Bayesian inference for Johnson’s SB and Weibull distributions

Mahdi Teimouri a, and Rafal Podlaski b

aDepartment of Statistics, Faculty of Science and Engineering, Gonbad Kavous University, Gonbad Kavous, Iran; bDepartment of Environmental Biology, Institute of Biology, Jan Kochanowski University, Kielce, Poland

ABSTRACT
The four-parameter Johnson’s SB (JSB) and three-parameter Weibull distributions have received significant attention in the field of forestry for characterising diameters at breast height (DBH). This study suggests the Bayesian method for estimating parameters of the JSB distribution. The maximum likelihood approach uses iterative methods such as a Newton–Raphson (NR) algorithm for maximising the logarithm of the likelihood function. However, there is no guarantee that the NR method converges. Through simulation, this study verified that the NR method for estimating the parameters of the JSB distribution sometimes fails to converge. Further, the Bayesian estimators presented herein were shown to be robust with respect to the initial values and estimate the parameters of the JSB distribution efficiently. The performance of the JSB and three-parameter Weibull distributions was compared in a Bayesian paradigm when these models were fitted to DBH data of three plots randomly selected from a study established in 107 plots of mixed-age ponderosa pine (Pinus ponderosa Dougl. ex Laws.) with scattered western juniper (Juniperus occidentalis Hook.) at the Malheur National Forest on the south end of the Blue Mountains near Burns, Oregon, USA. The Bayesian paradigm demonstrated that JSB was superior to the three-parameter Weibull for characterising the DBH distribution when these models were fitted to the DBH data of the three plots. Moreover, the Bayesian approach outperformed the moment, conditional maximum likelihood, and two-percentile methods when the JSB distribution was fitted to DBH data of three plots and all 107 plots simultaneously.

Introduction
Statistical modelling for the distribution of the diameter at breast height (DBH) is becoming increasingly popular to characterise the forest height structure, forest dynamics, and comparing the forest stands (Gorgoso-Varela and Grandas-Arias 2007; Mateus and Tomé 2011a; Özçelik et al. 2016). The statistical characterisation or modelling of the DBH distribution has a long history in both managed and natural forest stands. Among all statistical models, the preferred model shows more flexibility, that is, capturing the DBH distribution well. This is because different types of the forest stands show different shapes for DBH distribution. For example, two main types of the forest stands include even-aged which are usually unimodal (one peak) and roughly symmetric, and uneven-aged, whose DBH distributions often have a reverse-J shape. Among all statistical distributions, the Johnson’s SB (JSB) and Weibull have received significant attention in the context of forest management. Numerous efforts have been made in the literature for modelling the trees’ DBH. Among them are Bailey and Dell (1973), Malamats et al. (1995), Malamats et al. (2000), Pretzsch (2009), Teimouri et al. (2020b) and Zhang et al. (2010) for two- or three-parameter Weibull distribution, Fonseca et al. (2009), Hafley and Buford (1985), Kiviste et al. (2003), Knoebel and Burkhart (1991), Kudus et al. (1999), Marto et al. (2009), Mateus and Tomé (2011a), Mateus and Tomé (2011b), Pogoda et al. (2020), Ogana et al. (2017), Ogana (2018), Özçelik et al. (2016), Parresol (2003), Rennolls and Wang (2005) and Zhou and McTague (1996) for JSB distribution, and Gorgoso-Varela et al. (2012), Gorgoso-Varela et al. (2014), Gorgoso-Varela et al. (2021), Hafley and Schreuder (1977), Palahí et al. (2007) and Zhang et al. (2003) for both.

As the most popular estimation method, the maximum likelihood (ML) approach is obtained with the aid of mathematical optimisation tools. These tools maximize the logarithm of the likelihood (log-likelihood) function using an iterative algorithm such as Newton–Raphson (NR) and therefore need the initial values. If the initial values are far away from the true parameter, which is where the log-likelihood function reaches its global maximum, or when the log-likelihood function at the initial values becomes large, then there is no guarantee that the NR method will converge (Teimouri and Cao 2020, pp. 101). This means that the ML approach is sensitive to the initial values and can be considered as a weakness of the approach. In addition, the ML approach may break down when the regularity conditions fail to exist. The above criticisms may occur when one is interested in estimating the parameters of JSB and three-parameter Weibull distributions. Other methods such as moment-based estimators are not as efficient as the ML approach. For example, moment-based estimators of the
three-parameter Weibull distribution overcome the weaknesses in the ML approach, but their existence, uniqueness, and consistency are still open questions (Nagatsu et al. 2013) or in the case of JBS distribution, the moment-based estimator is not as efficient as regression-type estimators (Scalfaro et al. 2003). This study derives the Bayesian estimators for the parameters of the JSB and three-parameter Weibull distributions. To the best of our knowledge, the Bayesian estimators of the parameters of JSB distribution have never been used in the forestry literature for modelling DBH distributions. In what follows, we give some preliminaries.

