

Bayesian meta-analysis of regional biomass factors for *Quercus mongolica* forests in South Korea

Tzeng Yih Lam^{1,2} · Xiaodong Li¹ · Rae Hyun Kim¹ · Kyeong Hak Lee¹ · Yeong Mo Son¹

Received: 22 November 2013 / Accepted: 30 September 2014 / Published online: 21 June 2015
 © Northeast Forestry University and Springer-Verlag Berlin Heidelberg 2015

Abstract Indirect approaches to estimation of biomass factors are often applied to measure carbon flux in the forestry sector. An assumption underlying a country-level carbon stock estimate is the representativeness of these factors. Although intensive studies have been conducted to quantify biomass factors, each study typically covers a limited geographic area. The goal of this study was to employ a meta-analysis approach to develop regional biomass factors for *Quercus mongolica* forests in South Korea. The biomass factors of interest were biomass conversion and expansion factor (BCEF), biomass expansion factor (BEF) and root-to-shoot ratio (RSR). Our objectives were to select probability density functions (PDFs) that best fitted the three biomass factors and to quantify their means and uncertainties. A total of 12 scientific publications were selected as data sources based on a set of criteria. From

these publications we chose 52 study sites spread out across South Korea. The statistical model for the meta-analysis was a multilevel model with publication (data source) as the nesting factor specified under the Bayesian framework. Gamma, Log-normal and Weibull PDFs were evaluated. The Log-normal PDF yielded the best quantitative and qualitative fit for the three biomass factors. However, a poor fit of the PDF to the long right tail of observed BEF and RSR distributions was apparent. The median posterior estimates for means and 95 % credible intervals for BCEF, BEF and RSR across all 12 publications were 1.016 (0.800–1.299), 1.414 (1.304–1.560) and 0.260 (0.200–0.335), respectively. The Log-normal PDF proved useful for estimating carbon stock of *Q. mongolica* forests on a regional scale and for uncertainty analysis based on Monte Carlo simulation.

Project fund: This study was supported by Korea Forest Research Institute.

The online version is available at <http://www.springerlink.com>

Corresponding editor: Chai Ruihai

Electronic supplementary material The online version of this article (doi:10.1007/s11676-015-0089-x) contains supplementary material, which is available to authorized users.

✉ Tzeng Yih Lam
 tylam@ntu.edu.tw

¹ Division of Forest Management, Department of Forest and Climate Change, Korea Forest Research Institute, 57 Hoegiro, Dongdaemun-gu, Seoul 130-712, South Korea

² Present Address: School of Forestry and Resource Conservation, National Taiwan University, No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan, People's Republic of China

Keywords Uncertainty analysis · Monte Carlo simulation · Bayesian hierarchical model · Nesting structure · Biomass estimation

Introduction

Estimation of carbon flux has become important for countries seeking to comply with their agreements under the United Nations Framework Convention on Climate Change (UNFCCC) and to implement their commitments under the Kyoto Protocol (Van Camp et al. 2004). Unlike in other sectors, direct measurement of carbon flux in the land use, land use change and forestry (LULUCF) sector is difficult, and indirect methods often must be used (Somygi et al. 2007). One common indirect method is to first estimate forest biomass and then convert it to carbon stock based on the carbon fraction. Calculation of forest biomass

involves converting forest inventory estimates of the volume of growing stock by using conversion techniques such as biomass factors and biomass equations (Jalkanen et al. 2005). Biomass factors convert, expand or reduce stand volume estimates to biomass estimates, and biomass equations predict biomass as a function of stand structural variables (Somogyi et al. 2007). The factor approach is the Tier 1 method in the Intergovernmental Panel on Climate Change (IPCC) guidelines (IPCC 2006b), which is adopted by many countries in their reporting. Hence, reliable estimates of growing stock and biomass factors are crucial for accurately estimating changes in forest biomass and carbon stock.

An implicit assumption underlying a country-level carbon stock estimate is the representativeness of the growing stock estimate and biomass factors. Growing stock is generally derived from a national forest inventory (NFI), whose design is usually probability-based and covers a large geographic area (Schroeder et al. 1997). Thus, the estimate is statistically valid and representative on the country-level. However, most country NFIs are primarily intended for monitoring economically interesting wood volumes for the purpose of making sound management plans (Van Camp et al. 2004). Furthermore, NFIs generally do not measure non-commercial tree components such as branches, foliage and roots (Schroeder et al. 1997). Thus, biomass factors are relied upon to adequately quantify above- and below-ground biomass.

Developing representative biomass factors is a challenging task. Intensive studies have been conducted to quantify biomass and develop factors, but each study normally covers a limited geographical range (Schroeder et al. 1997). Due to logistic and cost constraints, most countries undertake only a few studies at few sites (Van Camp et al. 2004). In some instances, forests with greater than average biomass for a region tend to be selected for study, which results in an overestimation of national biomass (Brown and Lugo 1982; Fang and Wang 2001). Furthermore, biomass factors can vary greatly depending on location and forest structure (Brown 2002; Návar 2010). Therefore, applying biomass factors from any single study to predict country-level biomass is hindered by lack of representativeness. In contrast, combining information from all available studies could, by increasing the scope of inference, enhance the reliability of the biomass factors. The method of meta-analysis offers an opportunity for such an analysis.

Meta-analysis is a statistical synthesis of results from a series of studies (Borenstein et al. 2009). Its application is widespread in medical and social sciences but less so in natural science. The process is similar to conducting any research, i.e., research formulation, data collection, evaluation, analysis and interpretation (Hartung et al. 2008). An

important aspect is unequivocal definition of the population to which generalization will be made (Hedges 1994). This will drive the search for relevant sources of information, many of which will be scientific publications. However, publication bias due to a lack of reporting on research with insignificant conclusions has to be taken into account (Hartung et al. 2008), and analysis will only be meaningful if information sources have been collected systematically (Borenstein et al. 2009). A wide range of statistical models is available to combine published results, and the choice of model depends on the nature of the underlying problem (Hartung et al. 2008). In short, successful application of meta-analysis requires adequate understanding of these stages.

Several studies illustrated the effort to access a large number of sources to construct a dataset for meta-analysis of biomass factors and equations. Teobaldelli et al. (2009) developed generalized biomass equations as a function of age, growing stock and site index for various species groups from two extensive datasets containing a total of 6392 plots. The datasets were from Cannell (1982) covering temperate forest stands from 46 countries and from Usoltsev (2001) covering 37 countries in northern Eurasia. Wirth et al. (2004) compiled a dataset of 688 trees sampled in 102 forest stands by 19 authors to develop biomass equations for Norway spruce in central Europe. In an extensive effort to develop national-scale biomass equations for 100 tree species in United States, Jenkins et al. (2003) assembled 318 equations of total biomass and 389 equations of component biomass from 104 publications. Their generalized biomass equations were implemented in the FORCARB model to develop the U.S. carbon budget (Jenkins et al. 2003). In U.K., Levy et al. (2004) developed landscape-level biomass factors and prediction functions for 13 coniferous tree species from 2000 sampled trees covering a range of ages, soil types and site cultivation methods.

