Bayesian Classification and Survival Analysis of Fire Size Data PHP2530

Rachel Yost

May 12, 2025

1. Introduction

The intensity and frequency of wildfires is increasing, with the frequency of extreme events doubling between 2003 and 2023. Certain regions, such as the boreal and temperate conifer biomes, are especially at risk (Cunningham et al., 2024). Wildfires can pose a risk to human health and the surrounding environment (Bowman et al., 2017). Notably, in January 2025, the Los Angeles wildfires burned 60,000 acres and took 29 human lives (Kilkenny et al., 2025). Consequences of these fires will continue to increase as we continue to develop near wooded areas at-risk of fire. Understanding what causes forest fires, what causes them to increase in size, and how fire management agencies can best combat these fires, will be integral to preventing and controlling future fires.

A 2018 paper used classification and survival analysis to describe forest fire size (Tremblay et al., 2018). Using survival analysis to quantify factors that affect fire size distribution is a novel idea from these researchers. They suggest using survival analysis since it uses "time to event" as a response variable, which are positive random variables with a skewed distribution, much like fire sizes, and also exhibit censoring and truncation (Strauss et al., 1989). Fire data would be censored when the fire is still growing when it is measured at the state of "being held". "Being held" is a term used when "with currently committed resources, sufficient suppression action has been taken that the wildfire is not likely to spread beyond existent or predetermined boundaries under prevailing and forecasting conditions" according to the Canadian Wildland Fire Glossary. The data is also truncated because the fire has a positive size at its first measurement.

The fire size distribution could potentially be affected by the type of fuel, the moisture content of the fuel and the fire management practices (Cumming, 2001; Martell and Sun, 2008; Arienti et al., 2006), so in Tremblay et al. (2018), they use survival analysis to test the effects of certain types of intervention and fire weather information on fire size distribution. They use data from fires in northern Alberta, Canada. They also used classification to classify which fires grew between their initial measurement (size at initial attack (IA)) and their size at being held (BH), and which did not grow at all. Only those who grew were included in the survival analysis.

The original authors used frequentist methods to perform their analyses, but we propose that repeating their analyses with a Bayesian framework would be useful. Bayesian inference combines prior information about the model parameters with the likelihood of the data, to produce an updated posterior distribution of the parameters.

We have information from a prior survival analysis study about the effects of certain parameters on the duration of fire (Morin et al., 2015), which is closely related to fire size (Tremblay et al., 2018). Additionally, the authors notes that the variable, fire load, which measures how busy the management system was at the time of the fire, was not significant in their analysis, but could potentially have an interaction with attack method, but they did not have appropriate sample sizes to investigate this factor. Since we will use informative priors, we can include this interaction, since Bayesian survival analysis can handle small sample sizes (Van De Schoot et al., 2015).

We will use the same dataset as the original authors and perform classfication using Bayesian logistic regression and Bayesian Generalized Additive Models (GAMs). We will then perform a survival analysis using a Bayesian Weibull Accelerated Failure Time model, including the interaction between fire management method and fire load. The data will be divided into a test and training set, and we will assess classification model accuracy and auc/roc. From the survival models, we will compare the cumulative hazard of a fire ”being held”. We will also investigate the effect of fire control method on the survival curves.

2. Data Structure and Notation

The dataset used for this study was originally used by the paper, ”Survival analysis and classification methods for forest fire size”, by Tremblay et al. 2018. It contains a sample of 889 fires from Alberta, Canada. Each fire has two size measurements, once at its initial assessment, and one at its state of ”being held”, which means that they do not expect the size of the fire to increase. Other variables include the ”fire weather index”(FWI) from the day the fire started, ”initial spread index”, which describes the expected rate of fire spread, the fire load, if it started in the AM or PM, fuel type, method of intervention, method of initial detection, response time, and month.

A fire that did not grow was classified with $Y_i = 1$, and a fire that did grow was classified with $Y_i = 0$, with i representing the i_{th} fire. This representation seems backwards, but it is what is used in the original study.