The pdf of the Johnson SB (JSB) is given by Johnson (1949) and Norman et al. (1994):

$$g(x|\theta) = \frac{\delta \lambda}{\sqrt{2\pi}(\lambda + \xi - x)} \exp\left\{ -\frac{1}{2} \left[ \gamma + \delta \log\left( \frac{x - \xi}{\lambda + \xi - x} \right) \right]^2 \right\},$$

(1)

where $x$ is DBH observation, $\Theta = (\delta, \gamma, \lambda, \xi)^T$, $\xi < x < \lambda + \xi$, $\delta > 0$, $\lambda > 0$, $-\infty < \gamma < \infty$, and $-\infty < \xi < \infty$. Here, $\xi$ and $\lambda$ are the location and scale parameters, respectively. Both of parameters $\gamma$ and $\delta$ affect the shape of distribution so that increasing magnitude of $\gamma$, the skewness increases and kurtosis of distribution increases when $\delta$ increases. Hereafter, functions $g(.)$ and $G(.)$ account for pdf and cumulative distribution function (cdf) of the given distribution, respectively. As can be seen, the cdf of the JBS distribution has no closed-form expression. The pdf of three-parameter Weibull distribution is given by

$$g_{w}(x|\Theta) = \frac{\alpha}{\beta} \left( \frac{x - \mu}{\beta} \right)^{\alpha - 1} \exp\left\{ -\left( \frac{x - \mu}{\beta} \right)^{\alpha} \right\},$$

(2)

and where $\Theta = (\alpha, \beta, \mu)^T$, $\mu < x$, $\alpha > 0$, and $\beta > 0$. Now, $\alpha$, $\beta$, and $\mu$ are the shape, scale, and location parameters, respectively.

In the Bayesian framework, we assume that the unknown parameter vector $\Theta$ follows a distribution with pdf $\pi(\Theta)$. Using information available in random observations $x = (x_1, \ldots, x_n)^T$, a revision will be made regarding known $\pi(\Theta)$ using the well-known Bayes’ theorem as $\pi(\Theta|x_i)$. We have

$$\pi(\Theta|x) = \frac{g(x|\Theta)\pi(\Theta)}{g(x)},$$

The expressions $\pi(\Theta)$ and $\pi(\Theta|x)$ are known as prior pdf and posterior pdf of $\Theta$, respectively. Here, $g(x)$ is normalising constant and the Bayes’ theorem therefore can be written as

$$\pi(\Theta|x) \propto g(x|\Theta)\pi(\Theta),$$

(3)

### Materials and methods

#### Materials

Since our research motivated by the widespread use and application of the statistical distributions in forest management, we conducted a study in the context of forestry. A study was established in 107 plots of mixed-age ponderosa pine ($\text{Pinus ponderosa}$ Doug.). with scattered western juniper ($\text{Juniperus occidentalis}$ Hook.) located in Malheur National Forest on the south end of the Blue Mountains near Burns, Oregon, USA (Kerns et al. 2017). The data include tree height, diameter, and growth for a prescribed burning study with unburned controls. Of these variables, we only used the DBH (measured at a height of 1.3 m) of all live trees in three randomly selected plots (plots 9, 44, and 73) each of size 0.08 ha for statistical validation of the Bayesian approach. The plots summary statistics are given in Table 1.

#### Methods

This section has two parts. In the first part, we carried out a simulation study to check the convergence rate of the NR algorithm. The second part has been devoted for describing and implementing the Bayesian paradigm for modelling DBH data of three plots.

#### The NR algorithm for JSB distribution

As previously mentioned, the NR algorithm may fail to converge. Unfortunately, this happens when finding the ML estimators of the JSB distribution is desired. A simulation study was performed to prove the claim. A number of 10,000 samples with different sizes including 20, 50, 100, 250, 500, 1000, and 5000 were simulated from the JSB distribution with pdf given in (1). The ML estimators of the parameters were obtained using the command optim(.) in R (R Core Team 2018) environment. In each of 10,000 runs, the parameters $\delta, \gamma, \lambda, \text{and} \xi$ were generated from uniform distribution $(0.05,10), (20,20), (1,100), \text{and} (50,50), \text{respectively}$. While $x_{100}^{(i)}$ and $x_{50}^{(i)}$ denote, respectively, the smallest and largest values of the $i$th generated sample, for $i = 1, \ldots, 10,000$, the initial values of $\delta, \gamma, \lambda, \text{and} \xi$ were generated from uniform distribution $(0.05,10), (-20,20), (x_{100}^{(i)} - x_{50}^{(i)}, 100)$, and $(50, x_{100}^{(i)})$, respectively. The results of simulation are given in Table 2. As can be seen, the percentage of failed attempts to reach convergence through the NR algorithm is considerable (e.g. on the average 32%).

#### Bayesian approach

Here, some preliminaries about the three Markov chain Monte Carlo (MCMC) techniques are given in brief. Then,
the Bayesian paradigm for the JSB and Weibull distributions are described, respectively.

Due to the complicate nature of the posterior pdf $\pi(\theta|x)$, we have to sample from the posterior pdf and then the Bayesian estimators are obtained as the average of the last motions (samples) of the sampler when sampler is run for a long term. In practice, the exploitation of $\pi(\theta|x)$ needs the use of MCMC techniques. The Gibbs sampler is one of the MCMC techniques that enables us to sample from full conditional pdf, that is, the pdf of each element of the parameter vector given the other elements and observed data $x = (x_1, \ldots, x_n)^T$. For a comprehensive account on Gibbs sampler (sampling), we refer reader to Zhang et al. (2005), Braswell et al. (2005), Link and Eaton (2012), Robert and Casella (2010) and Dorazio and Rodriguez (2012). The Metropolis–Hastings (MH) algorithm is an efficient tool for drawing samples from a given posterior distribution. Assume that we want to simulate sample from pdf given by (3). First, we need to choose a proposal distribution $q(\cdot|\cdot)$ that changes the location of the chain at each iteration of the algorithm. The proposal distribution is arbitrary and can be chosen so that is easy to simulate from. Since realisation generated at each iteration is used to generate the sample at the next step, the chain constitutes a correlated stochastic process, but after $N$ numbers of generations, we hope that the chain produces uncorrelated samples and converges to the target distribution $\pi(\theta|x)$ as desired. For a given sample $x = (x_1, \ldots, x_n)^T$, suppose we are interested in sampling from posterior pdf $\pi(\theta|x)$. The well-known accept–reject sampling needs a suitable upper bound, known as $M$ that satisfies the following inequality:

