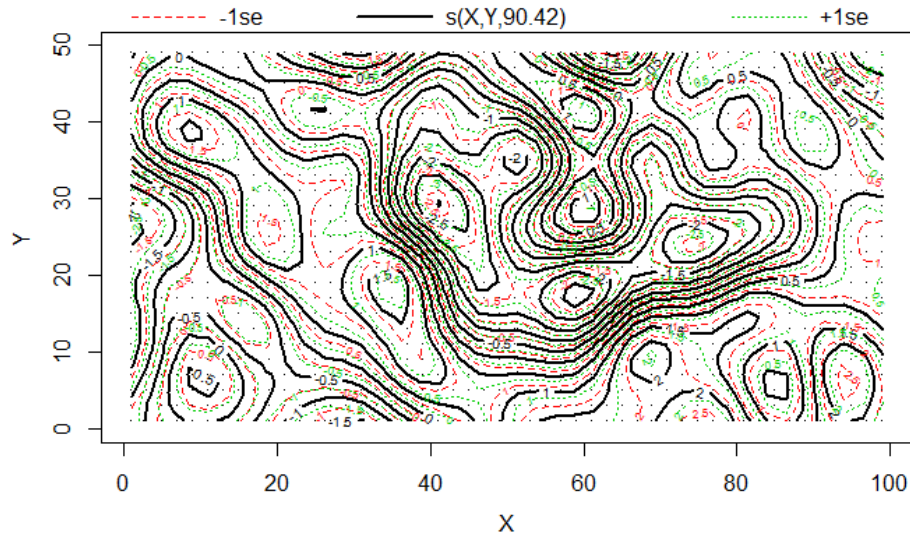
593

594 **Fig. 4** Estimated $f_2(X_i,\ Y_i)$ spatial smooth function (continuous black contour lines) and standard errors

595 (dashed red and green contour lines) on the scale of the linear predictor at intermediate stages of the

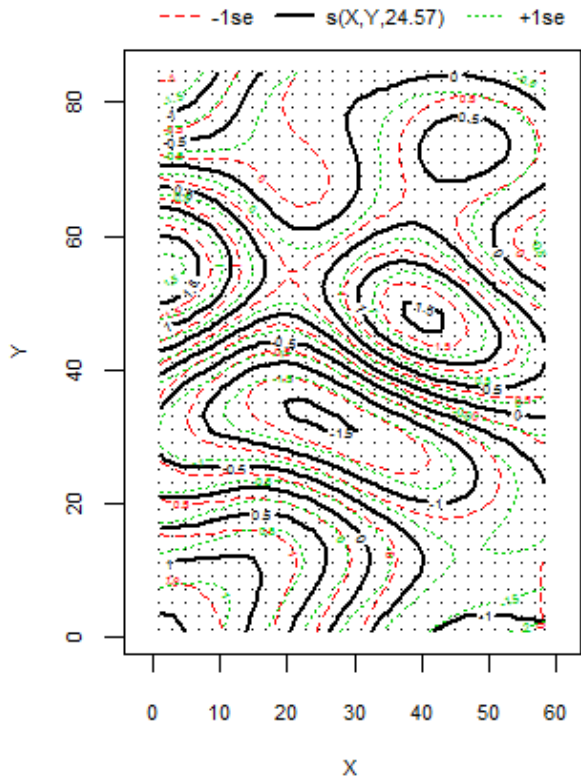596 regeneration period. Large values of $f_2(X_i,\ Y_i)$ indicate large number of saplings
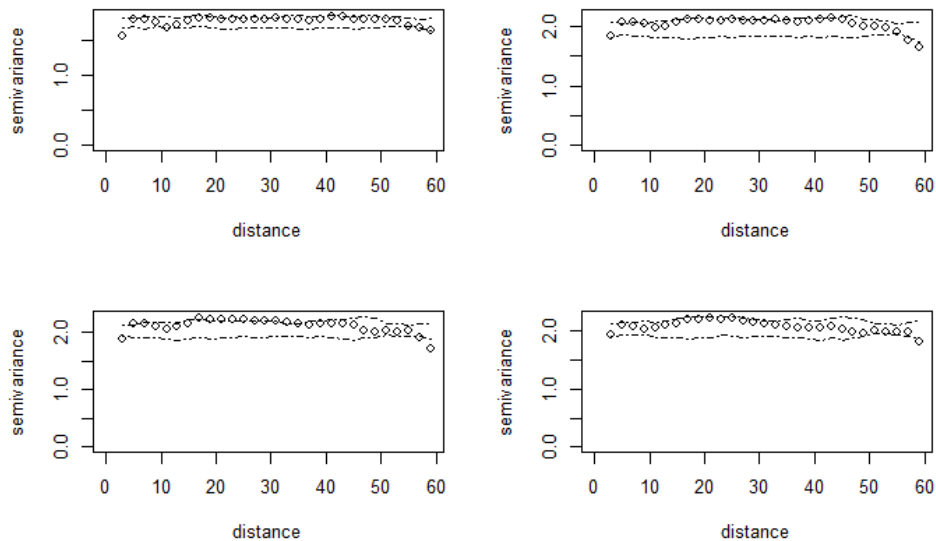
597

