USING A NONPARAMETRIC METHOD TO DESCRIBE DIAMETER DISTRIBUTIONS OF BIRCH (*BETULA PUBESCENS* EHRH.) STANDS IN NORTHWEST SPAIN

**J. Javier Gorgoso-Varela**
Departamento de Biología de Organismos y Sistemas, Universidad de Oviedo, Escuela Politécnica de Mieres

**Alberto Rojo-Alboreca, Juan Gabriel Álvarez-González**
Unidade de Xestión Forestal Sostible, Departamento de Enxeñaría Agroforestal, Escola Politécnica Superior, Universidade de Santiago de Compostela - Lugo

**Abstract**

In this study, the nonparametric k-nearest neighbour method was used to describe diameter distributions of birch stands in Northwest Spain. It was applied using the following essential steps: (i) estimation of the distance between target stand and reference stands (kth-nearest neighbours); (ii) use of weight function defined by estimated distance, for each reference stand (to enable construction of a linear combination that defines the diameter distribution of target stand); and (iii) identification of the optimum number of neighbours that minimizes the error associated with statistics used to evaluate the results (bias, mean absolute error, mean square error and relative error index).

The best results were obtained by measuring the distance between distributions with a function defined by arithmetic mean diameter and quadratic mean diameter, with a weight function defined by inverse of the distance between distributions and with ideal number of neighbours equal to 8.

**Key words:** k-nearest neighbour, distance between plots, weight function

**INTRODUCTION**

Nonparametric methods can be used to describe diameter distributions. In this case, for a specific stand, distribution of the number of stems or basal area in diameter classes is determined as a linear combination of the known distributions in other stands. These reference distributions are either weighted according to a distance that measures the divergence between the target and known distributions, or they are assumed to be similar to distribution of the target stand. Although the methods are very flexible and enable description of multimodal
Several nonparametric methods are available for estimating diameter distributions or distributions of other stand variables (basal area, volume, biomass, etc.): Kernel estimation (Droessler, Burk, 1989); most similar neighbour inference \((k-MSN)\) (Moeur, Stage, 1995; Muinonen et al., 2001), which is a multivariate method used in forest inventories and for interpretation of digitalized aerial photographs whereby the distance function is obtained by canonical correlation; and the non-parametric grid method (Puimalainen, 1994), which is similar to the previous one, except that the reference stands are weighted in the same way and the number of reference stands varies.

The \(k\)-nearest neighbour regression \(k\)-NN is a nonparametric method used to estimate diameter distributions. It was first described as a nonparametric method for discriminant analysis (Cover, Hart, 1967), but has subsequently been used in numerous forest applications in Central and Northern Europe (Korhonen, Kangas, 1997; Haara et al., 1997; Maltamo, Kangas, 1998; Niggemeyer, Schmidt, 1999; Tommola et al., 1999; Sironen et al., 2003, etc.). The best known application of \(k\)-NN technique in the field of forestry research is probably for estimation of inventory results based on multi-source techniques developed by Tomppo (1993, 2006; Tomppo et al., 2011) and many others.

In \(k\)-NN method, the diameter distribution of a stand is estimated as a weighted linear combination of the known diameter distributions of a certain number of stands or plots \(k\) that are closest to the target plot. The closeness is defined as a distance that depends on the form of the distributions of the linear combination and is obtained by use of a parametric model. The form of the weight function is based on the measured distance, and the neighbours (i.e. \(k\), the reference diameter distributions) must be selected from a database of measured observations.

The main advantage of nonparametric methods is that they retain the entire range of variation in the data as well as the structure of the population covariance (Moeur, Stage, 1995). Furthermore, they do not produce unrealistic predictions and estimations are given from available data (Maltamo, Eerikäinen, 2001). The disadvantages are that they require a large amount of reference material, in this case diameter distributions of a large number of stands (although these will also serve as reference stands at the model application stage) (Maltamo, Eerikäinen, 2001), and there is no guarantee that the estimations will be unbiased (Altman, 1992).

