Received: 1 December 2016 | Revised: 10 March 2017 | Accepted: 14 May 2017

DOI: 10.1002/env.2450

RESEARCH ARTICLE

WILEY

A model-based approach to wildland fire reconstruction using sediment charcoal records

Malcolm S. Itter[1] | Andrew O. Finley[1,2] | Mevin B. Hooten[3,4,5] | Philip E. Higuera[6] | Jennifer R. Marlon[7] | Ryan Kelly[8] | Jason S. McLachlan[9]

[1]Department of Forestry, Michigan State University, East Lansing, MI, U.S.A.

[2]Department of Geography, Michigan State University, East Lansing, MI, U.S.A.

[3]Colorado Cooperative Fish and Wildlife Research Unit, U.S. Geological Survey, Fort Collins, CO, U.S.A.

[4]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO, U.S.A.

[5]Department of Statistics, Colorado State University, Fort Collins, CO, U.S.A.

[6]Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, MT, U.S.A.

[7]Yale School of Forestry and Environmental Studies, Yale University, New Haven, CT, U.S.A.

[8]Neptune and Company Inc., Durham, NC, U.S.A.

[9]Department of Biological Sciences, University of Notre Dame, South Bend, IN, U.S.A.

Correspondence
Malcolm S. Itter, Department of Forestry, 480 Wilson Road, East Lansing, Michigan, U.S.A.
Email: ittermal@msu.edu

Funding information
National Science Foundation, Grant/Award Number: DMS-1513481, EF-1137309, EF-1241874, MSB-1241874, EF-1253225, MSB-1241856, MSB-1241846, BCS-1437074 and, MSB-1241870

Lake sediment charcoal records are used in paleoecological analyses to reconstruct fire history, including the identification of past wildland fires. One challenge of applying sediment charcoal records to infer fire history is the separation of charcoal associated with local fire occurrence and charcoal originating from regional fire activity. Despite a variety of methods to identify local fires from sediment charcoal records, an integrated statistical framework for fire reconstruction is lacking. We develop a Bayesian point process model to estimate the probability of fire associated with charcoal counts from individual-lake sediments and estimate mean fire return intervals. A multivariate extension of the model combines records from multiple lakes to reduce uncertainty in local fire identification and estimate a regional mean fire return interval. The univariate and multivariate models are applied to 13 lakes in the Yukon Flats region of Alaska. Both models resulted in similar mean fire return intervals (100–350 years) with reduced uncertainty under the multivariate model due to improved estimation of regional charcoal deposition. The point process model offers an integrated statistical framework for paleofire reconstruction and extends existing methods to infer regional fire history from multiple lake records with uncertainty following directly from posterior distributions.

KEYWORDS

Bayesian hierarchical model, fire history, fire return interval, paleoecology, Poisson point process

1 | INTRODUCTION

Charcoal particles deposited in lake sediments during and following wildland fires serve as records of local to regional fire history. Sediment charcoal records are used in paleoecological analyses to identify individual fire events and to estimate

fire frequency and regional biomass burned (i.e., the amount of organic plant matter consumed due to fire) at centennial to millennial time scales (Clark, 1988, 1990; Long, Whitlock, Bartlein, & Millspaugh, 1998; Power et al., 2008; Whitlock & Millspaugh, 1996). When combined with sediment pollen records, charcoal deposits can be used to infer relationships between changing climate, vegetation, and fire regimes, including fire frequency, size, and severity (Carcaillet et al., 2001; Clark & Royall, 1996; Clark, Royall, & Chumbley, 1996; Higuera, Brubaker, Anderson, Hu, & Brown, 2009; Kelly et al., 2013; Long et al., 1998). In particular, combined sediment charcoal records from multiple lakes have been used to correlate changes in regional biomass burned with shifts in regional vegetation and/or climate (Higuera et al., 2009; Kelly et al., 2013; Marlon et al., 2012; Power et al., 2008).

Charcoal deposits in lake sediments arise from several different sources. Large charcoal particles ($> 100\ \mu$m) have small dispersal distances and exhibit strong correlation with fire occurrence within roughly 500–1,000 m of lakes (Clark, 1988; Gavin, Brubaker, & Lertzman, 2003; Lynch, Clark, & Stocks, 2004; Peters & Higuera, 2007; Whitlock & Millspaugh, 1996). Small charcoal particles ($< 50\ \mu$m) have larger dispersal distances (typically 1–20 km) and are indicators of regional biomass burned (Clark, 1988; Clark et al., 1996). Sediment charcoal deposits arise from primary sources—direct transport during a fire—and secondary sources, including surface transport via wind and water of charcoal deposited within a lake catchment (Higuera, Peters, Brubaker, & Gavin, 2007; Whitlock & Millspaugh, 1996). Further, lake sediments mix over time, redistributing charcoal particles vertically and concentrating charcoal in the lake center (Whitlock & Millspaugh, 1996). The different depositional sources and sediment mixing increase the variability in sediment charcoal records and make inferences regarding the size and location of individual fire events difficult (Higuera et al., 2007). Despite the noise present in sediment charcoal records, the use of such data to accurately identify local fire events has been consistently demonstrated (Clark, 1990; Gavin et al., 2003; Higuera et al., 2007; Lynch et al., 2004).

Charcoal deposition is often expressed in terms of charcoal accumulation rate to account for different sedimentation rates over time (CHAR; particles $\cdot\ \mathrm{cm}^{-2}\cdot\ \mathrm{yr}^{-1}$). Analytical approaches to identify individual, local fire events based on sediment charcoal records decompose CHAR into background and peak components. The background component captures low-frequency variability associated with time-varying charcoal production rates (e.g., changes in biomass burned), secondary charcoal deposition, sediment mixing, and charcoal arising from regional sources. The peak component captures high-frequency variability associated with local fire events as well as measurement and random error (Clark et al., 1996; Long et al., 1998). Approaches to estimate background accumulation include low-pass filters applied to Fourier-transformed CHAR (Carcaillet et al., 2001; Clark et al., 1996) and locally weighted regression models (Gavin, Hu, Lertzman, & Corbett, 2006; Higuera et al., 2009; Long et al., 1998). Charcoal peaks are defined as the residuals resulting from raw CHAR series minus background CHAR or the ratio between raw and background CHAR. A threshold is used to distinguish charcoal peaks indicative of local fire events from false peaks attributable to elevated background deposition (Clark & Royall, 1996). Optimal thresholds, in terms of correct identification of local fires, are estimated using sensitivity analysis (Clark & Royall, 1996) or upper quantiles of a Gaussian mixture model (Gavin et al., 2006), lacking independent fire records to identify and validate optimal threshold values (Higuera et al., 2009).

Although methods to identify local fire events based on sediment charcoal records have been well developed over the past 30 years, an integrated statistical framework for fire identification is still lacking (Higuera, Whitlock, & Gage, 2010). We build upon existing charcoal analysis methods to develop a hierarchical Bayesian point process model for fire identification and estimation of fire return intervals (FRIs). The point process model offers a fully model-based approach to charcoal analysis with several important properties. The model operates on charcoal counts directly, using an offset term to control for sedimentation rate. We generate an explicit estimate of the probability of fire for each charcoal count. The hierarchical Bayesian approach makes for tractable error propagation allowing for a complete treatment of uncertainty sources in sediment charcoal records, including uncertainty associated with sediment age models. The model is easily extended to multivariate data sets, allowing for pooling of sediment charcoal records among lakes. Although methods currently exist to pool charcoal records (Power et al., 2008), the point process model requires no transformation or interpolation of charcoal counts, improving interpretability of results and avoiding potential introduction of nonquantifiable error to charcoal data sets. The modeling approach objectively identifies parameter values controlling the decomposition of sediment charcoal into background and peak components via regularization. Most importantly, our hierarchical Bayesian point process model provides an integrated probabilistic framework to identify local fires and estimate FRIs across multiple lake records with explicit uncertainty quantification.

The remainder of this article is organized as follows. In Section 2, we develop a point process model for charcoal deposition using data from a single lake (Section 2.1) and a regional network of lakes (Section 2.2). Estimation of local fire probability and mean FRIs are described in the context of developing the single-lake model and the multiple-lake model. Implementation of the two models applying Bayesian inference is described in Section 2.3. We demonstrate the application of the single- and

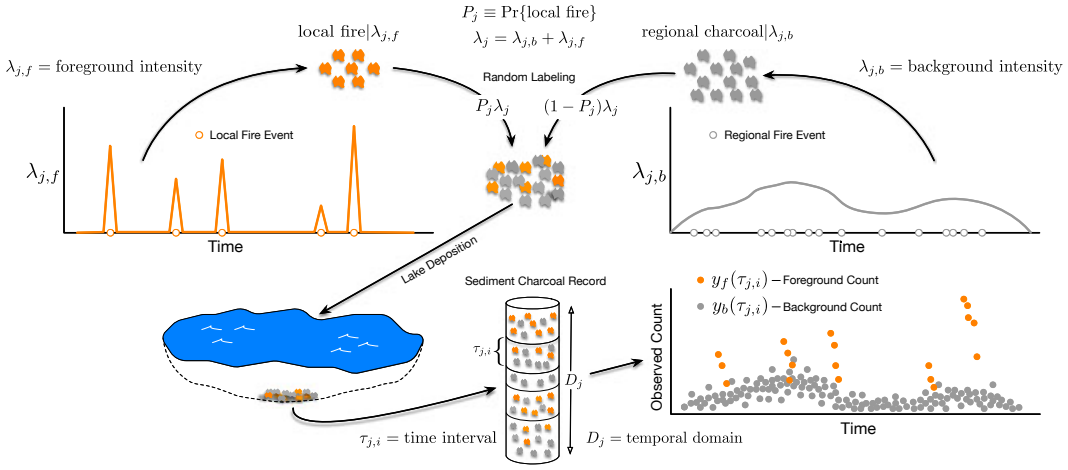

FIGURE 1 Illustration of theoretical charcoal deposition to a lake if charcoal particles arising from regional fires were distinguishable from particles arising from local fires (in practice, charcoal particles from different sources are indistinguishable, i.e., we observe only $y(\tau_{j,i}) = y_b(\tau_{j,i}) + y_f(\tau_{j,i})$). The figure also depicts the formation of a sediment charcoal record from deposited charcoal and the structure of the charcoal data (i.e., counts of charcoal over discrete, nonoverlapping time intervals). Note the figure does not depict charcoal arising from secondary sources such as surface water runoff or sediment mixing

multiple-lake models to both simulated and observed data from a regional network of lakes (Section 3). We conclude with a discussion of modeling properties and results (Section 4).

