

Grad Student Data Analysis Project

Introduction

On June 1st, 2011, a category EF3 tornado touched down in south-central Massachusetts, traveling east from Monson, MA to Sturbridge, MA, just south of Interstate 90. The tornado caused widespread destruction, notably blowing down a half-mile swath of mature hemlock-hardwood forest through the entire length of Brimfield State Forest, owned by the Massachusetts Department of Conservation and Recreation (MA DCR) in Brimfield, MA. Due to recent state conservation programming, the MA DCR decided against salvage logging the blown-over trees on their property, leaving those trees instead where they had fallen. Nearby private landowners quickly implemented salvage logging operations on their land to recoup some costs from their blown down wood and clean up their property.

This set up an interesting natural experiment surrounding the idea of single and compound disturbance effects in northern hardwood forests: how will the forest regenerate in Brimfield State Forest, where the fallen trees remained on the ground, versus salvage-logged properties, where most down wood was removed? How will this compare to tree regeneration and composition in the parts of the forest untouched by the tornado? Salvage logging, a type of compound disturbance, is not well-studied in New England forests, and its long-term ecological effects post-wind disturbance in this region are relatively unknown. Because increasing global temperatures suggest an increase in the frequency and severity of natural disturbance events in New England, there may be an increased need for implementation of salvage logging practices in the future. Therefore, it is pertinent to study early effects of salvage logging on tree regeneration so that future forest management actions can better inform salvaging efforts to focus on diverse regeneration and resilient forests.

This analysis uses data collected from Brimfield State Forest across three treatment types: tornado-damaged/blowdown (areas hit by the tornado), salvaged (areas hit by the tornado and subsequently salvage logged), and control (areas unaffected by either the tornado or any type of logging). Using Bayesian statistical approaches, this analysis investigates the effect of compound disturbance on the diversity of northern hardwood forest tree regeneration. Specifically, this study asks: does regeneration species diversity, as measured by Shannon's Index, differ by disturbance history (tornado damaged/blowdown, salvaged, and control) and volume of coarse woody debris (CWD) on the ground in 2017?

Methods

Data Description:

In 2012, researchers at the University of Massachusetts Amherst, in collaboration with MA DCR, established 72 permanent research plots (29 tornado-damaged/blowdown, 13 salvaged, and 30 control) across the landscape to study the effects of the tornado and subsequent salvage

logging over time. Later, an additional 9 plots were added to the study, resulting in 27 tornado-damaged/blowdown, 24 salvaged, and 30 control plots for a total of 81 plots. Two tornado-damaged/blowdown plots were lost because they experienced salvage logging and thus switched treatment categories. In 2017, University of Vermont researchers revisited these 81 plots as a five-year follow-up to monitor overstory condition, tree regeneration, and coarse woody debris decay in tornado-damaged plots, salvaged plots, and control plots.

Of the 81 total inventory plots in and around Brimfield State Forest, MA, 70 plots were used in this study. This is because only the original 72 plots experienced two measurements, first in 2012 and second in 2017. The other 9 plots only have one recorded measurement. Additionally, two plots that were classified as tornado-damaged/blowdown in 2012 were salvage logged before re-inventory in 2017. Due to switching treatment categories, these two plots cannot be compared accurately and thus were eliminated from this analysis. Remaining are 70 plots, broken down into 27 tornado-damaged/blowdown, 13 salvaged, and 30 control plots.

Numerous data were collected on-plot, but this analysis focuses on tree regeneration data. This is defined as any woody tree species (i.e. red oak, white pine, red maple) that was beyond first year growth but smaller than 5" in diameter. While overstory data were collected on circular fifth-acre plots, regeneration data were collected on two subplots north and south of the overstory plot center. These subplots were six feet in radius. All tree regeneration present was tallied on each subplot by species and count of stems. Additionally, coarse woody debris (CWD) data were collected on one 100-foot transect per plot at a random azimuth bearing. Every piece of CWD crossed by the transect was identified by species. Diameter at the transect crossing and length of the CWD piece were also recorded.

