Predicting Diameter at Breast Height from Stump Measurements of Removed Trees to Estimate Cuttings, Illegal Loggings and Natural Disturbances

Lucio Di Cosmo[1*], Patrizia Gasparini[1]

(1) Research Centre for Forestry and Wood, Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Piazza M. Nicolini 6, I-38123 Trento, Italy

* Correspondence: e-mails: lucio.dicosmo@crea.gov.it

ABSTRACT

Predicting diameter at breast height (DBH) of trees from stump information may be necessary to reconstruct silvicultural practices, to assess harvested timber and wood, or to estimate forest products' losses caused by illegal cuttings or natural disasters (disturbances). A model to predict DBH of felled trees was developed by the first Italian National Forest Inventory in 1985 (IFNI85). The model distinguished between the two broad groups of conifers and broadleaves and used stump diameter as the sole quantitative variable. Using an original dataset containing data from about 1200 trees of sixteen species recorded throughout Italy, we assessed the performance of that model. To improve the prediction of the DBH of removed trees over large areas and for multiple species, we developed new models using the same dataset. Performance of the new models was tested through indices computed on cross-validated data obtained through the leave-one-out method. A new model that performs better than the old one was finally selected. Compared to the old NFI model, the selected model improved DBH prediction for fourteen species up to 31.28%. This study proved that species specification and stump height are variables needed to improve the models' performance and suggested that data collection should be continued to get enhanced models, accurate for different ecological and stand conditions.

Keywords: tree diameter prediction; DBH; stump diameter; stump height; tree species

INTRODUCTION

Diameter at breast height (DBH) is one of the most important tree variables measured in forestry (Corral-Rivas et al. 2007, Bueno-López and Bevilacqua 2013). Its frequency distribution is used to describe stand structure (Pond and Froese 2015), but above all, it allows the computing of basal area, a variable highly correlated with tree volume, biomass and increment; it is also an attribute easy to measure with sufficient accuracy. In some cases, DBH of removed trees may also be of interest. Estimating DBH of trees from stump dimension may help to reconstruct silvicultural practices and to assess harvested timber and wood, when pre-operation data are not available. Pre-operation data are surely not available in case of illegal cuttings, and estimating DBH is the only way to assess the loss, through a two-step procedure

(Corral-Rivas et al. 2007, Özçelík et al. 2010), unless volume functions based on stump information are available. Another application of diameter prediction occurs when a harvest crosses ownership boundary (Chhetri et al. 1996, Pond and Froese 2014). The need to estimate DBH may also arise after natural disasters. In case of forest fires, stumps might be the only remaining part and consequently losses can be estimated based on their dimension. Other disturbances may lead to a variety of conditions that depend on type and intensity. For example, depending on the species, local windstorms might uproot trees still valuable for timber supply, which hold economic value and are for this reason still worth to be measured individually. On the contrary, hurricanes over wide regions may result in a mixture of a huge number of trees uprooted and broken, of different sizes and values that are too difficult and expensive to be measured individually,

also considering that exploitation might be urgent to free the roads and the watercourses as well as to avoid pests and wood deterioration. In such cases, estimates of losses are conducted with less accuracy and using stump dimension to DBH relationships after the removal is a possible method. The example of an event that did not allow the measuring of all downed trees is the recent Vaia storm, which occurred on 29th October 2018 and caused considerable damages in the southern side of the central and eastern Alps (Italy, Austria, Switzerland, Slovenia). In the North-East of Italy (Trentino Alto-Adige, Veneto, Friuli Venezia Giulia) it downed about eight million cubic meters of wood and left hundreds of kilometres of forest roads to be restored. In forest inventory surveys, stump measurements are carried out because they are connected to net increment calculation (Gschwantner et al. 2016).

The Italian National Forest Inventory (NFI) measures stump diameter of the trees cut in the last 12 months to estimate the current annual removals (Gasparini et al. 2010, Gasparini and Di Cosmo 2016). Whether the felling was carried out in the 12 months before the survey or earlier is generally assessed visually, based on stump conditions, and the volume and biomass of each felled tree are estimated through a two-step procedure. In the first two Italian NFIs the DBH of felled trees was predicted using a model developed by the first Italian NFI in 1985 (MAF-ISAFA 1988). The model is a linear equation, using stump diameter (SD) as the only predictive quantitative variable and with distinct regression parameters for the conifers and the broadleaves. To our knowledge, it is the only model developed in Italy for predicting DBH from stump information over wide areas, but, unfortunately, regression metadata was only partially reported. The model was built based on about 650 standing sample trees on which two right angle diameters were measured at stump height (SH) and at breast height. MAF-ISAFA (1988) did not report which species were measured, the regression goodness of fit and the reason why equations were developed separately for the two groups of species, e.g. if after an in-depth analysis, based on expert assessment or because of the limited number of observations by species. Concern about suitable use in any condition arises especially from the fact that sample trees were selected only in high forests, when stumps were found in the NFI plots; the sample might be unbalanced towards high trees and, given the longer rotation, also towards large diameters, potentially resulting in biased estimates, particularly in coppice stands.

The aim of this work was to assess the performance of the model developed by the first Italian NFI using a new dataset containing data from about 1200 trees collected throughout Italy and possibly provide a new, enhanced model for predicting DBH from stump measurements of removed trees. Especially, the new dataset contained information about the species and stump height, two variables not available or not used by the first Italian NFI.

