

Spatiotemporal modelling of extreme wildfires and severe thunderstorm environments

Présentée le 14 janvier 2022

Faculté des sciences de base
Chaire de statistique
Programme doctoral en mathématiques

pour l'obtention du grade de Docteur ès Sciences

par

Jonathan KOH BOON HAN

Acceptée sur proposition du jury

Prof. C. Hongler, président du jury
Prof. A. C. Davison, directeur de thèse
Prof. B. Shaby, rapporteur
Dr T. L. Thorarinsdottir, rapporteuse
Prof. A. Berne, rapporteur

Malu bertanya sesat jalan,
malu berkayuh perahu hanyut.
— Malay proverb

Acknowledgements

I am grateful to my supervisor Anthony Davison: for giving me the intellectual freedom throughout this PhD to pursue my interests; for his patience, care and guidance; and above all for his warmth and kindness. Anthony has been a crucial parental figure in my journey to be a better statistician, and I have benefitted greatly from his vast statistical and editorial expertise.

I thank my jury members for their time and effort: Clément Hongler for presiding the jury; Alexis Berne, Ben Shaby and Thordis Thorarinsdottir for agreeing to assess and examine my manuscript.

I thank the coauthors of my papers: Chiara Lepore, Michael Tippet, François Pimont and Jean-Luc Dupuy for their insights and references for the applied part of our work; Erwan Koch and Thomas Opitz, for the many discussions and your efforts that helped steer our work in the right direction.

Teaching assistantships were a big part of my EPFL experience, and I have learnt a lot from students, as well as instructors I assisted: Guillaume Dehaene, Guillaume Obozinski, and Emeric Thibaud.

I am grateful to everyone from the STAT group (past and present) for sharing lunch times, coffee breaks and your life stories with me: Léo, Raphaël, Hélène, Thomas, Yousra, Sebastian, Soumaya, Tim, Mario, Ophélie and Marcelo. Similar thanks go to the other colleagues from other groups across the hallway: Alessia, Zhiwen, Tomas R., Tomas M., Laya, Kartik, Jana and Moritz. Thank you for making my time at EPFL a memorable one. To Nadia, thank you for your professionalism, and for always having an open door whenever I had administrative queries.

Merci Sonia d'avoir le double rôle d'être ma collègue de bureau et mon dictionnaire français, and also for sharing my obsession with coffee. Grazie mille Stefano, for your valuable friendship which no doubt matured after the onset of the Covid-19 pandemic.

I could not have crossed the finish line without support from friends outside of EPFL. There are too many to thank, but those from Malaysia: Gloria; Jason, Ben and Kent. Thank you for all the close calls. From London: Damien, Vince, Pavel, Kuan Haw, and Zen. From Switzerland, I thank all the coaches and teammates I played with at Fly High and the pickup group; Vaibhav and Fernando; Martin; Geraldine, Justin, Mel and Sam for providing the warmth from home I so desperately needed at times.

Acknowledgements

My family played another important role throughout this journey. Thank you to Ko Tim and Vera, Ko Dan and Rachel, and my beautiful niece and nephews: Evie, Tobias, Ezra and Liyanna, for all the zoom calls from far away.

Thank you to Walter and Susanne, for inviting me over to your home in Zurich countless times; for your warmth and motivation, and for treating me like your own son. Vielen Dank für eure Gastfreundschaft.

I am forever indebted to my parents: ma and pa. Thank you for your prayers, your constant encouragement, and your never-ending love. I would not be here today without you, and I am grateful you instilled in me the importance of hard work and education, which helped frame the blueprint of my quest for knowledge.

To Caro, my partner in crime. Thank you for your love and support; for journeying with me through my ups and downs these four years; and for making this place my home.

Lausanne, December 9, 2021

J. K.

Abstract

Environmental extreme events can have devastating impacts on society when they interact with vulnerable human and natural systems. Such events can result from natural causes, like phenomena related to the El Niño-Southern Oscillation or decadal/multi-decadal climate variations. These causes can follow an increase in human activity, e.g., through land-use changes or anthropogenic climate change, that can influence the frequency, intensity, spatial extent and timing of these events, and spur unprecedented extremes. To accurately understand and quantify the risks associated with these events, it is important to identify trends related to these causes, which may be measured or unmeasured.

Fitting models for rare events is inherently difficult because of the paucity of data available. The most destructive extreme events are rarely isolated in space and time, so one must account for their spatial and temporal dependencies. This thesis deals with the parametric modelling of severe thunderstorms and wildfires using models motivated from limiting probabilistic results.

The first part of this thesis explores influences on the magnitude and spatial extent of extremes of environments related to severe US thunderstorms. Our results show that the risk from severe thunderstorms in April and May is increasing in parts of the US where it was already high, and that the risk from storms in February increases during La Niña years. We also show that these extremes are more localized during spring/summer seasons than in the winter, and find that some of these seasonal differences are more pronounced during El Niño years.

The second part of the thesis deals with predicting and explaining the spatial extent, frequency, intensity and timing of wildfires using meteorological and land-use covariates. Our first approach uses ideas from extreme-value theory in a machine learning context to give good prediction of the distributional tails of our data. The second approach uses a novel Bayesian hierarchical model designed specifically for extreme wildfires. We show that wildfire risk on the French Mediterranean basin is affected by significant random effects related to land-use and policy changes, and a seasonally-varying fire-weather index.

Keywords: Bayesian hierarchical model, environmental statistics, extreme values, generalized extreme value distribution, generalized Pareto distribution, gradient boosting, max-stability, model validation, severe thunderstorms, wildfire modelling

Résumé

Les événements environnementaux extrêmes peuvent avoir des effets dévastateurs sur la société lorsqu'ils interagissent avec des systèmes humains et naturels vulnérables. Ces événements peuvent résulter de causes naturelles, comme les phénomènes liés à l'oscillation australe El Niño ou les variations climatiques décennales/multidécennales. Ces causes peuvent faire suite à une augmentation de l'activité humaine, par exemple par le biais de changements dans l'utilisation des terres ou de changements climatiques anthropiques, qui peuvent influencer la fréquence, l'intensité, l'étendue spatiale et le moment de ces événements, et provoquer des extrêmes sans précédent. Pour comprendre et quantifier avec précision les risques associés à ces événements, il est important d'identifier les tendances liées à ces causes, qui peuvent être mesurées ou non.

L'ajustement de modèles pour les événements rares est intrinsèquement difficile en raison de la rareté des données disponibles. Les événements extrêmes les plus destructeurs sont rarement isolés dans l'espace et le temps, il faut donc tenir compte de leurs dépendances spatiales et temporelles. Cette thèse traite de la modélisation paramétrique des orages violents et des feux de forêt en utilisant des modèles motivés par des résultats probabilistes limitatifs.

La première partie de cette thèse étudie les facteurs qui influencent sur la magnitude et l'étendue spatiale des orages sévères aux états-unis. Nos résultats montrent que le risque d'orages violents en avril et mai augmente dans des régions des États-Unis où il est à priori élevé, et que le risque d'orages en février augmente pendant les années La Niña. Nous montrons également que ces extrêmes sont plus localisés pendant le printemps et l'été par rapport à l'hiver, et nous constatons que certaines de ces différences saisonnières sont plus prononcées pendant les années El Niño.

La deuxième partie de la thèse traite de la prédiction et de l'explication de l'étendue spatiale, de la fréquence, de l'intensité et du moment des incendies de forêt en utilisant des covariables météorologiques et d'utilisation des terres. Notre première approche utilise les notions de la théorie des valeurs extrêmes dans un contexte d'apprentissage automatique pour donner une bonne prédiction des queues de distribution de nos données. La seconde approche utilise un nouveau modèle hiérarchique bayésien conçu spécifiquement pour les incendies de forêt extrêmes. Nous montrons que le risque d'incendie dans le bassin méditerranéen français est affecté par des effets aléatoires significatifs liés à l'utilisation des terres et aux changements de politique, ainsi que par un indice de temps d'incendie variant selon les saisons.

Résumé

Mots-clés : Distribution généralisée de Pareto, distribution généralisée des valeurs extrêmes, modèle hiérarchique bayésien, modélisation des incendies, orages violents, renforcement du gradient, stabilité maximale, statistiques environnementales, valeurs extrêmes, validation du modèle.

Contents

Acknowledgements	i
Abstract (English/Français)	iii
Introduction	1
I Thunderstorms	5
1 Trends in the extremes of environments associated with severe US thunderstorms	7
1.1 Abstract	7
1.2 Introduction	8
1.3 Data and exploratory analysis	10
1.4 Methodology	14
1.4.1 Modelling of maxima	14
1.4.2 Assessment of GEV fit	16
1.4.3 Testing procedure	17
1.5 Results	19
1.6 Conclusion	26
2 Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments	27
2.1 Abstract	27
2.2 Introduction	27
2.3 Preliminaries	30
2.3.1 Max-stable random fields	30
2.3.2 Estimation of max-stable fields	32
2.3.3 Data and exploratory analysis	34
2.4 New methodology	35
2.4.1 Model	35
2.4.2 A max-stability test in an applied setting with unknown margins	38
2.4.3 Bootstrap-based uncertainty assessment and model selection	41
2.5 Case study	46
2.5.1 Choice of validation locations	46
2.5.2 Results	47
	vii

Contents

2.5.3	Model validation	51
2.5.4	Meteorological explanation	52
2.6	Discussion	53
2.7	Supplement	55
2.7.1	Algorithms	55
2.7.2	Plots of trend surfaces	55
II	Wildfires	59
3	Gradient boosting with extreme-value theory for wildfire prediction	61
3.1	Abstract	61
3.2	Introduction	61
3.3	Data and exploratory analyses	64
3.4	Methodology	68
3.4.1	Extreme-value theory	68
3.4.2	Gradient tree boosting	68
3.4.3	Loss functions	71
3.4.4	A spatiotemporal cross-validation scheme	74
3.5	Models	76
3.5.1	Fitting procedure	76
3.5.2	Results	78
3.6	Discussion	82
3.7	Supplement	83
3.7.1	Terms in the gradients and hessians	83
3.7.2	Priors and SPDE triangulation	84
4	Spatiotemporal wildfire modeling through point processes with extreme marks	87
4.1	Abstract	87
4.2	Introduction	88
4.3	Wildfire data	91
4.4	Methods for point patterns with extreme marks	92
4.4.1	Extreme-value theory	92
4.4.2	Mark-dependent thinning of point processes	93
4.4.3	Spatiotemporal log-Gaussian Cox processes	94
4.4.4	Data aggregation and subsampling schemes	95
4.4.5	Fully Bayesian inference using INLA-SPDE	96
4.5	Point processes with moderate and extreme marks	98
4.5.1	Bayesian hierarchical multi-response regression	98
4.5.2	Sharing latent effects	99
4.5.3	Prior structure of linear predictors	100
4.5.4	Alternative model specifications	101
4.6	Results	102

4.6.1	Model selection and comparison	102
4.6.2	Visual inspection of posterior predictive densities	104
4.6.3	Principal results of the main model M1	104
4.7	New insights for wildfire science	108
4.7.1	FWI and seasonal effects	108
4.7.2	Time trends during the study period	109
4.7.3	Shared spatial effects for improved regionalized predictions	109
4.8	Conclusion	110
4.9	Supplement	111
4.9.1	Plots for the inspection of posterior predictive densities	111
4.9.2	Plots showing regionalized predictions	111
4.9.3	Kernel intensity plot	111
4.9.4	Mean excess plots	112
4.9.5	Subsampling experiments	113
4.9.6	Other hyperpriors	114
4.9.7	Spatial effects in model M1	115
4.10	A brief overview of INLA	118
4.10.1	Approximating the posterior marginals for the hyperparameters	118
4.10.2	Approximating the posterior marginals for the latent field	119
4.11	A brief overview of the SPDE approach	121
5	Conclusion	123
	Bibliography	125
	Curriculum Vitae	139

Introduction

An elderly woman from the island of Evia in Greece holds her right hand over her heart in despair. Behind her, conflagrations engulf the forest and threaten her home. A photograph documenting her reaction to unprecedented wildfires so perfectly captures the general feeling of distress after the release of the sixth assessment report by The Intergovernmental Panel on Climate Change (IPCC) two days later, that it appeared on the front pages of the Daily Mail, the Financial Times and the Guardian.

A ‘code-red for humanity’, was UN Secretary-General António Guterres’s description of the report, which for the first time includes a chapter dedicated to environmental extremes (Seneviratne et al., 2021). ‘The alarm bells are deafening, and the evidence is irrefutable’. The chapter concludes that human-caused emissions of greenhouse gases have caused an increased frequency and/or intensity of some environmental extremes.

Yet the photograph encapsulates the surprised reaction of the woman to the wildfires: ‘Why should the forest burn? Isn’t it a sin?’ Although Greece has not experienced a heatwave quite like this in more than three decades, there are indeed other unmeasured causes for the blazes that had incinerated over 460,000 acres of forest in Evia by the 8th of August 2021. Several suspected arsonists were arrested in connection with the fires, and in fact roughly 90% of all wildfires are caused by human activity, which is difficult to quantify. Greece’s prime minister has also apologised for ‘weaknesses’ in efforts to quell wildfires, as local officials complained about the lack of firefighting resources, especially from the air. To assess the risks associated with environmental extremes such as these fires, it is crucial to incorporate both measured and unmeasured causes.

Identifying trends related to measured causes is important for policymakers and aids risk mitigation, but these trends are uncertain. The IPCC found that even a small incremental increase in global warming (+0.5°C) worsens droughts in some regions, but findings such as this are always subject to uncertainty. Accurately quantifying trends and their variability requires well-specified statistical models grounded in probability theory and expert knowledge.

Apart from human casualties, environmental extremes can also lead to major economic losses. Global insured claims due to wildfire events increased from below \$10 billion in 2000–2009 to \$45 billion in the subsequent decade¹. Annual losses from severe thunderstorms in the United

¹<https://www.swissre.com/risk-knowledge/mitigating-climate-risk/yet-more-wildfires.html>



Figure 1 – Ritsopi Panagiota, 81, reacting to wildfires reaching her house in the village of Gouves on Evia, Greece. Photograph by Konstantinos Tsakalidis.

States alone have exceeded \$10 billion in recent years². These economic and human impacts are a strong motivation to study how and why environmental extremes vary from season to season and region to region.

Extreme-value theory provides tools to study such events, which are often characterized by their complex dependencies in both space and time. The wildfires in Evia lasted two weeks and were not spatially isolated; other parts of Greece and neighbouring Turkey were simultaneously battling wildfires as temperatures rose to 45°C. Merging extreme-value theory with methods from spatial statistics and machine learning can help to model and explain such extremal dependencies.

The big data revolution has contributed to the boom of the aforementioned fields in the last decades, partly prompted by the impending climate crisis and the increasing desire to use data to assess risk. However, modelling extreme events is inherently difficult due to the paucity of data available, and computational challenges arise when the number of covariates or data locations increases.

This thesis deals with the analyses of measured and unmeasured causes of environmental extremes to better understand their associated risks. Chapter 1, 2 and 3 focuses on quantifying relevant trends from measured causes, which are of primary interest to policymakers or intergovernmental bodies such as the IPCC. Chapter 4 advocates estimating and incorporating unmeasured causes to better model complex processes, and we illustrate our method with an application to wildfires.

²[http://www.willisre.com/Media_Room/Press_Releases_\(Browse_All\)/2017/WillisRe_Impact_of_ENSO_on_US_Tornado_and_Hail_frequencies_Final.pdf](http://www.willisre.com/Media_Room/Press_Releases_(Browse_All)/2017/WillisRe_Impact_of_ENSO_on_US_Tornado_and_Hail_frequencies_Final.pdf)

Outline of the thesis

Severe thunderstorms are associated with concurrently elevated values of 0–180hPa Convective Available Potential Energy (CAPE; J kg^{-1}), which measures the amount of energy available for convection in the atmosphere, and 0–3km Storm Relative Helicity (SRH; m^2s^{-2}), which measures vertical wind shear. For the first half of the thesis, we model the extremes of SRH, CAPE and $\text{PROD} = \sqrt{\text{CAPE}} \times \text{SRH}$.

In Chapter 1, we present the generalized extreme value distribution, which describes the univariate limiting distribution of block maxima, and use it to perform trend analyses of all variables for all months. We consider time and a well known El Niño-Southern Oscillation index as covariates, and account for multiple testing. Our results suggest that the risk from severe thunderstorms in April and May is increasing in parts of the contiguous US where it was already high, and that the risk from storms in February tends to be higher over large parts of our spatial domain during La Niña years. This work has been published in the *Journal of Climate* and is available as Koch et al. (2021).

Chapter 2 introduces max-stable fields as the functional limits of component-wise block maxima, used to model the spatial dependence of extreme values. We propose three new tools for use when fitting these fields to gridded pointwise monthly maxima, and apply them to the data from Chapter 1. These tools are an out-sample selection scheme anchored on a local max-stability bootstrap test based on empirical likelihood, a model selection metric based on the block bootstrap which improves on the current state of the art, and a Brown–Resnick model with smooth spline bases incorporated into its parameters. We show that the extremes of thunderstorm-conducive environments tend to be more localized during spring and summer than in the winter, and find that these seasonal differences are more pronounced during El Niño events. This chapter will shortly be submitted to the *Journal of the American Statistical Association, Applications and Case Studies*.

The second half of the thesis concerns wildfire modelling. Chapter 3 deals with the prediction of wildfire counts and sizes over the contiguous US using meteorological and land-use covariates. We introduce gradient tree boosting, and propose new loss functions motivated by extreme-value theory that focus on good prediction of the tails of our data. One should always have the real-world prediction scenario in mind when validating models, and we appeal to ideas from spatial statistics to validate model predictions on spatially-dependent data. The predictions outperform the benchmark and other approaches. Equally important is our assessment of the importance and marginal effects of various covariates on the response, which could prompt national wildfire predictive services to rethink the design of fire danger warning systems across the contiguous US. This chapter has been submitted to *Extremes*, and is available as Koh (2021).

Chapter 4 deals with summer wildfires in the French Mediterranean basin. We develop a joint model for the occurrence intensity and the wildfire size distribution by combining extreme-value theory and tools from point processes within a novel Bayesian hierarchical

Introduction

model. Inference is performed by integrated nested Laplace approximation, with stratified subsampling of counts to handle computational issues arising from the large sample size. We introduce the log-Gaussian Cox process and use it to model wildfire ignitions. Burnt areas are numerical marks attached to points and are considered extreme if they exceed a high threshold. We capture the non-linear influences of covariates, such as a well-known fire weather index, on wildfire activity. The timing of periodic events in plant life cycles and stomatal control under drought help explain some of our estimates of these effects. We estimate various unmeasured drivers of different aspects of wildfire activity, and show that including them in our model leads to good predictions. This chapter has been submitted to the *Annals of Applied Statistics*, and is available as Koh et al. (2021).

We conclude with a discussion and outline possible extensions in Chapter 5.

Thunderstorms **Part I**

1 Trends in the extremes of environments associated with severe US thunderstorms

This chapter is a postprint of the article written with Anthony C. Davison, Erwan Koch, Chiara Lepore and Michael K. Tippett, published in Journal of Climate (Koch et al., 2021). Parts of this chapter may overlap with other chapters, but we keep it self-contained for clarity. The contributions of the doctoral candidate were in producing the results, simulations, plots and tables, and writing a substantial part of the paper.

1.1 Abstract

Severe thunderstorms can have devastating impacts. Concurrently high values of convective available potential energy (CAPE) and storm relative helicity (SRH) are known to be conducive to severe weather, so high values of $PROD = \sqrt{CAPE} \times SRH$ have been used to indicate high risk of severe thunderstorms. We consider the extreme values of these three variables for a large area of the contiguous United States (US) over the period 1979–2015, and use extreme-value theory and a multiple testing procedure to show that there is a significant time trend in the extremes for PROD maxima in April, May and August, for CAPE maxima in April, May and June, and for maxima of SRH in April and May. These observed increases in CAPE are also relevant for rainfall extremes and are expected in a warmer climate, but have not previously been reported. Moreover, we show that the El Niño-Southern Oscillation explains variation in the extremes of PROD and SRH in February. Our results suggest that the risk from severe thunderstorms in April and May is increasing in parts of the US where it was already high, and that the risk from storms in February is increased over the main part of the region during La Niña years. Our results differ from those obtained in earlier studies using extreme-value techniques to analyze a quantity similar to PROD.

Key words: Convective available potential energy; El Niño-Southern Oscillation; Generalized extreme-value distribution; Multiple testing; Severe weather; Storm relative helicity; Time trend.

1.2 Introduction

Annual losses from severe thunderstorms in the United States (US) have exceeded \$10 billion in recent years.¹ In addition to economic losses, 2011 was marked by 552 deaths caused by tornadoes. These economic and human impacts are a strong motivation to study how and why US thunderstorm activity varies from year to year and region to region. Two important aspects are trends potentially related to climate change or multi-decadal variability, and modulation by the El Niño-Southern Oscillation (ENSO). However, inadequacies in the length and quality of the thunderstorm data record present substantial challenges to addressing these questions directly (Verbout et al., 2006; Allen and Tippett, 2015; Edwards et al., 2018).

In the US, a severe thunderstorm is defined to be one that produces a tornado, hail greater than one inch in diameter, or wind gusts in excess of 50 kts. Supercell storms are responsible for a large fraction of severe thunderstorm reports (e.g., 79% of tornadoes according to Trapp et al. (2005)), even though only about 10% of thunderstorms are supercells (Doswell III, 2015), and a key element in forecasting severe thunderstorms is the prediction of where and when supercells will occur (Corfidi, 2017). A supercell is a thunderstorm with a deep, long-lived rotating updraft (mesocyclone). The presence of buoyancy, i.e., convective available potential energy (CAPE), and deep-layer vertical wind shear are important determinants for supercell development. In addition to the magnitude of the vertical shear, the angle between surface and upper-level winds is important for mesocyclone development and persistence. A key quantity is atmospheric helicity, which is computed relative to storm motion and is proportional to vertical wind shear and the amount of wind direction turning from the surface to upper levels (often 0–3 km).

Several recent studies of US tornado reports have concluded that annual numbers of reliably observed tornadoes, i.e., those rated E/EF1 and greater, show slight but statistically insignificant trends downward over time (Brooks et al., 2014; Elsner et al., 2015), whereas measures of tornado outbreaks or clusters show upward trends (Brooks et al., 2014; Elsner et al., 2015; Tippett et al., 2016). Changes in regional tornado activity have also been reported (Agee et al., 2016; Gensini and Brooks, 2018), but there is less evidence for changes in hail and damaging straight-line wind, perhaps due to the poorer quality of the relevant databases.

In view of the limitations of the historical storm record, a valuable alternative is the analysis of meteorological environments associated with severe thunderstorms. As mentioned above, severe thunderstorms, especially supercell storms, are more likely in the presence of high values of CAPE and of certain measures of vertical wind shear (see, e.g., Brooks et al., 2003; Brooks, 2013) such as storm relative helicity (SRH). Weather forecasters have routinely used such quantities for two decades to interpret observations and the output of numerical weather prediction models (Johns et al., 1993; Rasmussen and Blanchard, 1998; Doswell III et al., 1996), and they are also useful in climatological studies, especially in areas outside the US without

¹[http://www.willisre.com/Media_Room/Press_Releases_\(Browse_All\)/2017/WillisRe_Impact_of_ENSO_on_US_Tornado_and_Hail_frequencies_Final.pdf](http://www.willisre.com/Media_Room/Press_Releases_(Browse_All)/2017/WillisRe_Impact_of_ENSO_on_US_Tornado_and_Hail_frequencies_Final.pdf)

extensive historical reports (Brooks et al., 2003). The environmental approach can also provide an indication of expected severe thunderstorm activity in a warmer climate based on climate projections that do not resolve thunderstorms explicitly (Trapp et al., 2009; Diffenbaugh et al., 2013). On time-scales between weather forecasts and climate projections, this approach has provided a clearer picture of how ENSO modulates US hail and tornado activity (Allen et al., 2015; Lepore et al., 2017).

However, there are lacunae in previous statistical studies of environments associated with severe thunderstorms. For instance, relationships with ENSO were diagnosed based on *monthly* averages, which are at best indirect proxies for behaviour on the time-scale of weather. Similarly Gensini and Brooks (2018) computed monthly accumulations of daily maxima of a significant tornado parameter. Tippet et al. (2016) used submonthly environmental data but aggregated the results on an annual and US-wide basis. These gaps motivate the present work, which focuses on extremes of the environmental values rather than on monthly averages, and presents results that are spatially and temporally resolved. The framework that we use is statistical extreme-value theory.

Gilleland et al. (2013) apply the conditional extreme-value framework of Heffernan and Tawn (2004) to the product $WS \times W_{\max}$, where WS is a measure of wind shear and $W_{\max} = \sqrt{2 \times \text{CAPE}}$, by conditioning on the 75th percentile of that variable computed across the spatial domain. This approach has the advantage of allowing the study of real spatial patterns under severe conditions, as opposed to approaches looking at pointwise maxima. They show some temporal variations in the mean simulated values from their model.

Mannshardt and Gilleland (2013) perform an unconditional univariate analysis in which they fit the generalized extreme-value (GEV) distribution to the annual maxima of $WS \times W_{\max}$ and establish the existence of a time trend in the GEV location parameter. Heaton et al. (2011) consider three Bayesian hierarchical extreme-value models based on exceedances over a high threshold for $WS \times W_{\max}$, their third model being based on a Poisson point process with a yearly time trend. Neither paper clarifies whether this trend is attributable to both CAPE and WS or only to one of them. Moreover, both articles consider trends in annual quantities and thus cannot detect month-specific features, and they do not account for multiple testing, though this issue is briefly addressed in Gilleland et al. (2008). Finally, they consider only time as a covariate.

Our study covers a large part of the contiguous US for individual months from 1979 to 2015 and we consider CAPE, SRH (0–3 km) and the combined variable $\text{PROD} = \sqrt{\text{CAPE}} \times \text{SRH}$ separately. To motivate our use of PROD, we consider the discriminant line defined in Brooks et al. (2003, Equation (1)), which is one of the first thresholds used to distinguish low and high likelihoods of severe thunderstorm occurrence using a function of CAPE and vertical shear. This equation can be rewritten as $S6 \times \text{CAPE}^{0.62} = 18.60$, where $S6$ is the 0–6 km shear. Replacing $S6$ with 0–3 km SRH and approximating the power 0.62 by 0.5 leads to a discriminant line of the form $\text{SRH} \times \sqrt{\text{CAPE}} = c$, i.e., $\text{PROD} = c$, where c is a real constant, and shows that high values of

Chapter 1. Trends in the extremes of environments associated with severe US thunderstorms

PROD can be expected to be indicative of an elevated risk of severe thunderstorms. PROD has already been used as a proxy for severe thunderstorms (e.g., Tippett et al., 2016) and the plot of Figure 1 in Brooks et al. (2003) is little changed by replacing S6 with 0–3 km SRH (not shown). More generally, the product of CAPE and two shear-related variables (different or not), or equivalently its square root, is commonly used as an indicator of the likelihood of severe thunderstorm occurrence. For instance, the significant tornado parameter (STP) and the supercell composite parameter (SCP) involve the product of CAPE, S6 and 0–1 km SRH, and the product of CAPE, S6 and 0–3 km SRH, respectively (e.g., Thompson et al., 2003).

To ensure the soundness of our results we carefully check the suitability of the GEV model and the use of time and ENSO as explanatory variables in its location parameter, and we account for multiple testing by implementing the false discovery rate procedure of Benjamini and Hochberg (1995). As stated in Gilleland et al. (2013, Section 1), in addition to studying PROD, it is insightful to consider its components separately. Furthermore, accounting for multiple testing is essential when testing many hypotheses simultaneously, as stressed by Gilleland et al. (2013, Section 4).

We find a significant time trend in the GEV location parameter for PROD maxima in April, May and August (and to a lesser extent in June and December), in CAPE maxima in April, May and June (and to a lesser extent in August, November and January), and in SRH maxima in May (and to a lesser extent in April). The trends in CAPE maxima are striking, because CAPE is expected to increase in a warming climate (Del Genio et al., 2007; Van Kloooster and Roebber, 2009) and are relevant to rainfall extremes (Lepore et al., 2015), but have not previously been observed over the US. April and May are important months for PROD, as severe thunderstorms are frequent at this period. The corresponding time slope is positive in regions of the US where severe thunderstorms are already common, which may have implications for risk assessment and management. Our study also reveals that ENSO can explain variation in the GEV location parameter for PROD and SRH maxima in February. The corresponding slope is negative over most of the region we consider, possibly suggesting an increased risk of high storm impacts in February during La Niña years. Our results differ from those of Heaton et al. (2011), Mannshardt and Gilleland (2013) and Gilleland et al. (2013), but are fairly consistent with those obtained by Gensini and Brooks (2018), who inter alia consider the numbers of tornado reports.

The remainder of the chapter is organized as follows. §1.3 presents the data and a brief exploratory analysis. We describe our statistical approach and demonstrate its relevance in §1.4. §1.5 details our main results, and §1.6 summarises our findings and discusses them.

1.3 Data and exploratory analysis

The data we investigate consist of 3-hourly time-series of 0–180 hPa convective potential energy (CAPE, J kg^{-1}) and 0–3 km storm relative helicity (SRH, $\text{m}^2 \text{s}^{-2}$) from 1 January 1979 at 00:00 to 31 December 2015 at 21:00. The region covered is a rectangle over the contiguous US

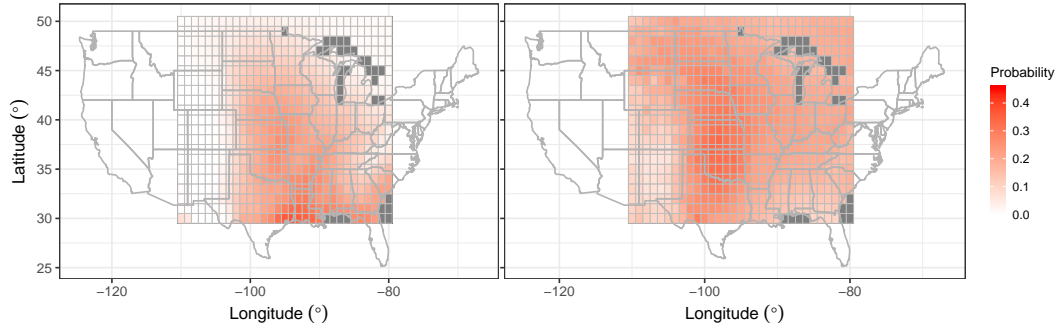


Figure 1.1 – Empirical pointwise probabilities of 3-hourly CAPE exceeding 1400 Jkg^{-1} (left) and SRH exceeding $170 \text{ m}^2\text{s}^{-2}$ (right) for the entire period 1979–2015. Dark grey corresponds to grid points where no observations are available.

from -110° to -80° longitude and 30° to 50° latitude and the resolution is 1° longitude and 1° latitude. These data constitute a coarse version of reanalysis data from the North American Regional Reanalysis (NARR); the original resolution is 32 km longitude and 32 km latitude (see, e.g., Mesinger et al., 2006). The region contains 651 grid points, with no data available for 32 grid points over the sea or lakes. Using these time series, we build 3-hourly time series of $\text{PROD} = \sqrt{\text{CAPE}} \times \text{SRH}$, measured in m^3s^{-3} .

As a physical covariate we use monthly values of the NINO 3.4 index ($^\circ\text{C}$) from 1979 to 2015, taken from the ERSSTv5 data set available on the NOAA Climate Prediction Center website.

Figure 1.1 shows the empirical pointwise probabilities that CAPE and SRH exceed thresholds corresponding to roughly the 90th percentile of each variable across the entire region. There is a clear North-South gradient for CAPE probabilities, while the regional spatial pattern for SRH suggests that the high values cluster towards the centre of the region.

Figure 1.2 shows an increase in the exceedance probabilities for PROD at many grid points over the decades; a similar result is visible for SRH, but less so for CAPE. This increase is of interest for risk assessment, especially in regions with a high risk of severe thunderstorms. Figure 1.2 strongly suggests the presence of a temporal trend in the maxima, but there seems to be no geographical shift, notwithstanding the results of Gilleland et al. (2013).

The top left panel of Figure 1.3 shows a positive correlation between PROD April maxima and time for many grid points, and the middle panels show a positive linear time trend for April maxima of PROD, CAPE and SRH in the subregion indicated. The top right panel shows strong negative correlation between PROD February maxima and ENSO at many grid points, while the scatter-plots in the bottom panels show a roughly linear negative trend for all variables. These analyses underscore the need to incorporate ENSO into our statistical modelling of maxima.

Chapter 1. Trends in the extremes of environments associated with severe US thunderstorms

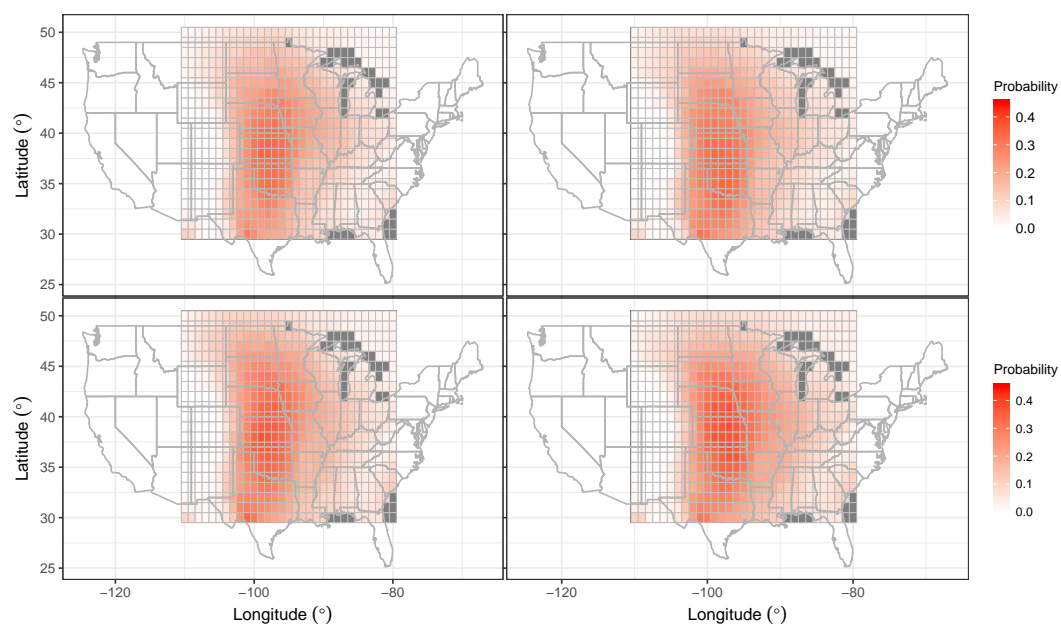


Figure 1.2 – Empirical pointwise probabilities of 3-hourly PROD exceeding $3300 \text{ m}^3 \text{ s}^{-3}$ during the periods 1979–1987 (top left), 1988–1996 (top right), 1997–2005 (bottom left) and 2006–2015 (bottom right).

1.3. Data and exploratory analysis

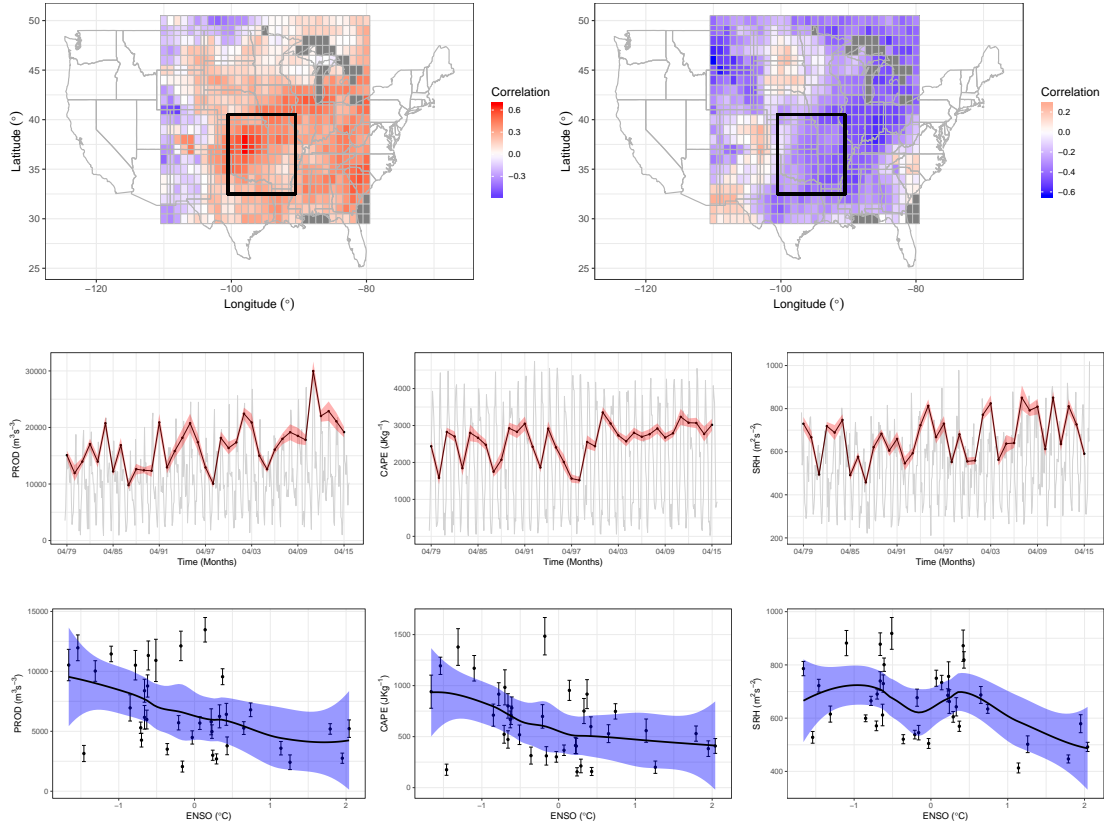


Figure 1.3 – Exploratory analysis for monthly maxima: The top panels show the correlation map with time (in years from 1 to 37) for PROD April maxima (left) and the correlation map with ENSO for PROD February maxima (right). The middle and bottom panels display PROD (left), CAPE (centre) and SRH (right) analyses on a subregion indicated by the black rectangle drawn on the correlation maps. The middle panels show the region-averaged monthly maxima time series across all 444 months in light grey, the region-averaged April maxima time series in black and its 95% confidence interval bounds indicated by the red shaded region. Every point in the time series is the averaged maxima across all grid points in the subregion indicated before, for a particular month and a particular year. The bottom panels show scatter-plots of the region-averaged February maxima with ENSO, along with the 95% confidence interval bounds at each point indicated by the whiskers. The black line represents the best fitted local regression trend estimate, with its 95% confidence interval bounds indicated by the shaded blue region.

1.4 Methodology

1.4.1 Modelling of maxima

Risk assessment entails the estimation of return levels associated with very high return periods and of the probabilities of observing events so extreme that they have never occurred before. Extreme-value theory provides a solid framework for the extrapolation needed to perform these tasks for the maxima of PROD, CAPE and SRH. Here we present the statistical background to the results in §1.5; for further explanation and references see Coles (2001) or Davison and Huser (2015).

Let M_n denote the maximum of independent and identically distributed random variables X_1, \dots, X_n . The extremal types theorem states that if there exist sequences $\{a_n\} > 0$ and $\{b_n\} \in \mathbb{R}$ such that $(M_n - b_n)/a_n$ has a non-degenerate limiting distribution as $n \rightarrow \infty$, then this distribution must be of generalized extreme-value form,

$$\text{GEV}_{\eta, \tau, \xi}(x) = \begin{cases} \exp \left[- \{1 + \xi(x - \eta)/\tau\}_+^{-1/\xi} \right], & \xi \neq 0, \\ \exp \left[- \exp \{-(x - \eta)/\tau\}_+ \right], & \xi = 0, \end{cases} \quad x \in \mathbb{R},$$

where ξ and η are real-valued, $\tau > 0$ and, for any real a , $a_+ = \max\{a, 0\}$. If n is large enough, this theorem suggests using the approximation

$$\mathbb{P}(M_n \leq x) \approx \text{GEV}_{\eta, \tau, \xi}(x), \quad x \in \mathbb{R}, \quad (1.1)$$

for suitably chosen η , τ and ξ , which are location, scale and shape parameters. The latter defines the type of the distribution: $\xi > 0$, $\xi < 0$ and $\xi = 0$ correspond to the Fréchet, Weibull and Gumbel types and allow quite different statistical behaviours, with the first giving a heavy upper tail with polynomial decay, the second modelling bounded variables, and the third an intermediate case, unbounded with an exponentially-decaying upper tail.

The GEV approximation for maxima remains valid if the variables are dependent, provided that distant extremes are “nearly independent”, or, more formally, Leadbetter’s $D(u_n)$ condition is satisfied. We shall see below that near-independence is credible for our time series, so it is plausible that (1.1) applies.

The results above provide a natural model for maxima of stationary sequences. To apply this model we split the data into blocks of equal lengths and compute the maximum of each block. Assume that we have T blocks of length n and let $M_n^{(1)}, \dots, M_n^{(T)}$ denote the corresponding maxima. If n is large enough, the distribution of the $M_n^{(t)}$ is approximately (1.1), and inference from fitting this model to the $M_n^{(t)}$ is commonly called the block maximum method. As noted in §1.3, PROD, CAPE and SRH maxima exhibit a time trend and/or a relation with ENSO for some months, and we can allow the GEV parameters to depend upon these variables. Figure 1.4 and results in §1.5 show that the temporal or ENSO effects only appear for certain months. For instance, time trends for PROD, CAPE and SRH are mainly present in April and

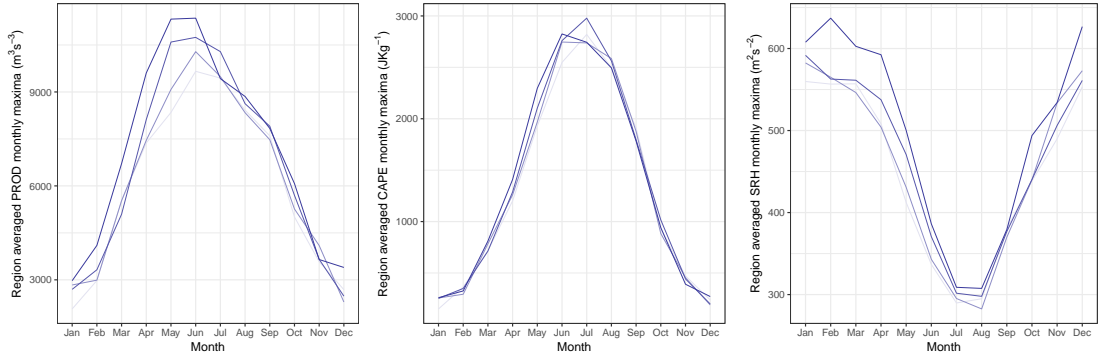


Figure 1.4 – Whole region-averaged monthly maxima of PROD (left), CAPE (centre) and SRH (right). The four lines coloured from light blue to dark blue correspond to the time periods 1979–1987, 1988–1996, 1997–2005 and 2006–2015, respectively.

May, April to June and April and May, respectively. We therefore choose our blocks to be the months and study each month separately, fitting the models

$$M_n^{(t)} \sim \text{GEV}_{\eta_{ti}(t), \tau_{ti}, \xi_{ti}}, \quad \eta_{ti}(t) = \eta_{0,ti} + \eta_{1,ti} t, \quad t = 1, \dots, T, \quad (1.2)$$

and

$$M_n^{(t)} \sim \text{GEV}_{\eta_{en}(t), \tau_{en}, \xi_{en}}, \quad \eta_{en}(t) = \eta_{0,en} + \eta_{1,en} \text{ENSO}_t, \quad t = 1, \dots, T, \quad (1.3)$$

where the subscripts ti and en refer to the dependence on time and on ENSO respectively, $\eta_{0,ti}$, $\eta_{1,ti}$, $\eta_{0,en}$, $\eta_{1,en}$, ξ_{ti} and ξ_{en} are real-valued, τ_{ti} and τ_{en} are positive, ENSO_t is the value of ENSO in that month for year t , and n equals 224, 232, 240 or 248, depending on the number of days in the month, as we have eight observations per day. Figure 1.3 suggests that effects of time and ENSO on maxima are roughly linear and impact the location parameter η only, so we consider constant scale and shape parameters; it is generally inappropriate to allow the shape parameter to depend on a covariate owing to the large uncertainty of its estimate. The time trend induces non-stationarity between the blocks (i.e., across years) but does not violate the within-block stationarity assumption; see below. Figure 1.4 suggests that the time trend does not stem from a shift of seasonality.

We compute the monthly maximum for each month and a given grid point and thereby obtain the maxima $M_{31}^{(1)}, \dots, M_{31}^{(37)}$ for January, say. We then fit the models (1.2) and (1.3) by numerical maximum likelihood estimation for each month and grid point.

Recall that, provided the block size n is large enough, within-block stationarity and the $D(u_n)$ condition ensure the validity of (1.1) and hence allow us to consider the models (1.2) and (1.3). To check the plausibility of these two properties, we consider the 3-hourly time series of PROD, CAPE and SRH at 50 representative grid points. For each block (associated with a triplet grid point-month-year), we fit several autoregressive-moving average (ARMA) processes to the corresponding time series, choose the fit that minimizes the Akaike information criterion (AIC), and use a Box–Pierce procedure to assess the independence of the corresponding

Chapter 1. Trends in the extremes of environments associated with severe US thunderstorms

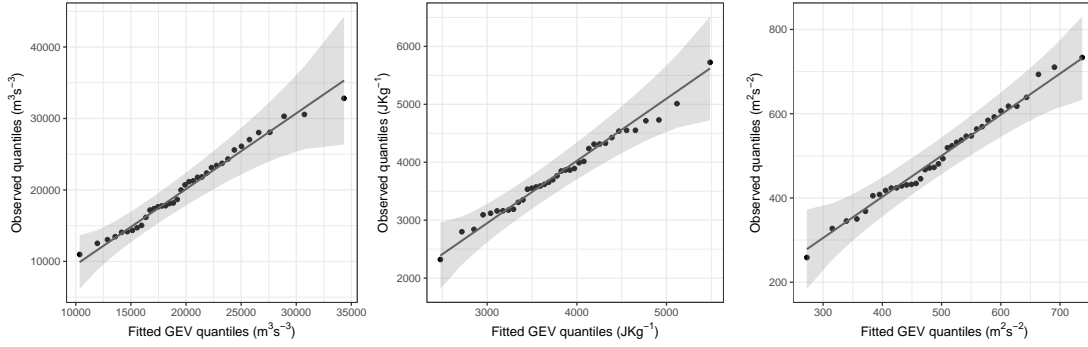


Figure 1.5 – Assessment of the in-sample fit of the GEV: QQ plots for PROD (left), CAPE (centre) and SRH (right) May maxima at the grid point whose South-West corner has coordinates 32° latitude and -99° longitude. The shaded regions indicate the 95% confidence bounds.

residuals; we find no systematic departure from independence or stationarity. Often the residual distribution appears to lie in the Fréchet or Gumbel maximum-domains of attraction, and Embrechts et al. (1997, Section 5.5) show that in such cases convergence of the maxima to the GEV limit occurs even for ARMA processes. Hence the time series of data within the months seem to satisfy both stationarity and the $D(u_n)$ condition. Choosing the months as blocks thus appears reasonable, as is confirmed by our analysis in the following section, whereas choosing seasons or years as blocks would mask many interesting features, and the sample size associated with day- or week-long blocks is too low for the GEV approximation (1.1) to be reasonable.

1.4.2 Assessment of GEV fit

At each grid point i and month j , we fit the GEV model to the monthly maxima, as described in §1.4.1, resulting in location, scale and shape parameter estimates $\hat{\eta}_{i,j}$, $\hat{\tau}_{i,j}$ and $\hat{\xi}_{i,j}$. We use the Kolmogorov–Smirnov test to assess the distributional proximity between this fitted model and the empirical distribution of the 37 observed monthly maxima. For PROD, CAPE and SRH, in most months, the fit appears acceptable at the 5% level at all grid points. These good in-sample fits of the GEV distribution for all variables are confirmed by the quantile-quantile (QQ) plots, which are displayed for one grid point in Figure 1.5.