2 | BAYESIAN POINT PROCESS MODEL FOR CHARCOAL DEPOSITION

We construct a Bayesian point process model that relates charcoal deposition in lake sediments to local and regional fire occurrence. Central to the model is the separation of charcoal arising from regional and secondary sources from charcoal attributable to local fires (as depicted in Figure 1). In practice, charcoal particles arising from different sources (i.e., regional and secondary sources versus local fire events) are indistinguishable in sediment charcoal records (apart from the size distinction noted earlier). We separate background from peak deposition by assuming that charcoal particles are generated by independent processes in time: a smooth background process, exhibiting low-frequency changes in charcoal deposition rates over time, and a highly variable foreground (or peak) process, exhibiting high-frequency changes in charcoal deposition rates associated with local fire events. Total charcoal deposition is proportional to the sum of the background and foreground processes (Clark & Royall, 1996). The separation of total charcoal deposition into background and foreground processes provides the necessary analytical mechanism to identify local fire events from noisy sediment charcoal records. We begin this section by defining a univariate point process model to identify local fire events. We then extend the univariate model to accommodate sediment charcoal records from a regional network of lakes using a multivariate model.

2.1 | Univariate model

Charcoal counts are observed over time intervals spanned by the bottom and top ages of a sediment core section: $\tau_{j,i} = t_{j,i}^{(b)} - t_{j,i}^{(a)}$, where $t_{j,i}^{(b)}$ and $t_{j,i}^{(a)}$ are the bottom and top ages of sediment core section i ($i = 1, \ldots, n_j$) from lake j ($j = 1, \ldots, k$), respectively. The $\tau_{j,i}$ correspond to nonoverlapping time intervals, such that $\cup_{i=1}^{n_j} \tau_{j,i} = D_j$, where D_j is the temporal domain of lake j. Throughout this article, we use $\boldsymbol{\tau}_j$ to denote the set of observed time intervals, the temporal support, for lake j (i.e., $\boldsymbol{\tau}_j = (\tau_{j,1}, \tau_{j,2}, \ldots, \tau_{j,n_j})'$). Let $y(\tau_{j,i})$ equal the observed charcoal count for $\tau_{j,i}$ defining $\mathbf{y}_{\boldsymbol{\tau}_j} = (y(\tau_{j,1}), y(\tau_{j,2}), \ldots, y(\tau_{j,n_j}))'$. We model the accumulated charcoal particles over each observed time interval using a Poisson likelihood conditional on a latent continuous intensity process mapped to the observed temporal support. Specifically, we let

$$\mathbf{y}_{\boldsymbol{\tau}_j} | \lambda_j \sim \prod_{i=1}^{n_j} \text{Poisson}\left(\mu(\tau_{j,i})\right), \tag{1}$$

where $\mu(\tau_{j,i})$ arises as the temporal aggregation of the continuous intensity process,

$$\mu(\tau_{j,i}) = \int_{\tau_{j,i}} \lambda_j(t) dt, \qquad t \in \tau_{j,i},$$

for $i = 1, \dots, n_j$ and $j = 1, \dots, k$. We assume the latent continuous intensity process λ_j can be decomposed into additive continuous background and foreground intensity processes: $\lambda_j = \lambda_{j,b} + \lambda_{j,f}$, where $\lambda_{j,b}$ and $\lambda_{j,f}$ denote the background and foreground intensities, respectively. This assumption allows us to model the arrival of charcoal particles from regional/secondary sources and local fire events as separate, independent Poisson point processes conditional on $\lambda_{j,b}$ and $\lambda_{j,f}$ (Figure 1).

The background and foreground intensities both vary with changes in environmental conditions, but at different frequencies. The background intensity varies with climate and regional vegetation patterns on a centennial-to-millennial scale; the foreground intensity varies with local fuel loads and weather patterns on an annual-to-decadal scale. We model changes in the background and foreground intensities at the observed temporal support $\boldsymbol{\tau}_j$, assuming intensity values are constant within each observed time interval. This represents a discretization of the continuous intensity functions. The length of observed time intervals, however, is extremely short relative to the span of sediment charcoal records ($|\tau_{j,i}|/|D_j| \approx 1.0 \times 10^{-3}$, where $|\cdot|$ represents the length) such that the approach mimics an inhomogeneous Poisson point process over D_j. Importantly, the environmental conditions driving changes in the intensities are unlikely to vary substantially in the observed time intervals.

We apply a basis expansion in time to approximate the temporal dynamics of the background and foreground intensities as a proxy for changing environmental conditions. Specifically, we model the background and foreground intensity processes at temporal support $\boldsymbol{\tau}_j$ on the log scale as

$$\begin{aligned} \log\left(\lambda_{j,b}(\tau_{j,i})\right) &= \beta_{0,j}^{(b)} + \mathbf{x}(\tau_{j,i})'\boldsymbol{\beta}_j^{(b)} \\ \log\left(\lambda_{j,f}(\tau_{j,i})\right) &= \beta_{0,j}^{(f)} + \mathbf{x}(\tau_{j,i})'\boldsymbol{\beta}_j^{(f)}, \end{aligned} \tag{2}$$

where $\beta_{0,j}^{(b)}$ and $\beta_{0,j}^{(f)}$ are intercept terms, $\boldsymbol{\beta}_j^{(b)}$ and $\boldsymbol{\beta}_j^{(f)}$ are p-dimensional vectors of regression coefficients, and $\mathbf{x}(\tau_{j,i})$ is a p-dimensional set of known covariate values corresponding to basis function values at p knots. We assume the latent intensity processes $\lambda_{j,b}$ and $\lambda_{j,f}$ are independent conditional on their respective regression coefficients. We apply natural cubic splines as our basis functions in Equation 2, although alternative spline or predictive process basis functions may also be used. Following from the previous discussion of the environmental factors driving changes in the two intensities over time, we assume the background intensity process is smooth, whereas the foreground process is highly variable. We regularize $\boldsymbol{\beta}_j^{(b)}$ and $\boldsymbol{\beta}_j^{(f)}$ to control the relative smoothness and volatility of the two processes (Hooten & Hobbs, 2015). Specifically, we apply separate scalar penalties to $\boldsymbol{\beta}_j^{(b)}$ and $\boldsymbol{\beta}_j^{(f)}$ (equivalent to informed prior variances; see Section 2.3.1) to constrain the background intensity to be smooth and allow the foreground intensity to be sufficiently flexible to capture short periods with high charcoal counts attributable to local fires. Optimal penalties are identified based on out-of-sample prediction (Section 2.3.3). Lacking observations of background and foreground charcoal counts (as depicted in Figure 1), the coefficients in Equation 2 would be unidentifiable given identical covariate values, without the aforementioned assumption and corresponding regularization.

Following from Equation 2, we can express the mean background (μ_b) and foreground (μ_f) charcoal counts at temporal support $\boldsymbol{\tau}_j$ as

$$\begin{aligned} \mu_b(\tau_{j,i}) &= \int_{\tau_{j,i}} \lambda_{j,b}(\tau_{j,i})\, dt = e^{\beta_{0,j}^{(b)} + \mathbf{x}(\tau_{j,i})'\boldsymbol{\beta}_j^{(b)}} |\tau_{j,i}| \\ \mu_f(\tau_{j,i}) &= \int_{\tau_{j,i}} \lambda_{j,f}(\tau_{j,i})\, dt = e^{\beta_{0,j}^{(f)} + \mathbf{x}(\tau_{j,i})'\boldsymbol{\beta}_j^{(f)}} |\tau_{j,i}|. \end{aligned} \tag{3}$$

The mean in Equation 1 can then be expressed as $\mu(\tau_{j,i}) = \mu_b(\tau_{j,i}) + \mu_f(\tau_{j,i})$, defining the total charcoal count in $\tau_{j,i}$ as a function of independent, homogeneous Poisson point processes conditional on $\lambda_{j,b}(\tau_{j,i})$ and $\lambda_{j,f}(\tau_{j,i})$. The point process model applied is similar to a Cox process in that charcoal counts are observed over disjoint time intervals ($\tau_{j,i}$) that span D_j conditional on the sum of variable intensity functions. Unlike a Cox process, however, the latent background and foreground intensities are nonstochastic. For example, adding normally distributed random error terms to the background and foreground intensity functions in Equation 2 results in two independent log-Gaussian Cox processes. The variances associated with such random error terms are unidentifiable in the current application, given we do not observe background and foreground charcoal counts.

2.1.1 | Probability of fire

Charcoal influx at time t arising from local fire events is distinguished from regional/secondary sources according to $P_j(\tau_{j,i}) \equiv \Pr\{\text{fire event local to lake } j \text{ at time } t\}$ for $t \in \tau_{j,i}$. That is, a charcoal particle arriving at lake j at time t generated conditional on $\lambda_j(\tau_{j,i}) = \lambda_{j,b}(\tau_{j,i}) + \lambda_{j,f}(\tau_{j,i})$, for $t \in \tau_{j,i}$, is labeled as a background or foreground particle according to independent Bernoulli trials with probability $P_j(\tau_{j,i})$ (Diggle, 2014). It follows that $\lambda_{j,f}(\tau_{j,i}) = P_j(\tau_{j,i})\lambda_j(\tau_{j,i})$ and $\lambda_{j,b}(\tau_{j,i}) = (1 - P_j(\tau_{j,i}))\lambda_j(\tau_{j,i})$ (Ross, 2010; see Figure 1), so that

$$P_j(\tau_{j,i}) = \frac{\lambda_{j,f}(\tau_{j,i})}{\lambda_{j,f}(\tau_{j,i}) + \lambda_{j,b}(\tau_{j,i})}. \tag{4}$$

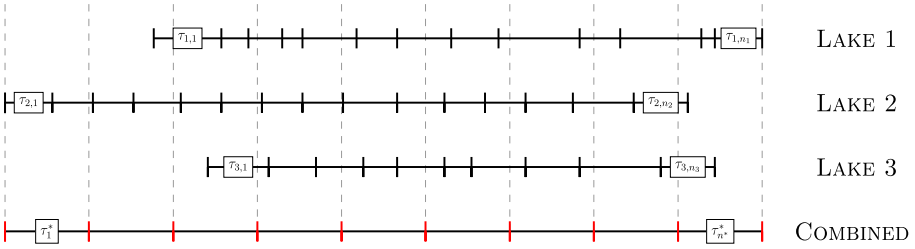

FIGURE 2 Example creation of new temporal support $\boldsymbol{\tau}^*$ from three individual lake records. Note the temporal misalignment across lakes and $\boldsymbol{\tau}^*$

Although the coefficients in Equation 2 may not all be statistically identifiable, even with the constraints placed on the background and foreground intensities, the resulting probability of fire as defined in Equation 4 is identifiable (see the Appendix).

