Development of an individual tree volume model for Irish Sitka spruce and comparison with existing UK Forestry Commission and Irish GROWFOR models

Sarah O'Rourke[a*], Máirtín Mac Siúrtáin[b] and Gabrielle Kelly[a]

Abstract

The two most common models used to estimate tree volume in Ireland are the British Forestry Commission (FC) models, developed using British data and the Irish Growfor models developed from Irish stand data. Here, a model analogous to the British FC model for Sitka spruce (*Picea sitchensis* (Bong.) Carr.) is developed using individual Irish tree data and all models are compared. Sitka spruce data from Coillte Teo (Irish Forestry Board) stands in the Wexford and Waterford regions, consisting of 81 thinned plots with a total of 4,315 volume-sampled trees (5,419 trees including repeated measurements), were used for the study. Stepwise regression was carried out to select the model and a Box-Cox transformation was used to improve model fit. A 2k-fold cross validation was used to check the validity of the selected model across different subsets of the data. The selected model slightly over-predicts volume in its mean volume per ha estimate by 0.38%, whereas the GROWFOR estimate under-predicts by 4.6% and the British FC model under-predicts by 8.6%.

Keywords: *Stepwise regression, 2k-fold cross validation, subset comparison, Box-Cox transformation, mean squared predictive error.*

Introduction

Sitka spruce (*Picea sitchensis* (Bong.) Carr.) is the main tree species in Irish forests accounting for approximately 52% of total forest area (National Forest Inventory 2007). Models are particularly useful when estimating individual tree volumes as they do not necessitate tree felling or extensive height and diameter measurements. The two main existing models used for estimating individual tree volume of Irish Sitka spruce are the British Forestry Commission (FC) single tree model for Sitka spruce (Hamilton 1985, Matthews and Mackie 2006) and the Sitka spruce model used in the Irish GROWFOR software that computes volume on a stand basis (Broad and Lynch 2006a).

More recently, dynamic models and spatio-temporal models have been developed to model tree growth, for example those developed by Fox (2007a, 2007b). This has arisen in response to more extensive data becoming available, for example through aerial photography (Dubayah 2000), GIS methodology (Plant 2012) and developments in statistical modelling. However, the models are not in a

a UCD School of Mathematical Sciences, University College Dublin, Belfield, Dublin 4.
b UCD Forestry, UCD School of Agriculture and Food Science, Belfield, Dublin 4.
* Corresponding author: sarah.o-rourke@ucdconnect.ie

form that is readily available or usable by industry, e.g. via commercial software packages. Use of such models is also contingent on the provision of training regarding their use.

The Forestry Commission model for estimating individual tree volume is widely used in the UK and Ireland by commercial, government and industrial organisations. The model parameter estimates are published in the FC tables. The FC model is a single tree tariff system with the dependent variable, the single tree tariff number (TN) expressed in the British quarter girth system units, while the independent variables, diameter at breast height (DBH) and total height, are both expressed in metric units. The quarter girth units are converted to metric units that can then be used to read the tables (see Materials and Methods section).

The FC model and the model developed in this paper are static models in that they assume the stands are managed according to a prescribed thinning pattern as given by Matthews and Mackie (2006). The GROWFOR model is dynamic in that it does not assume prescribed management regimes. It relies on modelling incremental changes in the variables of interest over time. The multivariate extension of the Bertalanffy-Richards model (García 1994) is used for estimation and just requires data on a stand basis rather than an individual tree basis. The model takes into account the relationship between the stand and its variables by classifying the stand by its top-height – age relationship (i.e. site index class). The GROWFOR model for Sitka spruce was developed by Broad and Lynch (2006a) with the aim of developing an "alternative growth projection mechanism for Sitka spruce that is amenable to simulating a wide range of management alternatives specific to Irish conditions" (Ibid.).

Until recently, forest growth and yield modelling in Ireland was carried out using Forestry Commission Yield Models for Forest Management (Hamilton 1985). In the absence of Irish models, these models served the Irish forestry sector well. However, given the higher growth rates in Ireland compared to Britain, the use of these tables in Ireland is questionable. The inadequacy of these tables was further highlighted by Keogh (1985). Since 1999, the Irish forest industry, with support from COFORD (Council of Forest Research and Development) and Coillte Teo (Irish Forestry Board), led a project to develop dynamic yield models which are based on Irish research data (Purser and Lynch 2012). When Irish dynamic yield models were compared directly with British Forestry Commission (FC) models it was found that current volumes (i.e. where there is no growth projection) produced by Irish dynamic yield models are frequently greater. This may be attributable to a combination of factors including: 1. Irish stands show improved upper stem diameters due to improved growing stock and/or growth conditions; 2. FC volume estimates may have been prepared on a conservative basis due to mensurational techniques employed (*Ibid*.).

The objective of this study was to develop a model for estimating the individual tree volume of Irish Sitka spruce that was comparable to the FC model for individual tree volume, but developed specifically using Irish data and to compare this with the existing FC and GROWFOR models. Permanent sample plot data from Coillte stands in the Wexford and Waterford region were used in this study.

Materials and methods

The data

Coillte Teo (Irish Forestry Board) maintains the most extensive crop structure database on Sitka spruce in the Irish Republic. The database includes many silvicultural thinning and spacing experiments that have been repeatedly remeasured during the period 1963 to 2006. These experiments were established using repeated measurement experimental designs with the randomized block design being the most widely adopted approach. Plots that were measured repeatedly were considered permanent.