However, these results do not take into account the fitting of the model to the data, which systematically decreases the values of the Kolmogorov–Smirnov statistic. In order to make an informal allowance for this decrease, for each grid point i and month j we:

1. fit the GEV distribution using the pooled observations from the eight grid points nearest to i to obtain $\hat{\eta}_{po,i,j}$, $\hat{\tau}_{po,i,j}$ and $\hat{\xi}_{po,i,j}$;
2. and then use a Kolmogorov–Smirnov test to check the agreement between the “out-sample” GEV fit with parameters $\hat{\eta}_{po,i,j}$, $\hat{\tau}_{po,i,j}$ and $\hat{\xi}_{po,i,j}$, and the empirical distribution

Variable	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
PROD	42	27	50	29	52	58	67	71	55	27	39	34
Sim PROD 5%	37	34	33	33	35	35	37	38	38	36	36	37
Sim PROD 95%	57	57	55	55	57	58	57	58	57	58	58	57
CAPE	59	33	48	34	60	64	71	90	67	43	58	74
Sim CAPE 5%	41	36	37	36	36	36	38	42	39	36	36	37
Sim CAPE 95%	59	58	60	56	57	58	61	63	61	60	58	59
SRH	36	23	24	21	22	42	42	34	35	26	24	36
Sim SRH 5%	36	36	34	35	34	36	36	36	34	36	36	34
Sim SRH 95%	60	59	59	53	57	57	57	58	58	58	61	57

Table 1.1 – Assessment of the out-sample fit of the GEV: Number of rejections from our out-sample Kolmogorov-Smirnov test (at the 5% level and without accounting for multiple testing) for each variable and each month. For each part (corresponding to one variable), the first row gives the observed number of rejections whereas the second and third ones provide the 5% and 95% quantiles of the empirical distributions of the number of rejections obtained from the simulation study.

of the 37 observed monthly maxima at grid point i .

We then repeat these two steps 100 times with data simulated from independent GEV models fitted at each grid point and compute the 5% and 95% quantiles of the empirical distribution of the number of rejections. Table 1.1 shows that, for all variables, the observed numbers of rejections are low compared to the number of grid points (619), especially as we did not account for multiple testing. Moreover, they are not hugely different from those obtained in the simulation study, although often slightly above the 95% quantile in the case of CAPE and slightly below the 5% quantile for SRH and PROD. These discrepancies may be explained by the substantial spatial dependence present in the data but not in the simulations. This analysis supports the use of the GEV model at grid points at which no data are available and thus goes beyond the initial goal of assessment of its fit.

As it fits them adequately, we conclude that the GEV distribution provides a suitable model for the monthly maxima of PROD, CAPE and SRH.

1.4.3 Testing procedure

General

In §1.5, we assess whether time and ENSO affect the GEV location parameter for PROD, CAPE and SRH at each of the $m = 619$ grid points.

Chapter 1. Trends in the extremes of environments associated with severe US thunderstorms

We first discuss the statistic used to test the significance of time and ENSO, respectively, in (1.2) and (1.3), at a single grid point. In the first case, we have to test the null and alternative hypotheses

$$H_0 : \eta_{1,\text{ti}} = 0 \quad \text{versus} \quad H_A : \eta_{1,\text{ti}} \neq 0,$$

by comparing the fits of the models

$$\mathcal{M}_0 : \eta_{\text{ti}}(t) = \eta_{0,\text{ti}}, \quad \mathcal{M}_1 : \eta_{\text{ti}}(t) = \eta_{0,\text{ti}} + \eta_{1,\text{ti}}t, \quad t = 1, \dots, 37,$$

and similarly for ENSO. We let $\ell_0(\mathcal{M}_0)$ and $\ell_1(\mathcal{M}_1)$ denote the maximized log-likelihoods for the models \mathcal{M}_0 and \mathcal{M}_1 and compute the signed likelihood ratio statistic $\tilde{T} = \text{sgn}(\hat{\eta}_{1,\text{ti}})[2\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\}]^{1/2}$, where $\text{sgn}(\hat{\eta}_{1,\text{ti}})$ is the sign of the estimated trend under model \mathcal{M}_1 ; \tilde{T} has an approximate standard Gaussian distribution under H_0 , and the corresponding p -value is $p = 2\Phi(-|\tilde{t}|)$, where \tilde{t} is the observed value of \tilde{T} and Φ denotes the standard Gaussian distribution function.

This test would be valid if applied at a single pre-specified grid point, but we must make allowance for the facts that we shall perform the same test at m grid points, and that spatial patterns in the effects of time and ENSO are likely to induce correlation among the p -values for nearby grid points. We now discuss how to obtain a valid testing procedure despite these facts.

Multiple testing

A popular approach for multiple testing in climatology is the field significance test of Livezey and Chen (1983), but unfortunately this gives little insight about where the results are significant, which is of high interest to us. The regression approach of DelSole and Yang (2011) has the same drawback. Among methods to identify which of the grid points have significant results are those, such as the Bonferroni procedure, that bound the probability that the number of falsely rejected null hypotheses exceeds unity. When the number of hypotheses is high, however, such methods are so stringent that their power is very low.

Benjamini and Hochberg (1995) introduce the false discovery rate (FDR), namely the expected proportion of incorrect rejections out of all rejections, and propose a procedure to ensure that the FDR is below a given level q when performing multiple testing. Their approach, which we call the BH procedure, would reject H_0 at all grid points i such that $p_i \leq p_{(k)}$, where $p_{(1)} \leq \dots \leq p_{(m)}$ are the ordered p -values and

$$k = \max \left\{ i : p_{(i)} \leq q \frac{i}{m} \right\}.$$

The BH procedure ensures that the FDR is below qm_0/m , where m_0 denotes the unknown number of grid points at which H_0 is true, and is said to control the FDR at level qm_0/m .

For a chosen q , let S_q be the number of grid points at which a particular covariate is declared

significant by the BH procedure. Then we expect the true number of grid points where the relation is significant, m_A , to satisfy

$$m_A \geq (1 - q)S_q. \quad (1.4)$$

As the BH procedure ensures that the FDR is not more than qm_0/m , we may argue a posteriori that we have controlled the FDR at level

$$q^{(1)} = \frac{q\{m - (1 - q)S_q\}}{m} \leq q \frac{m_0}{m},$$

which entails that $m_A \geq (1 - q^{(1)})S_q$. If we iterate this argument by defining

$$q^{(n+1)} = \frac{q\{m - (1 - q^{(n)})S_q\}}{m}, \quad n = 1, 2, \dots,$$

we see that the effective level at which we have controlled the FDR is $q_{\text{lim}} = \lim_{n \rightarrow \infty} q^{(n)}$. The limit is often well-approximated after just a few iterations, and then we have

$$m_A \geq (1 - q_{\text{lim}})S_q. \quad (1.5)$$

The BH procedure was originally shown to be valid for independent test statistics, but Benjamini and Yekutieli (2001, Theorem 2.1) prove that it controls the FDR at level qm_0/m if the statistics have a certain form of positive dependence. Ventura et al. (2004) apply this procedure to simulations representative of climatological data and covering the range of correlation scales likely to be encountered in practice, and find empirically that it controls the FDR at level qm_0/m . Yekutieli and Benjamini (1999) and Benjamini and Yekutieli (2001) propose two modifications to account for more general dependence between the test statistics. The first is complex and does not much improve on the BH procedure, whereas the second is applicable for any dependence structure but has greatly reduced power, so Ventura et al. (2004) recommend the use of the BH procedure. Test statistics and p -values based on the data there are clearly dependent, but as our data resemble those considered in Ventura et al. (2004), applying the BH procedure at level q should control the FDR at level qm_0/m , where m_0 denotes the unknown number of grid points at which H_0 is true from m p -values.

1.5 Results

In this section we quantify the effects of time and ENSO in the GEV location parameter and study their significance, using $q = 0.05$ and $q = 0.2$, corresponding to control of the FDR at the nominal levels 5% and 20%. In each case we first discuss PROD, which is the main variable of interest for severe thunderstorm risk, and then consider CAPE and SRH.

We begin with the effect of time. Table 1.2 shows that many of the 619 grid points exhibit a significant time trend for PROD in April, May and August (and to a lesser extent in June and

Chapter 1. Trends in the extremes of environments associated with severe US thunderstorms

December). In April, this number equals 313 at the 20% level, so (1.4) implies that at least 250 of these grid points indeed have a trend; with (1.5), this number rises to 278. Figure 1.6 indicates that, in April, the North-East, a very wide South-East corner and the South-West, show significant time trends. In the first two regions, $\hat{\eta}_{1,ti}$ is positive, corresponding to an increasing risk of severe thunderstorm impacts, particularly in already risky regions. Similar conclusions may be drawn from Figure 1.7 in the case of May, though the South-East is less prominent. The highest slope value corresponds to an annual increase of PROD maxima of about 3% of the corresponding PROD maximum recorded in 1979. Mannshardt and Gilleland (2013) and Heaton et al. (2011) do not find such a significantly positive time trend over the entire region most at risk, sometimes called tornado alley, nor do they find significantly positive trends in the North-East of our region, whereas they find a significant positive trend towards the West. These differences probably arise because these earlier papers consider a less recent period (1958–1999), their product variable is slightly different than ours, and they study annual instead of monthly maxima. The discrepancies with Heaton et al. (2011) may also be due to their use of a different, Bayesian hierarchical, approach. The evolution found by Gilleland et al. (2013) between the second (1979–1992) and the third (1993–1999) period is quite consistent with our trends in Spring; for the other seasons, however, the results differ appreciably. There are also many dissimilarities in the changes between the first (1958–1978) and the second (1979–1992) periods, but the first period does not belong to the time range we consider. Gilleland et al. (2013) consider the mean simulated values conditional on the total amount of energy being large, and then not all grid point values need be extreme, whereas we analyse maxima at each grid point. Moreover, the trends we find account for the year-to-year variation, whereas in Gilleland et al. (2013), changes can only be assessed by comparing values for three successive periods of about 15 to 20 years. The positive time trends we detect in Spring appear quite consistent with the results of Gensini and Brooks (2018), who use much more recent data than the papers previously described. The remaining differences, especially for Texas, may arise for the following reasons. First, as PROD is only an indicator of severe weather, there are necessarily discrepancies with results based on effective tornado reports. Second, PROD slightly differs from STP, so the corresponding results may differ somewhat. Furthermore, the findings of Gensini and Brooks (2018) about reports concern the total number of tornadoes per year, and those about STP are not based on the maxima of that variable.

Regarding CAPE, April, May and June (and to a lesser extent, August, November and January) show many grid points with a significant time trend. For April and May, Figures 1.6 and 1.7 show significantly negative $\hat{\eta}_{1,ti}$ in the West, contrasting with a significantly positive trend in the center and the East. As pointed out by Trapp et al. (2009) and Diffenbaugh et al. (2013), a positive time trend for CAPE is expected in a context of climate change. However, to the best of our knowledge, an *observed* trend has not been previously reported in the literature.

For SRH, May and to a lesser extent April have many significantly positive grid points spread approximately uniformly except in a large South-West corner in April and a large South-East corner in May. The significance for PROD in April and May comes from both CAPE and SRH.

Figures 1.6 and 1.7 suggest that the significant positive time trend in the riskiest part of the US stems mainly from CAPE in April and from SRH in May. Overall, no seasonal pattern appears: CAPE seems to drive PROD in January, April, August, November and December, whereas SRH seems to drive it in February, May, June and September. For March, July and October, there is no clear driver. Anyway, trying to relate the behaviour of PROD maxima with that of CAPE and SRH maxima has its limitations. Indeed, the maximum of PROD may not equal the product of the square root of CAPE maximum and the maximum of SRH, as their maxima may not coincide.

We now comment on the effect of ENSO. For PROD, Table 1.2 reveals that many grid points exhibit a significant relation in February. Figure 1.8 indicates that $\hat{\eta}_{1, \text{en}}$ is negative at those and that the main regions concerned are the North-East, the South-Center and the North-West; we expect higher PROD maxima during La Niña years in these regions. The highest slope absolute value corresponds to a decrease of PROD maxima per unit of ENSO of about 10% of the corresponding basic level of PROD maximum.

There is no strikingly significant result for CAPE, although Allen et al. (2015) found ENSO signals in CAPE seasonal averages for winter and spring, not accounting for multiple testing.

For SRH, Figure 1.8 shows that almost all grid points in the region exhibit significance in February, apart from a strip in the North and a tiny diagonal strip in the South-East corner. The estimate $\hat{\eta}_{1, \text{en}}$ is highly negative in most of the region but very positive in the extreme South-East, with a very rapid change in sign, presumably due to proximity with the Gulf of Mexico. There is a significant negative relation in regions at risk of thunderstorms or large-scale storms, for which SRH plays an essential role. The risk of large impacts may increase during La Niña years. A relationship between seasonal averages of SRH and ENSO in winter was noticed by Allen et al. (2015). Finally, Figure 1.8 suggests that CAPE contributes more than SRH to PROD in terms of significance, though the relation with ENSO is more pronounced for SRH than for CAPE.

We also consider the residuals of PROD, CAPE and SRH maxima after accounting for ENSO or temporal effects. For instance, if we observe a time trend, the idea of considering the residuals after accounting for ENSO is to determine whether the time trend is explained by ENSO. This allows us to determine whether the time and ENSO effects are “independent”.

In the case of PROD, Table 1.2 shows that removing ENSO does not much decrease the number of grid points exhibiting a significant time trend; there is a slight decrease for April but a small increase for some other months. Accounting for the time trend, on the other hand, can slightly increase the number of grid points showing a significant relation with ENSO. For CAPE, removing ENSO decreases the number of grid points exhibiting a significant time trend for March, but there is a slight increase for other months, whereas accounting for time slightly decreases the number of grid points showing a significant relation with ENSO in January and March only, with a slight increase in other months. Regarding SRH, removing ENSO decreases the number of grid points exhibiting a significant time trend in February but there is little

Chapter 1. Trends in the extremes of environments associated with severe US thunderstorms

Variable	Covariate	q	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
PROD	Time	0.05	7	0	1	41	36	0	0	36	2	0	0	22
	Time	0.2	40	0	4	313	203	81	13	148	23	0	0	98
	ENSO	0.05	0	58	10	0	0	1	0	0	0	0	0	1
	ENSO	0.2	1	172	26	0	3	3	0	0	0	0	0	1
CAPE	Time	0.05	37	13	28	109	60	89	18	55	4	0	30	1
	Time	0.2	92	37	73	268	273	206	75	133	35	40	134	16
	ENSO	0.05	15	0	0	0	0	2	2	0	0	0	1	1
	ENSO	0.2	27	11	21	0	0	3	16	14	0	1	6	13
SRH	Time	0.05	0	1	0	7	43	2	1	7	0	0	0	0
	Time	0.2	15	44	4	138	230	14	50	45	6	0	0	27
	ENSO	0.05	0	255	0	0	1	0	0	0	0	0	0	0
	ENSO	0.2	3	384	59	18	4	0	8	7	4	1	0	82
PROD res.	Time	0.05	7	0	2	30	88	0	0	41	2	0	0	38
	Time	0.2	50	16	6	274	221	86	21	137	18	0	2	100
CAPE res.	Time	0.05	35	20	15	87	96	89	25	59	9	0	19	2
	Time	0.2	88	46	51	219	267	223	91	139	54	41	120	29
SRH res.	Time	0.05	0	0	0	7	38	2	1	7	0	0	0	0
	Time	0.2	20	1	6	126	241	7	46	41	1	0	0	60
PROD res.	ENSO	0.05	1	66	8	0	0	1	0	0	0	0	0	7
	ENSO	0.2	1	178	26	0	49	3	0	0	0	0	0	33
CAPE res.	ENSO	0.05	1	0	0	0	0	3	5	1	0	0	0	2
	ENSO	0.2	21	38	0	1	0	4	17	16	0	0	1	21
SRH res.	ENSO	0.05	0	209	0	0	4	0	0	0	1	0	0	0
	ENSO	0.2	1	359	20	38	14	0	3	7	2	1	0	63

Table 1.2 – Number of grid points where $\hat{\eta}_{1,ti}$ and $\hat{\eta}_{1,en}$ are significant for PROD, CAPE and SRH maxima for each month (top); number of grid points where $\hat{\eta}_{1,ti}$ is significant for PROD, CAPE and SRH maxima residuals after accounting for the relation with ENSO (middle); number of grid points where $\hat{\eta}_{1,en}$ is significant for PROD, CAPE and SRH maxima residuals after accounting for the relation with time (bottom). We have accounted for multiple testing using the BH procedure with the values of q displayed.

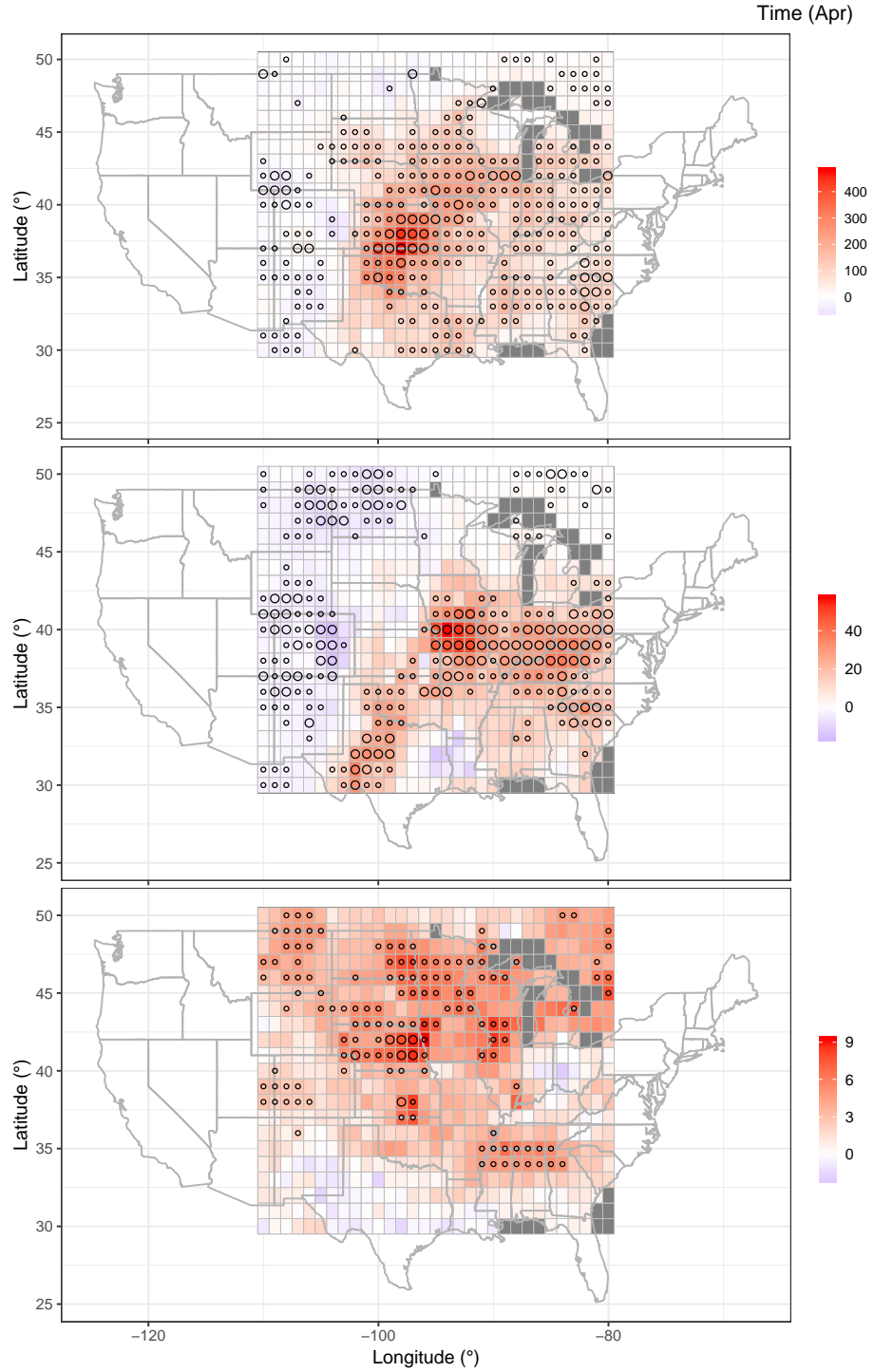


Figure 1.6 – Values and significance of the slope $\hat{\eta}_{1,ti}$ for PROD (top), CAPE (middle) and SRH (bottom) maxima in April. Large and small circles indicate significance (after accounting for multiple testing using the BH procedure) at any level not lower than 5% and 20%, respectively. The units of $\hat{\eta}_{1,ti}$ are $\text{m}^3\text{s}^{-3}\text{yr}^{-1}$, $\text{Jkg}^{-1}\text{yr}^{-1}$ and $\text{m}^2\text{s}^{-2}\text{yr}^{-1}$ for PROD, CAPE and SRH, respectively. Dark grey corresponds to grid points where no observations are available.

Chapter 1. Trends in the extremes of environments associated with severe US thunderstorms

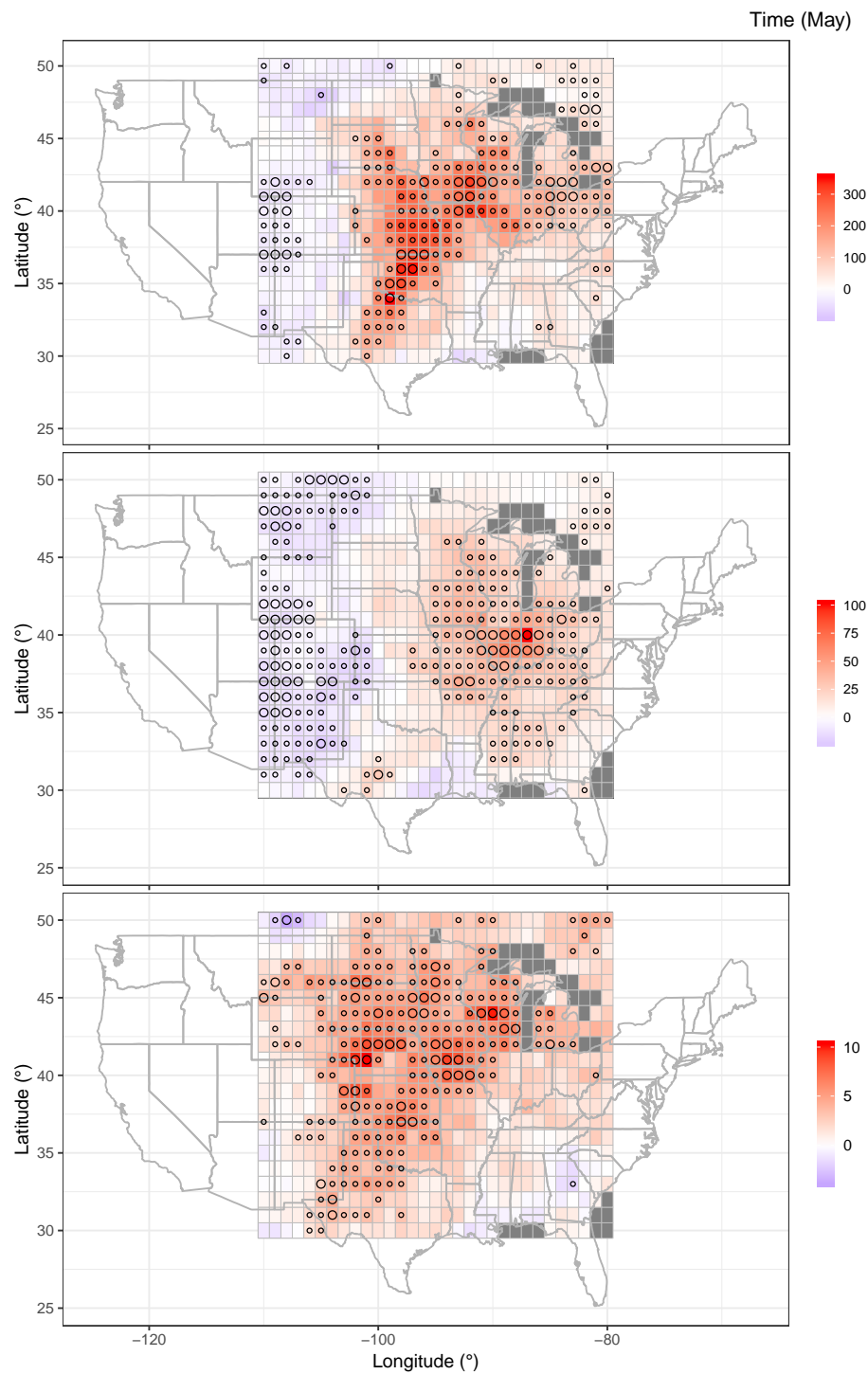


Figure 1.7 – Same content as in Figure 1.6 in the case of May.

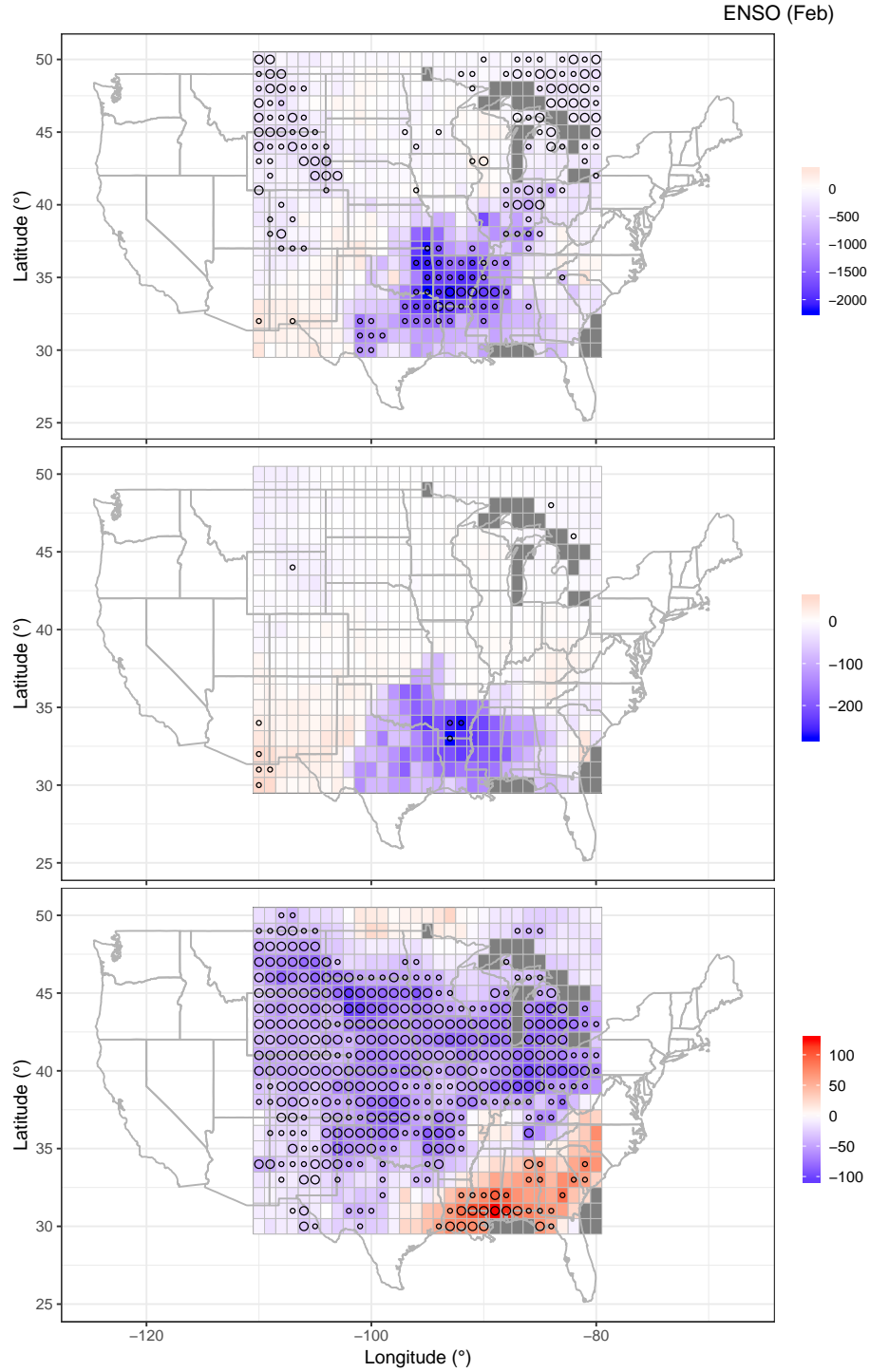


Figure 1.8 – Values and significance of the ENSO coefficient $\hat{\eta}_{1,en}$ for PROD (top), CAPE (middle) and SRH (bottom) maxima in February. Large and small circles indicate significance (after accounting for multiple testing using the BH procedure) at any level not lower than 5% and 20%, respectively. The units of $\hat{\eta}_{1,en}$ are $\text{m}^3 \text{s}^{-3} \text{C}^{-1}$, $\text{Jkg}^{-1} \text{C}^{-1}$ and $\text{m}^2 \text{s}^{-2} \text{C}^{-1}$ for PROD, CAPE and SRH, respectively.

impact for other months. The conclusions are similar when accounting for the time trend and studying the ENSO effect. The maps of the residuals (not shown) indicate that when removing a covariate has little impact on the number of grid points at which the relation with the other covariate is significant, it has almost no impact on their positions either. In summary, the effects of time and ENSO appear “independent”, except for CAPE in January and March and SRH in February.

1.6 Conclusion

This article quantifies the effects of time and ENSO on the distribution of monthly maxima of PROD, CAPE and SRH, which are highly relevant to the risk of severe thunderstorms. The use of the GEV distribution appears justified in our setting. After allowance for multiple testing we detect a significant time trend in the GEV location parameter for PROD maxima in April, May and August, CAPE maxima in April, May and June and SRH maxima in April and May. The observed upward time trend for CAPE, although expected in a warming climate, has not been reported before. April and May are prominent for PROD, as severe thunderstorms are common at this period, and the corresponding trend is positive in parts of the US where the risk is already high, which may have important consequences. We also find ENSO to be a good covariate in the GEV location parameter for PROD and SRH maxima in February. The corresponding relationship is negative over most of the region we consider, suggesting that the risk of storm impacts in February increases during La Niña years. Our results differ from those of Heaton et al. (2011), Mannshardt and Gilleland (2013) and Gilleland et al. (2013), but are quite consistent with those obtained by Gensini and Brooks (2018), perhaps in part because these authors consider a period similar to ours, more recent than in the earlier studies.

We investigate the effects of time and ENSO on the marginal (at each grid point) extremal behaviour of PROD, CAPE and SRH. Quantifying the potential impacts of these covariates on the local spatial extremal dependence of these variables would also be useful for risk assessment. Modelling the extremal dependence between CAPE and SRH might also be informative.

Finally, although concurrently high values of environments such as CAPE and SRH are conducive to severe weather, they do not guarantee that severe thunderstorms will occur. The degree to which changes in environmental characteristics result in changes in thunderstorm properties is also uncertain (Hoogewind et al., 2017). Hence, an interesting issue is the precise implication of an increase of PROD (or SRH) maxima in terms of risk. PROD can be seen as a proxy for the probability of severe thunderstorm occurrence, so it is natural to think that PROD maxima may be aligned with the maximum number of severe thunderstorms daily. If so, then those days with the highest values of PROD would tend to also have the most severe thunderstorm impacts. Better insight into this potential relationship would be valuable.

2 Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

This chapter is based on a preprint of a paper, jointly written with Anthony C. Davison and Erwan Koch. The chapter is self-contained and may overlap with other chapters. The contributions of the doctoral candidate were in producing the results, simulations, plots and tables, and writing the bulk of the paper.

2.1 Abstract

Concurrently high values of convective available potential energy (CAPE) and storm relative helicity (SRH) are conducive to hazardous convective weather associated with severe thunderstorms, so it is valuable to have probability models for the extremes of both variables to account for their spatial dependence. We propose three new tools for fitting max-stable random fields with unknown margins to gridded pointwise monthly maxima: an out-sample selection scheme based on a local max-stability bootstrap test, a model selection procedure that improves on the current state of the art, and a Brown–Resnick model with smooth spline bases incorporated into its parameters, such that non-stationary dependencies can be modelled and estimation strength can be shared across months. Our results show differences in the parameters for CAPE, SRH and a product variable, in the spring and summer, so the corresponding extremes are more localized during these seasons than in the winter. We also find that these seasonal differences tend to be more pronounced during El Niño events.

Keywords: Bootstrap; Brown–Resnick field; environmental statistics; information criteria; max-stability test; splines

2.2 Introduction

In the United States (US), severe thunderstorms are responsible for a substantial fraction of economic and human losses due to natural disasters. Hence it is imperative to model the time evolution of US thunderstorm activity, important drivers of which include climate change and the El Niño–Southern Oscillation (ENSO). A severe US thunderstorm is defined to be one

Chapter 2. Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

that produces tornadoes, hailstones greater than one inch (2.54 cm) in diameter, or wind gusts in excess of 50 kts (1 kt corresponds to approximately 0.51 m s^{-1}). Supercells, which are thunderstorms with a deep and persistent rotating updraft (mesocyclones), are responsible for a large fraction of severe thunderstorm reports (e.g., 79% of tornadoes according to Trapp et al., 2005), even though only about 10% of thunderstorms are supercells (Doswell III, 2015).

The limitations of the available thunderstorm data record (Verbout et al., 2006; Allen and Tippett, 2015; Edwards et al., 2018) due, for example, to observational bias, make its use for modelling non-trivial, so it is worthwhile to consider meteorological environments which are conducive to severe thunderstorms. Such storms, especially supercell storms, are more probable in the presence of elevated values of convective available potential energy (CAPE) and of certain measures of vertical wind shear (e.g., Brooks et al., 2003; Brooks, 2013) such as storm relative helicity (SRH), which have been used by weather forecasters and climatologists for over two decades. High values of the combined variable $\text{PROD} = \sqrt{\text{CAPE}} \times \text{SRH}$ are favorable to severe thunderstorms, and PROD has been used, e.g., by Tippett et al. (2016) and Koch et al. (2021), as a proxy of severe thunderstorm activity, for justification, see Brooks et al. (2003, Equation 1) and Koch et al. (2021, Section 1). In addition to the absence of observational bias, an advantage of using thunderstorm environments instead of thunderstorm reports is that their reanalysis values are available at high and regular spatio-temporal resolution (typically 1° longitude and 1° latitude every hour or 3 hours), which allows the use of many techniques from extreme-value theory and geostatistics. Moreover, these environments allow one to draw conclusions from numerical climate models that do not resolve thunderstorms explicitly.

Our study focuses on severe thunderstorms over a large rectangle over the contiguous US that contains the most risky region of the US, the famous Tornado Alley. We use reanalysis data from the North American Regional Reanalysis (NARR), which consists of 3-hourly time-series of CAPE and SRH from 1979 to 2015. These variables have been used to study how decadal/multi-decadal climate variations or large-scale weather signals, such as ENSO, modulate severe US thunderstorm activity. Mannshardt and Gilleland (2013) fitted the generalized extreme-value (GEV) distribution to the annual maxima of a related product variable in an unconditional univariate analysis and found a time trend in the location parameter. Heaton et al. (2011) considered a Bayesian hierarchical extreme-value model for the same product variable, based on a Poisson point process with a yearly time trend. Allen et al. (2015) and Lepore et al. (2017) showed a clear link between ENSO and seasonal or monthly means of environments during winter and spring in the US, and Koch et al. (2021) investigated the link between ENSO and monthly maxima of PROD and showed that PROD maxima in February are significantly larger during La Niña years.

The aforementioned papers focused on marginal effects (grid point by grid point) only, but data and physical arguments suggest that ENSO and other covariates may also affect how strongly maxima are dependent in space and time. For instance, El Niño episodes tend to warm the northern part of the US (including Alaska) and to make the southern part wetter (Halpert and Ropelewsk, 1992), a combination that is expected to lead to larger areas of instability,

and thus to increase the spatial extent of individual intense events of PROD. As pointwise maxima arise as a mixture of several individual events, the characteristic range of spatial dependence in the field of pointwise maxima is an increasing function of the spatial extent of the individual extreme events (Dombry et al., 2018). Thus, if the spatial extent of extreme events of PROD tends to increase with ENSO, the same will apply to the spatial dependence of its field of pointwise maxima. Apart from being influenced by large-scale weather signals such as ENSO, extreme values of CAPE, SRH and PROD depend on the month, and interactions between weather signals and seasonality are expected, as the effect of ENSO differs from one month to the next. In order to model these features in a flexible way, we propose max-stable random fields whose spatial dependence parameters are linked to covariates that account for seasonality and large-scale weather signals.

Max-stable fields (e.g., de Haan, 1984; de Haan and Ferreira, 2006; Davison et al., 2012) constitute a natural extension of multivariate extreme-value distributions to the infinite-dimensional setting. They are well-suited to modeling spatial extremes, as they arise as the only possible non-degenerate limits of appropriately rescaled pointwise maxima of independent replications of a random field. Commonly-used max-stable models, such as the extremal t (Opitz, 2013) and Brown–Resnick (Brown and Resnick, 1977; Kabluchko et al., 2009) models, involve an underlying Gaussian field whose correlation function or variogram contains a range parameter. We propose to introduce covariates into the range parameter. Although incorporating covariates with trend surfaces in the marginal parameters (GEV parameters) of a max-stable field has been proposed (e.g., Davison et al., 2013, and references therein), to the best of our knowledge, doing so in their dependence parameters has been less explored in the multivariate and spatial contexts. To model temperature maxima, Huser and Genton (2016) developed non-stationary max-stable dependence structures with covariates, while Mhalla et al. (2017) proposed a semi-parametric methodology to estimate non-stationary Pickands functions, which characterise the dependence of max-stable random vectors.

We present a general model and consider the month and ENSO as covariates. We show by simulation that the model parameters can be estimated rather accurately even in a complex setting, and that there are differences in the parameters for PROD, CAPE and SRH in the spring and summer, so the corresponding extremes are more localized during these seasons than in the winter. These seasonal differences are more pronounced during El Niño and La Niña events, though the associated uncertainty can be high.

In environmental applications, data may exhibit asymptotic independence, implying that max-stable fields are unsuitable, and several subasymptotic models have been proposed to alleviate this (e.g., Huser et al., 2021; Huser and Wadsworth, 2019). Thus, one should always assess the suitability of max-stable models in applications. Gabda et al. (2012) and Buhl and Klüppelberg (2016) proposed graphical diagnostics, and here we add a max-stability test based on empirical likelihood that is applicable when data have unknown margins.

Owing to the complexity of the model we consider, estimating the marginal generalized

Chapter 2. Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

extreme-value (GEV) parameters at each grid point and the dependence parameters of the max-stable field in a single step (optimizing the likelihood with respect to all parameters) is too computationally intensive. Thus, estimation must be performed in two steps: we estimate the GEV parameters at each grid point separately, transform the data to have standard Fréchet margins and then fit the dependence parameters using a pairwise likelihood (e.g., Padoan et al., 2010). We show that in such a two-step procedure, it is inappropriate to use the sandwich matrix to compute confidence intervals, and the non-parametric bootstrap gives better coverage. We also show that using the composite likelihood information criterion (CLIC) computed from the sandwich matrix in this two-step procedure is not optimal, and we propose estimating the composite Kullback–Leibler divergence with the non-parametric bootstrap.

The remainder of the chapter is organized as follows. In Section 2.3, we briefly review max-stable fields and their estimation by pairwise likelihood, and present the data and some exploratory analyses. Section 2.4 presents our main methodological contributions: the proposed model, the max-stability test, and the bootstrap-based confidence intervals and CLIC. Section 2.5 is dedicated to the case study: we use the model and methodologies developed in Section 2.4 to answer our concrete question. Section 2.6 summarizes our main contributions and findings and proposes some future perspectives.

2.3 Preliminaries

2.3.1 Max-stable random fields

In the following, $\stackrel{d}{=}$ and $\stackrel{d}{\rightarrow}$ denote equality and convergence in distribution, respectively; in the case of random fields, distribution has to be understood in terms of the set of all finite-dimensional multivariate distributions. A random field $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ is said to be max-stable if there exist sequences of functions $\{a_n(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}_{n \geq 1} > 0$ and $\{b_n(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}_{n \geq 1} \in \mathbb{R}$ such that, for any $n \geq 1$,

$$\left\{ \frac{\max_{i=1}^n Z_i(\mathbf{s}) - b_n(\mathbf{s})}{a_n(\mathbf{s})} : \mathbf{s} \in \mathbb{R}^d \right\} \stackrel{d}{=} \left\{ Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d \right\},$$

where the equality is in distribution and Z_1, \dots, Z_n are independent replicates of Z . Let $\tilde{T}_1, \dots, \tilde{T}_n$ be independent replications of a random field $\{\tilde{T}(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$, and let $\{c_n(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}_{n \geq 1} > 0$ and $\{d_n(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}_{n \geq 1} \in \mathbb{R}$ be sequences of functions. If there exists a non-degenerate random field $\{G(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ such that

$$\left\{ \frac{\max_{i=1}^n \tilde{T}_i(\mathbf{s}) - d_n(\mathbf{s})}{c_n(\mathbf{s})} : \mathbf{s} \in \mathbb{R}^d \right\} \stackrel{d}{\rightarrow} \left\{ G(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d \right\}, \quad n \rightarrow \infty, \quad (2.1)$$

then G is necessarily max-stable (de Haan, 1984). This explains the relevance of max-stable fields in the modelling of pointwise maxima of random fields. If $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ is a max-stable field, then, for any $\mathbf{s} \in \mathbb{R}^d$, $Z(\mathbf{s})$ follows the GEV distribution with location, scale and shape

parameters $\eta_s, \tau_s > 0$ and ξ_s . Max-stable fields having standard Fréchet margins, i.e., such that $\mathbb{P}(Z(\mathbf{s}) \leq z) = \exp(-1/z)$, $z > 0$, $\mathbf{s} \in \mathbb{R}^d$, are said to be simple.

Any simple max-stable field can be written as (de Haan, 1984)

$$Z(\mathbf{s}) = \max_{i=1}^{\infty} R_i U_i(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d, \quad (2.2)$$

where the $(R_i)_{i \geq 1}$ are the points of a Poisson point process on $(0, \infty)$ with intensity function $r^{-2} dr$ and the $(U_i)_{i \geq 1}$, are independent replicates of a non-negative random field $\{U(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ such that $E\{U(\mathbf{s})\} = 1$ for any $\mathbf{s} \in \mathbb{R}^d$. Any field defined by (2.2) is simple max-stable, moreover, and this has enabled the construction of parametric models of max-stable fields. The best known are the Smith (Smith, 1990), Schlather (Schlather, 2002), Brown–Resnick (Brown and Resnick, 1977; Kabluchko et al., 2009), and extremal- t (Opitz, 2013) models; the last two are flexible models that have been found to capture extremes well, and in our case study (Section 2.5) we use the Brown–Resnick model. Write $W(\mathbf{s}) = \exp\{\varepsilon(\mathbf{s}) - \text{Var}(\varepsilon(\mathbf{s}))/2\}$, $\mathbf{s} \in \mathbb{R}^d$, where Var denotes variance, and $\{\varepsilon(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$ is a centred Gaussian random field with stationary increments and semivariogram γ . Using $U = W$ in (2.2) leads to the Brown–Resnick random field associated with the semivariogram γ . A frequently used isotropic semivariogram is $\gamma(\mathbf{s}) = (\|\mathbf{s}\|/\rho)^\alpha$, $\mathbf{s} \in \mathbb{R}^d$, where $\rho \geq 0$ and $\alpha \in (0, 2]$ are the range and smoothness parameters, respectively, and $\|\cdot\|$ is the Euclidean distance. Such an unbounded semivariogram implies that the field is mixing, which is appropriate if the extreme events are spatially localized. It is possible to account for geometric anisotropy by introducing a symmetric positive definite matrix (e.g., Blanchet and Davison, 2011), which, for $d = 2$, is

$$A = \begin{pmatrix} \cos \kappa & -\sin \kappa \\ r \sin \kappa & r \cos \kappa \end{pmatrix}, \quad (2.3)$$

where $r > 0$ and $\kappa \in [0, \pi]$ are the scaling and rotation parameters, respectively. The resulting semivariogram is

$$\gamma(\mathbf{s}) = (\|A\mathbf{s}\|/\rho)^\alpha, \quad \mathbf{s} \in \mathbb{R}^d. \quad (2.4)$$

For any simple max-stable field, we have, for $\mathbf{s}_1, \dots, \mathbf{s}_D \in \mathbb{R}^d$ and $z_1, \dots, z_D > 0$,

$$\mathbb{P}\{Z(\mathbf{s}_1) \leq z_1, \dots, Z(\mathbf{s}_D) \leq z_D\} = \exp\{-V_{\mathbf{s}_1, \dots, \mathbf{s}_D}(z_1, \dots, z_D)\}, \quad (2.5)$$

where (Pickands, 1981)

$$V_{\mathbf{s}_1, \dots, \mathbf{s}_D}(z_1, \dots, z_D) = \int_{\mathcal{S}_D} \max\left\{\frac{w_1}{z_1}, \dots, \frac{w_D}{z_D}\right\} dM(w_1, \dots, w_D),$$

where $M_{\mathbf{s}_1, \dots, \mathbf{s}_D}$ is a measure on the D -dimensional simplex \mathcal{S}_D satisfying

$$\int w_k dM_{\mathbf{s}_1, \dots, \mathbf{s}_D}(w_1, \dots, w_D) = 1,$$

Chapter 2. Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

for each $k \in \{1, \dots, D\}$. The function V_{s_1, \dots, s_D} , called the exponent measure of the max-stable random vector $(Z(s_1), \dots, Z(s_D))'$, entirely characterizes its dependence and is homogeneous of order -1 . Several summaries of the dependence (so-called dependence measures) have been proposed for max-stable fields/vectors. One is the extremal coefficient (Schlather and Tawn, 2003). The bivariate distribution function of a simple max-stable field Z satisfies

$$\mathbb{P}(Z(s_1) \leq u, Z(s_2) \leq u) = \exp\left(-\frac{\theta(s_1, s_2)}{u}\right), \quad s_1, s_2 \in \mathbb{R}^d, \quad (2.6)$$

where $u > 0$ and $\theta(s_1, s_2)$ is the bivariate extremal coefficient. By homogeneity of the exponent measure, $\theta(s_1, s_2) = V_{s_1, s_2}(1, 1)$. Furthermore, $\theta(s_1, s_2) \in [1, 2]$ for any $s_1, s_2 \in \mathbb{R}^d$, with values 1 and 2 indicating perfect dependence and independence, respectively. The lower the value of $\theta(s_1, s_2)$, the higher the dependence. The extremal coefficient can be estimated without assuming any parametric model, as it has a one-to-one relation with the F-madogram v_F (Cooley et al., 2006); this allows efficient non-parametric estimation of $\theta(s_1, s_2)$. If Z in (2.6) is a Brown–Renick field associated with the semivariogram γ , then (e.g., Davison et al., 2012)

$$\theta(s_1, s_2) = 2\Phi\left(\sqrt{\gamma(s_2 - s_1)}/2\right), \quad s_1, s_2 \in \mathbb{R}^d, \quad (2.7)$$

where Φ denotes the standard univariate Gaussian distribution function.

If $\{Z(s) : s \in \mathbb{R}^d\}$ is a simple max-stable field, then (e.g., Davison et al., 2012)

$$\lim_{z \rightarrow \infty} \Pr\{Z(s_1) > z \mid Z(s_2) > z\} = 2 - \theta(s_1, s_2), \quad s_1, s_2 \in \mathbb{R}^d. \quad (2.8)$$

Thus, unless Z is a purely independent field, it is asymptotically dependent, as there exist s_1, s_2 such that the limit in (2.8) is strictly positive. A possible way to assess the suitability of max-stable models consists in testing if the left hand-side of (2.8) is zero for different pairs of grid points, e.g., using the tests reviewed by de Carvalho and Ramos (2012); see for instance Bacro et al. (2010). However, evidence of asymptotic dependence in data does not entail suitability of max-stable models, as data may show asymptotic dependence without being max-stable. In this chapter, we instead explicitly test the null hypothesis of max-stability; see Section 2.4.2.

2.3.2 Estimation of max-stable fields

Let $s_1, \dots, s_D \in \mathbb{R}^2$ denote grid points at which we observe a field of pointwise maxima $\{M(s) : s \in \mathbb{R}^2\}$ at a specific time scale. Due to the results in Section 2.3.1, it is classical to take a max-stable field as a model for M . As we will justify in Section 2.4.1, we model the margin related to each grid point s_i , $i = 1, \dots, D$, with a GEV distribution having location, scale and shape parameters η_{s_i} , τ_{s_i} and ξ_{s_i} , which entails having $3 \times D$ (i.e., $3 \times 619 = 1857$ in our setting) marginal parameters to estimate. Due to computational constraints, the classical approach first estimates these GEV parameters by maximum likelihood, and then estimates the dependence parameters of the max-stable field by maximizing the composite log-likelihood

in a second step; this section is dedicated to the latter.