MATERIALS AND METHODS

Data

The data used for this study came from the database created by Tabacchi et al. (2011) to construct tree volume and above-ground biomass models. Data collection started in 2001 thanks to a project launched to support the estimation procedure within the second Italian National Forest Inventory (INFC 2005). The project aimed to construct national models, applicable country-wide, and covering a relevant number of forest tree species. Based on 1289 sample trees measured over large areas of the country, prediction equations were finally made available for 26 species or groups of species (Tabacchi et al. 2011). Collection of sample trees has slowed down in recent years due to financial constrains; nevertheless, the overall goal of measuring 2800 trees in total has not been shelved and since 2011 the database has been enriched with about 150 additional sample trees. The data recorded include measurements of DBH, SD and SH. Details about the selection and number of sample trees, as well as about the measurements protocol, may be found in Tabacchi et al. (2011). Relevant to this article, the dataset includes trees from 11 out of 21 administrative regions of Italy and 58 species. A subset of the data came from the existing forest harvesting operations and consequently recorded on trees subjectively selected following the criterion of each operation. Other subsets came from the felling of trees undertaken for development operations (e.g. road construction, natural disasters) with no subjective selection due to silvicultural criteria. Finally, felling was commissioned specifically for the aim of the project, to fill the gaps in sampling for some species or diameter and/or height classes.

For this article, we decided to consider the species with at least forty sample trees. The available data are still limited and do not allow the analysis of the possible benefits from stratifying the sample by variables other than the species (e.g. management system, high trees vs coppice shoots). In this respect, while the analysis of the first NFI model's performance conducted in this study may be considered deep and complete, other models developed and tested should be better considered as preliminary results within a long-term research. Table 1 shows the number of sample trees used in this study, by species and species group, their DBH, SD and SH range and variation.

The DBH of each tree was measured before felling, with the first diameter measured from the uphill side in case of sloping terrain. Once the tree was felled, the diameter of the cut section and the stump height were measured. Diameters were measured with reference to two right-angled directions and expressed in centimetres with a millimetre approximation, while SH was measured in centimetres.

Analysis

Firstly, the dataset described in Table 1 was used to assess the performance of the model developed by the first Italian NFI (MAF-ISAFA 1988) (here named Model 0) and to re-calibrate it with the available data to obtain Model 1. Secondly, the data in Table 1 were used to study the relationship between DBH and stump dimensions through OLS regression analysis. A set of potential predictors was tested: SD, SH, 130/SH, 130-SH, and their transformed values (squares and/or natural logarithm). Both the species and the species group (conifers or broadleaves) were tested as dummy variables. During the analysis, results showed the existence of heteroscedasticity that could not be solved through feasible weighted least squares estimation. Therefore, we followed two alternative

Table 1. Number of sample trees (n.obs), diameter at breast height (DBH), stump diameter (SD) and stump height (SH) used in the study. For each quantitative variable, the table shows the minimum (Min), mean and maximum (Max) value recorded by species, as well as the standard deviation (StDev).

	Species		DBH (cm)				SD (cm)				SH (cm)			
No.	Name	n.obs	Min	Mean	Max	StDev	Min	Mean	Max	StDev	Min	Mean	Max	StDev
1	*Abies alba* Mill.	70	6	30.5	68	16.61	8	39.4	86	22.22	2	13.5	75	11.79
2	*Cupressus* spp.	44	5	17.7	38	8.45	8	22.7	46	9.60	1	10.1	29	6.31
3	*Larix decidua* Mill.	62	6	31.4	90	19.94	11	41.3	98	21.89	2	10.6	30	5.84
4	*Picea abies* K.	157	8	33.7	82	16.77	9	43.1	95	20.49	1	16.4	86	12.90
5	*Pinus cembra* L.	49	8	28.3	64	15.15	10	37.1	78	19.57	2	10.9	79	12.13
6	*Pinus halepensis* Mill.	43	5	22.8	56	12.03	8	28.7	59	13.21	1	7.6	15	3.67
7	*Pinus nigra* Arn.	70	9	27.6	64	12.34	11	33.4	74	13.58	1	9.5	22	4.66
8	*Pinus pinea* L.	43	13	31.3	61	10.89	18	39.9	73	11.59	4	10.0	33	4.57
9	*Pinus sylvestris* L.	53	8	24.4	59	12.51	11	30.0	64	13.87	3	9.6	28	6.12
	Total conifers	591	5	29.9	90	15.61	8	36.8	98	18.87	1	12.0	86	9.79
10	*Castanea sativa* Mill.	98	7	27.5	57	12.46	11	42.3	98	20.97	1	13.1	57	8.46
11	*Fagus sylvatica* L.	110	7	26.6	61	11.60	11	40.3	108	16.92	1	14.7	70	10.98
12	*Ostrya carpinifolia* Scop	64	6	17.6	32	6.60	7	26.9	51	10.27	1	10.4	64	8.26
13	*Quercus cerris* L.	68	8	28.5	50	10.38	15	42.6	72	15.38	3	11.5	43	7.94
14	*Quercus ilex* L.	83	5	20.7	55	12.02	7	29.6	84	18.65	1	8.3	22	3.96
15	*Quercus pubescens* Willd.	123	6	24.0	56	9.91	8	36.2	74	15.14	2	7.8	30	4.45
16	*Robinia pseudoacacia* L.	50	8	22.9	40	8.79	10	31.1	55	12.37	5	9.5	21	3.68
	Total broadleaves	596	5	24.3	61	11.16	7	36.3	108	17.35	1	10.9	70	7.88

approaches: i) evaluating the ongoing results also based on the heteroscedastic-robust standard errors (Models 2 and 3), and ii) running a logarithmic regression assuming the natural logarithm of DBH as the variable to regress (Models 4 and 5). In all cases, the quantitative explanatory variables were selected through a backward stepwise approach with tolerance ≤0.2 (Variance Inflation Factor ≤5). The best model within each approach was selected based on the following model parameters: adjusted coefficient of determination (R^2_{adj}), root mean squared error (RMSEmod), Akaike's information criterion (AIC), and Bayesian information criterion (BIC) computed on the fitting data.

