Statistical Models of Key Components of Wildfire Risk

Article *in* Annual Review of Statistics and Its Application · April 2017

DOI: 10.1146/annurev-statistics-031017-100450

Annual Review of Statistics and Its Application

Statistical Models of Key Components of Wildfire Risk

Dexen D.Z. Xi,[1] Stephen W. Taylor,[2] Douglas G. Woolford,[1] and C.B. Dean[3]

[1]Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario N6A 5B7, Canada; email: dxi@uwo.ca, dwoolfor@uwo.ca

[2]Pacific Forestry Centre, Natural Resources Canada, Victoria, British Columbia V8Z 1M5, Canada; email: steve.taylor@canada.ca

[3]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada; email: cdean@uwaterloo.ca

Annu. Rev. Stat. Appl. 2019. 6:197–222

First published as a Review in Advance on December 3, 2018

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-031017-100450

Keywords

fire duration, fire load, fire occurrence, joint modeling, wildfire management, wildfire science, fire survival, wildland fire

Abstract

Fire danger systems have evolved from qualitative indices, to process-driven deterministic models of fire behavior and growth, to data-driven stochastic models of fire occurrence and simulation systems. However, there has often been little overlap or connectivity in these frameworks, and validation has not been common in deterministic models. Yet, marked increases in annual fire costs, losses, and fatality costs over the past decade draw attention to the need for better understanding of fire risk to support fire management decision making through the use of science-backed, data-driven tools. Contemporary risk modeling systems provide a useful integrative framework. This article discusses a variety of important contributions for modeling fire risk components over recent decades, certain key fire characteristics that have been overlooked, and areas of recent research that may enhance risk models.

1. INTRODUCTION

The global average annual area burned due to wildfires was recently estimated (Giglio et al. 2013) to be approximately 3.48 million km^2 for the 1997–2011 period, about the area of India. Wildfire characteristics such as the number of fires, their size, their severity, the season during which they occur, and the annual area burned in a region vary considerably with climate, vegetation, topographic controls, and human influence at local (Heyerdahl et al. 2007), regional (Parks et al. 2012), and global (Krawchuk et al. 2009b) scales. Fires are moderately rare events at a daily scale. For example, while about 5,000 fires occur in Canada every year, this translates to a background rate of approximately one new fire per ten million hectares per day during an approximately five-month fire season. However, this may be punctuated by surges in the number of fire ignitions associated with high pressure systems or lightning storms at local or regional scales, resulting in many dozens to hundreds of fire starts being discovered within a few hours to days. Consumption of biomass, smoke emissions, and changes in land cover associated with vegetation fires have an important influence on global atmospheric chemistry, the global carbon budget, and energy balance, as well as the structure and function of affected ecosystems. As well, unwanted fires may also cause loss of life and property, impacts on air quality and human health, and loss of business revenue.

The field of fire science evolved over approximately 100 years from early descriptive studies (e.g., Plummer 1912) to the development of complex models of fire spread and other physical processes (e.g., Linn et al. 2007). Research has followed two streams: basic research to enhance understanding of wildfire as an Earth system and ecological process, and applied research to inform fire management decision making. Contemporary fire management organizations follow the four pillars of emergency management: prevention and mitigation, planning and preparedness, response, and recovery. Thus, to inform preparedness and response actions, wildfire managers would like to know, at a daily to weekly scale, how many fires will likely occur, whether and how fast they will spread, how intense and how large they will grow, and how long they will last. To inform prevention and mitigation activities, they would also like to know the long-term likelihood of a vegetated area burning. Because of the close connection between weather and fire, much early effort was devoted to the development of fire danger rating systems to predict fuel flammability, fire occurrence, and fire behavior in different vegetation types with changing weather conditions to support preparedness and suppression response decision making (Taylor & Alexander 2006, Hardy & Hardy 2007, Fujioka et al. 2008). An important historical development is that independent systems have been developed to portray fire danger in different countries; no single global fire danger system has emerged. Examples include the Canadian Fire Weather Index (FWI) System, a subsystem of the Canadian Forest Fire Danger Rating System; the National Fire Danger Rating System in the United States; and the McArthur Forest Fire Danger Index in Australia.

Because wildfire is a natural process that cannot be completely eliminated from some environments, even where unwanted, fire management is increasingly being recognized as a form of risk management. Natural hazard risk, the expected loss or impact arising from a natural event, is considered to have three components: hazard, vulnerability, and exposure (Cardona et al. 2012), which can be visualized as a risk triangle (Crichton 1999). It is important to note that this is not a statistical representation of risk but one that has been developed and utilized by the natural hazards community. We present it here because of its common usage in environmental science. Scott (2006) adapted this concept, defining the wildland fire risk triangle as including the three components: fire probability (i.e., the hazard or the risk of fire occurrence), fire behavior (i.e., the severity or potential behavior of a fire if it occurs), and fire effects (i.e., the exposure or potential impact of the fire). Statistical science has made many contributions to modeling some of the components of

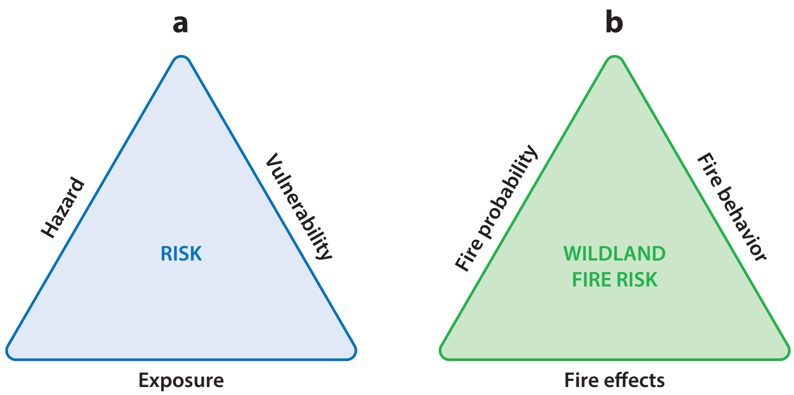

Figure 1

The risk triangle concept from the insurance and wildland fire perspectives. (*a*) The general risk triangle framework in insurance (adapted from Crichton 1999). (*b*) The risk triangle concept as it applies to assessing wildland fire risk (modified from Scott 2006). Risk in general, as well as in the context of wildland fires, can be viewed as having three connected components, as highlighted by the sides of the risk triangle.

this risk triangle, as we discuss later. **Figure 1** illustrates the concept of the general risk triangle and its adaptation to wildland fire risk.

Consequently, there has been increasing focus on the development of quantitative risk analysis methods (Miller & Ager 2013). Quantitative risk assessment to inform management decision making has its foundations in decision theory and utility theory (Morgan et al. 1992). **Figure 2** illustrates a number of factors that contribute to wildfire risk, including the likelihood and severity

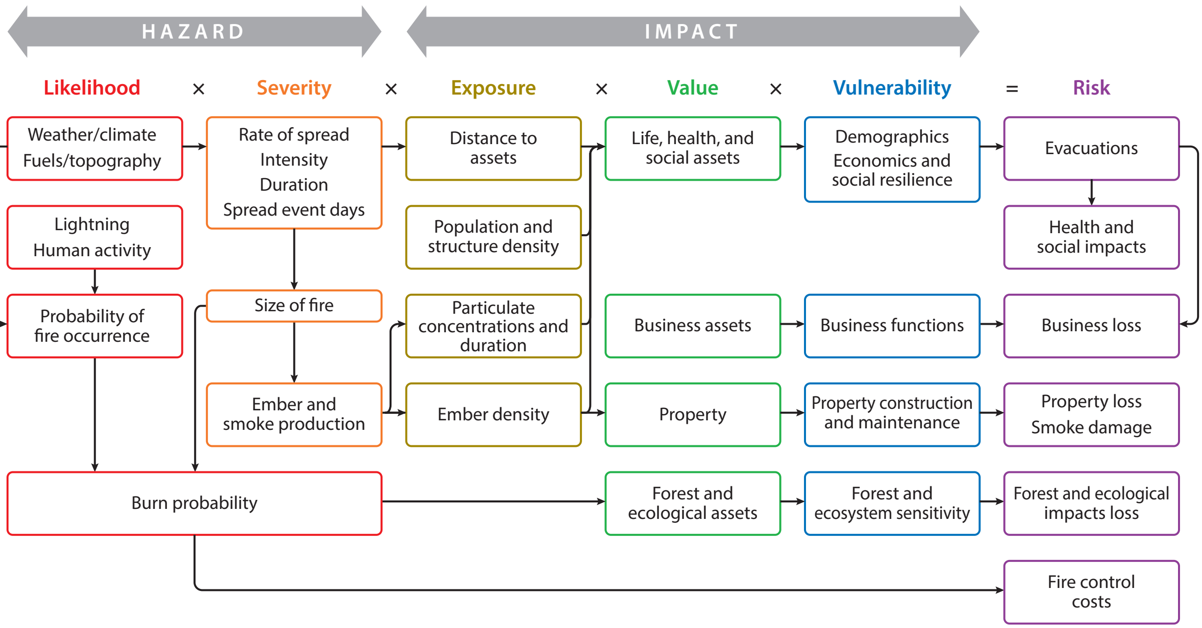

Figure 2

Some factors contributing to wildfire hazard and risk are estimated with various qualitative, deterministic, and stochastic models.

of fires in a region and the exposure, vulnerability, and value of valued assets. Finney (2005) defined wildfire risk as the expected change in net present value obtained from the aggregate losses and benefits in n values or assets over all N possible fire behaviors (under all weather conditions from all ignition locations):

$$\sum_{i=1}^{N} \sum_{j=1}^{n} p(F_i)[B_{ij} - L_{ij}],$$

where $p(F_i)$ is the probability of the ith fire behavior, and B_{ij} and L_{ij} are the benefits and losses resulting from the effects of the ith fire behavior on the jth asset type, respectively.

More recently, Papakosta et al. (2017) defined wildfire risk to the jth asset as

$$\int_{\substack{\text{Hazard}\\ \text{scenarios } h}} f_H(h) \int_{\substack{\text{Damage}\\ \text{scenarios } h}} f_{D_j|H}(d|h)\, C_j(d, h)\, \mathrm{d}d\, \mathrm{d}h,$$

where $f_{D_j|H}(d|h)$ is the conditional density of damage d given a wildfire event h (vulnerability), $C_j(d, h)$ is the cost associated with the damage by the wildfire, and $f_H(h)$ is the probability density of a wildfire event. In a further refinement of the model, the density $f_H(h)$ may be related to a certain fire severity characteristic, for example, the likelihood of a particular type of fire, fire intensity, duration, size, or incident complexity. These key characteristics represent positive continuous outcomes (or marks in the context of a point process; e.g., Daley & Vere-Jones 2003).

The precise definition of each fire severity characteristic may vary in different contexts. As an example, the characteristic may be the size being larger than some specific value, for instance, a fire being classified as a large fire as defined by Stocks et al. (2002). In this case, the risk on assets relates to risks specific to large fire events. Alternatively, a certain epoch of the fire's duration may be of interest (e.g., Morin et al. 2015). In that case, the risk on assets refers to the risk during this epoch of the fires. Note that this framework can be further extended by decomposing the fire characteristic into components, such as decomposing fire size into the probability of a large fire given fire occurrence and then modeling the size distribution for large fires. This approach was utilized in Preisler et al. (2011), where size was decomposed even further by coupling to a model for cost per acre in order to forecast future suppression costs.

