DISSERTATION

BAYESIAN APPROACHES TO EXTREME VALUE MODELING, WITH APPLICATIONS TO WILDFIRES

Submitted by

Elizabeth S. Lawler

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2025

Doctoral Committee:

Advisor: Benjamin Shaby

Daniel Cooley
Tianjian Zhou
Hussam Mahmoud

ABSTRACT

BAYESIAN APPROACHES TO EXTREME VALUE MODELING,
WITH APPLICATIONS TO WILDFIRES

The growing frequency and size of wildfires across the US necessitates accurate quantitative assessment of evolving wildfire behavior to predict risk from future extreme wildfires. In Chapter 2, we build a joint model of wildfire counts and burned areas, regressing key model parameters on climate and demographic covariates. We use extended generalized Pareto distributions to model the full distribution of burned areas, capturing both moderate and extreme sizes, while leveraging extreme value theory to focus particularly on the right tail. We model wildfire counts using a zero-inflated negative binomial model and join the wildfire counts and burned areas sub-models via a temporally varying shared random effect. Our model successfully captures the trends of wildfire counts and burned areas. By investigating the predictive power of different sets of covariates, we find that fire indices are better predictors of wildfire burned area behavior than individual climate covariates, whereas climate covariates are influential drivers of wildfire occurrence behavior.

Recent advances in multivariate extreme value modeling leverage a geometric perspective, using the shape of the multivariate point cloud and its connection to the Lebesgue joint density, to make inference on joint tail probabilities. While the original statistical framework was fully parametric, relying on a gauge function that uniquely defines the shape for a given density, newer methods have introduced semi- and non-parametric alternatives to increase flexibility. In Chapter 3, we propose a modeling approach that retains the simplicity of the parametric framework but adds flexibility by using Bayesian model averaging (BMA) to improve prediction of tail risk probabilities. In contrast to previous works that rely solely on a truncated radial likelihood, we propose using a censored likelihood, which we find consistently outperforms the truncated radial likelihood, particularly in small-sample settings. To generate predictions, we use a simple

importance sampling scheme that matches the accuracy of more complex methods at a fraction of the computational cost. Finally, we apply our approach to two fire weather indices, which are designed to capture somewhat orthogonal aspects of fire risk, to illustrate the practical utility of our method in environmental applications.

ACKNOWLEDGEMENTS

First and foremost, I'm incredibly grateful for my advisor, Ben Shaby. His expertise in extreme value theory and Bayesian modeling has been instrumental in my development and growth as a statistician and researcher. His thoughtful critiques and questions—not only in our meetings, but also in seminars, conferences, and reading group—have helped me begin to develop my own critical eye as a researcher. My graduate school experience was peppered with many non-academic challenges, from the disruptions of COVID to mental health struggles, and Ben's care and understanding made a real difference. His empathy and flexibility gave me the space to succeed on my own terms, and I'm incredibly thankful for that.

I'd like to thank my committee members: Dan Cooley and Tianjian Zhou from Statistics, and Hussam Mahmoud from Civil and Environmental Engineering. Given Dan's extensive background in extremes, especially multivariate extremes, he offered valuable critiques of my research and challenged me to think about broader questions encompassing my work. Tianjian's Bayesian course provided a strong theoretical foundation that underpinned much of my applied work, and as a committee member, he encouraged me to explore other ways in which Bayesian model averaging could be applied in my research. Hussam challenged me to think more deeply about other factors contributing to fire behavior, beyond the specific modeling focus of my work.

My family's constant love and encouragement were essential throughout these past six years. I'm grateful to my parents, Mark and Sally, for their unwavering support and for their efforts to understand what exactly it is I do. To my sister, Emily, for being my confidante, a source of laughter, and a shoulder to lean on. Sithee, for always sending me back to Colorado with a stack of za'atar for sustenance and a reminder of home. And to my aunts, uncles, and cousins—thank you for the steady stream of encouragement and love along the way.

My friends, both near and far, kept me grounded and reminded me there was life outside of grad school. I'm so grateful for such an incredible "chosen family", and I couldn't have asked for a better support system.

TABLE OF CONTENTS

Chapter 1

Introduction

The Los Angeles wildfires of January 2025 serve as a stark reminder of the destructive impacts of climate change and the unpreparedness of many communities for events like these. The fires led to 29 deaths and caused widespread damage. A mix of strong Santa Ana winds and dry, flammable vegetation created the perfect conditions for disaster. In the weeks and months that followed, the scale and suddenness of the event prompted numerous responses, from climate attribution analyses (World Weather Attribution, 2025) to the launch of a decade-long health study: the Los Angeles Fire Human Exposure and Long-Term Health Study (L.A. Fire HEALTH Study, lafirehealth.org).

The increasing frequency and intensity of wildfires—both in the United States (U.S.) and globally—show no sign of slowing down. This makes it all the more critical that scientific research plays a role in shaping policy and fire risk management across local, regional, and national levels. Extreme value theory (EVT) is well-suited for modeling these types of events and making predictions that extend beyond the range of observed data. In Chapter 2, we explore this in a univariate setting, modeling wildfire burned area across the U.S. using meteorological variables, housing density, and fire weather indices. In Chapter 3, we extend this work to the multivariate setting, focusing on the joint tail behavior of two fire weather indices that capture different aspects of fire behavior and can serve as strong predictors of occurrence and spread.

1.1 Univariate extreme value theory

Univariate EVT provides a comprehensive framework for modeling rare events. Of particular relevance to this dissertation is the threshold exceedances approach, in which the distribution of values above a high threshold is well-approximated by a generalized Pareto distribution (GPD). This allows us to model the tail behavior of continuous, positive-valued phenomena like wildfire burned area. We provide a more thorough overview of this framework in Chapter 2, where we model wildfire burned area using an extension of the GPD.

1.2 Multivariate extreme value theory

There are multiple approaches to modeling multivariate extremes, with the multivariate regular variation (MRV) framework being one of the most widely used. In this setting, data are typically transformed to Fréchet margins, and a pseudo-polar coordinate transformation yields independent radial and angular components. This representation is informative when variables exhibit asymptotic dependence (AD), as the mass of extreme observations lies in the interior of the positive orthant. In contrast, under asymptotic independence (AI), mass accumulates on the coordinate axes, rendering the angular measure degenerate and the MRV framework less informative. These behaviors are illustrated in Figure 1.1.

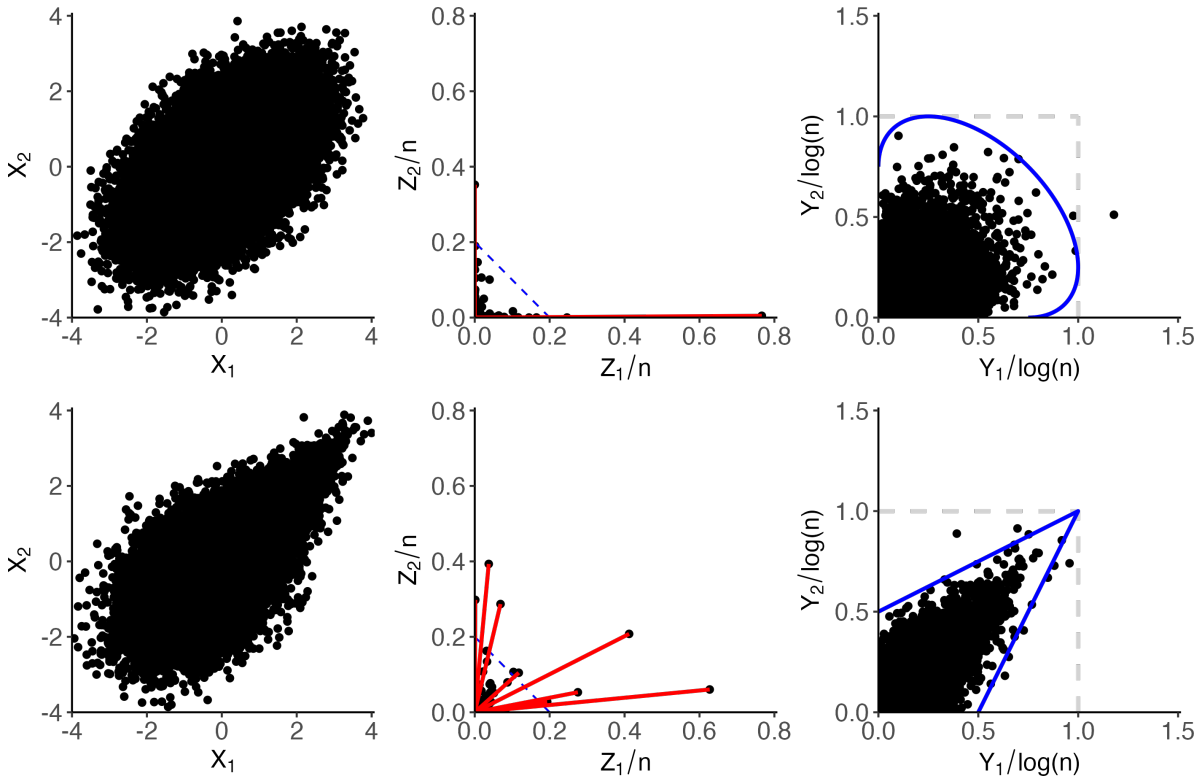


Figure 1.1: Simulated observations from a bivariate Gaussian (top) and bivariate logistic (bottom) distribution with mid-range dependence ($\rho = \lambda = 0.5$), shown under Gaussian (left), Fréchet (middle), and exponential (right) marginal transformations. Red lines in the middle panel extend beyond a threshold. The blue curves in the right panel illustrate the boundary of the limit set.

This dissertation uses an alternative framework—*geometric extremes*—which remains informative under both AD and AI. The geometric framework leverages the structure seen in the third panel of Figure 1.1, using limit sets to provide a more flexible characterization of joint tail behavior. We provide a brief overview of limit sets and their relationship to multivariate random vectors (and their associated extremes) in Chapter 3, where we develop a flexible extension of current parametric methods and apply our method to fire weather indices.

Chapter 2

Anthropogenic and meteorological effects on the counts and sizes of moderate and extreme wildfires[1]

2.1 Introduction

Wildfires have increased in frequency and size across the United States (US) (Dennison et al., 2014; Goss et al., 2020; Iglesias et al., 2022). In the last few years alone, the world has seen an unprecedented speed, intensity, and breadth in wildfires (Duane et al., 2021), with the likelihood of large wildfires expected to increase over the long term (Stavros et al., 2014). As regions continue to develop risk mitigation, land management, and resource allocation policies, it is imperative to quantitatively assess the changing wildfire risk factors and be able to predict future extreme wildfire behavior at a national (and global) level (Goss et al., 2020). To address this problem, we build a joint model for wildfire counts and burned areas in the contiguous United States (CONUS), conditional on meteorological and demographic predictors. We model the entire distribution of burned areas to simultaneously capture extreme and moderate events, leveraging extreme value theory to focus on the right tail, since extreme burn events are the most critical for risk assessment. Our model borrows strength spatially using the concept of ecoregions, which are geographic areas with common topographic features and vegetation types. We link the sub-models for wildfire counts and burned areas with shared random effects.

The proportion of total wildfire burned area in the Western US that can be attributed to the top 1% of the largest wildfires has been shown to range between 80–96% (Strauss et al., 1989), and as large fires increase in frequency across parts of the world, so too will the large proportion of burned area that they cause (Barbero et al., 2014). Climate models have demonstrated that climate change

[1]This chapter is based on the article: Lawler, E. S. and Shaby, B. A. (2024). Anthropogenic and meteorological effects on the counts and sizes of moderate and extreme wildfires. *Environmetrics*, 35(7), e2873. doi:10.1002/env.2873. Reproduced with permission of the publisher.

has and will continue to worsen wildfire weather and lengthen the wildfire season and is anticipated to cause a large increase in the number of expected wildfires through the end of the century (Liu et al., 2010; Martinuzzi et al., 2019; Pimont et al., 2022; Jones et al., 2022). With temperature extremes becoming more common due to climate change, there is a higher risk of seeing extreme wildfire seasons (Parente et al., 2018), particularly like those seen with the Australian wildfires of 2019–2020 (Van Oldenborgh et al., 2021).

When considering the effects of climate change, climate scientists have stressed the importance of studying and understanding the relationships among climate, climate extremes, humans, and wildfire activity to appropriately predict wildfire activity and protect against fires' risks (Jones et al., 2022). Early studies suggest that human factors have led to more than half of the observed increase in wildfire weather and danger (Abatzoglou and Williams, 2016; Williams et al., 2019; Barbero et al., 2020), suggesting that anthroprogenic covariates should be incorporated in wildfire modeling. And while humans have impacted climate change, their presence has also modified wildfire ignition patterns and landscapes so that, when coupled with extreme weather conditions in areas with wildfire fuel sources, there is an increased probability of wildfire ignitions (Pausas and Keeley, 2021; Hawbaker et al., 2013).

Like some previous studies assessing wildfire risk, we use ecoregions to pool information across space according to characteristics that are directly relevant to wildfire behavior (unlike, e.g. political boundaries). Ecoregions provide a convenient way to incorporate differences in covariates along with differences in soil, vegetation, and land-use (Hawbaker et al., 2013; Littell et al., 2009). The feedback process between wildfires and the vegetation they influence further demonstrates the need to look at wildfire patterns by ecoregions (Malamud et al., 2005).

Given this multifaceted nature of the occurrence and burned area of wildfires, we endeavor to tackle this problem by incorporating climate, anthropogenic, and wildfire risk covariates in a joint model that captures zero-inflated counts along with moderate and extreme sized burned areas.

Many previous works have studied the distribution of wildfires. Joseph et al. (2019) also use ecoregions as their areal units, independently modeling wildfire counts using zero-inflated

models and burned areas using a variety of response distributions. However, while their interpretation focuses on extremes, they eschew established best practices by fitting a generalized Pareto distribution to the entire burn size dataset, rather than just the tail, and conversely draw inferences about the tail based on distributions that are not suitable for extremes. Opitz et al. (2020) incorporate a log-Gaussian Cox process model into a Bayesian framework to regress wildfires on a set of land use and weather covariates using Laplace approximations. Pimont et al. (2021) develop a two-component flexible Bayesian model called 'firelihood' to fit the full distribution of wildfire burned areas and wildfire counts. Their approach requires categorizing the wildfire burned areas into four size categories through logistic regression, then fitting Pareto distributions to each. Drawbacks to this approach are that the piecewise specification is discontinuous, requires many parameters, and does not allow sharing of covariate and wildfire information across model pieces. Castel-Clavera et al. (2022) extend 'firelihood' by adding a spatio-temporal component. Becker et al. (2022) use a joint model to look at the counts and sizes of wildfires in a Canadian province, restricting their study period to the months of April–November across four decades, and linking count and burned area sub-models using a shared random effect similar to our model. Koh et al. (2023) builds upon Opitz et al. (2020) and Pimont et al. (2021) by developing a two-component mixture model for wildfire sizes that allows sharing of information between the moderate and extreme components. Their approach requires selecting a threshold selection prior to model fitting. Cisneros et al. (2023) develop a complex four-stage approach using stochastic partial differential equations to model wildfires in the US, which does not allow for sharing of information across stages. Also somewhat related is Richards and Huser (2022), who propose a neural-network model with extreme quantile regression to examine burned fractions of spatial grid boxes in the CONUS. However, their model focuses only on predicting extreme wildfires and does not capture the full distribution of wildfires.

Our approach is novel in three main respects. First, it models the full distribution of wildfire burned areas with a continuous distribution that leverages extreme value theory, considering moderate wildfires while emphasizing inference on the tails, without requiring threshold selection.

These models are taken from the literature but, to our knowledge, have not been previously applied in a hierarchical model for wildfire data. Second, we share information across space using an easily interpretable random effect that partitions the spatial domain by ecological features rather than the more typical political boundaries. This random effects structure reflects the geographical hierarchical nesting of smaller ecoregions within larger ecoregions within even larger ecoregions. Finally, we demonstrate the utility of two fire weather indices, which together capture somewhat orthogonal aspects of fire risk, as effective substitutes for more complicated climate and weather covariates for predicting wildfire burned areas.

2.2 Overview of univariate extreme value theory

We present a brief overview of threshold models in univariate extreme value theory (EVT) (Coles, 2001). In the context of wildfire burned area, denoted as the random variable Y below, the asymptotic characterization of the threshold exceedance model states that the amount of area burned, given that it exceeds a high threshold u, is well-approximated by generalized Pareto distribution (GPD), as the threshold u gets large. That is,

$$P(Y > y \,|\, Y > u) \approx \bar{H}_{\sigma,\xi}(y),$$

where $\bar{H}_{\sigma,\xi}$ is the survival function of the GPD and is defined as

$$\bar{H}_{\sigma,\xi}(y) = \begin{cases} (1 + \xi(y-u)/\sigma)_+^{(-1/\xi)}, & \xi \neq 0 \\ \exp(-(y-u)/\sigma), & \xi = 0 \end{cases} \tag{2.1}$$

where $a_+ = \max(a, 0)$, $\xi \in \mathbb{R}$ is the shape parameter which controls the thickness of the right tail, and $\sigma > 0$ is the scale parameter. The GPD has support $\{y \geq u : 1 + \xi(x-u)/\sigma \geq 0\}$, so that if $\xi < 0$, the upper tail of the distribution is bounded. If $\xi = 0$, the GPD reduces to an exponential distribution, which has finite moments. If $\xi > 0$, the distribution has a heavy upper tail. Empirically, the distribution of wildfire burned area over the CONUS appears to have a heavy

tail (see Figure 2.1), which is consistent with observations in the literature that a small number of fires cause most of the damage (e.g. Strauss et al., 1989). Therefore, we limit the scope of this research to the latter two cases of the GPD, where $\xi \geq 0$.

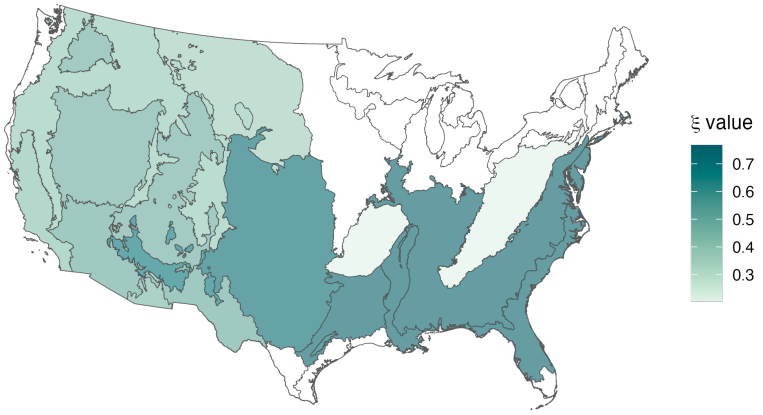


Figure 2.1: Map of maximum likelihood estimates (MLE) of ξ over the contiguous US, split by level 2 (L2) ecoregion. A univariate GPD was fit to every L2 ecoregion that had 20 or more burns at or above its respective 90th quantile. Shades correspond to the estimated shape (ξ) parameter. Higher values of ξ correspond to heavier tails.

In practice, implementation of the GPD model requires choosing a threshold u, subsequently fitting the GPD model only to those values in excess of u. In a standard extreme value analysis (EVA), where interest lies solely in the most extreme events, discarding observations below u is desirable; this "lets the tail speak for itself" and avoids bias that would occur by including moderate and small values in the fitting procedure. In the current analysis, however, moderate and small values are important, even if the extreme values are of primary interest—wildfire burned areas that would fall in the bulk of a distribution still cause damage, socially and ecologically. We therefore seek a model that can accurately model the full distribution of fire sizes, with particular care taken to capture key features of the tail.

2.2.1 Recent extensions of univariate EVT

Several recent works have proposed models for continuous random variables that preserve the properties of the GPD in the tail, while also modeling the full distribution. One approach is to

"stitch" one or more non-extreme distributions to a GPD tail. A subset of these methods require threshold selection prior to model fitting (Castro-Camilo et al., 2019; MacDonald et al., 2011), while others allow for the threshold to be treated as an unknown parameter within the model (Frigessi et al., 2002; Behrens et al., 2004; Tancredi et al., 2006; Carreau and Bengio, 2009; do Nascimento et al., 2012). Another series of approaches models the bulk of the distribution while achieving GPD-like tails through the composition of cumulative distribution functions (cdfs), an idea introduced by Papastathopoulos and Tawn (2013). One flavor of this general strategy involves the composition of only two cdfs and is intended to model positive data where the right tail is of primary interest (Naveau et al., 2016; Gamet and Jalbert, 2022; Tencaliec et al., 2020). A more flexible version allows for the composition of more than two cdfs and can accommodate distributions with more general support and with GPD-like behavior in both the left and right tails, permitting detailed inference on both the very large and very small values (Stein, 2021b,a; Krock et al., 2022). Since our wildfire size data is strictly positive, and primary interest lies in the right tail, we use extended generalized Pareto distribution (EGPD) models from Naveau et al. (2016).

The EGPD (Papastathopoulos and Tawn, 2013; Naveau et al., 2016; Gamet and Jalbert, 2022) is motivated by the probability integral transform and takes the general form

$$F(y) = G\{H_{\sigma,\xi}(y)\}, \tag{2.2}$$

where $H_{\sigma,\xi}(\cdot)$ is the cdf corresponding to the survival function defined in (2.1), and $G(\cdot)$ is a continuous cdf with support on $[0, 1]$, which Naveau et al. (2016) call a *carrier function*. Naveau et al. (2016) propose a set of constraints on G to preserve the upper tail behavior of $H_{\sigma,\xi}$ and give the distribution power law behavior near 0.

These conditions result in a Pareto-like right tail and ensure that the lower portion of the distribution is driven by the carrier function $G(\cdot)$, which can depend on parameters that are estimated from the data. Naveau et al. (2016) propose four parametric families of carrier functions that satisfy the specified constraints:

$$G_1(v) = v^{\kappa},\ \kappa > 0$$

$$G_2(v) = pv^{\kappa_1} + (1-p)v^{\kappa_2},\ \kappa_1, \kappa_2 > 0,\ p \in [0,1]$$

$$G_3(v) = 1 - Q_{\delta}\{(1-v)^{\delta}\},\ \delta > 0$$

$$G_4(v) = [G_3(v)]^{\kappa/2},\ \kappa > 0,$$

where Q_{δ} in the definition of $G_3(v)$ is a Beta cdf, with $\alpha = 1/\delta$ and $\beta = 2$. In our model, we follow de Carvalho et al. (2022) and build upon the EGPD by adding a regression link between covariates and the parameters of the carrier functions.

2.3 United States wildfire data

2.3.1 Wildfire counts and burned area data

We use wildfire data from the Monitory Trends in Burn Severity (MTBS) program from 1990–2020 (Eidenshink et al., 2007), yielding T=372 monthly timesteps. The MTBS provides a robust dataset including wildfire discovery date, wildfire occurrence location, and final burned area of the wildfire, across the entire CONUS. Fire occurrence location is determined by the geographic centroid of the delineated burned area boundary for each wildfire. Fires are included in the MTBS dataset if they are over 1,000 acres and 500 acres in the Western and Eastern United States, respectively. For consistency, we include only fires that are larger than 1,000 acres in our model. This necessitates truncated versions of EGPDs, which are easily accommodated in our model (see Section 2.4.3). To allow for better model mixing, we scale down the burned areas by a factor of 1,000, working in units of thousands of acres.

2.3.2 Meteorological and anthroprogenic data

We include four meteorological covariates and one anthropogenic covariate in our model, broadly following the approach of Joseph et al. (2019). We use the surface meteorological data

aggregated by gridMET (Abatzoglou, 2013), which is provided in daily timesteps at 4km resolution. The gridMET dataset contains daily relative minimum humidity, maximum air temperature, wind velocity, and precipitation accumulation, which we then aggregate to monthly timesteps, by taking the monthly mean for the first three variables and the monthly sum for precipitation. We then further aggregate precipitation on an annual scale to provide a measure of the previous year's accumulated precipitation, giving us five total meteorological variables. We use housing density estimates captured from the 1990, 2000, 2010, and 2020 census surveys as a proxy for anthroprogenic drivers, provided by the SILVIS lab at the block group level (Mockrin et al., 2023).

2.3.3 Regional data

The Environmental Protection Agency (EPA) provides a hierarchical framework of ecological regions (ecoregions) in North America, based on characteristics like vegetation, soil type, and land use (Omernik and Griffith, 2014). While the EPA provides four ecoregion levels (level one being the coarsest, level four the finest), we only use the first three levels due to computational limitations and data sparsity at the finest level. Within the contiguous US, there are S=84 total level three (L3) ecoregions, 20 L2, and 10 L1 (see Figure 2.2).

2.3.4 Fire indices

There is a tradition in wildfire research to combine several (usually weather-related) variables into a single index that is intended to summarize wildfire risk (e.g. Fosberg, 1978). We were curious whether a univariate wildfire index had similar explanatory power in a counts model or an EGPD burned area model, relative to the combination of several weather variables. We therefore experimented with replacing the weather covariates with some combination of two indices, which are designed to capture different aspects of wildfire risk. Fire weather and fuel indices, like the Energy Release Component (ERC, Bradshaw et al., 1984) and Canadian Forest Fire Weather Index (FWI, Turner and Lawson, 1978) have shown conflicting results in their ability to predict extreme wildfires.

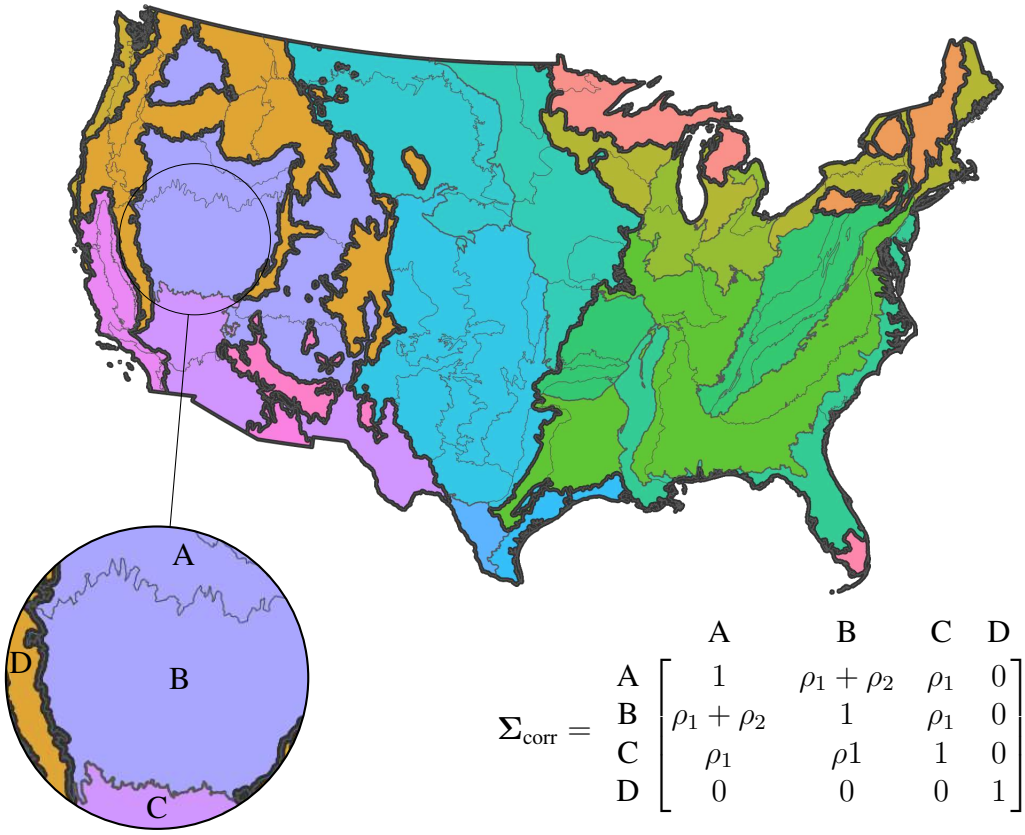


Figure 2.2: Hierarchical nesting of ecoregions in the contiguous United States. The EPA's L3 ecoregions are delineated by the thin light grey lines; L2 by the different colors; and L1 by the thick dark grey lines. A small section is magnified to illustrate the structure of the correlation matrix in the prior, as detailed in Section 2.4.1.

Fire indices may capture the relationships and interactions between meteorological variables better than the variables alone and therefore be better able to predict their subsequent effects on wildfire activity (Preisler et al., 2008; Abatzoglou and Kolden, 2013). Abatzoglou and Kolden (2013) demonstrated that, when compared to individual weather components like temperature and precipitation, ERC had a more direct link to fuel flammability and conditions habitable to fires burning large areas. This suggests that ERC may provide more valuable information than the four weather covariates we initially fit our model on, particularly regarding burned area size. However, some wildfire indices (FWI specifically) consistently under-represent the occurrence of large wildfires and their associated burned areas (Pimont et al., 2021).

ERC is a calculated fuel moisture index from the National Fire Danger Rating System, which considers the cumulative drying effect of temperature, precipitation, humidity, and solar radiation as an overall measure of wildfire fuel (Abatzoglou and Kolden, 2013). The ERC is a relative index, meaning that an ERC value of 30 represents a potential heat release that is twice that of an ERC value of 15. We use the daily ERC data which is derived by gridMET under a conifer forest model (Abatzoglou, 2013). We then aggregate these data to a monthly timestep.

FWI is a wildfire danger index that incorporates the effects of fuel moisture and weather conditions into the projected behavior of a wildfire. This index is often used as the main covariate in wildfire models (see Abatzoglou and Williams, 2016; Castel-Clavera et al., 2022; Pimont et al., 2021). We use the `R` package `cffdrs` (Wang et al., 2017) to calculate the FWI on a daily timestep using the meteorological raster data described in Section 2.3.2. We then aggregate these data to a monthly timestep.

2.4 Model

2.4.1 Ecoregion nesting

In building our models for wildfire counts and burned areas, we leverage the concept of ecoregions to borrow strength across space. Rather than using generic spatially autoregressive areal models, we observe that drivers of spatial correlation in wildfire behavior might include things like common vegetation types, common topographical features, and so forth. Furthermore, the ecoregions defined by EPA have an exploitable hierarchical structure, where the coarsest division is further divided into successively finer partitions, according to levels of ecological similarity. To account for and incorporate the hierarchical nature of the three levels of ecoregions, we build a 84×84 correlation matrix $\mathbf{\Sigma}_{\text{corr}}$, with each row and column corresponding to an L3 ecoregion. This matrix contains ones along the main diagonal (each L3 ecoregion is perfectly correlated with itself) and values between 0 and 1 elsewhere. Any two L3 ecoregions $\{i, j\}$ that are in the same L2 ecoregion will have the correlation coefficient ρ_2. This same pair of L3 ecoregions will also be in the same L1 ecoregion, therefore adding a correlation coefficient ρ_1, for a total correlation of

$\rho_1 + \rho_2$ assigned to this pair of L3 ecoregions. Any two L3 ecoregions $\{i, j\}$ that are in the same L1, but not the same L2 ecoregion, will have the correlation coefficient ρ_1. L3 ecoregion pairs that do not share any nesting will have a value of 0. See Figure 2.2 for an illustration of a subset of this matrix. We use this spatial nested structure throughout our model-building process.

2.4.2 Wildfire counts

We model wildfire counts as a zero-inflated (ZI) negative binomial (NB) random variables. This model defines the following likelihood for $n_{s,t}$, the number of wildfires in spatial unit s at time t, as

$$n_{s,t} \,|\, \lambda_{s,t}, \pi_{s,t}, \delta_s \sim \begin{cases} \pi_{s,t} + (1 - \pi_{s,t}) \times \text{NB}(0 \,|\, \lambda_{s,t}, \delta_s) & \text{if } n_{s,t} = 0 \\ (1 - \pi_{s,t}) \times \text{NB}(n_{s,t} \,|\, \lambda_{s,t}, \delta_s) & \text{if } n_{s,t} > 0, \end{cases}$$

where $s = 1, \ldots, S$ is the spatial unit (L3 ecoregion) and $t = 1, \ldots, T$ is the monthly timestep. Different levels of complexity are possible within this model. We tested versions of the model (see Section 2.5.1) where different parameters depended on covariates or not, where parameters varied spatially or not, and so forth. In the most general version, the rate parameter $\lambda_{s,t}$ and the probability of extra zeroes $\pi_{s,t}$ depend on spatially and temporally varying covariates, while the overdispersion parameter δ_s varies spatially according to the structured spatial prior described in Section 2.4.1. We also tried fixing $\delta_s = 0$ for all $s = 1, \ldots, S$, which yields a zero-inflated Poisson model. To ensure that $\delta_s, \lambda_{s,t} > 0$ and $\pi_{s,t} \in (0, 1)$, we modeled these parameters with log and logit links, respectively, as

$$\begin{aligned} \log(\boldsymbol{\lambda}_s) &= \mathbf{X}_s \boldsymbol{\beta}_s^{(\lambda)} + \boldsymbol{\Phi}_s^{(\lambda)} + a_s \\ \text{logit}(\boldsymbol{\pi}_s) &= \mathbf{X}_s \boldsymbol{\beta}_s^{(\pi)} + \boldsymbol{\Phi}_s^{(\pi)} \\ \log(\boldsymbol{\delta}) &\sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{corr}}^{(\delta)}), \end{aligned}$$

where $\boldsymbol{\lambda}_s$ and $\boldsymbol{\pi}_s$ are vectors of length T for L3 ecoregion s, $\boldsymbol{\delta}$ is a vector of length S, and a_s is a known offset equal to the area of the L3 ecoregion s. $\mathbf{X}_s$ is a $T \times (p+1)$ design matrix, $\boldsymbol{\beta}_s^{(\lambda)}$ and $\boldsymbol{\beta}_s^{(\pi)}$ are vectors of length $(p+1)$ with the priors discussed in Section 2.4.5, and $\boldsymbol{\Sigma}_{\text{corr}}^{(\delta)}$ is a structured correlation matrix described in Section 2.4.1. To account for residual structured variability, we include the spatio-temporal random effects $\boldsymbol{\Phi}^{(\lambda)}$ and $\boldsymbol{\Phi}^{(\pi)}$, which are $(T \times S)$ matrices, each composed of a spatial intrinsic conditional autoregressive component and a temporal autoregressive component, as

$$\begin{aligned} \boldsymbol{\phi}_{t=1}^{(\lambda)} &\sim \text{N}(\mathbf{0}, [\tau^{(\lambda)}(\mathbf{D} - \mathbf{W})]^{-1}) \\ \boldsymbol{\phi}_t^{(\lambda)} \,|\, \boldsymbol{\phi}_{t-1}^{(\lambda)} &\sim \text{N}(\eta^{(\lambda)}\boldsymbol{\phi}_{t-1}^{(\lambda)}, [\tau(\mathbf{D} - \mathbf{W})]^{-1}), \end{aligned} \tag{2.3}$$

and analogously for $\boldsymbol{\Phi}^{(\pi)}$, where $\boldsymbol{\phi}_t^{(\lambda)}$ is a row vector of length S. $\mathbf{D}$ is a fixed $(S \times S)$ diagonal matrix, with $D_{i,i}$ equal to the number of neighbors that region i has. The matrix $\mathbf{W}$ is a fixed $(S \times S)$ adjacency matrix, where $W_{i,j} = 1$ if regions i and j are neighbors and 0 otherwise. The parameters η and τ are given weakly informative priors

$$\begin{aligned} \eta^{(\lambda)} &\sim \text{Beta}(2, 8) \\ \tau^{(\lambda)} &\sim \text{Exp}(1). \end{aligned} \tag{2.4}$$

2.4.3 Wildfire burned area

We model the extreme and non-extreme wildfire burned areas simultaneously using the family of EGPD models defined in Section 2.2.1. Taking, for example, the carrier function $G_1(v) = v^\kappa$, yields the cdf $F_1(y)$ and density $f_1(y)$

$$\begin{aligned} F_1(y) &= \{1 - [1 + \xi(\frac{y}{\sigma})]^{-1/\xi}\}^\kappa \\ f_1(y) &= \frac{\kappa}{\sigma}[1 + \xi(\frac{y}{\sigma})]^{-(1/\xi+1)}\{1 - [1 + \xi(\frac{y}{\sigma})]^{-1/\xi}\}^{\kappa-1} \end{aligned} \tag{2.5}$$

Since all of the fires in our dataset are greater than 1,000 acres, we construct a truncated density from (2.5), as

$$\text{t-EGPD}_1(x) := \frac{f_1(x)}{1 - F_1(1)}, \tag{2.6}$$

where we normalize by $1 - F_1(1)$ because x has units of thousands of acres. (2.6) defines the following likelihood for y_{s_i,t_i}, the burned area of the ith wildfire event in spatial unit s at timestep t, where s and t are defined as in Section 2.4.2:

$$y_{s_i,t_i} \mid \kappa_{s_i,t_i}, \sigma_{s_i,t_i}, \xi_s \sim \text{t-EGPD}_1(y_{s_i,t_i} \mid \kappa_{s_i,t_i}, \sigma_{s_i,t_i}, \xi_s).$$

In most cases, $i = 1$, except when more than one wildfire occurs in the same spatial unit s in the same timestep t, therefore leading to multiple burned areas with otherwise identical s and t. Again, several levels of complexity are possible by allowing the EGPD parameters to depend on spatially and temporally varying covariates, to be spatially varying, to be constant across space, etc. For this example of the carrier function $G_1(\cdot)$, the most general model we considered allows κ, σ, and ξ to depend on spatially and temporally varying covariates, according to the same nested spatial prior described in Section 2.4.1. To ensure that all three parameters remained positive as required, we again use a log link, analogous to the links used in the count model (see Section 2.4.2), as

$$\begin{aligned}
\log(\boldsymbol{\kappa}_s) &= \mathbf{X}_s\boldsymbol{\beta}_s^{(\kappa)} + \boldsymbol{\Phi}_s^{(\kappa)} \\
\log(\boldsymbol{\sigma}_s) &= \mathbf{X}_s\boldsymbol{\beta}_s^{(\sigma)} + \boldsymbol{\Phi}_s^{(\sigma)} \\
\log(\boldsymbol{\xi}_s) &= \mathbf{X}_s\boldsymbol{\beta}_s^{(\xi)} + \boldsymbol{\Phi}_s^{(\xi)}
\end{aligned}$$

The priors on $\boldsymbol{\Phi}^{(\kappa)}$, $\boldsymbol{\Phi}^{(\sigma)}$, and $\boldsymbol{\Phi}^{(\xi)}$ are analogous to (2.3). The priors on the hyperparameters $\eta^{(\kappa)}$, $\eta^{(\sigma)}$, $\eta^{(\xi)}$, $\tau^{(\kappa)}$, $\tau^{(\sigma)}$, and $\tau^{(\xi)}$ are analogous to those in (2.4).