The Sitka spruce permanent sample plots were classified into seven regions throughout the country. The data used to develop the models were extracted from Sitka spruce thinned plots in counties Wicklow and Wexford. There were 81 plots in total distributed over eight locations (Table 1).

All of the data arose from thinned permanent sample plots. Any models fit to the data refer only to trees and stands where thinning has occurred. However, neither the thinning type, thinning intensity or thinning cycle were considered in this study.

Data from the 1971 to 2006 period were used. There were a total of 320 plot "visits", which included repeated measures on 81 plots. Plot sizes ranged from 0.0405 ha to 0.0809 ha. The trees in these plots were planted from 1943 to 1966. The plots were homogenous and contained only Sitka spruce.

The diameter at breast height (cm), measured at 1.3 m above ground-level, of all the trees in each plot were recorded at each assessment. The variable mean DBH is the quadratic mean diameter, i.e the square root of the sum of the DBH's squared (Matthews and Mackie 2006). Height (m) was measured on a subsample of the thinned trees and the maincrop trees after thinning. Within each plot there were trees for which data on upperstem diameters were also recorded at a specified number of points along the stem. The number of combined upperstem diameter-height measurements along the stem varied from 5 to 10, with 10 measurements being the norm and shorter trees having fewer measurements. From these the volume of the tree could be calculated. The total tree height (TOTH) of all these was also recorded. These trees are referred to as volume-sample trees. The age of each tree when measured or remeasured was also recorded.

Table 1: *Details of the Sitka spruce permanent sample plots in Wicklow and Wexford.*

	Location	Planting year	Property	No. plots
1.	Avoca	1943	Ballinvally	4
2.	Avoca	1953	Ballymoyle Hill	12
3.	Ballinglen	1966	Askakeagh	12
4.	Clonegal	1957	Coolmelagh	16
5.	Coolgraney	1947	Raheenleagh	4
6.	Coolgraney	1949	Raheenleagh	3
7.	Forth	1962	Cools	16
8.	Shillelagh	1951	Barnamuinga	14

For each plot a number of top height (m) measurements were made, where a "top height" tree "is the tree of largest DBH in a 0.01 ha sample plot" (Edwards and Christie 1981).

Stocking density (n ha^{-1}) was recorded, and the easting (E) (m), northing (N) (m) and elevation above sea-level (Z) (m) for each permanent sample plot were determined using mapping software ArcPad 10 and ArcGIS 10.

Model fitting

The 81 sampled plots consisted of 14,283 trees in total. The volume-sample tree data containing the sectional lengths and diameters were extracted from the dataset for 4,315 trees, including repeated measurements. This consisted of 5,419 measurements in total. The sectional diameters (d) were converted to cross-sectional areas ($CSA = \pi(d/200)^2$) (m^2). The sectional volumes (m^3) were calculated using Smalian's formula for the volume of the frustum of a paraboloid (Avery and Burkhart 1994). The sum of the sectional volumes is considered a reasonable estimate of the true individual tree volume.

A series of regression models were fitted with the aim of identifying the variables that are significant for the individual tree volume (m^3) estimates. The dependent variable was individual tree volume (VOL) and the explanatory variables considered were DBH, tree height (HT), tree age (years), plot stocking density (n ha^{-1}), approximate plot eastings, northings and elevation. Following a graphical analysis of the relationships between VOL and these variables, the quadratic and cubic terms of DBH and HT and the interaction between DBH and HT were also considered in the model. Nested models were fitted by adding and dropping covariates in succession. These models were compared using their adjusted coefficient of determination (adj. R^2) where R^2 is adjusted for the number of explanatory terms in the model. Unlike R^2, the adjusted R^2 increases only if the new term improves the model more than would be expected by chance. The preferred model was chosen as the model with the highest adjusted R^2 value.

A Box-Cox transformation (Box and Cox 1964) was used to estimate an appropriate power transformation (λ) for the dependent variable, individual tree volume, to improve model fit. Volume estimates were also calculated from the data using the Forestry Commission and Irish GROWFOR models for Sitka spruce. The models were compared using their adjusted coefficient of determination (adj. R^2) and by comparing the approximate volume (m^3 ha^{-1}) estimates of the predicted values.

A 2k-fold cross validation was used to check the validity of the preferred model. The dataset was split retrospectively randomly into approximately two halves; the model was re-fit to one half and used to predict tree volume estimates on the other. Stepwise regression was carried out on the two halves to determine if the same independent variables were selected using the two half datasets. In addition, a regression was carried out using the variables selected for the original model, a model with new parameter estimates was fit to one half of the data, the training dataset. These parameter estimates were then used to predict tree volume estimates for the remainder of the data, the test dataset. The mean squared error of the training dataset was compared to the mean squared predictive error (MSPR) of the test dataset (Kutner et al. 2004):

$$\mathrm{MSPR} = \frac{\sum_{i=1}^{n^*}(Y_i - \hat{Y}_i)^2}{n^*} \tag{1}$$

where Y_i is the value of the response variable in the i^{th} validation case, $\hat{Y}_i$ is the predicted value for the i^{th} validation case based on the model-building dataset, and n^* is the number of cases in the validation dataset.

The process was repeated, using the training data instead of the test data and the test data instead of the training data.