$$
\sup_{\theta} \frac{\pi(\theta|x)}{g(\theta)} \leq M,
$$

where $g(\theta)$ is an arbitrary pdf that is easy to sample from and its support includes the support of $\pi(\theta|x)$. If a suitable choice for $M$ exists, then the accept–reject sampling is efficient otherwise we refer to another Monte Carlo simulation technique known as the adaptive rejection sampling (ARS) algorithm. The ARS algorithm is used to simulate realisation when posterior pdf is log-concave (i.e. the second derivative of $\pi(\theta|x)$ with respect to $\theta$ is negative). In such a case, the ARS algorithm developed by Gilks and Wild (1992) and Gilks et al. (1995) is highly efficient.

Details of Bayesian paradigm for the JSB distribution are given in Appendix 1. The Bayesian inference for the three-parameter Weibull distribution originally was developed by Smith and Naylor (1987) and Green et al. (1994). Here, we give a slightly different version of the Bayesian paradigm developed by Green et al. (1994). We mention that the only difference occurs in updating the location parameter at each iteration of the chain. Details for Bayesian inference of the three-parameter Weibull distribution are given in Appendix 2.

Based on Bayesian paradigm described in Appendices 1 and 2, we compared the performance of the JBS and three-parameter Weibull distributions for modelling DBH data when the parameters of both models were estimated using the Bayesian approach. Figures 1 and 2 display histograms of the samples drawn from the full conditionals and the pairwise scatterplots of the sampler output for the JSB and three-parameter Weibull distributions, respectively, when these distributions are fitted to DBH data of the plot 9.

For Bayesian approach, we assumed that the sampler’s convergence has been attained before 5000 iterations in all three plots for both distributions. Therefore, we removed the first 5000 samples from the sampler output when it was repeated for 10,000 times (Figure 5). Based on the average

![Figure 1](image-url)

**Figure 1.** Left-hand side: Pairwise scatterplots of trimmed output of the Gibbs sampler for estimation parameters of the JSB distribution fitted to the DBH observations in plot 9. These outputs suggest that there is little dependence between $\delta$ and $\lambda$. Right-hand side: Histograms of the full conditionals (a) $\pi(\delta|\gamma, \lambda, \xi, x)$, (b) $\pi(\gamma|\delta, \lambda, \xi, x)$, (c) $\pi(\lambda|\delta, \gamma, \xi, x)$, and (d) $\pi(\xi|\delta, \gamma, \lambda, x)$ produced by the Gibbs sampler for 10,000 runs.

| Sample size | 20 | 50 | 100 | 250 | 500 | 1000 | 5000 |
|-------------|----|----|-----|-----|-----|------|------|
| Percentage  | 68.3% | 68.7% | 68.3% | 68.3% | 68.1% | 68.4% | 67.9% |

![Table 2](image-url)

**Table 2.** Percentage of runs that NR method truly converged for the JSB distribution.
The computed goodness-of-fit statistics are given in Table 4. The estimated parameters of the JSB distribution through fitting the JSB distribution to DBH of three plots and simulation. The results of comparison were given in Tables 5 and 6. Moreover, we compared these methods through a simulation study. For this purpose, the Bayesian paradigm was adapted for the case that location and scale parameters are known since the MM, CML, and KB methods require the location and scale parameters to be known. For settings $\delta = 0.9$, $\gamma = 0.2$, $\lambda = 64$, and $\xi = 3.5$ of the parameters, the bias and root of mean squared errors (RMSE) of estimators, given in Table 7, have been computed based on $N=500$ runs. Assuming that estimator $\hat{\theta}$ is computed based on $N$ samples for estimating unknown parameter $\theta$, the bias and RMSE are defined as

$$
\text{bias} = \frac{1}{N} \sum_{i=1}^{N} (\theta - \hat{\theta}), \quad \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\theta - \hat{\theta})^2}.
$$

The package ForestFit (Teimouri et al. 2020a) developed for R (R Core Team 2018) environment that uploaded.

**Table 3.** Bayesian estimators for the parameters of the JSB and three-parameter Weibull distributions.

| Plot | $\delta$ | $\gamma$ | $\lambda$ | $\xi$ | $\alpha$ | $\beta$ | $\mu$ |
|------|----------|----------|-----------|-------|----------|--------|-------|
| 9    | 0.772    | 0.545    | 52.311    | 8.719 | 1.682    | 23.120 | 7.436 |
| 44   | 0.875    | 0.203    | 64.162    | 3.642 | 2.145    | 34.496 | 2.171 |
| 73   | 0.641    | 0.978    | 88.592    | 8.182 | 1.005    | 22.746 | 8.278 |

**Table 4.** Computed goodness-of-fit statistics for modelling DBH data using JBS and three-parameter Weibull distributions.

| Plot | AD    | CM    | KS    | LL    | AD    | CM    | KS    | LL    |
|------|-------|-------|-------|-------|-------|-------|-------|-------|
| 9    | 0.167 | 0.023 | 0.059 | -196.808 | 0.210 | 0.048 | 0.082 | -199.579 |
| 44   | 0.216 | 0.028 | 0.100 | -120.072 | 0.302 | 0.040 | 0.115 | -121.846 |
| 73   | 0.105 | 0.031 | 0.075 | -141.751 | 0.257 | 0.045 | 0.084 | -143.299 |