The goal of this study was to use the meta-analysis approach to develop regional stand-level biomass factors for *Quercus mongolica* (Fisch.) forests in South Korea. This oak species is found in natural deciduous and mixed forests throughout South Korea (Son et al. 2007). It plays an important role in increasing biodiversity and wood production, and has cultural significance across the country (Li et al. 2010). Hence, estimating carbon stocks in *Q. mongolica* forests is relevant to developing the country carbon budget. The three biomass factors of interest are biomass conversion and expansion factor (BCEF), biomass expansion factor (BEF) and root-to-shoot ratio (RSR). The factors are defined according to IPCC (2006b). BCEF is defined as the ratio of above-ground biomass to stand volume (ton/m^3), BEF is the ratio of above-ground biomass to stem biomass, and RSR is the ratio of root biomass to

above-ground biomass. The latter two factors are dimensionless. The biomass factors are positive continuous variables, and their distributions are generally skewed towards low values (Levy et al. 2004). Therefore, a normal probability density function (PDF) may not be proper and the biomass factors are usually assumed to be log-normally distributed. Specification of an appropriate PDF for a biomass factor is important if one wishes to carry out uncertainty analysis using the Monte Carlo simulation approach recommended by the IPCC (IPCC 2006a). Hence, the two specific objectives of this study were (1) exploring PDFs and selecting the one that provided the best fit for each biomass factor, and (2) quantifying their means and uncertainties.

Materials and methods

Data

A literature search was carried out to collect publications on stand-level biomass estimates of *Q. mongolica* forests. A set of criteria was laid out to select appropriate publications to be included in the meta-analysis: (1) forest type, (2) plot configuration, (3) sample tree selection procedure, and (4) measurement method. A total of 12 publications consistent in these selection criteria were included in our meta-analysis (Table 1). The number of study sites per publication ranged from 1 to 16 and totalled 52 (Table 1). The study sites were located either in pure *Q. mongolica* forests or in mixed forests with high density of the oak species. They were distributed across the country (Fig. 1) and covered a range of elevations (260–1300 m) and stand ages (6–70 years). All publications established 20 m × 20 m (0.04 ha) plots with the exception of Park and Moon (1994) and Park et al. (1996), who established smaller plots of 10 m × 10 m (0.01 ha).

The sample tree selection and measurement procedures were generally identical among the 12 publications. All trees with diameter at breast height (DBH) ≥ 6 cm in a plot were measured for DBH and height, and 3–6 trees were destructively sampled for biomass based on the observed DBH distribution. This ensured the sampled trees represented stand structure in the plot. Stem, branch and foliage biomasses were measured, and above-ground biomass was the sum of the three biomass components. Root biomass was measured for a subset of the sampled trees. Biomass equations relating stem, above-ground and root biomasses to DBH and height were developed and were used to predict biomasses for the remaining trees in the plot. Stand-level stem, above-ground and root biomass were estimated as the sum of all tree biomasses respectively and were scaled to biomass densities (ton/ha). Not every publication

reported stand volume (m^3/ha), which was needed to estimate one of the biomass factors. For those without the information, stand volume was estimated by multiplying stand density and mean tree volume, where the latter was estimated from mean tree dbh and height reported in the publications using a published volume table (Korea Forest Service 2009). The stand-level above-ground and stem biomass for each study site of each publication are presented in Table S1 (Supplementary Material). The three stand-level biomass factors—BCEF, BEF and RSR—were then estimated from the stand biomasses and volume. A total of 12, 11 and 9 publications were available for estimating BCEF, BEF and RSR, respectively because not all 12 publications reported stem and root biomasses.

Statistical models

The statistical model for the meta-analysis of BCEF, BEF and RSR was formulated as a multilevel model under the Bayesian hierarchical model (BHM) framework. A multilevel model accommodates nesting structure in data. In this study, publication was chosen as the nesting factor, and respective study sites were nested under a publication (Table 1). This accounted for the variation between study sites within a publication. The multilevel model under BHM framework also incorporated between-publication variation. Difference in data collection methods was minimized by application of the criteria for selection of publications. However, other subtle factors specific to each publication might influence the biomass factor estimates, e.g., equipment used, field crew experience and field work conditions. In short, the multilevel BHM model helps us to derive an average estimate of a biomass factor on a regional level while accounting for variation between publications and between study sites within a publication (Gelman and Hill 2007). The BHM framework has several advantages. For example, complex random effects can be easily constructed for a variety of PDFs, and BHM can accommodate small sample sizes of some publications and use all available data to gather inferences for them (Congdon 2006; Gelman and Hill 2007).

BCEF, BEF and RSR are positive continuous variables, and their frequency distributions indicate positive skew (Fig. 2). Fitting a Normal PDF to their frequency distributions was unsatisfactory because the fitted PDF failed to capture the asymmetry and portions of the fitted PDF took on negative values (Fig. 2). Thus, a PDF that meets the distributional characteristics of the three biomass factors is more appropriate. In this study, we chose the Gamma, Log-normal and Weibull PDFs because of their ability to accommodate skewed and strictly positive variables. These PDFs were then specified as multilevel BHM models to derive regional level BCEF, BEF and RSR.

Table 1 List of publications used for meta-analysis of BCEF, BEF and RSR and average characteristics of the study sites in the publications

Publication	No. of study sites	No. of sample trees ^a	Plot dimension (m)	Mean age (year)	Mean stand density (no/ha)	Mean DBH (cm)	Mean height (m)	Mean stand volume (m ³ /ha)
Kwon and Lee (2006a)	6	37	20 × 20	52	1325	17.7	13.6	221.8
Kwon and Lee (2006b)	4	25	20 × 20	50	1920	14.8	10.4	196.0
Kwon and Lee (2006c)	12	18	20 × 20	54	1700	16.4	11.9	207.3
Lee and Kwon (2006)	3	23	20 × 20	47	1600	15.5	13.4	193.0
Lee and Park (1987)	1	10	10 × 10	22	1600	12.4	10.2	96.0
Park (2003)	3	27	20 × 20	43	594	18.7	11.0	240.0
Park and Moon (1994)	1	10	20 × 20	36	1040	12.9	9.7	72.0
Park et al. (1996)	1	10	10 × 10	34	705	15.0	11.6	76.0
Park et al. (2005a)	3	30	20 × 20	48	908	14.7	14.5	112.7
Park et al. (2005b)	1	10	20 × 20	36	3175	9.0	12.7	158.0
Song and Lee (1996)	1	10	20 × 20	67	875	24.0	11.8	131.0
KFRI (2010)	16	70	20 × 20	48	–	17.4	11.6	–

