Bayesian Analysis of Montesinho Forest Fire Data

Spencer Ebert

December 9, 2019

1 Introduction

In the past couple years, forest fires have seemed more prevalent especially in the West. Experiences like the campfire in Paradise California in 2018 cause drastic economic and ecological damage. Understanding factors that play in forest fires will help mitigate damage caused by the fire, and help prevent future fires. The goal of this analysis is to better understand the factors that play into forest fires and what causes them to be more harmful.

2 Data

Data were gathered in Portugal for forest fires in the Montesinho natural park from January 2000 to December 2003. Montesinho natural park is one of the largest natural parks in Portugal and covers 74,230 hectares of land. One hectare is equivalent to 100 square meters. The park contains many species of wildlife and many outdoor recreational activities. Protection of the park is very important to the ecosystems and livelihood of those that live around it.

Every time a fire occurred several variables were recorded. The dataset consists of 517 observations and twelve measurements for each fire. Four measurements from a Canadian system for rating fire danger called Fire Weather Index (FWI) were recorded. They are Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), and Initial Spread Index (ISI). These measurements give information on the moisture content of the soil, the amount of burnable fuel on the forest floor, and expected speed of spread. High values for each of these measurements indicates ideal conditions for a forest fire to spread. The higher the FFMC, the drier the fuel. In addition to these measurements, we have variables for temperature in Celcius, relative humidity percentage (RH), wind speed in km/hr, outside rain in mm/m2, month (Jan, Feb, etc.), day (Sunday, Monday, etc.), and total area burned measured in hectares.

The response used in this analysis is the total area burned. Interestingly, the total area burned is zero for 247 observations, and for the remaining 270 observations the area burned is highly skewed towards zero. Every observation in the dataset corresponds to an actual fire occurence, but a measurement of zero indicates that the fire was less than 0.01 hectares. Because of the skewness in the data, a lognormal distribution is used. Figure 1 illustrates the skewness and effect of a long transform.

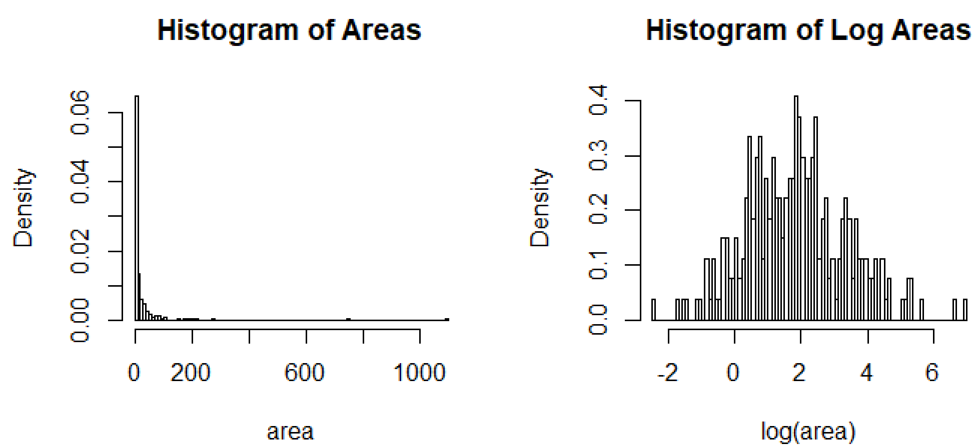

Figure 1

3 Model

Since half of the areas are zero and the rest have the property of continuous skewed data, I fit a mixture model with a probabilty of $1 - \pi_i$ that the area burned is zero and probability π_i that it follows a lognormal distribution. The formula for the model is given in equation 1.