**Comparison study**

Here, we compared the performance of the Bayesian and other non-Bayesian (or frequentist) methods in estimating parameters of the JSB distribution through fitting the JSB distribution to DBH data of three plots and simulation. The non-Bayesian competitors are method of moment (MM) (Fonseca et al. 2009), method of conditional maximum likelihood (CLM) (Schreuder and Haefly 1977), and two-percentile method proposed by Knoebel and Burkhart (1991) (called here KB). For this purpose, we compared the performance of the above four methods when these approaches were used for fitting JSB distribution to DBH of three plots given in Table 1 and also all 107 plots simultaneously. The results of comparison were given in Tables 5 and 6. Moreover, we compared these methods through a simulation study. For this purpose, the Bayesian paradigm was adapted for the case that location and scale parameters are known since the MM, CML, and KB methods require the location and scale parameters to be known. For settings $\delta = 0.9$, $\gamma = 0.2$, $\lambda = 64$, and $\xi = 3.5$ of the parameters, the bias and root of mean squared errors (RMSE) of estimators, given in Table 7, have been computed based on $N=500$ runs. Assuming that estimator $\hat{\theta}$ is computed based on $N$ samples for estimating unknown parameter $\theta$, the bias and RMSE are defined as

$$
\text{bias} = \frac{1}{N} \sum_{i=1}^{N} (\theta - \hat{\theta}), \quad \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\theta - \hat{\theta})^2}.
$$

The package ForestFit (Teimouri et al. 2020a) developed for R (R Core Team 2018) environment that uploaded.
to CRAN (Comprehensive R Archive Network) at https://cran.r-project.org/web/packages/ForestFit/index.html was used to obtain the results given in Tables 3 and 4.

**Results**

Plot 9 was characterised by the DBH distribution with trees ranging in size between 10.4 and 55.9 cm in DBH (Table 2). The DBH distribution indicates a forest patch composed of trees of different height and age. The youngest trees were the most common; with the increase in age the share of trees in individual bins decreased. The outline of empirical values suggests a concave polygonal chain (Figure 3). There are two peaks around 15 and 55 cm in DBH (Figure 3). The JSB distribution captured the DBH distribution better than the Weibull model (Table 4). The shapes of the fitted distributions to the DBH data of plot 9 were different in the right tail (Figure 3). In fact, the JSB characterised the tail of the DBH distribution better than the Weibull distribution. When the JSB distribution was fitted to DBH data of three sample plots of Table 1, the Bayesian approach outperformed the MM, CML, and KB methods in terms of AD, CM, and KS goodness-of-fit statistics (Table 5). The fitted JSB distribution through Bayesian, MM, CML, and KB methods is shown in Figure 4.

Plot 44 was characterised by the DBH distribution with trees ranging in size between 9.4 and 59.4 cm in DBH (Table 1). This plot represented a one-storied forest that was characterised by a DBH distribution with three modes around 15, 35, and 55 cm in DBH. Ignoring these modes, the DBH empirical distribution was characterised by a broadly uniform distribution between 9 and 60 cm in DBH (Figure 3). Overall, the JSB model was the superior model for DBH distribution in the sense of AD, CM, KS, and LL measures (Table 4). The shape of the fitted densities corresponding to the JSB and Weibull distributions has differences in the middle and right tails (Figure 3). This is largely because the JSB has two threshold parameters that makes its pdf bounded from the left and right. The MM outperforms other methods in sense of the KS statistic. This is while, the CML and KB approaches show the best performance in terms of AD and CM statistics. In this case, the Bayesian approach shows the weakest performance (Table 5).

Plot 73 was characterised by the DBH distribution with trees ranging in size between 9.1 and 88.6 cm in DBH (Table 1). The DBH distribution indicates a forest patch composed of trees of different height and age. The youngest trees were the most common; with the increase in age the share of trees in individual bins decreased. The outline of empirical values suggests a convex polygonal chain (Figure 3). There is one peak around 15 cm in DBH (Figure 3). In contrast to

| Plot | Method | AD   | CM   | KS   | LL     |
|------|--------|------|------|------|--------|
| 9    | Bayesian | 0.167 | 0.023 | 0.059 | −196.808 |
|      | MM     | 0.237 | 0.043 | 0.076 | −196.974 |
|      | CML    | 0.198 | 0.032 | 0.065 | −196.963 |
|      | KB     | 0.474 | 0.073 | 0.098 | −196.793 |
| 44   | Bayesian | 0.215 | 0.028 | 0.100 | −120.072 |
|      | MM     | 0.180 | 0.023 | 0.071 | −119.410 |
|      | CML    | 0.164 | 0.021 | 0.080 | −119.319 |
|      | KB     | 0.163 | 0.022 | 0.082 | −118.689 |
| 73   | Bayesian | 0.105 | 0.031 | 0.075 | −141.751 |
|      | MM     | 0.407 | 0.053 | 0.107 | −142.681 |
|      | CML    | 0.253 | 0.039 | 0.085 | −142.280 |
|      | KB     | 0.720 | 0.100 | 0.128 | −143.495 |

| Plot | Method | AD   | CM   | KS   | LL     |
|------|--------|------|------|------|--------|
| 9    | Bayesian | 5.201 | 0.831 | 0.031 | −19455.81 |
|      | MM     | 6.453 | 1.280 | 0.042 | −19458.15 |
|      | CML    | 5.218 | 0.867 | 0.032 | −19450.08 |
|      | KB     | 27.784 | 3.896 | 0.057 | −19497.83 |