^a Number of trees sampled for destructive sampling of biomass

Let x_{ij} be a biomass factor (BCEF, BEF or RSR) of j th study site of i th publication, m be the number of publications for a biomass factor ($m = 12, 11$ and 9 for BCEF, BEF and RSR, respectively). Assuming x_{ij} followed a Gamma PDF, let α_i and β_i be the shape and rate parameters for the i th publication, respectively; thus, the PDF of x_{ij} was $f(x_{ij}) = \beta_i^{\alpha_i} x_{ij}^{\alpha_i-1} \exp(-\beta_i x_{ij}) / \Gamma(\alpha_i)$ (Ntzoufras 2009). The multilevel BHM based on the Gamma PDF was specified as,

$$\begin{aligned} x_{ij} | \alpha_i, \beta_i &\sim \Gamma(\alpha_i, \beta_i) \\ \alpha_i | A_\alpha, B_\alpha &\sim \Gamma(A_\alpha, B_\alpha) \\ \beta_i | A_\beta, B_\beta &\sim \Gamma(A_\beta, B_\beta) \\ A_\alpha, B_\alpha, A_\beta, B_\beta &\sim \Gamma(0.001, 0.001) \end{aligned} \tag{1}$$

where, $\Gamma()$ was a Gamma function. Equation (1) implied that a Gamma PDF was fitted to each publication with biomass factor values (x_{ij}) from that publication; hence, the nesting structure. Under the BHM framework, α_i and β_i were assumed to follow independent Gamma distributions. A non-informative Gamma prior distribution was independently assigned to each of the hyperparameters ($A_\alpha, B_\alpha, A_\beta, B_\beta$). In short, the posterior distribution of the parameter set $\Omega = (\alpha, \beta, A_\alpha, B_\alpha, A_\beta, B_\beta)$ was,

$$\begin{aligned} P(\Omega | \mathbf{x}) &\propto L(\mathbf{x} | \alpha, \beta) P(\alpha | A_\alpha, B_\alpha) P(\beta | A_\beta, B_\beta) \\ &P(A_\alpha) P(B_\alpha) P(A_\beta) P(B_\beta) \end{aligned} \tag{2}$$

where, \mathbf{x} for all x_{ij} , $\alpha = (\alpha_1, \dots, \alpha_m)$, $\beta = (\beta_1, \dots, \beta_m)$, $L(\mathbf{x} | \alpha, \beta)$ was a likelihood function, and $P(.|.)$ and $P(.)$ were prior distributions for parameters and hyperparameters, respectively. The posterior mean biomass factor across all publications was estimated as (Ntzoufras 2009),

$$\tilde{\mu}_x = (\tilde{A}_\alpha / \tilde{B}_\alpha) / (\tilde{A}_\beta / B_\beta) \tag{3}$$

with the posterior variance of the biomass factor as,

$$\tilde{\sigma}_x^2 = (\tilde{A}_\alpha / \tilde{B}_\alpha) / (\tilde{A}_\beta / \tilde{B}_\beta)^2 \tag{4}$$

Secondly for Log-normal PDF, let μ_i be the mean parameter on a logarithmic-scale and τ be the precision parameter for the i th publication. Thus, the PDF of x_{ij} was specified as $f(x_{ij}) = \sqrt{\tau / 2\pi} x_{ij}^{-1} \exp(-\tau(\log x_{ij} - \mu_i)^2 / 2)$ (Ntzoufras 2009). The multilevel BHM based on the Log-normal PDF was specified following Ntzoufras (2009) as,

$$\begin{aligned} x_{ij} | \mu_i, \tau &\sim LN(\mu_i, \tau) \\ \mu_i | \theta, \varphi &\sim N(\theta, \varphi) \\ \theta &\sim N(0, 1000000) \\ 1/\sqrt{\tau}, 1/\sqrt{\varphi} &\sim U(0, 100) \end{aligned} \tag{5}$$

where, $LN()$ was Log-normal PDF, $N()$ was Normal PDF, $U()$ was Uniform PDF. Equation (5) implied that only the mean parameter μ was subjected to the nesting structure. We attempted to add the nesting structure on the precision parameter τ , but the model was unstable and failed to converge satisfactorily for the three biomass factors. Hence, we assumed that τ was not influenced by between-publication variation. The parameter μ_i was assumed to be normally distributed. Non-informative prior distributions were independently assigned to θ, τ and φ with θ assuming a Normal distribution, the inverse and square-root of precision parameter τ and hyperparameter φ were assumed to have Uniform distributions following the convention from Gelman and Hill (2007). In short, the posterior distribution of the parameter set $\Omega = (\mu, \tau, \theta, \varphi)$ was,