$$\begin{aligned}
y_i|\pi_i, \mu_i, \lambda &\overset{ind}{\sim} (1-\pi_i)\delta_0(y_i) + \pi_i \text{Lognormal}(\mu_i, \lambda) \\
\mu_i &= x_i \boldsymbol{\beta} \\
\pi_i &= \frac{1}{1+e^{-\tilde{x}_i \boldsymbol{\alpha}}} \\
\lambda &\sim \text{Gamma}(2, 2) \\
\beta_j &\overset{iid}{\sim} \text{N}(0, 0.001) \quad \text{for} \quad j = 1, 2, ..., 12 \\
\alpha_k &\overset{iid}{\sim} \text{N}(0, 0.001) \quad \text{for} \quad k = 1, 2, ..., 5
\end{aligned} \tag{1}$$

The x_i vector in this model consists of a column for the intercept, two columns for the season, one column for weekday, and the rest for FMC, DMC, DC, ISI, RH, temperature, wind, and rain. The $\tilde{x}_i$ vector consists of a column for the intercept, FMC, wind, and two columns for season. For the other variables, λ is the precision, y_i is the area burned, β_j is the linear effect of the j^{th} predictor and represents the effect of the predictor on the size of the fire. α_k is the effect the k^{th} predictor has on the probability and represents the probability of having a substantial fire.

There are two categorical variables in this data for month and day. Including a column for each month and day of the week causes columns in the X matrix to be a linear combination of the others. Because of this, I split month into three categories of fire season (July, August, and September), fall/winter (October, November, December, January, Febuary, and March), and spring/early summer (March,April, May, June). Most fires occur in the fire season months so I split up the months based on difference from fire season. From the research I did on forest fires, I found that fires tend to be caused more frequently on weekends when people are out and about, thus I split up days into weekdays (Monday, Tuesday, Wednesday, Thursday, Friday), and weekends (Saturday, Sunday).

I chose non-informative priors for β_j, α_k, and λ because I don't have the expertise to know how these predictors affect forest fires and this allows the data to have a stronger influence on the outcome. Choosing a gamma prior for λ allows conjugacy in the model. Discussion on sensitivity to different priors is given later.

The main goal of the analysis is to understand the parameters λ, $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$ so I take draws from the posterior of these parameters given the data. Formulas for the posterior distribution, joint distribution of $\boldsymbol{y}$, and the full conditional of λ are given in equations 2, 3, and 4 respectively.

$$p(\lambda, \boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\lambda, \boldsymbol{\alpha}, \boldsymbol{\beta}) p(\lambda) \prod_i^5 p(\alpha_i) \prod_j^{12} p(\beta_j) \tag{2}$$

$$p(\boldsymbol{y}|\lambda, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i:y_i=0} \left(\frac{e^{-\tilde{x}\alpha}}{1+e^{-\tilde{x}\alpha}} \right) \prod_{i:y_i>0} \left(\frac{1}{1+e^{-\tilde{x}\alpha}} \frac{\sqrt{\lambda}}{y_i\sqrt{2\pi}} e^{-\lambda(log(y_i)-x_i\beta)^2/2} \right) \tag{3}$$

$$p(\lambda|\cdot) \propto \lambda^{0.5\sum_{i=1}^n (I(y_i>0))+2-1} \exp\left[-\lambda \left(2 + 0.5 \sum_{i:y_i>0} (\log(y_i) - x_i\beta)^2 \right) \right] \tag{4}$$

The full conditional of λ follows a gamma distribution with $\alpha_{post} = 2 + 0.5\sum_{i=1}^n (I(y_i > 0))$ and $\beta_{post} = 2 + 0.5\sum_{i:y_i>0}(\log(y_i) - x_i\beta)^2$. This fact allows me to use Gibbs sampling to update λ.

4 Drawing From Posterior

To sample draws from the posterior distribution, I utilized the Metropolis-Hastings alorithm, and Gibbs sampling. One of the biggest challenges in getting the algorithm to converge was the correlation between the β_j's and the correlation between the α_k's. Univariate updates don't perform well in this situation so I used draws from a multivariate normal to update $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ separately. To help with the correlation, I utilized the fact that the β_j's should be roughly correlated with each other following the matrix $(X'X)^{-1}$. The X matrix contains a row for each observation that had an area burned greater than zero and the columns correspond to the columns from x_i in equation 1. Likewise, I assumed that the α_k's would be roughly correlated with each other following the matrix $(\tilde{X}'\tilde{X})^{-1}$, with a row for each observation that had an area burned of zero and columns that correspond to the columns in $\tilde{x}_i$ in equation 1. The form of the covariance matrices stayed the same. I only had to adjust the constant to multiply the matrices by to get efficient convergence. The process used to get draws from the posterior is shown below.