Although nonparametric methods are less commonly used in forest modelling than those based on probably density functions (PDFs) or cumulative distribution functions (CDFs), the efficiency of parametric methods can be combined with the flexibility of nonparametric methods. Maltamo, Kangas (1998) compared the performance of methods based on the \(k\)-nearest
neighbour and of parametric methods based on Weibull function (predicting the parameters with regression models) for describing basal area distributions of *Pinus sylvestris* L. stands in Finland. The nonparametric methods proved more accurate than the parametric.

Other applications of nonparametric methods in forestry research include studies involving the development of an individual tree growth model (Sironen et al., 2003), estimation of stand characteristics for planning timber harvests (Malinen, 2003; Malinen et al., 2001; 2003), and estimation of individual tree biomass (Ferhmann, Kleinn, 2005).

The aim of the present study was to use the nonparametric k-nearest neighbour method to describe diameter distributions of birch (*Betula pubescens* Ehrh.) stands in Northwest Spain.

**MATERIAL AND METHODS**

**Collection of field data**

Diameter distributions of the stands under study were described using a method based on a single inventory of a network of 125 research plots. These plots were installed with the primary aim of obtaining data for constructing site quality curves and yield tables for the species in Galicia (NW Spain). The plots were selected as representative of the existing ages, densities and site qualities of forest stands of *B. pubescens* in Galicia. Plots with less than 90% standing birch trees were excluded, and therefore all plots represent monospecific stands. Other species that appear in mixed stands with birch in Galicia include *Alnus glutinosa* (L.) Gaertn., *Quercus robur* L., *Salix atrocinera* Brot. and *Frangula alnus* Mill.

Once the stands of interest were identified, inventory plots were established. The surface area of these rectangular plots ranged between 200 m$^2$ and 1000 m$^2$, depending on the stand density. In each plot, the diameter at breast height of all standing trees of diameter larger than 5 cm, was measured by taking two measurements at right angles to each other with a tree caliper (mm scale). Dominant trees were measured in each plot, and dominant height was calculated from the percentage of 100 thickest trees per ha (Assmann, 1970). Stand age was estimated by counting the growth rings in a series of samples (obtained with a Pressler drill) from the base of several trees and calculating the mean value.

The statistics that define the most representative variables of 125 inventory plots are shown in Table 1.

**The k-nearest neighbour method**

The k-nearest neighbour method is a widely used nonparametric method for density estimation and regression. The method was developed with the aim of adjusting the data bandwidth (window size) used for estimations at one point to the density of surrounding points. This is the main difference with respect to the
kernel-type smoother, defined as the weighted average of the response variables in a fixed environment for any point $x$, so that the range of the environment is determined by the value of smoothed parameter or bandwidth.

In the k-nearest neighbour method, the range of environments is varied on the basis of a distance separating point $x$ from the target sample that occupies the k-th place in order of distance from $x$. Application of this method to diameter distributions gives a distribution resulting from the linear combination of other nearby or similar distributions ($k$), the coefficients of which are the weights assigned to each distribution according to the distance separating the distributions.

Maltamo, Kangas (1998) proposed the following three steps for applying the k-nearest neighbour method to diameter distributions:

a) Definition of a distance function for selecting the most similar reference stands;

b) Identification of the number of nearest neighbours that should be used;

c) Calculation of a weight function for assigning weights to the reference stands.

**Definition of distance between distributions.** The success of application of nonparametric methods is mainly determined by the correct selection of the distance function. In previous studies, different criteria have been used to measure the distances that characterize the divergence between stands or between the diameter distributions of several stands.