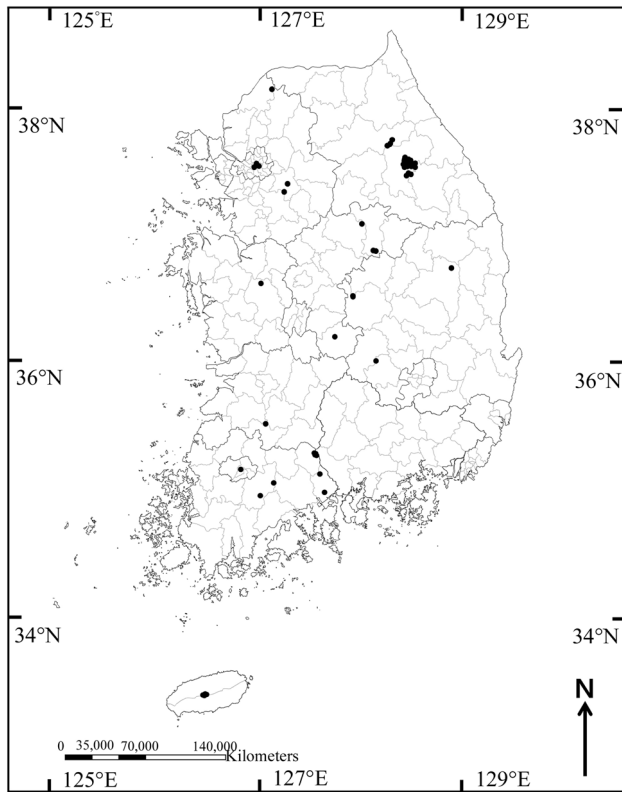


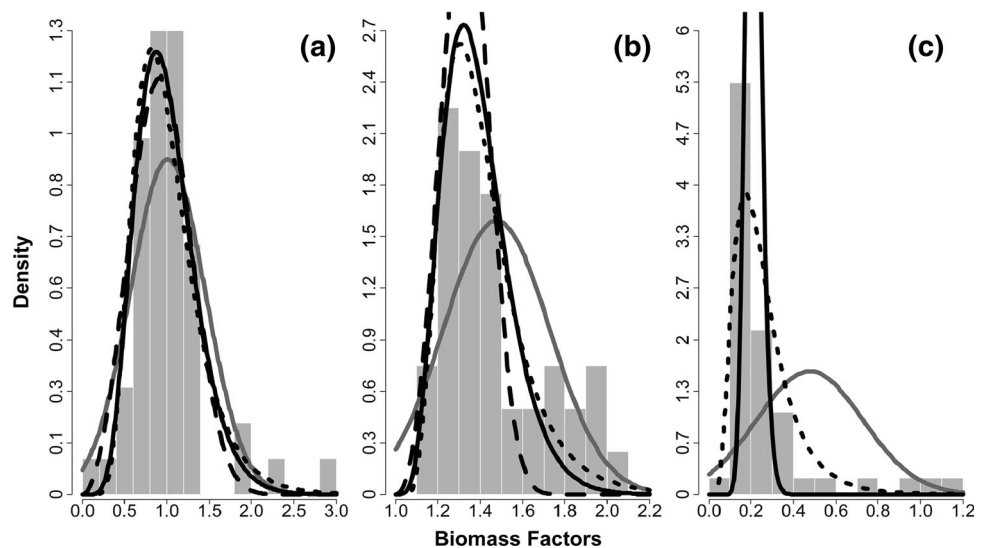
Fig. 1 Locations of 52 study sites (black dots) from the 12 publications used for meta-analysis in the Republic of Korea

$$P(\mathbf{\Omega}|\mathbf{x}) \propto L(\mathbf{x}|\mathbf{\mu}, \tau)P(\mathbf{\mu}|\theta, \varphi)P(\theta)P(1/\sqrt{\tau})P(1/\sqrt{\varphi}) \quad (6)$$

where, $\mathbf{\mu} = (\mu_1, \dots, \mu_m)$. The posterior mean biomass factor across all publications was estimated as (Ntzoufras 2009),

$$\tilde{\mu}_x = \exp\left(\tilde{\theta} + 1/2\tilde{\tau}\right) \quad (7)$$

Fig. 2 Frequency distribution of published **a** BCEF, **b** BEF, and **c** RSR. The fitted PDFs are Normal (gray solid line), Gamma (black solid line), Log-normal (black dotted line) and Weibull (black dashed line). The peaks of the fitted Weibull and Gamma PDFs for BEF and RSR, respectively, are beyond the scale of the vertical axis and are not shown



with the posterior variance of the biomass factor as,

$$\tilde{\sigma}_x^2 = (\exp(1/\tilde{\tau}) - 1) \exp\left(2\tilde{\theta} + 1/\tilde{\tau}\right) \quad (8)$$

Lastly assuming x_{ij} was distributed as a Weibull PDF, let v_i and λ_i be the parameters associated with the PDF for the i th publication. The PDF of x_{ij} was specified as $f(x_{ij}) = v_i \lambda_i x_{ij}^{v_i-1} \exp(-\lambda_i x_{ij}^{v_i})$ (Ntzoufras 2009). The multilevel BHM based on the Weibull PDF was,

$$\begin{aligned} x_{ij}|v_i, \lambda_i &\sim WB(v_i, \lambda_i) \\ v_i|A_v, B_v &\sim \Gamma(A_v, B_v) \\ \lambda_i|A_\lambda, B_\lambda &\sim \Gamma(A_\lambda, B_\lambda) \\ A_v, B_v, A_\lambda, B_\lambda &\sim \Gamma(0.001, 0.001) \end{aligned} \quad (9)$$

where, $WB()$ was Weibull PDF. Equation (9) assumed that both parameters of the Weibull PDF were Gamma distributed. A non-informative Gamma prior distribution was independently assigned to each of the hyperparameters ($A_v, B_v, A_\lambda, B_\lambda$). The resulting posterior distribution of the parameter set $\mathbf{\Omega} = (v, \lambda, A_v, B_v, A_\lambda, B_\lambda)$ was,

$$\begin{aligned} P(\mathbf{\Omega}|\mathbf{x}) &\propto L(\mathbf{x}|\mathbf{v}, \lambda)P(\mathbf{v}|A_v, B_v)P(\lambda|A_\lambda, B_\lambda) \\ &P(A_v)P(B_v)P(A_\lambda)P(B_\lambda) \end{aligned} \quad (10)$$

where, $\mathbf{v} = (v_1, \dots, v_m)$ and $\lambda = (\lambda_1, \dots, \lambda_m)$. The posterior mean biomass factor across all publications was estimated as (Ntzoufras 2009),

$$\tilde{\mu}_x = (\tilde{A}_\lambda/\tilde{B}_\lambda)^{-1/(\tilde{A}_v/\tilde{B}_v)} \Gamma(1 + 1/(\tilde{A}_v/\tilde{B}_v)) \quad (11)$$

with the posterior variance of the biomass factor as,

$$\tilde{\sigma}_x^2 = \left[\Gamma(1 + 2/(\tilde{A}_v/\tilde{B}_v)) - \Gamma(1 + 1/(\tilde{A}_v/\tilde{B}_v))^2 \right] (\tilde{A}_\lambda/\tilde{B}_\lambda)^{-2/(\tilde{A}_v/\tilde{B}_v)} \quad (12)$$

Model fitting

Modeling was done using the Monte Carlo Markov Chain (MCMC) simulation in WinBUGS (Lunn et al. 2000). A common algorithm to run MCMC simulation is the Gibbs sampler introduced by Metropolis et al. (Metropolis et al. 1953). The general process of the algorithm is: (1) start with initial values for all parameters, (2) choose the number of iterations, (3) at each iteration, update one parameter at a time by taking a random simulation draw given the data and the current estimates of all other parameters, and (4) once all iterations are completed, evaluate the chain for convergence (Gelman and Hill 2007). For this study, two disparate sets of initial values were used. The MCMC simulation was run with 300,000 iterations and a lag of 5 to reduce possible autocorrelation between the MCMC runs. At the end of the run, values from the initial 10,000 iterations were dropped, and MCMC samples for some parameters were thinned to further reduce autocorrelation.

The model convergence was judged from: (1) trace and history plots of the MCMC samples, (2) the Brooks–Gelman–Rubin (BGR) statistic (Brooks and Gelman 1998), and (3) the Monte Carlo (MC) error of an estimated parameter. Convergence of a fitted model was confirmed if: (1) the BGR statistic was <1.05 , (2) the MC error was $<2\%$ of the posterior standard deviation of the estimated parameter, and (3) the trace and history plots showed MCMC samples to be within a parallel zone without strong seasonality (Ntzoufras 2009).