Unless D is small, the D -dimensional multivariate density of max-stable random fields is often intractable, as the exponent measure in (2.5) can be difficult to characterize and the exponential leads to a combinatorial explosion of the number of terms in the density. Full likelihood inference is thus impossible and composite likelihood techniques (e.g., Varin et al., 2011) have been extensively used. Pairwise composite likelihoods are most common (e.g., Padoan et al., 2010; Blanchet and Davison, 2011; Davison et al., 2012), but higher order composite likelihoods have also been considered (e.g., Huser and Davison, 2013; Castruccio et al., 2016). Under mild regularity conditions, the maximum pairwise likelihood estimator is strongly consistent and asymptotically normal but does not attain the Cramér–Rao lower bound. Padoan et al. (2010) and Sang and Genton (2014) showed that truncating the pairwise likelihood by ignoring pairs of sites that are far apart can improve the statistical efficiency of that estimator; for similar findings in other settings, see the references in Sang and Genton (2014). Ignoring some pairs also decreases the computational burden, which is valuable for large values of D , as in our setting with $D = 619$. Castruccio et al. (2016) showed that the benefit of truncation in terms of statistical efficiency is larger for pairwise or triplewise likelihoods than for higher order composite likelihoods.

For any $\mathbf{s} \in \mathbb{R}^2$, let $Z(\mathbf{s})$ denote the transformed version of $M(\mathbf{s})$ that approximately follows the standard Fréchet distribution. We let $z_{\mathbf{s},t}$ denote the realization of $Z(\mathbf{s})$ during the t -th period, $\boldsymbol{\psi}$ denote the vector of dependence parameters of the max-stable model, and $f_{\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\psi}}$ denote the corresponding pairwise density for sites $\mathbf{s}_i, \mathbf{s}_j \in \mathbb{R}^2$. The truncated pairwise log-likelihood is

$$l(\boldsymbol{\psi}) = \sum_{t=1}^T \sum_{i=1}^{D-1} \sum_{j=i+1}^D \mathbb{I}_{\{\|\mathbf{s}_i - \mathbf{s}_j\| \leq \sqrt{2}c\}} \log f_{\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\psi}}(z_{\mathbf{s}_i, t}, z_{\mathbf{s}_j, t}), \quad (2.9)$$

where c is the truncation distance and $\mathbb{I}_{\{\cdot\}}$ is the indicator function.

We choose a value of c adapted to the context of our chapter through a simulation study. We consider a Brown–Resnick field having semivariogram (2.4) with $r = 1$, $\kappa = 0$, $\alpha = 1$ and $\rho \in \{1, 2, 4, 8, 12\}$, on squares containing 25 (5×5), 100 (10×10), and 225 (15×15) grid points. In each of these 15 settings, we simulate 444 (37×12) independent realizations of the field, and this 400 times independently. For each of these 400 experiments, we estimate ρ using the truncated pairwise log-likelihood (2.9) with $c = 1, 2, 3, 4$. The scale $r = 1$ and angle $\kappa = 0$ (corresponding to an isotropic semivariogram), and smoothness $\alpha = 1$, are close to those for our dataset. We let the value of the range parameter and the spatial domain vary. Figure 2.1 shows that the lower ρ is and the higher the D is, the more accurate the estimation; a similar result was reported by Sang and Genton (2014). In most settings, $c = 2$ leads to the most accurate estimation, so this is the value we use below.

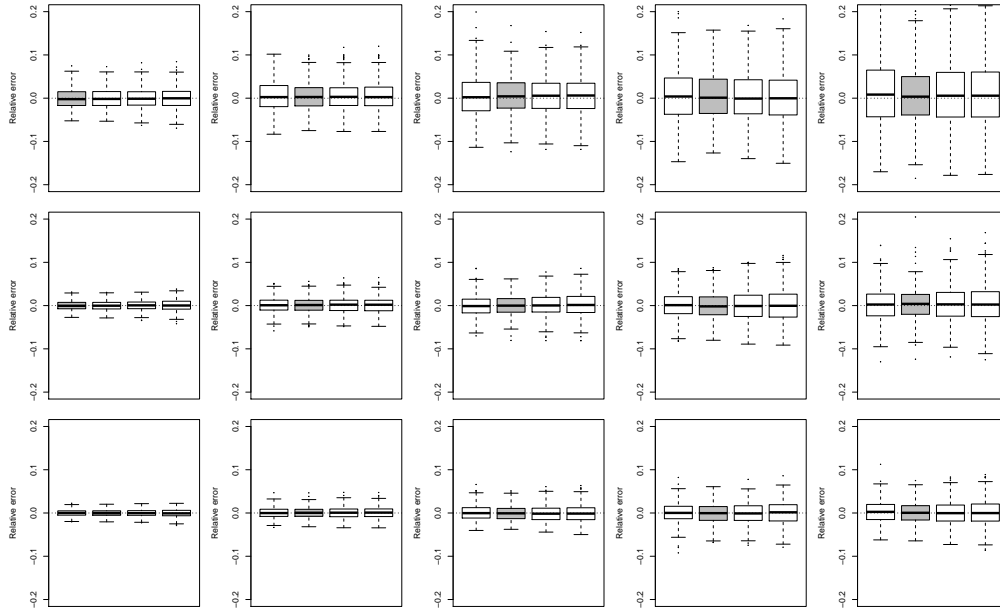


Figure 2.1 – Boxplots of the relative errors of the estimates of ρ for $c = 1, 2, 3, 4$ (from left to right). The boxplot associated with the lowest relative root mean squared error is highlighted in grey. The rows correspond to $D = 25$ (top), 100 (middle), 225 (bottom) and each column corresponds to a value of ρ : 1, 2, 4, 8, 12 (from left to right).

2.3.3 Data and exploratory analysis

The data we study are used in Koch et al. (2021) and constitute a coarse version of reanalysis data from the North American Regional Reanalysis (NARR). They consist of 3-hourly time-series of 0–180 hPa CAPE (J kg^{-1}) and 0–3 km SRH ($\text{m}^2 \text{s}^{-2}$) from 1 January 1979 at 00:00 Coordinated Universal Time (UTC) to 31 December 2015 at 21:00 UTC. For consistency, we removed data recorded on February 29; this does not impact our findings. The area considered is a rectangle over the contiguous US from -110° to -80° longitude and 30° to 50° latitude (see Figure 2.7), thus containing Tornado Alley, the most risky region of the US in terms of severe thunderstorms. The resolution of the data set is 1° longitude and 1° latitude, leading to 651 grid points in our region; no data are available for 32 grid points over water. We use the time series of CAPE and SRH to build 3-hourly time series of $\text{PROD} = \sqrt{\text{CAPE}} \times \text{SRH}$ ($\text{m}^3 \text{s}^{-3}$). Finally, as a measure of ENSO, we use monthly values of the Niño-3.4 index ($^\circ\text{C}$) from 1979 to 2015, taken from the ERSSTv5 data set publicly available from the NOAA Climate Prediction Center.

Figure 2.2 shows that for PROD maxima in April, the bivariate extremal coefficient function tends to be lower during El Niño episodes than during neutral ENSO states, which indicates that the spatial extent of individual PROD events is higher during El Niño episodes. This confirms what could be expected based on physics: El Niño events tend to trigger larger areas of instability over the US. Similarly, we observe lower extremal coefficients in January than in

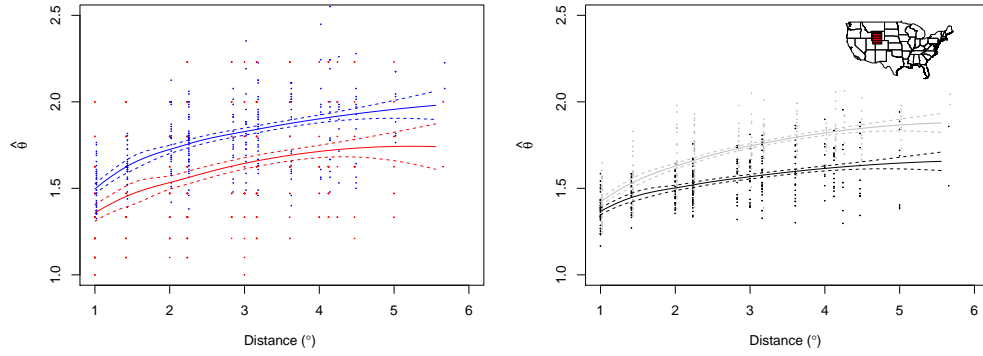


Figure 2.2 – The dots correspond to the empirical pairwise extremal coefficients for PROD using all grid points in the square from -110° to -106° longitude and 40° to 44° latitude (indicated by the red region on the map in the right panel). The solid lines depict the best fitted local polynomial regression curves and the dotted lines define the related 95% pointwise confidence intervals (assuming that the estimates are independent and known). The left panel concerns April maxima during positive (red) and small (blue) ENSO states, corresponding to $\text{ENSO} > +0.5^\circ\text{C}$ and $|\text{ENSO}| \leq 0.5^\circ\text{C}$, respectively. The right panel concerns January (black) and July (grey) maxima.

July, suggesting larger PROD events in January. This agrees with the fact that weather events tend to be more localized in summer than in winter (convective versus large-scale events). The curves depicted in Figure 2.2 were obtained on a rectangle from -110° to -106° longitude and 40° to 44° latitude.

2.4 New methodology

2.4.1 Model

General version

We propose a space-time random field model that is simple max-stable at each time point. Let T be the number of time points and let the time index represent the month. Let $\mathbf{x} = (x_1, \dots, x_p)'$ be a vector of p covariates at the time points, leading to $\mathbf{x}_t = (x_{1,t}, \dots, x_{p,t})$, $t = 1, \dots, T$. Examples of covariates are large-scale weather signals such as ENSO and the North Atlantic Oscillation (NAO), or the month associated with time t . Our space-time model arises by considering independent simple max-stable fields $\{Z(\mathbf{s}, t) : \mathbf{s} \in \mathbb{R}^2, t = 1, \dots, T\}$ whose spatial dependence varies with t through the vector of covariates \mathbf{x}_t .

Many parametric max-stable models such as the Brown–Resnick or extremal t random fields have range and smoothness parameters that appear in the correlation function of the underlying standard stationary Gaussian field. Typical examples include the powered exponential, Cauchy and Whittle–Matérn functions. Both parameters are tightly linked to extremal dependence: the range accounts for the characteristic dimension of individual events and the

Chapter 2. Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

smoothness for the roughness of the field. We could allow these parameters to depend on the covariates \mathbf{x}_t through a regression model, writing the parameter's value a_t at time t in terms of a linear basis expansion of \mathbf{x}_t ,

$$a_t = g \left(\sum_{m=1}^M \beta_m h_m(\mathbf{x}_t) \right), \quad (2.10)$$

where g is a monotonic link function, h_1, \dots, h_M are functions from \mathbb{R}^p to \mathbb{R} , and β_m are real numbers. A natural choice for g in the case of the range parameter is the exponential function, which ensures that $a_t > 0$. Other constraints on the β_m and/or h_m may be necessary for other parameters. Modelling both the smoothness and range parameters with (2.10) can lead to identifiability issues (both parameters capturing similar effects) and, for parsimony, we keep the scaling and rotation parameters r and κ constant. Depending on the choice of h_m , (2.10) can represent a generalized linear model or a more flexible and non-linear model (when the h_m are, e.g., splines or wavelets). We refer the reader to Wood (2017) for a detailed review of splines. Cyclic P-splines are appropriate when using the month as covariate, as they allow a smooth transition between the first and the last month. Radial cubic splines are useful for continuous variables such as weather variables (e.g., ENSO) and tensor product splines allow one to capture interactions between different covariates.

Version for the case study and simulations

Finding appropriate and parsimonious trend surfaces for the marginal parameters is challenging for the large and meteorologically heterogeneous region considered here. Using incorrect trend surfaces may lead to a bias in the dependence, so we model the monthly maxima using the GEV parameters fitted separately at each grid point, and we subsequently model the field obtained after transforming the data to standard Fréchet using those estimated parameters. The space-time model we propose for the standardized versions of CAPE, SRH and PROD is in the spirit of Section 2.4.1. We choose as max-stable field the Brown–Resnick model with variogram (2.4) as it often fits environmental data well (Davison et al., 2012). Based on physical knowledge and our exploratory analysis (Figure 2.2), we model the range parameter ρ as in (2.10) with month and ENSO as covariates, and keep the parameters α , r and κ constant. Considering the unit of time to be the month, (i.e., $t = 1$ means January 1 00:00 UTC of the 1st year, $t = 2$ means February 1 00:00 UTC of the first year, ...), our vector of covariates at t is

$$\mathbf{x}_t = (\text{ENSO}_t, t \bmod 12)', \quad t = 1, \dots, 444, \quad (2.11)$$

where \bmod denotes the modulo operation and ENSO_t is the value of ENSO for the month following t . The effect of these covariates on ρ appears to be non-linear and the covariates interact, so we choose the functions h_m as a tensor product spline basis between a cyclic P-spline basis in the month direction and a radial cubic spline basis in the ENSO direction, which allows us to borrow strength from neighbouring months and ENSO states. Finally, our model at time $t = 1, \dots, T$, denoted $\{Z(\mathbf{s}, t) : \mathbf{s} \in \mathbb{R}^2\}$, is the Brown–Resnick field associated with

the variogram

$$\gamma(\mathbf{s}, t) = \left(\frac{\|A\mathbf{s}\|}{\exp(\sum_{m=1}^M \beta_m h_m(\mathbf{x}_t))} \right)^\alpha, \quad \mathbf{s} \in \mathbb{R}^2, \quad (2.12)$$

where A is the matrix (2.3) with $r > 0$ and $\kappa \in [0, \pi]$. Fitting a classical Brown–Resnick model for each month separately showed no evidence for month-specific smoothness, scaling and rotation parameters (not shown), for any of PROD, CAPE or SRH. To choose M and the knots, we fix a number of knots in each direction and place them in a regular way. We fit models with different numbers of knots in each direction and then choose the best one using the bootstrap-based CLIC developed in Section 2.4.3.

We now check by simulation that the model parameters are identifiable and can be estimated reasonably well in a setting close to the case study (in terms of number of parameters and data amount). Such a check is especially important for our model in (2.12) owing to its complexity and the large number of parameters, and detecting temporal non-stationarity in the extremal dependence appears challenging without enough data. We consider (2.12) with $T = 444$ (37×12) on a square containing 625 (25×25) grid points and with parameters $r = 0.72$, $\kappa = -0.08$, and $\alpha = 1.26$. We choose four knots in the month direction, located at 0.5, 4.5, 8.5, 12.5, and three knots in the ENSO direction whose coordinates, $-1.06, 0.05, 1.16$, correspond to the 10% and 90% quantiles of observed ENSO values and their mid-point. As we use a circular P spline basis in the month direction, the values of the spline are the same at 0.5 and 12.5. Consequently, we have three distinct knots in both the month and ENSO directions, giving nine knots over the space of covariates. The corresponding coefficients are $\beta_0 = 0.52$ (intercept), $\beta_1 = -0.03$, $\beta_2 = 0.02$, $\beta_3 = 0.07$, $\beta_4 = 0.11$, $\beta_5 = -0.07$, $\beta_6 = -0.23$, $\beta_7 = -0.03$, $\beta_8 = 0.02$, and $\beta_9 = 0.04$. We simulate 100 independent replicates of this field with standard Fréchet margins and, using the month and ENSO covariates in the data of Section 2.3.3, estimate all parameters. Figure 2.3 suggests that all estimators are unbiased and that all parameters are recovered in a satisfactory way, but that the variability of the estimates varies: the errors tend to be larger for the parameters associated with the spline basis than for the others. The smoothness parameter α and the scaling factor r are very well estimated. Figure 2.4 also shows that signals (in March here) and non-signals (in September) are equally well-identified.

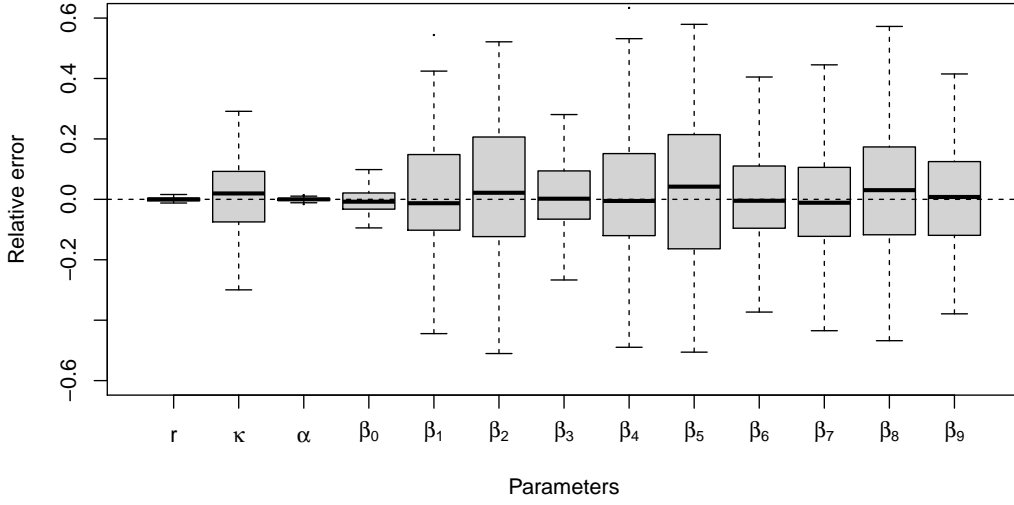


Figure 2.3 – Relative error for the parameters in the simulation study described in Section 2.4.1.

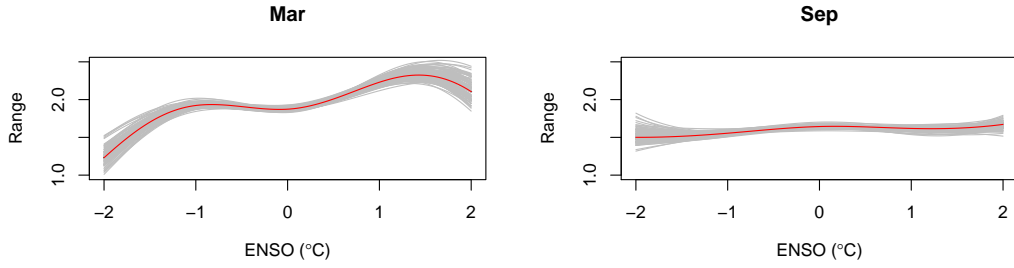


Figure 2.4 – The estimated range parameter with form given in (2.10), based on the simulation study giving Figure 2.4, for $t \bmod 12 = 3$ (March, left) and $t \bmod 12 = 9$ (September, right).

2.4.2 A max-stability test in an applied setting with unknown margins

Assume that a max-stable field with standard Gumbel margins, $\{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$, is observed in a testing region $\mathcal{T} \subseteq \mathbb{R}^d$ consisting of a finite set of grid points $\{\mathbf{s}_1, \dots, \mathbf{s}_D\}$, and let $\Delta = \{1, \dots, D\}$. The homogeneity of order -1 of the exponent measure in (2.5) implies that for any $S \subseteq \Delta$, the variable $\eta_S = \max_{j \in S} Z(\mathbf{s}_j)$ has a Gumbel distribution with location parameter $\log V_{\mathbf{s}_j, j \in S}$ (Gabda et al., 2012), where $V_{\mathbf{s}_j, j \in S}$ is the exponent measure of the random vector $\{Z(\mathbf{s}_j), j \in S\}$. In practice, the cardinality of S , labelled $|S|$, must be chosen, and a Gumbel distribution, with location $\log \hat{V}_{\mathbf{s}_j, j \in S}$ estimated from temporal realizations of η_S , can be compared with the empirical distribution of η_S using a probability-probability plot (e.g., Gabda et al., 2012; Buhl and Klüppelberg, 2016).

Here we extend this idea to a formal statistical test more appropriate in applications, where environmental fields do not usually have standard Gumbel margins and we would first need to estimate and then transform them before performing the test. We assess evidence against the

null hypothesis that the multivariate distribution is max-stable by using an Anderson–Darling statistic to measure the distance between the empirical distribution of η_S and a Gumbel distribution with location $\log \hat{V}_{s_j, j \in S}$. The null distribution of our test statistic should include the uncertainty arising from the estimation of both $\log V_{s_j, j \in S}$ and the margins, and we achieve this using a bootstrap. In the following, we fix $S = \Delta$ without loss of generality.

Our proposed bootstrap test is parametric for the margins and non-parametric for the dependence, as we do not want to assume a parametric model for the latter. Let $(z_{1,m}^*, \dots, z_{D,m}^*)'$ denote the vector of the m -th observed maxima at all grid points in S , where $m \in \{1, \dots, M\}$. For each grid point $d \in \{1, \dots, D\}$, we first fit the GEV distribution using $z_{d,1}^*, \dots, z_{d,M}^*$ to obtain estimators $\hat{\eta}_d^*$, $\hat{\tau}_d^*$ and $\hat{\xi}_d^*$. To find the approximate distribution of the Anderson–Darling statistic under the null hypothesis of max-stability, we repeat the following procedure B times:

1. Using the approach with empirical likelihood (Owen, 2001) outlined in the Supplementary Material 2.7.1, simulate M replicates of a max-stable vector, denoted $(\tilde{z}_{1,1}, \dots, \tilde{z}_{D,1})', \dots, (\tilde{z}_{1,M}, \dots, \tilde{z}_{D,M})'$, such that $\tilde{z}_{d,1}, \dots, \tilde{z}_{d,M}$, $d \in \{1, \dots, D\}$, are drawn from a GEV distribution with location, scale and shape parameters $\hat{\eta}_d^*$, $\hat{\tau}_d^*$ and $\hat{\xi}_d^*$;
2. For each $d \in \{1, \dots, D\}$, fit a GEV distribution to $\tilde{z}_{d,1}, \dots, \tilde{z}_{d,M}$, by maximum likelihood. Then, transform them with these GEV parameter estimates such that $z_{d,1}, \dots, z_{d,M}$ are approximately standard Gumbel distributed, yielding $(z_{1,1}, \dots, z_{D,1})', \dots, (z_{1,M}, \dots, z_{D,M})'$;
3. Evaluate $\eta_{S,m} = \max_{j \in S} z_{j,m}$, $m = 1, \dots, M$, and fit a Gumbel distribution to $\eta_{S,1}, \dots, \eta_{S,M}$ using maximum likelihood, giving the location parameter estimate $\log \hat{V}_{s_j, j \in S}$;
4. Calculate the Anderson–Darling statistic measuring the distance between the empirical distribution of $\eta_{S,1}, \dots, \eta_{S,M}$ and a Gumbel distribution with location parameter $\log \hat{V}_{s_j, j \in S}$.

We perform two experiments with $D = 25$, and $M = 40$ block maxima of size 240. We first generate from a max-stable multivariate logistic distribution

$$G(y_1, \dots, y_D) = \exp \left\{ - \left(\sum_{i=1}^D y_i^{-1/\lambda} \right)^\lambda \right\}, \quad y_1, \dots, y_D > 0,$$

with dependence parameter $\lambda \in \{0.1, 0.5, 0.9\}$, and perform our max-stability test on these observations with $B = 200$. Secondly, we perform the same experiment but generate from a multivariate Gaussian distribution with common pairwise correlation $\zeta \in \{0.1, 0.5, 0.9, 0.99\}$. To assess the empirical size and power of the test, we repeat both experiments 1000 times. The field of pointwise maxima of Gaussian fields converges in (2.1) to the degenerate independent max-stable field, but a block size of 240 is insufficient for convergence. As such, our approach using the multivariate normal distribution is an approximate but reasonable way to assess the power of our test.

Chapter 2. Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

	5%	20%	p-val AD	p-val KS
Max-stable logistic, $\lambda = 0.1$	5.20	20.90	0.77	0.94
Max-stable logistic, $\lambda = 0.5$	4.70	20.30	0.17	0.29
Max-stable logistic, $\lambda = 0.9$	6.12	22.20	0.08	0.20
Normal, $\zeta = 0.1$	5.20	20.20	0.02	0.04
Normal, $\zeta = 0.5$	9.30	26.30	0.00	0.00
Normal, $\zeta = 0.9$	16.20	34.50	0.00	0.00
Normal, $\zeta = 0.99$	7.02	24.02	0.01	0.03

Table 2.1 – Empirical size (%) (top three) and power (%) (bottom three) for tests at the 5% and 20% nominal level with $(B, p) = (200, 0.9)$, for 1000 samples of size $n = 240 \times 40$, and different levels of dependence. The last two columns show the p-values for the Anderson–Darling and Kolmogorov–Smirnov tests of uniformity for the 1000 p-values.

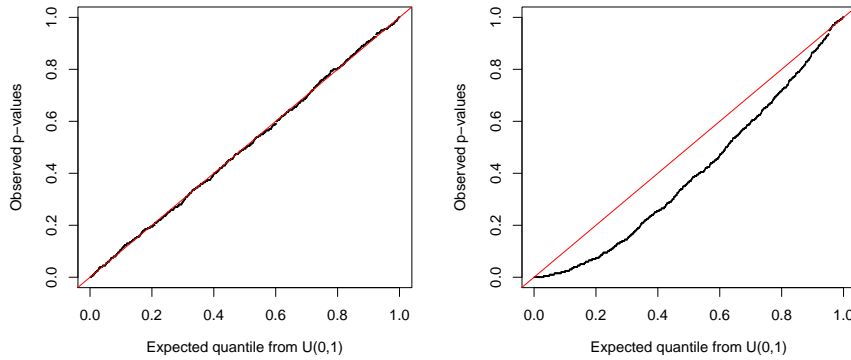


Figure 2.5 – Quantile-quantile plots for the 1000 p-values with the simulation setting involving the max-stable logistic, $\lambda = 0.1$ (left), and normal, $\rho = 0.9$ (right), distributions.

Table 2.1 shows that the empirical size of the test is controlled reasonably well in this setting, with the 1000 Anderson-Darling and Kolmogorov-Smirnov p-values correctly showing no departure from uniformity at the 5% level for the logistic max-stable cases. All tests for uniformity of the p-values are rejected at the 5% level for the multivariate normal cases. The power rises first as dependence increases, but falls when $\zeta = 0.99$, probably because this setting gives simulated empirical distributions that are very close to the perfectly dependent max-stable distribution. The power is similarly low when $\zeta = 0.1$, perhaps because the empirical distributions here are close to the independent max-stable distribution. Figure 2.5 shows quantile-quantile plots of the p-values in two simulation settings, and illustrates the departure from uniformity for the p-values in the multivariate normal case with $\zeta = 0.9$.

Selection of data used for validation

To validate a model, one should analyse whether its fit deteriorates non-negligibly when applied to pertinent new data not used for model fitting, so it is imperative to put a validation set aside. In our spatial extremes setting, it is common to exclude a full time-series of observations at randomly chosen locations (e.g., Huser and Wadsworth, 2019; Davison et al., 2012) before fitting models, but we would like to avoid omitting data in an extreme-value setting where data are scarce.

We split our spatial region into 24 small windows of 25–36 locations each. For each window and month, we implement the max-stability test developed in Section 2.4.2 for each window, i.e., $D = |K| = 25\text{--}36$. To choose our validation set, we select data from locations and time points in window-month combinations that failed the max-stability test. This ensures that we only exclude data for which we do not observe max-stability and that might corrupt our max-stable model fit. Since our test for the spatial dependence is non-parametric, no bias arises if the procedures we use to validate our models is model-independent. If validation is based on any notion of max-stability however, then this out-sample selection approach may be conservative.

Our procedure can be applied to large heterogeneous space and time frames in practical settings where the margins must be estimated. The small spatial testing windows we test allow us to identify data from small regions where a max-stable model may be unsuitable. However in our context we instead discard these data before estimating our models and only use the omitted data for validation. This ‘Swiss cheese’ approach to out-sample selection, where we omit spatial holes in the data for given months, also encourages us to consider models that borrow estimation strength across months.

2.4.3 Bootstrap-based uncertainty assessment and model selection

Let $\mathcal{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_T)'$ be a data matrix, where $\mathbf{Y}_1, \dots, \mathbf{Y}_T$ are independent replicates of a D -dimensional random vector \mathbf{Y} . In our case, \mathbf{Y}_t would be the vector of maxima at time t at all grid points in our region. We model the dependence in \mathbf{Y} , and the ultimate goal of this section is to propose a way to adequately quantify uncertainty of the dependence parameters' estimates and select the best dependence model when using composite likelihood when the margins have been modelled and estimated in a first step. Assume that we have a parametric model for each margin of \mathbf{Y} and that the marginal parameters of all components of \mathbf{Y} are gathered in $\boldsymbol{\lambda}$. Below, we consider both the virtual case where the exact marginal models and $\boldsymbol{\lambda}$ are known, and the realistic one where $\boldsymbol{\lambda}$ is estimated by $\hat{\boldsymbol{\lambda}}$ in a first step. Let $t_{\mathbf{a}}$ be a function depending on a vector \mathbf{a} that transforms a data matrix having modelled margins with parameters gathered in \mathbf{a} into a data matrix with specified common parameter-free margins. Let $\mathbf{Z}_1, \dots, \mathbf{Z}_T$ and \mathbf{Z} be transformed versions of $\mathbf{Y}_1, \dots, \mathbf{Y}_T$ and \mathbf{Y} so that they have the specified margins, and let $\mathcal{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_T)'$. We have $\mathcal{Z} = t_{\boldsymbol{\lambda}}(\mathcal{Y})$ if $\boldsymbol{\lambda}$ is known and $\mathcal{Z} = t_{\hat{\boldsymbol{\lambda}}}(\mathcal{Y})$ if $\boldsymbol{\lambda}$ has to be estimated by $\hat{\boldsymbol{\lambda}}$ in a first step.

Chapter 2. Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

Assume that we model \mathcal{Z} using a family of density functions with common parameter-free margins (typically standard Fréchet in the spatial extremes setting) and dependence parameter $\boldsymbol{\psi}$, denoted $\mathcal{F} = \{f(\mathbf{z}, \boldsymbol{\psi}) : \mathbf{z} \in \mathbb{R}^D, \boldsymbol{\psi} \in \Psi \subseteq \mathbb{R}^p\}$, and that we use composite likelihood for inference. The composite likelihood is $L_C(\boldsymbol{\psi}; \mathcal{Z}) = \prod_{t=1}^T f_C(\boldsymbol{\psi}; \mathbf{Z}_t)$, where f_C is defined through the density f and characterizes the composite likelihood (see, e.g., Varin and Vidoni, 2005, Definition 1), and we denote by $\hat{\boldsymbol{\psi}}$ the maximum composite likelihood estimator. In the case of the truncated pairwise likelihood (2.9), f_C is the sum over all pairs of the tapered bivariate densities.

Provided that the true marginal models and $\boldsymbol{\lambda}$ are known, under mild regularity assumptions, $\hat{\boldsymbol{\psi}} \sim N_p\{\boldsymbol{\psi}, \mathbf{I}(\boldsymbol{\psi})^{-1}\}$ for T large, where $\mathbf{I}(\boldsymbol{\psi}) = \mathbf{H}(\boldsymbol{\psi})\mathbf{J}(\boldsymbol{\psi})^{-1}\mathbf{H}(\boldsymbol{\psi})$ with

$$\mathbf{H}(\boldsymbol{\psi}) = \mathbb{E}\{-\nabla_{\boldsymbol{\psi}}^2 \log L_C(\boldsymbol{\psi}; \mathcal{Z})\}, \quad \mathbf{J}(\boldsymbol{\psi}) = \mathbb{V}\{\nabla_{\boldsymbol{\psi}} \log L_C(\boldsymbol{\psi}; \mathcal{Z})\},$$

where $\nabla_{\boldsymbol{\psi}}^2$ and $\nabla_{\boldsymbol{\psi}}$ denote the Hessian and gradient operators with respect to $\boldsymbol{\psi}$, and \mathbf{V} indicates the covariance matrix operator; $\mathbf{I}(\boldsymbol{\psi})$ is the sandwich information matrix and we call its inverse the sandwich variance matrix. Approximate confidence intervals can be computed from the estimated sandwich variance matrix $\hat{\mathbf{I}}(\hat{\boldsymbol{\psi}}) = \hat{\mathbf{H}}(\hat{\boldsymbol{\psi}})^{-1}\hat{\mathbf{J}}(\hat{\boldsymbol{\psi}})\hat{\mathbf{H}}(\hat{\boldsymbol{\psi}})^{-1}$ (Padoan et al., 2010), where

$$\hat{\mathbf{H}}(\hat{\boldsymbol{\psi}}) = -\nabla_{\boldsymbol{\psi}}^2 \log L_C(\hat{\boldsymbol{\psi}}; \mathcal{Z}), \quad \hat{\mathbf{J}}(\hat{\boldsymbol{\psi}}) = \sum_{t=1}^T \{\nabla_{\boldsymbol{\psi}} \log f_C(\hat{\boldsymbol{\psi}}; \mathbf{Z}_t)\} \{\nabla_{\boldsymbol{\psi}} \log f_C(\hat{\boldsymbol{\psi}}; \mathbf{Z}_t)\}'.$$

After fitting several models in \mathcal{F} using composite likelihood, it is standard to select that with the highest observed value of the composite likelihood information criterion (Varin and Vidoni, 2005)

$$\text{CLIC} = \log L_C(\hat{\boldsymbol{\psi}}, \mathcal{Z}) - \text{tr}\{\hat{\mathbf{J}}(\hat{\boldsymbol{\psi}})\hat{\mathbf{H}}(\hat{\boldsymbol{\psi}})^{-1}\}, \quad (2.13)$$

or equivalently the lowest observed value of (Padoan et al., 2010)

$$-2\log L_C(\hat{\boldsymbol{\psi}}, \mathcal{Z}) + 2\text{tr}\{\hat{\mathbf{J}}(\hat{\boldsymbol{\psi}})\hat{\mathbf{H}}(\hat{\boldsymbol{\psi}})^{-1}\}, \quad (2.14)$$

where tr denotes the trace; in the following, CLIC refers to (2.14) and not (2.13), as commonly done in the spatial extremes literature.

But if $\boldsymbol{\lambda}$ has been estimated in a first step, as is often the case in the field of spatial extremes, using respectively $\hat{\mathbf{I}}(\boldsymbol{\psi})$ and CLIC for uncertainty assessment of $\hat{\boldsymbol{\psi}}$ and model selection within \mathcal{F} is suboptimal, as the effect of estimating the marginal parameters in $\boldsymbol{\lambda}$ is ignored. The dependence parameters' uncertainty is typically underestimated, and an alternative approach that addresses this (see our simulation study in Section 2.4.3) uses the non-parametric bootstrap. In spatial extremes, this method for uncertainty assessment is employed by, e.g., Davison et al. (2013, 2018) and Huser and Wadsworth (2019), showing that researchers are aware of the shortcomings of using the estimated sandwich variance matrix. The same cannot be said of the use of CLIC for model selection in a two-step setting, as most studies in this context (such as, e.g., Davison et al., 2013, 2018; Huser and Genton, 2016; Huser et al., 2021) use it only after

the margins have been estimated and then transformed.

In order to account for the effect of marginal estimation on model selection when using composite likelihood, we propose a model selection criterion based on bootstrap estimation of the non-normalized composite Kullback–Leibler divergence (Varin and Vidoni, 2005). Although we deal with max-stable fields in this chapter, the methodology we develop is valid in a general setting. First, we consider that the margins are known and extend the results by Shibata (1997) and Cavanaugh and Shumway (1997) to the composite likelihood setting; see Section 2.4.3. Elaborating on this, we then define our new criterion, which enables one to account for the marginal effects (Section 2.4.3). Section 2.4.3 illustrates the benefits of this method through a simulation study. We do not show the dependence of \mathcal{Y} , \mathcal{Z} , and $\hat{\boldsymbol{\psi}}$ on T for notational simplicity.

Case of known margins

We assume that λ in t_λ is known and we wish to find the best model for $\mathcal{Z} = t_\lambda(\mathcal{Y})$. To this end, we estimate the non-normalized composite Kullback–Leibler divergence of a model to the truth using non-parametric bootstrap, as Cavanaugh and Shumway (1997) and Shibata (1997) did for the non-normalized Kullback–Leibler divergence.

Let $g(\mathbf{z})$, $\mathbf{z} \in \mathbb{R}^D$, be the true generating density of \mathbf{Z} . The non-normalized composite Kullback–Leibler divergence for a model with density in \mathcal{F} is $d_T(\boldsymbol{\psi}) = E_o\{-\log L_C(\boldsymbol{\psi}; \mathcal{Z})\}$, where E_o is the expectation under the true density g . The divergence of the model estimated by maximum composite likelihood (with $\hat{\boldsymbol{\psi}}$ as estimated parameter) to the truth is thus

$$d_T(\hat{\boldsymbol{\psi}}) = E_o\{-\log L_C(\boldsymbol{\psi}; \mathcal{Z})\} |_{\boldsymbol{\psi}=\hat{\boldsymbol{\psi}}}, \quad (2.15)$$

but this is impossible to evaluate unless we know g . Varin and Vidoni (2005) showed that a biased estimator of (2.15) is $-\log L_C(\hat{\boldsymbol{\psi}}; \mathcal{Z})$, and adjusting for the bias with a first-order correction leads to (2.13). Now, suppose that $\hat{\boldsymbol{\psi}}^*$ is a bootstrap replicate of $\hat{\boldsymbol{\psi}}$, and let E^* denote the expectation with respect to the bootstrap distribution of $\hat{\boldsymbol{\psi}}$. With arguments similar to those in Cavanaugh and Shumway (1997), one can show under the usual regularity conditions that

$$\text{Bias}_T^* = -2 [E^* \{-\log L_C(\hat{\boldsymbol{\psi}}^*; \mathcal{Z})\} + \log L_C(\hat{\boldsymbol{\psi}}; \mathcal{Z})] \quad (2.16)$$

converges almost surely to the bias of $-\log L_C(\hat{\boldsymbol{\psi}}; \mathcal{Z})$ as $T \rightarrow \infty$. A Monte Carlo estimator from B bootstrap replicates yields a strongly consistent (as $B \rightarrow \infty$) estimator of Bias_T^* ,

$$\hat{\text{Bias}}_T^* = -\frac{2}{B} \sum_{b=1}^B [-\log L_C\{\hat{\boldsymbol{\psi}}_b^*; t_\lambda(\mathcal{Y})\} + \log L_C\{\hat{\boldsymbol{\psi}}; t_\lambda(\mathcal{Y})\}]. \quad (2.17)$$

Chapter 2. Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

Thus, a natural estimator of twice the quantity in (2.15) is

$$-2\log L_C\{\hat{\boldsymbol{\psi}}; t_{\boldsymbol{\lambda}}(\mathcal{Y})\} - 2\text{Bias}_T^* = \frac{1}{B} \sum_{b=1}^B [2\log L_C\{\hat{\boldsymbol{\psi}}; t_{\boldsymbol{\lambda}}(\mathcal{Y})\} - 4\log L_C\{\hat{\boldsymbol{\psi}}_b^*; t_{\boldsymbol{\lambda}}(\mathcal{Y})\}]. \quad (2.18)$$

For T and B large enough, we expect model selection based on CLIC and (2.18) to be equivalent.

Case of unknown margins

We now assume that $\boldsymbol{\lambda}$ in $t_{\boldsymbol{\lambda}}$ is unknown and is estimated in a first step. Its estimate is denoted $\hat{\boldsymbol{\lambda}}$, and we are interested in finding the best model within \mathcal{F} for $\mathcal{Z} = t_{\hat{\boldsymbol{\lambda}}}(\mathcal{Y})$. An attractive property of the bootstrap-based estimator of the non-normalized composite Kullback–Leibler divergence developed in Section 2.4.3 is that the effects of estimating the margins can be accounted for in model selection. More precisely, in order to compute the maximum composite likelihood estimate of $\boldsymbol{\psi}$ for the b -th bootstrap replicate, $b = 1, \dots, B$, we do not treat the marginal parameters as fixed and given by $\hat{\boldsymbol{\lambda}}$, but we estimate them on the bootstrapped data, yielding the estimate $\hat{\boldsymbol{\lambda}}_b^*$. We make this explicit by writing the estimates $\hat{\boldsymbol{\psi}}$ and $\hat{\boldsymbol{\psi}}_b^*$ as functions of $\hat{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\lambda}}_b^*$, respectively. The expectation E^* in (2.16) with respect to the bootstrap distribution of $\hat{\boldsymbol{\psi}}$ takes the effect of estimating the margins into account. Following (2.18), our new criterion for model selection is therefore

$$\text{CLIC}^b = \frac{1}{B} \sum_{b=1}^B [2\log L_C\{\hat{\boldsymbol{\psi}}(\hat{\boldsymbol{\lambda}}); t_{\hat{\boldsymbol{\lambda}}}(\mathcal{Y})\} - 4\log L_C\{\hat{\boldsymbol{\psi}}_b^*(\hat{\boldsymbol{\lambda}}_b^*); t_{\hat{\boldsymbol{\lambda}}}(\mathcal{Y})\}], \quad (2.19)$$

and the model minimizing CLIC^b should be chosen. As a full likelihood is a particular instance of composite likelihood, the approach proposed in this section can also be used in the case of full likelihood inference.

The matrices $\hat{H}(\hat{\boldsymbol{\psi}})$ and $\hat{J}(\hat{\boldsymbol{\psi}})$, required for the calculation of sandwich-based (i.e., based on the estimated sandwich variance matrix) confidence intervals or CLIC, are often cumbersome to compute and careful application of pseudo-inverse procedures may be needed if $\hat{H}(\hat{\boldsymbol{\psi}})$ is singular, especially for complex models with many parameters such as those introduced in Section 2.4.1. This provides an additional argument supporting the use of CLIC^b , on top of its ability to account for the effect of estimating the margins in a first step. The bootstrap is anyway needed for an appropriate uncertainty quantification of the model parameters, so the calculation of CLIC^b comes at no extra computational cost.

Expressions asymptotically equivalent to (2.16) could be used, similarly as in Shibata (1997, Section 2), leading to different but asymptotically equivalent specifications of CLIC^b .

True/Alternative	$P^k/P^u/P^b$		
	$D = 25$	$D = 100$	$D = 225$
SM_0/BR_1	93/82/89	90/54/90	94/20/81
BR_0/BR_1	84/44/80	84/30/76	85/28/76

Table 2.2 – Frequency (in %, computed over 200 repetitions) of selection of the true (simpler) model for each experiment, procedure and value of D .

Simulation study

We perform two experiments with three procedures: P^k , where the margins are assumed (correctly) known when fitting the models and CLIC is subsequently used for selection; P^u , where the margins are supposed to be GEV distributed, estimated in a first step, and then transformed before fitting the dependence models and using CLIC for selection; and P^b , as P^u , but using $CLIC^b$ in (2.19) for model selection with a non-parametric block bootstrap ($B = 200$) where each replicate is a block. In all procedures, the dependence models are fitted using the approach of Section 2.3.2. Procedure P^k is associated with the ideal but virtual setting, and serves as an approximation to the best that CLIC can do.

In the first experiment, we generate 40 independent replicates at $D \in \{25, 100, 225\}$ grid points of a Smith field (Smith, 1990) with common standard Fréchet margins and twice the 2×2 identity matrix as covariance matrix, and apply P^k , P^u and P^b to choose between an isotropic Smith model labelled SM_0 and a two-parameter Brown–Resnick model labelled BR_1 ; the latter has one parameter too many to estimate, as the Smith field corresponds to the Brown–Resnick field with $\alpha = 2$ (e.g., Huser and Davison, 2013). We repeat this 200 times, and Table 2.2 shows that P^k correctly chooses SM_0 approximately 92% of the time for any D . This figure is much lower for P^u , dropping rapidly to as low as 20% when D increases, probably due to more unaccounted uncertainty from the estimation of the margins. By incorporating the effects of estimating the margins, P^b achieves true selection frequency that is close to that of P^k .

In the second experiment, we generate 40 independent replicates at $D \in \{25, 100, 225\}$ grid points of a Brown–Resnick field with common standard Fréchet margins, $\rho = 2$ and $\alpha = 1$ (to reflect a configuration that could be realistic in an environmental application), and apply P^k , P^u and P^b to choose between BR_1 and a simpler Brown–Resnick model labelled BR_0 with $\rho = 2$ fixed and α estimated. As above, we repeat this 200 times, and Table 2.2 shows that P^k correctly chooses BR_0 approximately 84% of the time for every D ; this number would be expected in the full likelihood inference setting with n large. The frequency of true selection ranges from 28% and 44% with P^u , and is between 76% and 80% for P^b .

According to the paired proportions test of McNemar (1947), all tests of differences between P^u and P^b are significant at the 95% level for both experiments (not shown). Thus, if the marginal and dependence parameters have to be estimated in two distinct steps and if composite likelihood is used, we strongly advocate the use of $CLIC^b$ in (2.19) instead of the commonly used CLIC in (2.14).

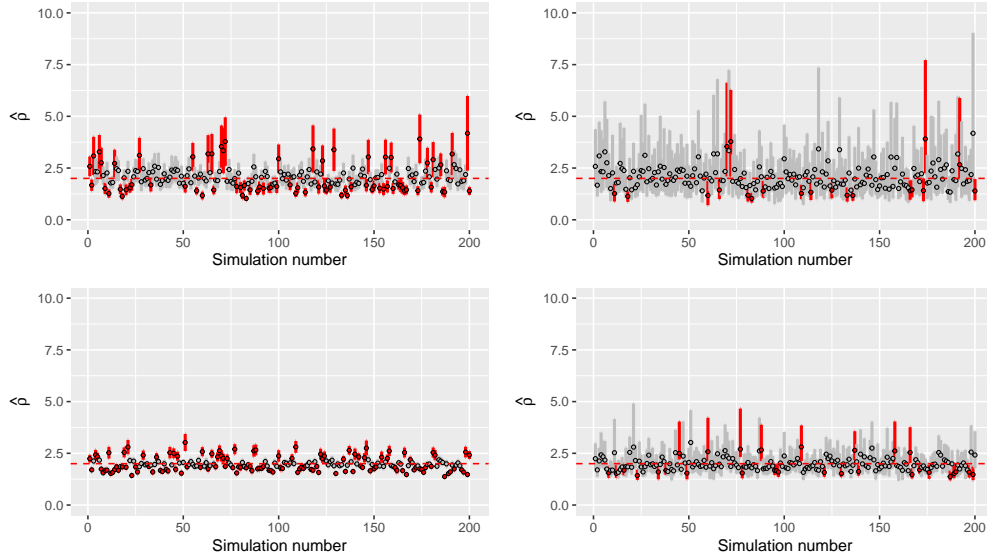


Figure 2.6 – Nominal 95% sandwich (left) and bootstrap (right) confidence intervals for the range parameter ρ (with logarithm as variance stabilizing transform) in the model BR_1 using the two-step estimation procedure, from 200 simulations with 40 replications, $D = 25$ (top) and $D = 225$ (bottom). The red confidence intervals do not cover the true value represented by the red dashed line.

In the second experiment, we also compare the 95% confidence intervals of the range parameter estimates of BR_1 calculated using the estimated sandwich variance matrix and the non-parametric block bootstrap (Davison and Hinkley, 1997, basic intervals in Chapter 5.2) with logarithm as variance stabilizing transform; see Figure 2.6 for $D \in \{25, 225\}$. The coverage of the sandwich-based intervals deteriorates from 61% to 39% as D increases from 25 to 225, probably as more marginal parameter estimates are treated as known, and is much below 95%, contrary to the coverages of the bootstrap-based confidence intervals (respectively 90% and 85%).

2.5 Case study

2.5.1 Choice of validation locations

Figure 2.7 shows the windows where the max-stability test rejects the null hypothesis for PROD. Similar results hold for CAPE and SRH. The maximum number of months rejected per region is four at the north-west. The problematic areas tend to cluster around the western part of our region, implying that max-stability is spatially dependent. There is also time dependence, as there are more late winter and spring months (e.g., February) where the test is rejected for SRH, late summer (e.g., August) rejections for CAPE, and early winter (e.g., November) rejections for PROD. Leaving out window-month combinations that fail the max-stability test leads to our omitting approximately 15% of data.

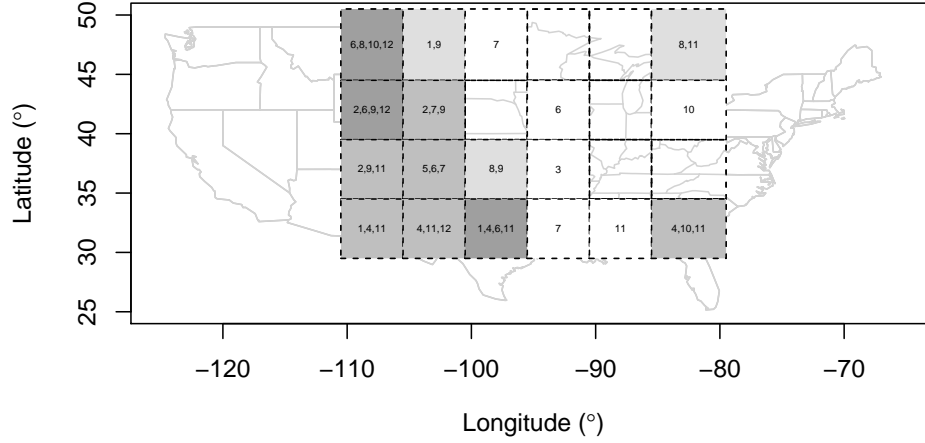


Figure 2.7 – The chosen validation sites from the failure of the max-stability test for PROD. The number in each window indicates the months 1, ..., 12 (January, ..., December) for which we choose the locations in the window for our validation set. The shade is proportional to this number (the more months, the darker) .