Statistical science has an important role in modeling and quantifying uncertainties in the various components that make up wildland fire risk, through such conditional and marginal models, in order to better understand wildland fire science and inform wildland fire management (Preisler & Ager 2013, Taylor et al. 2013). The latter review also includes a thorough commentary on the history of statistical modeling of fire occurrence, starting with the pioneering work of Bruce (1963) and Cunningham & Martell (1973), followed by the seminal work of Brillinger et al. (2003) and Preisler et al. (2004) that led to substantial developments over the next decade.

This article reviews key statistical models that have been used to predict wildfire risk components, including some very recently developed novel modeling strategies. Section 2 reviews models for fire occurrence prediction, while Section 3 discusses deterministic models for fire spread, intensity, and growth. Section 4 presents models for fire duration, and Section 5 examines models for estimating fire size. Sections 2 through 5 discuss past research investigating each of those characteristics of fire regimes separately or through the conditional framework as discussed above. In Section 6, we introduce the use of joint modeling methods using duration and size as an illustration, which we believe has the potential for gaining further insight into fire behavior. In

Section 7, we turn to an alternative conditional framework for modeling fire hazard and its utilizations through computer simulations. We conclude with a discussion.

2. OCCURRENCE

2.1. A Point Processes Viewpoint

It is important to note that not all wildland fire ignitions may appear in fire management agency records (Taylor et al. 2013). Ignitions that lead to sustained fire spread may be detected by wildland fire management agencies (e.g., aerial detection or stationary towers), by the public, or by satellite-borne sensors. Detected fires that are subsequently reported and then recorded by a fire management agency are referred to as fire occurrences. Observed patterns in fire occurrences can be viewed as realizations of a spatio-temporal point process. Examples of applying methodology from the point process literature include Podur et al. (2003), Wang & Anderson (2011), and Turner (2009). The spatio-temporal point process underlying the generation of fire occurrence is denoted $N(x, y, t)$, where x and y are location variables and t is time. Since the rate of wildfire occurrences depends on environmental conditions favorable for ignition, the presence of an external ignition source, and detection capability, the point process can be assumed to have an inhomogeneous conditional intensity function λ that depends on parameter $\theta = \theta(\boldsymbol{z})$, where $\boldsymbol{z}$ is a vector of such predictors. The log-likelihood of the spatio-temporal point process is

$$L(\theta) = \int_0^T \int_x \int_y \log[\lambda(x, y, t \,|\theta)]\, \mathrm{d}N(x,y,t) - \int_0^T \int_x \int_y \log[\lambda(x, y, t|\theta)]\, \mathrm{d}x\mathrm{d}y\mathrm{d}t.$$

A discretized approach to approximating this likelihood has been the preferred framework for modeling fire occurrences with the underlying conditional intensity function approximated by a Bernoulli probability of a fire occurrence; the response and covariates are recorded on a set of discrete space-time voxels, chosen to be at a fine enough scale so that the counts of fire occurrence are reduced to presence/absence of a fire occurrence in any given voxel. A common scale for dividing space-time is 1 km × 1 km by daily cells (voxels). Dynamic covariates, such as weather and fire weather indices, or lightning counts for lightning-caused fires, are interpolated to the centroid of that voxel. Static covariates, including measures of key predictors related to human-caused fires, such as measures related to roads, railways, or population density, are integrated over each voxel. For more details on the connection between the underlying spatio-temporal point process likelihood function and discretized approximations in the context of modeling fire occurrence and an example of such a model, readers are directed to Brillinger et al. (2003) or the review discussion in Taylor et al. (2013).

2.2. Logistic Models as a Discretized Approach to Estimate a Point Process

Using the discretized approach, the most widely employed method for modeling fire occurrence appears to be logistic regression or related extensions such as logistic generalized additive models (GAMs); sometimes models with random effects are also considered. Separate models, stratified by the cause of the fire, are commonly developed due to differences between the underlying processes generating the different types of ignitions. For example, different types of ignition sources can lead to different lag periods between the ignition of a fire and its eventual arrival to a fire management agency as a reported wildfire. This was reflected in the set of models characterizing the probabilities of ignition and eventual arrival (i.e., occurrence) of lightning fires in Ontario as developed by Wotton & Martell (2005). Lightning ignitions and their subsequent arrivals as reported forest fires are modeled separately. Then, the probability of a lightning strike igniting a fire at time s and that fire being reported at time $s+t$ is estimated by

$$\begin{aligned}&\mathrm{P}(\text{lightning strike at time } s \text{ leads to a lightning fire occurrence at time } s+t) = \\ &\mathrm{P}(\text{occurrence at time } s+t \mid \text{ ignition at time } s)\ \ \mathrm{P}(\text{ignition at time } s).\end{aligned}$$

There are also commonly highly nonlinear relationships between the probability of fire occurrence and other predictors, such as seasonality or spatial effects. These nonlinear relationships on the log-odds scale are commonly modeled by spline-based smoothers using logistic GAMs. Wood (2006), for example, provides a general discussion of GAMs, and Preisler & Ager (2013) provide a high-level overview of GAMs in the fire occurrence context. Prior to the introduction of GAMs, nonlinear seasonality components were modeled using periodic functions (e.g., Martell et al. 1989).

Let $Y_i, i = 1, \ldots, n$ be a set of random variables representing an indicator for fire occurrence ($\text{Yes} = 1, \text{No} = 0$) in voxel i, assumed to be independently distributed as Bernoulli(p_i), conditional on observed covariates. Here, $p_i = \mathrm{P}(Y_i = 1|\boldsymbol{x_i})$ where $\boldsymbol{x}_i$ denotes a vector of covariates for the ith voxel. We may model p_i through

$$\text{logit}(p_i) = \beta_0 + \sum_{p=1}^{P} g_p(x_{ip}),$$

where logit $p_i = \log(\frac{p_i}{1-p_i})$; β_0 is an intercept; $x_{ip},\ p = 1, \ldots, P$ are covariates; and g_p are corresponding zero-mean smoothers of these covariates. The terms in the model may include multidimensional smoothers [e.g., $g(\boldsymbol{s}, \boldsymbol{t})$, where $\boldsymbol{s_i} = (s_{i1},\ s_{i2})$ represents the location and $\boldsymbol{t_i} = (t_{i1}, t_{i2})$ represents day of year and year] to model spatial and temporal effects, where the latter allows for trends both within (e.g., seasonality) and across (e.g., climate change or other trends) years, as well as other nonlinear and/or linear effects of other key predictors, such as measures of fuel moisture and human-land use characteristics.

As noted by Woolford et al. (2011), the volume of data can present computational difficulties when modeling on a fine spatio-temporal scale such as the discretized space-time voxel approach as outlined previously. For example, in their case study of the Romeo Malette Forest in Ontario, Canada, Woolford et al. (2011) noted that discretizing the data to a set of 1 km $\times$ 1 km $\times$ daily voxels led to nearly 90 million records. For larger-scale studies, such as developing provincial or national modeling frameworks, this problem compounds immensely. Interestingly, however, the solution to this problem lies at the heart of one of the underlying dogmas of statistical inference: Rather than trying to fit a model to all data, a representative sample is used for model fitting. Since fires are an extremely rare event on any fine space-time scale, a response-dependent sampling scheme is commonly employed, where all the voxels with fire occurrences are kept, along with a simple random sample of the nonfire voxels.

From a decision-support point of view, a key contribution of spatio-temporal fire occurrence modeling is that it produces a relative occurrence probability map where cells with higher probability of fire occurrence are identified, which can aid decision support, such as aerial detection routing. The expected number of fire occurrences in a given region on a given day can then be estimated. We also note that it is common to achieve greater specificity (correct identification of cells without fire occurrences) than sensitivity (correct prediction of cells with fire occurrences) because of the stochastic nature of the ignition process and because an overwhelming number of voxels refer to nonfire day and areas. The mathematical details of this framework are discussed in depth by Brillinger et al. (2003) and Taylor et al. (2013). The latter also summarizes the connections between this technique and logistic retrospective case-control studies.

2.3. Changes in Fire Occurrence

Whether and where fire occurrence is changing with climate is of considerable interest to fire managers. For example, fire occurrence has been shown to be increasing under a warming climate in the western United States (e.g., Westerling et al. 2006). Observed increases in fire occurrence have been found to be associated with anomalies in fire weather indices (Woolford et al. 2014). The fire season as measured by fire occurrence probability has been getting longer (Woolford et al. 2010, Albert-Green et al. 2013), and, based on results from studies using global climate model data under various scenarios, fire occurrence probability is predicted to increase under a warming climate, (e.g., Krawchuk et al. 2009a,b; Wotton et al. 2003, 2010).

As commented on by Taylor et al. (2013), difficulties with historical analyses (e.g., Woolford et al. 2010, Albert-Green et al. 2013) arise because fire detection system effectiveness can change over time, leading to potential confounding with any climate change effect. Woolford et al. (2010) found that the median size at detection for lightning-caused fires had been decreasing over the 42 years of their study, suggesting that the detection system may have become more effective over time, under the assumption that a fire would continue to grow after ignition.

Woolford et al. (2010) noticed three dominant characteristics in the lightning-caused fire occurrence records they analyzed, namely, regular seasonal patterns (as commonly quantified in other fire occurrence work such as Martell et al. 1989; Brillinger et al. 2003; Preisler et al. 2004; Woolford et al. 2009, 2011); large deviations from these patterns, including zero-heavy behavior where no fires were observed even though fires were typically observed around such a period; and extreme behavior where many more fires were observed than what is typical.

The mixture framework of Woolford et al. (2014) identified significant increases to the lightning-caused fire occurrence probability that were associated with temperature and fire-weather index anomalies. Their study monitored long-term trends in a set of historical fire records for the period 1963–2009 for a region of northwestern Ontario, Canada, using a three-component mixture of logistic GAMs with the component densities representing seasonal, zero-heavy, and extreme behavior as discussed above. They noted that potential confounders, such as improved wildland fire detection systems, make it difficult to tease out climate change trends and that longer than a half-century of historical records is required to have strong confidence in correctly concluding significance in such a study monitoring temporal changes to fire occurrence. They also noted that determining power to detect trends in fire occurrence probability as a function of the number of years in the historical records was a "key, yet commonly overlooked, point in many quantitative scientific investigations of trends that may be related to climate change" (Woolford et al. 2014, p. 407). This power evaluation offers an opportunity to consider how many years of records are required to detect changes in environmental effects with high confidence.