While every timepoint and spatial unit in the count model contains a value in the dataset (either with $n = 0$ or $n > 0$), this is not the case for the burned area model. To account for this, we

treat any spatial unit and timepoint pair that does not have a wildfire as "missing" and model these missing data as latent parameters within the overall Bayesian model.

2.4.4 Joint model

To share information between the wildfire counts and burned area sub-models, we link the two using a shared random effect. Coupling the two models has the potential to increase performance, but it moreover has interesting interpretive potential. For example, one conjecture that could be investigated using such a coupled model is that a region experiencing few fires may be more likely to experience very large fires. Similar coupled count and magnitude models for marked point processes have appeared recently (see Koh et al., 2023; Becker et al., 2022; Yadav et al., 2023).

We include a temporally varying shared random effect $\boldsymbol{\theta}$, a vector of length T, in the regression components of the model parameters as

$$\begin{aligned} \log(\boldsymbol{\lambda}_s) &= \mathbf{X}_s\boldsymbol{\beta}_s^{(\lambda)} + \boldsymbol{\Phi}_s^{(\lambda)} + a_s + \boldsymbol{\theta} \\ \log(\boldsymbol{\kappa}_s) &= \mathbf{X}_s\boldsymbol{\beta}_s^{(\kappa)} + \boldsymbol{\Phi}_s^{(\kappa)} + \gamma_s\boldsymbol{\theta}. \end{aligned} \tag{2.7}$$

We include a spatially-varying link parameter γ_s to spatially modulate the relationship between the rate of counts and the distribution of wildfire sizes. To accommodate temporal correlation, the shared parameter $\boldsymbol{\theta}$ has an AR(1) prior, while the modulating parameter $\boldsymbol{\gamma}$ is modeled as a spatial random effect, as

$$\begin{aligned} \theta_t \,|\, \theta_{t-1} &\sim \mathrm{N}(\eta^{(\theta)}\theta_{t-1}, \epsilon^2) \\ \boldsymbol{\gamma} &\sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{corr}}^{(\gamma)}), \end{aligned}$$

along with the weakly informative hyperprior $\epsilon \sim \mathrm{t}_4(0, 1)$.

2.4.5 Nonlinear effects

To capture nonlinear relationships between our covariates and wildfire counts and burned areas across the large spatial domain (Brillinger et al., 2003; Preisler et al., 2004; Preisler and Westerling, 2007; Woolford et al., 2011, 2021), we introduce cubic B-splines with five basis functions for each of our covariates. To smooth the relationship among the basis functions within each covariate, we included a Gaussian auto-regressive of order 1 (AR(1)) prior for each coefficient vector (Lang and Brezger, 2004). This prior induces a penalty that is the Bayesian analogue of the classical penalized-spline approach of Eilers and Marx (1996).

In addition, it is advantageous to borrow strength across space for the covariate effects, which we accomplish using the spatial nesting scheme described in Section 2.4.1. This results in a Normal prior for each parameter vector $\boldsymbol{\beta}$ in the models defined above, centered around zero, with a covariance matrix constructed as the Kronecker product of an AR(1) covariance matrix and a nested spatial correlation matrix, as

$$\boldsymbol{\beta}_s^{(\lambda)} \sim \mathrm{N}(\mathbf{0}, \Sigma_{\mathrm{corr}}^{(\lambda)} \otimes \Sigma_{\mathrm{AR(1)}}^{(\lambda)}), \tag{2.8}$$

with analogous specifications for $\boldsymbol{\beta}_s^{(\pi)}$, $\boldsymbol{\beta}_s^{(\kappa)}$, $\boldsymbol{\beta}_s^{(\sigma)}$, and $\boldsymbol{\beta}_s^{(\xi)}$.

2.4.6 Training and validation

We first assessed multiple candidate models for both the wildfire counts and wildfire burned areas, separately, on a dataset that included five climate covariates and housing density, to determine the best model for each component. Then, we determined the best collection of covariates (either the meteorological covariates or one or both of the wildfire indices) for each model component. Each model was trained on 21 years of data (1995–2015, inclusive) and tested on 10 years of data (1990–1994, 2016–2020), yielding 252 monthly timesteps in the training dataset and 120 monthly timesteps in the test dataset.

We estimate all parameters through Monte Carlo Markov Chain (MCMC) sampling in `Stan`, using the default NUTS sampler (Stan Development Team, 2023; Gabry et al., 2023). We run three chains for each model, with 2000 iterations per chain (1000 for warm-up, 1000 for sampling).

2.5 Results

2.5.1 Model comparison

We developed multiple candidate models for both the wildfire counts and the wildfire burned area models. We then compared the various models on the test dataset using log-likelihood scores (log scores) for the counts and the burned areas, along with threshold-weighted continuous ranked probability scores (twCRPS) for the burned area models. We first give a brief description of these scores, and then we present the scores for the various candidate models.

Log scores are proper scoring rules based on the model's predictive density (Gneiting and Raftery, 2007). We calculate the log score at every MCMC iteration for each observation in the test dataset, and then sum all of the scores within each iteration. We then compare the median log scores across the iterations among each model. The CRPS (Matheson and Winkler, 1976) is another proper scoring rule, defined in terms of the probabilistic forecast (as opposed to the predictive density), and is sensitive to the distance between predictions and the true realizations. With $F(z)$ the predictive cdf associated with the probabilistic forecast and y the realization, the CRPS is defined as:

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(z) - \mathbf{1}\{z \geq y\})^2 \, dz. \tag{2.9}$$

Gneiting and Ranjan (2011) propose the twCRPS, which incorporates a weighting function to emphasize a specific region of interest while still being a proper scoring rule. With the same notation as (2.9) and weighting function $w(z)$, the twCRPS is defined as:

$$\text{twCRPS}(F, y) = \int_{-\infty}^{\infty} (F(z) - \mathbf{1}\{z \geq y\})^2 w(z) \, dz. \tag{2.10}$$

Since we are most interested in accurately capturing the upper tails of wildfire burned areas while still maintaining good predictive accuracy in the bulk of the burned area distribution, we use a Normal cdf as our weight function, with a mean of 21 thousand acres and standard deviation of nine thousand acres. This weight function is centered around the 90th quantile of the burned area sizes, thereby emphasizing fires above the 90th quantile while still giving some weight to fires below the 90th quantile. We denote this weight function as $\Phi_{21,9}(z)$. We use Monte Carlo approximation to evaluate the twCRPS through the discretization

$$\begin{aligned} \text{twCRPS}(F, y) &\approx \frac{\max\{y\} - \min\{y\}}{I - 1} \sum_{i=1}^{I} \Phi_{21,9}(z_i)(F(z_i) - \mathbf{1}\{z_i \geq y\})^2, \text{ where} \\ z_i &= \min\{y\} + i\frac{\max\{y\} - \min\{y\}}{I} \end{aligned} \tag{2.11}$$

(Gneiting and Ranjan, 2011). We generate I=5,000 equally spaced points z_i at which we evaluate the predictive cdf, $F(z_i)$, corresponding to the predictive density in (2.6). We perform this approximation for each burned area, y, in the test dataset for every MCMC iteration. The score for each iteration is the mean of the scores calculated for each y. We then compare the means across iterations for each model, with values closer to zero indicating better performance.

Ideally, we would compare all joint combinations of the counts and burned area sub-models in a full factorial design. Unfortunately, doing so would be computationally prohibitive due to the large number of potential models under consideration. Therefore, we used a multi-stage approach to select the best-performing sub-models before creating the joint model. We developed six candidate models for wildfire counts. We fit a ZINB model with overdispersion parameter δ held constant throughout space, as well as a version where δ is spatially-varying according to the nesting structure in Section 2.4.1. We also fit a zero-inflated Poisson (ZIP) model, which fixes $\delta = 0$. For these three models, we allowed the probability of inflated zeroes π to either be spatially-varying or have a logit regression link with the covariates, yielding six candidate models in total. We did not consider parametrizations of these models with a regression link on δ or π out of concern for overparametrizing the models. Table 2.1 shows the median difference of each model's scores,

along with the standard deviation (SD) of that difference, when compared with the best performing model, which was the ZINB model with spatially-varying δ parameter and spatially-varying π.

Table 2.1: Comparison of log scores on six candidate wildfire counts models. A '✓' in the 'Regression' column indicates that the respective parameter had a regression link with the covariates. An 's' indicates the parameter was spatially-varying according to the nesting structure described in Section 2.4.1, with no regression link. A 'c' indicates the parameter was spatially and temporally constant. An empty space indicates that the parameter was not involved in that particular model.

Model	Regression			Difference	
	λ	π	δ	Median	SD
ZI-Negative Binomial	✓	s	s	0.00	0.00
	✓	✓	s	-45.93	274.42
	✓	s	c	-502.91	314.63
	✓	✓	c	-544.13	307.83
ZI-Poisson	✓	✓		-1140.01	320.45
	✓	s		-1145.58	349.93

We then ran two additional versions of our best performing model by replacing the five climate covariates with either the ERC or the FWI (see Section 2.3.4). The best performing set of covariates was the original set with all five climate covariates (see Table 2.2).

Table 2.2: Comparison of log scores from the best count model, run on three different sets of covariates.

	Difference	
Dataset	Median	SD
Climate	0.00	0.00
ERC-FWI	-617.69	348.48
ERC	-738.98	342.80
FWI	-1030.65	395.42

We compare burned area models composed of the four EGPD carrier functions listed in Section 2.2.1, along with a lognormal model. We include the lognormal model, despite its inflexibility with respect to tail characteristics, because it was found to be the best-performing model in Joseph

et al. (2019). With the five model families (four EGPD carrier functions and lognormal), various permutations of which parameter(s) contains a regression link versus a nested spatial structure, and four different sets of covariates, we have 68 candidate models for wildfire burned area. To pare down these potential candidate models, we locally optimized in three stages. We first determined which parameter(s) would contain a regression link by running the four parameter permutations with carrier function G_1 on the climate covariate dataset. Table 2.3 shows the mean and median difference, along with the SD of that difference, of each permutation's twCRPS and log scores, respectively, on the test dataset. Despite the large SD, relative to the median and mean differences of the scores, we chose to proceed with the most parsimonious and easily interpretable model, which is the model with regression on κ only.

Table 2.3: Comparison of twCRPS and log scores for four parameter combinations of the G_1 EGPD model. Scores shown for twCRPS are multiplied by a factor of 10^2 to more easily see the magnitude of differences. Symbols are analogous to those in Table 2.1.

	Regression			twCRPS		Log score	
Model	κ	σ	ξ	Mean	SD	Median	SD
G_1	✓	s	s	0.00	0.00	0.00	0.00
	s	✓	s	5.39	14.03	-80.52	935.30
	✓	✓	s	12.45	17.96	-199.66	971.43
	✓	✓	✓	21.84	22.69	-432.41	1144.93

We then ran this version of the G_1 model on three additional sets of covariates to determine the best combination. Table 2.4 shows that while all four sets of covariates perform similarly on the test dataset, the combination of ERC-FWI outperforms them all (although the differences have very large standard errors) and lends itself nicely to interpretation by capturing different aspects of wildfire behavior.

Finally, we assessed the best model family from among the EGPD models with different carrier functions and the lognormal model by running the analogous versions of each EGPD carrier function and the lognormal model. We show the performance of all five models on the "best" dataset in Table 2.5. The additional 'Time' column shows that while G_3 and G_4 may perform similarly

Table 2.4: Comparison of twCRPS and log scores from the best G_1 parametrization, run on four different combinations of covariates.

	twCRPS		Log score	
Dataset	Mean	SD	Median	SD
ERC-FWI	0.00	0.00	0.00	0.00
ERC	0.31	9.45	-8.40	90.86
FWI	0.74	9.28	-11.18	80.96
Climate	1.92	11.78	-34.33	934.27

to G_1, they take approximately 13 and 19 times longer, respectively, than G_1 to complete 1,000 iterations in the sampling phase. We also note that some of the MCMC chains in G_3 and G_4 did not mix well, and we did not encounter this issue with G_1, G_2, and the lognormal models. We therefore proceed with G_1 for the joint model.

Table 2.5: Comparison of twCRPS and log scores for one-parameter versions of five families of burn models run on the ERC-FWI dataset. Symbols and scale of scores are analogous to those in Table 2.4. The 'Time' column indicates how long the model took to complete 1000 MCMC iterations in the sampling phase (i.e. excluding warmup phase) using `Stan`.

	Regression			twCRPS		Log score		
Model	$\kappa(\mu)$	σ	ξ	Mean	SD	Median	SD	Time (hrs)
G_1	✓	s	s	0.00	0.00	0.00	0.00	6.90
G_3		✓	s	1.65	9.05	-64.56	80.44	89.21
G_4	✓	s	s	3.19	10.85	-16.70	100.07	129.09
G_2	✓	s	s	8.17	12.76	-39.83	70.60	11.55
Lognormal	(✓)	s		2.45×10^3	9.64×10^2	-2.74×10^3	8.17×10^3	5.23

2.5.2 Model checking

With the best model and covariates in hand, we turn to assessing model adequacy for capturing key features of the wildfire data. To assess overall model performance, we combine the best-performing wildfire counts model with the best-performing burned area model using the shared random effects scheme in (2.7), and compare the predictive distribution to the observed wildfire counts and burned areas.

To assess the predictive ability of the wildfire counts sub-model, we first calculate the expected number of wildfires at each timepoint t for each L3 ecoregion s:

$$E[n_{s,t} \mid \pi_s, \lambda_{s,t}] = (1 - \pi_s) \times (\lambda_{s,t}). \tag{2.12}$$

We do this calculation for every iteration of each chain, for a total of 3,000 values of $E[n_{s,t} \mid \pi_s, \lambda_{s,t}]$, for each $t = 1, \ldots, T$ and $s = 1, \ldots, S$. For ease of visual representation, we then aggregate these values by summing over each year and across L1 ecoregion. Figure 2.3 displays the boxplots of these predictions, along with the observed wildfire counts.

We next look at the predictive ability of the wildfire burned areas sub-model. We estimate the expected burned area at each timepoint t for each L3 ecoregion s through the following procedure. At each MCMC iteration, given the current parameter values,

1. Randomly draw $n_{s,t}$ from the ZINB model.

2. If $n_{s,t} = 0$, then let the expected burned area $E[y_{s,t} \mid \kappa_{s,t}, \sigma_s, \xi_s] = 0$. If $n_{s,t} > 0$, randomly draw $n_{s,t}$ observations from the EGPD model and sum these observations to get $w_{s_{i=1},t_{i=1}}$, the total expected burned area at time t for region s for the first repetition.

3. Repeat steps one and two for $i = 2, \ldots, 500$ times for each unique pair of $\{s, t\}$. Then take the average of the $i = 1, \ldots, 500$ values w_{s_i,t_i} to yield an estimate of $E[y_{s,t} \mid \kappa_{s,t}, \sigma_s, \xi_s]$, the expected burned area at timepoint t in L3 ecoregion s.

We repeat this procedure for each chain and every iteration, for a total of 3,000 values of $E[y_{s,t} \mid \kappa_{s,t}, \sigma_s, \xi_s]$, for each t and s. We then aggregate the these values by year and across L1 ecoregion. Figure 2.4 displays the boxplots of these predictions, along with the observed total burned areas. While our model captures temporal trends in total burned areas well, it overestimates total burned areas in some of the L1 ecoregions. We have investigated this issue and conjecture that the tail parameter ξ may be poorly estimated due to a lack of flexibility in the EGPD model. We thought we may have overestimated ξ due to the standard lognormal prior shrinking towards a

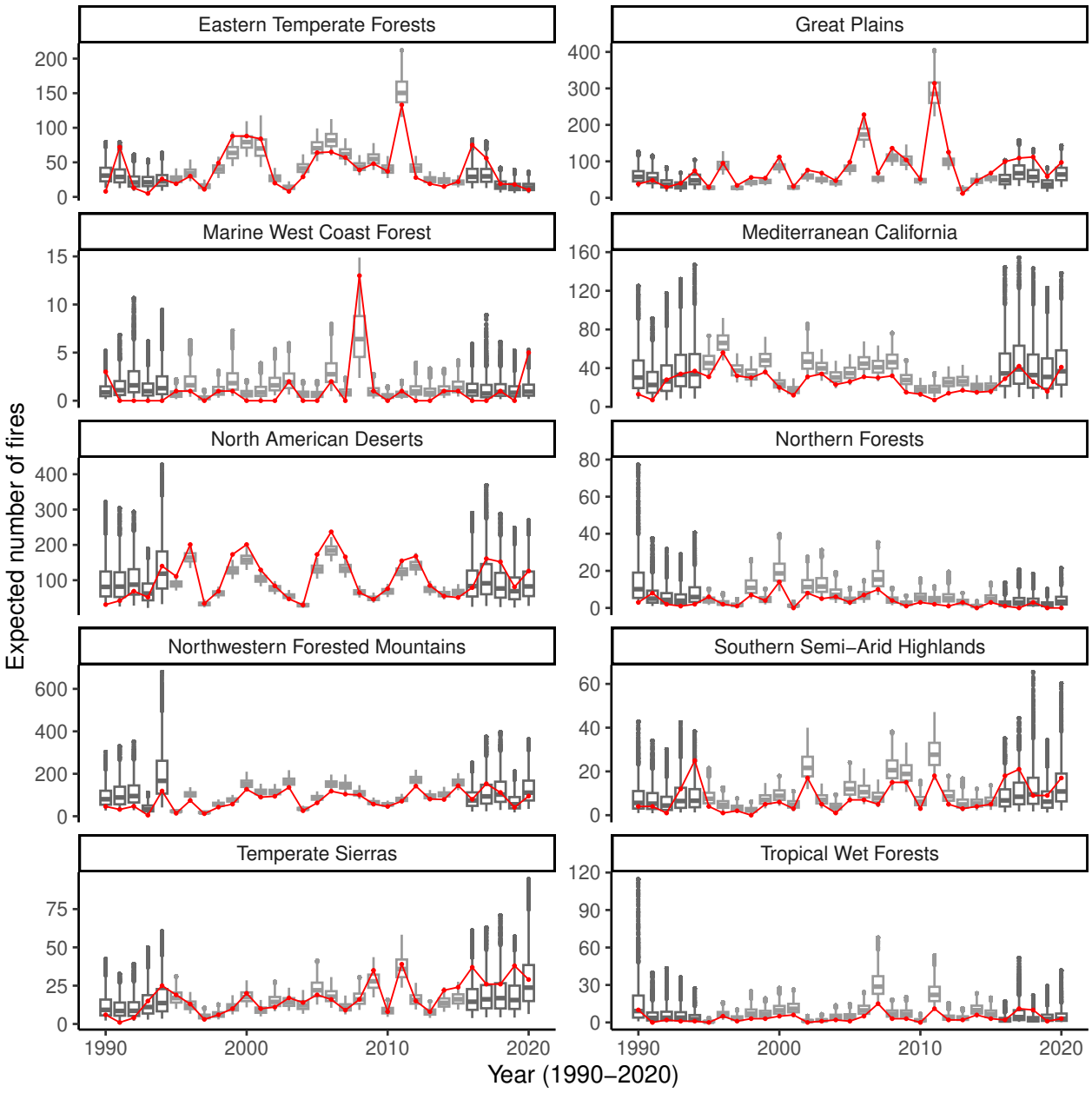


Figure 2.3: Expected wildfire counts compared with the truth. The red points represent the observed wildfire counts in a particular year for each L1 ecoregion. The dark grey boxplots indicate the test dataset years (1990–1994 and 2016–2020). The values shown are after applying a 95% winsorization to exclude extreme outliers and better visualize the trends across time and by region. Note the different scalings of the y-axes.

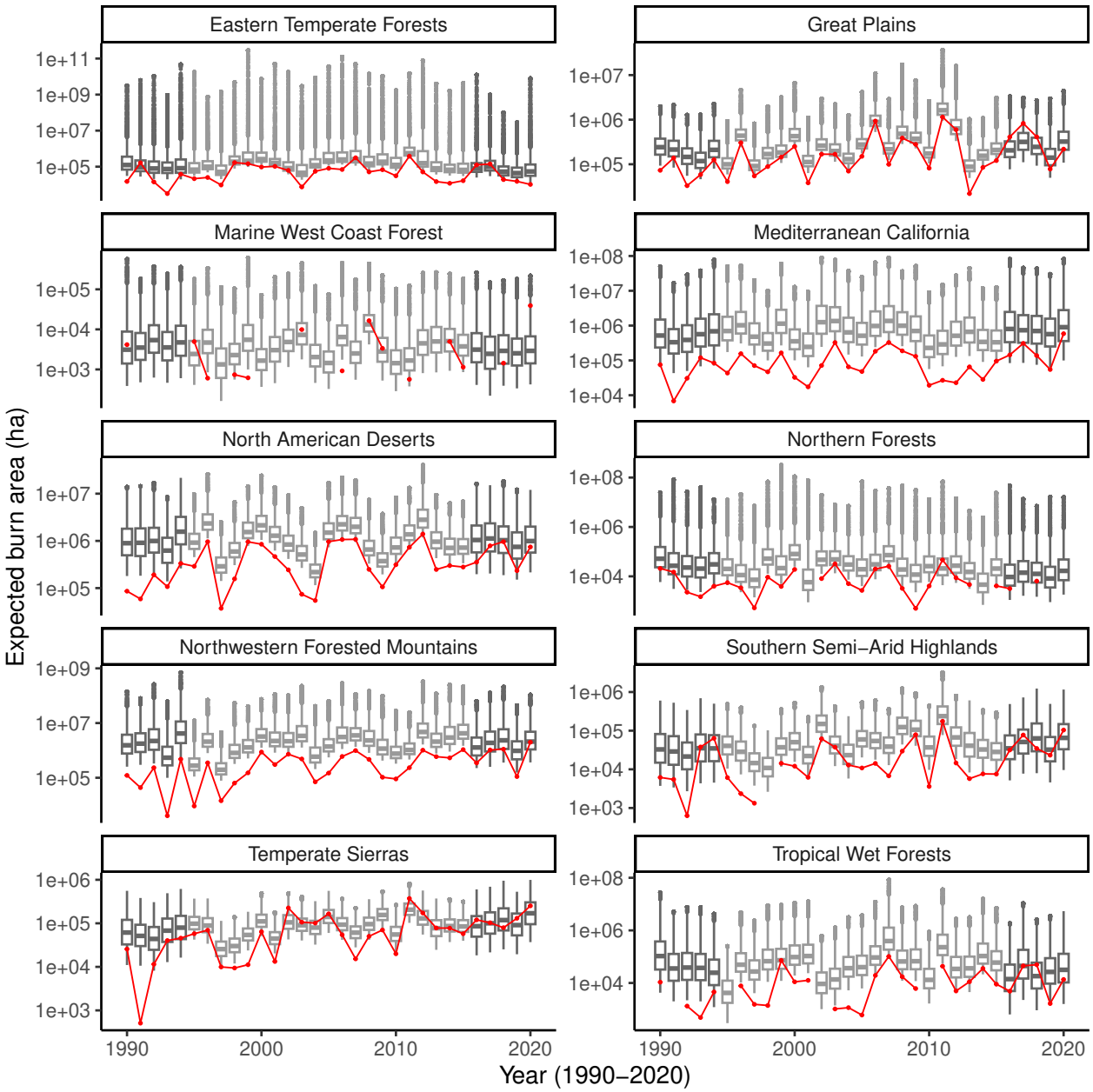


Figure 2.4: Predicted expected total wildfire burned area compared with the observed wildfire burned areas. Analogous to Figure 2.3, the red points are the observed wildfire burned area and the dark grey boxplots indicate the test dataset years. The values shown are on the log scale and are post 95% winsorization.

value of one. To investigate this, we tried placing a logit-normal prior on ξ with support $(0, 1.5)$, which yielded indistinguishable results. This remains a topic for further follow-up.

2.5.3 Covariate partial effects

We plot the covariate partial effects for both the wildfire count and burned area components of the joint model to determine which covariates have the strongest effect in the parameter with the regression link. The colors in each partial effects plot correspond to unique L2 ecoregions (see Figure 2.2, which has the same coloring scheme).

The two covariates that exhibited the greatest effect on the rate parameter λ within the negative binomial component of the joint model were minimum relative humidity and maximum air temperature (Figure 2.5). Wind speed also has a prominent effect in the Marine West Coast (L1 and L2 ecoregion of the same name). Interestingly, precipitation in the same month has a strong negative relationship with wildfire counts in this same region. However, the Marine West Coast did have many years with zero fires, so the lack of data may be contributing to these perceived effects. Precipitation in the same month has a positive relationship with wildfire counts in the Everglades (L2 ecoregion within the Tropical Wet Forests L1 ecoregion), which potentially reflects ignitions due to lightning strikes, which are common in this region.

The ERC partial effects plot for κ suggests that the importance of ERC in wildfire burned area begins at values of around 20 or above (Figure 2.6). Recall from Section 2.3.4 that the ERC is a relative index, so an ERC value of 20 indicates a heat potential double that of 10, similarly for an ERC value of 60 compared to ERC of 30. In the South Central Semi-Arid Prairies (part of the Great Plains L1 ecoregion), the Western Cordillera (part of the Northwestern Forested Mountains), and the Mediterranean California, ERC has a strong prominent effect on wildfire burned area. In contrast, the FWI partial effects on burned area are minimal in these three ecoregions, suggesting that heat potential is a more important driver of larger burned areas than wildfire weather. In the Cold and Warm Deserts (both part of the North American Deserts L1 ecoregion), FWI has a strong effect on wildfire burned area, while the ERC partial effects are much smaller in these

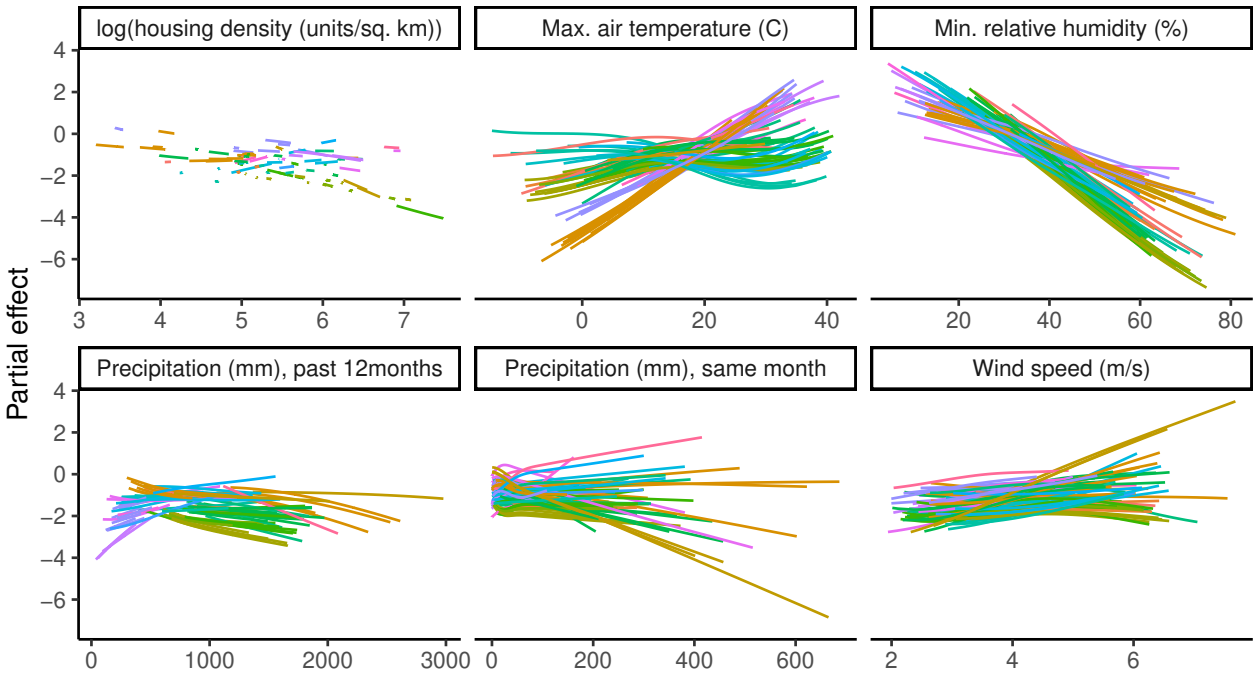


Figure 2.5: Partial effects on log-transformed rate parameter λ of negative binomial piece of overall model. These partial effects are solely the effects from the linear and basis functions of the covariates. These plots do not incorporate the spatio-temporal ICAR parameter $\mathbf{\Phi}$, the shared random effect θ_t, nor the area offset. Lines shown are posterior medians for each L3 ecoregion, colored by L2 ecoregion.

regions. This suggests that wildfire weather is a more important determinant of burned area in the North American Deserts than the heat potential of a wildfire. In the Marine West Coast and in the Northern Forests (corresponding to the two L2 ecoregions Atlantic Highlands and Boreal Plain), changes in ERC and FWI have minimal effects on the κ parameter. Again, this is likely due to lower numbers of wildfires in these ecoregions.

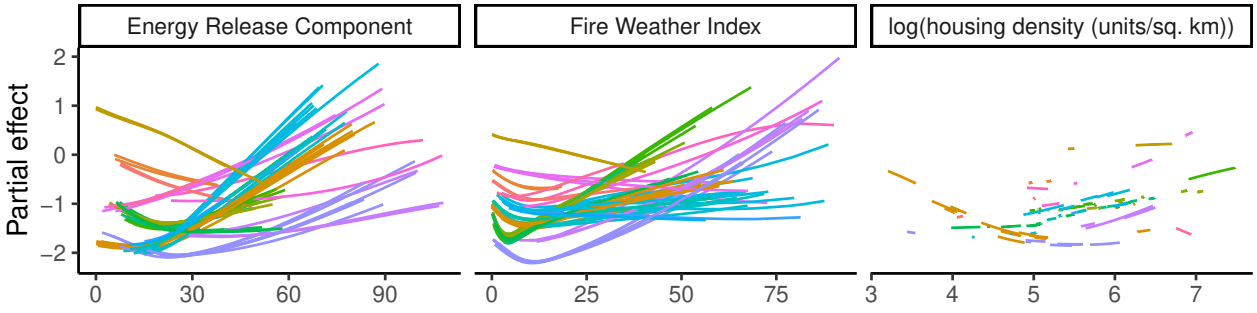


Figure 2.6: Partial effects on log-transformed parameter κ of the EGPD piece of overall model. Lines shown are posterior medians for each L3 ecoregion, colored by L2 ecoregion.

2.5.4 Burned area return levels and high quantiles

We now turn to investigating predicted high quantiles of burned areas in L1 ecoregions in two ways. These plots are colored by L1 ecoregion, matching the map of the US in Figure 2.7.

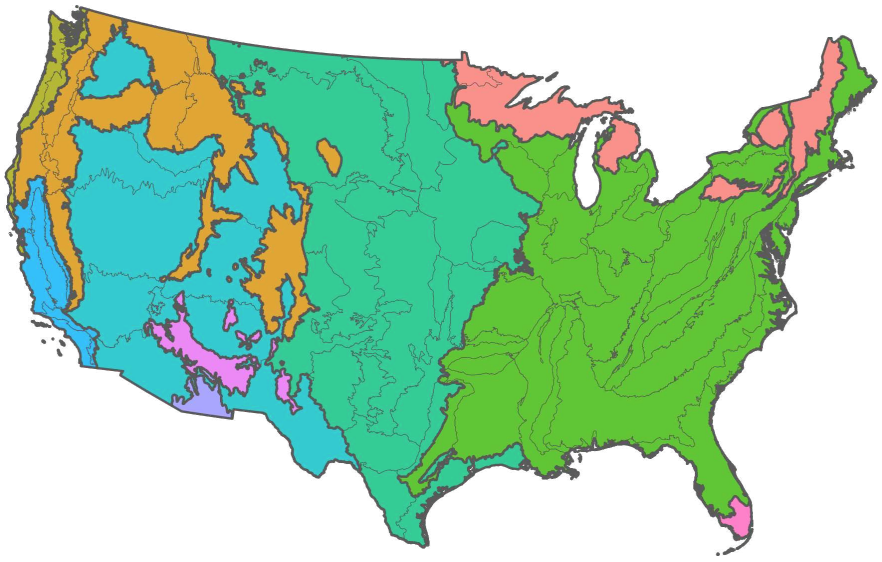


Figure 2.7: Map of the contiguous United States colored by L1 ecoregion.

First, we look at the 50-year return levels of wildfire burned areas by L1 ecoregion. We determine these burned area levels u through the following equation:

$$F_{s,t}^{-1}\{q^{1/(12\eta_{s,t})}\} = u_{s,t}, \tag{2.13}$$

where $F_{s,t}^{-1}$ is the quantile function corresponding to the predictive density of (2.6) for L3 ecoregion s at time t. The variable $\eta_{s,t}$ is the expected number of wildfires, as calculated from (2.12). We set $q = \frac{49}{50}$ to correspond to the 50-year return levels. From Figure 2.8, we see that Mediterranean California, North American Deserts, and Northwestern Forested Mountains have the most extreme wildfire burned areas. These predictions align with what we have experienced in the Western US over the past decade.