Maltamo, Kangas (1998) measured the similarity between stands in order to predict the basal area distribution with a distance function that considered weighting the absolute differences between stand characteristics (age, basal area, quadratic mean diameter and height). These authors also mentioned the possibility of using other types of distance, such as Mahalanobis distance (Mahalanobis, 1936),

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{Variable} & \text{Mean} & \text{Maximum} & \text{Minimum} & \text{Stand. Dev.} \\
\hline
N \text{ (trees/ha)} & 1793.3 & 6000.0 & 390.0 & 1134.3 \\
\hline
\bar{t} \text{ (years)} & 30.1 & 60.0 & 12.0 & 9.4 \\
\hline
H_o \text{ (m)} & 15.0 & 23.8 & 7.2 & 3.6 \\
\hline
G \text{ (m$^2$/ha)} & 24.6 & 66.5 & 3.3 & 10.7 \\
\hline
\bar{d}_q \text{ (cm)} & 14.1 & 23.2 & 7.4 & 3.54 \\
\hline
\bar{d} \text{ (cm)} & 13.5 & 22.5 & 7.1 & 3.49 \\
\hline
\end{array}
\]

$N$: density; $\bar{t}$: age; $H_o$: dominant height; $G$: basal area, $d_q$: quadratic mean diameter, $\bar{d}$: arithmetic mean diameter.

Table 1
Descriptive statistics for main stand variables in 125 inventory plots
which is commonly used in the most similar neighbour inference \( k\text{-MSN} \) (Moeur, Stage, 1995), and Euclidean distance (Tokola et al., 1996, Tommola et al., 1999).

Furthermore, Niggemeyer, Schmidt (1999) used the so-called genetic distance, modified from the original defined by Gregorius (1974), to measure the divergence between diameter distributions. A variable called ‘distribution level’ can also be used (Malinen, 2003). Gorgoso (2003) measured the Euclidean distance between diameter distributions of the reference stands, understood as the distance between two vectors in the vectorial space, \( \mathbb{R}^n \), where \( n \) equals the number of diameter classes that exist once the ranges of all the distributions are equalled, and a value of zero was assigned to diameter classes for which there were no standing trees. Each vector that represents a distribution therefore comprises the number of trees in each diameter class; a value of zero was assigned to classes for which there were no trees.

In this study, we measured the divergence between two distributions as Euclidean distance between two vectors in the vectorial space \( \mathbb{R}^2 \), in which the coordinates of each vector are given by the quadratic mean diameter \( (d_g) \) and by the arithmetic mean diameter \( (\bar{d}) \) corresponding to each distribution. The distance between the distributions \( X = [d_{g1}, \bar{d}_1] \) and \( Y = [d_{g2}, \bar{d}_2] \) is therefore given by the following expression:

\[
d_{xy} = \sqrt{(d_{g1} - d_{g2})^2 + (\bar{d}_1 - \bar{d}_2)^2} \quad (1),
\]

where \( d_g \) is the quadratic mean diameter and \( \bar{d} \) is the arithmetic mean diameter.

This distance is considered a good indicator of the difference between distributions due to the relationship between diameter variance, quadratic mean diameter and arithmetic mean diameter: \( s^2 = d_g^2 - \bar{d}^2 \). Therefore, these two variables indirectly represent the first two moments of diameter distribution, which have been widely used to characterize diameter distributions using parametric approaches (Burk, Burhart, 1984; Shifley, Lentz, 1985; Knowe et al., 1997; Gorgoso et al., 2012).

If quadratic mean diameter is known, the arithmetic mean diameter can be derived from other stand variables according to Frazier (1981): \( \bar{d} = d_g - e^{Xb} \) where \( X \) is the vector of independent stand variables in a fixed instant and \( b \) is the vector of parameters to be estimated.

**Optimum number of neighbours.** The optimum number of neighbours or distributions that form part of the linear combination obtained as a result of the target distribution will be the number that gives the minimum values of the statis-
tics considered: bias, absolute mean error (MAE), mean square error (MSE) and relative error index (rEI).

Sensitivity analysis was used to determine the variation in these statistics on varying the number of neighbours from 2 to 15. According to other authors, this range of values is suitable for reference stands (Maltamo, Kangas, 1998; Maltamo, Eerikäinen, 2001).