For the survival analysis, we need an indicator for whether or not the fire size is censored, meaning that it grew beyond its last measured size. However, our data did not have any censored observations, so our censoring indicator variable is equal to one for all of the fires. For the rest of our paper, t_i represents fire size for fire i, and $\mathbf{x}_i$ is the covariate matrix for fire i.

3. Methodology

3.1 Bayesian Logistic Regression

We considered Bayesian logistic regression and Bayesian GAMs to classify whether or not a fire grew. In a Bayesian logistic regression, the outcome is binary 0,1, with conditional density $f_{y|x}(y_i|x_i, \omega_{y|x})$, which is a bernoulli distibrution, where the parameters $\omega_{y|x}$ are β. The mean of the bernoulli $\eta(x) = g^{-1}(x_i'\beta)$ is in [0,1]. We can use the inverse logistic function, which is $g^{-1}(z) = e^z/(1 + e^z)$, for some real number, z. This means that the model can be $y|x_i \sim Ber(\eta(x_i))$ where $\eta(x_i) = e^{x_i'\beta}/(1+e^{x_i'\beta})$. This results in the likelihood $L_{Y|X}(\omega_{Y|X}|D) \propto \prod_{i=1}^{n}(e^{x_i'\beta}/(1+e^{x_i'\beta})^{y_i}(1-(e^{x_i'\beta}/(1+e^{x_i'\beta}))^{1-y_i})$. Thus, the posterior is $f(\omega_{Y|X}|D) \propto f(\beta) * \prod_{i=1}^{n}(e^{x_i'\beta}/(1+e^{x_i'\beta})^{y_i}(1-(e^{x_i'\beta}/(1+e^{x_i'\beta}))^{1-y_i})$. We use the package stan_glm() for this implementation from the rstanarm package Goodrich et al. (2024).

To implement the frequentist logistic regression that we use to compare results, we used the glm() function. The original paper used backwards selection based on p values to obtain the predictors used in the logistic regression, but when we performed backwards selection, we ended up with fewer predictors. We decided to go with our selected model, which was

$$logit(\text{fire does not grow}) = \beta_0 + \beta_1 log(IA\_Size) + \beta_2 FWI$$

β_0 was 0.91657, β_1 was -0.35797 , and β_2 was -0.03263.

For the Bayesian logistic regression, we have our prior intercept set to normal N(0,2). Higher values of Fire Weather Index (FWI) indicate a greater potential for fire growth, so we expect that as FWI increases, the probability of not growing will decrease, and based on the range of values available for FWI, it makes sense to set the prior to normal (-.05,1). We also set the prior for log initial attack size to be normal(-.3,1), since we expect that an increased initial size would decrease the probability of a fire not growing.

3.1.1 Bayesian GAM

Generalized Additive Models (GAMs) are a type of model that allow both parametric and nonparametric relationships between a predictor and outcome. The form of the additive model is as follows.

$$E(Y) = \beta_0 + f_1(x_1) + f_2(x_2) + f_3(x_3) + f_4(x_4) + \ ...$$

The model can include splines, which smooth the data. The model can also be adjusted to be used for binary data using a link function, thus we can use it for classification of fire growth, and the GAM will allow for more flexibility than a traditional logistic regression (White et al., 2020). Each smoothed f(x) can be modeled as:

$$f(x) = \sum_{k=1}^{L} \beta_k b_k(x)$$

, where $b_k(x)$ is a fixed basis function and β_k are coefficients. To avoid overfitting, the smoothed f(x) is penalized with

$$\sum_{m=1}^{M} \lambda_m \beta^T S_m \beta$$

, where S_m is a matrix of the fixed parts of the penalty and λ_m controls the amount of penalty. This results in a penalized log-likelihood:

$$\hat{\beta} = argmax_{\beta}(l(\beta) - \sum_{m=1}^{M} \lambda_m \beta^T S_m \beta))$$