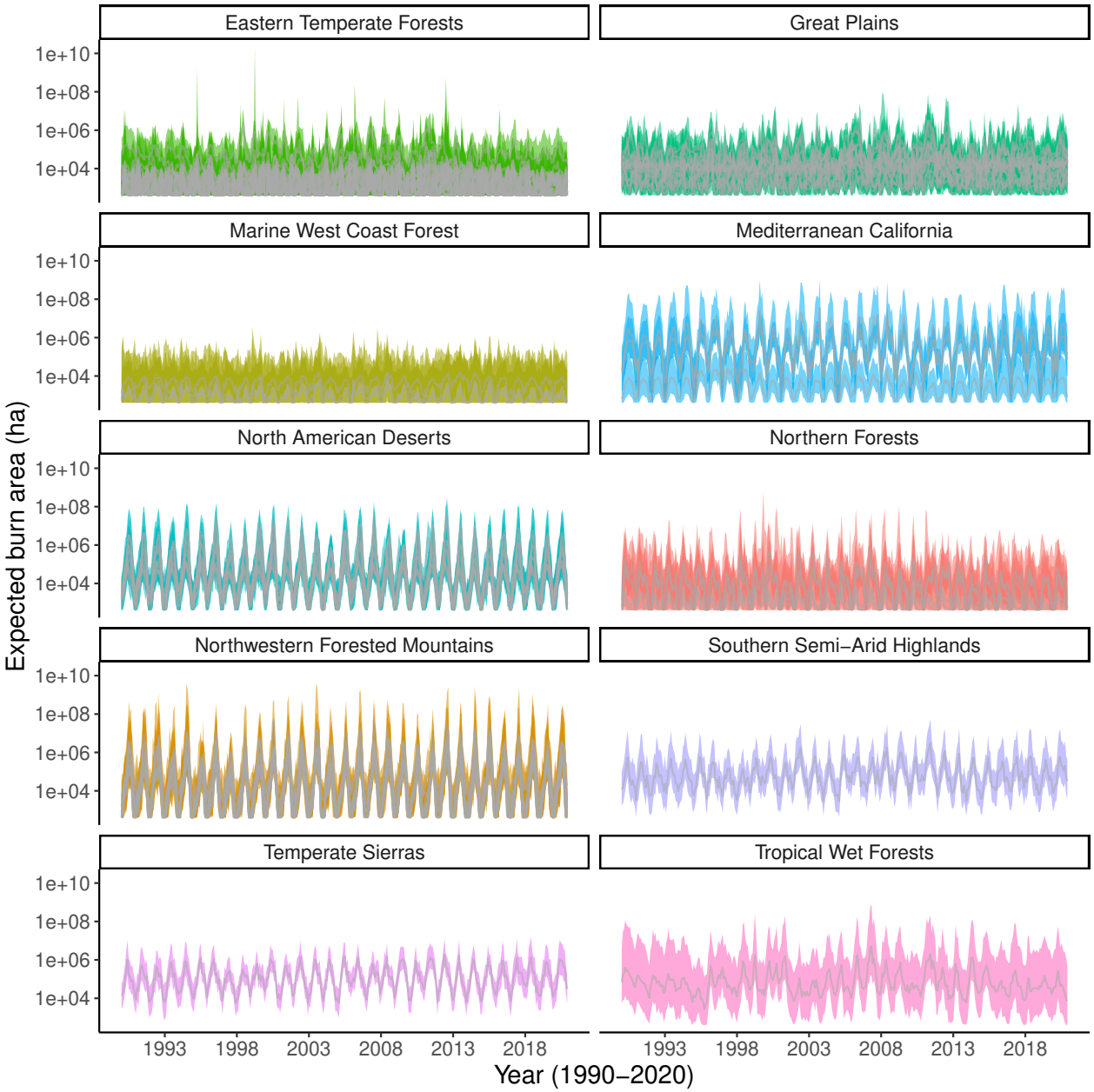


Figure 2.8: 50-year return levels for wildfire burned areas by L1 ecoregion. Each grey line represents an L3 ecoregion and the colored ribbons represent that 95% credible interval for each line. The values shown are on the log scale.

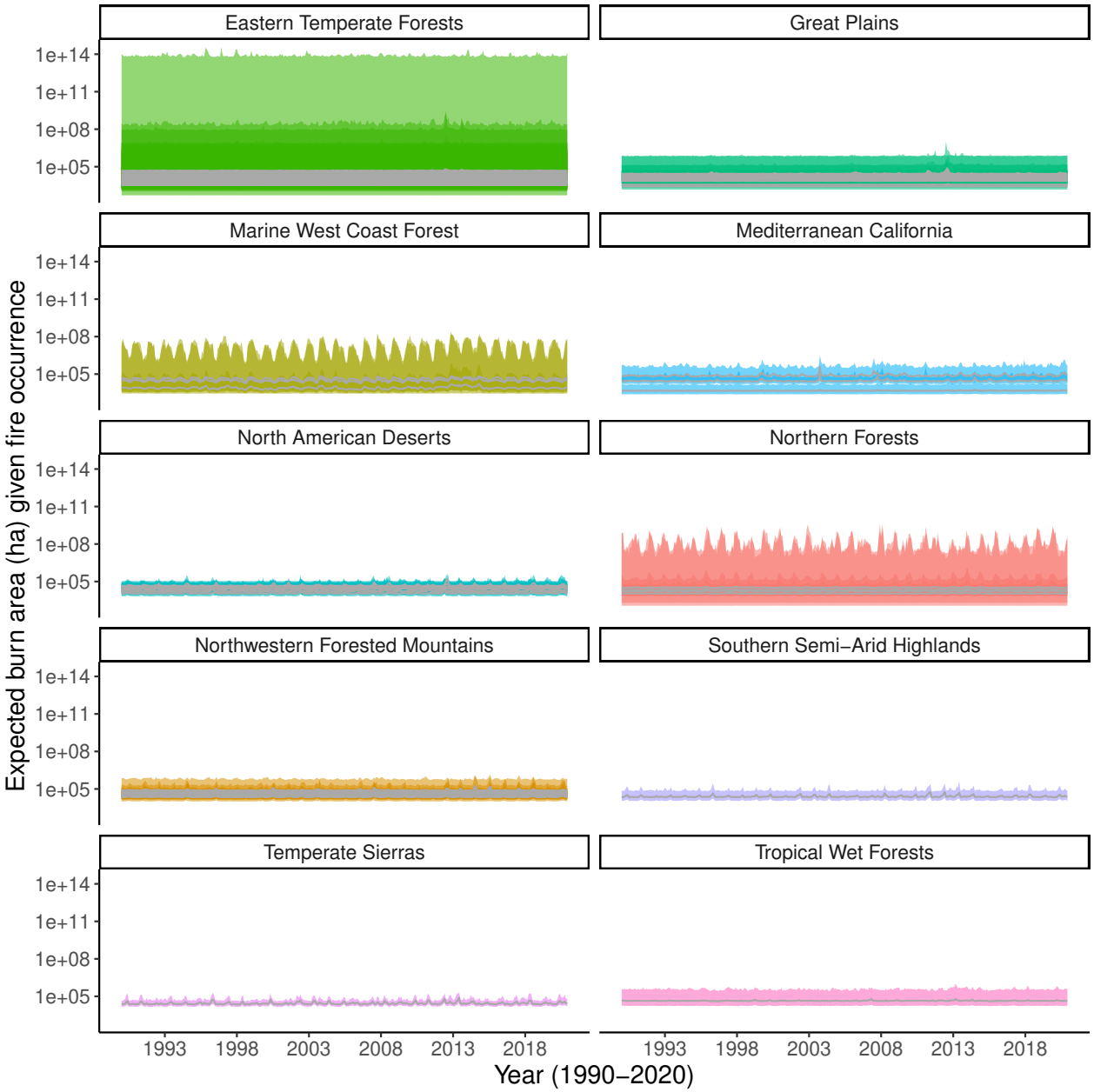


Figure 2.9: 98th quantile of wildfire burned areas by L1 ecoregion. The lines and ribbons are analogous to those in Figure 2.8.

We then look at the 98th quantile of wildfire burned areas by L1 ecoregions. We calculated this by setting $q = \frac{49}{50}$ in the following equation:

$$F_{s,t}^{-1}\{q\} = u_{s,t}. \tag{2.14}$$

While (2.14) describes the 98th quantile of the distribution of wildfire burned area for any individual wildfire, given that such a wildfire occurs, (2.13) describes the 98th quantile of the distribution of total burned areas, marginalized over the distribution of wildfire counts. From Figure 2.9, we see that Marine West Coast Forest and Eastern Temperate Forests have the largest burned areas at the 98th quantile, but their credible intervals are also quite wide. Northwestern Forested Mountains and North American deserts have the least variability in their predictions.

2.5.5 Maps of posterior parameter estimates

We compare the univariate GPD estimates of ξ with the posterior medians of ξ from our model (see Figure 2.10). Estimates of ξ from our model are larger in the Western US than in the Eastern US, which is consistent with the proliferation of high-impact wildfire events in the West. In addition, the spatial borrowing of strength allows us to obtain estimates of ξ in areas where there were not enough observations to fit a GPD in the exploratory stage.

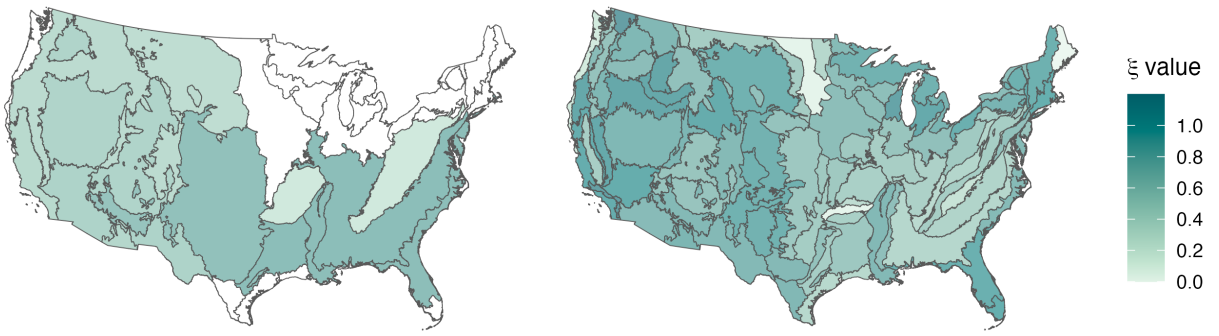


Figure 2.10: Posterior median of ξ by L3 ecoregion (right) compared to estimates ξ obtained from the MLE using the GPD by L2 ecoregion, with a the threshold set at the 90th percentile (left). The map on the left is the same as in Figure 2.1 but with an adjusted scale to be comparable to the posterior values of ξ on the right.

Table 2.6: Interpretation of the spatially varying values of the scaling parameter γ combined with the temporally varying values of the shared random effect θ.

	$\theta_t < 0$	$\theta_t > 0$
$\gamma_s < 0$	fewer counts, larger burned area	more counts, smaller burned area
$\gamma_s > 0$	fewer counts, smaller burned area	more counts, larger burned area

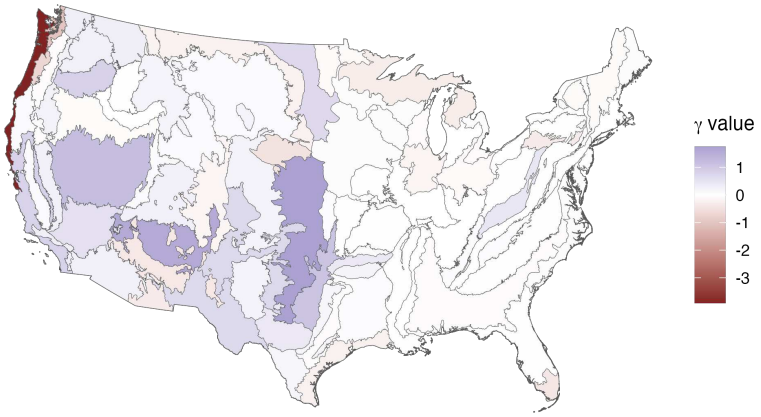


Figure 2.11: Posterior medians of γ across the United States by L3 ecoregion. Red regions have a negative γ value, indicating an inverse relationship between counts and burned areas. Blue regions have a positive γ value, indicating that counts and burned areas have a positive relationship. Darker shades of these two colors represent a stronger relationship between the count and burn parameters.

Figure 2.11 visualizes the spatially-varying modulating parameter γ of the joint model across the US by L3 ecoregion. Marine West Coast Forest ecoregions have a strong negative correlation between counts and burned areas, so the relationship between wildfire counts and burned area will follow the patterns in the first row of Table 2.6, while ecoregions in the Great Plains have a strong positive correlation (second row of the same table).

2.6 Discussion

We have implemented a joint Bayesian hierarchical model featuring a truncated EGPD sub-model, incorporating nonlinear effects in the covariates and accounting for any residual spatio-temporal variation with a Gaussian random effect. The flexibility of the model allows us to

incorporate all of the wildfire burned area data in the estimation of our spatially-varying shape and scale parameters, ξ and σ. The EGPD allows us to model the moderate and extreme wildfire burned areas, while allowing us to exploit EVT in extrapolating burned areas at high quantiles.

Our investigation into three datasets revealed a distinction in the covariates influencing wildfire occurrences compared to those affecting burned sizes. Notably, individual climate covariates excel in predicting wildfire counts, while fire indices, specifically ERC and FWI, outperform in predicting burned areas. This distinction underscores the nuanced nature of wildfire behavior and necessitates a tailored approach for accurate prediction. In our burned areas sub-model, the combined use of ERC and FWI surpasses individual fire indices, revealing regional variations in their effects. ERC significantly influences burned areas in the Great Plains, the Northwestern Forested Mountains, and the Mediterranean California, while FWI emerges as more crucial in the North American Deserts. These regional differences highlight the need for a nuanced understanding of fire indices in different ecoregions.

While our model adeptly captures temporal trends of wildfire burned areas, potential overestimation in certain ecoregions warrants attention. We suspect that the tail parameter ξ may be poorly estimated due to the EGPD model structure, suggesting a need for further exploration. Additionally, variations in vegetation and forestry across the CONUS raise concerns about potential overestimation stemming from the use of a coniferous forest fuel model in the calculation of ERC. Future research should consider obtaining ERC values calculated from diverse fuel models that best represent each ecoregion.

As the threat of more extensive wildfires looms with changing climate patterns, our study contributes vital insights for improved wildfire management within the CONUS. Recognizing the nonlinear drivers of wildfire occurrences and burned areas is essential for developing effective mitigation strategies and resource allocation policies. Our multifaceted approach, considering individual climate covariates, fire weather risk, and fire burn potential, provides an improved understanding of the drivers of wildfire dynamics.

Chapter 3

Bayesian model averaging of risk set probabilities using a geometric representation of multivariate extremes

3.1 Introduction

Predicting multivariate risk set probabilities requires understanding the shape and behavior of the joint tail. Classical multivariate extreme value theory relies on frameworks like multivariate regular variation (MRV) and max-stable models, both of which are limited to use in asymptotically dependent (AD) models (Coles et al., 1999), as they do not apply in asymptotically independent (AI) cases. In environmental applications, as we consider here, statisticians have increasingly argued against the use of max-stable models (Huser et al., 2025), in favor of more flexible models.

More recent multivariate extreme value modeling has emphasized working in light-tailed margins—as opposed to heavy-tailed in the MRV framework—allowing for more nuance in modeling the joint tail behavior and the ability to capture characteristics of both AD and AI classes. Most of this work is grounded in the limit set representation of multivariate extremes (Davis et al., 1988; Kinoshita and Resnick, 1991; Balkema et al., 2010; Balkema and Nolde, 2010, 2012), exploiting a radial-angular decomposition of the data. Other approaches, which are not based in limit set theory, also exploit this decomposition in light-tailed margins to develop semi-parametric angular-radial models (Mackay and Jonathan, 2024; Murphy-Barltrop et al., 2024). We restrict our attention to the limit set representation.

Building on earlier work of Nolde (2014) and Simpson et al. (2021), Nolde and Wadsworth (2022) leveraged the limiting shape of the point cloud to provide theoretical connections to key concepts in multivariate extremes, including tail dependence measures, multivariate regular variation (MRV), and conditional extremes. Wadsworth and Campbell (2024) introduced a parametric

framework for statistical inference and risk set probability estimation using limit sets, proposing a Gamma approximation of the radial distribution in polar coordinates. Semi-parametric methods followed, with some exploiting the aforementioned Gamma approximation (Majumder et al., 2024; Campbell and Wadsworth, 2024) and others leveraging the generalized Pareto distribution (Simpson and Tawn, 2024a; Papastathopoulos et al., 2025).

Majumder et al. (2024) and Simpson and Tawn (2024a) focus solely on estimating the shape of the limiting point cloud, while others pursue predictive inference approaches that rely on model selection procedures, introducing potential bias and model uncertainty (Wadsworth and Campbell, 2024; Campbell and Wadsworth, 2024; Papastathopoulos et al., 2025). The latter three propose methods of estimating risk set probabilities, either through a bespoke importance sampling scheme (Wadsworth and Campbell, 2024; Campbell and Wadsworth, 2024) or through other numerical and Monte Carlo integration (Papastathopoulos et al., 2025). Notably, all of these approaches discard observations below some suitably high threshold and make inference using truncated likelihoods. With small datasets, as is often the case in applied settings, incorporating data below the threshold may improve model fitting and subsequent prediction and inference.

In this work, we also use a radial-angular decomposition of the multivariate point cloud, necessitating a model for the angles and a model for the radii given the angles. We develop a method that maintains the simplicity of the parametric framework of Wadsworth and Campbell (2024), while adding flexibility through Bayesian model averaging (BMA).

The radial-angular decomposition yields a representation of the joint density in terms of two independent components: an angular component and a radial component given the angles. Wadsworth and Campbell (2024) show that the distribution of radial exceedances, conditioned on the angle, can be approximated by a truncated Gamma distribution. The rate parameter of the Gamma is connected to the limit set, which we discuss further in Section 3.2. To avoid introducing potential bias by including non-extremal data points, the authors fit large radii using a likelihood that consists of the right tail of a Gamma distribution, left-truncated at a high quantile. However, as angles corresponding to data below the threshold may be useful in informing the shape of the

multivariate point cloud, we introduce an alternative likelihood that uses all of the angles and assumes the corresponding radii come from a Gamma distribution, which is left-censored at a high threshold. We conjecture that this censored likelihood (Ch. 9 of Beirlant, 2004) makes more efficient use of the data, which is particularly important with smaller sample sizes.

We leverage several parametric characterizations of limit sets within the Gamma approximation to fit a suite of parametric models for the radii conditional on the angles. We use BMA to avoid selecting any one model and provide an optimal predictor (Madigan and Raftery, 1994). BMA circumvents many limitations that arise when using standard model selection procedures. With standard modeling procedures, one selects a model based on a metric like AIC, potentially choosing the "wrong" model. These procedures also overlook the uncertainty inherent in model selection, leading to overconfidence in our predictions (Madigan and Raftery, 1994). BMA optimizes over the posterior likelihoods of all models under consideration, and appropriately weights each model to maximize the combined likelihood. The resulting weights can then be used similarly to the weights of a mixture model, but without the potential mixing or identifiability issues that may arise in standard mixture models.

The use of BMA in multivariate extremes remains sparse; Apputhurai and Stephenson (2011) combine various AD and AI models through a weighted version of the summary measure ψ, Sabourin et al. (2013) average over the spectral measure of two AD models, and Vettori et al. (2020) average over various depths of the nested logistic model (Tawn, 1990). Our implementation differs from these previous works in that we apply BMA to limit set modeling to improve the prediction of extremely small risk set probabilities.

The remainder of this chapter is arranged as follows. Section 3.2.1 provides a brief overview of limit sets and how we can use them to perform statistical inference. Section 3.3 develops our modeling approach of the radial and angular components. Section 3.4 details our importance sampling scheme and provides a brief overview of BMA and how we incorporate it in our predictions of risk set probabilities. We demonstrate the performance of our approach through a comprehensive simulation study in Section 3.5. We apply our method to fire weather indices and discuss various

diagnostics in Section 3.6. And finally, we discuss the utility of our methods and proposed avenues of future research in Section 3.7.

3.2 Statistical inference

3.2.1 Limiting shape of the point cloud

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be n independent random vectors in $\mathbb{R}^d$ with exponential margins. If we consider the scaled point cloud, $N_n = \{\mathbf{X}_1/\log(n), \ldots, \mathbf{X}_n/\log(n)\}$, created by scaling these random vectors by $\log(n)$, then N_n converges onto the limit set G as $n \to \infty$ (Davis et al., 1988; Kinoshita and Resnick, 1991). G has the following characteristics:

1. G is star-shaped: $\mathbf{x} \in G \implies t\mathbf{x} \in G$ for $t \in (0, 1)$.

2. G can be characterized by a continuous gauge function, g: $G = \{\mathbf{x} \in \mathbb{R}^d_+ : g(\mathbf{x}) \leq 1\}$, meaning ∂G is the unit level set of g (i.e. $\partial G = \{\mathbf{x} \in \mathbb{R}^d_+ : g(\mathbf{x}) = 1\}$).

3. g is 1-homogeneous: $g(c\mathbf{x}) = cg(\mathbf{x}), c > 0$.

The shape of G is directly related to the asymptotic dependence classes mentioned in Section 3.1. In the bivariate case, AD and AI are commonly understood through the measure $\chi = \lim_{u \to 1} P(F_1(X_1) > u \,|\, F_2(X_2) > u)$, or the probability of one variable exceeding a suitably high threshold u given that the other exceeds u. AI arises when $\chi = 0$, and AD arises when $0 < \chi \leq 1$. Visually, one can easily distinguish between the limit sets of these classes (Figure 3.1): AI (top row) yields a *blunt* shape, meaning it does not reach the upper right corner of the unit box (apart from a few edge cases), while AD shapes touch the point $(1, 1)$ in two dimensions (Balkema and Nolde, 2010, 2012).

A sufficient condition for convergence of N_n onto G conveniently relates the joint Lebesgue density of $\mathbf{X}$, if it exists, to the gauge function (Nolde, 2014; Nolde and Wadsworth, 2022):

$$\lim_{t \to \infty} \frac{-\log f_{\mathbf{X}}(t\mathbf{x})}{t} = g(\mathbf{x}), \quad \mathbf{x} \in [0, \infty)^d. \tag{3.1}$$

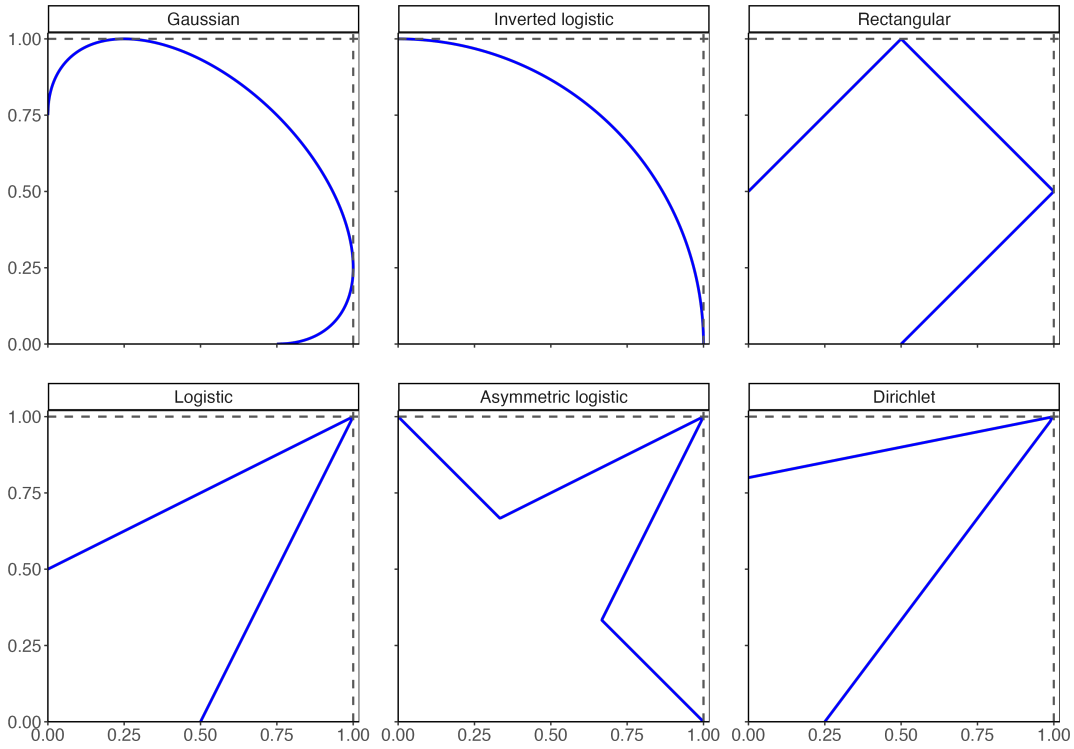


Figure 3.1: Unit level sets of six parametric gauge functions in blue. The boundary of the unit box is in gray. Top row corresponds to examples of AI gauge functions, and bottom to AD.

Assuming a tractable joint density and continuous gauge functions, one can determine the appropriate form of g and the associated G from (3.1) (Nolde and Wadsworth, 2022). Notably, each joint density has a unique gauge function, but the converse is not true.

For illustrative purposes, we provide the Gaussian and logistic gauge functions below and direct the reader to Wadsworth and Campbell (2024) and Nolde and Wadsworth (2022) for associated derivations and other parametric gauge functions. In exponential margins, the gauge function corresponding to data with a Gaussian dependence structure and correlation parameter ρ takes the general formula

$$g(\mathbf{x} \,|\, \rho) = \frac{x_1 + x_2 - 2\rho\sqrt{x_1 x_2}}{1 - \rho^2}$$

in two dimensions. For data with a logistic dependence structure and dependence parameter λ,

$$g(\mathbf{x} \,|\, \lambda) = \frac{1}{\lambda}(x_1 + x_2) + \min(x_1, x_2)(1 - \frac{2}{\lambda}).$$

In two dimensions, all of the gauge functions in Figure 3.1 have one dependence parameter except for Dirichlet, which has two. Throughout this work, we use θ when discussing dependence parameters generally, and use ρ or λ as required for specificity.

3.2.2 Radial-angular decomposition

When working with star-shaped sets, it is often convenient to transform the data from Euclidean to pseudo-polar coordinates: $R = \sum_{j=1}^{d} X_j$ and $\mathbf{W} = \mathbf{X}/R$. In two dimensions, $\mathbf{W}_2 = 1 - \mathbf{W}_1$. The joint density of $\mathbf{X}$ can then be re-expressed in terms of independent radial and angular components:

$$f_{\mathbf{X}}(\mathbf{x}) \to f_{\mathbf{W},R}(\mathbf{w}, r) = f_{R\,|\,\mathbf{W}}(r \,|\, \mathbf{w}) f_{\mathbf{W}}(\mathbf{w}), \tag{3.2}$$

necessitating models for both the radii and the angles.

3.2.2.1 Radial likelihood

First, we detail our modeling approach for the radial component of (3.2). Wadsworth and Campbell (2024) show the conditional density of $R \,|\, \mathbf{W} = \mathbf{w}$ satisfies

$$f_{R\,|\,\mathbf{W}}(r \,|\, \mathbf{w}) \propto r^{d-1} \exp\{-r g(\mathbf{w})[1 + o(1)]\}, \;\; r \to \infty. \tag{3.3}$$

They further show that for important parametric cases, the $[1 + o(1)]$ term can be moved out of the exponent and is, in most cases, negligible. Thus, for suitably large thresholds $r_\tau(\mathbf{w})$, where $\tau \in (0, 1)$ is some quantile level, a truncated Gamma distribution can approximate (3.3):

$$R \,|\, [\mathbf{W} = \mathbf{w}, R > r_\tau(\mathbf{w})] \sim \text{truncGamma}(\alpha, g(\mathbf{w}; \theta)), \tag{3.4}$$

where θ is an unknown dependence parameter and $\alpha = d$ in the limit (in most cases), but can more generally be treated as an unknown shape parameter.

To establish the asymptotic behavior of $f_{R\,|\,\mathbf{W}}(r \,|\, \mathbf{w})$ as $r \to \infty$ for various parametric dependence structures on exponential margins, the authors derive a suite of parametric gauge functions,

of which we will use six—three corresponding to AI and three to AD dependence structures (Figure 3.1). Since several known parametric AD models share the same shape of their limit set, the other known AD gauge functions are reparametrizations of the three we use and are therefore not included in our models. Throughout this work, we use the Gaussian, inverted logistic, and rectangular gauge functions for AI, and logistic, asymmetric logistic, and Dirichlet for AD (Figure 3.1).

Let $f(\cdot)$ and $F(\cdot)$ be the density and cumulative distribution function (cdf) of a Gamma distribution, respectively. The truncated Gamma model defines the following likelihood for a radius $r_i,\ \ i = \{1, \ldots, n\}$, where $\boldsymbol{\theta}_1$ is written as a vector account for the Dirichlet parametrization:

$$r_i \,|\, \mathbf{w}_i, r_i > r_\tau(\mathbf{w}_i), \boldsymbol{\theta}_1 \sim \frac{f(r_i \,|\, \alpha, g(\mathbf{w}_i; \boldsymbol{\theta}_1))}{1 - F(r_\tau(\mathbf{w}_i) \,|\, \alpha, g(\mathbf{w}_i; \boldsymbol{\theta}_1))}. \tag{3.5}$$

We also wanted to investigate whether using a censored likelihood approach could improve model fit. In contrast to typical censored likelihood approaches in multivariate extremes (Ch. 9 of Beirlant, 2004), the censored likelihood in this radial-angular decomposition uses all of the angular information below the threshold, which may raise concerns about introducing additional bias into the model. We address these concerns later in this section. Using similar notation as above, the censored model defines the following likelihood for a radius r_i:

$$r_i \,|\, \mathbf{w}_i, r_\tau(\mathbf{w}_i), \boldsymbol{\theta}_1 \sim \begin{cases} f(r_i \,|\, \alpha, g(\mathbf{w}_i; \boldsymbol{\theta}_1)) & \text{if } r_i \geq r_\tau(\mathbf{w}_i) \\ F(r_\tau(\mathbf{w}_i) \,|\, \alpha, g(\mathbf{w}_i; \boldsymbol{\theta}_1)) & \text{if } r_i < r_\tau(\mathbf{w}_i) \end{cases} \tag{3.6}$$

In both the truncated and censored cases, we rely on the asymptotic Gamma approximation, begging the question of how well this approximation performs in finite-sample settings—specifically in recovering the unknown dependence parameter θ_1. To investigate this, we simulated data from Gaussian and logistic dependence structures at varying dependence levels, as described in Section 3.5.1, and fit the two likelihoods with the data-generating gauge function. We then compared the posterior coverage rate of θ_1 to the nominal coverage level by calculating credible intervals at 19 levels, from 0.05 to 0.95. We also examined boxplots of the posterior parameter

estimates in comparison to the true value. We present results for two scenarios below (Figures 3.2 and 3.3), with additional results in Appendix A.1.

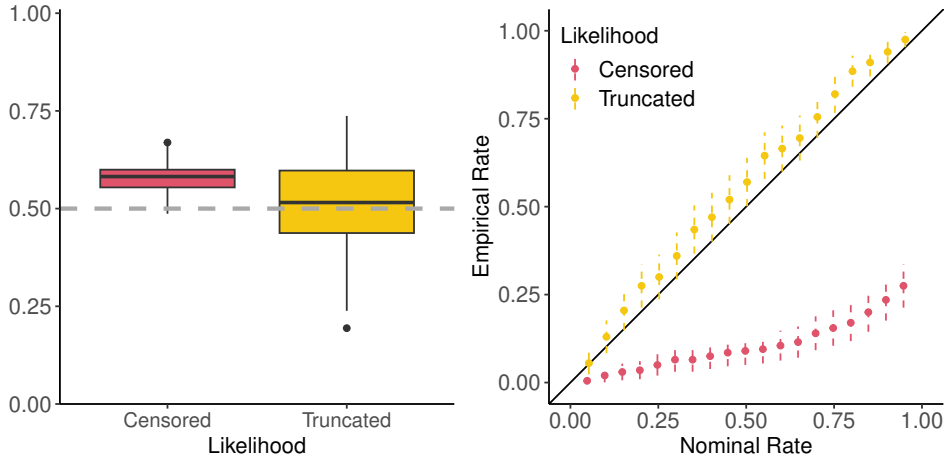


Figure 3.2: Posterior estimates (left) of the dependence parameter for 200 datasets generated from a Gaussian dependence structure with $\rho = 0.5$. Posterior coverage of the true ρ value is on the right.

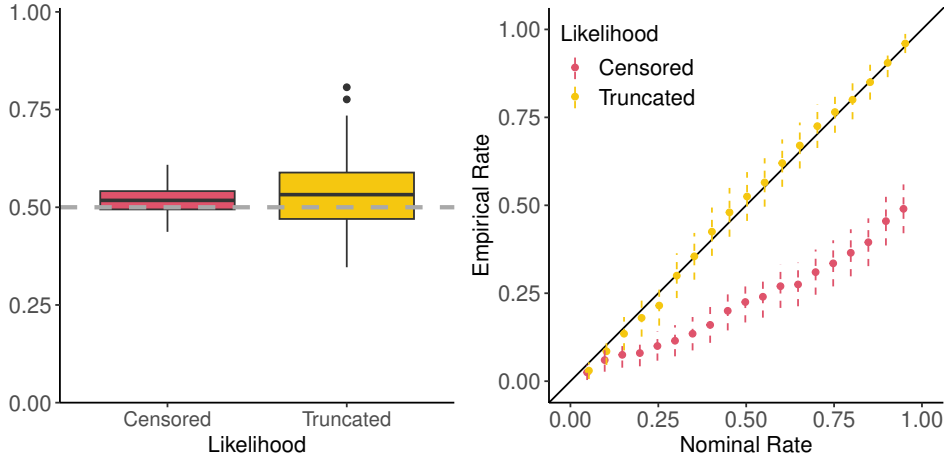


Figure 3.3: Posterior estimates (left) of the dependence parameter for 200 datasets generated from a logistic dependence structure with $\lambda = 0.5$. Posterior coverage of the true λ value is on the right.

Overall, the truncated likelihood provides substantially better coverage of the true parameter compared to the censored likelihood. While the censored likelihood tends to exhibit bias, it yields less variable estimates and may still be advantageous in small-sample settings.

Another potential drawback of the censored approach is that it tells the model that all data below the τ quantile threshold follow a Gamma distribution, thereby assuming that $(\tau \times 100)\%$ of the data come from a Gamma distribution. The truncated version ignores observations below the threshold, so is less sensitive to deviations from the gamma approximation. To explore this potential issue, we visualize the ratio of the true cdf arising from the conditional density in (3.3) to the cdf under the Gamma approximation, i.e., Gamma(shape $= 2,$ rate $= g(\mathbf{w}; \boldsymbol{\theta}_1)$), where $\boldsymbol{\theta}_1$ is the true data-generating parameter. The cdf is the second case in (3.6). Figure 3.4 illustrates how this ratio varies across angles under Gaussian and logistic dependence at three levels of dependence.

If the Gamma approximation were perfect, the ratio would be one. Ratios less than one mean the Gamma model *overestimates* the cdf; ratios greater than one mean it *underestimates*. Across all dependence scenarios, the ratio is close to one, with the largest deviations seen under the mid-dependence Gaussian setting at angles close to 0 and 1, and under the high-dependence Gaussian and low-dependence logistic settings at angles close to 0.5. These deviations could lead to bias in estimating the third term of (3.10).

Despite these potential concerns, we proceed with both likelihoods in our modeling approach. In particular, we include the censored likelihood to assess whether incorporating data below the threshold can yield more stable probability estimates, which we revisit in Section 3.6.