2.1.2 | Mean FRI

A chronology of local fire events is derived after fitting the point process model by establishing a probability of fire threshold, which indicates a local fire event when exceeded. Specifically, we transform $P_j(\tau_{j,i})$ into a binary variable $Z_j(\tau_{j,i})$, with 1 indicating a local fire event, according to $P_j(\tau_{j,i}) \overset{f}{\rightarrow} Z_j(\tau_{j,i})$, where

$$f\left(P_j(\tau_{j,i})|\xi\right) = \begin{cases} 1, \text{ if } P_j(\tau_{j,i}) > \xi, \\ 0, \text{ if } P_j(\tau_{j,i}) \leqslant \xi, \end{cases}$$

for a defined probability of fire threshold ξ. The fire chronology generated depends on the specific probability of fire threshold applied. We discuss selection of an optimal threshold in Section 2.3.4.

The series of estimated fire events resulting from the application of the probability of fire threshold constitute a temporal Poisson process defined by a rate parameter (α^{-1}). A property of Poisson processes is the interarrival times, in this case, the time intervals between local fire events, are independently and identically distributed as exponential random variables with mean α. The exponential mean represents the mean FRI, whereas its inverse, the Poisson rate parameter, represents the frequency of local fires. The maximum likelihood estimate (MLE) for α is equal to the observation period divided by the number of fire events observed for a given lake, for example,

$$\hat{\alpha}_j = \frac{|D_j|}{\sum_{i=1}^{n_j} Z_j(\tau_{j,i})}.$$

We apply Bayesian inference to estimate the exponential mean parameter (α) as described in Section 2.3.4, rather than calculating the MLE, to allow for estimation of a regional mean FRI (see Section 2.2.1). It is common in fire ecology to apply a Weibull likelihood function to estimate the mean FRI. Applying a Weibull likelihood function allows for the probability of a fire event to increase as a function of time elapsed since the last fire event. We did not apply a Weibull likelihood function in the current analysis because sediment charcoal records were of relatively short length for a number of study lakes, leading to poor estimation of the two Weibull likelihood parameters.

2.2 | Multivariate model

The multivariate model follows directly from the univariate model and allows for joint estimation of fire probabilities and mean FRIs across multiple lakes. We combine observations from all lakes $\mathbf{y}_{\boldsymbol{\tau}} \equiv (\mathbf{y}'_{\boldsymbol{\tau}_1}, \mathbf{y}'_{\boldsymbol{\tau}_2}, \ldots, \mathbf{y}'_{\boldsymbol{\tau}_k})'$ and define new temporal support $\boldsymbol{\tau}^* = \left(\tau_1^*, \tau_2^*, \ldots, \tau_{n^*}^*\right)'$ to accommodate the temporal misalignment among individual lake records (Figure 2). $\boldsymbol{\tau}^*$ is defined by nonoverlapping intervals of equal length $|\tau_i^*| = |\tau_{i'}^*|, \forall\, i, i' = 1, \ldots, n^*$, that span D^*, the temporal domain spanned by all lakes combined $D^* = \bigcup_{i=1}^{n^*} \tau_i^*$.

The background intensity process for each lake is defined at temporal support $\boldsymbol{\tau}^*$, allowing charcoal counts to be pooled across lakes to estimate a regional mean background intensity. Comparable with the univariate model, we model the background intensity process for each lake on the log scale but apply a new set of cubic regression splines defined for p^* knots corresponding to temporal support $\boldsymbol{\tau}^*$. Specifically,

$$\log\left(\lambda_{j,b}(\tau_i^*)\right) = \beta_{0,j}^{(b)} + \mathbf{x}(\tau_i^*)'\boldsymbol{\beta}_j^{(b)},$$

where $\mathbf{x}(\tau_i^*)$ is a p^*-dimensional set of known cubic regression spline covariate values equal to the ith row of the $n^* \times p^*$ matrix $\mathbf{X}$.

The background intensity process defined at temporal support $\boldsymbol{\tau}^*$ is mapped back to the observed temporal support for each lake $\boldsymbol{\tau}_j$ by an $N \times n^*$ matrix $\mathbf{A}$, where $N = \sum_{j=1}^{k} n_j$. Specifically, given $\mathbf{A} \equiv (\mathbf{A}_1', \ldots, \mathbf{A}_k')'$, where each $\mathbf{A}_j$ is an $n_j \times n^*$dimensional matrix,

$$\lambda_{j,b}(\tau_{j,i}) = \mathbf{a}_{j,i}' \exp\left(\beta_{0,j}^{(b)}\mathbf{1} + \mathbf{X}\boldsymbol{\beta}_j^{(b)}\right),$$

where $\mathbf{a}_{j,i}$ is an n^*-dimensional vector equal to the ith row of $\mathbf{A}_j$ and $\mathbf{1}$ is an n^*-dimensional vector of ones. The lth entry of $\mathbf{a}_{j,i}$ is equal to

$$a_{j,i}(\tau_l^*) = |\tau_{j,i} \cap \tau_l^*|, \quad l = 1, \ldots, n^*,$$

such that

$$\sum_{l=1}^{n^*} a_{j,i}(\tau_l^*) = |\tau_{j,i}|, \; i = (1, \ldots, n_j), \; j = (1, \ldots, k).$$

A joint, regional background intensity would ideally be modeled as a spatiotemporal process across multiple lakes at temporal support $\boldsymbol{\tau}^*$. Given the time and challenges associated with collecting sediment charcoal records, however, sediment samples from regional lake networks are rarely of sufficient size (in terms of the number of lakes sampled) to allow estimation of a spatiotemporal background intensity process. In the current analysis, charcoal counts are pooled across lakes to estimate a regional mean background intensity process by assigning the $\beta_{0,j}^{(b)}$ and $\boldsymbol{\beta}_j^{(b)}$ exchangeable normal prior distributions, as described in Section 2.3, in lieu of estimating a spatiotemporal background intensity process. The foreground intensity process is modeled exactly as in the univariate model. Specifically, the foreground process is modeled at the observed temporal support for each lake ($\boldsymbol{\tau}_j$) using lake-specific foreground coefficients ($\boldsymbol{\beta}_j^{(f)}$) corresponding to a set of p knots. Probability of fire estimates are calculated for each lake independently according to Equation 4.

2.2.1 | Regional mean FRI

We seek inference regarding the regional mean FRI in addition to individual-lake mean FRIs under the multivariate model. We apply a partial pooling approach to estimate the regional mean FRI across lakes. Specifically, individual-lake mean FRI values (α_j) are assigned exchangeable lognormal priors centered on the log of a regional mean FRI (log α^*) with variance σ_{fri}^2. The partial pooling approach allows charcoal records from each lake to inform the regional average but penalizes lakes with large uncertainty in their mean FRI value estimate. The regional mean FRI variance parameter (σ_{fri}^2) quantifies the deviation of individual-lake mean FRI values from the regional average.

2.3 | Bayesian implementation

The univariate and multivariate models are completed by specifying prior distributions for remaining unknown parameters. These include background and foreground regression coefficients and mean FRI parameters.

2.3.1 | Univariate model

We assigned normal priors to regression coefficients under the univariate model: $\beta_{0,j}^{(b)} \sim \text{N}(0, \sigma_0^2)$, $\boldsymbol{\beta}_j^{(b)} \sim \text{N}(\mathbf{0}, \sigma_{b,j}^2\mathbf{S}_j^{-1})$, $\beta_{0,j}^{(f)} \sim \text{N}(0, \sigma_0^2)$, and $\boldsymbol{\beta}_j^{(f)} \sim \text{N}(\mathbf{0}, \sigma_{f,j}^2\mathbf{S}_j^{-1})$. In our specification, σ_0^2 is fixed at a large value defining a diffuse normal prior. The $\sigma_{b,j}^2$ and $\sigma_{f,j}^2$ terms represent scalar penalties (or regulators) that ensure the background process is smooth while the foreground process is sufficiently flexible to capture irregular charcoal counts subject to the constraint $\sigma_{b,j}^2 < \sigma_{f,j}^2$ (identification of optimal penalties is described in Section 2.3.3). The $p \times p$ matrix $\mathbf{S}_j$ consists of known coefficients defined as a function of the selected knot values (Wood, 2006). Note that $\mathbf{S}_j$ is not full column rank; rather, its rank is $p - 2$ given the second derivative of the boundary knots is equal to zero for natural cubic splines. Thus, the priors for $\boldsymbol{\beta}_j^{(b)}$ and $\boldsymbol{\beta}_j^{(f)}$ are improper, but can be shown to result in proper posterior distributions. Combining the likelihood from Equation 1 with the priors for the regression parameters, the joint posterior distribution for a single lake under the univariate model, using notation similar to Gelman et al., (2014), is proportional to

$$\prod_{i=1}^{n_j} \text{Pois}\left(y(\tau_{j,i})|\mu(\tau_{j,i})\right) \times \text{N}\left(\beta_{0,j}^{(b)}|\sigma_0^2\right) \times \text{N}\left(\boldsymbol{\beta}_j^{(b)}|\sigma_{b,j}^2, \mathbf{S}_j\right) \times \text{N}\left(\beta_{0,j}^{(f)}|\sigma_0^2\right) \times \text{N}\left(\boldsymbol{\beta}_j^{(f)}|\sigma_{f,j}^2, \mathbf{S}_j\right). \quad (5)$$