. If we take the exponential of this log liklihood to get the likelihood, it is in the form of Bayes Theorem.

$$L_p(\beta, \lambda) = L(\beta)exp(-\beta^T S_\lambda \beta)$$

$L_p(\beta, \lambda)$ is the posterior, $L(\beta)$ is $p(y|\lambda, \beta)$, and $exp(-\beta^T S_\lambda \beta)$ is the prior on β (Miller, 2025). The prior beliefs can be used to specify the smoothness of the spline functions. For the Bayesian GAM, we use the brms package (Bürkner, 2017). From this package, we use the get_prior() function to specify the priors, but with greater domain knowledge, we could specify more informative priors. The function uses student_t(3, 0, 2.5) for the intercept and flat priors for β.

The frequentist GAM that we use for comparison was fit with the gam() function, and followed the same model used in the paper: Growth $\sim$ Period + s(logIA_Size) + s(FWI) + Method + Fuel_type + Detection + s(logResp_time) + Month + s(logNumber_of_fire), with s() indicating the smoothed terms.

3.2 Bayesian AFT Model

To quantify the effect of fire management, fuel, and weather on fire size, the original paper used Cox PH and accelerated failure time (AFT) (weibull, lognormal, and loglogistic) models. Fire size at "being held" is the event of interest. Theoretically, the data could be censored because fires could still be changing in size at the time of the last measurement, but the data did not have any censoring. Although the Cox-proportional hazards method performed better than the AFT methods, we decided to use the AFT method that performed best in the original study, which was the Weibull model. The accelerated failure time model is a simpler choice for a Bayesian analysis since the Cox PH model does not require any assumptions about the baseline hazard, which makes computing the likelihood difficult.

The AFT model does not assume proportional hazards and assumes that the covariates have a multiplicative affect on survival time, and thus an additive affect on the log of the survival time. The covariates either speed up or slow down the time until death. The model takes the form:

$$log(T) = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + \sigma\epsilon = \mathbf{x}'\beta + \sigma\mathbf{Z}$$

The β are coefficients, the x_j are covariates, and σ is a scale parameter. Z is the random error. When exponentiated, the following model is obtained:

$$T = \exp(\beta_0 + \sigma Z)\exp\left(\sum_j \beta_j x_j\right) = T_0 \exp\left(\sum_j \beta_j x_j\right)$$

The transformation log(T) is in location-scale form and the survival function can be written as:

$$S(t) = Pr(T > t) = Pr(Z > (\frac{log(t) - \mu}{\sigma}))$$

$$S(t) = Pr(T > t) = S_0(\frac{log(t) - \mu}{\sigma})$$

$$= S_0^*(\frac{t}{exp(x'\beta)})^{1/\sigma}$$

and $S_0^* = S_0 \log(t)$. The covarites change the scale but not the location of the survival time distribution.

If t are survival times, or in this case, fire sizes, δ_i is a censoring indicator (0=censored), h is the hazard function, S is the survival function, and D is the data, then the likelihood function is:

$$\mathcal{L}(\sigma, \beta \mid D) = \prod_{i=1}^{n} f(t_i \mid \sigma, \beta)^{\delta_i} S(t_i \mid \sigma, \beta)^{1-\delta_i} = \prod_{i=1}^{n} \left[\frac{f(t_i \mid \sigma, \beta)}{S(t_i \mid \sigma, \beta)}\right]^{\delta_i} S(t_i \mid \sigma, \beta) = \prod_{i=1}^{n} h(t_i \mid \sigma, \beta)^{\delta_i} S(t_i \mid \sigma, \beta)]$$

The log likelihood is:

$$\ell(\sigma, \beta \mid \mathcal{D}) = \sum_{i=1}^{n} (\delta_i \log h(t_i \mid \sigma, \beta) + \log S(t_i \mid \sigma, \beta))$$

Assuming α is the shape and λ is the scale, the weibull AFT survival model has the form: $S(t|\sigma, \beta) = exp(-(t/x'\beta)^{1/\sigma})$. The hazard model is: $h(t \mid \sigma, \beta) = \left(\frac{\sigma^{-1}}{exp(x'\beta)}\right)\left(\frac{t}{exp(x'\beta)})\right)^{1/\sigma - 1}$.