1. Begin with initial center of 0.01 for λ and 0 for the other parameters.
2. Update λ by using Gibbs sampling drawing from the full conditional $\lambda|\cdot$.
3. Update $\boldsymbol{\beta}$ by using the Metropolis-Hastings algorithm. The proposal distribution is a multivariate normal with mean the previous draw and covariance matrix $1.5(X'X)^{-1}$.
4. Update $\boldsymbol{\alpha}$ by using the Metropolis-Hastings algorithm. The proposal distribution is a multivariate normal with mean the previous draw and covariance matrix $1.5(\tilde{X}'\tilde{X})^{-1}$.
5. Repeat steps 2-4 for 50,000 draws.
6. Burn the first 1000 draws.

This algorithm is not computationally expensive on my machine and runs for about 15 seconds. The time adds up when I run the algorithm multiple times to compare different models, but overall the computational burden is fairly low.

5 Diagnostics

5.1 Different Center Values

I wanted to check whether my posterior draws were possibly getting stuck in a local mode so I ran the algorithm two more times but with different starting values. I used center values for extremes with the first being $\lambda_{start} = 0.001$, all the starting β's equal to 0.1, and the α's equal to 0. The next starting values I used were $\lambda_{start} = 3$, the β's equal to -0.1, and the α's equal to 0. A value of 0.1 is quite extreme for some values of β_j because of the scale, and I kept the α's the same because they were very difficult to tune. Figure 2 shows trace plots for the different starting values. Different starting values cause the algorithm to jump around for quite awhile, but about halfway through the iterations they finally converge to the same point where the well behaved starting values converge. This makes me feel comfortable that I am not getting stuck in a local mode during the algorithm.

5.2 Trace Plots and ACF Plots

To determine whether my posterior draws were converging and how many draws I need to burn, I looked at trace plots and acf plots for the stable starting center value. Figure 3 shows trace plots for the log posterior with and without burn-in and figure 4 shows autocorrelation for some of the variables.

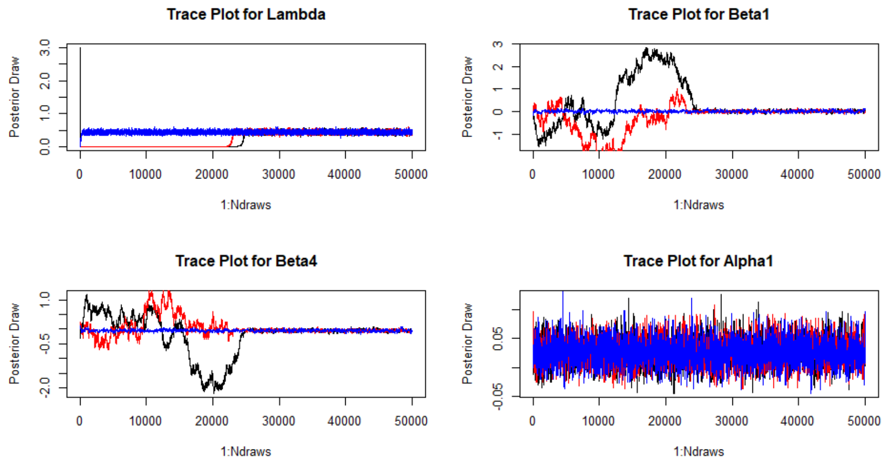

Figure 2: Blue lines indicate the normal starting values, red lines indicate the second set of starting values, and black lines indicate the third set of starting values.