2.3.2 | Multivariate model

We specified identical normal priors for the foreground regression coefficients under the multivariate model as in the univariate model, and exchangeable normal priors for the background coefficients: $\boldsymbol{\beta}_0^{(b)} \sim \mathrm{N}(\mu_{0,b}\mathbf{1}, \tau_b^2\mathbf{I}_k)$ and $\boldsymbol{\beta}^{(b)} \sim \mathrm{N}(\mathbf{R}\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$, where $\boldsymbol{\beta}_0^{(b)} \equiv \left(\beta_{0,1}^{(b)}, \beta_{0,2}^{(b)}, \ldots, \beta_{0,k}^{(b)}\right)'$, $\boldsymbol{\beta}^{(b)} \equiv \left(\boldsymbol{\beta}_1^{(b)\prime}, \boldsymbol{\beta}_2^{(b)\prime}, \ldots, \boldsymbol{\beta}_k^{(b)\prime}\right)'$, $\mu_{0,b}$ and τ_b^2 are univariate mean and variance parameters, $\mathbf{1}$ is a k-dimensional vector of ones, $\mathbf{I}_k$ denotes a k-dimensional identity matrix, $\boldsymbol{\mu}_b$ is a p^*-dimensional mean vector, $\boldsymbol{\Sigma}_b$ is a $kp^* \times kp^*$ covariance matrix, and $\mathbf{R}$ is a $kp^* \times p^*$ incidence matrix equal to $\mathbf{1} \otimes \mathbf{I}_{p^*}$. The univariate variance τ_b^2 quantifies interlake variation in the intercept of the background intensity. The covariance matrix $\boldsymbol{\Sigma}_b$ can be decomposed into inter- and within-lake covariance in background coefficients. Defining a $kp^* \times kp^*$ block diagonal matrix $\mathbf{S}_b \equiv \mathrm{Diag}(\sigma_{b,1}^2\mathbf{S}^{*-1}, \sigma_{b,2}^2\mathbf{S}^{*-1}, \ldots, \sigma_{b,k}^2\mathbf{S}^{*-1})$, where $\mathbf{S}^*$ is a $p^* \times p^*$ matrix of known coefficients associated with the p^* knots defined for $\boldsymbol{\tau}^*$, we can express the background coefficient covariance matrix as $\boldsymbol{\Sigma}_b = \mathbf{L}\mathbf{S}_b\mathbf{L}'$, where $\mathbf{L}\mathbf{L}'$ is the Cholesky decomposition of $\mathbf{H} \otimes \mathbf{I}_{p^*}$, and $\mathbf{H}$ is a k-dimensional covariance matrix. The matrix $\mathbf{S}_b$ accounts for covariance in background coefficients within lakes, whereas $\mathbf{H}$ captures covariance among lakes. We apply a spatial covariance function to construct $\mathbf{H}$, although any valid covariance function can be used. Specifically, $\mathbf{H} \equiv \mathbf{H}_s(\boldsymbol{\theta}_b)$, where $(\mathbf{H}_s(\boldsymbol{\theta}_b))_{j,j'} = c(||\mathbf{s}_j - \mathbf{s}_{j'}||; \boldsymbol{\theta}_b)$, given $\mathbf{s}_j$ indicates the geographic location of lake j, and $\boldsymbol{\theta}_b$ are unknown spatial covariance parameters.

In the current analysis, we assigned normal priors to the mean intercept $\mu_{0,b} \sim \mathrm{N}\left(0, \sigma_0^2\right)$ and mean regression coefficients $\boldsymbol{\mu}_b \sim \mathrm{N}\left(\mathbf{0}, \psi^2\mathbf{I}\right)$, where σ_0^2 is fixed at a large value, while ψ^2 is set to an appropriate order of magnitude for $\boldsymbol{\beta}_j^{(b)}$ based on univariate model results. The univariate among-lake standard deviation parameter τ_b was assigned a uniform prior: $\tau_b \sim \mathrm{Unif}(a_\tau, b_\tau)$. The scalar penalty on the background deposition process for each lake $(\sigma_{b,j}^2)$ was set equal to the optimal background penalty value from the univariate model $\boldsymbol{\sigma}_b^2 \equiv (\sigma_{b,1}^2, \sigma_{b,2}^2, \ldots, \sigma_{b,k}^2)$. We applied an exponential spatial covariance function to form $\mathbf{H}$: $(\mathbf{H}_s(\boldsymbol{\theta}_b))_{j,j'} = \sigma_s^2 \exp\left(-\phi||\mathbf{s}_j - \mathbf{s}_{j'}||\right)$, where $\boldsymbol{\theta}_b \equiv (\sigma_s^2, \phi)'$. The partial sill (σ_s^2) represents spatial variance in background regression coefficients among lakes and has the potential, if it is large, to generate highly variable, unconstrained background deposition processes for individual lakes. To avoid the generation of overly flexible background deposition processes, we fixed the partial sill at one $(\sigma_s^2 = 1)$. The spatial decay parameter was treated as a free parameter and estimated applying a diffuse uniform prior: $\phi \sim \mathrm{Unif}(a_\phi, b_\phi)$. Combining the joint likelihood for charcoal counts from all lakes with the priors for the multivariate regression model parameters, the joint posterior distribution for all lakes under the multivariate model is proportional to

$$\begin{aligned}\prod_{j=1}^{k}\prod_{i=1}^{n_j} \mathrm{Pois}\left(y(\tau_{j,i})|\mu(\tau_{j,i})\right) \times \mathrm{N}\left(\boldsymbol{\beta}_0^{(b)}\middle|\mu_{0,b}, \tau_b^2\right) \times \mathrm{N}\left(\boldsymbol{\beta}^{(b)}\middle|\boldsymbol{\mu}_b, \boldsymbol{\sigma}_b^2, \mathbf{S}^*, \phi\right) \\ \times \prod_{j=1}^{k} \mathrm{N}\left(\beta_{0,j}^{(f)}\middle|\sigma_0^2\right) \times \prod_{j=1}^{k} \mathrm{N}\left(\boldsymbol{\beta}_j^{(f)}\middle|\sigma_{f,j}^2, \mathbf{S}_j\right) \times \mathrm{N}\left(\mu_{0,b}\middle|\sigma_0^2\right) \\ \times\ \mathrm{N}\left(\boldsymbol{\mu}_b\middle|\psi^2\right) \times \mathrm{Unif}(\tau_b|a_\tau, b_\tau) \times \mathrm{Unif}(\phi|a_\phi, b_\phi).\end{aligned} \tag{6}$$

We use a Metropolis-within-Gibbs Markov chain Monte Carlo (MCMC) algorithm (Robert & Casella, 2004) to sample from the posterior distributions in Equations 5 and 6.

2.3.3 | Identification of optimal penalties

Optimal background $(\sigma_{b,j}^2)$ and foreground $(\sigma_{f,j}^2)$ penalties (prior variances) were identified based on out-of-sample model prediction. Specifically, we conducted a gridded search over a range of penalty values based on initial exploratory modeling with five unique penalty values assigned to the background and foreground: 25 total combinations. The gridded search was conducted applying the univariate point process model to a single validation data set for each lake with 25% of observations held out. Prediction of held-out charcoal counts was carried out via composition sampling using posterior samples of background and foreground coefficients $(\beta_{0,j}^{(b)}, \boldsymbol{\beta}_j^{(b)}, \beta_{0,j}^{(f)}, \boldsymbol{\beta}_j^{(f)})$ to generate $\lambda_{j,b}(\tau_{j,i})$ and $\lambda_{j,f}(\tau_{j,i})$. We sampled $y_{\mathrm{ho}}(\tau_{j,i}) \sim \mathrm{Pois}(\mu(\tau_{j,i}))$ in a one-for-one fashion, where $y_{\mathrm{ho}}(\tau_{j,i})$ indicates a held-out observation. The resulting samples of $y_{\mathrm{ho}}(\tau_{j,i})$ represent the posterior predictive distribution of $y_{\mathrm{ho}}(\tau_{j,i})$. Optimal penalty terms were identified as the background and foreground variances that minimized the posterior predictive loss calculated for the hold-out data (Gelfand & Ghosh, 1998). The posterior predictive loss rewards accuracy of predictions with a penalty for large variance in predictions indicative of overparameterization. For the multivariate model, we applied the optimal background and foreground penalty values identified for each lake under the univariate model.

2.3.4 | Mean FRI

The mean FRI under the univariate and multivariate models is estimated using FRIs calculated after applying a probability of fire threshold to sampled posterior probability of fire values at each iteration of the Gibbs sampler. Specifically, for the ℓth iteration of the Gibbs sampler, we obtain a set of fire event times $\left(t_{j,1}^{(\text{fire})}, t_{j,2}^{(\text{fire})}, \ldots, t_{j,m_j}^{(\text{fire})}\right)^{(\ell)}$, where m_j is the total number of fire events observed for lake j, by conditioning samples on $Z(\tau_{j,i})^{(\ell)} = 1$. A given FRI is equal to the elapsed time between two consecutive fire events. For example, for lake j and iteration ℓ, the rth FRI is given by $\text{FRI}_{j,r}^{(\ell)} = \left(t_{j,r+1}^{(\text{fire})} - t_{j,r}^{(\text{fire})}\right)^{(\ell)}$, for $r = (1, 2, \ldots, m_j - 1)$.

We assigned a semi-informative, conjugate inverse-gamma prior to the mean FRI parameter $\alpha_j \sim \text{InvGamma}(a_\alpha, b_\alpha)$ under the univariate model centering its density over the possible range of FRIs for study lakes based on previous analyses (Kelly et al., 2013). Combining the prior with the exponential likelihood of the FRIs (see Section 2.1.2), we obtain an inverse-gamma posterior distribution for α_j conditional on the derived set of FRIs. Specifically, for the ℓth iteration of the Gibbs sampler, $\mathbf{FRI}_j^{(\ell)} = \left(\text{FRI}_{j,1}, \text{FRI}_{j,2}, \ldots, \text{FRI}_{j,m_j-1}\right)'$ and $\alpha_j^{(\ell)} | \mathbf{FRI}_j^{(\ell)} \sim \text{InvGamma}(a_\alpha + m_j^{(\ell)} - 1, b_\alpha + \sum_{r=1}^{m_j^{(\ell)}-1} \text{FRI}_{j,r}^{(\ell)})$.