Finally, to complete the model, we specify the following priors for the truncated and censored likelihoods:

$$\begin{aligned} \alpha &\sim \text{Gamma}(4, \text{rate} = 2) \\ \boldsymbol{\theta}_1 &\sim \begin{cases} \text{Half-Student-t}(\nu = 4, \text{scale} = 4 \text{ or } 2) & \theta_{1,1}, \theta_{1,2}, \text{ Dirichlet} \\ \text{Uniform}(0, 1) & \theta_{1,1} \text{ otherwise} \end{cases} \end{aligned} \tag{3.7}$$

3.2.2.2 Angular likelihood

Next, we detail our modeling approach for the angular component of (3.2). Papastathopoulos et al. (2025) and Campbell and Wadsworth (2024) propose an angular density which depends on

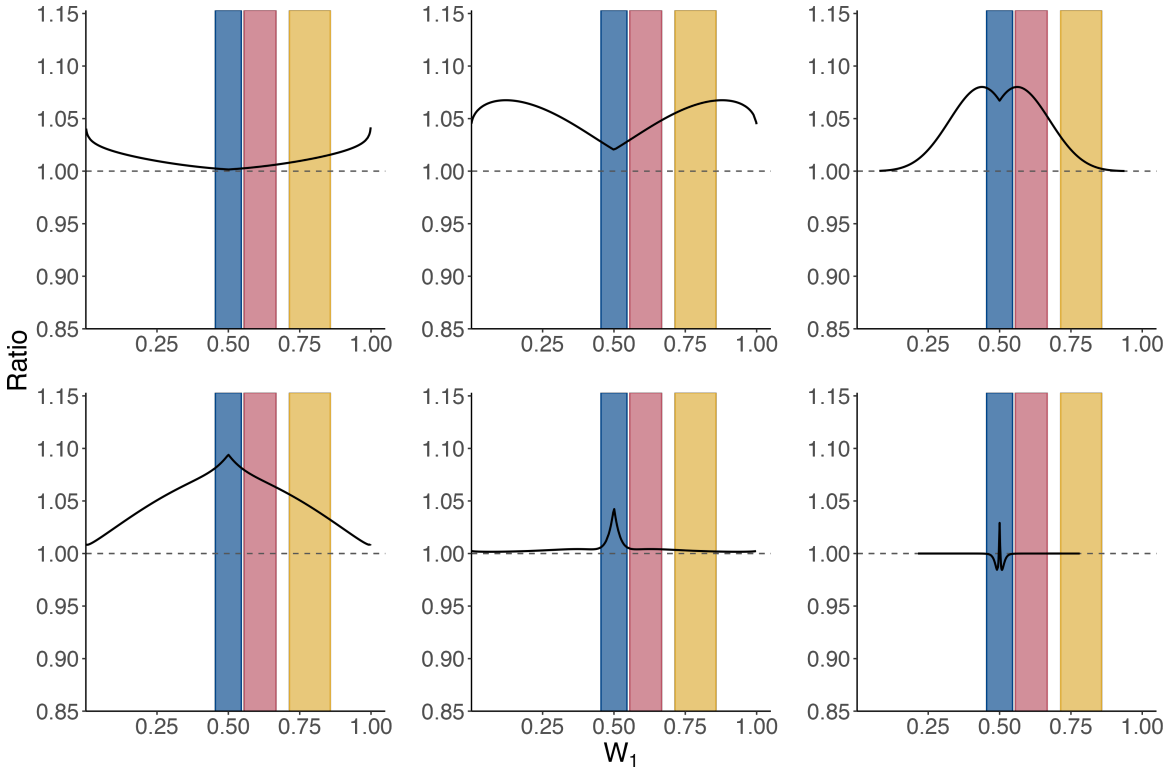


Figure 3.4: Ratios of the true cdf to the Gamma-based cdf, under Gaussian (top row) and logistic (bottom row) dependence structures. Columns show low, mid, and high-dependence (low: $\rho = 0.1,\ \lambda = 0.9$, mid:$\rho = \lambda = 0.5$, and high: $\rho = 0.9,\ \lambda = 0.1$, for Gaussian correlation and logistic dependence parameters, respectively). Colored regions correspond to risk sets B from Section 3.5. Gray dotted line indicates a ratio of one, i.e., no difference between the two probabilities.

the volume of some star-shaped set $\mathcal{L}$ and its associated gauge function ((3.8)), stemming from the work of Balkema and Nolde (2010). Since $\mathbf{w} = [\mathbf{w}_1 \quad \mathbf{w}_2]$ is fully specified by the first angular component $\mathbf{w}_1 \in (0, 1)$, we also propose a mixture of Beta densities, with weights $\mathbf{p}$ and $J = 10$ elements ((3.9)). We investigate both forms of the angular density in our approach:

$$f_{\mathbf{W}}(\mathbf{w}) = \begin{cases} \dfrac{g_{\mathcal{L}}(\mathbf{w})^{-2}}{2|\mathcal{L}|} & \text{Star-shaped} \qquad (3.8) \\ \displaystyle\sum_{j=1}^{J} p_j \, \text{Beta}(\mathbf{w}_1 \,|\, \alpha_j, \beta_j) & \text{Mixture} \qquad (3.9) \end{cases}$$

The star-shaped density relationship holds for *some* star-shaped set $\mathcal{L}$ and its associated gauge function $g_{\mathcal{L}}$, but does not necessarily apply to all multivariate point clouds. Specifically, when the joint density of the multivariate point cloud is not homothetic, as is often the case when

transforming from original to exponential margins, the relationship fails to capture the full density of angles (see Section 2.6 of Papastathopoulos et al., 2025, for a more comprehensive explanation). We include the Beta mixture density as a flexible alternative to this restriction. However, the higher dimension analogue of the Beta mixture is a mixture of Dirichlet densities, which could introduce additional computational limitations due to the larger parameter space.

To illustrate the potential differences arising from these densities, Figure 3.5 shows the empirical distribution of angles at varying dependence levels for 5000 points randomly generated from Gaussian and logistic dependence structures. Overlayed on each histogram is the posterior star-shaped density using the data-generating gauge function (Gaussian and logistic, respectively), along with the posterior Beta mixture density. In the Gaussian setting, the densities are fairly similar, and both can capture the shape of the distribution quite well at varying levels of dependence. On the other hand, the star-shaped density struggles to capture the angular distribution in the logistic setting, while the Beta mixture does not.

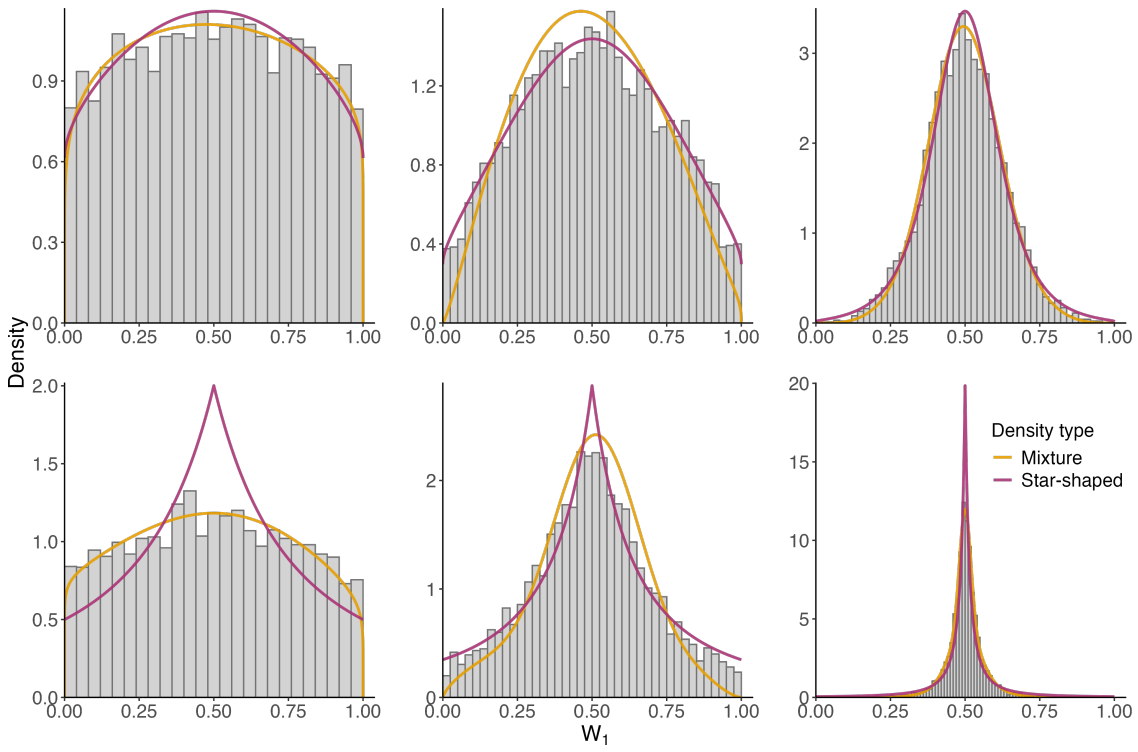


Figure 3.5: Posterior angular densities, as defined by (3.9) and sampled according to Section 3.2.2.2, for Gaussian (top) and logistic (bottom) dependence structures. Rows and columns presented as in Figure 3.4.

We use all of the angles to estimate the angular densities, despite using two different likelihood approaches for the radial density. It seems reasonable to assume that using all angles, instead of only angles associated with radii above the threshold, would allow for a better fitting posterior angular density. We discuss additional considerations surrounding this in Section 3.4.2, along with potential drawbacks in Section 3.6.4. Thus, the star-shaped angular density defines the following likelihood for $\mathbf{w}_i$:

$$\mathbf{w}_i \,|\, \boldsymbol{\theta}_2 \sim \frac{g_{\mathcal{L}}(\mathbf{w}_i \,|\, \boldsymbol{\theta}_2)^{-2}}{2|\mathcal{L}|}.$$

The prior on $\boldsymbol{\theta}_2$ is analogous to that in (3.7). The above holds for some star-shaped set $\mathcal{L}$ and a potentially unknown gauge function, g. In our implementation, we use the same family for this g as in the radial density, but allow the dependence parameters to differ. Papastathopoulos et al. (2025) estimate the volume of $\mathcal{L}$ through a latent parameter using Gaussian random fields, while Campbell and Wadsworth (2024) leverage the piecewise-linear nature of their gauge function estimation and, hence, calculate the volume in closed form. However, we have found that the volume of $\mathcal{L}$ can be estimated quite accurately and quickly using numerical integration methods, as we do here: generate a $(0,1) \times (0,1)$ grid of $n = 10,000$ equally spaced points, $\mathbf{y}_i$, and evaluate the gauge function at those points. Then, $|\mathcal{L}| \approx \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{g(\mathbf{y}_i) \leq 1\}$.

The Beta mixture defines the following likelihood for $w_{1,i}$, with weights $\mathbf{p}$:

$$w_{1,i} \,|\, \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\tau} = \sum_{j=1}^{J} p_j \text{Beta}(w_{1,i} \,|\, \mu_j \tau_j, (1-\mu_j)\tau_j),$$

In practice, imposing the convexity restriction on the mixing weights $\mathbf{p}$, $\sum_{j=1}^{J} p_j = 1$, poses mixing and identifiability challenges. To effectively implement this restriction, we place a stick-breaking prior on the weights and incorporate a latent cluster assignment $z_i \in j = \{1, \ldots, J\}$, as proposed by Ishwaran and James (2003). We also reparametrize the Beta density using location and variance-like parameters, μ and τ, to allow for efficient mixing (Kottas, 2006).

Now, the likelihood for $w_{1,i}$ becomes:

$$w_{1,i} \mid z_i, \boldsymbol{\mu}, \boldsymbol{\tau} \sim \text{Beta}(w_{1,i} \mid \mu_{z_i}\tau_{z_i}, (1-\mu_{z_i})\tau_{z_i}),$$

with priors specified as:

$$\begin{aligned}
\mu_j &\sim \text{Uniform}(0,1) \\
\tau_j &\sim \text{Inv-Gamma}(2,\beta) \\
z_i &\sim \text{Categorical}(\mathbf{p}) \\
p_1 = v_1, \quad p_j = v_j \prod_{k=1}^{j-1}(1-v_k), \quad p_J &= \prod_{k=1}^{J-1}(1-v_k) \\
v_k &\sim \text{Beta}(1,\alpha), \quad k \in \{1,\ldots,J-1\} \\
\alpha &\sim \text{Gamma}(2,2) \\
\beta &\sim \text{Exponential}(\frac{1}{8})
\end{aligned}$$

3.3 Model fitting procedures

3.3.1 Threshold selection

The gamma approximation in (3.3) applies to large values, so we must first choose a high threshold, $r_\tau(\mathbf{w})$, above which we model $R > r_\tau(\mathbf{w})$ according to (3.4). Wadsworth and Campbell (2024) use additive quantile regression to determine $r_\tau(\mathbf{w})$, with $\mathbf{W}$ treated as a covariate. Campbell and Wadsworth (2024) use a kernel density estimation approach, as a higher-dimensional analogue of the additive quantile regression approach. Papastathopoulos et al. (2025) use Gamma quantile regression in a Bayesian framework to find 20 sets of exceedances.

Similar to Wadsworth and Campbell (2024) and Campbell and Wadsworth (2024), we let $\tau = 0.95$ and select a threshold $r_\tau(\mathbf{w})$ that yields the top 5% of points overall (≈ 250 when $n = 5000$). However, we use a *marginal* threshold—with roughly balanced exceedances across both margins—as opposed to a quantile regression threshold. There are two key advantages to this. In contrast

to the quantile regression approaches, implementing a marginal threshold is simple and fast, even in higher dimensions. Secondly, the marginal threshold does not jointly depend on both variables, so transformation from standard to exponential margins could potentially be implemented into a hierarchical Bayesian model, along with the radial and angular components, thereby reducing the compounded uncertainty inherent in the marginal transformation, which is not accounted for when independently performing the transformation.

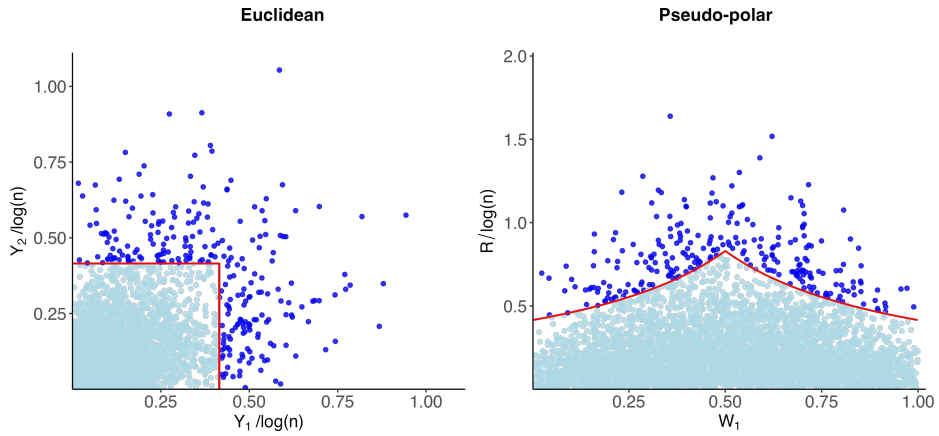


Figure 3.6: Marginal threshold, indicated by the red line in both plots, shown in both Euclidean and pseudo-polar coordinates. Truncated likelihood only uses the points above the threshold (dark blue points) in evaluating (3.5), while censored likelihood uses the dark blue points in evaluating the first case of (3.6) and the associated point on the threshold line for points below the threshold in evaluating the second.

Majumder et al. (2024) provide a comprehensive comparison of inference using a marginal threshold versus a quantile regression threshold, and they find that the marginal threshold is comparable to the quantile threshold in estimating the residual tail coefficient. While not included in this dissertation, we performed a similar comparison in early simulation studies and noted no perceivable advantage to using the quantile threshold over the marginal threshold for model fitting, evaluated against the accuracy of subsequent prediction of risk set probabilities. Thus, for quantile τ, the marginal threshold is (Figure 3.6):

$$r_\tau(\mathbf{w}) = \begin{cases} \frac{\tau}{w_1} & \text{if } w_1 > 0.5 \\ \frac{\tau}{1-w_1} & \text{if } w_1 < 0.5. \end{cases}$$

3.3.2 Joint model

Once we have determined $r_\tau(\mathbf{w})$, we can fit both radial likelihoods. Since $R \,|\, \mathbf{W}$ and $\mathbf{W}$ are independent, we model the angular likelihoods separately and join the radial and angular components after sampling. The various combinations of radial likelihoods and angular likelihoods yield four joint predictive models for evaluation. Letting $l \in \{\text{cens}, \text{trunc}\}$ denote the radial likelihood type (censored or truncated) and $a \in \{\text{mix}, \text{star}\}$ the angular density (mixture or star-shaped), we define $P_{l,a}(\mathbf{X} \in B)$ as the probability that $\mathbf{X} \in B$, under the combination of posterior parameters under radial likelihood l and angular density a.

For the censored and truncated radial likelihoods, along with the star-shaped angular likelihood, we fit $K = 6$ variations of the respective likelihoods discussed in Section 3.2.2, corresponding to each of the six gauge functions. We fit the mixture angular likelihood once, as its form does not depend on the limit set. We run these models in parallel, using `nimble` for their stick-breaking implementation of the Beta mixture likelihood, and `Rcpp` for the radial and the star-shaped angular models, implemented via an adaptive Metropolis-Hastings algorithm (Shaby and Wells, 2010).

This procedure results in K joint models under consideration in each of the four scenarios, with $k \in \{1, \ldots, K\}$ corresponding to one of the six gauge functions. More specifically, under the $\{l = \text{cens},\ a = \text{star}\}$ scenario, we have a censored radial likelihood parametrized with gauge_k coupled with a star-shaped angular likelihood with gauge_k as one model; under $l = \text{cens},\ a = \text{mix}$, a censored radial likelihood parametrized with gauge_k coupled with the mixture angular likelihood; and analogous versions for $l = \text{trunc}$. Recall that we use all angles to fit the angular likelihoods (Section 3.2.2.2, so additional manipulation is required when $l = \text{trunc}$, to ensure only the angles corresponding to the radial exceedances are incorporated. But we note this additional bookkeeping step is trivial.

3.4 Risk set probability predictions

3.4.1 Bayesian model averaging

With the various models within each scenario described above, the question remains of how (if at all) to pick the "best" model. With standard modeling procedures, we would select a model based on a metric like AIC. As mentioned in Section 3.1, model selection procedures introduce uncertainty and bias through the choice of comparison metric and potentially excluding (unknowingly) the data-generating model from consideration. We use BMA to overcome the limitations inherent in model selection.

To concretely illustrate the mechanics of BMA, we detail the concept in the context of our problem, considering the prediction of quantity $P_{l,a}(\mathbf{X} \in B)$. As described previously, we have access to models $\mathcal{M} = \{M_1, \ldots, M_K\}$, each corresponding to one of the gauge functions, and data $\mathbf{y} := (r, \mathbf{w}, r_\tau(\mathbf{w}))$. The average posterior prediction under each model, averaged across the models, is then

$$P_{l,a}(\mathbf{X} \in B \,|\, \mathbf{y}) = \sum_{k=1}^{K} P_{l,a}(\mathbf{X} \in B \,|\, M_k, \mathbf{y}) P_{l,a}(M_k \,|\, \mathbf{y}).$$

Predictions are weighted by the posterior probability of their corresponding models. This weighted average has superior predictive performance than any single model M_k on its own (Madigan and Raftery, 1994) and is the optimal predictor (in terms of log predictive score; Hoeting et al., 1999).

The posterior probability under scenario $\{l, a\}$ of model M_k is

$$P(M_k \,|\, \mathbf{y}) = \frac{f(\mathbf{y}|M_k)P(M_k)}{\sum_{k=1}^{K} f(\mathbf{y} \,|\, M_k)P(M_k)},$$

and the marginal likelihood of M_k is

$$f(\mathbf{y} \,|\, M_k) = \int f(\mathbf{y} \,|\, \boldsymbol{\beta}_k, M_k)\pi(\boldsymbol{\beta}_k \,|\, M_k)d\boldsymbol{\beta}_k,$$

where $\boldsymbol{\beta}_k$ are the parameters for M_k under $\{l, a\}$. This integral is often intractable, so several approximations have been proposed (Hoeting et al., 1999; Yao et al., 2018), including Pseudo-BMA

and Pseudo-BMA+. Pseudo-BMA is an AIC-type weighting that uses the product of Bayesian leave-one-out cross-validated predictive densities in place of the marginal likelihoods. Pseudo-BMA+ stabilizes these estimates using Bayesian bootstrapping.

Yao et al. (2018) also propose stacking weights, a predictive distribution analog of *stacking of means*:

$$\mathbf{s} \in \overset{\max}{[0,1]^K} \left[\frac{1}{n}\sum_{i=1}^{n} \log \sum_{k=1}^{K} s_k f(y_i|\mathbf{y}_{-i}, M_k)\right]; \ \sum_{k=1}^{K} s_k = 1.$$

Stacking weights use the corresponding leave one out predictive distribution as an empirical approximation of the full predictive distribution. These weights are asymptotically optimal (with respect to the predictive log score) compared to generic BMA weights, even when the data-generating model is not in $\mathcal{M} = \{M_1, \ldots, M_K\}$.

We calculate all three variations of BMA weights from the pointwise joint radial-angular log-likelihood using the `loo` package.

3.4.2 Importance sampling scheme

Before defining the estimation procedure of $P_{l,a}(\mathbf{X} \in B)$, we first lay out a key assumption:

Assumption 1. *Consider all angles* $\mathbf{W}$. *For any* $q \in (0,1)$, $\mathbb{P}(R \geq r_q(\mathbf{W})) = 1 - q$, *where* $r_q(\mathbf{w})$ *is the* q *conditional quantile of the radii at angle* $\mathbf{w}$. *Then, as* $q \to 0$, $\mathbf{W} \mid R \geq r_q(\mathbf{W}) \overset{d}{=} \mathbf{W}$ *(Prop. 5, Papastathopoulos et al. (2025)).*

Assumption 1 is valid when $r_q(\mathbf{w})$ is determined via perfect quantile regression. While our model fitting procedure does not use a quantile regression approach to find $r_q(\mathbf{w})$, our importance sampling schema below *does*. Figure 3.7 visualizes the density of angles above this quantile threshold (Figure 3.8), calculated from a Gamma distribution quantile parametrized with the posterior estimates of the radial parameters. The assumption seems to hold in most scenarios except for the logistic dependence structure with low dependence, where the assumption breaks down when w_1 is close to 0.5, aligning with the dip in the cdf ratios seen in Figure 3.4.

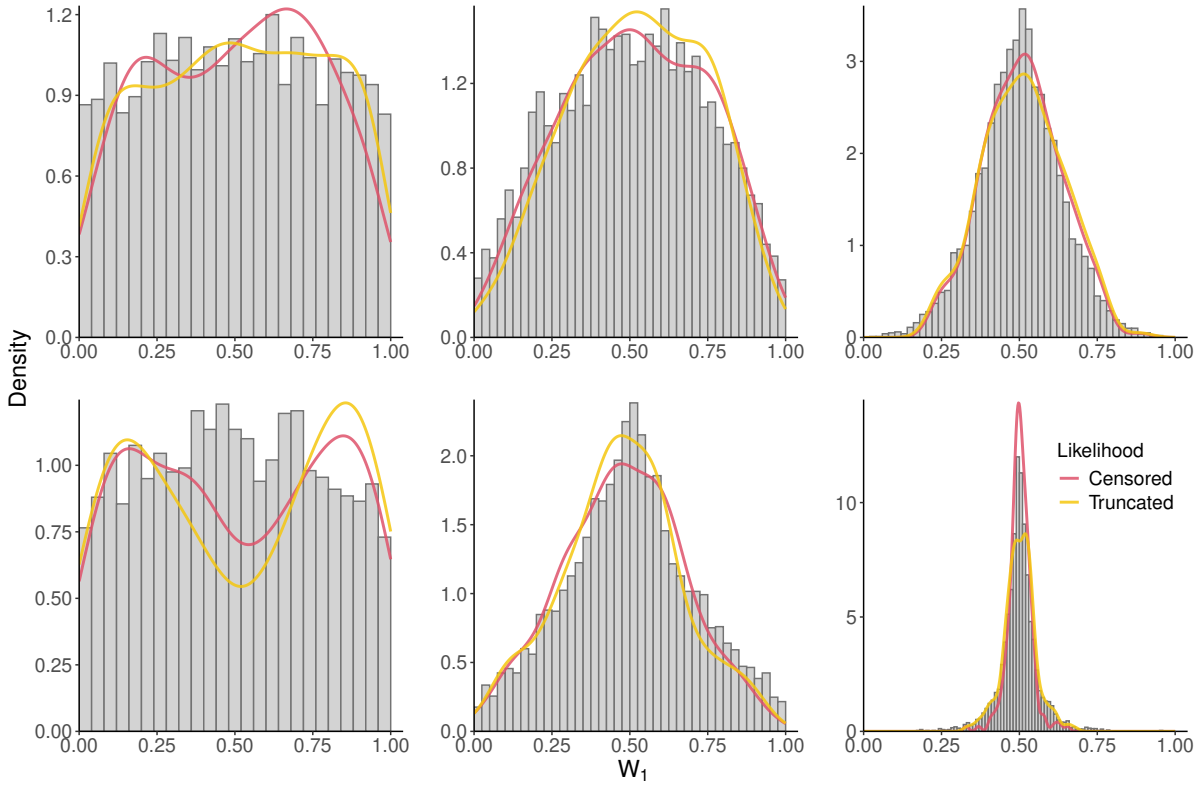


Figure 3.7: Histogram of empirical angles for one realization of the 200 simulated datasets generated for each dependence type and level. Top and bottom rows follow the same schema as Figure 3.4. The two lines overlayed on each histogram are the kernel density estimates of angles over the threshold, using the posterior estimates of α and $\boldsymbol{\theta}_1$ to find $r_q(\mathbf{w})$ as the 95th quantile of a Gamma(α_{post}, $\boldsymbol{\theta}_{1,\text{post}}$). See Figure 3.8 for visualization of these thresholds.

To estimate the probability $P_{l,a}(\mathbf{X} \in B)$ in regions beyond the range of our data, we use a simple importance sampling procedure. Let $S_B = \{\mathbf{w} \in \mathcal{S}^{d-1} : r\mathbf{w} \in B, r \in (0, \infty]\}$, $I_B(\mathbf{W}) = \{r > 0 : r\mathbf{w} \in B, \mathbf{w} \in S_B\}$. This implies that if $R\mathbf{W} \in B$, then $\mathbf{W} \in S_B$ and $R \in I_B(\mathbf{W})$. With this decomposition, we write:

$$
\begin{aligned}
P_{l,a}(\mathbf{X} \in B) &= P_{l,a}(\mathbf{X} \in B \,|\, R \geq r_q(\mathbf{w}))P_l(R \geq r_q(\mathbf{w})) \\
&= P_{l,a}(R\mathbf{W} \in B \,|\, R \geq r_q(\mathbf{w}))P_l(R \geq r_q(\mathbf{w})) \\
&= P_l(R \in I_B(\mathbf{W}) \,|\, \mathbf{W} \in S_B, R \geq r_q(\mathbf{w}))P_a(\mathbf{W} \in S_B \,|\, R \geq r_q(\mathbf{w}))P_l(R \geq r_q(\mathbf{w}))
\end{aligned}
$$

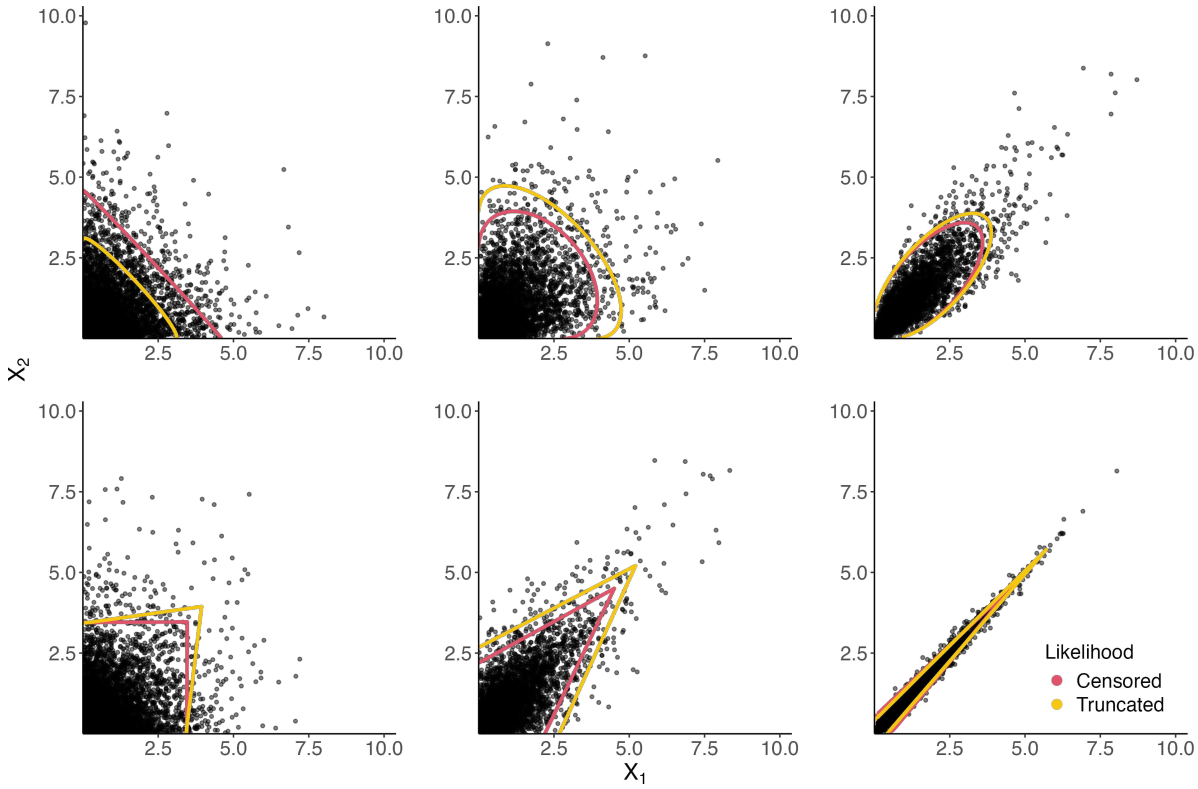


Figure 3.8: Posterior $r_q(\mathbf{w})$ determined as the 95th quantile of a Gamma distribution, using the posterior estimates of α and $\boldsymbol{\theta}_1$. Top and bottom rows follow the same schema as Figure 3.4.

Using Assumption 1,

$$P_{l,a}(\mathbf{X} \in B) = P_l(R \in I_B(\mathbf{W}) \,|\, \mathbf{W} \in S_B, R \geq r_q(\mathbf{w}))P_a(\mathbf{W} \in S_B)P_l(R \geq r_q(\mathbf{w})) \tag{3.10}$$

In two dimensions, we choose a bivariate normal distribution as our importance or proposal distribution, H, but we note that this can easily be extended to higher-dimensional settings with the multivariate normal analogue. We sample $\mathbf{X}^*$ from $H := \text{MVN}(\boldsymbol{\mu} = \text{midpoint of } B,\ \boldsymbol{\Sigma} = \mathbf{c}_{\text{dist}} \times \mathbf{I}_2)$, where $\mathbf{c}_{\text{dist}}$ is a 1×2 vector: the first element corresponds to the width of the box in the x-direction and the second to the y-direction. In our simulation study, we set $\mathbf{c}_{\text{dist}} = (2,\ 2)$, but allow it to vary in our data analyses.

Define $r_q(\mathbf{w}^*)$ as the q quantile of a Gamma distribution with shape α_{post} and rate $g(\mathbf{w}^*; \boldsymbol{\theta}_{1,\text{post}})$, where $\mathbf{w}^*$ are the corresponding angles of $\mathbf{X}^*$ (and r^* the corresponding radii). Let $h_{R,\mathbf{W}}$ denote

the importance density transformed to the L_1-norm. Then, suppressing the l, a for readability,

$$
\begin{aligned}
(3.10) = \frac{1}{n}\sum_{i=1}^{n} & \mathbb{1}\{r_i^* \in I_B(\mathbf{w}_i^*) \,|\, \mathbf{w}_i^* \in S_B, r_i^* \geq r_q(\mathbf{w}_i^*)\} \times \mathbb{1}\{\mathbf{w}_i^* \in S_B\} \\
& \times \frac{f_{R\,|\,\mathbf{W}}(r_i^* \,|\, \mathbf{w}_i^*, r_q(\mathbf{w}_i^*), \alpha_{\text{post}}, \boldsymbol{\theta}_{1,\text{post}}) f_{\mathbf{W}}(\mathbf{w}_i^* \,|\, \boldsymbol{\theta}_{2,\text{post}})}{h_{R,\mathbf{W}}(r_i^*, \mathbf{w}_i^*)} \times (1-q) \\
= \frac{1}{n}\sum_{i=1}^{n} & \mathbb{1}\{\mathbf{x}_i^* \in B \,|\, r_i^* \geq r_q(\mathbf{w}_i^*)\} \\
& \times \frac{f_{R\,|\,\mathbf{W}}(r_i^* \,|\, \mathbf{w}_i^*, r_q(\mathbf{w}_i^*), \alpha_{\text{post}}, \boldsymbol{\theta}_{1,\text{post}}) f_{\mathbf{W}}(\mathbf{w}_i^* \,|\, \boldsymbol{\theta}_{2,\text{post}})}{h_{\mathbf{X}^*}(\mathbf{x}_i^*)\; r_i^*} \times (1-q)
\end{aligned}
$$

As mentioned in Section 3.2.2.1, the Gamma approximation is not perfect, introducing the potential for bias in our probability prediction estimates. The most obvious place for this bias to arise is in the $(1-q)$ term of (3.10); when the ratio of true cdf to the approximated cdf (Figure 3.4) is less than one, our estimate of $(1-q)$ may be underbiased, while overbias may occur in the opposite situation. We tried to account for the variation seen in this ratio across angles by using self-normalized importance sampling (Owen, 2013). However, we did not see any improvement in predictive performance, and in some cases, the estimates were worse.

3.5 Simulation study

We evaluate different combinations of radial likelihoods, angular models, and BMA weighting against two existing semi-parametric approaches: BezELS (Majumder et al., 2024) and C-W (Campbell and Wadsworth, 2024). Majumder et al. (2024) do not use their approach to calculate risk set probability predictions, so we adopt their existing mechanics for the truncated radial likelihood and couple that with our Beta mixture angular density, subsequently generating predictions using our procedure detailed in Section 3.4.2. Campbell and Wadsworth (2024) propose multiple model variants, so, for the sake of comparison, we implement their best-fitting model (SS4) and use their proposed importance sampling scheme for prediction. We compare all methods based on their ability to capture the true probability of lying in three regions (Figure 3.9): $B_1 = (10, 12) \times (10, 12),\ B_2 = (10, 12) \times (6, 8),$ and $B_3 = (10, 12) \times (2, 4)$.

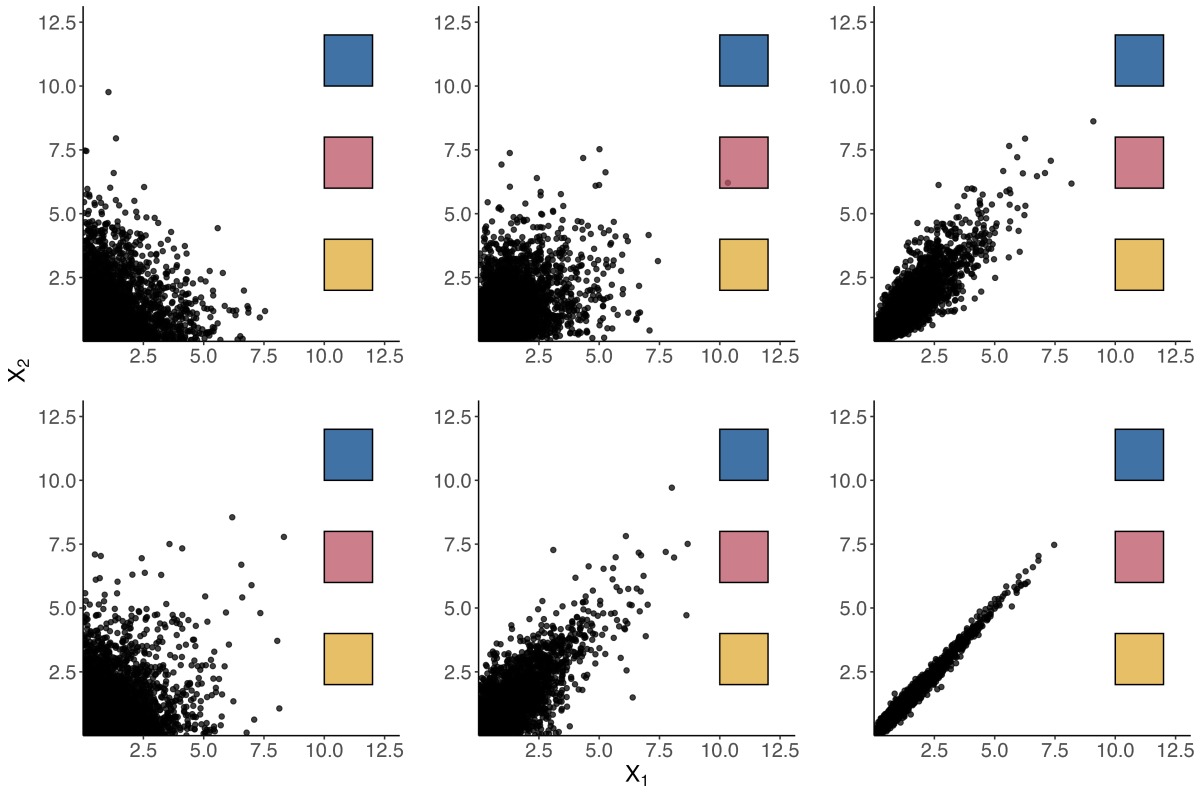


Figure 3.9: Prediction regions under the simulation study: $B_1 = (10, 12) \times (10, 12)$, $B_2 = (10, 12) \times (6, 8)$, and $B_3 = (10, 12) \times (2, 4)$, represented by the blue, red, and yellow boxes, respectively. Panels follow the scheme as Figure 3.4. Data are generated according to Section 3.5.1.