The performance of Model 0 to predict the DBH of the trees in the dataset and that of the five newly developed models (Table 2) was compared by assessing their predictive accuracy through the following indices of fit: the mean error (ME), the mean absolute error (MAE), root mean squared error (RMSE), the mean percent deviation (MPD) and the mean percent standard error (MPSE) (Zeng, 2015). The indices were calculated as follows:

$$ME = \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i \tag{1}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{e}_i| \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{e}_i)^2}{n}} \tag{3}$$

$$MPD = 100 \left(\sum_{i=1}^{n} \hat{y}_{(i)} - \sum_{i=1}^{n} \hat{y}_i \right) \Big/ \sum_{i=1}^{n} \hat{y}_i \tag{4}$$

$$MPSE = \sum_{i=1}^{n} \left(100 |\hat{e}_i| \big/ \hat{y}_i \right) \Big/ n \tag{5}$$

where $\hat{e}_i = y_i - \hat{y}_i$; y_i are observed values, $\hat{y}_i$ are values estimated by the model.

For models 1 to 5, the indices were calculated using the leave-one-out cross-validation method. Performance assessment included the evaluation of the relationship between the observed values plotted vs the predicted ones. Specifically, we tested whether the intercept of the regression line equalled zero (i.e. the line passed through the axes origin) and whether its slope was 45° (the 1:1 slope condition), that is whether the predictions were unbiased.

In order to assess the models' performance on real scale values (centimetres) the Log-normal DBHs predicted by Models 3 and 4 in Table 2 were back-transformed following Sprugel (1983) before the cross validation.

The analyses were carried out using R version 3.5.3 (R Core Team 2019).

RESULTS

Table 2 shows the models developed, their residual standard error, Adjusted R-squared, the Akaike and Bayesian information criteria.
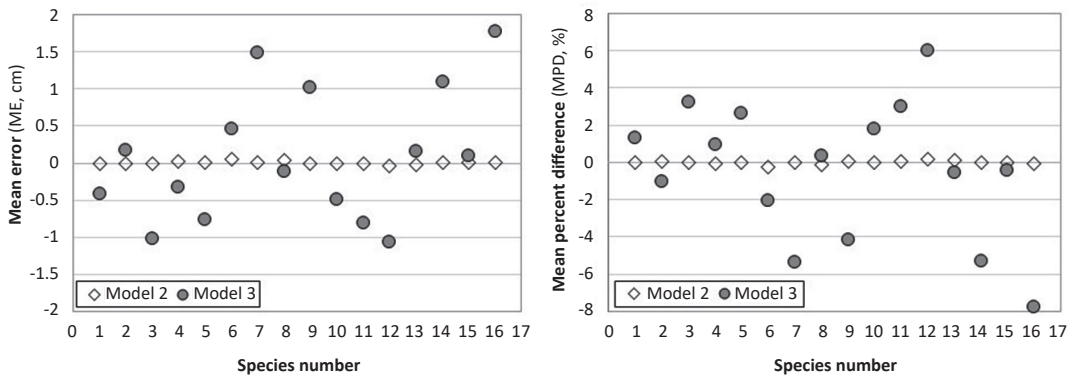
All the models provided a good level of fit as the adjusted *R*-squared value ranged between 0.93 and 0.95. The models that account for the species showed lower RSE compared to those that used the species group as a dummy variable. Table 3 shows the calculated performance indices of all the models tested and reports about the existence of bias when applying the model to independent data. Model 0 showed bias for both conifers and broadleaves. Model 1 showed that the re-calibration of Model 0 with the available data improved its performance consistently and the predictions for the two groups are unbiased; however, bias is still present in eight species. In general, model performance was improved by considering the species among the predictors. This is shown comparing the pairs of models with the same structure, that is, Model 2 vs Model 3 and Model 4 vs Model 5. For both pairs, RMSE was lower in eleven species and MPSE in nine species, out of sixteen. The importance of accounting for the species is also shown by the gains in the ME, the MAE and the MPD.

In Figure 1, ME and MPD for Models 2 and 3 are compared.

Based on the data in Table 3, similar patterns exist between Models 4 and 5. This shows that adopting the species as dummy variable instead of their two broad groups allowed the obtaining of more accurate estimates. MPD computed for Model 2 never exceeded 0.3 in absolute terms, while for Model 3 it exceeded 2% in nine species over sixteen. Based on the MPD, estimates of Model 3 were particularly poor for *Robinia pseudoacacia* (-7.8%), *Ostrya carpinifolia* (+6%), *Pinus nigra* (-5.4%), *Quercus ilex* (-5.3%), *Pinus sylvestris* (-4.1%). MPD of Model 2 is lower for all the species except *Pinus halepensis*, compared to the MPD of Model 3. Accounting for the species also reduced the presence of bias. Estimates of Model 5 are biased for ten species, while those obtained from Model 4 are biased only for three species. However, Model 2 is the only one that allowed unbiased estimates for all the species; its correspondent Model 3 was biased in six cases. Figure 2 shows the observed values plotted versus the estimated ones by Model 2, for four selected species, two conifers (*Abies alba* and *Larix decidua*) and two broadleaves, one deciduous (*Fagus sylvatica*) and one evergreen (*Quercus ilex*).

