7-14-2014 12:00 AM

A Spatial Analysis of Forest Fire Survival and a Marked Cluster Process for Simulating Fire Load

Amy A. Morin, *The University of Western Ontario*

A SPATIAL ANALYSIS OF FOREST FIRE SURVIVAL AND A MARKED CLUSTER PROCESS FOR SIMULATING FIRE LOAD

(Thesis format: Monograph)

by

Amy <u>Morin</u>

Graduate Program in Statistical and Actuarial Sciences

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

The duration of a forest fire depends on many factors, such as weather, fuel type and fuel moisture, as well as fire management strategies. Understanding how these impact the duration of a fire can lead to more effective suppression efforts as this information can be incorporated into decision support systems used by fire management agencies to help allocate suppression resources. This thesis presents a thorough survival analysis of lightning and people-caused fires in the Intensive fire management zone of Ontario, Canada from 1989 through 2004. The analysis is then extended to investigate spatial patterns across this region using proportional hazards Gaussian shared frailty models. The resulting posterior estimates suggest spatial patterns across this zone. A fire load model is also developed by coupling a fire occurrence model with a survival model and is explored via simulation. Marked cluster processes were found to nicely capture the overall fire load trend over the fire season.

Co-Authorship Statement

Portions of this thesis are used in a paper which will be submitted to the International Journal of Wildland Fire with Alisha Albert-Green, Dr. Douglas Woolford and Dr. David Martell, who provided editing and comments on these sections. A second paper is also planned for submission to Environmetrics; co-authors of this paper will include Dr. Douglas Woolford and Dr. David Martell.

Acknowledgements

I am grateful to the faculty, staff and students in the Department of Statistical and Actuarial Sciences at Western for a warm and welcoming learning environment. I would also like to thank my committee members: Dr. Braun, Dr. Kulperger, Dr. Hong and Dr. Zitikis, for dedicating time to my thesis examination and for providing helpful revisions.

Helpful conversations and revisions from our collaborators Alisha Albert-Green and Dr. David Martell are also much appreciated. In addition, Dr. Charmaine Dean is thanked for welcoming me to the lab, and for providing computing equipment and funding.

Dr. Douglas Woolford is particularly recognized for devoted supervision and patience throughout this degree. He has gone out of his way to help me get to this point and his guidance continues to lead me to new opportunities.

I wish to sincerely thank my office mates: Alisha, Erin, Mark, and Phil, for always being open to share insightful advice and knowledge. I will forever cherish my time at Western with these friends.

I am eternally grateful to my family, especially my parents, Jean-Pierre and Gisèle, for lovingly and unselfishly supporting me throughout my life. Lastly, I owe my deepest gratitude to John for always being there and believing in me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The duration of a forest or wildland fire influences the area burned in Canadian forests. To reduce the area burned, forest fires which are difficult to extinguish are of particular interest as they pose a significant threat to forest resources across much of the Province of Ontario, Canada. A better understanding of the variables which contribute to longer lasting forest fires could be incorporated into decision support systems used by fire management agencies to prioritize fires and allocate suppression resources.

In Canada, forest fires are ignited by lightning or by people. The ignition risk of lightning-caused fires depends heavily on the moisture within the top layer of the forest floor, which is often referred to as the duff layer, where a fire can smoulder until the surface of the forest floor becomes dry enough to sustain its spread (Wotton and Martell, 2005). By contrast, the ignition of people-caused fires is most dependent on the moisture of the fine litter fuels on the surface of the forest floor (Wotton, 2009). Consequently, lightning and people-caused fires have been modelled separately in the literature (Wotton and Martell, 2005; Wotton et al., 2003). The applications in this paper model lightning-caused fires and people-caused fires separately.

In this thesis, we use survival analysis methods to explore how fire weather variables affect the duration of a forest fire. For the purpose of this study, we define the *duration* of a forest fire to be the time interval from the beginning of initial attack action to the time that the fire is

declared as being under control, which we will hereafter refer to as the survival time. Survival analysis is often employed to investigate time to some significant event (Lawless, 2003), often in the presence of censoring and/or truncation, commonly in biostatistics and engineering. In the biostatistics literature, the event of interest is often death or the recurrence of a disease (e.g., Fleming and Lin, 2000), while time to product failure is often the event of interest in engineering applications (e.g., Tsai et al., 2003). In our application, the event of interest is the time at which a fire is declared under control.

In addition to the traditional applications, survival analysis has been applied to the lifetimes of living organisms, including animals (Johnson et al., 2004) and trees (Ritchie et al., 2007), extensively in ecology. However, the application of survival analysis to wildland fires has been focused on estimating fire frequency and fire cycles involving the time required to burn an area equal in size to the studied area (Larsen, 1997; Senici et al., 2010) by modelling the time since a fire occurred at specific points on the landscape. In this thesis, we investigate the associations between fire weather variables and the distribution of the survival time of a forest fire by the means of a thorough survival analysis. We also explore the data, looking for spatial patterns in the duration of forest fires. The objective of this research is to provide forest fire managers with an alternative fire weather modelling technique which could be incorporated in their strategic, tactical and operational planning systems.

To forest fire managers, the *fire load* represents the magnitude of the suppression efforts and resources corresponding to fires that occur in a specific area and over a specified time interval (Martell, 2007). In this thesis, we will specify the fire load on a given day of the fire season to be the number of fires burning on the landscape. A simulation of the fire load is performed by coupling a fire occurrence model with one of our survival models for a single region. The fire occurrence is modelled using a non-homogeneous Poisson process and special forms of cluster point processes (Cox and Isham, 1980). The survival time is then coupled to one of

the stochastic cluster processes as a mark. These types of models have been used extensively to model ecological occurrences including rainfall (Rodriguez-Iturbe et al., 1987) and earthquakes (Ogata, 1988).

The remainder of this thesis is organized as follows. A description of the study area, data and Ontario's fire management system is given in Chapter 2. Exploratory data analyses are also performed. In Chapter 3, we present definitions, statistical concepts and methods which will be applied in the study. Simple parametric log-location-scale models are then fit to the data in Chapter 4, along with survival models that incorporate the effects of other covariates. In addition, spatial patterns across our study region are analyzed using proportional hazards Gaussian shared frailty models and illustrated using choropleth maps. We also present the development of a stochastic fire load model via simulation. We conclude the study with a discussion of our results and future work in Chapter 5.

Chapter 2

The Data and Study Area

The Aviation, Forest Fire and Emergency Services (AFFES) Branch of the Ontario Ministry of Natural Resources (OMNR) is responsible for forest fire management on Crown land in the fire region of the Province of Ontario. Prior to 2004, this fire region (coloured area) was partitioned into the three fire management zones shown in Figure 2.1, each of which received different levels of protection. The strategy in the Intensive zone was to suppress all fires as soon as suppression resources were available, in the Measured zone fires were initially attacked and re-evaluated if extended attack was needed for containment, and in the Extensive zone fires were left to burn so long as they did not threaten a community. Recently, the number of zones and their corresponding management strategies have been modified (see OMNR, 2004). Consequently, we focus our analysis on historical records of fires prior to this change in the provincial fire management strategy.

2.1 Forest Fire Data

We studied the lifetimes of 18,183 forest fires in the Intensive zone of the Province of Ontario, for the period 1989 through 2004, using data provided by the OMNR. That fire archive includes many fire attributes including the dates and times at which each fire was reported, when suppression action began and when the fire was declared under control. It also includes other
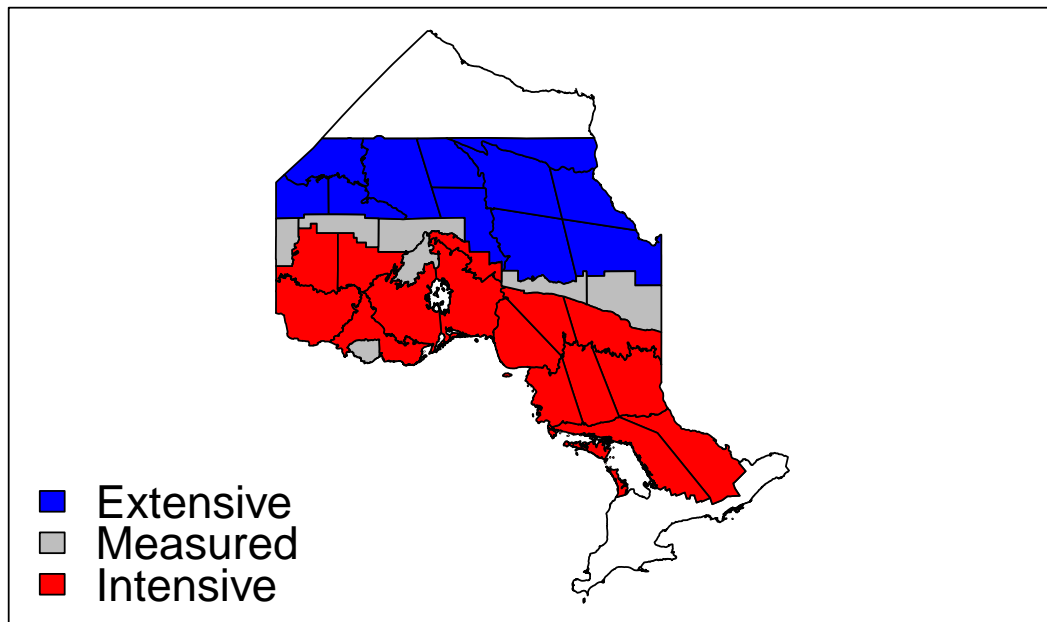
Figure 2.1: Ontario's fire management zones prior to 2004.

information such as the weather, vegetation or fuel type and fuel moisture associated with each fire. Among the fires we studied, 43% were caused by lightning and 57% were caused by people.

Martell (2007) described how fire management agencies track a forest fire's progress through several distinct phases. The life cycle of a fire begins with ignition, sometime later it is detected and later reported to the fire management agency. Once the fire has been reported, the duty officer or dispatcher prioritizes it, places it in the initial attack (IA) queue and dispatches initial attack resources (e.g., airtankers and fire fighters) to begin initial attack action. A fire is declared to be in a state of being held when it is no longer spreading but may possibly resume spreading, and subsequently deemed under control when adequate control lines have been established around the fire's perimeter. The fire crews then work on extinguishing the fire from the perimeter inwards until the fire crews have sufficiently extinguished any smoldering fuels,

at which point the fire is then declared as out. Figure 2.2 displays the importance of effective suppression efforts in terms of area burned over a typical fire's lifetime (Parks, 1964). From this figure, it is clear that understanding survival time is important, because area burned is proportional to this quantity.
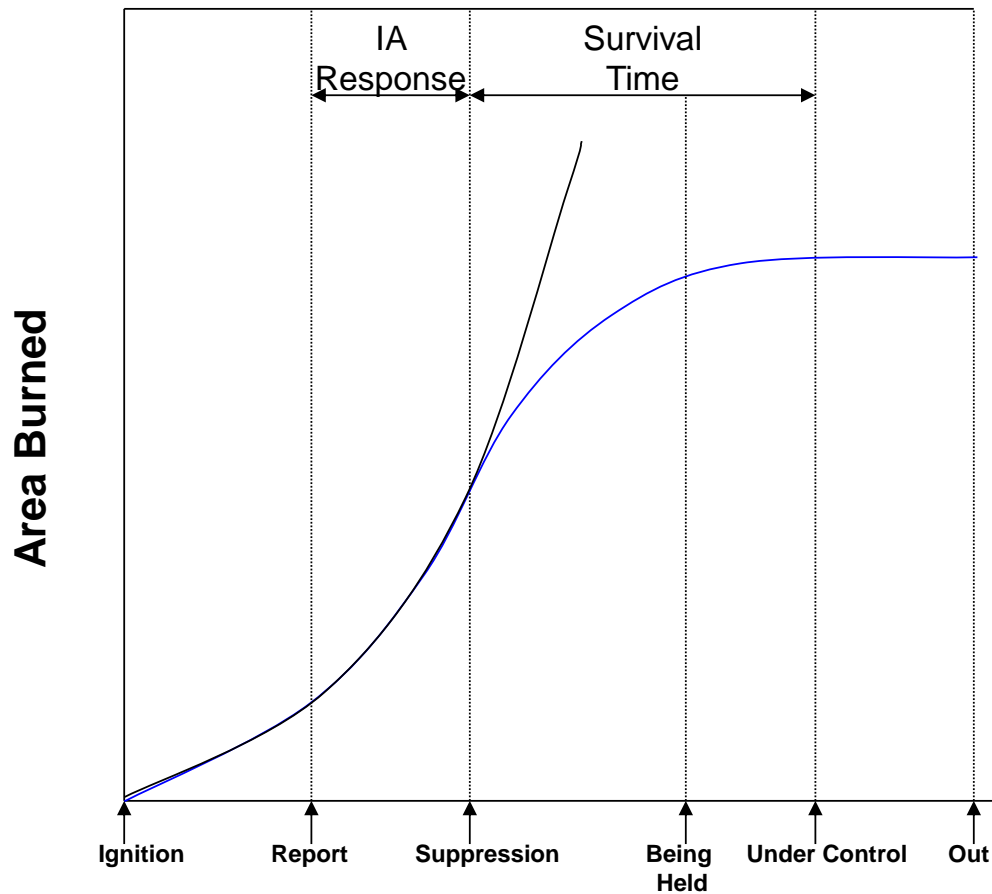
Figure 2.2: The typical timeline of the area burned by a fire through its progression phases with (blue) and without (black) suppression efforts.

As mentioned previously in this thesis, we define a fire's survival time as the time between when initial attack begins and when the fire is declared under control. These times are used because the initial suppression time is known (the start date is less accurate since it is often

estimated) and the amount of time that it takes to bring the fire under control is of most interest
and use to fire management agencies. This survival time is measured in hours, where decimal
places represent minutes, for our analysis. Fires with negative survival times (870 fires) that
resulted from data coding errors are removed, as well as those where the initial attack and
under control dates and times are the same (133 fires). Apart from simple data entry error,
negative survival times might have occurred because the time at which a fire was declared as
under control was not specified along with the date. In this situation, the time is recorded as
midnight which leads to problems when working with time differences between date-stamps
of short fires. Fires which have survival times of length 0 may be caused by similar mistakes
or by a fire crew arriving at a fire and deciding that suppression is not needed due to the low
intensity of the fire and expected weather (rain). These fires are removed from the dataset as in
this study we are only interested in fires which need to be suppressed.

The OMNR considers an initial attack to have been successful if a fire is brought to a state
of being held by noon on the day following the day the fire was reported or if final size is less
than 4 hectares. Roughly 97.5% of fires in Ontario are successfully contained by the initial
attack force. For this reason, the 3 fires with survival times which are greater than 1,000 hours
(just under 42 days) are removed. Some summary statistics for the data we analyze are dis-
played in Table 2.1 and histograms of the survival time of lightning and people-caused fires
which are declared under control within 2 days are displayed in Figure 2.3.

Statistic	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Lightning	0.02	1.25	3.33	14.66	16.08	745.40
People	0.02	0.48	1.17	7.65	4.00	570.00

Table 2.1: Descriptive statistics of the survival time (in hours) of fires from the Intensive zone,
by ignition cause.