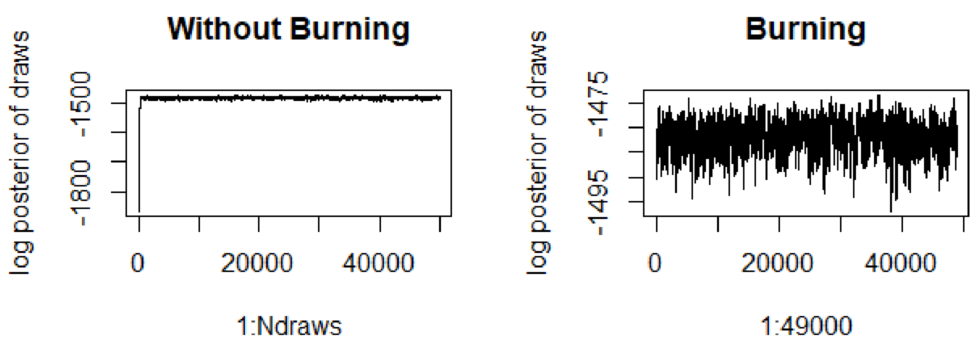

Figure 3

According to the trace plots in figure 3, it seems that the algorithm doesn't converge until about 500 draws. Because of this, I burn the first 1000 draws to be on the safe side leaving me with 49000 good draws.

The acf plot for λ looks really good but β_1 seems to have a higher correlation. All of the other β_j acf plots look similar and α_k acf plots fall between λ and β. The high correlation makes it harder to get more effective samples for β_j as can be seen in the plot. Because of this, I increased the number of draws to 50000. I initially had the draws at 10000, but wanted to get more effective draws from the posterior to overcome the autocorrelation.

5.3 Numerical Diagnostics

Looking at the trace plots and acf plots indicates sufficient convergence for the algorithm, but I also look at the effective sample size for each of the parameters. The effective sample sizes are given in table 1. These results show that λ has a fairly high effective sample size compared to the other parameters, and the effective size for the other parameters is sufficient for the analysis. I also calculated the acceptance rate for the multivariate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ draws: 0.35 and 0.18 respectively. They are not perfect acceptance rates, but in combination with the other diagnostics I feel confident my chain converged to

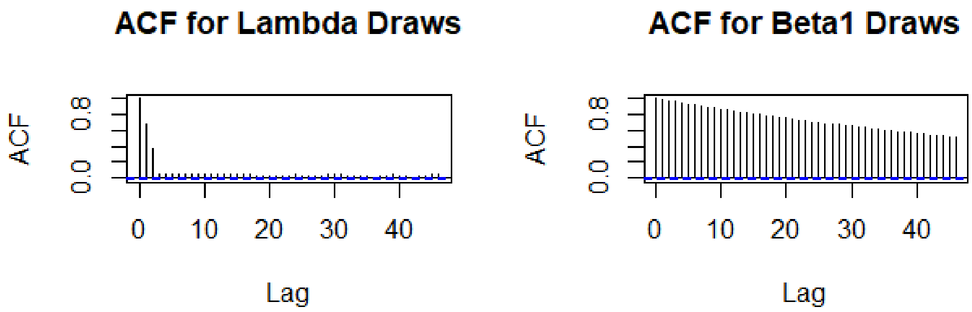

Figure 4

Table 1: Effective Sample Sizes

λ	β_0	β_1	β_2	β_3	β_4	β_5
7126.80	361.58	357.84	399.21	328.84	403.21	369.44
β_6	β_7	β_8	β_9	$\beta_1 0$	$\beta_1 1$	α_0
412.29	400.08	483.93	376.69	372.96	375.40	1314.48
α_1	α_2	α_3	α_4			
1331.58	702.99	879.25	958.34			

the posterior.

6 Model Checks

To check the model fit, I used a posterior predictive test for the the general goodness of fit, and looked at plots of the residuals. The computation for the general goodness of fit is shown below.

1. Simulate the posterior predictive distribution $p(y^{rep}|y)$ by sampling data based on each draw from the posterior distribution.