After field data collection, several forestry metrics were calculated from the raw data. Included in these were Shannon's Index and CWD volume in m³/hectare. Shannon's Index was calculated in R using the 'vegan' package on a plot-wise basis using a matrix of regeneration tree tally data by species. One Shannon Index value was calculated per plot. CWD volume was translated from a transect sample to an area (m³/ha) sample per plot using the equation:

$$Volume = 10,000 * \left(\frac{\pi^2}{8L} \right) \sum d_i^2$$

Where L is the length of the transect (in meters) and d is diameter of each log (in meters).

Data were appended to a spreadsheet that included (1) plot ID number, (2) treatment type (blowdown, control, salvage), (3) Shannon's Index, and (4) CWD Volume.

Data Analysis:

All data analysis was performed in R statistical software and JAGS (using the 'rjags' and 'runjags' packages in R). Initially the original data were used to parameterize the models; however, for this write-up, simulated data were used in place of the original for ease of data transfer. All following tables and figures are made from the simulated data (120 plots, 40 per treatment),

not the original data (70 plots). Simulated data were based on the distribution and parameters of the original data.

This analysis considered three Bayesian JAGS models in order to determine the effect of CWD volume and treatment/plot type on regeneration species diversity, as measured by Shannon's Index (Figure 1): (1) a pooled model, resembling an ordinary least-squares regression, where all three treatment types were "pooled" together for one effect, thereby ignoring any potential unique effects of treatment type; (2) a separate intercepts model, where each of the three treatment types were assigned a separate intercept based on an uninformative prior; and (3) a random intercepts model, where each of the treatment types were assigned a separate intercept based on a more informed hyperprior. Data were modeled with normal distributions, and all parameters have positive values.