Lightning–caused Fires

People–caused Fires
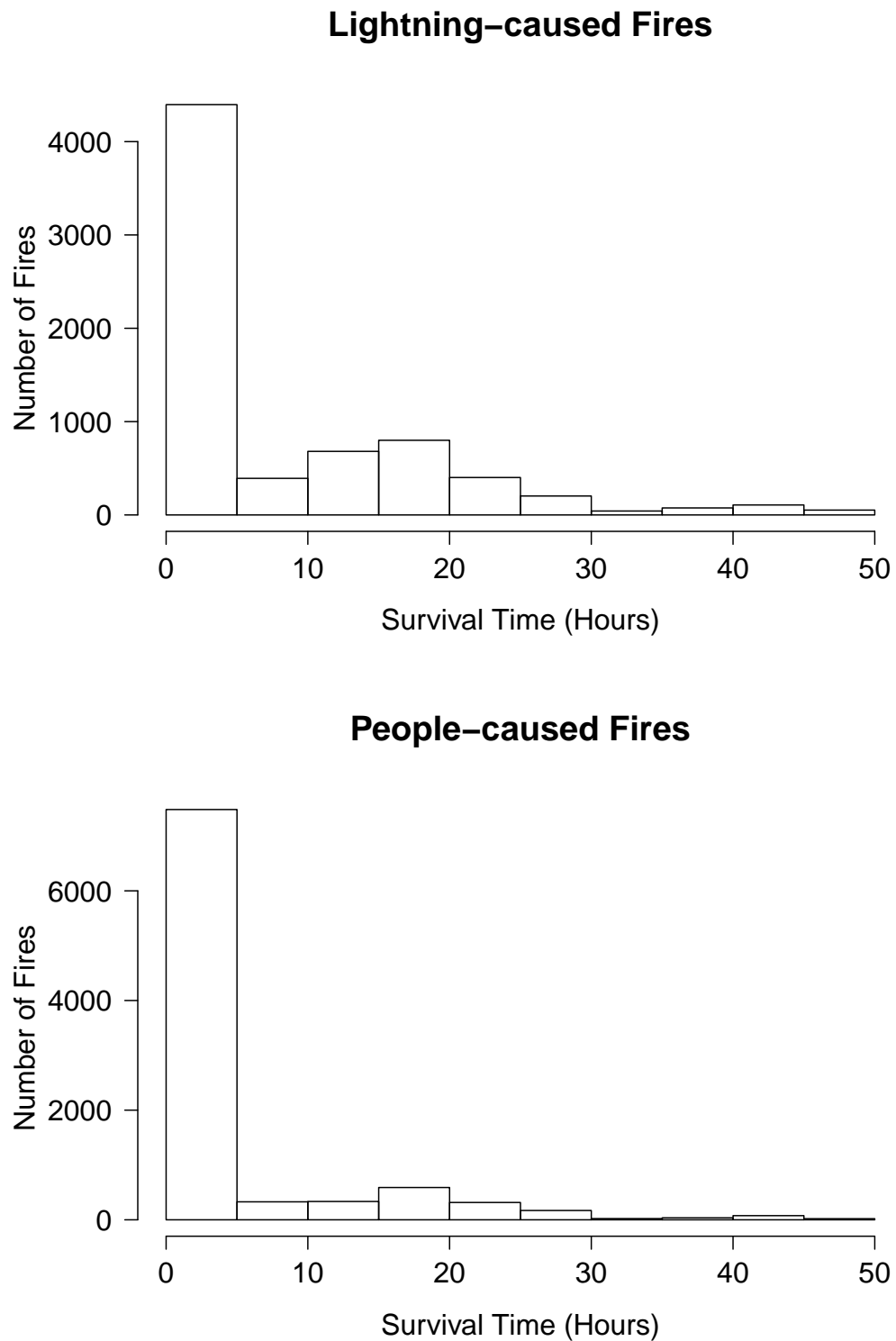
Figure 2.3: Histograms of the survival time, in hours, of lightning (top panel) and people-caused fires (bottom panel) which are declared under control within 2 days.

2.2 Fire Management and Weather Variables

The *initial attack response* is the time between when a fire is reported and the time that initial attack action begins. Figure 2.4 displays the histograms of the response time (in hours) for lightning and people-caused fires. Investigation of the effect of this covariate on the survival time of forest fires is motivated by the potential impact it may have on the final size of fires. The size (in hectares) of a fire at the start of initial attack is also considered.

Canadian forest fire managers use the Canadian Forest Fire Danger rating System (CFFDRS) to assess the impact of weather on fire occurrence and behaviour processes. We will be using 5 of the codes and indices from the CFFDRS that represent the cumulative impact of weather on the moisture content of different components of a forest fuel complex and potential fire behaviour. A thorough overview of the CFFDRS for the purpose of modelling is provided by Wotton (2009). The Fine Fuel Moisture Code (FFMC) is a numerical rating of the moisture in the small, readily consumed, fine fuels on the forest floor. It increases with increasing dryness varying from saturation to a completely dry surface layer. The Initial Spread Index (ISI) is a rating of the potential rate of spread of a fire and is based upon the wind speed and the FFMC. The larger the ISI, the greater the fire spread rate potential. The Duff Moisture Code (DMC) is a measure of the moisture content of the top 7 cm of the forest floor where litter starts to decay and ranges from 0 (full saturation) from where it can increase with no upper bound. The Drought Code (DC) is a measure of moisture content of deep layers of the forest floor and of dead woody debris approximately 18 cm thick and accounts for long-term drying. The higher the DC, the drier the deep organic fuel. The Build-Up Index (BUI) is a weighted mean of DMC and DC and serves as a measure of the potential fuel available for consumption on and in the forest floor. It is also a useful indicator of difficulty in extinguishing smouldering fires. These fire weather variables have been historically used in forest fire occurrence prediction models; notable examples include the logistic regression models of Woolford et al. (2011) and Wotton and Martell (2005), as well as the Poisson regression models of Wotton et al. (2003).
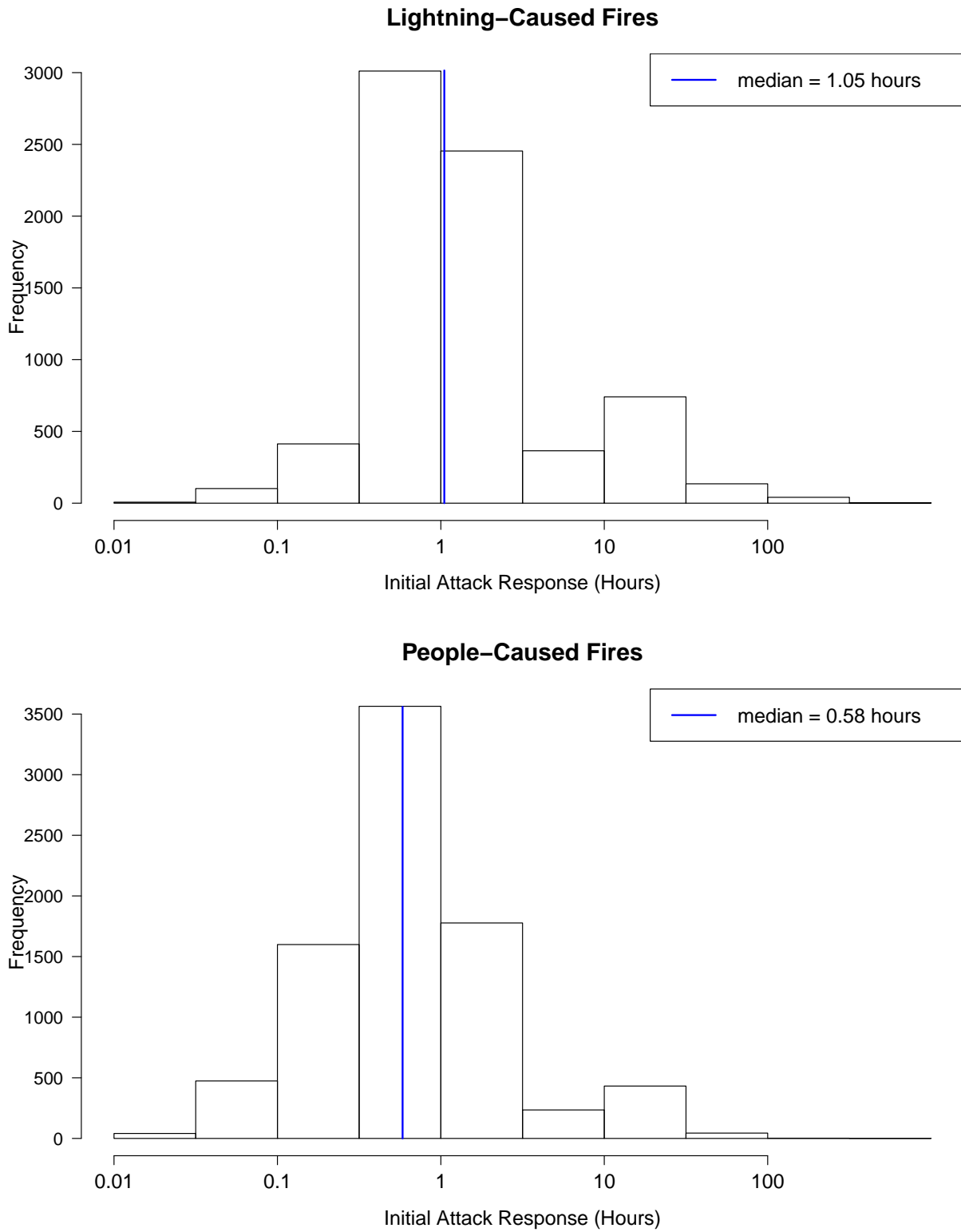
Figure 2.4: Histograms of the initial attack response time, in hours, of lightning (top panel) and people-caused fires (bottom panel).

Chapter 3

Methodology

3.1 Survival Analysis

3.1.1 The Survival Time, T

The survival time, T, is a continuous and non-negative random variable which represents the time to an event. In the analysis of lifetimes, one is most interested in the probability of survival, $S(t) = P(T > t)$, and the hazard rate, $\lambda(t) = \frac{f(t)}{S(t)}$, where $f(t)$ is the probability density function of T. In the context of our research, we analyze the probability distribution for the time until a fire is declared under control, given that it is being suppressed. Here, the hazard or failure rate can be viewed as the probability that a fire would "die" at time t, conditional on its survival to that point. In this sense, it can be viewed as the instantaneous rate that the fire becomes under control. The hazard rate is important because the survival function can be expressed as a function of this rate as

$$S(t) = \exp\left\{-\int_0^t \lambda(u)du\right\}.$$

The remainder of Section 3.1 reviews common non-parametric, parametric and semi-parametric methods to estimate survival probabilities or hazard rates. The proportional hazards shared

frailty model and inference procedures are also introduced.

3.1.2   The Kaplan-Meier Estimator

The Kaplan-Meier (KM) estimator approximates the underlying continuous lifetime distribution using a discrete distribution (Lawless, 2003). It is a non-parametric (i.e., no assumptions about the underlying distribution are made) technique. Estimates are made at each of the observed event times. The KM estimator of the survival function is defined as

$$\hat{S}^{KM}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{Y_j}\right), \quad t \geq 0,$$

where t_j denotes the observed event times, d_j is the number of events which occurred at time t_j and Y_j is the number of subjects at risk at time t_j, $j = 1, ..., n$. Note that d_j/Y_j is a non-parametric estimator of the hazard function since d_j/n estimates $f(t_j)$ and Y_j/n estimates $S(t_j)$.

3.1.3   Log-Location-Scale Models

Log-location-scale models are a class of parametric models commonly used to model lifetime data. The lifetime random variable, T, has a log-location-scale model if its survival function can be written in the form

$$S(t) = S_0 \left(\frac{\log(t) - u}{b}\right),$$

where S_0 is a fully specified survival function with positive support and $u \in \mathbb{R}$ and $b > 0$ are the location and scale parameters which shift and affect the spread of the distribution, respectively. The three most commonly used log-location-scale models are the Weibull, lognormal, and loglogistic. They are obtained when $S_0(t)$ is the extreme value, normal, or logistic survival function, respectively. The location and scale parameters are estimated from the data using maximum likelihood. First, we define the likelihood function, L, as a function of the unknown

parameters for a given parametric model. For complete data, the likelihood function of n observations $x_1, ..., x_n$ is

$$L \equiv L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}),$$

where f is the density function of the parametric model with parameter vector $\boldsymbol{\theta}$. The maximum likelihood estimates of the parameters are obtained by maximizing the log-likelihood function, $l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$.

Non-parametric survival estimates can be used as a method of comparing models (Lawless, 2003). Using this approach, the appropriateness of each log-location-scale parametric model can be determined by examining how closely the fitted survival probabilities follow the KM estimates. Non-parametric models are useful comparison tools as they are true to the data and only make the assumption that the data are independent; they do not impose a distribution upon the data.

The KM estimator and log-location-scale methods produce simpler models in the sense that they don't incorporate effects of other independent variables or covariates. The *accelerated failure time* (AFT) model is a parametric regression model which takes into account the effect of covariates on the survival function. It is a log-location-scale distribution of the form

$$S(t \mid \mathbf{x}) = S_0\left(\frac{\log t - u(\mathbf{x})}{b}\right),$$

where the location parameter is dependent on the covariate vector, $\mathbf{x}$, and is often assumed to be a linear function of p covariates. That is, $u(\mathbf{x}) = \sum_{i=1}^{p} \beta_i x_i$, where the β's are parameters to be estimated. The AFT model makes the assumption that the effect of covariates is equivalent to altering the rate at which time passes. If the linear function of the covariates, $u(\mathbf{x})$, is greater than 0 then the covariate vector is decelerating the time scale, or equivalently, increasing the

survival probability across time. Similarly, if $u(\mathbf{x}) < 0$ then the covariate vector is accelerating the time scale or decreasing the survival probability across time.

3.1.4   The Cox Proportional Hazards Model

The Cox proportional hazards (PH) model is a widely used semi-parametric method for survival analysis that can incorporate covariate effects. The hazard rate is modelled as

$$\lambda_j(t \mid \mathbf{x}_j) = \lambda_0(t)e^{x'_j\beta}, \, j = 1, ..., n$$

where j indexes the observations, λ_0 is the baseline hazard rate, $\mathbf{x}_j$ is the covariate vector and β is the associated parameter vector of length p. The baseline hazard rate can be viewed as the hazard function when all covariates are 0. The covariates then have a multiplicative effect on this baseline. In this application of the Cox PH model, our main interest is in quantifying the effects of the covariates. The Cox PH model is commonly built using forward selection of covariates, details of which will be provided in Section 3.1.6. The effect of the covariates, $e^{x'_j\beta}$, is parametrically estimated without considering the baseline by maximizing Cox (1972, 1975)'s partial likelihood,

$$L(\beta) = \prod_{j=1}^{n} \frac{e^{\mathbf{x}'_j\beta}}{\sum_{k\epsilon R_j} e^{\mathbf{x}'_k\beta}},$$

where R_j is the set of observations which are at risk for an event (fires not under control) at the time of the jth observation. This likelihood is partial as it involves only the parameters of interest and the baseline hazard is considered a nuisance function which subsequently may be non-parametrically estimated using estimators given by Breslow (1972) or Efron (1977), for example. If parameterized in terms of the survival function, the Cox PH model takes the form $S_j(t \mid \mathbf{x}_j) = S_0(t)^{e^{x'_j\beta}}$.