The Forestry Commission (FC) model

The Forestry Commission single tree tariff system for pure even-aged stands allow tree volume to be calculated. The principle underlying this approach is that there is a linear relationship between basal area and volume; therefore, volume can be estimated from DBH (Matthews and Mackie 2006). It does not provide an estimate of individual tree volume directly but rather estimates the tariff number for any individual Sitka spruce tree as a function of the DBH and the total height. The domains over which the single tree tariff model is applicable are:

- $8 \leq \mathrm{DBH} \leq 80$ (cm)
- $8 \leq \mathrm{TOTH} \leq 40$ (m)

A fundemental issue persists with the FC single tree tariff system in that the dependent variable, the single tree tariff number (TN), is expressed in the British quarter girth system units, while the independent variables DBH and total height are both expressed in metric units.

The TN refers to a pre-constructed volume-basal area equation each of which has two predefined volume-basal area coordinates. The first predefined volume-basal area coordinates correspond to minimum volume and minimum basal area i.e. (0.005 m^3, 0.03848 m^2). The second predefined volume-basal area coordinates correspond to the volume (m^3) associated with the tariff number and the basal area associated with one square foot quarter girth, i.e. (0.036054 TN m^3, 0.118288 m^2).

Having estimated the TN, the slope (b_1) (m^3/m^2) and intercept (b_0) (m^3) associated with the pre-constructed local volume-basal area model can be estimated. The volume of an individual tree is estimated using the estimated parameters of the local volume-basal area model and the basal area of the individual tree. Using information from Matthews and Mackie (2006), the Forestry Commission preconstructed metric single tree model for estimating Sitka spruce individual tree volume (m^3) can be calculated as follows:

$$\text{Individual tree volume} = b_0 + b_1 \mathrm{BA}(m^2) \tag{2}$$

where basal area (BA) = $\left(\frac{\text{DBH}}{200}\right)^2$

$$b_0 = -0.004882 - 0.002147(\text{HT}) + 0.000505(\text{DBH})$$
$$b_1 = 2.568687 + 0.558008(\text{HT}) - 0.131221(\text{DBH})$$

The FC individual tree volume was estimated for each volume-sample tree using Equation 2.

Irish GROWFOR software

GROWFOR is Irish forest dynamic yield software that estimates current timber volume (m^2 ha^{-1}) when a specific plot input vector is provided. The input variables per plot are: species, age, stocking density (n ha^{-1}), top height and basal area (m^2 ha^{-1}), or mean DBH - the mean DBH over all trees in a plot.

Input vector estimates from 320 Sitka spruce permanent sample plots, 81 plots with repeated measurements, were input individually into GROWFOR. Note GROWFOR estimates are made on a unit area basis and not an individual tree basis.

Results

Model fitting

Table 2 shows eight models that were fit to the data subset consisting of volume sample trees only (n = 5,419 volume sample trees) and their adjusted R^2. The independent variables in Model 2 (FC_{IRISH}) are the same as those that occur in the reparamaterized Forestry Commission single tree volume model derived previously.

Diameter at breast height is a good predictor of tree volume; the inclusion of the quadratic terms DBH and tree height in model 5 gave a high adj. R^2 of 0.9798. Including HT (tree height), age and interaction terms between HT and DBH led to

Table 2: *The table shows individual tree volume models for Irish Sitka spruce for a selection of models with associated adjusted R^2.*

Model	Variables in the model	Adj. R^2	
1	VOL = DBH, DBH^2, DBH^3, HT, HT×DBH, HT×DBH^2, AGE, AGE^2	0.9889	
2	VOL = DBH, DBH^2, DBH^3, HT, HT×DBH, HT×DBH^2	0.9885	FC_{IRISH}
3	VOL = DBH, DBH^2, HT, HT×DBH	0.9875	
4	VOL = mDBH, DBH^2, mDBH×DBH^2, mHT, mHT×DBH^2	0.9833	
5	VOL = DBH, DBH^2, HT	0.9798	
6	VOL = DBH, HT	0.8690	
7	VOL = AGE, AGE^2, E, N, Z	0.6360	
8	VOL = AGE, AGE^2	0.6313	

where VOL = tree volume , DBH = diameter at breast height (cm), HT = total height (m), AGE = tree age (years), mDBH = mean diameter at breast height of all volume sample trees in a plot (m), mHT = mean height of all volume sample trees in a plot (m), E = easting (m), N = northing (m) and Z = elevation above sea level (m).

an improved fit. Only variables with a significant p-value were retained in any of the models. Model 2 (FC_{IRISH}), which contains the same variables as the Forestry Commission individual tree model, but not the same parameter estimates, has an adj. R^2 of 0.9885. There is an increase from 0.9813 to 0.9885 using the newly estimated coefficients instead of the FC coefficients (Table 5).

When HT and DBH were in the model, E, N and Z were not significant and therefore were omitted. Similarly, HT^2 and the stocking density were not significant.

Model 1 is similar to model 2, except it includes the linear and quadratic term of tree age. This only led to a slight increase in the adj. R^2, but the variables were significant and were retained. Although the differences in the adjusted R^2 values are small, these become important, in practical terms, when tree volume is converted to volume per hectare.

The residual variance for these models (Table 2) increased as volume increased (Figure 1). Figure 2 shows the model fit less well for the smallest and largest volumes. A Box-Cox transformation (Box and Cox 1964) was applied to transform the dependent variable volume. The "best" lambda (λ) was estimated at 0.25 for all models containing DBH and height (models 1–6) and at 0 for models excluding DBH and HT (models 7 and 8). Table 3 shows the back-transformed adjusted R^2 for models 1 and 2, with the highest adjusted R^2 values and model 8, to show the effect of the transformation on a model excluding DBH and HT.

