

# Smooth additive mixed models for predicting aboveground biomass

Mariola Sánchez-González<sup>1</sup>, María Durbán<sup>2\*</sup>, Dae-Jin Lee<sup>3</sup>,  
Isabel Cañellas<sup>1</sup>, Hortensia Sixto<sup>1</sup>

<sup>1</sup> Dept. of Silviculture and Forest Systems Management, CIFOR-INIA  
Ctra. de la Coruña, km 7,5. 28040, Madrid, Spain

<sup>2</sup> Dept. of Statistics, Universidad Carlos III de Madrid  
Avda. Universidad 30, 28911, Leganés, Madrid, Spain

\* Corresponding author: mdurban@est-econ.uc3m.es

<sup>3</sup> BCAM - Basque Center for Applied Mathematics, Applied Statistics  
Mazarredo 14, 48009, Bilbao, Bizkaia, Spain

November 22, 2016

## Abstract

Aboveground biomass estimation in short-rotation forestry plantations is an essential step in the development of crop management strategies as well as allowing the economic viability of the crop to be determined prior to harvesting. Hence, it is important to develop new methodologies that improve the accuracy of predictions, using only a minimum set of easily obtainable information i.e. diameter and height. Many existing models base their predictions only on diameter (mainly due to the complexity of including further covariates), or rely on complicated equations to obtain biomass predictions. However, in tree species, it is important to include height when estimating aboveground biomass because this will vary from one genotype to another. This work proposes the use of a more flexible and easy to implement model for predicting aboveground biomass (stem, branches and total) as a smooth function of height and diameter using smooth additive mixed models which preserve the additive property necessary to model the relationship within wood fractions, and allows the inclusion of random effects and interaction terms. The model is applied to the analysis of three trials carried out in Spain, where nine clones at three different sites are compared. Also, an analysis of slash pine data is carried out in order to compared with the approach proposed by Parresol (2001)

KEYWORDS: Populus hybrids; additivity; tree biomass; penalized splines; mixed models

## 1 Introduction

The importance of short rotation forest (SRF) plantations to obtain biomass for renewable energy production has become widely accepted in recent years (Christeson, 2010). The *Populus* genus is considered one of the most suitable woody perennial crops for this purposes, due to its high productivity, ease of establishment from cuttings, high tolerance to coppicing and the abundance of clonal supply. Additionally, the quality of the biomass produced is highly suitable for energy purposes (Dickmann et al., 2001) in comparison with herbaceous crops (Guidi et al., 2008). All these factors contribute to making *Populus* the most interesting woody species in the short term for use in short rotation and high density plantations.

In southern Europe, the market for biomass, particularly from crops, is emerging strongly. The development of this market requires, among other factors, an improved ability to predict the aboveground biomass production in order to calculate economic gains or to decide on the timing of harvests versus continued management. For these reasons, developing suitable aboveground biomass equations may be considered one of the most important elements in short rotation coppice modeling.

The content in bark is a feature that can affect the woody biomass quality (Guidi et al., 2008). Therefore, a single aboveground biomass equation that estimates total biomass without taking into account the different wood fractions (stem and branches) is not appropriate for plantation management. When more than one component is considered, it is desirable that the biomass equations have the property of additivity, that is, that the predictions for the tree components (stem and branches) sum to the prediction for the total tree (Parresol, 1999). However, many published biomass equations of poplar short-rotation coppice were not additive (see for example Dillen et al., 2007; Dowell et al., 2009; Zabek and Prescott, 2006). The traditional methods for forcing additivity of component biomass equations to sum to total tree mass can be grouped into three categories, according to the form of the prediction equations: linear (with additive error), non-linear (with additive error) and non-linear (with multiplicative error). These approaches are commonly used, although some of them may lead to singular covariance matrices that will not ensure a unique solution for the parameters. Some solutions have been proposed based on *Non-linear Seemingly Unrelated Regression* (Parresol, 2001). However, they rely on a complicated selection procedure among a large set of non-linear equations (Ruiz-Peinado et al., 2011). Furthermore, data for estimating biomass are generally taken from trees growing in plots that are located in different sites. Such clustered

data are characterized by a lack of independence between observations which come from the same sampling cluster (Fox et al., 2001). Satisfying the additivity property in the presence of random effects is not a straightforward task, some authors use a nonlinear mixed effects model, but no constraints are imposed for additivity (Wutzler T. et al., 2008), and others, such as Parresol (2001) or Carvalho and Parresol (2003), use weighted or non-linear seemingly unrelated regression, that can deal with additivity, but no random effects are included, and rely on complicated non-linear equations that need to be chosen among a large number of possible options. Hence, it is necessary to find a model that is capable of providing accurate aboveground biomass predictions, satisfies the additivity property, and includes random effects to account for the different sources of variability in the data.

We propose the use of smooth additive mixed effects models (Lee et al., 2013; Wood et al., 2013) for the estimation and prediction of aboveground biomass data. This is a novel application of this family of models, and a simpler alternative to traditional approaches based on complex nonlinear biomass equations that have to be chosen ad-hoc for each data set. Furthermore, our proposal is more flexible, it can deal with all issues specified above, and can be applied in many different situations (different experimental designs, models with several covariates, trials carried out along time, etc.). A further advantage of this approach is that is easily implemented **R** (using the `lme` procedure or the `gam` function in the `mgcv` package), but also in other packages, such as **SAS**, by means of the `PROC MIXED` function.

Smooth additive models are widely used in many areas of research. However, until now, smoothing models have scarcely been used in forestry applications. Guan et al. (2006) used semiparametric mixed effects smoothing-spline models to analyze microclimate-monitoring data from a thinning experiment, and Jordan et al. (2008) used this approach to model the relationship between wood specific gravity and time for loblolly pine plantations in the United States. More recently, Goicoa et al. (2011) proved that these models satisfied the additivity property. This approach offers the flexibility to model non-linear relationships within components without including biases related to data transformation. However, from a management point of view, it is also important to be able to predict biomass with accuracy using only a minimum set of easily measurable information, i.e., diameter and height. Hence, the main reason for using this approach is that it is the best option when the same explanatory variables are considered for modeling all the components including total biomass. Goicoa et al. (2011) used diameter as an explanatory variable to compare alternative methods for obtaining biomass predictions which satisfy the additivity property, but they did not

approach the problem of having some sort of clustering and different genotypes in the data. Also, it is important to include height when estimating aboveground biomass in the *Populus* genus because differences exist between genotypes. Genotypes with greater development allocate more stem biomass to radial growth than to height growth, whereas the smaller size genotypes tend to exhibit greater height growth at the cost of diameter growth in order to gain access to light (Wu and Stettler, 1996). The inclusion of another explanatory variable implies that a smooth additive mixed model based on penalized splines (P-splines) by Eilers and Marx (1996) could be used, with the consequent difficulty involved in dealing with height-diameter interaction. An additional factor to consider is that Poplar hybrids present differences in growth strategies, which might result in significant differences in biomass production rates (Telenius and Verwijst, 1995). Biomass allocation patterns (stem and branches) differ among genotypes, but the environment also has a notable effect. Some authors (Dillen et al., 2007; Pontailier et al., 1997) have reported that the genotype effect is dominant over the environmental effect with regard to biomass production so, clone-specific models are more suitable for biomass estimations (in this study, we take this into account by using factor-by-surface interactions).

The rest of this paper is organized as follows: Section 2 describes P-splines and their representation as mixed models, and introduced several smooth additive mixed models for biomass prediction. The models proposed are illustrated in Section 3 with the analysis of data from three trials in Spain. We also analyze the dataset in Parresol (2001) in order to compare different methods. Finally, the paper concludes with a discussion in Section 5. Some technical details are available in the Supplementary Material.

## 2 Additive smooth mixed models for aboveground biomass data

We give first, a brief introduction to the mixed model representation of penalized splines, and then we introduced several smooth additive mixed models for prediction of biomass data

### 2.1 Penalized splines as mixed models

Suppose the variable  $\mathbf{y} = (y_1, \dots, y_n)$  depends smoothly on a single variable  $\mathbf{x} = (x_1, \dots, x_n)$ , then the nonparametric model for  $\mathbf{y}$  can be written  $\mathbf{y} = f(\mathbf{x}) + \boldsymbol{\epsilon}$  where  $f(\cdot)$  is a smoothly varying function and  $\boldsymbol{\epsilon}$  is the vector

of errors that could be assumed independent, or with variance  $\sigma^2\mathbf{W}^{-1}$ , and  $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$  is a diagonal matrix of weights to account for heteroscedasticity. We use the approach proposed by Eilers and Marx (1996), based on two assumptions: (i)  $\mathbf{E}(\mathbf{y}) = \mathbf{B}\boldsymbol{\theta}$  where  $\mathbf{B}$  is a matrix of B-splines, and  $\boldsymbol{\theta}$  is the vector of regression coefficients; (ii) modify the likelihood function by adding a penalty term over the adjacent coefficients to control the smoothness of the fit (details of this representation can be found in the Supplementary Material). Therefore, the model becomes:

$$\mathbf{y} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{W}^{-1}), \quad (1)$$

and the coefficients  $\boldsymbol{\theta}$  are chosen to minimize

$$S(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{B}\boldsymbol{\theta})'\mathbf{W}(\mathbf{y} - \mathbf{B}\boldsymbol{\theta}) + \boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta},$$

where  $\mathbf{P}$  is the penalty matrix. We will use second order difference penalties between adjacent coefficients as the penalty, i.e.  $\mathbf{P} = \lambda\mathbf{D}'\mathbf{D}$  ( $\mathbf{D}$  is the matrix of second order differences), and  $\lambda$  is the smoothing parameter that controls the influence of the penalty on the fit. Other basis/penalties could also be considered. We choose to reparameterize model in Equation (1) using the representation of a penalized spline model as a mixed model, as introduced by Durbán et al. (2003) (see further details in the Supplementary Material). The use of this representation is not always necessary, although as we shall see in the analysis of the data set, its use is crucial in this case since we need to include random effects in the model (site and block), and therefore, need a unified approach for the estimation of the fixed and random effects in the model. Then, model (1) becomes:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2\mathbf{W}^{-1}) \quad \boldsymbol{\alpha} \sim N(0, \mathbf{G}), \quad (2)$$

The reparametrization is based on the singular value decomposition (SVD) of the penalty matrix  $\mathbf{D}'\mathbf{D}$  (Currie et al., 2006; Lee, 2010), which yields.

$$\mathbf{X} = [\mathbf{1}_n : \mathbf{x} : \mathbf{x}^2 : \dots : \mathbf{x}^{q-1}], \quad (3)$$

where  $\mathbf{1}_n$  is a column vector of ones, and  $q$  is the order of the penalty. For the random part, the random effects matrix  $\mathbf{Z}$  is defined as:

$$\mathbf{Z} = \mathbf{B}\mathbf{U}_r\tilde{\boldsymbol{\Sigma}}^{-1/2}, \quad (4)$$

where  $\tilde{\boldsymbol{\Sigma}}^{-1/2}$  are the non-zero eigenvalues of the singular value decomposition (SVD) of the penalty, and  $\mathbf{U}_r$  are the corresponding eigenvectors.