A total of 9 analyses were carried out using the combination of three PDFs (Gamma, Log-normal and Weibull) and three biomass factors (BCEF, BEF and RSR). Modeling of BEF raised a location issue because the lower bound of BEF was 1 while that of the three PDFs was 0. To resolve this issue and avoid specifying complex truncated PDFs, the BEF values were subtracted by 1 prior to analysis. Shifting the lower bound of BEF to 0 did not affect the shape of its frequency distribution. The model fitting showed that model convergence was achieved for 8 of the 9 combinations, except the Weibull PDF and RSR combination. Hence, RSR was only fitted with the Gamma and Log-normal PDFs. The WinBUGS codes along with the data used for analyses are made available in the Supplementary Material (S1–S3; Supplementary Material).

The median, 2.5 and 97.5 % sample quantiles of the posterior mean and variance of a biomass factor were reported for the eight combinations. Comparison of fit between PDFs for a biomass factor was based on: (1) deviance information criterion (DIC, Spiegelhalter et al. 2002), (2) Kolmogorov–Smirnov test, and (3) Q–Q plot. The DIC was analogous to Akaike information criterion (AIC) with smaller values indicating better fit. The

Kolmogorov–Smirnov test compared the fit between the distributions of the observed biomass factor values and that of a fitted PDF from posterior median parameters associated with the PDF. The null hypothesis of the test was no difference between the two, and significance level was set at 0.05. Finally, the Q–Q plot was constructed by plotting the quantiles of the observed biomass factor values on the horizontal axis against the quantiles from a fitted PDF on the vertical axis. For a good fit, the plotted quantiles would fall on line of slope 1:1.

Results

Biomass conversion and expansion factor (BCEF)

The median posterior mean BCEF across all publications estimated by Gamma, Log-normal and Weibull PDFs were 1.003, 1.016 and 0.937, respectively (Table 2). Although the values appeared to be different, the 95 % credible intervals (range in 2.5 and 97.5 % sample quantiles) showed that the difference was insignificant (Table 2). The goodness-of-fit statistics suggested that Gamma PDF had the best fit followed by Log-normal and Weibull PDFs. The DIC for the fitted Gamma PDF was lowest (45.17) while that of the Weibull PDF was highest (48.84, Table 2). The Kolmogorov–Smirnov test suggested that all PDFs fitted the observed BCEF distribution adequately with Gamma PDF having the best fit (p value = 0.933, Table 2). This was supported by the fitted PDF curves overlaying the observed BCEF distribution (Fig. 2a).

The diagnostic plots suggested that the fits of the Gamma and Log-normal PDFs were comparable while that of the Weibull PDF was poorer (Fig. 3a, b, c). The Q–Q plot of the Gamma PDF indicated that the fitted PDF could not adequately capture the longer right tail of the observed BCEF distribution, i.e., quantiles were below the 1:1 line for BCEF larger than 1.5 (Fig. 3a). On the other hand, the fitted Log-normal PDF performed better at the right tail but slightly underestimated when BCEF was between 1.0 and 1.5 (Fig. 3b). The Weibull PDF overestimated BCEF when its value was less than 0.8 and greater than 1.5 (Fig. 3c), which was also supported by the fitted Weibull curve over the BCEF frequency distribution (Fig. 2a).

Posterior estimate of BCEF varied greatly from publication to publication for all PDFs (Fig. 3a, b, c). The estimate could be twice as large depending on the publication, with values ranging from 0.6 to 1.5. Most estimates were either equal to or larger than the posterior mean estimate across all publications. The width of the 95 % credible intervals also depended on the publications. Median, 2.5 and 97.5 % sample quantiles of

Table 2 Median posterior mean and variance of the three biomass factors across all publications with corresponding 2.5 and 97.5 % sample quantiles in parenthesis for each PDF

PDF	BCEF	BEF ^a	R
Gamma			
Posterior mean	1.003 (0.786, 1.236)	1.387 (1.277, 1.513)	0.217 (0.077, 0.564)
Posterior variance	0.125 (0.065, 0.231)	0.025 (0.011, 0.058)	0.001 (0.0002, 0.007)
DIC	45.17	-15.91	-103.10
p value ^b	0.933	0.030*	0.012*
Log-normal			
Posterior mean	1.016 (0.800, 1.299)	1.414 (1.304, 1.560)	0.260 (0.200, 0.335)
Posterior variance	0.160 (0.079, 0.343)	0.038 (0.016, 0.106)	0.021 (0.010, 0.049)
DIC	46.85	-13.20	-60.68
p value ^b	0.841	0.132	0.016*
Weibull			
Posterior mean	0.937 (0.710, 1.118)	1.332 (1.231, 1.438)	na ^c
Posterior variance	0.114 (0.052, 0.227)	0.012 (0.004, 0.031)	na
DIC	48.84	-31.51	na
p value ^b	0.617	0.002*	na

* p value <0.05 was indicated

^a BEF was expressed in its original scale

^b Kolmogorov–Smirnov test

^c Weibull model failed to converge for RSR

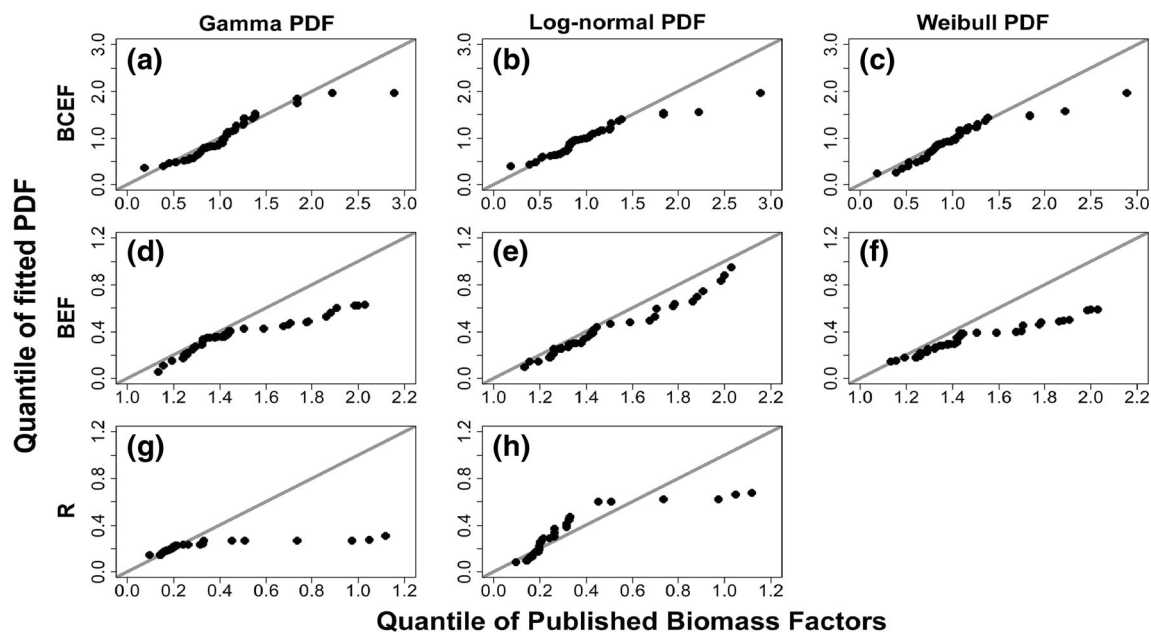


Fig. 3 Q–Q plots for the combinations of the three biomass factors and three PDFs. The three biomass factors are BCEF (a, b, c), BEF (d, e, f) and RSR (g, h), and the three PDFs are Gamma (a, d, g), Log-normal (b, e, h) and Weibull (c, f). The gray line depicts the 1:1 line

hyperparameters for each of the three fitted PDFs are presented in Table S2 (Supplementary Material).