Regardless of the underlying model and study, goodness-of-fit checking is a key step in any model-building framework. Typically, the goodness-of-fit logistic-based models can be assessed by comparing observed counts versus those expected under the model where such counts are aggregated on various scales. Ideally, cross-validation (Wood 2006, James et al. 2013) is also used

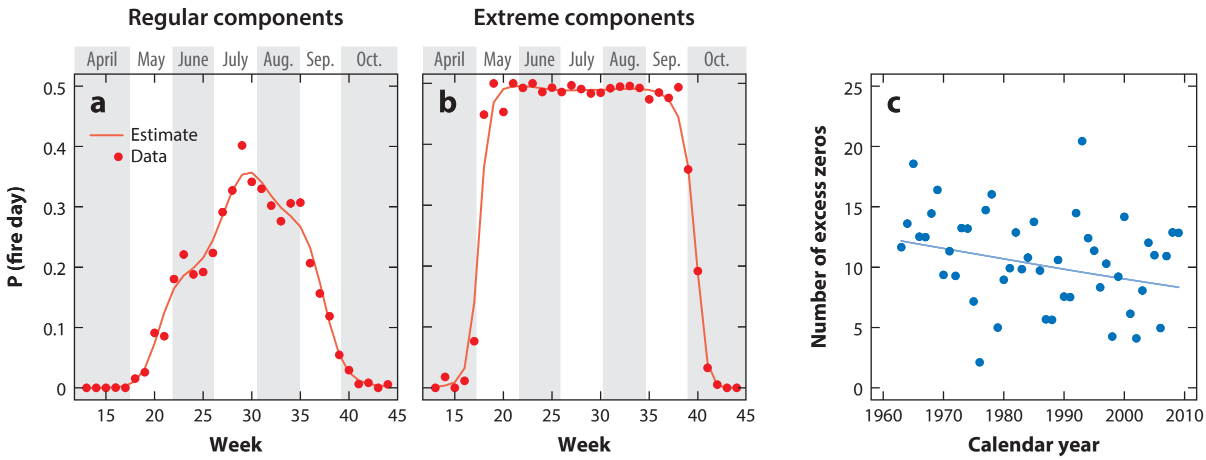

Figure 3

Panels *a* and *b* plot the estimated component-specific fire occurrence probability curves (*red lines*) for the regular and extreme components, respectively. A fire day refers to a day when one or more wildland fires are reported. Overlaid on each of these curves are the observed empirical weighted proportion of the number of fire days per week over all years (*red dots*), where the observed data were weighted by the posterior probabilities of membership for the corresponding regular or extreme component. Panel *c* compares observed and expected frequencies of excess zeros. The light blue line is the expected number of zeros from the zero-heavy component plotted versus year. The blue circles are the empirical number of excess zeros: the number of observed zeros minus the number of zeros expected to arise from both the regular and extreme seasonal components. Adapted with permission from Woolford et al. (2014).

to assess predictive accuracy. Vilar et al. (2010) provide simple examples of this in the context of fine-scale spatio-temporal fire occurrence prediction. However, assessing goodness of fit in the mixture modeling framework is more complicated. Woolford et al. (2014) examined goodness of fit by comparing observed versus expected counts for each subcomponent of their mixture model. This is illustrated in **Figure 3**. Such an assessment approach offers the opportunity to determine which of the subcomponents are not appropriate and could find wide application in other mixture modeling frameworks.

2.4. The Two Cultures of Fire Occurrence Prediction Modeling: Statistical Modeling Versus Algorithmic Methods

Breiman (2001) noted that the objective of a statistical analysis is to use data to make inferences, observing that the two dominant cultures for doing so are "statistical" and "algorithmic," where the latter focuses on finding a function to predict the response as a function of other variables without assuming a specific stochastic model. The preceding subsections have focused on summarizing key developments from a statistical modeling standpoint. However, algorithmic modeling approaches have also been used in the context of fire occurrence prediction.

For example, Vega-Garcia et al. (1996) developed artificial neural networks for wildfire occurrence and compared them to those of Vega-Garcia et al. (1995), who had utilized logistic regression models to analyze the data. This early study of fire occurrence using algorithm methods found reasonable predictive accuracy with neural nets; it correctly predicted 85% of the nonfire days and 78% of the fire days. They also commented that the improvement in predictions over traditional logistic regression modeling results were "not as dramatic as it has been in other applications" (Vega-Garcia et al. 1996, p. 14) that compared neural nets to logistic regression methods. The total percentage correctly predicted by the neural net model was found to improve by only

2% when compared with the logistic regression model for the same independent validation data set. They postulated that this could be due to the limited amount of data used in these studies (only five fire seasons). Ongoing research for large fire prediction in Canada (e.g., Nadeem et al. 2016) is exploring extensions to that work using lasso logistic regression, as well as algorithmic approaches, such as random forests. Hastie et al. (2001) and James et al. (2013) provide details on lasso and random forests.

3. FIRE SPREAD, INTENSITY, AND GROWTH

Once ignited, a fire will continue to spread from fuel particle to particle as a self-sustaining process as long as the heat produced by combustion is sufficient to heat the adjacent particles to the ignition temperature, or the fuel is exhausted. The rate of fire spread is influenced by fuel properties, particularly the moisture content and temperature, and the ambient atmospheric conditions, particularly wind speed. Over the past decades, several dozen mathematical fire spread models have been developed using approaches varying from simple empirically based nonlinear regressions to detailed computational fluid dynamics models (Sullivan 2009a,b). Fire intensity, the amount of energy released per unit length of fire front, is usually modeled as a function of fire spread rate, fuel consumption, and heat content of the fuel. A number of empirical models have been developed (assuming local homogeneity) to model the two-dimensional spread of the fire perimeter through heterogeneous fuel and topographic conditions at landscape scales (<1–100 km) (Sullivan 2009c). The study area and time period of interest represent a set of voxels, with vegetation, topographic and other geographic covariates varying between points in a two-dimensional grid, and weather and other covariates varying spatially across the grid and temporally for each time step in the period. Early models represented fire spread as cellular automata, where a cell along a fire perimeter composed of grid cells could ignite adjacent grid cells in a time step, depending on vegetation, topographic, and weather covariates in the adjacent cell. Higher-resolution models simulate fire spread as a wave process, projecting the angular velocity of a number of discrete points around the fire perimeter as a vector over a discrete time step, depending on vegetation, topographic and weather covariates, but where the spread distance and directions are unconstrained by the grid resolution. The fire perimeter after each time step is remapped as the convex hull of the new points. However, although fire prediction is inherently probabilistic (because of the difficulty in accurately representing fuel properties and assessing and predicting atmospheric conditions; Taylor et al. 2013), most fire spread models are deterministic. Recently several authors have used ensemble methods to introduce stochasticity to fire spread (Cruz 2010) and fire growth models (Braun & Woolford 2013, McLoughlin & Gibos 2016, Pinto et al. 2016) to better represent uncertainty.

Statistical models may provide important alternative risk measures or adjuncts to deterministic models. Noting that large, intense fires are rare events, Hernandez et al. (2015) fitted generalized extreme value (GEV) distributions to remote-sensing based observations of fire intensity (fire radiative power) and size for fires in Portugal and used a nearest neighbor procedure to estimate the parameters of the distribution from meteorological covariates. They suggested that this approach provides an important estimate of uncertainty beyond qualitative fire danger indices. Price et al. (2015) fitted binomial regression models of large fire spread distances through cells with varying fuel conditions and weather conditions for 677 large fires in the Sydney region of Australia. They used the models to estimate the likely spread distance and the probability that a fire starting from the 677 ignition points would reach one of 26,000 3.4-hectare receiver points in the study area. The heuristic of modeling potential fire spread from an ignition to a receiver point of interest provides a simpler alternative to explicitly modeling fire spread from all points on the fire perimeter, which is computationally demanding.

4. DURATION

Although the probability of fire occurrence has been well studied using statistical models for over half of a century, quantifying the survival distribution of fires during its containment for management purposes has not received much attention until more recently. Finney et al. (2009) categorized stages of containment of a fire into spreading intervals based on fire occurrences during 2001 to 2005 in the United States. The number of days in each interval were modeled using generalized linear mixed models using the same framework as for a repeated measures problem. Other quantitative studies also have been carried out for studying fire containment elsewhere with various foci in their statistical methods, such as for Italy (Marchi et al. 2014), Spain (Costafreda-Aumedes et al. 2015), Canada (Xiong 2015), Mediterranean Europe (DaCamara et al. 2014), and Portugal (Fernandes et al. 2016). The latter three modeled duration directly, while the latter two considered the survival probabilities of duration. The former two studies used analysis of variance and regression trees, respectively. However, integrating these models into a unique framework incorporating fire occurrence models to describe and predict the complete dynamics of wildfires remains a challenge.

The duration of a fire may also be modeled directly as a survival outcome using statistical survival models. These models exist commonly in industrial and medical research; they work in similar ways as logistic regression under fire occurrence modeling but differ in that the quantity of interest is the survivor function or the hazard function of an outcome (typically a time quantity, obtained by measuring from an origin to an event). Since the survivor and hazard functions represent, respectively, the probability that the individual can survive more than a certain time and the instantaneous rate of death at a certain time given survival up to that time, they seem well-suited to describing the dynamics of wildfires.

Let $t_i,\ i = 1, \ldots, n$ be the duration of fire i, assumed to be independent and identically distributed. We may model t_i through a log-location-scale model, sometimes referred to as the accelerated failure time (AFT) model:

$$\log(t_i) = \mu + \boldsymbol{\beta}^T \boldsymbol{x}_i + \sigma \varepsilon_i,$$

where μ and σ are the location and scale parameters, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{iP})^T$ is a vector of P covariates and $\boldsymbol{\beta}^T = (\beta_1, \ldots, \beta_P)$ are the corresponding coefficients, and ε_i are random errors. The survivor function of the outcome is

$$S(t_i \mid \boldsymbol{x}_i) = S_0\left(\frac{\log(t_i) - (\mu + \boldsymbol{x}_i^T \boldsymbol{\beta})}{\sigma}\right),$$

where $S_0(t)$ is the survivor function of the random error, termed the baseline survivor function. Since fire duration tends to be heavily right-skewed, the survivor functions may be modeled through parametric right-skewed distributions. Three common such baseline survivor functions are standard Gumbel, standard normal, and standard logistic, which correspond respectively to Weibull, log-normal, and log-logistic distributions for the outcome. As the covariate effects are multiplicative on time, the model assumes that different covariate values will scale the time axis of the survivor function. Covariates that are exogenous environmental variables (e.g., wind speed and temperature) then serve as stress factors that accelerate/decelerate the time to containment of a fire but keep the shape of the survivor function the same.

The hazard function may also be modeled directly using a Cox proportional hazards (PH) model:

$$h(t_i \mid \boldsymbol{x}_i) = h_0(t_i) \exp\left(\boldsymbol{\beta}^T \boldsymbol{x}_i\right),$$

where $h(t)$ is the hazard function of the outcome; $h_0(t)$ is the baseline hazard function corresponding to the hazard when $\boldsymbol{x}_i = 0$, which can be either parametrically specified or unspecified to capture potential irregular features. In the PH framework, the covariate effects act multiplicatively on the baseline hazard rate. The model assumes that fires with different covariate values will result in hazard functions that are proportional to each other. This modeling strategy particularly lends itself to covariates such as endogenous fire characteristic variables (e.g., initial size and drought indices), where the intrinsic tendency of burning is different for fires with different values of these covariates. In particular, fires with large initial size occurring during drought conditions will have less steep hazard curves.

Recently, Morin et al. (2015) used survival techniques to model the duration of forest fires in Ontario's intensive fire management zone using data on more than 18,000 fires recorded during 1989 through 2004. Fire management zones are partitions of a study region that are assumed to be approximately internally homogeneous with respect to ecological characteristics such as fuel, weather, topography, and fire management strategy, and so may have a similar range or pattern of fire characteristics including size, duration, intensity, frequency, and season (Morin et al. 2015). They restricted their analysis to a period up to 2004 due to a change to Ontario's fire management strategy that led to a change in the number and location of fire management zones in the province after 2004. Response time, initial size, and several other FWI System indices were considered as covariates. The duration of each fire was defined to be the time interval from the start of initial attack to the time that a fire was declared as being under control, measured in hours. To capture changes in shapes of the hazard that were observed in nonparametric estimates of the survivor function, and to ensure that the requirement for proportional covariate effects was not violated, a nonparametric stratified PH model was used to model survival times of lightning-caused fires. Their work appears to be the first of its kind to model duration on a fine timescale using a stratified PH model, demonstrating that survival models that include covariate effects, such as the PH model, can be used as building blocks for more complicated structures in wildfire modeling.

