Spatiotemporal prediction of wildfire extremes with Bayesian finite sample maxima

Maxwell B. Joseph*[1], Matthew W. Rossi[1], Nathan P. Mietkiewicz[1], Adam L. Mahood[1], Megan E. Cattau[1], Lise Ann St. Denis[1], R. Chelsea Nagy[1], Virginia Iglesias[1], John T. Abatzoglou[2], and Jennifer K. Balch[1]

[1]Earth Lab, University of Colorado Boulder

[2]Department of Geography, University of Idaho

Abstract

Wildfires are becoming more frequent in parts of the globe, but predicting where and when wildfires occur remains difficult. To predict wildfire extremes across the contiguous United States, we integrate a 30 year wildfire record with meteorological and housing data in spatiotemporal Bayesian statistical models with spatially varying nonlinear effects. We compared different distributions for the number and sizes of large fires to generate a posterior predictive distribution based on finite sample maxima for extreme events (the largest fires over bounded spatiotemporal domains). A zero-inflated negative binomial model for fire counts and a lognormal model for burned areas provided the best performance. This model attains 99% interval coverage for the number of fires and 93% coverage for fire sizes over a six year withheld data set. Dryness and air temperature strongly predict extreme wildfire probabilities. Housing density has a hump-shaped relationship with fire occurrence, with more fires occurring at intermediate housing densities. Statistically, these drivers affect the chance of an extreme wildfire in two ways: by altering fire size distributions, and by altering fire frequency, which influences sampling from the tails of fire size distributions. We conclude that recent extremes should not be surprising, and that the contiguous United States may be on the verge of even larger wildfire extremes.

*maxwell.b.joseph@colorado.edu

Key words: fire; wildfire; Bayesian; spatiotemporal; extremes; climate.

Introduction

Wildfire frequency and burned area has increased over the past couple decades in the United States (Dennison et al. 2014; Westerling 2016), and elsewhere (Krawchuk et al. 2009; Pechony and Shindell 2010). In addition to the ecological and smoke impacts associated with increased burned area, there has been an increasing interest in extreme wildfires (Williams 2013) given their impact on human lives and infrastructure (Kochi et al. 2010; Diaz 2012). While case studies of particular extremes provide insight into what caused past events (Peterson et al. 2015; Nauslar, Abatzoglou, and Marsh 2018), predictions of future extremes at a national level could inform disaster related resource allocation. The term "extreme" has multiple meanings with respect to wildfires (Tedim et al. 2018), and in this paper we take consider an extreme wildfire to be a fire with the largest burned area over a bounded spatiotemporal domain, i.e., the block maximum within a spatial region and a temporal interval (Coles et al. 2001). For example, the block maxima for widlfires across the contiguous U.S. can be defined on a yearly basis (Figure 1).

Factors driving wildfire extremes vary in space and time (Barbero et al. 2014), but it is unclear how best to account for this in a predictive model. Previous efforts have used year or region-specific models, aggregating over space or time (Bermudez et al. 2009), temporally or spatially explicit models (Mendes et al. 2010), and spatial models with year as a covariate (Díaz-Avalos, Juan, and Serra-Saurina 2016). Recently, rich spatiotemporal models have been described with linear, spatially constant covariate effects (Serra, Saez, Juan, et al. 2014; Serra, Saez, Mateu, et al. 2014). However, linear, spatially constant effects are suboptimal over large spatial domains with nonlinear drivers (Fosberg 1978, Goodrick (2002), Preisler et al. (2004); Preisler and Westerling 2007; Balshi et al. 2009; Krawchuk et al. 2009; Pechony and Shindell 2009; Vilar et al. 2010; Woolford et al. 2011; Woolford et al. 2014). For example, global wildfire probability shows a hump-shaped relationship with temperature and moisture (Moritz et al. 2012). Interactions among drivers also impose nonlinearity, e.g., in hot and dry climates fires are fuel limited (McLaughlin and Bowers 1982), but in cold and wet climates fires are energy limited (Krawchuk and Moritz 2011).

Prediction is also complicated by uncertainty in which distribution(s) to use to assign probabilities to extreme events. The generalized Pareto distribution (GPD) has frequently been used (Bermudez et al. 2009; Jiang and Zhuang 2011), but the GPD requires a threshold to delineate extreme events (Davison and Smith 1990, Coles (2014)). The utility and validity of a threshold for extremes in a heterogeneous region is debatable (Tedim et al. 2018). Recently proposed metastatistical extreme value (MEV) approaches do not require such a threshold, and are based on the statistical distribution of finite sample maxima, i.e., the probability distribution of the maximum value for a finite number of events (Marani and Ignaccolo 2015; Zorzetto, Botter, and Marani 2016). In the MEV framework, the occurrence and size of future events, and the parameters of their distributions are treated as random variables which together imply a distribution for extremes. This approach has roots in compound distributions (Dubey 1970; Wiitala 1999), doubly stochastic processes (Cox and Isham 1980), superstatistics (Beck and Cohen 2003), and the Bayesian posterior predictive distribution (Gelman et al. 2013). The link to Bayesian inference is particularly useful, as it provides an easy way to propagate uncertainty forward to to predictions of extremes (Coles, Pericchi, and Sisson 2003).

Here, we extend the finite sample maximum approach to account for non-linear, spatially varying covariate effects with the goal of predicting extreme wildfire events from a statistical perspective across the contiguous United States. Specifically, we aim to predict occurrence (where and when), and magnitude (burned area) of large wildfires at a monthly time scale and regional spatial scale across the contiguous United States.

Methods

Data description

We acquired wildfire event data for the contiguous United States from the Monitoring Trends in Burn Severity (MTBS, www.mtbs.gov) program (Eidenshink et al. 2007), which includes spatiotemporal information on the occurrence of wildfires in the United States from 1984 to 2016. The MTBS data contain fires greater than 1000 acres ($\approx$ 405 hectares) in the western U.S. and greater than 500 acres ($\approx$ 202 hectares) in the eastern U.S. For consistency across the U.S., we discarded all records in the MTBS data less than 1000 acres, retaining 10,736 fire events (Figure 2A). Each event in the MTBS data has a discovery date, spatial point location, and final size.

To explain fire size and occurrence, we used a combination of meteorological variables including humidity, air temperature, precipitation, and wind speed. These variables were selected on the basis of previous work, and also with an aim to drive a predictive model with interpretable meteorological quantities. Meteorological layers were acquired from the gridMET data (Abatzoglou 2013) that blends monthly high-spatial resolution (~4-km) climate data from the Parameter-elevation Relationships on Independent Slopes Model (Daly et al. 2008) with high-temporal resolution (hourly) data from the National Land Data Assimilation System (NLDAS2) using climatologically aided interpolation. The resultant products are a suite of surface meteorological variables summarized at the daily time step and at a 4-km pixel resolution. Daily total precipitation, minimum relative humidity, mean wind speed, and maximum air temperature were averaged at a monthly time step for each of 84 Environmental Protection Agency level 3 (L3) ecoregions for each month from 1984 to 2016 (Omernik 1987; Omernik and Griffith 2014). We also computed cumulative monthly precipitation over the previous 12 months for each ecoregion-month combination. We chose to segment the U.S. with level 3 ecoregions as a compromise between the more numerous (computationally demanding) level 4 ecoregions, and the coarser level 2 ecoregions.

We used publicly available housing density estimates that were generated based on the U.S. 2000 decennial census as explanatory variables that may relate to human ignition pressure (Radeloff et al. 2010). These are provided at decadal time steps, and spatially at the level of census partial block groups. To generate approximate measures of housing density at monthly time intervals, we used a simple linear interpolation over time for each block group, then aggregated spatially across block groups to compute mean housing density for each ecoregion in each month.

Model development

We built two types of models: one describing the occurrence of fires within each L3 ecoregion over time (i.e., the total number of fires occurring in each ecoregion for each month from 1984 - 2016), and another describing the size of each wildfire in each ecoregion and month. For occurrence models, the response variable was a count (number of fires), and for burned area models, the response was a continuous positive quantity (size of each fire event). We used the period from 1984 to 2009 for training, witholding the period from 2010 to 2016 to evaluate predictive performance.

Fire occurrence

We constructed four models for fire occurrence and compared their predictive performance based on test-set log likelihood and posterior predictive checks for the proportion of zeros, maximum count, and total count. The models differed in the distributions used in the likelihood, representing counts as a Poisson, negative binomial, zero-inflated Poisson, or zero-inflated negative binomial random variable. The Poisson distribution is a common choice for counts, and the negative binomial distribution provides an alternative that can account for overdispersion. The zero-inflated versions of these distributions include a component to represent extra zeros, which might be expected to work well if there are independent processes that determine whether nonzero counts are possible (Lambert 1992).

For spatial units (ecoregions) $s = 1, ..., S$ and time steps (months) $t = 1, ..., T$, each model defines a probability mass function for $n_{s,t}$: the number of fires over 405 hectares in ecoregion s and time step t. For each of the four count distributions under consideration, location parameters $\mu_{s,t}$ and (for zero-inflated models) structural zero inflation parameters $\pi_{s,t}$ were allowed to vary in space and time. We used a log link function to ensure that $\mu_{s,t} > 0$, and a logit link function to ensure that $\pi_{s,t} \in (0, 1)$. Concatenating over spatial and temporal units, so that $\boldsymbol{\mu} = (\mu_{s=1,t=1}, \mu_{s=2,t=1}, ..., \mu_{s=S,t=1}, \mu_{s=S,t=2}, ..., \mu_{s=S,t=T})$, and similarly for $\boldsymbol{\pi}$, we modeled location and (when applicable) zero inflation parameters as:

$$\log(\boldsymbol{\mu}) = \alpha^{(\mu)} + \mathbf{X}\boldsymbol{\beta}^{(\mu)} + \boldsymbol{\phi}^{(\mu)} + \log(\boldsymbol{a}),$$

$$\text{logit}(\boldsymbol{\pi}) = \alpha^{(\pi)} + \mathbf{X}\boldsymbol{\beta}^{(\pi)} + \boldsymbol{\phi}^{(\pi)},$$

where $\alpha^{(\mu)}$ and $\alpha^{(\pi)}$ are scalar intercept parameters, $\mathbf{X}$ is a known $(S \times T) \times p$ design matrix, where p is the number of input features, $\boldsymbol{\beta}^{(\mu)}$ and $\boldsymbol{\beta}^{(\pi)}$ are column vector parameters of length p, $\boldsymbol{\phi}^{(\mu)}$ and $\boldsymbol{\phi}^{(\pi)}$ are column vector parameters of length $S \times T$ containing spatiotemporal adjustments, and $\boldsymbol{a}$ is a known offset vector of areas for spatial unit $s = 1, 2, ..., S$, repeated T times.

Burned area

We developed five candidate models for fire size, each of which specified a different distribution for the size (burned area) of individual fire events (Reed and McKelvey 2002; Hernandez et al. 2015), including the generalized Pareto (Hosking and Wallis 1987), tapered Pareto (Schoenberg, Peng, and Woods 2003), lognormal, gamma, and Weibull distributions. We evaluated each model in terms of test set log likelihood, and posterior predictive checks for fire size extremes. We defined the response y_i as the number of hectares burned over 405 for the i^{th} fire event, which occurred in spatial unit s_i and time step t_i.

Because each burned area distribution has a different parameterization, we included covariate effects in a distribution-specific way. For the generalized Pareto distribution

(GPD), we assumed a positive shape parameter, leading to a Lomax distribution for exceedances (Bermudez et al. 2009). The GPD and Lomax shape parameters are related by $\kappa^{(GPD)} = 1/\kappa^{(L)}$, and the GPD scale parameter is related to the Lomax scale and shape parameters by $\sigma^{(GPD)} = \sigma^{(L)}/\kappa^{(L)}$. We introduced covariate dependence via the Lomax scale parameter using a log link. For event i, $\log(\sigma_i^{(L)}) = \alpha + \boldsymbol{X}_{(s_i,t_i)}\boldsymbol{\beta} + \phi_{s_i,t_i}$, where α is an intercept parameter, $\boldsymbol{\beta}$ is a length p vector of coefficients, $\boldsymbol{X}_{(s_i,t_i)}$ is a row vector from $\mathbf{X}$, and ϕ_{s_i,t_i} is a spatiotemporal adjustment for s_i and t_i. For the tapered Pareto model, we modeled the shape parameter as $\log(\kappa_i) = \alpha + \boldsymbol{X}_{(s_i,t_i)}\boldsymbol{\beta} + \phi_{s_i,t_i}$. The lognormal model included covariate dependence via the location parameter: $\mu_i = \alpha + \boldsymbol{X}_{(s_i,t_i)}\boldsymbol{\beta} + \phi_{s_i,t_i}$. The gamma model used a log link for the expected value: $\log(E(y_i)) = \alpha + \boldsymbol{X}_{(s_i,t_i)}\boldsymbol{\beta} + \phi_{s_i,t_i}$. Last, we modeled the Weibull scale parameter as $\log(\sigma_i) = \alpha + \boldsymbol{X}_{(s_i,t_i)}\boldsymbol{\beta} + \phi_{s_i,t_i}$. More detail on the parameterization of each burned area distribution is provided in the Appendices.

Accounting for nonlinear forcing

The design matrix $\mathbf{X}$ was constructed to allow for spatially varying nonlinear effects of housing density and meteorological drivers. We used B-splines to account for nonlinearity (Figure 3) and allowed the coefficients for each basis vector to vary spatially (Wood 2017). First, we constructed univariate B-splines for log housing density, wind speed, same month precipitation, previous 12 month precipitation, air temperature, and humidity, with five degrees of freedom (including an intercept) for each variable. This step generated 30 basis vectors (five for each of six variables).

To allow for spatial variation in these nonlinear effects, we added interaction effects between each of the basis vectors and ecoregions (Brezger and Lang 2006; Kneib, Hothorn, and Tutz 2009). The hierarchical nesting of ecoregion designations (Figure 2B-D) lends itself to such interactions. Conceptually, coefficients in a level 3 ecoregion may be related to coefficients in the level 2 ecoregion containing the level 3 region, the level 1 ecoregion containing the level 2 region, and a global effect. The coefficient associated with a basis vector for any level 3 ecoregion is treated as a sum of a global effect, a level 1 ecoregion adjustment, a level 2 ecoregion adjustment, and a level 3 ecogregion adjustment. Thus, for every univariate basis vector, we included interaction effects with ecoregion at each of the three ecoregion levels. This allows borrowing of information across space (level 3 ecoregions in a level 2 ecoregion are often adjacent), and for regions that are ecologically similar. We also included adjustments on the global intercept for each level 1, 2, and 3 ecoregion to account for spatial variation that is unrelated to climate or housing density. This specification induces sparsity in $\mathbf{X}$ that we exploit to increase the efficiency of computing $\boldsymbol{\mu}$ and $\boldsymbol{\pi}$. In total, $\mathbf{X}$ has $p = 3{,}472$ columns, with 97% zero entries.

Prior specification

To avoid overfitting, we used a regularized horseshoe prior on the coefficients associated with the spatially varying nonlinear effects described above (Piironen, Vehtari, and others 2017). This prior places high probability close to zero, while retaining heavy enough tails that nonzero coefficients are not shrunk too strongly toward zero. This is consistent with our prior expectation that most of the coefficients associated with the columns in $\mathbf{X}$ were close to zero. For the zero inflated count models, we used a multivariate horseshoe to allow information sharing between the zero inflated and distribution specific location parameters (Peltola et al. 2014). For the remaining count models and all burned area models, this was a univariate horseshoe prior. Spatiotemporal random effects were constructed using a temporally autoregressive, spatially intrinsically autoregressive formulation (Besag and Kooperberg 1995; Banerjee, Carlin, and Gelfand 2014). Details of these priors and the resulting joint distributions are provided in the Appendices.

Posterior predictive inference for finite sample maxima

We used the posterior predictive distribution to check each model and make inference on extremes. The posterior predictive distribution provides a distribution for replications of observed data (y^{rep}), and predictions of future data (Gelman et al. 2013). Conceptually, for a good model, y^{rep} should be similar to observed training data y, and future predictions should be similar to future data. Distributions over both quantities can be obtained by conditioning on y and marginalizing over model parameters θ, e.g., $[y^{\text{rep}}|y] = \int [y^{\text{rep}}|\theta][\theta|y]d\theta$.