Biomass expansion factor (BEF)

The median posterior mean BEF across all publications estimated by Gamma, Log-normal and Weibull PDFs were

1.387, 1.414 and 1.332, respectively (Table 2). The 95 % credible intervals showed that the differences in the median posterior means were insignificant (Table 2). The goodness-of-fit statistics suggested that the Log-normal PDF had the best fit to the observed BEF distribution (DIC = -13.20 and p value = 0.132, Table 2) while the fit of the Weibull PDF was considerably poorer

(DIC = -31.51 and p value = 0.002 , Table 2). The Gamma PDF had an intermediate fit but the Kolmogorov–Smirnov test showed significant difference between the fitted PDF and the observed BEF distribution (p value = 0.030 , Table 2).

The diagnostic plots showed a consistent result that the three PDFs fitted poorly to the right tail of the observed BEF distribution (Fig. 3d, e, f). For the Gamma and Log-normal PDFs, the Q–Q plots indicated overestimation for BEF greater than 1.5 with the quantiles from the fitted PDF lower than expected (Fig. 3d, e). Lastly, the Q–Q plot indicated a poor fit of the Weibull PDF such that the quantiles were smaller than expected for most BEF values (Fig. 3f). Furthermore, the peak of the fitted curve over the BE frequency distribution was beyond the scale (Fig. 2b).

Posterior estimates of BEF showed little variation among publications for all PDFs except for two publications (Fig. 4e, f, g). Furthermore, most estimates approached the posterior mean estimate across all publications. The width of 95 % credible intervals was generally quite similar among the posterior estimates of BEF by publication. Median, 2.5 and 97.5 % sample quantiles of hyperparameters for each of three fitted PDFs are presented in Table S3 (Supplementary Material).

Root-to-shoot ratio (RSR)

The median posterior means for RSR across all publications were 0.217 and 0.260 for Gamma and Log-normal PDFs, respectively (Table 2). The 95 % credible interval of the Gamma PDF was extremely wide, even covering the 95 % credible interval of the Log-normal PDF (Table 2). This suggested that the estimate of the Gamma PDF was highly uncertain.

The goodness-of-fit statistics suggested that neither PDF fit the observed RSR distribution well, but the Log-normal PDF had a better fit (Table 2). DIC from the Gamma PDF was almost twice that of the Log-normal PDF (-103.10 and -60.68 , respectively, Table 2). The Kolmogorov–Smirnov test showed the fits of both PDFs were similar (p values = 0.012 and 0.016 , respectively, Table 2).

The results from Q–Q plots were in agreement with the test statistics (Fig. 3g, h). Neither PDF generally yielded a good fit with observed RSR values greater than 0.2, but the lack of fit differed slightly between them. The Gamma PDF underestimated the quantiles of RSR values greater than 0.2 (Fig. 3g). In contrast, the Log-normal PDF overestimated the quantiles for values between 0.2 and 0.5 but behaved similarly to the Gamma PDF for values greater 0.5 (Fig. 3h).

Posterior estimates of RSR by publications differed between the Gamma and Log-normal PDFs (Fig. 4g, h). For the Gamma PDF, the median posterior estimates for

three publications were significantly or nearly significantly larger than those of the other publications (Fig. 4g). Furthermore, the 95 % credible interval was highly variable among publications. On the contrary, the median posterior estimates and 95 % credible intervals were similar among the publications for the Log-normal PDF (Fig. 4h). Median, 2.5 and 97.5 % sample quantiles of hyperparameters for each of two fitted PDFs are presented in Table S4 (Supplementary Material).

Discussion

The choice of the most suitable PDF for a biomass factor depends on quantitative and qualitative evaluations. For BCEF, the fit of the Gamma and Log-normal PDFs was comparable, but we favored the Log-normal PDF because it had a better fit to the right tail of the observed BCEF distribution. The likelihood of the observed large BCEF values (i.e., greater than 1.5) being outliers was lowered due to the predefined set of criteria for selection of publications. These values likely represented specific *Q. mongolica* forest stands as part of a larger population in the region, e.g., a young and dense forest stand with high amount of non-commercial tree components relative to stem volume. Thus, a PDF that could capture these values would be preferable. However, it should be noted that the Log-normal PDF performed less than ideally for BCEF values between 1.0 and 1.5, which was the price for capturing the long right tail of the observed distribution. This perhaps caused the goodness-of-fit statistics to be lower for the Log-normal PDF than for the Gamma PDF.

The Log-normal PDF was also recommended for BEF. The goodness-of-fit statistics and the qualitative Q–Q plots supported the choice of the Log-normal PDF. Nonetheless, the lack of fit was apparent for BEF values greater than 1.5 due to higher frequency of these values than expected under the Log-normal PDF. The lack of fit, which was consistent among the three PDFs, suggests that the observed BEF distribution may be a mixture of multiple PDFs. If so, two sub-populations of BEF might be present with one sub-population peaking around BEF of 1.2 and another sub-population peaking at 1.8. The two sub-populations might represent different *Q. mongolica* stand structures on the landscape with the former group having smaller non-commercial tree component biomass than the latter group relative to stem biomass. A finite mixture model might be more appropriate for modeling a distribution with sub-populations. A finite mixture model mixes a number of PDFs to generate a model that is flexible in accommodating a variety of random phenomena (McLachlan and Peel 2000). For BEF, it could be a mixture of two Normal PDFs to represent the two sub-populations