3.5.1 Data-generating procedure

To provide a comprehensive comparison, we examine nine different combinations of dependence structures and dependence levels:

- Gaussian: $\rho = 0.1$ (low), $\rho = 0.5$ (mid), and $\rho = 0.9$ (high)
- Logistic: $\lambda = 0.9$ (low), $\lambda = 0.5$ (mid), and $\lambda = 0.1$ (high)
- Hüsler-Reiss: $\lambda = 0.1$ (low), $\lambda = 1$ (mid), and $\lambda = 3$ (high) (Figure A7)

Under each dependence structure and level, we generate 200 datasets of $5,000$ points each, picking the top 250 ($\approx 5\%$ or $1 - \tau = 0.05$) points as the marginal threshold.

3.5.2 Results

With three boxes of interest for each dependence type and level, there are a total of 27 sets of comparison scenarios. As an initial overview of the results, we first compare the various methods using RMSE. We calculate RMSE for each of our four modeling scenarios under the three BMA weighting methods, along with the two semi-parametric methods, using the 200 sets of predictions generated within each approach. Since the true and predicted probabilities are incredibly small, the resulting RMSE is much smaller than one and difficult to interpret. We therefore standardize each RMSE value by the corresponding truth to yield a unitless metric that may be interpreted as percent difference from the truth. Tables 3.2, 3.3, and 3.4 provide these comparisons, with a corresponding key in Table 3.1.

Table 3.1: Key for understanding the RMSE comparisons among our models, BezELS, and C-W. All models developed in this work are represented by a combination of one aspect from each of the first three columns. The semi-parametric methods are depicted as shown in the last column.

BMA Method	Likelihood	Ang. Dens.	Other Methods
Pseudo-BMA	*Censored*	Mixture: ◎	C-W
Pseudo-BMA+	**Truncated**	Star-shaped: ★	◎ BezELS
Stacking			

Within our censored models, the best-performing angular density and BMA weight method varies by dependence structure. In the Gaussian setting, most models favored the star-shaped angular density, followed closely by the mixture; Pseudo-BMA and Pseudo-BMA+ outperform stacking weights. In the logistic setting, the mixture angular density is heavily preferred, and stacking performs best in the low-dependence setting. The mixture density has a slight advantage over the star-shaped in the Hüsler-Reiss scenarios, and stacking performs best under high dependence.

Our censored model outperforms BezELS and C-W in all but three of the 27 scenarios. C-W slightly outperforms our censored model with the angular mixture density using stacking weights in the logistic low-dependence setting for B_3 (RMSE: 0.810 vs 0.983). C-W's RMSE is about half that of our best model (truncated, angular mixture, Pseudo-BMA weights) under the

Table 3.2: Comparisons across low, mid, and high levels of Gaussian dependence structures, in three regions of interest (Figure 3.9, top row).

Low			Mid			High		
B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
★ *1.135*	◎ *0.557*	★ *0.211*	★ *0.637*	◎ *0.318*	★ *0.160*	★ *0.188*	◎ *0.184*	★ *0.670*
★ *1.138*	◎ *0.608*	◎ *0.225*	★ *0.637*	★ *0.336*	★ *0.160*	★ *0.191*	◎ *0.185*	★ *0.670*
4.830	★ *0.656*	◎ *0.230*	◎ *1.090*	★ *0.336*	◎ *0.196*	★ *0.191*	★ *0.203*	◎ *0.689*
★ **10.960**	★ *0.658*	★ *0.233*	1.201	◎ *0.356*	◎ *0.198*	◎ *0.199*	★ *0.203*	◎ *0.825*
★ **11.085**	◎ *1.738*	★ *0.233*	◎ *1.333*	◎ *0.394*	◎ *0.199*	◎ *0.202*	◎ *0.206*	◎ *3.233*
◎ *13.588*	2.536	◎ *0.240*	◎ *1.886*	★ *0.411*	★ *0.225*	◎ *0.207*	★ *0.224*	12.863
◎ *17.190*	★ **3.428**	0.827	★ **2.300**	0.658	0.466	0.684	0.624	★ *41.065*
◎ *46.196*	★ **3.435**	★ **0.847**	★ **2.322**	★ **1.338**	★ **0.841**	◎ 1.463	◎ **1.431**	◎ 140.260
★ **76.072**	★ **4.515**	★ **0.847**	★ *3.676*	★ **1.338**	★ **0.878**	◎ **1.893**	◎ **1.484**	◎ **245.925**
★ *335.088*	◎ **11.312**	★ **0.850**	★ **4.859**	★ **1.388**	★ **0.879**	◎ **1.969**	★ **1.584**	◎ **248.301**
◎ 412.551	◎ **12.729**	◎ **0.888**	◎ 8.620	◎ **1.549**	◎ **0.958**	◎ **2.064**	★ **1.584**	◎ **292.333**
◎ **514.489**	★ *12.974*	◎ **0.892**	◎ **12.112**	◎ **1.576**	◎ **0.995**	★ **2.573**	◎ 1.585	★ **314.683**
◎ **540.811**	◎ **14.493**	◎ **1.037**	◎ **12.475**	◎ **1.581**	◎ 1.082	★ **2.574**	★ **1.585**	★ **315.351**
◎ **778.821**	◎ 18.575	◎ 1.516	◎ **12.655**	◎ 1.722	◎ **1.193**	★ **2.581**	◎ **1.649**	★ **315.390**

Table 3.3: Comparisons across low, mid, and high levels of logistic dependence structures, in three regions of interest (Figure 3.9, bottom row). The probability for B_3 under the high scenario is smaller than R's machine precision, yielding zero as the true probability; as a result, the RMSE values are not informative and are excluded from this table.

Low			Mid			High		
B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
◎ *0.467*	★ *0.399*	0.810	◎ *0.194*	◎ *0.335*	◎ *0.805*	◎ *0.182*	★ *40.125*	–
★ *0.524*	◎ *0.485*	◎ *0.983*	◎ *0.207*	◎ *0.418*	◎ *0.821*	◎ *0.191*	★ *42.381*	–
0.635	◎ *0.503*	◎ *1.099*	◎ *0.287*	◎ *1.084*	◎ *1.527*	◎ *0.198*	◎ *60.439*	–
◎ *0.696*	◎ *0.541*	◎ *1.211*	0.376	★ **1.889**	★ *3.429*	★ *0.314*	◎ *88.483*	–
★ **0.770**	0.550	★ *1.274*	★ *0.436*	★ **1.952**	★ *3.610*	★ *0.493*	★ *4.412e+04*	–
★ **0.777**	★ *0.708*	◎ **1.454**	★ *0.560*	★ *2.051*	5.682	★ *0.522*	◎ *1.172e+05*	–
★ **0.781**	★ **0.720**	★ *1.528*	★ *0.623*	★ **2.177**	★ *10.734*	0.530	★ **1.042e+06**	–
◎ *0.878*	★ **0.721**	★ *1.544*	◎ 1.035	2.302	★ **16.002**	◎ 0.660	★ **1.598e+06**	–
★ *0.961*	★ *0.723*	◎ **1.582**	◎ **1.423**	★ *2.898*	★ **16.287**	◎ **0.752**	★ **2.354e+06**	–
★ *0.966*	★ **0.732**	◎ **1.618**	◎ **1.479**	◎ **3.189**	★ **16.571**	◎ **0.781**	◎ **3.214e+06**	–
◎ 2.132	◎ **1.261**	★ **1.740**	◎ **1.628**	★ *3.275*	◎ **26.386**	◎ **0.808**	◎ **5.305e+06**	–
◎ **2.235**	◎ **1.328**	★ **1.823**	★ **2.112**	◎ **3.540**	◎ **27.720**	★ **0.958**	◎ **7.779e+06**	–
◎ **2.436**	◎ **1.332**	★ **1.876**	★ **2.170**	◎ **3.754**	◎ **28.014**	★ **0.990**	$1.922e+07$	–
◎ **2.655**	◎ 2.009	◎ 1.920	★ **2.216**	◎ 4.070	◎ 29.302	★ **1.002**	◎ $2.013e+07$	–

Hüsler-Reiss mid-dependence case for B_3 and is about 25% better than ours under the Hüsler-Reiss high-dependence scenario in B_2. Notably, this is also the only scenario in which our truncated likelihood outperforms our censored likelihood.

Table 3.4: Comparisons across low, mid, and high levels of Hüsler-Reiss dependence structures, in three regions of interest (Figure A7). As in Table 3.3, the probability for B_3 under the high scenario is smaller than R's machine precision, yielding zero as the true probability; as a result, the RMSE values are not informative and are excluded from this table.

Low			Mid			High		
B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
★ *1.265*	◎ *0.555*	◎ *0.240*	◎ *0.441*	★ *0.167*	5.644	★ *0.219*	22.498	–
★ *1.267*	◎ *0.578*	◎ *0.245*	◎ *0.448*	★ *0.170*	◎ **10.007**	0.221	◎ **30.273**	–
4.508	★ *0.737*	◎ *0.262*	0.449	★ *0.223*	★ **10.276**	◎ *0.236*	◎ **31.948**	–
★ **20.243**	★ *0.737*	★ *0.275*	★ *0.450*	◎ *0.260*	◎ *10.297*	◎ *0.241*	★ **33.644**	–
★ **20.285**	1.909	★ *0.275*	◎ *0.478*	◎ *0.270*	◎ **10.356**	◎ *0.242*	★ **34.231**	–
◎ *57.390*	◎ *2.441*	★ *0.506*	★ *0.761*	◎ *0.299*	◎ **10.364**	★ *0.333*	★ **35.076**	–
◎ *62.626*	★ **5.934**	1.014	★ *0.799*	0.398	★ **10.602**	★ *0.333*	★ *36.018*	–
★ **72.494**	★ **6.234**	◎ **1.151**	◎ 2.471	★ **1.620**	★ **10.839**	◎ 0.877	◎ *36.157*	–
◎ *94.396*	★ **6.242**	◎ **1.174**	◎ **3.368**	★ **1.633**	◎ *11.006*	◎ **1.076**	◎ **36.805**	–
★ *1.470e+03*	◎ **17.149**	◎ **1.181**	◎ **3.537**	★ **1.654**	◎ *11.191*	◎ **1.162**	◎ *39.026*	–
◎ **1.612e+03**	◎ **22.004**	★ **1.202**	◎ **3.574**	◎ **2.535**	★ *12.932*	◎ **1.187**	◎ *42.181*	–
◎ **1.802e+03**	◎ **22.167**	★ **1.205**	★ **6.155**	◎ **2.669**	◎ 13.189	★ **1.807**	★ *46.009*	–
◎ $1.909e+03$	◎ 38.714	★ **1.205**	★ **6.608**	◎ **2.743**	★ *16.924*	★ **1.954**	★ *46.009*	–
◎ **2.712e+03**	★ *38.782*	◎ 1.745	★ **7.198**	◎ 2.986	★ *17.359*	★ **2.031**	◎ 84.377	–

The RMSE comparisons allow us to tangibly evaluate the bias-variance tradeoff across these methods, yet they do not tell the whole story. To investigate the comparisons on a more granular level, we rely on boxplots of the results by dependence structure and level. For readability, we present the results from one Gaussian and one logistic setting here; the remaining 25 (including all Hüsler-Reiss settings) are provided in Appendix A.2. All boxplots use a 95% winsorization to exclude extreme outliers and provide a better visual comparison.

Figures 3.10 and 3.11 provide detailed comparisons of the mid-dependence scenarios in B_2 for Gaussian and logistic structures, respectively. In both, predictions from the censored models exhibit substantially lower variability than those from the truncated, BezELS, or C-W approaches. In the Gaussian case, all of our censored models (along with the truncated angular mixture and BezELS) are slightly overbiased, while the truncated model with the star-shaped density is essentially unbiased, and C-W is slightly underbiased. In the logistic case, all approaches are heavily overbiased, except for the censored mixture model using Pseudo-BMA+ weights (unbiased) and its Pseudo-BMA counterpart (underbiased). This nuance is lost when looking at the RMSE

comparisons alone, so it is crucial to recognize the bias-variance tradeoff and weigh what is more important for the application setting.

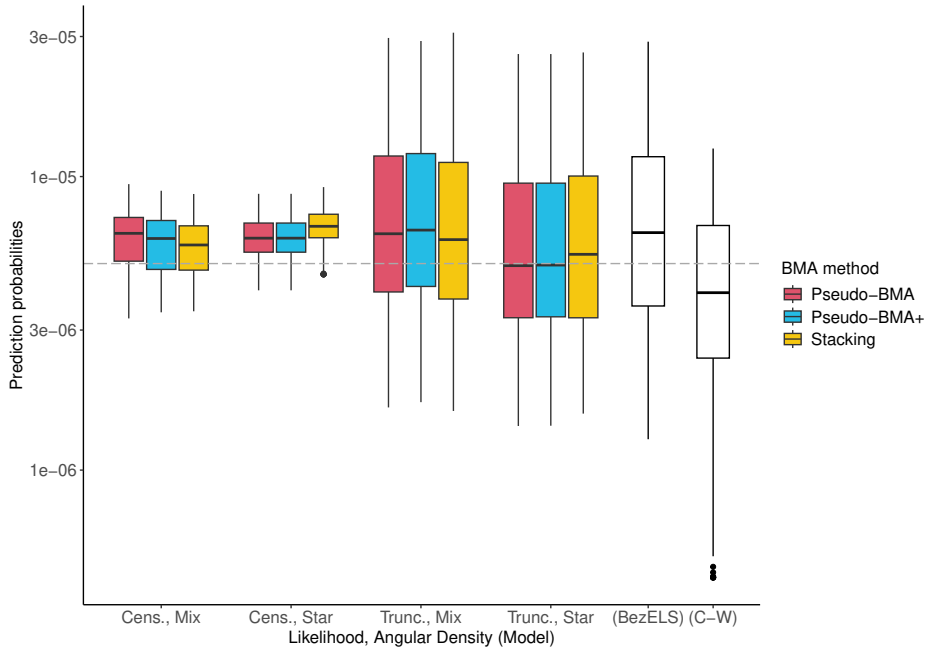


Figure 3.10: Predicted probabilities of lying in B_2, under a Gaussian dependence structure with $\rho = 0.5$ (see Figure 3.9), across 200 simulated datasets. The dashed grey line is the true probability. Y-axis is shown on the log scale.

3.6 Data analyses

We apply our modeling and prediction procedures in analyzing the relationship between two fire weather indices: the Energy Release Component (ERC, Bradshaw et al., 1984) and Canadian Forest Fire Weather Index (FWI, Turner and Lawson, 1978). Lawler and Shaby (2024) find regional variations and tradeoffs between the two indices, specifically in predicting the size of wildfire burned areas. Their analysis, based on surface meteorological data at a $4\text{km} \times 4\text{km}$ resolution and a single forest fuel model (obtained from GridMET, Abatzoglou, 2013), shows that when one index is extreme in a specific region, the other often is not. We use our modeling approach to further investigate this relationship at a more localized scale, using forest fuel models specific to the region and higher-resolution weather data.

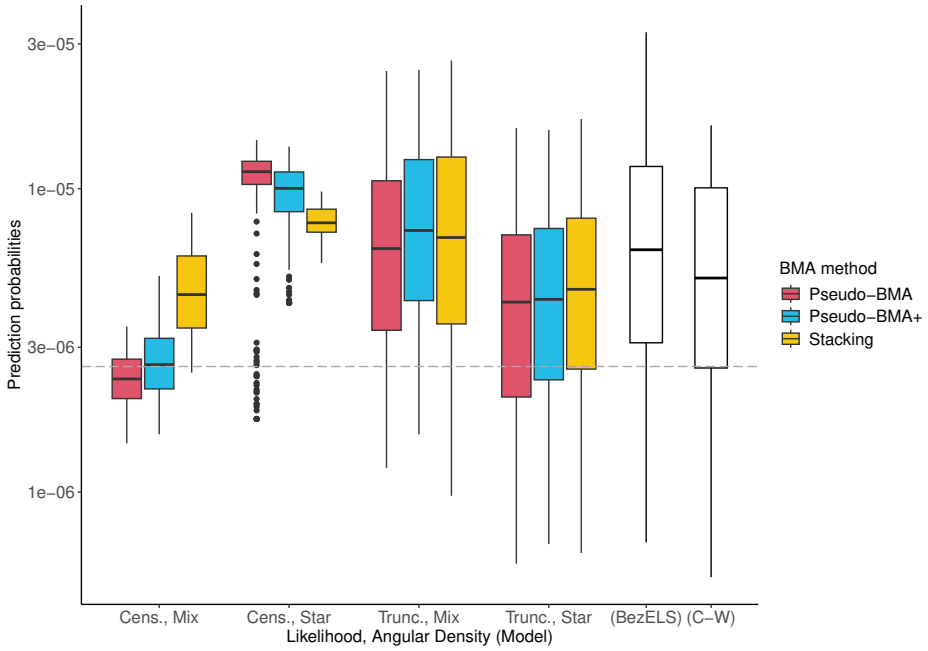


Figure 3.11: Predicted probabilities of lying in B_2, under a logistic dependence structure with $\lambda = 0.5$ (see Figure 3.9), across 200 simulated datasets. The dashed grey line is the true probability. Y-axis is shown on the log scale.

3.6.1 Fire weather indices

ERC is a calculated fuel moisture index from the National Fire Danger Rating System, which considers the cumulative drying effect of temperature, precipitation, humidity, and solar radiation as an overall measure of wildfire fuel (Abatzoglou and Kolden, 2013). As a relative index, an ERC value of 30 represents twice the potential heat release of a value of 15. FWI captures somewhat orthogonal fire behavior, combining fuel moisture and weather conditions into an index of potential wildfire activity.

We calculate daily ERC and FWI values using FireFamilyPlus (FF+,version 5; Bradshaw and McCormick, 2000), based on hourly weather data from the Program for Climate, Ecosystem, and Fire Applications and the Wildland Fires Application portal. After calculating the indices, we restrict these data to the fire season (April - October). We focus on two Remote Automatic Weather Stations (RAWS): Friend Mountain in California and Redstone in Colorado. We selected these sites due to their proximity to the August Complex and Cameron Peak fires, the largest wildfires in each state's history, both of which ignited in August 2020.

Notably, FF+'s methodology for calculating FWI differs slightly from packages like `cffdrs`, yielding only integer-valued outputs. While this results in a somewhat discrete appearance in parts of the FWI distribution, the number of observed values is large enough to justify applying a continuous marginal transformation.

3.6.2 Marginal transformation

Before applying our model to these fire indices, we must first transform the data to standard exponential margins. We want to capture the tail behavior of each index appropriately while also fitting the bulk of the distribution, so we fit an extended generalized Pareto distribution (EGPD, Naveau et al., 2016) separately to each fire index. For FWI, we incorporate the EGPD into a mixture model with an upper-truncated exponential distribution (truncated at 45 and 100 for the two RAWS, respectively). We truncate the exponential component so the upper tail behavior is driven by the EGPD component.

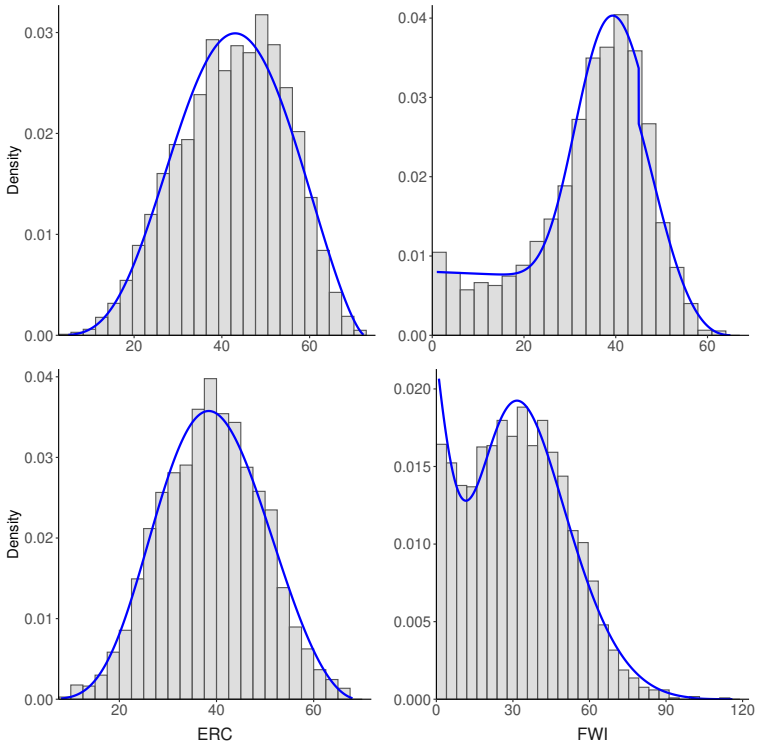


Figure 3.12: Empirical distribution of the fire indices, with their posterior marginal distributions overlayed in blue. Friend Mountain is the top row, and Redstone the bottom.

The usefulness of this mixture is evident in Figure 3.12, as the mixture marginal distribution estimates the FWI distribution quite well. The marginal EGPD distributions also appear to fit the ERC distributions well, with a potential underestimation around ERC values of 55 and 50, respectively, for each RAWS.

Once transformed onto standard exponential margins, the indices appear to follow a standard exponential distribution, apart from a few points in the tails of FWI at both RAWS (Figure 3.13).

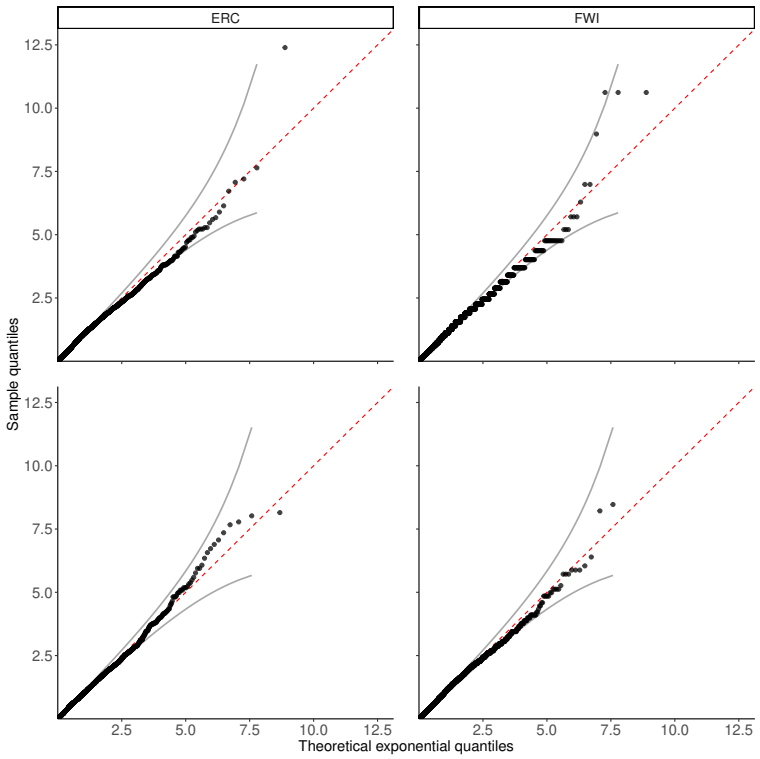


Figure 3.13: QQ plot comparing the fire indices on exponential margins to the theoretical quantiles of a standard exponential distribution. The identity line is in red and the 95% confidence band in gray. As in Figure 3.12, top is Friend Mountain and bottom is Redstone.

3.6.3 Threshold selection and model fitting

These datasets have fewer points than our simulated datasets, with approximately 3600 and 3000 data points, yielding 184 and 155 points in the top 5% of the datasets, respectively. The limitation of the smaller sample size is immediately apparent when comparing the unit level sets of

the posterior gauge function fits of the truncated and censored likelihoods (Figure 3.14), showing a clear distinction between the two approaches.

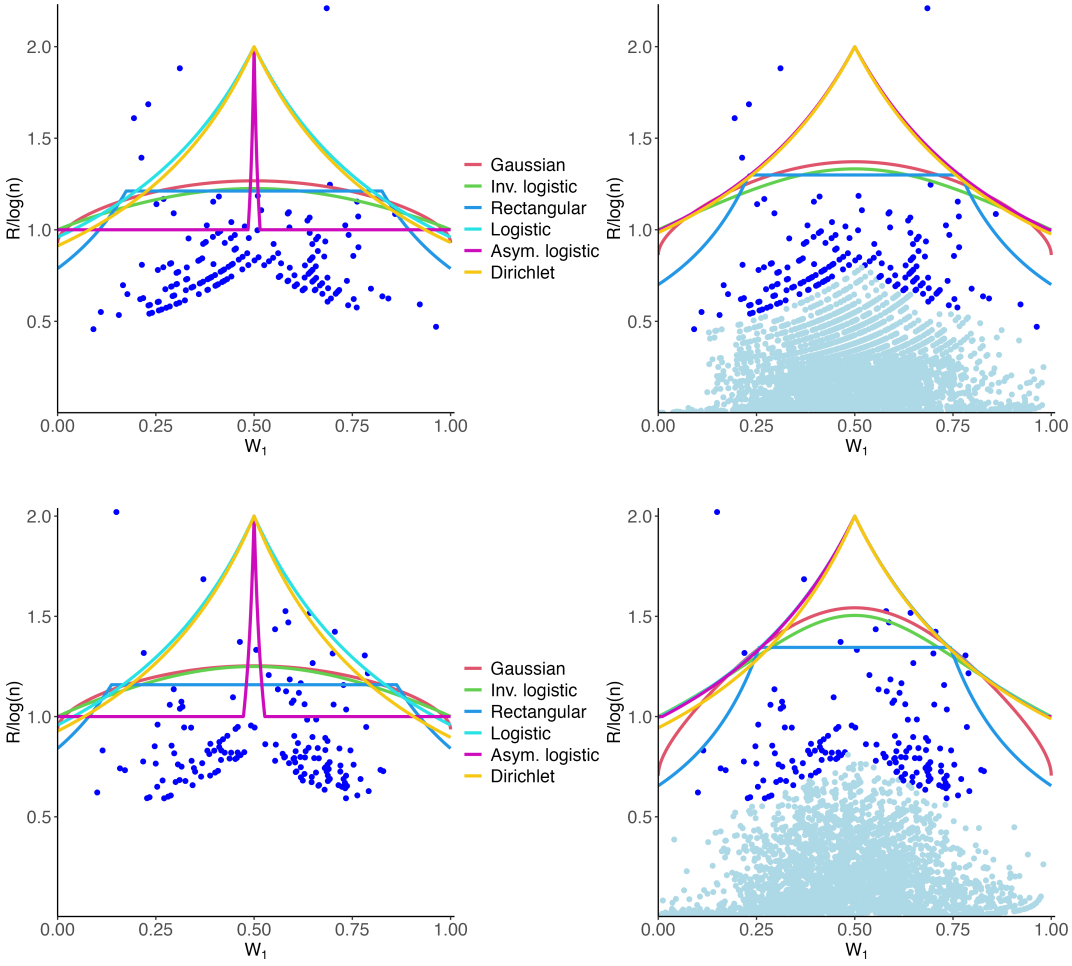


Figure 3.14: Posterior unit level sets of the six gauge functions fit to Friend Mountain (top) and Redstone (bottom), under the truncated (left) and censored (right) likelihoods.

3.6.4 Diagnostics

3.6.4.1 Radial density

We assess model fit for the truncated Gamma (tg) models in Equations (3.5) and (3.6) using probability-probability (PP) and quantile-quantile (QQ) plots, similar to Wadsworth and Campbell (2024) and Majumder et al. (2024). We define the fitted distribution function under scenario $\{l, a\}$:

$$\begin{aligned} F_{\text{tg},l,a}(r \mid r_\tau(\mathbf{w}), \mathbf{w}, \boldsymbol{\theta}_{1,\text{post}}, \alpha_{\text{post}}) &:= P_{l,a}(R \leq r \mid \mathbf{W} = \mathbf{w}, \boldsymbol{\theta}_{1,\text{post}}, \alpha_{\text{post}}) \\ &= 1 - \frac{\bar{F}_{l,a}(r; \alpha_{\text{post}}, g(\mathbf{w}; \boldsymbol{\theta}_{1,\text{post}}))}{\bar{F}_{l,a}(r_\tau(\mathbf{w}); \alpha_{\text{post}}, g(\mathbf{w}; \boldsymbol{\theta}_{1,\text{post}}))} \end{aligned}$$

For each of the n_0 observations with $R_i \geq r_\tau(\mathbf{w}_i)$, we compute:

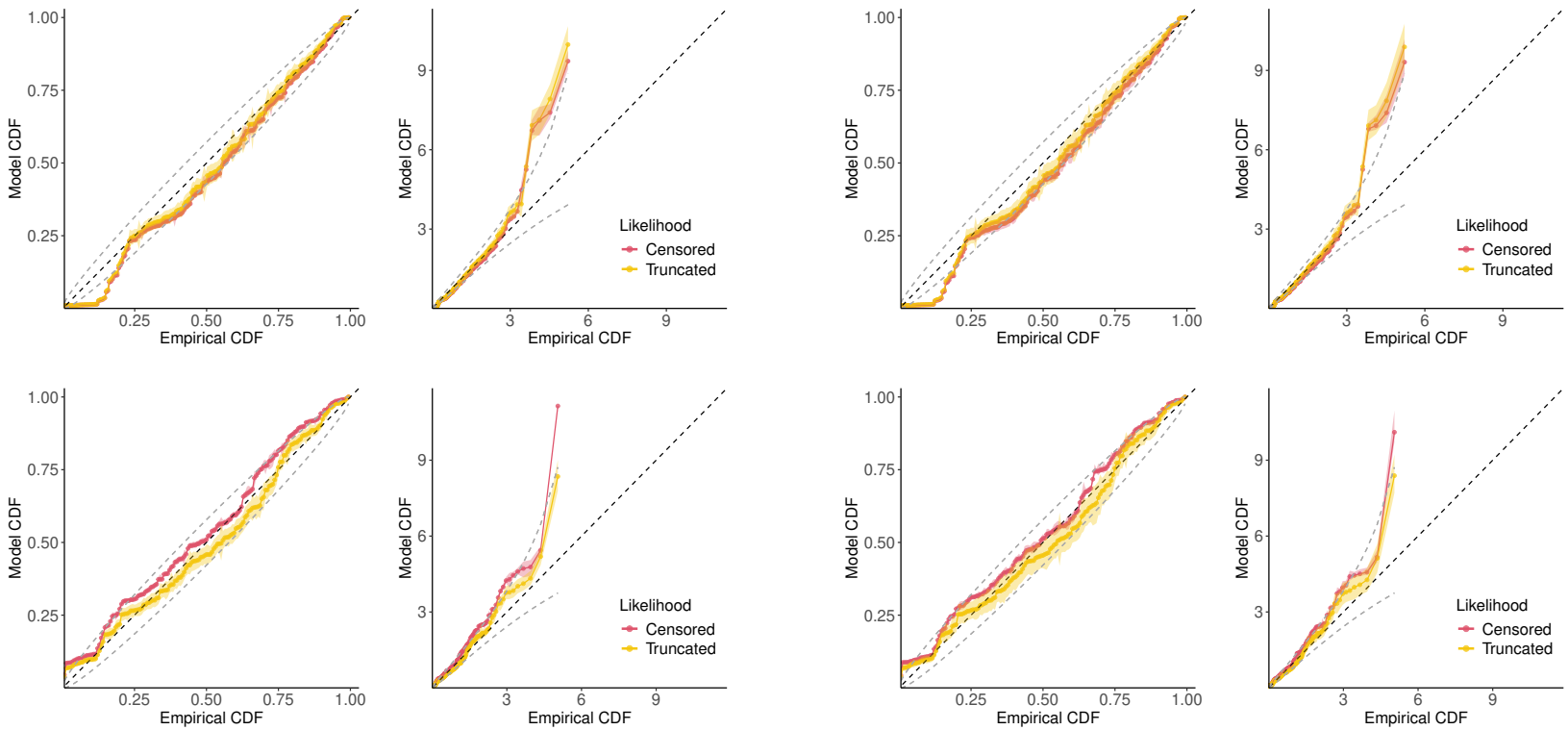
$$u_i = F_{\text{tg},l,a}(r_i \mid r_\tau(\mathbf{w}_i), \mathbf{w}_i, \boldsymbol{\theta}_{1,\text{post}}, \alpha_{\text{post}})$$

and compare it to the empirical estimate $\frac{i}{n_0+1}$ to generate the PP plot.

We compute u_i for all four combinations of radial and angular likelihoods, and construct Bayesian model-averaged diagnostics by weighting them accordingly. The resulting BMA-based u_i correspond to a true marginal posterior predictive distribution *only* when the angular density is held fixed, as is the case with the Beta mixture shared across the six parameterizations of the radial density. When angular densities vary across models—as is the case with the star-shaped densities—the BMA-weighted u_i do not correspond to a marginal posterior but can still be interpreted as a model-weighted diagnostic reflecting each model component's contribution to the joint density.

Figure 3.15 suggests an overall adequate fit across models, with no critical deviations from the diagonal. The only notable signs of lack of fit appear in the upper tail of the Friend Mountain data, where a few observations are heavier-tailed than expected when shown on exponential margins.

Another potential goodness-of-fit diagnostic is based on return-level curves (Papastathopoulos et al., 2025; Murphy-Barltrop et al., 2023; Simpson and Tawn, 2024b; Campbell and Wadsworth,



(a) True marginal posterior, constant angular density.

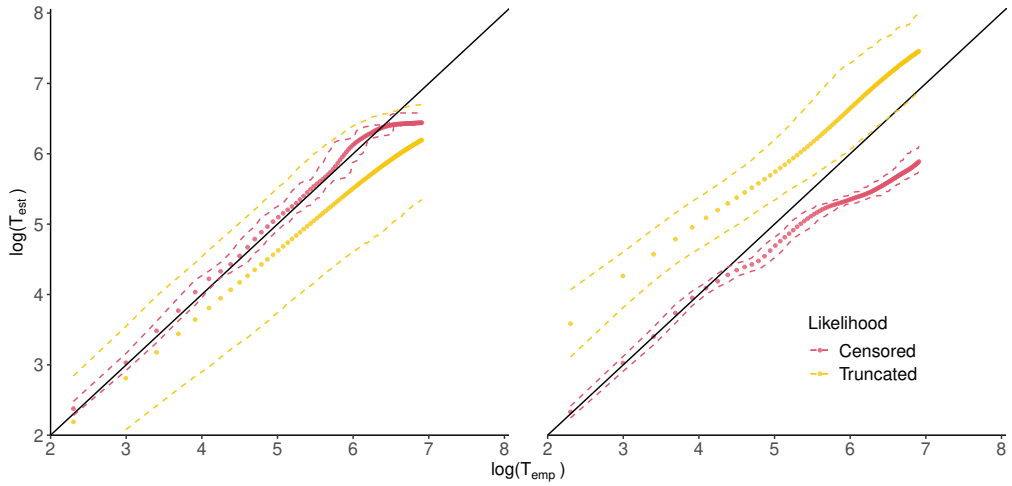
(b) Model-weighted diagnostic, varying star-shaped density.