**Assignation of weights to the reference stands.** Three possibilities were considered in relation to weighting the reference stands:

a) Not assigning any type of weighting;

b) Weighting with an inverse function of the distance that separates the distributions of the stands. The function is expressed as follows:

\[
W_{ij} = \frac{1}{1 + d_{ij}} \quad (2),
\]

where \(W_{ij}\) is the weight of reference stand \(i\) relative to target stand \(j\), and \(d_{ij}\) is the distance between the reference and target stand;

c) Weighting with an inverse function of the square of the distance that separates the distributions of the stands. In this case, the function is expressed as follows:

\[
W_{ij} = \frac{1}{1 + d_{ij}^2} \quad (3),
\]

where \(W_{ij}\) is the weight of reference stand \(i\) relative to target stand \(j\), and \(d_{ij}\) is the distance between the reference and target stand.

**Goodness of fit determinations.** Goodness of fit was evaluated by considering the bias, the mean absolute error (MAE) which avoids the final error of the results because uses the absolute values of the residuals, the usual mean square error (MSE) and the relative error index (rEI_{k}) for the number of neighbours \(k\) (Reynolds et al., 1988), which were calculated with the following expressions:

\[
Bias = \frac{\sum_{i=1}^{N} Y_i - \hat{Y}_i}{N} \quad (4)
\]

\[
MAE = \frac{\sum_{i=1}^{N} |Y_i - \hat{Y}_i|}{N} \quad (5)
\]

\[
MSE = \frac{\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}{N} \quad (6)
\]
where $Y_i$ is the observed value of relative frequencies of trees, $\hat{Y}_i$ is the theoretical value predicted by the method, $N$ is the total number of data, $w(x_i)$ is the volume of a tree of diameter class $i$, estimated using the equations proposed for this species in Galicia (Diéguez-Aranda et al., 2009) and $\hat{f}(x_i)$ and $f(x_i)$ are the predicted and observed relative frequency of diameter class $i$, respectively.

Bias, MAE and MSE were calculated for each case as the mean relative frequency of trees, for all diameter classes and all plots.

**RESULTS**

The arithmetic mean diameter was derived from other stand variables disaggregated from the yield tables with the following equation:

$$d = d_g - e^{(-2.125+0.066H_d+0.021t)} (8),$$

with $R^2 = 0.9409$, and where $d_g$ is the quadratic mean diameter, $H_d$ is the dominant height and $t$ is the age. The inclusion of age and dominant height in the equations indirectly takes into account the effect of site quality. The number of trees per hectare was also included in the model, but was not significant at the 5% level. The distance between sample plots, obtained using equation (1), ranged from 0 to 22.06 cm (mean value 5.68 cm). The mean values of bias, MAE, MSE, expressed as mean relative frequencies of the number of trees, and the rEI, obtained on varying the number of neighbours or reference stands between 2 and 15, are shown in Table 2. In this case, a weight function was not applied to the reference stands, and the target distribution was expressed as the mean of the distributions.

The same statistics obtained when the reference stands were weighted using the inverse of the distance separating the distributions (equation (2)) are shown in Table 3. Finally, the values of same statistics obtained when the inverse of the squared distance separating the distributions (equation (3)) was used as the weight function are shown in Table 4.

The variations in the mean values of bias and MSE in each diameter class when the number of neighbours or reference stands was equal to 8 and the inverse of the distance that separates the distributions was used as the weight function are shown in Fig. 1.