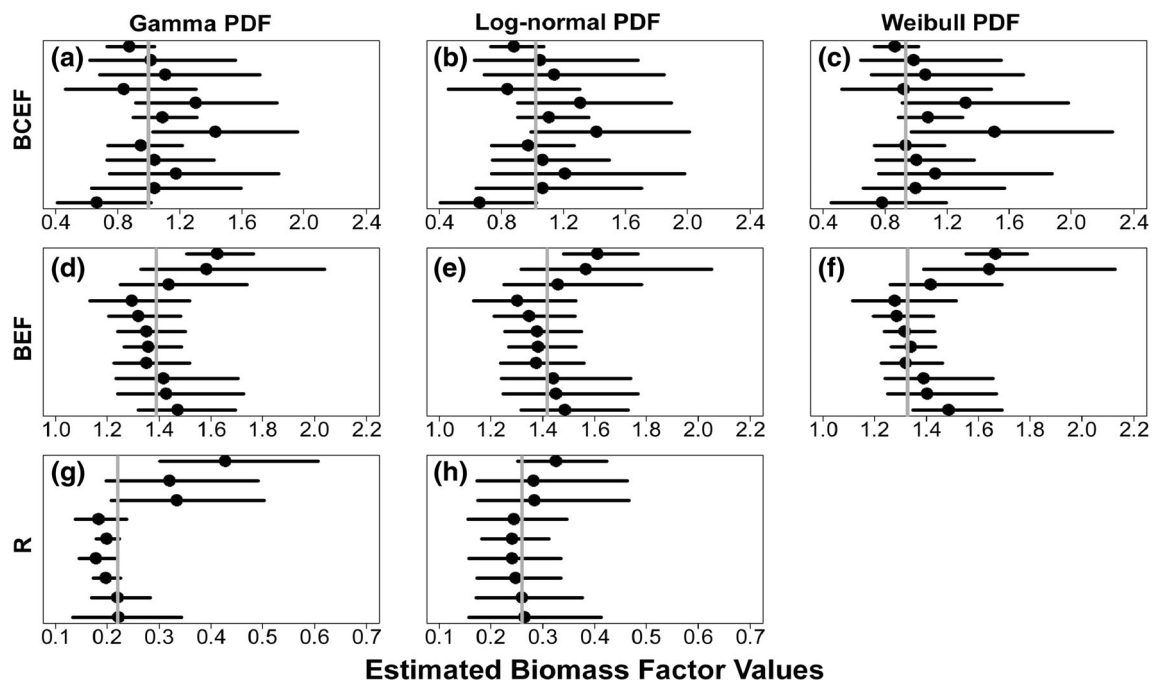


Fig. 4 Posterior estimates of the three biomass factors for each publication under the three PDFs. The biomass factors are BCEF (a, b, c), BEF (d, e, f) and RSR (g, h), and the three PDFs are Gamma (a, d, g), Log-normal (b, e, h) and Weibull (c, f). A black dot corresponds

to a publication depicting median posterior estimate of a biomass factor. The horizontal line depicting the 95 % credible interval (range in 2.5 and 97.5 % sample quantiles)

of *Q. mongolica* forests. Specification of a finite mixture model is considerably more complex, and to the best of our knowledge, it has not been attempted for modeling biomass factors.

For RSR, we recommend the Log-normal PDF because it performed better in comparison to the Gamma PDF despite the poor fit indicated by the goodness-of-fit statistics and diagnostic plots. The lack of fit was due to the heavy-tailed property of the observed RSR distribution, i.e., extreme positive skew with large RSR values. Uncertainty associated with estimating root biomass is high partly due to the difficulty of excavating and accurately measuring root systems. Undoubtedly, this uncertainty contributes to the large variation in the observed RSR distribution. RSR greater than 1.0 implied that root biomass is larger than above-ground biomass; thus, it could be that these values are outliers. However with little research on *Q. mongolica* root systems, this claim is inconclusive. Heavy-tailed PDF specifications such as the Pareto PDF are alternative models for the observed RSR distribution (Falk et al. 2010), but their application in modeling biomass factors is limited.

This study accounted for variation between and within publications by specifying publication as a nesting factor in a multilevel model. This strategy effectively accounted for causes of uncertainty specific to a publication such as conditions at study sites and methods of field measurement,

and its effect is similar to blocking in experimental design. As a result, the precision of estimates of a biomass factor by publications and across all publications would be higher. The posterior mean estimates of biomass factors across all publications from the meta-analysis have a wider scope of inference and could be applied for estimating carbon flux of *Q. mongolica* forests on a regional level. Furthermore, the recommended PDFs and associated posterior parameter estimates of the biomass factors could be used in the Monte Carlo simulation approach to uncertainty analysis as recommended by IPCC (2006a).

Priors are important components in Bayesian analysis. They are updated with newly collected information to derive posterior inferences that reflect both current and past knowledge. To take advantage of this feature, informative priors are often desirable. In this study, non-informative priors were applied as the analyses was unprecedented for *Q. mongolica* forests in South Korea. One could attempt to derive potential informative priors under certain assumptions. For example, informative priors can be derived from other *Quercus* species under the assumption that tree species with the same genus share similar life histories, tree form, allometry, and survival strategies. Furthermore, any future study attempts to refine estimation of regional biomass factors for *Q. mongolica* forests using the Bayesian approach could utilize the results from this study to construct informative prior distributions. In a similar manner,

Li et al. (2012) set the estimated posterior distributions as new priors to predict local parameters for a new dataset when modeling the tree height-age relationship for balsam fir in eastern Maine, USA.

Each PDF captures different aspects of the distribution of a biomass factor. For example, the model fit of the Gamma and Log-normal PDFs was comparable for BCEF, but the Log-normal PDF better captured the right tail of the distribution. Rather than having to choose the best fit model based on a set of criteria, one could use the Bayesian model average (BMA) method that offers a way to combine different distinct models into a single model (Zhang et al. 2014). Zhang et al. (2014) combined stand-level, tree-level and distribution models via BMA to predict stand basal area for even-aged Chinese fir plantations in southern China. Picard et al. (2012) combined multiple allometric equations into a single predictive model through BMA to predict tree above-ground biomass in tropical moist forests. The potential application of BMA to develop regional biomass factors is promising, and it requires further study.

Kwon and Lee (2006a) showed that biomass production of *Q. mongolica* forests depended on elevation but not on aspect. Therefore, biomass factors could be related to environmental factors and stand structures through regression models. This approach would further expand the scope of inference because biomass factors could then be predicted from any forest stands with measured predictors. A generalized linear mixed model (GLMM) would be suitable for this approach. GLMM requires specification of an appropriate PDF and random effects for a biomass factor. The recommended choice of PDF from this study and using publication as a nesting factor could be a starting point for building the GLMM. In short, this meta-analysis study has expanded the applicability of the findings reported in earlier studies, and the developed statistical model for the meta-analysis could be applied similarly to other tree species in the country.

Acknowledgements We would like to express our sincere thanks to Jeong-seon Kim for assistance in translating the Korean literature. We also thank Korea Forest Research Institute for the funding to conduct this research, and the many researchers who have conducted the studies used for the meta-analysis. Lastly, we would like to express sincere thanks to the anonymous reviewers for comments and suggestions that have greatly improve the publication.

