Compatible taper-volume models of *Quercus variabilis* Blume forests in north China

Conghui Zheng (1-2), Yuzhong Wang (1), Liming Jia (2), Euan Gordon Mason (3), Songpo We (3), Caowen Sun (3), Jie Duan (4)

Introduction

Compatible taper-volume models are flexible tools for estimating total and merchantable tree volume that can meet the demands of market trends as product specifications change. A compatible taper-volume estimation system contains a taper equation and a total volume equation. The taper equation can provide estimations of diameter at a given height up a tree and merchantable tree volume (Diéguez-Aranda et al. 2006), and the total volume equation can easily estimate the total volume of a tree. Both models require diameter at breast height over bark (DBH₀) and height as inputs. Compatible taper-volume estimation systems allow the volume computed by integration of the taper equation from the ground to the top of the tree to equal that calculated by a total volume equation. Taper and volume estimation systems can be divided into two types: Type (1), the total volume model is directly derived through integration of the taper equation; Type (2), equation form of the total volume model is independent from the taper equation. For type (1), two methods can be used to estimate parameters of the two models: Method (1), firstly fit the taper equation, then the volume model with its parameters can be directly obtained by integration (Martin 1981); Method (2), after obtaining the total volume equation through integration of the taper equation, the two models are fitted simultaneously to get their parameters by seemingly unrelated regression (SUR) or full information maximum likelihood (FIML) procedures (Jiang et al. 2005, Brooks et al. 2008, Pompa-García et al. 2009, Ozcêlîk & Brooks 2012). For type (2), there are also two methods to estimate the parameters: Method (1), firstly estimate the total volume equation using the total volume observations, then substitute the estimated total volume from the volume model and a specific parameter restriction into a taper model, so that a compatible taper model can be estimated (Goulding & Murray 1976, Malimbwi & Philip 1989, Fang & Bailey 1999, Muhairwe 1999); Method (2), simultaneous estimation of parameters of taper and volume models using SUR or FIML (Diéguez-Aranda et al. 2006). Type (2) can provide an easily applied total volume model which can rapidly estimate tree volume (Diéguez-Aranda et al. 2006), so this type is often preferred. Method (1) for Type (2) was especially useful when simultaneous estimation caused difficulty in achieving convergence, while Method (2) for Type (2) could make a reasonable compromise among the components in the system in the process of minimizing the sum of square errors (Fang et al. 2000, Diéguez-Aranda et al. 2006). Method (2) for Type (2) is more difficult when equations in the system have different numbers of observations. In this case, weights may be needed (Fang & Bailey 1999, Diéguez-Aranda et al. 2006). Some researchers compared the

Keywords: *Quercus variabilis* Blume, Dummy Variable, Box-Cox Transformation, Linear Mixed Effects Models, Compatible Taper-Volume Model

© SISEF http://www.sisef.it/iforest/ 567

iForest 10: 567-575
two methods and found that similar results were obtained (Fang et al. 2000). Method (1) for Type (2) also could make the system more flexible in application, i.e., for users who would like to use an existing volume table or volume equation to estimate volume; they can just use the taper model to obtain the diameter predictions (Fang & Bailey 1999). A large number of compatible taper-volume systems based on type (1) have been created for oak species in Greece, America, Denmark, Spain and Mexico (Hilt 1980, Thomas & Parresol 1991, Tarp-Johansen et al. 1997, Pompa-García et al. 2009, Kitikidou 2010). Simple equations (Hilt 1980, Thomas & Parresol 1991, Kitikidou 2010), variable exponent equations (Tarp-Johansen et al. 1997) or segmented-polynomial equations (Pompa-García et al. 2009) were chosen for taper modelling. There has been no similar research for Quercus variabilis Blume, an important broadleaf tree species in North China, prior to the study reported here. The objective of our study was to develop a suitable compatible taper and volume estimation system including a polynomial taper equation and a total volume model.

The objective of our study was to develop a suitable compatible taper and volume estimation system including a polynomial taper equation and a total volume model (2) using Method (1), which can describe the stem profile well and provide accurate estimates of the stem volume of Quercus variabilis Blume (cork oak) forests in Northern China.