With this transformation  $\mathbf{G}$  becomes  $\sigma_\alpha^2 \mathbf{I}_{c-q}$ , and the smoothing parameter  $\lambda = \sigma^2/\sigma_\alpha^2$ . Estimation of variance components is carried out by the Restricted Maximum Likelihood (REML) approach proposed by Patterson and Thompson (1971), and the vector of parameters  $\boldsymbol{\beta}$  and the random effects coefficients  $\boldsymbol{\alpha}$  are determined by the best linear unbiased prediction:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ \hat{\boldsymbol{\alpha}} &= \sigma_\alpha^2 \mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}),\end{aligned}$$

where  $\mathbf{V} = \mathbf{W}^{-1} + \mathbf{Z}\mathbf{G}\mathbf{Z}'$ .

## 2.2 Smooth models for aboveground biomass data

The use of P-splines for predicting aboveground biomass was first introduced by Goicoa et al. (2011), although the model used there only allowed for a single covariate (diameter of the tree). In this study, we propose the use of the model introduced in Lee et al. (2013) which allows the use of several covariates (in this case diameter and height) and its interactions. The model is flexible enough to estimate the main effect of these two variables, and the interaction between the two, while preserving the additivity property, and also allowing the estimation of random effects. The model is based on the P-spline ANOVA mixed model of Lee and Durbán (2011):

$$\mathbf{y} = f(\mathbf{x}_1) + f(\mathbf{x}_2) + f_{1,2}(\mathbf{x}_1, \mathbf{x}_2) + \boldsymbol{\epsilon}, \quad (5)$$

where  $f(\mathbf{x}_1)$  and  $f(\mathbf{x}_2)$  are functions that represent the main effects of diameter and height, and  $f_{1,2}$  is a smooth function for the interactions between them. The regression matrix is then defined by blocks as:

$$\mathbf{B} = [\mathbf{B}_1 | \mathbf{B}_2 | \mathbf{B}_{[1,2]}], \quad (6)$$

with marginal bases of the covariates,  $\mathbf{B}_1$  (for diameter), and  $\mathbf{B}_2$  (for height), with dimensions  $n \times c_1$  and  $n \times c_2$  respectively. The methodology proposed above can be extended to include any number of covariates by defining a new basis and penalty. In this case the basis is constructed from the tensor product of the marginal B-splines basis defined in Eilers et al. (2006):

$$\mathbf{B}_{[1,2]} = \mathbf{B}_1 \square \mathbf{B}_2 = (\mathbf{B}_1 \otimes \mathbf{1}'_{c_1}) \odot (\mathbf{1}'_{c_2} \otimes \mathbf{B}_2), \quad (7)$$

where the operator  $\odot$  is the *element-wise* matrix product, and  $\mathbf{1}$  is a column vector of ones. The operation in expression (7) is such that each row of

$\mathbf{B}_{[1,2]}$  is the Kronecker product for the corresponding rows of  $\mathbf{B}_2$  and  $\mathbf{B}_1$ . The penalty matrix associated with model (5) has a block-diagonal form:

$$\mathbf{P} = \text{blockdiag}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_{[1,2]}), \quad (8)$$

where  $\mathbf{P}_i = \lambda_i \mathbf{D}'_i \mathbf{D}_i$ , ( $i = 1, 2$ ) are the marginal difference penalties for height and diameter defined previously, and

$$\mathbf{P}_{[1,2]} = \lambda_3 \mathbf{D}'_1 \mathbf{D}_1 \otimes \mathbf{I}_{c_2} + \lambda_4 \mathbf{I}_{c_1} \otimes \mathbf{D}'_2 \mathbf{D}_2. \quad (9)$$

However, the interaction term  $f_{1,2}(\mathbf{x}_1, \mathbf{x}_2)$  may be not flexible enough, for example when the interaction between covariates also includes a varying coefficient component. Also, fitting model (5) might be computationally very demanding since there are four smoothing parameters to be chosen. We take the approach described in Lee et al. (2013) and model the interaction  $f_{[1,2]}(\mathbf{x}_1, \mathbf{x}_2)$  as the sum of three components: two linear-by-smooth interactions,  $g_1(\mathbf{x}_1)\mathbf{x}_2$  and  $\mathbf{x}_1 g_2(\mathbf{x}_2)$  (which allow smooth functions of clone and height to vary smoothly along the other covariate) and a smooth-by-smooth interaction, i.e,

$$f_{[1,2]}(\mathbf{x}_1, \mathbf{x}_2) = g_1(\mathbf{x}_1)\mathbf{x}_2 + \mathbf{x}_1 g_2(\mathbf{x}_2) + h(\mathbf{x}_1, \mathbf{x}_2). \quad (10)$$

Then, model (5), called PS-ANOVA (for P-splines Smooth ANOVA) becomes

$$\mathbf{y} = f(\mathbf{x}_1) + f(\mathbf{x}_2) + g_1(\mathbf{x}_1)\mathbf{x}_2 + \mathbf{x}_1 g_2(\mathbf{x}_2) + h(\mathbf{x}_1, \mathbf{x}_2) + \boldsymbol{\epsilon}, \quad (11)$$

with associated B-spline basis:

$$\mathbf{B} = [\mathbf{B}_1 | \mathbf{B}_2 | \mathbf{B}_3 | \mathbf{B}_4 | \mathbf{B}_5], \quad (12)$$

where  $\mathbf{B}_1$  and  $\mathbf{B}_2$  are the marginal basis functions for diameter and height, and

$$\mathbf{B}_3 = \mathbf{B}_1 \square \mathbf{x}_2, \quad \mathbf{B}_4 = \mathbf{x}_1 \square \mathbf{B}_2, \quad \text{and} \quad \mathbf{B}_5 = \mathbf{B}_1 \square \mathbf{B}_2,$$

The penalty matrix becomes:

$$\mathbf{P} = \text{blockdiag}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4, \mathbf{P}_5), \quad (13)$$

where,  $\mathbf{P}_i = \lambda_i \mathbf{D}'_i \mathbf{D}_i$ , ( $i = 1, 2$ ) (corresponding to the main effects),  $\mathbf{P}_3 = \lambda_3 \mathbf{D}'_1 \mathbf{D}_1$  and  $\mathbf{P}_4 = \lambda_4 \mathbf{D}'_2 \mathbf{D}_2$  penalize the varying coefficient terms  $g_1(\mathbf{x}_1)\mathbf{x}_2$  and  $\mathbf{x}_1 g_2(\mathbf{x}_2)$ , and  $\mathbf{P}_5 = \lambda_5 (\mathbf{D}'_1 \mathbf{D}_1 \otimes \mathbf{I}_{c_2} + \mathbf{I}_{c_1} \otimes \mathbf{D}'_2 \mathbf{D}_2)$  penalizes the interaction term (but with a single smoothing parameter). The fact that



the smoothing parameters are independent of each other greatly improves the computational time (even though 5 smoothing parameters are used). However, model (11) is not identifiable, since matrix  $\mathbf{B}$  in Equation (12) is not of full rank. The mixed model representation of this model will solve easily the identifiability problem. After appropriate transformation (details are given in the Supplementary Material), model (11) can be expressed as (2), where the fixed and random effects matrices are:

$$\mathbf{X} = [\mathbf{1}_n | \mathbf{x}_1 | \mathbf{1}_n | \mathbf{1}_n | \mathbf{x}_2 | \mathbf{x}_1 | \mathbf{x}_2] \quad (14)$$

$$\mathbf{Z} = [\mathbf{Z}_1 | \mathbf{1}_n | \mathbf{1}_n | \mathbf{Z}_2 | \mathbf{Z}_1 | \mathbf{x}_2 | \mathbf{x}_1 | \mathbf{Z}_2 | \mathbf{Z}_1 | \mathbf{Z}_2]. \quad (15)$$

The vector of random effects  $\boldsymbol{\alpha} \sim N(\mathbf{O}, \mathbf{G})$  can be expressed as  $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_5)'$ , where  $\boldsymbol{\alpha}_k \sim N(\mathbf{O}, \mathbf{G}_k)$  with covariance matrices:

$$\mathbf{G}_1 = \sigma_1^2 \mathbf{I}_{c_1-2}, \mathbf{G}_2 = \sigma_2^2 \mathbf{I}_{c_2-2}, \mathbf{G}_3 = \sigma_3^2 \mathbf{I}_{c_1-2}, \mathbf{G}_4 = \sigma_4^2 \mathbf{I}_{c_2-2}, \text{ and } \mathbf{G}_5 = \sigma_5^2 \mathbf{I}_{(c_1-2)(c_2-2)},$$

then,  $\mathbf{G} = \text{blockdiag}(\mathbf{G}_1, \dots, \mathbf{G}_5)$ .

Confidence intervals for the fitted smooth terms in model (11) can be calculated using the covariance matrix associated with each of the components:

$$\text{Var}(\hat{f}_j | \boldsymbol{\alpha}) = \sigma^2 \mathbf{H}_j,$$

where  $\mathbf{H}_j$  is part of the hat matrix of the model,  $\mathbf{H}$ , for the  $j$ -th smooth term:

$$\mathbf{H}_j = \mathbf{C}_j \mathbf{H} \quad \mathbf{H} = \mathbf{M}(\mathbf{M}'\mathbf{M} + \mathbf{R})^{-1} \mathbf{M},$$

where  $\mathbf{M} = [\mathbf{X} : \mathbf{Z}]$  and  $\mathbf{R}$  is the matrix  $\sigma^2 \mathbf{G}^{-1}$  augmented with 4 zero in the diagonal corresponding to the unpenalized coefficients of  $\mathbf{X}$ .  $\mathbf{C}_j$  is a diagonal matrix with ones in the diagonal positions corresponding to the  $j$ -th smooth term. Also,  $\mathbf{H}_j$  can be used to give a measure of the complexity of the smooth curve/surface (in a similar way to a linear model) by defining the effective dimension of a smooth term as:

$$e.d.(\hat{f}_j) = \text{trace}(\mathbf{H}_j).$$

Lee et al. (2013) showed through simulations and with real data sets, that model in Equation (11) presents better results in terms of computation and AIC than model in Equation (5) in most situations when interaction between covariates is present.