| Plot | Method | AD   | CM   | KS   | LL     |
|------|--------|------|------|------|--------|
| 9    | Bayesian | 0.044 | 0.148 | 0.014 | 0.104 | −0.001 | 0.081 |
|      | MM     | 0.995 | 5.541 | 0.930 | 4.680 | 0.896 | 4.201 |
|      | CML    | 0.531 | 4.193 | 0.492 | 3.553 | 0.478 | 3.214 |
|      | KB     | 0.242 | 0.177 | 0.138 | 0.127 | 0.081 | 0.095 |
| 44   | Bayesian | 0.099 | 0.239 | 0.076 | 0.178 | 0.060 | 0.135 |
|      | MM     | 1.026 | 5.755 | 0.967 | 4.948 | 0.927 | 4.423 |
|      | CML    | 0.556 | 4.343 | 0.520 | 3.741 | 0.502 | 3.364 |
|      | KB     | 0.354 | 0.336 | 0.227 | 0.234 | 0.150 | 0.166 |

**Figure 3.** Histograms of DBH data in plots 9, 44, and 73. Superimposed in each subfigure are estimated pdf of the JSB (blue solid line) and Weibull (red dashed line) distributions.
the plots 9 and 44, the DBH distribution of the plot 73 was much wider with a long tail skewed to the right. The JSB model characterised the DBH distribution better than the Weibull model (Table 4). Evidently, the JSB was the superior model than the Weibull model for DBH observations at the first peak and right tail (Figure 3). In all three plots, the JSB outperformed the Weibull model in terms of all goodness-of-fit statistics (Table 4).

When the JSB model was fitted to DBH data of three plots in Table 1, the Bayesian paradigm and KB presented the best and weakest performances, respectively (Table 5). Also, when the JSB distribution was fitted to DBH data of all 107 plots, the Bayesian approach presented the best performance (Table 6). Simulation study revealed that the Bayesian approach outperformed MM and KB methods in terms of all four goodness-of-fit statistics for estimating the parameters of the JSB distribution (Table 7). Moreover, the Bayesian method outperformed the CML method for estimating the parameters of the JSB distribution in terms of all AD, CM, and KS goodness-of-fit statistics (Table 7).

Discussion

The DBH distributions in all analysed plots proved to be multimodal. Existing peaks are most likely associated with partially overlapping tree cohorts. Both species forming the studied stands, P. ponderosa and J. occidentalis are shade intolerant, and younger trees of various sizes grew in the gaps, and not under the canopy of older trees. In two plots, there were clearly large numbers in the lowest DBH bins, characteristic for stands in which there is a continuous process of regeneration. In this type of stands, over time, individual cohorts of trees overlap and the multimodality of the DBH distributions gradually disappears. If the individual cohorts are still clearly visible, it is necessary to use the mixture distributions (Teimouri and Podlaski 2020). Then, as multimodality disappears, single distributions, such as the JSB model should be used.

In the case of the JSB distribution, it is known that the first-order statistic, that is, $x(1)$ is a sufficient statistic for $\xi$. Therefore, choosing $\xi^{(0)} = x(1) - 1/n$ in which $n$ is the sample size would be quite reasonable as the initial value for $\xi$. Likewise, since $\xi < x < \xi + \lambda$, a good initial value for $\lambda$ is given by $\lambda^{(0)} = x(n) - x(1) + 2/n$ in which $x(n)$ is the maximum value of DBH in the sample, i.e. $x_{(n)}$. Here, the constants $1/n$ and $2/n$ for $\xi^{(0)}$ and $\lambda^{(0)}$ have been used to avoid the possible singularity problems. A suitable initial value for $\gamma$ is obtained by considering the relation $\gamma = \log(1/y_{0.5} - 1)$ where $y_{0.5}$ is median of the transformation $y = (x - \xi)/\lambda$ with $x = (x_1, \ldots, x_n)^T$ (Özçelik et al. 2016). Therefore, $\gamma^{(0)} = \log(1/y_{0.5} - 1)$ where $y_{0.5}$ is median of the transformation $y = (x - \xi^{(0)})/\lambda^{(0)}$. It should be noted that we took $\delta^{(0)} = 1$ as the initial value for $\delta$. Additionally, our study revealed that the Bayesian paradigm presented herein is robust with respect to $\delta^{(0)}$, so that it can be started well away from the true value of $\delta$. In the case of the three-parameter Weibull distribution, similar to the JSB distribution, we used $\xi^{(0)} = x(1) - 1/n$ as the initial value for $\mu$. The initial values of the shape and scale parameters were obtained using the method of moments (Norman et al. 1994). We also performed a simulation study to check the robustness of the Bayesian paradigm with respect to the initial values for estimating the parameters of the JSB distribution. For this purpose, we confine ourselves to the case in which we have simulated 300 samples each of size 100 from JSB distribution with parameter vector $\Theta = (2, 2, 20, 0)^T$, i.e. $\delta = 2$, $\gamma = 2$, $\lambda = 20$, and $\xi = 0$. The results of simulation are given in Table 3. We note that in each of 300 runs, the initial values were not chosen by the method suggested above. Instead, the initial values were generated randomly from uniform distribution. We used this scenario to check the robustness of the Gibbs sampler. The initial values for $\delta$, $\gamma$, $\lambda$, and $\xi$ were generated from uniform distribution $(0.1, 15)$, $(-15, 15)$, $(20, 60)$, and $(-10, 10)$, respectively. For example, the general motion of the Gibbs sampler has been shown in Figure 5, when the initial values were chosen as $\delta^{(0)} = 15$, $\gamma^{(0)} = -15$, $\lambda^{(0)} = 60$, and $\xi^{(0)} = -10$ to show the robustness of the Bayesian paradigm.