3.1.5   The Proportional Hazards Shared Frailty Model

The purpose of frailty models is to describe the excess risk, or frailty, for distinct categories. The main idea is that there are greater correlations among data within these categories or clusters. In the application of the Cox Proportional Hazards model, these within cluster dependencies cause non-proportionality of the hazards and loss of precision of the estimated parameters. This problem is remedied by implementing a multivariable mixed-effects survival model, namely the shared frailty model. The shared frailty model is such that each observation belongs to only one of the distinct categories, all observations within a category share a common frailty, and the frailties from different categories are independent (Therneau et al., 2003). The hazard rate from the proportional hazards shared frailty model is

$$\lambda_{ij}(t) = \lambda_0(t)e^{\mathbf{x}'_{ij}\beta + \omega_i}, \quad j = 1, ..., n_i, \ \ i = 1, ..., s$$

where j indexes the observations in category i, $\mathbf{x}_{ij}$ is the covariate vector of length p with associated parameter vector β and ω_i is the random effect term of category i. The distribution and mean of the random effects must be specified, while the variance is left unspecified. The choice of the distribution of the random effects is based on the dependence structure present in the data being modeled. In practice, the most common choices include the gamma distribution with mean 1 and the normal distribution with mean 0, the latter of which will be applied in this thesis as it allows flexibility in modelling data with various dependence structures. We will also demonstrate that this normal frailty term is appropriate for our data later in this thesis. In the case of shared frailty models, where each observation belongs to only one category and the observations within a category have a common random effect term, $\sum\limits_{i=1}^{s} n_i = n$.

Penalized regression models are used to estimate Cox PH models with frailty terms (Ripatti and Palmgren, 2000). When fitting such models, the random effects are added to survival models as constrained, unobserved continuous covariates which are assumed to follow a probability

distribution. The penalized partial log-likelihood, ℓ_{ppl}, consists of adding a penalty function, $\ell_{penalty}$, to the partial log-likelihood, $\ell_{partial}$, as a constraint as follows:

$$\ell_{ppl} = \ell_{partial} - \ell_{penalty}$$

$$= \sum_{i=1}^{s} \sum_{j=1}^{n_i} \log \left(\frac{e^{\mathbf{x}'_{ij}\boldsymbol{\beta}+\omega_i}}{\sum\limits_{qw \in R_{ij}} e^{\mathbf{x}'_{qw}\boldsymbol{\beta}+\omega_q}} \right) - \ell_{penalty}.$$

In the case of normally distributed random effects with mean 0 and variance γ, the penalty function is $\frac{1}{2} \sum_{i=1}^{s} \left[\frac{\omega_i^2}{\gamma} + \log(2\pi\gamma) \right]$. The tuning parameter of the penalty function is the random effect variance, γ, which penalizes for extremely large absolute values of ω_i.

The parameter estimation is done by starting with an initial guess of $\hat{\gamma}$, and using the Newton-Raphson method to solve for $\hat{\boldsymbol{\beta}}$ and $\hat{\omega}_i$ from the estimating equations based on the first partial derivatives of the penalized partial log-likelihood above. Then, $\boldsymbol{\beta}$ and ω_i are fixed at these maximized values and an estimating equation based on the approximate marginal log-likelihood is solved for a new value of $\hat{\gamma}$, more details of which may be found in Ripatti and Palmgren (2000). These two steps are iterated until convergence.

3.1.6 Model Selection

Model selection aims to find the subset of covariates which will result in the best fitting model. Forward selection is a model building approach which applies a stepwise search algorithm (see e.g., Venables and Ripley, 2002) in order to select covariates through the use of the Akaike information criterion (AIC), where a smaller AIC represents a better fit (Therneau and Grambsch, 2000). Beginning with the null model, the covariate which decreases the AIC the most is added to the model, one at a time, until the addition of any remaining covariates does not decrease the AIC. In the case of close fits, this method establishes a preference for simpler models by using the number of parameters as a penalty term.

3.2   Point Processes

This section introduces point process methods and estimation procedures with emphasis on simulation procedures.

3.2.1   Homogeneous Poisson Processes

A point process is a stochastic process of point occurrences. When the points are restricted to vary only temporally, the process is called "on the line" which refers to the time-axis. The homogeneous Poisson process is the simplest point process where the points occur completely randomly in time. This process has independent non-overlapping time increments and the number of events in any such increment of length a, $N(a)$, has a Poisson distribution with rate λa, where λ is the intensity of the Poisson process. The times between any 2 points, the increment lengths, are independent and exponentially distributed with rate λ. The simulation of a homogeneous Poisson process is therefore straightforward.

3.2.2   Non-Homogeneous Poisson Processes

A non-homogeneous Poisson process has an intensity which varies temporally, $\lambda(t)$. For this process, the number of events in the increment $(a, b]$ has a Poisson distribution with rate $\int_{a}^{b} \lambda(t)dt$ and non-overlapping increments are independent. In contrast to the homogeneous Poisson process, the number of events in an increment is dependent on its start and end points, along with its length; this makes the process non-stationary.

Thinning, introduced by Lewis and Shedler (1979), is a convenient method for simulating a non-homogeneous Poisson process. It involves simulating a homogeneous Poisson process with constant rate function, λ_u, which dominates the desired rate function, $\lambda(t)$, for all t in the desired interval. A generated point at time t is then independently rejected with probability $\left\{1 - \frac{\lambda(t)}{\lambda_u}\right\}$. A simple way to achieve this rejection rate is to generate an independent standard

uniform random variable, u_i, for each generated point i and reject the point if $u_i > \frac{\lambda(t_i)}{\lambda_u}$.

3.2.3 Parent-Child Cluster Processes

A *cluster process* consists of a point process which generates cluster origins (parents) and another point process which associates points (offspring) to each parent with specified location about the parent (Cox and Isham, 1980). The overall cluster process may or may not include the parents as points in practice, but it is of note that these parents are purely conceptual and do not have physical representations.

The *Bartlett-Lewis* process is a special cluster process in which a Poisson process generates the parents, a point process generates the offspring of each parent, and the offspring are positioned such that the increments between points are independent and identically distributed according to some distribution, starting from the parent. The *Neyman-Scott* process differs from the Bartlett-Lewis only in that the offspring are independently and identically distributed away from the parent. For both cluster processes, there is typically a stopping rule either for the total number of points generated, or for the number of offspring generated for each cluster parent.

3.2.4 Estimating $\lambda(t)$

The intensity of the non-homogeneous Poisson process will be estimated using a Poisson *generalized additive model* (GAM), first introduced by Hastie et al. (1986). This regression model relates the conditional mean of the response, $\mu = \mu(\mathbf{x})$, to a linear combination of smooth functions of covariates, $\mathbf{x}$, using the log link function. For a single covariate, x, this model takes the form:

$$\log(\mu) = \beta_0 + f(x),$$

where β_0 is the intercept which is estimated parametrically and f is a smooth function, which, in our application is estimated using a thin plate regression spline (Duchon, 1977) basis. Penalized iteratively reweighted least squares, conditional on a smoothing parameter, is implemented to fit the model, details of which may be found in Wood (2006). The smoothing parameter is estimated by generalized cross-validation.

The results from the implementation of the methods introduced in this chapter will be presented in Chapter 4.

Chapter 4

Results

4.1 Survival Analysis

The methods discussed in Section 3.1 are implemented in R using the `survival` package of Therneau (2014) and the results are presented in the remainder of this section.

4.1.1 Kaplan-Meier Estimator

The KM estimates of the survival curves of both lighting-caused and people-caused fires are displayed in Figure 4.1. Examination of these curves clearly shows that people-caused fires tend to have smaller survival probabilities and reveals two noticeable periods at which the curves exhibit a short-lived plateau rather than continually decreasing as expected. These plateaux occur near the 8 and 32 hour marks which suggests that they may be caused by suppression being stalled overnight.

4.1.2 Log-Location-Scale Models

The Weibull, the lognormal, and the loglogistic survival models are plotted for lightning-caused fires and people-caused fires in Figure 4.2. These results suggest that the fits of these parametric models are reasonable, however they do not adequately capture the features demonstrated by
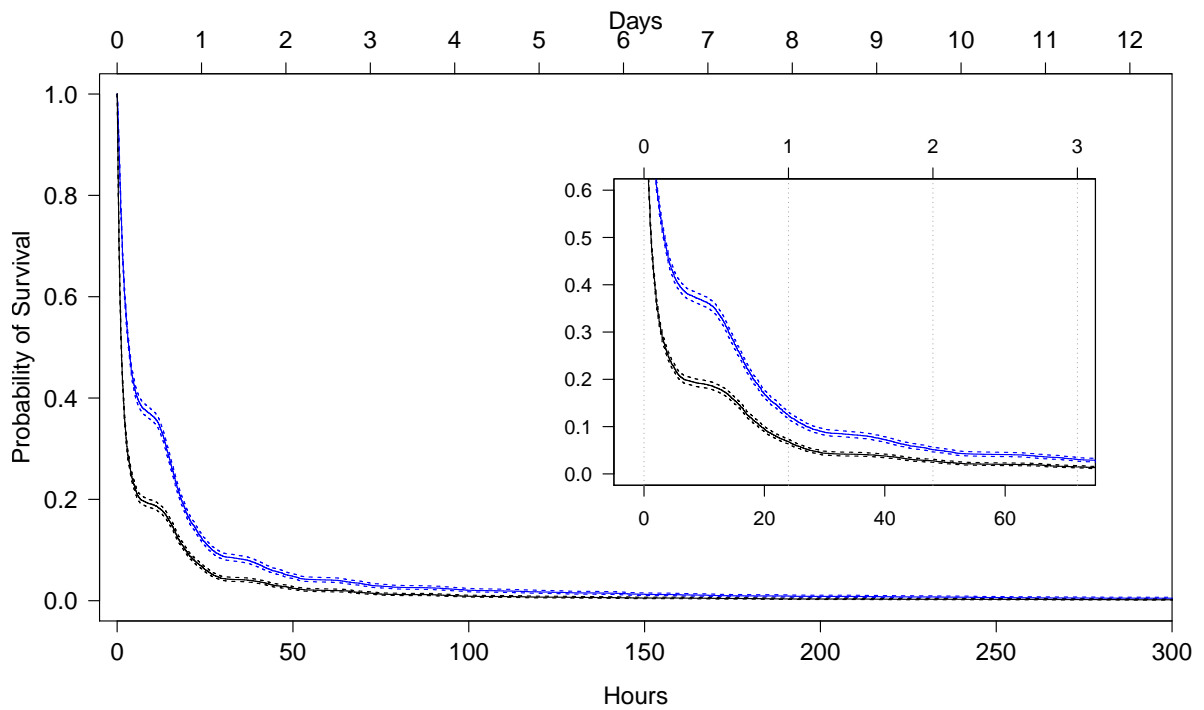
Figure 4.1: Kaplan-Meier survival curves of lightning (blue lines) and people-caused fires (black lines) with 95% confidence limits (dashed lines).
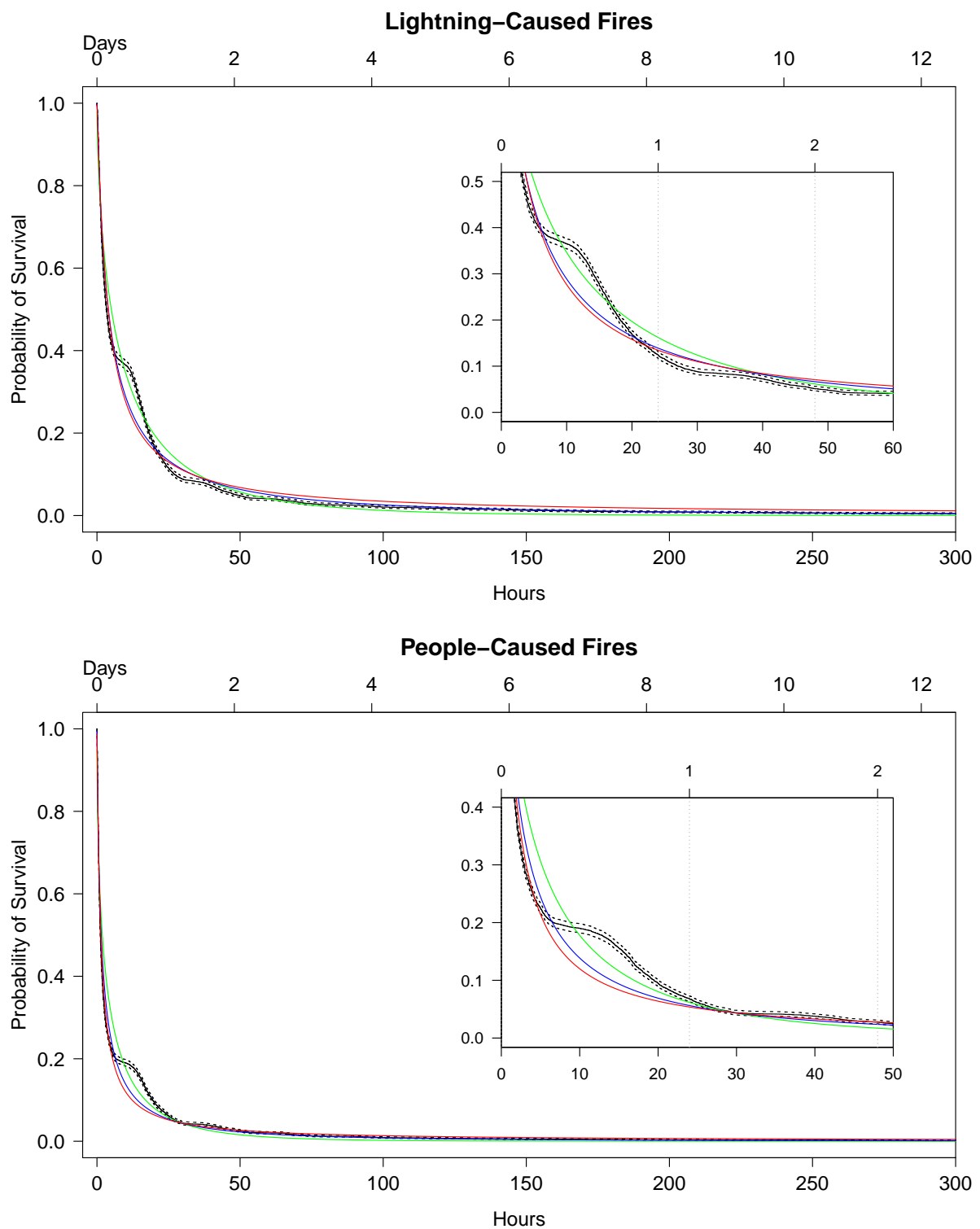
Figure 4.2: KM estimates (solid black lines) with 95% confidence limits (dashed black lines) and Weibull (green lines), lognormal (blue lines) and loglogistic (red lines) estimates of survival probabilities of lightning-caused fires (top panel) and people-caused fires (bottom panel).

the KM estimates in the left-tails of the distributions. This suggests that models which account for the effect of covariates should be explored. Examining the AIC values for each of these fitted models suggests that the Weibull distribution is the preferred log-location-scale model for both types of fires. Therefore, this distribution is employed to fit AFT models, the results of which are summarized in Table 4.1.

Lightning-Caused Fires		
Parameter	**Estimate (Std. Error)**	**P-Value**
IA Response	0.0122 (0.0012)	<0.001
FFMC	0.0140 (0.0018)	<0.001
ISI	0.0220 (0.0080)	0.0058
Size at IA	0.0006 (0.0002)	<0.001
DC	0.0009 (0.0003)	0.0102
DMC	0.0287 (0.0049)	<0.001
BUI	-0.0258 (0.0047)	<0.001

People-Caused Fires		
Parameter	**Estimate (Std. Error)**	**P-Value**
FFMC	0.0093 (0.0023)	<0.001
ISI	0.0518 (0.0068)	<0.001
Size at IA	0.0587 (0.0044)	<0.001
DC	0.0028 (0.0004)	<0.001
DMC	0.0242 (0.0065)	<0.001
BUI	-0.0207 (0.0061)	<0.001

Table 4.1: Parameter estimates, standard errors (Std. Errors) and p-values from the fitted AFT Weibull models of lightning (top panel) and people-caused (bottom panel) fires.