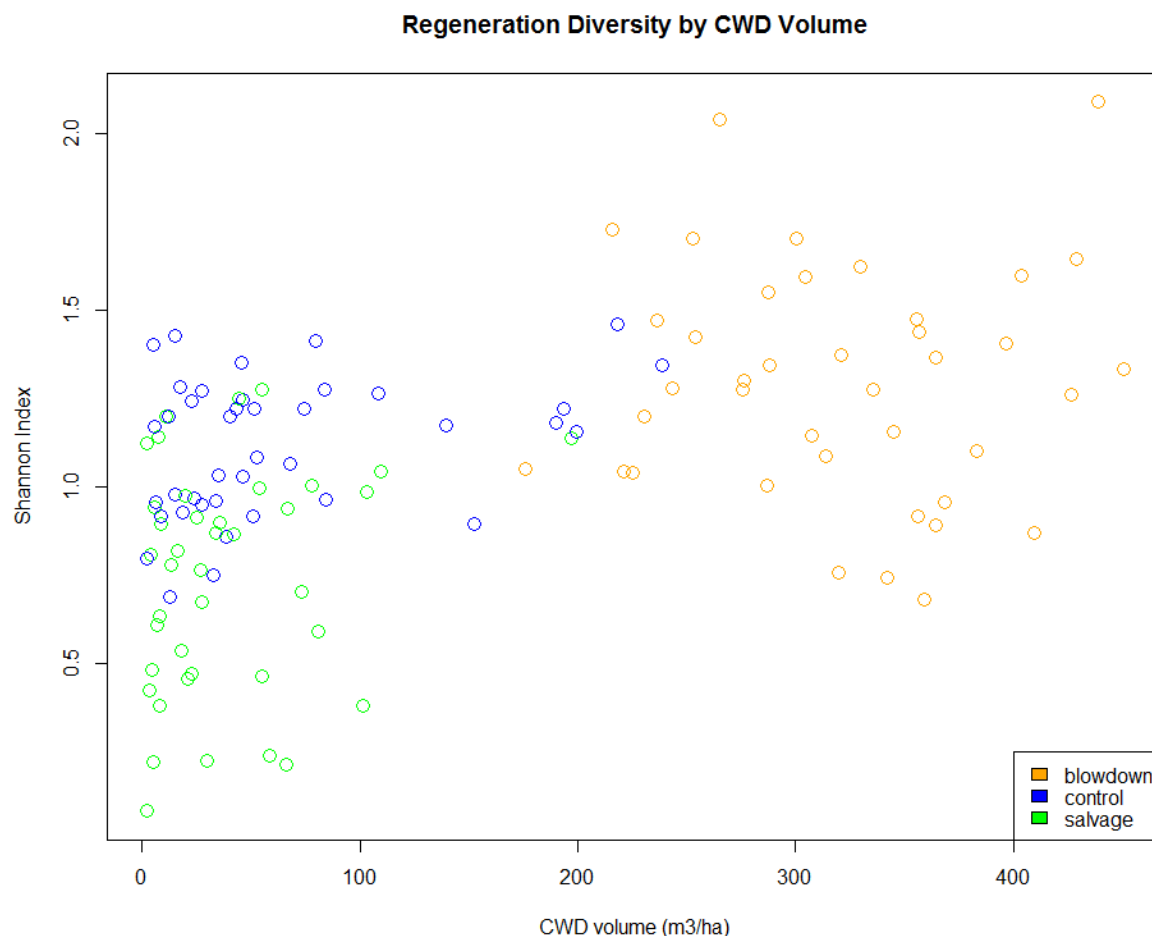


Figure 1: Regeneration Diversity, as measured by Shannon's Index, plotted against Coarse Woody Debris volume (m³/ha). Plot treatments are color-coded. Note: this graph displays simulated data for 120 plots based on distribution parameters of the original raw data.

These data are hierarchical in sampling design. In JAGS, variations on random intercepts models were used to account for potential differences in diversity measures based on treatment type (Table 1). This takes the design of a Bayesian linear mixed-effects model allowing for random effects based on treatment. Model (1), the pooled model, ignores any effect of treatment and essentially describes a Bayesian ordinary least-squares regression of CWD volume against Shannon Diversity. Model (2), the separate intercepts model, includes no random effects, but instead sets a different starting point (or intercept) for each of the three treatments based on uninformative priors. Because there are only three treatments, and these treatments are indicative of a larger ecological phenomenon, this could be a viable mode. Model (3) modifies the second model to include random intercepts; the key difference between this and the second model is that this model includes slightly informative priors, as indicated by a hyperprior on the treatment intercept that makes this a model parameter. Overall, all three models use vague priors because there is no previous ecological knowledge from this dataset to inform any priors.

All models were executed in 'rjags' within RStudio and output DIC values were compared. Each model ran for four chains and was sampled 10,000 times. The 'coda' library was used to plot and assess model fit, chain convergence, and summarize model parameters and quantiles.

Table 1: JAGS model string code for each of the three tested models. The variable 'y' indicates Shannon Diversity, the variable 'x' indicates CWD volume, and the variable 'a' indicates the intercept for treatment type.

Model 1: Pooled Model	Model 2: Separate Intercepts	Model 3: Random Intercepts
<pre> model { # Likelihood Model for (i in 1:n) { y[i] ~ dnorm(y.hat[i], tau.y) y.hat[i] <- a + b1 * x[i] } # Priors a ~ dnorm(0, 100) b1 ~ dnorm(0, 100) tau.y <- pow(sigma.y, -2) sigma.y ~ dunif(0, 10000) } </pre>	<pre> model { # Likelihood Model for (i in 1:n) { y[i] ~ dnorm(y.hat[i], tau.y) y.hat[i] <- a[trt[i]] + b1 * x[i] } for (j in 1:J) { a[j] ~ dnorm(0, 100) } # Priors b1 ~ dnorm(0, 100) tau.y <- pow(sigma.y, -2) sigma.y ~ dunif(0, 10000) } </pre>	<pre> model { # Likelihood Model for (i in 1:n) { y[i] ~ dnorm(y.hat[i], tau.y) y.hat[i] <- a[trt[i]] + b1 * x[i] } for (j in 1:J) { a[j] ~ dnorm(a.hat[j], tau.a) a.hat[j] <- mu.a } # Priors mu.a ~ dnorm(0, 0.000001) tau.a <- pow(sigma.a, -2) sigma.a ~ dunif(0, 10000) b1 ~ dnorm(0, 100) tau.y <- pow(sigma.y, -2) sigma.y ~ dunif(0, 10000) } </pre>

Results and Discussion

Results

In comparing DIC mean deviance values between the three tested models, Model 3 (random intercepts) seems to have performed the best, with the lowest mean deviance DIC value of 70.18 and penalized deviance of 75.34 (Table 2). This indicates that the slightly informative hyperpriors used in this model assisted with better representation of these data. For all models, Gelman-Rubin statistics indicated chain convergence.