Within fire management zones, the durations of fires within the same zone are dependent. An important extension of univariate regression type models to account for such dependence is the inclusion of a shared random effect z_i to explain the variation in homogenous space polygon $i = 1, \ldots, n$ for fires $k = 1, \ldots, n_i$ occurring in that polygon. A typical modeling framework in survival models is

$$S\left(t_{ik} \mid \boldsymbol{x}_{ik},\ z_i\right) = S(t_{ik} \mid \boldsymbol{x}_{ik})^{z_i},$$

or equivalently,

$$h\left(t_{ik} \mid \boldsymbol{x}_{ik}, z_i\right) = z_i h\left(t_{ik} \mid \boldsymbol{x}_{ik}\right),$$

where $S(t_{ik} \mid \boldsymbol{x}_{ik},\ z_i)$ is the conditional survivor function for fire k of polygon i with covariate vector $\boldsymbol{x}_{ik} = (x_{ik1}, \ldots, x_{ikP})^T$ and $h(t_{ik} \mid \boldsymbol{x}_{ik}, z_i)$ is the conditional hazard function. The term z_i is commonly referred to as a shared frailty, and the framework is referred as a shared frailty model. The frailty extension of the PH model has been discussed in many texts because of the popularity of PH frailty models in medical studies (Hougaard 2000, Therneau & Grambsch 2000, Duchateau & Janssen 2008, Wienke 2010). Using the fact that $h(t) = -\frac{d}{dt}\log S(t)$ and the formulation of the PH model, the above expression leads to

$$h\left(t_{ik} \mid \boldsymbol{x}_{ik}, z_i\right) = z_i h_0\left(t_{ik}\right) \exp\left(\boldsymbol{\beta}_k^T \boldsymbol{x}_{ik}\right) = h_0\left(t_{ik}\right) \exp\left(\boldsymbol{\beta}_k^T \boldsymbol{x}_{ik} + b_i\right),$$

where the term $b_i = \log(z_i)$ can be interpreted as a latent covariate.

In her study of the lifetimes of forest fires in Ontario, Morin (2014) developed a set of PH frailty models to explore and quantify spatial differences in duration across a set of fire management compartments (FMC). The FMC partition was developed by Martell & Sun (2008). Morin (2014) found that a Gaussian frailty term b_i, representing an FMC effect, had an estimated variance significantly different from zero for lightning-caused fires, which is evidence in favor of a positive dependence between the durations of fires in the same FMC. Mapping posterior estimates of the frailties showed that the western region of Ontario experiences lightning-caused fires with shorter survival times (**Figure 4**).

It is worth noticing that the AFT model cannot be easily extended to a shared frailty model of the form described earlier. For example, if the outcomes follow a Weibull distribution, with location parameters $\lambda_k = \exp(-\frac{\mu_k + \boldsymbol{x}_{ik}^T \boldsymbol{\beta}_k}{\sigma_k})$ and scale parameters $\upsilon_k = \frac{1}{\sigma_k}$, including the term b_i as

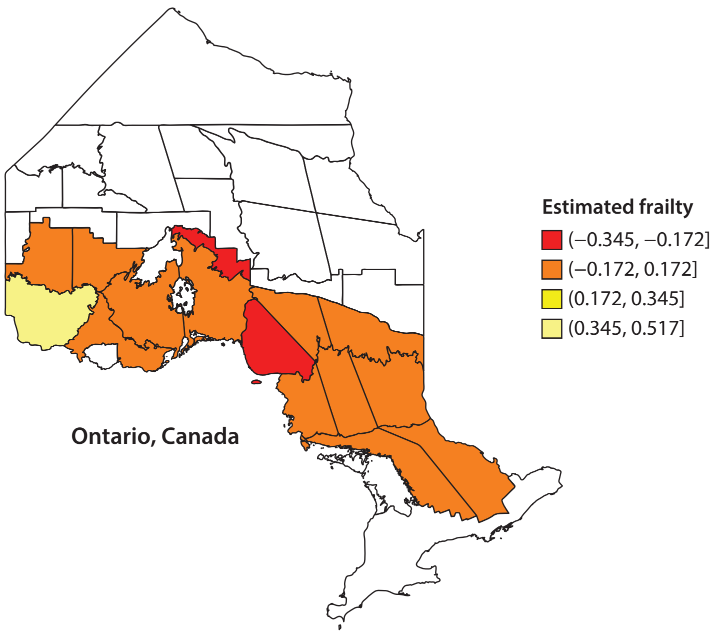

Figure 4

Choropleth map of the frailty terms for lightning-caused fires in the former intensive fire management zone of Ontario, Canada, which was partitioned into a set of fire management compartments (FMCs). Each FMC polygon was assigned a heat map color based on the estimate of the latent effect of the FMC (i.e., the frailty). FMCs that are outside of the study region are white. Exponentiated values of posterior frailty estimates can be viewed as multiplicative factors on the hazard function of fire lifetimes. Negative estimates imply an increase in survival probability via a reduction in hazard rate. Adapted with permission from Morin (2014).

an additive latent covariate yields

$$h\left(t_{ik} \mid \boldsymbol{x}_{ik}, b_i\right) = \lambda_k \upsilon_k t^{\upsilon_k - 1} = \exp\left(-\frac{\mu_k + \boldsymbol{x}_{ik}^T \boldsymbol{\beta_k} + b_i}{\sigma_k}\right) \frac{1}{\sigma_k} t^{\frac{1}{\sigma_k} - 1} = \exp\left(-\frac{b_i}{\sigma_k}\right) h\left(t_{ik} \mid \boldsymbol{x}_{ik}\right).$$

However, the AFT model may be an appropriate alternative when modeling long fires with environmental variables (i.e., having smooth survivor functions and exogenous covariates). We come back to this formulation in Section 6.

The frailty model and the copula model serve as important pieces of the foundation of modeling multivariate survival outcomes. Here, we briefly note the connection between shared frailty models and Archimedean copulas. For simplicity, we model without covariates. Shared frailty models assume that, conditional on z_i, t_{ik} are independent across all fires within the same polygon; thus, the joint survivor function conditioning on z_i is

$$\begin{aligned}
&S\left(t_{i1}, \ldots, t_{in_i} \mid z_i\right) = S\left(t_{i1} \mid z_i\right) \ldots S\left(t_{in_i} \mid z_i\right) \\
&= \exp\{-H(t_{i1} \mid z_i)\} \cdots \exp\left\{-H\left(t_{in_i} \mid z_i\right)\right\} \\
&= \exp\{-z_i H(t_{i1})\} \cdots \exp\left\{-z_i H\left(t_{in_i}\right)\right\} \\
&= \exp\left\{-z_i \left[H(t_{i1}) + \cdots + H\left(t_{in_i}\right)\right]\right\},
\end{aligned}$$

where $H(t) = -\log S(t)$ is the cumulative hazard function. Taking the expectation of the right-hand side of the expression above over z_i, the joint survivor function, yields

$$\begin{aligned}
&S\left(t_{i1}, \ldots, t_{in_i}\right) = E\, S\left(t_{i1}, \ldots, t_{in_i}\right) \\
&= E \exp\{-z_i \left[H(t_{i1}) + \cdots + H(t_{i2})\right]\} \\
&= \mathcal{L}\left[H(t_{i1}) + \cdots + H(t_{i2})\right],
\end{aligned}$$

where $\mathcal{L}(a) = Ee^{-xa}$ is the Laplace transformation of the random variable x. Using the fact that $S(t) = e^{-H(t)} = Ee^{-zH(t)} = \mathcal{L}H(t)$, we have

$$S\left(t_{i1}, \ldots, t_{i2}\right) = \mathcal{L}\left[H(t_{i1}) + \cdots + H(t_{i2})\right] = \mathcal{L}\left[\mathcal{L}^{-1}\, S\left(t_{i1}\right) + \cdots + \mathcal{L}^{-1}\, S\left(t_{i2}\right)\right],$$

which yields the so-called Archimedean copula family (Nelsen 2006, Liu 2012, Joe 2014). Embrechts & Hofert (2014) provide a detailed overview of the connections between the two frameworks, as well as the development of copulas in a quantitative risk management perspective.

5. SIZE

The increase in fire size during the life of a fire and its ultimate size at extinction are important to the difficulty of control, and fire size also has many other impacts of scientific and economic concern. Qualitative publications based on physical/process models derived in the natural sciences historically dominated the study of this phenomenon; however, quantitative studies based on empirical/statistical models have been appearing in the literature at an increasing rate since approximately the late 1990s (Cui & Perera 2008). Early empirical models assumed a power-law (i.e., Pareto) distribution for wildfire sizes:

$$f_X\left(x; b\right) \propto x^{-b},$$

where X is the random variable representing fire size, and its density function $f_X(x; b)$ depends on a parameter b. An example using this approach was presented by Schoenberg et al. (2003), who considered several parametric models for the distribution of wildfire sizes in Los Angeles County, California. Using visual diagnostics and nonparametric tests for comparing distributions, they advocated for the use of a tapered Pareto distribution for modeling size distributions in that area. Cumming (2001) modeled the survivor function of the size of fires in the province of Alberta, Canada, using a right-truncated exponential distribution under the assumption that there was a maximum size a fire could grow to, based on characteristics of the study area. Recent models for fire size include environmental variables. Butry et al. (2008) incorporated environmental variables using linear regression for modeling the logarithm of the size of large fires in northeast Florida from 1981–2001. Chen et al. (2014) used quantile regression to study the effect of precipitation on fires in southwestern China. A comprehensive review of fire size models appears in Cui & Perera (2008).

Here, we review two key threads of research in the development of statistical methods for modeling fire size. Power-law behaviors are commonly observed in nature. If fire growth follows a preferential attachment or Yule process (Gibrat's Law), the distribution of randomly killed states (or states observed once) under stochastic processes follows a power law in one or both tails (Reed 2001, Reed & Hughes 2002). Using percolation theory, Reed (1999) observed that a piecewise probability distribution, partitioned at the percolation threshold, fits the distribution of forest fire size reasonably well. Reed & McKelvey (2002) derived the density function and survivor function of the killed state. Let the fire size at time t be $X(t) = \exp(\mu t)$ and the growth rate at size X be $\mu(X) = \mu X$, proportional to size by a constant, μ. We further assume that the killing rate $k(t)$ takes the form of

$$k(t) = \lim_{dt \to 0} \frac{P(T < t + \mathrm{d}t \mid T > t)}{\mathrm{d}t} = v(X(t)),$$

where $v(x)$ is a nonincreasing function referred to as the extinguishment rate. Let $\bar{X}$ denote the killed state, and then it can be shown that the density function of $\bar{X}$ is

$$f_{\bar{X}}(x) = \rho(x) \exp\left(- \int_{x_0}^{x} \rho\,(x')\,\mathrm{d}x' \right),$$

where $\rho(x) = v(x)/\mu(x)$ is the hazard rate function. The survival function of $\bar{X}$ is

$$S_{\bar{X}}(x) = \exp\left(- \int_{x_0}^{x} \rho\,(x')\,\mathrm{d}x' \right).$$

The plot of empirical $\log S_{\bar{X}}(x)$ against $\log(x)$ demonstrates a linear trend if the data exhibit power-law behavior. The authors suggest plotting the extinguishment growth-rate ratio (EGRR) against $\log x$, which is expressed as

$$\mathrm{EGRR} = R(x) = \frac{x v(x)}{\mu(x)} = \frac{x f_{\bar{X}}(x)}{S_{\bar{X}}(x)}.$$

Power-law behavior will occur on an interval over which EGRR is constant. Conditions when EGRR is not constant that may lead to thin- or thick-tailed distributions include the following:

(a) In regions where the fire season is limited to a portion of the year (e.g., by winter), the distribution of killing times for fires starting later in the season may be right truncated; this also applies for regions where fire size may be limited by available fuel. (b) In a managed environment where all fires are suppressed but where occasional extreme weather such as Santa Ana winds favors large fire growth (Moritz 1997), distributions may be thick tailed; this may also occur when climate over a long sampling period is nonstationary. When power-law behavior does not occur, the authors recommended using non-power-law distributions such as a 3-parameter Weibull distribution for certain cases. The theoretical foundation of the work above connects well with methods developed to model the distribution of size in other fields of science (Reed & Hughes 2002, 2004; Reed & Jorgensen 2004; Reed 2011, 2012).