Therefore, root 4 of tree-volume ($VOL^{0.25}$) was used as the dependent variable for models 1 and 2, and log(VOL) for model 8. The improvement in the residual variance can be seen in Figures 3 and 4. Four observations were removed from the

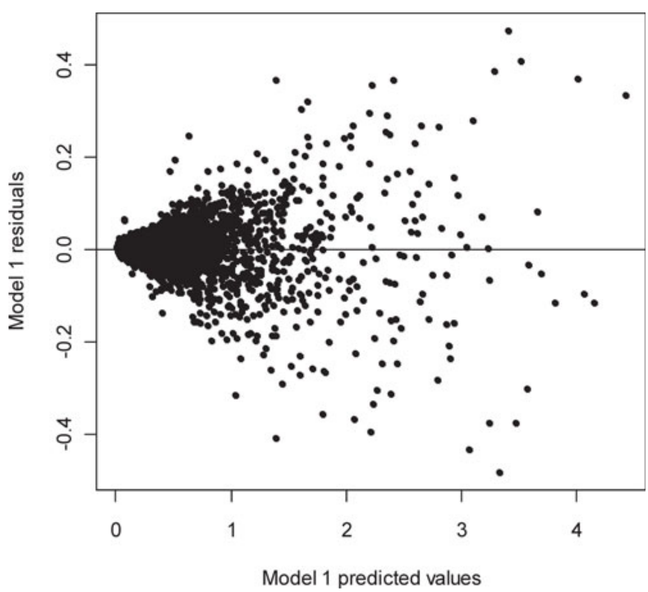

Figure 1: *The plot shows the residuals versus the predicted values of individual tree volume (m^3) for model 1 defined in the text. Model 1 is the best fitting model to observed tree volumes of Sitka spruce in the Wicklow region, based on diameter at breast height, height and age.*

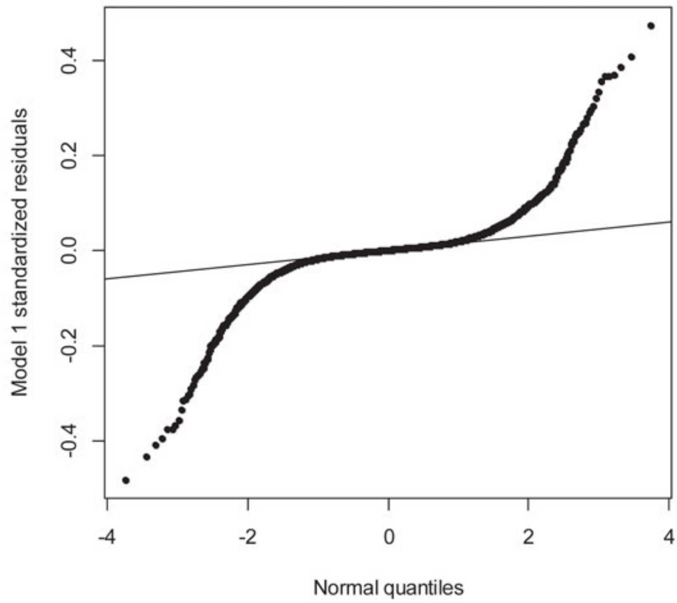

Figure 2: *A normal q-q plot of the standardised residuals from model 1 defined in the text. Model 1 is the best fitting model to observed tree volumes of Sitka spruce in the Wicklow region, based on diameter at breast height, height and age.*

data as outliers. The adjusted R^2 and back transformed adjusted R^2 for the transformed models are presented in Table 3.

The Box-Cox model 1 contains the same independent variables as model 1, except the dependent variable is volume$^{0.25}$. The back-transformed adjusted R^2 was 0.9901, which is a slight increase from 0.9889 in Table 2. Similarly, the Box-Cox model 2 has the same variables as model 2 and the back-transformed adjusted R^2 of 0.9898, which is a slight increase from 0.9885. There were very small increases in the back-transformed adjusted R^2; however, the variance of the residuals did not increase in proportion to the mean (Figure 3) and thus inference using least squares estimation, which requires constant variance, is valid.

There is a decrease in the back-transformed adjusted R^2 from model 8 compared

Table 3: *Models with the dependent variable (volume) transformed.*

Box-Cox model	**Model variables**	**Adj. R^2 (not back transformed)**	**Adj. R^2 (back transformed)**
1.	$VOL^{0.25}$ = DBH, DBH^2, DBH^3, HT, HT×DBH, DBH^2×HT age age^2	0.9942	0.9901
2.	$VOL^{0.25}$ = DBH, DBH^2, DBH^3, HT, DBH×HT, DBH^2×HT	0.9940	0.9898
8.	Log(VOL) = age, age^2	0.6233	0.6078

where DBH = diameter at breast height (cm), HT = total height (m), AGE = tree age (years) and VOL = tree volume (m^3).