598 **Fig. 5** Semivariograms (circles) and envelopes (dashed lines) of the Pearson residuals from the sapling

599 distribution model at intermediate stages of the regeneration in 2001 (upper left), 2006 (upper right), 2010
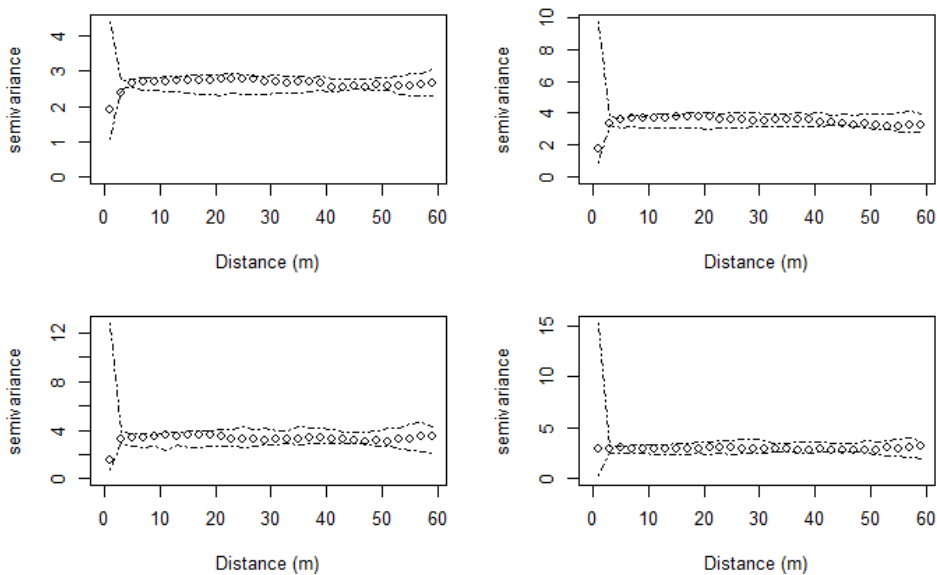
600 (lower left) and 2014 (lower right)

