Pete Clark

PBIO 294 – Ecological Monitoring

Graduate Data Analysis Project

The diameter distribution of a forest is a simple, versatile measure for understanding properties of a stand. For example, tree diameter is correlated to important biological and economic measures including volume, biomass/C, and potential merchantability. The “reverse-J” shape diameter distribution, also known as the logarithmic rotated sigmoid, is regularly referred to by forest practitioners to inform harvests objectives over a rotation period (Leak 2002). Models that capture this diameter distribution are also used for growth-yield models and to forecast long-term stand dynamics in managed and naturally regenerated forests (McGarrigle et al. 2011).

The Weibull distribution, a common distribution used to model diameter classes, was first introduced to forestry by Bailey and Dell (1972) and has been used frequently due to its flexibility (Rennolls et al. 1985; Grissino-Mayer 1999; Green et al. 1994; McGarrigle et al. 2011). The Weibull distribution is a flexible model that can capture tends in a wide range of unimodal distributions, including skewed and mound-shaped curves (Little 1983). The Weibull distribution probability density function (PDF) is most commonly written as a three-parameter function:

where β = is the shape parameter (e.g., slope), η is the scale parameter, and γ is the location parameter. In forestry and elsewhere, where γ = 0 or not given, a two-parameter Weibull is used (cites). Here, the pdf equation reduces to:

The use of the two-parameter versus three parameter Weibull on forests diameter distribution has been inconsistent in the forestry literature and the effects of each has not yet been evaluated. Using tree diameter data aggregated from fifteen 100-400 m2 sized forest plots in Jericho, VT, this paper tests the performance of both model distributions using packaged functions in the computer program R, likelihood estimates, and through a Bayesian framework using a Markov Chain Monte Carlo (MCMC) simulations to examine differences in model performance.

**Methods**

The data and model parameters were first analyzed using the fitdist function in the fitdistrplus package in program R. This package uses the Weibull distribution and allows for rapid assessment of likelihood estimates for the Weibull scale and shape parameters, model assumptions, and goodness-of-fit. Model performance was then assessed using using Akaike information criterion (AIC). Shape and scale parameters were then used as priors in Bayesian models.

Log likelihood functions were then written in R to examine parameter likelihood estimates and model performance for a two- and three-parameter Weibull. Rather than using the Weibull PDF, a cumulative density function (CDF) was calculated for each observation. The three-parameter Weibull CDF is written as:

whereas in the two-parameter Weibull CDF, γ = 0, or is omitted from the equation. Like the Weibull PDF, β = is the shape parameter (e.g., slope), η is the scale parameter, and γ is the location parameter. A CDF is useful because it allows for real-valued objects (here, tree diameters) to be evaluated so that the probability of that value is less than or equal to itself. To assess the CDF, densities for data must be derived. Using ECDF (Empirical Cumulative Distribution Function in R, the data were plotted to visually examine the distribution, then actual density values for each observation were written to a vector in R where all unique values were assigned to corresponding tree diameters. After ECDF values were obtained, a likelihood function was performed on each model. The R script for the log likelihood 3 parameter Weibull model is:

weiCBF3DataLik<-function(parVec, densityData, dbhData, sd){

density<- densityData #ECDF

dbh<- dbhData # DBH

scale<-parVec[1]

shape<-parVec[2]

location<-parVec[3]

three2PWei<-1-e^(-(dbh-location/scale)^shape) #CDF

llik<- sum(dnorm(x=density,mean=three2PWei,sd=sd,log=TRUE))

return(llik)

}

The primary difference in syntax for the 2 parameter Weibull script is that that the location parameter is removed. Function optim was then used on the likelihood functions to examine the best outputs for the parameters and the negative log likelihood for the model. AIC was then examined on each model MLE.

Lastly, both two- and three-parameter Weibull CDFs were examined using JAGS (Just Another Gibbs Sampler) through R packaged rjags. JAGS is a program for the analysis of Bayesian models using Markov Chain Monte Carlo (MCMC) simulations. The following is example code for the model string used for the two-parameter Weibull CDF and priors:

model{

#likelihood

for(i in 1:N) {

density[i]~dnorm(y.hat[i],tau.y) ##### similar to llik

y.hat[i] <- 1-exp(-(dbh[i]/scale)^shape) #2 param weibull cdf

}

#priors

scale~dnorm(0,.0001) #confirm the parameters for these. >0?

shape~dnorm(0,.0001)

tau.y<-1/(sigma.y\*sigma.y) #pow(sigma.y,-2)

sigma.y~dunif(0,100)

}

After the model is compiled, initialized, and adapted, the MCMC generates samples from the posterior distribution for the Weibull model parameters. The coda package was then used to summarize the output from the MCMC simulations including parameter means, quantiles, and plot density of posterior distributions for the parameters simulated. Diagnostic tests of autocorrelation and convergence were conducted in coda and using the Gelman-Rubin statistic.