### Additivity property

Using a separate P-spline model for the different tree components, i.e. for stem ( $\mathbf{ws}$ ), branches ( $\mathbf{wb}$ ), and total aboveground biomass ( $\mathbf{wt} = \mathbf{ws} + \mathbf{wb}$ ), will not give estimates that satisfy the additivity property. To satisfy this requirement, Goicoa et al. (2011) proposed the use of the same B-spline basis and the same smoothing parameter for all the components (in the context of mixed models this would mean using the same variance of the random effects for all components). Firstly, the P-spline model is fitted for the total aboveground biomass of the tree, and estimates of the smoothing parameters are obtained. The fact that the choice of the amount of smoothness is done using the total biomass is not a problem, since, in most cases, the other measurements present similar optimal smoothing parameters. The P-spline model is then fitted for the tree components setting the smoothing parameter to the value previously obtained. Given matrices  $\mathbf{B}$  and  $\mathbf{P}$  given in (12) and (13) respectively (or their mixed model equivalent):

$$\hat{\mathbf{w}}\mathbf{t} = \mathbf{B}(\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1}\mathbf{B}'\mathbf{w}\mathbf{t} = \mathbf{B}(\mathbf{B}'\mathbf{B} + \mathbf{P})^{-1}\mathbf{B}'(\mathbf{w}\mathbf{s} + \mathbf{w}\mathbf{b}) = \hat{\mathbf{w}}\mathbf{s} + \mathbf{1}\mathbf{b}$$

It is worth noticing that, although the smoothing parameters used in the estimation of each component must remain fixed, this does not mean that the parameters of the model or the functional form of the curve are the same, is only amount of smoothness that is fixed. Also, if other random effects are included in the model (site and block) their variance parameters should also be fixed in order to ensure additivity.

## 3 Application

Three trials were established in three different Mediterranean thermotypes (Rivas Martínez et al., 2002), using similar design and maintenance procedures. Nine clones belonging to the European Catalog of Basic Materials for the *Populus* genus, and used in southern Europe were included in the trials. These clones belong to different inter-specific hybrids of *Populus*: five of *P. deltoides* March. x *P. nigra* L. (AF2, Guardi, I214, MC and 2000 verde); two *P. trichocarpa* T. & G. x *P. deltoides* (Unal and USA 49-177) and two (*P. deltoides* x *P. trichocarpa*) x *P. nigra* (Monviso and Pegaso). Clone I214 is considered a control clone because of its widespread cultivation in the study area. A multi-environment trial in a four randomized block design was employed at each site. Site (tree levels) was the main factor, with four blocks and three replications per block at each site (12 subplots); clone

was the secondary factor (nine levels), hence 324 plots were measured in total. Each subplot comprised a total of 25 ramets per clone and site but only 9 were evaluated in order to avoid the border effect. Hence, a total of 2916 trees were evaluated (the model proposed in the previous section is particularly suitable for large data sets as it happens in this case). Several variables were measured outside the vegetative period and at the end of the rotation (third year of the rotation) in order to assess the performance of the clone at each site. These variables were diameter (in mm) at 1.30 m (dbh), measured using a digital calliper, and total height (in cm) measured using a pole (h). Production was measured by recording total aboveground dry biomass (Mg DM ha<sup>-1</sup> - BT), estimating the dry weight of a subsample (oven-dried to constant weight at 100 °C) taken from the nine plants contained in 12 subplots per clone and site. Stem and branch biomass were weighted separately and the stem/branch biomass ratio (S/B ratio) was used to assess the biomass distribution in the different clones. Figure 1 shows the relationship between diameter, height and the three measurements (total, stem and branch weight). The total aboveground dry biomass for all the genotypes and environments tested was 40.45 Mg.ha<sup>-1</sup> for the whole rotation (3 years). The production ranged from 48.77 Mg.ha<sup>-1</sup> for the most productive site (S2) and 27.86 Mg.h<sup>-1</sup> for the least productive one (S3). At clonal level, production ranged from 53.31 Mg.ha<sup>-1</sup> (AF2) to 22.34 Mg.ha<sup>-1</sup> (Pegaso). Regarding the proportion of stem biomass, this was 84.72% for all the genotypes and environments tested. At clonal level, this proportion ranged from 80.56% (Monviso) to 86.27% (Pegaso), while at site level it ranged from 87.37% (S1) to 83.12% (S3).

[Figure 1 about here.]

We fitted and compared the following models:

**Model 1:**

The basic model, corresponding to Goicoa et al. (2011), but adding random effects for site and block:

$$wt_{ijk} = \boldsymbol{\beta} \cdot clone + f(d_{ijk}) + u_i + v_{ij} + \epsilon_{ijk}, \quad (17)$$

$wt_{ijk}$  is total above-ground biomass measured at the  $i$ -th site within the  $j$ -th block on the  $k$ -th tree,  $\boldsymbol{\beta}$  is the clone fixed-effects parameter vector,  $f(d_{ijk})$  is a smooth function of the diameter,  $u_i$  and  $v_{ij}$  are random components

common to the observations taken from the same site and within-site block respectively, and  $\epsilon_{ijk}$  is the residual error term. This model fits a common smooth trend for all clones, which means that the relationship between diameter and total weight is a curve that only differs from one clone to another by a quantity given by the corresponding clone-parameter ( $\beta_l, l = 1, \dots, 9$ ).

### Model 2:

The simplest way to include another covariate in the model would be by using an additive mixed model for diameter and height:

$$wt_{ijk} = \beta \cdot clone + f(d_{ijk}) + f(h_{ijk}) + u_i + v_{ij} + \epsilon_{ijk}, \quad (18)$$

Figure 2 shows the smooth fitted trend for diameter and height (plots are done using partial residuals, i.e. after taking into account the other terms in the model). As expected, the effect of diameter is much stronger than height, but both smooth terms are significantly different from zero, and the effective dimensions (e.d.) for each term are 6.7 (for diameter) and 4.3 (for height). This also means that they are far from being a straight line (this would correspond to an e.d. of 2) which highlights the need for a flexible non-parametric model.

[Figure 2 about here.]

This model (like the previous one) assumes that the total weight will increase with height (or diameter) in a similar way for each clone. This is unrealistic, since the individual characteristics of each clone would modify the way in which the weight changes with respect to each covariate.

### Model 3:

Model 2 is extended to fit a separate curve for each clone, by using a factor-by-curve interaction model (Durbán et al., 2005), in which a categorical variable (clone) interacts with a continuous predictor (diameter and height), i.e.:

$$wt_{ijk} = \beta \cdot clone + f_{clone}(d_{ijk}) + f_{clone}(h_{ijk}) + u_i + v_{ij} + \epsilon_{ijk}, \quad (19)$$

$f_{clone}(d_{ijk})$  and  $f_{clone}(h_{ijk})$  are nine different smooth functions for diameter and height, respectively, depending on the value of the clone. For simplicity, we have assumed a common variance parameter for all nine curves of each variable. This means that all curves of each covariate have equivalent smoothness. However, the random effects are different from function to function, i.e., the curves are different but have the same amount of smoothing.

[Figure 3 about here.]

We give an example in Figure 3, where we plot the smooth trend for diameter obtained after fitting models 2 and 3, for clone *2000 verde*; note that the trend obtained from model 2 does not fit the data very well, this is due to the fact that we are imposing the same curve on all clones, while with model 3, the curve is adapted to the specific characteristics of each clone.

Figure 4 shows the fitted surfaces for the additive model in Equation (11) for total, stem and branch biomass for Unal clone.

[Figure 4 about here.]

#### Model 4:

An additive model for height and diameter assumes that there is no interaction between both variables. However, this is not the case since the effect of diameter on total biomass is not the same for all values of height; therefore, we need to include an interaction term in model (19) :

$$wt_{ijk} = \beta \cdot clone + f_{clone}(d_{ijk}) + f_{clone}(h_{ijk}) + f_{clone}(d_{ijk}, h_{ijk}) + u_i + v_{ij} + \epsilon_{ijk}, \quad (20)$$

where  $f_{clone}(d_{ijk}, h_{ijk})$  is a smooth function of the diameter-height interaction vector for tree  $k$  within block  $j$  in site  $i$ . This interaction function is expressed as in Equation (10):

$$f_{clone}(d_{ijk}, h_{ijk}) = g_{1_{clone}}(d_{ijk})h_{ijk} + d_{ijk}g_{2_{clone}}(h_{ijk}) + t_{clone}(d_{ijk}, h_{ijk}). \quad (21)$$

We have chosen the more complex expression for the interaction given above instead of a simpler interaction term (without the varying coefficients terms,  $g_{1_{clone}}(d_{ijk})h_{ijk}$  and  $d_{ijk}g_{2_{clone}}(h_{ijk})$ ), based preliminary empirical results. We have analyzed the data (for total biomass) and compared models (11) and (5). The value of AIC for model (11) was -1216.15 compared with -1178.10 for model (5). Similar results were obtained when the models were fitted to each clone separately (further details on these results and sample code can be found in the Supplementary Material). Indeed, the varying coefficient terms accounts for the smooth-by-linear interaction effects between diameter and height.

It is worth mentioning the problem of identifiability present in all models presented in this paper. In particular, smooth additive models are not identifiable due the fact that the intercept is confounded with the main effects ( $f_1(d_{ijk})$  and  $f_2(h_{ijk})$ ) and main effects are confounded with the two-way interactions ( $f_{1,2}(d_{ijk}, h_{ijk})$ ). Some authors (Wood, 2006) identify and impose the constraints numerically, but this is difficult to extend to the case of more than 2-way interactions. We take a simpler approach, and use the mixed model representation of P-splines to remove the linearly dependent columns from the basis (Lee and Durbán, 2011; Wood et al., 2013). When there are interaction between factors and curves or surfaces (as in models (19) to (21)) further identifiability constraints need to be imposed. Several options are available, depending on the type of contrast of interest: sum of effects equal to zero, effects corresponding to first level of the factor equal to zero, etc. In this case, the identifiability problem is solved modifying the fixed and random effects matrices as follows:

$$\widetilde{\mathbf{X}} = \mathbf{C} \square \mathbf{X} \quad \widetilde{\mathbf{Z}} = \mathbf{C} \square \mathbf{Z},$$

where  $\mathbf{C}$  is the usual constraint matrix for the *clone* effect in an ANOVA model, and  $\square$  is the row-wise Kronecker product defined in Equation (7). Then,

$$\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{X}_2 & \mathbf{X}_2 & \dots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_9 & \mathbf{O} & \dots & \mathbf{X}_9 \end{bmatrix}, \quad \widetilde{\mathbf{Z}} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{Z}_2 & \mathbf{Z}_2 & \dots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Z}_9 & \mathbf{O} & \dots & \mathbf{Z}_9 \end{bmatrix},$$

and  $\mathbf{X}_i$  and  $\mathbf{Z}_i$  are defined in (14) and (15) respectively (for each clone).