For applying the MM, CML, and KB methods, in simulation study whose results are given in Table 5, the location and scale parameters have been determined using the method suggested by Ogana (2018). We set $\xi$ and $\lambda$ to be minimum of diameter minus 1.34 and maximum of diameter, respectively. In the case of method KB, we set $\xi = \min(DBH) - 1.3$ and $\lambda = \max(DBH) - \xi + 3.8$. The setting of parameters for simulation study was $\delta = 0.9$, $\gamma = 0.2$, $\lambda = 64$, and $\xi = 3.5$ that are enough close to the Bayesian estimations of plot 44 (Table 3). These settings were chosen deliberately to
check the performance of the Bayesian method if DBH observations really follow the JSB distribution. The Bayesian paradigm is the most efficient in this case (Table 6) and weakness of this method in the presence of the MM, CML, and KB methods for modelling plot 44 could be assigned to this fact that the DBH observations in plot 44 do not follow the JSB distribution.

Conclusion

We derived the Bayesian estimators for the four-parameter Johnson’s SB (JSB) distribution. The maximum likelihood (ML) approach is the most commonly used method for estimating the model parameters, but simulation studies have shown that on average, 32% of attempts to reach the convergence through this method fail. We therefore suggest using the Bayesian paradigm to estimate the parameters of the JSB distribution. We demonstrated that the proposed Bayesian approach works efficiently and is robust with respect to the initial values. Furthermore, we considered the Bayesian estimators for parameters of the three-parameter Weibull distribution suggested by Green et al. (1994). The proposed algorithm in this work for sampling from full conditional of the location parameter is faster than that of Green et al. (1994). We fitted both the JSB and three-parameter Weibull distributions to the diameters at breast height (DBH) obtained from 3 out of 107 plots of size 0.08 ha established in mixed-age ponderosa pine (Pinus ponderosa Dougl. ex Laws.) forests with scattered western juniper (Juniperus occidentalis Hook.) located in the Malheur National Forest on the south end of the Blue Mountains near Burns, Oregon, USA. The estimation results indicated that the JSB model outperformed the three-parameter Weibull distribution and therefore characterised the DBH distribution more accurately. Simulation study revealed that the Bayesian paradigm is more efficient than other methods for estimating parameters of the JSB distribution including method of moment, method of conditional maximum likelihood, and two-percentile method. Furthermore, fitting these four methods to the DBH data of three plots, the Bayesian paradigm showed almost the best performance in fitting the JSB distribution to DBH data of two plots. As a possible future work, we are interested in estimating the parameters of the bivariate JSB distribution using the Bayesian method.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Polish National Science Centre [UMO-2016/21/B/NZ9/02749].

References

Bailey RL, Dell T. 1973. Quantifying diameter distributions with the Weibull function. Forest Sci. 19:97–104.

Braswell BH, Sacks WJ, Linder E, Schimel DS. 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. Glob Chang Biol. 11:335–355.

Dorazio RM, Rodriguez DT. 2012. A Gibbs sampler for Bayesian analysis of site-occupancy data. Methods Ecol Evol. 3(6):1093–1098.

Fonseca TF, Marques CP, Parresol BR. 2009. Describing maritime pine diameter distributions with Johnson’s SB distribution using a new all-parameter recovery approach. Forest Sci. 55:367–373.

Gils JR, Best NG, Tan KKC. 1995. Adaptive rejection metropolis sampling within Gibbs sampling. J R Stat Soc Ser C (Appl Stat). 44:455–472.

Gils WR, Wild P. 1992. Adaptive rejection sampling for Gibbs sampling. J R Stat Soc Ser C (Appl Stat). 41:337–348.

Gorgoso-Varela JJ, Grandas-Arias J. 2007. Modelling diameter distributions of Betula alba L. stands in northwest Spain with the two-parameter Weibull function. For Syst. 16:113–123.

Gorgoso-Varela JJ, Ponce RA, Cámaras-Obregón A, Rodríguez-Puerta F. 2021. Modeling diameter distributions with six probability density functions in Pinus halepensis Mill. plantations using low-Density airborne laser scanning data in aragon (Northeast Spain). Remote Sens. 13(12):2307.

Gorgoso-Varela JJ, Rojo A, Cámaras-Obregón A, Diéguez-Aranda U. 2012. A comparison of estimation methods for fitting Weibull, Johnson’s SB and beta functions to Pinus pinaster, Pinus radiate and Pinus sylvestris stands in northwest Spain. For Syst. 21:446–459.

Gorgoso-Varela JJ, Rojo A, Cámaras-Obregón A, Diéguez-Aranda U. 2014. A comparison of estimation methods for fitting Weibull and Johnson’s SB functions to pedunculate oak (Quercus robur) and birch (Betula pubescens) stands in northwest Spain. For Syst. 23 (3):500–505.

Green EJ, Roesch Jr FA, Smith AF, Strawderman WE. 1994. Bayesian estimation for the three-parameter Weibull distribution with tree diameter data. Biometrics. 50 (1), 254–269.

Hafley W, Buford M. 1985. A bivariate model for growth and yield prediction. For Sci. 31:237–247.

Hafley W, Schreuder H. 1977. Statistical distributions for fitting diameter and height data in even-aged stands. Can J For Res. 7(3):481–487.

Jeffreys H. 1961. Theory of probability. Oxford: Oxford University Press.

Johnson NL. 1949. Systems of frequency curves generated by methods of translation. Biometrika. 36(1–2):149–176.