Table 2. The models developed and tested in the study. Variables: diameter at breast height (DBH), stump diameter (SD), stump height (SH); dummy variables: species (SP) or species group (SPgroup).

Model	Formula	RSE	R^2_{adj}	AIC	BIC
1	$DBH=(a_0+ b_1 \times SD) \times SPgroup + u_j$	3.723	0.93	6494.9	6520.3
2	$DBH=(a_0+ b_1 \times SD + b_2 \times SD^2 + b_3 \times SH) \times SPgroup + u_j$	3.306	0.94	6271.5	6601.7
3	$DBH=(a_0+ b_1 \times SD + b_2 \times SD^2 + b_3 \times SH) \times SPgroup + u_j$	3.563	0.93	6394.7	6440.4
4	$\ln(DBH)=a_0+b_1 \times \ln(SD)+b_2 \times (130-SH) \times SP+u_j$	0.122	0.95	-1586.2	-1337.3
5	$\ln(DBH)=a_0+b_1 \times \ln(SD)+b_2 \times (130-SH) \times SP+u_j$	0.129	0.94	-1480.6	-1445.0

Figure 1. Mean error (ME) and mean percent difference (MPD) computed on cross validated data for Model 2 and Model 3. The species number is shown in Table 1.

Table 3. Indices of fit computed on the complete dataset for Model 0 and on cross-validated data for the Models 1 to 5, by species. Crosses in the bias columns show cases when the line regressing the observed values plotted vs the predicted ones does not source form the axes origin, or its slope differs from 45°.

Species (N., name) \ Model	Mean error (ME, cm)						Mean absolute error (MAE, cm)						Root mean square error (RMSE, cm)					
	0	1	2	3	4	5	0	1	2	3	4	5	0	1	2	3	4	5
1 Abies alba Mill.	-1.455	-0.418	-0.001	-0.407	0.190	-0.690	3.434	2.856	2.762	2.825	2.741	2.985	5.415	4.465	4.035	4.368	4.134	4.626
2 Cupressus spp.	0.274	0.158	-0.007	0.181	0.083	0.228	1.455	1.596	1.533	1.503	1.612	1.591	1.905	1.955	1.931	1.883	2.050	1.949
3 Larix decidua Mill.	-2.331	-1.145	-0.006	-1.020	-0.145	-1.207	3.806	3.551	2.791	3.652	2.759	3.424	4.670	4.629	3.729	4.831	3.772	4.365
4 Picea abies K	-1.541	-0.239	0.028	-0.317	0.074	-0.623	3.147	2.802	2.889	2.842	2.856	2.909	4.431	3.874	3.993	3.947	3.942	4.078
5 Pinus cembra L	-1.685	-0.803	0.004	-0.755	0.098	-0.846	3.184	2.798	2.543	2.726	2.606	2.765	4.847	4.086	3.910	3.981	3.911	4.095
6 Pinus halepensis Mill	0.083	0.389	0.057	0.463	0.019	0.535	1.835	2.039	1.973	2.001	1.957	2.027	2.491	2.758	2.756	2.759	2.683	2.769
7 Pinus nigra Arn.	0.850	1.482	0.003	1.483	0.195	1.603	1.991	2.426	1.544	2.365	1.646	2.422	2.342	2.896	1.987	2.835	2.172	2.901
8 Pinus pinea L.	-1.162	-0.078	0.042	-0.115	0.093	0.032	2.424	2.342	2.151	2.342	2.052	2.256	3.038	3.031	2.758	3.019	2.664	2.914
9 Pinus sylvestris L.	0.596	0.995	-0.012	1.011	-0.025	1.083	1.938	2.036	1.874	2.051	1.966	2.032	2.552	2.891	2.536	2.882	2.616	2.835
10 Castanea sativa Mill.	-4.001	-0.455	-0.001	-0.490	0.159	-0.651	5.093	3.045	2.493	2.495	2.552	2.610	7.258	4.184	3.284	3.311	3.522	3.631
11 Fagus sylvatica L.	-3.352	-0.151	-0.014	-0.800	-0.052	-0.668	4.695	3.858	3.558	3.454	3.475	3.451	6.602	5.040	4.746	4.591	4.764	4.622
12 Ostrya carpinifolia Scop	-1.870	-1.000	-0.030	-1.063	0.013	-0.821	2.477	1.823	1.820	2.002	1.808	1.779	3.409	2.445	2.352	2.838	2.334	2.451
13 Quercus cerris L.	-3.214	0.389	-0.026	0.152	0.147	0.217	4.203	3.012	2.814	2.736	2.972	2.799	5.577	3.951	3.557	3.465	3.912	3.545
14 Quercus ilex L.	-0.935	0.416	0.006	1.096	0.095	0.795	3.132	2.523	2.388	2.501	2.448	2.441	4.689	3.525	3.353	3.388	3.619	3.467
15 Quercus pubescens Willd.	-2.694	-0.208	0.007	0.098	-0.031	0.067	3.622	2.365	2.392	2.361	2.384	2.333	4.835	3.242	3.358	3.318	3.358	3.271
16 Robinia pseudoacacia L.	0.163	1.768	0.015	1.776	0.068	1.954	2.100	2.120	1.914	2.171	1.925	2.178	2.582	3.074	2.475	2.976	2.462	3.113
All conifers	-0.870	0.000	0.013	0.003	0.070	-0.099	2.734	2.595	2.364	2.591	2.372	2.614	3.944	3.637	3.364	3.652	3.382	3.641
All broadleaves	-2.517	-0.002	-0.005	-0.003	0.047	-0.005	3.809	2.769	2.570	2.593	2.588	2.581	5.473	3.832	3.518	3.541	3.636	3.597
All dataset	-1.697	-0.001	0.004	0.000	0.059	-0.052	3.274	2.682	2.467	2.592	2.481	2.597	4.773	3.736	3.442	3.597	3.512	3.554