Posterior predictive distributions facilitate model checks that compare predicted and observed test statistics (Gelman, Meng, and Stern 1996). To evaluate whether models captured tail behavior, we compared empirical maxima ($T(y) = \max(y)$) to the predicted distribution of maxima $T(y^{\text{rep}})$. We also include predictive checks for the proportion of zero counts, and totals for count and burned area models. Posterior predictive inference for finite sample maxima is similar in spirit to the MEV approach. Both obtain a distribution over maxima by marginalizing over unknowns including the number of events, size of each event, and parameters of their distributions (Marani and Ignaccolo 2015). However, a Bayesian approach explicitly conditions on the observed data to obtain a posterior distribution of parameters.

Seeing this connection is useful in the context of including priors and propagating uncertainty to derived parameters. For any ecoregion s and timestep t, if we define a particular maximum fire size conditional on a fire having occurred as $z_{s,t}$, and let $Z_{s,t}$ represent the random variable of maximum fire size, then the cumulative distribution function (CDF) for $z_{s,t}$ is given by $\Pr(Z_{s,t} \leq z_{s,t}) = \text{F}(y_{(s,t)})^{n_{s,t}}$, where $\text{F}(y_{(s,t)})$ is the CDF of fire size, and $n_{s,t}$ is the number of wildfire events. Thus, $\Pr(Z_{s,t} \leq z_{s,t})$ is the distribution function for the finite sample maximum. The CDF for $z_{s,t}$ can be inverted to produce a quantile function that permits computation of prediction intervals for maximum fire sizes, conditional on fires having occurred. Given a collection of posterior draws from a burned area model that parameterize $\text{F}(y_{(s,t)})$, and a collection of posterior draws of $n_{s,t}^{\text{rep}}$ from the posterior predictive distribution of a wildfire count model, a posterior distribution for the CDF or quantile function of maximum fire size can be generated which combines the two models to facilitate inference on the distribution of extremes.

Parameter estimation

We used a combination of variational approximations and Hamiltonian Monte Carlo methods to sample from the posterior distributions of count and burned area models. A variational approximation (Kucukelbir et al. 2015) was used for count models to quickly identify a preferred model. The best performing count model and all burned area models were fit using the No-U-Turn Sampler (Hoffman and Gelman 2014). Models were fit in the Stan probabilistic programming language using the `rstan` package (Carpenter et al. 2016; Stan Development Team 2018). We ran four chains for 1000 iterations each, discarding the first 500 iterations as warmup. Convergence was assessed using visual inspection of trace plots, with potential scale reduction statistic values $\hat{R} \geq 1.1$ as an indicator convergence failure (Brooks and Gelman 1998).

Implementation

All data processing, model fitting, and visualization were implemented with open source software, primarily in the R programming language (R Core Team 2017), and wrapped in a reproducible workflow via GNU Make and Docker (Stallman, McGrath, and Smith 2004; Boettiger 2015). Data cleaning and transformation required the R packages assertthat (Wickham 2017a), lubridate (Grolemund and Wickham 2011), Matrix (Bates and Maechler 2018), pbapply (Solymos and Zawadzki 2018), splines (R Core Team 2018), tidyverse (Wickham 2017b), and zoo (Zeileis and Grothendieck 2005). Spatial data were processed with raster (Hijmans 2017), rgdal (Bivand, Keitt, and Rowlingson 2018), sf (Pebesma 2018), and spdep (Bivand and Piras 2015). Finally, we used cowplot (Wilke 2017), ggrepel (Slowikowski 2018), ggthemes (Arnold 2018), patchwork (Pedersen 2017), and RColorBrewer (Neuwirth 2014) for visualization. The manuscript was written in R Markdown (Allaire et al. 2018). Analyses were run on an Amazon Web Services m5.2xlarge EC2 instance with four physical cores and 32 GB of RAM, and the whole workflow requires $\approx$ 72 hours. All code to reproduce the analysis is available on GitHub at https://github.com/mbjoseph/wildfire-extremes (Joseph 2018).

Results

Wildfire occurrence

The zero-inflated negative binomial distribution performed best on the held-out test set (Table 1), and was able to recover the proportion of zeros, count maxima, and count totals in posterior predictive checks for both the training and test data (Figure 4). All of the other count models that we considered exhibited lack of fit to at least one of these statistics in posterior predictive checks. Hereafter, we report results from the zero-inflated negative binomial model.

Minimum relative humidity and maximum air temperature had the strongest effects on both the zero-inflation component and the expected value of the negative binomial component (Figure 5, posterior median for ρ: 0.665, 95% credible interval (CI): 0.319 - 0.861). The model uncovered unique effects of meteorological variables at level 1, 2, and 3 ecoregions (Figure 6). For example, a positive interaction effect between the second air temperature basis vector and the L1 Great Plains ecoregions indicates that the expected number of wildfires in plains ecoregions with cold conditions is high relative to other ecoregions. The Ozark/Ouachita-Appalachian forest and Ozark Highlands were also identified as having region-specific temperature effects (Figure 6). Twelve month total precipitation also had region specific effects in the Mississippi Alluvial and Southeast Coastal Plains ecoregion, where it was associated with lower expected fire counts (Figure 6). In contrast, increasing cumulative twelve month precipitation was associated with higher counts in desert ecoregions (Figure 5). Housing density showed a unimodal relationship to expected count (Figure 5), with lower expected counts in sparsely populated ecoregions, and higher expected counts with moderately populated ecoregions.

Posterior 95% credible interval coverage for the number of fires over 405 hectares in the test set was 98.8%. The lowest test set interval coverage was 89.3%, in the Cross Timbers L3 ecoregion. When observed counts fell outside the 95% prediction interval, counts were larger than predicted 100% of the time. The largest difference between observed numbers and predicted 97.5% posterior quantiles (the upper limit for the 95% credible interval) occurred for the Columbia Mountains/Northern Rockies L3 ecoregion in August 2015, when 36 fires over 405 hectares occurred and at most 22 were predicted. For nearly half of the level 3 ecoregions (43 of 85), accounting for 39.7% of the land area of the contiguous U.S., the zero-inflated negative binomial model had 100% test set prediction interval coverage.

Wildfire burned areas

The lognormal distribution performed best on the test set (Table 2), and captured tail-behavior better than other burned area distributions (Figure 7). The GPD model was too heavy-tailed to adequately capture the pattern in the empirical data, predicting fires far larger than those observed in the training and test sets (Figure 7). The tapered Pareto distribution was too light-tailed (Figure 7). The gamma and Weibull models performed very poorly overall on the test set (Table 2), apparently due to a lack of congruence between the shapes of these distributions and the actual burned area distribution. Despite a poor fit to the bulk of the wildfire burned area distribution, both performed adequately in the upper tails (Figure 7). Hereafter we present results for the lognormal model, which had the highest test set log likelihood and captured tail behavior of the empirical fire size distribution.

Relative humidity was the primary driver of expected burned area for a fire event (Figure 8A). The first basis vector for mean daily minimum relative humidity was the only coefficient with a 95% credible interval that did not include zero (posterior median: 1.68, 95% CI: (0.8 - 2.29)). This nonlinear effect can be observed in Figure 8B as an increase in the expected burned area below 20% mean daily minimum humidity. This leads to a seasonality gradient among ecoregions of expected fire sizes, with little or no seasonal signal in typically humid ecoregions such as Marine West Coast Forests of the Pacific Northwest, and seasonal oscillations in ecoregions that have periodic fluctuations between dry and humid conditions such as the Temperate Sierras (Figure 8C). There was not strong evidence that meteorological variables had spatially variable effects on expected wildfire burned area.

Overall, 95% posterior predictive interval coverage in the test set for burned areas was 93%. The lowest test set coverage was 0%, for the Eastern Great Lakes Lowlands L3 ecoregion, followed by 50%, for the Central California Valley L3 ecoregion, though these ecoregions had just 1 and 2 wildfire events in the test set. When observed fire sizes fell outside the 95% prediction interval, 24.9% of wildfires were smaller than predicted, and 75.1% of wildfires were larger than predicted. The largest discrepancy between the actual size of a wildfire and the predicted 97.5% posterior quantile was observed with the Wallow Fire in 2011 which burned 228,107 hectares, but the predicted upper limit for size was 20,756. We investigate this discrepancy further in the case study below. The lognormal burned area model achieved 100% interval coverage in 24 of 67 ecoregions that had wildfire events in the test set, accounting for 26% of the land area of the contiguous U.S.

Inference on extremes

By combining the output of the event count and burned area models, we derived posterior prediction intervals for the size of the largest fire in a month for each region (the "burned area maximum"), integrating over uncertainty in the number of fires, as well as the lognormal mean and standard deviation for burned area. We evaluated the posterior distribution for the quantile function of the finite sample maximum of a lognormal distribution ($\exp(\mu + \sigma\sqrt{2}\text{erf}^{-1}(2P^{1/n} - 1))$, where n is the number of wildfire events, erf^{-1} is the inverse error function, and P is a probability) to generate prediction intervals for maximum fire sizes by month and ecoregion, conditional on one or more fires having occurred. In the holdout period from 2010 to 2016, a 99% prediction interval achieved 77.4% interval coverage, with 14.8% of the burned area maxima (140 fire events) being larger than predicted (Figure 9). As an additional check, we used the posterior distribution of the model to predict the total area burned by wildfires in the test set. The model predicted the total area burned over the entire contiguous United States in test period from 2010 to 2016 to be 30,339,123 (95% CI: (20,496,551 - 50,446,932) and the actual value was 30,440,173.

While fires over a million acres ($\approx$ 404,686 hectares) in size have happened historically in the contiguous U.S. (Pernin 1971), no such fires were represented in in the training or test sets. If we extrapolate, the probability of at least one fire this large in the period from 2010 to 2016 was estimated to be between 0.191 and 0.651 (95% CI), with a posterior median of 0.348. The highest probability for such an event was 0.014 (posterior median), with a 95% CI of (0, 0.237) seen for the Southwestern Tablelands ecoregion in June 2011. The second highest probability was 0.004 (posterior median), with a 95% CI of (0, 0.056) seen for the Arizona/New Mexico Mountains ecoregion in June 2011. Aggregating spatially, we estimated monthly probabilities of a million acre wildfire. These probabilities show seasonal signals corresponding to peak fire seasons, with a shift toward higher and broader peaks beginning in the 21st century (Figure 10).

Error analysis case study: the 2011 Wallow Fire

To better understand how well the model could or could not anticipate notable extreme events, and why, we used the largest fire in the test set as a case study. The Wallow Fire was accidentally ignited on May 29, 2011 by two campers in the L3 Arizona/New Mexico Mountains ecoregion. It burned through the month of June and into early July. The model underpredicted the total burned area of the Wallow Fire. Integrating over uncertainty in the predicted number of fires and expected fire size, the 99% credible interval for the maximum fire size for May 2011 was (730 - 107,419) hectares, but the Wallow Fire is recorded as 228,103 hectares.

We evaluated the contribution of each covariate to the linear predictor functions of the three model components (lognormal mean for burned areas, negative binomial mean for counts, and the logit probability of the zero-inflation component) to understand why these predictions differed. We defined the contribution of a variable as the dot product of the elements in the design matrix $\mathbf{X}$ corresponding to a particular driver variable (e.g., humidity), and the estimated coefficients in $\boldsymbol{\beta}$ corresponding to that variable. This provides a quantitative measure of how each input variable contributes to the linear predictor for an ecoregion, and incorporates the overall, level 1, level 2, and level 3 ecoregion adjustments on these effects. Humidity is the primary driver of variation in the model's predictions overall, and June 2011 - the month after ignition - favored more large fires, with drier, hotter conditions (Figure 11). The 99% credible interval for June 2011 was (4,258 - 428,765) hectares, which contains the true value. Had the Wallow Fire ignited two days later, the true final size would have been contained in the prediction interval. Evidently, conditions in May that drove (under)predictions of maximum burned area were not representative of the conditions over most of the Wallow Fire's duration.

Temporal mismatch aside, meteorological conditions local to the Wallow Fire differed from the monthly regional means (Figure 12). In particular, wind speeds in the Wallow Fire vicinity exceeded the regional monthly mean values on the date of ignition and in the weeks following ignition. Over the majority of the duration of the Wallow Fire (May 29 to July 8), local daily conditions were drier and hotter on average than regional mean monthly conditions in May, which were used to drive the statistical model. This local variability is not represented in the regional models developed here. The failure of the model to correctly predict the size of the Wallow fire suggests potential avenues for improvement, discussed below.

Discussion

Extreme wildfires are often devastating, but perhaps they need not be surprising. By allowing the non-linear effects of weather and housing density to vary across space, this model achieves good predictive accuracy for fire extremes at a regional scale over a six year prediction window. This model predicts that extremely large wildfires, perhaps even over one million acres (404,686 hectares), have a non-negligible probability of occurrence in the contiguous United States. Such predictions can support regional wildfire management and probabilistic hazard assessment.

Driving a model with meteorological features raises challenges related to predictive uncertainty and covariate shift - a change in the underlying distribution of forcing variables, potentially outside of the historic range. Ideally, this uncertainty would be propagated forward in a predictive model, possibly through stacking of predictive distributions that are generated from multiple models of future climate dynamics (Yao et al. 2017). But, even if one had a perfect forecast, novel conditions present a challenge for predictive modeling (Quionero-Candela et al. 2009). For example, the High Plains ecoregion had its highest mean monthly precipitation, lowest 12 month running precipitation, driest, hottest, and windiest conditions in the test set period. Extrapolating beyond the range of training inputs is generally difficult, but the hierarchical spatial effect specification used here allows partial pooling among climatically similar ecoregions that can inform such predictions, unlike models fit separately to disjoint spatial regions.

Similar issues could arise when making predictions for observed but rare meteorological conditions. For example, mean daily minimum humidity values over 60% accounted for just 3.76% of the ecoregion-months in the training data, and 0 fires occurred in such months. As a consequence, there is relatively little data that can be used to inform the model for such conditions, and the prior distribution which shrinks coefficients toward zero may dominate the likelihood in the posterior distribution. In this case, the posterior distribution for the last basis coefficient for the partial effect of humidity is likely to be close to zero. This could explain why the estimated partial effect of humidity on the expected counts was less negative at the upper end of the observed humidity range, although previous work has found similarly nonlinear partial effects (Preisler et al. 2004). The count model performed extremely well in this range, with 100% interval coverage for the 299 ecoregion-months with mean daily minimum humidity values greater than 60% in the withheld test data. The model nearly always predicted zero counts with high confidence when conditions were this humid: 298 of 299 predictions made for such conditions were 95% credible intervals of (0, 0). The remaining prediction had a posterior median of zero, along with a 95% credible interval from 0 to 1. Monotonicity constraints could be incorporated into these models via monotonic spline bases (Ramsay and others 1988), or an ordered prior distribution for basis coefficients (Brezger and Steiner 2008). In this case, the count model performs well under humid conditions without monotonicity constraints, and there seems to be little room for performance improvements that might result from such constraints.

Human-caused climate change is expected to increase fire activity in the western U.S.

(Rogers et al. 2011; Westerling et al. 2011; Moritz et al. 2012; Abatzoglou and Williams 2016) and elsewhere (Flannigan et al. 2009), but the nonlinear effect of housing density could provide additional insight into future expectations. While housing density is increasing over time in most U.S. ecoregions, some of these ecoregions are in the range of values in which this increases the expected number of large fires, while others are so populated that further increases would reduce the chance of a large fire. The hump-shaped effect of human density on the expected number of large fires is likely driven by ignition pressure and fire suppression (Balch et al. 2017). As human density increases from zero, ignition pressure increases, but eventually landscapes become so urbanized, fragmented, and/or fire-suppressed that wildfire risk decreases (Syphard et al. 2007; Bowman et al. 2011; Bistinas et al. 2013; Knorr et al. 2013; Mcwethy et al. 2013; Syphard et al. 2017; Nagy et al. 2018). At intermediate density, wildfire regimes respond to human ignition and altered fuel distributions (Guyette, Muzika, and Dey 2002), but these responses depend on environmental context and characteristics of the human population (Marlon et al. 2008; Li et al. 2009). This model indicates that the combination of moderate to high human density and dry conditions would nonlinearly increase the chance of an extreme fire event. Both human density and dryness are expected to increase in the future across large swaths of the U.S. (Lloyd, Sorichetta, and Tatem 2017; Stavros et al. 2014, Radeloff et al. (2010)), with potential implications for human mortality, health risks from smoke and particulate emission, and the financial burden of wildfire management (Reid et al. 2016; Radeloff et al. 2018).