Material and methods

Measurements

174 trees from 104 plots with an area of 20 × 20 m for cork oak natural forests and plantations in North China were used in this study, 57 of those trees were from plantations (including 51 average trees and 26 dominant trees), and the other 117 trees were from natural forests (including 60 average trees and 57 dominant trees). These plots were created in the following locations with different site conditions and age distributions: Gao-Luo forest station, Qi-Jiahe forest station, Bei-Tan forest station and Heng-He forest station of Zhong-tiaoshan region in Shanxi province, collective forests of Da-Geliao village in Xingtai city of Hebei province, Si-Zuoou forestry station and Xi-Shan forestry station in Beijing. Measured and computed variables were as follows: (1) single tree variables including diameter at breast height under bark (DBHb), total tree height (H); (2) two perpendicular diameters inside-bark (dib) of every five rings of each disc, starting with the outermost ring, at 0.0, 0.5, 1.3 and 1.5 m above the ground and then every 1.0 m along the remainder of the stem which were measured and averaged. For the outermost layer of those stem analyzed trees, diameters outside-bark at those same heights were also measured; (3) log volumes were calculated using the Huber’s formula (Figueiredo-Filho et al. 2000) where the top section was treated as a cone. Inside-bark total stem volumes (vib) were obtained by summing the log volumes and the volume of the top of the tree. Each tree contributed to the data set with as many height-diameter measurements from the stem analysis data as possible. Total stem volume (vib) diameter at breast height over bark (DBHao) and height (H) were repeated for each analyzed stem defined by 5 ring measurements (Nunes et al. 2010). The data with DBHao equals to 0 (or H < 1.3 m) were deleted from the data set because the total volume under bark model (vib) would rely on DBHao as an independent variable. Finally, a total of 2358 bark thickness observations, 12814 diameter-height observations and 1299 volume observations from the 174 trees were obtained.

Model building

Four alternative modelling strategies were tried, because (1) a few analyzed stems (“trees”) had ramincorns and (2) some of the small analyzed stems had very high values of relative diameter (Rd), which is equal to dib/DBHao, where dib is the diameter under bark at height h in cm, DBHao is the diameter at breast height under bark in cm, and h is the height from ground in m). Therefore, four sets of compatible bark thickness-taper-volume model systems were built, one for each dataset type: (i) System 1, using all the data of the analyzed stems; (ii) System 2, using data of analyzed stems without ramicorns; (iii) System 3, using data of analyzed stems with Rd less than 1.5; (iv) System 4, using data of the analyzed stems without ramincorns and simultaneously with a Rd less than 1.5. Descriptive statistics for those data sets are shown in Tab. 1. Each model system included three models, i.e., a bark thickness model, a volume model and a taper model. Two dummy variables were used in System 1: Rddummy (Rddummy = 1 when Rd < 1.5, and 0 otherwise) and branch (branch = 1 when the tree has ramicorns, and branch = 0 otherwise). The dummy variable Rddummy was used in System 2, and dummy variable branch was used in System 3.

For bark thickness models and volume models from the above-mentioned four systems, all data were used for model fitting. Before modelling, some exploratory and response variables (V) were transformed to new variables (V') by Box-Cox transformation to make frequency distributions of those variables (V') as close to normal distributions as possible. The following equation expresses a Box-Cox transformation (Sakia 1992 – eqn. 1):

\[ V' = \begin{cases} V^{\lambda} - 1 & \text{if } \lambda \neq 0, \\ \log(V) & \text{if } \lambda = 0 \end{cases} \]

where V was the original response or explanatory variable, V' was the response or explanatory variable after Box-Cox transformation, λ was the parameter in the Box-Cox transformation.

To establish bark thickness models and volume models, linear mixed effects analyses were used on transformed variables with the tree number (tree.no) as random effect. Model form was expressed as in eqn.2 (see below – Peng & Lu 2012, Su et al. 2012). The autocorrelation was addressed using three residual autocorrelation structures: a first-order autoregressive structure [AR(1)], a moving average structure [MA(1)] and combination of first-order autoregressive and moving average structures [ARMA(1,1)]. For bark thickness models of the four systems, tbl6, Rh, natural, dominant, branch (for System 1 and System 3) and their interaction terms were chosen as possible explanatory variables and tbt is

---

**Tab. 1 - Summary statistics of four data sets used for modelling.** (bt): bark thickness; (vib): stem total volume under bark; (dib): diameter under bark at height h; (H): height from ground; (H): total tree height; (DBHao): diameter at breast height over bark, breast height is 1.3 m height above the ground; (Rd): relative diameter, equal to dib/DBHao.