**Results**

Both the shape and scale parameter estimates from the function fitdist were higher than the estimates reported in the other tests (Table 1). Fitdist only produces an output for two-parameter Weibull parameters, so no location parameter could be obtained. The model AIC deviance was highest compared to the two likelihood functions, indicating that the fitdist model underperformed the data fit (Figure 1). Of the two likelihood functions, the two-parameter model fit the data the best, indicated by the lowest AIC from the three models tested. The three-parameter likelihood model produced a similar scale parameter compared to the other models, but interestngly the shape and location parameter estimates were highly spurious. The two-parameter Wiebull MCMC model simulated parameter posteriors that supported the estimates from the two-parameter likelihood function. MCMC model simulation criterion were met where autocorrelation was very low and upper confidence limits from the Gelman-Rubin statistic for both parameters were 1, indicating chain convergence (Figure 2).

When the empirical cumulative distribution functions were plotted against the modelled (theoretical) cumulative distribution function, the two-parameter Weibull models visually produced the best fits (Figure 3). The three-parameter did not fit the model, which is supported by the results presented in Table 1.

Unfortunately, at the time of this writing, the three-parameter Weibull MCMC code will not run despite best efforts and therefor cannot be assessed.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Shape parameter** | **Scale**  **parameter** | **Location parameter** | **Neg log likelihood** | **AIC** | **Gelman-Rubin**  **Upper C.I.** |
| Fitdistrplus  two-param | 1.820925 | 28.031346 | n/a | -9451.054 | 18906.11 | n/a |
| Likelihood func two-param | 1.627746 | 26.025059 | n/a | 463.4907 | 930.9815 | n/a |
| Likelihood func three-param | 0.09498651 | 25.99994458 | 1.00144076 | 476.924 | 957.848 | n/a |
| JAGS MCMC  two-param | 1.628 (posterior) | 26.025  (posterior) | n/a | n/a | n/a | Shape: 1  Scale: 1 |
| JAGS MCMC  three-param |  |  |  |  |  |  |