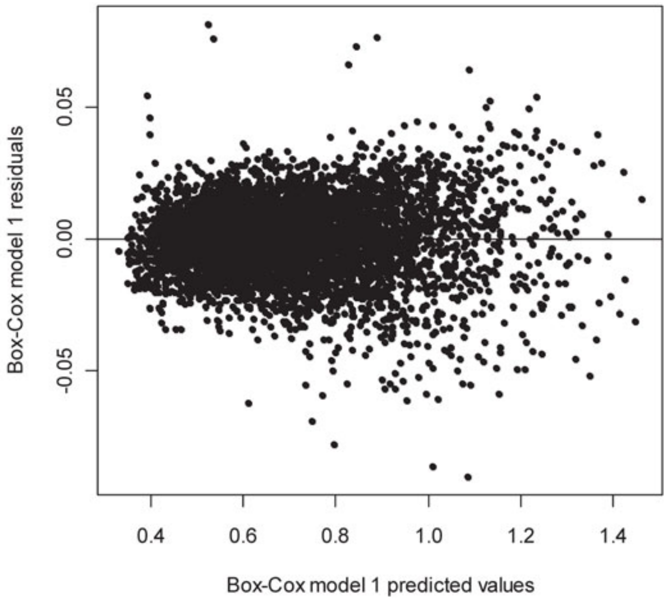

Figure 3: *The plot shows the residuals versus the predicted values of individual tree volume (m^3) for Box-Cox model 1 defined in the text. Box-Cox model 1 is the best fitting model, with a Box-Cox transformation, to root 4 of the observed tree volumes of Sitka spruce in the Wicklow region, based on diameter at breast height, height and age.*

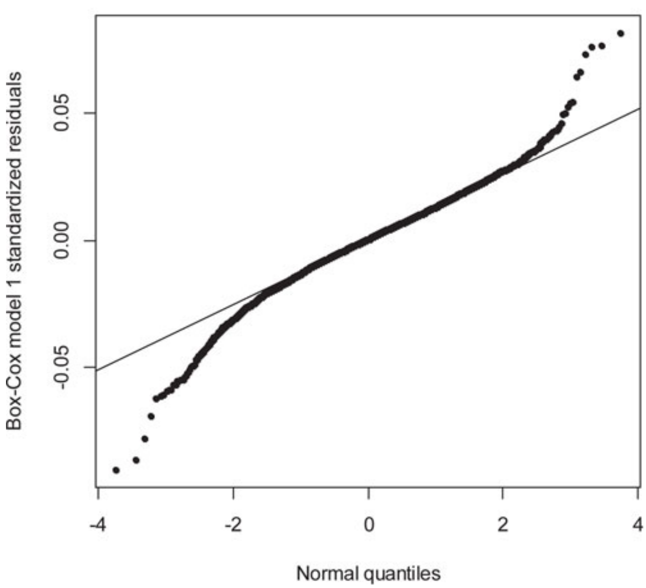

Figure 4: *A normal q-q plot of the standardised residuals from Box-Cox model 1 defined in the text. Box-Cox model 1 is the best fitting model, with a Box-Cox transformation, to root 4 of the observed tree volumes of Sitka spruce in the Wicklow region, based on diameter at breast height, height and age.*

Table 4: *Box-Cox single tree model 1 parameter estimates.*

	Parameter	Estimate	Standard error	t-value	Pr > |t|
β_0	Intercept	0.0376366284	0.00399242	9.43	<0.0001
β_1	DBH	0.0308911823	0.00033019	93.56	<0.0001
β_2	HT	0.0109354768	0.00053631	20.39	<0.0001
β_3	DBH^2	−0.0003524884	0.00001775	−19.86	<0.0001
β_4	DBH^3	0.0000008792	0.00000031	2.80	0.0051
β_5	DBH×HT	−0.0001484999	0.00003902	−3.81	0.0001
β_6	DBH^2×HT	0.0000041729	0.00000067	6.18	<0.0001
β_7	age	0.0030090328	0.00019115	15.74	<0.0001
β_8	age^2	−0.0000453409	0.00000294	−15.42	<0.0001

where DBH = diameter at breast height (cm), HT = total height (m) and AGE = tree age (years).

to the Box-Cox model 8; however as before, the distribution of the residual variance is random.

Box-Cox model 1 has the highest adj. R^2 and the residual variance does not increase in proportion to the mean, therefore, it is considered the "best" model of the models considered above. The parameter estimates are shown in Table 4 and the equation is as follows:

$$(\text{Tree Volume})^{0.25} = \beta_0 + \beta_1(\text{DBH}) + \beta_2(\text{DBH}^2) + \beta_3(\text{DBH}^3) + \beta_4(\text{HT}) + \beta_5(\text{DBH}\times\text{HT}) + \beta_6(\text{DBH}^2\times\text{HT}) + \beta_7(\text{age}) + \beta_8(\text{age}^2) \qquad (3)$$

The Box-Cox model 1 predicted versus the actual tree volumes are plotted in Figure 5. Clearly there still is more variability associated with larger volume estimates. It appears the model does not over- or under-estimate volume, given that there is an even distribution on either side of the fitted line.

The Forestry Commission mean volume and volume per hectare estimates

The percentage differences between actual tree volumes and the FC tree volume estimates (VOL_EST_{UK}) were calculated as a percentage of the actual tree volumes. The average difference was 12%; however, this varied from year to year. It should be noted that the average difference decreased in later years, i.e. as tree-age increased.

Figure 6 shows a plot of VOL_EST_{UK} versus VOL. Many of the tree volume estimates from the FC model are below the line of equality suggesting the FC model under-estimates tree volume. The adj. R^2 for the FC_{UK} model is 0.9813.

The 320 mean volume per ha (mVOL_ha) estimates are shown in Table 5 to further compare the models. The volumes estimated using Box-Cox model 1 were closest to the actual volumes from the volume-sample trees. The Forestry Commission model per hectare estimates were approximately 8.6% lower than the volume-sample tree estimates.