2. Compute expected value of the $y's$ for each draw from the posterior distribution. It can be shown $E(y_i|\lambda, \alpha, \beta) = \frac{1}{1+e^{-\tilde{x}_i\alpha}} e^{x_i\beta+0.5/\lambda}$

3. Compute test statistic for the predictive distribution and actual observation. $T(y, \theta) = \sum_{i=1}^{n} \frac{(E(y_i|\theta_i)-y)^2}{E(y_i|\theta_i)}$

4. Compare number of times the predictive statistic is greater than the actual and report p-value.

After following these steps, I got a p-value of 0.16. This p-value is fairly low, but it is not too extreme, so I don't reject the fit of my model. I also plotted the residuals and checked for any odd patterns. Figure 5 shows the residuals from the model. They don't seem to follow a particular pattern, which makes me feel comfortable with the fit of the model.

The last thing I checked was the model sensitivity to different priors. The initial priors I started with are shown in equation 1. For the first test prior, I made the hyperparameters for λ be 0.1 and the precision for both β and α 1. The other prior I tested was a hyperparameter of 5 for λ and precision of 0.00001 for β and α. Results for the parameters using these priors are shown in table 3. The parameter estimates are very similar to the final model so I feel good about the sensitivity of my model to different priors. This indicates that the model is being largely driven by the data.

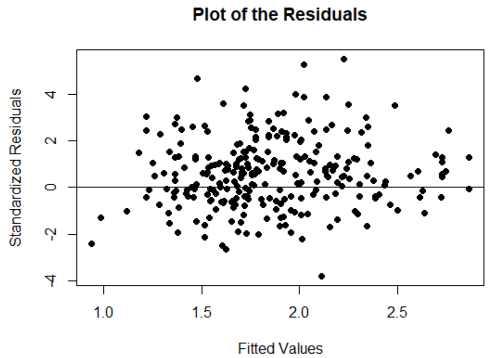

Figure 5

7 Model Comparison

The criteria I used in this analysis to compare models was DIC. The lower the DIC the better. Initially, I used a model with π not related to any of the predictors x. Since I wanted to understand how the predictors influenced the probability of having a substantial fire I made π_i equal to the function shown in equation 1 except $\tilde{x}_i$ included all the predictors. Interestingly, the model where π wasn't influenced by any of the predictors had a better DIC than the model using all of the predictors. The reason for this is that a lot of the parameters didn't have a significant effect on the probability. To help combat this issue I used backward selection based on DIC. The steps I used for my backward selection are shown below.

1. Include all predictors in the sample space. The length of $\boldsymbol{\alpha} = 12$.
2. Eliminate the α_i that has the lowest significance. Significance calculated as the mean of the α_k that is greater than 0 from the posterior draws.
3. Compute DIC on the new model.
4. Repeat steps 2 and 3 until the DIC does not get better.

Results using this method are shown in table 2. After eliminating unessential variables I ended up with a better DIC than using none of the variables.

Table 2: DIC using different models. Each row represents which predictor was eliminated.

	DIC
All Variables	2725.69
Eliminate DC	2724.96
RH	2721.23
DMC	2720.21
rain	2718.78
weekend	2716.80
ISI	2715.02
temp	2713.69
season	2717.32
No variables	2717.98

8 Results

After investigating the model fit and comparing to other models, I used the model in equation 1. Results for the parameters are found in table 3. Interestingly, most of the parameters in this model are not significant. The parameters that have the most significant impact are β_{DMC}, $\beta_{Weekend}$, α_{wind}, and $\alpha_{Spring/Summer}$. For higher DMC (drier deep organic layers) the total area burned goes up, and if the fire occurred on a weekend then the area burned goes up. The probability of having a substantial fire is primarily driven by wind and time of year. If the wind is higher on a particular day, then it is more likely that the fire will spread quickly and become substantial. Also, if the time of year is in the spring and early summer then there is a lower chance of having a substantial fire than during the fire season. I found it interesting that wind is significant for the probability of having a substantial fire, but it is not significant for β_{wind}. This indicates that high winds have a greater affect on creating a substantial fire through faster spread, but they do not necessarily make the area burned larger. Park rangers need to be aware of these factors so they can help mitigate the effect fire has. They need to be extra watchful on windy days to stop a fire before it becomes out of control and try to reduce DMC so fires won't cause more damage if it does become substantial.