For a Bayesian interpretation, we set priors on σ and β. The survival and hazard functions from the weibull AFT model can be used to get the posterior below using the previously described likelihood function (Ashraf-Ul-Alam and Ali Khan, 2021).

$$f(\sigma, \beta \mid t, X) \propto \mathcal{L}(\sigma, \beta \mid \mathcal{D}) \times p(\sigma) \times p(\beta)$$

We do not have any strong beliefs about the parameters in the survival model, so we set the scale prior to be half-Cauchy (0,25) and the β prior to be normal (0,10).

4. Computation

For the classification methods, we use packages in R to perform the MCMC algorithms that obtain the posterior distributions. However, for the survival analysis, we use stan. Stan uses the No-U-Turn sampling algorithm. This is similar to the Metropolis Hastings algorithm, which allows us to sample from a posterior distribution using only the un-normalized posterior (the product of the likelihood and prior). For Metropolis-Hastings, we start with an initial values for the parameter vector ω^0, propose a candidate from a proposal distribution (ω^*) with a density that depends on only the previous parameter values (ω^{m-1}), then get the ratio of relative posterior density based on ω^* and the posterior density based on ω^{m-1}. If this ratio is greater than 1, then the sample is accepted, but if the ratio is less than 1, then the acceptance of the sample is determined by a Uniform (0,1) distribution. This causes a shift towards the normalized posterior distribution by taking a "random walk" from the previous distribution. The size of the step depends on the proposal variance. Tuning the proposal variance is important. If the proposal variance is too large, the sample is more likely to be from a lower-density area, which would have a low acceptance probability and the parameter vectors would not change, leading to a high correlation betwen draws. If the proposal variance is too small, then it could also lead to high correlation between draws.

The No-U-Turn algorithm is a Hamiltonian Monte Carlo algorithm, which uses first-order gradients of the log-posterior to take steps, rather than the random walk used by MH. At the start of the algorithm, an initial momentum variable r^0 is sampled from a standard normal distribution. For each sample, a "leapfrog" step is performed L times. For each of the L steps, r and the parameter vector, ω, are updated based on the the log-posterior gradient. After the L steps, the new ω is either accepted or rejected. This algorithm is sensitive to the number of steps, L, and the step size specified by the user. The No-U-Turn sampler simplifies this, so that the user does not have to specify the number of steps (Hoffman et al., 2014).

In stan, we specify the log hazard function, the log survival function, and from these we specify the log likelihood of the survival function. These functions are shown in the *Methodology* section. We also specify the prior distributions of our parameters.

5. Data Analysis

After exploring the data, we feel it is important to note that the size of a fire at "being held" follows a very skewed distribution (Table 1). Further examination of the data reveals that there are several fires of this large size, and that it is not just a single outlier.

Quantile	**Size (x)**
0%	0.01
25%	0.05
50%	0.20
75%	1.50
100%	27486.50

Table 1: Quantiles of Fire Size When Being Held

5.1 Classification Results

We split the data 70/30 into a training and test set for the classification analysis. We ran the models with 4 chains and 10000 iterations with 5000 of those being warmup. The posterior means produced on the training set are 0.9235 for the intercept, -.3603 for log initial attack size, and -0.033 for FWI. These are very similar to the coefficients produced by the frequentist logsitic regresssion (which also used the training set). The 95% credible intervals are (0.47414629,1.36877295), (-0.45647561,-0.26729146), and (-0.05554765, -0.01071775) for the intercept, log initial attack size, and FWI respectively. The credible interval tells us the that there is a 95% probability that the true parameter value is within the interval. This suggests the model is fairly certain about FWI because it has a small credible interval. However, both credible intervals do not contain zero, suggesting that both initial attack size and FWI are associated with an increased probability of fire growth. Trace plots of the parameters show that the algorithm successfully converged.