Another thread of research in modeling fire size has developed in engineering. To examine extreme fire size, Holmes et al. (2008) utilized GEV methods for analyzing fire sizes with heavy-tail distributions. GEV methods play important roles in engineering and actuarial science because of their focus on rare but extreme events (Castillo 2012, Longin 2016). Foss et al. (2011) provide a probabilistic perspective of heavy-tailed distributions. Intuitively, the distribution of a random variable Y is said to be heavy-tailed if, for any $u > 0,\ v > 0$,

$$\lim_{u\to\infty} \mathrm{P}(Y > u + v \mid Y > u) = \lim_{u\to\infty} \frac{S(u+v)}{S(u)} = 1.$$

That is, if the observation already exceeds a large value u, then it will likely exceed a larger value $u + v$. The maximum value of a sample of observations is traditionally used for parameter estimation under GEV methods. To overcome the limitation of information loss, the authors used all the observations beyond a threshold (e.g., size >200 ha) instead. Thus, the survivor function of the observations beyond a threshold is

$$S(u + y_i \mid u) = \mathrm{P}(Y_i > u + y_i \mid Y_i > u) = \frac{S(u+y_i)}{S(u)},$$

where Y_i is the size and u is the threshold. It follows that the resulting distribution function follows a generalized Pareto distribution (Davison & Huser 2015):

$$f(y_i \mid \mu, \xi,\ \sigma_i) = \frac{1}{\sigma_i}(1 + \xi\,\frac{(y_i - \mu)}{\sigma_i})^{-(1+\frac{1}{\xi})},$$

where μ, σ_i, and ξ are the location, scale, and shape parameters. Covariates $\boldsymbol{z}_i$ can be included through $\sigma_i = \sigma(\boldsymbol{z_i}) = \mu + \boldsymbol{\beta}^T \boldsymbol{z}_i$ to model and simulate fire sizes given environmental variables. The model leads to a set of integrated frameworks (e.g., Preisler & Westerling 2007; Westerling & Bryant 2008; Preisler et al. 2011; Westerling et al. 2011a,b; Bryant & Westerling 2014) that can be used to understand, for example, the impact of climate change and human development on fire-related losses in different regions.

6. MODELING DURATION AND SIZE AS JOINT OUTCOMES

As mentioned in the section on fire occurrence modeling, it is possible to combine models for fire occurrence with other models, such as those for duration, those to model fire load (e.g., Morin 2014), or those for fire size and cost distributions, to develop spatially explicit forecasts for suppression costs (e.g., Preisler et al. 2011). These frameworks commonly decompose the problem through a multi-stage approach, developing separate, independent models for each component as building blocks for the overall model, such as an occurrence model coupled to an independent survival model (Morin 2014), or coupling occurrence models to independent models for fire size and cost distributions (Preisler et al. 2011). However, components may be linked.

For example, marked point process models have been proposed for wildfire modeling: The point process identifies the occurrence of the fire, with size as the mark. However, the marks may not be separable from the points. This was illustrated by Schoenberg (2004), who found evidence of a lack of separability between fire occurrences and sizes in Los Angeles County, California, due to small-scale clustering. Moreover, even outside of the context of developing marked point processes models, key wildfire characteristics are likely linked. An obvious example of this is fire duration and size, under the axiom that the longer a fire lasts, the larger it grows. Such situations motivate the need to consider alternative modeling frameworks where outcome characteristics are modeled jointly. In this section, we give an overview of joint modeling of two random variables, using fire duration and size as an illustration.

Jointly modeling the duration and size of fires with environmental variables as covariates offers a potential novel direction for effectively quantifying these outcomes. Since smaller-sized fires' (<2 hectares) lifetimes are usually short (<2 days), while larger ones are usually long (days to months), such modeling accounts for the dependence between duration and size. In managed regions, more than 90% of fires are contained during an initial attack, and for fires that escape extended attack, there is a clear connection between the time to containment and fire size (Fried & Gilless 1989). The two-dimensional framework for bivariate extreme value models (e.g., Weibull, log-normal, logistic) has been recently adopted in some pioneering work because duration and size are often weakly correlated with heavy tails (Yoder & Gebert 2012, Sun 2013). Bayham (2013), in his dissertation, modeled the duration, size, and cost of containment on 3,829 US fires using a tri-outcome PH frailty model with environmental and geographical variables as covariates. Endogenous time-varying covariates were lagged for one period, and the median value was used instead of the complete covariate trajectories.

Past work modeling duration and size as a function of environmental variables shares four common features. First, although the survivor or hazard functions of the outcomes are often of interest, they can be obtained easily from estimates of the distribution of these outcomes and hence do not need to be modeled directly. Nevertheless, they need to be measured from the same origin to the same event (e.g., from the start of initial attack to the time of final control). Second, heavy-tailed distributions may be used to model both outcomes. Although power-law or extreme value distributions have received much attention in the context of wildfire science, basic location-scale distributions also fit well, and such empirical approaches have been overlooked. Third, though AFT frameworks have been commonly used, they do not necessarily lead to a model where the frailty can be interpreted as a latent covariate acting multiplicatively on the hazard. Finally, we note that covariate coefficients in the PH frailty model may not be estimated well when the PH assumption does not hold (He & Lawless 2005, He 2014), which may be of concern in the use of these frailty models in wildfire science. As a result, placing the random effect additively as a latent covariate in an AFT model would provide a compromise to both frameworks (Lambert et al. 2004, Komárek & Lesaffre 2008) and a foundation to model duration and size jointly.

To illustrate a joint modeling framework that addresses the issues above, we consider a simple model that has been discussed extensively in the literature. Assuming that both the duration and size of fires follow a location-scale distribution, AFT models can be linked to model the two outcomes jointly:

$$\log(t_{ik}) = \mu_k + \boldsymbol{\beta}_k^T \boldsymbol{x}_{ik} + b_{ik} + \sigma_k \varepsilon_{ik},$$

where $\boldsymbol{b}_i = (b_{i1}, b_{i2})^T$ is a random effect with components that are dependent. Here, k equals 1 and 2, for duration and size, respectively; t_{ik} is the outcome; $\boldsymbol{x}_{ik}$ are covariates with associated coefficients β_k; μ_k is the intercept term (the mean of the logarithm of t_{ik} when $\boldsymbol{x}_{ik} = 0$); ε_{ik} is the outcome-specific error with unit variance, associated with outcome k for fire i; and σ_k are variance parameters associated with outcome k.

Various forms of $\boldsymbol{b}_i$ have been discussed in the literature. He & Lawless (2005) and Duchateau & Janssen (2008), among many others, note that $\boldsymbol{b}_i$ may be parameterized with $b_{i1} = b_{i2} = b_i$, as a shared frailty acting additively on the logarithm of the outcomes. To account for the scale difference between the two outcomes, an additional parameter, γ, often called the factor loading parameter, can be introduced by letting $\boldsymbol{b}_i = (b_i, \gamma b_i)^T$ with $\boldsymbol{b} = (b_1, \ldots b_n)^T \sim MVN(\boldsymbol{0},\ \Sigma_b)$, with b_i and ε_{ik} independent. The term b_i can be viewed as an individual-specific error that is shared across the two outcomes. With the assumption that individual fire lifetimes and sizes are independent, a simple form for Σ_b is $\sigma_b^2 I$ (Renouf et al. 2016, Juarez-Colunga et al. 2017). Having σ_b significantly different from 0 suggests that there is dependence between the two outcomes. When $\gamma = 1$ the shared random effect influences the two outcomes identically. This is not likely in situations where the two outcomes, such as duration and size, have very different scales. In general, having γ significantly different from 1 suggests that the terms b_i have different scales by which they act on the outcomes. If prior knowledge suggests that the frailty is correlated, for example, spatially, then Σ_b may take more complicated forms (Feng & Dean 2012). Additional constraints are required (i.e., removing ε_{i1} from the model) to ensure that the model is identifiable. Alternative forms such as assuming $\boldsymbol{b}_i = (b_{i1}, b_{i2})^T$ with $\boldsymbol{b}_i \sim N_2(0, \boldsymbol{D})$ have also being considered in Komárek & Lesaffre (2008) and Bogaerts et al. (2018). For other methods that also use random effects or latent variables to model multiple outcomes jointly, readers are directed to, for example, Verbeke & Molenberghs (2017).

7. EVENT SETS AND BURN PROBABILITY

Many fire characteristics, such as ignition probability, spread rate and duration, and fire size contribute to the fire hazard. Reed (2006) defined the local hazard of burning at a point x in a study area, at time t, as

$$\lambda(t;\ x) = \lim_{\mathrm{d}t \to 0} \{\mathrm{P}\,(\text{fire at location } x \text{ in } [t, t + \mathrm{d}t])\,/\mathrm{d}t\},$$

the area-wide hazard of burning as

$$\Lambda(t) = \lim_{\mathrm{d}t \to 0} \left\{\mathrm{P}\left(\text{fire ignited somewhere in the area in } \left[t, t + \mathrm{d}t\right]\right)/\mathrm{d}t\right\},$$

and the relationship between local and area-wide hazard of burning as

$$\lambda\,(t : x) = \Lambda\,(t) \int_A h\,(x, y; t)\, f\,(y; t)\,\mathrm{d}y,$$

where $h(x, y; t)$ is the conditional probability of a fire ignited at point y spreading to x at time t and $f(y; t)$ is the probability density function of where ignitions will occur over the area A given that a fire occurs in time t. The integral above can be simplified, where $p(t, x)$ is the conditional probability of a fire occurring at x given that a fire starts somewhere in the area, as

$$\lambda(t; x) = \Lambda(t)p(t, x).$$

It is noteworthy that most of the local hazard of burning at a point x obtains from incursions of fire from adjacent locations. This model is analogous to population system epidemiology (Koopman & Lynch 1999), where infection connections between individuals and joint effects of possible multiple exposures are incorporated into infectious disease spread analysis. The local hazard of burning in a region has been estimated from empirical data from fire scars and forest stand age data (see summary in Taylor et al. 2013) in a region, assuming spatial homogeneity, while the area-wide hazard can be estimated from administrative fire records or remote sensing data. However, a number of authors have found that the local hazard of burning may vary within a landscape at scales important to fire and land management due to topographic and vegetation conditions. For example, forest stands on warm slopes in the Rocky Mountains have a greater likelihood of burning (Rogeau & Armstrong 2017), while those adjacent to nonvegetated areas such as large lakes have a lower likelihood of burning (Bergeron 1991).