References

- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) Introduction to meta-analysis. Wiley, Chichester
- Brooks S, Gelman A (1998) General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 7:434–456
- Brown S (2002) Measuring carbon in forests: current status and future challenges. *Environ Pollut* 116:363–372
- Brown S, Lugo AE (1982) The storage and production of organic matter in tropical forests and their role in global carbon cycle. *Biotropica* 14:161–187
- Cannell MGR (1982) World forest biomass and primary production data. Academic Press, London
- Congdon P (2006) Bayesian statistical modelling. Wiley, Chichester
- Falk M, Hüslér J, Reiss R-D (2010) Laws of small numbers: extremes and rare events. Springer, Basel
- Fang J-Y, Wang ZM (2001) Forest biomass estimation at regional and global levels, with special reference to China's forest biomass. *Ecol Res* 16:587–592
- Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge
- Hartung J, Knapp G, Sinha BK (2008) Statistical meta-analysis with applications. Wiley, Hoboken
- Hedges LV (1994) Fixed effects models. In: Cooper H, Hedges LV (eds) The handbook of research synthesis. Russell Sage Foundation, New York, pp 285–301
- IPCC (2006a) IPCC guidelines for national greenhouse gas inventories: volume 1 general guidance and reporting. In: Eggleston S, Buendi L, Miwa K, Ngara T, Tanab K (eds), Intergovernmental Panel on Climate Change (IPCC) IPCC/IGES Hayama
- IPCC (2006b) IPCC guidelines for national greenhouse gas inventories: volume 4 agriculture, forestry and other land use. In: Eggleston S, Buendia L, Miwa K, Ngara T, Tanabe K (eds), Intergovernmental Panel on Climate Change (IPCC) IPCC/IGES Hayama
- Jalkanen A, Mäkipää R, Ståhl G, Lehtonen A, Petersson H (2005) Estimation of the biomass stock of trees in Sweden: comparison of biomass equations and age-dependent biomass expansion factors. *Ann For Sci* 65:845–851
- Jenkins JC, Chojnacky DC, Heath LS, Birdsey RA (2003) National-scale biomass estimators for United States tree species. *For Sci* 49:12–35
- KFRI (2010) Study on the basis of forest carbon accounting in Korea. Korea Forest Service, Seoul
- Korea Forest Service (2009) Stem volume table. Korea Forest Service, Seoul
- Kwon K-C, Lee DK (2006a) Biomass and annual net production of *Quercus mongolica* stands in Mt. Joongwang with respect to altitude and aspect. *J Korean For Soc* 95:398–404
- Kwon K-C, Lee DK (2006b) Biomass and energy content of *Quercus mongolica* stands in Gwangyang and Jeju areas. *J Korean Wood Sci Soc* 34:54–65
- Kwon K-C, Lee DK (2006c) Above- and below-ground biomass and energy content of *Quercus mongolica*. *J Korean For Energy Soc* 25:31–38
- Lee DK, Kwon K-C (2006) Biomass and annual net production of *Quercus mongolica* stands in Pyungchang and Jecheon areas. *J Korean For Soc* 95:309–315
- Lee KJ, Park IH (1987) Primary production and nutrients distribution in 22-year-old *Pinus koraiensis* and *Quercus mongolica* stands in Kwangju district. *J Korean For Energy Soc* 7:11–21
- Levy PE, Hale SE, Nicoll BC (2004) Biomass expansion factors and root:shoot ratios for coniferous tree species in Great Britain. *Forestry* 77:421–430
- Li X, Yi MJ, Son Y, Jin G, Lee KH, Son YM, Kim RH (2010) Allometry, biomass and productivity of *Quercus* forests in Korea: a literature-based review. *J Korean For Soc* 99:726–735
- Li R, Stewart B, Weiskittel A (2012) A Bayesian approach for modelling non-linear longitudinal/hierarchical data with random effects in forestry. *Forestry* 85:17–25
- Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 10:325–337

- McLachlan G, Peel D (2000) Finite mixture models. Wiley-Interscience, New York
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21:1087–1092
- Návar J (2010) Measurement and assessment methods of forest aboveground biomass: a literature review and the challenges ahead. In: Momba M, Bux F (eds) biomass. Croatia, Sciyo, pp 27–64
- Ntzoufras I (2009) Bayesian modeling using WinBUGS. Wiley, Hoboken
- Park G-S (2003) Biomass and net primary production of *Quercus mongolica* stands in Kwangyang, Pyungchang, and Youngdong areas. *J Korean For Soc* 92:567–574
- Park IH, Moon GS (1994) Biomass, net production and biomass estimation equations in some natural *Quercus* forests. *J Korean For Soc* 83:246–253
- Park IH, Lee DK, Lee KJ, Moon GS (1996) Growth, biomass and net production of *Quercus* species (I). *J Korean For Soc* 85:76–83
- Park I-H, Kim D-Y, Son Y-H, Yi M-J, Jin H-O, Choi Y-H (2005a) Biomass and net production of a natural *Quercus mongolica* forest in Namsan, Seoul. *Korean J Environ Ecol* 19:299–304
- Park IH, Son Y, Kim DY, Jin HO, Yi MJ, Kim RH, Hwang JO (2005b) Biomass and production of a naturally regenerated oak forest in southern Korea. *Ecol Res* 20:227–231
- Picard N, Henry M, Mortier F, Trotta C, Saint-André L (2012) Using Bayesian model averaging to predict tree aboveground biomass in tropical moist forests. *For Sci* 58:15–23
- Schroeder P, Brown S, Mo J, Birdsey R, Cieszewski C (1997) Biomass estimation for temperate broadleaf forests of the United States using inventory data. *For Sci* 43:424–434
- Somogyi A, Cienciala E, Mäkipää R, Muukkonen P, Lehtonen A, Weiss P (2007) Indirect methods of large-scale forest biomass estimation. *Eur J For Res* 126:197–207
- Son Y, Kim DY, Park IH, Yi MJ, Jin HO (2007) Production and nutrient cycling of oak forests in Korea: a case study of *Quercus mongolica* and *Q. variabilis* stands. Kangwon National University Press, Chuncheon
- Song CY, Lee SW (1996) Biomass and net primary productivity in natural forests of *Quercus mongolica* and *Quercus variabilis*. *J Korean For Soc* 85:443–452
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit. *J R Stat Soc Ser B* 64:583–640
- Teobaldelli M, Somogyi Z, Migliavacca M, Usoltsev VA (2009) Generalized functions of biomass expansion factors for conifers and broadleaved by stand age, growing stock and site index. *For Ecol Manag* 257:1004–1013
- Usoltsev VA (2001) Forest biomass of northern Eurasia: database and geography. Russian Academy of Sciences, Ural Branch Yekarinenburg
- Van Camp N, Walle IV, Mertens J et al (2004) Inventory-based carbon stock of Flemish forests: a comparison of European biomass expansion factors. *Ann For Sci* 61:677–682
- Wirth C, Schumacher J, Schulze E-D (2004) Generic biomass functions for Norway spruce in Central Europe—a meta-analysis approach toward prediction and uncertainty estimation. *Tree Physiol* 24:121–139
- Zhang X, Duan A, Dong L, Cao QV, Zhang J (2014) The application of Bayesian model averaging in compatibility of stand basal area for even-aged plantations in southern china. *For Sci* 60:645–651