In view of the study findings, we concluded that the best option was to weight neighbouring or reference stands by the inverse of the distance that separates the target and reference stands (equation (2)), as in previous studies (Ma-
### Table 2
Mean values of relative frequencies for different number of neighbours when the reference stands were not weighted

<table>
<thead>
<tr>
<th>Number of neighbours</th>
<th>Bias</th>
<th>MAE</th>
<th>MSE</th>
<th>rEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.00111</td>
<td>0.01694</td>
<td>0.00072</td>
<td>216.50</td>
</tr>
<tr>
<td>3</td>
<td>0.00119</td>
<td>0.01570</td>
<td>0.00064</td>
<td>121.60</td>
</tr>
<tr>
<td>4</td>
<td>0.00116</td>
<td>0.01501</td>
<td>0.00061</td>
<td>73.48</td>
</tr>
<tr>
<td>5</td>
<td>0.00118</td>
<td>0.01463</td>
<td>0.00059</td>
<td>47.32</td>
</tr>
<tr>
<td>6</td>
<td>0.00120</td>
<td>0.01419</td>
<td>0.00057</td>
<td>17.64</td>
</tr>
<tr>
<td>7</td>
<td>0.00089</td>
<td>0.01376</td>
<td>0.00056</td>
<td>22.05</td>
</tr>
<tr>
<td>8</td>
<td>0.00062</td>
<td>0.01340</td>
<td>0.00055</td>
<td>22.33</td>
</tr>
<tr>
<td>9</td>
<td>0.00059</td>
<td>0.01333</td>
<td>0.00056</td>
<td>38.44</td>
</tr>
<tr>
<td>10</td>
<td>0.00061</td>
<td>0.01315</td>
<td>0.00055</td>
<td>31.05</td>
</tr>
<tr>
<td>11</td>
<td>0.00063</td>
<td>0.01311</td>
<td>0.00055</td>
<td>35.73</td>
</tr>
<tr>
<td>12</td>
<td>0.00059</td>
<td>0.01303</td>
<td>0.00056</td>
<td>48.50</td>
</tr>
<tr>
<td>13</td>
<td>0.00062</td>
<td>0.01296</td>
<td>0.00055</td>
<td>54.49</td>
</tr>
<tr>
<td>14</td>
<td>0.00061</td>
<td>0.01279</td>
<td>0.00055</td>
<td>59.99</td>
</tr>
<tr>
<td>15</td>
<td>0.00061</td>
<td>0.01268</td>
<td>0.00055</td>
<td>66.08</td>
</tr>
</tbody>
</table>

### Table 3
Mean values of relative frequencies for different number of neighbours when the reference stands were weighted by the inverse of the distance separating them

<table>
<thead>
<tr>
<th>Number of neighbours</th>
<th>Bias</th>
<th>MAE</th>
<th>MSE</th>
<th>rEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.00108</td>
<td>0.01692</td>
<td>0.00072</td>
<td>210.94</td>
</tr>
<tr>
<td>3</td>
<td>0.00114</td>
<td>0.01563</td>
<td>0.00064</td>
<td>111.83</td>
</tr>
<tr>
<td>4</td>
<td>0.00114</td>
<td>0.01493</td>
<td>0.00060</td>
<td>63.83</td>
</tr>
<tr>
<td>5</td>
<td>0.00116</td>
<td>0.01454</td>
<td>0.00058</td>
<td>36.11</td>
</tr>
<tr>
<td>6</td>
<td>0.00116</td>
<td>0.01411</td>
<td>0.00056</td>
<td>7.50</td>
</tr>
<tr>
<td>7</td>
<td>0.00084</td>
<td>0.01363</td>
<td>0.00055</td>
<td>4.70</td>
</tr>
<tr>
<td>8</td>
<td>0.00059</td>
<td>0.01325</td>
<td>0.00054</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>0.00055</td>
<td>0.01314</td>
<td>0.00055</td>
<td>8.90</td>
</tr>
<tr>
<td>10</td>
<td>0.00056</td>
<td>0.01297</td>
<td>0.00054</td>
<td>2.32</td>
</tr>
<tr>
<td>11</td>
<td>0.00057</td>
<td>0.01292</td>
<td>0.00054</td>
<td>6.14</td>
</tr>
<tr>
<td>12</td>
<td>0.00054</td>
<td>0.01282</td>
<td>0.00054</td>
<td>16.00</td>
</tr>
<tr>
<td>13</td>
<td>0.00056</td>
<td>0.01276</td>
<td>0.00054</td>
<td>20.37</td>
</tr>
<tr>
<td>14</td>
<td>0.00054</td>
<td>0.01259</td>
<td>0.00054</td>
<td>25.06</td>
</tr>
<tr>
<td>15</td>
<td>0.00054</td>
<td>0.01246</td>
<td>0.00053</td>
<td>27.64</td>
</tr>
</tbody>
</table>
This gave the lowest values of MSE, MAE and rEI, although the values were very similar to those obtained when no weight factor was applied. However, the lowest values of bias were obtained when the data were weighted by the inverse of the squared distance (equation (3)), and the values were very similar to those obtained when the inverse of the distance (equation (2)) was used as the weight function and the number of neighbours was 8 (the value considered most suitable in light of the results). Nonetheless, the highest values of MAE, MSE and rEI were obtained when equation (3) was used as the weight function, and they were even higher than those obtained when no weighting was applied. These statistics are more useful than bias for choosing the best method, as in bias estimation, errors of different signs may compensate each other and give a low mean value.