Over the past decade, a body of work termed burn probability modeling (Miller et al. 2008) has developed to estimate the local hazard of burning while incorporating the influences of varying vegetation types and topographic positions within regional landscapes. Briefly, burn probability is estimated geometrically by modeling fire event sets, where an event is a spatially referenced fire perimeter map, the final fire perimeter obtained from the cumulative spread over a fire's lifetime. Monte Carlo methods are used to simulate a large number of fire events in a study area. As with the spread modeling described earlier, the study area and simulation period represent a set of voxels, with vegetation, topographic, and other geographic covariates varying spatially and weather covariates varying temporally for each day in the simulation period. In one approach, a fire is ignited at a point using a conditional spatial point process model of fire occurrence (Woo et al. 2017) including covariates at the grid points; the fire spreads between points using a deterministic fire growth model, depending on weather, vegetation, and topographic covariates at the grid points on a particular day; and individual fire events are modeled for a number of days informed by duration models. The burn probability of a given cell is estimated as the empirical proportion of fire events in that cell over the number of simulation iterations (usually years), as shown in **Figure 5**. The size distribution of fires in resulting event sets can be compared against fire size models to assess goodness of fit (e.g., appendix S.6 in Wang et al. 2016). Several systems have developed around different fire spread models. Parisien et al. (2013) provide a flowchart of the modeling process for an application of the BurnP3 system that uses the Prometheus spread model; similar procedures are used for simulations with the FSim system, which uses the FarSite fire growth simulator (Finney et al. 2011). Further challenges may include closer integration of joint models of size and duration and covariance of numbers of fires and fire size, as well as explicit representation of fire suppression.

8. DISCUSSION

Fire danger and risk research has evolved from the development of qualitative indices, to deterministic models of fire characteristics, to stochastic models of fire characteristics. It is an ongoing challenge to integrate these models and approaches in probabilistic, quantitative hazard and risk models. Improving prediction of daily wildfire dynamics is critical for proactive rather than reactive fire management decision making. Effective prediction for decision making requires (*a*) an understanding of the physical and management processes influencing ignition, growth, and

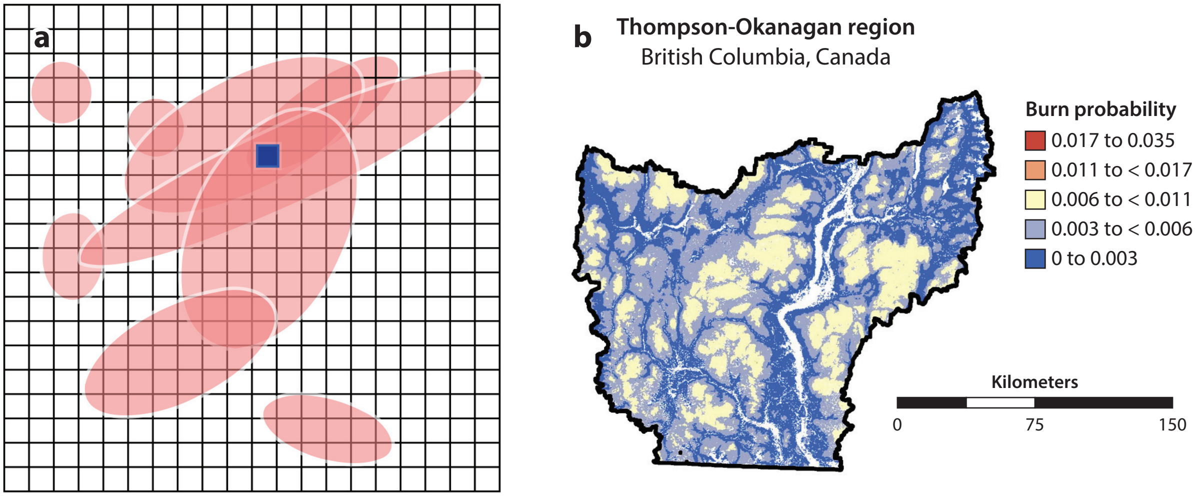

Figure 5

(*a*) Wildfire event sets can be generated with stochastic point process models and fire growth simulations of specified duration (ellipses used for illustration only; figure courtesy of Carol Miller, US Department of Agriculture Forest Service). Elements within a cell are homogenous with respect to weather, fuels, and topography. The ellipses are individual fires, and the blue square represents a sampling point. (*b*) Monte Carlo methods have been used to map burn probability by simulating large event sets representing many thousands of potential outcomes in modeled landscapes, such as the Thompson-Okanagan region of southern British Columbia (Wang et al. 2016).

survival; (*b*) the acquisition and assembly of historical, longitudinal data on daily fire starts, size, management actions, extinguishment, and covariates for building empirical models; (*c*) the development of appropriate statistical models; and (*d*) the implementation of predictive models in fire management decision support systems, preferably with an ability to use new data on fires for continuous improvement of predictive models.

Much work has focused on developing models to understand ignition and growth processes, and far fewer studies have considered containment and extinction. Even so, there are few models that consider changes in fire size by day while also accounting for resources allocated to suppression; this is due in part to limited availability of such longitudinal data on a daily scale. More work is also needed in the development of appropriate models that take into account fire-pest interaction effects and the health of trees in the path of a fire. New remote-sensing products may assist here, but there will be considerable work involved in building historical archives. Importantly, there are substantial challenges associated with mounting investigations and developing predictive models because of the massive effort involved in the assembly of historical fire databases, validation of these databases, linkage and data fusion across regions and across governmental agencies that record environmental and management variables associated with fire, and management of differences in spatial and temporal resolution that are associated with each database.

Verification of historical records can be very difficult, as can homogenization of long-term series of environmental data, when monitoring stations change location over time. There are also challenges associated with appropriately accommodating the differences in fire suppression management protocols over time, changes in detection efficiency over time, and the differences in tools and techniques for fire suppression that have evolved over recent decades. These large data issues are not inconsequential, especially when developing provincial/national models at high spatial/temporal resolution. Finally, accurate prediction models require incorporating variability associated with the differential use of fire suppression resources between fires and variability associated with future weather conditions. Importantly, we note that few statisticians are willing to expend the effort necessary to take models and methods to an implementation or knowledge translation stage as identified in item *d* above, but this is a key critical process step for impact. Further work is also needed to better represent uncertainty in models of spread, growth, and intensity and also for visualization of these characteristics.

Comprehensive, fine-scale fire occurrence modeling over a large study area introduces specific challenges. For example, the province of Ontario uses a suite of person-caused and lightning-caused fire occurrence prediction models operationally on a daily basis (Woolford et al. 2016). For this decision support tool, human-caused and lightning-caused fire occurrence predictions based on models need to be integrated into a single probability scale. This can be challenging because occurrence probabilities for lightning-caused fires in a given cell can be much larger than the probabilities for human-caused fires. This is because lightning-caused fire occurrence models incorporate lightning strike observations, which have high daily variability (e.g., Wotton & Martell 2005), whereas indicators of human presence or activity in human-caused fire models don't have strong daily variation (e.g., Woolford et al. 2011). This difference in scale may occur because lightning-caused fire occurrence models incorporate information about the observed strikes that are recorded by a network of sensors (e.g., Wotton & Martell 2005), whereas human-caused fire occurrence prediction models summarize historical patterns in fire ignitions without incorporating information about potential ignition sources (e.g., Woolford et al. 2011). In addition, outputs from fitting complex models in statistical software, such as a logistic GAM model object fit in R software, need to be summarized (as, e.g., a set of lookup tables for each partial effect in the model) for easy implementation into a non-R-based fire management operations decision support tool.

Simulation systems that are currently used to estimate the annual local burn probability use statistical models to represent stochastic components in that complex system. However, there are many components that are modeled as separate subprocesses. In order to enhance quantitative risk assessment models, a joint modeling framework should be considered when key characteristics may not necessarily be independent. Further development of quantitative risk assessment methods across all temporal scales will require, as in statistical physics, hybrid approaches that combine mathematical and statistical models with simulation methods to estimate very complex processes. For example, key stakeholders such as fire management agencies and property insurers are interested not only in annual burn probability maps but also in burn probabilities at other temporal scales, such as the probability that a fire may be ignited and spread into a nearby town on a given day.

We comment that it would be very interesting to compare simulation methods with other means of estimating these complex processes, which may be effective at some spatial scales. It would be also useful to compare modern statistical learning algorithmic techniques to the well-established logistic-based modeling techniques. Developing methodology for combining these and other models together in an ensemble framework, thereby building on the benefits of each of these approaches, would be particularly helpful.

It is challenging to incorporate estimates of uncertainty in fire management strategies, in part because it is a highly dynamic and multiscalar decision environment. Although advances have been made in developing stochastic models of characteristics such as fire occurrence, medium-term fire spread, and burn probability, few studies have connected hazard measures, including uncertainty, with damage functions and impacts (e.g., Preisler et al. 2011). Implementation of new models within a fire management decision environment presents special challenges at the interface between data analytics and human factors (that are not unique to the fire community). These include:

1. The time available for decision making decreases in the series of activities: mitigation/prevention, planning/preparedness, and response. At the sharp end of fire response, the time for decision making may be reduced to a few minutes or less (e.g., Alexander et al. 2016). Models have to be simple to use and easy to interpret; visualization techniques should be used whenever possible.
2. Decision makers within an operations background tend to be "men (or women) of action, rather than men of letters" (Macleod 1964, p. 8) coming from an institutional culture that values fast, intuitive decision making over slower, rational decision processes (e.g., Kahneman 2011) and have a healthy skepticism of models. It is important to validate models and provide case studies showing the value of information. In counterpoint, fire managers with long experience with weather dependent fire phenomena may have an intuitive appreciation for the stochastic nature of fire characteristics. Current machine learning algorithms based on historical data cannot adequately replicate such experience.

Collaborative approaches have proved successful in developing and implementing the models currently used in fire management. Whereas commonly the statistician's goal is finding a useful application, it is important at a project's outset to set common goals, find champions who are influential members of the user community, create relationships, and seek to understand the decision maker's way of doing business and constraints.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the Canadian Statistical Sciences Institute through its Collaborative Research Team program and the Natural Sciences and Engineering Research Council of Canada through its Discovery Grants program. We also thank B.M. Wotton for helpful discussions on the science underlying lightning fire ignitions and occurrences and A.K. Beaver for discussions on wildfire risk incorporated in **Figure 2**.

LITERATURE CITED

Albert-Green A, Dean C, Martell DL, Woolford DG. 2013. A methodology for investigating trends in changes in the timing of the fire season with applications to lightning-caused forest fires in Alberta and Ontario, Canada. *Can. J. For. Res.* 43:39–45

Alexander ME, Taylor SW, Page WG. 2016. Wildland firefighter safety and fire behavior prediction on the fireline. In *Proceedings of the 13th Annual Wildland Fire Safety Summit & 4th Human Dimensions of Wildland Fire Conference, April 20–24, 2015, Boise, ID*. Missoula, MT: Int. Assoc. Wildland Fire

Bayham J. 2013. *Characterizing incentives: an investigation of wildfire response and environmental entry policy*. PhD Thesis, Washington State University, Pullman, WA

Bergeron Y. 1991. The influence of island and mainland lakeshore landscapes on boreal forest fire regimes. *Ecology* 72:1980–92

Bogaerts K, Komárek A, Lesaffre E. 2018. *Survival Analysis with Interval-Censored Data: A Practical Approach with Examples in R, SAS, and BUGS*. Boca Raton, FL: CRC

Braun WJ, Woolford DG. 2013. Assessing a stochastic fire spread simulator. *J. Environ. Inform.* 22:1–12

Breiman L. 2001. Statistical modelling: the two cultures (with comments and a rejoinder by the author). *Stat. Sci.* 16(3):199–231

Brillinger DR, Preisler HK, Benoit JW. 2003. Risk assessment: a forest fire example. In *Statistics and Science: A Festschrift for Terry Speed*, ed. DR Goldstein, pp. 177–96. Beachwood, OH: Inst. Math. Stat.