This work points to promising directions for future predictive efforts. Default choices such as Poisson and GPD distributions should be checked against alternative distributions. Further, the predictive skill of this model seems to suggest that ordinary events provide information on extremes, which would not be the case if the generative distribution of extremes was completely unique. Previous case studies have identified that extremes or anomalies in climatological drivers play a role in the evolution of extreme wildfires (Peterson et al. 2015), but for this work, monthly averages of climatological drivers over fairly large spatial regions were used, which may smooth over anomalous or extreme conditions. Enhancing the spatiotemporal resolution of predictive models could better represent climatic and social drivers and provide localized insights to inform decision-making. This raises computational challenges, but recent advances in distributed probabilistic computing (Tran et al. 2017), efficient construction of spatiotemporal point processes (Shirota and Banerjee 2018), and compact representations of nonlinear spatial interactions (Lee and Durbán 2011) may provide solutions.

The Wallow Fire case study reveals at least one limitation of increasing the spatiotemporal resolution. When the model predictions are driven by covariates that are summarized in space and time (e.g. a mean across an ecoregion in a month), summary values may not represent conditions that are most relevant to an event. With a discrete space-time segmentation, events can occur at the boundary of a spatiotemporal unit, e.g., if a fire spreads into an adjacent ecoregion or ignites on the last day of the month. Large wildfires can span months, and a model that only uses conditions upon ignition to predict total burned area can fail to account for conditions that change over the course of the event. Modeling ignitions as a point process in continuous space and time (Brillinger, Preisler, and Benoit 2003), and explicitly modeling subsequent fire duration and spread could better separate conditions that ignite fires from those that affect propagation. Such an approach might be amenable to including information on fuel continuity, which is likely to limit the size of extremely large fires and did not factor into the current predictions (Rollins, Morgan, and Swetnam 2002; Hargrove et al. 2000).

To the extent that a model reflects the generative process for extreme events, the decomposition of contributions to the model's predictions may provide insight into attribution for meteorological and anthropogenic drivers of extremes. However, a model trained to represent a region-wide distribution of fire sizes will inevitably fail to capture local factors that are relevant to specific events such as the Wallow Fire. If predicting the dynamics of particular fire events is a goal, process-based models designed to model fire spread are likely to be more appropriate than regional statistical models such as those developed here.

This paper presents and evaluates a statistical approach to explain and predict extreme wildfires that incorporates spatially varying non-linear effects. The model reveals consider-able differences among ecoregions spanning the mountain west to the great plains, deserts, and eastern forests, and suggests a non-negligible chance of extreme wildfires larger than those seen in over the past 30 years in the contiguous U.S. Predictive approaches such as this can inform decision-making by placing probabilistic bounds on the number of wildfires and their sizes, while provide deeper insights into wildfire ecology.

Acknowledgments

We thank Mitzi Morris, Kyle Foreman, Daniel Simpson, Bob Carpenter, and Andrew Gelman for contributing to the implementation of an intrinsic autoregressive spatial prior in Stan. We also thank two anonymous reviewers for their thoughtful comments, which greatly improved the paper.

Literature cited

Abatzoglou, John T. 2013. "Development of Gridded Surface Meteorological Data for Ecological Applications and Modelling." *International Journal of Climatology* 33 (1). Wiley Online Library: 121–31.

Abatzoglou, John T., and A. Park Williams. 2016. "Impact of Anthropogenic Climate Change on Wildfire Across Western Us Forests." *Proceedings of the National Academy of Sciences* 113 (42). National Academy of Sciences: 11770–5. doi:10.1073/pnas.1607171113.

Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, and Winston Chang. 2018. *Rmarkdown: Dynamic Documents for R.* https://CRAN.R-project.org/package=rmarkdown.

Arnold, Jeffrey B. 2018. *Ggthemes: Extra Themes, Scales and Geoms for 'Ggplot2'.* https://CRAN.R-project.org/package=ggthemes.

Balch, Jennifer K, Bethany A Bradley, John T Abatzoglou, R Chelsea Nagy, Emily J

Fusco, and Adam L Mahood. 2017. "Human-Started Wildfires Expand the Fire Niche Across the United States." *Proceedings of the National Academy of Sciences* 114 (11). National Acad Sciences: 2946–51.

Balshi, Michael S, A DAVID McGUIRE, Paul Duffy, Mike Flannigan, John Walsh, and Jerry Melillo. 2009. "Assessing the Response of Area Burned to Changing Climate in Western Boreal North America Using a Multivariate Adaptive Regression Splines (Mars) Approach." *Global Change Biology* 15 (3). Wiley Online Library: 578–600.

Banerjee, Sudipto, Bradley P Carlin, and Alan E Gelfand. 2014. *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press.

Barbero, R, JT Abatzoglou, EA Steel, and Narasimhan K Larkin. 2014. "Modeling Very Large-Fire Occurrences over the Continental United States from Weather and Climate Forcing." *Environmental Research Letters* 9 (12). IOP Publishing: 124009.

Bates, Douglas, and Martin Maechler. 2018. *Matrix: Sparse and Dense Matrix Classes and Methods*. https://CRAN.R-project.org/package=Matrix.

Beck, Christian, and EGD Cohen. 2003. "Superstatistics." *Physica A: Statistical Mechanics and Its Applications* 322. Elsevier: 267–75.

Bermudez, P de Zea, J Mendes, JMC Pereira, KF Turkman, and MJP Vasconcelos. 2009. "Spatial and Temporal Extremes of Wildfire Sizes in Portugal (1984–2004)." *International Journal of Wildland Fire* 18 (8). CSIRO Publishing: 983–91.

Besag, Julian, and Charles Kooperberg. 1995. "On Conditional and Intrinsic Autoregressions." *Biometrika* 82 (4). Oxford University Press: 733–46.

Bistinas, Ioannis, Duarte Oom, Ana C. L. Sá, Sandy P. Harrison, I. Colin Prentice, and José M. C. Pereira. 2013. "Relationships between Human Population Density and Burned Area at Continental and Global Scales." *PLoS ONE*. doi:10.1371/journal.pone.0081188.

Bivand, Roger, and Gianfranco Piras. 2015. "Comparing Implementations of Estimation Methods for Spatial Econometrics." In. American Statistical Association.

Bivand, Roger, Tim Keitt, and Barry Rowlingson. 2018. *Rgdal: Bindings for the 'Geospa-*

tial' Data Abstraction Library. https://CRAN.R-project.org/package=rgdal.

Boettiger, Carl. 2015. "An Introduction to Docker for Reproducible Research." *ACM SIGOPS Operating Systems Review* 49 (1). ACM: 71–79.

Bowman, David MJS, Jennifer Balch, Paulo Artaxo, William J Bond, Mark A Cochrane, Carla M D'antonio, Ruth DeFries, et al. 2011. "The Human Dimension of Fire Regimes on Earth." *Journal of Biogeography* 38 (12). Wiley Online Library: 2223–36.

Brezger, Andreas, and Stefan Lang. 2006. "Generalized Structured Additive Regression Based on Bayesian P-Splines." *Computational Statistics & Data Analysis* 50 (4). Elsevier: 967–91.

Brezger, Andreas, and Winfried J Steiner. 2008. "Monotonic Regression Based on Bayesian P–Splines: An Application to Estimating Price Response Functions from Store-Level Scanner Data." *Journal of Business & Economic Statistics* 26 (1). Taylor & Francis: 90–104.

Brillinger, David R, Haiganoush K Preisler, and John W Benoit. 2003. "Risk Assessment: A Forest Fire Example." *Lecture Notes-Monograph Series*. JSTOR, 177–96.

Brooks, Stephen P, and Andrew Gelman. 1998. "General Methods for Monitoring Convergence of Iterative Simulations." *Journal of Computational and Graphical Statistics* 7 (4). Taylor & Francis: 434–55.

Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2016. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 20: 1–37.

Coles, S. 2014. *An Introduction to Statistical Modeling of Extreme Values*. Springer. https://books.google.com/books?id=G-D-sgEACAAJ.

Coles, Stuart, Joanna Bawa, Lesley Trenner, and Pat Dorazio. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Vol. 208. Springer.

Coles, Stuart, Luis Raúl Pericchi, and Scott Sisson. 2003. "A Fully Probabilistic Approach to Extreme Rainfall Modeling." *Journal of Hydrology* 273 (1-4). Elsevier: 35–50.

Cox, David Roxbee, and Valerie Isham. 1980. *Point Processes*. Vol. 12. CRC Press.

Daly, Christopher, Michael Halbleib, Joseph I Smith, Wayne P Gibson, Matthew K Doggett, George H Taylor, Jan Curtis, and Phillip P Pasteris. 2008. "Physiographically Sensitive Mapping of Climatological Temperature and Precipitation Across the Conterminous United States." *International Journal of Climatology* 28 (15). Wiley Online Library: 2031–64.

Davison, Anthony C, and Richard L Smith. 1990. "Models for Exceedances over High Thresholds." *Journal of the Royal Statistical Society. Series B (Methodological)*. JSTOR, 393–442.

Dennison, Philip E, Simon C Brewer, James D Arnold, and Max A Moritz. 2014. "Large Wildfire Trends in the Western United States, 1984–2011." *Geophysical Research Letters* 41 (8). Wiley Online Library: 2928–33.

Diaz, John M. 2012. "Economic Impacts of Wildfire." *Southern Fire Exchange*.

Díaz-Avalos, Carlos, Pablo Juan, and Laura Serra-Saurina. 2016. "Modeling Fire Size of Wildfires in Castellon (Spain), Using Spatiotemporal Marked Point Processes." *Forest Ecology and Management* 381. Elsevier: 360–69.

Dubey, Satya D. 1970. "Compound Gamma, Beta and F Distributions." *Metrika* 16 (1). Springer: 27–31.

Eidenshink, J, B Schwind, K Brewer, ZL Zhu, B Quayle, and S Howard. 2007. "A Project for Monitoring Trends in Burn Severity." *Nutrition and Cancer* 58 (1): 28–34.

Flannigan, Mike D., Meg A. Krawchuk, William J. de Groot, B. Mike Wotton, and Lynn M. Gowman. 2009. "Implications of Changing Climate for Global Wildland Fire." Journal Article. *International Journal of Wildland Fire* 18 (5): 483–507. doi:http://dx.doi.org/10.1071/WF08187.

Fosberg, Michael A. 1978. "Weather in Wildland Fire Management: The Fire Weather

Index." *US for Serv Reprints of Articles by FS Employees.*

Gelman, A., J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. 2013. *Bayesian Data Analysis, Third Edition.* Chapman & Hall/Crc Texts in Statistical Science. Taylor & Francis. https://books.google.com/books?id=ZXL6AQAAQBAJ.

Gelman, Andrew, Xiao-Li Meng, and Hal Stern. 1996. "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies." *Statistica Sinica.* JSTOR, 733–60.

Goodrick, Scott L. 2002. "Modification of the Fosberg Fire Weather Index to Include Drought." *International Journal of Wildland Fire* 11 (4). CSIRO: 205–11.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. http://www.jstatsoft.org/v40/i03/.

Guyette, R. P., R. M Muzika, and D. C. Dey. 2002. "Dynamics of an Anthropogenic Fire Regime." *Ecosystems* 5: 472–86. doi:10.1007/s10021-002-0115-7.

Hargrove, William W, RH Gardner, MG Turner, WH Romme, and DG Despain. 2000. "Simulating Fire Patterns in Heterogeneous Landscapes." *Ecological Modelling* 135 (2-3). Elsevier: 243–63.

Hernandez, Charles, C Keribin, P Drobinski, and S Turquety. 2015. "Statistical Modelling of Wildfire Size and Intensity: A Step Toward Meteorological Forecasting of Summer Extreme Fire Risk." In *Annales Geophysicae*, 33:1495–1506. 12.

Hijmans, Robert J. 2017. *Raster: Geographic Data Analysis and Modeling.* https://CRAN.R-project.org/package=raster.

Hoffman, Matthew D, and Andrew Gelman. 2014. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *Journal of Machine Learning Research* 15 (1): 1593–1623.

Hosking, Jonathan RM, and James R Wallis. 1987. "Parameter and Quantile Estimation for the Generalized Pareto Distribution." *Technometrics* 29 (3). Taylor & Francis: 339–49.

Jiang, Yueyang, and Qianlai Zhuang. 2011. "Extreme Value Analysis of Wildfires in Canadian Boreal Forest Ecosystems." *Canadian Journal of Forest Research* 41 (9). NRC Research Press: 1836–51.

Joseph, Max. 2018. "Mbjoseph/Wildfire-Extremes: First Release." doi:10.5281/zenodo.1326858.

Kneib, Thomas, Torsten Hothorn, and Gerhard Tutz. 2009. "Variable Selection and Model Choice in Geoadditive Regression Models." *Biometrics* 65 (2). Wiley Online Library: 626–34.

Knorr, W, T Kaminski, A Arneth, and U Weber. 2013. "Impact of human population density on fire frequency at the global scale Impact of human population density on fire frequency at the global scale Impact of human population density on fire frequency at the global scale." *Biogeosciences Discuss* 10: 15735–78. doi:10.5194/bgd-10-15735-2013.

Kochi, Ikuho, Geoffrey H Donovan, Patricia A Champ, and John B Loomis. 2010. "The Economic Cost of Adverse Health Effects from Wildfire-Smoke Exposure: A Review." *International Journal of Wildland Fire* 19 (7). CSIRO: 803–17.

Krawchuk, Meg A, Max A Moritz, Marc-André Parisien, Jeff Van Dorn, and Katharine Hayhoe. 2009. "Global Pyrogeography: The Current and Future Distribution of Wildfire." *PloS One* 4 (4). Public Library of Science: e5102.

Krawchuk, Meg A., and Max A. Moritz. 2011. "Constraints on global fire activity vary across a resource gradient." *Ecology* 92 (1): 121–32. doi:10.1890/09-1843.1.

Kucukelbir, Alp, Rajesh Ranganath, Andrew Gelman, and David Blei. 2015. "Automatic Variational Inference in Stan." In *Advances in Neural Information Processing Systems*, 568–76.

Lambert, Diane. 1992. "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing." *Technometrics* 34 (1). Taylor & Francis: 1–14.

Lee, Dae-Jin, and María Durbán. 2011. "P-Spline Anova-Type Interaction Models for Spatio-Temporal Smoothing." *Statistical Modelling* 11 (1). SAGE Publications Sage India:

New Delhi, India: 49–69.

Li, Li-Ming, Wei-Guo Song, Jian Ma, and Kohyu Satoh. 2009. "Artificial neural network approach for modeling the impact of population density and weather parameters on forest fire risk." *International Journal of Wildland Fire* 18 (6). CSIRO PUBLISHING: 640. doi:10.1071/WF07136.

Lloyd, Christopher T, Alessandro Sorichetta, and Andrew J Tatem. 2017. "High Resolution Global Gridded Data for Use in Population Studies." *Scientific Data* 4. Nature Publishing Group: 170001.

Marani, Marco, and Massimiliano Ignaccolo. 2015. "A Metastatistical Approach to Rainfall Extremes." *Advances in Water Resources* 79. Elsevier: 121–26.

Marlon, J. R., P. J. Bartlein, C. Carcaillet, D. G. Gavin, S. P. Harrison, P. E. Higuera, F. Joos, M. J. Power, and I. C. Prentice. 2008. "Climate and human influences on global biomass burning over the past two millennia." *Nature Geoscience*. doi:10.1038/ngeo313.

McLaughlin, Steven P, and Janice E Bowers. 1982. "Effects of Wildfire on a Sonoran Desert Plant Community." *Ecology*. JSTOR, 246–48.