Table 1: Parameter estimates and statistics for each model tested

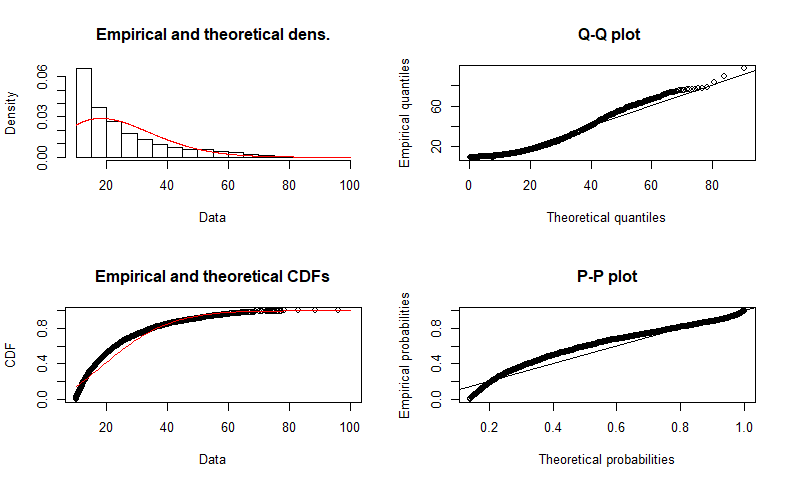


Figure 1: fitdistrplus output including ECDF versus theoretical CDF

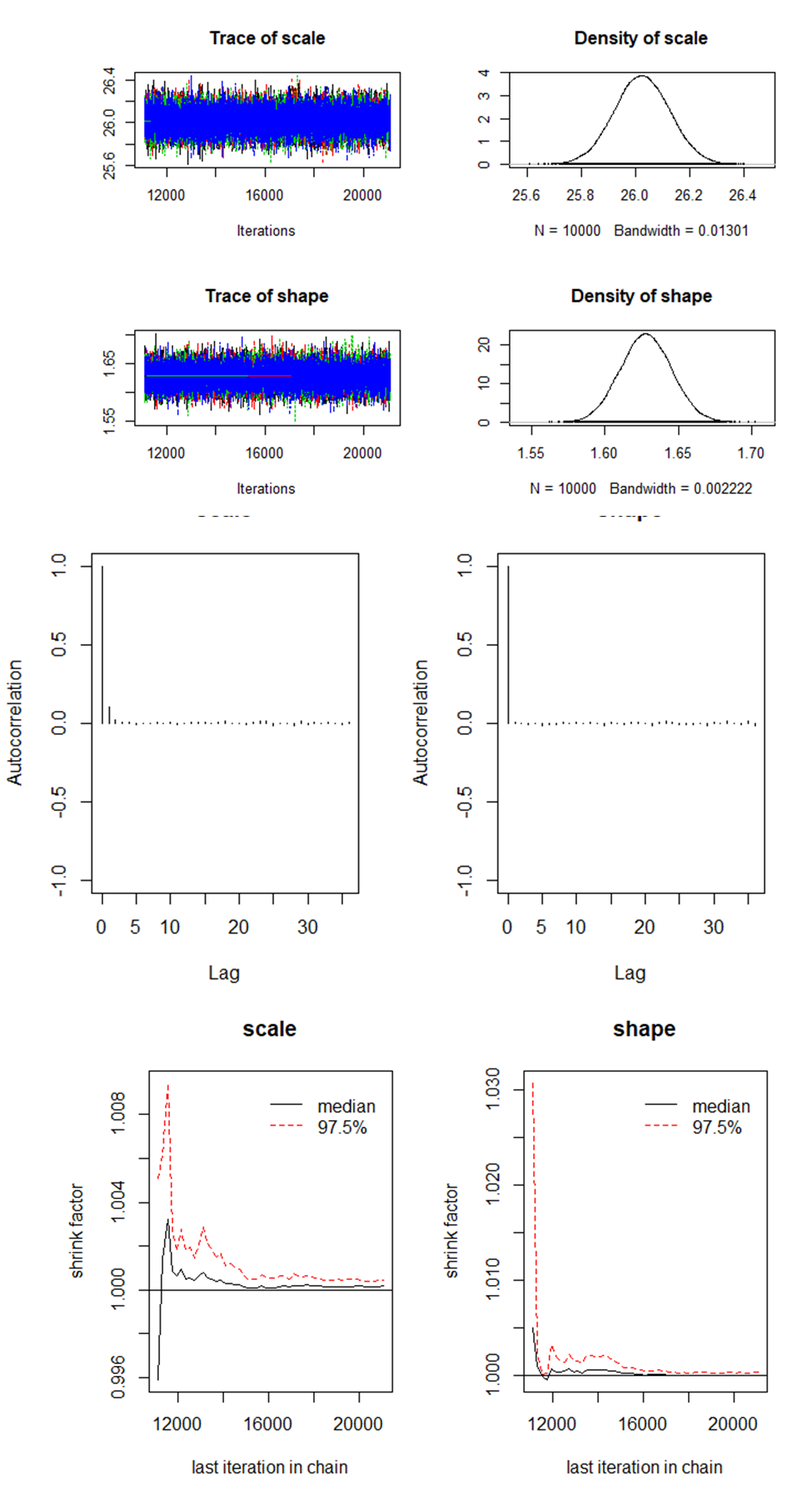


Figure 2: two-Parameter MCMC model performance for the shape and scale parameters. Upper left panel: trace plot convergence for four model chains. Upper right panel: Posterior densities for parameters. Middle panel: lack of autocorrelation in models. Bottom panel: Gelman-Rubin-Brooks plot showing the evolution of the shrink factor as the number of iterations increase