For the GAM we used 4000 iterations with 1000 of those being warmup.

AUC_logistic	AUC_logistic_Bayes	AUC_GAM	AUC_GAM_Bayes
0.7628	0.763	0.7579	0.7675

Table 2: Comparison of AUC values

In Table 2 we can see the Bayesian versions performed better in terms of AUC than the frequentist versions. The highest AUC was achieved by the Bayesian GAM, although the differences between all the methods are fairly small and would likely change upon using a different random seed.

With a threshold of .5, the model with the highest accuracy was the frequentist GAM (Table 3). The logistic regressions had the same accuracy. As with the AUC, these differences are slight.

	logistic	logistic_Bayes	GAM	GAM_Bayes
1	0.7453	0.7453	0.7669	0.764

Table 3: Model Accuracy

5.2 Survival Analysis Results

For survival analysis, we used a Weibull AFT model. Size at being held is the response, and the data is left truncated at the initial attack size. First, we filtered to only fires that grew. We included covariates for FWI, method, and fuel type in the frequentist analysis. For the Bayesian AFTs we had one model with the same covariates, and then another where we added an interaction between method and fire load, since the original analysis wanted to include it, but the sample size wasn't large enough. For each model, we plot the cumulative hazard, which is the cumulative risk of "being held" at a certain fire size. Table 4 shows the coefficients from the models not including the interaction term between fire management method and fire load. Overall, they are fairly similar, especially FWI. Most coefficients are in the same direction.

Model Term	**Freq**	**Bayes**
scale	-0.9540673	-0.4289579
FWI	0.1679640	0.1596040
factor(Method)Air Tanker	1.6895139	1.7060326
factor(Method)Ground trained	-1.2227525	-0.3886482
factor(Method)Helitanker	1.5705466	1.5759933
factor(Method)Other ground	-0.1288762	0.0662635
factor(Fuel_type)C2	0.5657822	0.5435710
factor(Fuel_type)M2	-0.2706039	0.1759365
factor(Fuel_type)Other	-1.2307478	-1.1540804

Table 4: Posterior means for model coefficients (Freq vs Bayes)