All models presented in this paper can be implemented in standard mixed model software (using `lme()` function in R, or `PROC MIXED` macro in SAS), or using the library `mgcv` (versions 1.7.28 onward) in R. In particular, the function `gamm()` fits smooth additive mixed models. The constraints needed for models (20) and (21) are obtained specifying the smooth terms by the function `ti()`.

The models were fitted using B-splines of degree 3, and second order penalties, 20 knots were used to construct the basis in models 1 to 3, and basis of 10 knots were used for the interaction term in model 4, due to computational restrictions. The variance of biomass may increase with tree age and size, leading to heteroscedasticity. These models can easily be adapted to cope with this by including the appropriate weight matrix in the error term. However, as our data are from three years old trees, most of which have a diameter under 5cm, there were no trends in residuals as a function of fitted values or explicative variables in any of the tested models.

Comparison of the proposed models (models 2 to 4), was done using the Akaike Information Criterion Akaike (1973), and the adjusted  $R^2$ , defined as the proportion of variance explained, where original variance and residual variance are both estimated using unbiased estimators. As we can see in Table 1, the surface by factor model (model 4) gives the best results for all variables: total, branch and stem weight, although it has a larger effective dimension (this could be expected since it is the most complex model).

[Table 1 about here.]

In Table 2 we show the effective dimension of each smooth component of the tested models. When the diameter and height interaction is not taken into account, the main effect of diameter is much greater than that of height. Once the interaction was included, the height smooth term has an effective dimension of zero and the smooth interaction effect of diameter and height was the most important. The linear-by-smooth interaction effects of diameter and height and the main diameter effect have similar values of effective dimension. This means that the smooth effect of height is only significant through the interaction with diameter.

[Table 2 about here.]

Figure 5 plots predicted versus observed values for the total aboveground, stem and branch biomass for the AF2 clone, derived from the four tested models. Model 1 tends to overestimate biomass, while model 2 underestimates it, specially in the case of stem biomass. The predictions were very similar for models 3 and 4 in the three cases, although model 4 performed better in terms of AIC and  $R^2$ .

[Figure 5 about here.]

Figure 6 shows the fitted surfaces of model in (20) for total, stem and branch biomass for Unal and I214 clones. Notice that compared to the additive surfaces shown in Figure 4, now the interaction terms in Equation (21) allows for clone by diameter and height effects simultaneously. This fact could be expected due to genotype differences. Figure 7 shows the predictions (given by Model 4) for the tree components (stem and branches) and the total biomass for one of the clones, while models 1 and 2 tend to underestimate biomass, especially in the case of branch weight. To ensure additivity, the smoothing/variance parameters were chosen from the fit of total biomass, and they were used to predict the stem and branch weight. Variance components for site and block within site were significant for all models, but their value decreased as the model became more complex.

[Figure 6 about here.]

[Figure 7 about here.]

### **Analysis of slash pine data in Parresol (2001)**

In order to compare our methodology with other existing approaches, we analyze the data in Parresol (2001), which correspond to a sample of 40 slash pine trees in the state of Louisiana, U.S.A. Several measurements were taken, including diameter, height, live grown length and age, but only diameter and height were used. Parresol (2001) used two different procedures depending on how the separate components of the model are aggregated: i) The total biomass function is the sum of the separate regression function of each component of the model, and ii) Each component has its own



independent variables, and additivity is ensured by imposing constraints on the coefficients. We compare our results with method ii), which is the one that performs better in most cases. The system of equations used was (see equation 18 in Parresol (2001)):

$$\begin{aligned}\hat{y}_{wood} &= b_{11} (D^2 H)^{b_{12}} \\ \hat{y}_{bark} &= b_{21} D^{b_{12}} \\ \hat{y}_{crown} &= b_{31} D^{b_{32}} H^{b_{33}} \\ \hat{y}_{total} &= b_{11} (D^2 H)^{b_{12}} + b_{21} D^{b_{12}} + b_{31} D^{b_{32}} H^{b_{33}}.\end{aligned}$$

The parameters were estimated by non-linear seemingly unrelated regression, and weights we included in the model fitting to account for heteroscedasticity. We used a multidimensional P-spline for height and diameter (see Equation (5)) for each of the components of the models, and the weights used were the reciprocal of the smoothed squared residuals, as in Currie and Durbán (2002). In Table 3, we compare the models using the same measures of goodness of fit that appear in Parresol (2001): the *Fit Index* (pseudo  $R^2$ ) and the Root Mean Square Error (RMSE).

[Table 3 about here.]

The P-spline mixed model approach improves the fitting in all cases, and specially in the case of the total biomass, where the RMSE is 20% smaller, with the advantage that no functional form needs to be imposed in the tree component of the total biomass.

## 4 Discussion

In this paper we present a novel application of multidimensional P-spline mixed models to obtain biomass equations in the presence of more than one predictor variable and random effects. The additivity property is automatically satisfied if the same smoothing/variance parameters are used for all components of the tree, and this estimation approach involves no added difficulty. Furthermore, it provides us with the possibility of estimating a nonlinear relationship between the response and the explanatory variables using linear mixed model theory. In comparison with other procedures that satisfy the additivity property, such as non-linear regression, these models are easily implemented and provide good results, while the non-linear

models can produce singular covariance matrices that do not guarantee a unique solution (Goicoa et al., 2011). The only drawback of using penalized spline models to satisfy the additivity property is the fact that the same explanatory variables are considered for modelling all the components of the tree.

The approach presented in this paper is based on penalized regression, therefore, the models only need to specify the type of basis and penalties. Other possible options would be the use of wavelets with a ridge penalty, as described in Angelini (2003). This approach yields directly a mixed model representation, as it does, for example, the use of truncated power functions (Ruppert et al., 2003), in which there is no need for reparametrization. However, wavelets are more suitable for cases where the true signal in the data is not so smooth. Although the reparametrization of the model might seem inconvenient, it is a standard procedure implemented in statistical packages that deal with penalized regression model. Also, it does not only yields the mixed model representation, but also allows for an easy implementation of the necessary constraints when dealing with additive and/or factor-by-function terms in the model. Furthermore, all models used in the paper can be fitted in R (using the `lme` function, or `gam`, in the `mgcv` package), but also in other packages, such as SAS, by means of the `PROC MIXED` function.

In this study, the variables measured were height and diameter. These variables have a relatively high predictive power, diameter being the most common variable used to estimate tree biomass or its components, whereas height is often used to differentiate growth conditions at different sites and commonly serves as a basis for expressing site index for the purpose of forest management planning. Branch biomass generally displays greater variability than stem and total aboveground biomass; hence some authors suggest that other variables besides diameter and height should be included in the branch biomass model in order to improve its accuracy. These variables might be related to the architecture or geometry of the tree. However, the possible improvement in estimation accuracy could not justify the added cost involved in measuring such variables in poplar short rotation coppice.

The models used in this study comprise two steps in order to satisfy the additivity property. Firstly, estimates of the smoothing/variance parameters are obtained by fitting the P-spline model for the total aboveground biomass of the tree. This gives the best fit, and also puts greater emphasis on total aboveground biomass, as recommended by Cienciala et al. (2006). The P-spline model is then fitted for the tree components, setting the parameters to the value previously obtained. When the results of fitting different curves/surfaces to the data for each clone were compared,

highly significant differences were found between clones. Furthermore, the estimates of variance parameters in the models were reduced when clone-specific curves/surfaces were included in the model. Our results confirm the necessity to use clone-specific models to estimate biomass in short-rotation plantations, as previously stated by Telenius and Verwijst (1995). According to our data, the models differ even for genotypes belonging to the same taxonomic group such as Unal and USA 49-177 (*P.x generosa*) or AF2 and I214 (*P.x canadensis*) for example. The additive plus interactions model (Model 4) gave the best fit to the data, indicating that the interaction of diameter and height has a significant influence on biomass production, and that this interaction is different among clones. This result is consistent with the genetic variability found by Wu and Stettler (1996) in the height and diameter growth patterns of poplars.

Finally, the models presented are very flexible and can be easily extended to more general cases, for example, for discrete response variables by the use of generalized linear mixed models.

## Acknowledgments

The study was carried out within the framework of INIA funded project RTA2005-00182-C02-01 and RTA2008-00025-C02-00, co-financed with funds from FEDER. The work Maria Durban and Dae-Jin Lee was funded by the Spanish Ministry of Economy and Competitiveness grant MTM2014-52184-P. The research by Dae-Jin Lee was supported by the Basque Government through the BERC 360 2014-2017 and the Department of Education, Language Policy and Culture of the Basque Government IT-620-13 programs and Basque Government Industry Department under the ELKA-RTEK Program, and by the Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2013-0323.

## References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrof, B. and Csàki, editors, *Second International Symposium on Information*, pages 267–281, Akademia Kiadó, Budapest.
- Angelini, C., De Canditiis, D., Leblanc, F. (2003). Wavelet regression es-

- timation in nonparametric mixed effect models. *Journal of Multivariate Analysis*, 85: 267–291.
- Carvalho, J. P. and Parresol, B. (2003). Additivity in tree biomass components of Pyrenean oak (*Quercus pyrenaica* Willd) *For. Ecol. Manage.*, 179: 269-276.
- Christeson, L. (2010). Wood production potential in poplar plantations in Sweden. *Biomass and Bioenergy*, pages 1289–1299.
- Cienciala, E., Cerny, M., Tatarinov, F., Apltauer, J., and Exnerova, Z. (2006). Biomass functions applicable to Scots pine. *Trees - Structure and Function*, 20:483–495.
- Currie, I. D. and Durbán, M. (2002). Flexible smoothing with  $P$ -splines: A unified approach. *Statistical Modelling*, 2:333–349.
- Currie, I. D., Durbán, M., and Eilers, P. H. C. (2006). Generalized linear array models with applications to multidimensional smoothing. *J. R. Statist. Soc. B*, 68:1–22.
- Dickmann, D., J.G., I., Blake, T., Kosola, K., and Kort, J. (2001). Physiological ecology of poplars. In D.I., D., Isebrands, J., Eckenwalder, J., and Richardson, J., editors, *Poplar culture in North America*, pages 77–118. NRC Research Press.
- Dillen, S., Marron, M., Bastien, C., Ricciotti, L., Salani, F., Sabattic, M., Pinel, M., Rae, M., Taylor, G., and Ceulemans, R. (2007). Effects of environment and progeny on biomass estimations of five hybrid poplar families grown at three contrasting sites across Europe. *Forest Ecology and Management*, 250:12–23.
- Dowell, R., Gibbins, D., Rhoads, J., and Pallardy, S. (2009). Biomass production physiology and soil carbon dynamics in short-rotation-grown *populus deltoides* and *p. deltoides p. nigra* hybrids. *Forest Ecology and Management*, 257:134–142.
- Durbán, M., Hackett, C., McNicol, J. and Newton, A., Thomas, T., and Currie, I. D. (2003). The practical use of semi-parametric models in field trials. *Journal of Agric., Bio. and Env. Stats.*, 8:48–66.
- Durbán, M., Harezlak, J., Wand, M. P., and Carroll, R. J. (2005). Simple fitting of subject-specific curves for longitudinal data. *Statist. Med.*, 24:1153–1167.