2.5.2 Results

We apply the model introduced in Section 2.4.1 to our data, with $k_1 \in \{2, 3, 4\}$ knots in the ENSO direction and $k_2 \in \{4, 5\}$ knots in the month direction. The knots for ENSO are placed evenly between the 10% and 90% quantiles of the ENSO values, i.e., between -1.06°C and 1.16°C , and those for the month are positioned evenly between 0.5 and 12.5; both latter points represent mid-December and correspond to the same cyclic spline values. We also consider the case of a constant range parameter involving no covariate. For each variable, the calibration set corresponds to all corresponding data apart from the validation set. We fit all models to the calibration set using the truncated pairwise likelihood approach (Section 2.3.2) with $c = 2$.

We assess the uncertainty of our estimates with basic confidence intervals, using non-parametric block bootstrap with 200 replicates. For the range parameter, we use the logarithm as a variance stabilizing transformation (Davison and Hinkley, 1997, p. 195) and derive the basic confidence intervals of the log-transformed range before transforming them back onto the original scale. For model selection, we use CLIC^b in (2.19) using the same bootstrap replicates as above. Our block bootstrap resamples the 37 years of data from all grid points to retain the spatial structure of the observations. We choose 19 resampling blocks (18 consist of two consecutive years and one consists of one year), which partly preserves the dependence between December and January.

Table 2.3 shows that the best models for PROD and CAPE have two and four knots in the

Chapter 2. Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

Knots	PROD	CAPE	SRH
None	19'519'829	18'472'723	16'552'412
2×4	19'510'372	18'466'193	16'516'130
3×4	19'515'502	18'468'421	16'629'860
4×4	19'520'216	18'472'426	16'520'391
2×5	19'511'256	18'466'535	16'516'015
3×5	19'519'077	18'470'945	16'521'017
4×5	19'527'238	18'475'342	16'525'644

Table 2.3 – Values of CLIC^b for different configurations of $k_1 \times k_2$, where k_1 and k_2 are the number of knots in the ENSO and month direction, respectively.

Parameter	PROD		CAPE		SRH	
α^*	1.29	(1.27, 1.31)	1.33	(1.30, 1.36)	1.47	(1.44, 1.50)
r	0.72	(0.71, 0.73)	0.76	(0.75, 0.78)	0.61	(0.60, 0.62)
κ	-0.08	(-0.10, -0.06)	-0.13	(-0.15, -0.10)	-0.07	(-0.08, -0.05)

Table 2.4 – For each variable, estimates of the model parameters (an asterisk indicates that the estimate has been bootstrap bias-corrected) with their 90% confidence intervals in brackets.

ENSO and month directions, and that the best one for SRH involves two and five knots. These models clearly outperform that involving a constant range, suggesting that incorporating ENSO and the month is valuable. Graphically, two knots seem enough to capture the ENSO effect, as increasing their number does not substantially modify the main features of the obtained surface; compare Figure 2.8 and plots in Section 2.7.2 of the Supplement. Thus, such an increase seems to bring useless additional complexity, which may explain the systematic choice of two knots in the ENSO direction by CLIC^b. It also increases the uncertainty on the parameter estimates (not shown). Below, for each variable, by model we mean the best model.

Table 2.4 suggests that the models for PROD and CAPE tend to generate slightly rougher sample paths than for SRH, and that moderate anisotropy is present for all variables, with a non-negligible compression in the longitude direction (factor ranging from 0.6 to 0.8) but with a small rotation. Sandwich-based confidence intervals are much narrower for all parameters (not shown), as the non-parametric bootstrap leads to more honest uncertainty quantification since it appropriately accounts for the effects of estimating the margins. For the smoothness parameter, there is a small but systematic difference between the bootstrap estimates and the original one. We estimate the mean of this difference from bootstrap replicates and use it to bias-correct the original estimate.

Figure 2.8 shows that, for all variables, the range parameter ρ is lowest during summer or early fall (July–September) and is highest in winter and spring (December–May). High ENSO values are associated with a shift of the highest range period towards late winter and spring (February–May) and to larger range values during that same period; for each corresponding month, the curve of ρ with respect to ENSO exhibits a kink at 0°C and increases above that value. However, the uncertainty associated with these trends is high for PROD and CAPE.

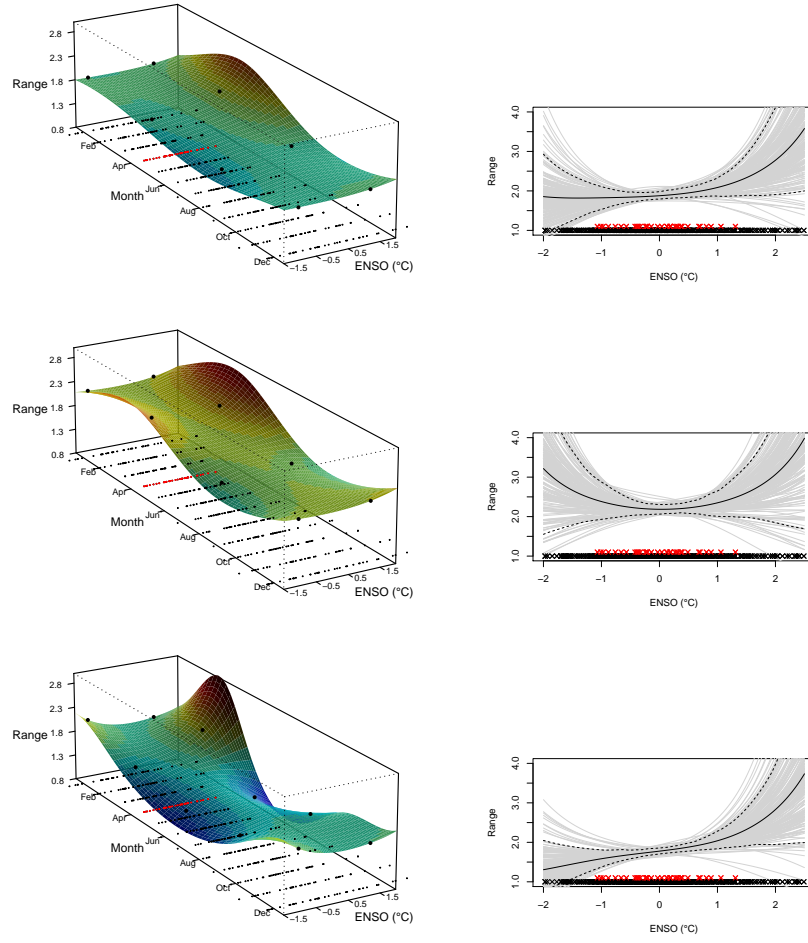


Figure 2.8 – Trend surface plots of the estimated range parameter $\hat{\rho}$ of the model for PROD (top), CAPE (middle) and SRH (bottom). The right panels show the two-dimensional slice of the surface at April in black with, in grey, the values of $\exp\{2\log(\hat{\rho}) - \log(\hat{\rho}_b^*)\}$, $b = 1, \dots, 200$, where $\hat{\rho}_b^*$ is the b -th bootstrap estimate, and, within the dashed lines, the 90% bootstrap pointwise confidence intervals. On all plots, the small red and black dots represent the ENSO values for April and the other months, respectively. The black dots on the trend surfaces indicate the knot positions.

Some of these trends are insignificant at the 90% level, so confirmation would be needed in future studies through the use of more data.

It follows from (2.4) and (2.7) that the bivariate extremal coefficient is decreasing in the range parameter, so our findings on the range parameter also apply to the extremal coefficient. For any value of ENSO and each month, we can compute the bivariate extremal coefficient of our model by combining (2.7) and (2.12). We build Figure 2.9 by doing so for each of the 200 sets of parameter estimates obtained from the 200 bootstrap replicates, and for grid points that are 1° longitude apart. For ENSO = 0°C , the estimates are rather stable across months, except

Chapter 2. Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

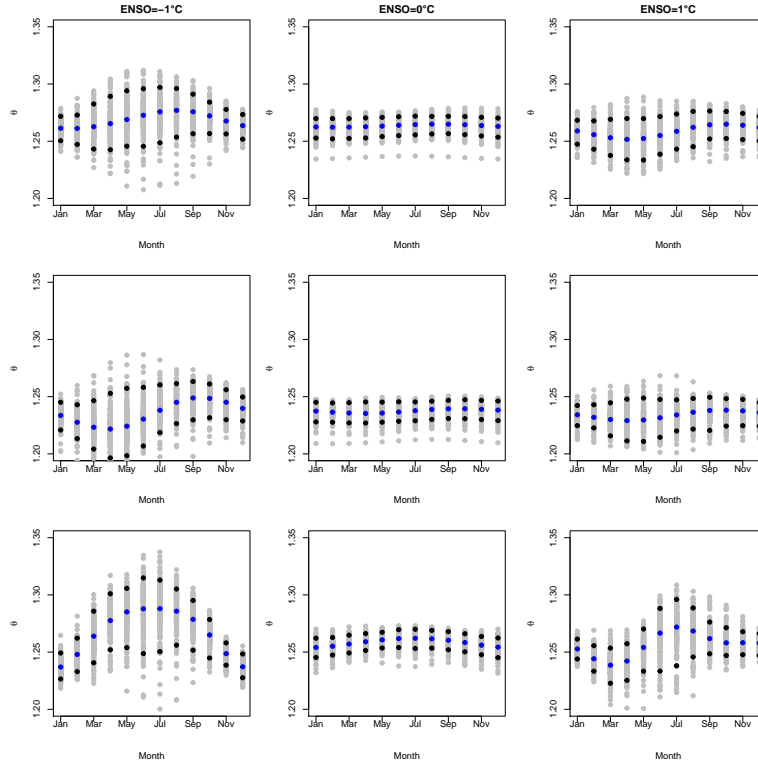


Figure 2.9 – In blue, bootstrap bias-corrected estimates of the model's bivariate extremal coefficient for PROD (top), CAPE (middle), SRH (bottom), when ENSO equals -1°C (left), 0°C (center) and 1°C (right), for two grid points 1° longitude apart. The grey dots display the 200 values of $h^{-1}\{2h(\hat{\theta}) - h(\hat{\theta}_b^*)\}$, $b = 1, \dots, 200$, where $\hat{\theta}_b^*$ is the b -th bootstrap estimate and $h(x) = \log\{(x-1)/(2-x)\}$, $x \in [1, 2]$. The black dots indicate the lower and upper confidence bounds of the 90% bootstrap confidence intervals.

for SRH for which they are higher in summer (June–August). The seasonal variation is more pronounced when $\text{ENSO} = -1^{\circ}\text{C}$ and $\text{ENSO} = 1^{\circ}\text{C}$ for all variables, and becomes significant at the 95% level for SRH; e.g., the estimated extremal coefficient during an El Niño event in July is significantly higher than in January–April, November and December. The decrease of the extremal coefficient (related to an increase of the range parameter) in late winter and spring during El Niño years is also clearly visible. For SRH, Figure 2.10 displays at each grid point the value of the estimated model bivariate extremal coefficient between that point and a reference point at the center of the map. The extremal coefficients are lower in April than August for the three chosen values of ENSO (-1°C , 0°C , 1°C) and decrease with ENSO in April; this indicates an increase of the extremal dependence, consistent with the increased range parameter. The contour lines in Figure 2.10 nicely illustrate the variation with ENSO of the spatial extent of the extreme events of SRH.

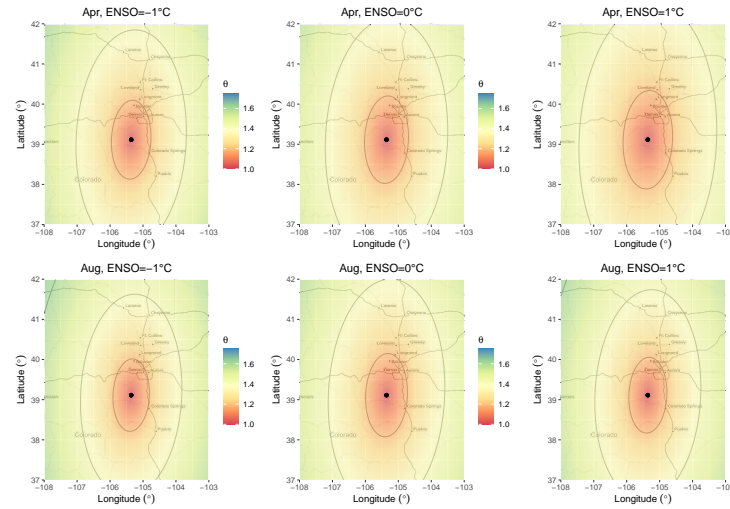


Figure 2.10 – Heat maps of the bootstrap bias-corrected estimates of the model’s bivariate extremal coefficient for SRH (with one reference fixed in Colorado, indicated by the black dot on each map) when ENSO equals -1°C (left), 0°C (center) and 1°C (right), in April (top) and August (bottom). The grey contour lines represent the 1.2, 1.4 and 1.6 levels.

2.5.3 Model validation

We assess our model’s performance on the validation set specified in Section 2.5.1. Figure 2.11 shows that the theoretical pairwise extremal coefficients computed from our model agree with the empirical ones, despite slight underestimation on the right panel. Our model appropriately captures the general dependence decrease as distance increases. Moreover, for PROD and SRH in the chosen window-month combinations, elevated values of ENSO are associated with lower empirical extremal coefficients than when ENSO is close to 0°C , which is not the case for CAPE in the chosen window in August; our model captures these effects well. Overall, the out-sample performance of our model is satisfactory.

The few departures between our model and the data likely stem from the spatially-constant extremal dependence in our model and the heterogeneity of the considered region in terms of weather influences. Koch et al. (2021, Figure 8) showed that the February maxima at each grid point are significantly increasing with ENSO around the Gulf of Mexico, although they significantly decrease with ENSO over the remainder of the region considered. If the marginal impact of ENSO varies across space, its effects on the dependence may also be space-varying. Finally, our model validation is not entirely fair as the model extremal coefficient is computed for a given distance and ENSO value, whereas the boxplots of Figure 2.11 are built using a range of distances and ENSO values.

Chapter 2. Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

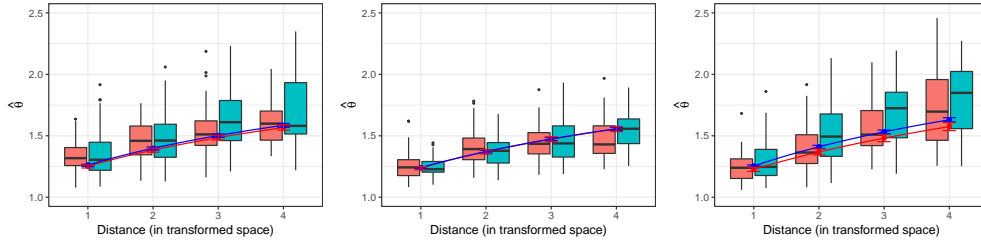


Figure 2.11 – Out-sample performance of the model for PROD (left), CAPE (center) and SRH (right), on the window-month combinations (8th, February), (17th, August) and (2nd, March), respectively, where the window's numbering is from West to East and North to South. The lines depict the bootstrap bias-corrected model's extremal coefficients computed at ENSO=1°C (red) and ENSO=0°C (blue) with respect to distance in the transformed space; the latter has been computed using (2.3) with the estimated κ and r . The whiskers indicate the pointwise basic 90% confidence intervals calculated with non-parametric bootstrap. The four pairs of boxplots summarize the empirical estimates of the extremal coefficient for pairs of grid points whose distance in the transformed space belongs to $[0.5, 1.5)$, $[1.5, 2.5)$, $[2.5, 3.5)$ and $[3.5, 4.5)$. The red and blue ones correspond to the values when ENSO > 0.5°C and $|\text{ENSO}| < 0.2^\circ\text{C}$, respectively.

2.5.4 Meteorological explanation

As mentioned in Section 2.3.1, the range parameter ρ in (2.4) can be interpreted for each variable as a characteristic spatial dimension of the individual extreme events of that variable, and may thus be viewed as a characteristic dimension of cyclones or thunderstorm systems (including multi-cell storms, squall lines, supercells and mesoscale convective systems). Thus, our results indicate that cyclones and thunderstorm systems are more localized in summer than in winter, in agreement with the fact that weather systems generally have a larger scale in winter than in summer. Our findings also suggest that the cyclones and thunderstorms systems are more spatially extended during El Niño events than during neutral ENSO states or La Niña events. On the other hand, Koch et al. (2021) found that, in late winter and spring, maxima of PROD, CAPE and SRH tend to be larger during La Niña years, possibly implying more intense thunderstorm systems. The combination of findings points to more spatially widespread but less intense thunderstorm systems during El Niño years than La Niña years. Possible rough physical explanations are as follows. El Niño winter events tend to be associated with positive and negative temperature anomalies in respectively the northern and southern part of the US (e.g., Ropelewski and Halpert, 1986; Zhang et al., 2011), and thus with a less pronounced north-south temperature gradient than during neutral ENSO states or La Niña episodes. El Niño periods are also related to positive rainfall anomalies in the southern part of the US, especially California (e.g., Ropelewski and Halpert, 1986; Jong et al., 2016). These anomalies tend to persist during the following spring, while progressively weakening (see https://www.oc.nps.edu/webmodules/ENSO/NA_EN.html). The temperature gradient is a key factor of instability as the shock between warm and cold air typically triggers rising motion, and one generally expects larger gradients to be linked with more localized and more

dynamic cyclones or mesoscale systems such as mesoscale convective systems. Another, less plausible, explanation is that the combination of the warming in the northern part of the US and increased rainfall amounts in the south associated with El Niño is expected to lead to larger areas of instability, as warm air and moisture are two important ingredients of instability.

The main drivers of weather are not the same across the region, and thus finding an interpretation that is true throughout it is delicate. More investigation as well as collaborations with climatologists would be needed to confirm these coarse interpretations and to thoroughly understand the very complex physical mechanisms involved. The link between ENSO and the spatial extent of weather phenomena has received relatively little attention from researchers; see, e.g., Lyon (2004) and Lyon and Barnston (2005) in the case of the drought and rainfall extremes, respectively.

2.6 Discussion

This chapter provides three tools for fitting max-stable fields to gridded pointwise maxima. The first, a max-stability test which we used for out-sample selection, allowed us to pinpoint regions where a max-stable model may be unsuitable, without relying on testing secondary properties like asymptotic independence (e.g., Bacro et al., 2010). Indeed, even if data show no evidence of asymptotic independence, this need not imply that max-stable models should be used; data could be asymptotically dependent but not max-stable. The development of subasymptotic models remains important, and the test developed here provides an additional diagnostic to assess whether these subasymptotic models are needed.

The second tool is a new model selection procedure based on the block bootstrap for use in situations where one uses composite likelihood in two-step inference, where the margins are first estimated and then standardized, and then the dependence structure is estimated. This fixes the way the CLIC is currently computed, though the relevance of the CLIC itself could be questioned if the primary focus is on specific regions (e.g., Tornado Alley) or months (e.g., in spring, where the risk of thunderstorms is highest), rather than on the full dataset. Scores for out-sample analysis remain an open question, especially for extreme values (Brehmer and Strokorb, 2019).

The third tool is a Brown–Resnick model that can capture non-stationarities in the dependence in the extremes of data via a spline-based methodology. An alternative is to adapt the covariance function associated with the Brown–Resnick model, and ideas could be borrowed from Huser and Genton (2016) and Risser and Calder (2015) to introduce other non-stationarities through covariates in this way, though care should be taken not to overparameterize spatial covariance functions.

Using the tools developed, our case study assessed ENSO and month-related signals that affect the spatial dependence of the extremes of PROD, CAPE and SRH. Our models appear justified in this setting: we use the max-stability test to filter data before model fitting, and model

Chapter 2. Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

validation shows satisfactory results. The models pool information from all months to estimate seasonal ENSO signals. We detect differences in the range parameter for all variables in the spring and summer; the corresponding extremes are more localized during these seasons than in the winter, especially during El Niño events. These conclusions are significant and clearer for SRH (and to a lesser extent, PROD), but less so for CAPE because of the wide bootstrap confidence intervals. These results add to those in Koch et al. (2021), who found ENSO to be a good covariate in the location parameter of the GEV for PROD and SRH maxima in February. The corresponding relationship is negative over the same region we consider, implying that the risk of storm impacts in February decreases during El Niño years. The results here suggest that the spatial extents of storm events in these years are also higher.

2.7 Supplement

2.7.1 Algorithms

Let $\mathbf{y}_j = (y_{1,j}, \dots, y_{D,j})'$ denote the j -th three-hourly observation, empirically transformed to be standard Fréchet-distributed, where D is the number of locations. Given n three-hourly observations, let $(\mathbf{y}_1, \dots, \mathbf{y}_n)' \in \mathbb{R}^{n \times D}$ denote the transformed dataset. The procedure for generating the max-stable vector is

1. Compute $\tilde{R}_j = \|\mathbf{y}_{1,j}, \dots, \mathbf{y}_{D,j}\|_1$, known as the radial coordinates, where $\|\cdot\|_1$ is the L_1 norm. Then calculate $\tilde{\mathbf{W}}_j = \mathbf{y}_j / \tilde{R}_j$, $j = 1, \dots, n$, commonly referred as angular coordinates. Keep those $\tilde{\mathbf{W}}_j$ for which $\tilde{R}_j > r_0$, with r_0 fixed to be the empirical $p \in (0, 1)$ quantile of $\tilde{R}_1, \dots, \tilde{R}_n$, so the retained number of observations is $n_0 = (1 - p)n$. Let R_i and \mathbf{W}_i ($i = 1, \dots, n_0$) denote the coordinates retained.
2. Following the empirical likelihood approach of Einmahl and Segers (2009), but extended to $D > 2$, we find the estimated angular probability measure

$$\hat{Q}(\mathbf{w}) = \sum_{i=1}^{n_0} q_i \mathbb{I}_{\{\mathbf{W}_i < \mathbf{w}\}}, \quad \mathbf{w} \in [0, 1]^D,$$

where $\mathbb{I}_{\{\cdot\}}$ is the indicator function and the tilting probabilities $\{q_i\}_{i=1}^{n_0}$ satisfy

$$q_i = \arg \max_{q_i} \prod_{i=1}^{n_0} q_i, \text{ such that } \sum_{i=1}^{n_0} q_i = 1 \text{ and } \sum_{i=1}^{n_0} q_i \mathbf{W}_i / R_i = \mathbf{D}^{-1}, \quad (2.20)$$

where $\mathbf{D}^{-1} = (1/D, \dots, 1/D)'$, and (2.20) is solved using Lagrange multipliers.

3. With $\{q_i\}_{i=1}^{n_0}$ and $\{\mathbf{W}_i\}_{i=1}^{n_0}$, generate a simple max-stable vector using Algorithm 1, based on Dombry et al. (2016).

2.7.2 Plots of trend surfaces

Figures 2.12, 2.13 and 2.14 show the same plots as the left panels of Figure 2.8, for the fitted models from Table 2.3.

Algorithm 1: Simulate max-stable vector using tilting weights

Input: Tilting weights $\{q_i\}_{i=1}^{n_0}$ and vectors $\{W_i\}_{i=1}^{n_0}$
Output: Max-stable D -dimensional vector with standard Frechét margins
Generate $E^* \sim \text{Exp}(1)$;
Set $R^* = D/E^*$;
Set $Z = (Z_1, \dots, Z_D) = (0, \dots, 0)$;
while $R^* > \min\{Z_1, \dots, Z_D\}$ **do**
 Draw $W' = (W'_1, \dots, W'_D)$ from the set $\{W_1, \dots, W_{n_0}\}$ with sampling probabilities $\{q_1, \dots, q_{n_0}\}$;
 for $j \leftarrow 1$ **to** D **do**
 Set $Z_j = \max(Z_j, R^* W'_j)$;
 end
 Generate $E^* \sim \text{Exp}(1)$;
 Set $R^* = \frac{1}{(1/R^* + E^*/D)}$;
end
return Z ;

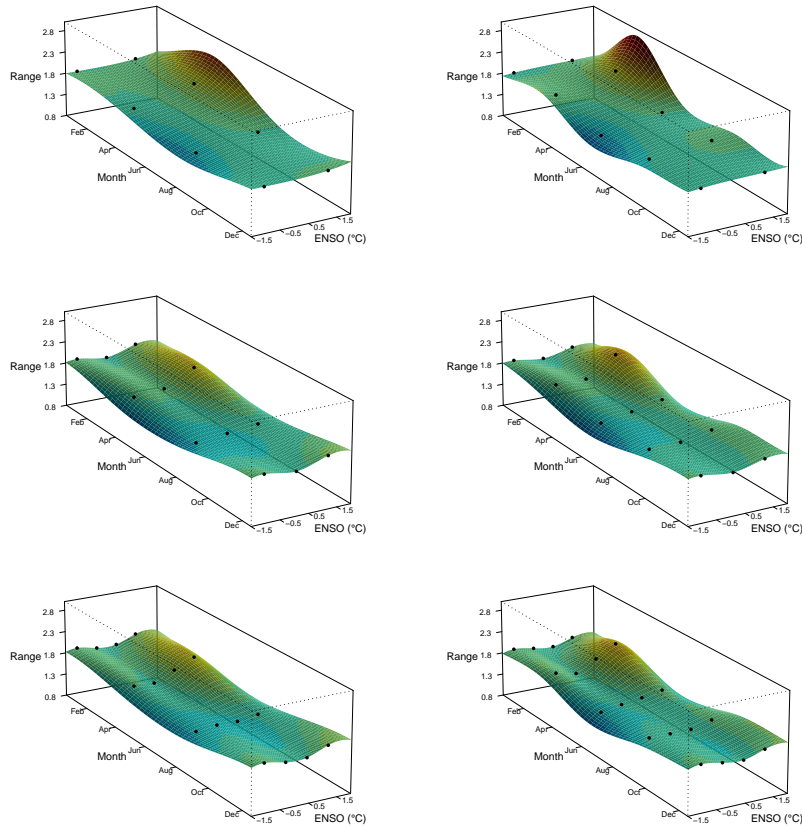


Figure 2.12 – Trend surface plots of the range parameter in the models with (from top left to bottom right) 2×4 , 3×4 , 4×4 , 2×5 , 3×5 , 4×5 knots in the ENSO and month direction for PROD. The black dots on the trend surfaces indicate the knots' positions.

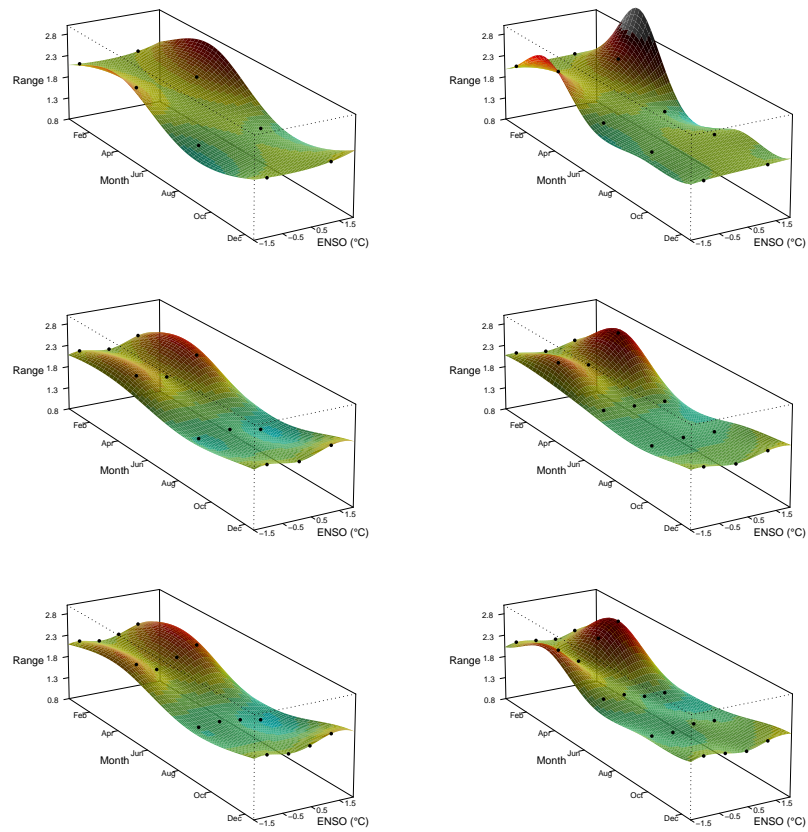


Figure 2.13 – Same as Figure 2.12, but for CAPE.

Chapter 2. Practical tools for fitting max-stable fields, applied to severe US thunderstorm environments

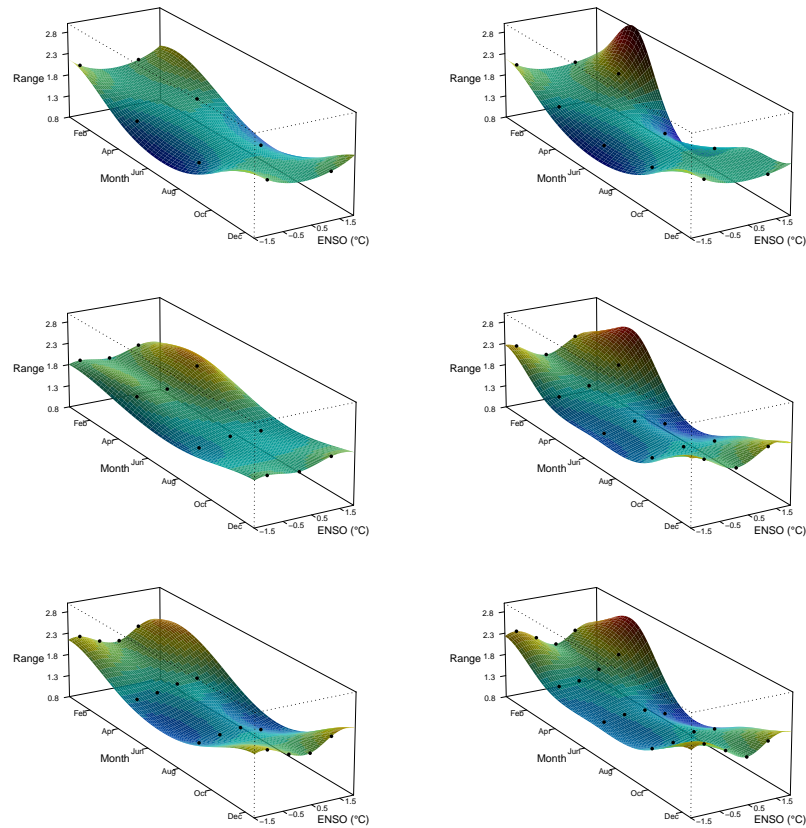


Figure 2.14 – Same as Figure 2.12, but for SRH.

Wildfires Part II

3 Gradient boosting with extreme-value theory for wildfire prediction

This chapter is a preprint of a paper that has been submitted to a peer-reviewed journal, so it is structured to be self-contained and may overlap with other chapters. It is available as Koh (2021). The doctoral candidate is entirely responsible for this work.

3.1 Abstract

This chapter details the approach of the team *Kohrrelation* in the 2021 Extreme Value Analysis data challenge, dealing with the prediction of wildfire counts and sizes over the contiguous US. Our approach uses ideas from extreme-value theory in a machine learning context with theoretically justified loss functions for gradient boosting. We devise a spatial cross-validation scheme and show that in our setting it provides a better proxy for test set performance than naive cross-validation. The predictions are benchmarked against boosting approaches with different loss functions, and perform competitively in terms of the score criterion, finally placing second in the competition ranking.

Keywords: Cross-validation; Generalized Pareto distribution; Gradient boosting; Loss likelihood; Wildfire prediction

3.2 Introduction

Wildfires occur in every season of the year and are a natural phenomenon of the forest ecosystem, important for clearing out decayed vegetation and helping plants to reproduce. However, they have the potential to become conflagrations—intense, destructive fires—that may have huge environmental and ecological impacts. Apart from human casualties, these fires can lead to substantial economic losses; global insured claims due to wildfire events have increased dramatically in recent years, from below \$10 billion in 2000–2009 to \$45 billion in the subsequent decade¹.

Wildfires are complex dynamic processes: their occurrences and behaviour are the product of interconnected factors that include the ignition source, fuel composition, topography and the

¹<https://www.swissre.com/risk-knowledge/mitigating-climate-risk/yet-more-wildfires.html>

weather. For example, the wind plays a big role in the spread and ease of fire containment, but its effect is magnified in the presence of accumulated biomass on a hilly boreal forest after a prolonged dry spell. The modelling of wildfires is made even more complicated by the need to model the Wildland-to-Urban interface (Stewart et al., 2007), as 90% of fires are caused by human activity.

An important measure of wildfire impact and size is the burned area of wildfire events, commonly used by government agencies and aggregated at different spatial and temporal scales for reporting purposes, e.g., National Interagency Fire Center (2021). There is a positive but non-linear relationship between wildfire counts and sizes in the contiguous US. In 2020, nearly 26,000 wildfires burned approximately 9.5 million acres (ac) in the west, compared with the over 33,000 fires that burned just under 0.7 million ac in the east. Similarly, although the numbers of wildfires have fallen since the 1990s, the average annual acreage burned since 2000 has more than doubled.

Many statistical approaches have been developed to aid in wildfire prevention and risk mitigation, with most studies modelling wildfire occurrences and sizes separately (Taylor et al., 2013; Xi et al., 2019; Pereira and Turkman, 2019; Jain et al., 2020), though models that identify latent factors affecting both have been proposed (e.g., Koh et al., 2021). Point processes are natural models for the spatiotemporal pattern of occurrences (Peng et al., 2005; Genton et al., 2006; Tonini et al., 2017; Opitz et al., 2020b). Cumming (2001), Cui and Perera (2008) and Pereira and Turkman (2019) suggested modelling fire sizes with various probability distributions. As data usually show heavy-tailed behaviour, only a small fraction of wildfires account for the vast majority of the area burned. Obvious candidates to capture this stem from extreme-value theory, such as the generalized Pareto distribution (GPD) for modelling threshold exceedances (De Zea Bermudez et al., 2009; Turkman et al., 2010; Pereira and Turkman, 2019).

Both Bayesian and frequentist methods have been used for explanatory modeling, the former predominantly for hierarchical mixed effect models (Koh et al., 2021; Joseph et al., 2019; Pimont et al., 2021) and the latter within the generalized additive modelling (GAM, Wood, 2017) framework (Preisler et al., 2004; Woolford et al., 2011; Brillinger et al., 2006; Vilar et al., 2010). Covariates include weather variables such as humidity, temperature, precipitation and meteorologically-based fire danger indices such as the Canadian Fire Weather Index (van Wagner, 1977). When available, land-use or locally observed anthropogenic variables like population density and the distance to the nearest train line are used to help to explain human-induced occurrences; spatiotemporal random effects have been incorporated as surrogates for these variables.

If accurate prediction is of primary interest, then machine learning (ML) techniques offer an attractive alternative to the statistical modelling approaches described above. Since the 1990s, the surge in the availability of data and covariates has spurred the use of these techniques to predict wildfire behaviour. Jain et al. (2020) found 127 journal papers or conference proceedings published up to the end of 2019 on ML applied to fire occurrence, susceptibility and risk;

of these adversarial neural networks (ANN) were the most prominent (*e.g.*, Liang et al., 2019; Dutta et al., 2013; Shidik and Mustofa, 2014). For wildfire occurrences, most studies focus on classification tasks instead of count modelling. Sakr et al. (2010) used meteorological variables with support vector machines to predict a four-class fire risk index based on the daily number of fires in Lebanon. Dutta et al. (2013) compared ten ANN based cognitive imaging systems to determine the relationship between monthly fire incidence and climate for Australia. Xie and Peng (2019) and Mitsopoulos and Mallinis (2017) showed that ensemble learning methods like random forests and boosting trees performed well in estimating area burned or classifying wildfire sizes in Portugal and Greece, respectively.

Gradient boosting techniques (Friedman, 2001) have exploded in popularity over the last decade, in part due to the development and dissemination of open-source packages such as *gbm* (Greenwell et al., 2020) and *xgboost* (Chen and Guestrin, 2016). A key ingredient of gradient boosting is the loss function used to train these models, and choices for these functions have largely been restricted to those that emphasize good prediction of the distributional bulk instead of the tails. For example, squared loss, the default when modelling wildfire sizes, implicitly presupposes normality of the response given the covariates, which may be inappropriate if the focus is predominantly on extreme values. The Poisson loss is popular for modelling wildfire counts within a grid cell, but the zero-inflated nature and potentially heavy tails of count distributions suggest that this loss may be unsuitable. Evaluation metrics should also reflect the non-linear impact of wildfire events; *e.g.*, in many cases, predicting a false negative occurrence is much costlier than predicting a false positive.

As ML methods are prone to overfitting, it is imperative to evaluate models with held-out datasets using robust validation schemes. A realistic approach in the forecasting context (when there are no trends) is to leave out the most recent portion of the dataset (*e.g.*, Dutta et al., 2013; Koh et al., 2021; Joseph et al., 2019; Woolford et al., 2011). Dutta et al. (2013) explored different combinations of training-testing splits to identify the best possible paradigm to maximize the generalization capability of their ANN architecture. *K*-fold cross-validation is also popular (De Angelis et al., 2015; Shidik and Mustofa, 2014; Xie and Peng, 2019; Mitsopoulos and Mallinis, 2017), but it may give overly optimistic evaluations for spatially dependent data (Roberts et al., 2017). An alternative is spatial cross-validation (Pohjankukka et al., 2017), but it is still unclear how best to construct spatial folds in this context, and doing so anyway ignores time dependencies.

Our work aims to tackle the limitations of the studies mentioned above, and does so in the context of the Extreme Value Analysis 2021 data challenge (Opitz, 2021). We develop novel gradient boosting models trained with loss functions appropriate for predicting extreme values. Our model for fire counts is a discrete generalized Pareto distribution (Shimura, 2012) relying on a covariate-dependent parameter that models a chosen high quantile of the distribution, and a shape hyperparameter selected by cross-validation. The model for fire sizes has three components and covariate-dependent probabilities. The first component models the probability of observing no fires, and the others model the probabilities of observing

medium-sized and extreme fires. We approximate the conditional distribution above a high threshold with a GPD, and the conditional distribution below the threshold with a truncated log-gamma distribution.

To improve our models, we also engineered new covariates that incorporate more spatial information into the climatic and land-use covariates provided by averaging them across neighbouring grid cells each month. With a smart imputation method for replacing missing data, we also use the wildfire counts as a covariate when predicting wildfire sizes, and vice versa.

We develop a spatiotemporal cross-validation scheme that provides a better proxy for our models' test set performance than the naive scheme. This involves fitting a space-time latent Gaussian model to pseudo-binary observations that indicate whether a grid cell was masked in a particular month, and then simulating from the fitted model to generate folds of train-test regimes.

In the remainder of the chapter, we first explore the data on wildfires and their covariates, and then introduce the problem set out by the data challenge in §3.3. We provide general background on extreme-value theory and gradient boosting and on how to combine them in §3.4. Our spatiotemporal cross-validation scheme is developed in §3.4.4 and the specific model structure is detailed in §3.5. We highlight the prediction of wildfire activity components in §3.5.2, and compare them to related and competing approaches. We conclude with a discussion and outlook in §3.6.

3.3 Data and exploratory analyses

The Extreme Value Analysis 2021 data challenge dealt with the prediction of monthly wildfire counts and burned areas at 3503 grid cells across the contiguous US over the period 1993–2015. As fuel moisture is an integral of past precipitation and evaporation mediated by soil field capacity, temporal scales longer than hourly or daily (e.g., monthly in our case) are appropriate for predicting fire risk from climatic covariates.

The data comprise the monthly numbers of wildfires (CNT) and the aggregated burned area (BA) in each grid cell based on a $0.5^\circ \times 0.5^\circ$ grid of longitude and latitude coordinates (roughly 55km \times 55km) covering the study area, from March to September each year. Figure 3.1 shows that the grid cells with the highest averaged CNT tend to be clustered towards the west (California) and southeast (North and South Carolina) corners of the study region, while clusters in the west (near the border of Idaho and Nevada), southwest (Arizona, New Mexico and Texas), and southeast (Florida) are observed for BA.

Thirty-five auxiliary variables related to land cover, weather and altitude are provided at the same spatial and temporal resolution, and can be used for modelling. Figure 3.2 hints at the importance of some of these variables, such as meteorological covariate 5 (clim5; potential

evaporation, measured in meter water equivalent – mwe) and land cover covariate 7 (lc7; tree needleleaf evergreen closed to open, measured in % of the grid cell), for predicting high BA. However it also indicates that their effect differs over space, and that the interactions between these covariates may be important for predicting large burned areas.

The Rocky Mountain Area and Great Basin, two regions defined formally as ‘Geographic Area Coordination Centers’ by the United States Department of Agriculture, have empirical exceedance probabilities that respond differently to both covariates. For instance, the probability tends to decrease, then increase (after the 75% quantile) with potential evaporation in the Great Basin, while the opposite holds for the Rocky Mountain Area, though the associated large uncertainties suggest that there is substantial heterogeneity within each region.

The original dataset has no gaps, but missing data were artificially created to compare predictive approaches; the full dataset was split into training and testing subsets to evaluate participants’ test scores. No data were masked in the odd years, but 80,000 observations of each variable were masked in the even years. Figure 3.3 shows that the spatial and temporal positions of test data are clustered in space, while there is little evidence for temporal dependence across months. Moreover, the test grid cells for BA and CNT are correlated. This masking is reminiscent of a real-world situation in which two related and spatially dependent processes could render both CNT and BA unobservable at small spatial clusters every month (e.g., from spatially dependent but temporally independent measurement error), and one could only use the available covariates and responses from the surrounding non-masked regions for prediction.

The evaluation metrics used for the competition (see Opitz, 2021) require estimates of the probability $\Pr(\text{BA} < u_{\text{BA}})$ and $\Pr(\text{CNT} < u_{\text{CNT}})$ for 28 thresholds u_{CNT} and u_{BA} . The metrics are variants of weighted ranked probability scores and put relatively strong weight on good prediction in the extremes of the distributions of counts and burned areas. As such, we expect that models that emphasize accurate modelling of the largest counts and burned areas will perform better, and we achieve this by appealing to extreme-value theory.

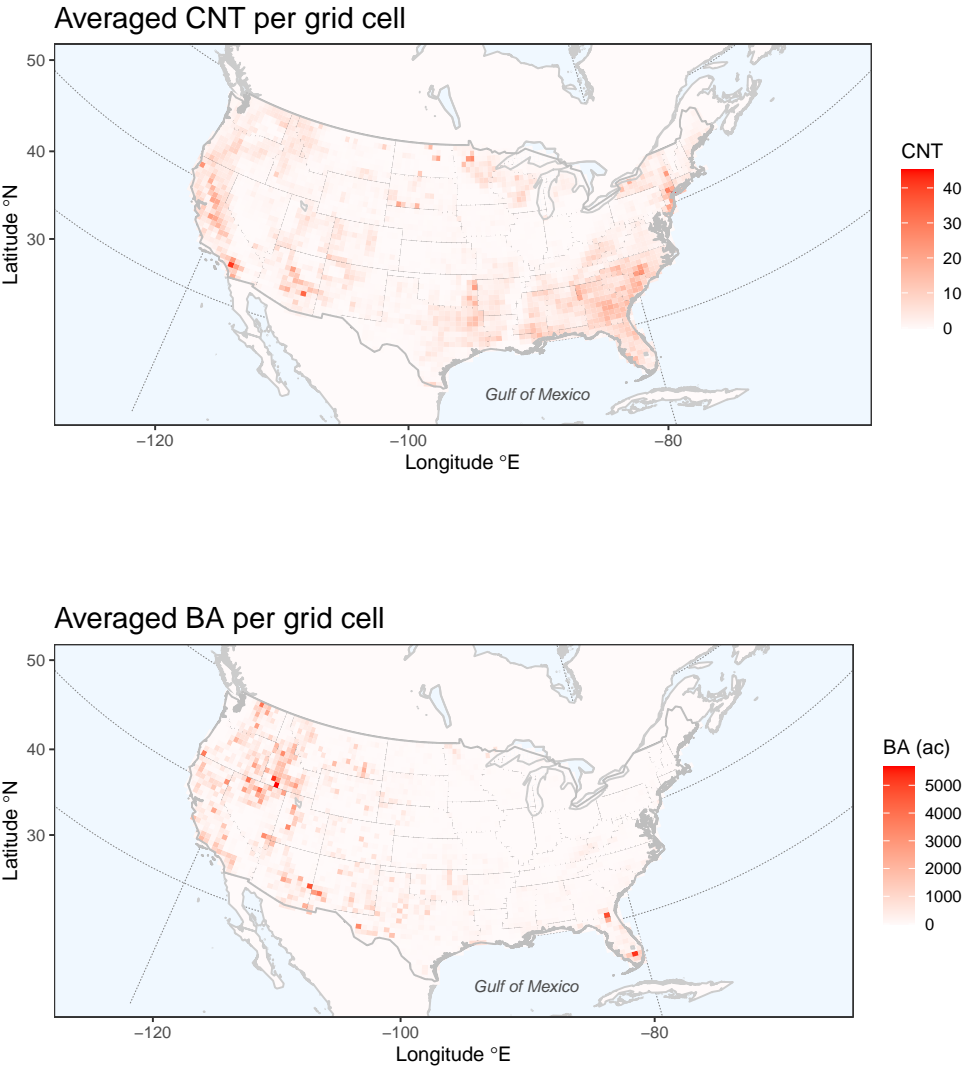


Figure 3.1 – Maps of CNT (top) and BA (bottom) averaged across all months and years for each grid cell.

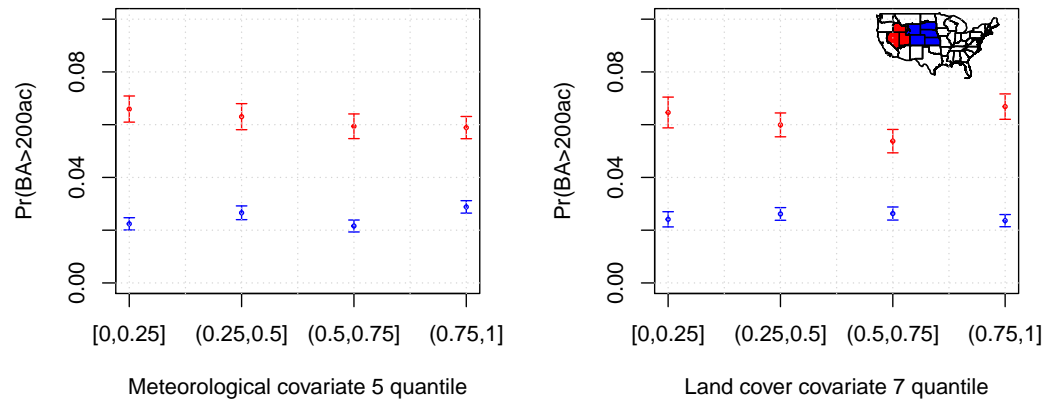


Figure 3.2 – Empirical frequency $\text{Pr}(\text{BA} > 200\text{ac})$ with 95% error bars, as a function of the potential evaporation (in m, left) and tree needleleaves evergreen closed to open (in %, right) covariates, grouped by observations within four empirical quantile ranges, for grid cells within the Rocky Mountain Area (blue) and Great Basin (red) coordination centers. The two regions are highlighted in blue and red in the map inserted in the right panel.

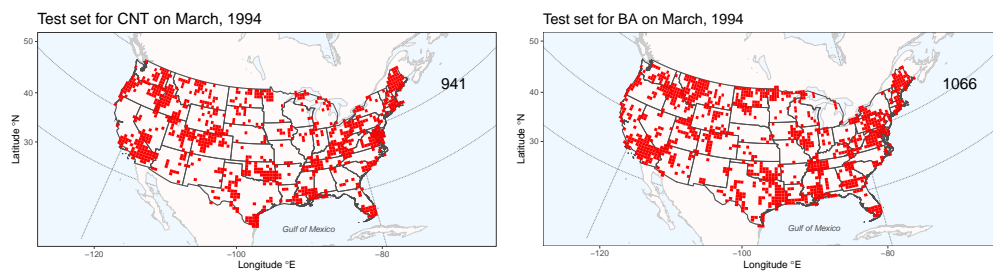


Figure 3.3 – The test set grid cells (in red) for CNT (left) and BA (right) in March 1994. The number at the top right indicates the number of masked grid cells.