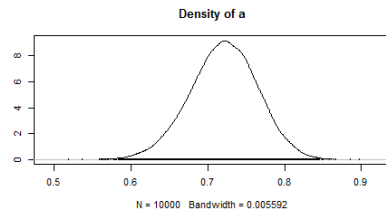
Table 2: Model comparison for three Bayesian models. Estimates of the mean are reported for each parameter in question. Note that because Model 1 considered treatment to be pooled, only one estimate of 'a' is reported for that model.

Parameter	Model 1: Pooled Model	Model 2: Separate Intercepts	Model 3: Random Intercepts
a[1] blowdown	0.7207 (pooled)	0.4144	0.8459
a[2] control		0.4645	0.9114
a[3] salvage		0.4022	0.8160
b1 cwd slope	0.0018	0.0029	0.0014
sigma.y	0.3411	0.4512	0.3262
DIC (mean deviance)	80.76	146.7	70.18
DIC (penalized deviance)	83.87	151.8	75.34

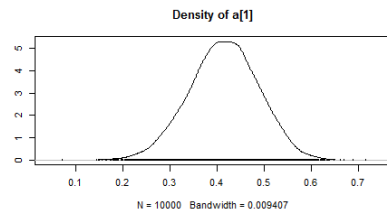
Overall, the three treatment types (blowdown, control, and salvage) had similar intercept values (Table 2), indicating little variability between treatment types in this dataset. Because all three treatment types were forced to take the same CWD slope value (b1), these results suggest that regeneration species diversity, as measured by Shannon's Index, does not vary much between treatment types across a range of CWD values. However, this fixed intercept could be masking additional variability in the data, and future research constructing variable/random slope models by treatment type may indicate different rates of diversity change given increasing volumes of CWD (Figure 3).

Posterior distributions from all models capture parameter means (Figure 2). However in Model 3, the sigma.a parameter does not seem fully represented by the posterior, suggesting that perhaps different priors are necessary. In Model 1, the posterior distribution does not capture the same set of values as Model 3 with random intercepts. This may indicate that Model 1 is not a good fit for these data, and separating out intercepts (and possibly slopes; data not tested here) for each of the three treatment types creates a more accurate model for these data.