- Eilers, P. H. C., Currie, I. D., and Durbán, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics and Data Analysis*, 50(1):61–76.
- Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with  $B$ -splines and penalties. *Stat. Sci.*, 11:89–121.
- Fox, J., Ades, P., and Bi, H. (2001). Stochastic structure and individual-tree growth models. *Forest Ecology and Management*, 154:261–273.
- Goicoa, T., Militino, A., and Ugarte, M. (2011). Modelling aboveground tree biomass while achieving the additivity property. *Environmental and Ecological Statistics*, 18:367–385.
- Guan, B., Kuo, S., Chang, T., and Hsu, C. S. (2006). Analyzing the effects of stand thinning on microclimates with semiparametric smoothing splines. *Canadian Journal of Forest Research*, 36:1641–1648.
- Guidi, W., Piccioni, E., Ginanni, M., and Bonari, E. (2008). Bark content estimation in poplar (*populus deltoides* l.) short-rotation coppice in central italy. *Biomass and Bioenergy*, 32:518–524.
- Jordan, L., Clark, A., Schimleck, L., Hall, D., and Daniels, R. (2008). Regional variation in wood specific gravity of planted loblolly pine in the United States. Canadian journal of forest research. *Canadian Journal of Forest Research*, 38:698–710.
- Lee, D.-J. (2010). *Smoothing mixed model for spatial and spatio-temporal data*. PhD thesis, Department of Statistics, Universidad Carlos III de Madrid, Spain.
- Lee, D.-J. and Durbán, M. (2011).  $P$ -spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling*, 1:49–69.
- Lee, D.-J., Durbán, M., and Eilers, P. H. C. (2013). Efficient two-dimensional smoothing with  $P$ -spline ANOVA mixed models and nested bases. *Computational Statistics & Data Analysis*, 61:22–37.
- Parresol, B. (1999). Assessing tree and stand biomass: a review with examples and critical comparisons. *Forest Science*, 13:573–793.
- Parresol, B. (2001). Additivity of nonlinear biomass equations. *Canadian Journal of Forest research*, 31:865–878.

- Patterson, H. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58:545–554.
- Pontailleur, J., Ceulemans, R., Guittet, J., and Mau, F. (1997). Linear and non-linear functions of volume index to estimate woody biomass in high density young poplar stands. *Annals of Forest Science*, 54:335–345.
- Rivas-Martínez, Díaz, T. E., Fernández-González, F., Izco, J., Loidi, J., Lousã, M., Penas, A. (2002). Vascular plant communities of Spain and Portugal. Addenda to the syntaxonomical checklist of 2001. *Itinera Geobotanica*, 15(1-2):5-922.
- Ruppert, D. and Wand, M.P. and Carroll, R.J. (2003). Semiparametric Regression. *Cambridge University Press*.
- Ruiz-Peinado, R., del Rio, M., and Montero, G. (2011). New models for estimating the carbon sink capacity of spanish softwood species. *Forest Systems*, 20:176–188.
- Telenius, B. and Verwijst, T. (1995). The influence of allometric variation, vertical biomass distribution and sampling procedure on biomass estimates in commercial short rotation forests. *Bioresource Technology*.
- Wood, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4):1025–1036.
- Wood, S. N., Scheipl, F. and Faraway, J. (2013) Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing*, 23(3):341–360.
- Wu, R. and Stettler, R. (1996). The genetic resolution of juvenile canopy structure and function in a three-generation pedigree of populus. *Trees*, 11:99–108.
- Wutzler T., Wirth C., Schumacher J. (2008). Generic biomass function for Common Beech (*Fagus sylvatica* L.) in central Europe: predictions and components of uncertainty. *Canadian Journal of Forest Research*, 38:1661–1675.
- Zabek, L. and Prescott, C. (2006). Biomass equations and carbon content of aboveground leafless biomass of hybrid poplar in coastal british columbia. *Forest Ecology and Management*, 233:291–302.

## List of Figures

1	Plot of weight (total, stem and branches) vs <i>diameter</i> and <i>height</i> . . . . .	24
2	Additive fits by model 2. . . . .	25
3	Smooth curves for <i>diameter</i> Model 3 vs Model 2 (clone 2000 verde). . . . .	26
4	Additive fits by model 3 for Unal and I214 clones. . . . .	27
5	Plot of predicted versus observed values for all models (AF2 clone). . . . .	28
6	Additive with interactions fits by model 4 for Unal and I214 clones. . . . .	29
7	Plot of predicted total, branch, and stem biomass for clone Unal . . . . .	30

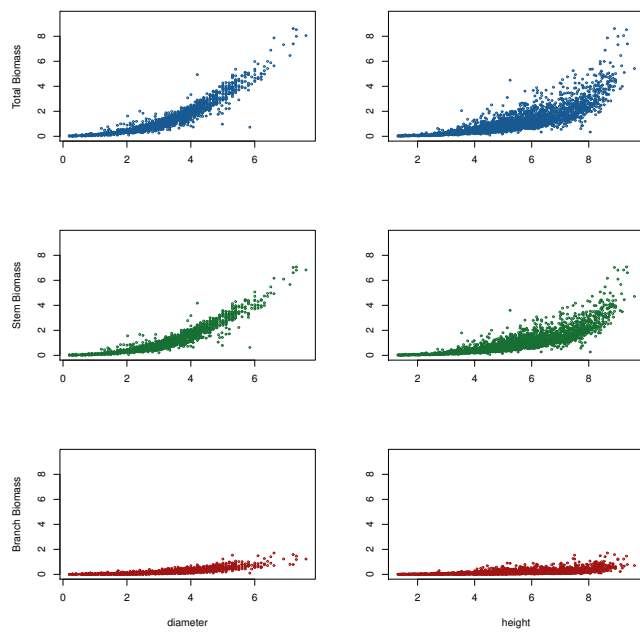
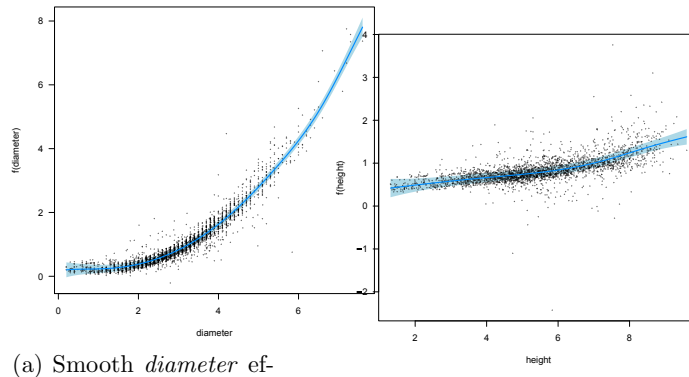


Figure 1: Plot of weight (total, stem and branches) versus *diameter* and *height*





(a) Smooth *diameter* effect.

(b) Smooth *height* effect.

Figure 2: Additive smooths terms for *diameter* and *height* by model 2.

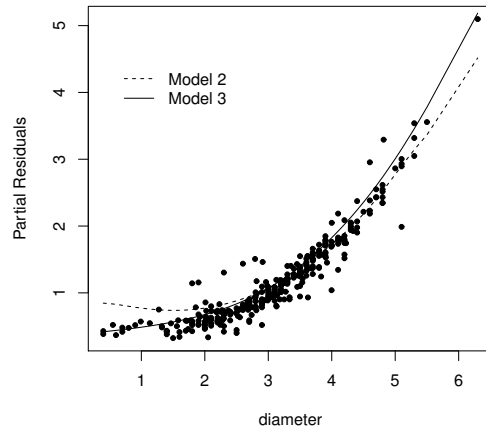


Figure 3: Fitted smooth curves for *diameter* using Model 3 (solid line) and Model 2 (dashed line) for clone 2000 verde.

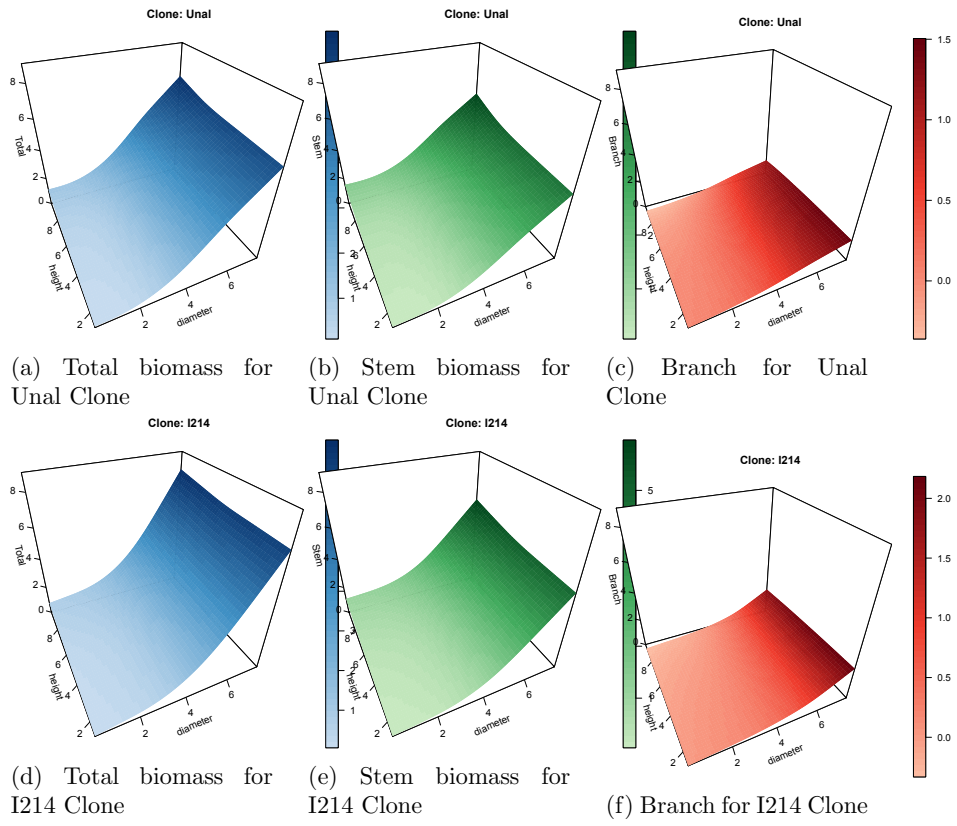


Figure 4: Fitted additive surface by model 3 for Unal (top) and I214 clones (bottom).