3.4 Methodology

3.4.1 Extreme-value theory

The generalized Pareto distribution (GPD) arises asymptotically for excesses above a large threshold of a random variable $X \sim F$, when the distribution F satisfies mild regularity conditions. Let $x^* = \sup\{x : F(x) < 1\}$. The excess distribution above $u < x^*$ can be approximated (Davison and Smith, 1990) as

$$\Pr(X > x + u \mid X > u) \approx 1 - \text{GPD}_{\sigma, \xi}(x) = \begin{cases} (1 + \xi x / \sigma)_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-x / \sigma), & \xi = 0, \end{cases} \quad x > 0, \quad (3.1)$$

with shape parameter $\xi \in \mathbb{R}$ and scale parameter $\sigma = \sigma(u) > 0$, where $a_+ = \max(a, 0)$. The shape parameter determines the rate of tail decay, with slow power-law decay for $\xi > 0$, exponential decay for $\xi = 0$, and polynomial decay towards a finite x^* , for $\xi < 0$. When the approximation (3.1) is exact asymptotically (i.e., when $u \rightarrow x^*$), we say that the random variable X lies in the maximum domain of attraction of a generalized Pareto distribution with shape parameter ξ , written $X \in \text{MDA}_\xi$.

Positive discrete random variables

We say that a discrete non-negative random variable Y lies in the discrete maximum domain of attraction, $Y \in \text{dMDA}_\xi$, if there exists a random variable $X \in \text{MDA}_\xi$ with $\xi \geq 0$ such that $\Pr(Y \geq k) = \Pr(X \geq k)$, for $k = 0, 1, \dots$. The random variable X is called the *version* of Y , and many popular discrete distributions such as the geometric, Poisson and negative binomial distributions lie in dMDA_ξ .

For $Y \in \text{dMDA}_\xi$ and large integers u ,

$$\begin{aligned} \Pr(Y - u = k \mid Y \geq u) &\approx \text{GPD}_{\sigma, \xi}(k + 1) - \text{GPD}_{\sigma, \xi}(k) \\ &= (1 + \xi k / \sigma)_+^{-1/\xi} - (1 + \xi(k + 1) / \sigma)_+^{-1/\xi}, \end{aligned} \quad (3.2)$$

where the last term is the probability mass function of the discrete generalized Pareto distribution (dGPD). Several studies have used this distribution to model count data; Prieto et al. (2014) modelled numbers of road accidents and Hitz et al. (2017) modelled numbers of extreme tornadoes per outbreak and multiple births.

3.4.2 Gradient tree boosting

The generic gradient boosting estimator (Bühlmann and Hothorn, 2007) is a sum of base procedure estimates. Regression trees (Breiman et al., 1984, CART) are popular base procedures, as they include non-linear covariate interactions by construction, and are invariant under monotone transformations of covariates, so the user need not search for good data

transformations.

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}$, $\mathbf{x}_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}$, $i = 1, \dots, n$, be a dataset with n observations and p covariates. A *binary split* of the covariate space uses a splitting variable indexed by $j \in \{1, \dots, p\}$ and a split point $v \in \mathbb{R}$ to partition the space into the pair of half-spaces $\{\mathbf{x} \in \mathbb{R}^p : x_j \leq v\}$ and $\{\mathbf{x} \in \mathbb{R}^p : x_j > v\}$, where x_j is the j -th component of \mathbf{x} .

By successive binary splits, a regression tree partitions the covariate space into a set of L disjoint regions A_1, \dots, A_L , and fits a simple model such as a constant in each region. The regions created by the splits are called nodes; a terminal node is called a *leaf* and an interior node is called a *branch*. We index leaves by $l \in \{1, \dots, L\}$, with leaf l representing region A_l . The simplest tree is one with two leaves, known as a stump. A learning algorithm needs to decide the tree structure, i.e., the splitting variables and split points.

Suppose that L leaves with regions A_1, \dots, A_L have been chosen and we model the response as a *score* $c_l \in \mathbb{R}$ in each region. A regression tree is a function

$$f(\mathbf{x}_i) = \sum_{l=1}^L c_l \mathbb{I}(\mathbf{x}_i \in A_l),$$

where \mathbb{I} is the indicator function. A gradient tree boosting model uses T such trees to model the *boosting estimate*

$$\hat{y}_i = \sum_{t=1}^T f_t(\mathbf{x}_i), \quad f_t \in \mathcal{T}, \quad (3.3)$$

where \mathcal{T} is the space of regression trees. The boosting estimate using stumps will be additive in the original covariates, because every base estimate is a function of a single covariate. A boosting model that has trees with at most L leaves has interactions of order at most $L-2$. Thus, constraining the maximum number of nodes in the base procedure controls the complexity of the model.

Gradient tree boosting learns the set of trees used in (3.3) by minimizing a regularized objective function in a greedy iterative fashion; at each iteration we add the tree that most improves our model according to an objective function \mathcal{O} . More precisely, let $\hat{y}_i^{(j)}$ be the boosting estimate for the i -th observation at the j -th iteration. We add a tree f_j at each iteration to minimize

$$\mathcal{O}^{(j)} = \sum_{i=1}^n \mathcal{L}\{y_i, \hat{y}_i^{(j-1)} + f_j(\mathbf{x}_i)\} + \Omega(f_j),$$

where $\Omega(f_j) = \eta L^{(j)} + \lambda \|\mathbf{c}\|^2 / 2$, \mathcal{L} is a differentiable convex loss function, $L^{(j)}$ is the number of leaves in the tree f_j and $\mathbf{c} \in \mathbb{R}^{L^{(j)}}$ is the corresponding vector of scores. The regularization term Ω is added to penalize the complexity of each tree, and the positive parameters η and λ control the penalization. The form of Ω is simple enough to allow parallel computational (Chen and Guestrin, 2016).

Using a second-order approximation for the objective (Friedman et al., 2000) gives

$$\mathcal{O}^{(j)} \simeq \sum_{i=1}^n \left\{ \mathcal{L}(y_i, \hat{y}_i^{(j-1)}) + g_i f_j(\mathbf{x}_i) + h_i f_j^2(\mathbf{x}_i) \right\} + \Omega(f_j), \quad (3.4)$$

where $g_i = \partial \mathcal{L}(y_i, \hat{y}_i^{(j-1)}) / \partial \hat{y}_i^{(j-1)}$ and $h_i = \partial^2 \mathcal{L}(y_i, z_i) / \partial z_i^2 |_{z_i = \hat{y}_i^{(j-1)}}$. We then minimize (3.4) with respect to the tree structure and weight vector \mathbf{c} .

Let $I_l = \{i : \mathbf{x}_i \in A_l\}$ denote the *instance set* of leaf l . For a fixed tree structure with regions $A_1, \dots, A_{L^{(j)}}$, the optimal weights \mathbf{c}^* can easily be found and have components

$$c_l^* = -\frac{\sum_{i \in I_l} g_i}{\sum_{i \in I_l} h_i + \lambda}, \quad l = 1, \dots, L^{(j)}.$$

Plugging the weights \mathbf{c}^* into (3.4) and removing the term that does not depend on f_j gives

$$\tilde{\mathcal{O}}^{(j)} = -\frac{1}{2} \sum_{l=1}^{L^{(j)}} \frac{(\sum_{i \in I_l} g_i)^2}{\sum_{i \in I_l} h_i + \lambda} + \eta L^{(j)}, \quad (3.5)$$

which can be used as a scoring function to measure the quality of the tree structure, a role similar to the impurity score in Breiman et al. (1984).

Assume that a split has been performed, and let I_L and I_R denote the instance sets of the left and right leaves from this split. Define $I = I_L \cup I_R$. The loss reduction from this split is

$$G = \frac{1}{2} \left\{ \frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right\} - \eta, \quad (3.6)$$

and (3.6) is used for evaluating candidate split variables and points.

As it is impossible to enumerate all possible tree structures, most existing tree boosting implementations, such as in `scikit-learn` (Pedregosa et al., 2011) and `gbm` (Greenwell et al., 2020), use greedy algorithms that start from a single leaf and iteratively add branches to the tree based on (3.6), until the gain for the best split is negative. Here we use the greedy algorithm implemented in `xgboost` (called the approximate algorithm with weighted quantile sketch Chen and Guestrin, 2016, Appendix A), that further reduces computational cost and parallelizes computations when the data do not fit into memory. This algorithm proposes candidate splitting points from the empirical quantiles of each covariate, instead of considering all possible splitting points for each variable. We also subsample a proportion $s \in [0, 1]$ of the covariates at each iteration, like in the random forest algorithm (Breiman, 2001), to further prevent overfitting and to accelerate parallel computation.

As gradient tree boosting is an ensemble method combining predictions from many trees, the exact relationship between covariates (or their interactions) and the response is difficult to determine. Suppose we treat the covariates as random, and let $\mathbf{X}_{\mathcal{S}}$ denote the random

subvector (of size $l < p$) of the full covariate vector $\mathbf{X} = (X_1, \dots, X_p)^T$, indexed by the set $\mathcal{S} \subset \{1, \dots, p\}$. Let \mathcal{C} be the set complementary to \mathcal{S} . For a predictive function f at a fixed point $\mathbf{x} \in \mathbb{R}^l$, the partial dependence function (Friedman, 2001),

$$f_{\mathcal{S}}(\mathbf{x}) = \mathbb{E}_{\mathbf{X}_{\mathcal{C}}} f\{(\mathbf{x}, \mathbf{X}_{\mathcal{C}})^T\},$$

can be estimated by Monte Carlo as

$$\hat{f}_{\mathcal{S}}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n f\{(\mathbf{x}, \mathbf{x}_{i\mathcal{C}})^T\}, \quad (3.7)$$

where $\mathbf{x}_{1\mathcal{C}}, \dots, \mathbf{x}_{n\mathcal{C}}$ are realisations of $\mathbf{X}_{\mathcal{C}}$ from the training data.

Metrics used to rank covariates in terms of their importance include the *coverage* for a chosen covariate, which is the sum of the second order gradients h_i from (3.5), in each node which uses this covariate, standardized by dividing by the sum of the metrics for all other covariates (so the resulting metric is a proportion). The *gain* metric represents the fractional contribution of a chosen covariate to the model based on the total gain of all the splits involving this covariate, measured by G in (3.6); it is the total improvement of the model in terms of the objective function, from the branches that include the covariate. In both cases a higher proportion implies a more important predictive variable.

The loss function \mathcal{L} in (3.4) strongly governs the type of models that we fit, and we discuss choices for this next.

3.4.3 Loss functions

For wildfire counts

Given n monthly wildfire counts in a grid cell, y_1, \dots, y_n , a simplified Poisson loss is

$$\mathcal{L}_{\text{Pois}}(y_i, \hat{y}_i) = y_i \log\{y_i / \exp(\hat{y}_i)\} - y_i + \exp(\hat{y}_i), \quad (3.8)$$

where Stirling's approximation $\log(y_i!) \approx y_i \log(y_i) - y_i$ is used and the boosting estimate \hat{y}_i models the log mean of the i -th Poisson count. Although (3.8) is the most common choice for modelling wildfire counts in the literature, the zero-inflation and potentially heavy tails of count distributions suggest that it may be unsuitable.

Instead, if we let $\alpha = 1/\xi > 0$ and $\lambda = \sigma\alpha$ in (3.2), we motivate a new loss function for counts from extreme-value theory, the discrete generalized Pareto (dGPD) loss

$$\mathcal{L}_{\text{dGPD}}(y_i, \hat{y}_i) = \{1 + \exp(\hat{y}_i)y_i\}^{-\alpha} - \{1 + \exp(\hat{y}_i)(y_i + 1)\}^{-\alpha}. \quad (3.9)$$

The boosting estimate \hat{y}_i models the logarithm of the rescaled scale parameter λ_i . If $\alpha > 1$, the

predicted mean of the i -th count is

$$\hat{m}_i = \sum_k^{\infty} 1 / \{1 + \exp(\hat{y}_i) k\}^{\alpha}, \quad (3.10)$$

but otherwise the mean does not exist.

The first and second derivatives of (3.9) with respect to the boosting estimate, i.e.,

$$\begin{aligned} g_i^{\text{dGPD}} &= -\alpha \{1 + \exp(\hat{y}_i) y_i\}^{-\alpha-1} \{\exp(\hat{y}_i) y_i\} \\ &\quad + \alpha \{1 + \exp(\hat{y}_i) (y_i + 1)\}^{-\alpha-1} \{\exp(\hat{y}_i) (y_i + 1)\}, \\ h_i^{\text{dGPD}} &= -\alpha(-\alpha-1) \{1 + \exp(\hat{y}_i) y_i\}^{-\alpha-2} \{\exp(\hat{y}_i) y_i\}^2 \\ &\quad - \alpha \{1 + \exp(\hat{y}_i) (y_i + 1)\}^{-\alpha-1} \{\exp(\hat{y}_i) y_i\} \\ &\quad + \alpha(-\alpha-1) \{1 + \exp(\hat{y}_i) y_i\}^{-\alpha-2} \{\exp(\hat{y}_i) (y_i + 1)\}^2 \\ &\quad + \alpha \{1 + \exp(\hat{y}_i) (y_i + 1)\}^{-\alpha-1} \{\exp(\hat{y}_i) (y_i + 1)\}, \end{aligned}$$

are used in the second-order approximation of the objective in (3.4) and are essential for determining split variable and split point candidates when building trees with (3.6).

For wildfire sizes

The squared loss is the dominant choice in the literature on predicting burned areas in this regression context, but it implicitly presupposes normality of the response conditional on the covariates, which may be inappropriate if the distributional tail decays more slowly than exponential. Moreover, burned areas cannot be negative. Modelling log-transformed burned areas addresses the latter issue, though doing so still excludes conditional distributions with Pareto-like tails.

Another approach to modelling the full distribution of wildfire sizes is to use a mixture, first choosing an appropriately high threshold and fitting the burned areas below and above that threshold with different loss functions.

To model the monthly burned area in a grid cell, y_i , below a chosen threshold $u > 0$, the negative log-loss likelihood of a truncated distribution could be used. The probability density function of a truncated gamma distribution is

$$f(x) = \frac{(\mu/k)^k x^{k-1} \exp(-xk/\mu)}{\gamma(k, ku/\mu)}, \quad x > 0,$$

where $\mu > 0$ is the rescaled scale parameter, $k > 0$ is the shape parameter and $\gamma(k, s) = \int_0^s t^{k-1} \exp(-t) dt$, $s > 0$, is the lower incomplete gamma function. Modelling $\log(\mu)$ with

the boosting estimate gives

$$\begin{aligned}
 \mathcal{L}_{\text{trGamma}}(y_i, \hat{y}_i) &= -\log\{(\exp(\hat{y}_i)/k)^k y_i^{k-1}\} \\
 &\quad + y_i k / \exp(\hat{y}_i) + \log\{\gamma\{k, ku / \exp(\hat{y}_i)\}\}, \\
 g_i^{\text{trGamma}} &= \exp(\hat{y}_i) \{k / \exp(\hat{y}_i) - y_i k / \exp(\hat{y}_i)^2 \\
 &\quad + \gamma'\{k, ku / \exp(\hat{y}_i)\} / \gamma\{k, ku / \exp(\hat{y}_i)\}\}, \\
 h_i^{\text{trGamma}} &= \exp(\hat{y}_i)^2 \left\{ -k / \exp(\hat{y}_i)^2 + 2y_i k / \exp(\hat{y}_i)^3 \right. \\
 &\quad \left. + \frac{\gamma\{k, ku / \exp(\hat{y}_i)\} \gamma''\{k, ku / \exp(\hat{y}_i)\} - \gamma'\{k, ku / \exp(\hat{y}_i)\}^2}{\gamma\{k, ku / \exp(\hat{y}_i)\}^2} \right\} \\
 &\quad + g_i^{\text{trGamma}},
 \end{aligned}$$

where $\gamma'\{k, ku / \exp(\hat{y}_i)\}$ and $\gamma''\{k, ku / \exp(\hat{y}_i)\}$ are the first and second derivatives of $\gamma\{k, ku / \exp(\hat{y}_i)\}$ with respect to \hat{y}_i , and have closed forms (see Supplement S3.7).

To model only the excesses above a threshold u , we can use the GPD in (3.1), and assume that $\xi > 0$, since burned areas tend to be heavy-tailed. If we reparameterize and model the logarithmic $\kappa \in [0, 1]$ quantile of the excesses with the boosting estimate, i.e., $\hat{y}_i = \log[(1 - \kappa)^{-\xi} - 1] \sigma_i / \xi$, we can define the generalized Pareto (GPD) loss and obtain its derivatives

$$\begin{aligned}
 \mathcal{L}_{\text{GPD}}(y_i, \hat{y}_i) &= \frac{\xi + 1}{\xi} \log \left\{ 1 + \frac{y_i \{(1 - \kappa)^{-\xi} - 1\} \sigma_i}{\exp(\hat{y}_i)} \right\}, \\
 g_i^{\text{GPD}} &= \frac{-f'\{y_i, \exp(\hat{y}_i), \xi\} \exp(\hat{y}_i)}{f\{y_i, \exp(\hat{y}_i), \xi\}}, \\
 h_i^{\text{GPD}} &= -\exp(\hat{y}_i)^2 \frac{f\{y_i, \exp(\hat{y}_i), \xi\} f''\{y_i, \exp(\hat{y}_i), \xi\} - f'\{y_i, \exp(\hat{y}_i), \xi\}^2}{f\{y_i, \exp(\hat{y}_i), \xi\}^2} \\
 &\quad - \exp(\hat{y}_i) [f'\{y_i, \exp(\hat{y}_i), \xi\} / f\{y_i, \exp(\hat{y}_i), \xi\}],
 \end{aligned}$$

where f' and f'' are the first and second derivatives of the reparameterised probability density function $f\{y_i, \exp(\hat{y}_i), \xi\}$ given in Supplement S3.7.

For wildfire size classification

Adopting the mixture modelling approach to wildfire sizes requires us to model the probability that a fire belongs to each size component $1, \dots, C$, where C is the number of components. Let $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,C})$ denote the vector of wildfire size component indicators, where $y_{i,c} = 1$ if the i -th fire size is in component c , and otherwise $y_{i,c} = 0$. We can model the probability of each class with the boosting estimate using the softmax function $\sigma : \mathbb{R}^C \rightarrow [0, 1]^C$ defined by

$$\sigma(\mathbf{z})_i = \frac{\exp(z_i)}{\sum_{j=1}^C \exp(z_j)}, \quad i = 1, \dots, C, \quad \mathbf{z} = (z_1, \dots, z_C) \in \mathbb{R}^C. \quad (3.11)$$

The generalization of the logistic (Cox, 1958) loss to multiple classes is the cross-entropy loss, which can be reweighted to give

$$\mathcal{L}_{\text{CE}}(\mathbf{y}_i, \hat{\mathbf{y}}_i) = -w_i \sum_{c=1}^C y_{i,c} \log \left\{ \exp(\hat{y}_{i,c}) / \sum_{d=1}^C \exp(\hat{y}_{i,d}) \right\}, \quad (3.12)$$

where the vector of component probabilities is modelled by applying (3.11) to the boosting estimate $\hat{\mathbf{y}}_i = (\hat{y}_{i,1}, \dots, \hat{y}_{i,C})^T$, and the weights w_1, \dots, w_n could be chosen to improve predictions in unbalanced classification tasks.

3.4.4 A spatiotemporal cross-validation scheme

The use of k -fold cross-validation generally presupposes independent replicates, so it would produce optimistic predictive performance estimates in our setting because data points that are geographically closer will have stronger dependencies. To address this, we first study the spatiotemporal process leading to grid cells being masked, which we call the *masking process*. Figure 3.3 hints at either a common or two inter-correlated spatially dependent latent processes governing the observed masking processes for CNT and BA, which we model with a common latent Gaussian process (Rasmussen and Williams, 2005). We then fit a Bernoulli model to observations arising from the masking process, and simulate observations from the model to generate cross-validation folds.

We consider only the months m with masked observations, and let M denote the number of those months and D denote the number of grid cells. Let $R_{d,m}^{\text{CNT}}$ and $R_{d,m}^{\text{BA}}$ denote the binary 0-1 observations indicating whether the grid cell $d \in \{1, \dots, D\}$ (with centroid \mathbf{s}_d) at month $m \in \{1, \dots, M\}$ was masked for the CNT and BA responses, respectively. Our hierarchical model for the masking processes is

$$\begin{aligned} R_{d,m}^{\text{CNT}} | \mu_{dm}^{\text{CNT}} &\sim \text{Bernoulli}\{\text{expit}(\mu_{dm}^{\text{CNT}})\}, \\ R_{d,m}^{\text{BA}} | \mu_{dm}^{\text{BA}} &\sim \text{Bernoulli}\{\text{expit}(\mu_{dm}^{\text{BA}})\}; \end{aligned}$$

$$\begin{aligned} \mu_{dm}^{\text{CNT}} &= \beta_0^{\text{CNT}} + g_m(\mathbf{s}_d) + \epsilon_m^{\text{CNT}}, \\ \mu_{dm}^{\text{BA}} &= \beta_0^{\text{BA}} + \beta g_m(\mathbf{s}_d) + \epsilon_m^{\text{BA}}; \end{aligned}$$

$$\begin{aligned} g_m(\cdot) &\sim \mathcal{GP}(\zeta), \\ \epsilon_m^{\text{CNT}}, \epsilon_m^{\text{BA}} &\sim \mathcal{N}(0, \phi), \\ \beta &\sim \mathcal{N}(0, \omega); \end{aligned}$$

$$\boldsymbol{\theta} = \{\beta_0, \beta, \zeta, \phi, \omega\} \sim \text{Priors},$$

where $\text{expit}(x) = \{1 + \exp(-x)\}^{-1}$ is the inverse logit function.

We fit this model using the integrated nested Laplace approximation (INLA, Rue et al., 2009), which is an approximate Bayesian inference technique well-suited for latent Gaussian models. The parameter β governs the degree of latent sharing between the two masking processes and we use a flat and independent zero-centered Gaussian hyperprior for it. Similar frameworks were used by Koh et al. (2021) for the joint modelling of different wildfire risk components and by Diggle et al. (2010) and Pati et al. (2011) to model preferential sampling. The spatial process g_t is independently replicated in time and each replicate has a Gaussian process prior \mathcal{GP} with a Matérn covariance structure governed by the parameter vector ζ . We represent these Gaussian processes via a numerically convenient Gauss–Markov random field approximation, constructed by solving a stochastic partial differential equation (Lindgren et al., 2011). Supplement §3.7 details the full procedure.

We generate samples from this Bayesian model by first sampling parameters from the posterior distribution, and then generating observations from the Bernoulli model with the sampled parameters. We do this for all months, including in those where observations were masked; if a location was already part of the test set, i.e., if it was already masked, then we removed it from validation set. Five samples were generated to obtain five folds for our cross-validation scheme. Figure 3.4 shows two samples from this model for March 1993. The degree of spatial and inter-variable dependencies resemble those of the masking processes in Figure 3.3, and the numbers of grid cells masked and chosen for validation in each month are also similar. The triplet (2.5% posterior quantile, posterior mean, 97.5% posterior quantile) for the scaling parameter β is (0.28, 0.42, 0.58).

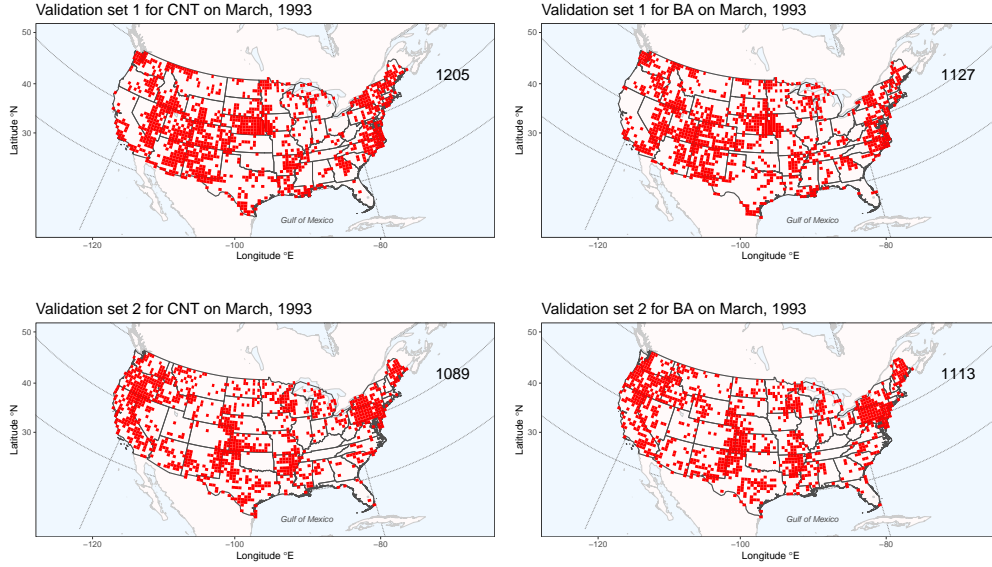


Figure 3.4 – The first (top) and second (bottom) validation folds (in red) for CNT (left) and BA (right) for March, 1993, from our spatiotemporal cross-validation scheme, generated from the Bayesian spatial model. The number at the top right indicates the sum of grid cells chosen.

3.5 Models

3.5.1 Fitting procedure

We fit our gradient tree boosting models with the approach outlined in §3.4.2, minimizing the loss functions described in §3.4.3.

We use the Poisson and dGPD loss functions to fit models on the full CNT distribution. We experimented with different high thresholds u but achieved the best prediction when we modelled the full distribution with the dGPD, i.e. $u = 0$.

For wildfire sizes, we fit models with the squared loss on $\log(1+BA_i)$, $i = 1, \dots, n$. We also consider mixture models which require split modelling of the distribution. For these, we first choose a sufficiently high (95% empirical quantile) threshold at 200ac. We then use the fire sizes above the threshold to fit a model with the GPD loss, and the log-transformed positive sizes below the threshold with the truncated gamma loss. Lastly, we fit a multi-class classifier to the wildfire size component indicators $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,3})$, defined in §3.4.3, using the cross-entropy loss from (3.12); here, $y_{i,1} = 1$ if we observe no fire, $y_{i,2} = 1$ if BA_i is a medium fire (between 0 to 200ac), and $y_{i,3} = 1$ if it is a large fire ($> 200ac$).

Given the covariates at the i -th observation \mathbf{x}_i , we combine the three model components to

get the prediction of the cumulative conditional probability for each observation

$$\begin{aligned}\hat{\Pr}(\text{BA}_i \leq b \mid \mathbf{x}_i) = & \hat{\Pr}(y_{i,1} = 1 \mid \mathbf{x}_i) + \hat{\Pr}(y_{i,2} = 1 \mid \mathbf{x}_i) \hat{\Pr}(\text{BA}_i \leq b \mid \mathbf{x}_i, y_{i,2} = 1) \\ & + \hat{\Pr}(y_{i,3} = 1 \mid \mathbf{x}_i) \hat{\Pr}(\text{BA}_i \leq b \mid \mathbf{x}_i, y_{i,3} = 1), \quad b \geq 0.\end{aligned}$$

We also engineered new covariates to improve our predictions. To incorporate more spatial information from our covariates (other than the longitude and latitude coordinates), we took the average value of the covariate across neighbouring grid cells for each month; this smooths the climatic variables across space. We also allow land-use covariates at neighbouring grid cells to help predictions at each grid cell.

The relationship between CNT and BA is positive but non-linear; a high CNT does not necessarily imply a high BA, though CNT=0 implies that BA=0. Thus, it is natural to consider the other response as a covariate when modelling a given response, or at least to use this information whenever possible. As the test grid cells for BA and CNT are correlated, there are instances where the BA response was masked but the CNT wasn't, and vice versa; for 39% of masked BA observations, their corresponding CNT observations were unmasked. Using the CNT/BA covariate to predict BA/CNT thus raises the question of how best to impute its value if it was masked for a given observation. The default way to handle a missing covariate is to impute its mean across all observations, as in algorithms such as `xgboost` or `gbm`, though this will be sub-optimal for predictions on a spatially heterogeneous dataset. Instead, we use an imputation method which fits a model for the covariate and then imputes the best estimate from this model.

When modelling BA, we first fit a gradient boosting model with the dGPD loss on the CNT response, without using BA as a covariate, and then use the estimated parameters to find the estimated mean CNT for each observation using (3.10); we then impute this estimate whenever CNT was masked. When modelling CNT, we first fit a gradient boosting model, without using CNT as a covariate, with the cross-entropy loss on the wildfire size component indicators \mathbf{y}_i , $i = 1, \dots, n$, and then impute the estimates of the probabilities $\Pr(y_{i,1} = 1)$, $\Pr(y_{i,2} = 1)$ and $\Pr(y_{i,3} = 1)$ from the fitted model, whenever BA was masked; when BA wasn't masked, we use the observed indicators \mathbf{y}_i as a covariate.

To assess models using these covariates, a cross-validation scheme should also reflect this imputation procedure; thus, it becomes even more important to appropriately model the inter-variable dependence between CNT and BA of the masking processes, i.e., the parameter β in our spatiotemporal cross-validation scheme described in §3.4.4.

Our models have hyperparameters that must be tuned by cross-validation. They include the regularization parameters λ and η , the proportion s of covariates subsampled at each iteration, the maximum number of leaves for each tree L , and the number of trees T (see §3.4.2). Other hyperparameters from the loss functions include k , ξ , and α , which govern the shape and tails of the fitted conditional distribution, and weights w_i ($i = 1, \dots, n$) which govern the importance of each observation in the cross entropy loss. Some of our models assume

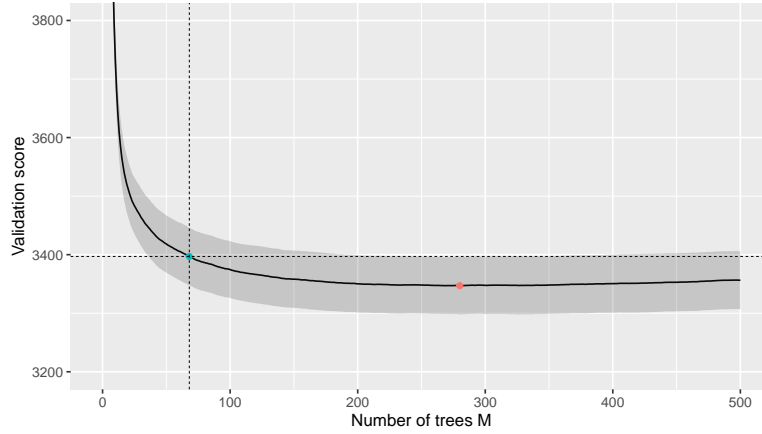


Figure 3.5 – The average rescaled evaluation score across all five folds of our spatiotemporal cross-validation scheme for a mixture model, as a function of the number of trees M . The shaded region shows the pointwise one-standard-error bound. The red point shows the minimum average validation score and the blue point shows the M chosen by the one-standard-error rule.

common shape parameters ξ and α governing the tails of wildfire sizes and counts across the whole sample space; our preliminary analysis suggests that this assumption is reasonable in space, as the frequentist estimates of the shape parameters from pooled data are relatively homogeneous across the wildfire coordination regions.

We use the Bayesian optimization procedure outlined in Snoek et al. (2012) to choose the parameters (excluding the number of trees M) minimizing the average evaluation metric, calculated on the five cross-validation folds generated in §3.4.4. This procedure treats the objective function as random and first places a Gaussian process prior on it. After gathering function evaluations, the prior is updated to form the posterior distribution over the objective function, which is then used to construct an infill sampling criterion. For the mixture model, we implement separate Bayesian optimization procedures for each of the three model components.

Given the other parameters, we choose M with the one-standard-error rule (Hastie et al., 2009); i.e., we select the largest M within one standard error of the parameter that achieves the minimum in terms of the evaluation metric. Figure 3.5 shows the evaluation metric on the validation folds as a function of M for the wildfire size classifier in a mixture model.

3.5.2 Results

The benchmark models are described in Opitz (2021). That for CNT corresponds to a generalized linear model (GLM, Wood, 2017) with a Poisson response distribution and log link linear predictor using all the original covariates. The benchmark for BA first fits a generalized linear model with Gaussian response and log-link using all of the original covariates. The probability

	Model	Ours	5-fold	Truth
Counts	Benchmark*	5172	5235	5565
	Poisson*	3413	3213	3302
	Poisson	3283	3102	3131
	dGPD*	3304	3092	3194
	dGPD	3215	2896	3068
Sizes	Benchmark*	3923	3834	4244
	Log-Normal*	3553	3450	3771
	Log-Normal	3473	3202	3501
	Mixture*	3480	3403	3582
	Mixture	3364	3089	3446

Table 3.1 – The averaged rescaled evaluation score for all models, according to our cross-validation scheme outlined in §3.4.4, the 5-fold cross-validation scheme with random partitioning, and the true score. The bold figures highlight the best model chosen by our cross-validation approach. The asterisks indicate models not using the CNT/BA covariate.

predictions, $\Pr(\text{BA}_i \leq u_{\text{BA}})$, are obtained by combining the log-Gaussian BA model with the estimated probability that $\text{CNT}_i=0$, obtained from the benchmark Poisson model for CNT.

We relied on our cross-validation scheme devised in §3.4.4, using the evaluation metrics used from the competition outlined in §3.3, to choose which model predictions to use for the competition. After the competition, we had access to the truth and could calculate how the predictions of every model would have performed on the test set.

Table 3.1 shows that incorporating the engineered CNT and BA covariates improves the scores of all models by up to 7%. According to our cross-validation scheme, the best model for wildfire sizes is the mixture model, and for the counts it is the dGPD model. The best mixture model and dGPD models from the Bayesian optimization procedure have $\alpha = 52$ and $\xi = 0.8$. This implies a fat tail for the size distribution, but a thinner (Gumbel-like) tail for the wildfire count distribution, though the parameter α in the dGPD loss provides additional flexibility to the model that gives slightly better predictions than the Poisson model. All our gradient boosting models outperform the benchmark by around 10–50%.

Our cross-validation scheme tends to perform better than the 5-fold cross-validation scheme as a proxy for the true test set performance; the scores from the 5-fold cross validation scheme are generally too optimistic compared to the true test error, especially for the wildfire size models. This optimism is especially pronounced when evaluating models that use the engineered CNT and BA covariates. Our scheme is better able to capture the inter-variable dependence between the CNT and BA masking processes, giving a better reflection of how the models that incorporate the engineered covariates would perform when predicting responses on the test set.

Figure 3.6 shows that the gain and coverage metrics introduced in §3.4.2 give similar orderings for the importance of covariates when predicting the probability of being in a given wildfire

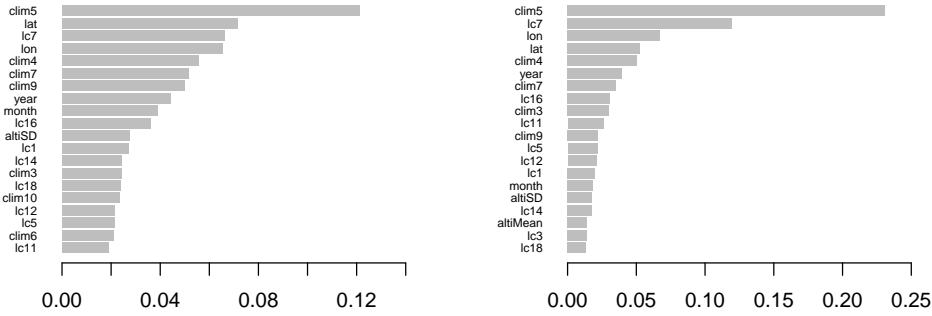


Figure 3.6 – The coverage (left) and gain (right) metrics for the top 20 covariates in the wildfire size classifier submodel of the best mixture model for BA (without using the engineered covariates). More information about the covariates is given in Opitz (2021).

size component with the best mixture model. As hypothesized in §3.3, clim5, lc7 and the spatial covariates of longitude (lon) and latitude (lat) are among the five most important variables for both metrics. The other variables (e.g., clim4, clim7, clim9, year, lc16, etc.) are relevant, but each is less than half as important as clim5, the most important covariate.

To evaluate the marginal effect of clim5 on the CNT response in the best dGPD model, we transformed the partial dependence estimate (3.7) by (3.10) to get the predicted mean count \hat{m}_i , and evaluated and plotted it with clim5 in the set of interest \mathcal{S} in (3.7). As it is computationally infeasible to evaluate all data points $\mathbf{x}_{i\mathcal{C}}$ in our setting with over 500,000 observations, we subsampled 10,000 observations to obtain our estimates.

Figure 3.7 shows that the marginal effect of clim5 on \hat{m}_i tends to be negative, especially above -0.03mwe . Figure 3.8 displays the joint marginal effects of clim5 and land cover covariate 12 (lc12; grassland, in %) on the predicted mean count, i.e., with clim5 and lc12 in the set of interest \mathcal{S} in (3.7). The figure hints at interaction between the two covariates; increasing lc12 tends to decrease the response CNT slightly if clim5 is low, but not when clim5 is high. Although the partial dependence plot is useful for showing the overall marginal trend of a covariate on the response across all observations considered, it is important to be honest about the uncertainty associated with the Monte Carlo estimate in (3.7). The estimates in Figures 3.7 and 3.8 are associated with high uncertainty throughout (not shown for the latter); our dataset is very heterogeneous and it is not possible to quantify the marginal effects of covariates with less uncertainty.

Our chosen models perform competitively when compared to the other teams' submission in the data challenge (Opitz, 2021), placing second out of 28 teams in the final ranking; the other top three teams used other popular prediction techniques such as random forests, hierarchical

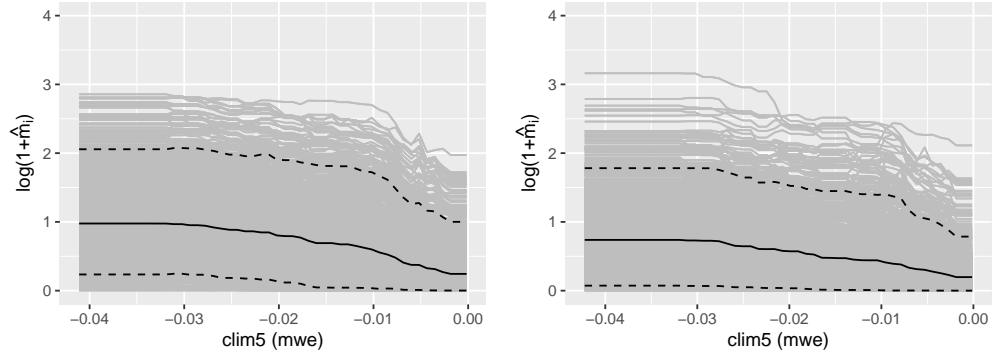


Figure 3.7 – The partial dependence plot with clim5 in the set of interest \mathcal{S} , transformed so the y -axis shows the predicted mean CNT from (3.10), for the observations in the Rocky Mountain Area (left) and Great Basin (right) regions. The empirical 95% and 5% pointwise quantiles from the Monte Carlo estimates are shown by the dotted lines.

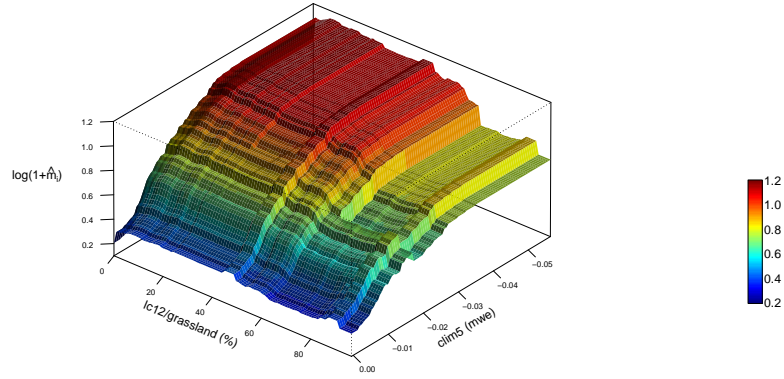


Figure 3.8 – The three-dimensional partial dependence function with clim5 and lc12 in the set of interest \mathcal{S} , transformed as in Figure 3.7 and plotted for all observations.

Bayesian modelling and ANN models with adapted loss functions (Opitz, 2022).

3.6 Discussion

We have implemented novel gradient boosting models for wildfire activity that are trained with loss functions motivated by extreme-value theory. Compared to models trained on the Poisson loss, our chosen model for wildfire counts has an additional parameter α that governs the tail of the count distribution, which, after tuning by cross-validation, enables the model to give better predictions. Our chosen model for burned areas has specific components for extreme fire sizes. According to the given score criteria which puts more weight on large fire sizes, this model improves on the models that do not discriminate between extreme and non-extreme fire sizes.

As the use of other data sources, e.g., covariates not included in the provided dataset, was strictly prohibited for the competition, we were not able to leverage other spatial information, such as the Geographic Area Coordination Center of a grid cell. Each center follows its own governing jurisdictions that could affect its fire mitigation and suppression strategies. Including this information, either by incorporating additional covariates, or by having a separate gradient boosting model for each coordination center, would improve predictions.

Our mixture model has the threshold fixed at $u = 200ac$. However, u could have also been allowed to as an additional parameter chosen by cross-validation (e.g., as in Opitz et al., 2018), though the computational cost would be significantly higher, as the optimization of the model components would need to be done jointly as they all rely on u .

Our best chosen dGPD model from the Bayesian optimization procedure has $\alpha = 52$, which implies a thin tail for the CNT distribution that is not too different from the Poisson distribution. This explains the slight improvement of the model using the dGPD compared to the Poisson loss. Had the tail of the CNT distribution been fatter, as in the case of another data application (e.g., insurance claim counts), we would have also noticed a larger improvement in predictions by the dGPD model.

We have implemented a spatial cross-validation scheme for our context which partly fixes the optimism when using traditional k -fold cross validation to evaluate complex models with engineered covariates over a spatially heterogeneous but dependent dataset. One should always have the real-world prediction scenario in mind when choosing a cross-validation scheme, and we appealed to tools from spatial statistics to aid the validation of model predictions on our test data.

Apart from cross-validation approaches, model comparison using other predictive scores, e.g., Continuous Ranked Probability Scores (CRPS, Matheson and Winkler, 1976) or tail-weighted CRPS (Gneiting and Ranjan, 2011), could be used to compare simulated predictive distributions of burned areas or counts. These scores, along with graphical summaries for validating models, are an important future topic of research, and have already been explored in the extreme wildfire prediction context by Joseph et al. (2019), Pimont et al. (2021) and Koh et al. (2021). Due to the time constraints of the competition however, this is out of the scope of

this chapter.

Our gradient boosting models have hyperparameters from the loss functions that govern the tails of the predictive distributions: ξ , and α , which, once chosen by cross-validation, are fixed in the model. To allow more flexible modelling of the distributional tails, one could incorporate them as an additional boosting estimate in §3.4.2 which would allow these parameters to depend on the covariates. The boosting estimate \hat{y}_i , gradient g_i and hessian h_i in (3.4) would then be vectors. The recent work on extreme quantile regression by Velthoen et al. (2021) has similarities to the described approach.

Apart from better predictions, our models improve decision support in wildfire management. The partial dependence plots in Figures 3.7 and 3.8 allow marginal and interaction effects of covariates to be assessed, though one should be aware of the large uncertainty associated with these estimates. Importance metrics like the gain and coverage in Figure 3.6 could be used for covariate selection and could prompt national wildfire predictive services to rethink designs of fire danger warning systems (e.g., indices) across the contiguous US.

3.7 Supplement

3.7.1 Terms in the gradients and Hessians

For the truncated gamma loss

The lower incomplete gamma function is $\gamma(k, s) = \int_0^s t^{k-1} \exp(-t) dt$, $s > 0$. Set $s = ku / \hat{y}_i$. By applying the fundamental theorem of calculus and the chain rule, taking derivatives with respect to \hat{y}_i gives

$$\begin{aligned} \gamma'\{k, ku / \exp(\hat{y}_i)\} &= \frac{\partial \gamma(k, ku / \hat{y}_i)}{\partial \hat{y}_i} \exp(\hat{y}_i) \\ &= \exp(\hat{y}_i) \left\{ - \left(\frac{ku}{\hat{y}_i} \right)^{k-1} \exp\left(-\frac{ku}{\hat{y}_i}\right) \frac{ku}{\hat{y}_i^2} \right\}, \quad u > 0, \hat{y}_i \in \mathbb{R}, \end{aligned}$$

and

$$\begin{aligned} \gamma''\{k, ku / \exp(\hat{y}_i)\} &= \frac{\partial^2 \gamma(k, ku / \hat{y}_i)}{\partial^2 \hat{y}_i} \exp(\hat{y}_i)^2 + \gamma'\{k, ku / \exp(\hat{y}_i)\} \\ &= \left[\left\{ - \left(\frac{ku}{\hat{y}_i^2} \right)^2 (k-1) \left(\frac{ku}{\hat{y}_i} \right)^{k-2} - 2 \left(\frac{ku}{\hat{y}_i} \right)^{k-1} \frac{ku}{\hat{y}_i^3} \right\} \exp\left(-\frac{ku}{\hat{y}_i}\right) \right. \\ &\quad \left. - \left(\frac{ku}{\hat{y}_i} \right)^{k-1} \left(\frac{ku}{\hat{y}_i^2} \right)^2 \exp\left(-\frac{ku}{\hat{y}_i}\right) \right] \exp(\hat{y}_i)^2 + \gamma'\{k, ku / \exp(\hat{y}_i)\}. \end{aligned}$$

For the GPD loss

Let $\kappa \in (0, 1)$. We first reparameterise the GPD probability density function as

$$f\{y_i, \exp(\hat{y}_i), \xi\} = \left\{ 1 + \frac{y_i\{(1-\kappa)^{-\xi} - 1\}\sigma_i}{\exp(\hat{y}_i)} \right\}^{-(\xi+1)/\xi}, \quad \xi > 0, \hat{y}_i \in \mathbb{R}. \quad (3.13)$$

The functions $f'\{y_i, \exp(\hat{y}_i), \xi\}$ and $f''\{y_i, \exp(\hat{y}_i), \xi\}$ are obtained by differentiating (3.13) with respect to \hat{y}_i . Write $A = y_i\{(1-\kappa)^{-\xi} - 1\}\sigma_i / \exp(\hat{y}_i)$, and notice that $\partial A / \partial \hat{y}_i = A' = -A$. Then

$$f'\{y_i, \exp(\hat{y}_i), \xi\} = A \frac{\xi + 1}{\xi} (1 + A)^{-(2\xi+1)/\xi},$$

and so

$$f''\{y_i, \exp(\hat{y}_i), \xi\} = -A \frac{(\xi + 1)(2\xi + 1)}{\xi} (1 + A)^{-(3\xi+1)/\xi} - f'\{y_i, \exp(\hat{y}_i), \xi\}.$$

3.7.2 Priors and SPDE triangulation

This section details the prior and SPDE triangulation specifications of the spatiotemporal Bernoulli model for the masking processes. Fuller details of INLA and the SPDE approach are outlined in §4.10 and §4.11.

The fixed effect coefficient β_0^{CNT} and β_0^{BA} in our model was assigned a flat Gaussian prior with zero mean and precision 0.001. The prior for the scaling parameter β is a zero-centered Gaussian distribution with precision $\omega = 1/20$. Lastly, we assigned a log-gamma hyperprior with mean unity and precision 0.0005 to the variance hyperparameter ϕ .

To have tractable conditional distributions of spatial Gaussian random effects g_m , we follow the approach described in §4.4.5.

The discretization points when triangulating the spatial domain in (4.9) are chosen as the nodes of a finite element representation which enables efficient inference for random effects representing spatial variation. Our spatial triangulation mesh in Figure 3.9 has 508 nodes. It is sparser in the extended zone around the study area to ensure that the SPDE boundary conditions have negligible influence on the study area.

For computational reasons, we only use observations of the masking processes from the first ten months, and subsample 30% of the data when fitting our model.