Mcwethy, D. B., P. E. Higuera, C. Whitlock, T. T. Veblen, D. M J S Bowman, G. J. Cary, S. G. Haberle, et al. 2013. "A conceptual framework for predicting temperate ecosystem sensitivity to human impacts on fire regimes." *Global Ecology and Biogeography* 22 (8): 900–912.

Mendes, Jorge M, Patrícia Cortés de Zea Bermudez, José Pereira, KF Turkman, and MJP Vasconcelos. 2010. "Spatial Extremes of Wildfire Sizes: Bayesian Hierarchical Models for Extremes." *Environmental and Ecological Statistics* 17 (1). Springer: 1–28.

Moritz, Max A, Marc-André Parisien, Enric Batllori, Meg A Krawchuk, Jeff Van Dorn, David J Ganz, and Katharine Hayhoe. 2012. "Climate Change and Disruptions to Global Fire Activity." *Ecosphere* 3 (6). Wiley Online Library: 1–22.

Nagy, R, Emily Fusco, Bethany Bradley, John T Abatzoglou, and Jennifer Balch. 2018. "Human-Related Ignitions Increase the Number of Large Wildfires Across Us Ecoregions."

Fire 1 (1). Multidisciplinary Digital Publishing Institute: 4.

Nauslar, Nicholas, John Abatzoglou, and Patrick Marsh. 2018. "The 2017 North Bay and Southern California Fires: A Case Study." *Fire* 1 (1). Multidisciplinary Digital Publishing Institute: 18.

Neuwirth, Erich. 2014. *RColorBrewer: ColorBrewer Palettes*. https://CRAN.R-project.org/package=RColorBrewer.

Omernik, James M. 1987. "Ecoregions of the Conterminous United States." *Annals of the Association of American Geographers* 77 (1). Taylor & Francis: 118–25.

Omernik, James M, and Glenn E Griffith. 2014. "Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework." *Environmental Management* 54 (6). Springer: 1249–66.

Pebesma, Edzer. 2018. *Sf: Simple Features for R*. https://CRAN.R-project.org/package=sf.

Pechony, O, and DT Shindell. 2009. "Fire Parameterization on a Global Scale." *Journal of Geophysical Research: Atmospheres* 114 (D16). Wiley Online Library.

Pechony, Olga, and Drew T Shindell. 2010. "Driving Forces of Global Wildfires over the Past Millennium and the Forthcoming Century." *Proceedings of the National Academy of Sciences* 107 (45). National Acad Sciences: 19167–70.

Pedersen, Thomas Lin. 2017. *Patchwork: The Composer of Ggplots*. https://github.com/thomasp85/patchwork.

Peltola, Tomi, Aki S Havulinna, Veikko Salomaa, and Aki Vehtari. 2014. "Hierarchical Bayesian Survival Analysis and Projective Covariate Selection in Cardiovascular Event Risk Prediction." In *Proceedings of the Eleventh Uai Conference on Bayesian Modeling Applications Workshop-Volume 1218*, 79–88. CEUR-WS. org.

Pernin, Peter. 1971. "The Great Peshtigo Fire: An Eyewitness Account." *The Wisconsin*

Magazine of History. JSTOR, 246–72.

Peterson, David A, Edward J Hyer, James R Campbell, Michael D Fromm, Johnathan W Hair, Carolyn F Butler, and Marta A Fenn. 2015. "The 2013 Rim Fire: Implications for Predicting Extreme Fire Spread, Pyroconvection, and Smoke Emissions." *Bulletin of the American Meteorological Society* 96 (2): 229–47.

Piironen, Juho, Aki Vehtari, and others. 2017. "Sparsity Information and Regularization in the Horseshoe and Other Shrinkage Priors." *Electronic Journal of Statistics* 11 (2). The Institute of Mathematical Statistics; the Bernoulli Society: 5018–51.

Preisler, Haiganoush K, and Anthony L Westerling. 2007. "Statistical Model for Forecasting Monthly Large Wildfire Events in Western United States." *Journal of Applied Meteorology and Climatology* 46 (7): 1020–30.

Preisler, Haiganoush K, David R Brillinger, Robert E Burgan, and JW Benoit. 2004. "Probability Based Models for Estimation of Wildfire Risk." *International Journal of Wildland Fire* 13 (2). CSIRO: 133–42.

Quionero-Candela, Joaquin, Masashi Sugiyama, Anton Schwaighofer, and Neil D Lawrence. 2009. *Dataset Shift in Machine Learning*. The MIT Press.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

———. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Radeloff, Volker C, David P Helmers, H Anu Kramer, Miranda H Mockrin, Patricia M Alexandre, Avi Bar-Massada, Van Butsic, et al. 2018. "Rapid Growth of the Us Wildland-Urban Interface Raises Wildfire Risk." *Proceedings of the National Academy of Sciences* 115 (13). National Acad Sciences: 3314–9.

Radeloff, Volker C, Susan I Stewart, Todd J Hawbaker, Urs Gimmi, Anna M Pidgeon, Curtis H Flather, Roger B Hammer, and David P Helmers. 2010. "Housing Growth in and Near United States Protected Areas Limits Their Conservation Value." *Proceedings of the*

National Academy of Sciences 107 (2). National Acad Sciences: 940–45.

Ramsay, James O, and others. 1988. "Monotone Regression Splines in Action." *Statistical Science* 3 (4). Institute of Mathematical Statistics: 425–41.

Reed, William J, and Kevin S McKelvey. 2002. "Power-Law Behaviour and Parametric Models for the Size-Distribution of Forest Fires." *Ecological Modelling* 150 (3). Elsevier: 239–54.

Reid, Colleen E, Michael Brauer, Fay H Johnston, Michael Jerrett, John R Balmes, and Catherine T Elliott. 2016. "Critical Review of Health Impacts of Wildfire Smoke Exposure." *Environmental Health Perspectives* 124 (9). National Institute of Environmental Health Science: 1334.

Rogers, Brendan M., Ronald P. Neilson, Ray Drapek, James M. Lenihan, John R. Wells, Dominique Bachelet, and Beverly E. Law. 2011. "Impacts of Climate Change on Fire Regimes and Carbon Stocks of the U.S. Pacific Northwest." *Journal of Geophysical Research: Biogeosciences* 116 (G3): n/a–n/a. doi:10.1029/2011JG001695.

Rollins, Matthew G, Penelope Morgan, and Thomas Swetnam. 2002. "Landscape-Scale Controls over 20th Century Fire Occurrence in Two Large Rocky Mountain (Usa) Wilderness Areas." *Landscape Ecology* 17 (6). Springer: 539–57.

Schoenberg, Frederic Paik, Roger Peng, and James Woods. 2003. "On the Distribution of Wildfire Sizes." *Environmetrics* 14 (6). Wiley Online Library: 583–92.

Serra, Laura, Marc Saez, Pablo Juan, Diego Varga, and Jorge Mateu. 2014. "A Spatio-Temporal Poisson Hurdle Point Process to Model Wildfires." *Stochastic Environmental Research and Risk Assessment* 28 (7). Springer: 1671–84.

Serra, Laura, Marc Saez, Jorge Mateu, Diego Varga, Pablo Juan, Carlos Díaz-Ávalos, and Håvard Rue. 2014. "Spatio-Temporal Log-Gaussian Cox Processes for Modelling Wildfire Occurrence: The Case of Catalonia, 1994–2008." *Environmental and Ecological Statistics* 21 (3). Springer: 531–63.

Shirota, Shinichiro, and Sudipto Banerjee. 2018. "Scalable Inference for Space-Time Gaussian Cox Processes." *arXiv Preprint arXiv:1802.06151*.

Slowikowski, Kamil. 2018. *Ggrepel: Automatically Position Non-Overlapping Text Labels with 'Ggplot2'*. https://CRAN.R-project.org/package=ggrepel.

Solymos, Peter, and Zygmunt Zawadzki. 2018. *Pbapply: Adding Progress Bar to '*Apply' Functions*. https://CRAN.R-project.org/package=pbapply.

Stallman, Richard M., Roland McGrath, and Paul D. Smith. 2004. *GNU Make: A Program for Directing Recompilation, for Version 3.81*. Free Software Foundation.

Stan Development Team. 2018. "RStan: The R Interface to Stan." http://mc-stan.org/.

Stavros, E Natasha, John T Abatzoglou, Donald McKenzie, and Narasimhan K Larkin. 2014. "Regional Projections of the Likelihood of Very Large Wildland Fires Under a Changing Climate in the Contiguous Western United States." *Climatic Change* 126 (3-4). Springer: 455–68.

Syphard, Alexandra D. AD, Volker C. VC Radeloff, Jon E. Keeley, Todd J. Hawbaker, Murray K. Clayton, Susan I. Stewart, and Roger B. Hammer. 2007. "Human influence on California fire regimes." *Ecological Applications* 17 (5): 1388–1402.

Syphard, Alexandra D., Jon E. Keeley, Anne H. Pfaff, and Ken Ferschweiler. 2017. "Human presence diminishes the importance of climate in driving fire activity across the United States." *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1713885114.

Tedim, Fantina, Vittorio Leone, Malik Amraoui, Christophe Bouillon, Michael R Coughlan, Giuseppe M Delogu, Paulo M Fernandes, et al. 2018. "Defining Extreme Wildfire Events: Difficulties, Challenges, and Impacts." *Fire* 1 (1). Multidisciplinary Digital Publishing Institute: 9.

Tran, Dustin, Matthew D Hoffman, Rif A Saurous, Eugene Brevdo, Kevin Murphy, and David M Blei. 2017. "Deep Probabilistic Programming." *arXiv Preprint arXiv:1701.03757*.

Vilar, Lara, Douglas G Woolford, David L Martell, and M Pilar Martín. 2010. "A Model for Predicting Human-Caused Wildfire Occurrence in the Region of Madrid, Spain." *Inter-*

national Journal of Wildland Fire 19 (3). CSIRO: 325–37.

Westerling, AL, BP Bryant, HK Preisler, TP Holmes, HG Hidalgo, T Das, and SR Shrestha. 2011. "Climate Change and Growth Scenarios for California Wildfire." *Climatic Change* 109 (1). Springer: 445–63.

Westerling, Anthony LeRoy. 2016. "Increasing Western Us Forest Wildfire Activity: Sensitivity to Changes in the Timing of Spring." *Phil. Trans. R. Soc. B* 371 (1696). The Royal Society: 20150178.

Wickham, Hadley. 2017a. *Assertthat: Easy Pre and Post Assertions*. https://CRAN.R-project.org/package=assertthat.

———. 2017b. *Tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse.

Wiitala, Marc R. 1999. "Assessing the Risk of Cumulative Burned Acreage Using the Poisson Probability Model." *Fire Economics, Planning, and Policy: Bottom Lines*. Citeseer, 51.

Wilke, Claus O. 2017. *Cowplot: Streamlined Plot Theme and Plot Annotations for 'Ggplot2'*. https://CRAN.R-project.org/package=cowplot.

Williams, Jerry. 2013. "Exploring the Onset of High-Impact Mega-Fires Through a Forest Land Management Prism." *Forest Ecology and Management* 294. Elsevier: 4–10.

Wood, S.N. 2017. *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman; Hall/CRC.

Woolford, DG, DR Bellhouse, WJ Braun, Ch B Dean, DL Martell, and J Sun. 2011. "A Spatio-Temporal Model for People-Caused Forest Fire Occurrence in the Romeo Malette Forest." *Journal of Environmental Statistics* 2: 2–16.

Woolford, Douglas G, CB Dean, David L Martell, Jiguo Cao, and BM Wotton. 2014. "Lightning-Caused Forest Fire Risk in Northwestern Ontario, Canada, Is Increasing and Associated with Anomalies in Fire Weather." *Environmetrics* 25 (6). Wiley Online Library:

406–16.

Yao, Yuling, Aki Vehtari, Daniel Simpson, Andrew Gelman, and others. 2017. "Using Stacking to Average Bayesian Predictive Distributions." *Bayesian Analysis*. International Society for Bayesian Analysis.

Zeileis, Achim, and Gabor Grothendieck. 2005. "Zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software* 14 (6): 1–27. doi:10.18637/jss.v014.i06.

Zorzetto, E, G Botter, and M Marani. 2016. "On the Emergence of Rainfall Extremes from Ordinary Events." *Geophysical Research Letters* 43 (15). Wiley Online Library: 8076–82.

780 # Tables

Table 1. Performance of count models on the test set in descending order. Posterior means are provided with standard deviations in parentheses.

Model	Holdout log likelihood
ZI Negative binomial	-3671 (70)
ZI Poisson	-4093 (77)
Negative binomial	-4298 (114)
Poisson	-4572 (139)

Table 2. Performance of burned area models on the test set in descending order. Posterior means are provided with standard deviations in parentheses.

Model	Holdout log likelihood
Lognormal	-26341 (43)
Generalized Pareto	-26377 (45)
Tapered Pareto	-26386 (49)
Weibull	-27592 (236)
Gamma	-30675 (993)

781 # Figures

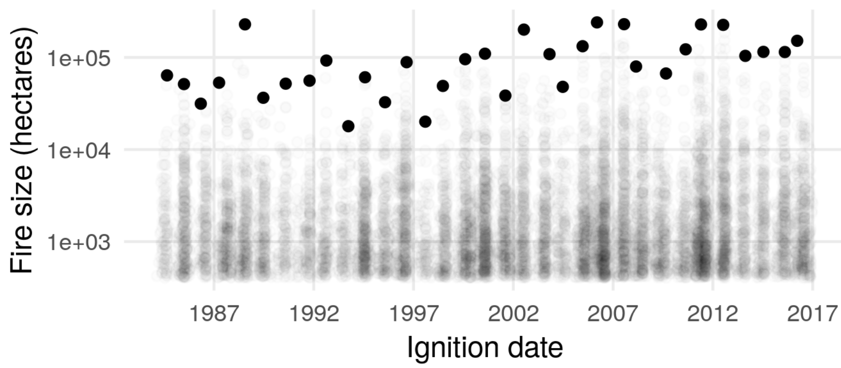

Figure 1. Sizes of wildfires over 405 hectares in the contiguous United States, from the Monitoring Trends in Burn Severity multiagency program. Each point represents a fire event, and the largest fires for each year (the block maxima) are shown as solid black points.

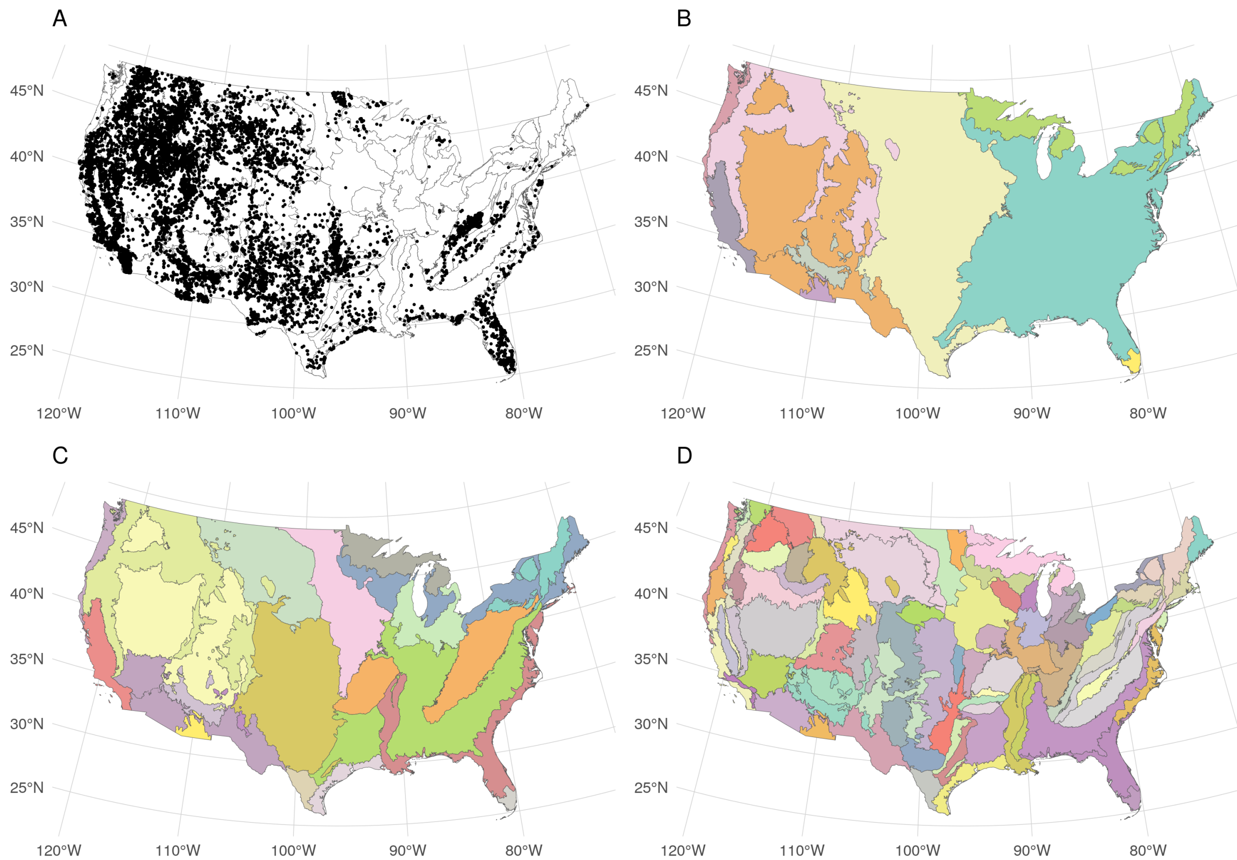

Figure 2. A. Large wildfire ignition locations are shown as points across the study region. Colors in panels B, C, and D show level 1, 2, and 3 ecoregions respectively.