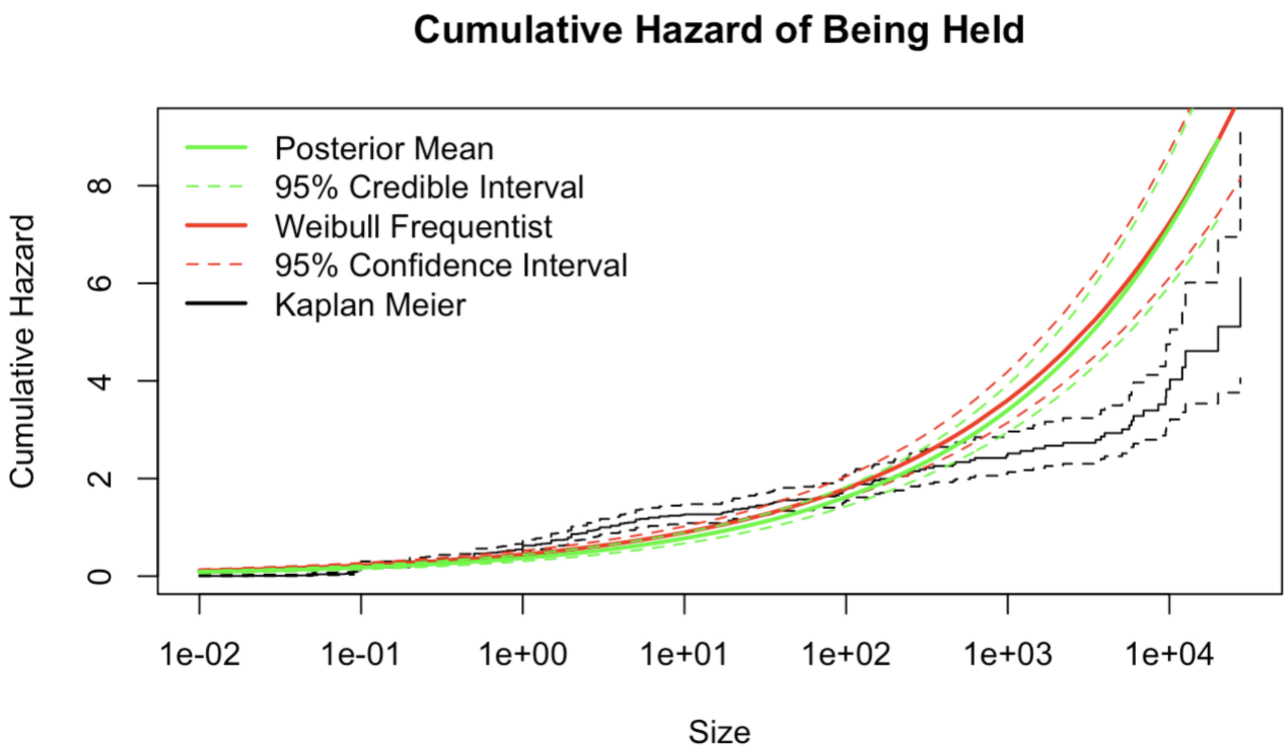

Figure 1: Cumulative Hazard Curves

The cumulative hazard plot (Figure 1) shows that the Cumulative Hazard curves were similar across both methods, but neither matched the nonparametric curve very well, especially for the larger sizes.

Table 5 shows the posterior means and 95% credible intervals from the AFT model that included an interaction between method and fire load. All of the credible intervals for the interactions include zero, suggesting that the interaction between number of fires and method of fire control is not important for predicting fire size. Fire weather index (FWI) appears to be an important predictor of the probability of fire growth. For every one unit increase in FWI for a fire that was controlled by air (the reference fire control method), the fire size increases by $\exp(0.16) = 1.17$ or 17%.

Figure 2 shows the stratified survival curves based on the method of controlling the fire. Based on the plot, it appears that helitanker, air, and air tanker controlled fires have a higher probability of growing than ground-controlled fires. For this plot, all other variables were set to their mean value.

Figure 3 shows that the algorithm converged well since the draws converged to mean parameter values.

Model Term	Posterior Mean	Lower CI	Upper CI
(scale)	-0.24663287	-2.08421539	1.74407005
FWI	0.16340507	0.11259254	0.21263244
MethodAir Tanker	1.58393070	-0.19557427	3.54451543
MethodGround trained	-1.76090558	-6.59548634	4.49027681
MethodHelitanker	0.74428509	-2.09385507	4.10038908
MethodOther ground	0.45396191	-1.24454165	2.32747098
Fuel_typeC2	0.50871196	-1.29780731	2.15409554
Fuel_typeM2	0.23977720	-2.15200924	2.52724306
Fuel_typeOther	-1.19162284	-3.56237279	1.09550435
number_of_fire	-0.01818649	-0.06186052	0.02845639
MethodAir Tanker:number_of_fire	0.01151927	-0.10376356	0.12762955
MethodGround trained:number_of_fire	0.17536927	-0.22896917	0.69198374
MethodHelitanker:number_of_fire	0.08813128	-0.15464188	0.32964083
MethodOther ground:number_of_fire	-0.01471594	-0.09812534	0.06927308