Model 1: Pooled Model



Model 2: Separate Intercepts



Model 3: Random Intercepts

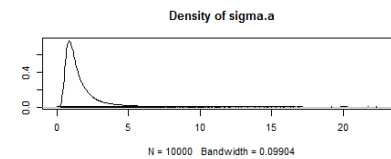
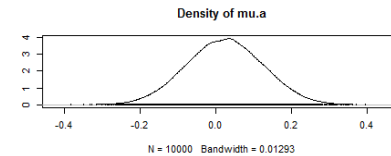
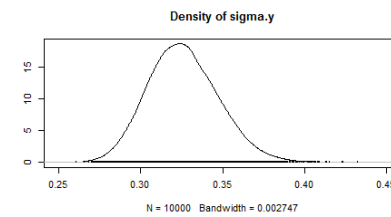
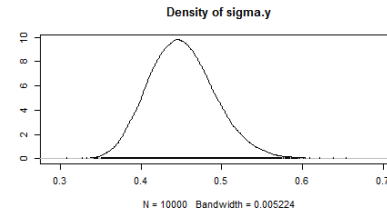
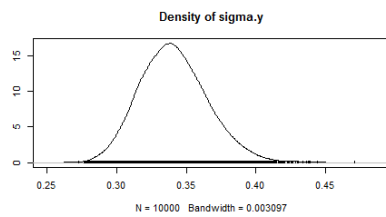
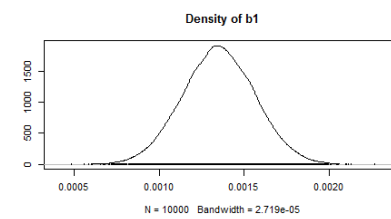
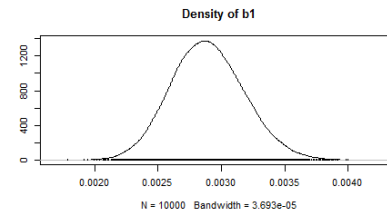
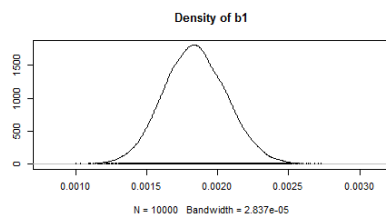
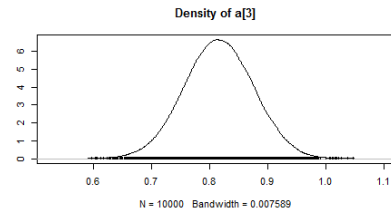
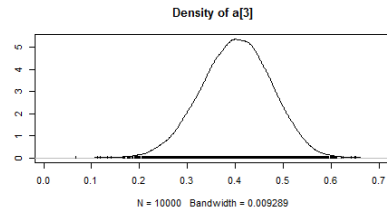
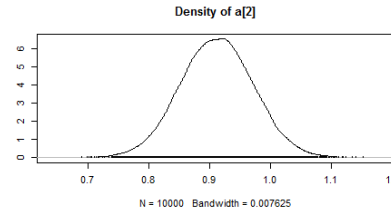
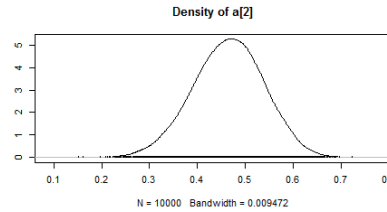
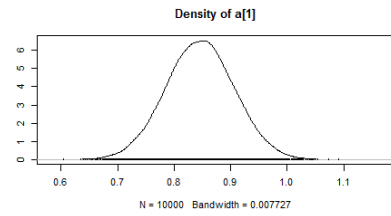


Figure 2: Posterior distributions of the parameters in the three tested models. From left to right: Model 1 (pooled model; note that only one 'a' parameter value is modeled due to the pooled effect of treatment), Model 2 (separate intercept model), and Model 3 (random intercept model).

Discussion

Overall, these models – particularly Model 3 – provide a start in describing how regeneration species diversity, as measured by Shannon's Index, varies based on CWD volume across treatment type. Even though the intercepts for all three treatments are very similar, this model indicates that the control plots have the highest regeneration species diversity values based on CWD volume, and that in general, regeneration species diversity tends to increase with increasing CWD volume remaining on the ground. This has implications for post-disturbance management: these data suggest leaving down wood on the ground (i.e. no or minimal salvage logging) may lead to an increase in initial post-disturbance regeneration species diversity (Figure 3). Constructing a model with variable/random slopes or non-linear slopes may drastically change the outputs of this model and provide a better indication of how regeneration species diversity changes with increasing CWD volumes and treatment type, and will better inform management recommendations.