We seek inference regarding the regional mean FRI in addition to individual-lake mean FRIs under the multivariate model. We assigned diffuse uniform priors to the regional mean FRI (α^*) and the interlake standard deviation (σ_{fri}). Combining the priors and the exponential likelihood of the FRIs, the joint posterior distribution conditional on the model parameters defined in Equation 6 is proportional to

$$\prod_{j=1}^{k} \prod_{r=1}^{m_j-1} \text{Exp}(\text{FRI}_{j,r} | \alpha_j) \times \prod_{j=1}^{k} \text{N}(\log\ \alpha_j | \log\ \alpha^*, \sigma_{\text{fri}}^2) \times \text{Unif}(\alpha^* | a_*, b_*) \times \text{Unif}(\sigma_{\text{fri}} | a_\sigma, b_\sigma).$$

We estimated mean FRI values under the univariate and multivariate models using a range of probability of fire thresholds: $\boldsymbol{\xi} = (0.50, 0.55, \ldots, 1.00)'$. Ideally, we would select the probability of fire threshold yielding the greatest accuracy in terms of identifying local fire events (i.e., the most accurate mean FRI estimate). A common challenge with the use of sediment charcoal records, however, is the lack of independent fire history data to conduct model validation. In the absence of independent fire history data, we sought an optimal probability of fire threshold in the sense of providing a precise mean FRI estimate based on a consistent fire chronology for each posterior sample of $\mathbf{P}_j(\boldsymbol{\tau}_j) \equiv (P(\tau_{j,1}), P(\tau_{j,2}), \ldots, P(\tau_{j,n_j}))'$. We estimated the coefficient of variation for posterior mean FRI samples (i.e., $\tilde{\sigma}/\tilde{\mu}$, where $\tilde{\sigma}$ and $\tilde{\mu}$ are the posterior sample standard deviation and mean, respectively). We selected the probability of fire threshold that minimized the coefficient of variation as the optimal threshold. The optimal threshold provided the most consistent mean FRI estimate based on posterior samples. This approach is similar to previous paleoecological approaches relying on sensitivity analysis to identify an optimal threshold (Clark et al., 1996).

3 | MODEL APPLICATION

We apply the point process model to both simulated data and to sediment charcoal records from a 13-lake network in interior Alaska. The use of simulated data allows for proper model validation, which is difficult using sediment charcoal records due to the lack of observed fire data to compare with the probability of fire estimates.

3.1 | Simulation study

We tested the accuracy of the point process model by applying it to simulated sediment charcoal records generated using CharSim (Higuera et al., 2007). CharSim is a semimechanistic model that generates fires on a landscape and maps the subsequent deposition of charcoal particles to a target lake. The amount of charcoal deposited in the target lake is proportional to the size of the fire, its proximity to the target lake, the atmospheric injection height of charcoal particles during the fire, and secondary charcoal deposition and sediment mixing within the lake (see Higuera et al., 2007, for additional details).

We applied the univariate point process model to a simulated sediment charcoal record generated by CharSim to mimic a fire regime consistent with historic fire regimes in the Alaskan boreal forest (as described in table 2 of Higuera et al., 2007). We defined a binary variable for each sample interval of the simulated record, with 1 indicating a local fire event within 100 m of the target lake within an interval and 0 indicating no local fires within an interval. Probability of fire estimates from the point process model were converted to binary fire occurrences, as described in Section 2.1.2, and compared with true fire occurrence values to determine the percentage of fires correctly identified.

The simulated record was 4,760 years long divided into 238 equal-length sample intervals (i.e., each interval was 20 years). There were 41 fires within 100 m of the target lake, leading to a true mean FRI of 116 years. The point process model correctly

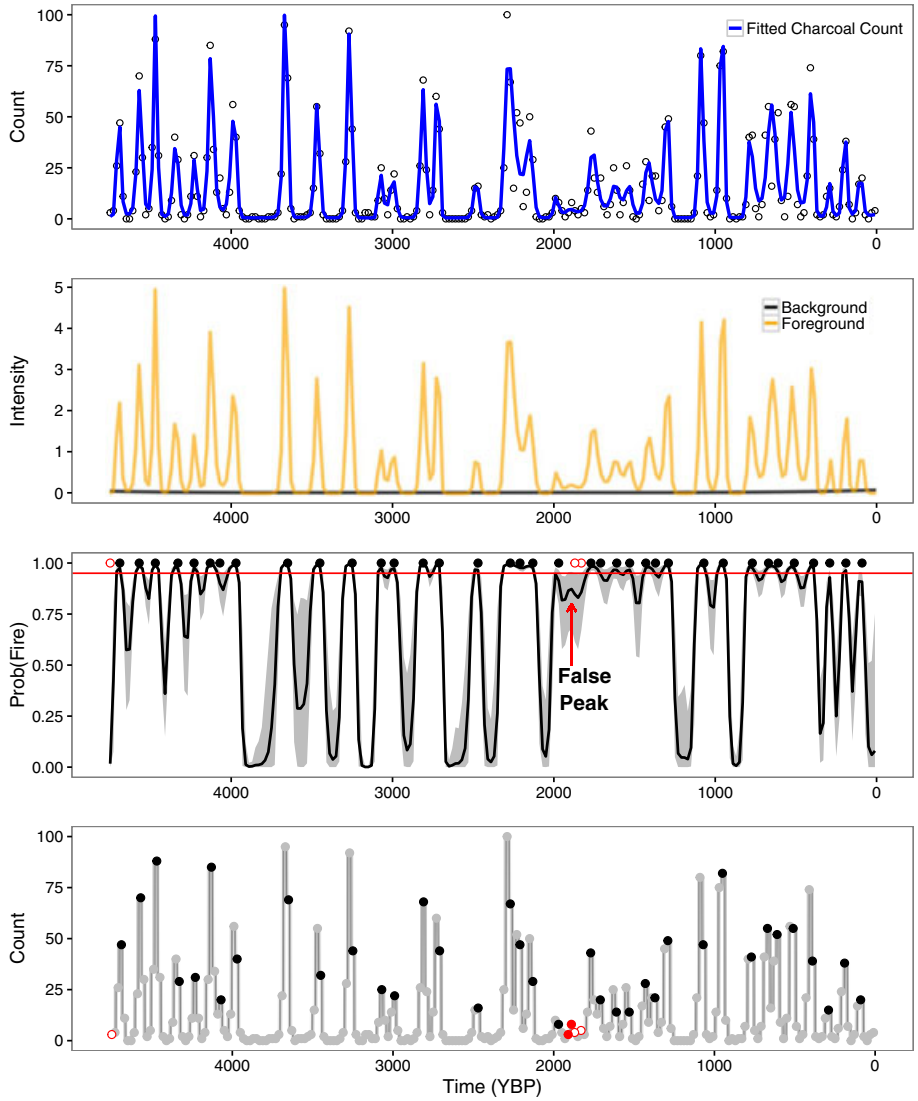

FIGURE 3 Univariate model results for simulated sediment charcoal record generated by CharSim. Upper panel indicates observed charcoal counts along with the posterior mean charcoal count (blue line). Second panel illustrates posterior mean foreground (orange line) and background (black line) intensities. Third panel plots posterior mean probability of fire estimates for each observed time interval (black line) along with the upper and lower bounds of the 95% credible interval (gray shading) and the optimal threshold (red line). The points in the third panel correspond to true fire events occurring during the sample interval, with black dots delineating correctly identified fires and red open dots delineating missed fires. The arrow highlights the single falsely identified fire during the simulated study period. Lower panel indicates observed charcoal counts, with the color and shape indicating whether the count was correctly identified as a true fire (black points), no fire (gray points), missed true fire (red open circles), or falsely identified fire (red points). YBP = years before present

identified 38 of the 41 fires occurring over the simulated study period, applying an optimal threshold value of 0.95, corresponding to a 93% fire identification rate (Figure 3). The model was accurate in its identification of local fires with only a single falsely identified fire between 1,880 and 1,920 years before present (YBP). Despite the accuracy of the point process model in identifying true local fires, the mean FRI was overestimated: posterior mean FRI equaled 191 years (95% credible interval: 135–271 years). The upward bias of the point process model in estimating the mean FRI was due to the model's inability to separate fires occurring in close temporal proximity as unique fire events. For example, there were 40 sample intervals that included a true fire in the simulated record (note that two fires occurred within a single sample interval), but only 25 unique threshold exceedances (or peaks) were estimated. A more in-depth discussion of source of bias is provided in Section 4. Model results were sensitive to the definition of a local fire event. Specifically, if a local fire is defined as occurring within 1,000 m of the target lake (rather than 100 m), the fire identification rate drops to 75%.