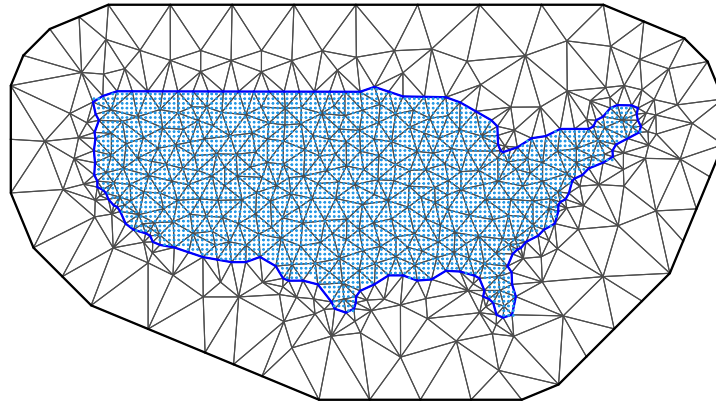


Figure 3.9 – Triangulation mesh of the spatial study region (blue contours) for the SPDE approach. Neumann boundary conditions are set on the exterior (black) boundary to obtain a unique solution. The blue points represent the centroids of the grid cells. The finite element solution defines a Gauss–Markov random vector with one variable in each node.

4 Spatiotemporal wildfire modeling through point processes with extreme marks

This chapter is based on a preprint of a paper that has been submitted to a peer-reviewed journal. It is jointly written with Jean-Luc Dupuy, François Pimont and Thomas Opitz, and is available as Koh et al. (2021). The chapter is structured to be self-contained and may overlap with other chapters. The contributions of the doctoral candidate were in producing the results, simulations, plots and tables, and writing the bulk of the paper.

4.1 Abstract

Accurate spatiotemporal modeling of conditions leading to moderate and large wildfires provides better understanding of mechanisms driving fire-prone ecosystems and improves risk management. Here we develop a joint model for the occurrence intensity and the wildfire size distribution by combining extreme-value theory and point processes within a novel Bayesian hierarchical model, and use it to study daily summer wildfire data for the French Mediterranean basin during 1995–2018. The occurrence component models wildfire ignitions as a spatiotemporal log-Gaussian Cox process. Burnt areas are numerical marks attached to points and are considered as extreme if they exceed a high threshold. The size component is a two-component mixture varying in space and time that jointly models moderate and extreme fires. We capture non-linear influence of covariates (Fire Weather Index, forest cover) through component-specific smooth functions, which may vary with season. We propose estimating shared random effects between model components to reveal and interpret common drivers of different aspects of wildfire activity. This increases parsimony and reduces estimation uncertainty, giving better predictions. Specific stratified subsampling of zero counts is implemented to cope with large observation vectors. We compare and validate models through predictive scores and visual diagnostics. Our methodology provides a holistic approach to explaining and predicting the drivers of wildfire activity and associated uncertainties.

Keywords: Bayesian hierarchical model; Cox process; Extreme-value theory; Forest fires; Shared random effects.

4.2 Introduction

Wildfires are defined as uncontrolled fires of combustible natural vegetation such as trees in a forest. Their activity usually shows seasonal cycles, as several conditions must coincide for their occurrence: the presence of combustible material as fuel, its easy flammability resulting from weather conditions such as droughts, and a trigger. Triggers include natural causes such as lightning, but the majority of occurrences in Europe are caused by human activity, either intentional (arson), neglectful (cigarette stubs) or accidental (agriculture).

Wildfires represent major environmental and ecological risks worldwide. They provoke many human casualties and substantial economic costs, and can trigger extreme air pollution episodes and important losses of biomass and biodiversity. While climate change is expected to exacerbate their frequency and extent (Jones et al., 2020), wildfires themselves contribute an important fraction of global greenhouse gases that can accelerate climate change. To aid in wildfire prevention and risk mitigation, one must identify the factors contributing to wildfires and predict their spatiotemporal distribution. Prediction maps of various components of wildfire risk are relevant for the study of historical periods, for short-term forecasting and for long-term projections.

The study of wildfire activity has led to a large body of statistical and machine learning literature on methods for identifying risk factors and producing risk maps (Preisler et al., 2004; Xi et al., 2019; Pereira and Turkman, 2019). Most studies focus on modeling either occurrence counts or sizes, the latter usually represented by the burnt areas of spatially and temporally contiguous wildfire events. In occurrence modeling, the spatial or spatiotemporal pattern of ignition points (or other representative points of separate wildfire events) can be analyzed with point process tools (Peng et al., 2005; Genton et al., 2006; Xu and Schoenberg, 2011; Serra et al., 2013; Tonini et al., 2017; Pereira and Turkman, 2019; Opitz et al., 2020b). Often, data are available as presence/absence or counts over dense spatial or spatiotemporal grids, or have been transformed to such representations to facilitate modeling and to harmonize different spatial-temporal scales of wildfire and predictor data such as weather conditions, land cover and land use.

Burnt area, a key measure of wildfire impact, usually provides a good proxy for biomass loss and greenhouse gas emissions. Many univariate probability distributions have been explored for modeling fire sizes (*e.g.*, Cumming, 2001; Schoenberg et al., 2003; Cui and Perera, 2008; Pereira and Turkman, 2019). Empirical distributions are usually heavy-tailed, as with the wildfire data we consider in Mediterranean France, and a few extreme wildfires account for a very large fraction of total burnt area. There is no consensus on which distribution provides the best fit (Pereira and Turkman, 2019). Distributions suggested by extreme-value theory, such as the generalized Pareto distribution (GPD), have been studied (*e.g.*, approaches by De Zea Bermudez et al., 2009; Mendes et al., 2010; Turkman et al., 2010; Pereira and Turkman, 2019).

Joint statistical analyses of wildfire occurrence and sizes have been proposed and often use

tools for marked point processes, where numerical marks represent burnt areas. Descriptive approaches (*e.g.*, Tonini et al., 2017) characterize different regimes of wildfire activity (*i.e.*, numbers, sizes, spatiotemporal autocorrelation) by taking into account weather, land cover, fire management and environmental factors. For explanatory and predictive modeling, Bayesian hierarchical models are useful; they can include latent Gaussian components to allow for observation and estimation uncertainty and capture nonlinear influences of covariates. One may consider only categorical information (*e.g.*, small and large wildfires) without attempting to model the continuous distribution of sizes; for example, Serra et al. (2014) construct a Bayesian spatiotemporal “hurdle” model for large wildfires. As to continuous distributions, Ríos-Pena et al. (2018) implement MCMC inference for zero-inflated Beta-regression to model the occurrence of wildfires in spatial units, with absence corresponding to zero-inflation, while the positive area fraction covered by wildfires is captured through the Beta distribution. Joseph et al. (2019) estimate separate regression models with random effects for occurrence numbers in areal units and for sizes, and study posterior predictive distributions for block maxima of wildfire sizes. Pimont et al. (2021) developed a marked spatiotemporal log-Gaussian Cox process model, called Firelihood, for daily data by applying the integrated nested Laplace approximation (INLA, Rue et al., 2009) for Bayesian inference on most components of the model. Their distribution of wildfire sizes over positive values is based on estimating exceedance probabilities and excess distributions over a range of severity thresholds. Weather information is included through a nonlinear effect of the Fire Weather Index (FWI, van Wagner, 1977), constructed to yield high correlation with wildfire activity.

In this work, we develop the following novelties to address key shortcomings of the works cited above. As large wildfires play a dominant and critical role for fire activity due to the heavy tails of burnt areas, we focus on accurate modeling of their distribution, and in particular its spatiotemporal variation. However, models constructed using only extreme wildfires would lead to high estimation uncertainty when inferring complex spatiotemporal structures. We therefore propose the joint estimation of extreme and non-extreme wildfires where the model borrows strength from the latter to help estimate the former; the large number of observations available for moderate fires improves the prediction of larger fires, so changes in extreme fire activity are better accounted for.

Complex models such as Firelihood require separate estimation of the occurrence and size model components, thus hampering inferences that exploit interactions between them. Temporal stochastic structures are often restricted to the spatiotemporal variability in covariates. In Pimont et al. (2021), simulated predictive distributions of wildfire activity for various divisions of the space-time domain failed to capture some very extreme events, specifically the year 2003. Here we increase the flexibility of the spatiotemporal structure, especially for extremes.

Our new approach leverages a combination of marked point processes defined over continuous space and time and extreme-value theory to represent the mechanisms leading to wildfires exceeding a high severity threshold for burnt areas. The point pattern of extreme fires

Chapter 4. Spatiotemporal wildfire modeling through point processes with extreme marks

is viewed as a thinning of the full pattern, and we select a suitable threshold before using the GPD model for threshold excesses.

We also advocate sharing spatial random effects that affect several model components simultaneously: these effects are estimated for one response variable (*e.g.*, wildfire counts) but we also include them with scaling coefficients in other response variables (*e.g.*, wildfire size exceedances). This approach decreases uncertainty in the estimation of those regression equations whose vector of observed responses carries too little information to estimate complex predictive structures. We will highlight the improved inferences through sharing in our wildfire application. Besides increasing model parsimony, sharing also provides new scientific insight by highlighting joint drivers of different wildfire components.

The FWI quantifies the influence of weather drivers on wildfire activity and is often mapped as an index for fire danger, for instance by the French weather service Météo France. Model diagnostics of Pimont et al. (2021) showed that the predictive power of FWI in France may diminish in some seasons, and we estimate a more sophisticated seasonal nonlinear FWI effect.

Predictive model validation is intricate because of heavy tails and high prediction uncertainty for individual wildfires. Customary validation scores, such as means of squared or absolute errors, are not useful. In addition to visual diagnostics, we tackle this difficulty through joint assessment of several numerical criteria, either through scores for binary data (*e.g.*, Area under the Curve, Fawcett, 2006) to assess the exceedance behavior over a relevant severity threshold, or through comparison of probabilistic scores for continuous predictions, such as the scaled Continuous Ranked Probability Score (Bolin and Wallin, 2020).

We estimate our marked log-Gaussian Cox process in a Bayesian setting using INLA (Illian et al., 2012) by adopting penalized complexity (PC) priors for hyperparameters (Simpson et al., 2017). Gaussian process priors follow the Matérn covariance function, and we use the Stochastic Partial Differential Equation (SPDE) approach of Lindgren et al. (2011) for numerically efficient Gauss–Markov approximation. Fully Bayesian inference is out of reach, as we have several million observations of wildfire counts, so we devise a specific subsampling scheme for zero counts that keeps a relatively larger proportion of observations with high FWI, for which most wildfires occur. This allows joint Bayesian inference on all components, and we ensure that our subsample sizes allow the fitting of models on standard personal computers, in contrast to other recent approaches (*e.g.*, Joseph et al., 2019; Pimont et al., 2021; Opitz et al., 2020b) requiring high computer memory.

In the remainder of the chapter, we first explore the available data on wildfires and predictors in §4.3. We provide general background on extreme-value theory and point processes and on how to combine them in a Bayesian hierarchical model using the INLA-SPDE method in §4.4. The specific hierarchical structure for the joint analysis of extreme and non-extreme wildfires is developed in §4.5. Estimation with subsampling of pixel-days without wildfire occurrences is detailed in §4.4.3. After a comparative analysis of models in §4.6.1, we highlight key findings

and prediction of wildfire activity components in §4.6.3 and §4.7, and we conclude in §4.8.

4.3 Wildfire data

Since 1973, wildfires occurring in the fire-prone French Mediterranean region have been recorded in the Prométhée database (www.promethee.com). Each wildfire occurrence is reported with its fire ignition cell in a $2 \times 2\text{km}^2$ grid, day of detection and burnt area in hectare (ha). Inconsistent reporting was found for small wildfires, especially smaller than 1 ha, and we keep only data with reported burnt area larger than 1 ha; *i.e.*, of escaped wildfires that could not quickly be extinguished. We use the observation period 1995–2018, for which gridded weather reanalysis data (SAFRAN model of Météo France) and information on forested area are available.

Figure 4.1 illustrates the heavy tails in the distribution of burnt areas and strong spatial variability in numbers and sizes of wildfires. It also shows the contours of administrative areas (“départements”) in the study region. Small to moderately large wildfires dominate the pie charts for wildfire counts, while large wildfires dominate the those for aggregated burnt area. Certain spatial patterns are similar in the distribution of numbers and sizes of wildfires (top and bottom display of Figure 4.1, respectively), but there are notable differences. For example, large wildfire numbers do not always entail large aggregated burnt areas, as we see for the Pyrénées-Orientales département in the southwest. The disparities show the need to model spatiotemporal structures in both wildfire numbers and sizes, as well as in their interaction. Figure 4.2 (left panel) shows a histogram of burnt area values. The sum of burnt areas exceeding the empirical 99%-quantile is larger than the corresponding sum for the remaining wildfires.

The SAFRAN model provides gridded weather reanalyses at 8km resolution. The joint influence of weather variables such as temperature, precipitation and wind speed on fire activity patterns is highly complex. Meteorological indices of fire danger have been constructed, such as the widely used unitless Fire Weather Index (FWI) that was originally defined for Canadian forests. Its values are often used for direct interpretation and fire danger mapping. Instead, we here study its relationship to components of fire risk, such as occurrence frequency and wildfire sizes. For our models, we preprocess SAFRAN data to daily FWI and use the SAFRAN grid by aggregating daily wildfire counts to its cells; Pimont et al. (2021) provide arguments to use this spatial-temporal resolution. Forest cover is another crucial explanatory variable. Around 60% of the study area has forested areas or vegetation types that ignite easily (shrubland; other natural herbaceous vegetation). Wildfires do not propagate easily through the other available land cover types. We consider relevant fuel material through the proportion covered by this vegetation in each SAFRAN grid cell (and day) based on CORINE Land Cover data (CLC). CLC dynamics are captured by linear temporal interpolation of several inventories. We refer to the resulting pixel-day predictor as forested area (FA), in %.

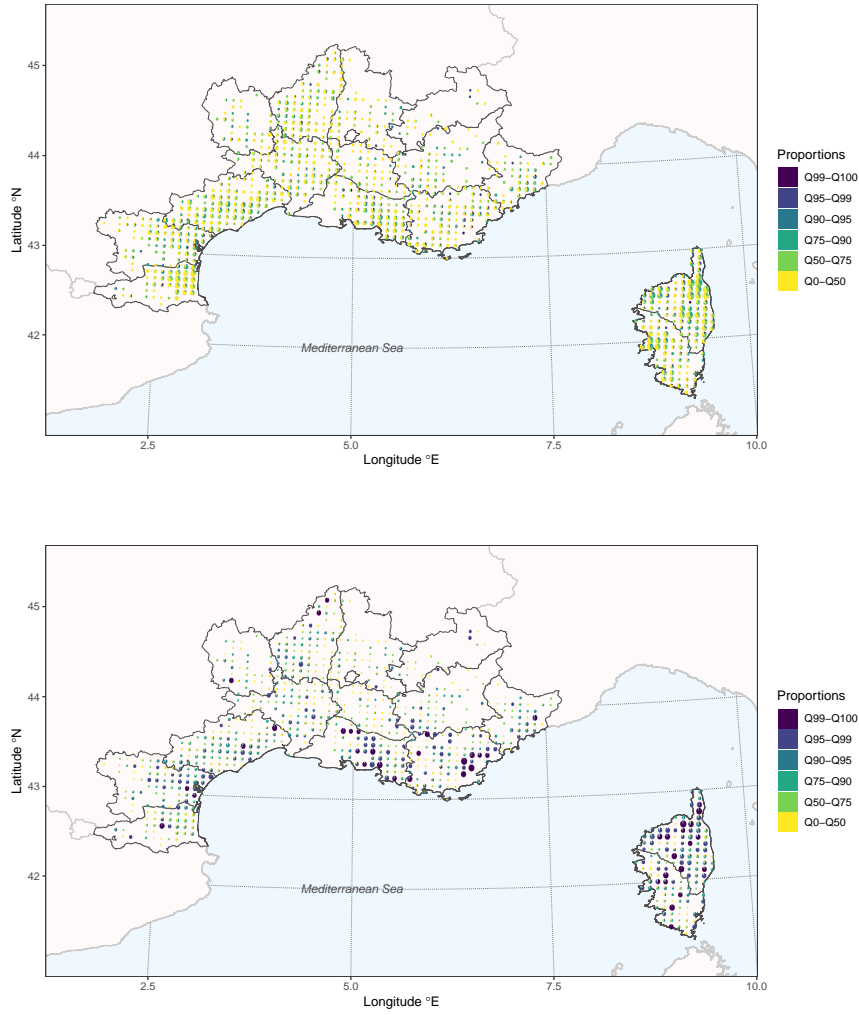


Figure 4.1 – Maps of Prométhée data aggregated to the SAFRAN grid at 8km resolution. The pie charts in the grid cells are based on 6 wildfire size classes with boundaries given by empirical quantile levels 0, 0.5, 0.75, 0.9, 0.95, 0.99, 1 of all burnt areas (June–October). Top display: pie charts show relative count proportions over the six classes and have size increasing with increasing counts. Bottom display: pie charts show relative burnt area proportions and have size increasing with increasing aggregated burnt area.

4.4 Methods for point patterns with extreme marks

4.4.1 Extreme-value theory

Given a random variable X whose distribution F satisfies mild regularity conditions, the generalized Pareto distribution (GPD) arises asymptotically for the positive excesses of X above a threshold increasing to $x^* = \sup\{x : F(x) < 1\}$ (Coles, 2001). Therefore, given a large threshold $u < x^*$, the tail behavior of a wide class of random variables X can be approximated

as

$$\Pr(X > x + u \mid X > u) \approx \text{GPD}_{\sigma, \xi}(x) = \begin{cases} (1 + \xi x / \sigma)_+^{-1/\xi} & \xi \neq 0, \\ \exp(-x/\sigma) & \xi = 0, \end{cases} \quad x > 0, \quad (4.1)$$

with shape parameter $\xi \in \mathbb{R}$ and scale parameter $\sigma = \sigma(u) > 0$, where $a_+ = \max(a, 0)$. The shape parameter determines the rate of tail decay, with slow power-law decay for $\xi > 0$, exponential decay for $\xi = 0$, and polynomial decay towards a finite upper bound for $\xi < 0$. Writing $p_{\text{exc}} = 1 - F(u)$ for the exceedance probability of X above u , we use (4.1) to approximate the cumulative distribution function F of X above the threshold u (Davison and Smith, 1990) as

$$F(x) \approx 1 - p_{\text{exc}} \text{GPD}_{\sigma, \xi}(x - u), \quad x > u, \quad (4.2)$$

where ξ, σ and p_{exc} are parameters to be estimated. We account for dependence and non-stationarity among observations by including auxiliary variables and Gaussian random effects in σ and p_{exc} . Nonstationarity in ξ is often hard to identify, and we therefore keep ξ stationary.

Based on (4.2), we model the conditional GPD of fire size excesses and p_{exc} . To explore the tail behavior of all fire sizes pooled together and choose an appropriate threshold u , we can use mean excess plots (see Supplement §4.9.4) or the following threshold stability plot of parameters, here considered for the GPD shape ξ , estimated by maximum likelihood for thresholds $v_m > \dots > v_1$. We use multiple statistical tests (Northrop and Coleman, 2014) to test the null hypotheses that the data come from a common truncated GPD on all intervals (v_k, v_{k+1}) , $k = 1, \dots, m$, where $v_{m+1} = \infty$. Using $m = 40$ equidistant intervals of length 5ha for fire sizes, Figure 4.2 provides evidence that stability is reached above approximately the 95% quantile (79ha), as we fail to reject the null hypothesis $\xi_k = \dots = \xi_m$ for intervals with $v_k > 79\text{ha}$ and estimated shape $\hat{\xi}_k \approx 0.7$.

Joseph et al. (2019) modeled fire sizes in the contiguous United States and concluded that the GPD leads to overestimation of extreme fire sizes. However, they fitted the GPD to the full distribution; Figure 4.2 shows that we would have obtained $\hat{\xi} \approx 1.4$ for $u = 1$, corresponding to extremely slow tail decay.

4.4.2 Mark-dependent thinning of point processes

We consider the point pattern of fire ignitions and burnt areas as a realization of a spatiotemporal marked point process; *i.e.*, of a random count measure N that attributes values $N(B) \in \{0, 1, 2, \dots\}$ to Borel sets $B \subset \mathcal{D} \subset \mathbb{R}^2 \times \mathbb{R}$. We model the intensity function $\lambda(x)$ of the point process in the observation window \mathcal{D} , defining the expected number of points

$$\Lambda(B) = \mathbb{E}N(B) = \mathbb{E} \sum_{i=1}^N 1(x_i \in B) = \int_B \lambda(x) \, dx.$$

We focus on Poisson point processes characterized by the counts $N(B) \sim \text{Pois}\{\Lambda(B)\}$. With two types of points, such as non-extreme and extreme points, the point pattern is a superposition

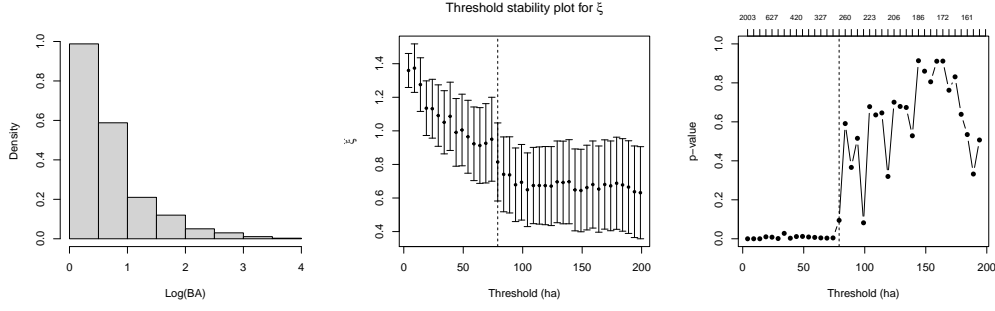


Figure 4.2 – Burnt area distribution. Left: Histogram of burnt areas (ha) in base-10-logarithm. Middle: Parameter stability of the tail index. Right: p-values for the null hypothesis of a GPD distribution above the threshold; tick labels on top indicate the number of fires above the thresholds.

of the two single-type patterns, and the intensity is a sum $\lambda = \lambda_1 + \lambda_2$. The points of a specific type, say type 2, are generated by thinning the full point pattern; *i.e.*, by removing the points of other types (here type 1) using the thinning probability $p(x) = \lambda_2(x)/\lambda(x)$, $x \in \mathcal{D}$. Extreme events, characterized as points x_i whose magnitude mark y_i exceeds a fixed high value $u(x_i)$ are obtained by thinning the full point pattern. Given a point pattern $\{x_1, \dots, x_N\}$, $N \geq 1$, we define variables $E_i = \mathbb{I}\{y_i > u(x_i)\} \sim \text{Bernoulli}\{p(x_i)\}$. An independently thinned Poisson process (*i.e.*, E_i are independent) is again a Poisson process.

4.4.3 Spatiotemporal log-Gaussian Cox processes

Log-Gaussian Cox processes (LGCPs) are Poisson processes with log-Gaussian intensity function $\lambda(x)$. This random specification of the intensity function can explain spatiotemporal variability not captured by deterministic parameters, and provides a natural framework for the Bayesian modeling of point processes with Gaussian process priors. Two major challenges arise for likelihood-based inference in LGCPs: (i) intensity functions are conceptually defined over continuous space; (ii) the Gaussian random effects lead to an intractable likelihood with no general closed-form expression. Challenge (ii) requires estimation techniques to handle latent variables; see §4.4.5. As to (i), without considering the marks, LGCPs have no general closed-form expression for their probability densities

$$(x_1, \dots, x_N) \mapsto \mathbb{E}_\lambda \exp \left(- \int_{\mathcal{D}} \lambda(x) dx \right) \prod_{i=1}^N \lambda(x_i). \quad (4.3)$$

Different approximation strategies allow numerical computation of the integral $\int_{\mathcal{D}} \lambda(x) dx$ for a given intensity function. We discretize the observation window using the SAFRAN grid, and assume that the intensity function does not vary within pixel-day grid cells. Conditional on λ , the number of points observed in a cell C_k , $k = 1, \dots, K$, is Poisson distributed, so estimating

the LGCP corresponds to performing a (mixed) Poisson regression with log-link;

$$N_k | \lambda_k \stackrel{\text{ind}}{\sim} \text{Pois}(|C_k| \lambda_k), \quad \log(\lambda_k) = \mu_k, \quad k = 1, \dots, K, \quad (4.4)$$

where λ_k is the constant intensity in cell C_k , $|C_k|$ is the Lebesgue volume of C_k , $\bigcup_k^K C_k = \mathcal{D}$ and $C_{k_1} \cap C_{k_2} = \emptyset$ if $k_1 \neq k_2$. The linear predictor μ_k is a sum of fixed and random effects. Likelihood-based inference for latent Gaussian processes is often based on Laplace approximation (Tierney and Kadane, 1986). In particular, the INLA framework assumes conditional independence of the observations given the latent Gaussian predictor and is thus well suited for LGCPs, where the Poisson observations N_k are conditionally independent given μ_k (Illian et al., 2012; Opitz et al., 2020a). Other approaches for numerically approximating the integral in (4.3) exist and typically use appropriately weighted sums $\sum_k \omega_k \lambda(\tilde{x}_k)$ with discretization points \tilde{x}_k and weights $\omega_k > 0$, leading to variants of Poisson and logistic regression (e.g., the Berman–Turner 1992 device); see Baddeley et al. (2010).

4.4.4 Data aggregation and subsampling schemes

Spatiotemporal hierarchical modeling is notoriously computer-intensive due to large datasets and numerical challenges with covariances. The R-INLA implementation (Rue et al., 2017) can handle up to several hundred thousand observations. Stable inferences may require compromises with respect to the complexity of the latent model and the number of observations, which jointly determine the size and sparsity of the Gaussian precision matrices, which influence computation times, memory requirements and well-conditioned numerical behavior. Even stronger restrictions arise with the use of methods such as Markov Chain Monte Carlo (MCMC) to achieve approximation quality comparable to INLA (Taylor and Diggle, 2014; van Niekerk et al., 2019). Krainski et al. (2018, §8.4) develop strategies for LGCPs by aggregating the events to larger mapping units and lowering spatial-temporal resolution of random effects to decrease computation times, though this impedes the modeling of structures with small spatiotemporal scales.

Another way to cope with this issue of having too many observations is subsampling (Baddeley and Turner, 2000; Rathbun et al., 2007; Baddeley et al., 2010; Rathbun, 2013; Baddeley et al., 2014), where the model is estimated using an appropriately reweighted subsample of data points, which keeps the loss of information small. Since maximum likelihood estimation is equivalent to maximizing the empirical expectation of the log-density of observations, a subsampling scheme is appropriate provided it ensures a faithful approximation of this expectation. Subsampling in likelihood-based estimation can be interpreted as importance sampling (Tokdar and Kass, 2010): the original sample with observation weight unity is replaced by a subsample with larger observation weights. Weighted subsampling theory goes back to Horvitz and Thompson (1952).

The Poisson intensities $\lambda_k = \exp(\mu_k)$ ($k = 1, \dots, K$) in (4.4) are the parameters to be estimated, and we need a subsample N_{k_j} with weights ω_j ($j = 1, \dots, J$) such that the subsample likelihood

Chapter 4. Spatiotemporal wildfire modeling through point processes with extreme marks

is close to the full density (4.3). The sample size K exceeds 5 million, as there are over 1000 daily-replicated spatial pixels. To enable R-INLA-based estimation, we devise a stratified subsampling scheme to reduce the number of observations one hundredfold. Observations $N_k > 0$ are not subsampled since they are rare and highly informative; we keep them with unit weights. For the zero wildfire occurrence counts, we link subsampling to Poisson additivity. The likelihood contribution $\exp(-\lambda_k)^{\omega_k} = \exp(-\omega_k \lambda_k)$ with weight $\omega_k \in \mathbb{N}$ is equal to the likelihood of the sum of ω_k observations with count 0; the size of the initial sample is divided by the factor ω_k . The predictors (covariates, random effects), and therefore of intensities λ_k , differ between different pixel-days k in our models, so Poisson additivity cannot be applied without additional approximations. However, the values of such predictors may often be very similar for cells located close in space and time, so we control the loss of information by subsampling that preserves a representative coverage of space and time.

We partition our data by years and pixels and then apply subsampling within each partition. The subsample contains two observations for each year-pixel combination. We thus obtain approximately 50,000 observations in the subsample, in line with the rule of thumb of Baddeley et al. (2014, 2015) that the subsample should be at least a factor four larger than the number of event points. The resulting models can be run on standard desktop computers (16Gb of memory). Within pixel-year combinations, we use non-uniform random sampling to overweight specific parts of the predictor space. For inference on the FWI-month interaction, we set different sampling probabilities for FWI values above and below the empirical FWI-quantile at p_{FWI} for each pixel-year. Values above the threshold are expected to correspond to more fire-prone conditions, and we over-represent them, by fixing sampling probabilities $p_{\text{SS}} = 0.9$ for FWI values below the threshold. To appropriately identify seasonal effects, we choose the month among June–October at random. For instance, high FWI values tend to be less frequent in October, but uniform subsampling of months gives them more weight. With this scheme, we obtain a positive sampling probability $p_k > 0$ for each observation N_k in (4.4), and likelihood weights are $\omega_k = 1/p_k$ for the selected observations. Simulation experiments (see Supplement §4.9.5) motivated taking $(p_{\text{FWI}}, p_{\text{SS}}) = (0.7, 0.9)$.

4.4.5 Fully Bayesian inference using INLA-SPDE

Integrated nested Laplace approximation (INLA Rue et al., 2009; Lindgren and Rue, 2015; Opitz, 2017) is a Bayesian technique for fitting generalized additive models with Gaussian random effects. It uses astutely designed deterministic approximations for accurate posterior inference on model parameters, random effects and predictions conditional on data. INLA enables transfer of information across components, appropriate uncertainty assessment and estimation of shared effects. A brief overview of INLA is given in the Supplement §4.10. We implement penalized complexity priors (PC priors, Simpson et al., 2017) in our models to control the complexity of model components. Such priors penalize the distance of the prior of a model component towards a simpler baseline at a constant rate.

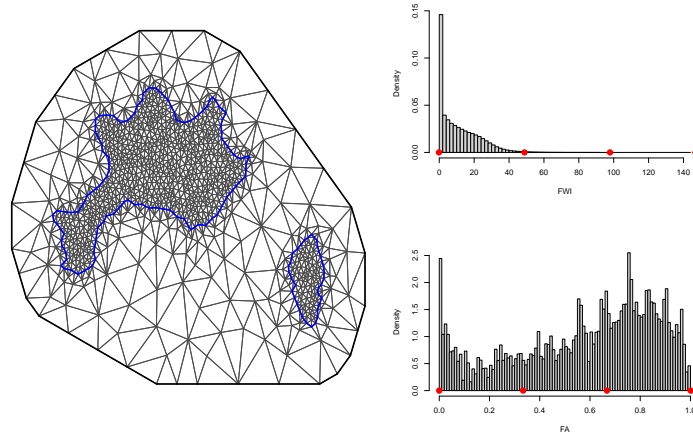


Figure 4.3 – Discretization of random effects with SPDE-based Gaussian prior processes. Left: Triangulation mesh of the study area (blue contours) for the SPDE approach. Neumann boundary conditions are set on the exterior (black) boundary to obtain a unique solution. The finite element solution defines a Gauss–Markov random vector with one variable in each node. Right: Histograms of FWI and FA values. The red points indicate where the spline knots are placed.

Due to the large number of pixels in our problem, spatial Gaussian random effects and their conditional distributions must be tractable in this setting. We use the Matérn covariance function for random effects (denoted g), given as follows for two points s_1 and s_2 :

$$\text{Cov}\{g(s_1), g(s_2)\} = \sigma^2 2^{1-\nu} (\kappa \|s_1 - s_2\|)^\nu K_\nu(\kappa \|s_1 - s_2\|) / \Gamma(\nu), \quad \sigma, \nu > 0,$$

with Euclidean distance $\|\cdot\|$, gamma function Γ , modified Bessel function of the second kind K_ν , and standard deviation and smoothness parameters σ and ν . The empirical range at which the correlation drops to approximately 0.1, is $r = \sqrt{8\nu}/\kappa$. Numerically convenient representations by approximating Gauss–Markov random fields (GMRF, characterized by sparse precision, *i.e.*, inverse covariance, matrices) are constructed by solving a stochastic partial differential equation (SPDE, Lindgren et al., 2011; Krainski et al., 2018), where we fix the smoothness ν at unity. The discretization points of the triangulation in the SPDE approach (full details in the Supplement §4.11) are chosen as the nodes of a finite element representation (*e.g.*, the triangulation of space for $d = 2$, or spline nodes for $d = 1$), which enables efficient inference for random effects representing spatial variation ($d = 2$) or nonlinear functions ($d = 1$ for the FWI and FA effects). Our spatial triangulation mesh in Figure 4.3 has 1114 nodes. It is less dense in the extended zone around the study area to ensure that SPDE boundary conditions have negligible influence on the study area. The four splines knots for FWI and FA are evenly spaced throughout the feature space.

4.5 Point processes with moderate and extreme marks

Point processes govern the space-time point patterns of occurrences; size processes govern the moderate-level and extreme quantitative marks. We write N_{it} for the number of wildfire occurrences on day $t \in \{1, \dots, n\}$ and over the $8\text{km} \times 8\text{km}$ grid cell $i \in \{1, \dots, 1143\}$ with centroid s_i , and $\mathcal{C}_{i,t} \subset \mathcal{D}$ for the space-time cell with volume $|\mathcal{C}_{i,t}| = 64 \text{ (km}^2 \times \text{day)}$. If $N_{it} > 0$, we let $\mathbf{Y}_{it} = (Y_{it,1}, \dots, Y_{it,N_{it}}) \in (1, \infty)^{N_{it}}$ denote the corresponding quantitative marks. We write $z_k(s, t)$ ($k = 1, \dots, K$) for known deterministic covariates.

We model data of escaped fires ($> 1 \text{ ha}$), whose occurrence structure is captured by a regression component COX defining a LGCP. A logistic regression component BIN is used to classify fires into moderate (0) and large (1) according to their exceedance or not above a fixed threshold u , *i.e.*, to provide the thinning of the point pattern and leave only extreme wildfires. Based on Figure 4.2, we consider a fire size $Y_{it,k}$ to be extreme if $Y_{it,k} > 79\text{ha}$ ($k = 1, \dots, N_{it}$); *i.e.*, we take $u = 79$. We write $\mathbf{R}_{it} = (R_{it,1}, \dots, R_{it,N_{it}}) \in \{0, 1\}^{N_{it}}$ for the vector of binary exceedance indicators $R_{it,k} = \mathbb{I}(Y_{it,k} > u)$. Moderate wildfire sizes $Y_{it,k} \in (1, u]$ are modeled through a Beta regression component BETA applied to $(Y_{it,k} - 1)/(u - 1)$. The Beta distribution, usually parametrized by two shape parameters $a, b > 0$, is here parametrized through a precision parameter $\phi = a + b > 0$ and the mean $\mu_{it}^{\text{BETA}} = a/(a + b) \in (0, 1)$ with logit-link function, such that $a = \mu_{it}^{\text{BETA}}\phi$ and $b = \phi(1 - \mu_{it}^{\text{BETA}})$; it is a flexible location-shape family for interval-valued data and can be used with INLA. For large wildfires, we use the extreme-value framework in §4.4.1 and model excesses $Y_{it} - u > 0$ above u through a GPD regression component GPD to characterize extreme wildfires. Following Opitz et al. (2018), we use a log-link function for the median μ_{it}^{GPD} of the GPD.

Some hyperparameters (*e.g.*, precision parameters of priors for fixed effects) are fixed a priori, but those that may heavily influence the posterior model structure are estimated. The priors are fully detailed in the Supplement §4.9.6.

4.5.1 Bayesian hierarchical multi-response regression

Our modeling assumptions in §4.4.3 give the linear COX predictor:

$$\mu_{it}^{\text{COX}} = \log \int_{\mathcal{C}_{i,t}} \lambda(s, t) d(s, t) = \log \lambda(s_i, t) + \log |\mathcal{C}_{i,t}|.$$

We construct the system of regression equations in a Bayesian generalized additive mixed model (GAMM):

$$\begin{aligned} N_{it} | \mu_{it}^{\text{COX}} &\sim \text{Poisson}\{\exp(\mu_{it}^{\text{COX}})\}, \\ R_{it,k} | \mu_{it}^{\text{BIN}} &\sim \text{Bernoulli}\{\text{logit}^{-1}(\mu_{it}^{\text{BIN}})\}, \quad k = 1, \dots, N_{it}, \\ \{Y_{it,k} - u | R_{it,k} = 1, \mu_{it}^{\text{GPD}}\} &\sim \text{GPD}\{\exp(\mu_{it}^{\text{GPD}}), \xi\}, \\ \{(Y_{it,k} - 1)/(u - 1) | R_{it,k} = 0, \mu_{it}^{\text{BETA}}\} &\sim \text{Beta}\{\text{logit}^{-1}(\mu_{it}^{\text{BETA}}), \phi\}; \end{aligned}$$

$$\mu_{it}^{\text{COMP}} = \sum_{k=1}^K g_k^{\text{COMP}} \{z_k(s_i, t); \boldsymbol{\theta}^{\text{COMP}}, \boldsymbol{\theta}^{\text{SHR}}\}, \quad \text{COMP} = \{\text{COX}, \text{BIN}, \text{GPD}, \text{BETA}\};$$

$$\boldsymbol{\theta} = (\xi, \phi, \boldsymbol{\theta}^{\text{COX}}, \boldsymbol{\theta}^{\text{BIN}}, \boldsymbol{\theta}^{\text{GPD}}, \boldsymbol{\theta}^{\text{BETA}}, \boldsymbol{\theta}^{\text{SHR}}) \sim \text{Hyperpriors},$$

where terms g_k^{COMP} capture linear or nonlinear influence of the covariates in the corresponding model component. The specifics of $\boldsymbol{\theta}$ are discussed below.

The intensity function λ_{exc} of the point process of large fires satisfies $\lambda_{\text{exc}}(s_i, t) \leq \lambda(s_i, t)$. The exceedance probability $\text{logit}^{-1} \mu_{it}^{\text{BIN}} = \lambda_{\text{exc}}(s_i, t) / \lambda(s_i, t)$ defines the independent Bernoulli probability of the full point pattern in COX. Since $\lambda_{\text{exc}}(s_i, t) = \exp(\mu_{it}^{\text{BIN}}) \exp(\mu_{it}^{\text{COX}}) / \{1 + \exp(\mu_{it}^{\text{BIN}})\}$ and typically $\exp(\mu_{it}^{\text{BIN}}) \approx 0$, we obtain $\log \lambda_{\text{exc}}(s_i, t) \approx \mu_{it}^{\text{BIN}} + \mu_{it}^{\text{COX}}$.

4.5.2 Sharing latent effects

For maximal flexibility, we could incorporate mutually independent spatial effects into all model components. However, models would become overly complex, with too many spatial effects and hyperparameters to estimate, and with high posterior uncertainties in the spatial effects of the BIN and GPD components due to the small number of large wildfires. We share spatial random effects between model components of the point and size processes, with a preliminary model selection procedure (see §4.6.1) that avoids compromising the quality of model fit and predictions. We assign SPDE-based spatial GMRF priors $g^{\text{COX-BETA}}$, $g^{\text{COX-BIN}}$ and $g^{\text{BIN-GPD}}$ (recall §4.4.5) for the shared spatial effects. We use superscripts to indicate the two components into which we jointly incorporate an effect, and write n to indicate the number of latent random variables for the corresponding effect (in superscript):

$$\begin{aligned} g^{\text{COX-BETA}}(s_i) &\sim \mathcal{GP}_{2\text{D-SPDE}}(\boldsymbol{\omega}_1), & n^{\text{COX-BETA}} &= 1114, \\ g^{\text{COX-BIN}}(s_i) &\sim \mathcal{GP}_{2\text{D-SPDE}}(\boldsymbol{\omega}_2), & n^{\text{COX-BIN}} &= 1114, \\ g^{\text{BIN-GPD}}(s_i) &\sim \mathcal{GP}_{2\text{D-SPDE}}(\boldsymbol{\omega}_3), & n^{\text{BIN-GPD}} &= 1114, \end{aligned}$$

where $\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$ and $\boldsymbol{\omega}_3$ consist of r and σ with PC priors (Fuglstad et al., 2018). Each shared effect is additively included in the linear predictor of the second component and then shared towards the first component with scaling factor $\beta \in \mathbb{R}$, with superscripts to denote the two components. We denote the vector of sharing-related hyperparameters by $\boldsymbol{\theta}^{\text{SHR}} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \boldsymbol{\omega}_3, \beta^{\text{COX-BETA}}, \beta^{\text{COX-BIN}}, \beta^{\text{BIN-GPD}})$, and use flat, independent zero-centered Gaussian hyperpriors for the scaling factors.

Sharing allows modeling of residual spatial effect components that jointly affect multiple model responses, such as landuse at the Wildland-to-Urban interface (Stewart et al., 2007), where human activities intermingle with wildland vegetation. Accurate sharing improves parsimony of the model and borrows estimation strength for random effects across model components by simultaneously using different types of data. Expert knowledge should guide

Chapter 4. Spatiotemporal wildfire modeling through point processes with extreme marks

the choice of spatial effects to be shared between specific components; shared coefficients that differ from zero provide novel insight into the interplay of spatial structures.

4.5.3 Prior structure of linear predictors

We let $z_{\text{FWI}}(s_i, t)$ and $z_{\text{FA}}(s_i, t)$ denote the average FWI and FA on day t in grid cell i , and by $a(t)$ and $m(t)$ the corresponding year and month of day t . Writing α for the intercept and g for the other GAMM components, the prior structure of the model component COX for escaped fire occurrences is

$$\begin{aligned}\mu_{it}^{\text{COX}} = & \alpha^{\text{COX}} + g_1^{\text{COX}}(s_i) + \beta^{\text{COX-BETA}} g^{\text{COX-BETA}}(s_i) + \beta^{\text{COX-BIN}} g^{\text{COX-BIN}}(s_i) \\ & + g_2^{\text{COX}}\{z_{\text{FA}}(s_i, t)\} + g_3^{\text{COX}}\{z_{\text{FWI}}(s_i, t); m(t)\} \\ & + g_4^{\text{COX}}\{a(t)\} + g_5^{\text{COX}}\{m(t)\};\end{aligned}$$

$$\begin{aligned}g_1^{\text{COX}}(s_i) & \stackrel{\text{iid}}{\sim} \mathcal{N}\{0, 1/\tau_1\}, & n_1^{\text{COX}} &= 1143, \\ g_2^{\text{COX}}(\cdot) & \sim \mathcal{GP}_{\text{1D-SPDE}}(\phi_1), & n_2^{\text{COX}} &= 4, \\ g_3^{\text{COX}}(\cdot; m) & \sim \mathcal{GP}_{\text{1D-SPDE}}(\phi_2), \\ g_3^{\text{COX}}(z_{\text{FWI}}; \cdot) & \sim \mathcal{GP}_{\text{RW1}}(1/\tau_2), & n_3^{\text{COX}} &= 4 \times 5 = 20, \\ g_4^{\text{COX}}(\cdot) & \sim \mathcal{GP}_{\text{RW1}}(1/\tau_3), & n_4^{\text{COX}} &= 20, \\ g_5^{\text{COX}}(\cdot) & \sim \mathcal{GP}_{\text{RW1}}(1/\tau_4), & n_5^{\text{COX}} &= 5;\end{aligned}$$

$$\theta^{\text{COX}} = \{\alpha^{\text{COX}}, \phi_1, \phi_2, \tau_1, \tau_2, \tau_3, \tau_4\} \sim \text{Hyperpriors}.$$

Spatial occurrence hot-spots (see Supplement §4.9.3) may arise due to time-invariant landuse. Moreover, spatial variation may be shared from patterns in the BETA and BIN components through the components $g^{\text{COX-BETA}}(s_i)$ and $g^{\text{COX-BIN}}(s_i)$. The month and year effects, g_4^{COX} and g_5^{COX} , capture spatially homogeneous temporal variation in occurrence intensities. They are assigned first-order random-walk priors $\mathcal{GP}_{\text{RW1}}$ with a sum-to-zero constraint for identifiability; *e.g.*, for the yearly effect and for $a = 1995, \dots, 2013$,

$$g_4^{\text{COX}}(a+1) - g_4^{\text{COX}}(a) \sim \mathcal{N}(0, 1/\tau_3), \quad \sum_{i=1995}^{2014} g_4^{\text{COX}}(i) = 0.$$

The quadratic B-spline functions of FWI and FA are assigned priors $\mathcal{GP}_{\text{1D-SPDE}}$, constrained to zero at the left boundary 0 and to sum to zero, respectively. Most wildfires in the region are caused by human activity, possibly leading to a nonlinear relationship between FA and occurrence intensity, as dense forest areas are often exposed to low human activity. We model monthly variation of the nonlinear FWI effect through separate $\mathcal{GP}_{\text{1D-SPDE}}$ -terms in g_3 for each month, linked across successive months with a $\mathcal{GP}_{\text{RW1}}$ -structure in the prior model.

The regression equation used for the Bernoulli process is

$$\begin{aligned}\mu_{it}^{\text{BIN}} = & \alpha^{\text{BIN}} + g^{\text{COX-BIN}}(s_i) + \beta^{\text{BIN-GPD}} g^{\text{BIN-GPD}}(s_i) + g_1^{\text{BIN}}\{z_{\text{FWI}}(s_i, t)\} \\ & + g_2^{\text{BIN}}\{z_{\text{FA}}(s_i, t)\} + g_3^{\text{BIN}}\{a(t)\};\end{aligned}$$

$$\begin{aligned}g_k^{\text{BIN}}(\cdot) & \sim \mathcal{GP}_{\text{1D-SPDE}}(\zeta_k), \quad k = 1, 2, & n_1^{\text{BIN}}, n_2^{\text{BIN}} &= 5, \\ g_3^{\text{BIN}}(\cdot) & \sim \mathcal{GP}_{\text{RW1}}(1/\tau_5), & n_3^{\text{BIN}} &= 5;\end{aligned}$$

$$\theta^{\text{BIN}} = \{\alpha^{\text{BIN}}, \zeta_1, \zeta_2, \tau_5\} \sim \text{Hyperpriors}.$$

The linear predictor of the Bernoulli probability has a simpler form than the occurrence component but still allows the capture of specific nonlinear effects of FWI and FA. In Figure 4.1, we discern hot-spot areas of large fire occurrences that differ substantially from the overall occurrence structure, and we aim to capture these residual effects through the shared spatial effects.

The prior structure for the two mixture components of quantitative marks is

$$\begin{aligned}\mu_{it}^{\text{BETA}} = & \alpha^{\text{BETA}} + g^{\text{COX-BETA}}(s_i) + g_1^{\text{BETA}}\{z_{\text{FWI}}(s_i, t)\} + g_2^{\text{BETA}}\{z_{\text{FA}}(s_i, t)\}, \\ \mu_{it}^{\text{GPD}} = & \alpha^{\text{GPD}} + g^{\text{BIN-GPD}}(s_i) + g_1^{\text{GPD}}\{z_{\text{FWI}}(s_i, t)\} + g_2^{\text{GPD}}\{z_{\text{FA}}(s_i, t)\} \\ & + g_3^{\text{GPD}}\{a(t)\};\end{aligned}$$

$$\begin{aligned}g_k^{\text{BETA}}(\cdot), g_k^{\text{GPD}}(\cdot) & \sim \mathcal{GP}_{\text{1D-SPDE}}(\kappa_k), \quad k = 1, 2, & n_1^{\text{GPD}}, n_2^{\text{GPD}}, n_1^{\text{BETA}}, n_2^{\text{BETA}} &= 5, \\ g_3^{\text{GPD}}(\cdot) & \sim \mathcal{GP}_{\text{RW1}}(1/\tau_6), & n_3^{\text{GPD}} &= 5;\end{aligned}$$

$$\theta^{\text{MARK}} = \{\alpha^{\text{GPD}}, \alpha^{\text{BETA}}, \kappa_1, \kappa_2, \tau_6\} \sim \text{Hyperpriors}.$$

We assigned random-walk priors to year effects included in some of the components (COX, BIN, GPD). In all components (BETA, BIN, COX, GPD), we model non-linear relationships with respect to FWI or FA.