Table 5: Posterior means and 95% credible intervals for model coefficients

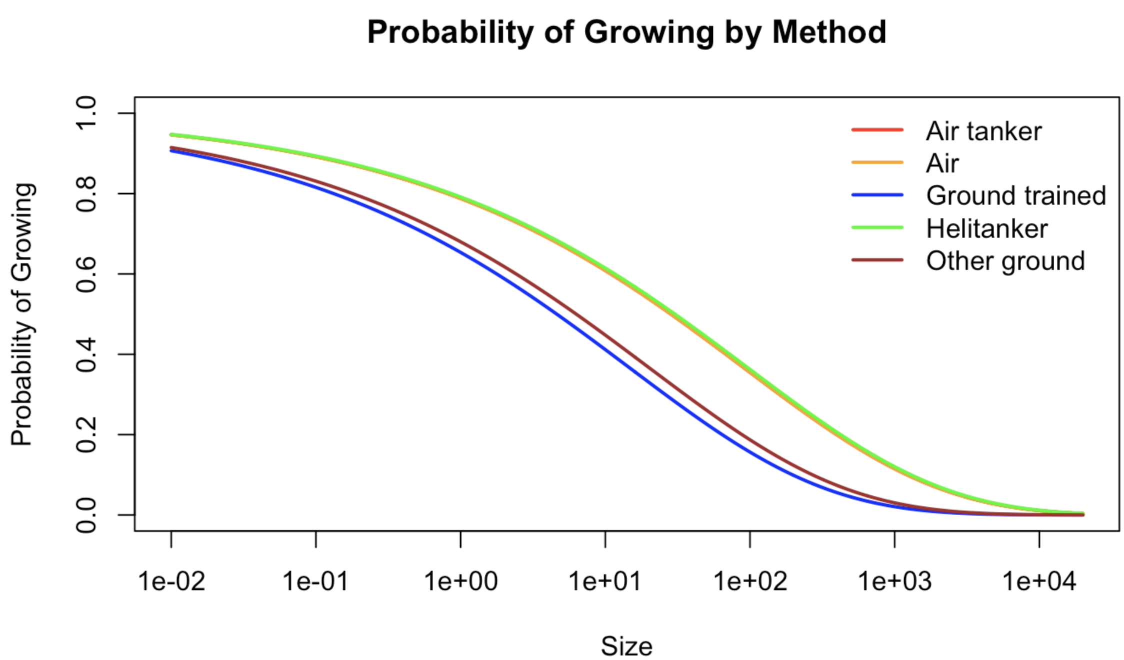

Figure 2: Survival Curves by Method for the Model with Interaction between fire load and method