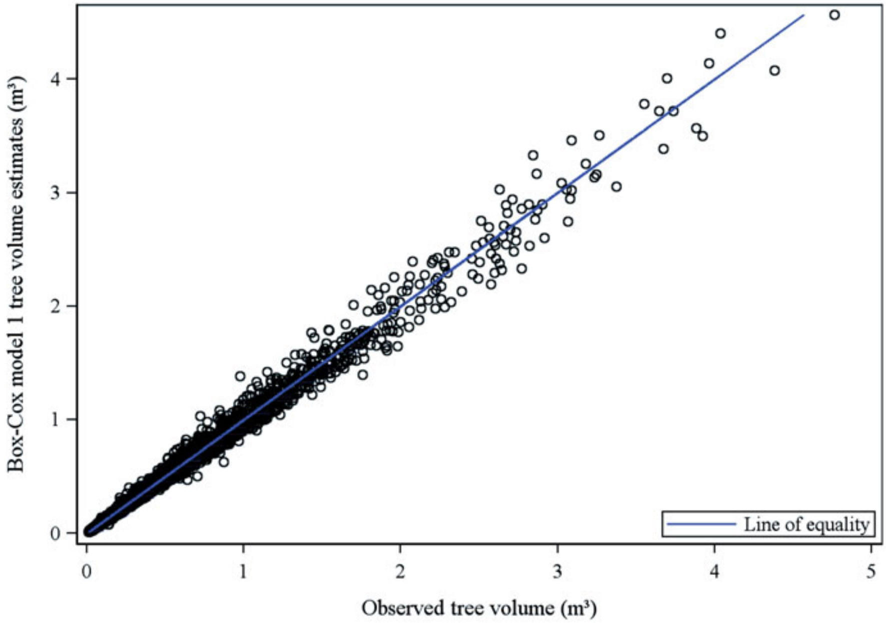

Figure 5: *Predicted values of the best fitting model (Box-Cox model 1), with a Box-Cox transformation, to root 4 of the observed tree volumes of Sitka spruce in the Wicklow region, based on diameter at breast height, height and age were obtained. In the figure these are plotted against observed tree volumes (VOL).*

GROWFOR estimates

Using the GROWFOR input vector for a plot, i.e. age, stocking, top-height and mean DBH, the output, i.e. the VOL_ha estimates, for each plot were computed and stored in a separate file.

The mean and total volume per hectare estimates from GROWFOR were approximately 4.6% lower than the observed per hectare estimates. It may be argued that these are not comparable since the projections from GROWFOR are stand-based. However, it is worth noting that the GROWFOR estimates did better than the Forestry Commission volume per hectare estimates. Figure 7 shows GROWFOR volume per hectare estimates versus the observed volume per hectare estimates; the points below the fitted model illustate that the GROWFOR estimates tended to be lower than the observed volumes.

Table 5: *Volume estimates by model.*

Model	Adj. R^2	Mean Vol_ha estimate (SE) (m^3 ha^{-1})
Actual observed volumes		397.1 (10.7)
Box-Cox model 1	0.9901	398.6 (10.8)
GROWFOR		378.7 (10.3)
Forestry Commission	0.9813	363.1 (11.0)

where SE = standard error.

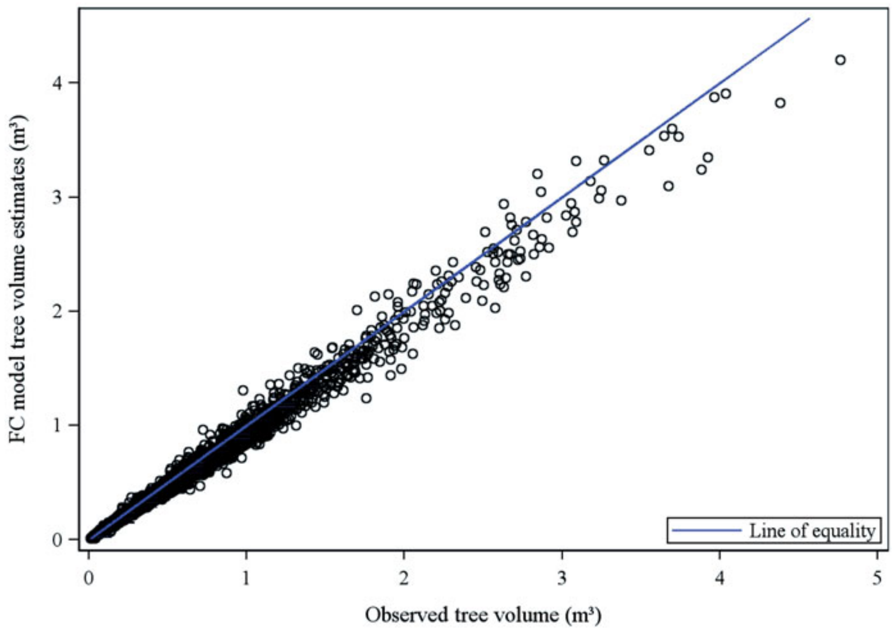

Figure 6: *Predicted values of the forestry commission model of the response variable observed tree volume (VOL_EST_{UK}) of Sitka spruce in the Wicklow region, were obtained. In the figure these are plotted against observed tree volumes (VOL).*

Subset comparison

The Irish individual tree models above were fitted to the entire dataset. As a validation method for Box-Cox model 1, the dataset was split retrospectively and randomly into approximately two halves to carry out cross validation of the data. Each half contained 160 plots and will be referred to as D_A (2,157 trees) and D_B (2,187 trees).