Figure 5. Output of (a) δ, (b) γ, (c) λ, and (d) ξ from Gibbs sampler for the robust analysis. Each subfigure shows the motion of the Gibbs sampler across 10,000 iterations.
Kerns BK, Westlind DJ, Day MA. 2017. Season and interval of burning and cattle exclusion in the southern Blue Mountains, Oregon: Overstory tree height, diameter and growth. USDA Forest Service Research Data Archive, https://doi.org/10.2737/RDS-2017-0041.

Kiviste A, Nilson A, Hordo M, Merenik K. 2003. Diameter distribution models and height-diameter equations for Estonian forests. In: Amaro A, Reed D, Soares P, editors. Modelling Forest Systems. CAB International; p. 169–179.

Knoebel BR, Burkhart HE. 1991. A bivariate distribution approach to modeling forest diameter distributions at two points in time. Biometrics. 3(241–253.

Kudus KA, Ahmad M, Lapongan J. 1999. Nonlinear regression approach to estimation Johnson SB parameters for diameter data. Can J For Res. 29(3):310–314.

Link WA, Eaton MJ. 2012. On thinning of chains in MCMC. Methods Ecol Evol. 3(1):112–115.

Maltamo M, Kangas A, Uuttera J, Torniainen T, Saramäki J. 2000. Functions for modelling basal area diameter distribution in stands of Pinus sylvestris and Picea abies. Scand J For Res. 10(1–4):284–295.

Marto M, Palma J, Mateus A, Tomé M. 2009. Computer program for estimation of Johnson’s SB parameters using a parameter recovery method. Publicações Científicas Forchage PC-X/2009. Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Lisboa.

Mateus A, Tomé M. 2011a. Modelling the diameter distribution of eucalyptus plantations with Johnson’s SB probability density function: parameters recovery from a compatible system of equations to predict stand variables. Ann For Sci. 68(2):325–335.

Mateus A, Tomé M. 2011b. Estimating the parameters of the Johnson SB distribution using an approach of method of moments. AIP Conf Proc. 1389:1483–1485.

Nagatsuka H, Kamakura T, Balakrishnan N. 2013. A consistent method of estimation of the three-parameter Weibull distribution using an approach of method of moments. AIP Conf Proc. 1534:1329–1332.

Ogana FN, Itam ES, Osho JSA. 2017. Modeling diameter distributions of Gmelina arborea plantation in Omo forest reserve, Nigeria with Johnson’s SB. J Sustain For. 36(2):121–133.

Özçelik R, Fidalgo Fonseca TJ, Parresol BR, Eler U. 2016. Modeling the diameter distributions of Brutian Pine stands using Johnson’s SB distribution. Forest Science. 62(6):587–593.

Palihi M, Pukkala T, Blasco E, Trasobares A. 2007. Comparison of beta, Johnson’s SB, Weibull and truncated Weibull functions for modeling the diameter distribution of forest stands in Catalonia (north-east of Spain). Eur J For Res. 126(4):563–571.

Parresol BR. 2003. Recovering parameters of Johnson’s SB distribution. Res. Pap. SRS-31. Asheville, NC: US Department of Agriculture, Forest Service, Southern Research Station. 9 p. 31.

Pogoda P, Ochal W, Orzel S. 2020. Performance of kernel estimator and Johnson SB function for modeling diameter distribution of black alder (Alnus glutinosa (L.) Gaertn.) stands. Forests. 11(6):634.

Presztch H. 2009. Forest dynamics, growth, and yield. In: Forest dynamics, growth and yield. Springer, Berlin, Heidelberg.

R Core Team. 2018. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Remmols IL, Wang M. 2005. A new parameterization of Johnson’s SB distribution with application to fitting forest tree diameter data. Can J For Res. 35(3):575–579.

Robert CP, Casella G. 2010. Introducing Monte Carlo methods with R. New York: Springer.

Schoenfeld JRS, Tabai FCV, de Macedo RLG, Acerbi Jr FW, de Assis AL. 2003. SB distribution’s accuracy to represent the diameter distribution of Pinus taeda, through five fitting methods. For Ecol Manage. 175(1–3):489–496.

Smith RL, Naylor J. 1987. A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. J Royal Statist Soc Ser C Appl Statist. 36:358–369.

Teimouri M, Abdolalinezhad K, Ghalandarayeshi Sh. 2020a. Evaluation of estimation methods for parameters of the probability functions in tree diameter distribution modeling. Environ Resour Res. 8:25–40.

Teimouri M, Cao VC. 2020. Statistical inference for Bimbbaum–Saunders and Weibull distributions fitted to grouped and ungrouped data. Environ Resour Res. 8:97–108.

Zhou B, McTague JP. 1996. Comparison and evaluation of estimation methods for parameters of the probability functions in tree diameter distributions at two points in time. Biometrics. 3:241–253.