Species (N., name) \ Model	Mean percent difference (MPD, %)						Mean percent standard error (MPSE, %)						Bias					
	0	1	2	3	4	5	0	1	2	3	4	5	0	1	2	3	4	5
1 Abies alba Mill.	4.765	1.369	0.005	1.332	-0.618	2.259	11.594	9.491	9.418	9.850	9.182	9.684	×	×	-	×	×	×
2 Cupressus spp.	-1.553	-0.893	0.040	-1.024	-0.469	-1.293	9.893	10.310	9.822	9.714	10.467	10.315	-	-	-	-	-	-
3 Larix decidua Mill.	7.430	3.649	0.018	3.251	0.465	3.849	12.437	12.117	10.736	12.250	10.139	11.862	×	×	-	×	-	×
4 Picea abies K	4.573	0.708	-0.083	0.940	-0.220	1.848	9.268	8.319	8.773	8.462	8.715	8.426	×	-	-	-	-	-
5 Pinus cembra L	5.949	2.833	-0.013	2.667	-0.346	2.986	10.842	9.480	9.673	9.452	9.687	9.217	×	-	-	-	-	×
6 Pinus halepensis Mill	-0.364	-1.709	-0.252	-2.032	-0.082	-2.350	8.563	9.102	9.135	8.830	8.946	9.084	-	×	-	×	-	×
7 Pinus nigra Arn.	-3.079	-5.369	-0.011	-5.374	-0.701	-5.809	8.457	9.426	6.907	9.236	6.827	9.447	-	×	-	×	-	×
8 Pinus pinea L.	3.711	0.249	-0.134	0.369	-0.297	-0.102	7.939	7.711	7.657	7.694	7.485	7.468	-	×	-	×	×	×
9 Pinus sylvestris L.	-2.441	-4.074	0.048	-4.140	0.101	-4.436	9.614	8.689	7.633	8.980	8.288	8.709	-	×	-	×	-	×
10 Castanea sativa Mill.	14.555	1.655	0.004	1.781	-0.574	2.368	14.207	10.950	9.498	8.945	9.300	9.247	×	×	-	×	-	×
11 Fagus sylvatica L.	12.618	0.570	0.054	3.011	0.197	2.516	14.543	14.039	14.001	12.525	13.845	12.512	×	-	-	-	-	-
12 Ostrya carpinifolia Scop	10.620	5.676	0.170	6.037	-0.071	4.661	12.504	10.303	10.794	11.066	11.385	9.965	×	-	-	-	-	-
13 Quercus cerris L.	11.283	-1.367	0.091	-0.535	-0.514	-0.762	12.702	10.686	9.910	9.610	10.518	9.922	×	-	-	-	-	-
14 Quercus ilex L.	4.518	-2.012	-0.028	-5.297	-0.459	-3.842	14.643	12.485	11.567	13.975	11.253	11.769	×	-	-	×	-	-
15 Quercus pubescens Willd.	11.223	0.868	-0.030	-0.408	0.130	-0.279	13.458	10.310	10.414	10.442	10.749	10.184	×	-	-	-	-	-
16 Robinia pseudoacacia L.	-0.711	-7.719	-0.066	-7.758	-0.295	-8.531	9.855	9.732	8.463	10.960	8.253	10.400	×	×	-	-	-	×
All conifers	3.008	0.001	-0.045	-0.011	-0.243	0.341	9.845	9.405	8.8615	9.385	8.860	9.357	×	-	-	-	×	×
All broadleaves	10.451	0.110	0.101	0.120	-0.284	0.128	13.130	11.215	10.664	11.075	10.757	10.571	×	-	-	-	×	×
All dataset	6.377	0.005	-0.014	0.000	-0.220	0.195	11.655	10.343	9.872	10.202	9.895	9.959	×	-	-	-	-	×
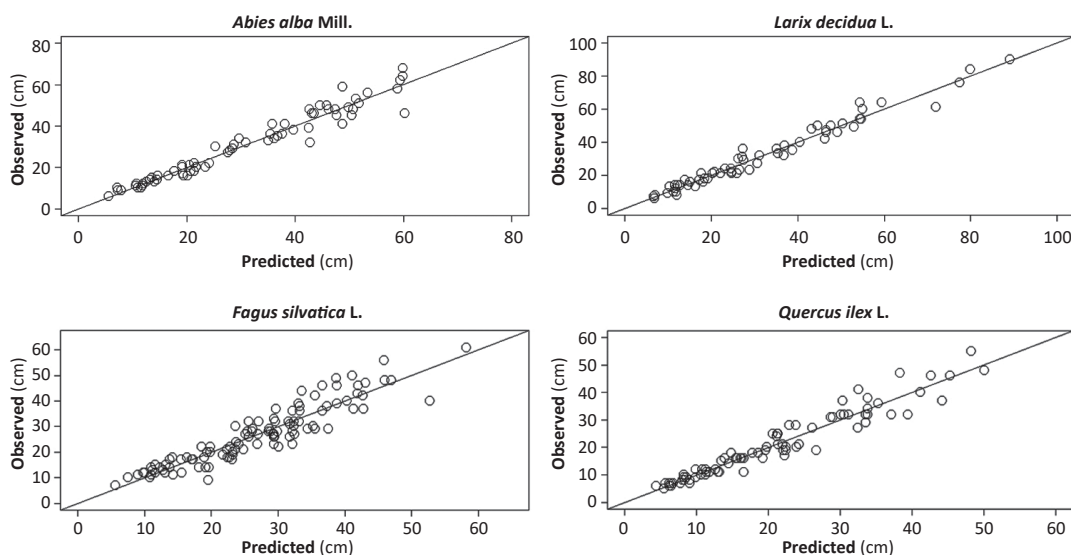
Figure 2. Examples of observed DBH values plotted against cross-validated values predicted by Model 1. Regression lines intercept does not differ from 0 and the slope does not differ from line 1:1.