The mean value of bias remained stable until the number of neighbours reached 6, after which there was a strong decrease in the value of the statistic with 7-8 neighbours. The value remained more or less constant with larger numbers of references stands (Table 3).

**DISCUSSION**

The distance indicating the divergence between distributions was calculated in the present study by using equation (1), which was derived from the
variables quadratic mean diameter ($d^q$) and arithmetic mean diameter ($\bar{d}$). The advantage of using this distance is that the value of $d^q$ can be obtained directly from yield tables for birch in Galicia (Rojo et al., 2005) and the arithmetic mean diameter can be derived from other stand variables included in these tables (Frazier, 1981). Dominant height and age were used in present study. These variables have been widely used in modelling the diameter distributions of different species by parametric methods because they indirectly represent the first two moments of diameter distribution (Gorgoso et al., 2007; Gove, 2003).

The method of Niggemeyer, Schmidt (1999) relates the so-called genetic distance (Gregorius, 1974) between two stands to the difference between their quadratic mean diameters. We ruled out use of this method because of the lack of a good model fit to the available data (Gorgoso, 2003). We also ruled out the use of Euclidean distance across all diameter classes, proposed by Gorgoso (2003), because of the difficulty in relating this distance form to stand variables and because of the poor results obtained with this method.

The optimum number of neighbours was 8. The information provided by the k-nearest neighbour method increases with the number of reference stands used; however, if this number is too large, the smoothing effect of weighted mean can reduce the accuracy of estimates (Malinen, 2003). In other forestry studies that have applied nonparametric methods, the optimum number of neighbours ranged from 2 (or 3) to 10 (Maltamo, Kangas, 1998; Malinen, 2003).

MSE decreased until the number of neighbours reached 8 and stabilized thereafter (Table 3) and the minimum value of the relative error index was reached for 8 neighbours. Therefore, this number of neighbours (8) can be considered optimal for application of this nonparametric method to birch stands in NW Spain. With lower numbers of neighbours, the estimates were very similar to those obtained with the sampled data, and they were not very accurate. With higher numbers, the estimates will be smoothed and may be biased and less accurate (Altman, 1992). In general, the bias and the MSE are higher in the diameter class

![Fig. 1. Variation in the mean values of bias and mean square error (MSE) in each diameter class for 8 neighbours or reference stands weighted by the inverse of the distance that separates the distributions](image)
where the number of trees is bigger for 8 neighbours or reference stands weighted by the inverse of the distance that separates the distributions (Fig. 1).