Figure 3.15: PP and QQ plots of Friend Mountain (top) and Redstone (bottom) for the fitted truncated gamma distribution, with uniform margins in the first panel of each subfigure and exponential margins in the second panel. Colored ribbons around estimates represent the 95% credible interval, dotted grey lines represent the 2.5% and 97.5% confidence bands, and the black line is the identity.

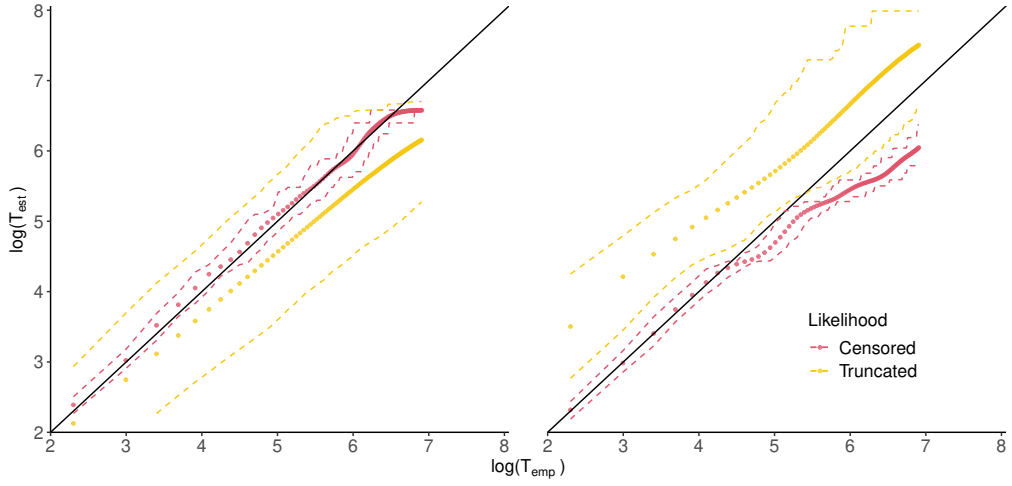
2024), which generalize univariate return levels to the multivariate setting. For a given return period T, Campbell and Wadsworth (2024) derive the return curve in the truncated gamma setting:

$$\mathcal{R}(T) = \{\mathbf{x} \in \mathbb{R}_+^d \,|\, \mathbf{x} = F_{\text{Ga}}^{-1}(1 - T^{-1}; d, g(\mathbf{w}, \boldsymbol{\theta})), \mathbf{w} \in \mathcal{S}_{d-1}\}.$$

Similarly, we compute return curves for 100 values of $T = \{10, 20, \ldots, 1000\}$, and compare the observed proportion of exceedances $\mathcal{R}(T)$ to the expected value of T^{-1}. As with the PP and QQ plots, we construct Bayesian model-averaged return curves, with analogous caveats and interpretations with respect to the angular densities.



(a) True estimated return curve, constant angular density



(b) Diagnostic return curves, varying star-shaped density

Figure 3.16: Estimated proportion of exceedances compared with the empirical for Friend Mountain (left) and Redstone (right). Dashed colored lines are the 95% credible intervals, and the identity is in black. Axes are on the log scale.

Figure 3.16 shows good agreement between Friend Mountain's estimated return-level sets and the empirical expected exceedances, with the censored likelihood outperforming the truncated. Redstone exhibits poorer performance, particularly under the truncated likelihood; the censored likelihood performs better, though not as well as Friend Mountain. However, the fit only begins to diverge after the 99th percentile ($\log(T) \approx 4.6$).

3.6.4.2 Angular density

The empirical distribution of angles can be compared to the fitted angular densities using density plots. As with the PP, QQ, and return-level plots, we construct Bayesian model-averaged diagnostics for the star-shaped density, with analogous caveats and interpretation. The form of the true marginal mixture density requires no additional manipulation. Figure 3.17 shows good agreement between the empirical and fitted angular distributions for both weather stations.

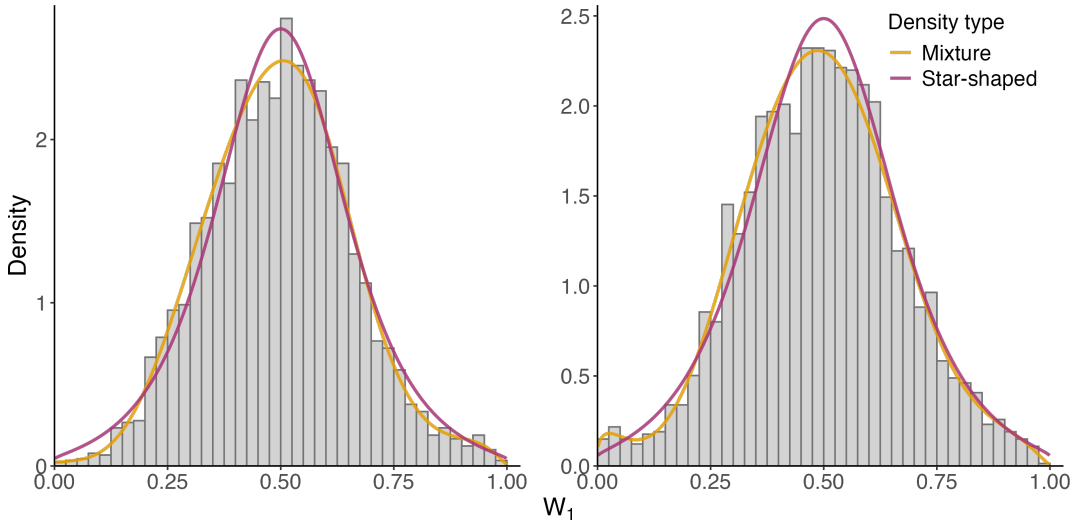


Figure 3.17: Empirical distribution of angles for Friend Mountain (left) and Redstone (right), with the marginal mixture density and the diagnostic model-weighted star-shaped density overlayed.

Finally, we assess Assumption 1, as we did in Section 3.4.2. While Figure 3.7 checks this assumption using the data-generating process, we now do so using the fitted radial models. For both weather stations and both likelihoods, we compute a gamma-based quantile threshold for each of the six gauge functions. We then plot a kernel density estimate of the angles that exceed

each threshold (Figure 3.18). Some discrepancies are observed between the fitted densities and the empirical angular distribution, particularly for Redstone. The deviations appear less pronounced for Friend Mountain, possibly due to the slightly higher number of observations exceeding the 95th percentile threshold. This diagnostic suggests we may be introducing some additional bias in our prediction procedures by relying on Assumption 1.

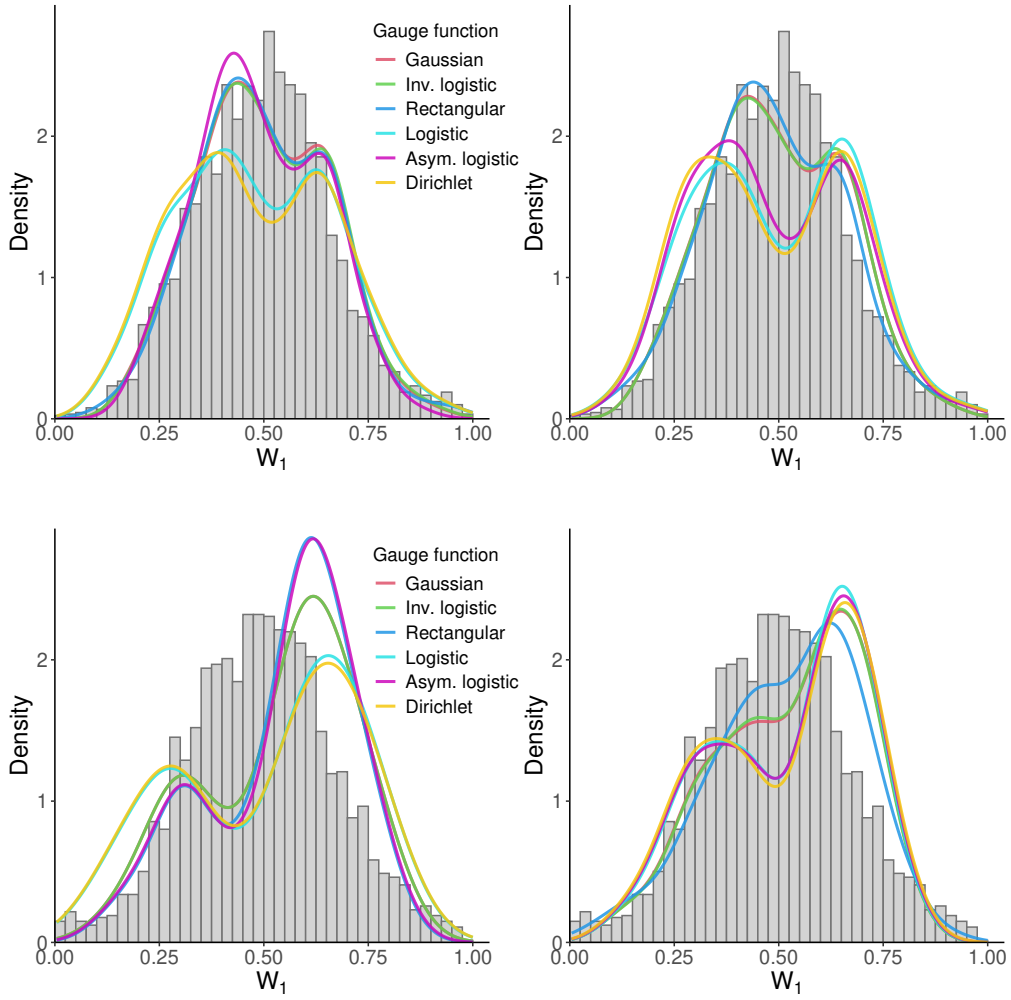


Figure 3.18: Histograms of the empirical distribution of angles for Friend Mountain (top) and Redstone (bottom). Within each location, the histogram remains the same across columns, as all angles were used in model fitting. Each column visualizes the density of $\mathbf{W}$ above a high threshold determined via gamma quantile estimation. Since there are six radial fits (one per gauge function), we obtain six thresholds and thus six resulting densities for $\mathbf{W}$. The left column corresponds to the truncated likelihood, and the right to the censored.

3.6.5 Predictions

We evaluate predictive performance in three distinct regions for both RAWS locations, corresponding to scenarios where one or both indices are extreme (Figure 3.19). Specifically:

- B_1 (teal): captures cases with extreme fire weather but less extreme burn potential
- B_2 (gold): corresponds to joint extremes in both fire weather and burn potential
- B_3 (pink): captures high burn potential with less extreme fire weather

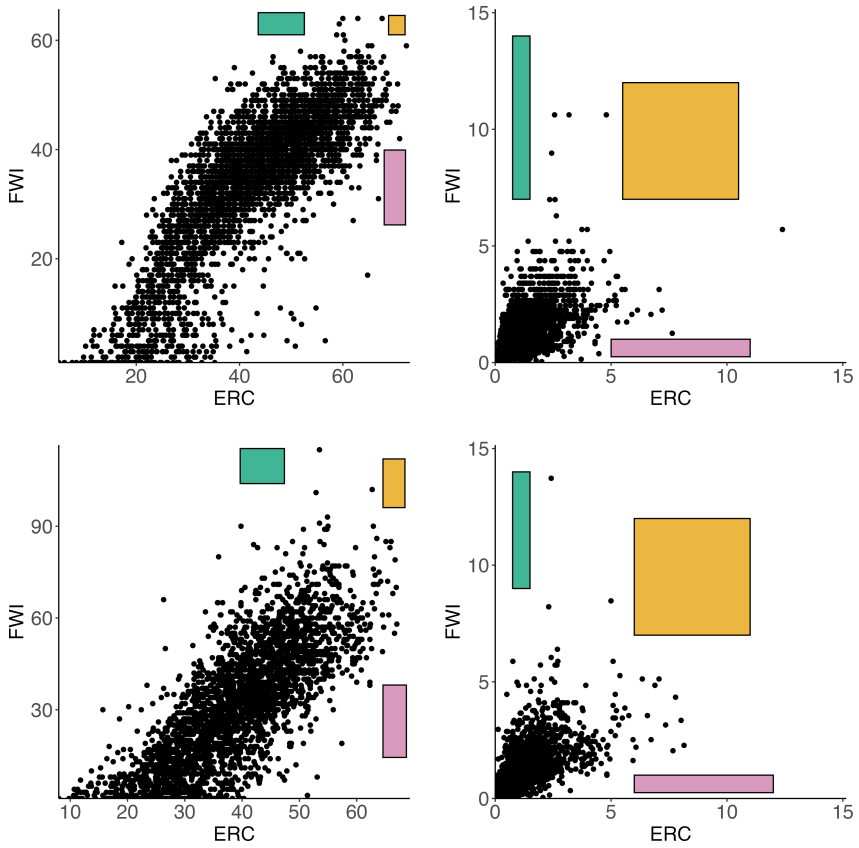


Figure 3.19: Illustration of the three sets we predict the probability of being in for Friend Mountain (top) and Redstone (bottom). The left column is in original margins, the right in exponential margins.

Notably, no observed point falls in any of these boxes, so using a classical approach (instead of an extreme value approach) would yield an empirical estimate of 0 for the probability of lying in the region.

We generate full posterior distributions of predicted probabilities for each region, under all combinations of likelihood, angular density, and BMA weighting method. Tables 3.5 and 3.6 report the median and 95% credible intervals. Given the simulation study results and consistently unreliable performance of the truncated likelihood across both weather stations, we recommend disregarding those predictions entirely. The remainder of our discussion focuses on results from the censored likelihood.

Table 3.5: Probability predictions for Friend Mountain RAWS in the three regions illustrated in the top row of Figure 3.19.

			Median (95% credible interval) $\times 10^5$		
Box	Likelihood	Ang. Dens.	Pseudo-BMA	Pseudo-BMA+	Stacking
B_1	Censored	mix	1.36 (0.702, 3.22)	1.56 (0.928, 3.29)	1.33 (0.676, 3.20)
		star	6.08 (3.59, 8.84)	6.08 (3.59, 8.84)	6.02 (3.58, 8.68)
	Truncated	mix	4.26 (1.39, 13.5)	4.65 (1.67, 13.4)	3.01 (0.259, 17.7)
		star	6.49 (1.14, 21.8)	6.53 (1.20, 22.0)	6.64 (1.30, 21.8)
B_2	Censored	mix	7.59 (5.34, 10.5)	7.69 (5.62, 10.4)	7.57 (5.30, 10.5)
		star	9.38 (4.52, 16.5)	9.38 (4.52, 16.5)	9.29 (4.60, 16.2)
	Truncated	mix	6.84 (2.63, 16.6)	8.24 (3.09, 20.5)	4.22 (0.583, 19.5)
		star	5.71 (1.46, 17.6)	5.67 (1.49, 17.5)	5.71 (1.49, 17.6)
B_3	Censored	mix	15.1 (8.91, 29.6)	17.2 (11.1, 30.8)	14.9 (8.65, 29.4)
		star	39.6 (24.0, 59.2)	39.6 (24.0, 59.2)	39.6 (24.4, 58.6)
	Truncated	mix	41.7 (15.3, 120.)	44.9 (18.7, 119.)	28.2 (3.36, 157.)
		star	45.8 (9.84, 132.)	46.4 (10.5, 134.)	47.1 (11.0, 134.)

There is general agreement between the star-shaped density and the mixture when predicting in B_2, which centers around $w_1 \approx 0.5$—where the empirical angular distribution is most concentrated. In contrast, predictions under the star-shaped density are consistently larger than those from the mixture in regions B_1 and B_3, which correspond to w_1 near 1 or 0, respectively, i.e., close to the axes on exponential margins. Specifically, star-shaped predictions are 5.6 and 2.6 times larger than mixture predictions for Redstone, and 4.5 and 2.6 for Friend Mountain, in B_1 and B_3, respectively.

Table 3.6: Probability predictions for Redstone RAWS in the three regions illustrated in the bottom row of Figure 3.19.

			Median (95% credible interval) $\times 10^5$		
Box	Likelihood	Ang. Dens.	Pseudo-BMA	Pseudo-BMA+	Stacking
B_1	Censored	mix	0.0543 (0.0258, 0.123)	0.0876 (0.0438, 0.165)	0.136 (0.0541, 0.274)
		star	0.306 (0.100, 0.613)	0.304 (0.100, 0.606)	0.254 (0.0907, 0.494)
	Truncated	mix	6.01 (2.45, 13.3)	6.14 (2.61, 13.0)	4.84 (0.397, 19.2)
		star	6.17 (1.38, 18.4)	6.60 (1.56, 18.2)	5.49 (0.631, 19.6)
B_2	Censored	mix	6.91 (5.08, 8.88)	7.92 (5.93, 10.2)	9.81 (6.92, 13.9)
		star	13.3 (8.06, 20.7)	13.2 (8.03, 20.6)	11.7 (7.60, 17.6)
	Truncated	mix	64.6 (25.3, 159.)	74.4 (29.3, 193.)	27.4 (7.72, 77.7)
		star	33.5 (11.8, 93.7)	33.2 (12.6, 88.1)	34.2 (9.07, 107.)
B_3	Censored	mix	2.67 (1.50, 5.25)	3.50 (2.05, 5.98)	4.78 (2.38, 8.35)
		star	7.01 (2.97, 12.4)	6.96 (2.96, 12.3)	6.10 (2.88, 10.3)
	Truncated	mix	101. (50.2, 185.)	102. (53.7, 182.)	83.4 (14.2, 234.)
		star	89.6 (26.3, 214.)	93.1 (29.4, 210.)	84.2 (14.4, 228.)

However, these differences are actually on the scale of 10^{-5}, so in terms of the original magnitude, the predictions are still similar.

3.7 Discussion

Using BMA, we have introduced flexibility into a parametric modeling framework, providing a viable alternative to more complex semi-parametric approaches. BMA also allowed us to bypass the uncertainty and potential bias associated with model selection. Incorporating the censored likelihood led to superior predictive performance in nearly all simulation settings—and consistently better results with real data—highlighting the value of the censored approach in small sample size scenarios. Yet the increase in bias that potentially results from the censored approach should be considered in the context of the application. While we were initially concerned that the star-shaped angular density may be overly rigid, it performed quite well, likely due to BMA.

We were somewhat surprised by the poorer performance of stacking weights, especially under the Hüsler-Reiss dependence structure setting, since stacking should work well when the data-

generating process is not included in the suite of models being averaged over. While stacking is asymptotically optimal when the data-generating model is not under consideration, this advantage may not hold in finite sample scenarios such as ours.

Extensions to higher dimensions present some challenges within our framework. Each increase in dimension introduces more parameters for the gauge functions; as an extreme example, the asymmetric logistic model has 128 possible forms in three dimensions. The mixture angular model would generalize to a Dirichlet mixture, but the effectiveness of the stick-breaking process in higher dimensions is unclear. Despite these concerns, we plan to explore this framework in three dimensions.

Assumption 1 may break down in small sample sizes. One way to address this would be to fit angular densities only to angles exceeding a high radial threshold. If a marginal threshold is used initially, this approach would require fitting the radial model to then identify the quantile-based radial threshold $r_q(\mathbf{w})$, above which the associated angles are used to fit the angular model.

Currently, the marginal transformation, radial likelihoods, and angular likelihoods are modeled independently, even though the latter two depend on the transformation. Future work could focus on a hierarchical model that jointly incorporates all three components, allowing for information sharing across model components and accounting for uncertainty propagated forward by the marginal transformation.

Finally, although we attempted to account for the varying exceedance ratios (Figure 3.4) with self-normalized importance sampling, the BMA framework itself suggests another possibility. Yao et al. (2022) propose hierarchical stacking weights that vary with covariates. In our framework, this would entail allowing the stacking weights to depend on the angular values. Implementation requires discretization of the angular space, perhaps with groupings informed by the exceedance ratios. The hierarchical weights are then estimated in a Bayesian model using the discretized angles and the leave-one-out likelihoods as inputs. Though this approach is still new to us, it presents a promising direction for future research.

Chapter 4

Conclusions and future work

In this dissertation, we explored both univariate and multivariate approaches to modeling wildfire extremes. We note that in both methods, the covariates and factors we considered are not exhaustive in terms of their ability to model and predict wildfire behavior. Collaboration with fire scientists and environmental engineers could make these models and results more robust, and an interesting avenue for future work.

In Chapter 2, we implemented an extension of the GPD to model the full distribution of wildfire burned areas, recognizing that while extreme wildfires can have catastrophic impacts, moderately sized wildfires also carry significant societal and environmental consequences. Our approach does not require threshold selection, as is typical in standard extreme value analysis, allowing us to use information from all wildfire burned areas to inform the shape and scale of the distribution. We integrated this burned area model into a joint Bayesian hierarchical model, linking the burned areas to wildfire occurrences via a temporally varying random effect. Within the spatiotemporally varying parameters of the burned areas and occurrences sub-models, we incorporated nonlinear effects in the covariates and accounted for any residual spatio-temporal variation with a Gaussian random effect. Our model shared information across space using an easily interpretable random effect that partitions the spatial domain by hierarchical ecological features. We demonstrated the utility of FWI and ERC as effective substitutes for more complicated climate covariates in predicting wildfire burned areas.

While computationally prohibitive, it may be valuable to recalculate ERC using localized forest fuel models prior to spatial aggregation, to better align with regional variation in fuel availability. We found potential overestimation of ξ in some ecoregions and suspect it may be poorly estimated due to the EGPD model structure, suggesting a need for further exploration.

In Chapter 3, we focused on developing a method that builds upon recent work in the geometric framework of multivariate extremes and apply the method to weather station-level ERC and FWI

observations. By incorporating BMA into our tail risk probability predictions, we added flexibility while maintaining the simplicity inherent in the parametric approach, avoiding the bias and uncertainty that arise when selecting a single model. The BMA-weighted predictions outperformed the flexible semi-parametric models. The censored likelihood approach appeared to offer an advantageous bias-variance tradeoff in nearly all simulation scenarios, as well as in our real data analyses. By using an extremes-based approach, we could extrapolate beyond the available ERC and FWI data and predict various scenarios involving joint extremes or when one variable is extreme and the other is less so.

We are considering a few different elements to improve our method. More immediately, we plan to extend the method from two dimensions to three dimensions. The increased dimension may add some additional computational complexity given the increased parameter space of the asymmetric logistic gauge function, along with that of the Dirichlet mixture—the higher-dimensional analogue of the Beta mixture—used in modeling the angular density. We also would like to incorporate the marginal transformation into a fully joint hierarchical model, thereby appropriately accounting for and propagating any uncertainty arising from the transformation in the radial and angular modeling. This hierarchical model would also link the radial and angular components, no longer making them independent. Regarding the marginal transformation, it would also be valuable to develop a robust method that accounts for scenarios where the data have alternative shapes or highly concentrated values. We encountered some difficulties when applying our methods to other weather station locations, due to the varying shapes of the ERC and FWI marginal distributions.

Bibliography

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1):121–131.

Abatzoglou, J. T. and Kolden, C. A. (2013). Relationships between climate and macroscale area burned in the Western United States. *International Journal of Wildland Fire*, 22:1003–1020.

Abatzoglou, J. T. and Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences*, 113(42):11770–11775.

Apputhurai, P. and Stephenson, A. G. (2011). Accounting for uncertainty in extremal dependence modeling using Bayesian model averaging techniques. *Journal of Statistical Planning and Inference*, 141(5):1800–1807.

Balkema, A. A., Embrechts, P., and Nolde, N. (2010). Meta densities and the shape of their sample clouds. *Journal of Multivariate Analysis*, 101(7):1738–1754.

Balkema, G. and Nolde, N. (2010). Asymptotic independence for unimodal densities. *Advances in Applied Probability*, 42(2):411–432.

Balkema, G. and Nolde, N. (2012). Asymptotic Dependence for Light-Tailed Homothetic Densities. *Advances in Applied Probability*, 44(2):506–527.

Barbero, R., Abatzoglou, J. T., Pimont, F., Ruffault, J., and Curt, T. (2020). Attributing increases in fire weather to anthropogenic climate change over France. *Front. Earth Sci.*, 8.

Barbero, R., Abatzoglou, J. T., Steel, E. A., and Larkin, N. K. (2014). Modeling very large-fire occurrences over the continental United States from weather and climate forcing. *Environ. Res. Lett.*, 9(12).

Becker, D. G., Woolford, D. G., and Dean, C. B. (2022). Assessing dependence between frequency and severity through shared random effects. *PLoS ONE*, 17(8):e0271904.

Behrens, C. N., Lopes, H. F., and Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Stat. Modell.*, 4(3):227–244.

Beirlant, J., editor (2004). *Statistics of extremes: theory and applications*. Wiley series in probability and statistics. Wiley, Hoboken, NJ.

Bradshaw, L. and McCormick, E. (2000). *FireFamily Plus user's guide, Version 2.0*.

Bradshaw, L. S., Deeming, J. E., Burgan, R. E., and Cohen, J. D. (1984). The 1978 National Fire-Danger Rating System: technical documentation. Technical Report INT-GTR-169, U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, UT.

Brillinger, D. R., Preisler, H. K., and Benoit, J. W. (2003). Risk assessment: A forest fire example. In *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, pages 177–196. Institute of Mathematical Statistics, Beachwood, OH.

Campbell, R. and Wadsworth, J. (2024). Piecewise-linear modeling of multivariate geometric extremes.

Carreau, J. and Bengio, Y. (2009). A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes*, 12(1):53–76.

Castel-Clavera, J., Pimont, F., Opitz, T., Ruffault, J., Rivière, M., Dupuy, J.-L., Castel-Clavera, J., Pimont, F., Opitz, T., Ruffault, J., Rivière, M., and Dupuy, J.-L. (2022). Disentangling the factors of spatio-temporal patterns of wildfire activity in south-eastern France. *International Journal of Wildland Fire*, 32(1):15–28.

Castro-Camilo, D., Huser, R., and Rue, H. (2019). A spliced gamma-generalized Pareto model for short-term extreme wind speed probabilistic forecasting. *J. Agric. Biol. Environ. Stat.*, 24(3):517–534.

Cisneros, D., Gong, Y., Yadav, R., Hazra, A., and Huser, R. (2023). A combined statistical and machine learning approach for spatial prediction of extreme wildfire frequencies and sizes. *Extremes*, 26(2):301–330.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer, London.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence Measures for Extreme Value Analyses. *Extremes*, 2:339–365.

Davis, R. A., Mulrow, E., and Resnick, S. I. (1988). Almost sure limit sets of random samples in $\mathbb{R}^d$. *Advances in Applied Probability*, 20(3):573–599.

de Carvalho, M., Pereira, S., Pereira, P., and de Zea Bermudez, P. (2022). An extreme value Bayesian lasso for the conditional left and right tails. *J. Agric. Biol. Environ. Stat.*, 27(2):222–239.

Dennison, P. E., Brewer, S. C., Arnold, J. D., and Moritz, M. A. (2014). Large wildfire trends in the western United States, 1984–2011. *Geophys. Res. Lett.*, 41(8):2928–2933.

do Nascimento, F. F., Gamerman, D., and Lopes, H. F. (2012). A semiparametric Bayesian approach to extreme value estimation. *Stat. Comput.*, 22(2):661–675.

Duane, A., Castellnou, M., and Brotons, L. (2021). Towards a comprehensive look at global drivers of novel extreme wildfire events. *Clim. Change*, 165(3):43.

Eidenshink, J., Schwind, B., Brewer, K., Zhu, Z.-L., Quayle, B., and Howard, S. (2007). A project for monitoring trends in burn severity. *Fire Ecol.*, 3(1):3–21.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Stat. Sci.*, 11(2):89–121.

Fosberg, M. A. (1978). Weather in wildland fire management: the fire weather index. *US For Serv Reprints of articles by FS employees*.

Frigessi, A., Haug, O., and Rue, H. (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5(3):219–235.

Gabry, J., Češnovar, R., and Johnson, A. (2023). *cmdstanr: R Interface to 'CmdStan'*.

Gamet, P. and Jalbert, J. (2022). A flexible extended generalized Pareto distribution for tail estimation. *Environmetrics*, 33(6).

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.*, 102(477):359–378.

Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *J. Bus. Econom. Statist.*, 29(3):411–422.

Goss, M., Swain, D. L., Abatzoglou, J. T., Sarhadi, A., Kolden, C. A., Williams, A. P., and Diffenbaugh, N. S. (2020). Climate change is increasing the likelihood of extreme autumn wildfire conditions across California. *Environ. Res. Lett.*, 15(9).

Hawbaker, T. J., Radeloff, V. C., Stewart, S. I., Hammer, R. B., Keuler, N. S., and Clayton, M. K. (2013). Human and biophysical influences on fire occurrence in the United States. *Ecol. Appl.*, 23(3):565–582.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14(4):382–401.

Huser, R., Opitz, T., and Wadsworth, J. L. (2025). Modeling of spatial extremes in environmental data science: time to move away from max-stable processes. *Environmental Data Science*, 4:e3.

Iglesias, V., Stavros, N., Balch, J. K., Barrett, K., Cobian-Iñiguez, J., Hester, C., Kolden, C. A., Leyk, S., Nagy, R. C., Reid, C. E., Wiedinmyer, C., Woolner, E., and Travis, W. R. (2022). Fires that matter: Reconceptualizing fire risk to include interactions between humans and the natural environment. *Environ. Res. Lett.*, 17(4).

Ishwaran, H. and James, L. F. (2003). Some Further Developments for Stick-Breaking Priors: Finite and Infinite Clustering and Classification. *Sankhyā: The Indian Journal of Statistics (2003-2007)*, 65(3):577–592.

Jones, M. W., Abatzoglou, J. T., Veraverbeke, S., Andela, N., Lasslop, G., Forkel, M., Smith, A. J. P., Burton, C., Betts, R. A., van der Werf, G. R., Sitch, S., Canadell, J. G., Santín, C., Kolden, C., Doerr, S. H., and Le Quéré, C. (2022). Global and regional trends and drivers of fire under climate change. *Rev. Geophys.*, 60(3).

Joseph, M. B., Rossi, M. W., Mietkiewicz, N. P., Mahood, A. L., Cattau, M. E., St. Denis, L. A., Nagy, R. C., Iglesias, V., Abatzoglou, J. T., and Balch, J. K. (2019). Spatiotemporal prediction of wildfire size extremes with Bayesian finite sample maxima. *Ecol. Appl.*, 29(6).

Kinoshita, K. and Resnick, S. I. (1991). Convergence of Scaled Random Samples in Rd. *The Annals of Probability*, 19(4):1640–1663.

Koh, J., Pimont, F., Dupuy, J.-L., and Opitz, T. (2023). Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *Ann. Appl. Stat.*, 17(1):560–582.

Kottas, A. (2006). Dirichlet Process Mixtures of Beta Distributions, with Applications to Density and Intensity Estimation. *Proceedings of the Workshop on Learning with Nonparametric Bayesian Methods, 23rd*.

Krock, M., Bessac, J., Stein, M. L., and Monahan, A. H. (2022). Nonstationary seasonal model for daily mean temperature distribution bridging bulk and tails. *Weather Clim. Extremes*, 36.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *J. Comput. Graph. Statist.*, 13(1):183–212.

Lawler, E. S. and Shaby, B. A. (2024). Anthropogenic and meteorological effects on the counts and sizes of moderate and extreme wildfires. *Environmetrics*, 35(7):e2873.

Littell, J. S., McKenzie, D., Peterson, D. L., and Westerling, A. L. (2009). Climate and wildfire area burned in western U.S. ecoprovinces, 1916–2003. *Ecol. Appl.*, 19(4):1003–1021.

Liu, Y., Stanturf, J., and Goodrick, S. (2010). Trends in global wildfire potential in a changing climate. *For. Ecol. Manage.*, 259(4):685–697.

MacDonald, A., Scarrott, C., Lee, D., Darlow, B., Reale, M., and Russell, G. (2011). A flexible extreme value mixture model. *Comput. Stat. Data Anal.*, 55(6):2137–2157.

Mackay, E. and Jonathan, P. (2024). Modelling multivariate extremes through angular-radial decomposition of the density function.

Madigan, D. and Raftery, A. E. (1994). Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window. *Journal of the American Statistical Association*, 89(428):1535–1546.

Majumder, R., Shaby, B. A., Reich, B. J., and Cooley, D. (2024). Semiparametric Estimation of the Shape of the Limiting Bivariate Point Cloud.

Malamud, B. D., Millington, J. D. A., and Perry, G. L. W. (2005). Characterizing wildfire regimes in the United States. *Proc. Natl. Acad. Sci. U.S.A.*, 102(13):4694–4699.

Martinuzzi, S., Allstadt, A. J., Pidgeon, A. M., Flather, C. H., Jolly, W. M., and Radeloff, V. C. (2019). Future changes in fire weather, spring droughts, and false springs across U.S. National Forests and Grasslands. *Ecol. Appl.*, 29(5).

Matheson, J. E. and Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Manage. Sci.*, 22(10):1087–1096.

Mockrin, M. H., Radeloff, V. C., and Helmers, D. P. (2023). Census block level housing change 1990 - 2020 for the conterminous United States.

Murphy-Barltrop, C. J. R., Mackay, E., and Jonathan, P. (2024). Inference for bivariate extremes via a semi-parametric angular-radial model. *Extremes*.

Murphy-Barltrop, C. J. R., Wadsworth, J. L., and Eastoe, E. F. (2023). New estimation methods for extremal bivariate return curves. *Environmetrics*, 34(5):e2797.

Naveau, P., Huser, R., Ribereau, P., and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52(4):2753–2769.

Nolde, N. (2014). Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95.

Nolde, N. and Wadsworth, J. L. (2022). Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*, 54(3):688–717.

Omernik, J. M. and Griffith, G. E. (2014). Ecoregions of the conterminous United States: Evolution of a hierarchical spatial framework. *Environ. Manage.*, 54(6):1249–1266.

Opitz, T., Bonneu, F., and Gabriel, E. (2020). Point-process based Bayesian modeling of space–time structures of forest fire occurrences in Mediterranean France. *Spatial Stat.*, 40.

Owen, A. B. (2013). *Monte Carlo theory, methods and examples*. `https://artowen.su.domains/mc/`.

Papastathopoulos, I., de Monte, L., Campbell, R., and Rue, H. (2025). Statistical inference for radial generalized Pareto distributions and return sets in geometric extremes.

Papastathopoulos, I. and Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *J. Statist. Plann. Inference*, 143(1):131–143.

Parente, J., Pereira, M. G., Amraoui, M., and Fischer, E. M. (2018). Heat waves in Portugal: Current regime, changes in future climate and impacts on extreme wildfires. *Sci. Total Environ*, 631–632:534–549.

Pausas, J. G. and Keeley, J. E. (2021). Wildfires and global change. *Front. Ecol. Environ.*, 19(7):387–395.

Pimont, F., Fargeon, H., Opitz, T., Ruffault, J., Barbero, R., Martin-StPaul, N., Rigolot, E., RiviÉre, M., and Dupuy, J.-L. (2021). Prediction of regional wildfire activity in the probabilistic Bayesian framework of Firelihood. *Ecological Applications*, 31(5).

Pimont, F., Ruffault, J., Opitz, T., Fargeon, H., Barbero, R., Castel-Clavera, J., Martin-StPaul, N., Rigolot, E., and Dupuy, J.-L. (2022). Future expansion, seasonal lengthening and intensification of fire activity under climate change in southeastern France. *Int. J. Wildland Fire*, 32(1):4–14.

Preisler, H., Brillinger, D. R., Burgan, R. E., and Benoit, J. (2004). Probability based models for estimation of wildfire risk. *Int. J. Wildland Fire*, 13(2):133–142.

Preisler, H. K., Chen, S.-C., Fujioka, F., Benoit, J. W., and Westerling, A. L. (2008). Wildland fire probabilities estimated from weather model-deduced monthly mean fire danger indices. *Int. J. Wildland Fire*, 17(3):305–316.