RMSEs of Model 2 are equal or lower than those of Model 4 in eleven cases out of sixteen. Improvement ranged between 0% (*Quercus pubescens*) and 9.08% (*Quercus cerris*); in the remaining five cases, Model 4 showed RMSE reduced by from 0.51% (*Robinia pseudoacacia*) to 3.51% (*Pinus pinea*).

Table 4 provides the estimated parameters for Model 2 with their heteroscedastic-robust standard errors. For the

Table 4. Model 2: explanatory variables, estimated coefficients (Estimate), heteroscedastic-robust standard error (St. Error), *t*-value and *P*-value of the *t*-test.

Variable	Estimate	St.Error	*t*-value	Pr(>|t|)
(Intercept)	-2.16952	1.71786	-1.26292	0.206881
SD	0.93400	0.12200	7.65607	0.000000
SD^2	-0.00258	0.00150	-1.71868	0.085948
SH	0.08761	0.03796	2.30786	0.021188
SD x *Larix decidua*	-0.38716	0.14286	-2.71009	0.006829
SD x *Pinus silvestris*	-0.38447	0.17495	-2.19759	0.028182
SD x *Quercus pubescens*	-0.30830	0.15629	-1.97262	0.048784
SD^2 x *Larix decidua*	0.00608	0.00172	3.53088	0.000431
SD^2 x *Pinus halepensis*	0.00669	0.00288	2.31904	0.020572
SD^2 x *Pinus pinea*	0.00466	0.00217	2.14821	0.031911
SD^2 x *Pinus silvestris*	0.00739	0.00252	2.93199	0.003436
SH x *Picea abies*	-0.10348	0.04914	-2.10587	0.035437
SH x *Ostrya carpinifolia*	-0.09178	0.04332	-2.11838	0.034362
SH x *Castanea sativa*	0.16570	0.05426	3.05365	0.002314
SH x *Quercus cerris*	0.17576	0.06493	2.70670	0.006899
SH x *Pinus pinea*	0.17293	0.08152	2.12125	0.034119
SH x *Pinus nigra*	0.13766	0.06720	2.04851	0.040742

dummy variables only the coefficients significantly different from zero are showed. Complete coefficient estimates are given in Supplementary File 1.

The estimated parameters in Table 4 show that SD is the main explanatory variable. For example, applying the model with the parameters in the table first to the mean SD and SH values reported in Table 1 and then to SD and SH values increased in turn by 1 cm revealed that the higher relative importance of SH to SD is for *Quercus cerris*. Nevertheless, the 1 cm increase in SD enlarged the DBH estimate by 0.71 cm, while the 1 cm increase in SH enlarged the DBH estimate by 0.26 cm.
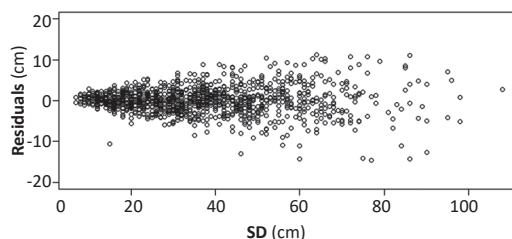
Figure 3 shows the residuals of Model 2 plotted against the stump diameters. Heteroscedasticity was confirmed by the Breush-Pagan test (BP = 267.29, df = 63, *p*-value < 2.2e-16).

Compared to Model 1, Model 2 RMSE is lower in fourteen species; the RMSE is reduced consistently for most species, up to 21.49% for *Castanea sativa* and 31.38% for *Pinus nigra*. In two cases (*Picea abies* and *Quercus pubescens*) Model 1 allowed estimates with reduced RMSE by from 3.07% to 3.58%.

In Table 5, the estimated parameters for Model 3 are given.

Figure 3. Residuals (cm) of Model 2 plotted against stump diameter (SD, cm).

Table 5. Model 3: explanatory variables, estimated coefficients (Estimate), heteroscedastic-robust standard errors (St. Error), *t*-value and *P*-value of the *t*-test.

Variable	Estimate	St.Error	*t*-value	Pr(>|t|)
(Intercept)	-1.51407	0.74383	-2.03550	0.042024
SD	0.83716	0.04916	17.03073	0.000000
SD^2	-0.00044	0.00067	-0.65952	0.509692
SH	0.02876	0.01651	1.74223	0.081730
Broadleaves	-0.49377	0.90772	-0.54397	0.586568
SD x Broadleaves	-0.05889	0.05932	-0.99275	0.321038
SD^2 x Broadleaves	-0.00176	0.00079	-2.23831	0.025387
SH x Broadleaves	0.12160	0.03475	3.49952	0.000483