Table 3: Results for final method, frequentist analysis, and different priors

	Expected Value	2.5%	97.5%	Frequentist	Prior 2	Prior 3
λ	0.4408	0.3678	0.5188	0.4368	0.4429	0.4408
β_{Int}	0.1250	-7.5567	8.0006	0.5021	0.0709	0.1250
β_{FFMC}	0.0184	-0.0676	0.1078	0.0150	0.0219	0.0184
β_{DMC}	0.0039	-0.0006	0.0082	0.0038	0.0036	0.0039
β_{DC}	0.0002	-0.0014	0.0018	0.0002	0.0001	0.0002
β_{ISI}	-0.0479	-0.1162	0.0192	-0.0458	-0.0503	-0.0479
β_{temp}	-0.0051	-0.0602	0.0474	-0.0074	-0.0083	-0.0051
β_{RH}	-0.0101	-0.0264	0.0064	-0.0107	-0.0114	-0.0101
β_{wind}	0.0406	-0.0605	0.1449	0.0427	0.0460	0.0406
β_{rain}	0.1075	-0.3468	0.5459	0.1045	0.1156	0.1075
$\beta_{weekend}$	0.4166	0.0047	0.8187	0.4047	0.4177	0.4166
$\beta_{Fall/Winter}$	0.8610	-0.1565	1.8931	0.8129	0.6627	0.8610
$\beta_{Spring/Summer}$	0.4904	-0.5097	1.5034	0.4796	0.3071	0.4904
α_{Int}	-2.4892	-6.3437	1.1486	-2.2950	-0.5879	-2.4892
α_{FFMC}	0.0255	-0.0137	0.0675	0.0235	0.0051	0.0255
α_{wind}	0.0934	-0.0099	0.1949	0.0917	0.0840	0.0934
$\alpha_{Fall/Winter}$	-0.0218	-0.7288	0.6746	-0.0445	-0.1446	-0.0218
$\alpha_{Spring/Summer}$	-0.6806	-1.2179	-0.1615	-0.6803	-0.6720	-0.6806

8.1 Comparison to Frequentist Analysis

To further test the conclusions from my Bayesian analysis, I used a combination of frequentist analyses to compare the parameter estimates. First, I utilized multiple linear regression on only the observations that had an area burned greater than zero. The response was the log of area burned. I also performed logistic regression on the indicator of area burned being greater than zero $I(y_i > 0)$. I included all the observations and the predictors included correspond to $\tilde{x}_i$ from equation 1. After performing these two analyses I compared the parameter estimates to those I got from the Bayesian analysis. Results are found in table 3. Most of the parameter estimates are the same as in the Bayesian analysis. The only big differences are the intercepts, but the intercepts have high variability in both models anyways.

One advantage to using the Bayesian approach is I can compare parameter estimates between α_k and β_j. Either way though the conclusions are the same.

9 Conclusion

Forest fires continue to be a threat and it is important to continually research the causes. This analysis identified key components to fire danger by utilizing Bayesian techniques. The results found here are not particularly surprising, but they help further emphasize the factors that influence more dangerous forest fires. Efforts to identify forest fires on particularly dangerous days can help reduce the number of substantial fires. Some methods that are already in use to reduce ground litter are effective and should be continued according to this analysis. Also, methods to determine a starting forest fire could be used in combination with this analysis to help combat the spread before it becomes unmanageable.

One of the main struggles encountered in this analysis was the collinearity between the predictors. FWI measurements are often based on other measures like wind and rain, which makes it difficult to distinguish what factor is having the greatest effect. Another weakness was the low p-value for the posterior predictive test. Moving forward, it would be interesting to investigate the collinearity between these variables and entertain further models. I am also interested in gathering forest fire data specifically from the West Coast because of its proximity.