The implementation of forward selection (e.g., Venables and Ripley, 2002) results in fitting the model for lightning-caused fires with all covariates included, while the model for people-caused fires includes all variables except the initial attack response time. All parameter estimates, with the exception of BUI, are positive which suggests that increases in these variables are associated with an increased survival probability. The fitted AFT survival curves, along

with the KM estimates and 95% confidence limits of the survival curves, are displayed in Figure 4.3. These graphs suggest that the parametric AFT models over-smooth the data and do not adequately capture the characteristics which were displayed by the KM estimates of the survival curves. These log-location-scale models do, however, appear to do an adequate job of describing the overall trend in survival probability.

4.1.3   Cox Proportional Hazards Model

The Cox PH model, as described in Section 3.1.4, is fit to the survival time of lightning and people-caused fires. The same covariates as in the AFT models are selected. The summaries of the fitted models for lightning-caused fires and people-caused fires are displayed in Table 4.2.

Lightning-Caused Fires		
Parameter	**Estimate (Std. Error)**	**P-Value**
IA Response	-0.0067 (<0.001)	<0.001
FFMC	-0.0093 (0.0011)	<0.001
ISI	-0.0127 (0.0052)	0.0147
Size at IA	-0.0003 (<0.001)	0.0040
DC	-0.0005 (<0.001)	0.0223
DMC	-0.0142 (0.0031)	<0.001
BUI	0.0133 (0.0031)	<0.001

People-Caused Fires		
Parameter	**Estimate (Std. Error)**	**P-Value**
Size at IA	-0.0252 (0.0024)	<0.001
FFMC	-0.0061 (0.0013)	<0.001
ISI	-0.0256 (0.0040)	<0.001
DC	-0.0014 (0.0002)	<0.001
DMC	-0.0129 (0.0038)	<0.001
BUI	0.0109 (0.0036)	0.0021

Table 4.2: Parameter estimates, standard errors (Std. Errors) and p-values from the fitted Cox PH models of lightning (top panel) and people-caused (bottom panel) fires.

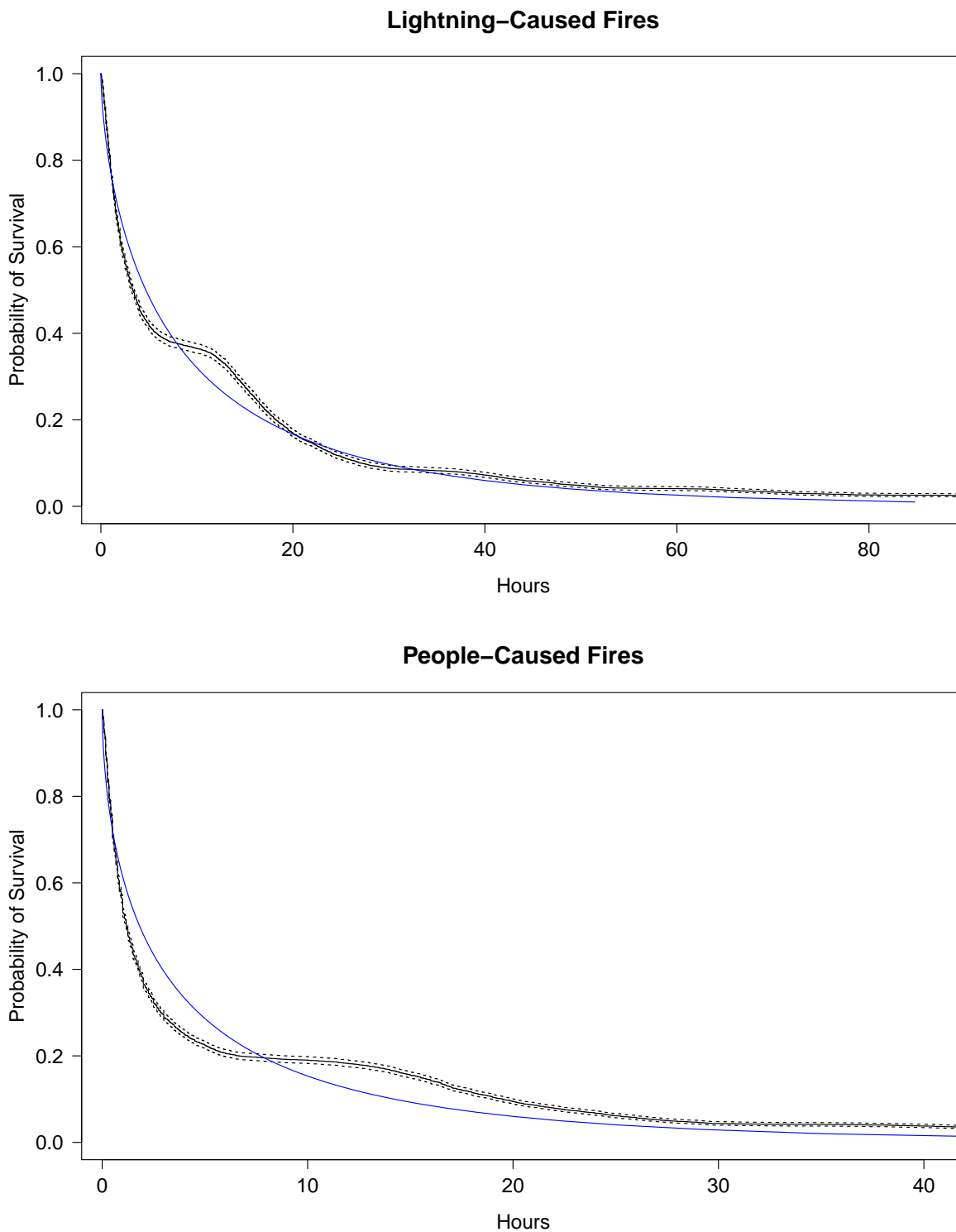**Lightning−Caused Fires**

People−Caused Fires

Figure 4.3: Fitted survival curves of lightning (top panel) and people-caused (bottom panel) fires at covariate values representative of a typical fire, as fit by the AFT Weibull model (blue lines) and the KM estimates (solid black lines) of survival probabilities with 95% confidence limits (dashed black lines).

Similarly to the AFT model, all parameter estimates, except BUI, are negative which indicates a decreased hazard rate, or equivalently, an increased survival time with increasing parameter values. The positive BUI parameter estimate is likely due to a dependence between it and another variable. Table 4.3 displays the marginal improvements in the AIC at each stage of the forward selection of covariates in these Cox PH models. The structure of the implemented stepwise selection guarantees an improvement in fit at each stage, however overfitting may occur. A general rule of thumb to avoid overfitting is that a marginal decrease in the AIC of less than 2 suggests that there is not substantial evidence which indicates a gain of information in the model with the additional covariate, relative to the previous model (Burnham and Anderson, 2004). Based on this rule, the Cox PH models of both lightning and people-caused fires would not include DMC and BUI.

Figure 4.4 displays the fitted Cox PH survival curves of lightning and people-caused fires (with all selected covariates), along with the KM estimates and 95% confidence limits of the survival curves. These graphs suggest that the Cox PH models capture the important characteristics which were displayed by the KM estimates of the survival curves. The Cox PH model is more flexible than the AFT model because its baseline distribution is left unspecified.

4.1.4 Proportional Hazards Shared Frailty Model

The Cox proportional hazards model assumes that the effects of covariates on the baseline hazard rate is proportional over time (with vertical shifts) and that the survival times of fires are independent; these assumptions will be violated when there is correlation among certain fires. Since the Province of Ontario has vast forest landscapes, the survival times of fires will be affected differently by the previously modeled covariates in dissimilar regions, thereby creating spatial correlation between fires within similar sections of Ontario. Martell and Sun (2008) partitioned the province into small polygons referred to as *fire management compartments* (FMCs) which are displayed in Figure 4.5. This partition was constructed by overlaying the map of fire

Lightning-Caused Fires		
Model	**AIC**	**Marginal Decrease**
Null	115304.7	-
IA Response	115108.6	196.1
FFMC	114939.8	168.8
Size at IA	114924.2	15.6
ISI	114916.5	7.7
DC	114913.6	2.9
DMC	114913.1	0.5
BUI	114893.9	19.2

People-Caused Fires		
Model	**AIC**	**Marginal Decrease**
Null	133389.9	-
Size at IA	133137.7	252.2
ISI	132988.0	149.7
DC	132874.5	113.5
FFMC	132854.1	20.4
DMC	132852.8	1.3
BUI	132845.1	7.7

Table 4.3: The AIC and marginal decrease in AIC at each stage of the forward selection of covariates in the Cox PH models of lightning (top panel) and people-caused (bottom panel) fires.

management zones in Ontario, displayed in Figure 2.1, with a digital map of the forest sections from the forest region classification system of Rowe (1972). Conditional on these compartments, the survival times of the fires may now be appropriately modeled as each FMC can be assumed to be approximately internally homogeneous with respect to ecological characteristics such as fuel, weather and topography as well as fire management strategy. The implementation of a shared frailty model, in which each fire belongs to only one compartment and the fires within a compartment have a common random effect, allows the assumption that the fires have independent survival times, conditional on the random effect terms.
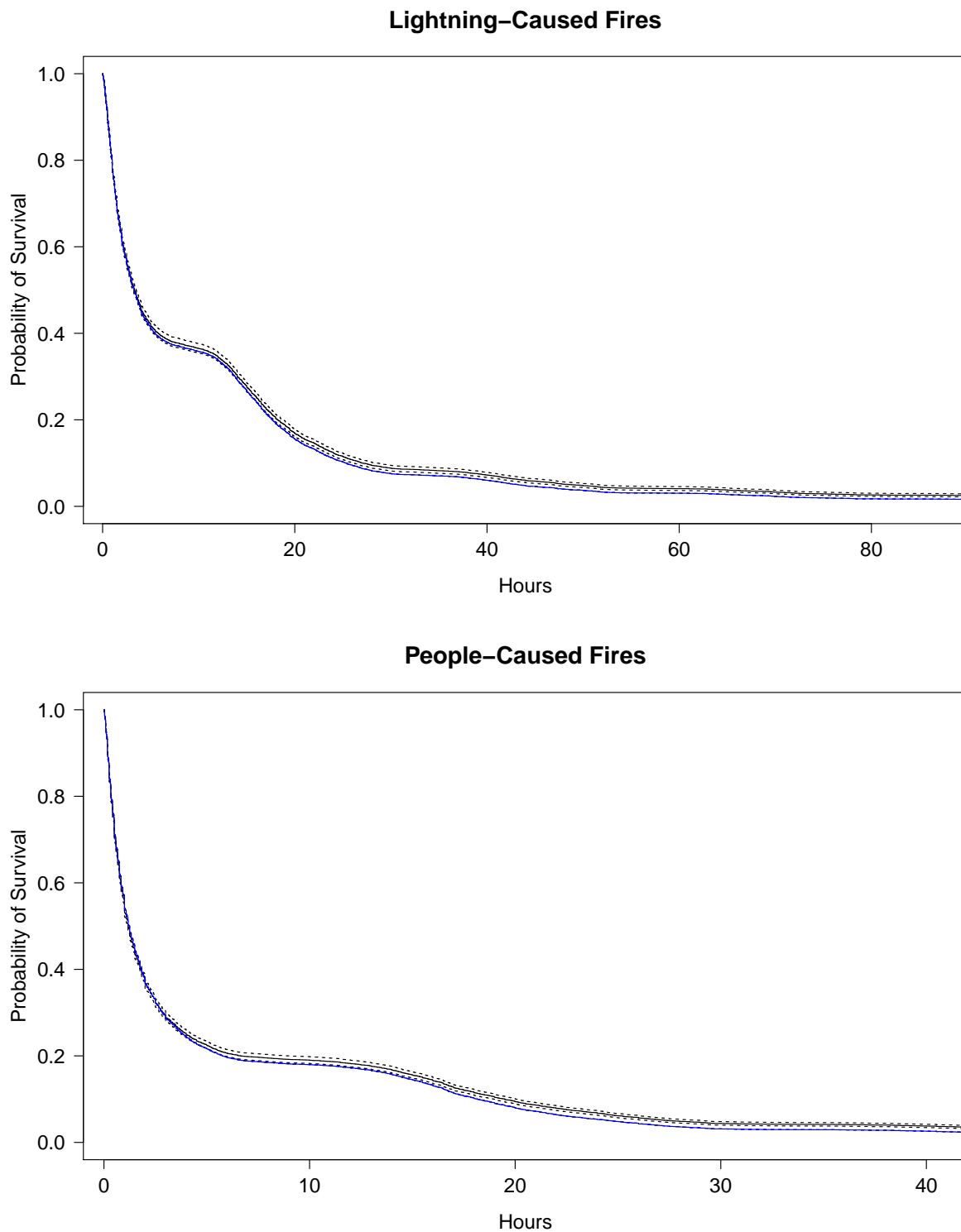
Figure 4.4: Fitted survival curves of lightning (top panel) and people-caused (bottom panel) fires at covariate values representative of a typical fire, as fit by the Cox PH model (blue lines) and the KM estimates (solid black lines) of survival probabilities with 95% confidence limits (dashed black lines).
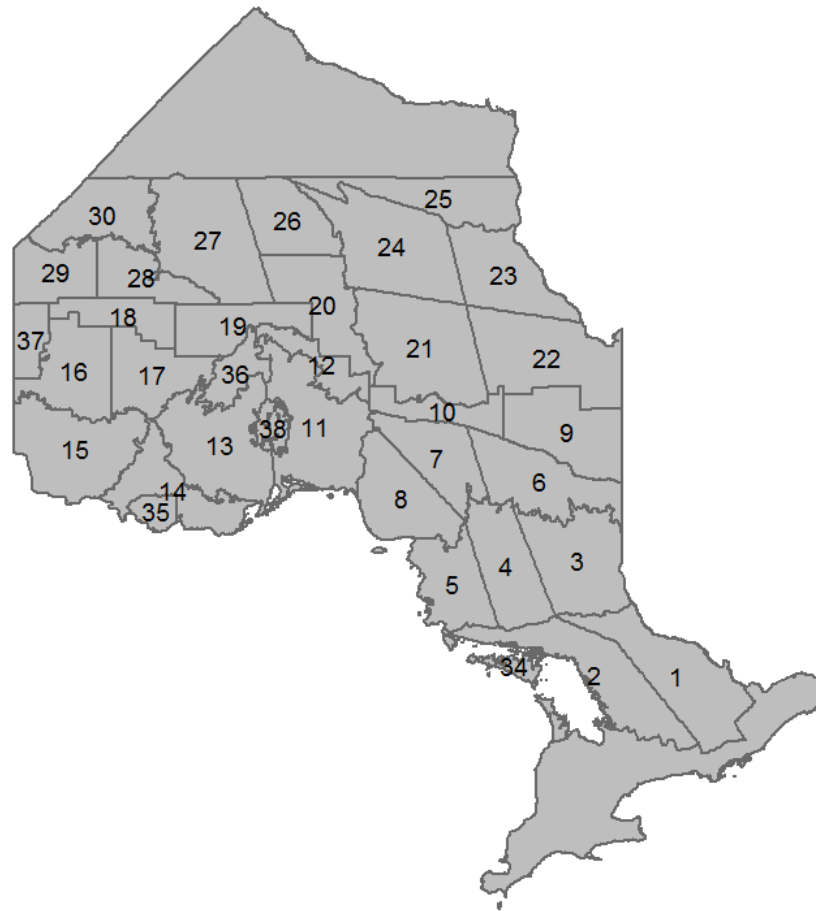
Figure 4.5: The partition of Ontario into fire management compartments.