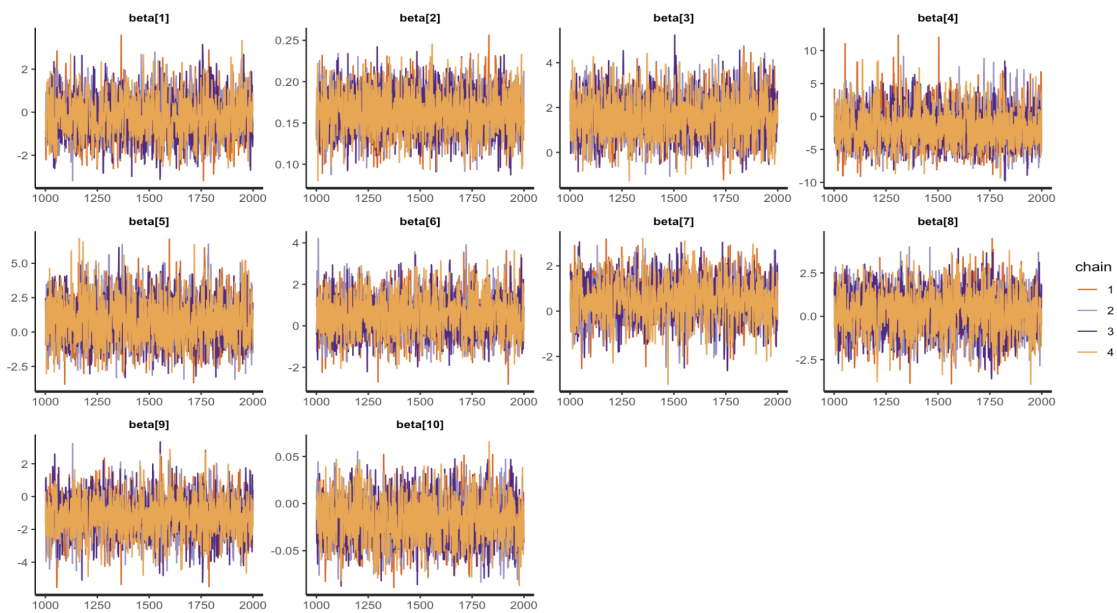

Figure 3: Traceplots-First 10 Parameters

6. Discussion and Conclusion

Fires appear to have slight differences in growth based on control method, but posterior means do not include zero, so these differences are insignificant. However, we did not have much prior knowledge about the parameters in the survival model, so perhaps with more domain knowledge and results from previous experiments, these results would change.

Overall, our Bayesian results are similar to the results described in the original paper. However, using Bayesian methods is still useful for this research area since it can incorporate prior historical knowledge, which should be especially useful for fires since data are limited. It also allows for improved quantification of uncertainty since we obtain posterior distributions of parameters. Although not used in this paper, models can also be averaged together into single models when appropriate.

6.1 Code Availability

The code for this paper is available at https://github.com/rachel-yost/Bayesian-Final-Proj

References

M Cecilia Arienti, Steven G Cumming, and Stan Boutin. Empirical models of forest fire initial attack success probabilities: the effects of fuels, anthropogenic linear features, fire weather, and management. *Canadian Journal of Forest Research*, 36(12):3155–3166, 2006.

Md Ashraf-Ul-Alam and Athar Ali Khan. Comparison of accelerated failure time models: a bayesian study on head and neck cancer data. *Journal of Statistics Applications & Probability*, 10(3):715–738, 2021.

David MJS Bowman, Grant J Williamson, John T Abatzoglou, Crystal A Kolden, Mark A Cochrane, and Alistair MS Smith. Human exposure and sensitivity to globally extreme wildfire events. *Nature ecology & evolution*, 1(3):0058, 2017.

Paul-Christian Bürkner. brms: An r package for bayesian multilevel models using stan. *Journal of statistical software*, 80:1–28, 2017.

SG Cumming. A parametric model of the fire-size distribution. *Canadian Journal of Forest Research*, 31(8):1297–1303, 2001.

Calum X Cunningham, Grant J Williamson, and David MJS Bowman. Increasing frequency and intensity of the most extreme wildfires on earth. *Nature ecology & evolution*, 8(8): 1420–1425, 2024.

Ben Goodrich, Jonah Gabry, Imad Ali, and Samuel Brilleman. *rstanarm: Bayesian Applied Regression Modeling via Stan*, 2024. URL `https://mc-stan.org/rstanarm`. R package version 2.32.1.

Matthew D Hoffman, Andrew Gelman, et al. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.

Katherine Kilkenny, William Frishman, and Joseph S Alpert. Los angeles wildfires: Getting to the heart of it. *The American Journal of Medicine*, 2025.

David L Martell and Hua Sun. The impact of fire suppression, vegetation, and weather on the area burned by lightning-caused forest fires in ontario. *Canadian Journal of Forest Research*, 38(6):1547–1563, 2008.

Amy A Morin, Alisha Albert-Green, Douglas G Woolford, and David L Martell. The use of survival analysis methods to model the control time of forest fires in ontario, canada. *International Journal of Wildland Fire*, 24(7):964–973, 2015.

David Strauss, Larry Bednar, and Romain Mees. Do one percent of the forest fires cause ninety-nine percent of the damage? *Forest science*, 35(2):319–328, 1989.

Pier-Olivier Tremblay, Thierry Duchesne, and Steven G Cumming. Survival analysis and classification methods for forest fire size. *PLoS one*, 13(1):e0189860, 2018.

Rens Van De Schoot, Joris J Broere, Koen H Perryck, Mariëlle Zondervan-Zwijnenburg, and Nancy E Van Loey. Analyzing small data sets using bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European journal of psychotraumatology*, 6(1):25216, 2015.