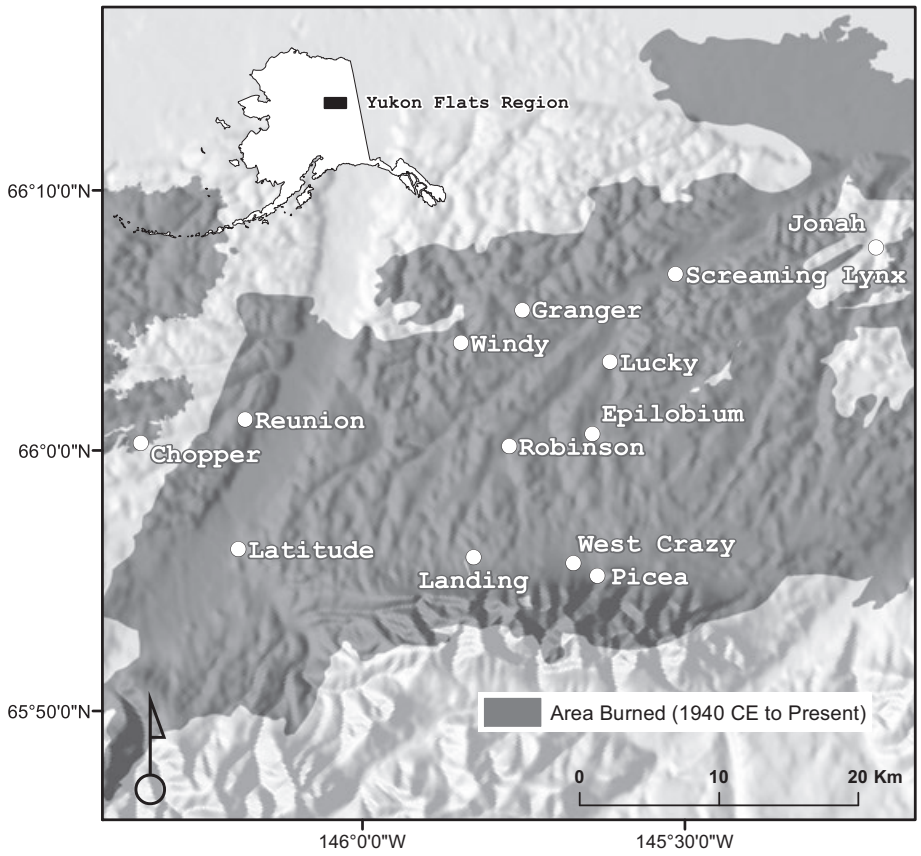

FIGURE 4 Location of study lakes within the Yukon Flats region of Alaska relative to areas burned in wildland fire since 1940 Common Era (CE)

3.2 | Yukon Flats

We applied the univariate and multivariate point process models to previously published sediment charcoal records from 13 lakes in the Yukon Flats region of Alaska (Figure 4; Kelly et al., 2013). The Yukon Flats region is dominated by boreal forests and has a fire regime characterized by stand-replacing fires with return intervals of several decades to centuries.

3.2.1 | Univariate model

We applied the univariate model to each of the 13 lakes in the Yukon Flats data set and the multivariate model to all lakes jointly. Mean FRI values from the univariate model varied from roughly 134 years for fires local to Chopper Lake to roughly 356 years for fires local to Screaming Lynx Lake. Table 1 provides mean FRI value estimates for each of the 13 lakes. Optimal threshold values providing the most precise mean FRI estimate for each lake varied from 0.50 to 0.75 (Table 1). Results of the univariate model for Chopper and Screaming Lynx Lakes are provided in Figure 5 (similar plots are provided for all lakes in the Supporting Information along with additional plots for Chopper and Screaming Lynx Lakes illustrating the identification of an optimal threshold and uncertainty in local fire chronologies; Chopper and Screaming Lynx Lakes are selected to illustrate the lakes with the shortest and longest mean FRI, respectively).

Geospatial fire perimeter data for the state of Alaska date back to 1940 Common Era (CE; Alaska Fire Service, 2016). Restricting the fire perimeter data to fires local to lakes within the Yukon Flats data set, there were five fires since 1940 within 100 m of one or more of the 13 study lakes (a threshold of 100 m from the lake edge was used to distinguish local fire events based on results of the simulation study) the earliest of which occurred in 1985. Although the local fire record is insufficient in length to conduct proper model validation (i.e., roughly 70 years of local fire data for a study period of over 10,000 years is less than 1% data coverage), we can compare the true occurrence of local fires to local fires identified by the point process model to assess model performance. Two of the five fires local to at least one study lake occurred after the end of the sediment charcoal records for local lakes: Big Creek Fire in 2009 and Discovery Creek Fire in 2013. The Preacher Creek Fire in 2004 occurred local to Picea and Epilobium Lakes; however, the point process model did not identify a local fire for either of these lakes in the most recent 50 years. There were two unnamed fires, the first in 1985 and the second in 1988, local to several study lakes.

TABLE 1 Summary of univariate and multivariate Poisson process model results for each lake in the Yukon Flats data set. Mean fire return interval (FRI) is equal to the posterior mean FRI with 95% credible intervals (Cred. Int.) in parentheses

	Univariate model			Multivariate model		
Lake	**Optimal threshold**	**Mean FRI**	**Cred. Int. width**	**Optimal threshold**	**Mean FRI**	**Cred. Int. width**
Chopper	0.60	134 (91,197)	106	0.85	144 (97,216)	119
Epilobium	0.50	197 (133,292)	159	0.50	201 (136,293)	157
Granger	0.55	286 (187,436)	249	0.50	246 (165,368)	203
Jonah	0.50	159 (111,228)	117	0.55	148 (107,207)	100
Landing	0.70	303 (206,445)	239	0.70	291 (201,419)	218
Latitude	0.75	158 (106,235)	129	0.75	137 (96,194)	98
Lucky	0.75	136 (92,200)	108	0.75	138 (93,206)	114
Picea	0.50	318 (227,448)	221	0.50	312 (223,429)	206
Reunion	0.70	244 (167,354)	186	0.70	237 (165,345)	180
Robinson	0.55	142 (94,213)	119	0.80	124 (87,180)	93
Screaming Lynx	0.60	356 (255,497)	242	0.55	357 (255,502)	247
West Crazy	0.50	204 (133,315)	182	0.70	200 (129,306)	177
Windy	0.50	162 (110,239)	128	0.65	155 (107,226)	119

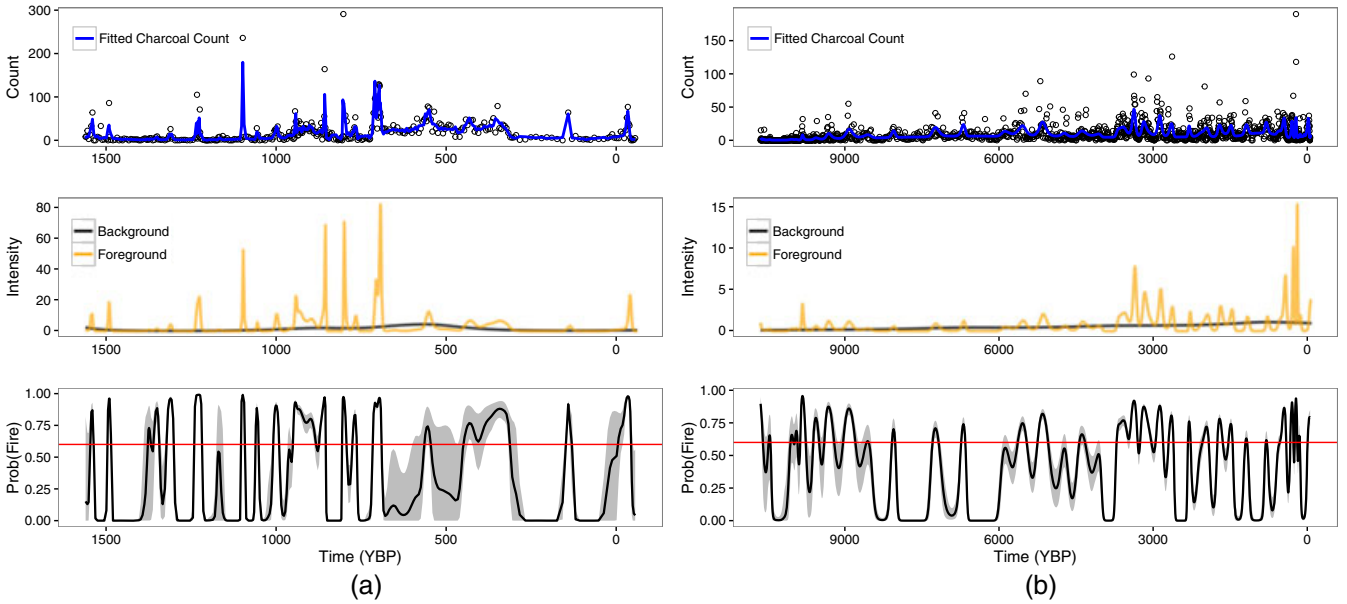

FIGURE 5 Univariate model results for (a) Chopper Lake and (b) Screaming Lynx Lake. Upper panel indicates observed charcoal counts along with the posterior mean charcoal count (blue line). Middle panel illustrates posterior mean foreground (orange line) and background (black line) intensities. Lower panel plots posterior mean probability of fire estimates for each observed time interval (black line) along with the upper and lower bounds of the 95% credible interval (gray shading) and the optimal threshold (red line). YBP = years before present

A fire event was identified by the point process model within the past 25 years in four out of seven lakes local to the 1985 fire and two of five lakes local to the 1988 fire.

3.2.2 | Multivariate model

Joint probability of fire estimates generated using the multivariate model varied in magnitude from the univariate model results. Figure 6 presents the multivariate model results for Chopper and Screaming Lynx Lakes. Comparing the results presented in Figure 6 to those from the univariate model (Figure 5), the probability of fire estimates for Chopper Lake are slightly higher under the multivariate model than the univariate model, whereas the probability of fire estimates for Screaming Lynx Lake are roughly consistent between the two models. In general, probability of fire estimates were higher under the multivariate model than the univariate model. The different magnitudes of probability of fire estimates under the multivariate model necessitated calculating new optimal fire thresholds for each lake (Table 1). Optimal threshold values ranged from 0.50 to 0.85 for the