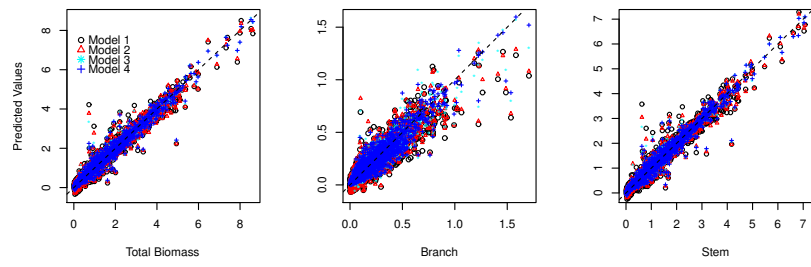


Figure 5: Plot of predicted versus observed values for all models (AF2 clone).

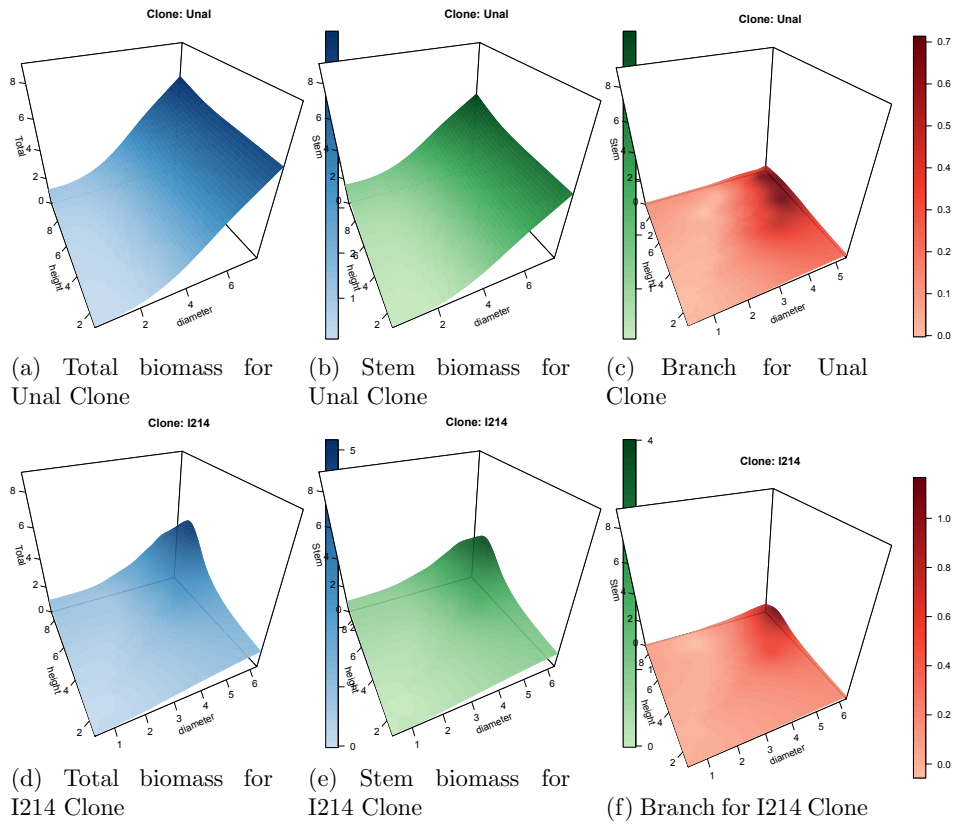


Figure 6: Fitted additive with interactions surface by model 4 for Unal (top) and I214 (bottom) clones.

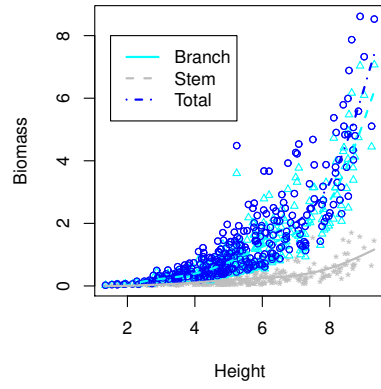


Figure 7: Plot of predicted total, branch, and stem biomass for clone Unal

## List of Tables

1	AIC, ED and adjusted $R^2$ model 1 to 4. . . . .	32
2	$ED_j$ for each smooth component, models 2 to 4. . . . .	33
3	NSUR vs P-splines results for slash pine biomass data . . . . .	34

	Total Biomass			Stem		Branches	
	AIC	ED	Adj. R <sup>2</sup>	AIC	Adj. R <sup>2</sup>	AIC	Adj. R <sup>2</sup>
Model 1	-395,2	37,5	94,1%	-350,7	93,6%	-299,1	75,8%
Model 2	-413,4	40,3	95,5%	-397,2	95,8%	-355,8	82,0%
Model 3	-678,1	116,6	96,5%	-615,3	96,4%	-572,2	85,8%
Model 4	-761,3	166,2	96,8%	-738,9	96,7%	-674,4	86,4%

Table 1: Values of Akaike Information Criterion (AIC), effective dimension (ED), and adjusted R<sup>2</sup> of models 1 to 4.



Smooth Component	Model 2	Model 3	Model 4
ED <sub>1</sub> (diameter)	14,67	55,44	28,11
ED <sub>2</sub> (height)	4,66	24,17	
ED <sub>3</sub> (diameter by height)			29,17
ED <sub>4</sub> (height by diameter)			27,87
ED <sub>5</sub> (diameter*height)			35,04

Table 2: Effective dimensions (ED<sub>*j*</sub>) associated with each smooth component of models 2 to 4. The interactions expressed with “*by*” correspond to the linear by non-linear interaction term in model 4.

Model	NSUR		P-spline	
	FI	RMSE	FI	RMSE
Wood	0,984	27,1	0,989	21,8
Bark	0,961	5	0,966	4,6
Crown	0,909	13	0,899	13,29
Tree	0,988	31,2	0,993	25,22

Table 3: Non-linear seemingly unrelated regression and P-spline mixed model results for the slash pine biomass data

## Supplementary Material

Smooth additive mixed models for predicting  
aboveground biomass

## Appendix A: Mixed model representation of penalized splines

The mixed model representation is motivated because it allows us to include smoothing in large class of models (from correlated data to longitudinal studies and survival analysis), and the use of the methodology and software already developed for mixed models for estimation and inference. Moreover, this representation allows the smoothing parameter to be determined as the ratio between the variance of the error terms and the variance of the random effects, that is,  $\lambda = \sigma^2/\sigma_\alpha^2$ . Both variance components can be estimated through REML procedure, and, therefore, it is not longer necessary to estimate  $\lambda$  via a cross-validation method or an information criterion. The reformulation of a P-spline into a mixed model can be viewed as a reparameterization of the original non-parametric model, for which we transform the model B-spline basis into a new model basis, i.e.,  $\mathbf{B} \rightarrow [\mathbf{X} : \mathbf{Z}]$ , and coefficients  $\boldsymbol{\theta} \rightarrow (\boldsymbol{\beta}, \boldsymbol{\alpha})'$ . Hence, this representation decomposes the fitted values as the sum of a polynomial (unpenalized) part  $(\mathbf{X}\boldsymbol{\beta})$  and a non-linear (penalized)  $(\mathbf{Z}\boldsymbol{\alpha})$  smooth term. There are several alternatives depending on the bases and the penalty used. Here, we use the B-spline basis and the penalty defined in section 2 to reparameterize the original model into a mixed model and, then, the aim is to find a transformation matrix  $\mathbf{T}$  to achieve this reparameterization.

We consider the singular value decomposition of the difference penalty matrix  $\mathbf{P} = \mathbf{D}'\mathbf{D}$ , that is:

$$\mathbf{D}'\mathbf{D} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}',$$

where  $\boldsymbol{\Sigma}$  is a diagonal matrix that contains the eigenvalues of  $\mathbf{D}'\mathbf{D}$  with  $d$  zero eigenvalues, and  $\mathbf{U}$  is the corresponding matrix of eigenvectors that can be decompose into two parts:  $\mathbf{U}_n$  of dimension  $c \times d$  containing the null-part eigenvectors and  $\mathbf{U}_s$  of dimension  $c \times (c - d)$  (where  $c$  is the rank of the basis and  $d$ , the order of the penalty) with non-null-part eigenvectors. Note that we can write  $\boldsymbol{\Sigma}$  as  $\boldsymbol{\Sigma} = \text{blockdiag}(\mathbf{0}_d, \tilde{\boldsymbol{\Sigma}})$ , where  $\tilde{\boldsymbol{\Sigma}}$  is a diagonal matrix that contains the non-zero eigenvalues of  $\mathbf{D}'\mathbf{D}$  and  $\mathbf{0}_d$  is a  $d \times d$  matrix of zeroes. Therefore, we can define a suitable transformation matrix  $\mathbf{T}$  as:

$$\mathbf{T} = [\mathbf{U}_n : \mathbf{U}_s \tilde{\boldsymbol{\Sigma}}^{-1/2}],$$

where the fixed and random effect matrices are  $\mathbf{X} = \mathbf{B}\mathbf{U}_n$ , and  $\mathbf{Z} = \mathbf{B}\mathbf{U}_s$ , respectively. Also, given this transformation matrix, the new coefficients are

$\boldsymbol{\beta} = \mathbf{U}'_n \boldsymbol{\theta}$  and  $\boldsymbol{\alpha} = \mathbf{U}'_s \tilde{\boldsymbol{\Sigma}}^{-1/2} \boldsymbol{\theta}$ . The fixed effect matrix  $\mathbf{X}$  may be replaced by any sub-matrix such that  $[\mathbf{X} : \mathbf{Z}]$  has full rank and  $\mathbf{X}'\mathbf{Z} = \mathbf{0}$  (that is,  $\mathbf{X}$  and  $\mathbf{Z}$  are orthogonal). This is possible since the fixed parameters are unpenalized. So, if we assume a second order penalty ( $d = 2$ ), the fixed effect matrix can be taken as  $\mathbf{X} = [\mathbf{1} : \mathbf{x}]$ , where  $\mathbf{1}$  is a vector of ones and  $\mathbf{x}$  is the explanatory variable. Also, the penalty term  $\boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta}$  becomes  $\boldsymbol{\alpha}'\mathbf{F}\boldsymbol{\alpha}$ , where  $\mathbf{F} = \lambda\mathbf{I}$ . This fact follows since  $\mathbf{T}$  is orthogonal and  $(\boldsymbol{\beta}, \boldsymbol{\alpha})' = \mathbf{T}'\boldsymbol{\theta}$ . Hence, given the new basis and the new penalty, the penalized sum of squares,

$$S(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{B}\boldsymbol{\theta})'\mathbf{W}(\mathbf{y} - \mathbf{B}\boldsymbol{\theta}) + \boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta},$$

becomes:

$$S(\boldsymbol{\beta}, \boldsymbol{\alpha}, \lambda) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha})'\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha}) + \sigma^2\lambda\boldsymbol{\alpha}'\mathbf{I}_{c-p}\boldsymbol{\alpha},$$

This corresponds to the likelihood of a linear mixed model with  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{W}^{-1})$  and  $\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{G})$  with  $\mathbf{G} = \sigma_\alpha^2\mathbf{I}_{c-p}$  and  $\lambda = \sigma^2/\sigma_\alpha^2$ . Differentiating the equation above with respect to  $\boldsymbol{\beta}$  and  $\boldsymbol{\alpha}$ , it is straightforward to obtain the standard mixed model equations given in section 2.