Stepwise regression was carried out separately on D_A and D_B, with all of the variables discussed in the results (model fitting) section included, with $VOL^{0.25}$ as the dependent variable. In the case of D_B, the same variables were selected as those in the Box-Cox model 1. In the case of D_A, all the variables that were in Box-Cox model 1 were chosen except DBH×HT, the interaction term between DBH and tree height. However, since DBH^2×HT was chosen, both terms were kept in the model in accordance with the hierarchy principle. In both cases, D_A and D_B, the models selected are consistent with Box-Cox model 1. Therefore, in the comparison below all variables of Box-Cox model 1 were used for prediction. The variables from Box-Cox model 1 were fitted to D_A, providing a new set of parameter estimates. These were used to calculate predicted estimates of tree volume for D_B. The mean squared error (MSE) from D_A and the mean squared predictive error (MSPR) from D_B are presented in Table 6.

The MSE for D_A is 0.002627 and the MSPR for D_B is 0.002552. The fact that the

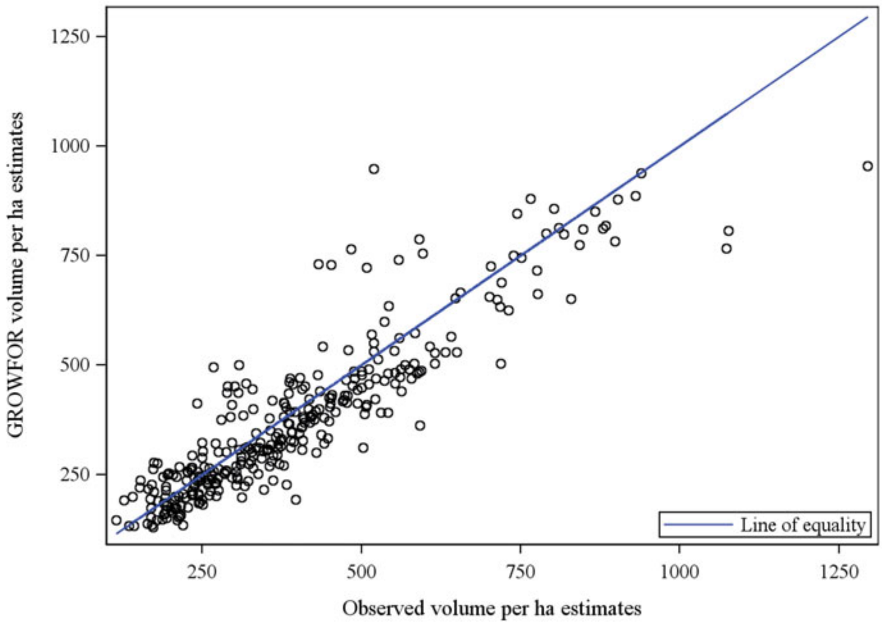

Figure 7: *Volume estimates (m^3 ha^{-1}) for 320 Sitka spruce plots in the Wicklow region were obtained using the GROWFOR software. In the figure these are plotted against the observed volumes.*

MSPR is close to MSE implies that the MSE based on the training data set is a valid indicator of the predictive ability of the fitted regression model.

The process was repeated with D_B as the training data and D_A as the test data. The variables from Box-Cox model 1 were fit to D_B, providing a new set of parameter estimates which were used to calculate predicted estimates of tree volume for D_A. The MSE from D_B and MSPR from D_A are shown in Table 7. When using D_B for the training data, the MSE of the model fit is 0.002311, the MSPR for D_A is 0.003164. As above, the fact that these values are close suggests the MSE for D_B is a reasonably valid indicator of the predictive ability of the fitted regression model. The fact that the MSE and MSPR values are not very different is not surprising

Table 6: *Fit statistics for the training data (D_A) and predicted fit statistics for the test data (D_B). The residual sum of squares (SSE), the mean squared error (MSE) and the mean squared predictive error (MPSR) are provided.*

	D_A training data	D_B test data
Adj. R^2	0.9889	0.9901
SSE	5.6441	5.5586
MSE	0.0026	
MSPR		0.0026

Table 7: *Fit statistics for the training data (D_B) and predicted fit statistics for the test data (D_A). The residual sum of squares (SSE), the mean squared error (MSE) and the mean squared predictive error (MPSR) are provided.*

	D_B training data	D_A test data
Adj. R^2	0.9911	0.9866
SSE	5.0344	6.7963
MSE	0.0023	
MSPR		0.0032

since the training and test data were subsets of the same large dataset. However, it supports the validity of the model fit to these data.

Discussion

The 81 plots from which the data were collected plots established for the purposes of research by Coillte. They were used by Broad and Lynch (2006a) to develop dynamic yield models for Sitka spruce, and made available in GROWFOR. In a subsequent paper, Broad and Lynch (2006b) found that sampling in Coillte research plots has been skewed towards higher productivity stands. Defining site index as the top height at age 30 (*Ibid.*), there were no experimental plots in the lower two site index classes in the Coillte estate. Therefore, any models fit to these data are representative of the site index distribution associated with research plots.