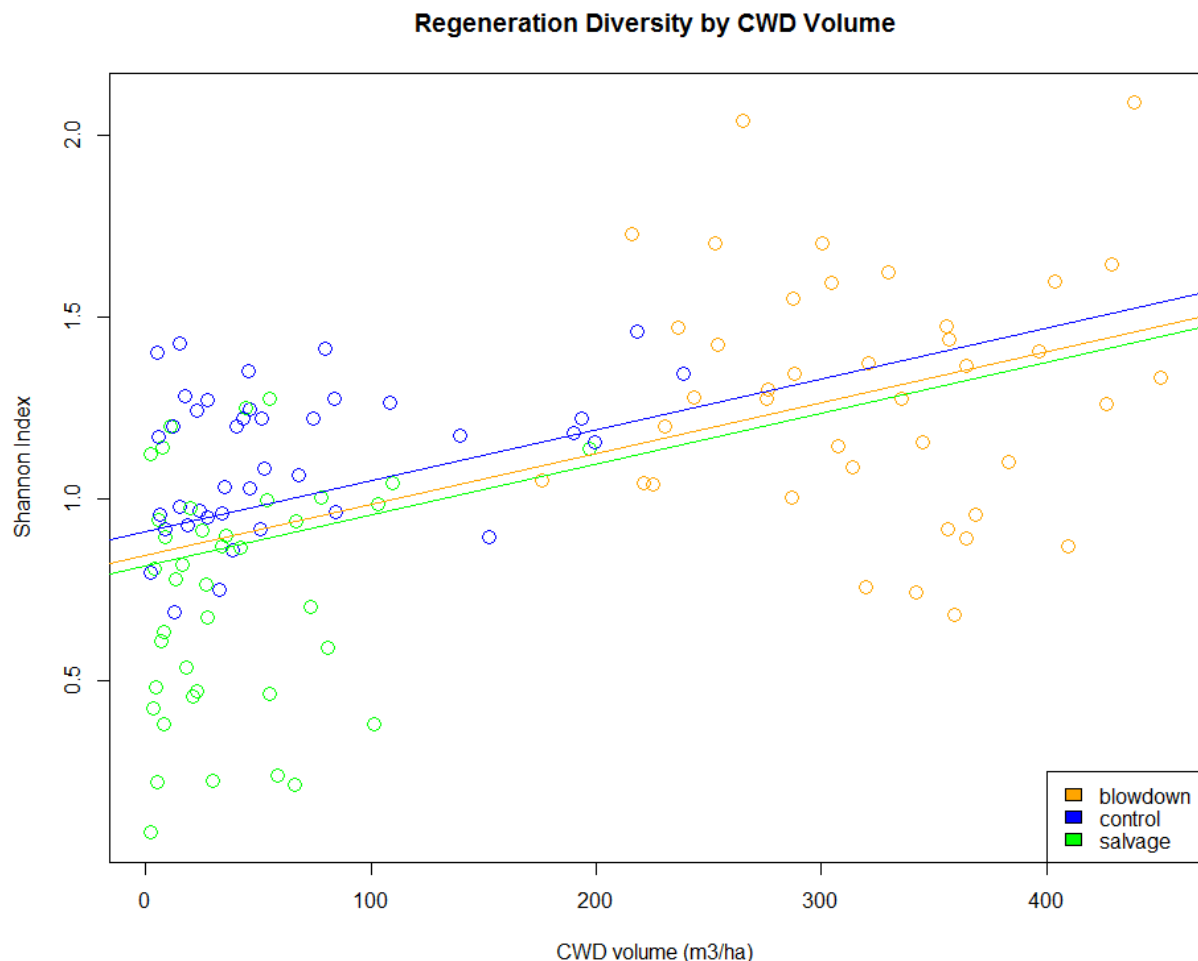


Figure 3: Regeneration Diversity, as measured by Shannon's Index, plotted against Coarse Woody Debris volume (m^3/ha). Plot treatments are color-coded and regression lines have been added based on the slope and intercept values from Model 3 (also color-coded). Note: this graph displays simulated data for 120 plots.

Future analyses will consider other covariates in these models (such as light availability) and the addition of year (2012 and 2017) as a temporal effect along with corresponding diversity and CWD measurements for those years. Additionally, performing this analysis in the 'rstan' package would be advantageous to utilize STAN's built-in limits (i.e. <lower>) to confirm prior distributions and parameter estimates do not dip below zero, which is ecologically impossible given the parameters in this dataset (both Shannon's Index and CWD volume must be positive numbers).

Future modeling efforts should consider a random slopes and intercepts model or a hierarchical model to account for potential differences by treatment type (Figure 3). This may help to tease apart potential differences in CWD and species diversity between treatment types. Finally, these models used purposely broad priors, as no previous ecological information was known before completing this study. Subsequent analyses could tweak these priors in an attempt to provide additional information on the dataset.

Appendix

Table 3: Quantiles for Model (3): random intercept model.

Parameter	2.5%	25%	50%	75%	97.5%
a[1] blowdown	0.7291	0.8061	0.8465	0.8869	0.9639
a[2] control	0.7932	0.8709	0.9111	0.9512	1.0283
a[3] salvage	0.6990	0.7761	0.8160	0.8556	0.9328
b1 cwd slope	0.0009	0.0012	0.0014	0.0015	0.0018
mu.a	-0.1743	-0.0451	0.0242	0.0924	0.2221
sigma.a	0.5021	0.8514	1.2099	1.8937	6.1077
sigma.y	0.2864	0.3109	0.3250	0.3402	0.3723