DISCUSSION

DBH prediction is rarely the ultimate goal of an estimation procedure. Particularly when the economic value of the loss of forest products is under estimation, caused by illegal logging or natural hazards, estimates are addressed to assess tree volume or biomass weight. In this respect, estimating accurately the DBH from stump information is a matter of no secondary importance because the relationship between DBH and those two variables is multiplicative (Pond and Froese 2014), and errors in predicting DBH result in amplified errors in volume and weight predictions. Westfall and McRoberts (2017) found that, within a sampling approach, a two-step procedure based first on reconstructing DBH and then on applying a tree-volume model based on it gave better estimates than a model estimating merchantable volume directly from stumps information, suggesting that DBH prediction is unavoidable also when stump information to tree volume models is available. Despite its importance, Westfall (2010) noticed that little work on the prediction of DBH from stump dimensions has been done since the 1990s; our impression is that since that period such kind of works have become even less frequent. One reason might be the availability of such models that were developed in the past, but are now used without caring about their characteristics. Based on the new data available we could assess the performance of a model developed in the 1980s, based on data collected across wide areas but poorly described. The analysis was important because it showed the accuracy one might expect using that model. Its parametrization with the new data improved the accuracy on cross-validated data and provided overall unbiased predictions, as well as unbiased predictions for the conifers and the broadleaves groups, but also showed that predictions might be biased if used for particular species. In this respect, although the new Model 2 developed in the study is a preliminary result within a long-term research, it allows improved predictions compared to the old model. The main gain from using Model 2 is to obtain unbiased predictions, a result of the flexibility given to the model structure by the increased number of dummies to take into account the species; however, in terms of accuracy the improvement was limited, as indicated by the RMSE values in Table 3. The accuracy of Model 2

predictions is low, compared to other studies (e.g. Pond and Froese 2014), suggesting that dataset enrichment is necessary to identify the strata in order to reduce variability and optimize the sample size. In fact, almost twice the data available for Model 0 were used, but the number of samples by species is still limited and exceeds one-hundred for only three of them (*Picea abies, Fagus sylvatica, Quercus pubescens*). Furthermore, our data introduced additional variability because they come from trees felled following the criteria that differed between the operations and, for the broadleaves, include coppice shoots too. Like in the old dataset, our data capture a great variety of environmental and silvicultural conditions in Italy. The country spans over 10° latitude, two biogeographic regions (Pignatti 1979), and trees grow from the Mediterranean vegetation at the sea level to the tree line both on the Apennines and in the Alps. Also, the measuring of individual stumps instead of groups in sampling plots is a source of variability because sampling stumps close to each other in sample plots could capture at least the effects of homogeneous growing conditions at the very local scale. Ercanli et al. (2014) found that even selecting a subset of the available dataset may improve calibration to get better estimates of DBHs of independent data, highlighting the importance of the sampling scenario and of having enough data to run a deep analysis. Besides the variability in the data, accuracy performance might have been affected by the modelling approach. An OLS regression analysis was performed because it was functional to investigate the importance of considering the species and, in such case, whether individually or grouped through nested models; modelling by species or species groups is common in this kind of analysis. Either non-linear models or species-specific models might have improved accuracy. For practical uses, there are some other reasons for recommending using Model 2 (or Model 3) instead of Model 0. Metadata have been clearly described and this provides awareness of its limits and qualities. The dataset used in the study contained data on stumps from 7 cm diameter, also recorded in coppice stands; within the NFI computations, that threshold is consistent with the reference used for the current NFI measurements, that is, a minimum DBH of 4.5 cm for calipered trees (Bosela et al. 2016). The model presented in this paper was developed for stumps with bark, so its use is recommended only for intact or weakly/medium decayed stumps.

In addition to providing a new model useful for forest practice, this study provided some directions about the continuation of the research. Model 2 allows the estimation of DBH for sixteen species which are among the most important forest species in Italy. Those species mark the Italian forest categories or sub-categories, as adopted by the second Italian NFI, and amount to a relevant percentage of the country forest area (INFC 2007). Nevertheless, there are other important tree species not covered by the model, and this is an additional reason for underlying the importance of pursing the continuation of data collection. This is a matter of no secondary importance for practical implications. Corral-Rivas et al. (2007) demonstrated that when a species-based model is incorrectly used for estimating the DBH of another species, errors may be severe and the use of more general models, i.e. fitted with multi-species data, is recommended

instead. Although the good performance for one species that was not in the dataset used for calibration should be proved, Model 3 might be used to predict DBH of species other than those used in our analysis. In fact, choosing a species-specific model developed for species other than the one in hand might result as even more unjustified, because it requires the particularly hard judging of similarities to disclose for relationships like those implicated with stem taper. More precisely, Model 3 might be preferred in case of doubts about the stump species in mixed forests composed by the species in our dataset and for this reason it may be considered an additional result of this study. SH is a variable not recorded during the first Italian NFI but recorded from the second Italian NFI onwards (Gasparini and Di Cosmo 2015). Although SD remained the main explanatory variable, our analysis showed that, besides the species, SH improves the predictive accuracy of models. This confirms the finding of Pond and Froese (2014), who compared models with and without SH and stressed about the importance of including such a variable among the predictors.

CONCLUSIONS

Based on the analysis carried out, the following conclusions can be drawn:
- Prediction of tree diameter at breast height (DBH) is improved when the stump height (SH) variable is considered among the quantitative explanatory variables, although stump diameter (SD) remains the most important predictive variable.
- Introducing the species among the explanatory variables allowed the model to perform unbiased estimates.
- Model accuracy was limited by the sample size, suggesting importance to collect additional data to accomplish the study.
- Future steps of the study should be addressed to the identification of statistical strata to direct new data acquiring and inclusion in the modelling.
- Comparison of model performance consistently benefitted from the assessment of possible bias in predictions.