Bruce D. 1963. How many fires. *Fire Control Notes* 24:45–50

Bryant BP, Westerling AL. 2014. Scenarios for future wildfire risk in California: links between changing demography, land use, climate, and wildfire. *Environmetrics* 25:454–71

Butry DT, Gumpertz M, Genton MG. 2008. The production of large and small wildfires. In *The Economics of Forest Disturbances: Wildfires, Storms, and Invasive species*, ed. TP Holmes, JP Prestemon, KL Abt, pp. 79–106. New York: Springer

Cardona OD, van Aalst MK, Birkmann J, Fordham M, McGregor G, et al. 2012. Determinants of risk: exposure and vulnerability. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*, ed. CB Field, V Barros, TF Stocker, D Qin, DJ Dokken, et al., pp. 65–108. Cambridge, UK: Cambridge Univ. Press

Castillo E. 2012. *Extreme Value Theory in Engineering*. New York: Elsevier

Chen F, Fan Z, Niu S, Zheng J. 2014. The influence of precipitation and consecutive dry days on burned areas in Yunnan Province, Southwestern China. *Adv. Meteorol.* 2014:748923

Costafreda-Aumedes S, Cardil A, Molina DM, Daniel SN, Mavsar R, Vega-Garcia C. 2015. Analysis of factors influencing deployment of fire suppression resources in Spain using artificial neural networks. *iForest Biogeosciences For*. 9:138–45

Crichton D. 1999. The risk triangle. In *Natural Disaster Management*, ed. J Ingleton, pp. 102–3. London: Tudor Rose

Cruz MG. 2010. Monte Carlo-based ensemble method for prediction of grassland fire spread. *Int. J. Wildland Fire* 19:521–30

Cui W, Perera AH. 2008. What do we know about forest fire size distribution, and why is this knowledge useful for forest management? *Int. J. Wildland Fire* 17:234–44

Cumming S. 2001. A parametric model of the fire-size distribution. *Can. J. For. Res.* 31:1297–303

Cunningham AA, Martell DL. 1973. A stochastic model for the occurrence of man-caused forest fires. *Can. J. For. Res.* 3:282–87

DaCamara CC, Calado TJ, Ermida SL, Trigo IF, Amraoui M, Turkman KF. 2014. Calibration of the Fire Weather Index over Mediterranean Europe based on fire activity retrieved from MSG satellite imagery. *Int. J. Wildland Fire* 23:945–58

Daley DJ, Vere-Jones D. 2003. *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. New York: Springer

Davison AC, Huser R. 2015. Statistics of extremes. *Annu. Rev. Stat. Appl.* 2:203–35

Duchateau L, Janssen P. 2008. *The Frailty Model*. New York: Springer

Embrechts P, Hofert M. 2014. Statistics and quantitative risk management for banking and insurance. *Annu. Rev. Stat. Appl.* 1:493–514

Feng C, Dean C. 2012. Joint analysis of multivariate spatial count and zero-heavy count outcomes using common spatial factor models. *Environmetrics* 23:493–508

Fernandes PM, Pacheco AP, Almeida R, Claro J. 2016. The role of fire-suppression force in limiting the spread of extremely large forest fires in Portugal. *Eur. J. For. Res.* 135:253–62

Finney MA. 2005. The challenge of quantitative risk analysis for wildland fire. *For. Ecol. Manag.* 211:97–108

Finney MA, Grenfell IC, McHugh CW. 2009. Modelling containment of large wildfires using generalized linear mixed-model analysis. *For. Sci.* 55:249–55

Finney MA, McHugh CW, Grenfell IC, Riley KL, Short KC. 2011. A simulation of probabilistic wildfire risk components for the continental United States. *Stoch. Environ. Res. Risk Assess.* 25:973–1000

Foss S, Korshunov D, Zachary S. 2011. *An Introduction to Heavy-Tailed and Subexponential Distributions*. New York: Springer

Fried JS, Gilless JK. 1989. Notes: expert opinion estimation of fireline production rates. *Forest Sci*. 35:870–77

Fujioka FM, Gill AM, Viegas DX, Wotton BM. 2008. Fire danger and fire behavior modelling systems in Australia, Europe, and North America. *Dev. Environ. Sci.* 8:471–97

Giglio L, Randerson JT, Werf GR. 2013. Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *J. Geophysical Res. Biogeosciences* 118:317–28

Hardy CC, Hardy CE. 2007. Fire danger rating in the United States of America: an evolution since 1916. *Int. J. Wildland Fire* 16:217–31

Hastie T, Tibshirani R, Friedman J. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer

He W. 2014. Analysis of multivariate survival data with Clayton regression models under conditional and marginal formulations. *Comput. Stat. Data Anal.* 74:52–63

He W, Lawless JF. 2005. Bivariate location–scale models for regression analysis, with applications to lifetime data. *J. R. Stat. Soc. B* 67:63–78

Hernandez C, Keribin C, Drobinski P, Turquety S. 2015. Statistical modelling of wildfire size and intensity: a step toward meteorological forecasting of summer extreme fire risk. *Ann. Geophys.* 33:1495–1506

Heyerdahl EK, Lertzman K, Karpuk S. 2007. Local-scale controls of a low-severity fire regime (1750–1950), southern British Columbia, Canada. *Ecoscience* 14:40–47

Holmes TP, Huggett RJ Jr., Westerling AL. 2008. Statistical analysis of large wildfires. In *The Economics of Forest Disturbances: Wildfires, Storms, and Invasive Species*, ed. TP Holmes, JP Prestemon, KL Abt, pp. 59–77. New York: Springer

Hougaard P. 2000. Shared frailty models. In *Analysis of Multivariate Survival Data*, ed. P Hougaard, pp. 215–62. New York: Springer

James G, Witten D, Hastie T, Tibshirani R. 2013. *An Introduction to Statistical Learning with Applications in R*. New York: Springer

Joe H. 2014. *Dependence Modeling with Copulas*. Boca Raton, FL: CRC

Juarez-Colunga E, Silva G, Dean C. 2017. Joint modelling of zero-inflated panel count and severity outcomes. *Biometrics* 73:1413–23

Kahneman D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux

Komárek A, Lesaffre E. 2008. Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions. *J. Am. Stat. Assoc.* 103:523–33

Koopman JS, Lynch JW. 1999. Individual causal models and population system models in epidemiology. *Am. J. Public Health* 89:1170–74

Krawchuk MA, Cumming SG, Flannigan MD. 2009a. Predicted changes in fire weather suggest increases in lightning fire initiation and future area burned in the mixedwood boreal forest. *Clim. Change* 92:83–97

Krawchuk MA, Moritz MA, Parisien M-A, Van Dorn J, Hayhoe K. 2009b. Global pyrogeography: the current and future distribution of wildfire. *PLOS ONE* 4:e5102

Lambert P, Collett D, Kimber A, Johnson R. 2004. Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Stat. Med.* 23:3177–92

Linn R, Winterkamp J, Edminster C, Colman JJ, Smith WS. 2007. Coupled influences of topography and wind on wildland fire behaviour. *Int. J. Wildland Fire* 16:183–95

Liu J. 2012. *Modelling dependence induced by a common random effect and risk measures with insurance applications*. MSc Thesis, Simon Fraser Univ., Burnaby, BC, Can.

Longin F. 2016. *Extreme Events in Finance: A Handbook of Extreme Value Theory and Its Applications*. New York: Wiley

Macleod JC. 1964. *Planning for forest fire control*. Dep. Publ. 1048, Can. Dep. For., Ottawa, Can.

Marchi E, Neri F, Tesi E, Fabiano F, Brachetti Montorselli N. 2014. Analysis of helicopter activities in forest fire-fighting. *Croat. J. For. Eng.* 35:233–43

Martell DL, Bevilacqua E, Stocks BJ. 1989. Modelling seasonal variation in daily people-caused forest fire occurrence. *Can. J. Forest Res.* 19:1555–63

Martell DL, Sun H. 2008. The impact of fire suppression, vegetation, and weather on the area burned by lightning-caused forest fires in Ontario. *Can. J. For. Res.* 38:1547–63

McLoughlin N, Gibos K. 2016. A 72-day probabilistic fire growth simulation used for decision support on a large mountain fire in Alberta, Canada. In *Proceedings for the 5th International Fire Behavior and Fuels Conference, April 11-15, 2016, Portland, OR*. Missoula, MT: Int. Assoc. Wildland Fire

Miller C, Ager AA. 2013. A review of recent advances in risk analysis for wildfire management. *Int. J. Wildland Fire* 22:1–14

Miller C, Parisien MA, Ager A, Finney M. 2008. Evaluating spatially-explicit burn probabilities for strategic fire management planning. *WIT Trans. Ecol. Environ.* 119:245–52

Morgan MG, Henrion M, Small M. 1992. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge, UK: Cambridge Univ. Press

Morin AA. 2014. *A spatial analysis of forest fire survival and a marked cluster process for simulating fire load*. Master's Thesis, Univ. West. Ontario, London, Can.

Morin AA, Albert-Green A, Woolford DG, Martell DL. 2015. The use of survival analysis methods to model the control time of forest fires in Ontario, Canada. *Int. J. Wildland Fire* 24:964–73

Moritz MA. 1997. Analyzing extreme disturbance events: fire in Los Padres National Forest. *Ecol. Appl.* 7:1252–62

Nadeem K, Taylor SW, Dean CB, Woolford DG, Magnussen S, Wotton BM. 2016. *Severe wildland fire risk prediction in Canada*. Poster presented at Wildland Fire Canada 2016: Building Resilience, Kelowna, BC, Oct. 24–28

Nelsen R. 2006. *An Introduction to Copulas*. New York: Springer

Papakosta P, Xanthopoulos G, Straub D. 2017. Probabilistic prediction of wildfire economic losses to housing in Cyprus using Bayesian network analysis. *Int. J. Wildland Fire* 26:10–23

Parisien MA, Walker GR, Little JM, Simpson BN, Wang X, Perrakis DD. 2013. Considerations for modeling burn probability across landscapes with steep environmental gradients: an example from the Columbia Mountains, Canada. *Nat. Hazards* 66:439–62

Parks SA, Parisien M-A, Miller C. 2012. Spatial bottom-up controls on fire likelihood vary across western North America. *Ecosphere* 3:1–20

Pinto RM, Benali A, Sá AC, Fernandes PM, Soares PM, et al. 2016. Probabilistic fire spread forecast as a management tool in an operational setting. *SpringerPlus* 5:1205

Plummer FG. 1912. *Forest fires: their causes, extent, and effects, with a summary of recorded destruction and loss*. Bull. 117, For. Serv., US Dep. of Agric., Washington, DC

Podur J, Martell DL, Csillag F. 2003. Spatial patterns of lightning-caused forest fires in Ontario, 1976–1998. *Ecol. Model.* 164:1–20