Preisler, H. K. and Westerling, A. L. (2007). Statistical model for forecasting monthly large wildfire events in western United States. *J. Appl. Meteorol. Climatol.*, 46(7):1020–1030.

Richards, J. and Huser, R. (2022). Regression modelling of spatiotemporal extreme U.S. wildfires via partially-interpretable neural networks.

Sabourin, A., Naveau, P., and Fougères, A.-L. (2013). Bayesian model averaging for multivariate extremes. *Extremes*, 16(3):325–350.

Shaby, B. and Wells, M. (2010). Exploring an Adaptive Metropolis Algorithm.

Simpson, E. S. and Tawn, J. A. (2024a). Estimating the limiting shape of bivariate scaled sample clouds: with additional benefits of self-consistent inference for existing extremal dependence properties.

Simpson, E. S. and Tawn, J. A. (2024b). Inference for New Environmental Contours Using Extreme Value Analysis. *Journal of Agricultural, Biological and Environmental Statistics*.

Simpson, E. S., Wadsworth, J. L., and Tawn, J. A. (2021). A geometric investigation into the tail dependence of vine copulas. *Journal of Multivariate Analysis*, 184:104736.

Stan Development Team (2023). *Stan Modeling Language Users Guide and Reference Manual*.

Stavros, E. N., Abatzoglou, J. T., McKenzie, D., and Larkin, N. K. (2014). Regional projections of the likelihood of very large wildland fires under a changing climate in the contiguous western United States. *Clim. Change*, 126(3):455–468.

Stein, M. L. (2021a). A parametric model for distributions with flexible behavior in both tails. *Environmetrics*, 32(2).

Stein, M. L. (2021b). Parametric models for distributions when interest is in extremes with an application to daily temperature. *Extremes*, 24(2):293–323.

Strauss, D., Bednar, L., and Mees, R. (1989). Do one percent of the forest fires cause ninety-nine percent of the damage? *For. Sci.*, 35(2):319–328.

Tancredi, A., Anderson, C., and O'Hagan, A. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes*, 9(2):87–106.

Tawn, J. A. (1990). Modelling Multivariate Extreme Value Distributions. *Biometrika*, 77(2):245–253.

Tencaliec, P., Favre, A.-C., Naveau, P., Prieur, C., and Nicolet, G. (2020). Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics*, 31(2).

Turner, J. and Lawson, B. (1978). *Weather in the Canadian Forest Fire Danger Rating System: A user guide to national standards and practices*. Fisheries and Environment Canada, Canadian Forest Service, Pacific Forest Research Centre, Victoria, British Columbia, Canada.

Van Oldenborgh, G. J., Krikken, F., Lewis, S., Leach, N. J., Lehner, F., Saunders, K. R., Van Weele, M., Haustein, K., Li, S., Wallom, D., Sparrow, S., Arrighi, J., Singh, R. K., Van Aalst, M. K., Philip, S. Y., Vautard, R., and Otto, F. E. L. (2021). Attribution of the Australian bushfire risk to anthropogenic climate change. *Nat. Hazards Earth Syst. Sci.*, 21(3):941–960.

Vettori, S., Huser, R., Segers, J., and Genton, M. G. (2020). Bayesian Model Averaging Over Tree-based Dependence Structures for Multivariate Extremes. *Journal of Computational and Graphical Statistics*, 29(1):174–190.

Wadsworth, J. L. and Campbell, R. (2024). Statistical inference for multivariate extremes via a geometric approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkae030.

Wang, X., Wotton, B. M., Cantin, A. S., Parisien, M.-A., Anderson, K., Moore, B., and Flannigan, M. D. (2017). cffdrs: an r package for the canadian forest fire danger rating system. *Ecological Processes*, 6(1):1–11.

Williams, A. P., Abatzoglou, J. T., Gershunov, A., Guzman-Morales, J., Bishop, D. A., Balch, J. K., and Lettenmaier, D. P. (2019). Observed impacts of anthropogenic climate change on wildfire in California. *Earth's Future*, 7(8):892–910.

Woolford, D. G., Bellhouse, D. R., Braun, W. J., Dean, C. B., Martell, D. L., and Sun, J. (2011). A spatio-temporal model for people-caused forest fire occurrence in the Romeo Malette Forest. *J. Environ. Stat.*, 2(1):2–25.

Woolford, D. G., Martell, D. L., McFayden, C. B., Evens, J., Stacey, A., Wotton, B. M., and Boychuk, D. (2021). The development and implementation of a human-caused wildland fire occurrence prediction system for the province of Ontario, Canada. *Can. J. For. Res.*, 51(2):303–325.

World Weather Attribution (2025). Climate change increased the likelihood of wildfire disaster in highly exposed los angeles area. Technical report, World Weather Attribution. Accessed: 2025-04-22.

Yadav, R., Huser, R., Opitz, T., and Lombardo, L. (2023). Joint modelling of landslide counts and sizes using spatial marked point processes with sub-asymptotic mark distributions. *J. R. Stat. Soc. Ser. C Appl. Stat.*, page qlad077.

Yao, Y., Pirš, G., Vehtari, A., and Gelman, A. (2022). Bayesian hierarchical stacking: Some models are (somewhere) useful. *Bayesian Analysis*, 17(4).

Yao, Y., Vehtari, A., Simpson, D., and Gelman, A. (2018). Using Stacking to Average Bayesian Predictive Distributions (with Discussion). *Bayesian Analysis*, 13(3):917–1007.

Appendix A

Supplement to *Bayesian model averaging of risk set probabilities using a geometric representation of multivariate extremes*

A.1 Asymptotic Gamma approximation

As described in Section 3.2.2.1, we conducted an experiment to determine how well the gamma approximation performs in finite settings. Recall that data were generated according to Section 3.5 and fit using the data-generating gauge function, looking at the posterior estimates of the unknown dependence parameter compared with the truth. For ease of reference, we again provide the two figures we showed previously, in this supplement, along with the remaining figures. See Section 3.2.2.1 for interpretation.

A.1.1 Gaussian joint density

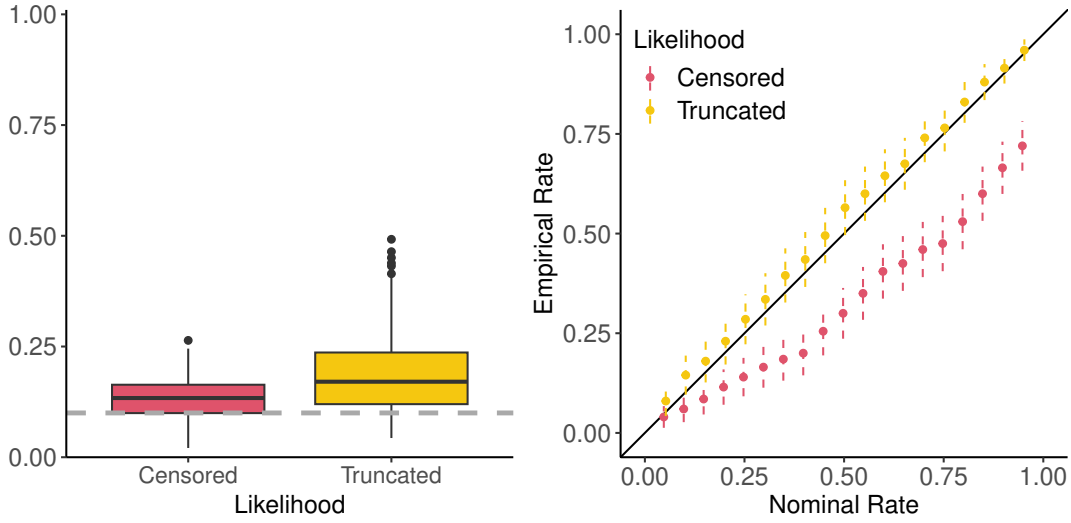


Figure A1: Gaussian dependence, $\rho = 0.1$. Grey dotted line is the true parameter value.

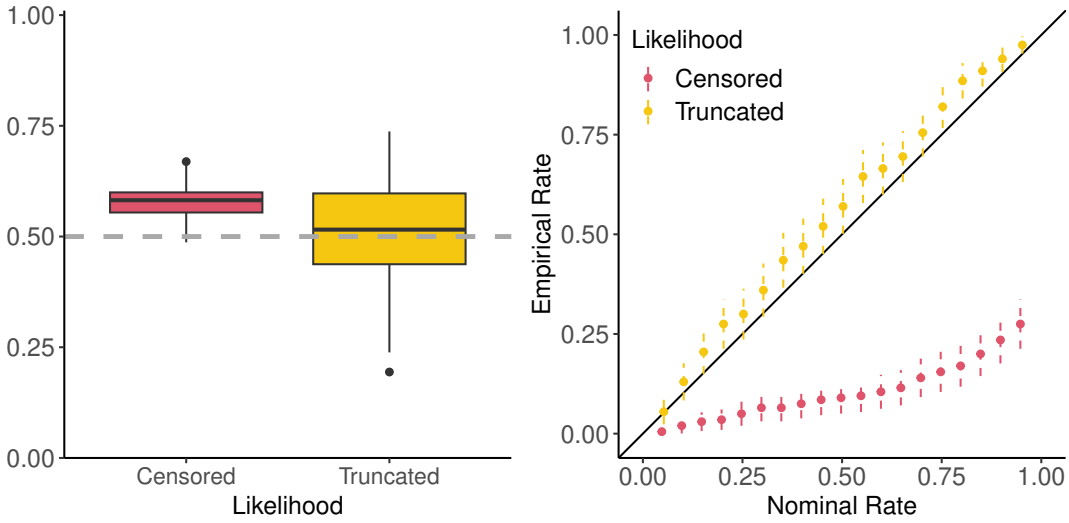


Figure A2: Gaussian dependence, $\rho = 0.5$. This plot is the same as Figure 3.2, but repeated here for convenience.

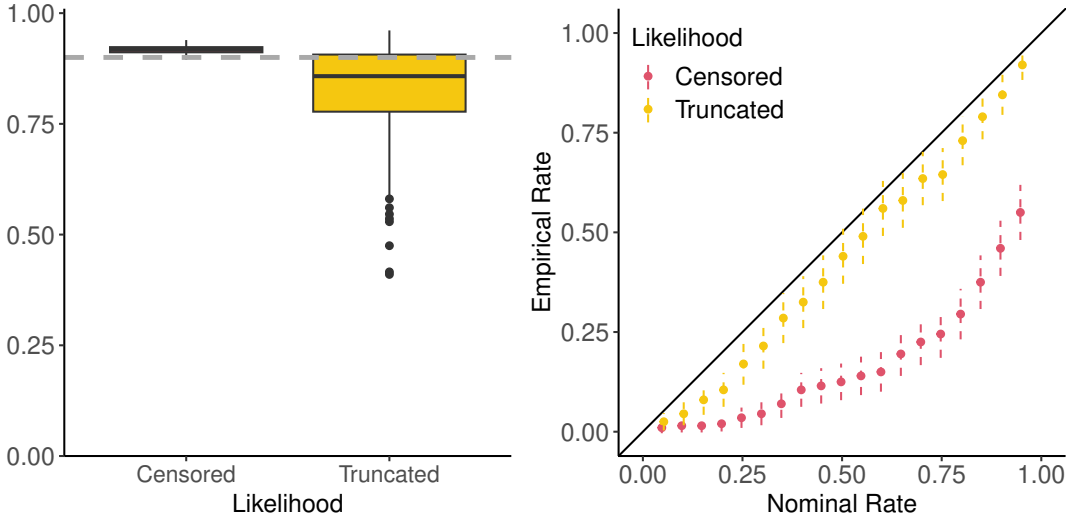


Figure A3: Gaussian dependence, $\rho = 0.9$.

A.1.2 Logistic joint density

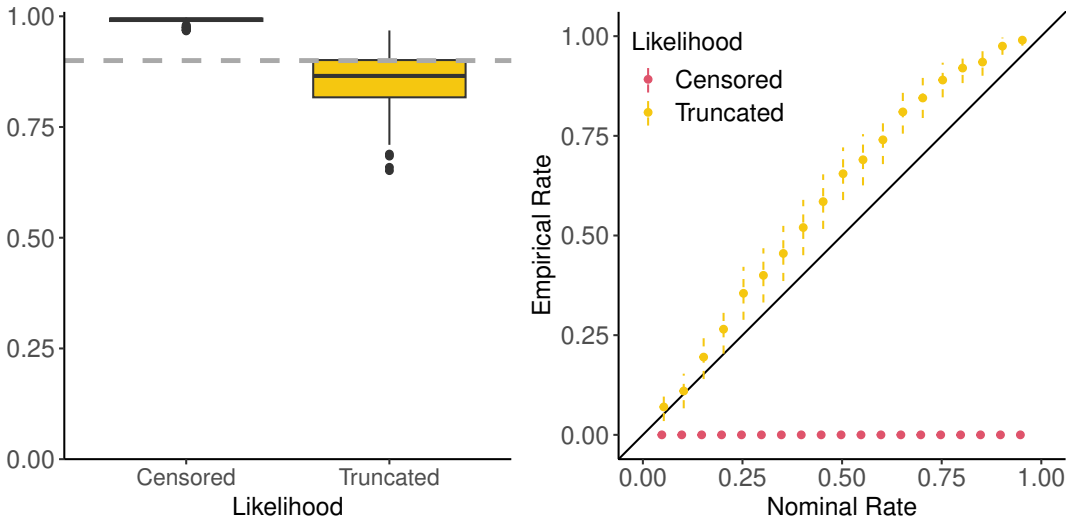


Figure A4: Logistic dependence, $\rho = 0.1$. Grey dotted line is the true parameter value.

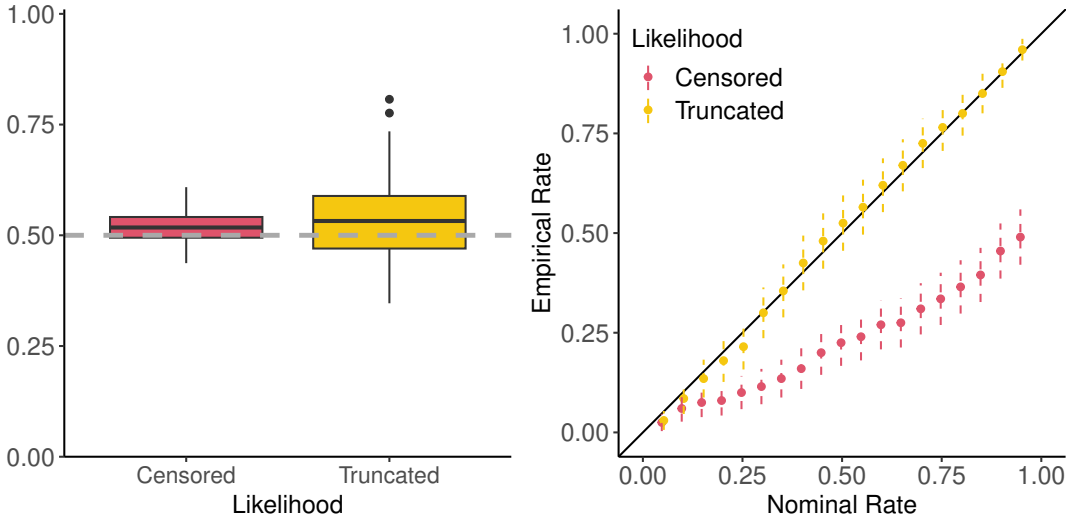


Figure A5: Logistic dependence, $\rho = 0.5$. This plot is the same as Figure 3.3, but repeated here for convenience.

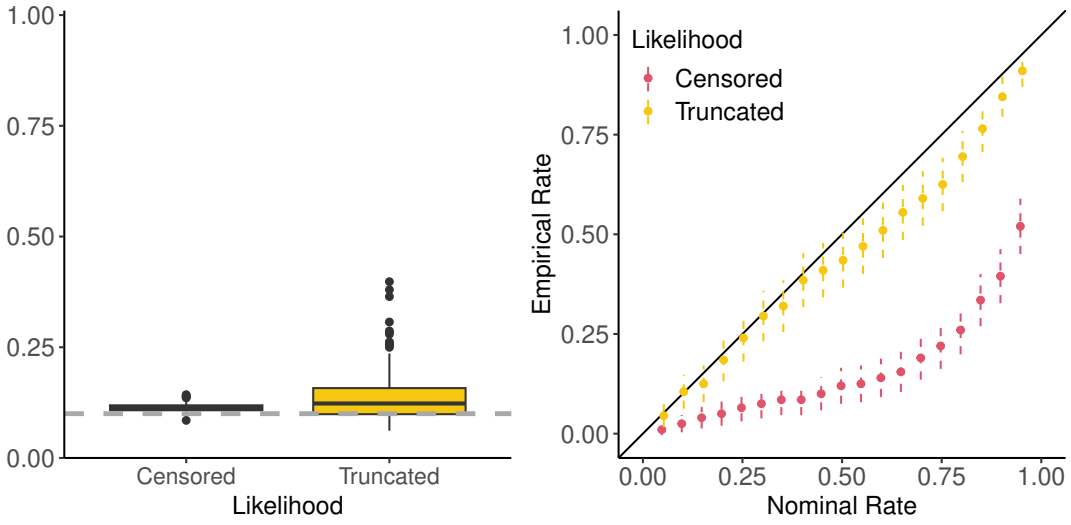


Figure A6: Logistic dependence, $\rho = 0.9$.

A.2 Simulation study results

In Section 3.5, we outline the dependence structures and levels we examine and give a visualization of the Gaussian and logistic datasets (Figure 3.9). We present the boxplots of the predictions resulting from all scenarios. For reference, Figure A7 visualizes the three dependence levels we consider for the Hüsler-Reiss structure.

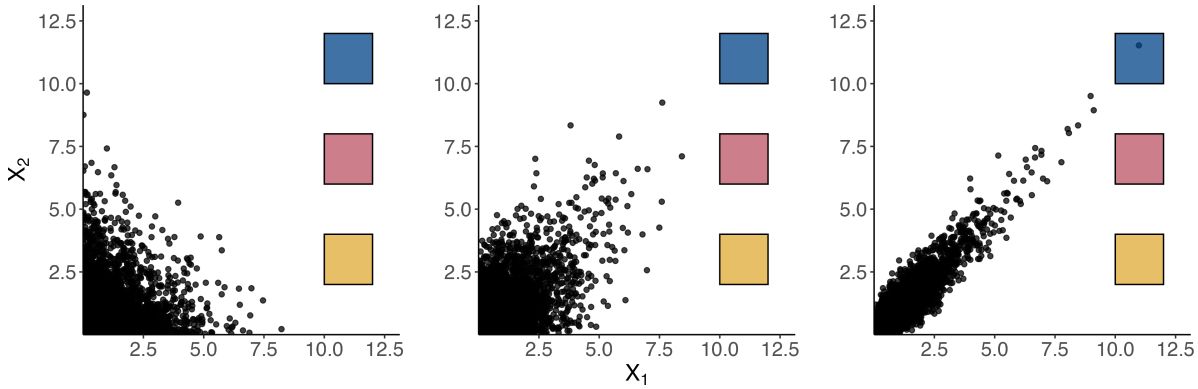


Figure A7: Prediction regions under the simulation study: $B_1 = (10, 12) \times (10, 12)$., $B_2 = (10, 12) \times (6, 8)$., and $B_3 = (10, 12) \times (2, 4)$., represented by the blue, red, and yellow boxes, respectively. Levels of dependence $\lambda = 0.1, 1,$ and 3, corresponding to low, mid, and high, are shown from left to right. Data are generated according to 3.5.1.

A.2.1 Gaussian dependence structures

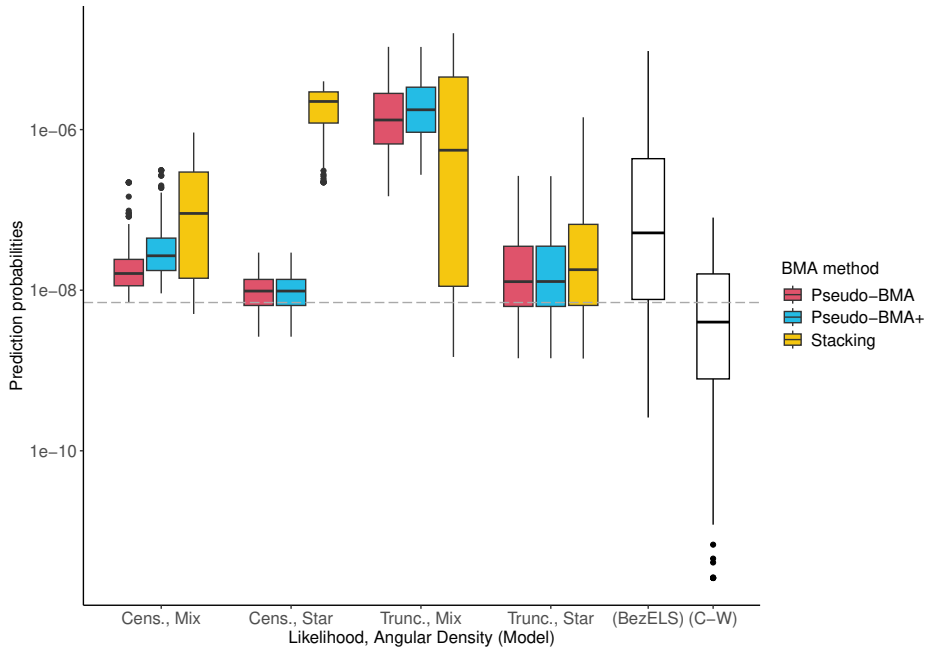


Figure A8: Gaussian dependence, $\rho = 0.1$. Predictions in $B_1 = (10, 12) \times (10, 12)$.

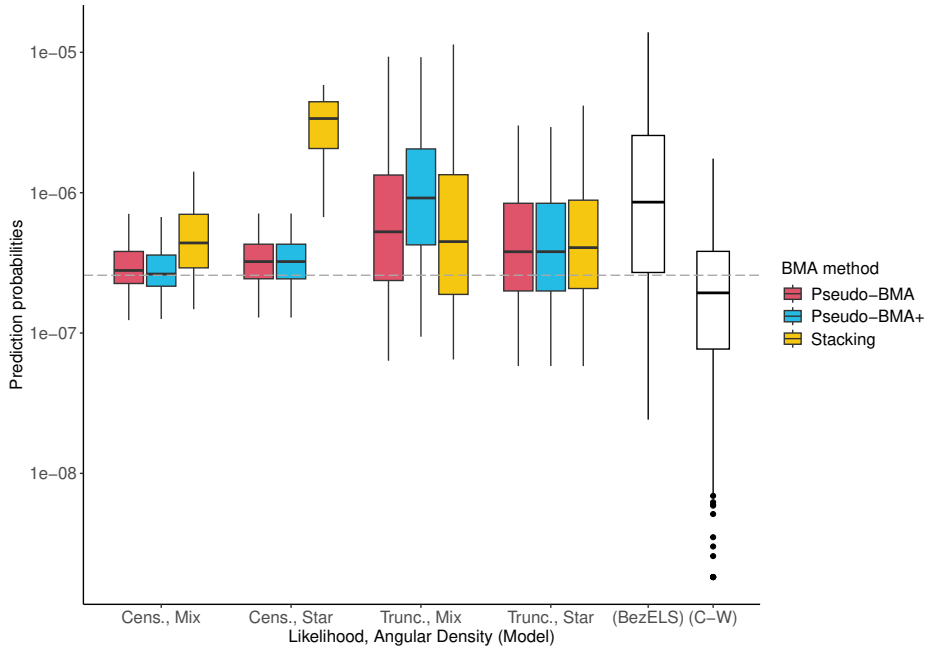


Figure A9: Gaussian dependence, $\rho = 0.1$. Predictions in $B_2 = (10, 12) \times (6, 8)$.

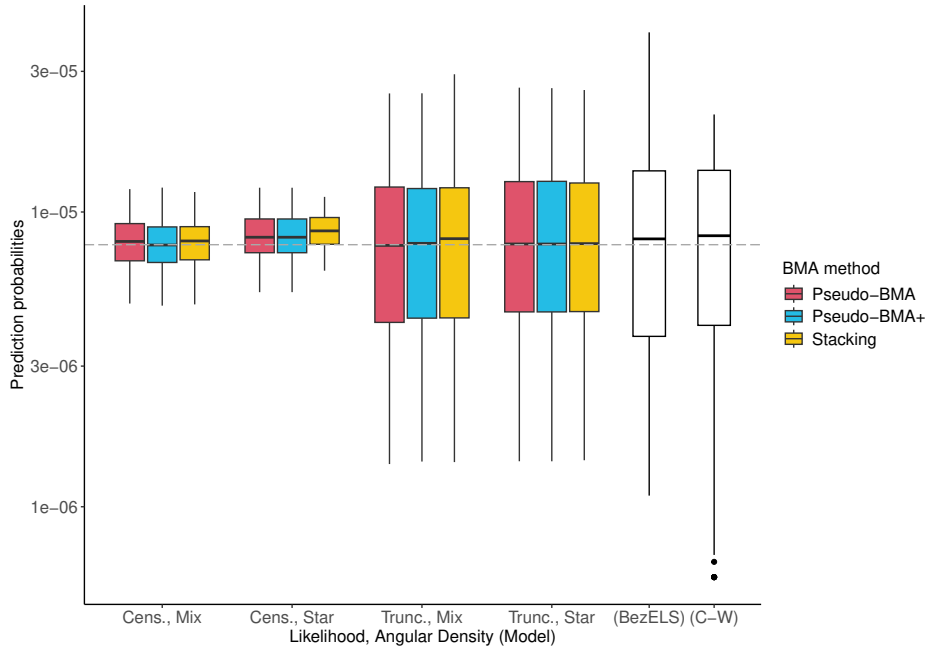


Figure A10: Gaussian dependence, $\rho = 0.1$. Predictions in $B_3 = (10, 12) \times (2, 4)$.

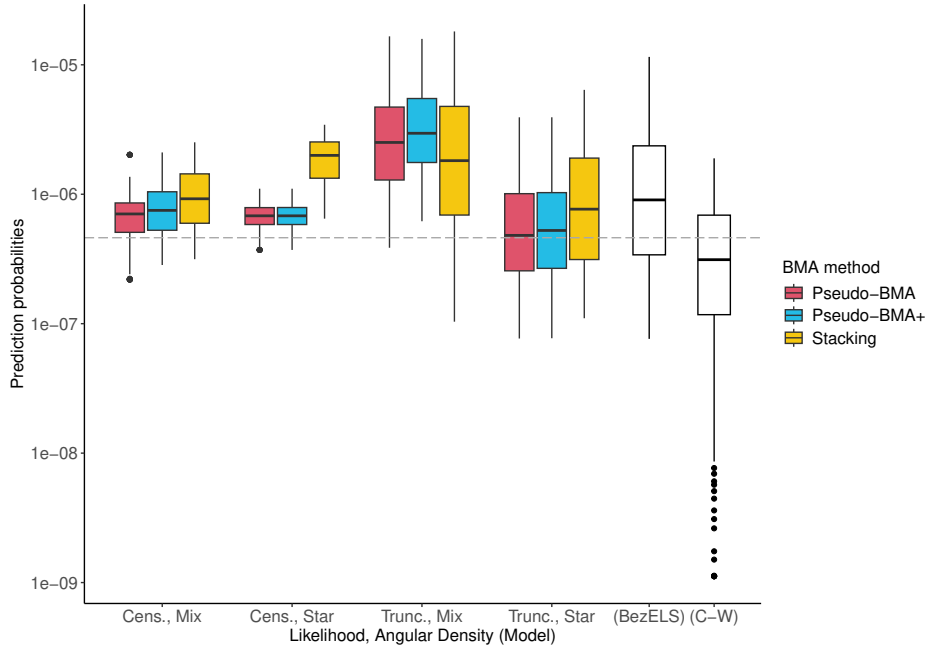


Figure A11: Gaussian dependence, $\rho = 0.5$. Predictions in $B_1 = (10, 12) \times (10, 12)$.

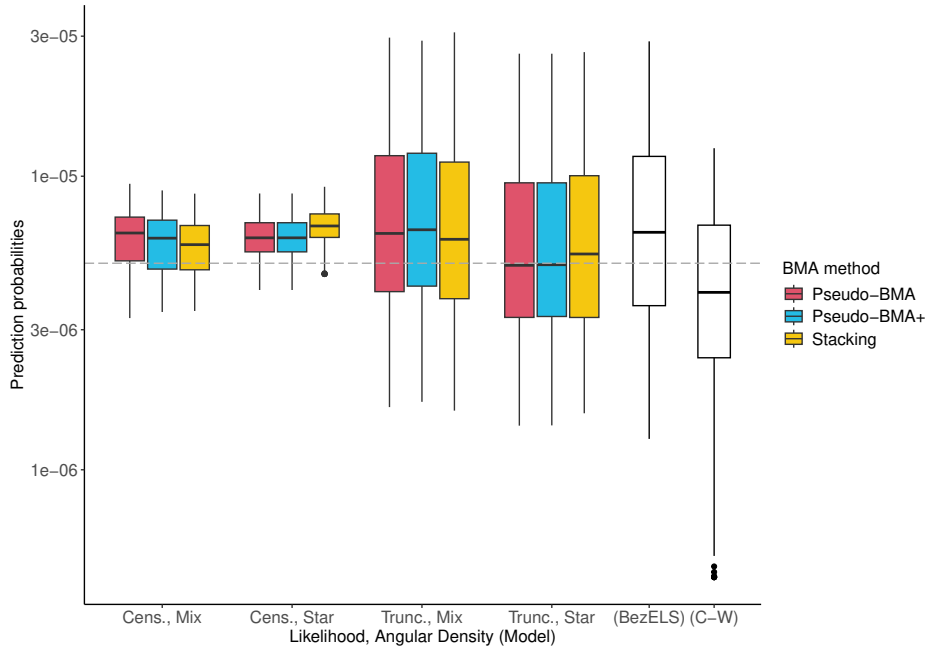


Figure A12: Gaussian dependence, $\rho = 0.5$. Predictions in $B_2 = (10, 12) \times (6, 8)$.

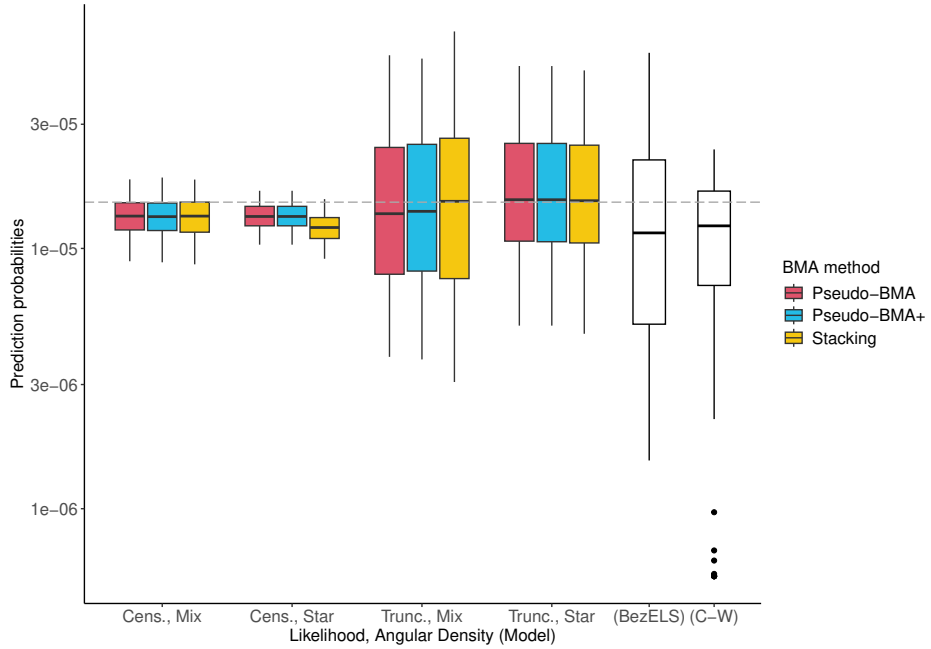


Figure A13: Gaussian dependence, $\rho = 0.5$. Predictions in $B_3 = (10, 12) \times (2, 4)$.

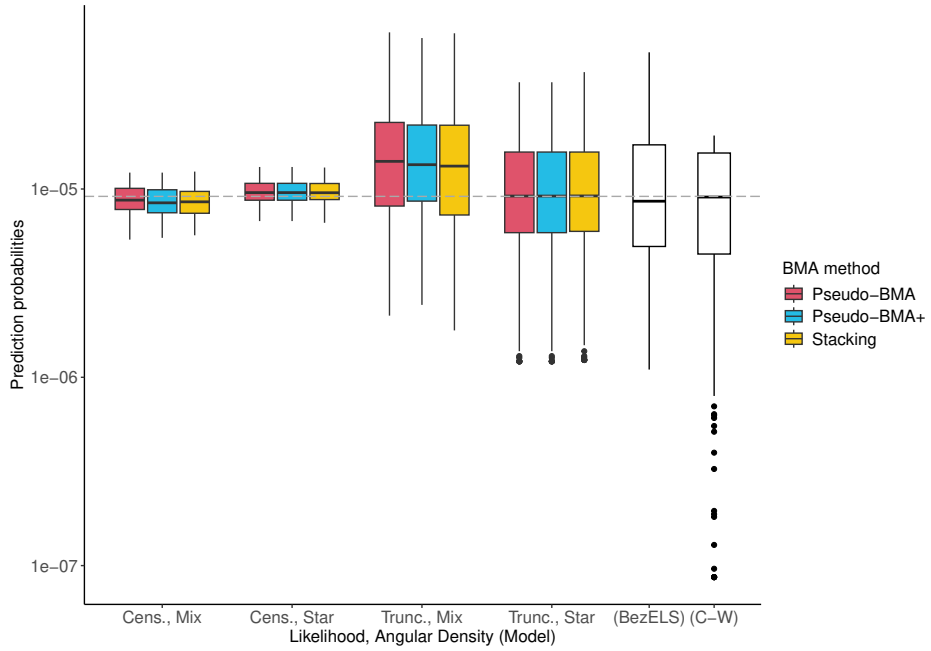


Figure A14: Gaussian dependence, $\rho = 0.9$. Predictions in $B_1 = (10, 12) \times (10, 12)$.

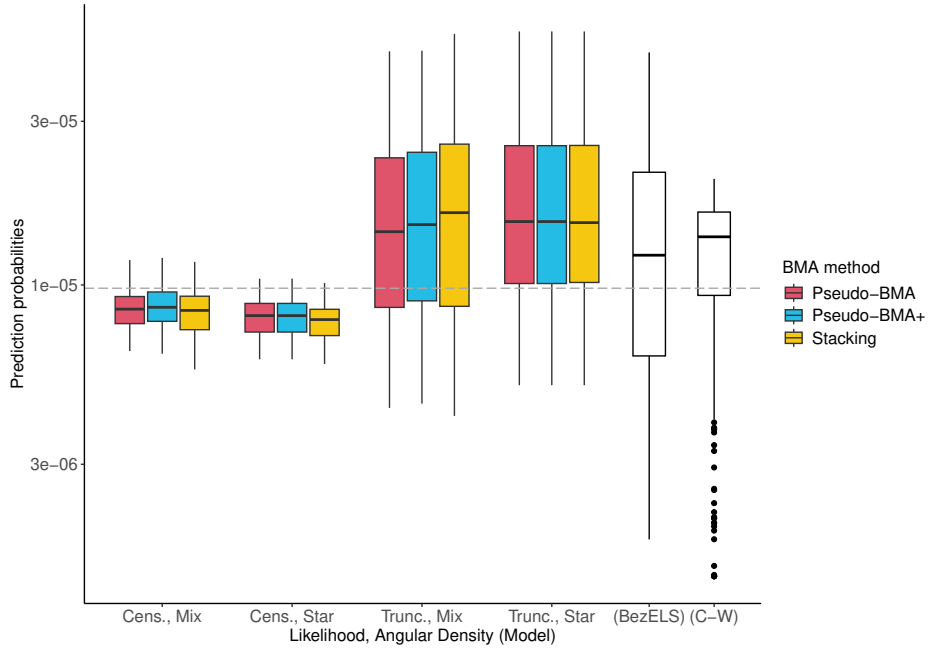


Figure A15: Gaussian dependence, $\rho = 0.9$. Predictions in $B_2 = (10, 12) \times (6, 8)$.