Our findings are similar to those obtained in other studies that used comparable methods. For example, Malinen et al. (2001), who used the k-nn MSN method and a weight function defined as the inverse of the distance between distributions, found that the root mean square error (RMSE) decreased as the number of neighbours increased up to 5, but increased when the numbers of neighbours was greater than 5. Maltamo, Eerikäinen (2001) obtained similar results with a volume production model for *Pinus kesiya* Royle ex Gordon in Zambia: the value of the RMSE decreased considerably as the number of neighbours increased from one to four, and remained constant thereafter. These authors obtained low bias values, which were not affected by the number of neighbours.

Niggemeyer, Schmidt (1999), who used the k-nearest neighbour method, obtained the best results with 10 neighbours and without weighting the distance used. However, Tommola et al. (1999), who used the same method and Euclidean distance, concluded that the optimal number of neighbours was between 5 and 10. Maltamo, Kangas (1998) applied methods based on the k-nearest neighbour method to real reference stands, to distributions estimated by Weibull function, and they compared the results with those obtained by parametric methods involving use of Weibull function to predict the basal area diameter distribution in stands of *P. sylvestris*: the lowest values of the RMSE were obtained when 14 empirical distributions were used as reference stands. This number is much higher than the optimal number found in the previously mentioned and the present study. Sironen et al. (2003) used different nonparametric methods to elaborate an individual tree growth model and found that for the k-nearest method and a weight function that affected the growth estimates with more than 5 neighbours (the inverse of the distance between trees), the error decreased greatly on increasing the number of neighbouring trees from 1 to 12 and stabilized thereafter.

Diameter distribution models based on theoretical distribution functions such as beta, Weibull and Johnson’s *S* distributions, which are defined by function-specific parameters, are often used in Spain (Álvarez-González, 1997; Condés, 1997; Palahí et al., 2007; Gorgoso et al., 2012). Probability functions are particularly suitable for unimodal distributions, although they can also be used for aggregated distributions with more than one maximum (Condés, Martínez-Millán, 1997; Westphal et al., 2006). However, their generalized use may oversimplify the description of stand structure, especially in natural and/or mixed stands (Maltamo, Kangas, 1998).

The best results obtained with this nonparametric method for birch stands are similar to those obtained by Gorgoso et al. (2007) in terms of bias, MAE and MSE by applying the two parameter Weibull distribution to the same sample.
Although nonparametric methods are less commonly used in forest modelling than those based on probably density functions (PDFs) or cumulative distribution functions (CDFs), the efficiency of parametric methods can be combined with the flexibility of nonparametric methods.

CONCLUSIONS

The nonparametric k-nearest neighbour method can be used as an alternative to parametric methods (or can be combined with these) to estimate diameter distributions of birch stands in NW Spain, as the accuracy of both methods is similar.

For this type of stand, the distance that represents the divergence between the distributions should be determined from stand variables that are available in yield tables or that can be modelled from other variables included in these tables or in density management diagrams.

The function that gave the best results for weighting the reference stands or nearest neighbours is the function defined as the inverse of the distance separating the stands. In this case, the optimum number of stands or neighbours was 8.

Acknowledgements: This study was financed by Asociación Galega Monte-Industria through project 2000/CG488: Elaboración de modelos de crecimiento y producción para la gestión sostenible de las masas de abedul; and by Ministerio de Educación y Ciencia and ERDF within the Project AGL2007-66739-CO2-01/FOR: Modelos dinámicos de crecimiento y capacidad de secuestro de carbono de bosques de frondosas autóctonas del noroeste peninsular. The ‘Unidade de Xestión Forestal Sostible’ (UXFS) is funded by Consellería de Cultura, Educación e Ordenación Universitaria of Xunta de Galicia through the Programa de Consolidación e Estructuración de Unidades de Investigación Competitivas, modalidade de Grupos de Referencia Competitiva for the period 2014-2015.

REFERENCES


in South and East Africa. Faculty of Forestry, University of Joensuu, Finland. Res. Notes 97, 195-209.


E-mail: gorgoso@uniovi.es