<table>
<thead>
<tr>
<th>Model type</th>
<th>Model system</th>
<th>Sample number</th>
<th>Range of age (year)</th>
<th>Range of Rd</th>
<th>Range of DBHao (cm)</th>
<th>Range of H (m)</th>
<th>Range of response variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>bt (cm)</td>
<td>1, 3</td>
<td>2358</td>
<td>16-84</td>
<td>0.01-1.50</td>
<td>3.8-39.9</td>
<td>5.0-21.0</td>
<td>0.0-3.5</td>
</tr>
<tr>
<td></td>
<td>2, 4</td>
<td>2059</td>
<td>16-84</td>
<td>0.02-1.50</td>
<td>3.8-22.6</td>
<td>5.0-21.0</td>
<td>0.0-3.0</td>
</tr>
<tr>
<td>vib (m³)</td>
<td>1</td>
<td>1299</td>
<td>5-84</td>
<td>1.00</td>
<td>0.3-39.9</td>
<td>1.4-21.0</td>
<td>0.00001-0.649</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1201</td>
<td>5-84</td>
<td>1.00</td>
<td>0.3-23.1</td>
<td>1.4-21.0</td>
<td>0.00001-0.224</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1035</td>
<td>5-84</td>
<td>1.00</td>
<td>1.6-39.9</td>
<td>1.4-21.0</td>
<td>0.0002-0.649</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>937</td>
<td>5-84</td>
<td>1.00</td>
<td>1.6-23.1</td>
<td>1.4-21.0</td>
<td>0.0002-0.224</td>
</tr>
<tr>
<td></td>
<td>dib (cm)</td>
<td>12814</td>
<td>5-84</td>
<td>0.01-12.00</td>
<td>0.3-39.9</td>
<td>1.4-21.0</td>
<td>0.1-41.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11326</td>
<td>5-84</td>
<td>0.01-12.00</td>
<td>0.3-23.1</td>
<td>1.4-21.0</td>
<td>0.1-41.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>11419</td>
<td>5-84</td>
<td>0.01-1.50</td>
<td>1.6-39.9</td>
<td>1.4-21.0</td>
<td>0.1-41.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>9942</td>
<td>5-84</td>
<td>0.01-1.50</td>
<td>1.6-23.1</td>
<td>1.4-21.0</td>
<td>0.1-41.0</td>
</tr>
</tbody>
</table>