Author Contributions

Both authors contributed equally to this manuscript.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Supplementary Materials

Supplementary File 1. Estimated parameters for Model 2.

REFERENCES

Bosela M, Gasparini P, Di Cosmo L, Parisse B, De Natale F, Esposito S, Scheer L, 2016. Evaluating the potential of an individual-tree sampling strategy for dendroecological investigations using the Italian National Forest Inventory data. *Dendrochronologia* 38: 90-97. https://doi.org/10.1016/j.dendro.2016.03.011.

Bueno-López S, Bevilacqua E, 2013. Diameter growth prediction for individual *Pinus occidentalis* Sw. trees. *iForest* 6: 209-203. https://doi.org/10.3832/ifor0843-006.

Chhetri DBK, Fowler GW, 1996. Estimating diameter at breast height and basal diameter of trees from stump measurements in Nepal's lower temperate broad-leaved forests. *Forest Ecol Manag* 81(1-3): 75-84. https://doi.org/10.1016/0378-1127(95)03663-6.

Corral-Rivas JJ, Barrio-Anta M, Aguirre-Calderón Oa, Diéguez-Aranda U, 2007. Use of stump diameter to estimate diameter at breast height and tree volume for major pine species in El Salto, Durango (Mexico). *Forestry* 80(1): 29-40. https://doi.org/10.1093/forestry/cpl048.

Ercanli I, Gunlu A, Başkent EZ, 2014. Mixed effect models for predicting breast height diameter from stump diameter of Oriental beech in Göldağ. *Sci Agric* 72(3): 245-251. https://doi.org/10.1590/0103-9016-2014-0225.

Gasparini P, Di Cosmo L, 2015. Forest carbon in Italian forests: stocks, inherent variability and predictability using NFI data. *Forest Ecol Manag* 337: 186-195. https://doi.org/10.1016/j.foreco.2014.11.012.

Gasparini P, Di Cosmo L, 2016. Italy. In: Vidal C, Alberdi I, Hernandez L, Redmond J (eds) National Forest Inventories – Assessment of Wood Availability and Use. Springer, Cham, Switzerland, pp. 485-506. https://doi.org/10.1007/978-3-319-44015-6_26.

Gasparini P, Tosi V, Di Cosmo L. 2010. National Forest Inventory Reports: Italy. In: Tomppo E, Gschwantner T, Lawrence M, McRoberts RE (eds) National Forest Inventories – Pathways for Common Reporting. Springer, Dordrecht, Netherlands, pp. 311-331. https://doi.org/10.1007/978-90-481-3233-1

Gschwantner T, Lanz A, Vidal C, Bosela M, Di Cosmo L, Fridman J, Gasparini P, Kuliešis A, Tomter S, Schadauer K, 2016. Comparison of methods used in European National Forest Inventories for the estimation of volume increment: towards harmonisation. *Ann For Sci* 73(4): 849-856. https://doi.org/10.1007/s13595-016-0554-5.

INFC, 2007. Le stime di superficie 2005 – Prima seconda. Autori G. Tabacchi, F. De Natale, L. Di Cosmo, A. Floris, C. Gagliano, P. Gasparini, L. Genchi, G. Scrinzi, V. Tosi. Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio. MiPAF – Corpo Forestale dello Stato – Ispettorato Generale, CRA - ISAFA, Trento. Available online: https://www.sian.it/inventarioforestale/jsp/home_en.jsp (15 January 2020).

MAF-ISAFA, 1988. Criteri di stima delle utilizzazioni legnose pregresse. In: Inventario forestale nazionale – IFNI1985. Sintesi metodologica e risultati. Ministero dell'Agricoltura e delle Foreste, Corpo forestale dello Stato. Istituto Sperimentale per l'Assestamento Forestale e per l'Alpicoltura, Trento, 463p.

Özçelík R, Brooks JR, Diamantopoulou MJ, Wiant HV Jr, 2010. Estimating breast height diameter and volume from stump diameter for three economically important species in Turkey. *Scand J Forest Res* 25(1): 32-45. https://doi.org/10.1080/02827580903280053.

Pignatti S, 1979. I piani di vegetazione in Italia. *Giorn Bot Ital* 113: 411-428. (in Italian).

Pond NC, Froese RE, 2014. Evaluating published approaches for modelling diameter at breast height from stump dimensions. *Forestry* 87(5): 683-696. https://doi.org/10.1093/forestry/cpu029

Pond NC, Froese RE, 2015. Interpreting stand structure through Diameter distributions. *For Sci* 61(3): 429-437. https://doi.org/10.5849/forsci.14-056.

R Core Team, 2019. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Sprugel DG, 1983. Correcting for bias in log-transformed allometric equations. *Ecology* 64(1): 209-210. https://doi.org/10.2307/1937343.

Tabacchi G, Di Cosmo L, Gasparini P, 2011. Aboveground tree volume and phytomass prediction equations for forest species in Italy. *Eur J Forest Res* 130(6): 911-934. https://doi.org/10.1007/s10342-011-0481-9.

Westfall JA, 2010. New models for predicting diameter at breast height from stump dimensions. *North J Appl For* 27(1): 21-27. https://doi.org/10.1093/njaf/27.1.21.

Westfall JA, Mcroberts RE, 2017. An assessment of uncertainty in volume estimates for stands reconstructed from tree stump information. *Forestry* 90(3): 404-412. https://doi.org/10.1093/forestry/cpw064.

Zeng WS, 2015. Using nonlinear mixed model and dummy variable model approaches to develop origin-based individual tree biomass equations. *Trees* 29: 275-283. https://doi.org/10.1007/s00468-014-1112-0.