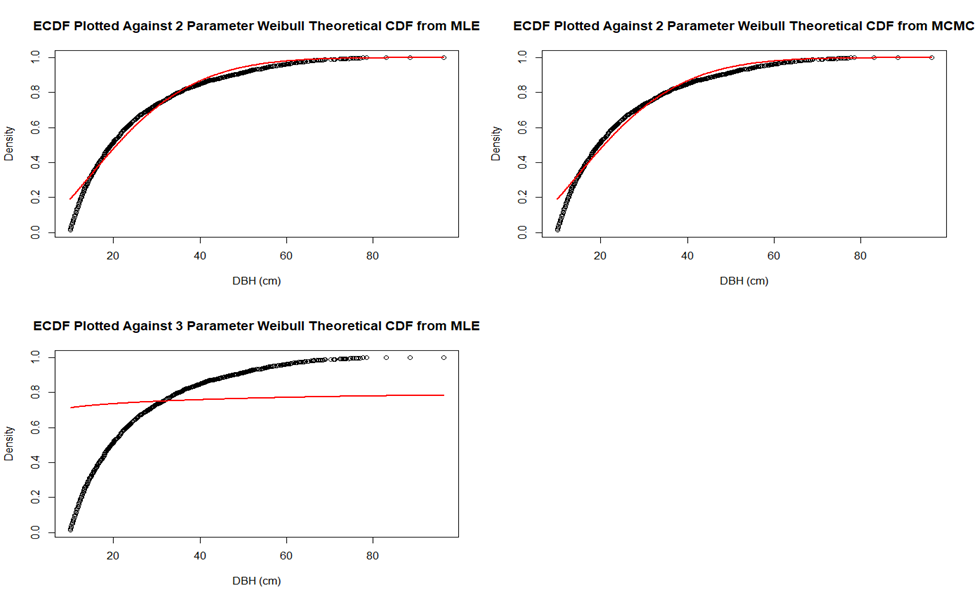


Figure 3: Goodness of theoretical model fit (red line) plotted over empirical cumulative density for two-parameter MLE (upper left), three-parameter MLE (lower left), and two-parameter MCMC (upper right)

**Conclusion**

The use of the two-parameter and three parameter Weibull distribution has been used to model forest stand diameter distributions, however the effects of the two models has yet to be evaluated. This paper used three models, fitdistrplus, MLE, and Bayesian MCMC to assess the performance of a two-parameter and three-parameter Weibull to estimate the cumulative density function for tree diameters. Model results indicate that the two-parameter MCMC and MLE performed the best to flexibly capture ECDF. The three-parameter models failed to improve fits compared to the more parsimonious two-parameter models, indicating that the Weibull location parameter may be superfluous for these data. This result support Grissino-Mayer (1999) who found that the three-parameter Weibull underperformed compared to a two-parameter model when using fire interval data from the American Southwest. The results from the three-parameter Weibull MLE were spurious and suggest a possible error in the model. Additionally, due to coding ineptitude, results could not be obtained for the three-parameter MCMC. Results from this work do indicate that an MLE or MCMC far outperform canned estimates from the fitdistrplus, suggesting that an MLE and MCMC should be used. However, further inquiry into model performance is required to adequately assess the question posed by this paper.

**References Cited**

Bailey, R.L. & Dell, R., 1972. Quantifying Diameter Distributions with the Weibull Function. *Forest Sciences*, 19(2), pp.97–104. Available at: http://ftp.forestaluchile.cl/dasometria/Lectura 001.pdf.

Green, E.J.. et al., 1994. Bayesian Estimation for the Three-Parameter Weibull Distribution with Tree Diameter Data. *International Biometric Society*, 50(1), pp.254–269.

Grissino-Mayer, H.D., 1999. Modeling fire interval data from the American Southwest with the Weibull distribution. *International Journal of Wildland Fire*, 9(1), pp.37–50.

Leak, W.B., 2002. Origin of sigmoid diameter distributions. *USDA Forest Service*, Res. Pap., p.10.

Little, 1983. Weibull diameter distributions for mixed stands of western conifers. *Canadian Journal of Forest Research*, 13.

McGarrigle, E. et al., 2011. Predicting the number of trees in small diameter classes using predictions from a two-parameter Weibull distribution. *Forestry*, 84(4), pp.431–439.

Rennolls, K., Geary, D.N. & Rollinson, T.J.D., 1985. Characterizing diameter distributions by the Weibull distribution. *Forestry*, 58(1), pp.57–66.

**Appendix**: two-parameter Weibull MCMC Coda Summary with parameter means and quantiles

----------

#1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

# Mean SD Naive SE Time-series SE

#scale 26.025 0.10217 5.109e-04 5.566e-04

#shape 1.628 0.01745 8.727e-05 8.825e-05

#2. Quantiles for each variable:

# 2.5% 25% 50% 75% 97.5%

#scale 25.824 25.956 26.025 26.09 26.226

#shape 1.594 1.616 1.628 1.64 1.662