A semiparametric PH shared frailty model with zero-mean normally distributed random effects is fit to lightning and people-caused fires by maximizing the penalized partial likelihoods as described in Section 3.1.5. The fixed effects used in these models are the same as in the non-frailty models to allow for appropriate comparisons. The parameter estimates, standard errors and p-values of the fixed effects from the fitted models are displayed in Table 4.4.

Lightning-Caused Fires		
Parameter	**Estimate (Std. Error)**	**P-Value**
IA Response	-0.0067 (0.0008)	< 0.001
FFMC	-0.0092 (0.0012)	< 0.001
ISI	-0.0103 (0.0052)	0.047
Size at IA	-0.0002 (0.0001)	0.009
DC	-0.0004 (0.0002)	0.067
DMC	-0.0150 (0.0031)	< 0.001
BUI	0.0127 (0.0031)	< 0.001

People-Caused Fires		
Parameter	**Estimate (Std. Error)**	**P-Value**
FFMC	-0.0060 (0.0013)	< 0.001
ISI	-0.0253 (0.0040)	< 0.001
Size at IA	-0.0249 (0.0024)	< 0.001
DC	-0.0014 (0.0002)	< 0.001
DMC	-0.0128 (0.0038)	< 0.001
BUI	0.0104 (0.0036)	0.004

Table 4.4: Parameter estimates, standard errors (Std. Errors) and p-values of the fixed effects from the fitted proportional hazards shared frailty model of lightning (top panel) and people-caused fires (bottom panel).

The posterior estimates of the random effects, $\hat{\omega}_i$, from each FMC are extracted from the model and applied to create the choropleth maps in Figures 4.6 and 4.7. The exponentiated posterior estimates are interpreted as multiplicative factors on the hazard rate. Negative posterior estimates imply a reduction in the hazard rate of fires, or analogously, an increase in the survival

**Frailty by Fire Management Compartment
of Lightning−Caused Fires**

Legend:
- (−0.35 , −0.25]
- (−0.25 , −0.15]
- (−0.15 , −0.05]
- (−0.05 , 0.05]
- (0.05 , 0.15]
- (0.15 , 0.25]
- (0.25 , 0.35]
- (0.35 , 0.45]

**Frailty by Fire Management Compartment
of Lightning−Caused Fires**

Legend:
- (−0.345,−0.172]
- (−0.172,0.172]
- (0.172,0.345]
- (0.345,0.517]
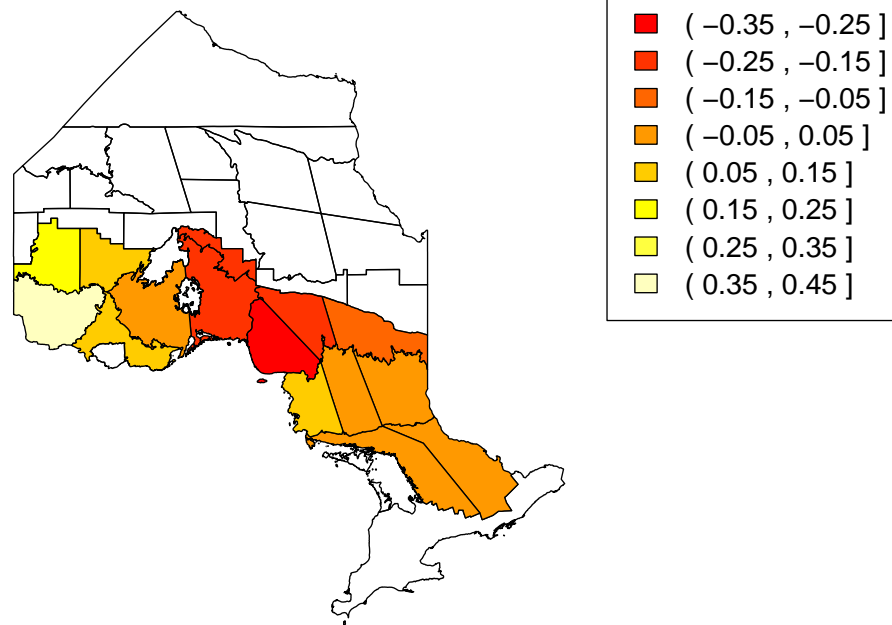
Figure 4.6: Choropleth maps of lightning-caused fires where each FMC is assigned a heat map colour based on its frailty term. The top panel uses evenly spaced intervals and the bottom panel uses an interval length equal to the standard deviation of the random effects.

Figure 4.7: Choropleth maps of people-caused fires where each FMC is assigned a heat map colour based on its frailty term. The top panel uses evenly spaced intervals and the bottom panel uses an interval length equal to the standard deviation of the random effects.

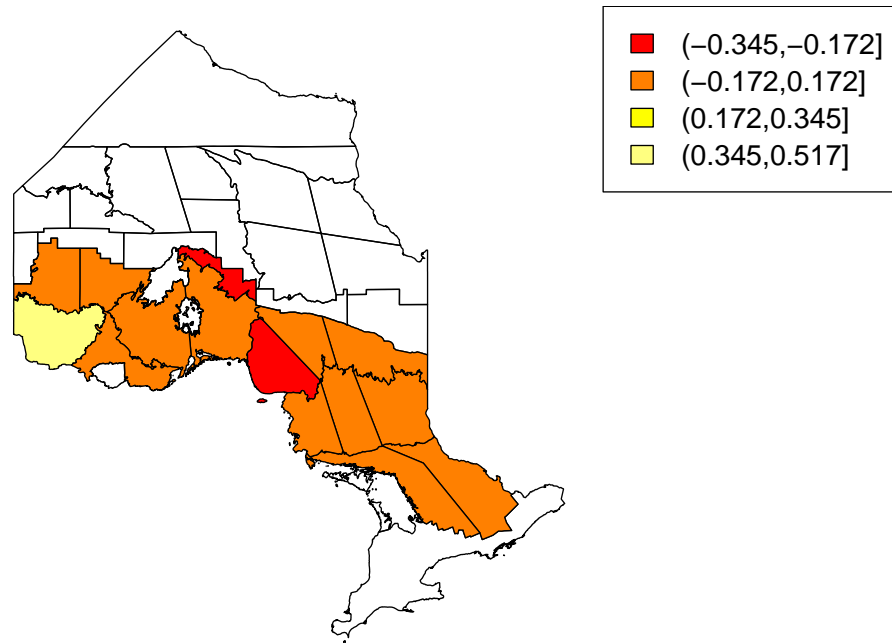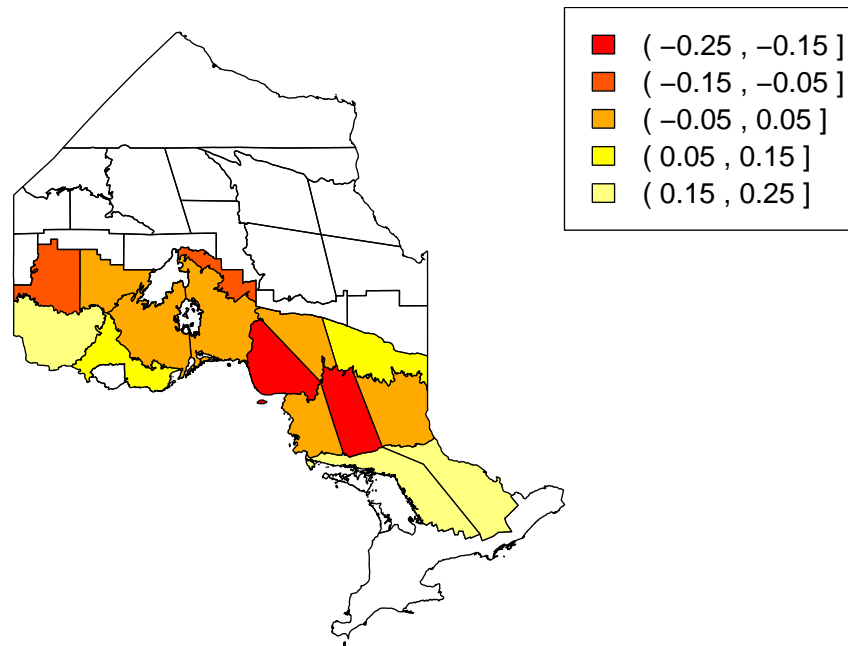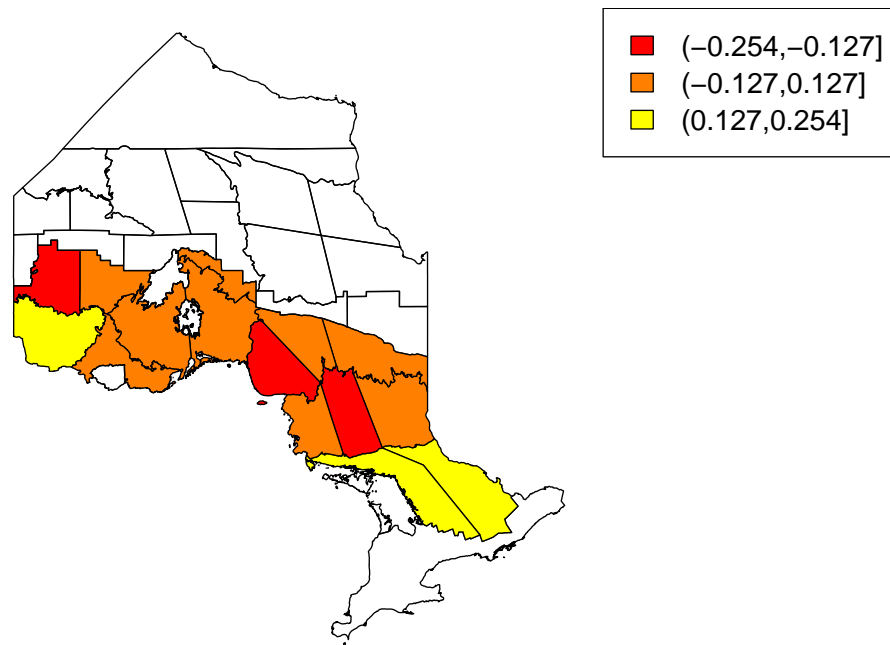probability. As such, the choropleth maps apply the brightest red color of the heat palette to compartments with the largest negative posterior estimates to imply greater fire danger through to the palest yellow being applied to compartments with the largest positive posterior estimates to imply the least fire danger.

The choropleth maps of people-caused fires do not display clear or easily interpretable trends, however the choropleth maps of lightning-caused fires have interesting patterns. The use of evenly-spaced intervals in the top panels allow the visualization of specific patterns. There is a slight west to east gradient which is noticeable in the Intensive zone of the evenly-spaced lightning-caused fires map (Figure 4.6, top panel). The bottom panels allow the visualization of more broad spatial patterns by using interval lengths which represent standard deviations of the random effects. The map of lightning-caused fires with standard deviation-based intervals (Figure 4.6, bottom panel) reinforces that in the Intensive zone, the western region experiences fires with shorter survival times.

4.1.5   Goodness of Fit

The construction of a profile likelihood-based confidence interval is used as a preliminary assessment of the frailty term's significance. For a sequence of random effect variances, the likelihood ratio (LR) test statistics are computed and plotted at each fixed variance. A horizontal line is added to the plot 3.84 units below the test statistic value of the fitted frailty model, which corresponds to a chi-squared test on 1 degree of freedom. The intersections of the horizontal line and the profile likelihood produce a profile likelihood-based 95% confidence interval for the variance of the random effects. The resulting plots, displayed in Figure 4.8, suggest that the frailty terms of the models are significant in terms of this preliminary assessment since the confidence intervals are relatively narrow. It is of note that the confidence interval for the variance of the random effects from the frailty model of lightning-caused fires is narrower than the confidence interval from the frailty model of people-caused fires.

Profile Likelihood of Lightning–Caused Fires

Profile Likelihood of People–Caused Fires
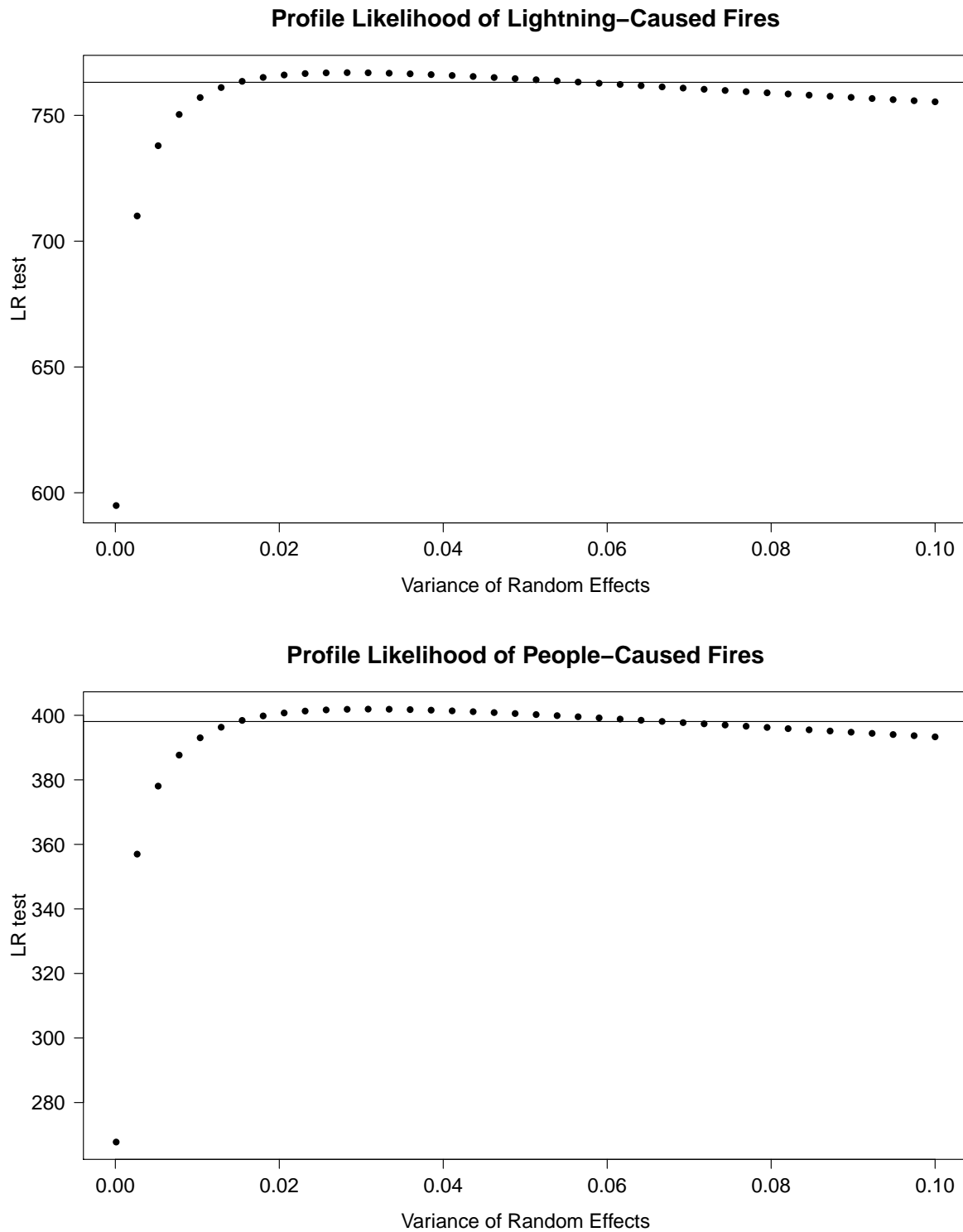
Figure 4.8: Profile likelihood-based 95% confidence intervals of the variance of the random effects.