---

Zheng C et al. - iForest 10: 567-575

iforest - Biogeosciences and Forestry

iForest: 567-575
the response variable, where $td_{ib}$ is equal to $(td_{ib})^\lambda$ ($n=1, 2, ..., 5$). $td_{ib}$ is the transformed diameter under bark at height $h$ by the Box-Cox method (cm), $R^2$ is equal to $(Rh)^\lambda$ ($n=1, 2, ..., 5$). $Rh$ is the relative height and equal to $h/H$, where $h$ is height from ground (m), $H$ is total tree height (m); natural and dominant are dummy variables to define the forest origin and the tree size, respectively: natural = 1 when the origin was a natural forest, natural = 0 in the case of plantations; dominant = 1 when the tree is dominant, dominant = 0 (otherwise); $bt$ is the transformed bark thickness by the Box-Cox method (cm). The dummy variable $R_{dummy}$ was not used in System 1 and System 2 because the data used for fitting the tbt model were all with a $Rd<1.5$ (Tab. 1). For volume models of the four systems, $DBH_{obs}$, $h_{obs}$, natural, dominant, branch (for System 1 and System 3), $R_{dummy}$ (for System 1 and System 2) and their interaction terms were chosen as possible explanatory variables, while $tvib$ was the response variable, where $DBH_{obs}$, $h_{obs}$ and $tvib$ are transformed diameter at breast height over bark, transformed total tree height and transformed stem total volume under bark by the Box-Cox method, respectively (cm, m, m³), $td_{h}$ is the transformed $dh$ by the Box-Cox method, $dh$ is equal to $DBH_{obs} + dBH_{obs} + h$; the other variables have the same specifications as in the bark thickness models. $DBH_{obs}$ of inner layers of the 174 trees were calculated by eqn. 3 (see below), where bark thickness ($bt$) could be obtained from the bark thickness model. Model parameters were estimated by the ordinary least squares method (OLS – eqn. 2): \[
Y = X \beta + Z \mu + \epsilon
\]
where $Y$ is the vector of the response variable; $X$ is the vector of fixed-effect regressors; $Z$ is the vector of random-effect regressors; $\beta$ is the vector of fixed-effect coefficients; $\mu$ is the vector of the random-effect coefficients; $\epsilon$ is the vector of errors. $DBH_{obs}$ (diameter at breast height over bark, cm) was obtained as follows (eqn. 3): \[
DBH_{obs} = DBH + 2 \cdot bt
\]
where $DBH_{obs}$ is the diameter at breast height under bark (cm), breast height is 1.3m height from ground, and $bt$ is the bark thickness (cm).

For each model system, an overall merit-based method was used to select model explanatory variables. Regression equations for bark thickness models and volume models with different variable combinations were compared. Four sets of optimal base equations were obtained by examining the coefficients of determination ($R^2$) and root mean square errors (RMSE); then the Akaike’s information criterion (AIC) was used to successively determine the best random-effects combination and the best residual autocorrelation structure for each optimal base model to obtain four sets of optimal tbt models and tvib models. For those optimal models, residual distribution homogeneity and model bias were visually checked by residual plots with loess regression lines overlaid on the plots. For an unbiased model, a loess line should be flat and located at the zero value on the vertical axis in the residual plot (Jacoby 2000). Normality of residuals was checked with histograms of residuals and by using a Shapiro-Wilk test (probabilities of type I error or $p$-values < 0.05 indicate a departure from a normal distribution). Finally, $bt$ models and $tvib$ models were obtained by back-transforming tbt models and tvib models (see eqn. 4), where $exp$ is the natural exponential function, other notations have the same meanings with those in eqn. 1) and the residuals were also examined (eqn. 4): \[
Y = \left(1 + tV^\lambda + c_1 + c_2 + c_3 + c_4 + c_5 + c_6 + c_7 \right) \exp(V^\lambda) \]
For taper modelling of the four systems, a subset of data (80%) from the analyzed data set. The residual ranges and prediction ranges of the models and $tvib$ models were tested. To sum up, four sets of data (with a bark thickness model, a volume model and a taper model in each of them) were used for modelling and the most suitable set was then selected using the above-mentioned statistics and residual plots.

**Optimal model system evaluation**

After the selection of the optimal model system, representing essentially the best dataset, the transformed bark thickness model ($tbt$) and the transformed volume model ($tvib$) in it were tested by the leave-one-out Jackknife method (Sánchez-González et al. 2005, 2007). The residual ranges and prediction ranges of the models and their corresponding Jackknife tests were compared. Mean biases (Bias) and mean absolute biases (MAD) of the back-transformed $tvib$ model were also assessed for each diameter classes.

For the taper model in the optimal model system, the predictive performance of $dib$ model was evaluated using the validation data set. The residual ranges and prediction ranges based on the validation data set were compared with those based on the fitting data set. Mean biases (Bias) and mean absolute biases (MAD) which were computed respectively using the fit data and validation data were assessed by position (percent relative height, i.e., 5%, 10%, 15%, ..., 95%). Finally, the fitting and validation datasets were combined, the taper model ($Y$ and $dib$) was refitted and the corresponding statistics and plots were examined again (Muhiarwe 1999).