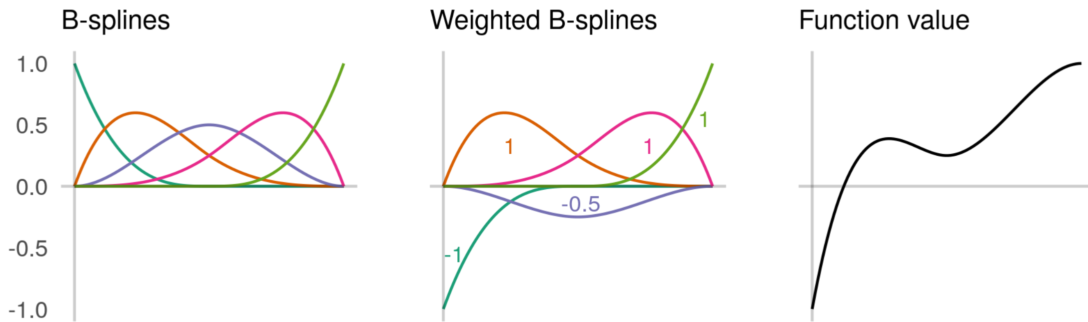

Figure 3. Conceptual figure to illustrate the use of B-splines to construct nonlinear functions. In the left panel, five B-spline vectors are shown, which map values of an input variable (on the x-axis) to a value on the y-axis. The middle panel shows the same B-spline vectors, but weighted (multiplied) by real numbers, with the weights illustrated as annotations. These weighted B-spline vectors are summed to produce the values of a nonlinear function (right panel).

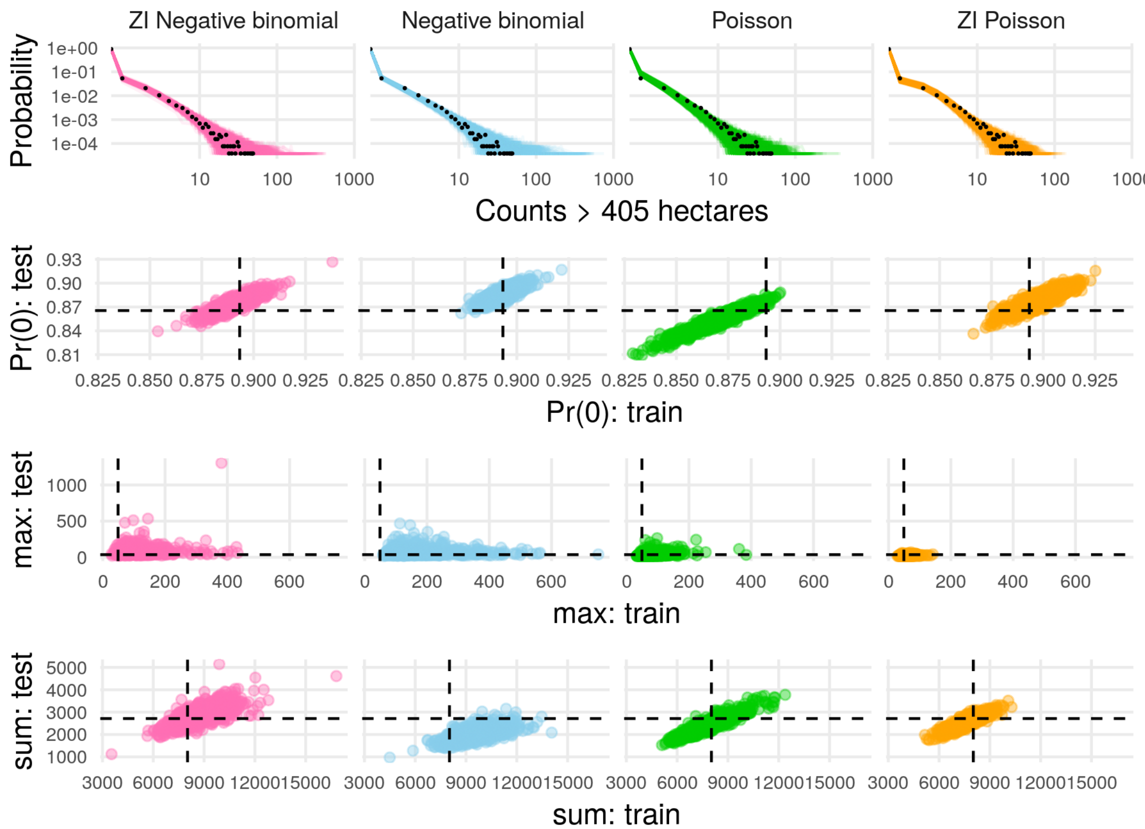

Figure 4. Count predictive checks. Row one shows observed count frequencies as black points and predicted frequencies as lines. Rows two, three, and four show predicted proportions of zeros, maxima, and sums (respectively) in the training and test data, with empirical values as dashed lines. Rows two through four facilitate comparison of performance on training and test sets. Ideally, model predictions cluster around the dashed lines for both the training (x-axis direction) and test (y-axis direction) sets, leading to a tight cluster of points at the intersection of the dashed lines.

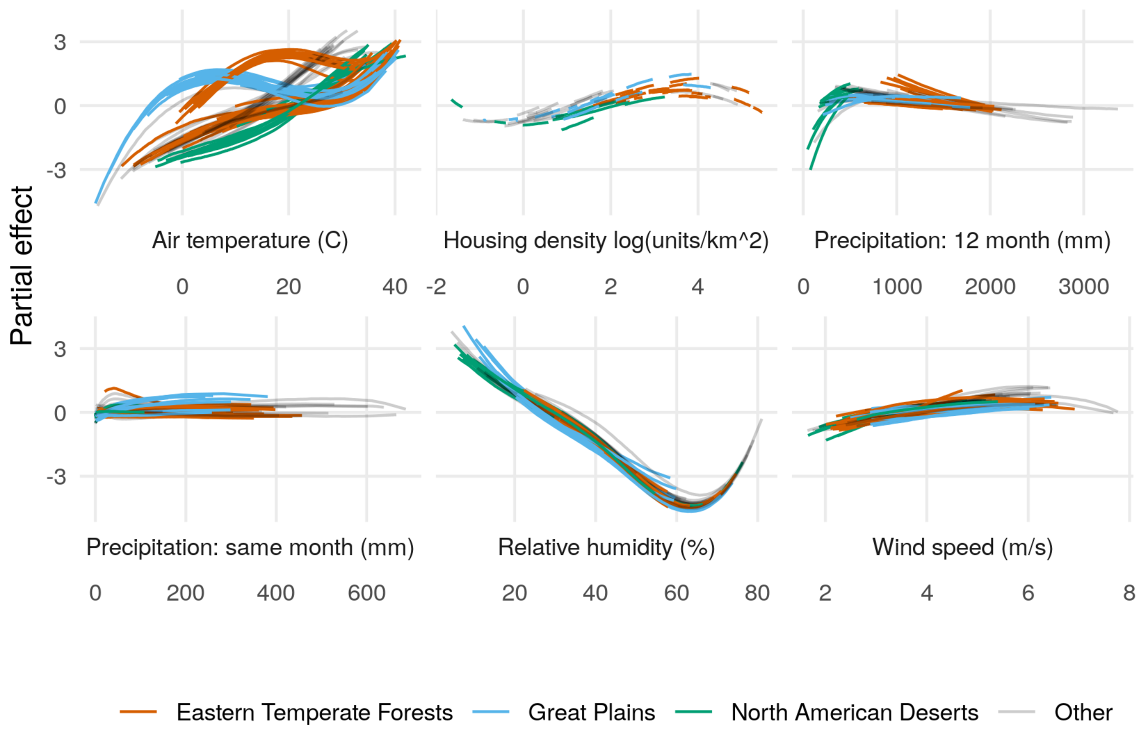

Figure 5. Partial effects on the log-transformed negative binomial mean component of the zero-inflated negative binomial model for each level 3 ecoregion, colored by level 1 ecoregion. Lines are posterior medians. Results are similar for the zero-inflation component.

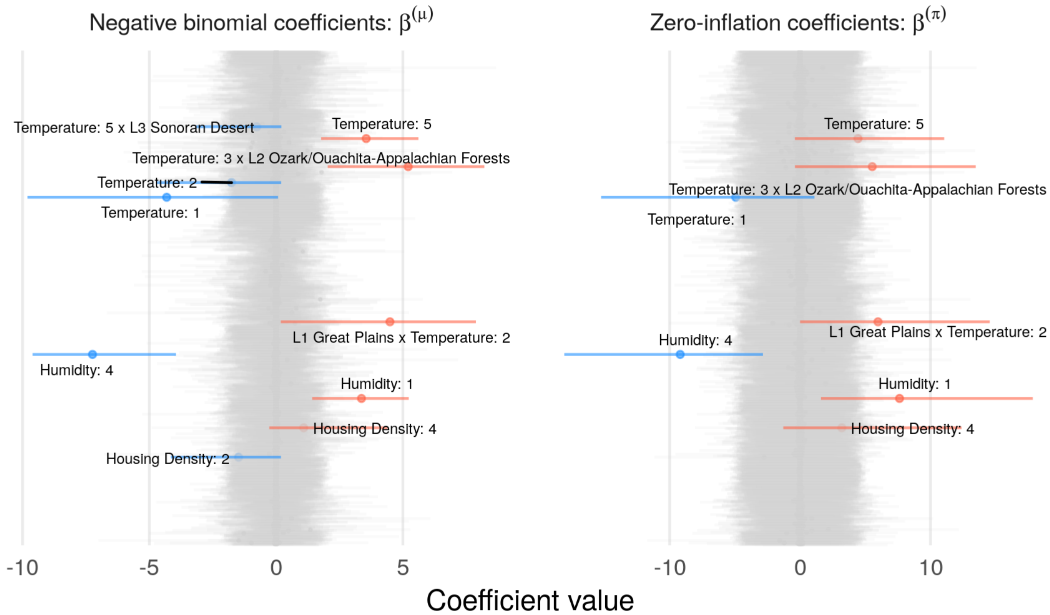

Figure 6. Caterpillar plots of zero inflated negative binomial model coefficients, $\beta^{(\mu)}$ (left) and $\beta^{(\pi)}$ (right). Horizontal line segments denote 95% credible intervals. Grey segments indicate coefficients with a less than 87% posterior probability of being positive or negative, and colored segments indicate coefficients that are probably positive (red) or negative (blue). B-spline vectors are indicated by colons, e.g., `Humidity:1` indicates the first basis vector corresponding to humidity. Interactions between variables `a` and `b` are represented as `a x b`. Level 1 ecoregions are represented by `L1 ecoregion name`, and `L2` and `L3` indicate level 2 and 3 ecoregions.

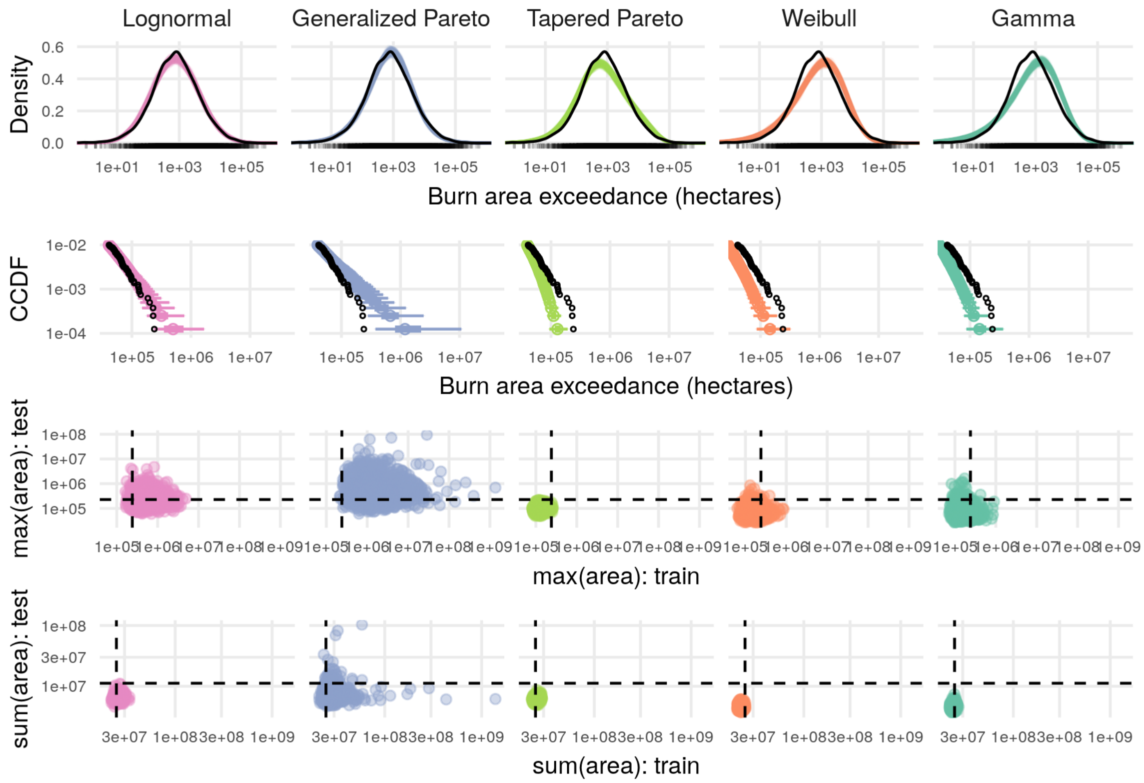

Figure 7. Predictive checks for burned area models. The top row shows predicted density in color and empirical density for the training set in black, which reveals overall lack of fit for the gamma and Weibull models. Row two shows the complementary cumulative distribution function (CCDF) at the tails, with 95% and 50% prediction intervals shown in color and observed data as black points, which shows that the Generalized Pareto distribution predicts values that are too extreme. The third and fourth rows show checks for maximum and total burned areas in the training and test set, with observed values as dashed lines and posterior draws as colored points. These final two rows facilitate checks for summary statistics on both the training and test set, with the ideal model generating predictions (colored points) clustered close to where the dashed lines intersect.