Therneau and Grambsch (2000) suggests that the significance of the frailty term is more strongly evaluated by a likelihood ratio test comparing the integrated-likelihood of the frailty model, where the frailty terms are integrated out of the likelihood, to the Cox PH fitted likelihood without a frailty term. Based on this comparison, both the lightning and people-caused frailty models are significant improvements on the respective proportional hazards models without frailty term. However, it is of note that this chi-squared test is conservative since the frailty terms, e^ω, are constrained to be greater than or equal to zero.

Next, the suitability of the normality assumption of the frailty model is verified by fitting the Cox PH model with additional fixed effects for the FMCs. FMC-15, the compartment in the western region which experiences the most lightning-caused fires, is chosen as the baseline compartment. To make the choropleth maps from the frailty model comparable to those from the fixed effects model in Figures 4.9 and 4.10, FMC-15's posterior estimate is subtracted from the other posterior estimates such that the values are representative of the multiplicative difference on the hazard rate of fires in FMC-15. The fixed effect parameter estimates are similar to the posterior random effect estimates which suggests that this modelling is robust to violation of the normal frailty assumption.

**Fixed Effect Parameter Estimates
of Lightning−Caused Fires**
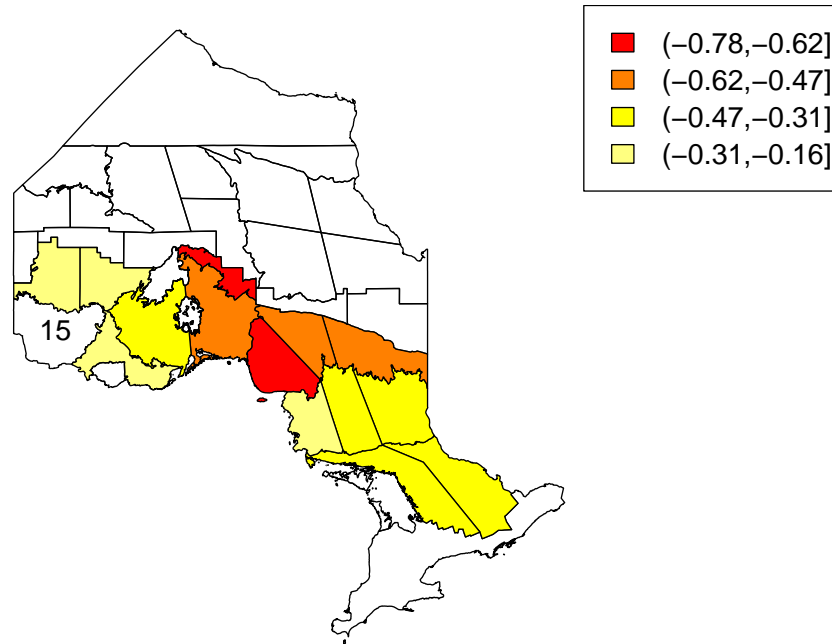
**Frailty Comparison to FMC15
of Lightning−Caused Fires**

Figure 4.9: The parameter estimates of the compartment-specific fixed effects (top panel) and the posterior estimates, in reference to FMC-15, from the frailty model (bottom panel) of lightning-caused fires.

**Fixed Effect Parameter Estimates
of People–Caused Fires**

🟥	(−0.56,−0.42]
🟧	(−0.42,−0.28]
🟨	(−0.28,−0.14]
🟨	(−0.14,0.02]

**Frailty Comparison to FMC15
of People–Caused Fires**

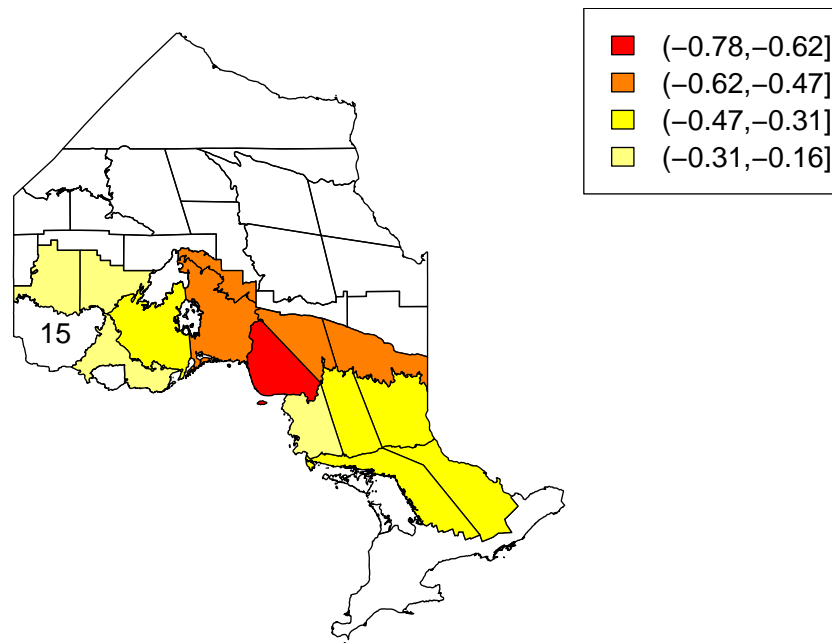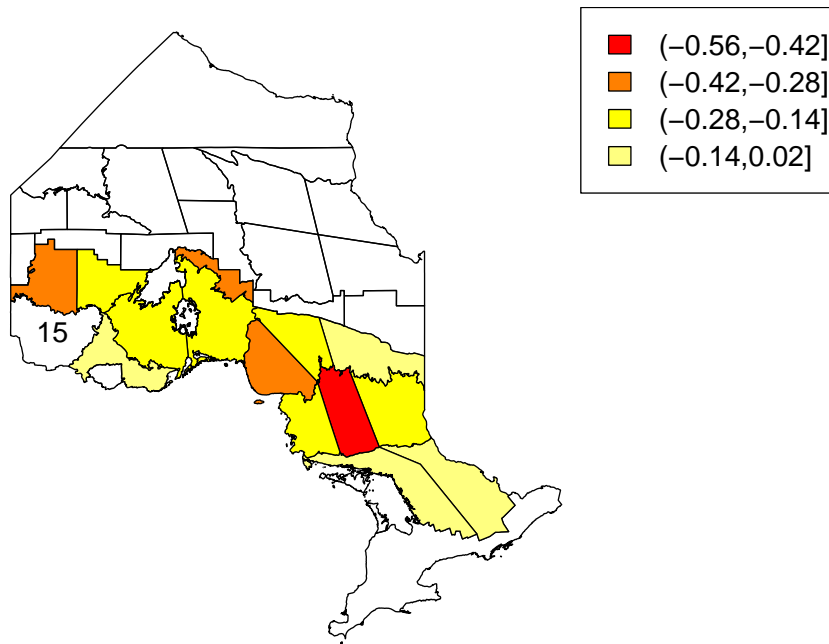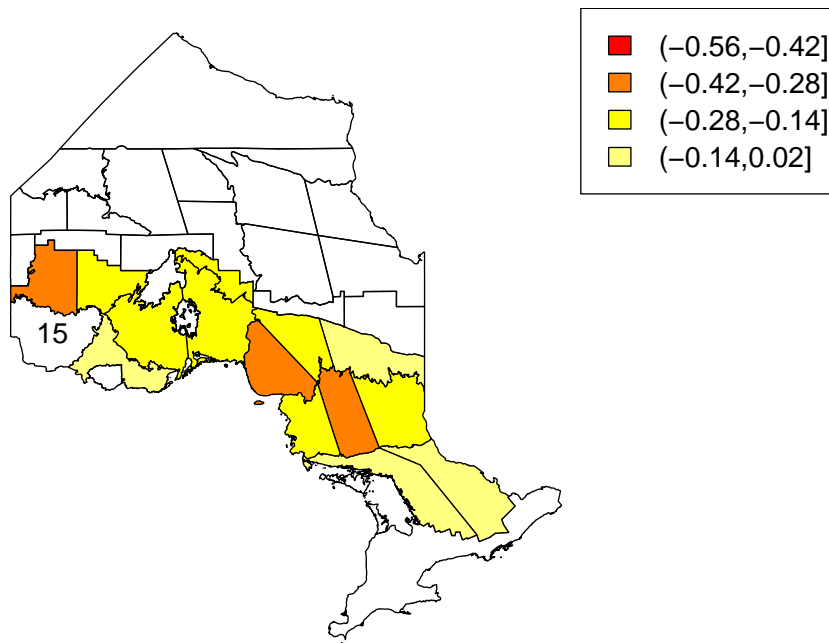|	
🟥	(−0.56,−0.42]
🟧	(−0.42,−0.28]
🟨	(−0.28,−0.14]
🟨	(−0.14,0.02]

Figure 4.10: The parameter estimates of the compartment-specific fixed effects (top panel) and the posterior estimates, in reference to FMC-15, from the frailty model (bottom panel) of people-caused fires.

4.2 Fire Arrival Modelling

The modelling in the remainder of this section will be implemented on lightning-caused fire occurrences from 1963 through 2004. Our study region is restricted to FMC-15 such that the modelling is purely temporal.

4.2.1 Non-Homogeneous Poisson Process

A Poisson GAM with a thin-plate spline for a day of year effect is fit to the daily number of fires. The resulting time-dependent rate is used to generate fire arrivals (fires which have been reported) according to a non-homogeneous Poisson process. This simulation is facilitated by thinning realizations from a homogeneous Poisson process as described in Section 3.2.2. The resulting 95th percentiles of daily fire arrivals over 1000 simulated fire seasons and the true data are displayed in Figure 4.11. Inspection of this figure suggests that the intensity of the Poisson process is too small, especially near the peak (middle) of the fire season, due to the large number of days with 0 fires in the data. Table 4.5 indicates that the average number of fires per fire season from the simulation is similar to that of the historical data, while the average number of *fire days*, defined as days with 1 or more fires, per fire season from the simulation is much larger than the historical data. This discrepancy implies that the generated fires are too spread out and that there are not enough fires generated on the same day. An intuitively appealing way of modelling the inherent clustering of the arrival of lightning-caused fires is by the means of a parent-child cluster process, which will be implemented in the next section.

	Average No. Fire Days	Average No. Fires
Historical	29.60 (15.97)	77.50 (78.76)
Poisson Process	58.09 (5.40)	78.23 (8.41)

Table 4.5: A comparison of the average number of fire days and fires per fire season, with standard deviations in parentheses, from the historical data and the Poisson process simulation.

Non–Homogeneous Poisson Process Simulation

Figure 4.11: The 95th percentiles of daily fire arrivals over the fire season from 1000 runs of a simulated non-homogeneous Poisson process (blue line) and from the true data (black line).

4.2.2 Parent-Child Cluster Processes

The implemented Bartlett-Lewis process is assumed to have the following two-component structure. A non-homogeneous Poisson process of rate $\lambda(t)$, which is dependent on the day of year, generates the cluster origins (parent fires), and for each parent, offspring are generated according to a homogeneous Poisson process of rate β within 1 day from the time of the parent fire arrival. The generated parents are also assumed to be fires, which implies that if 0 offspring are generated for a parent, the associated cluster consists of a single fire. This cluster process couples the two individual components independently; implications of this assumption are discussed in Chapter 5.

The rate of the non-homogeneous Poisson process, $\lambda(t)$, is estimated by fitting a Poisson GAM, with a thin-plate spline for the day of year effect, to the daily presence/absence of fires (i.e., fire

days). The resulting time-dependent rate is displayed in Figure 4.12. A simple Poisson model is fit to the daily fire counts from the zero-truncated data. The resulting estimated rate, $\hat{\beta}$, represents the offspring intensity, given a fire day has been observed. These estimated parameters are used to simulate the Bartlett-Lewis process 1000 times and the 95th percentiles of daily fire arrivals over the fire season from the simulation and true data are displayed in Figure 4.13.

Poisson GAM of Fire Days
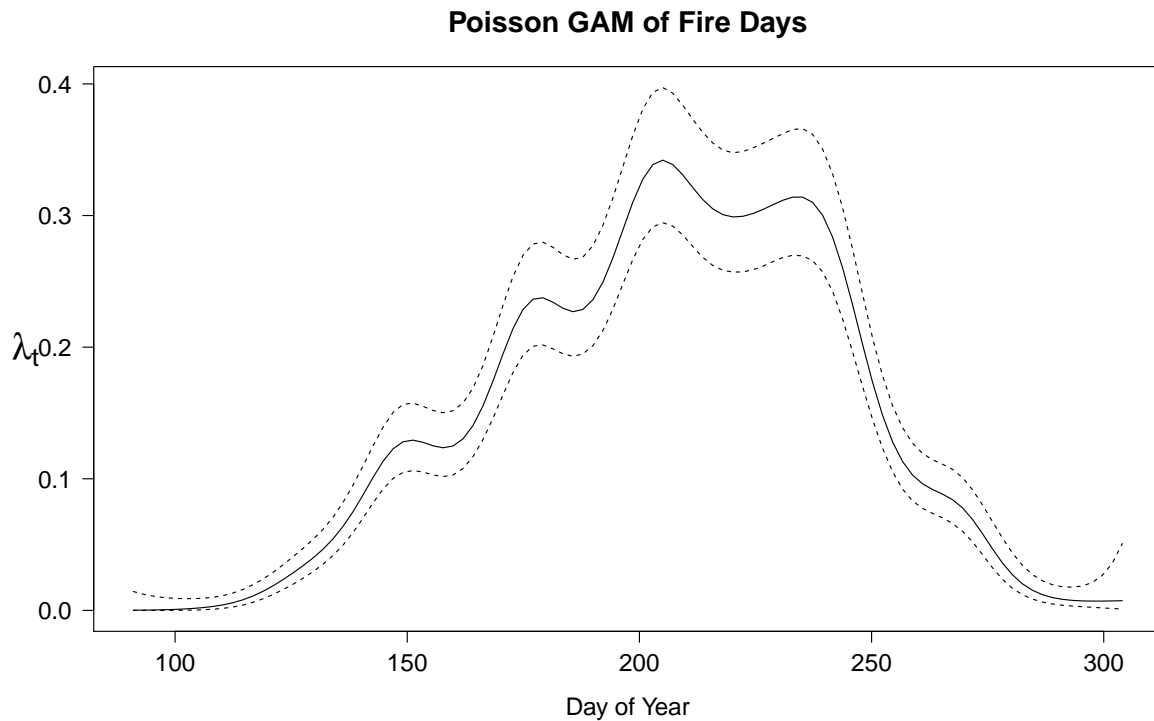
Figure 4.12: The estimated time-dependent rate from a Poisson GAM with day of year effect fit to the presence/absence of fires.