**Results**

**Four sets of models**

The $p$-value, $R^2$ and RMSE of eight models (tbt, tvib, Y, bt vib and dib) in each model system are shown in Tab. 2. $R^2$ values were higher than 0.85 for all the models and values of RMSE of all the models were low compared to the ranges of response variables (Tabs. 1, Tab. 2). Probabilities of type I errors of all the models were lower than 0.05 (Tab. 2). Overall, models in System 4...
Zheng C et al. - iForest 10: 567-575

**Tab. 2** - Values of fitting statistics for eight models in four modeling systems. (bt): bark thickness; (vib): stem total volume under bark; (dib): diameter under bark at height \( h \) (cm); (tbt and (tvib): transformed values of bt and vib by the Box-Cox method; (Y): calculated according to eqn. 8; (p-value): probability of type I error in Shapiro Wilks test (for the tbt and tvib models) and probability of type I error in Kolmogorov-Smirnov test (for the Y models); (R^2): coefficient of determination; (RMSE): root mean square errors.

<table>
<thead>
<tr>
<th>Models type</th>
<th>Model system</th>
<th>Transformed models ( bt ) (cm), ( tvib ) (m^3) or ( Y )</th>
<th>Back-transformed models ( bt ) (cm), ( vib ) (m^3) or ( dib ) (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p-value</td>
<td>( R^2 )</td>
<td>RMSE</td>
</tr>
<tr>
<td>bt (cm)</td>
<td>1, 3</td>
<td>(&lt; 2.2 \times 10^{-16})</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>2, 4</td>
<td>9.646 \times 13</td>
<td>0.94</td>
</tr>
<tr>
<td>vib (m^3)</td>
<td>1</td>
<td>3.343 \times 12</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9.195 \times 08</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6.921 \times 09</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.148 \times 06</td>
<td>0.99</td>
</tr>
<tr>
<td>dib (cm)</td>
<td>1</td>
<td>(&lt; 2.2 \times 10^{-16})</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9.195 \times 08</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.148 \times 06</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>(&lt; 2.2 \times 10^{-16})</td>
<td>0.96</td>
</tr>
</tbody>
</table>

**Tab. 3** - Summaries for the tbt and bt models in System 4. The type III sum of squares was used in those models. (bt): bark thickness (cm); (tbt) and (dib): transformed values of bt and dib by the Box-Cox method, i.e., \( bt = (bt^{1/0.21}/0.21) \) or \( dib = dib^{1/0.81}/0.81 \). (tbt): diameter under bark at height \( h \) (cm); (Rh): relative height and equal to \( h/t \); (H): height from ground (m); (tbt): total tree height (m); (natural): dummy variable (natural = 1 for natural forests; natural = 0 for plantations); (a_i): model parameters (i = 1, 2, 3, ..., n); (SE): standard errors of coefficients; (p): the parameter for first-order autoregressive structure [AR(1)]; (\( \sigma^2 \)): the residual variance; (\( \sigma^2_i \)): the variance for the random effects. (***): p < 0.001.

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation form</th>
<th>Coefficients (± SE)</th>
<th>Predicted value range/Residual range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Model (Entire data)</td>
</tr>
<tr>
<td>tbt</td>
<td>( bt = a_0 + a_1 \cdot tbt + 1^{/0.538} )</td>
<td>( a_0 = -0.881 \pm 0.058*** )</td>
<td>((\text{Entire data}))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( a_1 = -0.277 \pm 0.071*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( a_2 = 0.669 \pm 0.160*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( a_3 = -0.495 \pm 0.101*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( a_4 = 0.278 \pm 0.018*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( a_5 = -3.182 \pm 0.201*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( a_6 = 8.094 \pm 0.882*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( a_7 = -8.695 \pm 1.426*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( a_8 = 3.327 \pm 0.745*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( a_9 = 0.089 \pm 0.022*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \rho = 0.403 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \sigma^2 = 0.025 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \sigma^2_i = 0.013 )</td>
</tr>
</tbody>
</table>

**Tab. 4** - Summaries for the tvib and vib models in System 4. The type III sum of squares was used in those models; (tvib, (th) and (tdh)): transformed values of vib, H and \( d/h \) by the Box-Cox method; i.e., \( \text{tvib} = (vib^{1/0.467})/0.467, \text{th} = (h^{0.467})/0.467, \text{tdh} = (d/h)^{0.467} \); diameter at breast height over bark (cm); (H): total tree height (m); (vib): stem total volume under bark (m^3); (b): model parameters \( (i = 1, 2, 3, ..., n) \); (SE): standard errors of coefficients; (\( \theta \)): the parameter for moving average structure [MA(1)]; (\( \sigma^2 \)): the residual variance; (\( \sigma^2_i \)): the variance for the random effects. (***): p < 0.001.