## Appendix B: Derivation of the *PS*-ANOVA mixed model formulation

In this Appendix, we show how to obtain the mixed model formulation of the model given by:

$$\mathbf{y} = f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) + g_1(\mathbf{x}_1)\mathbf{x}_2 + \mathbf{x}_1g_2(\mathbf{x}_2) + h(\mathbf{x}_1, \mathbf{x}_2) + \boldsymbol{\epsilon}, \quad \text{with } \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2). \quad (1)$$

This model is known as *PS*-ANOVA mixed model introduced in Lee et. al (2013), where the interaction between the covariates is decomposed as a sum of three components:

- $g_1(\mathbf{x}_1)\mathbf{x}_2$  is a *smooth-by-linear* interaction between  $\mathbf{x}_1$  and  $\mathbf{x}_2$ ,
- $\mathbf{x}_1g_2(\mathbf{x}_2)$  is a *linear-by-smooth* interaction between  $\mathbf{x}_1$  and  $\mathbf{x}_2$  and
- $h(\mathbf{x}_1, \mathbf{x}_2)$  is a *smooth-by-smooth* interaction between  $\mathbf{x}_1$  and  $\mathbf{x}_2$ .

where  $f_i$  and  $g_i$  are smooth functions of the covariates  $\mathbf{x}_i$  for  $i = 1, 2$  and  $h$  a smooth function for the interaction.

In order to show the mixed model formulation of model in Eq. (1), firstly, let us consider a bivariate smooth additive with interaction model given by:

$$\mathbf{y} = f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) + f_{1,2}(\mathbf{x}_1, \mathbf{x}_2) + \boldsymbol{\epsilon}, \quad \text{with } \boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2), \quad (2)$$

where  $f_1$  and  $f_2$  are additive smooth terms of  $\mathbf{x}_1$  and  $\mathbf{x}_2$  respectively and  $f_{1,2}(\mathbf{x}_1, \mathbf{x}_2)$  is a bivariate interaction term. Model in Eq. (2) was proposed by Gu (2002) in the context of smoothing splines as *Smoothing-splines ANOVA* (SS-ANOVA) models. In the context of low-rank smoothers such as  $P$ -splines, the model in Eq. (2) is constructed as follows:

Given  $\mathbb{E}[\mathbf{y}] = \mathbf{B}\boldsymbol{\theta}$ , where the  $B$ -spline regression basis  $\mathbf{B}$  is defined block-wise as:

$$\mathbf{B} = (\mathbf{B}_1 | \mathbf{B}_2 | \mathbf{B}_1 \square \mathbf{B}_2), \quad (3)$$

where each  $\mathbf{B}_i$  is a  $B$ -spline basis of dimension  $n \times c_i$ ,  $i = 1, 2$ . The symbol  $\square$  denotes the row-tensor product of two matrices. Hence, the size of  $\mathbf{B}$  is  $n \times (c_1 + c_2 + c_1 c_2)$ . The vector of regression coefficients  $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_{[1,2]})'$ , where  $\boldsymbol{\theta}_1$  and  $\boldsymbol{\theta}_2$  are the  $c_1 \times 1$  and  $c_2 \times 1$ , vector of coefficients for the main (additive) effects and  $\boldsymbol{\theta}_{[1,2]}$ , of length  $c_1 c_2 \times 1$ , for the interaction. Then, model in Eq. (2) can be written as:

$$\mathbb{E}[\mathbf{y}] = \mathbf{B}_1 \boldsymbol{\theta}_1 + \mathbf{B}_2 \boldsymbol{\theta}_2 + (\mathbf{B}_2 \square \mathbf{B}_1) \boldsymbol{\theta}_{[1,2]}.$$

The penalty would have a block-diagonal structure of the form:

$$\mathbf{P} = \begin{pmatrix} \lambda_1 \mathbf{D}'_1 \mathbf{D}_1 & & \\ & \lambda_2 \mathbf{D}'_2 \mathbf{D}_2 & \\ & & \lambda_3 \mathbf{D}'_2 \mathbf{D}_2 \otimes \mathbf{I}_{c_1} + \lambda_4 \mathbf{I}_{c_2} \otimes \mathbf{D}'_1 \mathbf{D}_1 \end{pmatrix}, \quad (4)$$

of dimension  $(c_1 + c_2 + c_1 c_2) \times (c_1 + c_2 + c_1 c_2)$ , where each block corresponds to the penalty over each of the coefficients of the model. The penalty matrix (4) includes one-dimensional penalties of the additive smooth terms with smoothing parameters  $\lambda_1$  and  $\lambda_2$ , and a two-dimensional penalty for the interaction term with  $\lambda_3$  and  $\lambda_4$ . However, the regression matrix (3) is not of full rank, (in fact  $\text{rank}(\mathbf{B}) = c_1 c_2$ ), so there are  $(c_1 + c_2)$  linearly dependent columns. In other words, some elements of the basis  $\mathbf{B}_1$  and  $\mathbf{B}_2$  and are included in the basis for the interaction  $\mathbf{B}_2 \square \mathbf{B}_1$ . Hence, model in Eq. (2) should be modified in order to preserve the identifiability. The identifiability problem is also reflected in the fact that the penalty matrix in (4) is rank deficient (i.e.  $\text{rank}(\mathbf{D}'_i \mathbf{D}_i) = c_i - \text{pord}$ , where  $\text{pord}$  is the order

of the penalty, usually  $pard = 2$ ). Hence, matrix  $\mathbf{P}$  in Eq. (4) has rank  $(c_1 + c_2 + c_1c_2 - 8)$ .

A simple and elegant way to avoid identifiability problems is to reformulate the model as a mixed model as shown in **Appendix A**. The mixed model representation of model (2) allows us to find out that some terms are repeated in the model matrices  $\mathbf{X}_i$  and  $\mathbf{Z}_i$ ,  $i = 1, 2$  and solve the identifiability problem. Consider a second order penalty  $pard = 2$ , for each  $\mathbf{B}_i$  there exists a transformation matrix  $\mathbf{T}_i = [\mathbf{T}_{in} | \mathbf{T}_{is}]$  such that  $\mathbf{B}_i \mathbf{T}_i = [\mathbf{B}_i \mathbf{T}_{in} | \mathbf{B}_i \mathbf{T}_{is}] = [\mathbf{X}_i | \mathbf{Z}_i]$  of size  $n \times c_i$ . for the interaction term  $\mathbf{B}_1 \square \mathbf{B}_2$  the transformation matrix is given by the tensor product of two transformation matrices  $\mathbf{T}_1 \otimes \mathbf{T}_2$ . For second order penalties,  $pard = 2$ , let be the  $\mathbf{T}_{in} = [\mathbf{u}_{in}^{(1)} : \mathbf{u}_{in}^{(2)}]$  are the null-part eigenvectors of  $\mathbf{D}'_i \mathbf{D}_i$  and is of size  $c_i \times 2$  and  $\mathbf{T}_{is} = [\mathbf{U}_{is}]$  of dimensions  $c_i \times (c_i - 2)$  contains the non-null part eigenvectors. Then, it is straightforward to see that by removing the column vector of  $\mathbf{1}$ 's in the fixed effects matrices, the identifiability problem is easily solved. Hence, the model matrices for model in Eq. (2) are:

$$\begin{aligned} \mathbf{X} &= [\mathbf{1}_n | \mathbf{x}_1 | \mathbf{x}_2 | \mathbf{x}_2 \square \mathbf{x}_1] \\ \mathbf{Z} &= [\mathbf{Z}_1 | \mathbf{Z}_2 | \mathbf{x}_2 \square \mathbf{Z}_1 | \mathbf{Z}_2 \square \mathbf{x}_1 | \mathbf{Z}_2 \square \mathbf{Z}_1] \end{aligned}$$

Removing the column vector of ones also implies to remove the corresponding column vector of the transformation matrix  $\mathbf{T}$ . Finally, the matrix  $\mathbf{T}$  required to obtain the fixed and random effects matrices in  $\mathbf{X}$  and  $\mathbf{Z}$  is:

$$\mathbf{T}_n = \begin{bmatrix} 1 & \cdots & & 0 \\ \vdots & \mathbf{u}_{1n}^{(2)} & & \\ & & \mathbf{u}_{2n}^{(2)} & \\ 0 & & & \mathbf{u}_{2n}^{(2)} \otimes \mathbf{u}_{1n}^{(2)} \end{bmatrix} \quad \text{and} \quad (5)$$

$$\mathbf{T}_s = \begin{bmatrix} 0 & \cdots & & \\ \mathbf{U}_{1s} & & & \\ \vdots & \mathbf{U}_{2s} & & \\ & & \mathbf{u}_{2n}^{(2)} \otimes \mathbf{U}_{1s} : \mathbf{U}_{2s} \otimes \mathbf{u}_{1n}^{(2)} : \mathbf{U}_{2s} \otimes \mathbf{U}_{1s} & \end{bmatrix}, \quad (6)$$

where  $\mathbf{u}_{in}^{(2)}$  are the second columns of  $\mathbf{U}_{in}$ , and  $\mathbf{U}_{in}$  and  $\mathbf{U}_{is}$  are the matrices of eigenvectors corresponding to the null and non-null eigenvalues of the singular value decompositions of matrices  $\mathbf{D}'_i \mathbf{D}_i$  for  $i = 1, 2$ .