The Bartlett-Lewis process specifies that the intervals between successive points are iid, which was achieved by generating the offspring according to a homogeneous Poisson process. By contrast, the Neyman-Scott process positions points such that they are iid from the cluster origin. This is achieved by generating offspring arrival times away from the parent fires using independent truncated exponential($\hat{\beta}$) random variables (truncated above 1 day). In this case, a stopping rule which involves a count of offspring is more suitable as this process does not generate points sequentially. Instead, the associated number of offspring per parent is defined
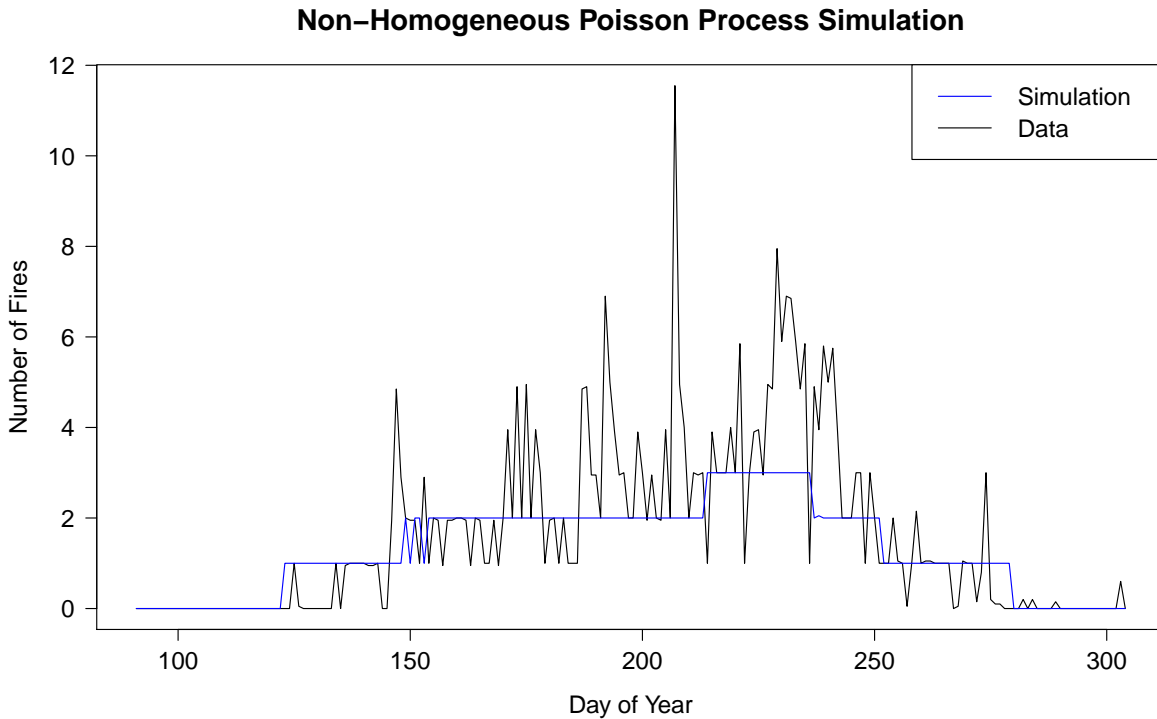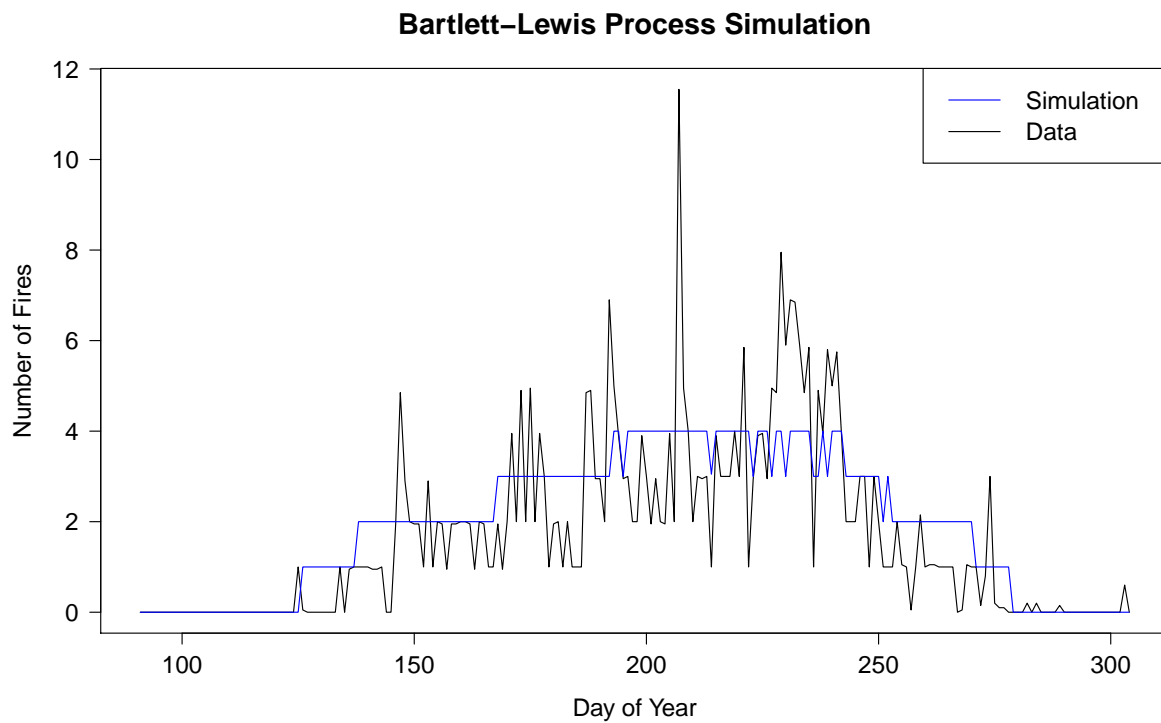
Figure 4.13: The 95th percentiles of daily fire arrivals over the fire season from 1000 runs of a simulated Bartlett-Lewis process (blue line) and from the true data (black line).

as $\hat{\beta}$. Since β is the expected number of offspring over 1 day from the homogeneous Poisson process in the Bartlett-Lewis process, the total expected number of fires from the two cluster processes are equivalent. Figure 4.14 displays 2 historical fire seasons (1983 and 1998), along with 2 fire seasons generated by each of the non-homogeneous Poisson, Bartlett-Lewis and Neyman-Scott processes. This Neyman-Scott process is simulated 1000 times and the 95th percentiles of daily fire arrivals over the fire season from the simulation and true data are displayed in Figure 4.15.

Figures 4.13 and 4.15 illustrate that parent-child cluster models improve the simulated arrivals during the peak of the fire season as compared to the simple arrival process. The clustering also significantly improves the average number of fire days per season, displayed in Table 4.6, as they are closer to the historical value than the single Poisson process simulation.

	Average No. Fire Days	**Average No. Fires**
Historical	29.60 (15.97)	77.5 (78.76)
Bartlett-Lewis	37.91 (6.27)	78.22 (15.94)
Neyman-Scott	37.17 (6.49)	78.45 (14.58)

Table 4.6: A comparison of the average number of fire days and fires per fire season, with standard deviations in parentheses, from the historical data, the Bartlett-Lewis simulation, and the Neyman-Scott simulation.

4.2.3   Fire Load Simulation

A simulation of the fire load, the number of fires burning on the landscape, over a fire season is accomplished by generating observations from a marked cluster process, where the survival times are independent marks associated with points from the cluster process. For each arrival time generated from the Bartlett-Lewis process, a survival time is simulated from the Weibull AFT model which was fit in Section 4.1.2 with a covariate vector which is representative of a
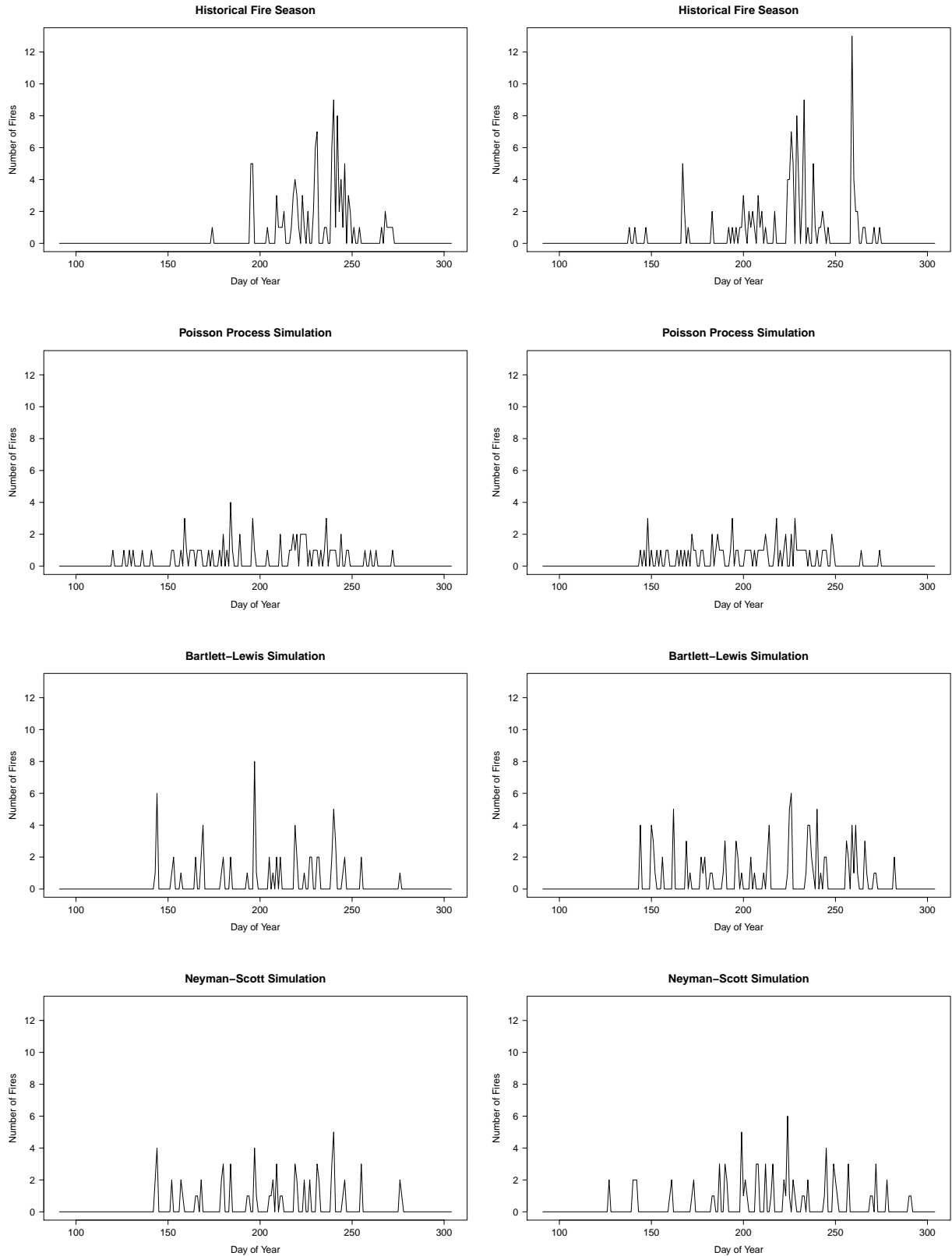
Figure 4.14: Historical fire seasons for 1983 and 1998 (top row) and fire seasons generated by the non-homogeneous Poisson (2nd row), Bartlett-Lewis (3rd row) and Neyman-Scott (bottom row) processes.
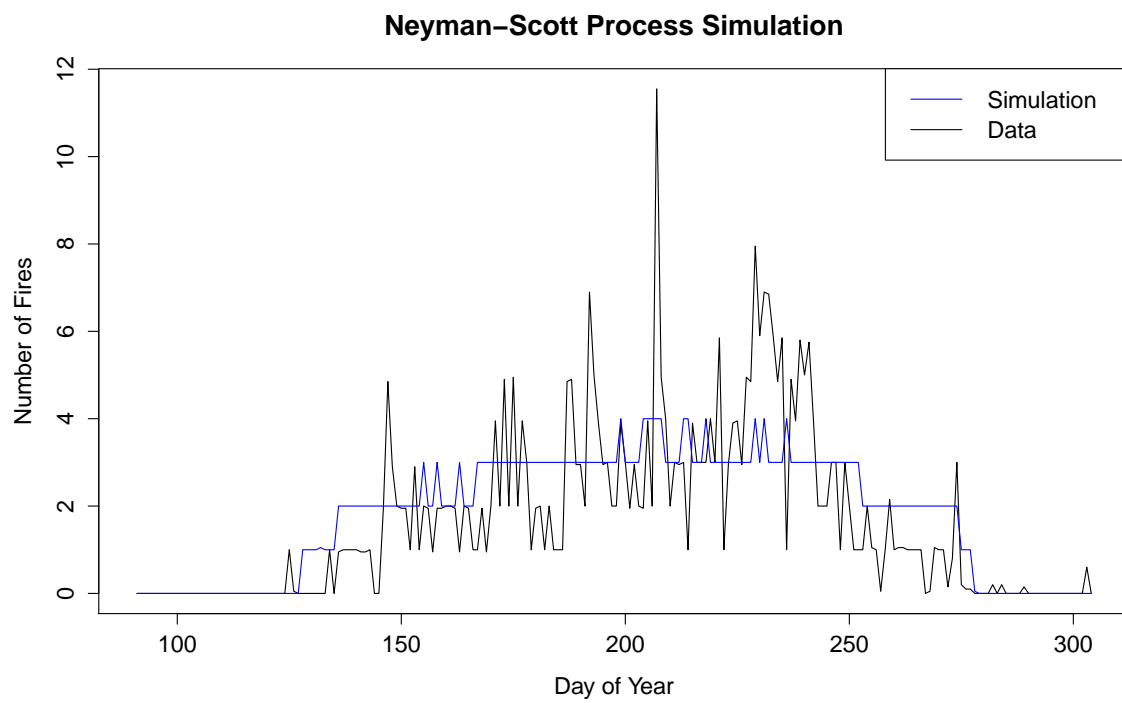
Figure 4.15: The 95th percentiles of daily fire arrivals over the fire season from 1000 runs of a simulated Neyman-Scott process (blue line) and from the true data (black line).

typical fire. The top panel of Figure 4.16 displays each fire arrival and its survival time for a single run of this simulation, along with the associated daily fire load over the fire season in the bottom panel.

The results presented in this chapter will be discussed, followed by future work, in the next chapter.