4.5.4 Alternative model specifications

We also consider size processes that do not model the moderate-level and extreme marks separately; *i.e.*, with no mixture representation of the size process. Similar models have been proposed in the literature (*e.g.*, Joseph et al., 2019), though without the sharing of random effects. We use either the Gamma distribution for the full range of marks: $Y_{it,k} \mid \mu_{it}^{\text{SIZE}} \sim \text{Gam}\{\exp(\mu_{it}^{\text{SIZE}}), \phi_{\text{Gam}}\}$, or the Normal distribution for the logarithmically transformed marks: $\log Y_{it,k} \mid \mu_{it}^{\text{SIZE}} \sim \mathcal{N}\{\exp(\mu_{it}^{\text{SIZE}}), \phi_{\mathcal{N}}\}$, where the distributions are parameterized by the link

Chapter 4. Spatiotemporal wildfire modeling through point processes with extreme marks

function μ_{it}^{SIZE} modeling the mean and precision parameters $\phi_{\text{Gam}} = \exp(\mu_{it}^{\text{SIZE})^2 / \text{Var}(Y_{it,k})$ and $\phi_{\mathcal{N}} = 1 / \text{Var}(\log Y_{it,k})$, respectively. In both cases

$$\begin{aligned} \mu_{it}^{\text{SIZE}} = & \alpha^{\text{SIZE}} + g^{\text{SIZE-COX}}(s_i) + g_1^{\text{SIZE}}\{z_{\text{FWI}}(s_i, t)\} + g_2^{\text{SIZE}}\{z_{\text{FA}}(s_i, t)\} \\ & + g_3^{\text{SIZE}}\{a(t)\} + g^{\text{SIZE}}(s_i); \end{aligned}$$

$$\begin{aligned} g_k^{\text{SIZE}}(\cdot) & \sim \mathcal{GP}_{1\text{D-SPDE}}(\boldsymbol{\mu}_k), \quad k = 1, 2, & n_1^{\text{SIZE}}, n_2^{\text{SIZE}} &= 5, \\ g_3^{\text{SIZE}}(\cdot) & \sim \mathcal{GP}_{\text{RW1}}(1/\tau_7), & n_3^{\text{SIZE}} &= 5; \end{aligned}$$

$$\boldsymbol{\theta}^{\text{SIZE}} = \{\alpha^{\text{SIZE}}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \tau_7\} \sim \text{Hyperpriors},$$

where the spatial effects $g^{\text{SIZE-COX}}(s_i)$ and $g^{\text{SIZE}}(s_i)$ are controlled by Matérn parameters ω_4 and ω_5 , similar to those in §4.5.2.

4.6 Results

4.6.1 Model selection and comparison

Estimation was carried out using the INLA-SPDE approach described in §4.4.5 by applying the subsampling scheme proposed in §4.4.4. In a preliminary analysis of the regression models described in §4.5, we used the Widely Applicable Information Criterion (WAIC, Watanabe, 2010) in a step-wise manner to compare nested models with different components in the regression equations (*e.g.*, linear vs nonlinear effects of explanatory variables) to choose their final forms.

We label the model with prior structure detailed in §4.5.3 M1, and the model without spatial effects in the size and extreme occurrence components M2. We also considered other models from the recent wildfire modeling literature. We refer to model M2 but without monthly variation in the FWI effect as M3, which is similar to the approach of Pimont et al. (2021). We let M4 and M5 denote the models with the point process model of M1 but with no mixture representation of the size process, for which we use the log-Normal and Gamma distribution for the size distribution in the model structure described in §4.5.4. These models do not differentiate between extreme and non-extreme fires, but their response distributions were found to be good modeling candidates in Joseph et al. (2019), though their approach does not use shared random effects.

For the observed individual fires in the validation (2015–2018) periods, we generated posterior predictive distributions of each model based on 500 posterior simulations. First, we evaluated the models' ability to predict exceedances above the empirical 90% quantile of burnt areas using the AUC (Fawcett, 2006) and the Brier score (Brier, 1950). The severity threshold chosen here is sufficiently high for extreme risk assessment, but low enough to retain enough observa-

	Score	Model				
		M1	M2	M3	M4	M5
Individual fires, $n = 823$	sCRPS	2.74	2.87	2.94	2.84	3.19
	p-value	-	< 5%	< 1%	< 5%	< 1%
	Brier _{q90}	0.0855	0.0868	0.0866	0.0944	0.0967
	p-value	-	< 5%	6%	< 1%	< 1%
	1 – AUC _{q90}	0.3052	0.3502	0.3516	0.3184	0.3122
	p-value	-	< 5%	< 5%	40%	41%
Dép-month, $n = 75$	sCRPS	3.55	3.62	3.64	3.62	3.58
	p-value	-	7%	7%	9%	39%

Table 4.1 – Comparison of models using predictive scores (averaged over n observations) calculated with data from the validation period: sCRPS, Brier and AUC scores for individual fires, and sCRPS for the spatiotemporally aggregated burnt areas at month-département scale, based on 500 simulations of the posterior models, with p-values for a permutation test comparison with the best model M1. A lower score is better.

tions to evaluate these scores with sufficient precision. As we considered average predictive score across all observations in the validation set, we also computed the scaled Continuous Ranked Probability Score (sCRPS Bolin and Wallin, 2020), which scales each observation's CRPS before calculating the average. For these analyses, we kept the original locations of observed fires, and simulated only from the size components. By combining posterior simulations of the occurrence and size components, we also evaluated burnt area predictions aggregated at the month-département scale.

Table 4.1 shows good relative performance of M1 for all scores when evaluating wildfire predictions on the validation period. To better grasp the uncertainty in scores, we show p-values of a permutation test assessing the significance of negative values in the differences of scores between M1 and the other models, based on 2000 permutations. For the sCRPS of individual fires, the score differences are all significant at the 5% level. A general finding is that using sophisticated structures such as the mixture representation of size processes, sharing and monthly variation of FWI effect, improves predictions; it further allows for the novel scientific insights presented in §4.7.

Comparison of M1 and M2 confirms the benefits of incorporating spatial random effects in the size model components in M1 using the sharing detailed in §4.5.2. M1 performs better than M2, and M2 and M3 give similar predictions for wildfire sizes and their aggregation. Model M1 performs better than M4 and M5, especially with respect to Brier and sCRPS scores, though in some cases improved scores have relatively high p-values. Models M4 and M5 perform better than M2 and M3 for some scores like the AUC and sCRPS at the month-département aggregation because of the additional sharing and spatial random effects in the size component, though they perform worse for the other scores due to having no components for extreme wildfires. Despite good scores of M4 and M5 on the training set (not shown), their

Chapter 4. Spatiotemporal wildfire modeling through point processes with extreme marks

worse results on the validation set suggest that the log-Normal and Gamma distributions for burnt areas do not predict the extremes in new data as well as M1, which does not show overfitting.

4.6.2 Visual inspection of posterior predictive densities

We also assess the predictive behavior of our chosen model M1 visually. First, we assess whether the size component correctly predicts extreme wildfires for specific départements. In the Supplement §4.9.1 (Figure 4.9), we use simulations from the posterior model at pixel-days where fires have been observed to compare empirical and predicted excess probabilities over high thresholds, starting at 100ha. Predictions are generally good, since most empirical exceedance probabilities fall within the inter-quantile range of simulations, except for the départements of Var and Haute-Corse, in which there is slight underestimation at very large thresholds. These two départements have large continuous forest areas and saw unusually many large wildfires in the summer of 2017. Much of their land has acidic soils that favor biomass production and is covered by tall and dense shrubland, so the 2017 fires were harder to contain due to their higher heat release. Overall, the tail behavior in fire-prone and less fire-prone regions is well modelled.

Next, we compare the numbers of simulated and observed fires aggregated by year over the study region (Supplement §4.9.1 : Figure 4.10, left display). Observed annual fire numbers for both test and training sets fall within the inter-quantile range of simulations for more than half of the study period. M1 captures the relatively high observed numbers of 2001, 2003 and 1998 (training) and 2017 (test), while it also accurately predicts the sharp decrease in 2018.

Lastly, we jointly evaluate the size and occurrence components of our model M1. We aggregated simulated burnt areas by year, over the whole spatial region in Figure 4.10 (right display), and over départements in Figure 4.11. The global time trend in observed burnt areas is well captured in Supplement §4.9.1, Figure 4.10, with inter-quantile coverage of 42%. M1 captures the exceptional peak in 2003, which is poorly predicted by M4 and M5 and the Firelihood model of Pimont et al. (2021). M1 also succeeds in accurately predicting the moderately high burnt areas in 2001 and 2017, and it generally discriminates well between fire conditions leading to small, moderate, large and very large fire numbers. Figure 4.11 further shows that regional differences across départements are well captured by M1. Overall, our model captures spatiotemporal variation and provides satisfactory regionalized forecasts for operational purposes.

4.6.3 Principal results of the main model M1

Covariate effects

For the COX component, Figure 4.4 shows that the month-specific FWI effect varies significantly across months. For easier comparison, we have centred the curves so that the posterior

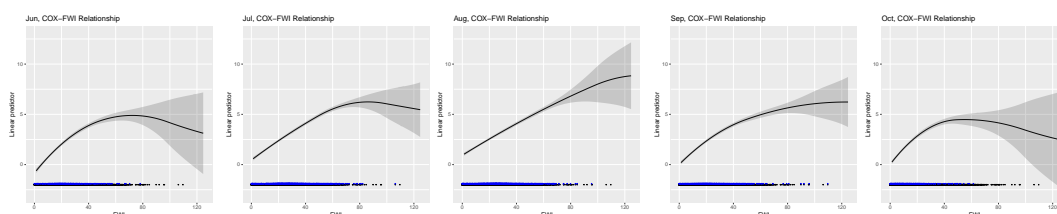


Figure 4.4 – Posterior estimates of $g_3^{\text{COX}}(\cdot; m) + g_5^{\text{COX}}(m)$, $m = 1, \dots, 5$, the joint FWI-month effect, for June–October in the linear predictor of the point process (COX) component. The blanket of black and blue points at the bottom of each plot shows FWI values for pixel-days with fires in any month and the specific month, respectively.

mean is 0 for FWI= 0 in September. All the posterior means increase monotonically up to FWI of 75, and then flatten for higher values, especially at the beginning and end of the wildfire season, with a slight decrease of the curve towards the highest FWI.

The posterior partial effect of FA on the COX component in Figure 4.5 indicates a “bump”-shaped effect of FA, which is significant based on pointwise credible intervals. High FA can be regarded as a good proxy for low human activity, while low FA means lack of fuel. Clearly, expected wildfire ignition numbers are not proportional to forest area.

As to temporal partial effects without spatial variation (Figure 4.5), the posterior year effect suggests a significant drop in wildfire activity after 2003, potentially related to policy changes after the exceptional 2003 events. The partial month effect (top right display of Figure 4.5, corresponding to the intercept of its combined effect with FWI in Figure 4.4) is lowest at the start of the wildfire season and peaks in August.

As to the probability of occurrence of large fires (BIN), Figure 4.5 (bottom middle display) highlights a strong positive posterior effect of FWI, increasing monotonically and significantly up to FWI values of around 75, before it dampens at very large FWI values, similar to the COX component: large wildfires are more frequent with moderate to high FWI values. The probability of large wildfires tends to increase with increasing FA in a grid cell (Figure 4.5, bottom left display), which is reasonable because more FA fuel is available over large areas. The pointwise credible bounds of yearly effects across the study period suggest that the occurrence of large events was significantly higher around the peak in 2003.

The additive effects in the GPD and BETA components of the size distribution, seen in Figures 4.6, show similar posterior effects of FWI and forest area for extreme and moderate sizes. The posterior estimates imply that fires become larger when FWI increases until 60 but this effect flattens for higher FWI. Increasing FA leads to increasing wildfire size in both components until 50%, then reaches a plateau after. For the year effect in the extreme component GPD, no clear trend arises, though 2003 has a significantly higher effect than 1998.

Chapter 4. Spatiotemporal wildfire modeling through point processes with extreme marks

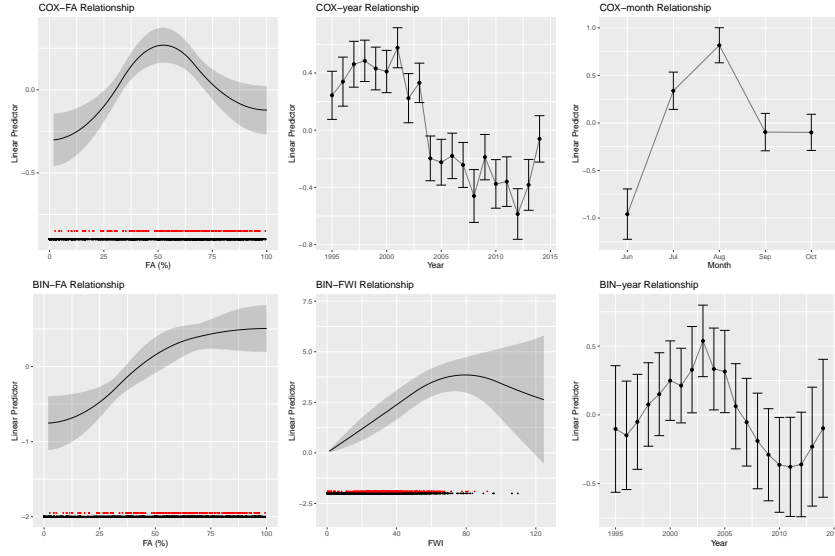


Figure 4.5 – Posterior estimates of $g_2^{\text{COX}}(\cdot)$ (FA effect, top left panel), $g_4^{\text{COX}}(\cdot)$ (year effect, top middle panel), $g_5^{\text{COX}}(\cdot)$ (month effect, top right panel), $g_2^{\text{BIN}}(\cdot)$ (FA effect, bottom left panel), $g_1^{\text{BIN}}(\cdot)$ (FWI effect, bottom middle panel) and $g_3^{\text{BIN}}(\cdot)$ (year effect, bottom right panel) in the linear predictor of the point process (COX) component and large wildfire probability component (BIN). At the bottom of some displays, the blanket of black and red points shows FA/FWI values for pixel-days with moderate and large fires, respectively.

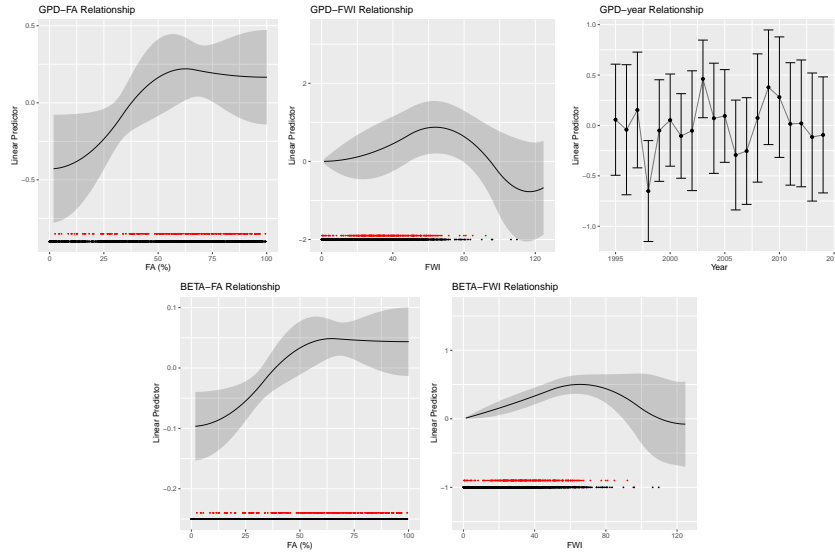


Figure 4.6 – Panels as in Figure 4.5. Posterior estimates of $g_2^{\text{GPD}}(\cdot)$ (FA effect, top left), $g_1^{\text{GPD}}(\cdot)$ (FWI effect, top middle), $g_3^{\text{GPD}}(\cdot)$ (year effect, top right), $g_2^{\text{BETA}}(\cdot)$ (FA effect, bottom left) and $g_1^{\text{BETA}}(\cdot)$ (FWI effect, bottom right) in the linear predictor of the large wildfire size component (GPD) and moderate wildfire size component (BETA).

Sharing effects induce correlated wildfire activity components

Here we focus only on the spatial effects shared between model components. The 95% credible intervals for the scaling parameters $\beta^{\text{COX-BETA}}$, $\beta^{\text{COX-BIN}}$ and $\beta^{\text{BIN-GPD}}$ do not cover 0; their

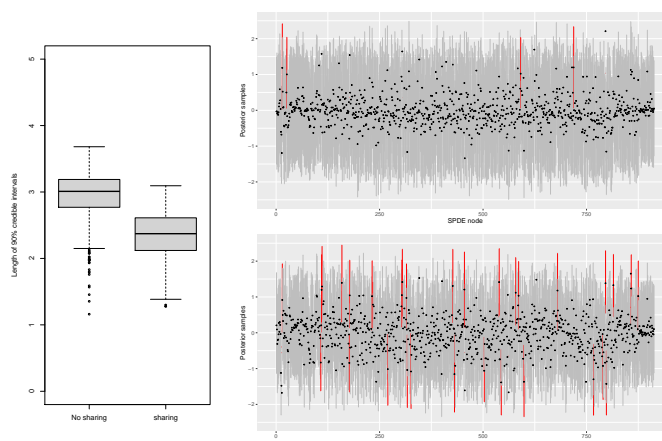


Figure 4.7 – Lengths of the 90% credible intervals of spatial random effect variables at the SPDE triangulation nodes within the study area in the BIN component, based on 500 posterior simulations. Boxplots (left), and error bar plots for the models without (top right) and with sharing (bottom right). Red error bars indicate nodes where the intervals do not include zero.

posterior estimates for the triplet (2.5% quantile, mean, 97.5% quantile) are (6.4, 10.3, 14.0), $(-3.1, -1.8, -0.9)$, and (0.5, 1.0, 1.6), respectively. The posterior mean of $\beta^{\text{COX-BETA}}$ is positive and that of $\beta^{\text{COX-BIN}}$ is negative, confirming significant positive and negative sharing between COX and BETA, and COX and BIN, respectively; these findings provide new spatial insights for fire risk management described in §4.7. The posterior means for the effective range parameters of the shared spatial fields, $r^{\text{COX-BETA}}$, $r^{\text{COX-BIN}}$ and $r^{\text{BIN-GPD}}$, are 34.3km, 26.2km and 156.9km, respectively. Posterior mean maps of their corresponding spatial random effects are shown in the Supplement §4.9.7.

Sharing decreases uncertainty by borrowing estimation strength between model components. The average lengths of 95% posterior credible intervals of variables constituting the random effect shrink by up to 30% (Figure 4.7) because of a higher observation-to-parameter ratio that enables us to better capture relevant spatial signals.

To identify the hot-spot regions of spatial random effects, we study credible sets for excursion regions (Bolin and Lindgren, 2015). We evaluate where the fields exceed or fall below the thresholds $u = 0.1$ and $-u$, respectively. These thresholds approximately correspond to a 10% increase and decrease, respectively, on the scale of the response when taking into account the log or logistic link. The u -excursion set with probability α , $E_{u,\alpha}^+(X)$, is defined as the largest set for which the level u is exceeded at all locations in the set with probability $1 - \alpha$. The negative u excursion set with probability α , $E_{u,\alpha}^-(X)$, is defined as the largest set for which the process remains below the level $-u$ at all locations in the set with probability $1 - \alpha$. This approach determines the largest set contained in the exceedance set with a minimum probability threshold, and it assumes a parametric family for the exceedance sets. To visualize excursion sets simultaneously for all values of α , Bolin and Lindgren (2015) introduced the positive and negative excursion functions $F_u^+(s) = 1 - \inf\{\alpha \mid s \in E_{u,\alpha}^+\} \in [0, 1]$

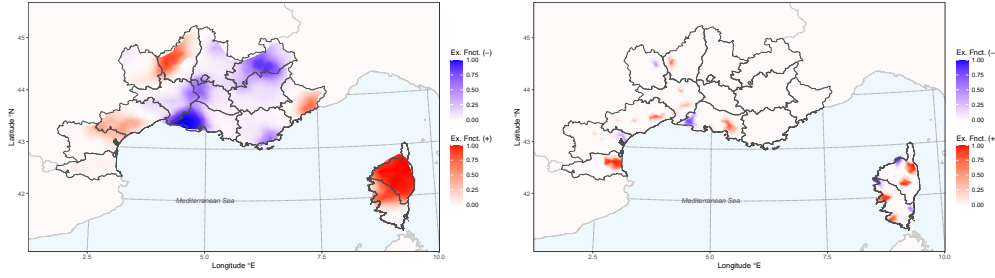


Figure 4.8 – Excursion functions of posterior latent fields above 0.1 and below -0.1 . Plots show $\max\{F_{0.1}^+(\cdot), F_{0.1}^-(\cdot)\}$ for the shared spatial random fields $g^{\text{COX-BETA}}$ (left panel) and $g^{\text{COX-BIN}}$ (right panel).

and $F_u^-(s) = 1 - \inf\{\alpha \mid s \in E_{u,\alpha}^-\} \in [0, 1]$. Figure 4.8 highlights several hot-spot regions for the shared spatial effects, which we interpret with respect to wildfire management in §4.7.

4.7 New insights for wildfire science

Pimont et al. (2021) pointed out several critical divergences between simulations of their model and observed wildfire activity, and they have proposed hypotheses to explain them. The novel models developed here, especially M1, include components to estimate the sources of space-time variability conjectured by Pimont et al. (2021), leading to a better fit and more reliable inferences and predictions. Here, we outline the new insights.

4.7.1 FWI and seasonal effects

The estimated FWI effect on all wildfire components (COX, BIN, BETA, GPD) is nonlinear with a strong increase when moving from $\text{FWI} = 0$ towards $\text{FWI} \approx 60-80$, followed by a dampening and a slight decrease for extreme FWI values, though with wide credible bounds. Moreover, seasonal patterns emerge in the joint FWI-month effect in the occurrence component COX. The common practice of using FWI directly as a proxy for wildfire activity, without a nonlinear and seasonally varying transfer function as estimated here, would predict extreme wildfires badly and miss seasonally varying response of fire activity to this index.

This non-linear, even decreasing, response to high FWI and seasonal biases can be attributed to the excessively sharp exponential response of FWI to wind speed in its upper range and to the limited ability of the Drought Code (a subcomponent of the FWI) to reproduce live fuel moisture dynamics in France (Ruffault et al., 2018). In spring, vegetation budburst produces new foliage with a high water content that is maintained until the onset of the summer drought, typically in early July. The timing of periodic events in plant life cycles and stomatal control under drought might also explain why dynamics of soil and vegetation water content are unsynchronized at certain times. In our COX component, we model not only a seasonal

effect but also different responses of FWI across the months. The shapes of these monthly responses vary greatly, so seasonal variations cannot be handled solely through separate random effects. The response in August did not exhibit any saturation in the upper part of the FWI range, suggesting that higher values in mostly dry conditions correspond to increased fire activity; the contribution of wind to FWI could be adequate in these already-dry conditions. On the contrary, a flattening and notable decrease of the COX response to FWI was observed at $\text{FWI} \approx 45\text{--}50$ for relatively moist conditions in June and October. This supports the hypotheses that the desynchronization of soil and fuel moistures caused by plant phenology in spring could be involved, and the response of the FWI to high wind would be inaccurate in such moist conditions. July and September, with their mixture of dry and moist days, show intermediate response levels to very high FWI. These findings confirm a need to develop better wildfire danger indices in the study region.

4.7.2 Time trends during the study period

The year 2003 was catastrophic in terms of fire sizes and burnt area. It has a pivotal role with a decrease of occurrence numbers and sizes afterwards, as highlighted by the year component of our posterior model that captures temporal trends not explained through weather and land-cover related predictors. In 2003, a heat wave coincided with severe drought conditions, leading to an unusually high number of escaped fires ($> 1\text{ha}$), and of fires larger than 10ha for several weeks, whose occurrence was not matched by very high values of FWI due to its weaknesses outlined in §4.7.1. The drop in the estimated yearly effect after 2003 could be due to official policy measures that have slightly evolved after 2003, and to better prevention or suppression policies by fire managers (Pimont et al., 2021).

The yearly effect from the BIN component should interest wildfire managers, as it shows that the probability of observing a large fire increases following a decade of continuous decrease. Our results also confirm those of Evin et al. (2018), who found no clear time trend for the probability of extreme fires.

4.7.3 Shared spatial effects for improved regionalized predictions

The shared spatial effects shown in §4.6.3 highlight regional differences in fire size distributions and provide quantitative interpretations of effects. They also reveal substantial regional variation in proportions of moderate and extreme fires. In particular, the sharing effect with significantly negative $\beta^{\text{COX-BIN}}$ allows for interpretation with respect to different wildland-to-urban interactions. The lowland area in the western Pyrénées-Orientales region, fairly densely populated with a large proportion of abandoned agricultural land intermixed with urban surfaces, appears to have high occurrence intensities, but its combustible area is strongly fragmented, so wildfires are mostly small. More fires than expected from weather/climate and forest area occur in densely populated or rural landscapes with significant human activities promoting fire ignitions, while landscape fragmentation and landscape management reduce

the likelihood of large fires. The COX-BETA sharing effect is highly positive in Corsica, where extreme fires become larger more often than elsewhere, perhaps due to longer arrival times of firefighters in remote Corsican forests and less frequent airborne firefighting. Moreover, extreme fires tend to be more frequent because of large contiguous forests. Further regional disparities in predictions are illustrated in the Supplement S4.9.2 where the right panel of Figure 4.12 highlights significant differences in threshold exceedance probabilities.

4.8 Conclusion

We have implemented a novel Bayesian spatiotemporal model for wildfire activity with specific components for extreme events, and with shared random effects to account for stochastic dependence among components not explained by covariates. The sophisticated structure of our fully Bayesian hierarchical model allows us to accurately disentangle the effects and interactions of various observed and unobserved drivers. The use of Gaussian random effects at high spatial resolution provides crucial benefits over frequentist generalized additive models, since fine-scale spatial variation and associated uncertainties can be identified properly.

Different sharing strategies respond to different considerations. If statistical stability is the focus, then sharing from well-identified model components towards those less informed by data is appropriate. If focus is on accurate inference of a specific component (*e.g.*, extremes), then it is sensible to share effects from this component towards others. In both cases, component-specific effects without sharing remain important and should be included as far as data allow them to be estimated. In some applications, however, introducing common components by sharing is the only way to incorporate spatial effects in certain response variables. For example, had we chosen a threshold larger than 79ha for large wildfires, we would have had even fewer observations available for the extreme fire size component. A separate spatial effect in this component would provide wider credible intervals than those in Figure 4.7 (top right), and be of less practical use. Our findings improve decision support in wildfire management: shared spatial effects explain how wildfire numbers and extreme sizes interact by providing maps of the significant disparities between regions. Moreover, FWI maps used for fire danger rating must be interpreted with care because of the strong nonlinear and seasonal effect on wildfire risk identified by our model. Future work could explore the spatial disparity in temporal trends due to changes in landuse practices and fire management, by incorporating space-varying temporal random effects.

Beyond wildfire modeling, our approach could be used to provide new insights and improved extreme-value predictions for other problems. Landslide inventories can be represented as point processes with heavy-tailed magnitude marks (Stark and Hovius, 2001; Lombardo et al., 2020). Another promising application is in modeling locations, times and values of high-impact events extracted from processes indexed over space and time, such as local extremes in gridded climate data. This would yield a parsimonious representation of extreme events in such processes. Models for preferentially sampled spatial data (Diggle et al., 2010) can be

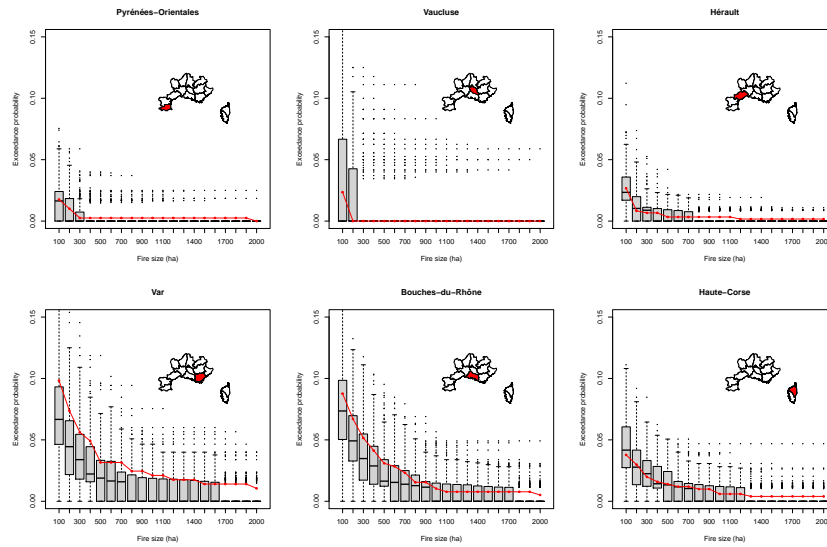


Figure 4.9 – Exceedance probability plots for six départements (in red on the maps) in the validation period (2015–2018). Boxplots are based on 200 posterior simulations. Red lines represent observed empirical exceedance probabilities.

viewed as marked point processes with shared effects, such that our approach would allow the capture of preferential sampling effects specifically in extreme values.

4.9 Supplement

4.9.1 Plots for the inspection of posterior predictive densities

Figures 4.9, 4.10 and 4.11 show our visual assessment of the predictive behavior of our chosen model M1.

4.9.2 Plots showing regionalized predictions

Figure 4.12 shows the regionalized predictions due to the spatial effects used in our model.

4.9.3 Kernel intensity plot

Figure 4.13 shows a map of the wildfire locations as recorded in the Prométhée database. The overlaid contour lines of a kernel intensity estimation highlight the strong spatial nonstationarity, with several relatively small hotspot areas characterized by high occurrence numbers.

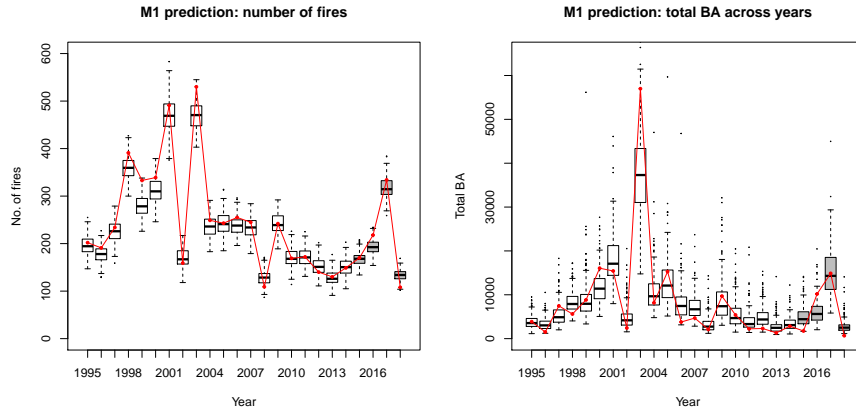


Figure 4.10 – Boxplots by year for the predicted numbers of fires (left) and total burnt area (right) across the whole region from 200 simulations of the posterior model. The grey boxplots indicate the out-sample years. The red lines represent the observed annual total number of fires and burnt area in the whole region.

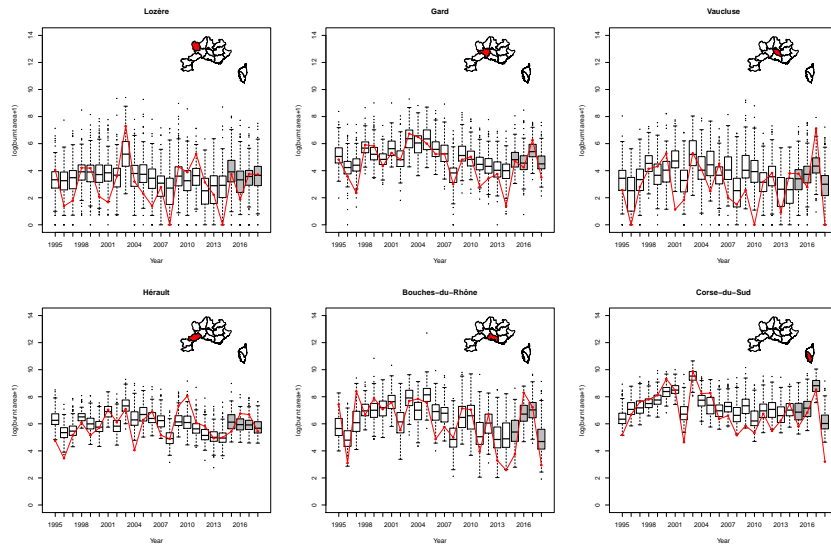


Figure 4.11 – Boxplots by year and département for the predicted logarithmic total burnt area from 200 simulations of the model. Red dots represent the observed annual log total burnt area in each département. The département used for each panel is shown in red on the maps.

4.9.4 Mean excess plots

We consider the mean excess plots of burnt areas and \log_{10} BA in the middle and right displays of Figure 4.14. Given a threshold value u set for a random variable Y , the mean excess corresponds to the conditional expectation $\mathbb{E}[Y - u \mid Y > u]$, *i.e.*, the expectation of the positive excess above the threshold. Mean excess plots report the corresponding empirical means. In case of exponential tail decay $\Pr(Y > y) = \exp\{-(y - \mu)/\lambda\}$ for $y \geq u_0$ with scale

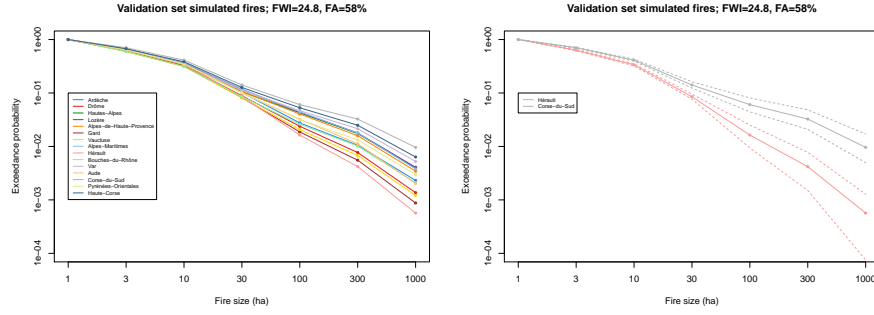


Figure 4.12 – Left: Exceedance probability plots by département from 1000 posterior samples of the model M1 on the validation period given fixed FWI and FA. Right: Same as the left panel, but only for two départements, with corresponding 95% credible intervals.

$\lambda > 0$ and an arbitrary shift $\mu \in \mathbb{R}$, the mean excess would be constant λ for thresholds u above u_0 . The mean excess plot for \log_{10} BA indicates approximately exponential tail decay for low thresholds, where mean excess values are relatively stable for threshold values in $(0, 1.5)$ except for rounding of burnt areas. However, the tail decay becomes faster at around 30 ha. Exponential decay on log-scale would correspond to power-law decay at the original scale; *i.e.*, to Pareto-like behavior. By contrast, the mean excess plot of original BA values becomes relatively stable for thresholds above 500 ha, such that the true, ultimate tail decay rate at very high quantiles could be exponential. These plots reveal the difficulty of choosing an appropriate probability distribution for burnt areas.

4.9.5 Subsampling experiments

We implemented several experiments to aid the choice of the parameters of the subsampling scheme detailed in §4.4.4. We fixed the sampling probability parameter to $p_{SS} = 0.9$ but allowed the empirical FWI probability p_{FWI} to equal $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. The case $p_{SS} = 1 - p_{FWI}$, *i.e.*, $p_{FWI} = 0.1$, corresponds to uniform subsampling, whereas higher values of p_{FWI} include a relatively larger number of high FWI observations in the subsample. In a first simulation experiment, we sampled from the COX model with log-linear intensity in §4.5.1,

$$\mu_{i,t}^{COX} = \alpha + \beta_1 z_{FWI}(s_i, t) + \beta_2 m(t),$$

with $\alpha = -11$, $\beta_1 = 0.15$ and $\beta_2 = 0.1$ to reflect intensities that could be realistic in a wildfire application, and fit this model with INLA. Figure 4.15 highlights the improvement in estimation quality by moving away from uniform subsampling, with lower root mean squared errors of the posterior means. In another experiment, we estimated the COX model with the linear predictor in §4.5.3 and evaluated the sCRPS scores for the annually aggregated predicted and observed number of fires over the whole spatial region in the training set with 500 posterior simulations and 50 different subsampling seeds. The left panel of Figure 4.16 shows that subsampling scheme with $(p_{FWI}, p_{SS}) = (0.7, 0.9)$ achieves the best score. Next, we repeated

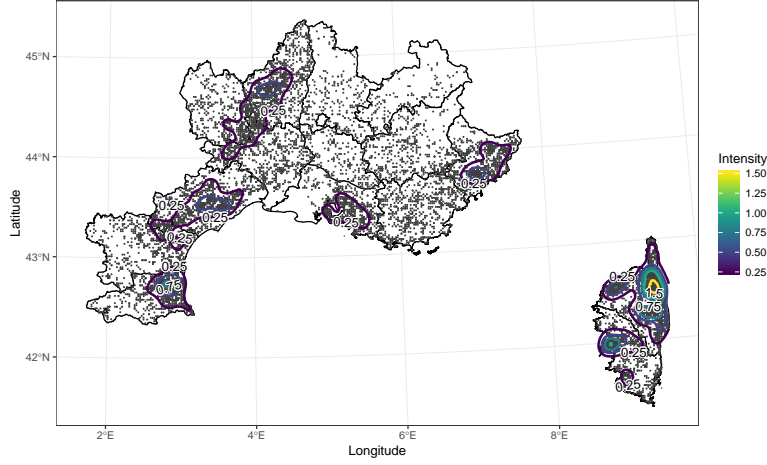


Figure 4.13 – Map of Prométhée data in Southern France with the island of Corsica on the lower right, based on the original DFCI grid used for recording wildfires. Black lines indicate boundaries of administrative regions (“départements”). Coloured lines correspond to intensities (*i.e.*, to average numbers of wildfires per km²) and highlight areas with many wildfires. Some of the gray points correspond to multiple wildfire occurrences.

the experiment with a fixed p_{FWI} and p_{SS} , but increased the number of subsamples taken within each pixel-year. The right panel of Figure 4.16 shows that there is little improvement in sCRPS score beyond two subsamples per pixel-year, while the computational time and memory requirements increase strongly and non-linearly with the number of subsamples (not shown).

4.9.6 Other hyperpriors

All fixed effect coefficients in our models (*e.g.*, α^{COX} , α^{BIN} , α^{GPD} and α^{BETA}) are assigned flat Gaussian priors with zero mean and precision 0.001. The prior for each of the scaling parameters $\beta^{COX-BETA}$, $\beta^{COX-BIN}$ and $\beta^{BIN-GPD}$ is a zero-centered Gaussian distribution with precision 1/20. To reduce the number of estimated hyperparameters, we fixed the hyperparameters associated with the priors $\mathcal{GP}_{1D-SPDE}$ to values guided by prior knowledge about the relationship between FWI/FA and the relevant aspects of wildfire risk. For the tail index parameter ξ in the GPD component, we assign an exponential distribution with rate unity, which corresponds to an approximate penalized complexity prior (Opitz et al., 2018) with moderate level of penalization from the base model ($\xi = 0$). Lastly, we assign a log-Gamma hyperprior with mean unity and precision 0.0005 to each of the random-walk hyperparameters $\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6$ and τ_7 .

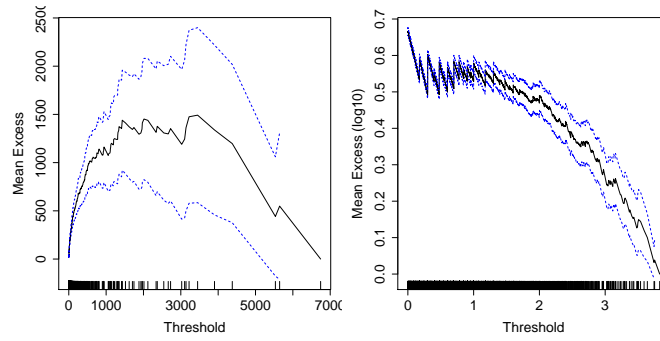


Figure 4.14 – Mean excess plots. Left: for burnt area (in ha). Right: for log10 of burnt area. Original observations are indicated at the bottom of the mean excess plots. Blue lines indicate symmetric pointwise confidence intervals at 95%.

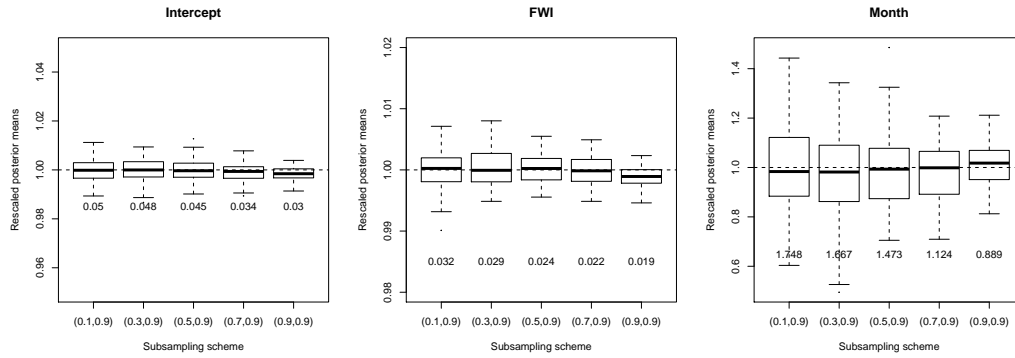


Figure 4.15 – Boxplots of posterior means (rescaled by dividing them by the true parameter value) of fixed effect coefficients from 100 simulations with different (p_{FWI} , p_{SS}) combinations. The relative root mean square errors (rRMSE) for each subsampling scheme are displayed below the corresponding boxplots.

4.9.7 Spatial effects in model M1

Figure 4.17 shows the same plot as Figure 4.8 but for the shared spatial random field $g^{\text{BIN-GPD}}$. Figure 4.18 shows the posterior means of all the spatial model M1, with priors detailed in §4.5.3.

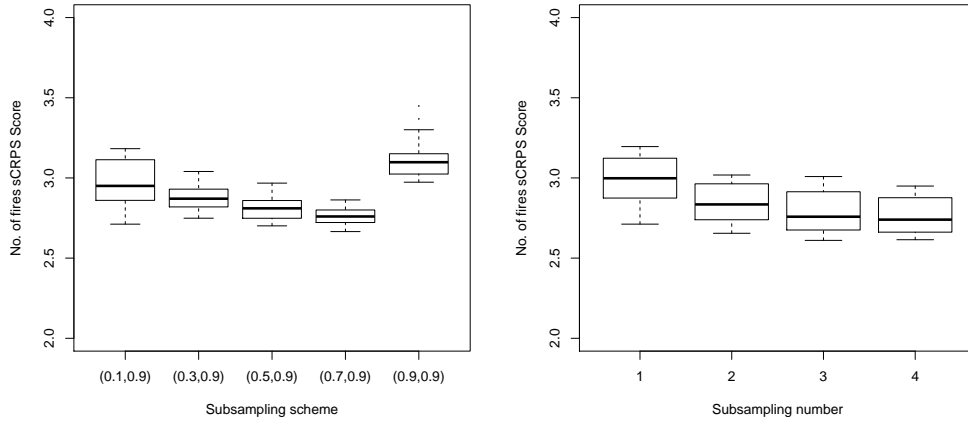


Figure 4.16 – sCRPS for the annually aggregated predicted and observed number of fires over the whole spatial region in the training set (1994–2014), with the subsampling scheme over 50 different seeds. Right: Boxplots of sCRPS scores for the subsampling scheme with one subsample per pixel-year and different (p_{FWI}, p_{SS}) combinations. Left: Boxplots of sCRPS scores for the subsampling with $(p_{FWI}, p_{SS})=(0.1,0.9)$, with different number of subsamples per pixel-year.

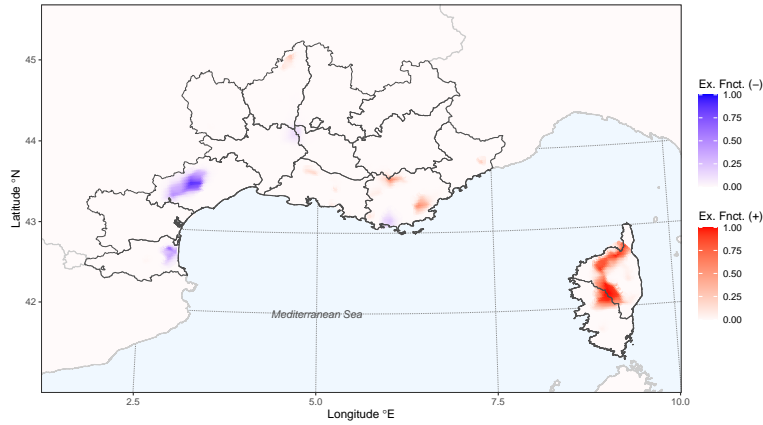


Figure 4.17 – Excursion functions of posterior latent fields above $u = 0.1$ and below $-u$. Plots show $\max\{F_{0.1}^+(\cdot), F_{0.1}^-(\cdot)\}$ for the shared spatial random field $g^{\text{BIN-GPD}}$.

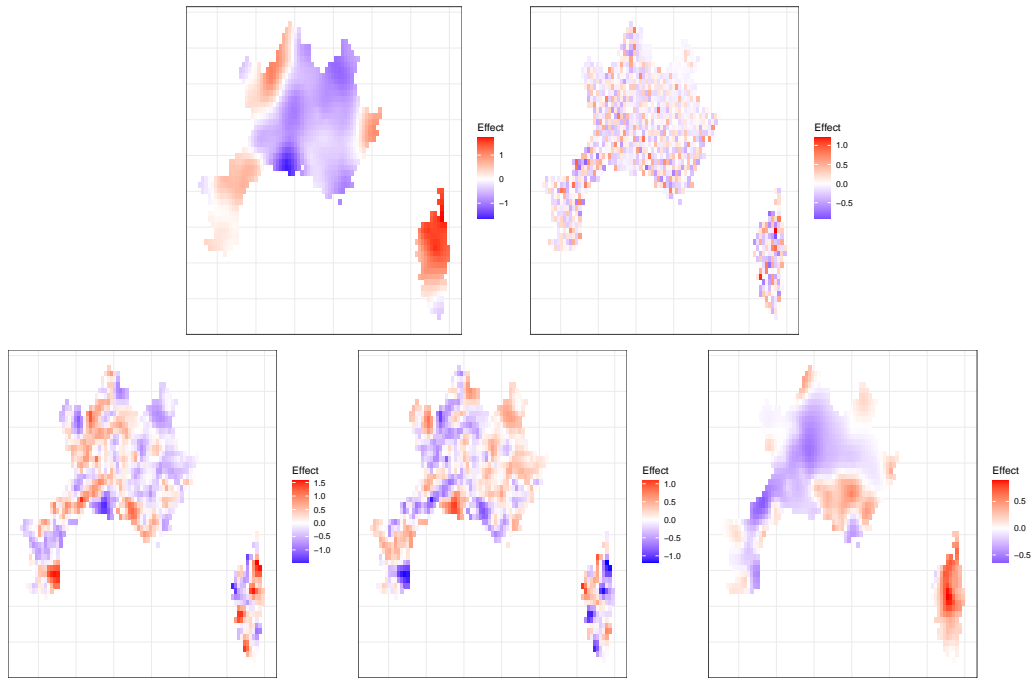


Figure 4.18 – Posterior means of (top left to bottom right): $\beta^{\text{COX-BETA}} g^{\text{COX-BETA}}(\cdot)$, $g_1^{\text{COX}}(\cdot)$, $\beta^{\text{COX-BIN}} g^{\text{COX-BIN}}(\cdot)$, $g^{\text{COX-BIN}}(\cdot)$ and $\beta^{\text{BIN-GPD}} g^{\text{BIN-GPD}}(\cdot)$ effects in the model.

4.10 A brief overview of INLA

The integrated nested Laplace approximation is a Bayesian technique for fitting hierarchical model with Gaussian random effects. It uses deterministic approximations for accurate posterior inference on model parameters, random effects and predictions conditional on data.

In a Bayesian hierarchical modelling framework with latent Gaussian components, we typically have

$$\begin{aligned} \mathbf{y} | \mathbf{x}, \boldsymbol{\theta} &\sim \prod_i \pi(y_i | \eta_i(\mathbf{x}), \boldsymbol{\theta}), \\ \mathbf{x} | \boldsymbol{\theta} &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta})^{-1}), \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}), \end{aligned}$$

where $\pi(y_i | \eta_i(\mathbf{x}), \boldsymbol{\theta})$ is the univariate density of the data point y_i , \mathbf{x} is the latent Gaussian components that is of dimension n , and $\boldsymbol{\theta}$ is a vector of hyperparameters, typically of dimension below 20 and typically 2–5. The matrix \mathbf{Q} is known as the precision matrix. An observation matrix \mathbf{A} links the latent Gaussian components \mathbf{x} to observations by

$$\boldsymbol{\eta}(\mathbf{x}) = \{\eta_1(\mathbf{x}), \dots, \eta_n(\mathbf{x})\} = \mathbf{A}\mathbf{x}.$$

For computational tractability when n is high (10^3 – 10^5), we require that there are only few non-zero entries in each row of \mathbf{A} and the data vector \mathbf{y} , and that the distribution of the latent field $\mathbf{x} | \boldsymbol{\theta}$ is a Gaussian Markov random field (GMRF), so \mathbf{Q} has mostly zero entries.