601
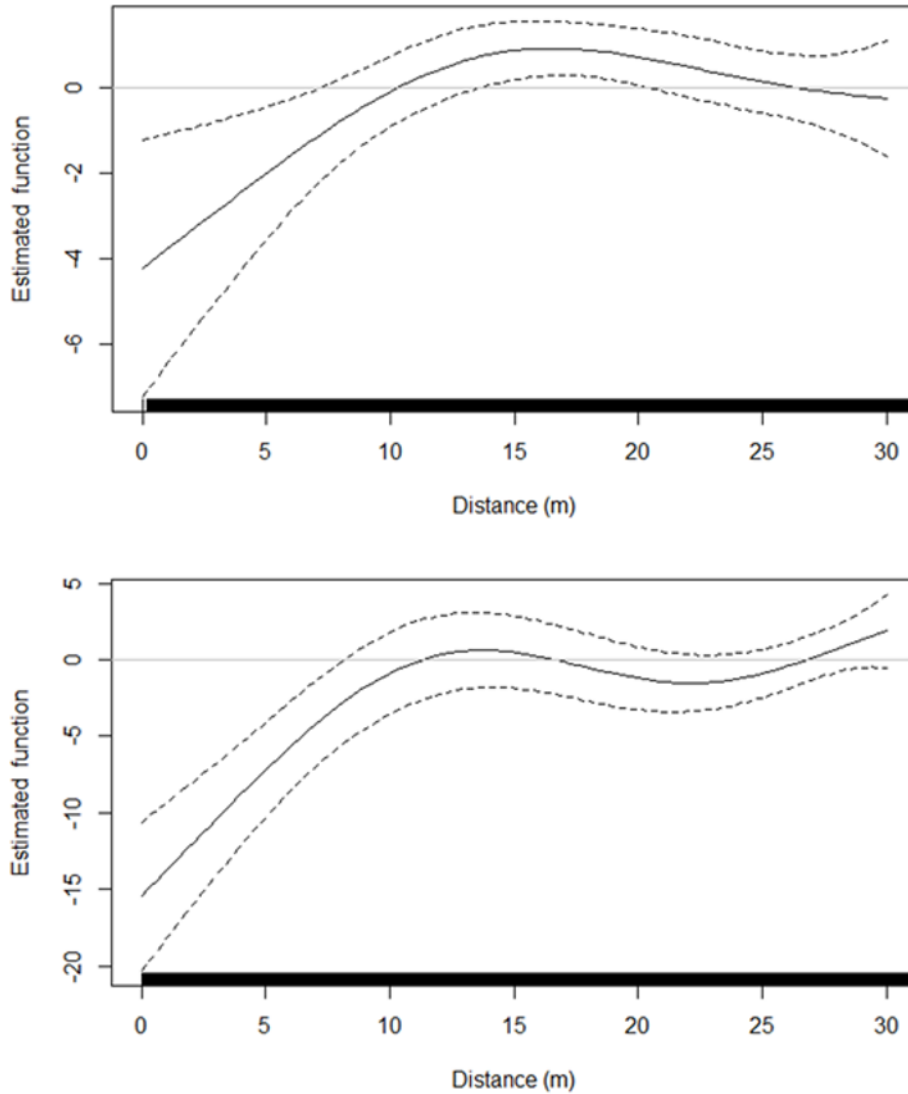
Fig. 6 Estimated $f_l(Dist_{in})$ smooth coefficient function of the diameter at breast height of adult trees over

the distance between adult trees – saplings (continuous lines) and 95% confidence intervals (dashed lines)

at intermediate stages (upper) and the end (lower) of the regeneration period. Positive values of $f_l(Dist_{in})$

indicate positive effects of the diameter at breast height of adult trees on the number of saplings

606

607   **Fig. 7** Estimated $f_2(X_i, Y_i)$ spatial smooth function (continuous black contour lines) and standard errors

608   (dashed red and green contour lines) on the scale of the linear predictor at the end of the regeneration period.

609   Large values of $f_2(X_i, Y_i)$ indicate large number of saplings

610

611   **Fig. 8** Semivariograms (circles) and envelopes (dashed lines) of the Pearson residuals from the sapling

612   distribution model at the end of the regeneration period in 2001 (upper left), 2006 (upper right), 2010 (lower

613   left) and 2014 (lower right)