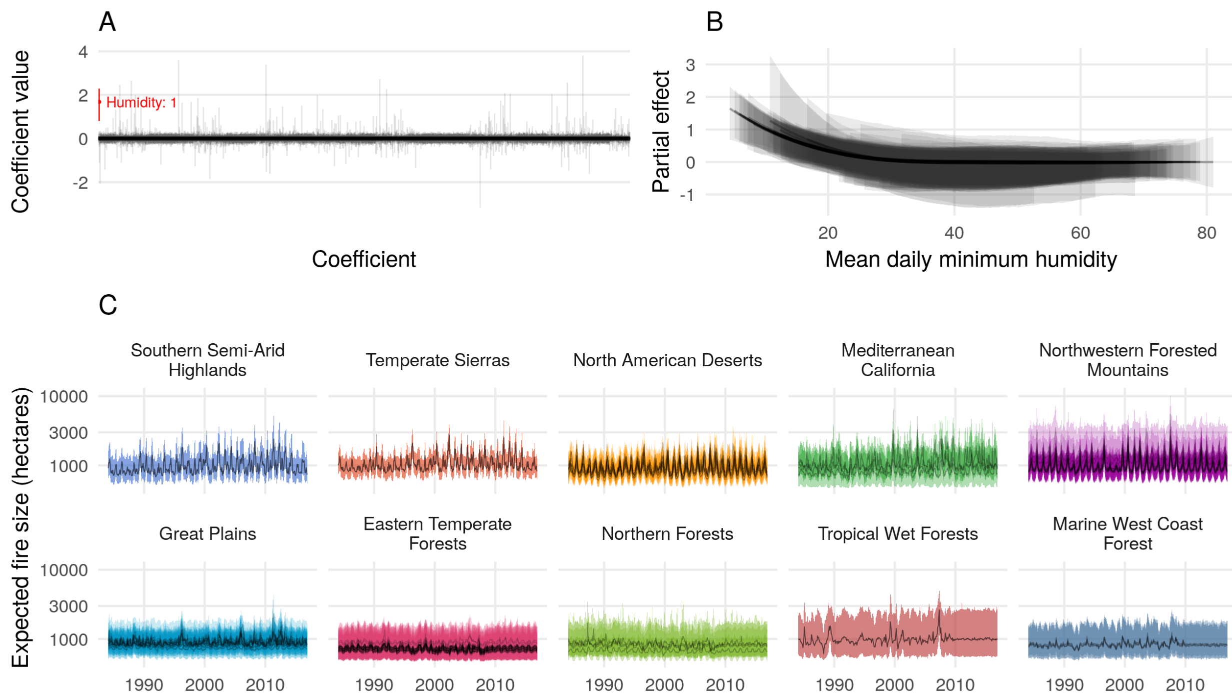

Figure 8. **A**. Estimated posterior medians and 95% credible intervals for each of the 3,473 coefficients associated with expected burned area. Only one coefficient - the first basis vector for humidity - had a 95% credible interval that excluded zero, shown in red. This effect is visualized in **B**. Partial effects of mean daily minimum humidity for each level 3 ecoregion, with posterior medians drawn as lines, and the 95% credible intervals as ribbons. **C**. Monthly time series of expected fire sizes for every level 3 ecoregion, faceted and colored by level 1 ecoregions sorted by mean humidity. Lines are posterior medians and ribbons are 95% credible intervals.

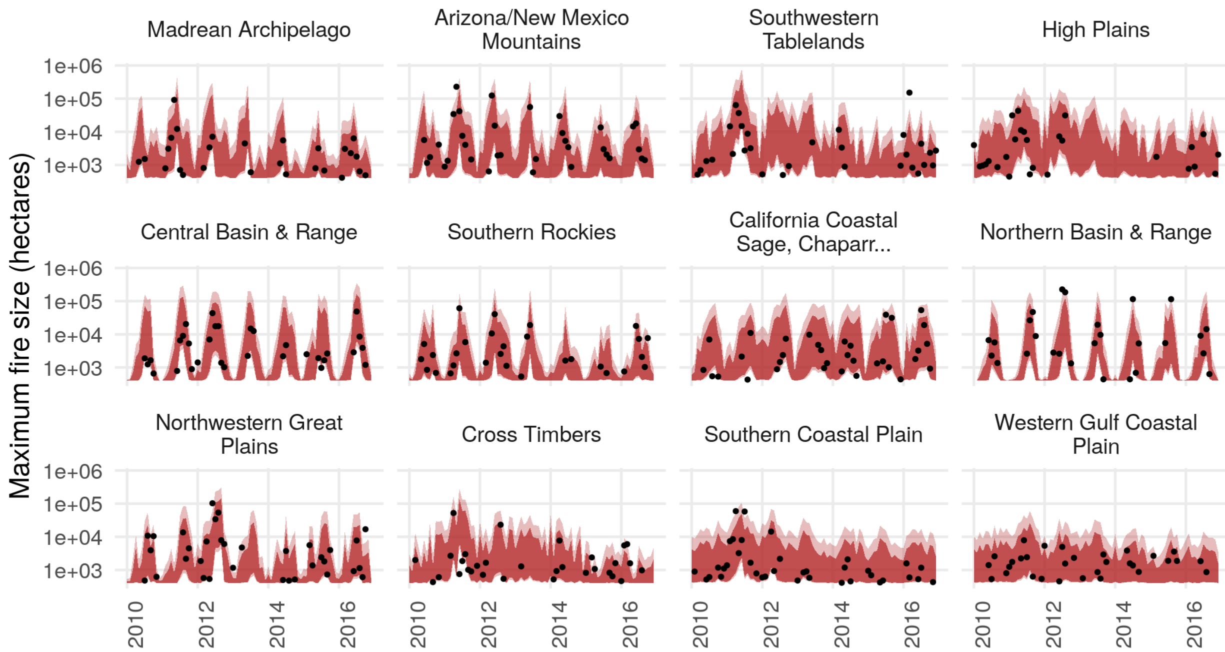

Figure 9. Posterior 99% (light red) and 95% (dark red) prediction intervals for the burned area of the largest fire event by month and level 3 ecoregion in the test set, shown for ecoregions with wildfires in more than 20 months. Empirical maxima are shown as black dots.

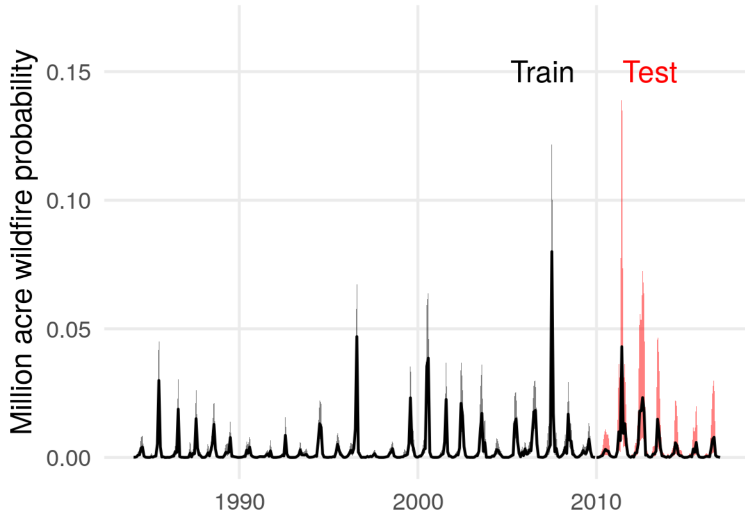

Figure 10. Estimated monthly posterior probabilities that one or more fire events exceed one million acres (404,686 hectares). The line represents the posterior median, and shaded region represents an 80% credible interval. The training period up to 2010 is shown in black, and the test period for which data were withheld during parameter estimation is shown in red.

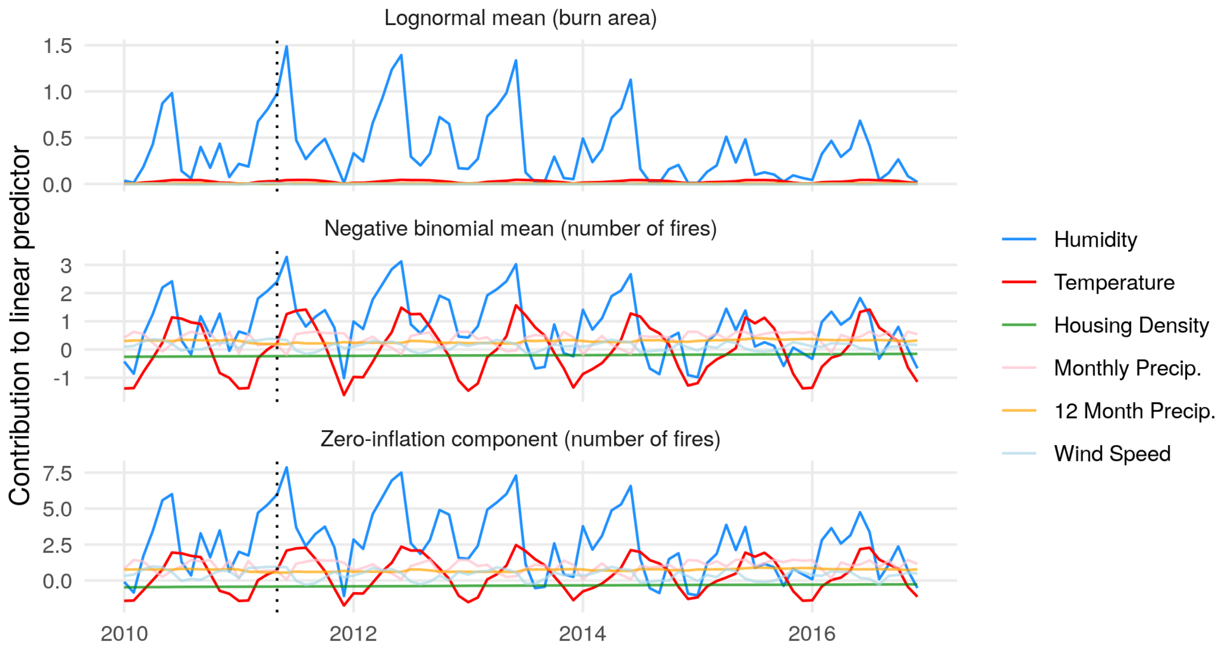

Figure 11. Posterior median contribution of each input variable to the linear predictor function of model components for the Arizona/New Mexico Mountains level 3 ecoregion from 2010-2016. A dotted vertical line marks May 2011, when the Wallow Fire ignited. Vertical positions of colored lines show contributions to the linear predictor function of each model component.

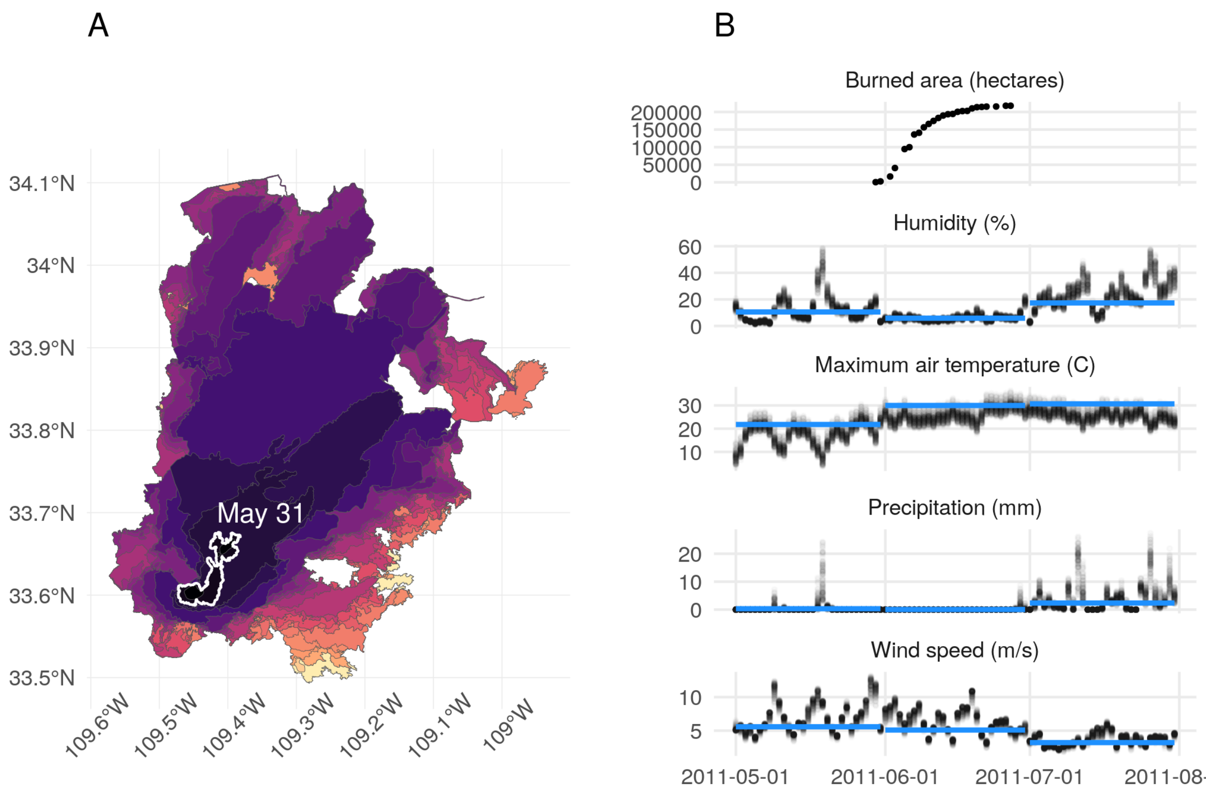

Figure 12. A. Progression of the Wallow Fire based on perimeter data collected from the GeoMAC database, spanning May 30, 2011 to June 27, 2011. The perimeter at the end of may is outlined in white, with brighter colors indicating later dates. B. Daily local meteorological conditions and monthly regional mean conditions for the Wallow Fire, along with the associated burned area over time. The blue line represents monthly averages of meteorological quantities computed over the entire Arizona/New Mexico Mountains ecoregion, and black points represent values extracted for "local" 4 km grid cells contained within the final burned area perimeter of the Wallow Fire.

Appendices

Prior specifications

Prior distributions were chosen to regularize coefficients on the distribution specific means $\beta^{(\mu)}$ and structural zero parameters $\beta^{(\pi)}$. We used a regularized horseshoe prior on these coefficients, which shrinks irrelevant coefficients towards zero, while regularizing nonzero coefficients (Piironen, Vehtari, and others 2017). For zero-inflated models, we used a multivariate version of the regularized horseshoe (Peltola et al. 2014):

$$\begin{pmatrix} \beta_j^{(\mu)} \\ \beta_j^{(\pi)} \end{pmatrix} \sim \mathrm{N}\left(\mathbf{0}, \begin{pmatrix} \tau_1^2 \tilde{\lambda}_{1,j}^2 & \rho\tau_1\tau_2\tilde{\lambda}_{1,j}\tilde{\lambda}_{2,j} \\ \rho\tau_1\tau_2\tilde{\lambda}_{1,j}\tilde{\lambda}_{2,j} & \tau_2^2\tilde{\lambda}_{2,j}^2 \end{pmatrix}\right),$$

$$\tilde{\lambda}_{m,j}^2 = \frac{c_m^2 \lambda_j^2}{c_m^2 + \tau_m^2 \lambda_j^2},$$

for each response dimension $m = 1, 2$ and coefficient $j = 1, ..., p$. Here ρ is a correlation parameter, τ_1 and τ_2 are global variance hyperparameters, c_1 and c_2 are hyperparameters that determine the amount of shrinkage on the largest coefficients, and λ_j is a local scale parameter drawn from a half-Cauchy distribution that control the amount of shrinkage applied to coefficient j (Piironen, Vehtari, and others 2017). With this prior specification, information can be shared across the two response dimensions through the correlation parameter ρ, and/or through the local scale parameters λ_j. For count models without structural zeros (the Poisson and negative binomial models), this multivariate prior simplifies to a univariate regularized horseshoe prior.