A key purpose of Bayesian inference is to find the posterior marginal densities $\pi(\theta_j | \mathbf{y})$ and $\pi(x_k | \mathbf{y})$, for chosen components j of $\boldsymbol{\theta}$ and k of \mathbf{x} , and to calculate summaries from these distributions, e.g., posterior means or 95% credible intervals.

4.10.1 Approximating the posterior marginals for the hyperparameters

Using Bayes' rule, the joint posterior density for the hyperparameters is

$$\begin{aligned} \pi(\boldsymbol{\theta} | \mathbf{y}) &= \int \pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y}) \, d\mathbf{x} = \frac{\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y})}{\pi(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})} \\ &= \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y}) / \pi(\mathbf{y})}{\pi(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})} \\ &\propto \frac{\pi(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}) \pi(\mathbf{x} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta})}{\pi(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})}. \end{aligned} \tag{4.5}$$

In the first step, INLA applies a Laplace approximation $\tilde{\pi}_L(\boldsymbol{\theta} | \mathbf{y})$ to (4.5). This is equivalent to approximating the conditional posterior in (4.5), by a multivariate Gaussian density function

$\tilde{\pi}_G$ centered at the mode $\mathbf{x}^* = \mathbf{x}^*(\boldsymbol{\theta})$ of $\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y})$, yielding

$$\tilde{\pi}_L(\boldsymbol{\theta} | \mathbf{y}) \propto \frac{\pi(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}) \pi(\mathbf{x} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta})}{\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})} \bigg|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})}. \quad (4.6)$$

The Gaussian approximation for the denominator in (4.6) is accurate due to the GMRF assumption imposed on \mathbf{x} , as

$$\pi(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x} + \sum_i \log \pi(y_i | x_i, \boldsymbol{\theta}) \right\},$$

while

$$\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y}) = (2\pi)^{-n/2} |\mathbf{P}(\boldsymbol{\theta})|^{1/2} \exp \left[-\frac{1}{2} \{\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})\}^T \mathbf{P}(\boldsymbol{\theta}) \{\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})\} \right], \quad (4.7)$$

where $\mathbf{P}(\boldsymbol{\theta}) = \mathbf{Q}(\boldsymbol{\theta}) + \text{diag}\{\mathbf{c}(\boldsymbol{\theta})\}$ and $\boldsymbol{\mu}(\boldsymbol{\theta})$ is the location of the mode. The vector $\mathbf{c}(\boldsymbol{\theta})$ contains the negative second derivatives of the log-likelihood at the mode with respect to x_i . If y_i is Gaussian conditional on x_i and $\boldsymbol{\theta}$, then $\log \pi(y_i | x_i, \boldsymbol{\theta})$ is quadratic and $\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})$ is exact. Otherwise, the approximation error is in the omitted third and higher-order terms.

Equation (4.7) is the key that enables fast computation. The Gaussian dependence structure is unchanged because (4.7) is a GMRF with respect to the same graph as $\pi(\mathbf{x} | \boldsymbol{\theta})$, i.e., \mathbf{P} and \mathbf{Q} have the same zero non-diagonal entries, as conditioning on the observations \mathbf{y} only shifts the mean and the diagonal of the precision matrix; the cross terms $x_i x_j \mathbf{Q}(\boldsymbol{\theta})$ remain untouched due to the conditional independence assumption.

As the dimension of $\boldsymbol{\theta}$ is small, a numerical integral approximation can be used in the second step to obtain the posterior marginal; the integral

$$\tilde{\pi}(\theta_j | \mathbf{y}) = \int \tilde{\pi}_L(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-j},$$

is approximated using either a grid approximation, the central composite design (Box and Wilson, 1951), or by using the mode as the sole integration point.

4.10.2 Approximating the posterior marginals for the latent field

The posterior marginal densities of the latent field are

$$\pi(x_j | \mathbf{y}) = \int \pi(x_j, \boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} = \int \underbrace{\int \pi(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y}) d\mathbf{x}_{-j}}_{\pi(x_j | \boldsymbol{\theta}, \mathbf{y})} \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}, \quad j = 1, \dots, n.$$

The term $\pi(\boldsymbol{\theta} | \mathbf{y})$ is approximated by (4.6); the term ‘nested’ in the acronym INLA stems from this step. The first term $\pi(x_j | \boldsymbol{\theta}, \mathbf{y})$ can be replaced by

1. another Laplace approximation

$$\tilde{\pi}_L(x_j | \boldsymbol{\theta}, \mathbf{y}) \propto \frac{\pi(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x} | \boldsymbol{\theta})}{\tilde{\pi}_G(\mathbf{x}_{-j} | x_j, \boldsymbol{\theta}, \mathbf{y})} \bigg|_{\mathbf{x}_{-j} = \mathbf{x}_{-j}^*(x_j, \boldsymbol{\theta})}, \quad j = 1, \dots, n,$$

though this involves n Laplace approximations for every configuration of $\boldsymbol{\theta}$ needed later for the numerical integration step, which is computationally very expensive;

2. the multivariate Gaussian joint conditional density $\tilde{\pi}_G(\mathbf{x} | \boldsymbol{\theta}, \mathbf{y})$ computed from (4.7). This is fast as it uses the univariate marginal distribution whose mean value can be read from $\mathbf{x}^*(\boldsymbol{\theta})$ and whose variance is easily calculated from a partial inversion of the precision $\mathbf{P}(\boldsymbol{\theta})$. However, as discussed in §4.10, this may be inaccurate because conditioning on the data can introduce skewness and heavy-tailed behaviour into the marginals of the joint conditional distributions; or
3. computing a Taylor expansion around the mode of the Laplace approximation, which provides linear and cubic correction terms to the standardized Gaussian (more details in Rue et al., 2009).

Once an approximation has been chosen and $\tilde{\pi}(x_j | \boldsymbol{\theta}, \mathbf{y})$ computed, the integrals

$$\tilde{\pi}(x_j | \mathbf{y}) = \int \tilde{\pi}(x_j | \boldsymbol{\theta}, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}, \quad j = 1, \dots, n.$$

are computed with a chosen numerical integration method, e.g., grid approximation, central composite design or sole integration point, similar to the last step of §4.10.1.

4.11 A brief overview of the SPDE approach

The main computational limitation of using high dimensional Gaussian fields is the $\mathcal{O}(N^3)$ cost of factorising dense $N \times N$ covariance matrices. Rue et al. (2009) solve this ‘big N ’ problem by using a Gaussian Markov random field (GMRF) representation. A GMRF is a Gaussian field $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1})$ such that $\mathbf{Q}_{i,j} \neq 0 \iff j \in \delta_i \cup \{i\}$ (Rue, 2005), for some set of neighbours δ_i for each component i of \mathbf{x} . Using Markov properties, GMRFs are discretely indexed representations of Gaussian fields that decrease the number of non-zero entries in the precision matrix; this typically reduces the computation cost to $\mathcal{O}(N^{3/2})$ for a two-dimensional field.

The Matérn covariance function for two points \mathbf{s} and \mathbf{s}' in \mathbb{R}^2 is given as follows:

$$\text{Cov}\{\mathbf{x}(\mathbf{s}), \mathbf{x}(\mathbf{s}')\} = \sigma^2 2^{1-\nu} (\kappa \|\mathbf{s} - \mathbf{s}'\|)^\nu K_\nu(\kappa \|\mathbf{s} - \mathbf{s}'\|) / \Gamma(\nu), \quad \sigma, \nu > 0, \quad (4.8)$$

with Euclidean distance $\|\cdot\|$, gamma function Γ , modified Bessel function of the second kind K_ν , scale κ , and standard deviation and smoothness parameters σ and ν .

Lindgren et al. (2011) showed that GMRF representations of Gaussian fields with Matérn covariances can be constructed through the solution to a linear fractional stochastic partial differential equation (SPDE)

$$(\kappa^2 - \Delta)^{\alpha/2} \mathbf{x}(\mathbf{s}) = W(\mathbf{s}), \quad \mathbf{s} = (s_1, s_2) \in \mathcal{D}, \quad (4.9)$$

where W is a spatial Gaussian white noise process, $\mathcal{D} \subset \mathbb{R}^2$ is the spatial domain, α controls the smoothness, κ is a scale parameter from the Matérn covariance function, and the Laplacian operator Δ is defined as

$$\Delta = \frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}.$$

Equations (4.8) and (4.9) are linked through the equality $\alpha = \nu + d/2$, where d is the dimension of the process (equal to 2 here), and the marginal variance

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2) (4\pi)^{d/2} \kappa^{2\nu}}.$$

Lindgren et al. (2011) propose to triangulate the spatial domain for irregular point data to solve (4.9). First, consider a set of test functions $\boldsymbol{\xi} = (\xi_1, \dots, \xi_m)$, where m is the number of vertices in the triangulation. The stochastic weak solution to the SPDE for any field $\mathbf{x}(\mathbf{s})$ is found by requiring that

$$\int_{\mathcal{D}} \xi_k(\mathbf{s}) (\kappa^2 - \Delta)^{\alpha-2} \mathbf{x}(\mathbf{s}) d\mathbf{s} \stackrel{d}{=} \int_{\mathcal{D}} \xi_k(\mathbf{s}) dW(\mathbf{s}), \quad (4.10)$$

where the equality is in distribution.

Chapter 4. Spatiotemporal wildfire modeling through point processes with extreme marks

We can construct a finite element representation of the solution to the SPDE as

$$\mathbf{x}(\mathbf{s}) = \sum_{k=1}^m \psi_k(\mathbf{s}) w_k, \quad (4.11)$$

where ψ_1, \dots, ψ_m are basis functions and $\mathbf{w} = (w_1, \dots, w_m)$ are Gaussian weights. If one uses piecewise linear basis functions such that $\psi_k = 1$ at vertex k and 0 at the other vertices, we get a Markov structure where values of \mathbf{x} at vertices are determined by the weights \mathbf{w} , and \mathbf{x} evaluated at any other point is found by linear interpolation from the three surrounding vertices; the joint distribution of \mathbf{w} thus determines the distribution of the continuous solution.

Given appropriate test functions ξ_1, \dots, ξ_k , the finite element solution (4.11) can be obtained by finding the distribution of \mathbf{w} that satisfies (4.10). Lindgren et al. (2011) choose $\xi_k = (\kappa^2 - \Delta)^{1/2} \psi_k$ for $\alpha = 1$ and $\xi_k = \psi_k$ for $\alpha = 2$. Define the $m \times m$ matrices \mathbf{C} , \mathbf{G} and \mathbf{K}_κ as

$$\mathbf{C}_{i,j} = \langle \psi_i, \psi_j \rangle, \quad \mathbf{G}_{i,j} = \langle \Delta \psi_i, \Delta \psi_j \rangle, \quad (\mathbf{K}_\kappa)_{i,j} = \kappa^2 \mathbf{C}_{i,j} + \mathbf{G}_{i,j},$$

where $\langle f, g \rangle = \int_{\mathcal{D}} f(\mathbf{s}) g(\mathbf{s}) d\mathbf{s}$.

Let $\mathbf{Q}_{\alpha,\kappa}$ (which depends on κ) be the precision matrix for the Gaussian weights \mathbf{w} for $\alpha = 1, 2$. The finite-dimensional representations of the solutions to (4.9) have precision matrices (full derivations in Lindgren et al., 2011)

$$\mathbf{Q}_{1,\kappa} = \mathbf{K}_\kappa, \quad \mathbf{Q}_{2,\kappa} = \mathbf{K}_\kappa \mathbf{C}^{-1} \mathbf{K}_\kappa,$$

and the authors further suggest replacing \mathbf{C} with the diagonal matrix $\tilde{\mathbf{C}} = \langle \psi_i, 1 \rangle$. Finally, we have a finite element solution to the SPDE that provides a mapping from the parameters of the Matérn Gaussian field to the elements of the GMRF's precision matrix.

5 Conclusion

The latest IPCC report considers compound extremes (Seneviratne et al., 2021), which can be characterized into four main categories, though with soft boundaries (Zscheischler et al., 2020). These events include amplified impacts that occur due either to a hazard on a precondition (*preconditioned*), to multiple concurrent drivers (*multivariate*), to a sequence of hazards (*temporally compounding*), or to spatially concurrent hazards (*spatially compounding*). This thesis discussed environmental extremes that fit into these categories.

In Chapter 2, we tackled spatially compounding extremes via the analysis of the spatial extremal dependencies of environments related to severe thunderstorms across the contiguous US using max-stable fields. Extreme wildfires are preconditioned compound events, where the initial fuel moisture heavily dictates the vegetation susceptibility to ignition. The Fire Weather Index (FWI) introduced in Chapter 4 incorporates the Fine Fuel Moisture Code, Duff Moisture Code and Drought Code, which are proxies of the fuel moisture at different soil depths and over different time scales (van Wagner, 1977). Large fire spreads may also be temporally and spatially compounding, and are partly governed by land-use, fuel composition and policy changes over the region; our model in Chapter 4 captured some of these features of wildfires by having spatially or temporally dependent random effects that were shared between different aspects of wildfire risk.

Another unifying theme of this thesis is the identification and incorporation of trends to better model and predict environmental extremes. The effect of the El Niño-Southern Oscillation on the magnitude and spatial extent of severe thunderstorm drivers was the main focus of Chapters 1 and 2, while Chapter 4 focused on the effect of FWI, forest area and spatial/temporal proxies of unmeasured effects on the spatial extent, frequency, intensity and timing of wildfire occurrences and sizes. Although the emphasis of Chapter 3 focused on good prediction, a by-product was the identification of the most important covariates for predicting wildfires.

Chapter 1 used false discovery control for multiple significance testing, based on the argument that our data resemble those considered in Ventura et al. (2004). A more rigorous argument should show that the test statistics are asymptotically jointly Gaussian and that Theorem 2.1 of Benjamini and Yekutieli (2001) can be applied, and this could be in the scope of future work.

Conclusion

Severe thunderstorms can also be regarded as multivariate compound extremes because elevated levels of both SRH and CAPE are necessary for their formation. The focus in Chapter 2 was on modelling these variables and a product variable separately. Future work could model the extremes of SRH and CAPE simultaneously with bivariate fields that account for their inter-variable dependence, while modelling their spatial dependencies (e.g., Genton et al., 2015); one should explore how to keep computational costs manageable when fitting such models with many data locations.

To further address the preconditioned nature of extreme wildfires, one could engineer new covariates based on existing meteorological ones with a time lag (e.g., of a month and up to a year) to allow covariates like past precipitation or temperature to provide information about the preconditions necessary for large fires. A machine learning technique such as the gradient boosting approach in Chapter 3 would select which covariates are most relevant.

Alternatives to the SPDE approach used in Chapters 3 and 4 for sparse modelling of high-dimensional Gaussian include the nearest-neighbour Gaussian processes (Datta et al., 2016); their predictive and computational performances have been compared in a case study competition (Heaton et al., 2019).

The random effects in Chapter 4 are spatially or temporally, but not spatio-temporally, dependent. Future work could address unmeasured trends over time with these random effects, linked perhaps to the changing wildland-to-urban interface in the region. While the focus in this chapter was on generative and predictive modeling, the adaptation of descriptive tools from stochastic geometry (K-functions, mark correlation functions, see Chiu et al., 2013) would further improve the analysis of point processes with extreme marks.

This research poses more questions than answers for the growing field of environmental extremes, where one should always exploit climate and weather-specific expertise to guide model choice. The model frameworks built in this thesis could be applied not only to wildfires and severe thunderstorms, but to other environmental extremes, such as landslides, floods and heatwaves, and using these frameworks could give insights into the mechanisms driving these events, thus helping to provide more accurate risk estimation and mitigation solutions.

Bibliography

- Agee, E., Larson, J., Childs, S., and Marmo, A. (2016). Spatial redistribution of U.S. tornado activity between 1954 and 2013. Journal of Applied Meteorology and Climatology, 55(8):1681–1697.
- Allen, J. T. and Tippett, M. K. (2015). The characteristics of United States hail reports: 1955–2014. Electronic Journal of Severe Storms Meteorology, 10(3):1–31.
- Allen, J. T., Tippett, M. K., and Sobel, A. H. (2015). Influence of the El Niño/Southern Oscillation on tornado and hail frequency in the United States. Nature Geoscience, 8:278–283.
- Bacro, J.-N., Bel, L., and Lantuéjoul, C. (2010). Testing the independence of maxima: from bivariate vectors to spatial extreme fields. Extremes, 13(2):155–175.
- Baddeley, A., Berman, M., Fisher, N. I., Hardegen, A., Milne, R. K., Schuhmacher, D., Shah, R., and Turner, R. (2010). Spatial logistic regression and change-of-support in Poisson point processes. Electronic Journal of Statistics, 4:1151–1201.
- Baddeley, A., Coeurjolly, J.-F., Rubak, E., and Waagepetersen, R. (2014). Logistic regression for spatial Gibbs point processes. Biometrika, 101(2):377–392.
- Baddeley, A., Rubak, E., and Turner, R. (2015). Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC Press, New York.
- Baddeley, A. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. Australian & New Zealand Journal of Statistics, 42(3):283–322.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1):289–300.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics, 29(4):1165–1188.
- Berman, M. and Turner, T. R. (1992). Approximating point process likelihoods with GLIM. Journal of the Royal Statistical Society: Series C (Applied Statistics), 41(1):31–38.
- Blanchet, J. and Davison, A. C. (2011). Spatial modeling of extreme snow depth. The Annals of Applied Statistics, 5(3):1699–1725.

Bibliography

- Bolin, D. and Lindgren, F. (2015). Excursion and contour uncertainty regions for latent Gaussian models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77(1):85–106.
- Bolin, D. and Wallin, J. (2020). Scale dependence: Why the average CRPS often is inappropriate for ranking probabilistic forecasts. arXiv preprint:1912.05642.
- Box, G. E. P. and Wilson, K. B. (1951). On the experimental attainment of optimum conditions. Journal of the Royal Statistical Society: Series B (Methodological), 13(1):1–45.
- Brehmer, J. R. and Strokov, K. (2019). Why scoring functions cannot assess tail properties. Electronic Journal of Statistics, 13(2):4015–4034.
- Breiman, L. (2001). Random forests. Machine Learning, 45(1):5–32.
- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification and Regression Trees. Wadsworth and Brooks, Monterey.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78:1–3.
- Brillinger, D. R., Preisler, H. K., and Benoit, J. W. (2006). Probabilistic risk assessment for wildfires. Environmetrics, 17(6):623–633.
- Brooks, H. E. (2013). Severe thunderstorms and climate change. Atmospheric Research, 123:129–138.
- Brooks, H. E., Carbin, G. W., and Marsh, P. T. (2014). Increased variability of tornado occurrence in the United States. Science, 346(6207):349–352.
- Brooks, H. E., Lee, J. W., and Craven, J. P. (2003). The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. Atmospheric Research, 67–68:73–94.
- Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. Journal of Applied Probability, 14(4):732–739.
- Buhl, S. and Klüppelberg, C. (2016). Anisotropic Brown–Resnick space-time processes: estimation and model assessment. Extremes, 19(4):627–660.
- Bühlmann, P. and Hothorn, T. (2007). Boosting algorithms: regularization, prediction and model fitting. Statistical Science, 22(4):477–505.
- Castruccio, S., Huser, R., and Genton, M. G. (2016). High-order composite likelihood inference for max-stable distributions and processes. Journal of Computational and Graphical Statistics, 25(4):1212–1229.
- Cavanaugh, J. E. and Shumway, R. H. (1997). A bootstrap variant of AIC for state-space model selection. Statistica Sinica, 7(2):473–496.

- Chen, T. and Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, pages 785–794, New York, NY, USA. ACM.
- Chiu, S. N., Stoyan, D., Kendall, W. S., and Mecke, J. (2013). Stochastic Geometry and its Applications. Wiley, Hoboken, 3rd edition.
- Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag, London.
- Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In Bertail, P., Soulier, P., and Doukhan, P., editors, Dependence in Probability and Statistics, pages 373–390. Springer, New York.
- Corfidi, S. (2017). Forecasting severe convective storms. In Oxford Research Encyclopedia of Climate Science. Oxford University Press.
- Cox, D. R. (1958). The regression analysis of binary sequences (with discussion). Journal of the Royal Statistical Society: Series B (Methodological), 20(2):215–232.
- Cui, W. and Perera, A. H. (2008). What do we know about forest fire size distribution, and why is this knowledge useful for forest management? International Journal of Wildland Fire, 17(2):234–244.
- Cumming, S. (2001). A parametric model of the fire-size distribution. Canadian Journal of Forest Research, 31(8):1297–1303.
- Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. Journal of the American Statistical Association, 111(514):800–812.
- Davison, A. C. and Hinkley, D. V. (1997). Bootstrap Methods and their Application. Cambridge University Press, Cambridge.
- Davison, A. C. and Huser, R. (2015). Statistics of extremes. Annual Review of Statistics and its Application, 2:203–235.
- Davison, A. C., Huser, R., and Thibaud, E. (2013). Geostatistics of dependent and asymptotically independent extremes. Mathematical Geosciences, 45(5):511–529.
- Davison, A. C., Huser, R., and Thibaud, E. (2018). Spatial extremes. In Gelfand, A. E., Fuentes, M., Hoeting, J. A., and Smith, R. L., editors, Handbook of Environmental and Ecological Statistics. CRC Press, Boca Raton.
- Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. Statistical Science, 27(2):161–186.

Bibliography

- Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). Journal of the Royal Statistical Society: Series B (Methodological), 52(3):393–442.
- De Angelis, A., Ricotta, C., Conedera, M., and Pezzatti, G. B. (2015). Modelling the meteorological forest fire niche in heterogeneous pyrologic conditions. PLOS ONE, 10(2):1–17.
- de Carvalho, M. and Ramos, A. (2012). Bivariate extreme statistics, II. Revstat-Statistical Journal, 10:81–104.
- de Haan, L. (1984). A spectral representation for max-stable processes. The Annals of Probability, 12(4):1194–1204.
- de Haan, L. and Ferreira, A. (2006). Extreme Value Theory: An Introduction. Springer-Verlag, New York.
- De Zea Bermudez, P., Mendes, J., Pereira, J. M., Turkman, K. F., and Vasconcelos, M. J. (2009). Spatial and temporal extremes of wildfire sizes in Portugal (1984–2004). International Journal of Wildland Fire, 18(8):983–991.
- Del Genio, A. D., Yao, M.-S., and Jonas, J. (2007). Will moist convection be stronger in a warmer climate? Geophysical Research Letters, 34(16):L16703.
- DelSole, T. and Yang, X. (2011). Field significance of regression patterns. Journal of Climate, 24(19):5094–5107.
- Diffenbaugh, N. S., Scherer, M., and Trapp, R. J. (2013). Robust increases in severe thunderstorm environments in response to greenhouse forcing. Proceedings of the National Academy of Sciences, 110(41):16361–16366.
- Diggle, P. J., Menezes, R., and Su, T.-I. (2010). Geostatistical inference under preferential sampling (with discussion). Journal of the Royal Statistical Society: Series C (Applied Statistics), 59(2):191–232.
- Dombry, C., Engelke, S., and Oesting, M. (2016). Exact simulation of max-stable processes. Biometrika, 103(2):303–317.
- Dombry, C., Ribatet, M., and Stoev, S. (2018). Probabilities of concurrent extremes. Journal of the American Statistical Association, 113(524):1565–1582.
- Doswell III, C. A. (2015). Mesoscale meteorology | severe storms. In Encyclopedia of Atmospheric Sciences, pages 361–368.
- Doswell III, C. A., Brooks, H. E., and Maddox, R. A. (1996). Flash flood forecasting: an ingredients-based methodology. Weather and Forecasting, 11(4):560–581.
- Dutta, R., Aryal, J., Das, A., and Kirkpatrick, J. B. (2013). Deep cognitive imaging systems enable estimation of continental-scale fire incidence from climate data. Scientific Reports, 3(1):3188.

- Edwards, R., Allen, J. T., and Carbin, G. W. (2018). Reliability and climatological impacts of convective wind estimations. Journal of Applied Meteorology and Climatology, 57(8):1825–1845.
- Einmahl, J. H. J. and Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. The Annals of Statistics, 37(5B):2953–2989.
- Elsner, J. B., Elsner, S. C., and Jagger, T. H. (2015). The increasing efficiency of tornado days in the United States. Climate Dynamics, 45(3–4):651–659.
- Embrechts, P., Mikosch, T., and Klüppelberg, C. (1997). Modelling Extremal Events for Insurance and Finance. Springer-Verlag, Berlin, Heidelberg.
- Evin, G., Curt, T., and Eckert, N. (2018). Has fire policy decreased the return period of the largest wildfire events in France? A Bayesian assessment based on extreme value theory. Natural Hazards and Earth System Sciences, 18(10):2641–2651.
- Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8):861–874.
- Friedman, J., Hastie, T., and Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). The Annals of Statistics, 28(2):337–407.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of Statistics, 29:1189–1232.
- Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2018). Constructing priors that penalize the complexity of Gaussian random fields. Journal of the American Statistical Association, 114(525):445–452.
- Gabda, D., Towe, R., Wadsworth, J., and Tawn, J. (2012). Discussion of “Statistical Modelling of Spatial Extremes” by A.C. Davison, S.A. Padoan and M. Ribatet. Statistical Science, 27(2):189–192.
- Gensini, V. A. and Brooks, H. E. (2018). Spatial trends in United States tornado frequency. npj Climate and Atmospheric Science, 1:number 38.
- Genton, M. G., Butry, D. T., Gumpertz, M. L., and Prestemon, J. P. (2006). Spatio-temporal analysis of wildfire ignitions in the St Johns River water management district, Florida. International Journal of Wildland Fire, 15(1):87–97.
- Genton, M. G., Padoan, S. A., and Sang, H. (2015). Multivariate max-stable spatial processes. Biometrika, 102(1):215–230.
- Gilleland, E., Brown, B. G., and Ammann, C. M. (2013). Spatial extreme value analysis to project extremes of large-scale indicators for severe weather. Environmetrics, 24(6):418–432.

Bibliography

- Gilleland, E., Pocerich, M., Brooks, H. E., Brown, B. G., and Marsh, P. (2008). Large-scale indicators for severe weather. <https://ral.ucar.edu/~ericg/GillelandEtAl2008.pdf>.
- Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–422.
- Greenwell, B., Boehmke, B., Cunningham, J., and Developers, G. (2020). *gbm: Generalized Boosted Regression Models*. R package version 2.1.8.
- Halpert, M. and Ropelewsk, C. (1992). Surface temperature patterns associated with the Southern Oscillation. *Journal of Climate*, 5(6):577–593.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York, 2nd edition.
- Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D. W., Sun, F., and Zammit-Mangion, A. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, 24(3):398–425.
- Heaton, M. J., Katzfuss, M., Ramachandar, S., Pedings, K., Gilleland, E., Mannshardt-Shamseldin, E., and Smith, R. L. (2011). Spatio-temporal models for large-scale indicators of extreme weather. *Environmetrics*, 22(3):294–303.
- Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546.
- Hitz, A., Davis, R., and Samorodnitsky, G. (2017). Discrete extremes. [arXiv preprint:1707.05033](https://arxiv.org/abs/1707.05033).
- Hoogewind, K. A., Baldwin, M. E., and Trapp, R. J. (2017). The impact of climate change on hazardous convective weather in the United States: insight from high-resolution dynamical downscaling. *Journal of Climate*, 30(24):10081–10100.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.
- Huser, R. and Davison, A. C. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika*, 100(2):511–518.
- Huser, R. and Genton, M. G. (2016). Non-stationary dependence structures for spatial extremes. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(3):470–491.
- Huser, R., Opitz, T., and Thibaud, E. (2021). Max-infinitely divisible models and inference for spatial extremes. *Scandinavian Journal of Statistics*, 48(1):321–348.
- Huser, R. and Wadsworth, J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 114(525):434–444.

- Illian, J. B., Sørbye, S. H., and Rue, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). The Annals of Applied Statistics, 6(4):1499–1530.
- Jain, P., Coogan, S. C., Subramanian, S. G., Crowley, M., Taylor, S., and Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. Environmental Reviews, 28(4):478–505.
- Johns, R. H., Davies, J. M., and Leftwich, P. W. (1993). Some wind and instability parameters associated with strong and violent tornadoes: 2. variations in the combinations of wind and instability parameters. In The Tornado: Its Structure, Dynamics, Prediction, and Hazards, pages 583–590. Wiley Online Library.
- Jones, M. W., Smith, A., Betts, R., Canadell, J. G., Prentice, I. C., and Le Quéré, C. (2020). ScienceBrief Review: Climate change increases the risk of wildfires. In Le Quéré, C., Liss, P., and Forster, P., editors, Critical Issues in Climate Change Science.
- Jong, B.-T., Ting, M., and Seager, R. (2016). El Niño's impact on California precipitation: Seasonality, regionality, and El Niño intensity. Environmental Research Letters, 11(5):054021.
- 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. Ecological Applications, 29(6):e01898.
- Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. The Annals of Probability, 37(5):2042–2065.
- Koch, E., Koh, J., Davison, A. C., Lepore, C., and Tippet, M. K. (2021). Trends in the extremes of environments associated with severe U.S. thunderstorms. Journal of Climate, 34(4):1259–1272.
- Koh, J. (2021). Gradient boosting with extreme-value theory for wildfire prediction. arXiv preprint:2110.09497.
- Koh, J., Pimont, F., Dupuy, J.-L., and Opitz, T. (2021). Spatiotemporal wildfire modelling through point processes with moderate and extreme marks. arXiv preprint:2105.08004.
- Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA. Chapman and Hall/CRC, Boca Raton.
- Lepore, C., Tippet, M. K., and Allen, J. T. (2017). ENSO-based probabilistic forecasts of March–May US tornado and hail activity. Geophysical Research Letters, 44(17):9093–9101.
- Lepore, C., Veneziano, D., and Molini, A. (2015). Temperature and CAPE dependence of rainfall extremes in the eastern United States. Geophysical Research Letters, 42(1):74–83.

Bibliography

- Liang, H., Zhang, M., and Wang, H. (2019). A neural network model for wildfire scale prediction using meteorological factors. IEEE Access, 7:176746–176755.
- Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with R-INLA. Journal of Statistical Software, 63(19).
- Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73(4):423–498.
- Livezey, R. E. and Chen, W. Y. (1983). Statistical field significance and its determination by Monte Carlo techniques. Monthly Weather Review, 111(1):46–59.
- Lombardo, L., Opitz, T., Ardizzone, F., Guzzetti, F., and Huser, R. (2020). Space-time landslide predictive modelling. Earth Science Reviews, 209:103318.
- Lyon, B. (2004). The strength of El Niño and the spatial extent of tropical drought. Geophysical Research Letters, 31(21).
- Lyon, B. and Barnston, A. G. (2005). ENSO and the spatial extent of interannual precipitation extremes in tropical land areas. Journal of Climate, 18(23):5095–5109.
- Mannshardt, E. and Gilleland, E. (2013). Extremes of severe storm environments under a changing climate. American Journal of Climate Change, 2(3A):47–61.
- Matheson, J. E. and Winkler, R. L. (1976). Scoring rules for continuous probability distributions. Management Science, 22(10):1087–1096.
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika, 12(2):153–157.
- Mendes, J. M., de Zea Bermudez, P. C., Pereira, J., Turkman, K. F., and Vasconcelos, M. J. (2010). Spatial extremes of wildfire sizes: Bayesian hierarchical models for extremes. Environmental and Ecological Statistics, 17(1):1–28.
- Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E. H., Ek, M. B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., and Shi, W. (2006). North American regional reanalysis. Bulletin of the American Meteorological Society, 87(3):343–360.
- Mhalla, L., Chavez-Demoulin, V., and Naveau, P. (2017). Non-linear models for extremal dependence. Journal of Multivariate Analysis, 159:49–66.
- Mitsopoulos, I. and Mallinis, G. (2017). A data-driven approach to assess large fire size generation in Greece. Natural Hazards, 88(3):1591–1607.

- National Interagency Fire Center (2021). Total wildfires and acres. Data retrieved from <https://www.predictiveservices.nifc.gov/intelligence/intelligence.htm>.
- Northrop, P. J. and Coleman, C. L. (2014). Improved threshold diagnostic plots for extreme value analyses. *Extremes*, 17(2):289–303.
- Opitz, T. (2013). Extremal t processes: elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, 122:409–413.
- Opitz, T. (2017). Latent Gaussian modeling and INLA: a review with focus on space-time applications. *Journal de la Société Française de Statistique*, 158(3):62–85.
- Opitz, T. (2021). EVA 2021 data challenge. <https://www.maths.ed.ac.uk/school-of-mathematics/eva-2021/competitions/data-challenge>. Accessed: 2021-08-17.
- Opitz, T. (2022). Editorial: EVA 2021 data competition on spatio-temporal prediction of wildfire activity in the United States. *Extremes*, to appear.
- Opitz, T., Bakka, H., Huser, R., and Lombardo, L. (2020a). High-resolution Bayesian mapping of landslide hazard with unobserved trigger event. *arXiv preprint:2006.07902*.
- Opitz, T., Bonneau, F., and Gabriel, E. (2020b). Point-process based modeling of space-time structures of forest fire occurrences in Mediterranean France. *Spatial Statistics*, 40:100429.
- Opitz, T., Huser, R., Bakka, H., and Rue, H. (2018). INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, 21(3):441–462.
- Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC, New York.
- Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.
- Pati, D., Reich, B. J., and Dunson, D. B. (2011). Bayesian geostatistical modelling with informative sampling locations. *Biometrika*, 98(1):35–48.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Édouard Duchesnay (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830.
- Peng, R. D., Schoenberg, F. P., and Woods, J. A. (2005). A space-time conditional intensity model for evaluating a wildfire hazard index. *Journal of the American Statistical Association*, 100(469):26–35.
- Pereira, J. M. C. and Turkman, K. F. (2019). Statistical models of vegetation fires: spatial and temporal patterns. In Gelfand, A. E., Fuentes, M., Hoeting, J. A., and Smith, R. L., editors, *Handbook of Environmental and Ecological Statistics*, pages 401–420. Chapman and Hall/CRC, Boca Raton.

Bibliography

- Pickands, J. (1981). Multivariate extreme value distributions. Bulletin of the International Statistical Institute: Proceedings of the 43rd Session of the International Statistical Institute, Vol. 2 (Buenos Aires), pages 859–878, 894–902.
- Pimont, F., Fargeon, H., Opitz, T., Ruffault, J., Barbero, R., Martin-StPaul, N., Rigolot, E. I., Rivière, M., and Dupuy, J.-L. (2021). Prediction of regional wildfire activity in the probabilistic Bayesian framework of Firelihood. Ecological Applications, 31(5):e02316.
- Pohjankukka, J., Pahikkala, T., Nevalainen, P., and Heikkonen, J. (2017). Estimating the prediction performance of spatial models via spatial k-fold cross validation. International Journal of Geographical Information Science, 31(10):2001–2019.
- Preisler, H. K., Brillinger, D. R., Burgan, R. E., and Benoit, J. (2004). Probability based models for estimation of wildfire risk. International Journal of wildland fire, 13(2):133–142.
- Prieto, F., Gómez-Déniz, E., and Sarabia, J. M. (2014). Modelling road accident blackspots data with the discrete generalized Pareto distribution. Accident Analysis & Prevention, 71:38–49.
- Rasmussen, C. E. and Williams, C. K. I. (2005). Gaussian Processes for Machine Learning. The MIT Press, Cambridge.
- Rasmussen, E. N. and Blanchard, D. O. (1998). A baseline climatology of sounding-derived supercell and tornado forecast parameters. Weather and Forecasting, 13(4):1148–1164.
- Rathbun, S. L. (2013). Optimal estimation of Poisson intensity with partially observed covariates. Biometrika, 100(1):277–281.
- Rathbun, S. L., Shiffman, S., and Gwaltney, C. J. (2007). Modelling the effects of partially observed covariates on Poisson process intensity. Biometrika, 94(1):153–165.
- Ríos-Pena, L., Kneib, T., Cadarso-Suárez, C., Klein, N., and Marey-Pérez, M. (2018). Studying the occurrence and burnt area of wildfires using zero-one-inflated structured additive beta regression. Environmental Modelling & Software, 110:107–118.
- Risser, M. D. and Calder, C. A. (2015). Regression-based covariance functions for nonstationary spatial modeling. Environmetrics, 26(4):284–297.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillerá-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography, 40(8):913–929.
- Ropelewski, C. F. and Halpert, M. S. (1986). North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). Monthly Weather Review, 114(12):2352–2362.
- Rue, H. (2005). Marginal variances for Gaussian Markov random fields. Technical report. Norwegian Institute of Science and of Technology, Trondheim.

- Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71(2):319–392.
- Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with INLA: a review. Annual Review of Statistics and Its Application, 4:395–421.
- Ruffault, J., Martin-StPaul, N., Pimont, F., and Dupuy, J.-L. (2018). How well do meteorological drought indices predict live fuel moisture content (LFMC)? An assessment for wildfire research and operations in Mediterranean ecosystems. Agricultural and Forest Meteorology, 262:391–401.
- Sakr, G. E., Elhadj, I. H., Mitri, G., and Wejinya, U. C. (2010). Artificial intelligence for forest fire prediction. In 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pages 1311–1316.
- Sang, H. and Genton, M. G. (2014). Tapered composite likelihood for spatial max-stable models. Spatial Statistics, 8:86–103.
- Schlather, M. (2002). Models for stationary max-stable random fields. Extremes, 5(1):33–44.
- Schlather, M. and Tawn, J. A. (2003). A dependence measure for multivariate and spatial extreme values: properties and inference. Biometrika, 90(1):139–156.
- Schoenberg, F. P., Peng, R., and Woods, J. (2003). On the distribution of wildfire sizes. Environmetrics, 14(6):583–592.
- Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S., Wehner, M., and Zhou, B. (2021). Weather and climate extreme events in a changing climate. In Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., M., H., Leitzell, K., E., L., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., editors, Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
- Serra, L., Juan, P., Varga, D., Mateu, J., and Saez, M. (2013). Spatial pattern modelling of wildfires in Catalonia, Spain 2004–2008. Environmental Modelling & Software, 40:235–244.
- Serra, L., Saez, M., Juan, P., Varga, D., and Mateu, J. (2014). A spatio-temporal Poisson hurdle point process to model wildfires. Stochastic Environmental Research and Risk Assessment, 28(7):1671–1684.
- Shibata, R. (1997). Bootstrap estimate of Kullback-Leibler information for model selection. Statistica Sinica, 7:375–394.

Bibliography

- Shidik, G. F. and Mustofa, K. (2014). Predicting size of forest fire using hybrid model. In Linawati, Mahendra, M. S., Neuhold, E. J., Tjoa, A. M., and You, I., editors, Information and Communication Technology, pages 316–327. Springer, Berlin, Heidelberg.
- Shimura, T. (2012). Discretization of distributions in the maximum domain of attraction. Extremes, 15(3):299–317.
- Simpson, D., Rue, H., Riebler, A., Martins, T. G., Sørbye, S. H., et al. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. Statistical Science, 32(1):1–28.
- Smith, R. (1990). Max-stable processes and spatial extremes. Unpublished manuscript.
- Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, NIPS’12: Proceedings of the 25th International Conference on Neural Information Processing Systems, volume 25. Curran Associates Inc., Red Hook.
- Stark, C. P. and Hovius, N. (2001). The characterization of landslide size distributions. Geophysical Research Letters, 28(6):1091–1094.
- Stewart, S. I., Radeloff, V. C., Hammer, R. B., and Hawbaker, T. J. (2007). Defining the Wild-land–Urban Interface. Journal of Forestry, 105(4):201–207.
- Taylor, B. M. and Diggle, P. J. (2014). INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. Journal of Statistical Computation and Simulation, 84(10):2266–2284.
- Taylor, S. W., Woolford, D. G., Dean, C. B., and Martell, D. L. (2013). Wildfire prediction to inform fire management: Statistical science challenges. Statistical Science, 28(4):586–615.
- Thompson, R. L., Edwards, R., Hart, J. A., Elmore, K. L., and Markowski, P. (2003). Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. Weather and Forecasting, 18(6):1243–1261.
- Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. Journal of the American Statistical Association, 81(393):82–86.
- Tippett, M. K., Lepore, C., and Cohen, J. E. (2016). More tornadoes in the most extreme U.S. tornado outbreaks. Science, 354(6318):1419–1423.
- Tokdar, S. T. and Kass, R. E. (2010). Importance sampling: a review. Wiley Interdisciplinary Reviews: Computational Statistics, 2(1):54–60.
- Tonini, M., Pereira, M. G., Parente, J., and Orozco, C. V. (2017). Evolution of forest fires in Portugal: from spatio-temporal point events to smoothed density maps. Natural Hazards, 85(3):1489–1510.

- Trapp, R. J., Diffenbaugh, N. S., and Gluhovsky, A. (2009). Transient response of severe thunderstorm forcing to elevated greenhouse gas concentrations. Geophysical Research Letters, 36(1):L01703.
- Trapp, R. J., Tessendorf, S. A., Godfrey, E. S., and Brooks, H. E. (2005). Tornadoes from squall lines and bow echoes. Part I: climatological distribution. Weather and Forecasting, 20(1):23–34.
- Turkman, K. F., Amaral Turkman, M. A., and Pereira, J. M. (2010). Asymptotic models and inference for extremes of spatio-temporal data. Extremes, 13(4):375–397.
- Van Klooster, S. L. and Roebber, P. J. (2009). Surface-based convective potential in the contiguous United States in a business-as-usual future climate. Journal of Climate, 22(12):3317–3330.
- van Niekerk, J., Bakka, H., Rue, H., and Schenk, L. (2019). New frontiers in Bayesian modeling using the INLA package in R. arXiv preprint:1907.10426.
- van Wagner, C. (1977). Conditions for the start and spread of crown fire. Canadian Journal of Forest Research, 7(1):23–34.
- Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. Statistica Sinica, 21(1):5–42.
- Varin, C. and Vidoni, P. (2005). A note on composite likelihood inference and model selection. Biometrika, 92(3):519–528.
- Velthoen, J., Dombry, C., Cai, J.-J., and Engelke, S. (2021). Gradient boosting for extreme quantile regression. arXiv preprint:2103.00808.
- Ventura, V., Paciorek, C. J., and Risbey, J. S. (2004). Controlling the proportion of falsely rejected hypotheses when conducting multiple tests with climatological data. Journal of Climate, 17(22):4343–4356.
- Verbout, S. M., Brooks, H. E., Leslie, L. M., and Schultz, D. M. (2006). Evolution of the U.S. tornado database: 1954–2003. Weather and Forecasting, 21(1):86–93.
- Vilar, L., Woolford, D. G., Martell, D. L., and Martín, M. P. (2010). Spatio-temporal analysis of wildfire ignitions in the St Johns River water management district, Florida. International Journal of Wildland Fire, 19(3):325–337.
- Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and Widely Applicable Information Criterion in singular learning theory. Journal of Machine Learning Research, 11(116):3571–3594.
- Wood, S. (2017). Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC, Boca Raton, 2nd edition.

Bibliography

- 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. Journal of Environmental Statistics, 2(1):1–26.
- Xi, D. D., Taylor, S. W., Woolford, D. G., and Dean, C. (2019). Statistical models of key components of wildfire risk. Annual Review of Statistics and Its Application, 6:197–222.
- Xie, Y. and Peng, M. (2019). Forest fire forecasting using ensemble learning approaches. Neural Computing and Applications, 31(9):4541–4550.
- Xu, H. and Schoenberg, F. P. (2011). Point process modeling of wildfire hazard in Los Angeles County, California. The Annals of Applied Statistics, 5(2A):684–704.
- Yekutieli, D. and Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. Journal of Statistical Planning and Inference, 82(1-2):171–196.
- Zhang, T., Hoerling, M. P., Perlwitz, J., Sun, D.-Z., and Murray, D. (2011). Physics of US surface temperature response to ENSO. Journal of Climate, 24(18):4874–4887.
- Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., Maraun, D., Ramos, A. M., Ridder, N. N., Thiery, W., and Vignotto, E. (2020). A typology of compound weather and climate events. Nature Reviews Earth & Environment, 1(7):333–347.

Jonathan Koh

+41 (0) 787995396 | jonathan.koh@epfl.ch

Education

EPF Lausanne, PhD Statistics

Sep 2017–Present

- Thesis: *Spatiotemporal modelling of extreme wildfires and severe thunderstorm environments*
- Supervised by Prof. Anthony C. Davison and funded by the Swiss National Science Foundation
- Completed 12 ECTS credits at the mathematics doctoral school (grade: 6/6)

ETH Zürich & Universität Zürich, MSc Quantitative Finance

Sep 2014–May 2017

- Graduated with summa cum laude (highest honour)
- Master thesis at RiskLab ETH, supervised by Prof. Paul Embrechts and Dr. Erwan Koch (grade: 6/6)
- Completed one-year internship with UBS Switzerland

London School of Economics, BSc Mathematics & Economics

Sep 2011–Jul 2014

- Graduated with first class honours (highest honour)
- Focussed on measure theoretic probability and financial mathematics

Taylor's College Malaysia, A-level

Jul 2009–Dec 2010

- Achieved 4 A*s, with an average of 96% across all subjects
- Awarded the Tan Sri Loy Hean Hong full scholarship

Publications

- *Koch, E., *Koh, J., Davison, A. C., Lepore, C., Tippet, M. K. (2021). Trends in the extremes of environments associated with severe US thunderstorms. *Journal of Climate*, 34(4):1259–1272
- Davison, A. C., Koch, E., Koh, J. (2019). Comment: Models are approximations! *Statistical Science*, 34(4):580–480

* indicates joint first author

Working or submitted papers

- Koh, J., Opitz, T, Pimont, F., Dupuy, J., -L. (2021). Spatiotemporal wildfire modelling through point processes with moderate and extreme marks. *ArXiv: 2105.08004*. In revision for *The Annals of Applied Statistics*
- Koh, J. (2021). Gradient boosting with extreme-value theory for wildfire prediction. *ArXiv: 2110.09497*. In revision for *Extremes*
- Koh, J., Koch, E., Davison, A. C. (2021). Space-time extremes of severe US thunderstorm environments.

Presentations

- Invited talk, German Statistical Week 2021 (virtual) September 2021
- Invited talk, 12th Extreme Value Analysis (virtual) July 2021
- Contributed talk, 12th Extreme Value Analysis (virtual) June 2021
- Contributed talk, 13th Workshop on Spatial Statistics and Image Analysis (virtual) May 2021
- Invited talk, 13th CMStat (virtual) Dec 2020
- Invited seminar, IMSV Institute Seminar Bern (virtual) Dec 2020
- Invited talk, 12th CMStat, London, United Kingdom Dec 2019
- Contributed talk, 62nd ISI World Statistics Congress, Kuala Lumpur, Malaysia Aug 2019
- Invited talk, 11th Extreme Value Analysis, Zagreb, Croatia Jul 2019

- Contributed talk, 14th IMSC, Toulouse, France *Jun 2019*
- Contributed poster, Climate Extremes 2019, Riederalp, Switzerland *Mar 2019*
- Contributed poster, 11th CMStat, Pisa, Italy *Dec 2018*
- Contributed talk, BCUR 2014, Nottingham, United Kingdom *Apr 2014*

Teaching assistantships

- Wildfire Modelling:** four MSc/BSc semester projects *Spring 2021*
- Statistical Machine Learning:** MATH-412 (MSc) *Fall 2017, 2018, 2019, 2020*
- Bayesian Computation:** MATH-432 (MSc) *Spring 2019, 2020*
- Probability and Statistics:** MATH-232 (BSc) *Spring 2018*
- Duties included: answering students' questions in class, planning class sessions, preparing exercises and solution sheets, holding office hours and review sessions, supervising and grading projects, and aiding in the writing and marking of quizzes, midterms and final exams
 - Awarded a PhD teaching award from The Mathematics Section (SMA) in 2019

Competitions

- 2nd place (£250), 12th Extreme Value Analysis Data Challenge *July 2021*
- 1st place (€1,250), 8th Europe Congress of Mathematics Data Challenge *June 2021*

Industry experience

- UBS Switzerland, Intern, Quantitative Risk Methodology** *Oct 2015–Sep 2016*
- Contributed to the refinement and improvement of the bank's credit risk methods (i.e., probability of default model) and their implementation in the mortgage business based on new regulatory requirements as well as new best practice standards
- Ernst & Young Malaysia, Intern, Lead Advisory in Transaction Advisory Services** *Jul–Sep 2012*
- Performed buy-side mergers & acquisition advisory research and analysed opportunities for corporate clients to grow inorganically and realise synergies through acquisition
- HSBC Malaysia, Intern, Sales in Global Payments & Cash Management** *Apr–Jun 2011*
- Assisted the sales team with pitches, company research and sales forecasting for payment and receivable solutions while spearheading the sales team's target of winning Euromoney's "Best Cash Management Bank" award for 2011

Languages	English (Native), Malay (Conversational), Mandarin (Conversational), French (B1), German (A2)
Programming	R, Python, C, SAS, VBA, Latex, Matlab