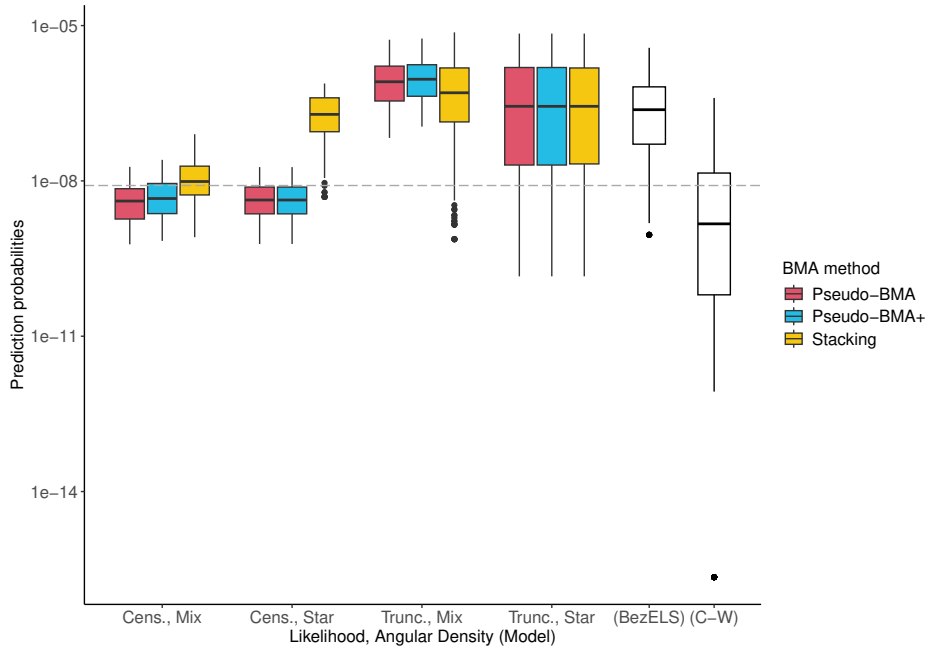


Figure A16: Gaussian dependence, $\rho = 0.9$. Predictions in $B_3 = (10, 12) \times (2, 4)$.

A.2.2 Logistic dependence structures

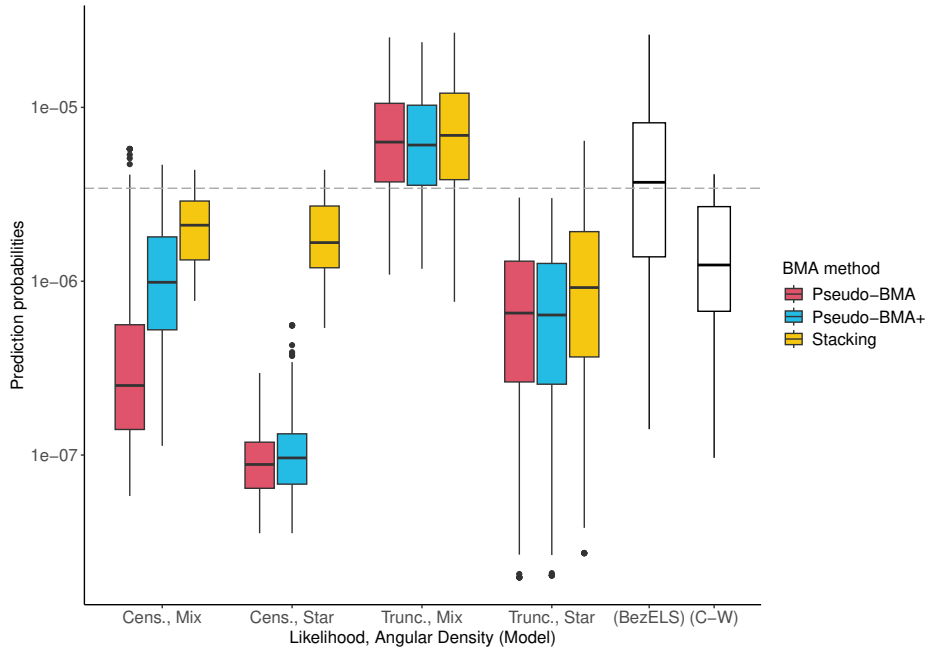


Figure A17: Logistic dependence, $\lambda = 0.9$. Predictions in $B_1 = (10, 12) \times (10, 12)$.

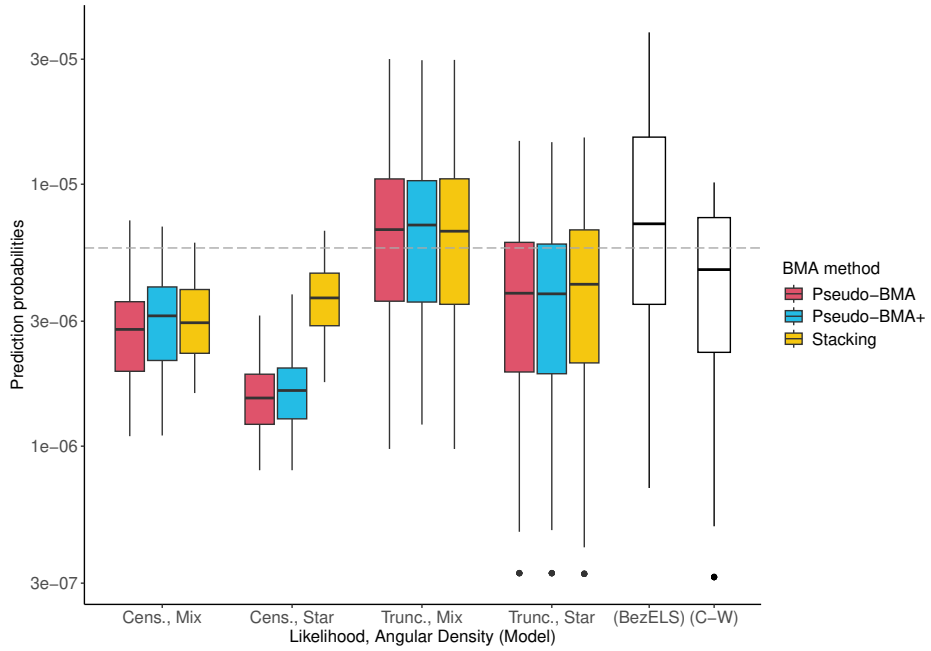


Figure A18: Logistic dependence, $\lambda = 0.9$. Predictions in $B_2 = (10, 12) \times (6, 8)$.

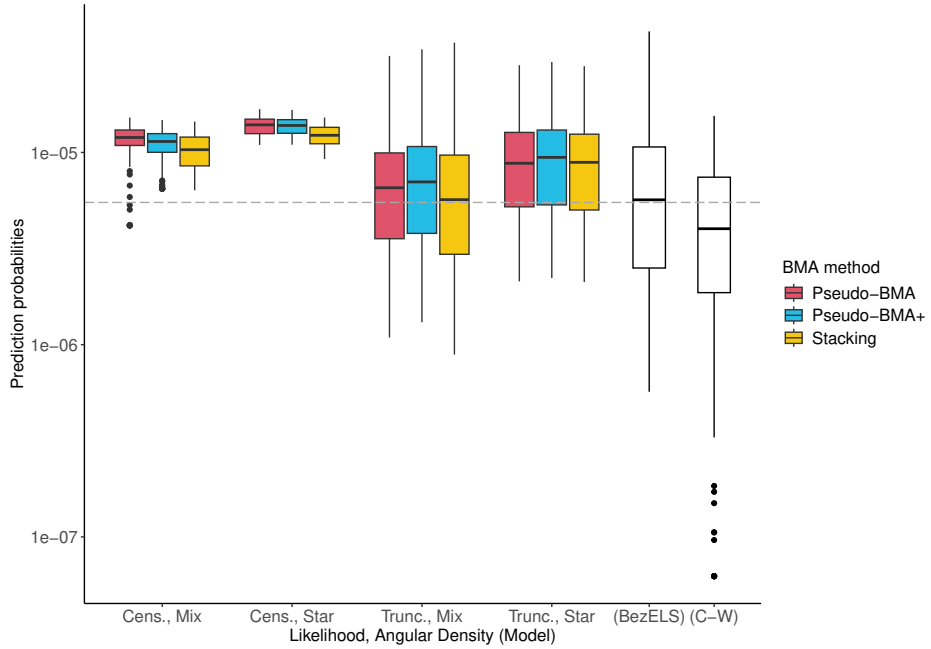


Figure A19: Logistic dependence, $\lambda = 0.9$. Predictions in $B_3 = (10, 12) \times (2, 4)$.

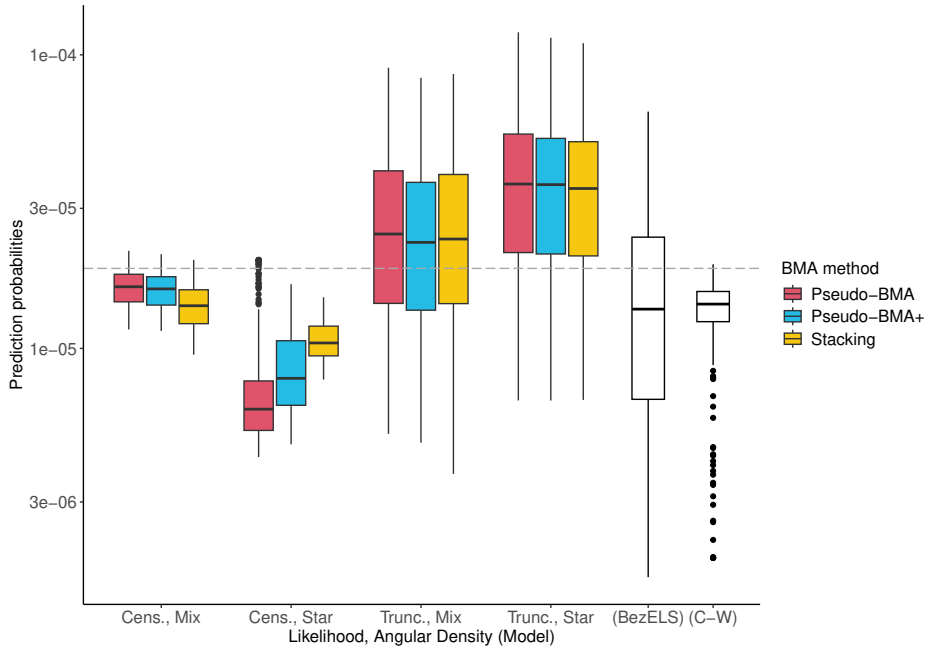


Figure A20: Logistic dependence, $\lambda = 0.5$. Predictions in $B_1 = (10, 12) \times (10, 12)$.

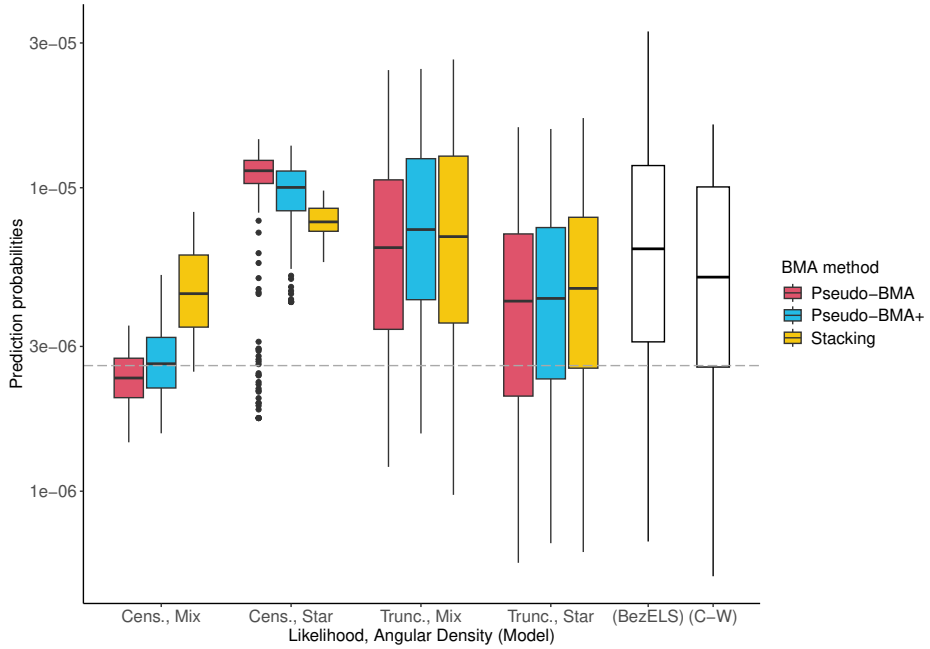


Figure A21: Logistic dependence, $\lambda = 0.5$. Predictions in $B_2 = (10, 12) \times (6, 8)$.

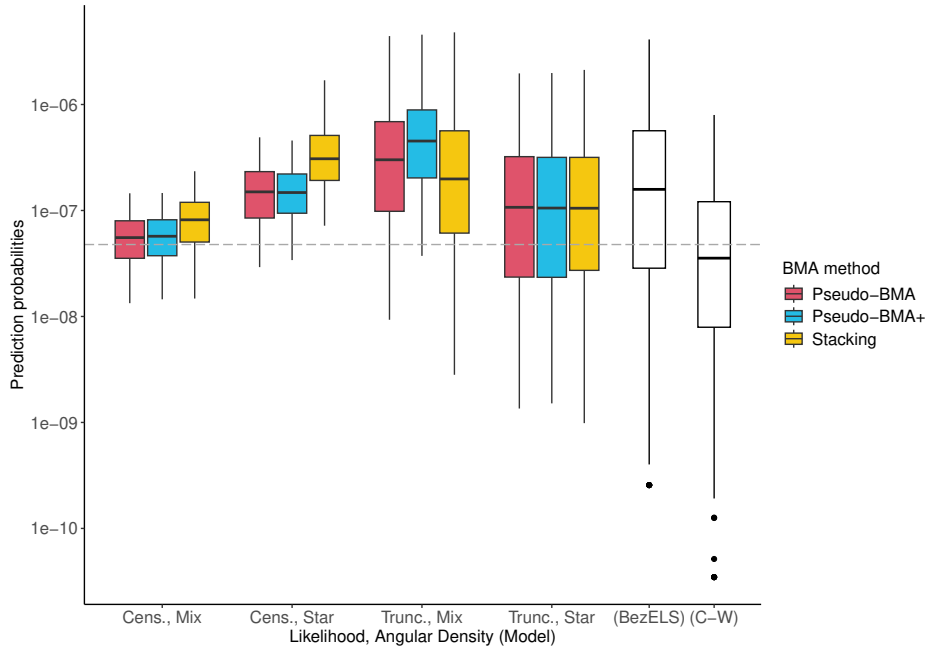


Figure A22: Logistic dependence, $\lambda = 0.5$. Predictions in $B_3 = (10, 12) \times (2, 4)$.

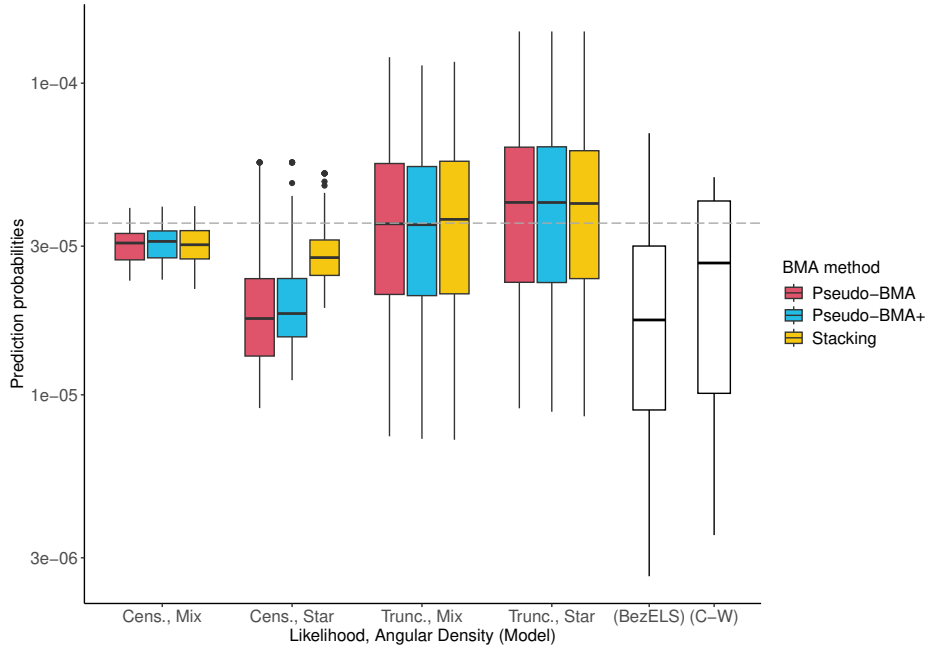


Figure A23: Logistic dependence, $\lambda = 0.1$. Predictions in $B_1 = (10, 12) \times (10, 12)$.

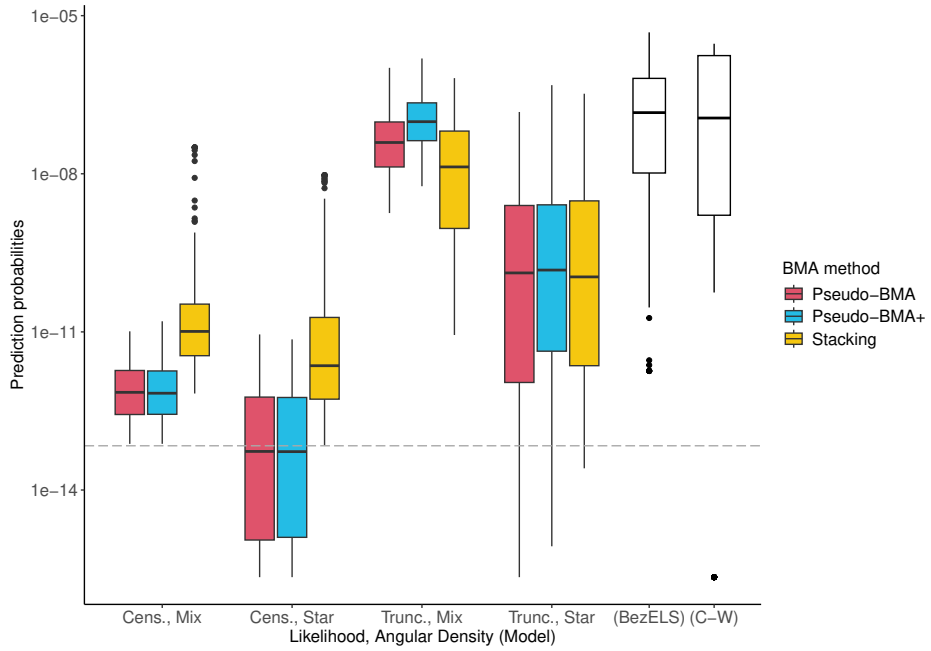


Figure A24: Logistic dependence, $\lambda = 0.1$. Predictions in $B_2 = (10, 12) \times (6, 8)$.

A.2.3 Hüsler-Reiss dependence structures

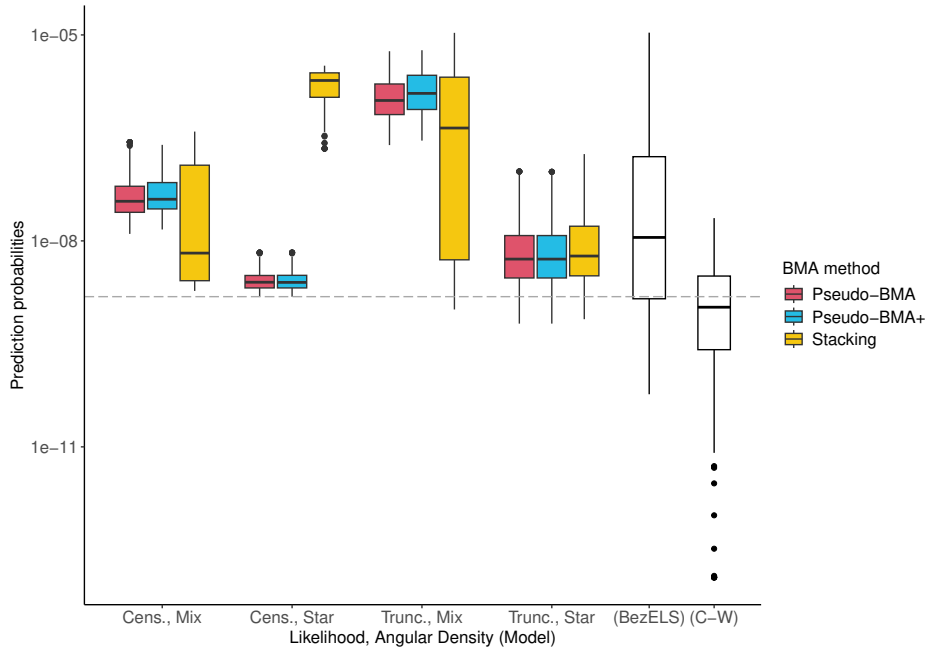


Figure A25: Hüsler-Reiss dependence, $\lambda = 0.1$. Predictions in $B_1 = (10, 12) \times (10, 12)$.

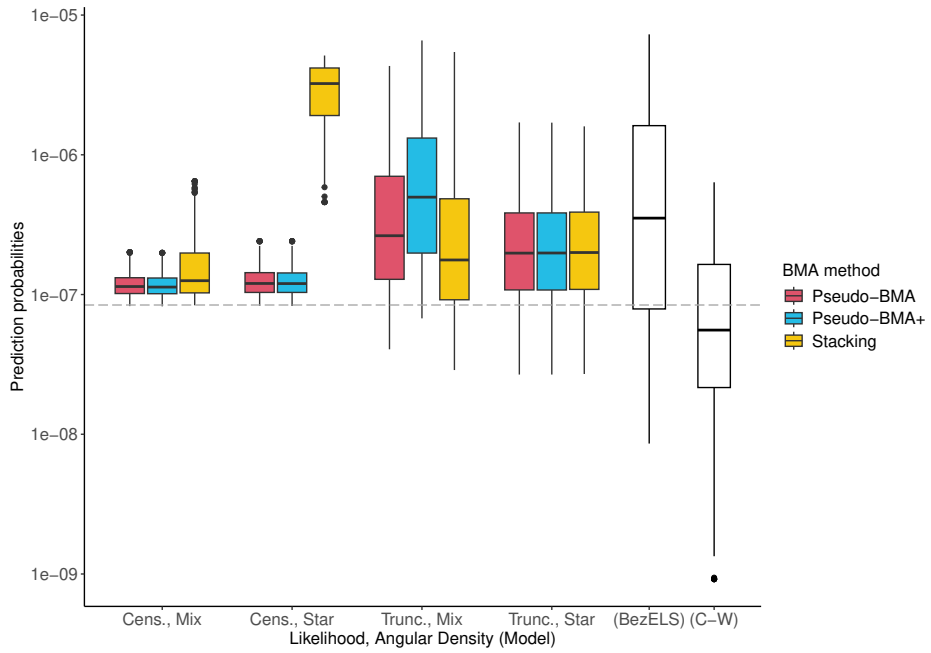


Figure A26: Hüsler-Reiss dependence, $\lambda = 0.1$. Predictions in $B_2 = (10, 12) \times (6, 8)$.

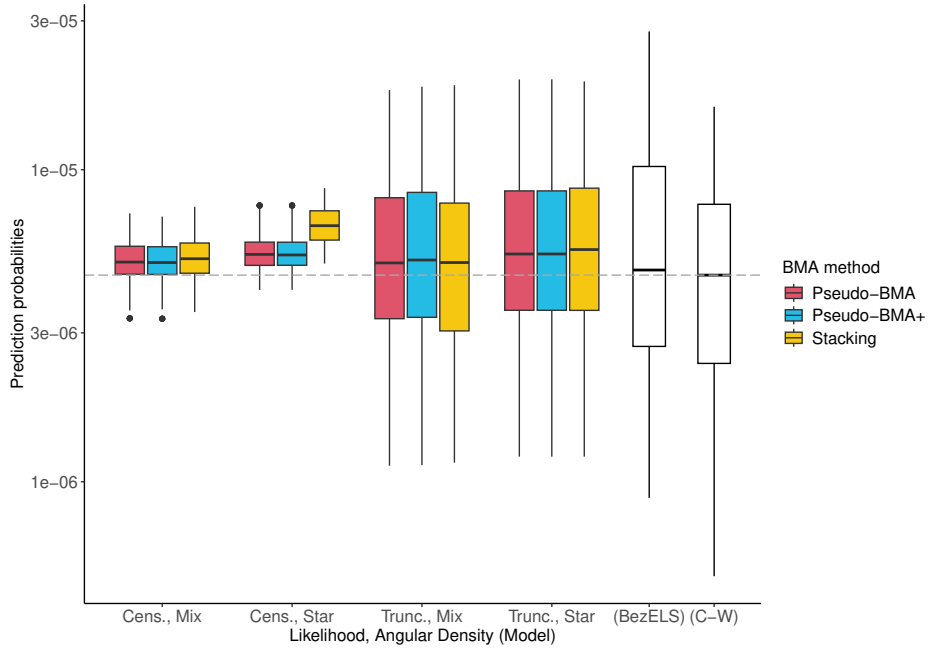


Figure A27: Hüsler-Reiss dependence, $\lambda = 0.1$. Predictions in $B_3 = (10, 12) \times (2, 4)$.

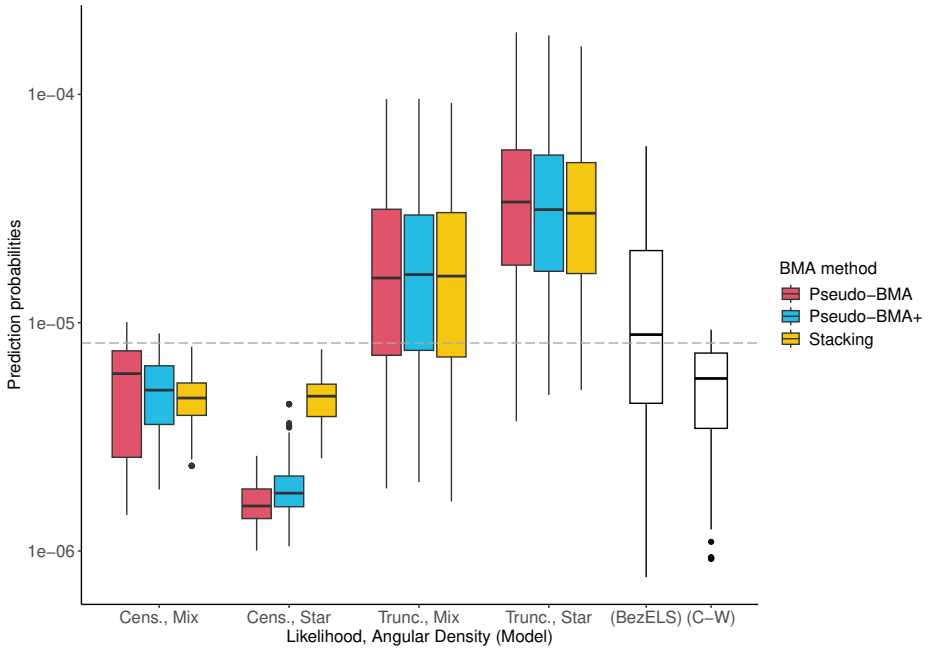


Figure A28: Hüsler-Reiss dependence, $\lambda = 1$. Predictions in $B_1 = (10, 12) \times (10, 12)$.

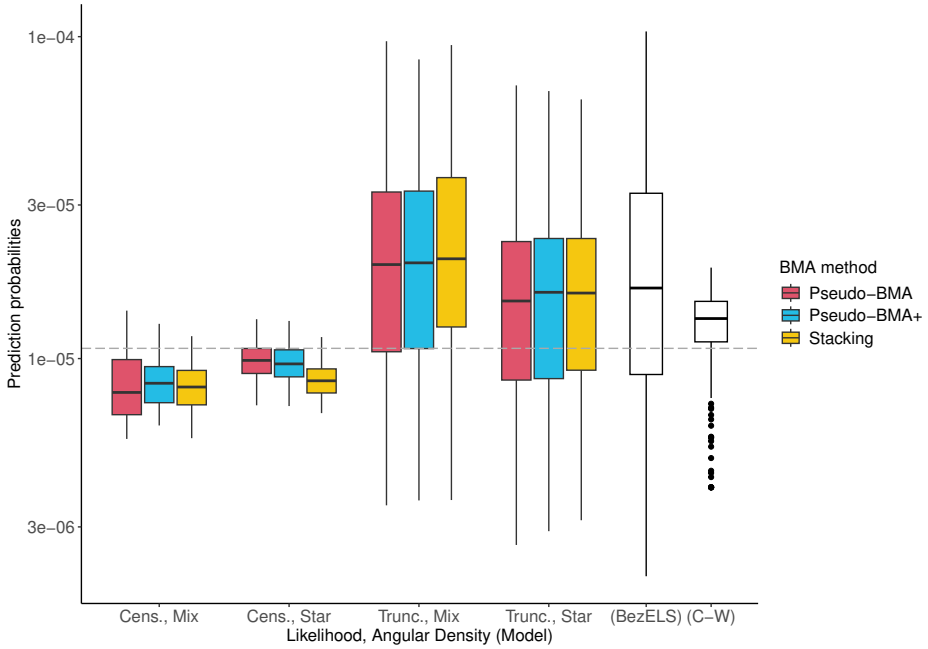


Figure A29: Hüsler-Reiss dependence, $\lambda = 1$. Predictions in $B_2 = (10, 12) \times (6, 8)$.

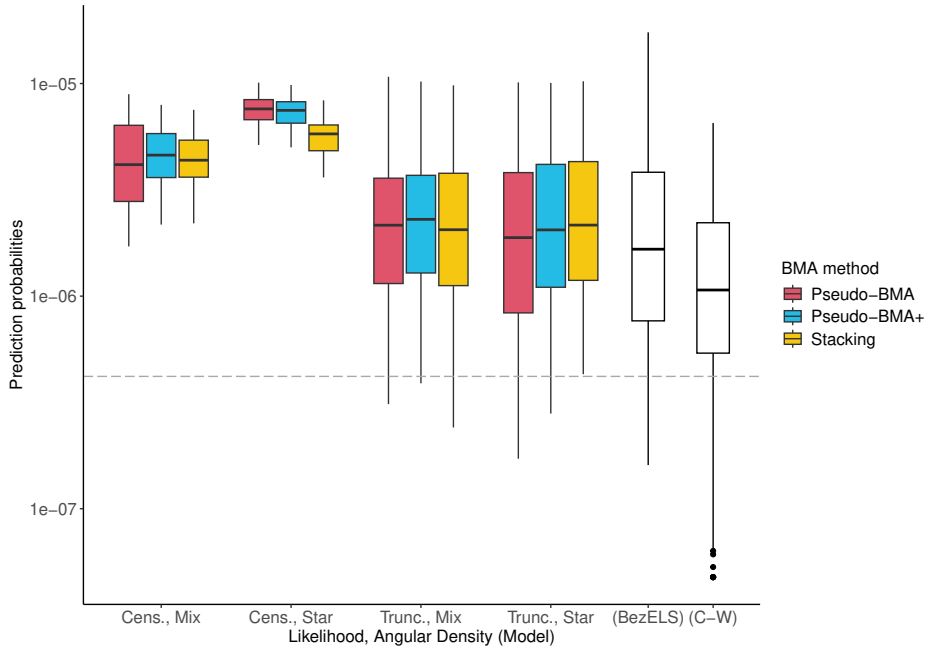


Figure A30: Hüsler-Reiss dependence, $\lambda = 1$. Predictions in $B_3 = (10, 12) \times (2, 4)$.

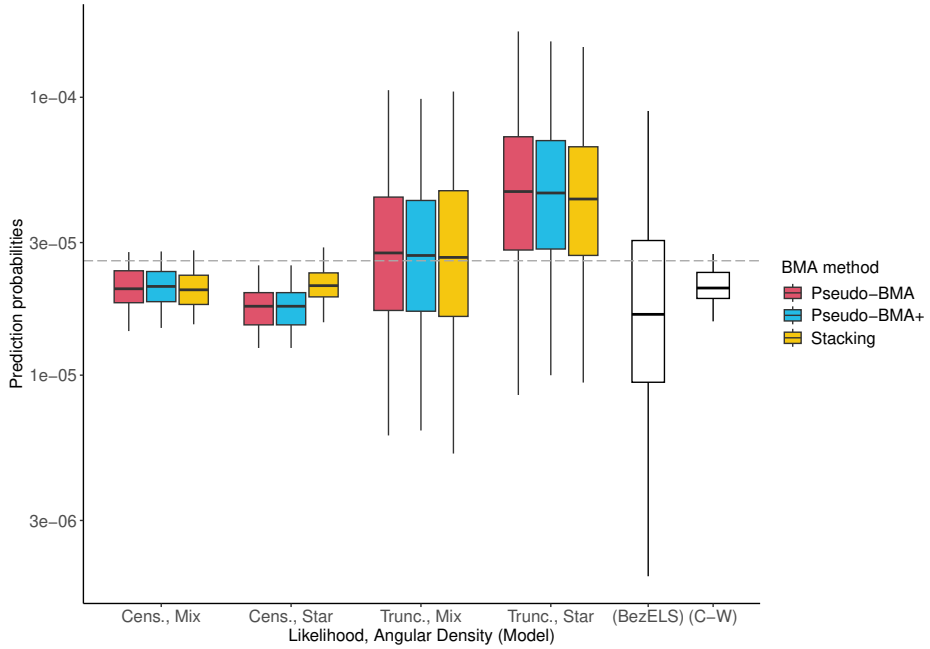


Figure A31: Hüsler-Reiss dependence, $\lambda = 3$. Predictions in $B_1 = (10, 12) \times (10, 12)$.

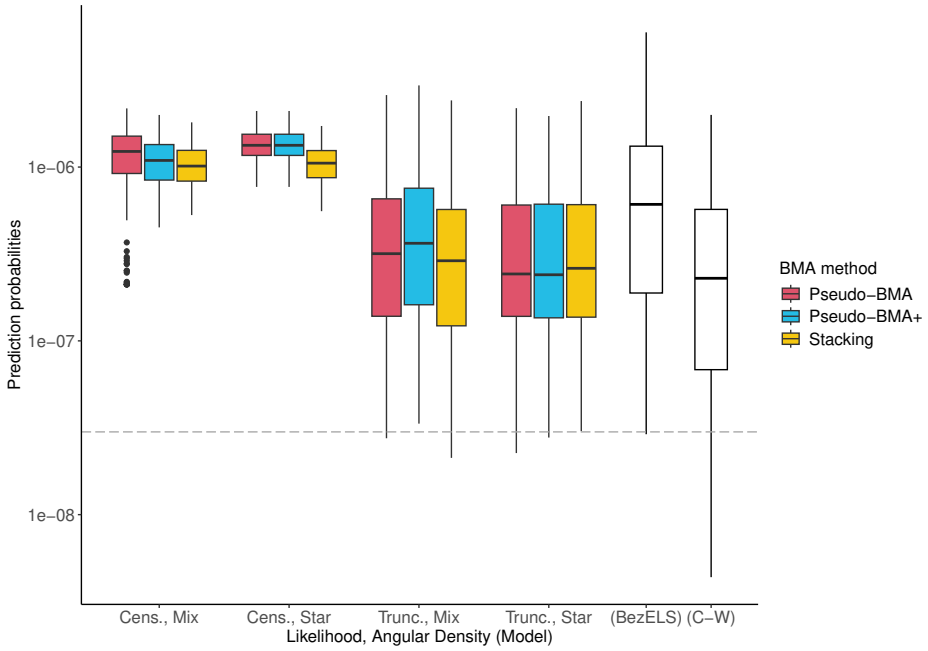


Figure A32: Hüsler-Reiss dependence, $\lambda = 3$. Predictions in $B_2 = (10, 12) \times (6, 8)$.