Spatiotemporal random effects were constructed using a temporally autoregressive, spatially intrinsically autoregressive formulation (Besag and Kooperberg 1995; Banerjee, Carlin, and Gelfand 2014). Temporarily suppressing the superscript that indicates whether the effects are on μ or π, and denoting column t from an $S \times T$ $\mathbf{\Phi}$ as $\boldsymbol{\phi}_t$ we have:

$$\boldsymbol{\phi}_{t=1} \sim \mathrm{N}(\mathbf{0}, (\tau^{(\phi)}(\mathbf{D} - \mathbf{W}))^{-1})$$

$$\boldsymbol{\phi}_t \sim \mathrm{N}(\eta \boldsymbol{\phi}_{t-1}, (\tau^{(\phi)}(\mathbf{D} - \mathbf{W}))^{-1}), \quad t = 2, ..., T$$

where η is a temporal dependence parameter, $\tau^{(\phi)}$ is a precision parameter, $\mathbf{D}$ is an $S \times S$ diagonal matrix with entries corresponding to the number of spatial neighbors for each spatial unit, and $\mathbf{W}$ is an $S \times S$ spatial adjacency matrix with nonzero elements only when spatial unit i is a neighbor of spatial unit j ($w_{i,j} = 1$ if i is a neighbor of j, and $w_{i,j} = 0$ otherwise, including $w_{i,i} = 0$ for all i). $\tau^{(\phi)}$ is a precision parameter. We imposed a soft identifiability constraint that places high prior mass near $\sum_{s=1}^{S} \phi^*_{t,s} = 0$ for all t.

We applied a univariate regularized horseshoe prior to all β coefficients in burned area models (Piironen, Vehtari, and others 2017):

$$\beta_j \sim \mathrm{N}\Big(0, \tau^2 \tilde{\lambda}_j^2\Big), \quad \tilde{\lambda}_j^2 = \frac{c^2 \lambda_j^2}{c^2 + \tau^2 \lambda_j^2},$$

Spatiotemporal random effects were constructed in the same way as for the count models.

Joint distributions

Here we provide the unnormalized posterior densities for each model. Square brackets represent a probability mass or density function. Parameterizations for model likelihoods are provided first, followed by the factorization of the joint distribution, with explicit priors.

Poisson wildfire count model

We used the following parameterization of the Poisson distribution:

$$[n|\mu] = \frac{\mu^n e^{-\mu}}{n!},$$

where μ is the mean and variance.

The unnormalized posterior density of this model is:

$$\begin{aligned}
[\boldsymbol{\beta}^{(\mu)}, \alpha^{(\mu)}, \boldsymbol{\phi}, \sigma^{(\phi)}, \eta, \boldsymbol{\lambda}, c, \tau \mid \mathbf{N}] \propto \\
\prod_{s=1}^{S} \prod_{t=1}^{T} [n_{s,t} | \boldsymbol{\beta}^{(\mu)}, \alpha^{(\mu)}, \phi_{s,t}] \times \\
[\boldsymbol{\phi}_1 | \sigma^{(\phi)}] \prod_{t=2}^{T} [\boldsymbol{\phi}_t | \boldsymbol{\phi}_{t-1}, \sigma^{(\phi)}, \eta] \times \\
\prod_{j=1}^{p} [\beta_j^{(\mu)} | \lambda_j, c, \tau][\lambda_j] \times \\
[\sigma^{(\phi)}][\eta][c][\tau][\alpha^{(\mu)}]
\end{aligned}$$

$$
\begin{aligned}
&= \prod_{s=1}^{S}\prod_{t=1}^{T} \mathrm{Poisson}(n_{s,t}|\exp(\alpha^{(\mu)} + \mathbf{X}_{(s,t)}\boldsymbol{\beta}^{(\mu)} + \phi_{s,t}))\times \\
&\mathrm{Normal}(\boldsymbol{\phi}_1|\mathbf{0}, ((\sigma^{(\phi)})^{-2}(\mathbf{D}-\mathbf{W}))^{-1})\times \\
&\prod_{t=2}^{T}\mathrm{Normal}(\boldsymbol{\phi}_t|\eta\boldsymbol{\phi}_{t-1}, ((\sigma^{(\phi)})^{-2}(\mathbf{D}-\mathbf{W}))^{-1})\times \\
&\prod_{j=1}^{p}\mathrm{Normal}\left(\beta_j^{(\mu)}|0, \frac{\tau^2 c^2 \lambda_j^2}{c^2+\tau^2\lambda_j^2}\right)\times \mathrm{Cauchy}^+(\lambda_j|0,1)\times \\
&\mathrm{Normal}^+(\sigma^{(\phi)}|0,1^2)\times \mathrm{Beta}(\eta|1,1)\times \text{Inv-Gamma}(c^2|2.5,10)\times \\
&\mathrm{Normal}^+(\tau|0,5^2)\times \mathrm{Normal}(\alpha^{(\mu)}|0,5^2).
\end{aligned}
$$

Negative binomial wildfire count model

We used the following parameterization of the negative binomial distribution:

$$[n|\mu, \delta] = \binom{n+\delta-1}{n}\Big(\frac{\mu}{\mu+\delta}\Big)^{n}\Big(\frac{\delta}{\mu+\delta}\Big)^{\delta},$$

where μ is the mean, and δ is a dispersion parameter.

The unnormalized posterior density of this model is:

$$\begin{aligned}
&[\boldsymbol{\beta}^{(\mu)}, \alpha^{(\mu)}, \boldsymbol{\phi}, \sigma^{(\phi)}, \eta, \boldsymbol{\lambda}, c, \tau, \delta \mid \mathbf{N}] \propto \\
&\prod_{s=1}^{S}\prod_{t=1}^{T}[n_{s,t}|\boldsymbol{\beta}^{(\mu)}, \alpha^{(\mu)}, \phi_{s,t}, \delta]\times \\
&[\boldsymbol{\phi}_1|\sigma^{(\phi)}]\prod_{t=2}^{T}[\boldsymbol{\phi}_t|\boldsymbol{\phi}_{t-1}, \sigma^{(\phi)}, \eta]\times \\
&\prod_{j=1}^{p}[\beta_j^{(\mu)}|\lambda_j, c, \tau][\lambda_j]\times \\
&[\sigma^{(\phi)}][\eta][c][\tau][\alpha^{(\mu)}][\delta]
\end{aligned}$$

$$\begin{aligned}
&= \prod_{s=1}^{S}\prod_{t=1}^{T}\text{Negative Binomial}(n_{s,t}|\exp(\alpha^{(\mu)} + \mathbf{X}_{(s,t)}\boldsymbol{\beta}^{(\mu)} + \phi_{s,t}), \delta)\times \\
&\text{Normal}(\boldsymbol{\phi}_1|\mathbf{0}, ((\sigma^{(\phi)})^{-2}(\mathbf{D}-\mathbf{W}))^{-1})\times \\
&\prod_{t=2}^{T}\text{Normal}(\boldsymbol{\phi}_t|\eta\boldsymbol{\phi}_{t-1}, ((\sigma^{(\phi)})^{-2}(\mathbf{D}-\mathbf{W}))^{-1})\times \\
&\prod_{j=1}^{p}\text{Normal}\Bigg(\beta_j^{(\mu)}|0, \frac{\tau^2 c^2 \lambda_j^2}{c^2+\tau^2\lambda_j^2}\Bigg)\times \text{Cauchy}^{+}(\lambda_j|0,1)\times \\
&\text{Normal}^{+}(\sigma^{(\phi)}|0, 1^2)\times \text{Beta}(\eta|1,1)\times \text{Inv-Gamma}(c^2|2.5, 10)\times \\
&\text{Normal}^{+}(\tau|0, 5^2)\times \text{Normal}(\alpha^{(\mu)}|0, 5^2)\times \text{Normal}^{+}(\delta|0, 5^2).
\end{aligned}$$

Zero-inflated Poisson wildfire count model

We used the following parameterization of the zero-inflated Poisson distribution:

$$[n|\mu, \pi] = I_{n=0}(1 - \pi + \pi e^{-\mu}) + I_{n>0}\pi \frac{\mu^n e^{-\mu}}{n!},$$

where μ is the Poisson mean, and $1 - \pi$ is the probability of an extra zero.

The unnormalized posterior density of this model is:

$$\begin{aligned}
[\boldsymbol{\beta}^{(\mu)}, \alpha^{(\mu)}, \boldsymbol{\beta}^{(\pi)}, \alpha^{(\pi)}, \boldsymbol{\phi}^{(\mu)}, \sigma^{(\phi,\mu)}, \eta^{(\mu)}, \boldsymbol{\phi}^{(\pi)}, \sigma^{(\phi,\pi)}, \eta^{(\pi)}, \boldsymbol{\lambda}, c, \tau, \rho \mid \mathbf{N}] \propto \\
\prod_{s=1}^{S} \prod_{t=1}^{T} [n_{s,t} | \boldsymbol{\beta}^{(\mu)}, \alpha^{(\mu)}, \boldsymbol{\beta}^{(\pi)}, \alpha^{(\pi)}, \phi_{s,t}^{(\mu)}, \phi_{s,t}^{(\pi)}] \times \\
[\boldsymbol{\phi}_1^{(\mu)} | \sigma^{(\phi,\mu)}] \prod_{t=2}^{T} [\boldsymbol{\phi}_t^{(\mu)} | \boldsymbol{\phi}_{t-1}^{(\mu)}, \sigma^{(\phi,\mu)}, \eta^{(\mu)}] \times \\
[\boldsymbol{\phi}_1^{(\pi)} | \sigma^{(\phi,\pi)}] \prod_{t=2}^{T} [\boldsymbol{\phi}_t^{(\pi)} | \boldsymbol{\phi}_{t-1}^{(\pi)}, \sigma^{(\phi,\pi)}, \eta^{(\pi)}] \times \\
\prod_{j=1}^{p} [\beta_j^{(\mu)}, \beta_j^{(\pi)} | \lambda_j, c, \tau, \rho][\lambda_j] \times \\
[\sigma^{(\phi,\mu)}][\sigma^{(\phi,\pi)}][\eta^{(\mu)}][\eta^{(\pi)}][\alpha^{(\mu)}][\alpha^{(\pi)}][\rho] \prod_{m=1}^{2} [c_m][\tau_m]
\end{aligned}$$

$$
\begin{aligned}
= \prod_{s=1}^{S}\prod_{t=1}^{T} \mathrm{ZIP}(n_{s,t}|e^{\alpha^{(\mu)}+\mathbf{X}_{(s,t)}\beta^{(\mu)}+\phi^{(\mu)}_{s,t}}, \mathrm{logit}^{-1}(\alpha^{(\pi)}+\mathbf{X}_{(s,t)}\boldsymbol{\beta}^{(\pi)}+\phi^{(\pi)}_{s,t}))\times \\
\mathrm{Normal}(\boldsymbol{\phi}^{(\mu)}_1|\mathbf{0},((\sigma^{(\phi,\mu)})^{-2}(\mathbf{D}-\mathbf{W}))^{-1})\times \\
\prod_{t=2}^{T}\mathrm{Normal}(\boldsymbol{\phi}^{(\mu)}_t|\eta^{(\mu)}\boldsymbol{\phi}^{(\mu)}_{t-1},((\sigma^{(\phi,\mu)})^{-2}(\mathbf{D}-\mathbf{W}))^{-1})\times \\
\mathrm{Normal}(\boldsymbol{\phi}^{(\pi)}_1|\mathbf{0},((\sigma^{(\phi,\pi)})^{-2}(\mathbf{D}-\mathbf{W}))^{-1})\times \\
\prod_{t=2}^{T}\mathrm{Normal}(\boldsymbol{\phi}^{(\pi)}_t|\eta^{(\pi)}\boldsymbol{\phi}^{(\pi)}_{t-1},((\sigma^{(\phi,\pi)})^{-2}(\mathbf{D}-\mathbf{W}))^{-1})\times \\
\prod_{j=1}^{p}\mathrm{N}\left(\left|\begin{pmatrix}\beta^{(\mu)}_j\\ \beta^{(\pi)}_j\end{pmatrix}\right|\mathbf{0},\left|\begin{matrix}\tau_1^2\frac{c_1^2\lambda_j^2}{c_1^2+\tau_1^2\lambda_j^2} & \rho\tau_1\tau_2\sqrt{\frac{c_1^2\lambda_j^2}{c_1^2+\tau_1^2\lambda_j^2}}\sqrt{\frac{c_2^2\lambda_j^2}{c_2^2+\tau_2^2\lambda_j^2}}\\ \rho\tau_1\tau_2\sqrt{\frac{c_1^2\lambda_j^2}{c_1^2+\tau_1^2\lambda_j^2}}\sqrt{\frac{c_2^2\lambda_j^2}{c_2^2+\tau_2^2\lambda_j^2}} & \tau_2^2\frac{c_2^2\lambda_j^2}{c_2^2+\tau_2^2\lambda_j^2}\end{matrix}\right)\right)\times \\
\prod_{j=1}^{p}\mathrm{Cauchy}^{+}(\lambda_j|0,1)\times \\
\mathrm{Normal}^{+}(\sigma^{(\phi,\mu)}|0,1^2)\times\mathrm{Normal}^{+}(\sigma^{(\phi,\pi)}|0,1^2)\times \\
\mathrm{Beta}(\eta^{(\mu)}|1,1)\times\mathrm{Beta}(\eta^{(\pi)}|1,1)\times \\
\mathrm{Normal}(\alpha^{(\mu)}|0,5^2)\times\mathrm{Normal}(\alpha^{(\pi)}|0,5^2)\times\mathrm{LKJ}(\rho|3)\times \\
\prod_{m=1}^{2}\mathrm{Inv\text{-}Gamma}(c_m^2|2.5,10)\times\mathrm{Normal}^{+}(\tau_m|0,5^2).
\end{aligned}
$$

Zero-inflated negative binomial wildfire count model

We used the following parameterization of the zero-inflated negative binomial distribution:

$$[n|\mu, \delta, \pi] = I_{n=0}(1 - \pi + \pi\Big(\frac{\delta}{\mu + \delta}\Big)^{\delta}) + I_{n>0}\binom{n + \delta - 1}{n}\Big(\frac{\mu}{\mu + \delta}\Big)^{n}\Big(\frac{\delta}{\mu + \delta}\Big)^{\delta},$$

where μ is the negative binomial mean, δ is the negative binomial dispersion, and , and $1 - \pi$ is the probability of an extra zero.