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation form</th>
<th>Coefficients (± SE)</th>
<th>Predicted value range/Residual range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Model (Entire data)</td>
</tr>
<tr>
<td>tvib</td>
<td>( tvib = b_0 + b_1 \cdot th + b_2 \cdot th^2 )</td>
<td>( b_0 = -4.488 \pm 0.007*** )</td>
<td>((\text{Entire data}))</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( b_1 = 0.125 \pm 0.002**)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( b_2 = -0.021 \pm 0.005*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \theta = 0.467 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \sigma^2 = 0.002 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \sigma^2_i = 0.002 )</td>
</tr>
<tr>
<td>vib</td>
<td>( vib = (0.18 \cdot tvib + 1)^{1/0.21} )</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

had higher \( R^2 \) and lower RMSE (Tab. 2). The loss curves of models in System 4 were closest to the x-axis, followed by those of System 2, then those of System 3, and models in System 1 with the biggest deviation from the x-axis; this means that models in System 4 had much lower bias than those in the other three systems. Therefore, System 4 was the optimal model system (only some of the residuals plots are shown in this paper, see below).

Equation forms, coefficients and standard errors of coefficients of the models in System 4 are shown in Tab. 3, Tab. 4 and Tab. 5. All the coefficients were significant at \( \alpha = 0.001 \) confidence level and standard errors of coefficients were small compared to the coefficients (Tab. 3, Tab. 4 and Tab. 5). A first-order autoregressive structure [AR(1)], a moving average structure [MA(1)] and a combination of first-order autoregressive and moving average struc-
turers [ARMA(1,1)] were respectively elected as the best to address the autocorrelation of the tbt model, tvb model and Y model, respectively (Tab. 3, Tab. 4 and Tab. 5). Predicted value ranges and residual ranges of those models are also shown in Tab. 3, Tab. 4 and Tab. 5. Residual plots of the bark thickness model (bt), volume model (vib) and taper model (dib) in System 4 fitted to the combined datasets are shown in Fig. 1, Fig. 2 and Fig. 3, respectively. In general, the loess lines are represented by flat lines located at the baseline, except for the trees with a relative height (Rh) higher than 0.9 for the residual plot of the taper model. Heteroscedasticity was not obvious in the transformed bark thickness model (tbt), transformed volume model (tvb) and the taper models (Y and dib), while a weak heteroscedasticity could be detected in the back-transformed thickness model (bt) and in the back-transformed volume model (vib). Low p-values of the Shapiro Wilks tests or Kolmogorov-Smirnov tests (Tab. 2) and histograms of residuals (data not shown) suggested that residuals of the models in System 4 did not have a normal distribution. Skewness was not detected in these distributions, but kurtosis was.

Evaluation of the models in System 4

Predicted value ranges and residual ranges from the Jackknife validations for the transformed bark thickness model (tbt) and the transformed volume model (tvb) in System 4 were similar with those obtained from fittings and are shown in Tab. 3 and Tab. 4. Scatter distribution in residual plots from fittings and that from the jackknife validations were similar, as well as the frequency distributions in histograms of residuals (data not shown). Moreover, frequencies, mean bias (Bias) and mean absolute deviation (MAD) of the back-transformed volume model (vib)
For unique forms of volume models have been reported in the literature, such as the model represented by eqn. 10 (see below) proposed by Schumacher & Hall in 1933 (Bailey 1994, Fang & Bailey 1999), eqn. 11 (Malimbwi & Philip 1989), eqn. 12 (Diéguez-Aranda et al. 2006), eq. 13 (Honner 1965) and eqn. 14 (Muhairee 1999). These equations are shown below (eqn. 10 to eqn. 14):

\[
V = a \cdot DBH^2 \cdot H
\]

\[
V = a + a' \cdot DBH^2 \cdot H
\]

\[
V = a \cdot DBH^2 \cdot H
\]