One issue with this dataset is that it contains repeated measures. Data from the 81 plots were collected over a 35-year period, with each plot revisited approximately four or five times. The selection procedure followed was always the same over time as specified by the experiment protocol. Some volume-sample trees were measured once only, while some were measured up to three times. Repeated measurements, from the same tree, are not independent and are therefore correlated. However, in like manner, repeated measurements from different trees are independent and uncorrelated. This issue was not accounted for in the present model and may lead to underestimation of standard errors in estimates in the fitted models. For comparison purposes, with the FC and GROWFOR models, the correlation from repeated measurements was not investigated.

The eastings, northings and elevations were excluded from models that included DBH and HT as they were not significant, suggesting a plot's location did not affect the volume of its trees. However, values of E, N and Z were the same for all trees in a plot and individual tree GIS co-ordinates were not available for analysis. Thus, while there were no large scale spatial effects, on a smaller scale they may have existed.

While the issues discussed above (i.e. bias in the data and repeated measures) need to be considered, for this research dataset Box-Cox model 1 had the highest adjusted R^2 and the best fit. It over-predicted mean volume per ha estimates (Table 5) by 0.38%, but was better than the GROWFOR estimates which under-predicted by 4.6% and the FC model which under-predicted by 8.6%. For a forest of 100 ha size, the estimated timber volume (using the observed tree volumes) would be 397,100 m^3. Using the Box-Cox model 1, the predicted volume would be 398,600 m^3. Using the GROWFOR and FC models the predicted volume would be 378,100 m^3 and 363,100 m^3, respectively. This leads to a difference of 20,500 m^3 and 35,500 m^3, respectively, which may be significant in terms of commercial timber value. Note that these models assume a pre-determined management regime.

The Box-Cox model 1 requires the age, DBH and height of a tree to estimate volume. While the age of a tree is usually recorded and the DBH is easily measured, height is usually only measured on a subsample of trees. This might no longer be the case with new methods of measurements, such as lidar remote sensing technology and other methods, which has demonstrated the capability to accurately estimate important forest structural characteristics (Dubayah 2000).

Further analysis could be carried out to investigate the presence of spatial effects and the effects of thinning, if individual tree locations were available. In the present model time is accounted for by age of the tree. The data could be modelled over time, as the trees grow, producing a more dynamic model. This is very important for forecasting and predicting tree volume.

Currently forests are seen as ecosystems and developments continue to progress in relation to their management application subject to associated changing rules from regulatory bodies. In addition, climate change plays an increasing role in forest management and associated production. Thus, while here only a static model has been considered it may provide a basis on which more general models may be built. It may also be useful in itself as changes may occur too rapidly to be incorporated in a general model.

Acknowledgements

The authors would like to thank Ted Lynch of Coillte Teo for helpful discussions. The first author was funded by the Council for Forest Research and Development (COFORD) and UCD School of Mathematical Sciences.

References

Avery, T.E. and Burkhart, H.E. 1994. *Forest Measurements*. 4th Ed. New York, McGraw-Hill.

Box, G.E.P. and Cox, D.R. 1964. An analysis of transformations. *Journal of the Royal Statistical Society, Series B* 26(2): 211–252.

Broad, L.R. and Lynch, T. 2006a. Growth models for Sitka spruce in Ireland. *Irish Forestry* 63: 53–79.

Broad, L.R. and Lynch, T. 2006b. Panel data validation using cross-sectional methods. *Irish Forestry* 63: 80–95.

Dubayah, R.O. and Drake, J.B. 2000. Lidar remote sensing for forestry. *Journal of Forestry* 98: 44–46.

Edwards, P.N. and Christie, J.M. 1981. Yield Models for Forest Management. *Forestry Commission Booklet 48*. HMSO.

ESRI (Environmental Systems Resource Institute). 2010. *ArcGIS 10.* ESRI, Redlands, California.

Forestry Commission. 1970. *Metric Tariff Tables. Forestry Commission*. Supplement No. 1 to Tariff Tables Forest Record No. 31. HMSO, London.

Fox, J.C., Bi, H. and Ades, P.K. 2007a. Spatial dependence and individual-tree growth models. I. Characterising spatial dependence. *Forest Ecology and Management* 245: 10–19.

Fox, J.C., Bi, H. and Ades, P.K. 2007b. Spatial dependence and individual-tree growth models II. Modelling spatial dependence. *Forest Ecology and Management* 245: 20–30.

García, O. 1994. The state-space approach in growth modelling. *Canadian Journal of Forest Research* 24:1894–1903.

Hamilton, G.J. 1985. *Forest Mensuration Handbook*. Booklet 39. HMSO.

Keogh, R.M. 1985. *Preliminary Stand Assortment Tables for Sitka Spruce in Ireland*. Research Communication No. 24. Forest and Wildlife Service, Crop Structure Section.

Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W. 2005 *Applied Linear Statistical Models.* 5th Ed. McGraw-Hill/Irwin Series.

Matthews, R.W. and Mackie. E.D. 2006. *Forest Mensuration: A Handbook for Practitioners*. Forestry Commission, Edinburgh.

National Forest Inventory. 2007. *National Forest Inventory of Ireland*: Forest Service, Department of Agriculture, Fisheries and Food.

Plant*, R.E. 2012.* Spatial Data Analysis in Ecology and Agriculture Using R. *CRC Press, Boca Raton, Florida.*

Purser, P. and Lynch, T. 2012. *Dynamic Yield Models used in Irish Forestry*. COFORD Connects, Silviculture/Management No. 20.