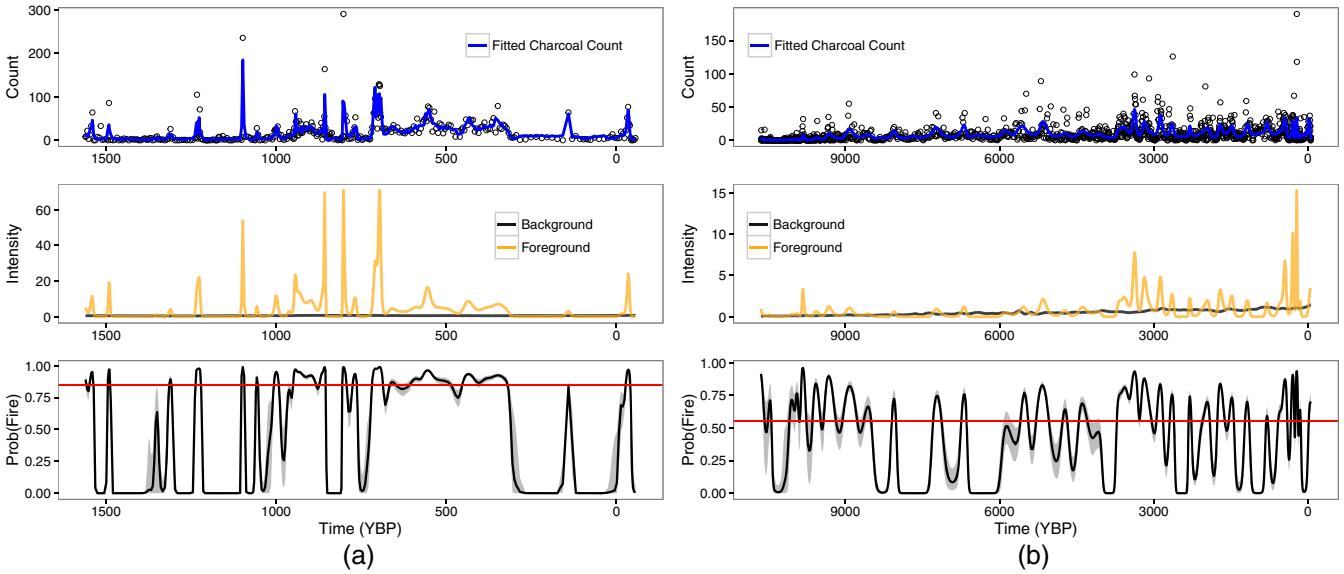

FIGURE 6 Multivariate model results for (a) Chopper Lake and (b) Screaming Lynx Lake. Upper panel indicates observed charcoal counts along with the posterior mean charcoal count (blue line). Middle panel illustrates posterior mean foreground (orange line) and background (black line) intensities. Lower panel plots posterior mean probability of fire estimates for each observed time interval (black line) along with the upper and lower bounds of the 95% credible interval (gray shading) and the optimal threshold (red line). YBP = years before present

multivariate model consistent with slightly higher probability of fire estimates compared to the univariate model. Mean FRI values estimated using the multivariate model were consistent with the mean FRI values estimated for each lake using the univariate model; however, the credible interval widths for the multivariate model were narrower than under the univariate model. Specifically, 10 out of 13 lakes had narrower credible intervals under the multivariate model than the univariate model (Table 1). The average credible interval width for the mean FRI was 156 years under the multivariate model versus 168 years under the univariate model. In addition to lake-specific, local mean FRIs, we applied the joint probability of fire estimates and optimal thresholds for each lake to estimate a joint regional mean FRI, as described in Section 2.2.1. We estimated a regional mean FRI over the study period (10,680 to −59 YBP, relative to 1950 CE) of roughly 187 years (95% credible interval: 136–261 years) for the Yukon Flats region based on the lake network data.

The multivariate model provides inference on regional background charcoal deposition. Applying an exponential spatial covariance function to describe spatial correlation in background regression coefficients (as described in Section 2.3.2) resulted in an estimated effective spatial range of approximately 16.5 km (95% credible interval: 14.4–19.1 km) where the effective spatial range is defined as the distance at which the correlation drops to 0.05. This suggests the parameters describing local, background charcoal intensity for individual lakes are similar for lakes within 20 km of each other. Finally, the multivariate model provides an estimate of the background charcoal deposition in each lake over the entire study period, although the sediment charcoal records for most lakes are shorter than the full study period. The background charcoal intensities for each lake in the Yukon Flats network are plotted together in Figure 7 along with a regional loess smooth function. The background charcoal deposition for most lakes exhibited a similar pattern, with a long-term increase in background charcoal deposition from 6,000 YBP to present, a sharp increase in background deposition roughly 3,000 YBP, and a secondary increase 1,000 YBP followed by a decrease in background deposition roughly 500 YBP.

4 | DISCUSSION

The use of sediment charcoal records to reconstruct past fire regimes is challenging given that charcoal counts rather than past fire occurrences are observed. Further, observed charcoal counts include charcoal generated during local fires as well as charcoal stemming from regional fire activity and secondary sources. The goal of our analysis was to construct an integrated statistical framework for local fire identification and the estimation of mean FRIs based on sediment charcoal records from individual lakes building on previous approaches to paleofire reconstruction. We further sought to advance existing approaches to reconstruct regional fire history through the development of a multivariate model, which combines sediment charcoal records from multiple lakes to identify local fires and jointly estimate mean FRIs at individual-lake and regional scales. Here, we discuss

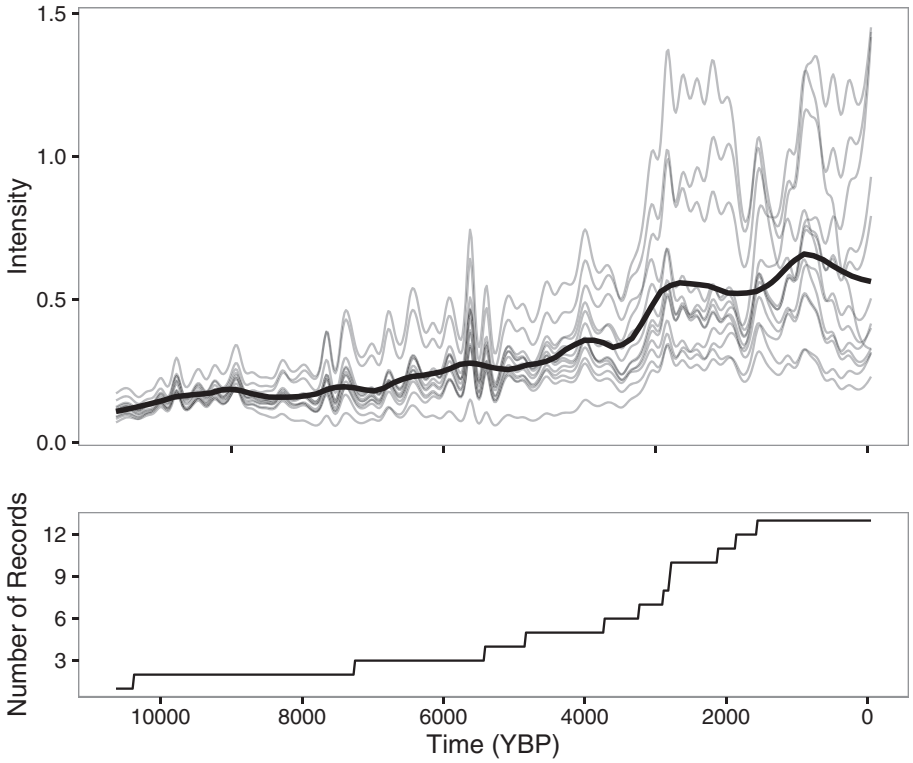

FIGURE 7 Background charcoal deposition intensity for each lake in the Yukon Flats data set over the entire study period (10,680 to −59 years before present [YBP], relative to 1950 Common Era [CE]) based on the multivariate point process model. The bold line in the upper panel is a regional loess smooth function reflecting mean changes in background charcoal deposition across all lakes (fit using a span of 0.15). The lower panel indicates the number of lake records used to estimate background charcoal deposition over time

the key results of the application of the univariate and multivariate point process models to the Yukon Flats data set and connect the results of the current analysis to previous studies in the same region.

Mean FRI estimates from the univariate and multivariate point process models applied to the Yukon Flats data set ranged from 100 to 350 years (Table 1). The lakes with the largest mean FRI estimates have sediment charcoal records dating back the longest among study lakes: Reunion, Granger, Landing, Picea, and Screaming Lynx Lakes all have records that date back at least 5,000 YBP. This pattern is consistent with previous interpretations of Holocene fire history in Alaskan boreal forests, which highlight increased fire activity over the last several thousand years beginning with the local arrival of black spruce between 6,000 and 3,000 YBP (Higuera et al., 2009; Kelly et al., 2013; Lynch et al., 2004). We observe a similar pattern of increased fire activity in plots of the background charcoal deposition intensity for each lake over the study period derived from the multivariate model (Figure 7). The increase in fires across the Yukon Flats region from 6,000 to 3,000 YBP reflected in the background intensity suggests that lakes with sediment charcoal records dating back prior to 3,000 YBP should have longer mean FRI values than lakes with relatively short records. A secondary peak in the background charcoal deposition intensity is observable around 500 YBP coincident with the Medieval Climate Anomaly, a period of increased temperatures and drought frequency (1,000–500 YBP), followed by the Little Ice Age, a period of cooler and wetter climatic conditions (500–80 YBP). Finally, modeled background charcoal intensities for individual lakes indicate an increase in biomass burned in recent decades, although the increase is not reflected in the regional loess smoother (Figure 7). The modeled background charcoal intensities are consistent with composite CHAR records (i.e., mean charcoal accumulation rate among lakes) calculated using the Yukon Flats data (Kelly et al., 2013).

The univariate Bayesian point process model provides a model-based approach to estimate probability of fire values associated with sediment charcoal records from a single lake and convert those probabilities into mean FRI estimates. The multivariate model allows for correlation among lakes in the parameters used to estimate the background charcoal deposition intensity. The background intensity reflects regional charcoal sources and exhibits low-frequency changes over time associated with factors such as species composition and climate. As such, background charcoal deposition should be similar among lakes in the same region with a high potential for correlation in background deposition process parameters. As expected, mean FRI estimates for each lake are similar based on the univariate and multivariate models (Table 1). The multivariate model, however, resulted in

mean FRI estimates with reduced uncertainty. Specifically, the 95% credible interval for the mean FRI was narrower in 10 out of 13 lakes in the Yukon Flats network, with a mean credible interval width of 156 years for the multivariate model versus 168 years for the univariate model. The reduced uncertainty in mean FRI estimates under the multivariate model provides some evidence that background charcoal deposition is indeed correlated among lakes located in the same region and that we can reduce uncertainty in estimates of background charcoal deposition by accounting for such correlation. The effective spatial range for the background process regression parameters ($\boldsymbol{\beta}_j^{(b)}$) was estimated to be 16.5 km and provides some indication of the distance within which background charcoal deposition is similar among lakes within the Yukon Flats region. This estimate is consistent with the previous analysis using the Yukon Flats data set, which found significant correlation between composite CHAR and regional area burned within a 20-km radius (Kelly et al., 2013).