Simulated Arrivals

Simulated Fire Load
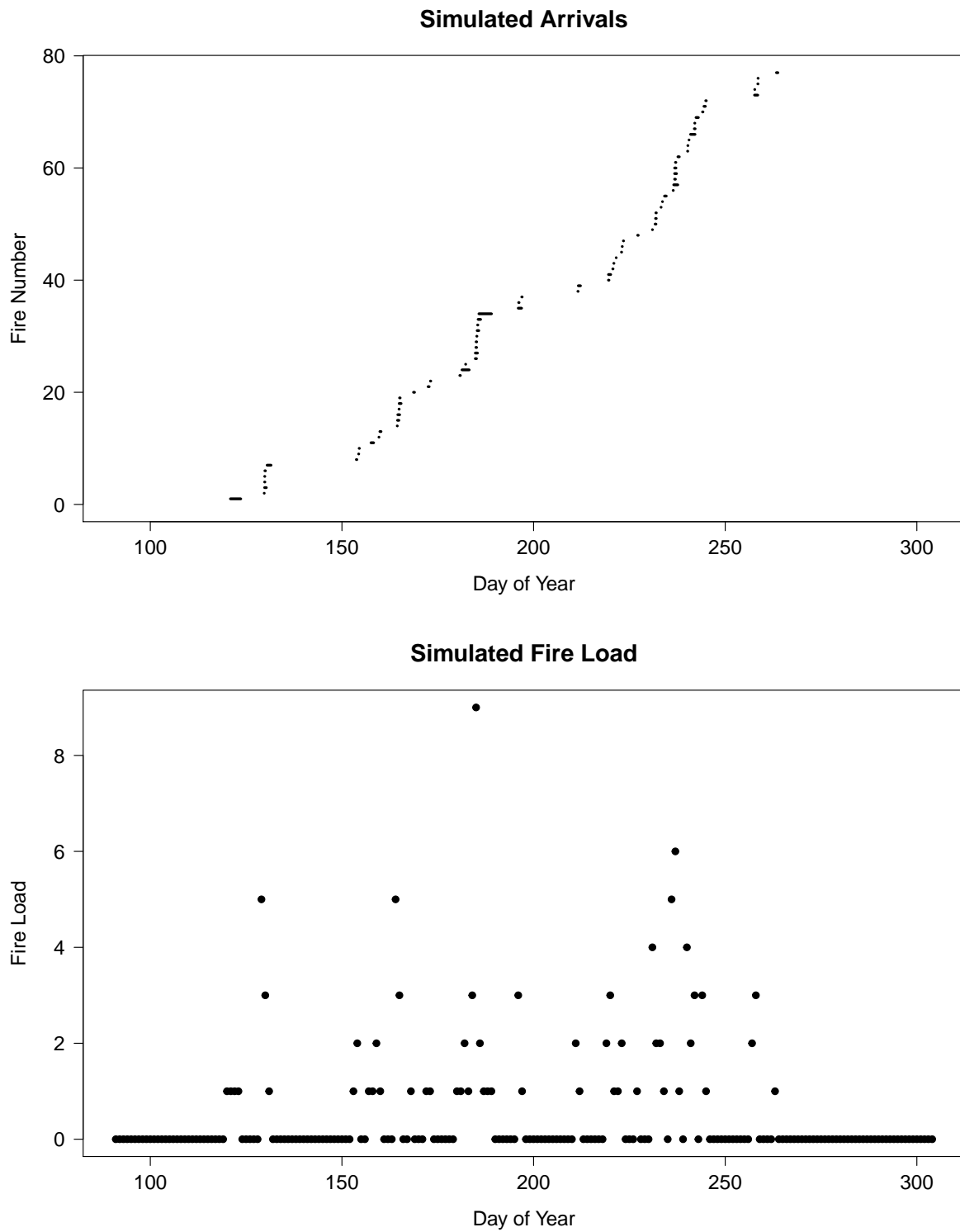
Figure 4.16: Fire arrivals (segments) and survival times (length of segments) over the fire season (top panel) and the associated daily fire load (bottom panel) from a single run of the simulation.

Chapter 5

Conclusion

5.1 Discussion

We have presented a thorough survival analysis in the context of exploring the associations between fire weather covariates and the distribution of the survival time of a forest fire, along with an investigation of spatial patterns. In doing so, we have illustrated that the survival times of forest fires are dependent on fire weather variables. The initial attack response time between a fire being reported and the start of initial fire suppression efforts, and the fire's size at initial attack were statistically significant when modelling hazard rates for forest fires in Ontario. The simple (i.e., no covariates) parametric log-location-scale models failed to capture the plateaux-like features observed in the left-tail of the survival distributions. Although incorporating the effect of covariates using Weibull AFT models led to improvements in fit, the estimated survival curves were quite smooth and again failed to capture the important characteristics displayed in the KM curves. The results from the Cox PH model were found to closely resemble the plateaux noticed in the KM curves and this was seen for both lightning and people-caused fires. Operationally, the left-tail of the distribution is important as most fires are declared under control within 1 or 2 days from the onset of initial attack.

In our analysis of historical forest fire data, all variables considered were included in the model for lightning-caused fires; the fitted model for people-caused fires did not include the initial attack response time. This supports the conjecture, and the basis of fire management, that the longer the initial attack response time, the larger a fire becomes and the more difficult it is to get under control. Figure 2.4 illustrated that the initial attack response time for people-caused fires, as compared to lightning-caused fires, does not tend to be as long. This may be a result of people-caused fires being smaller in size at the time of initial attack and because they tend to occur near populated areas which makes them easier to access by suppression crews and equipment. Lightning-caused fires also tend to arrive in clusters due to passing storm systems (Woolford and Braun, 2007). This may result in multiple fires on the landscape, in relatively close proximity to each other, which can overwhelm the fire management crews who have limited suppression resources, resulting in longer initial attack response times before suppression action begins on some of those fires.

Each of the parameter estimates from the Cox PH models, with the exception of BUI, had a negative value which indicated an increase in the survival probability as the covariates increased. This relation is consistent as a longer initial attack response time leads to a fire with a greater size at initial attack which makes the fire more difficult to control. Also, the fire weather variables from the CFFDRS are structured to indicate increased fire risk with increasing values, which may result in longer survival times. The positive value of the BUI parameter estimate may be the result of its dependence on at least one of the other variables in the models since, by definition, the BUI is a function of the DMC and the DC (Wotton, 2009). Inspection of the marginal improvements in the AIC at each stage of the forward selection of covariates suggested that both the models for lightning and for people-caused fires should not include DMC or BUI to avoid overfitting.

A lightning strike can cause the ignition of a fire, however the arrival of this fire, which is

defined once the fire is reported, is heavily dependent on the moisture within the top layer or on the surface of the forest floor (Wotton and Martell, 2005). This layer must exhibit conditions such that a fire can smoulder until the forest floor is dry enough to sustain its spread. The FFMC was a highly significant covariate in the fitted Cox PH model of lightning-caused fires. Since the FFMC represents the moisture content in surface fuels, its significance in modelling the survival time of lightning-caused fires is logical. In the fitted Cox PH model of people-caused fires, the ISI was a highly significant covariate. The significance of this variable is consistent with the nature of people-caused fires which are dependent on the moisture of the fine fuels on the surface of the forest floor, as the ISI is based on a combination of the FFMC and the wind speed.

The Cox PH model is based on the assumption that the effects of covariates on the base-line hazard rate is proportional over time. This PH assumption may be tested in R using the `cox.zph` function of the `survival` package (Therneau, 2014). Details of the formal test are found in Grambsch and Therneau (1994). Upon formally testing the selected covariates, FFMC was found to have violated the PH assumption for the model of lightning-caused fires and similarly with size at initial attack for both models of lightning and people-caused fires. The non-proportionalities of these covariates were further inspected, as suggested by Therneau and Grambsch (2000) when dealing with large datasets. For FFMC, the range of non-proportionality is relatively small in comparison to its range of values and the non-proportional test outcome for size at initial attack seems to be caused by a few long fires and therefore not a cause for concern. Both covariates were deemed appropriate in terms of interpretation and retained in the models. The models were re-fit to the data excluding these long fires to investigate whether they are driving the reported effects. The same covariates were found to be statistically significant, the associated parameter values had the same sign and, in general, they were of the same order of magnitude. We conclude that these long fires are not influential outliers.

To account for spatial variability, the Cox PH models were extended to incorporate random effects which improved the fits due to the presence of spatial correlation between the survival times of fires within a set of fire management compartments which formed a spatial partition of Ontario's fire region. The semi-parametric PH shared frailty models were fit with zero-mean normally distributed random effects for each fire management compartment. These frailty terms were assessed as significant by creating profile likelihood-based confidence intervals and by performing likelihood ratio tests which compared the integrated-likelihoods of the frailty models and the likelihoods of the models without frailty terms. Choropleth maps were created using the posterior estimates of the random effects from each fire management compartment which were extracted from the fitted models. The maps from the lightning-caused fires displayed much more prominent patterns than did the maps from the people-caused fires. The analysis of spatial patterns in the case of lightning-caused fires is of greater importance as lightning strokes tend to arrive in spatio-temporal clusters from storm-weather systems. From the choropleth maps of lightning-caused fires, a slight west to east gradient was visible in the Intensive fire management zone. Looking at more broad spatial patterns brought to light that the western region of the Intensive zone seems to have fires with shorter survival times. This highly populated region experiences a large proportion of lightning-caused fires and, as a result, has been provided with numerous resources, historically. The normal frailty assumption was verified by re-creating these choropleth maps using fixed effects to represent the compartment effect. The fixed effect parameter estimates were similar to the posterior random effect estimates which suggests that this modelling is robust to violation of the normal frailty assumption. In the future, these types of frailty models could be extended to regions with a greater number of compartments, in which using fixed effects would be inappropriate due to the increasing number of parameters to be estimated.

The arrivals of lightning-caused fires in a single FMC were investigated by the means of temporal simulations. A non-homogeneous Poisson process, which took into account the day of

year effect, was found to generate an appropriate number of fires per fire season on average, but too many fire days as compared to historical data. This implied that the inherent clustering in the arrival of lightning-caused fires was not represented by this single process, and as a result, parent-child cluster processes were considered. Special forms of the Bartlett-Lewis and Neyman-Scott processes generated an average number of fire days per fire season which was much closer to the historical data, as compared to the single non-homogeneous Poisson process. Due to the large number of days with 0 fires and a purely stochastic modelling approach, our simulations were assessed by comparing the 95th percentiles of the number of fires on each day of the fire season from 1000 runs and from the historical data. This comparison illustrated that the parent-child cluster processes capture the overall trend over the fire season quite well and could possibly benefit from a third process to allow the intensity of the offspring generation to vary with weather conditions and storm intensity. A fire load model was then developed by independently coupling survival times, generated from the previously fitted Weibull AFT, to the arrivals from the Bartlett-Lewis model.

5.2   Future Work

The survival time presented in this thesis was defined as the time period from the onset of initial attack to the fire being declared under control. The survival analysis presented could be applied to alternative time periods including using the time of ignition or report of the fire as the lower-bound of the time period. It is of note however, that the time of ignition is often estimated and using this lower-bound would therefore require the left-censoring be accounted for in the model fitting. A variation of the marked cluster process could also include using information other than the survival time as the mark such as area burned or cost of suppression.

Our future work also involves implementing refinements to both the survival and fire arrival models presented. The values of the fire weather variable from the report dates of the fires

were used as constant covariates in the survival models. These variables, however, are changing over the duration of the fire and could be implemented as time-varying covariates since they are recorded daily. In terms of predicting the survival time as a fire arrives, the time-varying covariate values could then be incorporated from weather forecasts. The arrival model could be refined by employing a clustering algorithm (e.g., Woolford and Braun, 2007) to find cluster origins, rather than modelling fire days. The practical interpretation, and possibly the quality of the model, would be improved as the clusters in the parent-child cluster processes are currently representative of fire days, while in reality, a cluster could be much shorter or possibly span multiple days. A further refinement could be to incorporate a correlation structure between the current fire load and the mark (survival time) associated with the arrival of an additional fire. The independence currently assumed in the coupling of the fire arrival and survival models does not take into account resources being tied up when there are multiple fires on the landscape which would lead to longer initial attack response times, and in turn, longer survival times. The refined models could be used to study the potential future changes to forest fire load under climate change scenarios.

Bibliography

Breslow, N. (1972). Discussion on regression models and life-tables(by dr cox). *J. Roy. Statist. Soc. Ser. B*, 34:216–217.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference understanding aic and bic in model selection. *Sociological methods & research*, 33(2):261–304.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Cox, D. R. and Isham, V. (1980). *Point processes*, volume 12. CRC Press.

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in sobolev spaces. In *Constructive theory of functions of several variables*, pages 85–100. Springer.

Efron, B. (1977). The efficiency of cox's likelihood function for censored data. *Journal of the American statistical Association*, 72(359):557–565.

Fleming, T. R. and Lin, D. (2000). Survival analysis in clinical trials: past developments and future directions. *Biometrics*, 56(4):971–983.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526.

Hastie, T., Tibshirani, R., et al. (1986). Generalized additive models. *Statistical science*, 1(3):297–310.

Johnson, C. J., Boyce, M. S., Schwartz, C. C., and Haroldson, M. A. (2004). Modeling survival: application of the andersen-gill model to yellowstone grizzly bears. *Journal of Wildlife Management*, 68(4):966–978.

Larsen, C. (1997). Spatial and temporal variations in boreal forest fire frequency in northern alberta. *Journal of Biogeography*, 24(5):663–673.

Lawless, J. F. (2003). *Statistical models and methods for lifetime data*. John Wiley & Sons.

Lewis, P. A. and Shedler, G. S. (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413.

Martell, D. L. (2007). Forest fire management. In *Handbook Of Operations Research In Natural Resources*, pages 489–509. Springer.

Martell, D. L. and Sun, H. (2008). 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.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27.

OMNR (2004). Forest Fire Management Strategy for Ontario. Queen's Printer for Ontario, Toronto.

Parks, G. M. (1964). Development and application of a model for suppression of forest fires. *Management Science*, 10(4):760–766.

Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56(4):1016–1022.

Ritchie, M. W., Skinner, C. N., and Hamilton, T. A. (2007). Probability of tree survival after wildfire in an interior pine forest of northern california: effects of thinning and prescribed fire. *Forest Ecology and Management*, 247(1):200–208.

Rodriguez-Iturbe, I., Cox, D., and Isham, V. (1987). Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 410(1839):269–288.

Rowe, J. S. (1972). Forest regions of Canada. *Canadian Forest Service, National Capital Region*, (1300).

Senici, D., Chen, H. Y., Bergeron, Y., and Cyr, D. (2010). Spatiotemporal variations of fire frequency in central boreal forest. *Ecosystems*, 13(8):1227–1238.

Therneau, T. M. (2014). *A Package for Survival Analysis in S*. R package version 2.37-7.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling survival data: extending the Cox model*. Springer.

Therneau, T. M., Grambsch, P. M., and Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of computational and graphical statistics*, 12(1):156–175.

Tsai, B.-W., Harvey, J. T., and Monismith, C. L. (2003). Application of Weibull theory in prediction of asphalt concrete fatigue performance. *Transportation Research Record: Journal of the Transportation Research Board*, 1832(1):121–130.

Venables, W. N. and Ripley, B. D. (2002). *Modern applied statistics with S*. Springer.

Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press.

Woolford, D., Bellhouse, D., Braun, W., Dean, C., Martell, D., and Sun, J. (2011). A spatio-temporal model for people-caused forest fire occurrence in the romeo malette forest. *Journal of Environmental Statistics*, 2:2–16.

Woolford, D. G. and Braun, W. J. (2007). Convergent data sharpening for the identification and tracking of spatial temporal centers of lightning activity. *Environmetrics*, 18(5):461–479.

Wotton, B. M. (2009). Interpreting and using outputs from the Canadian Forest Fire Danger Rating System in research applications. *Environmental and Ecological Statistics*, 16(2):107–131.

Wotton, B. M. and Martell, D. L. (2005). A lightning fire occurrence model for Ontario. *Canadian Journal of Forest Research*, 35(6):1389–1401.

Wotton, B. M., Martell, D. L., and Logan, K. A. (2003). Climate change and people-caused forest fire occurrence in Ontario. *Climatic Change*, 60(3):275–295.

Curriculum Vitae

Name: Amy A. Morin

Post-Secondary Wilfrid Laurier University
Education and   Waterloo, Ontario, Canada
Degrees: 2007 - 2012 B.Sc. Mathematics

                    The University of Western Ontario
                    London, Ontario, Canada
                    2012 - 2014 M.Sc. Statistics

Related Work   Teaching Assistant, and Research Assistant
Experience: The University of Western Ontario
                    2012 - 2014