Appendices

Appendix 1

For Bayesian inference of the JSB distribution parameters, we assume that all four priors are statistically independent and so the full Bayesian model (joint posterior pdf) up to proportionality becomes

\[ p(\theta | x) \propto \frac{\delta^3 \lambda^3}{(2\pi)^2} \prod_{i=1}^n \left( \phi_i - \delta \right)^{\gamma \lambda} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left( \gamma + \delta \log \left( \frac{x_i - \delta}{\lambda + \delta - x_i} \right) \right)^2 \right\} \cdot \pi(\delta | y) \pi(\lambda | \delta) \]  

We note that we produce the Gibbs sampler directly from the full conditionals by choosing improper priors, that is, bypassing the propriety of the posterior. This is because the full conditionals are well-defined and the posterior model under study is sufficiently complex (Robert and Casella 2010). It is easy to see that

\[ \pi(\delta | y, \lambda, \xi, x) \propto C \delta^3 \exp \left\{ \frac{k_2}{2} \delta + \frac{k_1}{k_2} \right\}, \]

where

\[ k_1 = \sum_{i=1}^n \log \left( \frac{x_i - \delta}{\lambda + \delta - x_i} \right) \]  

and

\[ k_2 = \sum_{i=1}^n \log \left( \frac{x_i - \delta}{\lambda + \delta - x_i} \right)^2. \]

and \( C \) is a normalizing constant independent of \( \delta \). And the full conditional pdf of \( \delta \) is log-concave since the first and second derivatives of \( \log p(\delta | y, \lambda, \xi, x) \) with respect to \( \delta \) are always negative. Assume that we are currently at \( t \)th iteration of the sampler, sampling from \( \pi(\delta | y^{(t)}, \lambda^{(t)}, \xi^{(t)}, x) \) is carried out through the ARS algorithm. Here, \( y^{(t)} \), \( \lambda^{(t)} \), and \( \xi^{(t)} \) denote the generated values, respectively, from \( y, \lambda \) and \( \xi \).
at $t$th iteration. For full conditional pdf of $y$, we have

\[
\pi(y|\delta, \lambda, \xi, x) \propto \exp \left\{ -\frac{1}{2} \sum_{j=1}^{n} \gamma + \delta \log \left( \frac{x_j - \xi}{\lambda + \xi - x_j} \right)^2 \right\}
\]

\[
\propto \exp \left( -\frac{n}{2} \left[ \gamma + \frac{\delta k_1^2}{n} \right] \right).
\]

where $k_1$ is defined in (A2). Assume that we are currently at $t$th iteration of the sampler, for sampling from full conditional pdf of $y$, it is enough to sample from Gaussian distribution. The full conditional of $\lambda$ (up to proportionality) is

\[
\pi(\lambda|\delta, y, \xi, x) = \lambda^a \Pi_{i=1}^{n} (\lambda + \xi - x_i)^{y_i-1} \exp \left\{ -\frac{\delta ^2}{2} \sum_{i=1}^{n} \left[ \log \left( \frac{x_i - \xi}{\lambda + \xi - x_i} \right) \right]^2 \right\},
\]

(A4)

The structure of (A4) is complicated. Additionally, $\pi(\lambda|\delta, y, \xi, x)$ is not always log-concave. So, we use the MH algorithm for sampling from $\pi(\lambda|\delta, y, \xi, x)$ by choosing $\exp \left( -\left( \lambda - x_{\text{m}} + \xi \right) \right)$ as the proposal pdf for $\lambda$. Also, $x_{\text{m}} = \max \{x_1, \ldots, x_n\}$. Similar to the full conditional pdf of $\lambda$ in (A4), the full conditional pdf of $\xi$ has complicated structure. We have

\[
\pi(\xi|\delta, \gamma, \lambda, x) \propto \Pi_{i=1}^{n} \left( \lambda + \xi - x_i \right)^{y_i-1} \exp \left\{ -\frac{\delta ^2}{2} \sum_{i=1}^{n} \log \left( \frac{x_i - \xi}{\lambda + \xi - x_i} \right)^2 \right\},
\]

where $x_{\text{m}} - \lambda < \xi < x_{(1)}$ in which $x_{(1)} = \min \{x_1, \ldots, x_n\}$. Since the full conditional pdf of $\xi$ is not always log-concave, we use the MH algorithm to sample from it. For this aim, we choose the uniform distribution on $[x_{\text{m}} - \lambda, x_{(1)}]$ as the proposal.

### Appendix 2

Assume that $x = (x_1, \ldots, x_n)^T$ denotes the vector of $n$ independent observations and each follows distribution with pdf given in (2). We consider the Jeffreys’ prior (Jeffreys 1961) for $a$ and $\beta$, i.e. $\pi(a) \propto 1/a$ and $\pi(\beta) \propto 1/\beta$. Also, we allow the prior for $\mu$ to be uniformly over $\mathbb{R}$. The full conditionals of $a$, $\beta$, and $\mu$ are obtained easily. As pointed out by Green et al. (1994), the full conditional of $\pi(a|\beta, \mu, x)$ is log-concave. In order to generate from the full conditional $\pi(\beta|a^{(s+1)}, \mu^{(s)}, x)$, it is sufficient to simulate a realization from gamma distribution with shape parameter $n$, say $z$, and then update $\beta^{(s)}$ as $\beta^{(s+1)}$ using the relation

\[
\beta^{(s+1)} = \left( \frac{\sum_{i=1}^{n} (x_i - \mu^{(s)}) a^{(s+1)}}{z} \right)^{1/2}.
\]

For full conditional $\pi(\mu|a^{(s+1)}, \beta^{(s)}, x)$, we do not follow the accept–reject sampling method proposed by Green et al. (1994). Our study revealed that when $a$ is small (say $a < 2$) the accept–reject sampling method does not work efficiently (the chin takes too much time for updating). Instead, we use the MH algorithm for generating from full conditional of $\mu$. To do so, we use the uniform distribution on $(x_{(1)} - \beta, x_{(1)})$ as the proposal. We note that the MH algorithm adopted for simulating from full conditional of $\mu$ is faster than the accept–reject sampling method proposed by Green et al. (1994) when $a < 2$. Also, by choosing a uniform proposal, we allow the location parameter to vary over real line. This feature of our proposed prior for the location parameter will appeal to a wide range of study fields in which the three-parameter Weibull distribution with negative location parameter is frequently used to model data.