\[
V = DBH^2 \cdot |a + a' \cdot H|
\]

\[
V = e^{\cdot} DBH^{a} \cdot |H^{b} (H - 1.3)|^{c}
\]

where \( V \) is the volume, DBH is the diameter at breast height, \( H \) is the total height, \( a \) terms are model parameters, \( e \) is the base of the natural logarithm. Some of the volume models have a conceptual basis in the geometry of solids of revolution and have a constant form factor (Bailey 1994). In fact, the form of a tree depends upon the actual
tree size, e.g., there was a downward trend of form factor along with increased tree height in our study. Some volume models in some studies explicitly represented the change of a form factor, which was defined as a function of diameter at breast height, total height, stem height at a predetermined fraction of diameter at breast height outside bark, or the ratio of this height to total height (Bailey 1994, Rustagi & Loveless 1991). However, there is always a problem of heterogeneity in those non-linear volume models. For removing heterogeneity, weighted non-linear regression was usually used to estimate parameters (Muhaire 1999). However, computing a suitable weighting variable was awkward. Another simple and common way of removing heterogeneity of residual variance is performing a transformation to stabilize variance. In our study, the Box-Cox transformation was used, then the linear mixed effect equation of the transformed volume was built and no heterogeneity was detected. Meanwhile, bias of the back-transformed volume model was found to be small, and no correction factor was used in this study. Additionally, an overall merit-based method was used to select model explanatory variables for the volume model, so the volume model did not have a conceptual basis in the geometry of solids of revolution and did not explicitly represent change of a form factor.

In the application of a mixed effect model, when a sub-sample of the dataset is available to calculate the random effects, users can calibrate the coefficients of the linear mixed effect model ("Ilme" - Temesgen et al. 2008) and then obtain unbiased predictions. However, predicting the random effects is hard for users. Actually, in our study the bias was found small enough, even though just the fixed effect was considered in prediction; thus, there was no need to calibrate the random effect before using the "Ilme" volume model. Population predictions of volume for a new tree can be obtained using fixed effect coefficients. Similar features can be found in the bark thickness model, which was also a linear mixed effect model using variables transformed by the Box-Cox method.

The simple taper model
According to several studies in the literature (Diéguez-Aranda et al. 2006, Oytunemre et al. 2008, Hjelm 2013), taper equations can be grouped in three types: (1) simple taper equations (Bruce et al. 1968, Kozak et al. 1969, Demaerschalk 1972, Demaerschalk 1975, Ormerod 1973, Goulding & Murray 1976, Fang & Bailey 1979, Sharma & Oderwald 2001); (2) segmented taper equations (Max & Burkhardt 1976, Fang et al. 1980, Parresol et al. 1987, Fang et al. 2000, Jiang et al. 2005); (3) variable exponent taper equations (Kozak 1988, Newham 1992, Bi 2000, Lee et al. 2003, Kozak 2004). Some researchers have pointed out that segmented taper equations and variable exponent taper equations can sometimes provide more flexible descriptions of tree profiles than simple taper equations: variable exponent taper equations usually have the least bias and best predictive abilities among the three kinds of models (Kozak 1988, Newham 1992, Muhaire 1999, Oytunemre et al. 2008). For simpler equations, the presence of larger residuals located in the lower bole (the stump region) is pronounced in some studies (Hjelm 2013). However, a shortcoming of variable exponent taper equations is that they cannot be analytically integrated to calculate total stem or log volumes (Diéguez-Aranda et al. 2006, Ozcêlîk & Brooks 2012). Additionally, segmented taper equations and variable exponent taper equations suffer from statistical complexity, difficulties in parameter estimation and difficulties of being understood and correctly used by forest managers. Therefore, when simplicity of use is an objective, the simple taper model would be a good choice, despite its lower accuracy in the lower bole (Martin 1981).

A polynomial taper equation (Goulding & Murray 1976) was used in this study. Larger residuals were only found at about 90° of stem height (Fig. 3, Fig. 5). The poorer performance observed in predictions at the stem top is negligible from a practical point of view (Figureredo-Filho et al. 1996, Hjelm 2013), as the top part of cork oak is usually collected for bio-fuel. As we are interested only in the middle part of the bole, a simple taper model can be used for practical purposes (Oytunemre et al. 2008).