The unnormalized posterior density of this model is:

$$\begin{aligned}
[\boldsymbol{\beta}^{(\mu)}, \alpha^{(\mu)}, \boldsymbol{\beta}^{(\pi)}, \alpha^{(\pi)}, \boldsymbol{\phi}^{(\mu)}, \sigma^{(\phi,\mu)}, \eta^{(\mu)}, \boldsymbol{\phi}^{(\pi)}, \sigma^{(\phi,\pi)}, \eta^{(\pi)}, \boldsymbol{\lambda}, c, \tau, \rho, \delta \mid \mathbf{N}] \propto \\
\prod_{s=1}^{S}\prod_{t=1}^{T}[n_{s,t}|\boldsymbol{\beta}^{(\mu)}, \alpha^{(\mu)}, \boldsymbol{\beta}^{(\pi)}, \alpha^{(\pi)}, \phi_{s,t}^{(\mu)}, \phi_{s,t}^{(\pi)}, \delta]\times \\
[\boldsymbol{\phi}_1^{(\mu)}|\sigma^{(\phi,\mu)}]\prod_{t=2}^{T}[\boldsymbol{\phi}_t^{(\mu)}|\boldsymbol{\phi}_{t-1}^{(\mu)}, \sigma^{(\phi,\mu)}, \eta^{(\mu)}]\times \\
[\boldsymbol{\phi}_1^{(\pi)}|\sigma^{(\phi,\pi)}]\prod_{t=2}^{T}[\boldsymbol{\phi}_t^{(\pi)}|\boldsymbol{\phi}_{t-1}^{(\pi)}, \sigma^{(\phi,\pi)}, \eta^{(\pi)}]\times \\
\prod_{j=1}^{p}[\beta_j^{(\mu)}, \beta_j^{(\pi)}|\lambda_j, c, \tau, \rho][\lambda_j]\times \\
[\sigma^{(\phi,\mu)}][\sigma^{(\phi,\pi)}][\eta^{(\mu)}][\eta^{(\pi)}][\alpha^{(\mu)}][\alpha^{(\pi)}][\rho][\delta]\prod_{m=1}^{2}[c_m][\tau_m].
\end{aligned}$$

$$
\begin{aligned}
= \prod_{s=1}^{S}\prod_{t=1}^{T} \mathrm{ZINB}(n_{s,t} | e^{\alpha^{(\mu)} + \mathbf{X}_{(s,t)}\beta^{(\mu)} + \phi_{s,t}^{(\mu)}}, \delta, \mathrm{logit}^{-1}(\alpha^{(\pi)} + \mathbf{X}_{(s,t)}\boldsymbol{\beta}^{(\pi)} + \phi_{s,t}^{(\pi)}))\times \\
\mathrm{Normal}(\boldsymbol{\phi}_1^{(\mu)} | \mathbf{0}, ((\sigma^{(\phi,\mu)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1})\times \\
\prod_{t=2}^{T} \mathrm{Normal}(\boldsymbol{\phi}_t^{(\mu)} | \eta^{(\mu)}\boldsymbol{\phi}_{t-1}^{(\mu)}, ((\sigma^{(\phi,\mu)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1})\times \\
\mathrm{Normal}(\boldsymbol{\phi}_1^{(\pi)} | \mathbf{0}, ((\sigma^{(\phi,\pi)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1})\times \\
\prod_{t=2}^{T} \mathrm{Normal}(\boldsymbol{\phi}_t^{(\pi)} | \eta^{(\pi)}\boldsymbol{\phi}_{t-1}^{(\pi)}, ((\sigma^{(\phi,\pi)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1})\times \\
\prod_{j=1}^{p} \mathrm{N}\left(\left(\begin{array}{c} \beta_j^{(\mu)} \\ \beta_j^{(\pi)} \end{array} \right) \middle| \mathbf{0}, \left(\begin{array}{cc} \tau_1^2 \frac{c_1^2\lambda_j^2}{c_1^2+\tau_1^2\lambda_j^2} & \rho\tau_1\tau_2\sqrt{\frac{c_1^2\lambda_j^2}{c_1^2+\tau_1^2\lambda_j^2}}\sqrt{\frac{c_2^2\lambda_j^2}{c_2^2+\tau_2^2\lambda_j^2}} \\ \rho\tau_1\tau_2\sqrt{\frac{c_1^2\lambda_j^2}{c_1^2+\tau_1^2\lambda_j^2}}\sqrt{\frac{c_2^2\lambda_j^2}{c_2^2+\tau_2^2\lambda_j^2}} & \tau_2^2 \frac{c_2^2\lambda_j^2}{c_2^2+\tau_2^2\lambda_j^2} \end{array} \right) \right)\times \\
\prod_{j=1}^{p} \mathrm{Cauchy}^{+}(\lambda_j | 0, 1)\times \\
\mathrm{Normal}^{+}(\sigma^{(\phi,\mu)} | 0, 1^2) \times \mathrm{Normal}^{+}(\sigma^{(\phi,\pi)} | 0, 1^2)\times \\
\mathrm{Beta}(\eta^{(\mu)} | 1, 1) \times \mathrm{Beta}(\eta^{(\pi)} | 1, 1)\times \\
\mathrm{Normal}(\alpha^{(\mu)} | 0, 5^2) \times \mathrm{Normal}(\alpha^{(\pi)} | 0, 5^2) \times \mathrm{LKJ}(\rho | 3) \times \mathrm{Normal}^{+}(\delta | 0, 5^2)\times \\
\prod_{m=1}^{2} \mathrm{Inv\text{-}Gamma}(c_m^2 | 2.5, 10) \times \mathrm{Normal}^{+}(\tau_m | 0, 5^2).
\end{aligned}
$$

Generalized Pareto/Lomax burned area model

We used the following parameterization of the GPD/Lomax distribution:

$$[y|\sigma, \kappa] = \frac{1}{\sigma}\Big(\frac{\kappa y}{\sigma} + 1\Big)^{-(\kappa+1)\kappa^{-1}},$$

where κ is a shape parameter and σ is a scale parameter.

The unnormalized posterior density of this model is:

$$\begin{aligned}
&[\boldsymbol{\beta}, \alpha, \boldsymbol{\phi}, \sigma^{(\phi)}, \eta, \kappa^{(L)}, \boldsymbol{\lambda}, c, \tau \mid \boldsymbol{y}] \propto \\
&\prod_{i=1}^{n_{\text{tot}}} [y_i | \boldsymbol{\beta}, \alpha, \phi_{s_i,t_i}, \kappa^{(L)}] \times \\
&[\boldsymbol{\phi}_1 | \sigma^{(\phi)}] \prod_{t=2}^{T} [\boldsymbol{\phi}_t | \boldsymbol{\phi}_{t-1}, \sigma^{(\phi)}, \eta] \times \\
&\prod_{j=1}^{p} [\beta_j | \lambda_j, c, \tau][\lambda_j] \times \\
&[\alpha][c][\tau][\kappa^{(L)}][\eta][\sigma^{(\phi)}]
\end{aligned}$$

$$\begin{aligned}
&= \prod_{i=1}^{n_{\text{tot}}} \text{Lomax}(y_i | \kappa^{(L)}, e^{\alpha + \mathbf{X}_{(s_i,t_i)}\boldsymbol{\beta} + \phi_{s_i,t_i}}) \times \\
&\text{Normal}(\boldsymbol{\phi}_1 | \mathbf{0}, ((\sigma^{(\phi)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1}) \times \\
&\prod_{t=2}^{T} \text{Normal}(\boldsymbol{\phi}_t | \eta \boldsymbol{\phi}_{t-1}, ((\sigma^{(\phi)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1}) \times \\
&\prod_{j=1}^{p} \text{Normal}\Big(\beta_j | 0, \frac{\tau^2 c^2 \lambda_j^2}{c^2 + \tau^2 \lambda_j^2}\Big) \times \text{Cauchy}^+(\lambda_j | 0, 1) \times \\
&\text{Normal}(\alpha | 0, 5^2) \times \text{Inv-Gamma}(c^2 | 2.5, 10) \times \text{Normal}^+(\tau | 0, 5^2) \\
&\text{Normal}^+(\kappa^{(L)} | 0, 5^2) \times \text{Beta}(\eta | 1, 1) \times \text{Normal}^+(\sigma^{(\phi)} | 0, 1^2).
\end{aligned}$$

Tapered Pareto burned area model

We used the following parameterization of the tapered Pareto distribution:

$$[y|\kappa,\nu] = \left(\frac{\kappa}{y} + \frac{1}{\nu}\right)\exp(-x/\nu),$$

where κ is a shape parameter and ν a taper parameter.

The unnormalized posterior density of this model is:

$$\begin{aligned}
[\boldsymbol{\beta}, \alpha, \boldsymbol{\phi}, \sigma^{(\phi)}, \eta, \nu, \boldsymbol{\lambda}, c, \tau \mid \boldsymbol{y}] \propto \\
\prod_{i=1}^{n_{\text{tot}}} [y_i|\boldsymbol{\beta}, \alpha, \phi_{s_i,t_i}, \nu] \times \\
[\boldsymbol{\phi}_1|\sigma^{(\phi)}] \prod_{t=2}^{T} [\boldsymbol{\phi}_t|\boldsymbol{\phi}_{t-1}, \sigma^{(\phi)}, \eta] \times \\
\prod_{j=1}^{p} [\beta_j|\lambda_j, c, \tau][\lambda_j] \times \\
[\alpha][c][\tau][\nu][\eta][\sigma^{(\phi)}]
\end{aligned}$$

$$\begin{aligned}
= \prod_{i=1}^{n_{\text{tot}}} \text{Tapered Pareto}(y_i|e^{\alpha + \mathbf{X}_{(s_i,t_i)}\boldsymbol{\beta} + \phi_{s_i,t_i}}, \nu) \times \\
\text{Normal}(\boldsymbol{\phi}_1|\mathbf{0}, ((\sigma^{(\phi)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1}) \times \\
\prod_{t=2}^{T} \text{Normal}(\boldsymbol{\phi}_t|\eta\boldsymbol{\phi}_{t-1}, ((\sigma^{(\phi)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1}) \times \\
\prod_{j=1}^{p} \text{Normal}\left(\beta_j|0, \frac{\tau^2 c^2 \lambda_j^2}{c^2 + \tau^2 \lambda_j^2}\right) \times \text{Cauchy}^{+}(\lambda_j|0,1) \times \\
\text{Normal}(\alpha|0, 5^2) \times \text{Inv-Gamma}(c^2|2.5, 10) \times \text{Normal}^{+}(\tau|0, 5^2) \times \\
\text{Cauchy}^{+}(\nu|0,1) \times \text{Beta}(\eta|1,1) \times \text{Normal}^{+}(\sigma^{(\phi)}|0, 1^2).
\end{aligned}$$

Lognormal burned area model

We used the following parameterization of the lognormal distribution:

$$[y|\mu, \sigma] = \frac{1}{y} \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(- \frac{(\log(y) - \mu)^2}{2\sigma^2} \Big),$$

where μ and σ are location and scale parameters, respectively.

The unnormalized posterior density of this model is:

$$\begin{aligned}
&[\boldsymbol{\beta}, \alpha, \boldsymbol{\phi}, \sigma^{(\phi)}, \eta, \sigma, \boldsymbol{\lambda}, c, \tau \mid \boldsymbol{y}] \propto \\
&\prod_{i=1}^{n_{\text{tot}}} [y_i|\beta, \alpha, \phi_{s_i,t_i}, \sigma] \times \\
&[\boldsymbol{\phi}_1|\sigma^{(\phi)}] \prod_{t=2}^{T} [\boldsymbol{\phi}_t|\boldsymbol{\phi}_{t-1}, \sigma^{(\phi)}, \eta] \times \\
&\prod_{j=1}^{p} [\beta_j|\lambda_j, c, \tau][\lambda_j] \times \\
&[\alpha][c][\tau][\sigma][\eta][\sigma^{(\phi)}]
\end{aligned}$$

$$\begin{aligned}
&= \prod_{i=1}^{n_{\text{tot}}} \text{Lognormal}(y_i|\alpha + \mathbf{X}_{(s_i,t_i)}\boldsymbol{\beta} + \phi_{s_i,t_i}, \sigma) \times \\
&\text{Normal}(\boldsymbol{\phi}_1|\mathbf{0}, ((\sigma^{(\phi)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1}) \times \\
&\prod_{t=2}^{T} \text{Normal}(\boldsymbol{\phi}_t|\eta\boldsymbol{\phi}_{t-1}, ((\sigma^{(\phi)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1}) \times \\
&\prod_{j=1}^{p} \text{Normal}\bigg(\beta_j|0, \frac{\tau^2 c^2 \lambda_j^2}{c^2 + \tau^2 \lambda_j^2}\bigg) \times \text{Cauchy}^+(\lambda_j|0, 1) \times \\
&\text{Normal}(\alpha|0, 5^2) \times \text{Inv-Gamma}(c^2|2.5, 10) \times \text{Normal}^+(\tau|0, 5^2) \times \\
&\text{Normal}^+(\sigma|0, 5^2) \times \text{Beta}(\eta|1, 1) \times \text{Normal}^+(\sigma^{(\phi)}|0, 1^2).
\end{aligned}$$

Gamma burned area model

We used the following parameterization of the gamma distribution:

$$[y|\kappa, \sigma] = \frac{1}{\Gamma(\kappa)\sigma^{\kappa}} y^{\kappa-1} \exp(-y/\sigma),$$

where κ is a shape parameter and σ a scale parameter.

The unnormalized posterior density of this model is:

$$\begin{aligned}
[\boldsymbol{\beta}, \alpha, \boldsymbol{\phi}, \sigma^{(\phi)}, \eta, \kappa, \boldsymbol{\lambda}, c, \tau \mid \boldsymbol{y}] \propto \\
\prod_{i=1}^{n_{\text{tot}}} [y_i | \beta, \alpha, \phi_{s_i, t_i}, \kappa] \times \\
[\boldsymbol{\phi}_1 | \sigma^{(\phi)}] \prod_{t=2}^{T} [\boldsymbol{\phi}_t | \boldsymbol{\phi}_{t-1}, \sigma^{(\phi)}, \eta] \times \\
\prod_{j=1}^{p} [\beta_j | \lambda_j, c, \tau][\lambda_j] \times \\
[\alpha][c][\tau][\kappa][\eta][\sigma^{(\phi)}]
\end{aligned}$$

$$\begin{aligned}
= \prod_{i=1}^{n_{\text{tot}}} \text{Gamma}(y_i | \kappa, \kappa / \exp(\alpha + \mathbf{X}_{(s_i, t_i)} \boldsymbol{\beta} + \phi_{s_i, t_i})) \times \\
\text{Normal}(\boldsymbol{\phi}_1 | \mathbf{0}, ((\sigma^{(\phi)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1}) \times \\
\prod_{t=2}^{T} \text{Normal}(\boldsymbol{\phi}_t | \eta \boldsymbol{\phi}_{t-1}, ((\sigma^{(\phi)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1}) \times \\
\prod_{j=1}^{p} \text{Normal}\left(\beta_j | 0, \frac{\tau^2 c^2 \lambda_j^2}{c^2 + \tau^2 \lambda_j^2}\right) \times \text{Cauchy}^+(\lambda_j | 0, 1) \times \\
\text{Normal}(\alpha | 0, 5^2) \times \text{Inv-Gamma}(c^2 | 2.5, 10) \times \text{Normal}^+(\tau | 0, 5^2) \times \\
\text{Normal}^+(\kappa | 0, 5^2) \times \text{Beta}(\eta | 1, 1) \times \text{Normal}^+(\sigma^{(\phi)} | 0, 1^2).
\end{aligned}$$

Weibull burned area model

We used the following parameterization of the Weibull distribution:

$$[y|\kappa, \sigma] = \frac{\kappa}{\sigma}\Big(\frac{y}{\sigma}\Big)^{\kappa-1}\exp\Big(-\Big(\frac{y}{\sigma}\Big)^{\alpha}\Big),$$

where κ is a shape parameter and σ is a scale parameter.

The unnormalized posterior density of this model is:

$$\begin{aligned}
[\boldsymbol{\beta}, \alpha, \boldsymbol{\phi}, \sigma^{(\phi)}, \eta, \kappa, \lambda, c, \tau \mid \boldsymbol{y}] \propto \\
\prod_{i=1}^{n_{\text{tot}}} [y_i|\beta, \alpha, \phi_{s_i,t_i}, \kappa] \times \\
[\boldsymbol{\phi}_1|\sigma^{(\phi)}] \prod_{t=2}^{T} [\boldsymbol{\phi}_t|\boldsymbol{\phi}_{t-1}, \sigma^{(\phi)}, \eta] \times \\
\prod_{j=1}^{p} [\beta_j|\lambda_j, c, \tau][\lambda_j] \times \\
[\alpha][c][\tau][\kappa][\eta][\sigma^{(\phi)}]
\end{aligned}$$

$$\begin{aligned}
= \prod_{i=1}^{n_{\text{tot}}} \text{Weibull}(y_i|\kappa, \exp(\alpha + \mathbf{X}_{(s_i,t_i)}\boldsymbol{\beta} + \phi_{s_i,t_i})) \times \\
\text{Normal}(\boldsymbol{\phi}_1|\mathbf{0}, ((\sigma^{(\phi)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1}) \times \\
\prod_{t=2}^{T} \text{Normal}(\boldsymbol{\phi}_t|\eta\boldsymbol{\phi}_{t-1}, ((\sigma^{(\phi)})^{-2}(\mathbf{D} - \mathbf{W}))^{-1}) \times \\
\prod_{j=1}^{p} \text{Normal}\Big(\beta_j|0, \frac{\tau^2 c^2 \lambda_j^2}{c^2 + \tau^2\lambda_j^2}\Big) \times \text{Cauchy}^+(\lambda_j|0, 1) \times \\
\text{Normal}(\alpha|0, 5^2) \times \text{Inv-Gamma}(c^2|2.5, 10) \times \text{Normal}^+(\tau|0, 5^2) \times \\
\text{Normal}^+(\kappa|0, 5^2) \times \text{Beta}(\eta|1, 1) \times \text{Normal}^+(\sigma^{(\phi)}|0, 1^2).
\end{aligned}$$