Given this transformation, the mixed model penalty for model in Eq. (2) is the block-diagonal matrix defined by:

$$\mathbf{F} = \mathbf{T}'_s \mathbf{P} \mathbf{T}_s = \text{blockdiag}(\mathbf{F}_{(1)}, \mathbf{F}_{(2)}, \mathbf{F}_{(1,2)}), \quad (7)$$

where

$$\begin{aligned}\mathbf{F}_{(1)} &= \lambda_1 \tilde{\boldsymbol{\Sigma}}_1, \\ \mathbf{F}_{(2)} &= \lambda_2 \tilde{\boldsymbol{\Sigma}}_2, \text{ and} \\ \mathbf{F}_{(1,2)} &= \text{blockdiag}(\lambda_3 \tilde{\boldsymbol{\Sigma}}_1, \lambda_4 \tilde{\boldsymbol{\Sigma}}_2, \lambda_3 \mathbf{I}_{c_2-2} \otimes \tilde{\boldsymbol{\Sigma}}_1 + \lambda_4 \tilde{\boldsymbol{\Sigma}}_2 \otimes \mathbf{I}_{c_1-2}),\end{aligned}$$

and  $\tilde{\boldsymbol{\Sigma}}_i$  are the non-zero eigenvalues of the marginal penalties  $\mathbf{D}'_i \mathbf{D}_i$ ,  $i = 1, 2$ . Then, the covariance of the random effects is given by  $\sigma^2 \mathbf{F}^{-1}$ . Notice that, for  $\lambda_1 = \lambda_2$  and  $\lambda_3 = \lambda_4$ , the model is equivalent to the bivariate anisotropic tensor product smooth.

Looking at the composition of  $\mathbf{X}$  and  $\mathbf{Z}$  above, it is almost immediate to think of a model with 5 smooth perms instead of 3, one per component of those matrices. This is equivalent to model the interaction  $f_{[1,2]}(\mathbf{x}_1, \mathbf{x}_2)$  as:

$$f_{[1,2]}(\mathbf{x}_1, \mathbf{x}_2) = g_1(\mathbf{x}_1)\mathbf{x}_2 + \mathbf{x}_1 g_2(\mathbf{x}_2) + h(\mathbf{x}_1, \mathbf{x}_2). \quad (8)$$

This yields model (1), with associated B-spline basis:

$$\mathbf{B} = [\mathbf{B}_1 | \mathbf{B}_2 | \mathbf{B}_3 | \mathbf{B}_4 | \mathbf{B}_5], \quad (9)$$

where  $\mathbf{B}_1$  and  $\mathbf{B}_2$  are the marginal basis functions for diameter and height, and

$$\mathbf{B}_3 = \mathbf{B}_1 \square \mathbf{x}_2, \quad \mathbf{B}_4 = \mathbf{x}_1 \square \mathbf{B}_2, \text{ and } \mathbf{B}_5 = \mathbf{B}_1 \square \mathbf{B}_2,$$

The penalty matrix becomes:

$$\mathbf{P} = \text{blockdiag}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4, \mathbf{P}_5), \quad (10)$$

where,  $\mathbf{P}_i = \lambda_i \mathbf{D}'_i \mathbf{D}_i$ , ( $i = 1, 2$ ) (corresponding to the main effects),  $\mathbf{P}_3 = \lambda_3 \mathbf{D}'_1 \mathbf{D}_1$  and  $\mathbf{P}_4 = \lambda_4 \mathbf{D}'_2 \mathbf{D}_2$  penalize the varying coefficient terms  $g_1(\mathbf{x}_1)\mathbf{x}_2$  and  $\mathbf{x}_1 g_2(\mathbf{x}_2)$ , and  $\mathbf{P}_5 = \lambda_5 (\mathbf{D}'_1 \mathbf{D}_1 \otimes \mathbf{I}_{c_2} + \mathbf{I}_{c_1} \otimes \mathbf{D}'_2 \mathbf{D}_2)$  penalizes the interaction term (but with a single smoothing parameter).

In this case, the transformation matrices required are as follows:  $\mathbf{T}_n$  is given by (5), and  $\mathbf{T}_s$  is a modified version of (6) given by:

$$\mathbf{T}_s = \begin{bmatrix} 0 & & \cdots & & & & & & & & & \\ \mathbf{U}_{1s} \tilde{\boldsymbol{\Sigma}}_1^{-1/2} & & & & & & & & & & & \\ \vdots & & \mathbf{U}_{2s} \tilde{\boldsymbol{\Sigma}}_2^{-1/2} & & & & & & & & & \\ & & & & \mathbf{u}_{2n}^{(2)} \otimes \mathbf{U}_{1s} \tilde{\boldsymbol{\Sigma}}_1^{-1/2} & & & & & & & \\ & & & & & & \mathbf{U}_{2s} \tilde{\boldsymbol{\Sigma}}_2^{-1/2} \otimes \mathbf{u}_{1n}^{(2)} & & & & & \\ & & & & & & & & \mathbf{U}_{2s} \tilde{\boldsymbol{\Sigma}}_2^{-1/2} \otimes \mathbf{U}_{1s} \tilde{\boldsymbol{\Sigma}}_1^{-1/2} & & & \\ & & & & & & & & & & & \end{bmatrix}. \quad (11)$$



Again:

$$\mathbf{F} = \mathbf{T}'_s \mathbf{P} \mathbf{T}_s = \text{blockdiag}(\mathbf{F}_1, \dots, \mathbf{F}_5),$$

where

$$\begin{aligned} \mathbf{F}_1 &= \lambda_1 \mathbf{I}_{c_1 2}, & \mathbf{F}_2 &= \lambda_2 \mathbf{I}_{c_2 2}, \\ \mathbf{F}_3 &= \lambda_3 \mathbf{I}_{c_1 2}, & \mathbf{F}_4 &= \lambda_4 \mathbf{I}_{c_2 2} \\ \mathbf{F}_5 &= \lambda_5 \mathbf{I}_{(c_1-2)(c_2-2)} \end{aligned}$$

using the fact that  $\mathbf{G} = \sigma^2 \mathbf{F}^{-1}$ , we obtain that the vector of random effects  $\boldsymbol{\alpha} \sim N(\mathbf{O}, \mathbf{G})$  can be expressed as  $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_5)'$ , where  $\boldsymbol{\alpha}_k \sim N(\mathbf{O}, \mathbf{G}_k)$  with covariance matrices:

$\mathbf{G}_1 = \sigma_1^2 \mathbf{I}_{c_1-2}$ ,  $\mathbf{G}_2 = \sigma_2^2 \mathbf{I}_{c_2-2}$ ,  $\mathbf{G}_3 = \sigma_3^2 \mathbf{I}_{c_1-2}$ ,  $\mathbf{G}_4 = \sigma_4^2 \mathbf{I}_{c_2-2}$ , and  $\mathbf{G}_5 = \sigma_5^2 \mathbf{I}_{(c_1-2)(c_2-2)}$ , then,  $\mathbf{G} = \text{blockdiag}(\mathbf{G}_1, \dots, \mathbf{G}_5)$ .

## Appendix C: Comparison of models for each clone

Simulations results given in Lee et al. 2013 show that a model with an a more complex interaction is superior to a model using the function  $f_{1,2}$  (which is a sub-model of the previous one). In the paper, we showed that The analysis of the data confirm this result (looking at the AIC values). In order to reinforce this finding we fit the two models to the total biomass data to each clone separately:

- (a) *PS*-ANOVA model with one interaction term:

$$wt_{ijk} = f(d_{ijk}) + f(h_{ijk}) + t(d_{ijk}, h_{ijk}) + u_i + v_{ij} + \epsilon_{ijk}$$

- (b) *PS*-ANOVA model, with 3 interaction terms:

$$wt_{ijk} = f(d_{ijk}) + f(h_{ijk}) + g_1(d_{ijk})h_{ijk} + d_{ijk}g_2(h_{ijk}) + t(d_{ijk}, h_{ijk}) + u_i + v_{ij} + \epsilon_{ijk}$$

The following table gives the results for the fitting of each clone:

Model	Clone								
	AF2	Guardi	I214	MC	Monviso	Pegaso	Unal	USA	verde
(a)	-839.87	-978.62	-941.69	-752.11	-808.52	-1200.95	-915.30	-1012.49	-969.90
(b)	-903.75*	-978.73*	-948.26*	-753.25*	-832.40*	-1217.06 *	-934.80*	-1012.92*	-1016.71*

Table 1: Akaike information Criterion (AIC) for models (a) and (b) fitted for each clone. Symbol \* indicates the best model.

In all cases, again model (b) outperformed model (a).

Bellow, we give an example of the code needed to fit model (a) and (b) using `gamm` function in R for one of the clones:

```

# Model (a) (Clone="Unal")

model4Unal <- gamm(wt ~ gamm(wt ~ ti(diameter,m=2,bs="ps",k=10)+
  ti(height,m=2,bs="ps",k=10)+
  t(diameter,m=2,bs="ps",k=c(10,10)),
  random=list(site=~1, sitbl=~1),
  data = biomass[Clone=="Unal",],method="REML")

# Model (b) (Clone="Unal")

model5Unal <- gamm(wt ~ ti(diameter,m=2,bs="ps", k=10)+ti(height,m=2,bs="ps",k=10)+
  ti(diameter,by=height,m=2,bs="ps",k=10)+
  ti(height,by=diameter,m=2,bs="ps",k=10)+
  ti(diameter,height,bs="ps",m=2,k=c(10,10)),
  random=list(site=~1, sitbl=~1),
  data = biomass[Clone=="Unal",],method="REML")

```

wt is the total weight and site and sitbl, are variables corresponding to the site and block within site design. We have used B-spline basis of size 10 (k=10) and difference penalties (bs='ps'), of order 2 (m=2). All terms are specified using ti() to produce tensor product interactions, appropriate when the main effects (and any lower interactions) are also present. In this case we have not specified the smoothing parameter, which are selected using REML. However, if the additive property needs to be satisfied, the smoothing parameters would be the same for each component of the tree (as explained at the end of section 2). If the fit is done for all clones, a further argument by=Clone would be included in each smooth component of the formula.

## Bibliography

- Gu, C. (2002). Smoothing Spline ANOVA Models. Springer Series in Statistics, Springer.
- Lee, D.-J., Durbán, M., and Eilers, P. (2013). Efficient two-dimensional smoothing with P-spline anova mixed models and nested bases. Computational Statistics & Data Analysis, 61:22–37.