Ramicorns and relative diameter
Branches are an important aspect of tree form because they affect stem utilization. A ramicorn branch is a steeply angled branch diverging less than 30° from the main stem and significantly smaller than the main stem (Xiong et al. 2014). In this study, the number of trees with ramicorns was very small. Additionally, the values of relative diameter for most of the computed trees in the data set were under 1.5, which were not common in practice. Therefore, models in System 2 were much better than those of System 1 and System 3, and just a little poorer than those of System 4. However, System 2 contained the data of stems with a Rd bigger than 1.5, which were not common in practical application. Therefore, models in System 4 were selected as the most appropriate in terms of precision, lack of bias and practical application. They can be used to predict bark thickness, diameter at a specific point along the stem, merchantable volume and total stem volume of cork oak forests in North China within the specific ranges of DBH (1.6-23.1 cm) or H (1.4-18.2 m).

In System 4, data from four big trees were removed because they had ramicorns. Due to the small sample size for big trees, more big trees should be measured in the future to obtain a compatible taper-volume model system with a larger useable diameter span. It should be noted that if models created using System 4 are used for predictions of stems with ramicorn branches, then errors would be likely greater than those reported here. Therefore, we suggest that models created with System 4 can be used for predictions of stems without ramicorn branches and simultaneously with a relative diameter less than 1.5.

Conclusion
Linear mixed effect equations with tree number as random factor were used for bark thickness and volume modelling using variables transformed by the Box-Cox method to minimise heteroscedasticity. Using the polynomial equation reported by Goulding & Murray (1976), linear mixed effect equations with tree number and natural (a dummy variable specifying the stand origin) as random factors were fitted during taper modelling.

Four sets of compatible taper-volume models systems using different data sets were established and compared. The models in System 4 had superior coefficients of determination (R²), root mean square errors (RMSE) and lack of bias than models from the other three systems and thus were selected as the most suitable in this study. Furthermore, the models in System 4 had good performances in jackknife validation or independent data set validation. Heteroscedasticity was not obvious in the transformed bark thickness model, transformed volume model and the taper
model, while a weak heteroscedasticity could be detected in the back-transformed bark thickness model and back-transformed volume model. Residuals of the models in System 4 did not follow normal distribution. Skewness was not detected in these distributions, but they were slightly kurtotic.

Within the specified ranges of DBH (1.6-23.1 cm) or H (1.4-18.2 m) tested in this study, the compatible taper-volume models system can be used for predicting diameter at a specific point along the stem, merchantable volume and total stem volume of cork oak forests in North China.

Acknowledgements
This research was jointly supported by scientific and research base construction projects of Beijing Municipal Education Commission (YSBL2009), forestry science promotion project of the State Forestry Bureau (2011-44), open fund project of Beijing Forestry University “985” advantage subject innovation platform (000-1180003), special fund project for forestry public service industry and research (201004201) and China Scholarship Council. We acknowledge the strong support from Zhong Tiaoshan National Forest Authority, Xingtai County Forestry Bureau, SI Zuolou forestry station and Xi Shan forestry station in Beijing.

Conghui Zheng and Yuzhong Wang have equally contributed to this work and should be regarded as co-first authors.

References
Ozcitlik R, Brooks JR (2012). Compatible volume and taper models for economically important tree species of Turkey. Annals of Forest Science 69 (1): 105-118. doi:10.1051/forest/201103-017-4
Supplementary Material

Tab. S1 – Summary statistics of four data sets used for modelling.

Tab. S2 – Values of fitting statistics for eight models (tbt, tvib, Y, bt, vib and dib).

Tab. S3 – Results of tbt and bt models.

Tab. S4 – Results of tvib and vib models.

Tab. S5 – Results of Y and dib models.

Tab. S6 – The estimated taper function (red dotted curve) and the basic taper data for each system.

Tab. S7 – Frequencies, Bias and MAD of the vib model in system 4.

Tab. S8 – Frequencies, Bias and MAD of dib model in system 4.

Link: Zheng_2114@suppl001.pdf