Preisler HK, Ager AA. 2013. Forest-fire models. In *Encyclopedia of Environmetrics*, ed. AH El-Shaarawi, WW Piegorsch. **https://doi.org/10.1002/9780470057339.vaf010.pub2**

Preisler HK, Brillinger DR, Burgan RE, Benoit J. 2004. Probability based models for estimation of wildfire risk. *Int. J. Wildland Fire* 13:133–42

Preisler HK, Westerling AL. 2007. Statistical model for forecasting monthly large wildfire events in western United States. *J. Appl. Meteorol. Climatol.* 46:1020–30

Preisler HK, Westerling AL, Gebert KM, Munoz-Arriola F, Holmes TP. 2011. Spatially explicit forecasts of large wildland fire probability and suppression costs for California. *Int. J. Wildland Fire* 20:508–17

Price O, Borah R, Bradstock R, Penman T. 2015. An empirical wildfire risk analysis: the probability of a fire spreading to the urban interface in Sydney, Australia. *Int. J. Wildland Fire* 24:597–606

Reed WJ. 1999. Forest fires and oilfields as percolation phenomena. Work. Pap., Dep. Math. Stat., Univ. Victoria, BC, Can. **https://www.math.uvic.ca/faculty/reed/Sci.paper.pdf**

Reed WJ. 2001. The Pareto, Zipf and other power laws. *Econ. Lett.* 74:15–19

Reed WJ. 2006. A note on fire frequency concepts and definitions. *Can. J. For. Res.* 36:1884–88

Reed WJ. 2011. A flexible parametric survival model which allows a bathtub-shaped hazard rate function. *J. Appl. Stat.* 38:1665–80

Reed WJ. 2012. Power-law adjusted survival models. *Commun. Stat. Theory Methods* 41:3692–703

Reed WJ, Hughes BD. 2002. From gene families and genera to incomes and internet file sizes: why power laws are so common in nature. *Phys. Rev. E* 66:067103

Reed WJ, Hughes BD. 2004. A model explaining the size distribution of gene and protein families. *Math. Biosci.* 189:97–102

Reed WJ, Jorgensen M. 2004. The double Pareto-lognormal distribution—a new parametric model for size distributions. *Commun. Stat. Theory Methods* 33:1733–53

Reed WJ, McKelvey KS. 2002. Power-law behaviour and parametric models for the size-distribution of forest fires. *Ecol. Model.* 150:239–54

Renouf E, Dean CB, Bellhouse DR, McAlister VC. 2016. Joint survival analysis of time to drug change and a terminal event with application to drug failure analysis using transplant registry data. *Int. J. Stat. Med. Res.* 5:198–213

Rogeau M-P, Armstrong GW. 2017. Quantifying the effect of elevation and aspect on fire return intervals in the Canadian Rocky Mountains. *For. Ecol. Manag.* 384:248–61

Schoenberg FP. 2004. Testing separability in spatial-temporal marked point processes. *Biometrics* 60(2):471–81

Schoenberg FP, Peng R, Woods J. 2003. On the distribution of wildfire sizes. *Environmetrics* 14:583–92

Scott JH. 2006. An analytical framework for quantifying wildland fire risk and fuel treatment benefit. In *Fuels Management—How to Measure Success: Conference Proceedings*, ed. PL Andrews, BW Butler, pp. 169–84. Washington, DC: US For. Serv.

Stocks B, Mason J, Todd J, Bosch E, Wotton B, et al. 2002. Large forest fires in Canada, 1959–1997. *J. Geophysical Res. Atmos.* **https://doi.org/10.1029/2001JD000484**

Sullivan AL. 2009a. Wildland surface fire spread modelling, 1990–2007. 1: Physical and quasi-physical models. *Int. J. Wildland Fire* 18:349–68

Sullivan AL. 2009b. Wildland surface fire spread modelling, 1990–2007. 2: Empirical and quasi-empirical models. *Int. J. Wildland Fire* 18:369–86

Sullivan AL. 2009c. Wildland surface fire spread modelling, 1990–2007. 3: Simulation and mathematical analogue models. *Int. J. Wildland Fire* 18:387–403

Sun C. 2013. Bivariate extreme value modelling of wildland fire area and duration. *For. Sci.* 59:649–60

Taylor SW, Alexander ME. 2006. Science, technology, and human factors in fire danger rating: the Canadian experience. *Int. J. Wildland Fire* 15:121–35

Taylor SW, Woolford DG, Dean C, Martell DL. 2013. Wildfire prediction to inform management: statistical science challenges. *Stat. Sci.* 28:586–615

Therneau TM, Grambsch PM. 2000. *Modelling Survival Data: Extending the Cox Model*. New York: Springer

Turner R. 2009. Point patterns of forest fire locations. *Environ. Ecol. Stat.* 16:197–223

Vega-Garcia C, Lee B, Woodard P, Titus S. 1996. Applying neural network technology to human-caused wildfire occurrence prediction. AI Appl. 10:9–18

Vega-Garcia C, Woodard P, Titus S, Adamowicz W, Lee B. 1995. A logit model for predicting the daily occurrence of human caused forest-fires. *Int. J. Wildland Fire* 5:101–11

Verbeke G, Molenberghs G. 2017. Modeling through latent variables. *Annu. Rev. Stat. Appl.* 4:267–82

Vilar L, Woolford DG, Martell DL, Martín MP. 2010. A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain. *Int. J. Wildland Fire* 19:325–37

Wang X, Parisien M-A, Taylor SW, Perrakis DD, Little J, Flannigan MD. 2016. Future burn probability in south-central British Columbia. *Int. J. Wildland Fire* 25:200–12

Wang Y, Anderson KR. 2011. An evaluation of spatial and temporal patterns of lightning- and human-caused forest fires in Alberta, Canada, 1980–2007. *Int. J. Wildland Fire* 19:1059–72

Westerling A, Bryant B. 2008. Climate change and wildfire in California. *Clim. Change* 87:231–49

Westerling A, Bryant B, Preisler H, Holmes T, Hidalgo H, et al. 2011a. Climate change and growth scenarios for California wildfire. *Clim. Change* 109:445–63

Westerling AL, Hidalgo HG, Cayan DR, Swetnam TW. 2006. Warming and earlier spring increase western US forest wildfire activity. *Science* 313:940–43

Westerling AL, Turner MG, Smithwick EA, Romme WH, Ryan MG. 2011b. Continued warming could transform Greater Yellowstone fire regimes by mid-21st century. *PNAS* 108:13165–70

Wienke A. 2010. *Frailty Models in Survival Analysis*. Boca Raton, FL: CRC

Woo H, Chung W, Graham JM, Lee B. 2017. Forest fire risk assessment using point process modeling of fire occurrence and Monte Carlo fire simulation. *Int. J. Wildland Fire* 26:789–805

Wood S. 2006. *Generalized Additive Models: An Introduction with R*. Boca Raton, FL: CRC

Woolford DG, Bellhouse D, Braun W, Dean CB, Martell D, Sun J. 2011. A spatio-temporal model for people-caused forest fire occurrence in the Romeo Malette Forest. *J. Environ. Stat.* 2:2–16

Woolford DG, Braun W, Dean C, Martell D. 2009. Site-specific seasonal baselines for fire risk in Ontario. *Geomatica* 63:355–63
Woolford DG, Cao J, Dean CB, Martell DL. 2010. Characterizing temporal changes in forest fire ignitions: looking for climate change signals in a region of the Canadian boreal forest. *Environmetrics* 21:789–800
Woolford DG, Dean C, Martell DL, Cao J, Wotton B. 2014. Lightning-caused forest fire risk in Northwestern Ontario, Canada, is increasing and associated with anomalies in fire weather. *Environmetrics* 25:406–16
Woolford DG, Wotton BM, Martell DL, McFayden C, Stacey A, et al. 2016. *Daily lightning- and person-caused fire prediction models used in Ontario*. Poster presented at Wildland Fire Canada Conf. 2016. **http://www.wildlandfire2016.ca/wp-content/uploads/2017/03/McFayden-Fire-Occurence-Prediction-Poster-Ontario-2016-10-17V2Final.pdf**
Wotton BM, Martell DL. 2005. A lightning fire occurrence model for Ontario. *Can. J. For. Res.* 35:1389–401
Wotton BM, Martell DL, Logan K. 2003. Climate change and people-caused forest fire occurrence in Ontario. *Climatic Change* 60:275–95
Wotton BM, Nock CA, Flannigan MD. 2010. Forest fire occurrence and climate change in Canada. *Int. J. Wildland Fire* 19:253–71
Xiong Y. 2015. *Analysis of spatio-temporal data for forest fire control*. Master's Thesis, Simon Fraser Univ., Burnaby, BC, Can.
Yoder J, Gebert K. 2012. An econometric model for ex ante prediction of wildfire suppression costs. *J. Forest Econ.* 18:76–89

Annual Review of Statistics and its Application

Volume 6, 2019

Contents

Stephen Elliott Fienberg 1942–2016, Founding Editor of the *Annual Review of Statistics and Its Application*
Alicia L. Carriquiry, Nancy Reid, and Aleksandra B. Slavković 1

Historical Perspectives and Current Directions in Hockey Analytics
Namita Nandakumar and Shane T. Jensen 19

Experiments in Criminology: Improving Our Understanding of Crime and the Criminal Justice System
Greg Ridgeway 37

Using Statistics to Assess Lethal Violence in Civil and Inter-State War
Patrick Ball and Megan Price 63

Differential Privacy and Federal Data Releases
Jerome P. Reiter 85

Evaluation of Causal Effects and Local Structure Learning of Causal Networks
Zhi Geng, Yue Liu, Chunchen Liu, and Wang Miao 103

Handling Missing Data in Instrumental Variable Methods for Causal Inference
Edward H. Kennedy, Jacqueline A. Mauro, Michael J. Daniels, Natalie Burns, and Dylan S. Small 125

Nonprobability Sampling and Causal Analysis
Ulrich Kohler, Frauke Kreuter, and Elizabeth A. Stuart 149

Agricultural Crop Forecasting for Large Geographical Areas
Linda J. Young 173

Statistical Models of Key Components of Wildfire Risk
Dexen D.Z. Xi, Stephen W. Taylor, Douglas G. Woolford, and C.B. Dean 197

An Overview of Joint Modeling of Time-to-Event and Longitudinal Outcomes
Grigorios Papageorgiou, Katya Mauff, Anirudh Tomer, and Dimitris Rizopoulos 223

Self-Controlled Case Series Methodology
Heather J. Whitaker and Yonas Ghebremichael-Weldeselassie 241

Precision Medicine
Michael R. Kosorok and Eric B. Laber 263

Sentiment Analysis
Robert A. Stine 287

Statistical Methods for Naturalistic Driving Studies
Feng Guo 309

Model-Based Learning from Preference Data
Qinghua Liu, Marta Crispino, Ida Scheel, Valeria Vitelli, and Arnoldo Frigessi 329

Finite Mixture Models
Geoffrey J. McLachlan, Sharon X. Lee, and Suren I. Rathnayake 355

Approximate Bayesian Computation
Mark A. Beaumont 379

Statistical Aspects of Wasserstein Distances
Victor M. Panaretos and Yoav Zemel 405

On the Statistical Formalism of Uncertainty Quantification
James O. Berger and Leonard A. Smith 433

Errata

An online log of corrections to *Annual Review of Statistics and Its Application* articles may be found at http://www.annualreviews.org/errata/statistics