We estimated a regional mean FRI of roughly 187 years (95% credible interval: 136–261 years) for the Yukon Flats over the study period, applying the partial pooling approach described in Section 2.2.1. The partial pooling approach also produces estimates of mean FRI values for individual lakes similar to the univariate and multivariate results presented in Table 1. However, we do not see the same reduction in uncertainty in individual-lake mean FRI estimates when conducting partial pooling. Specifically, credible interval widths were narrower in only six out of 13 lakes, with the remaining intervals comparable to univariate model results. The partial pooling approach adds two additional parameters ($\alpha^*, \sigma^2_{\text{fri}}$) and combines uncertainty in mean FRI values across lakes. As such, it is not surprising that the partial pooling approach does not lead to the same reductions in uncertainty as generating individual-lake mean FRI estimates based on the multivariate model. We envision the partial pooling approach being applied only in the setting where a researcher is interested in estimating a regional mean FRI; otherwise, the individual-lake approach is preferred.

The univariate point process model we developed achieves our goal of developing an integrated statistical framework for local fire identification and estimation of mean FRIs based on sediment charcoal records for individual lakes. The Bayesian hierarchical model structure allows for tractable propagation of additional uncertainty sources in paleofire reconstructions. In particular, uncertainty in sediment age models can be integrated by treating the ages of sediment core sections as unobserved, latent variables in the point process model. The multivariate extension of the point process model provides a novel approach for paleofire reconstruction applying multiple lake records to make inferences at both individual-lake and regional scales. Specifically, the multivariate model provides estimates of individual-lake mean FRIs, a regional mean FRI, and background charcoal deposition intensity indicative of regional biomass burned. When applied to the Yukon Flats data set, pooling of individual-lake records under the multivariate model led to reduced uncertainty in individual-lake mean FRIs. We expect the multivariate model to provide even greater reductions in uncertainty in individual-lake mean FRI values and improved estimates of regional parameters, including the regional mean FRI, when applied to larger regional lake networks (i.e., networks with more than 20 lakes), assuming that all lakes share a common regional fire regime.

ACKNOWLEDGEMENTS

The authors thank Shannon Kay and collaborators on the PalEON project as well as participants of the PAGES Paleofire Workshop, Harvard Forest, Sep. 27 to Oct. 2, 2015, Patrick Bartlein in particular, for their useful comments and feedback on the modeling approach. We thank Christopher Paciorek and two anonymous reviewers whose thoughtful comments helped to improve the paper. This work was supported by National Science Foundation Grants: DMS-1513481, EF-1137309, MSB-1241874 and EF-1253225 (M.S. Itter and A.O. Finley); MSB-1241856 (M.B. Hooten); MSB-1241846 (P.E. Higuera); BCS-1437074 and MSB-1241870 (J.R. Marlon). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

REFERENCES

Alaska Fire Service (2016). Alaska Fire History: Fire perimeters updated February 2, 2016.

Carcaillet, C., Bergeron, Y., Richard, P. J., Fréchette, B., Gauthier, S., & Prairie, Y. T. (2001). Change of fire frequency in the eastern Canadian boreal forests during the Holocene: Does vegetation composition or climate trigger the fire regime? *Journal of Ecology*, *89*(6), 930–946.

Clark, J. S. (1988). Particle motion and the theory of charcoal analysis: Source area, transport, deposition, and sampling. *Quaternary Research*, *30*(1), 67–80.

Clark, J. S. (1990). Fire and climate change during the last 750 yr in northwestern Minnesota. *Ecological Monographs*, *60*(2), 135–159.

Clark, J. S., & Royall, P. D. (1996). Local and regional sediment charcoal evidence for fire regimes in presettlement north-eastern North America. *Journal of Ecology*, *84*(3), 365–382.

Clark, J. S., Royall, P. D., & Chumbley, C. (1996). The role of fire during climate change in an eastern deciduous forest at Devil's Bathtub, New York. *Ecology*, *77*(7), 2148–2166.

Diggle, P. J. (2014). *Statistical analysis of spatial and spatio-temporal point patterns* (3rd ed.). Boca Raton, FL: CRC Press.

Gavin, D. G., Brubaker, L. B., & Lertzman, K. P. (2003). An 1800-year record of the spatial and temporal distribution of fire from the west coast of Vancouver Island, Canada. *Canadian Journal of Forest Research*, *33*(4), 573–586.

Gavin, D. G., Hu, F. S., Lertzman, K., & Corbett, P. (2006). Weak climatic control of stand-scale fire history during the late Holocene. *Ecology*, *87*(7), 1722–1732.

Gelfand, A. E., & Ghosh, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika*, *85*(1), 1–11.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Higuera, P. E., Brubaker, L. B., Anderson, P. M., Hu, F. S., & Brown, T. A. (2009). Vegetation mediated the impacts of postglacial climate change on fire regimes in the south-central Brooks Range, Alaska. *Ecological Monographs*, *79*(2), 201–219.

Higuera, P. E., Peters, M. E., Brubaker, L. B., & Gavin, D. G. (2007). Understanding the origin and analysis of sediment-charcoal records with a simulation model. *Quaternary Science Reviews*, *26*(1314), 1790–1809.

Higuera, P. E., Whitlock, C., & Gage, J. A. (2010). Linking tree-ring and sediment-charcoal records to reconstruct fire occurrence and area burned in subalpine forests of Yellowstone National Park, USA. *The Holocene*, *19*, 996–1014.

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, *85*(1), 3–28.

Kelly, R., Chipman, M. L., Higuera, P. E., Stefanova, I., Brubaker, L. B., & Hu, F. S. (2013). Recent burning of boreal forests exceeds fire regime limits of the past 10,000 years. *Proceedings of the National Academy of Sciences*, *110*(32), 13055–13060.

Long, C. J., Whitlock, C., Bartlein, P. J., & Millspaugh, S. H. (1998). A 9000-year fire history from the Oregon Coast Range, based on a high-resolution charcoal study. *Canadian Journal of Forest Research*, *28*(5), 774–787.

Lynch, J. A., Clark, J. S., & Stocks, B. J. (2004). Charcoal production, dispersal, and deposition from the Fort Providence experimental fire: Interpreting fire regimes from charcoal records in boreal forests. *Canadian Journal of Forest Research*, *34*(8), 1642–1656.

Marlon, J. R., Bartlein, P. J., Gavin, D. G., Long, C. J., Anderson, R. S., Briles, C. E., … Power, M. J. (2012). Long-term perspective on wildfires in the western USA. *Proceedings of the National Academy of Sciences*, *109*(9), E535–E543.

Peters, M. E., & Higuera, P. E. (2007). Quantifying the source area of macroscopic charcoal with a particle dispersal model. *Quaternary Research*, *67*(2), 304–310.

Power, M. J., Marlon, J., Ortiz, N., Bartlein, P. J., Harrison, S. P., Mayle, F. E., … Zhang, J. H. (2008). Changes in fire regimes since the Last Glacial Maximum: An assessment based on a global synthesis and analysis of charcoal data. *Climate Dynamics*, *30*(7), 887–907.

Robert, C., & Casella, G. (2004). *Monte Carlo statistical methods*. New York, NY: Springer-Verlag.

Ross, S. M. (2010). *Introduction to probability models* (10th ed.). Burlington, MA: Academic Press.

Whitlock, C., & Millspaugh, S. H. (1996). Testing the assumptions of fire-history studies: An examination of modern charcoal accumulation in Yellowstone National Park, USA. *The Holocene*, *6*(1), 7–15.

Wood, S. (2006). *Generalized additive models: An introduction with R*. Boca Raton, FL: CRC press.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Itter MS, Finley AO, Hooten MB, et al. A model-based approach to wildland fire reconstruction using sediment charcoal records. *Environmetrics*. 2017;28:e2450. https://doi.org/10.1002/env.2450

APPENDIX A: PROBABILITY OF FIRE IDENTIFACTION

The probability of fire for a given sample interval $\tau_{j,i}$ as defined in Equation 4 can be expressed as

$$\frac{e^{\beta_{0,j}^{(f)}+\mathbf{x}(\tau_{j,i})'\boldsymbol{\beta}_j^{(f)}}}{e^{\beta_{0,j}^{(f)}+\mathbf{x}(\tau_{j,i})'\boldsymbol{\beta}_j^{(f)}}+e^{\beta_{0,j}^{(b)}+\mathbf{x}(\tau_{j,i})'\boldsymbol{\beta}_j^{(b)}}},$$

which can be simplified to

$$\frac{1}{1+e^{-\left(\beta_{0,j}^{*}+\mathbf{x}(\tau_{j,i})'\boldsymbol{\beta}_j^{*}\right)}},$$

where $\beta_{0,j}^{*}=\beta_{0,j}^{(f)}-\beta_{0,j}^{(b)}$ and $\boldsymbol{\beta}_j^{*}=\boldsymbol{\beta}_j^{(f)}-\boldsymbol{\beta}_j^{(b)}$, thereby proving the identifiability of the probability of fire $P_j(\tau_{j,i})$. The simplified probability of fire is equivalent to the mean response of a logistic regression model fit using a binary variable indicating the occurrence of a local fire within a given sample interval. Specifically, we can express the odds of a local fire in a given sample interval as

$$\frac{\lambda_{j,f}(\tau_{j,i})}{\lambda_{j,b}(\tau_{j,i})}=\frac{e^{\beta_{0,j}^{(f)}+\mathbf{x}(\tau_{j,i})'\boldsymbol{\beta}_j^{(f)}}}{e^{\beta_{0,j}^{(b)}+\mathbf{x}(\tau_{j,i})'\boldsymbol{\beta}_j^{(b)}}}$$

such that the log odds are given by $\beta_{0,j}^{*}+\mathbf{x}(\tau_{j,i})'\boldsymbol{\beta}_j^{*}$. Thus, there is a direct connection between the mean probability of fire function defined using Poisson count data in Equation 4 and a logistic regression model used to estimate the probability of fire using Bernoulli observations of fire occurrence/nonoccurrence.